Eberhard Zeidler

Quantum Field Theory

Quantum Electrodynamics

A Bridge Between Mathematicians and Physicists

Springer

Quantum Field Theory II: Quantum Electrodynamics

Eberhard Zeidler

Quantum Field Theory II: Quantum Electrodynamics

A Bridge between Mathematicians and Physicists

Springer

Eberhard Zeidler

Max Planck Institute
for Mathematics in the Sciences
Inselstr. 22-26
04103 Leipzig
Germany

ISBN 978-3-540-85376-3 e-ISBN 978-3-540-85377-0

DOI 10.1007/978-3-540-85377-0

Library of Congress Control Number: 2006929535

Mathematics Subject Classification (2000): 35-XX, 47-XX, 49-XX, 51-XX, 55-XX, 81-XX, 82-XX

Cover design: WMXDesign GmbH

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

TO FRIEDRICH HIRZEBRUCH
IN GRATITUDE

Preface

> And God said, Let there be light; and there was light.
> Genesis 1,3

> Light is not only the basis of our biological existence, but also an essential source of our knowledge about the physical laws of nature, ranging from the seventeenth century geometrical optics up to the twentieth century theory of general relativity and quantum electrodynamics.
> Folklore

> Don't give us numbers: give us insight!
> A contemporary natural scientist to a mathematician

The present book is the second volume of a comprehensive introduction to the mathematical and physical aspects of modern quantum field theory which comprehends the following six volumes:

Volume I: Basics in Mathematics and Physics
Volume II: Quantum Electrodynamics
Volume III: Gauge Theory
Volume IV: Quantum Mathematics
Volume V: The Physics of the Standard Model
Volume VI: Quantum Gravitation and String Theory.

It is our goal to build a bridge between mathematicians and physicists based on the challenging question about the fundamental forces in

- macrocosmos (the universe) and
- microcosmos (the world of elementary particles).

The six volumes address a broad audience of readers, including both undergraduate and graduate students, as well as experienced scientists who want to become familiar with quantum field theory, which is a fascinating topic in modern mathematics and physics.

For students of mathematics, it is shown that detailed knowledge of the physical background helps to motivate the mathematical subjects and to

discover interesting interrelationships between quite different mathematical topics. For students of physics, fairly advanced mathematics are presented, which is beyond the usual curriculum in physics. The strategies and the structure of the six volumes are thoroughly discussed in the Prologue to Volume I. In particular, we will try to help the reader to understand the basic ideas behind the technicalities. In this connection, the famous ancient story of Ariadne's thread is discussed in the Preface to Volume I. In terms of this story, we want to put the beginning of Ariadne's thread in quantum field theory into the hands of the reader.

The present volume is devoted to the physics and mathematics of light.

It contains the following material:

We try to find the right balance between the mathematical theory and its applications to interesting physical effects observed in experiments. In particular, we do not consider purely mathematical models in this volume.

> *It is our philosophy that the reader should learn quantum field theory by studying a realistic model, as given by quantum electrodynamics.*

Let us discuss the main structure of the present volume. In Chapters 1 through 4, we consider topics from classical mathematics which are closely related to modern quantum field theory. This should help the reader to understand the basic ideas behind quantum field theory to be considered in this volume and the volumes to follow. In Chapter 1 on the mathematical principles of modern natural philosophy, we discuss

- the infinitesimal strategy due to Newton and Leibniz,
- the optimality principle for processes in nature (the principle of critical action) and the calculus of variations due to Euler and Lagrange, which leads to the fundamental differential equations in classical field theory,
- the propagation of physical effects and the method of the Green's function,
- harmonic analysis and the Fourier method for computing the Green's functions,
- Laurent Schwartz's theory of generalized functions (distributions) which is related to the idea that the measurement of physical quantities by devices is based on averaging,
- global symmetry and conservation laws,
- local symmetry and the basic ideas of modern gauge field theory, and
- the Planck quantum of action and the idea of quantizing classical field theories.

Gauge field theory is behind both

- the Standard Model in elementary particle physics and
- Einstein's theory of gravitation (i.e., the theory of general relativity).

In quantum field theory, a crucial role is played by *renormalization.* In terms of physics, this is based on the following two steps:

- the regularization of divergent integrals, and
- the computation of effective physical parameters measured in experiments (e.g., the effective mass and the effective electric charge of the electron).

Renormalization is a highly technical subject. For example, the full proof on the renormalizability of the electroweak sector of the Standard Model in particle physics needs 100 pages. This can be found in:

> E. Kraus, Renormalization of the electroweak standard model to all orders, Annals of Physics **262** (1998), 155–259.

Whoever wants to understand quantum field theory has to understand the procedure of renormalization. Therefore, the different aspects of renormalization theory will be studied in all of the six volumes of this series of monographs. This ranges from

- resonance phenomena for the anharmonic oscillator (classical bifurcation theory),
- the Poincaré–Lindstedt series (including small divisors) in celestial mechanics,

- and the Kolmogorov–Arnold–Moser (KAM) theory for perturbed quasi-periodic oscillations (e.g., in celestial mechanics) based on sophisticated iterative techniques (the hard implicit function theorem)

to the following fairly advanced subjects:

- the Feynman functional integral (the Faddeev–Popov approach),
- the Wiener functional integral (the Glimm–Jaffe approach),
- the theory of higher-dimensional Abelian integrals (algebraic Feynman integrals),
- Hopf algebras and Rota–Baxter algebras in combinatorics (the modern variant of the Bogoliubov–Parasiuk–Hepp–Zimmermann (BPHZ) approach due to Kreimer),
- the Riemann–Hilbert problem and the Birkhoff decomposition (the Connes–Kreimer approach),
- Hopf superalgebras (the Brouder–Fauser–Frabetti–Oeckl (BFFO) approach),
- characterization of physical states by cohomology and algebraic renormalization (the Becchi–Rouet–Stora–Tyutin (BRST) approach),
- the Riesz–Gelfand theory of distribution-valued meromorphic functions (construction of the Green's functions),
- wave front sets and Hörmander's multiplication of distributions (the Stueckelberg–Bogoliubov–Epstein–Glaser–Scharf approach),
- the Master Ward identity as a highly non-trivial renormalization condition and the generalized Dyson–Schwinger equation (the Dütsch–Fredenhagen approach),
- q-deformed quantum field theory (the Wess–Majid–Wachter–Schmidt approach based on the q-deformed Poincaré group, quantum groups, and the q-analysis on specific classes of q-deformed quantum spaces),
- deformation of bundles and quantization (the Weyl–Flato–Sternheimer–Fedosov–Kontsevich approach),
- microlocal analysis and renormalization on curved space-times (the Radzikowski–Brunetti–Fredenhagen–Köhler approach),
- renormalized operator products on curved space-times (the Wilson–Hollands–Wald approach to quantum field theory),
- natural transformations of functors in category theory and covariant quantum field theory on curved space-time manifolds (the Brunetti–Fredenhagen–Verch approach),

as well as

- one-parameter Lie groups and the renormalization group,
- attractors of dynamical systems in the space of physical theories (the Wilson–Polchinski–Kopper–Rivasseau approach to renormalization based on the renormalization group),
- the Master Ward Identity and the Stueckelberg–Petermann renormalization group (the Dütsch–Fredenhagen approach),
- motives in number theory and algebraic geometry, the Tannakian category, and the cosmic Galois group as a universal (motivic) renormalization group (the Connes–Marcolli approach),
- noncommutative geometry and renormalization (the Grosse–Wulkenhaar approach).

The recent work of Alain Connes, Dirk Kreimer, and Matilde Marcolli shows convincingly that renormalization is rooted in highly nontrivial mathematical structures. We also want to emphasize that the theory of many-particle systems (considered in statistical physics and quantum field theory) is deeply rooted in the theory of operator algebras. This concerns

- von Neumann algebras (the von Neumann approach),
- C^*-algebras (the Gelfand–Naimark–Segal approach),
- local nets of operator algebras (the Haag–Kastler approach) and,
- noncommutative geometry (the Connes approach).

As a warmup, we show in Chapter 2 that the regularization of divergent expressions represents a main subject in the history of mathematics starting with Euler in the eighteenth century. In this connection, we will consider

- the regularization of divergent series, and
- the regularization of divergent integrals.

In particular, in Sect. 2.1.3, we will discuss the classical Mittag–Leffler theorem on meromorphic functions f. If the function f has merely a finite number of poles, then the method of partial fraction decomposition works well. However, as a rule, this method fails if the function f has an infinite number of poles. In this case, Mittag–Leffler showed in the late nineteenth century that one has to subtract special terms, which are called subtractions by physicists.

The subtractions force the convergence of the infinite series.

This is the prototype of the method of iteratively adding subtractions in the Bogoliubov–Parasiuk–Hepp–Zimmermann (BPHZ) approach to renormalization theory. The corresponding iterative algorithm (called the Bogoliubov R-operation) has to be constructed carefully (because of nasty overlapping divergences). This was done by Nikolai Bogoliubov in the 1950s. An ingenious explicit solution formula for this iterative method was found by Wolfhart Zimmermann in 1969. This is the famous *Zimmermann forest formula*. In the late 1990s, it was discovered by Dirk Kreimer that the sophisticated combinatorics of the Zimmermann forest formula can be understood best in terms of a Hopf algebra generated by Feynman diagrams. By this discovery, the modern formulation of the BPHZ approach is based on both Hopf algebras and Rota–Baxter algebras.

As a warmup, in Chapter 3, we give an introduction to the modern combinatorial theory, which was founded by Gian-Carlo Rota (MIT, Cambridge, Massachusetts) in the 1960s. This includes both Hopf algebras and Rota–Baxter algebras.

Surprisingly enough, it turns out that the Zimmermann forest formula is closely related to methods developed by Lagrange in the eighteenth century when studying the solution of the Kepler equation for the motion of planets in celestial mechanics.

In modern terminology, the Lagrange inversion formula for power series expansions is based on the so-called Faà di Bruno Hopf algebra.[1] This will be studied in Sect. 3.4.3.

[1] The Italian priest and mathematician Francesco Faà di Bruno (1825–1888) was beatified in 1988.

In physics, symmetries are basic. For describing symmetries in terms of mathematics, there are two approaches based on

- groups and
- Hopf algebras.

In 1941, Heinz Hopf wanted to compute the cohomology of topological groups. Hopf discovered that the cohomology ring of topological groups is equipped with an additional structure which is called a Hopf algebra structure today. This additional algebraic structure is based on the notion of the coproduct. Roughly speaking, the concept of Hopf algebra is dual to the concept of group. Hopf algebras are intimately related to quantum groups. We will show in Chapter 3 that:

> *The product and the coproduct of a Hopf algebra model the fusion and the splitting of objects (e.g., elementary particles), respectively.*

In terms of analysis, the algebra of linear differential operators with constant coefficients can be equipped with the structure of a Hopf algebra. Here,

- the coproduct is related to the Leibniz product rule of differentiation, and
- the coinverse (also called the antipode) is related to the integration-by-parts formula (see Sect. 3.3.1).

The integration-by-parts formula is a special case of the general Stokes integral theorem, which lies at the heart of the duality between homology and cohomology in topology. This duality plays a key role for the mathematical description of processes in nature. In particular, cohomology is deeply rooted in Maxwell's theory of electrodynamics (see Sect. 4.4.7).

Incredible cancellations. When doing computations in renormalization theory, as a big surprise, physicists and mathematicians encounter incredible cancellations of a large amount of terms. This dramatically simplifies the final result. In terms of mathematics, a sophisticated combinatorics is behind these cancellations. The prototype for this is given by the Faà di Bruno Hopf algebra mentioned above.

The language of modern mathematics. We do not assume that the reader of this series of monographs is familiar with the language used in modern mathematics. In this connection, we want to help the reader. For example, many notions in advanced mathematics and hence in modern mathematical physics are based on mathematical operations applied to equivalence classes. For example, this concerns

- the construction of quantum states as equivalence classes of elements of Hilbert spaces (and the relation to projective geometry),
- the Gelfand–Naimark–Segal (GNS) construction for representing the elements of an abstract C^*-algebra as observables on a Hilbert space (the algebraic approach to quantum theory),
- the Wightman reconstruction theorem for axiomatically defined quantum fields (via the GNS-construction),

- moduli spaces of Riemann surfaces (modulo conformal equivalence) and physical states in string theory.

This leads to quotient spaces in algebra, analysis, geometry, and topology, which will be encountered again and again in this series of monographs (e.g., homology groups, cohomology groups, homotopy groups, and K-theory in topology). Chapter 4 serves as an introduction to quotient structures in mathematics and physics. The idea of the quotient ring (modulo a fixed integer) can be traced back at least to the *Disquisitiones arithmeticae* written by the young Gauss (1777–1855) in 1801.[2] In order to give the reader a feel for the usefulness of working with equivalence classes, we will consider the following examples:

- Gaussian quotient rings (modulo a prime number) and coding theory (as warmup for quantum information),
- quotient fields and Heaviside's symbolic method in electrical engineering (the Mikusiński operational calculus),
- physical fields, observers, bundles, and cocycles,
- deformation, mapping classes, and topological charges,
- loop spaces and higher homotopy groups,
- the projective and the injective limit of mathematical structures (e.g., topological spaces), and
- the rigorous approach to Leibniz's infinitesimals via ultrafilters (nonstandard analysis).

For the foundation of nonstandard analysis, one needs the construction of ultrafilers via Zorn's lemma based on the axiom of choice in set theory.

The language of theoretical physics. Chapters 5 through 9 are devoted to the basic ideas of

- classical geometric optics,
- classical mechanics, and
- quantum mechanics.

Here, we want to help mathematicians who are not familiar with theoretical physics. In Chapter 5, we study Carathéodory's royal road to geometrical optics based on the fundamental duality between

- light rays and
- wave fronts

which can be traced back to the work of Huygens in the seventeenth century. In string theory, Kähler manifolds play a crucial role. In Chapter 5, we will show how Poincaré's non-Euclidean geometry on the upper half-plane is related to both geometrical optics and Kähler geometry.

[2] The enormous influence of Gauss' first masterpiece on the development of mathematics is described in the monograph by C. Goldstein, N. Schappacher, and J. Schwermer: The Shaping of Arithmetic after Gauss' Disquisitiones Arithmeticae, Springer, Berlin 2007.

Since all the models of quantum fields are based on the study of an infinite number of (slightly perturbed) harmonic oscillators in the setting of perturbation theory, we use the harmonic oscillator as a paradigm for both classical mechanics and quantum mechanics. In Chapter 6 on classical mechanics, we will study the following topics:

- Newtonian mechanics,
- Lagrangian mechanics (the Euler–Lagrange equation, the Jacobi accessory eigenvalue problem and Morse theory),
- Hamiltonian mechanics (the canonical dynamical system and the Hamilton–Jacobi partial differential equation), and
- Poissonian mechanics.

In particular, this concerns

- the Legendre transformation and contact geometry,
- the Hamiltonian flow and symplectic geometry,
- the *tangent bundle* of the position space (the position-velocity space also called the state space),
- the *cotangent bundle* of the position space (the position-momentum space also called the phase space),
- the Legendre transformation as a transformation from the tangent bundle to the cotangent bundle; the latter is equipped with a natural symplectic structure.

In terms of mathematics, the fundamental relation between symmetry and conservation laws in physics is related to

- the Noether theorem, and
- Poisson brackets and Lie's momentum map.

Quantum mechanics. The comprehensive Chapter 7 lies at the heart of this series of monographs. This chapter should help the reader to understand the different aspects of the passage from classical physics to quantum physics, by using the different procedures of *quantization*. We will use the paradigm of the harmonic oscillator in order to explain the basic ideas of the following approaches:

- Heisenberg's quantum mechanic (via creation and annihilation operators),
- Schrödinger's quantum mechanics (via the Schrödinger partial differential equation),
- Feynman's quantum mechanics (via the path integral),
- von Neumann's functional-analytic approach (via the spectral theory for self-adjoint operators in Hilbert spaces),
- von Neumann's density operator in statistical physics (via trace class operators),
- Weyl's symbolic calculus for pseudo-differential operators (deformation quantization),
- the Poincaré–Wirtinger calculus and Bargmann's holomorphic quantization,
- the Stone-von Neumann uniqueness theorem (for the fundamental commutation relations in quantum mechanics) and the Weyl functor[3] based on symplectic geometry,
- supersymmetric quantization.

[3] At this place, the general theory of mathematical structures (also called category theory) enters the theory of quantization (also called quantum mathematics).

Concerning the Feynman path integral as a fundamental tool in quantum physics, we will study the following:

- Brownian motion and the infinite-dimensional rigorous Wiener integral based on measure theory,
- the rigorous Feynman–Kac formula for diffusion processes,
- rigorous finite-dimensional Gaussian integrals, the computation of correlations and moments, the Wick theorem, and Feynman diagrams,
- rigorous definition of infinite-dimensional Gaussian integrals via zeta function regularization,
- the Wentzel–Kramers–Brioullin (WKB) method of stationary phase for the computation of Gaussian integrals, and the approximate computation of Feynman path integrals.

The Feynman path integral can be obtained from the Wiener integral by using formal analytic continuation from real time to imaginary time. This corresponds to the fact that the Schrödinger equation describes diffusion processes in imaginary time. Furthermore, in Chapter 7, we discuss the basic ideas of the *algebraic approach* to quantum mechanics by using C^*-algebras and von Neumann algebras. In this connection, we consider:

- applications to statistical mechanics (Boltzmann statistics, Bose–Einstein statistics, and Fermi–Dirac statistics),
- thermodynamic equilibrium states (Kubo–Martin–Schwinger (KMS) states) and the Tomita–Takesaki theory for von Neumann algebras,
- the Murray–von Neumann classification of factors in the theory of von Neumann algebras,
- projection operators and the main theorem of quantum logic (Gleason's extension theorem for C^*-algebras).

The modern theory of operator algebras culminates in Alain Connes's noncommutative geometry, which represents the appropriate mathematical structure for a deeper understanding of the Standard Model in elementary particle physics. This will be investigated in Volume IV on quantum mathematics. For the interested reader, we refer to the following fundamental monograph:

> A. Connes and M. Marcolli, Noncommutative Geometry, Quantum Fields, and Motives, American Mathematical Society 2008.
> Internet: `http://www.math.fsu.edu/~ marcolli/bookjune4.pdf`

Chapters 8 and 9 serve as an introduction to scattering theory, which plays a crucial role in elementary particle physics. As a paradigm for general scattering theory, we consider the scattering of a quantum particle on the real line. We consider

- the energy levels of bound states,
- the energy levels of scattering states, and distributions as generalized eigenfunctions of the Schrödinger equation,
- the transition matrix,
- the unitary S-matrix and transition probabilities for scattering processes,
- the relation between the singularities of the S-matrix in the complex energetic plane and the energy levels of stable bound states,

- unstable particles (resonances) and the second sheet of the energetic Riemann surface (the Breit–Wigner formula for the energy levels and the mean lifetime of resonances),
- stationary scattering theory, the Green's function of the Helmholtz equation, and the Lippmann–Schwinger integral equation,
- instationary scattering theory, wave operators, the absolutely continuous spectrum of the Hamiltonian, and the S-matrix in functional analysis.

Here, we do not assume that the reader is familiar with

- von Neumann's functional-analytic spectral theory for self-adjoint operators in Hilbert spaces,
- the Gelfand–Kostyuchenko theory of generalized eigenfunctions for self-adjoint operators,
- the Møller–Kato theory of wave operators in scattering theory, and
- the Weyl–Kodaira theory for singular differential operators.

For the convenience of the reader, the necessary material will be summarized at the proper place when it is needed in Volumes II and III.

Quantum electrodynamics. In the present volume, it is our main goal to illustrate the beauty of quantum electrodynamics by proceeding pragmatically.

> *We do not start with an abstract approach, but with the computation of important physical effects which are observed in experiments, including radiative corrections in lowest order of renormalization theory.*

This should help the reader in getting a feel for the essential questions. More sophisticated approaches are postponed to later volumes of this series of monographs. In the introductory Chapter 10, we study creation and annihilation operators for electrons, positrons, and photons. In Chapter 11, we formulate the classical field equations of quantum electrodynamics on the interaction between electrons and photons, by coupling the Maxwell equations of the electromagnetic field to the Dirac equation of the electron wave function. This equation depends on the gauge fixing of the four-potential for the electromagnetic field. However, it turns out that physical effects measured in experiments are independent of the choice of the gauge fixing. The point is that:

> *The classical field equations of quantum electrodynamics have to be quantized.*

In this connection, we have to distinguish between

- the single free quantum fields for electrons, positrons, and photons, and
- the total interacting quantum field for electrons, positrons, and photons.

In Chapter 12, we construct free quantum fields by using the method of Fourier quantization based on creation and annihilation operators.

In Chapter 13, we study the interacting quantum field of electrons, positrons, and photons by using

- the magic Dyson formula for the S-matrix (scattering matrix), and
- the main Wick theorem for the S-matrix, which implies the Feynman diagrams.

This is Dyson's classical approach to understanding the Feynman diagrams.[4] Originally, Feynman invented his exciting diagram technique on the basis of ingenious physical intuition. In Dyson's mathematical setting, the Feynman diagrams are nothing other than graphical representations of well-defined analytic expressions, which are effectively produced by the main Wick theorem. Feynman's use of propagators and Dyson's magic formula for the S-matrix are closely related to Lagrange's variation-of-parameter formula in celestial mechanics. Many mathematicians complain about the following situation:

> *In the physics textbooks, one reads the Feynman rules for Feynman diagrams, but it is not clear where the Feynman rules come from.*

In the present textbook, we will thoroughly study the mathematical and physical origin of both the Feynman diagrams and the Feynman rules. We will also consider applications to interesting physical effects.

In Chapter 15, we investigate the following physical effects in lowest order of perturbation theory:

- the cross section for Compton scattering between photons and electrons (improvement of the Thomson formula in classical electrodynamics),
- the cross section for the scattering of electrons in an external electromagnetic field,
- the intensity of spectral lines for bound states in an external electromagnetic field, and
- the Cherenkov radiation.

For the computation of terms corresponding to higher order of perturbation theory, renormalization theory is needed. In Chapter 17, we discuss the physics behind the following radiative corrections in lowest possible order of renormalization theory:

- the screening of the Coulomb potential by vacuum polarization (the Uehling potential),
- the anomalous magnetic moment of the electron (the Schwinger formula),
- the anomalous magnetic moment of the muon, and
- the Lamb shift in the spectrum of the hydrogen atom.

Unfortunately, the explicit computations (in the framework of dimensional regularization in renormalization theory) are lengthy. We will postpone these detailed computations to Volume III.

In Chapter 18, we discuss the main result telling us that quantum electrodynamics can be renormalized to all orders of perturbation theory. The final result consists of getting finite expressions in each order of perturbation theory (e.g., cross sections for scattering processes), which depend on the two fundamental free parameters

[4] Dyson's discovery of this approach is described by himself in his book, F. Dyson, *Disturbing the Universe,* Harper & Row, New York, 1979 (see page 27 of Volume I for this fascinating story).

- m_{eff} (effective mass of the electron) and
- $-e_{\text{eff}}$ (effective electric charge of the electron).

Observe the crucial fact that:

> *The free parameters m_{eff} and e_{eff} cannot be determined theoretically by quantum electrodynamics.*

They have to be determined by physical experiments. In the SI system, one obtains the following values:

$$m_{\text{eff}} = 0.511 \text{ MeV}/c^2, \qquad e_{\text{eff}} = 1.602 \cdot 10^{-19} \text{ As}.$$

A reader who wants to become familiar with quantum electrodynamics as quickly as possible should start reading with Chapter 10.

The incredible effectiveness of perturbation theory in physics. Surprisingly enough, low-order radiative corrections are sufficient for getting a fantastic coincidence between theory and experimental data. For example, the anomalous magnetic moment of the electron measured in experiments is predicted very precisely by fourth-order radiative corrections (up to 9 digits). However, the necessary amount of computations is enormous. One has to evaluate high-dimensional integrals which correspond to 891 Feynman diagrams; this needs years of supercomputer time.

A warning to the reader. In summer 1976, Arthur Wightman (Princeton University) organized a famous conference in Erice (Sicily/Italy) on renormalization theory. In the introduction to the Proceedings of this conference, he writes:[5]

> Renormalization theory is a notoriously complicated and technical subject... I want to tell stories with a moral for the earnest student: Renormalization theory has a history of egregious errors by distinguished savants (see page 967). It has a justified reputation for perversity; a method that works up to 13th order in the perturbation theory fails in the 14th order. Arguments that sound plausible often dissolve into mush when examined closely. The worst that can happen often happens. The prudent student would do well to distinguish sharply between what has been proved and what has been plausible, and in general he should watch out!

In 1999 Gerardus 't Hooft and Martinus Veltman were awarded the Nobel prize in physics for their contributions to the renormalization of the theory of electroweak interaction and for the computation of radiative corrections in the Standard Model of particle physics.

Perspectives. More advanced approaches to renormalization theory will be systematically studied in the following volumes of this series of monographs. In particular, this concerns the new approaches to perturbative quantum field theory due to Connes and Kreimer (Hopf algebras), and Brunetti, Dütsch,

[5] A. Wightman, Orientation. In: Renormalization Theory, pp. 1–20. Edited by G. Velo and A. Wightman, Reidel, Dordrecht, 1976 (reprinted with permission).

and Fredenhagen (microlocal analysis and the Master Ward Identity). In order to give the reader an overview on the large variety of different approaches to renormalization theory, we summarize important references in Section 19.11, and in Chapter 19 we sketch some basic ideas.

The propagation of light in the universe, namely,

- the deflection of light at the sun, and
- the red shift of spectral lines as a consequence of the expansion of the universe (the Hubble effect)

will be investigated in Volume III in terms of Einstein's theory of general relativity.

The basic idea of our approach to quantum electrodynamics. As a rule, mathematicians have trouble with reading some textbooks on quantum field theory written by physicists. The point is that:

> *In mathematics, one never does computations with quantities which do not exist.*[6]

In order to respect this basic principle in mathematics, we will use the *lattice approach*. That is, roughly speaking, we will proceed in the following two steps.

Step 1: The discretized physical system. We put the physical system in a cubic box of finite side length L and volume $\mathcal{V} = L^3$. The boundary conditions are given by periodicity.

- We observe the physical system in a finite time interval $[-\frac{T}{2}, \frac{T}{2}]$.
- We choose a maximal energy $E_{\max}$.
- In the 3-dimensional momentum space, we introduce a finite lattice of spacing Δp and maximal momentum $P_{\max}$. In this setting, the Fourier integral transform is replaced by a discrete Fourier transform via finite Fourier series expansions.
- We define Dyson's S-matrix for this situation.
- The main Wick theorem allows us to compute the S-matrix elements (i.e., the transition amplitudes) in an elegant manner, by eliminating the creation and annihilation operators, and replacing them by propagators (i.e., correlation functions for free fields).
- The point is that the propagators are discrete algebraic Feynman integrals, which are indeed well-defined finite sums.
- The transition amplitudes can be graphically represented by Feynman diagrams.
- The Feynman rules allow us to translate the Feynman diagrams into well-defined finite sums.
- From the transition amplitudes, we obtain the transition probabilities which yield the cross sections for scattering processes. Note that cross sections can be measured in particle accelerators.

Step 2: The delicate continuum limit. Explicitly, we have to study the following limits:

- $L \to \infty$ (the volume L^3 of the cubic box becomes infinite),

[6] For example, this concerns infinite renormalization constants or ill-defined infinite-dimensional functional/path integrals.

- $T \to \infty$ (the time interval becomes infinite),
- $P_{\max} \to \infty$ (i.e., $E_{\max} \to \infty$) (high-energy limit),
- $\Delta p \to 0$ (low-energy limit).

In order to force the convergence of the discrete algebraic Feynman integrals to well-defined expressions, we modify the classical Lagrangian density by setting

$$m_{\text{eff}} := m_e + \delta m, \qquad e_{\text{eff}} := e + \delta e.$$

That is, we replace the so-called bare electron mass m_e and the so-called bare electron charge $-e$ in the Lagrangian density by

$$m_e = m_{\text{eff}} - \delta m, \qquad -e = -e_{\text{eff}} + \delta e,$$

respectively. This way, the classical Lagrangian density

$$\mathcal{L}(\psi, \partial\psi, A, \partial A; m_e, e)$$

is modified by the function

$$\mathcal{L}_{\text{modified}}(\psi, \partial\psi, A, \partial A; m_{\text{eff}}, e_{\text{eff}}; \delta m, \delta e).$$

The terms multiplied by $\delta m, \delta e$ are called counterterms of the classical Lagrangian density $\mathcal{L}$. Note that in this lattice approach, δe and δm are real parameters which depend on the shape of the lattice, that is, they depend on the maximal energy $E_{\max}$. Now consider the high-energy limit

$$E_{\max} \to +\infty.$$

Roughly speaking, we have to show that $\delta m(E_{\max})$ and $\delta e(E_{\max})$ can be chosen in such a way that the finite continuum limit exists for the S-matrix elements (i.e., the transition elements). This is the procedure of *renormalization*.

Observe the following peculiarity. By the Stone–von Neumann uniqueness theorem, a *finite* number of creation and annihilation operators is uniquely determined by the commutation relations (up to unitary equivalence). This is not true anymore for an *infinite* number of creation and annihilation operators, as was shown by Lars Gårding and Arthur Wightman in 1954. However, our approach avoids the latter ambiguity, since we only work with a finite number of creation and annihilation operators before passing to the continuum limit (of the vacuum expectation values). We also would like to emphasize that our approach differs only slightly from the usual approach used by physicists. In particular, we use a notation for discrete Fourier integrals such that the formal passage to the language used by physicists is possible at each stage of our procedure.

For the physical quantities which can be measured in experiments, our final formulas coincide with the formulas used by physicists.

Moreover, in each step of our procedure it is easy to pass formally to the expressions used by physicists, since the Feynman diagrams are the same.

> *This way, we hope to help mathematicians in getting a better understanding for the ingenious and beautiful approach invented by physicists.*

From the physical point of view, the modification of the classical Lagrangian density reflects the fact that:

> *Quantum effects have to be added to the classical theory.*

Intuitively, this means that the quantum fluctuations of the ground state of the quantum field of electrons, positrons, and photons influence physical effects observed in experiments. For example, this concerns the anomalous magnetic moment of the electron and the spectrum of the hydrogen atom (Lamb shift).

Convention. If we do not expressively state the opposite, we will use the SI system of physical units (international system of units) which can be found in the Appendix to Volume I. In particular, note that in Chapters 10–19 on quantum electrodynamics, we will use the energetic system with $c = 1$ (velocity of light in a vacuum), $\hbar = h/2\pi = 1$ (Planck's quantum of action), $k = 1$ (Boltzmann constant), $\varepsilon_0 = \mu_0 = 1$ (see page 790).

The Poincaré Seminar. The best way of getting information about recent developments in modern physics is to look at the books which report the lectures given at the Poincaré Seminar in Paris. Starting in 2002, this seminar has been organized by *l'Institut Henri Poincaré* in Paris (see page 1050). Bertrand Duplantier and Vincent Rivasseau write in the Foreword to *Quantum Spaces*, Birkhäuser, Basel, 2007:

> This book is the seventh in a series of lectures of the *Séminaire Poincaré*, which is directed towards a large audience of physicists and mathematicians.
> The goal of this seminar is to provide up-to-date information about general topics of great interest in physics. Both the theoretical and experimental aspects are covered, with some historical background. Inspired by the *Séminaire Bourbaki* in mathematics in its organization, hence nicknamed "Séminaire Bourbaphy," the Poincaré Seminar is held twice a year at the *Institut Henri Poincaré* in Paris, with contributions prepared in advance. Particular care is devoted to the pedagogical nature of the presentation so as to fulfill the goal of being readable by a large audience of scientists.

Two recent survey volumes. The following two volumes try to reflect the state of the art by summarizing the most important approaches used in modern quantum field theory:

- B. Fauser, J. Tolksdorf, and E. Zeidler (Eds.), Quantum Gravity: Mathematical Models and Experimental Bounds, Birkhäuser, Basel, 2006.
- B. Fauser, J. Tolksdorf, and E. Zeidler (Eds.), Quantum Field Theory – Competitive Methods, Birkhäuser, Basel, 2008.

Acknowledgements. I am very grateful to my collaborators Bertfried Fauser and Jürgen Tolksdorf for organizing two workshops in 2005 and 2007 on recent developments in quantum field theory, and for both doing an excellent editorial job and telling me new research strategies and new results in the algebraic and geometric approach to quantum field theory. In particular, I would like to thank Bertfried Fauser for drawing my attention to the relation between quantum field theory and the following topics: Gian-Carlo Rota's work on combinatorics, the ubiquitous Hopf algebras, Rota–Baxter algebras, quantum Clifford algebras, the Rota–Stein cliffordization process, Littlewood's work on the representation theory of the symmetric group, motivic number theory, topos theory, and category theory in algebraic geometry. For illuminating discussions about promising new approaches to (perturbative) quantum field theory in both flat and curved space-time, I would like to thank Klaus Fredenhagen and the late Julius Wess, as well as Detlev Buchholz, Christian Brouder, Romeo Brunetti, Michael Dütsch, Kurusch Ebrahimi-Fard, Joel Feldman, Felix Finster, Christian Fleischhack, Alessandra Frabetti, Stefan Hollands, Harald Grosse, Jerzy Kijowski, Dirk Kreimer, Elisabeth Kraus, Alexander Lange, Matilde Marcolli, Mario Paschke, Klaus Rehren, Gerd Rudolph, Manfred Salmhofer, Alexander Schmidt, Klaus Sibold, Rainer Verch, Hartmut Wachter, and Raimar Wulkenhaar.

I would like to thank Thomas Hahn from the Max Planck Institute for Physics, Werner Heisenberg, in Munich for informing me on the state of the art in automated multi-loop computations in perturbation theory. Such sophisticated computer programs are used for preparing the experiments at the LHC (Large Hadron Collider) of CERN (European Organization for Nuclear Research at Geneva, Switzerland) (see Sect. 18.4).

On the occasion of Professor Friedrich Hirzebruch's 80th birthday on October 17 in 2007, I would like to dedicate this volume to him in gratitude. His scientific work deeply influenced the development of mathematics in the second half of the twentieth century. Nowadays physicists frequently use Hirzebruch's results in order to study the topological structure of physical fields. In 1982, Friedrich Hirzebruch founded the Max Planck Institute for Mathematics in Bonn (Germany). Mathematicians from all over the world enjoy doing research in the relaxed and highly stimulating atmosphere of this institute. In 1996, the Max Planck Institute for Mathematics in the Sciences was founded in Leipzig. Friedrich Hirzebruch was the chairman of the Founders' Committee. The staff and the visitors of our institute are very grateful to Professor Hirzebruch for his efforts made as chairman.

For helping me to save a lot of time, I am very grateful to my secretary Regine Lübke (invaluable support), Katarzyna Baier (answering patiently almost infinitely many bibliographical questions), the library team (steadily support), Kerstin Fölting (graphics and tables), Micaela Krieger-Hauwede (answering patiently my LaTeX questions), Katrin Scholz (internet searching), and Thomas Heid (computer expert). I also would like to thank the staff of

the Springer publishing house in Heidelberg, Ruth Allewelt, Joachim Heinze, and Martin Peters, for the harmonious collaboration. Many years ago, my Czech colleague from Prague, the late Svatopluk Fučik, wrote the following sentence in the preface to his book on nonlinear partial differential equations: "Finally, I would like to thank myself for typing the manuscript." I resist the temptation of repeating this sentence with respect to my LaTeX file. But I would like to thank Donald Knuth for enjoying the mathematical community with the beautiful gift of the TeX tool.

In the Proverbs 31, 10 of the Bible, one reads:

> Who can find a virtuous woman? For her price is far above rubies. The heart of her husband does safely trust in her, so that he shall have no need of spoil. She will do him good and not evil all the days of her life.

I am very grateful to my beloved wife, Christine, who has been taking care of me for 40 years.

Leipzig, Summer 2008 *Eberhard Zeidler*

Contents

Prologue

One thing I have learned in a long life: that all our science, measured against reality, is primitive and childlike – and yet it is the most precise thing we have.

Albert Einstein (1879–1955)

The development of quantum mechanics in the years 1925 and 1926 had produced rules for the description of systems of microscopic particles, which involved promoting the fundamental dynamical variables of a corresponding classical system into operators with specified commutators. By this means, a system, described initially in classical particle language, acquires characteristics associated with the complementary classical wave picture. It was also known that electromagnetic radiation contained in an enclosure, when considered as a classical dynamical system, was equivalent energetically to a denumerably infinite number of harmonic oscillators. With the application of the quantization process to these fictitious oscillators, the classical radiation field assumed characteristics describable in the complementary classical particle language. The ensuing theory of light quantum emission and absorption by atomic systems[7] marked the beginning of quantum electrodynamics...

When it was attempted to quantize the complete electromagnetic field,[8] difficulties were encountered that stem from the gauge ambiguity of the potentials that appear in the Lagrangian formulation of the Maxwell equations...

From the origin of quantum electrodynamics, in the classical theory of point charges, came a legacy of difficulties.[9] The coupling of an electron with the electromagnetic field implied an infinite displacement, and, indeed, an infinite shift of all spectral lines emitted by an atomic system;[10]

[7] P. Dirac, The quantum theory of the emission and absorption of radiation, Proc. Royal Soc. Ser. A **14** (1927), 244–265.

[8] W. Heisenberg and W. Pauli, On the quantum electrodynamics of wave fields (in German), Z. Phys. **56** (1929), 1–61; **59** (1930), 108–190.

[9] H. Lorentz, Theory of Electrons, Dover, New York, 1915.

[10] R. Oppenheimer, Note on the interaction of field and matter, Phys. Rev. **35** (1930), 461–477.

> in the reaction of the electromagnetic field stimulated by the presence of the electron, arbitrary short wave lengths play a disproportionate and divergent role. The phenomenon of *electron-positron pair creation,* which finds a natural place in the relativistic electron field theory, contributes to this situation in virtue of the fluctuating densities of charge and current that occur even in the *vacuum state*[11] as the matter-field counterpart of the fluctuations in electric and magnetic field strengths.[12]
>
> In computing the energy of a single electron relative to that of the vacuum state, it is of significance that the presence of the electron tends to suppress the charge-current fluctuations induced by the fluctuating electromagnetic field. The resulting electron energy, while still divergent in its dependence upon the contributions of arbitrarily short wave lengths exhibits only a *logarithmic* infinity;[13] the combination of quantum and relativistic effects has destroyed all correspondence with the classical theory and its strongly structured-dependent electromagnetic mass.
>
> The existence of current fluctuations in the vacuum has other implications, since the introduction of an electromagnetic field induces currents that tend to modify the initial field; the "vacuum" acts as a *polarizable* medium.[14]
>
> New nonlinear electromagnetic phenomena appear, such as the *scattering of one light beam* by another, or by an electrostatic field...
>
> It is not likely that future developments will change drastically the practical results of the electron theory, which gives contemporary quantum electrodynamics a certain enduring value. Yet the real significance of the work of the past decade lies in the recognition of the ultimate problems facing electrodynamics, the problems of *conceptual consistency* and of *physical completeness.* No final solution can be anticipated until physical science has met the heroic challenge to comprehend the structure of the sub-microscopic world that nuclear exploration has revealed.[15]
>
> Julian Schwinger, 1958

This quotation is taken from a beautiful collection of 34 papers which played a fundamental role in the development of quantum electrodynamics. This volume was edited by Julian Schwinger from Harvard University who himself made fundamental contributions to this fascinating field of contemporary physics.

In the present volume, we will use Dyson's extremely elegant approach to quantum electrodynamics based on the Dyson series for the S-matrix (scat-

[11] The ground state of a quantum field is also called the vacuum state, by abuse of language. Note that the vacuum state is full of physics. In particular, quantum fluctuations of the vacuum state cause the essential physical effects observed in physical experiments.

[12] W. Heisenberg, On electric charge fluctuations caused by electron-positron pair creation, Sächsische Akademie der Wissenschaften, Leipzig, Vol. 86 (1934), 317–323 (in German).

[13] V. Weisskopf, On the self-energy and the electromagnetic field of the electron, Phys. Rev. **56** (1939), 72–86.

[14] P. Dirac, Théorie du positron, Rapport du 7^e Conseil Solvay de Physique 1934, pp. 203–212.

[15] J. Schwinger (Ed.), 34 Selected Papers on Quantum Electrodynamics, Dover Publications, New York, 1958 (reprinted with permission).

tering matrix). In the beginning of his *Selected Papers*, Freeman Dyson (born 1923) describes the history of quantum electrodynamics:[16]

> My first stroke of luck was to find Nicholas Kemmer in Cambridge (England) in 1946. He was the teacher I needed. He rapidly became a friend as well as a teacher.[17] Our friendship is still alive and well after 45 years. Kemmer gave two courses of lectures in Cambridge, one on nuclear physics and one on quantum field theory. In 1946, the only existing text-book on quantum field theory was the book "Quantentheorie der Wellenfelder", by Gregor Wentzel (1898–1978) written in Zürich and published in 1943 in Vienna in the middle of the war. Kemmer had been a student of Wentzel and possessed a copy of Wentzel's book. It was at that time a treasure without price. I believe there were then only two copies in England. It was later reprinted in America and translated into English.[18] But in 1946, few people in America knew of its existence and fewer considered it important. Kemmer not only possessed a copy, he also lent it to me and explained why it was important...
>
> In 1947, I arrived at Cornell as a student and found myself, thanks to Kemmer, the only person in the whole university who knew about quantum field theory. The great Hans Bethe (1906–2005) and the brilliant Richard Feynman (1918–1988) taught me a tremendous lot about many areas of physics, but when we were dealing with quantum field theory, I was the teacher and they were the students[19] ...
>
> Julian Schwinger (1918–1994) had known about quantum field theory long before. But he shared the American view that it was a mathematical extravagance, better avoided unless it should turn out to be essential. In 1948, he understood that it could be useful. He used it for calculations of the energy level shifts[20] revealed by the experiments of Lamb and Retherford,

[16] F. Dyson, Selected Papers of Freeman Dyson with Commentary, American Mathematical Society, Providence, Rhode Island, and International Press, Cambridge, Massachusetts (reprinted with permission).

[17] Nicholas Kemmer (1911–1998) was born in Saint Petersburg (Russia). In 1922, his family moved to Germany. He studied at the University of Göttingen. In 1940, he moved to Trinity College, Cambridge (England), to work on the wartime atomic energy project. He went to the University of Edinburgh from 1953–1979 as Tait Professor of Mathematical Physics, then Professor Emeritus. He was elected F.R.S. (Fellow of the Royal Society) in 1956.

[18] G. Wentzel, Quantum Theory of Wave Fields, Interscience, New York, 1949.

[19] Hans Bethe was born in Strasbourg (Alsace) in 1906. He studied at the University of Frankfurt/Main (Germany), and he obtained his Ph.D. at the University of Munich in 1928. In 1934, he emigrated to the United States, and he was appointed to a professorship at the Cornell University (Ithaca, New York). From 1943 until 1946, he worked in Los Alamos (New Mexico) (the Manhattan project for constructing the atomic bomb). In 1946, Bethe returned to Cornell and brought with him a group of brilliant young experimental and theoretical physicists. Among them was Richard Feynman. In 1967, Hans Bethe was awarded the Nobel prize in physics for his contributions to the theory of nuclear reactions, especially his discoveries concerning the energy production in stars. See H. Bethe, R. Bacher, and M. Livingstone, Basic Bethe: Seminal Articles on Nuclear Physics 1936–1937, American Institute of Physics, 1986.

[20] The first calculations of the Lamb shift were accomplished by Bethe in 1947; this was a highlight in quantum electrodynamics.

Foley and Kusch at Columbia.[21] But he used it grudgingly. In his publications, he preferred not to speak explicitly about quantum field theory. Instead, he spoke about *Green's Functions*. It turned out that the Green's Functions that Schwinger talked about and the quantum field theory that Kemmer talked about were fundamentally the same thing...

At Cornell, I was learning Richard Feynman's quite different way of calculating atomic processes. Feynman had never been interested in quantum field theory. He had his own private way of doing calculations. His way was based on things that he called "Propagators," which were probability amplitudes for particles to propagate themselves from one space-time point to another. He calculated the probabilities of physical processes by adding up the propagators. He had rules for calculating the propagators. Each propagator was represented graphically by a collection of diagrams. Each diagram gave a pictorial view of particles moving along straight lines and colliding with one another at points where the straight lines met. When I learned this technique of drawing diagrams and calculating propagators from Feynman, I found it completely baffling, because it always gave the right answer, but did not seem based on any solid mathematical foundation. Feynman called his way of calculating physical processes "the space-time approach," because his diagrams represented events as occurring at particular places and at particular times. The propagators described sequences of events in space and time. It later turned out that Feynman's propagators were merely another kind of Green's Functions. Feynman had been talking the language of Green's Functions all his life without knowing it.

Green's Functions also appeared in the work of Sin-Itiro Tomonaga (1906–1979), who had developed independently a new elegant version of relativistic quantum field theory. His work was done in the complete isolation of war-time Japan, and was published in Japanese in 1943. The rest of the world became aware of it only in the spring of 1948, when an English translation of it arrived at Princeton sent by Hideki Yukawa (1907–1981) to Robert Oppenheimer (1904–1967). Tomonaga was a physicist in the European tradition, having worked as a student with Heisenberg (1901–1976) at Leipzig before the war. For him, in contrast to Schwinger and Feynman, quantum field theory was a familiar and natural language in which to think.

After the war, Tomonaga's students had been applying his ideas to calculate the properties of atoms and electrons with high accuracy, and were reaching the same results as Schwinger and Feynman. When Tomonaga's papers began to arrive in America, I was delighted to see that he was speaking the language of quantum field theory that I had learned from Kemmer. It did not take us long to put the various ingredients of the pudding together. When the pudding was cooked, all three versions of the new theory of atoms and electrons turned out to be different ways of expressing the same basic idea. The basic idea was to calculate Green's Functions for all atomic processes that could be directly observed. Green's Functions appeared as the essential link between the methods of Schwinger and Feynman, and Tomonaga's relativistic quantum field theory provided the firm mathematical foundation for all three versions of quantum electrodynamics.

[21] Columbia University, New York

Dyson wrote two fundamental papers on the foundations of quantum electrodynamics, which are now classics:

> F. Dyson, The radiation theories of Tomonaga, Schwinger, and Feynman, Phys. Rev. **75** (1949), 486–502.
>
> F. Dyson, The S-matrix in quantum electrodynamics, Phys. Rev. **75** (1949), 1736–1755.

The fascinating story of the first paper can be found on page 27 of Vol. I. Dyson's second paper on renormalization theory starts as follows:

> The covariant (i.e., relativistically invariant) quantum electrodynamics of Tomonaga, Schwinger, and Feynman is used as the basis for a general treatment of scattering problems involving electrons, positrons, and photons. Scattering processes, including the creation and annihilation of particles, are completely described by the S-matrix (scattering matrix) of Heisenberg.[22] It is shown that the elements of this matrix can be calculated by a consistent use of perturbation theory to any order in the fine-structure constant. Detailed rules are given for carrying out such calculations, and it is shown that divergences arising from higher order radiative corrections can be removed from the S-matrix by a consistent use of the ideas of *mass and charge renormalization.*
> Not considered in this paper are the problems of extending the treatment to bound-state phenomena, and of proving the convergence of the theory as the order of perturbation itself tends to infinity.[23]

In 1950, John Ward published a short note where he used a highly formal argument in order to get a specific identity:[24]

> It has been recently proved by Dyson that all divergences in the S-matrix may be removed by a renormalization of mass and charge. Dyson defines certain fundamental divergent operators Γ_μ, S'_F, D'_F and gives a procedure for their finite parts $\Gamma_{\mu 1}, S'_{F1}, D'_{F1}$ by a process of successive approximation. It is then shown that
>
> $$\Gamma_\mu = Z_1^{-1}\Gamma_{\mu 1}(e_1), \quad S'_F = Z_2 S'_{F1}(e_1), \quad D'_F = Z_3 D'_{F1}(e_1),$$
> $$e_1 = Z_1^{-1} Z_2 Z_3^{1/2} e,$$
>
> where Z_1, Z_2, and Z_3 are certain *infinite* constants and e_1 is the (finite) renormalized electronic charge. Dyson conjectured that $Z_1 = Z_2$ and it is proposed here to give a *formal* proof of this relation.

[22] W. Heisenberg, The observable quantities in particle physics I–III, Z. f. Phys. **120** (1943), 513–538, 673–702; **123** (1944), 93–112 (in German).

[23] Bound states were considered in:
H. Bethe and E. Salpeter, A relativistic equation for bound-state problems. Phys. Rev. **84** (1951), 1232–1242.
H. Bethe and E. Salpeter, Quantum Mechanics of One- and Two-Electron Atoms, Springer, Berlin, 1957.

[24] J. Ward, An identity in quantum electrodynamics, Phys. Rev. **78** (1950), p. 182 (letter to the editor).

Nowadays this identity is called the Ward identity; it is the prototype of the crucial Ward–Takahashi–Slavnov–Taylor identities in gauge field theories, which are consequences of (local) gauge symmetry. In quantum electrodynamics, the Ward identity guarantees the unitarity of the S-matrix; this is a decisive ingredient of S-matrix theory. In fact, the unitarity is crucial for relating elements of the S-matrix to transition probabilities (see Sect. 7.15 of Vol. I). If the unitarity of the S-matrix is violated, then the theory becomes meaningless from the physical point of view.

After thinking about the convergence problem in quantum electrodynamics for a long time, Dyson published his paper *Divergence of perturbation theory in quantum electrodynamics,* Phys. Rev. **85** (1952), 631–632. The abstract of this paper reads as follows:

> An argument is presented which leads tentatively to the conclusion that all the power-series expansions currently in use in quantum electrodynamics are divergent after the renormalization of mass and charge. The divergence in no way restricts the accuracy of practical calculations that can be made with the theory, but raises important questions of principle concerning the nature of the physical concepts upon which the theory is built.

Dyson's heuristic argument can be found in Sect. 15.5.1 of Vol. I. Silvan Schweber writes the following in his excellent history on quantum electrodynamics entitled *QED and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga,* Princeton University Press, Princeton, New Jersey, 1994:[25]

> The importance of Schwinger's 1947 calculation of the anomalous magnetic moment of the electron cannot be underestimated. In the course of theoretical developments, there sometimes occur important calculations that alter the way the community thinks about particular approaches. Schwinger's calculation is one such instance...
> The papers of Tomonaga, Schwinger, and Feynman did not complete the renormalization program since they confined themselves to low order calculations. It was Dyson who dared to face the problem of high orders and brought the program to completion. In magnificently penetrating papers, he pointed out and resolved the main problems of this very difficult analysis... Whatever the future may bring, it is safe to assert that the theoretical advances made in the unravelling of the constitution of matter since World War II comprise one of the greatest intellectual achievements of mankind. They were based on the ground secured by the contributions of Bethe, Tomonaga, Schwinger, Feynman, and Dyson to quantum field theory and renormalization theory in the period from 1946 to 1951.

For creating quantum electrodynamics, Richard Feynman, Julian Schwinger, and Sin-Itiro Tomonaga were awarded the Nobel prize in physics in 1965. Freemann Dyson was awarded the Wolf prize in physics in 1981. Working at the Institute for Advanced Study in Princeton, Dyson is one of the most influential intellectuals of our time; his research concerns mathematics (number

[25] Reprinted by permission of Princeton University Press.

theory, random matrices), physics (quantum field theory, statistical mechanics, solid state physics, stability of matter), astrophysics (interstellar communication), biology (origin of life), history, and philosophy of the sciences. Much material can be found in the *Selected Papers of Freeman Dyson*, Amer. Math. Soc., Providence, Rhode Island and International Press, Cambridge, Massachusetts, 1996. In particular, we refer to the following beautiful essays and books written by Dyson:

Essays:

- Mathematics in the physical sciences, Scientific American **211** (1964), 129–164.
- Missed opportunities, Bull. Amer. Math. Soc. **78** (1972), 635–652.
- George Green and Physics, Physics World **6**, August 1993, 33–38.
- A walk through Ramanujan's garden. Lecture given at the Ramanujan (1887–1920) Centenary Conference in 1987, University of Illinois. In: F. Dyson, Selected Papers, pp. 187–208.
- Foreword to J. Havil, Gamma: Exploring Euler's Constant, Princeton University Press, 2003.
- Foreword to P. Odifreddi, The Mathematical Century: The 30 Greatest Problems of the Last 100 Years, Princeton University Press, 2004.
- The Scientist as Rebel, New York Review Books, 2007.

Books:

- Disturbing the Universe, Harper and Row, New York, 1979.
- From Eros to Gaia, Pantheon Books, New York, 1992.
- Origins of Life, Cambridge University Press, 1999.
- The Sun, the Genome and the Internet: Tools of Scientific Revolution, Oxford University Press, 1999.

Elliott Lieb (Princeton University) writes the following in the foreword to Dyson's *Selected Papers* (reprinted with permission):

> If any proof be needed that theoretical physics papers are not ephemeral and are not written on a blackboard that has to be erased every five years, then the papers in this volume will supply ample witness. The writings of Freeman Dyson are among the jewels that crown the subject and today even the earliest among them can be read with profit and much pleasure by beginners and experts...
> Dyson along with Feynman, Schwinger, and Tomonaga was a founder of quantum electrodynamics. When I started my graduate studies in the fifties, it was not easy to find a coherent pedagogical representation of the new field, but fortunately, Dyson had given lectures at Cornell in 1951 and these were available as notes. Thanks to their clarity many people, including me, were able to enter the field.

Recently, these classic notes were published:

> F. Dyson, Advanced Quantum Mechanics: Cornell Lectures on Quantum Electrodynamics 1951, World Scientific, Singapore, 2007.

Feynman's approach to quantum electrodynamics was elegantly based on the use of graphs called Feynman diagrams today. David Kaiser writes the following in his book *Drawing Theories Apart: The Dispersion of Feynman Diagrams in Postwar Physics* (the University of Chicago Press, Chicago and London, 2005 – reprinted with permission):

> For all of Feynman's many contributions to modern physics, his diagrams have had the widest and longest-lasting influence. Feynman diagrams have revolutionized nearly every aspect of theoretical physics since the middle of the twentieth century. Feynman first introduced his diagrams in the late 1940s as a bookkeeping device for simplifying lengthy calculations in one area of physics – quantum electrodynamics, physicist's quantum-mechanical description of electromagnetic forces. Soon the diagrams gained adherents throughout the fields of nuclear and particle physics. Not long thereafter, other theorists adopted – and subtly adapted – Feynman diagrams for many-body applications in solid-state physics. By the end of the 1960s, some physicists even wielded the line drawings for calculations in gravitational physics. With the diagrams' aid, entire new calculational vistas opened for physicists; theorists learned to calculate things that many had barely dreamed possible before World War II. With the list of diagrammatic applications growing ever longer, Feynman diagrams helped to transform the way physicists saw the world, and their place within it.

There is no doubt that quantum electrodynamics is one of the most beautiful theories in theoretical physics. The following quotation is taken from the forthcoming article "Quantum Theory and Relativity" written by Arthur Jaffe, Contemporary Mathematics, American Mathematical Society, Providence, Rhode Island, 2008, pp. 209–245 (reprinted with permission):[26]

> Two major themes dominated twentieth century physics: *quantum theory* and *relativity*. These two fundamental principles provide the cornerstones upon which one might build the understanding of modern physics. And today after one century of elaboration of the original discoveries by Poincaré, Einstein, Bohr, Schrödinger, Heisenberg, Dirac – and many others – one still dreams of describing the forces of nature within such an arena. Yet we do not know the answer to the basic question:
>
> *Are quantum theory, relativity, and interaction mathematically compatible?*
>
> Even if one restricts relativity to special relativity, we do not know the answer to this question about our four-dimensional world – much less about other higher-dimensional worlds considered by string theorists.
> Should quantum theory with relativity not qualify as logic? Physics suggests that a natural way to combine quantum theory, special relativity and interaction is through a nonlinear quantum field. Enormous progress on this problem has been made over the past forty years. This includes showing that theories exist in space-times of dimension two and three. Building this new mathematical framework and finding these examples has become known as the subject of *constructive quantum field theory*...
> For centuries, the tradition in physics has been to describe natural phenomena by mathematics. Eugene Wigner marveled on the relevance of mathematics in his famous essay: "On the Unreasonable Effectiveness of Mathematics in the Natural Sciences," Comm. Pure Appl. Math. **13** (1960), 1–14. Intuition can go a long way. But by endowing physics with a mathematical foundation, one also bestows physical laws with longevity. For mathematical ideas can be understood and conveyed more easily than conjectures, both from person to person, and also from generation to generation.

[26] I would like to thank Arthur Jaffe for sending me the electronic version of this beautiful article before publishing it.

In recent years, we have witnessed enormous progress in another direction – of transferring ideas from physics to mathematics: to play on Wigner's title, concepts from physics have had an unreasonable effectiveness in providing insight to formulate mathematical conjectures! The resulting infusion of new perspectives has truly blossomed into a mathematical revolution, which has been sufficiently robust to touch almost every mathematical frontier...

1. Mathematical Principles of Modern Natural Philosophy

The book of nature is written in the language of mathematics.
Galileo Galilei (1564–1642)

At the beginning of the seventeenth century, two great philosophers, Francis Bacon (1561–1626) in England and René Descartes (1596–1650) in France, proclaimed the birth of modern science. Each of them described his vision of the future. Their visions were very different. Bacon said, "All depends on keeping the eye steadily fixed on the facts of nature." Descartes said, "I think, therefore I am." According to Bacon, scientists should travel over the earth collecting facts, until the accumulated facts reveal how Nature works. The scientists will then *induce* from the facts the laws that Nature obeys. According to Descartes, scientists should stay at home and *deduce* the laws of Nature by pure thought. In order to deduce the laws correctly, the scientists will need only the rules of logic and knowledge of the existence of God. For four hundred years since Bacon and Descartes led the way, science has raced ahead by following both paths simultaneously. Neither Baconian empiricism nor Cartesian dogmatism has the power to elucidate Nature's secrets by itself, but both together have been amazingly successful. For four hundred years English scientists have tended to be Baconian and French scientists Cartesian.
Faraday (1791–1867) and Darwin (1809–1882) and Rutherford (1871–1937) were Baconians; Pascal (1623–1662) and Laplace (1749–1827) and Poincaré (1854–1912) were Cartesians. Science was greatly enriched by the cross-fertilization of the two contrasting national cultures. Both cultures were always at work in both countries. Newton (1643–1727) was at heart a Cartesian, using pure thought as Descartes intended, and using it to demolish the Cartesian dogma of vortices. Marie Curie (1867–1934) was at heart a Baconian, boiling tons of crude uranium ore to demolish the dogma of the indestructibility of atoms.[1]

Freeman Dyson, 2004

It is important for him who wants to discover not to confine himself to a chapter of science, but keep in touch with various others.[2]

Jacques Hadamard (1865–1963)

[1] From Dyson's foreword to the book by P. Odifreddi, The Mathematical Century: The 30 Greatest Problems of the Last 100 Years, Princeton University Press, 2004. Reprinted by permission of Princeton University Press.

[2] J. Hadamard, The Psychology of Invention, Princeton University Press, 1945.

> Mathematics takes us still further from what is human, into the region of absolute necessity, to which not only the actual world, but every possible world must conform.[3]
>
> Bertrand Russel (1872–1972)

1.1 Basic Principles

There exist the following fundamental principles for the mathematical description of physical phenomena in nature.

(I) The infinitesimal principle due to Newton and Leibniz: The laws of nature become simple on an infinitesimal level of space and time.[4]

(II) The optimality principle (or the principle of least action): Physical processes proceed in such an optimal way that the action is minimal (or at least critical). Such processes are governed by ordinary or partial differential equations called the Euler–Lagrange equations.

(III) Emmy Noether's symmetry principle: Symmetries of the action functional imply conservation laws for the corresponding Euler–Lagrange equations (e.g., conservation of energy).

(IV) The gauge principle and Levi-Civita's parallel transport: The fundamental forces in nature (gravitational, eletromagnetic, strong, and weak interaction) are based on the symmetry of the action functional under local gauge transformations. The corresponding parallel transport of physical information generates the intrinsic Gauss–Riemann–Cartan–Ehresmann curvature which, roughly speaking, corresponds to the acting force (also called interaction). Briefly: force = curvature.

(V) Planck's quantization principle: Nature jumps.

(VI) Einstein's principle of special relativity: Physics does not depend on the choice of the inertial system.

(VII) Einstein's principle of general relativity: Physics does not depend on the choice of the local space-time coordinates of an observer.

(VIII) Dirac's unitarity principle: Quantum physics does not depend on the choice of the measurement device (i.e., on the choice of an orthonormal basis in the relevant Hilbert space). This corresponds to the invariance under unitary transformations.[5]

[3] The Earl of Russel was awarded the Nobel prize in literature in 1950. He worked in philosophy, mathematical logic, social sciences, and politics.

[4] I. Newton, Philosophiae naturalis principia mathematica (Mathematical principles of natural philosophy) (in Latin), 1687. Translated into English by A. Motte, in 1729, edited by F. Cajori, University of California Press, Berkeley, California, 1946. See also S. Chandrasekhar, Newton's Principia for the Common Reader, Oxford University Press, Oxford, 1997.

[5] Newton (1643–1727), Leibniz (1646–1716), Euler (1707–1783), Lagrange (1736–1813), Laplace (1749–1828), Legendre (1752–1833), Fourier (1768–1830), Gauss (1777–1855), Poisson (1781–1840), Faraday (1791–1867), Green (1793–1841),

Geometrization of physics. In mathematics, the properties of geometric objects do not depend on the choice of the coordinate system. This is similar to the principles (VI)–(VIII). Therefore, it is quite natural that geometric methods play a fundamental role in modern physics.

Linearity and nonlinearity. We have to distinguish between

(i) linear processes, and
(ii) nonlinear processes.

In case (i), the superposition principle holds, that is, the superposition of physical states yields again a physical state. Mathematically, such processes are described by linear spaces and linear operator equations. The mathematical analysis can be simplified by representing physical phenomena as superposition of simpler phenomena. This is the method of harmonic analysis (e.g., the Fourier method based on the Fourier series, the Fourier integral, or the Fourier–Stieltjes integral).

In case (ii), the superposition principle is violated. As a rule, interactions in nature are mathematically described by nonlinear operator equations (e.g., nonlinear differential or integral equations). The method of perturbation theory allows us to reduce (ii) to (i), by using an iterative method.

Basic properties of physical effects. For the mathematical investigation of physical effects, one has to take the following into account.

(A) Faraday's locality principle: Physical effects propagate locally in space and time (law of proximity theory).
(B) Green's locality principle: The response of a linear physical system can be described by localizing the external forces in space and time and by considering the superposition of the corresponding special responses (method of the Green's function). Furthermore, this can be used for computing nonlinear physical systems by iteration.
(C) Planck's constant: The smallest action (energy × time) in nature is given by the action quantum $h = 6.626\,0755 \cdot 10^{-34}$Js.
(D) Einstein's propagation principle: Physical effects travel at most with the speed of light c in a vacuum. Explicitly, $c = 2.997\,92458 \cdot 10^{8}$m/s.
(E) Gauge invariance principle: Physical effects are invariant under local gauge transformations. Physical experiments are only able to measure quantities which do not depend on the choice of the gauge condition.

Riemann (1826–1866), Maxwell (1831–1879), Lie (1842–1899), Klein (1849–1925), Poincaré (1854–1912), Planck (1858–1947), Élie Cartan (1859–1951), Hilbert (1862–1943), Minkowski (1864–1909), Levi-Civita (1873–1941), Einstein (1879–1955), Emmy Noether (1882–1935), Weyl (1885–1955), Schrödinger (1887–1961), Heisenberg (1901–1976), Dirac (1902–1984), Ehresmann (1905–1979), von Neumann (1903–1957), Tomonaga (1906–1979), Landau (1908–1968), Laurent Schwartz (1915–2002), Feynman (1918–1988), Schwinger (1918–1994), Yang (born 1922), Dyson (born 1923), Salam (1926–1996), Gell-Mann (born 1929), Glashow (born 1932), Weinberg (born 1933), Fritzsch (born 1943).

(F) The Planck scale hypothesis: Physics dramatically changes below the Planck length given by $l = 10^{-35}$m.

In what follows, let us discuss some basic ideas related to all of the principles summarized above. To begin with, concerning Faraday's locality principle, Maxwell emphasized the following:[6]

> Before I began the study of electricity I resolved to read no mathematics on the subject till I had first read through Faraday's 1832 paper *Experimental researches on electricity*. I was aware that there was supposed to be a difference between Faraday's way of conceiving phenomena and that of the mathematicians, so that neither he nor they were satisfied with each other's language. I had also the conviction that this discrepancy did not arise from either party being wrong. For instance, Faraday, in his mind, saw lines of force traversing all space where the mathematicians (e.g., Gauss) saw centers of force attracting at a distance; Faraday saw a medium where they saw nothing but distance; Faraday sought the seat of the phenomena in real actions going on in the medium, where they were satisfied that they had found it in a power of action at a distance impressed on the electric fluids.
>
> When I had translated what I considered to be Faraday's ideas into a mathematical form, I found that in general the results of the two methods coincide... I also found that several of the most fertile methods of research discovered by the mathematicians could be expressed much better in terms of the ideas derived from Faraday than in their original form.

1.2 The Infinitesimal Strategy and Differential Equations

> Differential equations are the foundation of the natural scientific, mathematical view of the world.
>
> Vladimir Igorevich Arnold (born 1937)

The infinitesimal strategy due to Newton and Leibniz studies the behavior of a physical system for infinitesimally small time intervals and infinitesimally small distances. This leads to the encoding of physical processes into differential equations (e.g., Newton's equations of motion in mechanics, or Maxwell's equations in electrodynamics).

> *The task of mathematics is to decode this information; that is, to solve the fundamental differential equations.*

1.3 The Optimality Principle

It is crucial that the class of possible differential equations is strongly restricted by the optimality principle. This principle tells us that the fundamental differential equations are the Euler–Lagrange equations to variational

[6] J. Maxwell, A Treatise on Electricity and Magnetism, Oxford University Press, Oxford, 1873.

problems. In 1918, Emmy Noether formulated her general symmetry principle in the calculus of variations. The famous Noether theorem combines Lie's theory of continuous groups with the calculus of variations due to Euler and Lagrange. This will be studied in Section 6.6.

1.4 The Basic Notion of Action in Physics and the Idea of Quantization

The most important physical quantity in physics is not the energy, but the action which has the physical dimension *energy times time*. The following is crucial.

(i) The fundamental processes in nature are governed by the principle of least action

$$\boxed{S = \min!}$$

where we have to add appropriate side conditions. In fact, one has to use the more general principal $S =$ critical! (principle of critical action). For example, if we consider the motion $q = q(t)$ of a particle of mass m on the real line, then the action is given by

$$S[q] := \int_{t_0}^{t_1} \left(\tfrac{1}{2} m\dot{q}(t)^2 - U(q(t))\right) dt.$$

Here, the function $U = U(q)$ is the potential, and the negative derivative, $-U'$, describes the acting force. In this case, the principle of critical action reads as

$$\boxed{S[q] = \text{critical!}, \qquad q(t_0) = q_0,\ q(t_1) = q_1} \tag{1.1}$$

where we fix the following quantities: the initial time t_0, the initial position q_0 of the particle, the final time t_1, and the final position q_1 of the particle. The solutions of (1.1) satisfy the Euler–Lagrange equation

$$m\ddot{q}(t) = F(t), \qquad t \in \mathbb{R}$$

with the force $F(t) = -U'(q(t))$. This coincides with the Newtonian equation of motion (see Sect. 6.5).

(ii) In 1900 Planck postulated that there do not exist arbitrarily small amounts of action in nature. The smallest amount of action in nature is equal to the Planck constant h. In ancient times, philosophers said:

Natura non facit saltus. (Nature does never make a jump.)

In his "Noveaux essais," Leibniz wrote:

Tout va par degrés dans la nature et rien par saut. (In nature everything proceeds little by little and not by jumping.)

In contrast to this classical philosophy, Planck formulated the hypothesis in 1900 that the energy levels of a harmonic oscillator form a discrete set. He used this fact in order to derive his radiation law for black bodies (see Sect. 2.3.1 of Vol. I). This was the birth of quantum physics. More generally, the energy levels of the bound states of an atom or a molecule are discrete. The corresponding energy jumps cause the spectra of atoms and molecules observed in physical experiments (e.g., the spectral analysis of the light coming from stars). Nowadays, we say that:

Nature jumps.

This reflects a dramatic change in our philosophical understanding of nature.

(iii) In order to mathematically describe quantum effects, one has to modify classical theories. This is called the process of quantization, which we will encounter in this series of monographs again and again. As an introduction to this, we recommend reading Chapter 7. Now to the point. Feynman discovered in the early 1940s in his dissertation in Princeton that the process of quantization can be most elegantly described by path integrals (also called functional integrals) of the form

$$\boxed{\int \mathrm{e}^{\mathrm{i}S[\psi]/\hbar}\,\mathcal{D}\psi}$$

where we sum over all classical fields ψ (with appropriate side conditions). Here, $\hbar := h/2\pi$. For example, the quantization of the classical particle considered in (i) can be based on the formula

$$G(q_0, t_0; q_1, t_1) = \int \mathrm{e}^{\mathrm{i}S[q]/\hbar}\,\mathcal{D}q.$$

Here, we sum over all classical motions $q = q(t)$ which satisfy the side condition $q(t_0) = q_0$ and $q(t_1) = q_1$. The Green's function G determines the time-evolution of the wave function ψ, that is, if we know the wave function $\psi = \psi(x, t_0)$ at the initial time t_0, then we know the wave function at the later time t by the formula

$$\psi(x,t) = \int_{\mathbb{R}} G(x,t;y,t_0)\psi(y,t_0)dy.$$

Finally, the wave function ψ tells us the probability

$$\int_a^b |\psi(x,t)|^2 dx$$

of finding the quantum particle in the interval $[a, b]$ at time t.

(iv) In quantum field theory, one uses the functional integral

$$\int \mathrm{e}^{\mathrm{i}S[\psi]/\hbar}\, \mathrm{e}^{\mathrm{i}\langle\psi|J\rangle}\, \mathcal{D}\psi$$

with the additional external source J. Differentiation with respect to J yields the moments of the quantum field. In turn, the moments determine the correlation functions (also called Green's functions). The correlation functions describe the correlations between different positions of the quantum field at different points in time. These correlations are the most important properties of the quantum field which can be related to physical measurements.

Feynman's functional integral approach to quantum physics clearly shows that both classical and quantum physics are governed by the classical action functional S. This approach can also be extended to the study of many-particle systems at finite temperature, as we have discussed in Sect. 13.8 of Vol. I. Summarizing, let us formulate the following general strategy:

The main task in modern physics is the mathematical description of the propagation of physical effects caused by interactions and their quantization.

In Sect. 1.9 we will show that in modern physics, interactions are described by gauge theories based on local symmetries.

1.5 The Method of the Green's Function

Basic ideas. As a prototype, consider the motion $x = x(t)$ of a particle of mass $m > 0$ on the real line under the action of the continuous force $F : \mathbb{R} \to \mathbb{R}$. The corresponding Newtonian equation of motion reads as

$$\boxed{m\ddot{x}(t) = F(t) \qquad \text{for all} \quad t \in \mathbb{R}} \tag{1.2}$$

with the initial condition

$$x(0) = a, \quad \dot{x}(0) = v.$$

We are given the initial position a and the initial velocity v at time $t = 0$. For simplifying notation, we set $m := 1$. In order to discuss Green's locality principle in physics, let us summarize the following facts. The unique solution of (1.2) reads as

$$x(t) = a + vt + \int_0^t (t - \tau)F(\tau)d\tau \qquad \text{for all} \quad t \in \mathbb{R}.$$

Equivalently, this can be written as

$$\boxed{x(t) = a + vt + \int_{-\infty}^{\infty} G(t,\tau)F(\tau)d\tau \qquad \text{for all} \quad t \in \mathbb{R}.} \tag{1.3}$$

The function G is called the Green's function of the differential equation (1.2). Explicitly,

$$G(t,\tau) := \begin{cases} t-\tau & \text{if} \quad 0 \le \tau \le t, \\ \tau - t & \text{if} \quad t \le \tau < 0, \\ 0 & \text{otherwise.} \end{cases}$$

Let us discuss the physical meaning of the Green's function G. To this end, for fixed positive number Δt and all times $t \in \mathbb{R}$, we introduce the Dirac Δt-delta function

$$\delta_{\Delta t}(t) := \begin{cases} \frac{1}{\Delta t} & \text{if} \quad 0 \le t \le \Delta t, \\ 0 & \text{otherwise.} \end{cases}$$

Obviously, we have

$$\lim_{\Delta t \to +0} \delta_{\Delta t}(t) = \begin{cases} +\infty & \text{if} \quad 0 \le t \le \Delta t, \\ 0 & \text{otherwise,} \end{cases}$$

and the normalization condition $\int_{-\infty}^{\infty} \delta_{\Delta t}(t)dt = 1$ is satisfied.

(i) Localized force. We are given the parameters $\Delta t > 0$ and $F_0 \in \mathbb{R}$. For fixed time t_0, we choose the special force

$$F(t) := F_0 \cdot \delta_{\Delta t}(t - t_0) \qquad \text{for all} \quad t \in \mathbb{R}.$$

By (1.3), the corresponding motion reads as

$$x_{\Delta t}(t) = a + vt + F_0 \cdot \frac{1}{\Delta t} \int_{t_0}^{t_0+\Delta t} G(t,\tau)d\tau \qquad \text{for all} \quad t \in \mathbb{R}.$$

Letting $\Delta t \to +0$, we get the motion[7]

$$x(t) = a + vt + F_0 \cdot G(t, t_0) \qquad \text{for all} \quad t \in \mathbb{R}. \tag{1.4}$$

This can be considered as the motion of the particle under the influence of the kick force $t \mapsto F_0\delta_{\Delta t}(t - t_0)$ at time t_0, as $\Delta t \to +0$. For $t_0 \ge 0$, the motion (1.4) looks like

[7] In fact, it follows from $\lim_{\varepsilon \to +0} G(t, t_0 + \varepsilon) = G(t, t_0)$ for all $t, t_0 \in \mathbb{R}$ that

$$\lim_{\Delta \to +0} \frac{1}{\Delta t} \int_{t_0}^{t_0+\Delta t} G(t,\tau)d\tau = G(t, t_0).$$

$$x(t) = \begin{cases} a + vt & \text{if} \quad t < t_0, \\ a + vt + F_0 \cdot (t - t_0) & \text{if} \quad t \geq t_0. \end{cases} \tag{1.5}$$

That is, the velocity jumps at time t_0. For $t_0 < 0$, (1.4) looks like

$$x(t) = \begin{cases} a + vt & \text{if} \quad t > t_0, \\ a + vt + F_0 \cdot (t_0 - t) & \text{if} \quad t \leq t_0. \end{cases} \tag{1.6}$$

(ii) Superposition of the original force by kick forces (physical interpretation of the Green's function). Fix $\Delta t > 0$. Consider the discrete points in time $n\Delta t$ where $n = 0, \pm 1, \pm 2, \ldots$ In terms of physics, let us approximate the given force $F = F(t)$ by a step function F_{approx}. That is, we use the superposition

$$F_{\text{approx}}(t) := \sum_{n=-\infty}^{\infty} F_n(t), \qquad t \in \mathbb{R}$$

of the kick forces $F_n(t) := F(n\Delta t)\delta_{\Delta t}(t - n\Delta t)\Delta t$. Explicitly,

$$F_n(t) = \begin{cases} F(n\Delta t) & \text{if} \quad n\Delta t \leq t \leq (n+1)\Delta t, \\ 0 & \text{otherwise.} \end{cases}$$

If Δt is sufficiently small, then the kick force F_n generates the approximate motion

$$x_n(t) = F(n\Delta t)G(t, n\Delta t)\Delta t, \qquad t \in \mathbb{R}$$

with $x_n(0) = 0$ and $\dot{x}_n(0) = 0$ for all $n = \pm 1, \pm 2, \ldots$ That is, the particle rests at the initial time $t = 0$. Consequently, by superposition, the force F_{approx} generates the approximate motion

$$x_{\text{approx}}(t) = \sum_{n=-\infty}^{\infty} x_n(t) = \sum_{n=-\infty}^{\infty} G(t, n\Delta t)F(n\Delta t)\Delta t, \qquad t \in \mathbb{R}.$$

As $\Delta t \to 0$, we get the motion $x(t) = \int_{-\infty}^{\infty} G(t, \tau)F(\tau)d\tau$ for all $t \in \mathbb{R}$.

The motions (1.5) and (1.6) have the following properties:

(a) $t \mapsto x(t)$ is continuous on $\mathbb{R}$.
(b) $t \mapsto x(t)$ is smooth on $\mathbb{R} \setminus \{t_0\}$, and $\ddot{x}(t) = 0$ for all $t \neq t_0$.
(c) $\dot{x}(t_0 + 0) = \dot{x}(t_0 - 0) + F_0$ (jump of the velocity at time t_0).
(d) In the sense of distributions, we have the following equation of motion:[8]

$$\boxed{m\ddot{x}(t) = F_0\delta(t - t_0), \qquad t \in \mathbb{R}.}$$

[8] We choose $m := 1$.

Let us prove (d). We have to show that

$$\int_{-\infty}^{\infty} x(t)\ddot{\varphi}(t)dt = F_0\varphi(t_0) \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}).$$

Noting (b), integration by parts yields that $\int_{t_0}^{\infty} x(t)\ddot{\varphi}(t)dt$ is equal to

$$-x(t_0)\dot{\varphi}(t_0) - \int_{t_0}^{\infty} \dot{x}(t)\dot{\varphi}(t)dt = -x(t_0)\dot{\varphi}(t_0) + \dot{x}(t_0+0)\varphi(t_0).$$

Similarly, $\int_{-\infty}^{t_0} x(t)\ddot{\varphi}(t)dt = x(t_0)\dot{\varphi}(t_0) - \dot{x}(t_0-0)\varphi(t_0)$. Finally, use (c). □

Examples. Fix $t_0 := 0$ and $F_0 := 1$. If we choose $a = v := 0$, then the motion (1.5) looks like

$$x(t) = \theta(t)t \qquad \text{for all} \quad t \in \mathbb{R}.$$

Here, θ denotes the Heaviside function.[9] If we choose $a := 0$ and $v := -1$, then the motion (1.5) looks like

$$x(t) = -\theta(-t)t \qquad \text{for all} \quad t \in \mathbb{R}.$$

Finally, if we choose $a := 0, v := -\frac{1}{2}$, then the motion (1.5) looks like

$$x(t) = \tfrac{1}{2}(\theta(t)t - \theta(-t)t) = \tfrac{1}{2}|t| \qquad \text{for all} \quad t \in \mathbb{R}.$$

The relation between the theory of distributions and the method of averaging will be discussed in Sect. 1.7.

Iterative solution of nonlinear problems. The experience of physicists shows that

> *Interactions in nature lead to nonlinear terms in the corresponding differential equations.*

This explains the importance of nonlinear problems in physics. We want to show that the Green's function can also be used in order to investigate nonlinear problems. As a prototype, consider the differential equation

$$\boxed{m\ddot{x}(t) = -\kappa x(t)^3, \qquad t \in \mathbb{R}} \tag{1.7}$$

with the positive parameter κ called coupling constant, and the initial condition $x(0) = a, \dot{x}(0) = v$. This problem describes an anharmonic oscillator (see page 370). By (1.3), the initial-value problem (1.7) is equivalent to the nonlinear integral equation

$$\boxed{x(t) = a + vt - \kappa \int_{-\infty}^{\infty} G(t,\tau)x(\tau)^3 d\tau \qquad \text{for all} \quad t \in \mathbb{R},} \tag{1.8}$$

[9] Recall that $\theta(t) := 1$ if $t \geq 0$ and $\theta(t) := 0$ if $t < 0$.

by setting $F(t) := -\kappa x(t)^3$. The corresponding iterative method reads as

$$x_{n+1}(t) = a + vt - \kappa \int_{-\infty}^{\infty} G(t,\tau) x_n(\tau)^3 d\tau, \qquad n = -1, 0, 1, \ldots$$

with $x_{-1}(t) := 0$ for all $t \in \mathbb{R}$. This iterative method is also called the bootstrap method by physicists. In particular, $x_0(t) = a + vt$ for all $t \in \mathbb{R}$. The first approximation,

$$x_1(t) = a + vt - \kappa \int_{-\infty}^{\infty} G(t,\tau) x_0(\tau)^3 d\tau \qquad \text{for all} \quad t \in \mathbb{R}$$

is called the Born approximation by physicists. If the coupling constant κ is sufficiently small, then the iterative method converges to the solution of the original integral equation (1.8), that is, $\lim_{n\to\infty} x_n(t) = x(t)$ for all $t \in \mathbb{R}$.

> *The two problems* (1.7) *and* (1.8) *reflect a crucial duality between differential equations and integral equations. The kernel G of the integral equation* (1.8) *is the Green's function of the linearized differential equation* (1.2).

In this series of monographs, we will frequently use this duality. For example, in Sect. 8.6 we will study stationary scattering processes in quantum mechanics by replacing the Schrödinger differential equation by the dual Lippmann–Schwinger integral equation.

> *Therefore, nonlinear problems can be iteratively solved if the Green's function is known.*

This is the method of perturbation theory, which is basic for quantum field theory. For the computation of the Green's function, one can use Fourier's method. For the Newtonian motion (1.2), this will be studied in Sect. 2.2.14 in terms of the Fourier integral transform.

1.6 Harmonic Analysis and the Fourier Method

The superposition principle. In 1822 Fourier published his monograph *Théorie analytique de la chaleur* (analytic heat theory) where he used both the Fourier series and the Fourier integral in order to solve numerous problems in heat conduction. Let us sketch the basic ideas. For given time period $T > 0$, let us introduce the corresponding angular frequency

$$\Delta\omega := \frac{2\pi}{T}.$$

Fourier's method of harmonic analysis is the most important method for getting explicit solutions of linear partial differential equations in mathematical physics and for explicitly computing the corresponding Green's functions.[10]

[10] Much material can be found in P. Morse and H. Feshbach, Methods of Theoretical Physics, Vols. 1, 2, McGraw-Hill, New York, 1953. Fourier's method is intimately

(i) Discrete Fourier transformation. Let $f : \mathbb{R} \to \mathbb{C}$ be a smooth function of period $T > 0$. Then[11]

$$F(t) = \sum_{k=-\infty}^{\infty} a(k) \mathrm{e}^{\mathrm{i}tk\Delta\omega}, \qquad t \in \mathbb{R} \tag{1.9}$$

with the so-called Fourier coefficients

$$a(k) := \frac{1}{T} \int_{-T/2}^{T/2} F(t) \mathrm{e}^{-\mathrm{i}tk\Delta\omega} dt, \qquad k = 0, \pm 1, \pm 2, \ldots$$

Rigorously, the Fourier series (1.9) converges uniformly on the real line. Equation (1.9) tells us that the force function F can be represented by a superposition of special oscillating forces $t \mapsto a(k)\mathrm{e}^{\mathrm{i}tk\Delta\omega}$ of period T, angular frequency $k\Delta\omega$, and amplitude $a(k)$ with $k = 0, \pm 1, \pm 2, \ldots$ The map

$$F \mapsto \{a(k)\}_{k\in\mathbb{Z}}$$

is called the discrete Fourier transformation (with respect to the given period T).

(ii) Rescaling. Set $\hat{F}(k\Delta\omega) := Ta(-k)/\sqrt{2\pi}$ where $k = 0, \pm 1, \pm 2, \ldots$, and choose the angular frequencies

$$\omega := k\Delta\omega, \qquad k = 0, \pm 1, \pm 2, \ldots$$

Then

$$F(t) = \frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \hat{F}(k\Delta\omega) \mathrm{e}^{-\mathrm{i}tk\Delta\omega} \Delta\omega, \qquad t \in \mathbb{R}$$

with

$$\hat{F}(\omega) := \frac{1}{\sqrt{2\pi}} \int_{-T/2}^{T/2} F(t) \mathrm{e}^{\mathrm{i}t\omega} dt, \qquad k = 0, \pm 1, \pm 2, \ldots$$

(iii) Continuous Fourier transformation. Suppose that the period T goes to infinity, $T \to \infty$. Then we formally obtain the integral

$$F(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{F}(\omega) \mathrm{e}^{-\mathrm{i}t\omega} d\omega, \qquad t \in \mathbb{R} \tag{1.10}$$

related to special functions in mathematical physics based on symmetries. We refer to N. Vilenkin and A. Klimyk, Special Functions and Representations of Lie Groups, Vols. 1–4, Kluwer, Dordrecht, 1991, and to A. Wawrzyńczyk, Group Representations and Special Functions, Reidel, Dordrecht, 1984 (see also the references for further reading about special functions in quantum mechanics on page 762).

[11] Recall that, by definition, $\sum_{k=-\infty}^{\infty} b(k) := \sum_{k=0}^{\infty} b(k) + \sum_{k=-1}^{-\infty} b(k)$.

with

$$\hat{F}(\omega) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(t) \mathrm{e}^{\mathrm{i}t\omega} dt, \qquad \omega \in \mathbb{R}. \tag{1.11}$$

Rigorously, if the function $F : \mathbb{R} \to \mathbb{C}$ is smooth and rapidly decreasing at infinity, that is, $F \in \mathcal{S}(\mathbb{R})$,[12] then the function F can be represented by (1.10) where the function $\hat{F}$ is given by (1.11). Moreover, we have $\hat{F} \in \mathcal{S}(\mathbb{R})$. Equation (1.10) tells us that each function $F \in \mathcal{S}(\mathbb{R})$ is the superposition of harmonic waves $t \mapsto \hat{F}(\omega)\mathrm{e}^{-\mathrm{i}t\omega}$ of angular frequency ω, and the corresponding amplitude function $\omega \mapsto \hat{F}(\omega)$ lies in $\mathcal{S}(\mathbb{R})$. The map $F \mapsto \hat{F}$ is called the continuous Fourier transformation (or, briefly, the Fourier transformation) from the time space to the frequency space.

Terminology. Passing from frequency ω to energy $E = \hbar\omega$, we define the Fourier transformation from the time space to the energy space by setting

$$F(t) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \hat{F}_*(E) \mathrm{e}^{-\mathrm{i}Et/\hbar} dE, \qquad t \in \mathbb{R} \tag{1.12}$$

with

$$\hat{F}_*(E) := \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} F(t) \mathrm{e}^{\mathrm{i}Et/\hbar}\, dt, \qquad E \in \mathbb{R}. \tag{1.13}$$

Here, $\sqrt{\hbar}\hat{F}_*(\hbar\omega) = \hat{F}(\omega)$. This is also called the rescaled Fourier transformation. Motivated by the Fourier-Minkowski transformation in the 4-dimensional space-time (Minkowski space) in Einstein's theory of special relativity, we will distinguish between the Fourier transformation (1.12), (1.13) from the time space to the energy space and the Fourier transformation from the position space to the momentum space given by

$$F(x) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \hat{F}_{**}(p) \mathrm{e}^{\mathrm{i}xp/\hbar}\, dp, \qquad x \in \mathbb{R} \tag{1.14}$$

with

$$\hat{F}_{**}(p) := \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} F(x) \mathrm{e}^{-\mathrm{i}xp/\hbar}\, dx, \qquad p \in \mathbb{R}. \tag{1.15}$$

Note that $\sqrt{\hbar}\hat{F}_{**}(p) = \hat{F}(-p/\hbar)$ for all momenta $p \in \mathbb{R}$. This is discussed on page 538 of Vol. I. To simplify notation, we will frequently write $\hat{F}$ instead of $\hat{F}_*$ (resp. $\hat{F}_{**}$) if any misunderstanding is excluded.[13]

[12] The definition of the space $\mathcal{S}(\mathbb{R})$ can be found in Sect. 10.3.3 of Vol. I.

[13] In the literature, one also uses the asymmetrical formulas

$$F(x) = \int_{-\infty}^{\infty} \hat{F}_{\mathrm{asym}}(p) \mathrm{e}^{\mathrm{i}xp/\hbar}\, dp, \qquad x \in \mathbb{R}$$

Prototype of the Fourier method. Consider the differential equation

$$\boxed{\ddot{x}(t) = -\omega_0^2 x(t) + F(t), \qquad t \in \mathbb{R}.} \tag{1.16}$$

We are given the parameter $\omega_0 > 0$ and the periodic smooth force function $F : \mathbb{R} \to \mathbb{R}$ with the period $T > 0$. We are looking for a solution $x : \mathbb{R} \to \mathbb{R}$. In terms of physics, the function $x = x(t)$ describes the motion of a particle with mass $m = 1$ under the action of the external force $F(t)$ and the reactive force $-\omega_0^2 x(t)$ at time t. Physicists call this an harmonic oscillator. The angular frequency of the force F is given by $\Delta\omega := 2\pi/T$.

We postulate that $\omega_0 \neq k\Delta\omega$ for all integers k.

In terms of physics, this crucial condition means that the external force F is not in resonance with the eigenoscillations of the harmonic oscillator. The general solution of (1.16) reads as

$$\boxed{x(t) = \Re(a\mathrm{e}^{\mathrm{i}t\omega_0} + b\mathrm{e}^{-\mathrm{i}t\omega_0} + x_{\mathrm{special}}(t)), \qquad t \in \mathbb{R}} \tag{1.17}$$

with arbitrary complex numbers a, b. Furthermore, we have the special solution

$$x_{\mathrm{special}}(t) := \int_{-T/2}^{T/2} G(t,\tau)F(\tau)d\tau, \qquad t \in \mathbb{R},$$

and the Green's function

$$G(t,\tau) := \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \frac{\mathrm{e}^{\mathrm{i}(t-\tau)k\Delta\omega}}{\omega_0^2 - (k\Delta\omega)^2}\, \Delta\omega. \tag{1.18}$$

Let us prove this.

(I) Formal computation. In order to construct a special solution of (1.16), we start with the ansatz

$$x_{\mathrm{special}}(t) := \sum_{k=-\infty}^{\infty} b(k)\mathrm{e}^{\mathrm{i}tk\Delta\omega}.$$

Now insert this into (1.16) and use the Fourier series (1.9) for the force F. From $\ddot{x}(t) + \omega_0^2 x(t) - F(t) = 0$ we get

with

$$\hat{F}_{\mathrm{asym}}(p) = \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} F(x)\mathrm{e}^{-\mathrm{i}xp/\hbar}\, dx, \qquad p \in \mathbb{R}.$$

However, whereas the transformations $F \mapsto \hat{F}$ and $F \mapsto \hat{F}_*$ generate unitary operators on the Hilbert space $L_2(\mathbb{R})$, this is not the case for $F \mapsto \hat{F}_{\mathrm{asym}}$. Therefore, the choice of the transformation $F \mapsto \hat{F}_{\mathrm{asym}}$ has the disadvantage that it violates the fundamental unitary symmetry between position and momentum in quantum physics.

$$\sum_{k=-\infty}^{\infty} \Big(-(k\Delta\omega)^2 b(k) + \omega_0^2 b(k) - a(k)\Big) \cdot \mathrm{e}^{\mathrm{i}tk\Delta\omega} = 0.$$

Hence $-(k\Delta\omega)^2 b(k) + \omega_0^2 b(k) - a(k) = 0$. This implies

$$x_{\text{special}}(t) = \sum_{k=-\infty}^{\infty} \frac{a(k)}{\omega_0^2 - (k\Delta\omega)^2} \cdot \mathrm{e}^{\mathrm{i}tk\Delta\omega}. \tag{1.19}$$

Noting that $a(k) = \frac{\Delta\omega}{2\pi} \int_{-T/2}^{T/2} F(\tau)\mathrm{e}^{-\mathrm{i}\tau k\Delta\omega} d\tau$, we get the desired formula (1.17), by formally interchanging summation and integration.

(II) Rigorous proof. Let $N = 1, 2, \ldots$ Since the function F is smooth, we have

$$a(k) = O\left(\frac{1}{k^N}\right) \qquad \text{for all} \quad k \in \mathbb{Z}.$$

The same is true for $b(k)$. By the majorant criterion, all the Fourier series involving $a(k)$ and $b(k)$ converge uniformly on the real line, and hence term-by-term differentiation (resp. integration) is allowed. This shows that the function x_{special} given by (1.19) is indeed a special solution of the inhomogeneous differential equation (1.16). Finally, note that the general solution of the homogeneous equation (1.16) with $F = 0$ is given by $x(t) = a\mathrm{e}^{\mathrm{i}t\omega_0} + b\mathrm{e}^{-\mathrm{i}t\omega_0}$ with arbitrary complex numbers a and b. This finishes the proof. □

Resonances and the singularities of the Green's function. Suppose that $\omega_0 = k_0\Delta\omega$ for some nonzero integer k_0. Then it follows from (1.18) that the Green's function G has a singularity if we choose $k = k_0$. In the case where the function F satisfies the condition $a(k_0) \neq 0$, that is,

$$\int_{-T/2}^{T/2} F(\tau)\mathrm{e}^{-\mathrm{i}\tau\omega_0} d\tau \neq 0,$$

physicists say that the external force F is in resonance with the eigenfrequency ω_0 of the harmonic oscillator.

Resonance effects cause singularities of the Green's function.

In the present case, the difficulty disappears if we demand that $a(k_0) = 0$. Then the singularity drops out in (1.19).

Resonances are responsible for complicated physical effects.

For example, the observed chaotic motion of some asteroids is due to resonance effects in celestial mechanics (the Kolmogorov–Arnold–Moser theory). In quantum field theory, internal resonances of the quantum field cause special quantum effects (e.g., the Lamb shift in the spectrum of the hydrogen atom and the anomalous magnetic moment of the electron), which have to be treated with the methods of renormalization theory (see Chap. 17 on radiative corrections in quantum electrodynamics).

1.7 The Method of Averaging and the Theory of Distributions

In the early 20th century, mathematicians and physicists noticed that for wave problems, the Green's functions possess strong singularities such that the solution formulas of the type (1.3) fail to exist as classical integrals.[14] In his classic monograph

The Principles of Quantum Mechanics,

Clarendon Press, Oxford, 1930, Dirac introduced a singular object $\delta(t)$ (the Dirac delta function), which is very useful for the description of quantum processes and the computation of Green's functions. In the 1940s, Laurent Schwartz gave all these approaches a sound basis by introducing the notion of distribution (generalized function). In order to explain Laurent Schwartz's basic idea of averaging, consider the continuous motion

$$x(t) := |t| \qquad \text{for all} \quad t \in \mathbb{R}$$

of a particle on the real line. We want to compute the force $F(t) = m\ddot{x}(t)$ acting on the particle at time t. Classically, $F(t) = 0$ if $t \neq 0$, and the force does not exist at the point in time $t = 0$. We want to motivate that

$$\boxed{F(t) = 2m\delta(t) \qquad \text{for all} \quad t \in \mathbb{R}.} \tag{1.20}$$

(I) The language of Dirac. For the velocity, $\dot{x}(t) = 1$ if $t > 0$, and $\dot{x}(t) = -1$ if $t < 0$. For $t = 0$, the derivative $\dot{x}(0)$ does not exist. We define $\dot{x}(0) := 0$. Hence

$$\dot{x}(t) = \theta(t) - \theta(-t).$$

Since $\dot{\theta}(t) = \delta(t)$, we get

$$\ddot{x}(t) = \delta(t) + \delta(-t) = 2\delta(t) \qquad \text{for all} \quad t \in \mathbb{R}.$$

Formally, $\delta(t) = 0$ if $t \neq 0$, and $\delta(0) = \infty$ with $\int_{-\infty}^{\infty} \delta(t)dt = 1$. Obviously, there is *no classical function* δ which has such properties.[15]

(II) The language of Laurent Schwartz. Choose $\varepsilon > 0$. We first pass to the regularized motion $x = x_\varepsilon(t)$ for all $t \in \mathbb{R}$. That is, the function $x_\varepsilon : \mathbb{R} \to \mathbb{R}$ is smooth for all $\varepsilon > 0$ and

[14] For example, see J. Hadamard, The Initial-Value Problem for Linear Hyperbolic Partial Differential Equations, Hermann, Paris (in French). A modern version of Hadamard's theory can be found in P. Günther, Huygens' Principle and Hyperbolic Differential Equations, Academic Press, San Diego, 1988. See also C. Bär, N. Ginoux, and F. Pfäffle, Wave Equations on Lorentzian Manifolds and Quantization, European Mathematical Society 2007.

[15] See the detailed discussion of the formal Dirac calculus in Sect. 11.2 of Vol. I.

$$\lim_{\varepsilon\to+0} x_\varepsilon(t) = |t| \qquad \text{for all} \quad t \in \mathbb{R},$$

where this convergence is uniform on all compact time intervals.[16] We introduce the averaged force

$$\mathsf{F}_\varepsilon(\varphi) := \int_{-\infty}^{\infty} m\ddot{x}_\varepsilon(t)\varphi(t)dt$$

for all averaging functions $\varphi \in \mathcal{D}(\mathbb{R})$ (i.e., $\varphi : \mathbb{R} \to \mathbb{C}$ is smooth and vanishes outside some bounded interval. In other words, φ has compact support.) Since x_ε is smooth, integration by parts twice yields

$$\mathsf{F}_\varepsilon(\varphi) = \int_{-\infty}^{\infty} mx_\varepsilon(t)\ddot{\varphi}(t)dt.$$

Letting $\varepsilon \to +0$, we define the mean force by

$$\mathsf{F}(\varphi) := \lim_{\varepsilon\to+0} \mathsf{F}_\varepsilon(\varphi) = \int_{-\infty}^{\infty} mx(t)\ddot{\varphi}(t)dt.$$

Integration by parts yields $\int_0^\infty |t|\ \ddot{\varphi}(t)dt = -\int_0^\infty \dot{\varphi}(t)dt = \varphi(0)$. Similarly, $\int_{-\infty}^0 |t|\ddot{\varphi}(t)dt = -\int_{-\infty}^0 \dot{\varphi}(t)dt = \varphi(0)$. Summarizing, we obtain the averaged force

$$\boxed{\mathsf{F}(\varphi) = 2m\varphi(0) \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}).} \tag{1.21}$$

In the language of distributions, we have $\mathsf{F} = 2m\delta$, where δ denotes the Dirac delta distribution. A detailed study of the theory of distributions and its applications to physics can be found in Chaps. 11 and 12 of Vol. I. In particular, equation (1.20) is equivalent to (1.21), in the sense of distribution theory.

In terms of experimental physics, distributions correspond to the fact that measurement devices only measure averaged values. It turns out that classical functions can also be regarded as distributions. However, in contrast to classical functions, the following is true:

Distributions possess derivatives of all orders.

Therefore, the theory of distributions is the quite natural completion of the infinitesimal strategy due to Newton and Leibniz, who lived almost three hundred years before Laurent Schwartz. This shows convincingly that the development of mathematics needs time.

[16] For example, choose $x_\varepsilon(t) := r_\varepsilon(t)|t|$ for all $t \in \mathbb{R}$, where the regularizing function $r_\varepsilon : \mathbb{R} \to [0,1]$ is smooth, and $r_\varepsilon(t) := 1$ if $t \notin [-2\varepsilon, 2\varepsilon]$, as well as $r_\varepsilon(t) := 0$ if $t \in [-\varepsilon, \varepsilon]$.

1.8 The Symbolic Method

The symbol of the operator $\frac{d}{dt}$. Consider again the Fourier transformation

$$F(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{F}(\omega) \mathrm{e}^{-\mathrm{i}t\omega} d\omega, \qquad t \in \mathbb{R} \tag{1.22}$$

with

$$\hat{F}(\omega) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(t) \mathrm{e}^{\mathrm{i}t\omega t} dt, \qquad \omega \in \mathbb{R}. \tag{1.23}$$

Let $n = 1, 2, \ldots$ Differentiation of (1.22) yields

$$\frac{d^n}{dt^n} F(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (-\mathrm{i}\omega)^n \hat{F}(\omega) \mathrm{e}^{\mathrm{i}t\omega} d\omega, \qquad t \in \mathbb{R}. \tag{1.24}$$

The function

$$s(\omega) := -\mathrm{i}\omega \qquad \text{for all} \quad \omega \in \mathbb{R}$$

is called the *symbol* of the differential operator $\frac{d}{dt}$. For $n = 0, 1, 2, \ldots$, we have

$$\frac{d^n}{dt^n} F \Rightarrow s^n \hat{F}.$$

This means that the action of the differential operator $\frac{d^n}{dt^n}$, with respect to time t, can be described by the multiplication of the Fourier transform $\hat{F}$ by s^n in the frequency space.

This corresponds to a convenient algebraization of derivatives.

Over the centuries, mathematicians and physicists tried to simplify computations. The relation

$$\ln(ab) = \ln a + \ln b \qquad \text{for all} \quad a, b > 0 \tag{1.25}$$

allows us to reduce multiplication to addition. This fact was extensively used by Kepler (1571–1630) in order to simplify his enormous computations in celestial mechanics.

Similarly, the Fourier transformation allows us to reduce differentiation to multiplication.

Furthermore, there exists a natural generalization of the logarithmic function to Lie groups. Then the crucial formula (1.25) passes over to the transformation formula from the Lie group $\mathcal{G}$ to its Lie algebra $\mathcal{LG}$. This transformation is well defined for the group elements near the unit element (see Vol. III).

Pseudo-differential operators and Fourier integral operators. The modern theory of pseudo-differential operators (e.g., differential and integral

operators) and Fourier integral operators is based on the use of symbols of the form

$$s = s(\omega, t, \tau),$$

which depend on frequency ω, time t, and time τ. The expressions

$$(AF)(t) := \frac{1}{2\pi} \int_{\mathbb{R}^2} s(\omega, t, \tau) F(\tau) \mathrm{e}^{\mathrm{i}\omega(\tau - t)} \, d\tau d\omega, \quad t \in \mathbb{R} \tag{1.26}$$

and

$$(BF)(t) := \frac{1}{2\pi} \int_{\mathbb{R}^2} s(\omega, t, \tau) F(\tau) \mathrm{e}^{\mathrm{i}\varphi(\omega, t, \tau)} \, d\tau d\omega, \quad t \in \mathbb{R} \tag{1.27}$$

correspond to the pseudo-differential operator A and the Fourier integral operator B. If we choose the special phase function

$$\varphi(\omega, t, \tau) := \omega(t - \tau),$$

then the operator B passes over to A. If, in addition, the symbol s does not depend on τ, then integration over τ yields

$$(AF)(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} s(\omega, t) \hat{F}(\omega) \mathrm{e}^{-\mathrm{i}\omega t} \, d\omega.$$

In the special case where the symbol $s(\omega, t)$ only depends on the frequency ω, the pseudo-differential operator corresponds to the multiplication operator $\omega \mapsto s(\omega)\hat{F}(\omega)$ in the frequency space (also called Fourier space).

Long before the foundation of the theory of pseudo-differential operators and Fourier integral operators in the 1960s and 1970s, mathematicians and physicists used integral expressions of the form (1.26) and (1.27) in order to compute explicit solutions in electrodynamics (e.g., the Heaviside calculus and the Laplace transform applied to the study of electric circuits[17]), elasticity (singular integral equations), geometric optics (e.g., diffraction of light), and quantum mechanics.

> *The point is that the symbols know a lot about the properties of the corresponding operators, and an elegant algebraic calculus for operators can be based on algebraic operations for the symbols.*

As an introduction, we recommend:

> Yu. Egorov and M. Shubin, Foundations of the Classical Theory of Partial Differential Equations, Springer, New York, 1998 (Encyclopedia of Mathematical Sciences).
>
> Yu. Egorov, A. Komech, and M. Shubin, Elements of the Modern Theory of Partial Differential Equations, Springer, New York, 1999 (Encyclopedia of Mathematical Sciences).

[17] See E. Zeidler (Ed.), Oxford Users' Guide to Mathematics, Sect. 1.11, Oxford University Press, 2004.

F. Berezin and M. Shubin, The Schrödinger Equation, Kluwer, Dordrecht, 1991.

L. Faddeev and A. Slavnov, Gauge Fields, Benjamin, Reading, Massachusetts, 1980 (gauge theory, Weyl calculus, the Feynman path integral, and the Faddeev–Popov ghost approach to the Standard Model in particle physics).

We also refer to the following treatises:

L. Hörmander, The Analysis of Linear Partial Differential Operators. Vol. 1: Distribution Theory and Fourier Analysis, Vol. 2: Differential Operators with Constant Coefficients, Vol. 3: Pseudo-Differential Operators, Vol. 4: Fourier Integral Operators, Springer, New York, 1993.

R. Dautray and J. Lions, Mathematical Analysis and Numerical Methods for Science and Technology, Vols. 1–6, Springer, New York, 1988.

Heaviside's formal approach. Consider the differential equation

$$\frac{d}{dt}x(t) - x(t) = f(t). \tag{1.28}$$

We want to discuss the beauty, but also the shortcomings of the symbolic method due to Heaviside (1850–1925). Formally, we get

$$\left(\frac{d}{dt} - 1\right) x(t) = f(t).$$

Hence

$$x(t) = \frac{f(t)}{\frac{d}{dt} - 1}.$$

For complex numbers z with $|z| < 1$, we have the convergent geometric series $\frac{1}{z-1} = -1 - z - z^2 - z^3 + \ldots$ This motivates

$$x(t) = \left(-1 - \frac{d}{dt} - \frac{d^2}{dt^2} - \ldots\right) f(t). \tag{1.29}$$

If we choose $f(t) := t^2$, then

$$x(t) = -t^2 - 2t - 2. \tag{1.30}$$

Surprisingly enough, we get $\dot{x}(t) = -2t-2 = x(t)+t^2$. Therefore, the function $x(t)$ from (1.30) is a solution of (1.28). The same is true for all polynomials. To prove this, let f be a polynomial of degree $n = 0, 1, 2 \ldots$ Set

$$x(t) := -\sum_{k=0}^{n} \frac{d^k}{dt^k} f(t).$$

Then we get $\dot{x}(t) = -\sum_{k=0}^{n} \frac{d^{k+1}}{dt^{k+1}} f(t) = f(t) + x(t)$, since the $(n+1)$th derivative of f vanishes. However, the method above fails if we apply it to the exponential function $f(t) := \mathrm{e}^t$. Then

$$x(t) = \mathrm{e}^t + \mathrm{e}^t + \mathrm{e}^t + \dots,$$

which is meaningless. There arises the problem of establishing a more powerful method. In the history of mathematics and physics, formal (also called symbolic) methods were rigorously justified by using the following tools:

- the Fourier transformation,
- the Laplace transformation (which can be reduced to the Fourier transformation),
- Mikusiński's operational calculus based on the quotient field over a convolution algebra,
- von Neumann's operator calculus in Hilbert space,
- the theory of distributions,
- pseudo-differential operators and distributions (e.g., the Weyl calculus in quantum mechanics), and
- Fourier integral operators and distributions.

Mikusiński's elegant approach will be considered in Sect. 4.2 on page 191.

Motivation of the Laplace transformation via Fourier transformation. Consider the motion $x = x(t)$ of a particle on the real line with $x(t) = 0$ for all $t \le 0$. Suppose that the function $x : \mathbb{R} \to \mathbb{R}$ is continuous and bounded. The Fourier transform from the time space to the energy space reads as

$$\hat{x}(E) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} x(t)\mathrm{e}^{\mathrm{i}Et/\hbar}\, dt = \frac{1}{\sqrt{2\pi\hbar}} \int_{0}^{\infty} x(t)\mathrm{e}^{\mathrm{i}Et/\hbar}\, dt.$$

As a rule, this integral does not exist. To improve the situation, we fix the regularization parameter $\varepsilon > 0$, and we define the damped motion

$$x_\varepsilon(t) := x(t)\mathrm{e}^{-\varepsilon t} \qquad \text{for all} \quad t \in \mathbb{R}.$$

This is also called the adiabatic regularization of the original motion. Obviously, $\lim_{\varepsilon \to +0} x_\varepsilon(t) = x(t)$ for all $t \in \mathbb{R}$. The Fourier transform looks like

$$\hat{x}_\varepsilon(E) = \frac{1}{\sqrt{2\pi\hbar}} \int_{0}^{\infty} x(t)\mathrm{e}^{-\varepsilon t}\mathrm{e}^{\mathrm{i}Et/\hbar}\, dt = \frac{1}{\sqrt{2\pi\hbar}} \int_{0}^{\infty} x(t)\mathrm{e}^{\mathrm{i}\mathcal{E}t/\hbar}\, dt,$$

by introducing the complex energy $\mathcal{E} := E + \mathrm{i}\varepsilon$. To simplify notation, we set $\hbar := 1$.

Complex energies, damped oscillations, and the Laplace transform. The formal Heaviside calculus was justified by Doetsch in the 1930s by using the Laplace transform.[18] As a simple example, let us use the Laplace transformation in order to solve the differential equation (1.28). In particular, we will consider the case

[18] G. Doetsch, Theory and Applications of the Laplace Transform, Springer, Berlin, 1937 (in German). See also D. Widder, The Laplace Transform, Princeton University Press, 1944.

$$f(t) := \mathrm{e}^t$$

where the Heaviside method above fails. Let $x : [0, \infty[\to \mathbb{R}$ be a smooth function with the growth condition

$$|x(t)| \leq \text{const} \cdot \mathrm{e}^{\gamma_1 t} \qquad \text{for all} \quad t \geq 0$$

and fixed real number γ_1. The Laplace transform reads as

$$\boxed{\mathcal{L}(x)(\mathcal{E}) := \int_0^\infty x(t)\, \mathrm{e}^{\mathrm{i}\mathcal{E}t}\, dt, \qquad \Im(\mathcal{E}) > \gamma_1} \tag{1.31}$$

with the inverse transform

$$x(t) = \frac{1}{2\pi} PV \int_L (\mathcal{L}x)(\mathcal{E}) \mathrm{e}^{-\mathrm{i}\mathcal{E}t} d\mathcal{E}, \qquad t > 0 \tag{1.32}$$

on the real line $L := \{E + (\gamma_1 + 1)\mathrm{i} : \ E \in \mathbb{R}\}$ of the complex energy space. Here, we choose a system of units with $\hbar = h/2\pi := 1$ for Planck's action quantum.[19] The Laplace transform sends the function $t \mapsto x(t)$ on the time space to the function $\mathcal{E} \mapsto (\mathcal{L}x)(\mathcal{E})$ on the complex energy space. Here, it is crucial to use complex energies $\mathcal{E} = E - \Gamma\mathrm{i}$. In what follows, we will use the standard properties of the Laplace transformation which are proved in Sect. 2.2.6 of Vol. I. Let us start with an example. Choose the complex energy $\mathcal{E}_0 := E_0 - \Gamma_0\mathrm{i}$ with real values E_0 and Γ_0, and set[20]

$$\boxed{x(t) := \mathrm{e}^{-\mathrm{i}\mathcal{E}_0 t} = \mathrm{e}^{-\mathrm{i}E_0 t} \cdot \mathrm{e}^{-\Gamma_0 t}, \qquad t \in \mathbb{R}.} \tag{1.33}$$

Then, $\gamma_1 = -\Gamma_0 = \Im(\mathcal{E}_0)$, and we get

$$(\mathcal{L}x)(\mathcal{E}) = \frac{\mathrm{i}}{\mathcal{E} - \mathcal{E}_0}, \qquad \Im(\mathcal{E}) > \Im(\mathcal{E}_0).$$

Now to the point. We are given the smooth function $f : [0, \infty[\to \mathbb{C}$ with the growth condition

$$|f(t)| \leq \text{const} \cdot \mathrm{e}^{\gamma_0 t} \qquad \text{for all} \quad t \geq 0.$$

In order to solve the differential equation (1.28), we proceed as follows.

(I) Suppose first that the differential equation (1.28) has a smooth solution $x : [0, \infty[\to \mathbb{C}$ with $|x(t)| \leq \text{const} \cdot \mathrm{e}^{\gamma_1 t}$ for all $t \geq 0$ with $\gamma_1 \geq \gamma_0$. Then

[19] Set $\gamma_2 := \gamma_1 + 1$. The principal value of the integral is defined by

$$PV \int_L g(\mathcal{E}) d\mathcal{E} := \lim_{E_0 \to +\infty} \int_{-E^0 + \gamma_2 \mathrm{i}}^{E_0 + \gamma_2 \mathrm{i}} g(E + \gamma_2 \mathrm{i}) dE.$$

[20] If $E_0 > 0$ and $\Gamma_0 > 0$ then (1.33) is a damped oscillation with angular frequency $\omega_0 := E_0/\hbar = E_0$ and mean lifetime $\Delta t = \Gamma_0/\hbar = \Gamma_0$.

the Laplace transforms $\mathcal{L}x$ and $\mathcal{L}f$ exist for all $\mathcal{E} \in \mathbb{C}$ with $\Im(\mathcal{E}) > \gamma_1$. Furthermore,

$$(\mathcal{L}\dot{x})(\mathcal{E}) = -\mathrm{i}\mathcal{E}(\mathcal{L})(\mathcal{E}) - x(+0),$$

that is, the Laplace transforms converts differentiation into multiplication and translation in the complex energy space. By (1.28),

$$-\mathrm{i}\mathcal{E}(\mathcal{L}x)(\mathcal{E}) - (\mathcal{L}x)(\mathcal{E}) - x(+0) = (\mathcal{L}f)(\mathcal{E}), \qquad \Im(\mathcal{E}) > \gamma_1.$$

This yields the Laplace transform of the solution $t \to x(t)$, namely,

$$(\mathcal{L}x)(\mathcal{E}) = \frac{\mathrm{i}x(+0)}{\mathcal{E} - \mathrm{i}} + \frac{\mathrm{i}(\mathcal{L}f)(\mathcal{E})}{\mathcal{E} - \mathrm{i}}.$$

Setting $g(t) := \mathrm{e}^t$, we get $(\mathcal{L}g)(\mathcal{E}) = \frac{\mathrm{i}}{\mathcal{E}-\mathrm{i}}$. Therefore,

$$\mathcal{L}x = (\mathcal{L}g)x(+0) + (\mathcal{L}g)(\mathcal{L}f).$$

The convolution rule from Sect. 2.2.6 of Vol. I tells us that

$$x = gx(+0) + g * f.$$

Explicitly, this reads as

$$\boxed{x(t) = \mathrm{e}^t x(+0) + \int_0^t \mathrm{e}^{(t-\tau)} f(\tau) d\tau.} \tag{1.34}$$

Our argument shows that a solution of (1.28) has necessarily the form (1.34).

(II) Conversely, differentiation yields

$$\dot{x}(t) = \mathrm{e}^t x(+0) + f(t) + \int_0^t \mathrm{e}^{(t-\tau)} f(\tau) d\tau = x(t) + f(t)$$

for all $t \geq 0$. Consequently, the function $x = x(t)$ given by (1.34) is indeed a solution of the original differential equation (1.28) for all times $t \geq 0$. For example, if $f(t) := \mathrm{e}^t$, then

$$x(t) = \mathrm{e}^t x(+0) + t\mathrm{e}^t.$$

This is a solution of (1.28) for all times $t \in \mathbb{R}$.

The same method of the Laplace transformation can be applied to general systems of ordinary differential equations with constant coefficients. Such equations are basic for the investigation of electrical circuits. Therefore, the Laplace transformation plays a key role in electrical engineering.

1.9 Gauge Theory – Local Symmetry and the Description of Interactions by Gauge Fields

As we have discussed in Chap. 2 of Vol. I, the Standard Model in particle physics is based on

- 12 basic particles (6 quarks and 6 leptons), and
- 12 interacting particles (the photon, the 3 vector bosons W^+, W^-, Z^0 and 8 gluons).

This model was formulated in the 1960s and early 1970s. Note the following crucial fact about the structure of the fundamental interactions in nature.

> *The fields of the interacting particles can be obtained from the fields of the basic particles by using the principle of local symmetry (also called the gauge principle).*

Prototype of a gauge theory. Let us explain the basic ideas by considering the following simple model. To this end, let us choose the unit square $Q := \{(x,t) : 0 \le x, t \le 1\}$. We start with the principle of critical action

$$\boxed{\int_Q L(\psi, \psi_t, \psi_x; \psi^\dagger, \psi_t^\dagger, \psi_x^\dagger)\, dxdt = \text{critical!}} \tag{1.35}$$

with the boundary condition $\psi = \psi_0$ on ∂Q and the special Lagrangian

$$L := \psi^\dagger \psi_t + \psi^\dagger \psi_x. \tag{1.36}$$

Here, ψ_t (resp. ψ_x) denotes the partial derivative of ψ with respect to time t (resp. position x). We are given a fixed continuous function $\psi_0 : \partial Q \to \mathbb{C}$ on the boundary of the square Q. We are looking for a smooth function $\psi : Q \to \mathbb{C}$ which solves the variational problem (1.35).

By a basic result from the calculus of variations, we get the following. If the function ψ is a solution of (1.35), then it is a solution of the two Euler–Lagrange equations

$$\frac{\partial}{\partial t} L_{\psi_t^\dagger} + \frac{\partial}{\partial x} L_{\psi_x^\dagger} = L_{\psi^\dagger} \tag{1.37}$$

and

$$\frac{\partial}{\partial t} L_{\psi_t} + \frac{\partial}{\partial x} L_{\psi_x} = L_{\psi}. \tag{1.38}$$

Here, the symbol L_ψ (resp. $L_{\psi^\dagger}$) denotes the partial derivative of L with respect to the variable ψ (resp. $\psi^\dagger$). The proof can be found in Problem 14.7 of Vol. I. Explicitly, the two Euler–Lagrange equations read as

$$\psi_t + \psi_x = 0, \qquad \psi_t^\dagger + \psi_x^\dagger = 0. \tag{1.39}$$

If the function ψ is a solution of (1.39), then we have

$$(\psi\psi^\dagger)_t + (\psi\psi^\dagger)_x = 0, \tag{1.40}$$

which is called a conservation law. In fact, $\psi_t\psi^\dagger + \psi_t^\dagger\psi = -\psi_x\psi^\dagger - \psi_x^\dagger\psi$. This is equal to $-(\psi\psi^\dagger)_x$. Conservation laws play a fundamental role in all fields of physics, since they simplify the computation of solutions. In the 18th and 19th century, astronomers unsuccessfully tried to find $6N$ conservation laws for the motion of N bodies in celestial mechanics ($N \geq 3$), in order to compute the solution and to prove the stability of our solar system.[21]

Step by step, mathematicians and physicists discovered that

Conservation laws are intimately related to symmetries.

The precise formulation of this principle is the content of the Noether theorem proved in 1918 (see Sect. 6.6). We want to show that the invariance of the Lagrangian L (with respect to a global gauge transformation) is behind the conservation law (1.40).

(i) Global symmetry and the Noether theorem. Let α be a fixed real number. We consider the global symmetry transformation

$$\psi_+(x,t) := \mathrm{e}^{\mathrm{i}\alpha}\psi(x,t) \qquad \text{for all} \quad x,t \in \mathbb{R}, \tag{1.41}$$

that is, the field ψ is multiplied by the constant phase factor $\mathrm{e}^{\mathrm{i}\alpha}$, where α is called the phase. The transformation (1.41) is also called a global gauge transformation, by physicists. We also define the infinitesimal gauge transformation $\delta\psi$ by setting

$$\delta\psi(x,t) := \frac{d}{d\alpha}\left(\mathrm{e}^{\mathrm{i}\alpha}\psi(x,t)\right)_{|\alpha=0} = \mathrm{i}\psi(x,t).$$

This means that $\psi_+(x,t) = 1 + \alpha\cdot\delta\psi(x,t) + O(\alpha^2)$ as $\alpha \to 0$. Noting that $\psi_+^\dagger = \mathrm{e}^{-\mathrm{i}\alpha}\psi^\dagger$, the special Lagrangian L from (1.36) is invariant under the global gauge transformation (1.41), that is,

$$\psi_+^\dagger(\psi_+)_t + \psi_+^\dagger(\psi_+)_x = \psi^\dagger\psi_t + \psi^\dagger\psi_x.$$

Generally, the Lagrangian L is invariant under the global gauge transformation (1.41) iff

$$L(\psi_+, (\psi_+)_t, (\psi_+)_x; \psi_+^\dagger, (\psi_+^\dagger)_t, (\psi_+^\dagger)_x) = L(\psi, \psi_t, \psi_x; \psi^\dagger, \psi_t^\dagger, \psi_x^\dagger).$$

[21] See D. Boccaletti and G. Pucacco, Theory of Orbits, Vol 1: Integrable Systems and Non-Perturbative Methods, Vol. 2: Perturbative and Geometrical Methods, Springer, Berlin, 1996.
Y. Hagihara, Celestial Mechanics, Vols. 1–5, MIT Press, Cambridge, Massachusetts, 1976.
W. Neutsch and K. Scherer, Celestial Mechanics: An Introduction to Classical and Contemporary Methods, Wissenschaftsverlag, Mannheim, 1992.

Then a special case of the famous Noether theorem on page 387 tells us the following: If the function ψ is a solution of the variational problem (1.35), then

$$\frac{\partial}{\partial t}\left(L_{\psi_t}\delta\psi + L_{\psi_t^\dagger}\delta\psi^\dagger\right) + \frac{\partial}{\partial x}\left(L_{\psi_x}\delta\psi + L_{\psi_x^\dagger}\delta\psi^\dagger\right) = 0.$$

If we choose the special Lagrangian $L = \psi^\dagger\psi_t + \psi^\dagger\psi_x$, then we obtain the conservation law (1.40).

(ii) Local symmetry and the covariant derivative. We now replace the global gauge transformation (1.40) by the following local gauge transformation

$$\boxed{\psi_+(x,t) := \mathrm{e}^{\mathrm{i}\alpha(x,t)}\psi(x,t) \qquad \text{for all} \quad x,t \in \mathbb{R},} \tag{1.42}$$

where the phase α depends on space and time. We postulate the following crucial local symmetry principle:

(P) *The Lagrangian L is invariant under local gauge transformations.*

It can be easily shown that the function L from (1.36) does *not* possess this invariance property for arbitrary functions $\alpha = \alpha(x,t)$. This follows from

$$(\psi_+)_t = \mathrm{i}\alpha_t\mathrm{e}^{\mathrm{i}\alpha}\psi + \mathrm{e}^{\mathrm{i}\alpha}\psi_t.$$

Here, the appearance of the derivative α_t of the phase function α destroys the invariance property of L.

Our *goal* is to modify the function L in such a way that it is invariant under (1.42). To this end, we introduce the so-called covariant partial derivatives

$$\boxed{\nabla_t := \frac{\partial}{\partial t} + \mathrm{i}U(x,t), \qquad \nabla_x := \frac{\partial}{\partial x} + \mathrm{i}A(x,t),} \tag{1.43}$$

where $U, A : \mathbb{R}^2 \to \mathbb{R}$ are given smooth real-valued functions called gauge fields. The local gauge transformation of U and A is defined by

$$U_+ := U - \alpha_t, \qquad A_+ := A - \alpha_x.$$

Furthermore, we define the following transformation law for the covariant partial derivatives:

$$\nabla_t^+ := \frac{\partial}{\partial t} + \mathrm{i}U_+, \qquad \nabla_x^+ := \frac{\partial}{\partial x} + \mathrm{i}A_+. \tag{1.44}$$

The key relation is given by the following elegant transformation law for the covariant partial derivatives:

$$\boxed{\nabla_t^+\psi_+ = \mathrm{e}^{\mathrm{i}\alpha}\nabla_t\psi, \qquad \nabla_x^+\psi_+ = \mathrm{e}^{\mathrm{i}\alpha}\nabla_x\psi.} \tag{1.45}$$

Theorem 1.1 *There holds* (1.45).

This theorem tells us the crucial fact that, in contrast to the classical partial derivatives, the covariant partial derivatives are transformed in the same way as the field ψ itself. This property is typical for covariant partial derivatives in mathematics. Indeed, our construction of covariant partial derivatives has been chosen in such a way that (1.45) is valid.
Proof. By the product rule,

$$\left(\frac{\partial}{\partial t} + \mathrm{i}U_+\right)\psi_+ = \mathrm{e}^{\mathrm{i}\alpha}(\mathrm{i}\alpha_t\psi + \psi_t + \mathrm{i}U_+\psi) = \mathrm{e}^{\mathrm{i}\alpha}\left(\frac{\partial}{\partial t} + \mathrm{i}U\right)\psi.$$

This yields $\nabla_t^+\psi_+ = \mathrm{e}^{\mathrm{i}\alpha}\nabla_t\psi$. Similarly, we get $\nabla_x^+\psi_+ = \mathrm{e}^{\mathrm{i}\alpha}\nabla_x\psi$. □
Now let us discuss the main idea of gauge theory:

We replace the classical partial derivatives $\frac{\partial}{\partial t}, \frac{\partial}{\partial x}$ *by the covariant partial derivatives* ∇_t, ∇_x*, respectively.*

This is the main trick of gauge theory. In particular, we replace the Lagrangian

$$L = \psi^\dagger\frac{\partial}{\partial t}\psi + \psi^\dagger\frac{\partial}{\partial x}\psi$$

from the original variational problem (1.35) by the modified Lagrangian

$$\boxed{L := \psi^\dagger\nabla_t\psi + \psi^\dagger\nabla_x\psi.}$$

Explicitly, we have

$$L = \psi^\dagger\psi_t + \psi^\dagger\psi_x + \mathrm{i}\psi^\dagger U\psi + \mathrm{i}\psi^\dagger A\psi.$$

The corresponding Euler–Lagrange equations (1.37) and (1.38) read as

$$\nabla_t\psi + \nabla_x\psi = 0, \tag{1.46}$$

and $(\nabla_t\psi + \nabla_x\psi)^\dagger = 0$, respectively.

The local symmetry principle (P) above is closely related to the Faraday–Green locality principle, saying that physical interactions are localized in space and time.

Summarizing, the local symmetry principle (P) *enforces the existence of additional gauge fields* U, A *which interact with the originally given field* ψ*.*

In the Standard Model in particle physics and in the theory of general relativity, the additional gauge fields are responsible for the interacting particles.

Consequently, the mathematical structure of the fundamental interactions in nature is a consequence of the local symmetry principle.

In his search of a unified theory for all interactions in nature, Einstein was not successful, since he was not aware of the importance of the principle of local symmetry. In our discussion below, the following notions will be crucial:

- local gauge transformation,
- gauge force F,
- connection form $\mathcal{A}$,
- curvature form $\mathcal{F}$ (gauge force form), and
- parallel transport of information.

Gauge force. Covariant partial derivatives can be used in order to introduce the following notions:

(a) Gauge force (also called curvature): We define

$$\boxed{\mathrm{i}F := \nabla_x \nabla_t - \nabla_t \nabla_x.} \tag{1.47}$$

In physics, the function F is called the **force** induced by the gauge fields U, A. Explicitly, we get

$$F = U_x - A_t. \tag{1.48}$$

Relation (1.47) tells us that:

The "gauge force" F measures the non-commutativity of the covariant partial derivatives.

In particular, the force F vanishes if the gauge fields U, A vanish. The proof of (1.48) follows from

$$\begin{aligned}\nabla_t(\nabla_x\psi) &= \left(\frac{\partial}{\partial t} + \mathrm{i}U\right)(\psi_x + \mathrm{i}A\psi)\\ &= \psi_{tx} + \mathrm{i}A_t\psi + \mathrm{i}A\psi_t + \mathrm{i}U\psi_x - UA\psi\end{aligned}$$

and

$$\begin{aligned}\nabla_x(\nabla_t\psi) &= \left(\frac{\partial}{\partial x} + \mathrm{i}A\right)(\psi_t + \mathrm{i}U\psi)\\ &= \psi_{xt} + \mathrm{i}U_x\psi + \mathrm{i}U\psi_x + \mathrm{i}A\psi_t - AU\psi.\end{aligned}$$

Hence $(\nabla_x\nabla_t - \nabla_t\nabla_x)\psi = \mathrm{i}(U_x - A_t)\psi$. □

The transformation of the force F with respect to the gauge transformation $\psi_+(x,t) = \mathrm{e}^{\mathrm{i}\alpha(x,y)}\psi(x,t)$ is defined by

$$\mathrm{i}F^+ := \nabla_x^+\nabla_t^+ - \nabla_t^+\nabla_x^+.$$

Theorem 1.2 $F^+ = e^{\mathrm{i}\alpha}Fe^{-\mathrm{i}\alpha}$.

Proof. It follows from Theorem 1.1 on page 37 that

$$\begin{aligned} \mathrm{i}F^+\psi_+ &= (\nabla_x^+\nabla_t^+ - \nabla_t^+\nabla_x^+)\psi_+ = \nabla_x^+(\mathrm{e}^{\mathrm{i}\alpha}\nabla_t\psi) - \nabla_t^+(\mathrm{e}^{\mathrm{i}\alpha}\nabla_x\psi) \\ &= \mathrm{e}^{\mathrm{i}\alpha}(\nabla_x\nabla_t\psi - \nabla_t\nabla_x\psi) = \mathrm{e}^{\mathrm{i}\alpha}\mathrm{i}F\psi = (\mathrm{e}^{\mathrm{i}\alpha}\mathrm{i}F\mathrm{e}^{-\mathrm{i}\alpha})\psi_+. \end{aligned}$$

□

In the present case, we have the commutativity property $F\mathrm{e}^{-\mathrm{i}\alpha} = \mathrm{e}^{-\mathrm{i}\alpha}F$. Hence

$$F^+ = \mathrm{e}^{\mathrm{i}\alpha}\mathrm{e}^{-\mathrm{i}\alpha}F = F,$$

that is, the force F is gauge invariant. In more general gauge theories, the phase factor $\mathrm{e}^{\mathrm{i}\alpha(x,t)}$ is a matrix. In this case, the force F is not gauge invariant anymore. However, it is possible to construct gauge invariants which depend on F. This is the case for the Standard Model in particle physics (see Vol. III).

(b) Covariant directional derivative: Consider the curve

$$C : x = x(\sigma), \quad t = t(\sigma),$$

where the curve parameter σ varies in the interval $[0, \sigma_0]$. The classical directional derivative along the curve C is defined by

$$\frac{d}{d\sigma} := \frac{dx(\sigma)}{d\sigma}\frac{\partial}{\partial x} + \frac{dt(\sigma)}{d\sigma}\frac{\partial}{\partial t}.$$

Explicitly, we get

$$\frac{d}{d\sigma}\psi(x(\sigma),t(\sigma)) = \frac{dx(\sigma)}{d\sigma}\psi_x(x(\sigma),t(\sigma)) + \frac{dt(\sigma)}{d\sigma}\psi_t(x(\sigma),t(\sigma)).$$

Similarly, the covariant directional derivative along the curve C is defined by

$$\boxed{\frac{D}{d\sigma} := \frac{dx(\sigma)}{d\sigma}\nabla_x + \frac{dt(\sigma)}{d\sigma}\nabla_t.}$$

Explicitly,

$$\begin{aligned} \frac{D}{d\sigma}\psi(x(\sigma),t(\sigma)) = \frac{d}{d\sigma}\psi(x(\sigma),t(\sigma)) + \mathrm{i}A(x(\sigma)),t(\sigma))\frac{dx(\sigma)}{d\sigma} \\ +\mathrm{i}U(x(\sigma),t(\sigma))\frac{dt(\sigma)}{d\sigma}. \end{aligned} \tag{1.49}$$

(c) Parallel transport: We say that the field function ψ is parallel along the curve C iff

$$\boxed{\frac{D}{d\sigma}\psi(x(\sigma),t(\sigma)) = 0, \qquad 0 \le \sigma \le \sigma_0.} \tag{1.50}$$

By (1.49), this notion depends on the gauge fields U, A. In particular, if the gauge fields U, A vanish, then parallel transport means that the field ψ is constant along the curve C.

The following observation is crucial. It follows from the key relation (1.45) on page 36 that the equation (1.50) of parallel transport is invariant under local gauge transformations. This means that (1.50) implies

$$\frac{D^+}{d\sigma}\,\psi_+(x(\sigma),t(\sigma)) = 0, \qquad 0 \le \sigma \le \sigma_0.$$

Consequently, in terms of mathematics, parallel transport possesses a geometric meaning with respect to local symmetry transformations.

In terms of physics, parallel transport describes the transport of physical information in space and time.

This transport is local in space and time, which reflects the Faraday–Green locality principle.

The Cartan differential. The most elegant formulation of gauge theories is based on the use of the covariant Cartan differential. As a preparation, let us recall the classical Cartan calculus. We will use the following relations:

$$dx \wedge dt = -dt \wedge dx, \quad dx \wedge dx = 0, \quad dt \wedge dt = 0. \tag{1.51}$$

Moreover, the wedge product of three factors of the form dx, dt is always equal to zero. For example,

$$dx \wedge dt \wedge dt = 0, \quad dt \wedge dx \wedge dt = 0. \tag{1.52}$$

For the wedge product, both the distributive law and the associative law are valid. Let $\psi : \mathbb{R}^2 \to \mathbb{C}$ be a smooth function. By definition,

- $d\psi := \psi_x dx + \psi_t dt$.

The differential 1-form

$$\boxed{\mathcal{A} := \mathrm{i}A dx + \mathrm{i}U dt} \tag{1.53}$$

is called the Cartan **connection form**. By definition,

- $d\mathcal{A} := \mathrm{i}dA \wedge dx + \mathrm{i}dU \wedge dt$,
- $d(\psi\, dx \wedge dt) = d\psi \wedge dx \wedge dt = 0$ (Poincaré identity).

The Poincaré identity is a consequence of (1.52).

The covariant Cartan differential. We now replace the classical partial derivatives by the corresponding covariant partial derivatives. Therefore, we replace $d\psi$ by the definition

- $D\psi := \nabla_x \psi\, dx + \nabla_t \psi\, dt$.

Similarly, we define

- $D\mathcal{A} := \mathrm{i}DA \wedge dx + \mathrm{i}DU \wedge dt$,
- $D(\psi\, dx \wedge dt) = D\psi \wedge dx \wedge dt = 0$ (Bianchi identity).

The Bianchi identity is a consequence of (1.52). Let us introduce the Cartan **curvature form** $\mathcal{F}$ by setting

$$\boxed{\mathcal{F} := D\mathcal{A}.} \tag{1.54}$$

Theorem 1.3 (i) $D\psi = d\psi + \mathcal{A}\psi$.
(ii) $\mathcal{F} = d\mathcal{A} + \mathcal{A} \wedge \mathcal{A}$ *(Cartan's structural equation).*
(iii) $D\mathcal{F} = 0$ *(Bianchi identity).*

In addition, we have the following relations for the curvature form $\mathcal{F}$:

- $\mathcal{F} = d\mathcal{A} + [U, A]_-$.[22]
- $\mathcal{F} = \mathrm{i}F dx \wedge dt$, where $\mathrm{i}F = U_x - A_t$.

Proof. Ad (i). $D\psi = (\psi_x + \mathrm{i}A\psi)dx + (\psi_t + \mathrm{i}U\psi)dt$.
Ad (ii). Note that

$$\begin{aligned} DA &= \mathrm{i}(A_x + A^2)dx + \mathrm{i}(A_t + \mathrm{i}UA)dt, \\ DU &= \mathrm{i}(U_x + \mathrm{i}AU)dx + (U_t + \mathrm{i}U^2)dt, \end{aligned}$$

and

$$\mathcal{A} \wedge \mathcal{A} = -(Adx + Udt) \wedge (Adx + Udt) = (UA - AU)\, dx \wedge dt.$$

Hence

$$\begin{aligned} D\mathcal{A} &= \mathrm{i}DA \wedge dx + \mathrm{i}DU \wedge dt \\ &= \mathrm{i}(A_t + \mathrm{i}UA)\, dt \wedge dx + \mathrm{i}(U_x + \mathrm{i}AU)\, dx \wedge dt = d\mathcal{A} + \mathcal{A} \wedge \mathcal{A}. \end{aligned}$$

This yields all the identities claimed above. □

The results concerning the curvature form $\mathcal{F}$ above show that Cartan's structural equation (1.54) is nothing else than a reformulation of the equation

$$\mathrm{i}F = U_x - A_t,$$

which relates the force F to the potentials U, A. Furthermore, we will show in Sect. 5.11 on page 333 that Cartan's structural equation is closely related to both

- Gauss' theorema egregium on the computation of the Gaussian curvature of a classic surfaces by means of the metric tensor and its partial derivatives,
- and the Riemann formula for the computation of the Riemann curvature tensor of a Riemannian manifold by means of the metric tensor and its partial derivatives.

In the present case, the formulas can be simplified in the following way. It follows from the commutativity property $AU = UA$ that:

[22] Here, we use the Lie bracket $[U, A]_- := UA - AU$.

- $\mathcal{F} = dA$,
- $\mathcal{F} = \mathrm{i}F\ dx \wedge dt = \mathrm{i}(U_x - A_t)\ dx \wedge dt$.

A similar situation appears in Maxwell's theory of electromagnetism. For more general gauge theories, the symbols A and U represent matrices. Then we obtain the additional nonzero terms $[U, A]_-$ and $\mathcal{A} \wedge \mathcal{A}$. This is the case in the Standard Model of elementary particles (see Vol. III).

The mathematical language of fiber bundles. In mathematics, we proceed as follows:

- We consider the field $\psi : \mathbb{R}^2 \to \mathbb{C}$ as a section of the line bundle $\mathbb{R}^2 \times \mathbb{C}$ (with typical fiber $\mathbb{C}$) (see Fig. 4.9 on page 208).
- The line bundle $\mathbb{R}^2 \times \mathbb{C}$ is associated to the principal fiber bundle $\mathbb{R}^2 \times U(1)$ (with structure group $U(1)$ called the gauge group in physics).[23]
- As above, the differential 1-form $\mathcal{A} := \mathrm{i}A dx + \mathrm{i}U dt$ is called the connection form on the base manifold $\mathbb{R}^2$ of the principal fiber bundle $\mathbb{R}^2 \times U(1)$, and
- the differential 2-form

$$\mathcal{F} = d\mathcal{A} + \mathcal{A} \wedge \mathcal{A}$$

  is called the curvature form on the base manifold $\mathbb{R}^2$ of the principle fiber bundle $\mathbb{R}^2 \times U(1)$.
- Finally, we define

$$D\psi := d\psi + \mathcal{A}\psi. \tag{1.55}$$

  This is called the covariant differential of the section ψ of the line bundle $\mathbb{R}^2 \times \mathbb{C}$.

Observe that:

> *The values of the gauge field functions* $\mathrm{i}U, \mathrm{i}A$ *are contained in the Lie algebra* $u(1)$ *of the Lie group* $U(1)$*. Thus, the connection form* $\mathcal{A}$ *is a differential 1-form with values in the Lie algebra* $u(1)$*.*

This can be generalized by replacing

- the special commutative Lie group $U(1)$
- by the the general Lie group $\mathcal{G}$.

Then the values of the gauge fields $\mathrm{i}U, \mathrm{i}A$ are contained in the Lie algebra $\mathcal{LG}$ to $\mathcal{G}$. If $\mathcal{G}$ is a noncommutative Lie group (e.g., $SU(N)$ with $N \geq 2$), then the additional force term $\mathcal{A} \wedge \mathcal{A}$ does not vanish identically, as in the special case of the commutative group $U(1)$.[24] In Vol. III on gauge theory, we will show that the Standard Model in particle physics corresponds to this approach by choosing the gauge group $U(1) \times SU(2) \times SU(3)$. Here,

[23] Recall that the elements of the Lie group $U(1)$ are the complex numbers $\mathrm{e}^{\mathrm{i}\alpha}$ with real parameter α. The elements of the Lie algebra $u(1)$ are the purely imaginary numbers $\alpha\mathrm{i}$.

[24] The Lie group $SU(N)$ consists of all the unitary $(N \times N)$-matrices whose determinant is equal to one (special unitary group).

- the electroweak interaction is the curvature of a $(U(1) \times SU(2))$-bundle (Glashow, Salam and Weinberg in the 1960s), and
- the strong interaction is the curvature of a $SU(3)$-bundle (Gell-Mann and Fritzsch in the early 1970s).

Historical remarks. General gauge theory is equivalent to modern differential geometry. This will be thoroughly studied in Vol. III. At this point let us only make a few historical remarks.

> *In 1827 Gauss proved that the curvature of a 2-dimensional surface in 3-dimensional Euclidean space is an intrinsic property of the manifold.*

This means that the curvature of the surface can be measured without using the surrounding space. This is the content of Gauss' *theorema egregium*. The Gauss theory was generalized to higher-dimensional manifolds by Riemann in 1854. Here, the Gaussian curvature has to be replaced by the Riemann curvature tensor. In 1915 Einstein used this mathematical approach in order to formulate his theory of gravitation (general theory of relativity). In Einstein's setting, the masses of celestial bodies (stars, planets, and so on) determine the Riemann curvature tensor of the four-dimensional space-time manifold which corresponds to the universe. Thus, Newton's gravitational force is replaced by the curvature of a four-dimensional pseudo-Riemannian manifold $\mathcal{M}^4$. The motion of a celestial body (e.g., the motion of a planet around the sun) is described by a geodesic curve C in $\mathcal{M}^4$. Therefore, Einstein's equation of motion tells us that the 4-dimensional velocity vector of C is parallel along the curve C. Roughly speaking, this corresponds to (1.50) where ψ has to be replaced by the velocity field of C. In the framework of his theory of general relativity, Einstein established the principle

$$\textit{force} = \textit{curvature}$$

for gravitation. Nowadays, the Standard Model in particle physics is also based on this beautiful principle which is the most profound connection between mathematics and physics.

In 1917 Levi-Civita introduced the notion of parallel transport, and he showed that both the Gaussian curvature of 2-dimensional surfaces and the Riemann curvature tensor of higher-dimensional manifolds can be computed by using parallel transport of vector fields along small closed curves. In the 1920s, Élie Cartan invented the method of moving frames.[25] In the 1950s, Ehresmann generalized Cartan's method of moving frames to the modern curvature theory for principal fiber bundles (i.e., the fibers are Lie groups) and their associated vector bundles (i.e., the fibers are linear spaces). In 1963, Kobayashi and Nomizu published the classic monograph

[25] For an introduction to this basic tool in modern differential geometry, we refer to the textbook by T. Ivey and J. Landsberg, Cartan for Beginners: Differential Geometry via Moving Frames and Exterior Differential Systems, Amer. Math. Soc., Providence, Rhode Island, 2003. See also Vol. III.

Foundations of Differential Geometry,

Vols. 1, 2, Wiley, New York. This finishes a longterm development in mathematics.

In 1954, the physicists Yang and Mills created the Yang–Mills theory. It was their goal to generalize Maxwell's electrodynamics. To this end, they started with the observation that Maxwell's electrodynamics can be formulated as a gauge theory with the gauge group $U(1)$. This was known from Hermann Weyl's paper: Elektron und Gravitation, Z. Phys. **56** (1929), 330–352 (in German). Yang and Mills

- replaced the commutative group $U(1)$
- by the non-commutative group $SU(2)$.

The group $SU(2)$ consists of all the complex (2×2)-matrices A with $AA^\dagger = I$ and $\det A = 1$. Interestingly enough, in 1954 Yang and Mills did not know a striking physical application of their model. However, in the 1960s and 1970s, the Standard Model in particle physics was established as a modified Yang–Mills theory with the gauge group

$$\boxed{U(1) \times SU(2) \times SU(3).}$$

The modification concerns the use of an additional field called Higgs field in order to generate the masses of the three gauge bosons W^+, W^-, Z^0. In the early 1970s, Yang noticed that the Yang–Mills theory is a special case of Ehresmann's modern differential geometry in mathematics. For the history of gauge theory, we refer to:

L. Brown et al. (Eds.), The Rise of the Standard Model, Cambridge University Press, 1995.

L. O'Raifeartaigh, The Dawning of Gauge Theory, Princeton University Press, 1997.

C. Taylor (Ed.), Gauge Theories in the Twentieth Century, World Scientific, Singapore, 2001 (a collection of fundamental articles).

Mathematics and physics. Arthur Jaffe writes the following in his beautiful survey article *Ordering the universe: the role of mathematics* in the Notices of the American Mathematical Society **236** (1984), 589–608:[26]

> There is an exciting development taking place right now, *reunification* of mathematics with theoretical physics... In the last ten or fifteen years mathematicians and physicists realized that modern geometry is in fact the natural framework for gauge theory. The gauge potential in gauge theory is the connection of mathematics. The gauge field is the mathematical curvature defined by the connection; certain *charges* in physics are the topological invariants studied by mathematicians. While the mathematicians and physicists worked separately on similar ideas, they did *not*

[26] Reprinted by permission of the American Mathematical Society. This report was originated by the National Academy of Sciences of the U.S.A.

duplicate each other's efforts. The mathematicians produced general, far-reaching theories and investigated their ramifications. Physicists worked out details of certain examples which turned out to describe nature beautifully and elegantly. When the two met again, the results are more powerful than either anticipated... In mathematics, we now have a new motivation to use specific insights from the examples worked out by physicists. This signals the return to an ancient tradition.

Felix Klein (1849–1925) writes about mathematics:

Our science, in contrast to others, is not founded on a single period of human history, but has accompanied the development of culture through all its stages. Mathematics is as much interwoven with Greek culture as with the most modern problems in engineering. It not only lends a hand to the progressive natural sciences but participates at the same time in the abstract investigations of logicians and philosophers.

Hints for further reading:

S. Chandrasekhar, Truth and Beauty: Aesthetics and Motivations in Science, Chicago University Press, Chicago, Illinois, 1990.

E. Wigner, Philosophical Reflections and Syntheses. Annotated by G. Emch. Edited by J. Mehra and A. Wightman, Springer, New York, 1995.

G. 't Hooft, In Search for the Ultimate Building Blocks, Cambridge University Press, 1996.

R. Brennan, Heisenberg Probably Slept Here: The Lives, Times, and Ideas of the Great Physicists of the 20th Century, Wiley, New York, 1997.

J. Wheeler and K. Ford, Geons, Black Holes, and Quantum Foam: a Life in Physics, Norton, New York, 1998.

B. Greene, The Elegant Universe: Supersymmetric Strings, Hidden Dimensions, and the Quest for the Ultimate Theory, Norton, New York, 1999.

A. Zee, Fearful Symmetry: The Search for Beauty in Modern Physics, Princeton University Press, 1999.

G. Johnson, Strange Beauty: Murray Gell-Mann and the Revolution in Twentieth Century Physics, A. Knopf, New York, 2000.

G. Farmelo (Ed.), It Must be Beautiful: Great Equations of Modern Science, Granta Publications, London, 2003.

M. Veltman, Facts and Mysteries in Elementary Particle Physics, World Scientific, Singapore, 2003.

R. Penrose, The Road to Reality: A Complete Guide to the Laws of the Universe, Jonathan Cape, London, 2004.

J. Barrow, New Theories of Everything: The Quest for Ultimate Explanation, Oxford University Press, New York, 2007.

F. Patras, La pensée mathématique contemporaine (Philosophy of modern mathematics), Presses Universitaire de France, Paris, 2001 (in French).

H. Wußing, 6000 Years of Mathematics: a Cultural Journey through Time, Vols. I, II, , Springer, Heidelberg, 2008 (in German).

E. Zeidler, Reflections on the future of mathematics. In: H. Wußing (2008), Vol. II (last chapter) (in German).

The Cambridge Dictionary of Philosophy, edited by R. Audi, Cambridge University Press, 2005.

1.10 The Challenge of Dark Matter

> Although science teachers often tell their students that the periodic table of the elements shows what the Universe is made of, this is not true. We now know that most of the universe – about 96% of it – is made of dark matter that defies brief description, and certainly is not represented by Mendeleev's periodic table. This unseen 'dark matter' is the subject of this book...
>
> Dark matter provides a further remainder that we humans are not essential to the Universe. Ever since Copernicus (1473–1543) and others suggested that the Earth was not the center of the Universe, humans have been on a slide away from cosmic significance. At first we were not at the center of the Solar System, and then the Sun became just another star in the Milky Way, not even in the center of our host Galaxy. By this stage the Earth and its inhabitants had vanished like a speck of dust in a storm. This was a shock.
>
> In the 1930s Edwin Hubble showed that the Milky Way, vast as it is, is a mere 'island Universe' far removed from everywhere special; and even our home galaxy was suddenly insignificant in a sea of galaxies, then clusters of galaxies. Now astronomers have revealed that we are not even made of the same stuff as most of the Universe. While our planet – our bodies, even – are tangible and visible, most of the matter in the Universe is not. Our Universe is made of darkness. How do we respond to that?
>
> Ken Freeman and Geoff McNamarra, 2006

This quotation is taken from the monograph by K. Freemann and G. McNamarra, In Search of Dark Matter, Springer, Berlin and Praxis Publishing Chichester, United Kingdom, 2006 (reprinted with permission). As an introduction to modern cosmology we recommend the monograph by S. Weinberg, Cosmology, Oxford University, 2008.

2. The Basic Strategy of Extracting Finite Information from Infinities – Ariadne's Thread in Renormalization Theory

There is no doubt that renormalization is one of the most sophisticated procedures for obtaining significant numerical quantities by starting from meaningless mathematical expressions. This is fascinating for both physicists and mathematicians.[1]

Alain Connes, 2003

Quantum field theory deals with fields $\psi(x)$ that destroy and create particles at a spacetime point x. Earlier experience with classical electron theory provided a warning that a point electron will have infinite electromagnetic self-mass; this mass is

$$\frac{e^2}{6\pi ac^2}$$

for a surface distribution of charge with radius a, and therefore blows up for $a \to 0$. Disappointingly this problem appeared with even greater severity in the early days of quantum field theory, and although greatly ameliorated by subsequent improvements in the theory, it remains with us to the present day.

The problem of infinities in quantum field theory was apparently first noted in the 1929–30 papers by Heisenberg and Pauli.[2] Soon after, the presence of infinities was confirmed by calculations of the electromagnetic self-energy of a bound electron by Oppenheimer, and of a free electron by Ivar Waller.[3]

Steven Weinberg, 1995

[1] A. Connes, Symétries galoisiennes et renormalisation (in French). In: Poincaré Seminar, Paris, 2002: Vacuum energy – renormalization, pp. 241–264. Edited by B. Duplantier and V. Rivasseau, Birkhäuser, Basel, 2003.

[2] W. Heisenberg and W. Pauli, On the quantum electrodynamics of wave fields, Z. Phys. **56** (1929), 1–61; **59** (1930), 108–190 (in German).
J. Oppenheimer, Note on the theory of the interaction of field and matter, Phys. Rev. **35** (1930), 461–477.
I. Waller, Remarks on the role of the self-energy of electrons in the quantum theory of radiation, Z. Phys. **61** (1930), 721–730; **62** (1930), 673–676 (in German).

[3] S. Weinberg, Quantum Field Theory, Vol. 1, Sect. 1.3, The Problem of Infinities, Cambridge University Press, Cambridge, United Kingdom (reprinted with permission).

2.1 Renormalization Theory in a Nutshell

In renormalization theory, one has to clearly distinguish between

(I) mathematical regularization of infinities, and
(II) renormalization of physical parameters by introducing effective parameters which relate the mathematical regularization to physical measurements.

In order to help the reader to find his/her way through the jungle of renormalization theory, let us discuss a few basic ideas. (The details will be studied later on.) This concerns:

- Bare and effective parameters – the effective frequency and the running coupling constant of an anharmonic oscillator.
- The renormalized Green's function and the renormalization group.
- The zeta function, Riemann's idea of analytic continuation, and the Casimir effect in quantum electrodynamics.
- Meromorphic functions and Mittag-Leffler's idea of subtractions.
- Euler's gamma function and the dimensional regularization of integrals.

Behind this there is the general strategy of extracting finite information from infinities. For example, we will consider the following examples:

- Regularization of divergent integrals (including the famous overlapping divergences).
- Abel's adiabatic regularization of infinite series.
- Adiabatic regularization of oscillating integrals (the prototype of the Feynman path integral trick).
- Poincaré's asymptotic series, the Landau singularity, and the Ritt theorem.
- The summation methods by Euler, Frobenius, Hölder, and Borel.
- Tikhonov regularization of ill-posed problems.

In modern renormalization theory, the combinatorics of Feynman diagrams plays the crucial role. This is related to the mathematical notion of

- Hopf algebras and
- Rota–Baxter algebras.

The prototypes of these algebras appear in classical complex function theory in connection with the inversion of holomorphic functions (the relation between Lagrange's inversion formula and the coinverse (antipode) of a Hopf algebra) and the regularization of Laurent series (Rota-Baxter algebras). This will be studied in Chap. 3.

2.1.1 Effective Frequency and Running Coupling Constant of an Anharmonic Oscillator

In quantum electrodynamics, physicists distinguish between

- the bare mass of the electron, and
- its effective mass (resp. the bare charge of the electron and its effective charge).

The effective mass and the effective charge of the electron coincide with the physical quantities measured in physical experiments. In contrast to this, the bare parameters are only introduced as auxiliary quantities in renormalization theory for constructing the effective quantities in terms of a mathematical algorithm. To illustrate the crucial difference between bare parameters and effective physical parameters, let us summarize the main features of a simple *rigorous* model which is studied in Sect.

11. 5 of Vol. I in full detail. We consider the following nonlinear boundary-value problem

$$\boxed{\ddot{x}(t) + \omega^2 x(t) = -\kappa x(t)^3 + \mu \sin t, \quad t \in \mathbb{R}, \quad x(0) = x(\pi) = 0.} \tag{2.1}$$

In terms of physics, this equation describes the motion $x = x(t)$ of a point of mass $m = 1$ on the real line. The positive parameters ω and κ are called the angular frequency and the coupling constant, respectively. Moreover, the real parameter μ measures the strength of the external force $F(t) = \mu \sin t$. The term $-\kappa x(t)^3$ describes the self-interaction of the particle. The sign convention of κ is chosen in such a way that positive values of κ correspond to a repulsive force, by (6.16) on page 370. Physicists call this an anharmonic oscillator. We will assume that the coupling constant κ and the external force parameter μ are sufficiently small. Then it is possible to apply the methods of perturbation theory. We are looking for a smooth 2π-periodic odd solution $x : \mathbb{R} \to \mathbb{R}$ of (2.1). By the theory of Fourier series, the function $x(.)$ possesses the convergent representation

$$x(t) = b_1 \sin t + b_2 \sin 2t + \dots \qquad \text{for all} \quad t \in \mathbb{R} \tag{2.2}$$

with the Fourier coefficients $b_k := \sqrt{\frac{2}{\pi}} \int_0^\pi F(t) \sin kt \, dt$ where $k = 1, 2, \dots$ In what follows, we will also write $b_k(x)$ instead of b_k. For the mathematical analysis, it is crucial first to study the homogeneous linearized problem

$$\ddot{x}(t) + \omega^2 x(t) = 0, \qquad t \in \mathbb{R}, \qquad x(0) = x(\pi) = 0 \tag{2.3}$$

and the inhomogeneous linearized problem

$$\ddot{x}(t) + \omega^2 x(t) = F(t), \qquad t \in \mathbb{R}, \qquad x(0) = x(\pi) = 0 \tag{2.4}$$

with the given smooth, 2π-periodic, odd force function $F : \mathbb{R} \to \mathbb{R}$. Then the function F allows the convergent representation

$$F(t) = \beta_1 \sin t + \beta_2 \sin 2t + \dots, \qquad t \in \mathbb{R}$$

with the Fourier coefficients $\beta_k := b_k(F) = \sqrt{\frac{2}{\pi}} \int_0^\pi F(t) \sin kt \; dt, k = 1, 2, \dots$ If $x(t) = b_1 \sin t + b_2 \sin 2t + \dots$ is a solution of (2.4), then

$$(\omega^2 - 1^2) b_1 \sin t + (\omega^2 - 2^2) b_2 \sin 2t + \dots = \beta_1 \sin t + \beta_2 \sin 2t + \dots$$

This implies the key condition

$$\boxed{b_k(\omega^2 - k^2) = \beta_k, \qquad k = 1, 2, \dots} \tag{2.5}$$

In particular, this tells us that the homogeneous linearized problem (2.3) (i.e., $\beta_k = 0$ for all k) has precisely the following nontrivial solutions

$$\omega = k, \qquad x(t) = C \sin kt, \quad t \in \mathbb{R}, \qquad k = 1, 2, \dots$$

which are called eigenoscillations in physics. Here, C is a free complex parameter.

Let us now investigate the nonlinear problem (2.1). We have to distinguish between two cases, namely,

- the regular non-resonance case and
- the singular resonance case which corresponds to renormalization in physics.

We will show that the trouble comes from resonances between the external forces and the eigenoscillations.

To illustrate this, consider problem (2.4) with $\omega := 1$. By (2.5), we get $\beta_1 = 0$. Explicitly,

$$\int_0^\pi F(t) \sin t \, dt = 0. \tag{2.6}$$

This is a solvability condition for the external force F. The original problem (2.4) has a solution iff (2.6) is satisfied. In terms of physics, condition (2.6) means that the external force is not in resonance with the eigenoscillation $t \mapsto \sin t$. In particular, if we choose $\omega := 1$ and $F(t) := \sin t$, then the condition (2.6) is violated, and problem (2.4) has no solution.

In contrast to this, if $\omega := 1$ and $F(t) = \sin kt$ for fixed $k = 2, 3, \ldots$, then it follows from (2.5) that the original problem (2.4) has the general solution

$$x(t) = C \sin t + \frac{\sin kt}{1 - k^2} \tag{2.7}$$

where C is an arbitrary complex parameter. The point is that, in this critical case, the external force F does not uniquely determine the motion $x = x(t)$ of the mass point. Thus the mathematical theory does not determine the parameter C.

For a given motion $x = x(t)$, the specific value of C has to be determined by a physical experiment.

We expect that, in the singular resonance case, small perturbations of the external force transform the one-parameter family (2.7) of motions into a perturbed one-parameter family of motions. We will show below that this is true.

(i) Non-resonance case. Let $\omega > 0$ with $\omega \neq 1, 2, \ldots$. Then the inhomogeneous linearized problem (2.4) has the unique solution

$$x(t) = \int_0^\pi G(t, \tau) F(\tau) d\tau, \qquad t \in \mathbb{R}$$

with the Green's function

$$G(t, \tau) := \sqrt{\frac{2}{\pi}} \sum_{k=1}^{\infty} \frac{\sin t \sin \tau}{\omega^2 - k^2}, \qquad t, \tau \in \mathbb{R}. \tag{2.8}$$

This corresponds to $b_k(x) = \frac{b_k(F)}{\omega^2 - k^2}$ for all $k = 1, 2, \ldots$ by (2.5). Observe that the singularities of the Green's function correspond to the angular frequencies of the eigenoscillations. Our assumption $\omega \neq 1, 2, \ldots$ guarantees that the Green's function G is well-defined. The original nonlinear problem (2.1) is equivalent to the nonlinear integral equation

$$x(t) = -\kappa \int_0^\pi G(t, \tau) x(\tau)^3 d\tau + \mu \int_0^\pi G(t, \tau) \sin \tau \, d\tau. \tag{2.9}$$

This equation can be solved by the iterative method

$$x_{n+1}(t) = -\kappa \int_0^\pi G(t, \tau) x_n(\tau)^3 d\tau + \mu \int_0^\pi G(t, \tau) \sin \tau \, d\tau \tag{2.10}$$

for $n = 0, 1, 2, \ldots$ with $x_0 := 0$. This iterative method converges uniformly on the interval $[0, \pi]$ to the solution $x(.)$ of the original problem (2.1) if the

parameters κ and μ are sufficiently small (see Sect. 11. 5 of Vol. I). The first approximation

$$x_1(t) = \mu \int_0^\pi G(t,\tau) \sin\tau \, d\tau = \frac{\mu \sin t}{\omega^2 - 1}, \qquad t \in \mathbb{R}$$

is called the Born approximation by physicists.[4] In mathematics, the convergence of such iterative methods was systematically studied by Picard (1856–1941) in 1890. The abstract setting is the Banach fixed-point theorem (also called the contraction principle) which was formulated by Banach (1892–1945) in 1922.[5]

(ii) Resonance case. Let $\omega = 1$. If we formally use the integral equation (2.9) and the iterative method (2.10), then the approach breaks down because of the singularity of the Green's function (2.8) with $k = 1$. For example, the Born approximation $x_1(.)$ is infinite if $\omega = 1$.

This is the typical trouble in physics if resonance occurs.

To overcome the difficulties in the resonance case, we will use the method of the pseudo-resolvent introduced by Erhard Schmidt in 1908.[6]

Let us consider the anharmonic oscillator (2.1) in the case of the critical angular frequency $\omega = 1$. More precisely, we choose

$$\omega^2 = 1 + \varepsilon.$$

We also assume that the external force vanishes, that is, $\mu = 0$. This means that we want to study the problem

$$\boxed{\ddot{x}(t) + (1+\varepsilon)x(t) = -\kappa x(t)^3, \quad t \in \mathbb{R}, \quad x(0) = x(\pi) = 0.} \tag{2.11}$$

This problem has the trivial solution $x = 0$, $\varepsilon =$ arbitrary. We are interested in a nontrivial solution. We will show that, for small real parameters κ, s, there exists a nontrivial solution given by

$$\boxed{x(t;\kappa,s) = s \cdot \sin t + \ldots, \quad \omega(\kappa,s)^2 = 1 - \tfrac{3}{4}\kappa s + \ldots} \tag{2.12}$$

The dots denote terms of higher order with respect to the small parameters κ, s. In addition, we have the relation

$$s = b_1(x(.;\kappa,s)) = \sqrt{\frac{2}{\pi}} \int_0^\pi x(t;\kappa,s) \sin t \, dt. \tag{2.13}$$

The crucial point with respect to the interpretation in terms of renormalization theory. Before sketching the mathematical proof of (2.12), let us discuss the physical consequences.

[4] M. Born, On collision processes in quantum theory, Z. Phys. **37** (1926), 863–867 (in German).

[5] E. Picard, On partial differential equations and the method of successive approximations, J. Math. Pures et Appl. **6** (1890), 145–219 (in French).
S. Banach, On operations in abstract sets and their applications to integral equations, Fund. Math. **3** (1922), 133–181 (in French).

[6] E. Schmidt, On the theory of linear and nonlinear integral equations III, Math. Ann. **65** (1908), 370–399 (in German). Erhard Schmidt (1876–1959) was a Ph.D. student of Hilbert (1862–1943) in Göttingen. A detailed study of Schmidt's pseudo-resolvent method in bifurcation theory can be found in Zeidler (1986), Vol. I, quoted on page 1049.

The point is that we do not get a unique solution, but a family of solutions which depend on an additional parameter s. This parameter has to be determined by a physical experiment.

To explain this, suppose that we measure the motion $x = x(t)$ and the angular frequency ω of the motion. Then $x(t) = x(t; \kappa, s)$ and $\omega = \omega(\kappa, s)$.

- By (2.13), we obtain the parameter s.
- By (2.12), we obtain the value of the coupling constant

$$\kappa = \frac{4(1-\omega^2)}{3s}.$$

This tells us that the coupling constant κ depends on s. Therefore, we call this a running coupling constant.

This simple model corresponds to the (much more complicated) determination of the electron charge, the electron mass, and the running coupling constant in quantum electrodynamics. In particle physics, the running coupling constants of electromagnetic, weak, and strong interaction depend on energy and momentum of the scattering process observed in a particle accelerator.

Sketch of the proof of (2.12). The following arguments are typical for a branch of nonlinear analysis called bifurcation theory.[7] First let us replace the ill-posed linearized inhomogeneous problem (2.4) with $\omega = 1$ by the well-posed problem

$$\ddot{x}(t) + x(t) + b_1(x)\sin t = F(t), \quad t \in \mathbb{R}, \quad x(0) = x(\pi) = 0 \tag{2.14}$$

where $b_1(x) := \sqrt{\frac{2}{\pi}} \int_0^\pi x(t) \sin t \, dt$. The additional term $b_1(x) \sin t$ is chosen in order to guarantee that the homogeneous problem (2.14) has only the trivial solution $x = 0$. The point is that the key relation (2.5) passes over to $b_k(1 - k^2 + \delta_{k1}) = \beta_k$ for $k = 1, 2, \ldots$ Hence

$$b_1(x) = b_1(F), \qquad b_k(x) = \frac{b_k(F)}{1-k^2}, \quad k = 2, 3, \ldots$$

More precisely, for given smooth odd 2π-periodic function $F : \mathbb{R} \to \mathbb{R}$, problem (2.14) has the unique solution

$$x(t) = \int_0^\pi G_{\text{pseudo}}(t, \tau) F(\tau) \, d\tau, \qquad t \in \mathbb{R}$$

with the so-called pseudo-resolvent

$$G_{\text{pseudo}}(t, \tau) := \sqrt{\frac{2}{\pi}} \left(\sin t \sin \tau + \sum_{k=2}^{\infty} \frac{\sin t \sin \tau}{1-k^2} \right).$$

This represents a *regularization* of the Green's function (2.8) with $\omega = 1$. The original problem (2.11) is equivalent to the equation

[7] Bifurcation theory and its applications in astronomy, astrophysics, biology, chemistry, elasticity, fluid dynamics, and so on, are studied in Zeidler (1986), Vols. I and IV, quoted on page 1049. We also recommend the monograph by H. Kielhöfer, Bifurcation Theory: An Introduction with Applications to Partial Differential Equations, Springer, New York, 2004.

$$\ddot{x}(t) + x(t) + b_1(x)\sin t = -\kappa x(t)^3 - \varepsilon x(t) + s\sin t, \quad x(0) = x(\pi) = 0$$

with the side condition $s = b_1(x)$. Using the pseudo-resolvent, this problem is equivalent to

$$x(t) = \int_0^\pi G_{\text{pseudo}}(t,\tau)(-\kappa x(\tau)^3 - \varepsilon x(\tau) + s\sin\tau)\,d\tau$$

with the side condition $s = b_1(x)$. This means that we have to solve the equation

$$x(t) = s\sin t - \int_0^\pi G_{\text{pseudo}}(t,\tau)(\kappa x(\tau)^3 + \varepsilon x(\tau))\,d\tau \tag{2.15}$$

with the side condition $s = b_1(x)$. We will proceed as follows.

(a) First we solve the equation (2.15) by an iterative method. This way, we get the function $x = x(t;\kappa,s,\varepsilon)$ which depends on the small real parameters κ, s, ε.
(b) We solve the side condition

$$b_1(x(.;\kappa,s,\varepsilon)) = s$$

by the implicit function theorem.[8] This way, we get $\varepsilon = \varepsilon(\kappa, s)$.
(c) The final solution reads as $x = x(t;\kappa,s,\varepsilon(\kappa,s))$.

This yields the desired solution (2.12). The complete proof can be found in Vol. I, Sect. 11.5.

The renormalized Green's function. In quantum field theory, physicists use the notion of renormalized (or effective) Green's function. Let us explain this in the case of the present simple model. The solution $x = x(t;\kappa,s)$, $\omega = \omega(\kappa,s)$ satisfies the original equation

$$\ddot{x}(t;\kappa,s) + \omega(\kappa,s)^2 x(t;\kappa,s) = -\kappa x(t;\kappa,s)^3$$

with the boundary condition $x(0;\kappa,s) = x(\pi;\kappa,s) = 0$. If κ and s are different from zero and sufficiently small, then $\omega(\kappa,s) \neq 1$. That is, the angular frequency is different from the angular eigenfrequency $\omega = 1$. Therefore, we can use the Green's function G as in the non-resonance case (2.8). This yields the integral equation

$$x(t;\kappa,s) = -\kappa \int_0^\pi G(t,\tau)x(\tau;\kappa,s)^3\,d\tau$$

with

$$G(t,\tau) = \sqrt{\frac{2}{\pi}} \sum_{k=1}^{\infty} \frac{\sin t \sin\tau}{\omega(\kappa,s)^2 - k^2}, \qquad t,\tau \in \mathbb{R}.$$

This function depends on the effective angular frequency $\omega(\kappa,s)$ and is called the effective (or renormalized) Green's function. In the 1940s and 1950s, physicists invented approximative methods for computing effective Green's functions in quantum electrodynamics and in more general quantum field theories.

The renormalization group. It turns out that the parameter s of the solution is not fixed, but it can be rescaled (see Sect. 11.5.6 of Vol. I). This is the prototype of the method of renormalization group.

[8] In bifurcation theory, this side condition is called the bifurcation equation (see (2.13)).

2.1.2 The Zeta Function and Riemann's Idea of Analytic Continuation

For all complex numbers s with $\Re(s) > 1$, the definition

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}$$

of the zeta function makes sense, since the series is convergent.[9] In 1750 Euler proved that $\zeta(2) = \frac{\pi^2}{6}$ and

$$\zeta(2n) = (-1)^{n-1} \frac{(2\pi)^{2n}}{2(2n)!} B_{2n}, \qquad n = 1, 2, \ldots$$

with the Bernoulli numbers B_k defined by the generating function

$$\frac{x}{\mathrm{e}^x - 1} = \sum_{k=0}^{\infty} \frac{B_k}{k!} x^k \qquad \text{for all} \quad x \in \mathbb{R} \setminus \{0\}.$$

Explicitly, $B_0 = 1, B_1 = \frac{1}{2}, B_2 = \frac{1}{6}, B_3 = 0, B_4 = -\frac{1}{30}, \ldots$, and $B_{2k+1} = 0$ for all $k = 1, 2, \ldots$ For all complex numbers s with $\Re(s) < 1$, the series $\sum_{n=1}^{\infty} \frac{1}{n^s}$ is not convergent. In particular, if $s \leq 1$, then $\sum_{n=1}^{\infty} \frac{1}{n^s} = \infty$.

It is our goal to give this divergent series a well-defined finite meaning.

In 1859 Riemann proved that the zeta function can be analytically continued to a meromorphic function on the complex plane. This function has precisely one pole which is located at the point $s = 1$ with the Laurent series

$$\zeta(s) = \frac{1}{s-1} + \gamma + \sum_{k=1}^{\infty} a_k (s-1)^k \qquad \text{for all} \quad s \in \mathbb{C} \setminus \{1\}$$

where γ is the famous Euler constant given by

$$\gamma := \lim_{n \to \infty} \left(1 + \frac{1}{2} + \ldots + \frac{1}{n} - \ln n \right).$$

In particular, we have the Mellin transformation formula

$$\zeta(x) = \frac{1}{\Gamma(x)} \int_0^{\infty} \frac{u^{x-1}}{\mathrm{e}^u - 1}\, du$$

for all real numbers x different from $1, 0, -1, -2, \ldots$ Letting $f(x) := \frac{1}{x}$, the Euler–Maclaurin summation formula tells us that for all $n, m = 1, 2, \ldots$, we have

[9] In what follows, we are going to discuss important results obtained by the following mathematicians: Maclaurin (1698–1746), Euler (1707–1783), Mascheroni (1750–1800), Gauss (1777–1855), Cauchy (1789–1857), Abel (1802–1829), Liouville (1809–1882), Riemann (1826–1866), Mittag-Leffler (1846–1927), Frobenius (1849–1917), Poincaré (1854–1912), Otto Hölder (1859–1937), Hadamard (1865–1963), Tauber (1866–1942), Borel (1871–1956), Fejér (1880–1959), Marcel Riesz (1886–1969), Heinz Hopf (1894–1971), Rota (1933–1989).

$$\sum_{k=1}^{n} f(k) = \int_1^n f(x)dx + \frac{f(1)+f(n)}{2} + \sum_{k=1}^{m} \frac{B_{2k}}{(2k)!} \left(f^{(2k-1)}(n) - f^{(2k-1)}(0) \right) + R_{n,m}$$

with the remainder estimate

$$|R_{n,m}| \leq \frac{2}{(2\pi)^m} \int_1^n |f^{2m+1}(x)|\, dx.$$

Here, $f^{(k)}$ denotes the kth derivative of the function f. Explicitly,

$$\sum_{k=1}^{n} \frac{1}{k} = \ln n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + \frac{1}{120n^4} - \frac{1}{252n^6} + \dots$$

Euler used this series up to the term $\frac{1}{n^{14}}$ to compute the value

$$\gamma = 0.577\ 215\ 664\ 901\ 532\ 5\dots$$

In 1790 Mascheroni published an approximation of γ to 32 decimal places (with a mistake at the 20th decimal place). Therefore, the number γ is also called the Euler–Mascheroni constant.[10] Since

$$1 + \frac{1}{2} + \dots + \frac{1}{n} = \ln n + \gamma + o(1), \qquad n \to \infty,$$

the Euler constant γ is called the finite part of the divergent harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$. Furthermore, we get

$$\lim_{s \to 1} \left(\zeta(s) - \frac{1}{s-1} \right) = \gamma.$$

This motivates to say that the Euler constant is the finite part of the Riemann zeta function at the singular point $s = 1$. Motivated by these considerations, we define the regularization of the Euler series $\sum_{n=1}^{\infty} \frac{1}{n^s}$ in the following way:

$$\left[\sum_{n=1}^{\infty} \frac{1}{n^s} \right]_{\text{reg}} := \begin{cases} \zeta(s) & \text{if} \quad s \in \mathbb{C} \setminus \{1\}, \\ \gamma & \text{if} \quad s = 1. \end{cases}$$

For example, we obtain

$$[1 + 1 + 1 + \dots]_{\text{reg}} = \zeta(0) = -\frac{1}{2},$$

and $\left[\sum_{n=1}^{\infty} n^2\right]_{\text{reg}} = \zeta(-2) = 0$, as well as

[10] As standard texts on the Riemann zeta function, we recommend E. Titchmarsh and D. Heath-Brown, The Theory of the Riemann Zeta Function, Cambridge University Press, 1986, and H. Edwards, Riemann's Zeta Function, Academic Press, New York, 1974. The history of the Euler constant and its relation to numerous important developments in mathematics can be found in J. Havil, Gamma: Exploring Euler's Constant, Princeton University Press, 2003.

$$\left[\sum_{n=1}^{\infty} n^3\right]_{\text{reg}} = \zeta(-3) = \frac{1}{120}.$$

We have shown in Sect. 6.6 of Vol. I how to apply the idea of analytic continuation discussed above to the Epstein zeta function in order to compute the crucial Casimir effect in quantum electrodynamics. In the case of the Casimir effect, the physical experiment measures indeed the regularized mathematical quantity. This shows that

Nature sees analytic continuation.

2.1.3 Meromorphic Functions and Mittag-Leffler's Idea of Subtractions

The method of subtractions plays a fundamental role in the renormalization of quantum field theories according to Bogoliubov, Parasiuk, Hepp, and Zimmermann (called BPHZ renormalization).

The basic idea is to enforce convergence by subtracting additional terms called subtractions.

Let us explain the classical background which can be traced back to Mittag-Leffler (1846–1927). The prototype of the Mittag-Leffler theorem is the Euler formula

$$\pi \cot(\pi z) = \frac{1}{z} + \sum_{k=1}^{\infty} \left(\frac{1}{z+k} - \frac{1}{k}\right) + \sum_{k=-1}^{-\infty} \left(\frac{1}{z+k} - \frac{1}{k}\right) \tag{2.16}$$

for all complex numbers z different from the integers $0, \pm 1, \pm 2, \ldots$ More generally, let $f : \mathbb{C} \to \mathbb{C} \cup \{\infty\}$ be a meromorphic function. By definition, this means that the function f is holomorphic on the complex plane $\mathbb{C}$ up to an at most countable set of isolated poles $p_1, p_2, \ldots$ In the neighborhood of each pole $p = p_j$, we have the Laurent series expansion

$$f(z) = \mathcal{P}_p(z) + \sum_{k=0}^{\infty} a_k (z-p)^k \tag{2.17}$$

with the so-called principal part

$$\mathcal{P}_p(z) := \sum_{k=1}^{n} \frac{a_{-k}}{(z-p)^k}.$$

The formula (2.17) is valid in the largest pointed open disc

$$\{z \in \mathbb{C} : 0 < |z - p\,| < r\}$$

in which the given function f is holomorphic. The complex number a_{-1} is called the residue of the function f at the pole p. For example, the function $\cot(\pi z)$ is meromorphic with poles at the points $0, \pm 1, \pm 2, \ldots$ At the pole $p = 0$, the Laurent series reads as

$$\cot(\pi z) = \frac{1}{\pi z} - \sum_{k=1}^{\infty} \frac{|4^k B_{2k}|}{(2k)!} (\pi z)^{2k-1}, \qquad 0 < |z| < 1$$

with the principal part $\mathcal{P}(z) = \frac{1}{\pi z}$ and the residue $a_{-1} = \frac{1}{\pi}$ at the pole $z = 0$. Let us distinguish the following three cases.

(i) Entire function: If the function f is holomorphic on the complex plane, then it is called an entire function. In this case, we have the power series expansion

$$f(z) = \sum_{k=0}^{\infty} a_k z^k \qquad \text{for all} \quad z \in \mathbb{C}.$$

For example, the functions $f(z) := \mathrm{e}^z, \sin z, \cos z$ are entire functions.

(ii) Finite number of poles. Suppose that the function f has the poles $p_1, \ldots, p_m$. Then there exists an entire function g such that

$$f(z) = g(z) + \sum_{k=1}^{m} \mathcal{P}_{p_k}(z)$$

for all $z \in \mathbb{C} \setminus \{p_1, \ldots, p_m\}$. Here, $\mathcal{P}_{p_1}, \ldots, \mathcal{P}_{p_m}$ are the principal parts of f at the poles.

(iii) Infinite number of poles and subtractions. Suppose now that the function f has the poles $p_1, p_2, \ldots$ We assume that $|p_1| \leq |p_2| \leq \ldots$ Motivated by (i), we expect that

$$f(z) = g(z) + \sum_{k=1}^{\infty} \mathcal{P}_{p_k}(z) \tag{2.18}$$

where g is an entire function. Unfortunately, this is not always true. It may happen that the series $\sum_{k=1}^{\infty} \mathcal{P}_{p_k}(z)$ is not convergent.

> *The trouble with convergence frequently appears in physics when passing from a finite number of degrees of freedom to an infinite number of degrees of freedom.*

For example, in quantum field theory, this concerns the passage from lattice approximations to the continuum of the space-time manifold. It was the idea of Mittag-Leffler to enforce the convergence of (2.18) by using

$$\boxed{f(z) = g(z) + \sum_{k=1}^{\infty} (\mathcal{P}_{p_k}(z) - c_k(z))} \tag{2.19}$$

with suitable polynomials c_k called subtractions.[11] To discuss this, assume first that $p = 0$ is not a pole of the function f. Then $|p_k| > 0$ for all k. Introduce the discs

$$C_k := \{z \in \mathbb{C} : |z| \leq \tfrac{1}{2}|p_k|\}, \qquad k = 1, 2, \ldots$$

By Taylor expansion,

$$\mathcal{P}_{p_k}(z) = b_{k0} + b_{k1} z + \ldots + b_{kn_k} z^{n_k} + \ldots \qquad \text{for all} \quad z \in C_k.$$

Now choose the subtraction as the truncated Taylor expansion, that is,

$$c_k(z) := b_{k0} + b_{k1} z + \ldots + b_{kn_k} z^{n_k}.$$

The crucial Mittag-Leffler theorem tells us the following:

[11] These subtractions are also called counterterms. However, in this volume we will reserve the terminology "counterterm" for additional terms in the Lagrangian.

There exist positive integers $n_1, n_2, \ldots$ and an entire function g such that the relation (2.19) holds for all $z \in \mathbb{C} \setminus \{p_1, p_2, \ldots\}$.

This claim remains true if $p = 0$ is a pole of f. Then we choose $c_1 := 0$. Let us sketch the proof of the Mittag-Leffler theorem. The trick is to choose the number n_k in such a way that

$$|\mathcal{P}_{p_k}(z) - c_k(z)| \le \frac{1}{2^k} \qquad \text{for all} \quad z \in C_k, \quad k = 1, 2, \ldots$$

Then the convergence of the series (2.19) follows from the convergence of $\sum_{k=1}^{\infty} \frac{1}{2^k}$ by applying the majorant criterion (see Problem 2.1).

In specific situations, the general Mittag-Leffler formula (2.19) can be simplified. As an example, let us sketch the proof of the classical formula (2.16). We will use Cauchy's residue method. First introduce the function $f(z) := \pi \cot(\pi z) - \frac{1}{z}$, and choose the positively oriented circles

$$\mathcal{C}_m := \{z \in \mathbb{C} : |z| = m + \tfrac{1}{2}\}, \qquad m = 1, 2, \ldots$$

The function f has the poles $\pm 1, \ldots, \pm m$ inside the circle $\mathcal{C}_m$. Let z be a point inside the circle $\mathcal{C}_m$ which is not a pole of f. By Cauchy's residue theorem,

$$\frac{1}{2\pi \mathrm{i}} \int_{\mathcal{C}_m} \frac{f(\zeta)}{\zeta - z}\, d\zeta = f(z) - \sum_{k=1}^{m} \frac{1}{z+k} - \sum_{k=-1}^{-m} \frac{1}{z+k}.$$

In the special case where $z = 0$, we get

$$\frac{1}{2\pi \mathrm{i}} \int_{\mathcal{C}_m} \frac{f(\zeta)}{\zeta}\, d\zeta = f(0) - \sum_{k=1}^{m} \frac{1}{k} - \sum_{k=-1}^{-m} \frac{1}{k}.$$

Subtracting this, we obtain

$$J_m = f(z) - f(0) - \sum_{k=1}^{m} \left(\frac{1}{z+k} + \frac{1}{k} \right) - \sum_{k=-1}^{-m} \left(\frac{1}{z+k} + \frac{1}{k} \right)$$

with the integral $J_m := \frac{z}{2\pi \mathrm{i}} \int_{\mathcal{C}_m} \frac{f(\zeta)}{\zeta(\zeta - z)}$. Finally, $\lim_{m \to \infty} J_m = 0$. This follows from the fact that the function f is uniformly bounded on the circles $\mathcal{C}_1, \mathcal{C}_2, \ldots$ (see Problem 2.2).

2.1.4 The Square of the Dirac Delta Function

Formal approach. Physicists use the following formal definition

$$\delta(t)^2 := T\delta(t) \qquad \text{for all} \quad t \in \mathbb{R}. \tag{2.20}$$

Here, T is the length of a typical time interval which corresponds to the physical experiment under consideration. Similarly, physicists use

$$\delta(x, y, z)^2 = V\delta(x, y, z) \qquad \text{for all} \quad x, y, z \in \mathbb{R}.$$

Here, V is a typical volume. For scattering processes in particle physics, physicists set $\delta(x, y, z, t) := \delta(x, y, z)\delta(t)$, and they write

$$\delta(x,y,z,t)^2 = VT \cdot \delta(x,y,z,t) \qquad \text{for all} \quad x,y,z,t \in \mathbb{R},$$

where the scattering process takes place in a box of volume V during the time interval $[-\frac{T}{2}, \frac{T}{2}]$. Let us discuss (2.20) in rigorous terms. The remaining formulas can be discussed analogously.

Rigorous discretization of time. Choose $\Delta t > 0$. Define

$$\delta_{\Delta t}(t) := \begin{cases} \frac{1}{\Delta t} & \text{if} \quad t \in [0, \Delta t], \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 2.1 *In the sense of tempered distributions, we have the following two limits:*
(i) $\lim_{\Delta t \to +0} \delta_{\Delta t} = \delta$.
(ii) $\lim_{\Delta t \to +0} \Delta t \cdot (\delta_{\Delta t})^2 = \delta$.

Proof. Ad (i). For all test functions $\varphi \in \mathcal{S}(\mathbb{R})$,

$$\lim_{\Delta t \to +0} \delta_{\Delta t}(\varphi) = \lim_{\Delta t \to +0} \int_{-\infty}^{\infty} \delta_{\Delta t}(t)\varphi(t)dt = \lim_{\Delta t \to +0} \frac{1}{\Delta t} \int_0^{\Delta t} \varphi(t)dt.$$

This is equal to $\varphi(0)$. Finally, recall that $\delta(\varphi) = \varphi(0)$.

Ad (ii). For all $\Delta t > 0$,

$$(\delta_{\Delta t})^2 = \frac{1}{\Delta t} \cdot \delta_{\Delta t}.$$

This can be regarded as a motivation for (2.20) with $T := 1/\Delta t$. Furthermore,

$$\lim_{\Delta t \to +0} \Delta t \int_{-\infty}^{\infty} \delta_{\Delta t}(t)^2 \varphi(t)dt = \lim_{\Delta t \to +0} \int_{-\infty}^{\infty} \delta_{\Delta t}(t)\varphi(t)dt = \varphi(0).$$

□

This idea will be used in Chap. 15 in order to compute cross sections for scattering processes in quantum electrodynamics, by using limits of lattice approximations.

Rigorous extension. We are given a tempered distribution, $F \in \mathcal{S}'(\mathbb{R})$, which has the property that $F(\varphi) = 0$ for all test functions $\varphi \in \mathcal{S}(\mathbb{R})$ with $\varphi(0) = 0$.

Proposition 2.2 *Then there exists a complex number C such that $F = C\delta$.*

Proof. Choose a fixed test function $\chi \in \mathcal{S}(\mathbb{R})$ such that $\chi(0) = 1$. It follows from the decomposition $\varphi = \varphi - \chi\varphi(0) + \chi\varphi(0)$ for all test functions $\varphi \in \mathcal{S}(\mathbb{R})$ that

$$F(\varphi) = F(\varphi - \chi\varphi(0)) + F(\chi\varphi(0)) = F(\chi\varphi(0)) = \varphi(0)F(\chi) = F(\chi)\delta(\varphi).$$

□

This motivates the basic definition

$$\delta^2 := \{C\delta : \quad C \in \mathbb{C}\}.$$

By this definition, δ^2 is not a single distribution, but a one-parameter family of tempered distributions. This idea will be used in renormalization theory, in the Epstein–Glaser approach to quantum field theory. From the physical point of view, the constant C describes an additional degree of freedom which has to be fixed by physical experiment.

Proposition 2.2 is a special case of the Steinmann extension theorem for distributions (see Sect. 11.4.3 of Vol. I).

2.2 Regularization of Divergent Integrals in Baby Renormalization Theory

In order to distinguish between convergent and divergent integrals, we will use the method of power-counting based on a cut-off of the domain of integration. In terms of physics, this cut-off corresponds to the introduction of an upper bound for the admissible energies (resp. momenta). For the regularization of divergent integrals, we will discuss the following methods used by physicists in renormalization theory:

(i) the method of differentiating parameter integrals,
(ii) the method of subtractions (including the famous overlapping divergences),
(iii) Pauli–Villars regularization,
(iv) dimensional regularization by means of Euler's Gamma function,
(v) analytic regularization via integrals of Riemann–Liouville type, and
(vi) distribution-valued analytic functions.

2.2.1 Momentum Cut-off and the Method of Power-Counting

The prototype. Let α be a real number. Then

$$\int_1^\infty \frac{dp}{p^\alpha} := \begin{cases} \frac{1}{\alpha-1} & \text{if} \quad \alpha > 1 \\ +\infty & \text{if} \quad \alpha \leq 1. \end{cases} \tag{2.21}$$

In fact, introducing the positive cut-off $P_{\max}$, we get

$$\int_1^{P_{\max}} \frac{dp}{p^\alpha} = \begin{cases} \frac{P_{\max}^{1-\alpha}-1}{1-\alpha} & \text{if} \quad \alpha \neq 1 \\ \ln P_{\max} & \text{if} \quad \alpha = 1. \end{cases}$$

Finally, note that

$$\int_1^\infty \frac{dp}{p^\alpha} = \lim_{P_{\max}\to+\infty} \int_1^{P_{\max}} \frac{dp}{p^\alpha}.$$

In particular, if $\alpha = 1$, then we say that the integral (2.21) diverges logarithmically. In terms of physics, the variable p can be regarded as momentum, and the cut-off $P_{\max}$ is an upper bound for momentum. The limit $P_{\max} \to +\infty$ is called the ultraviolet limit in physics. In particular, if p is the momentum of a photon, then it has the energy $E = cp$. If this energy is high, then the corresponding light wave looks ultraviolet. Let $f : [0, \infty[\to \mathbb{R}$ be a continuous function with the asymptotic behavior[12]

$$f(p) \simeq \frac{C}{p^\alpha}, \qquad p \to \infty$$

for fixed real exponent α and fixed real nonzero constant C.

Proposition 2.3 *The integral* $\int_1^\infty f(p)dp$ *is convergent if* $\alpha > 1$ *(resp. divergent if* $\alpha \leq 1$*).*

[12] This means that $\lim_{p\to+\infty} f(p)p^\alpha = C$.

The proof follows from (2.21). This proposition tells us that we have to count the power α. For example, let $\alpha > 0$. The function

$$f(p) := \frac{1 + 4p^2}{1 + 2p^{\alpha+2}}, \qquad p \geq 1$$

is asymptotically equal to $\frac{2}{p^\alpha}$ as $p \to +\infty$. Thus, the integral $\int_1^\infty f(p)dp$ is convergent if $\alpha > 1$ (resp. divergent if $\alpha \leq 1$).

Order of divergence. Since

(a) $\int_1^\infty dp = \lim_{P_{\max}\to+\infty} \int_1^{P_{\max}} dp = \lim_{P_{\max}\to+\infty}(P_{\max} - 1) = +\infty$,

(b) $\int_1^\infty p\, dp = \lim_{P_{\max}\to+\infty} \frac{1}{2}(P_{\max}^2 - 1) = +\infty$,

(c) $\int_1^\infty \frac{dp}{p} = \lim_{P_{\max}\to+\infty} \ln P_{\max} = +\infty$,

we say that the integral (a), (b), (c) possesses linear, quadratic and logarithmic divergence, respectively.

Higher dimensions. Let $N = 1, 2, 3, \ldots$ Suppose that the continuous function $f : \mathbb{R}^N \to \mathbb{R}$ has the asymptotic behavior

$$f(p) \simeq \frac{C}{|p|^\alpha}, \qquad |p| \to +\infty$$

with the fixed real exponent α and the fixed real nonzero constant C.

Proposition 2.4 *The integral $\int_{\mathbb{R}^N} f(p)d^N p$ is convergent if $\alpha > N$ (resp. divergent if $\alpha \leq N$).*

This follows from Prop. 2.3 by using spherical coordinates. In fact, setting $r := |p|$, we get

$$\int_{\mathbb{R}^N} f(p)d^N p = \int_0^\infty \int_{|p|=1} f(p) r^{N-1} dr.$$

For example, let $N = 2$ and $\alpha > 0$. Choose

$$f(p) := \frac{1 + 4|p|^2}{1 + 2|p|^{\alpha+2}}, \qquad p \in \mathbb{R}^2$$

where $p = (p_1, p_2)$ and $|p| = \sqrt{p_1^2 + p_2^2}$. Then the integral $\int_{\mathbb{R}^2} f(p)d^2 p$ is convergent if $\alpha > 2$ (resp. divergent if $\alpha \leq 2$).

A sophisticated refinement of the results above is given by Weinberg's power-counting theorem (see Sect. 11.6.3 of Vol. I).

Asymptotic expansions and regularization of divergent integrals by subtraction. Let us consider two simple examples.

(i) First example (one subtraction): We want to regularize the divergent integral

$$J(P) := \int_1^\infty \frac{dp}{p + P} \tag{2.22}$$

with the fixed parameter $P > 0$. In physics, the parameter P describes the momentum of an incoming or outgoing particle of a scattering process. The integrand has the following asymptotic behavior

$$\frac{1}{p + P} \simeq \frac{1}{p}, \qquad p \to +\infty$$

which is responsible for the divergence of the integral by power-counting. In order to regularize the integral, we subtract some function from the integrand which cancels the bad asymptotic behavior as $p \to +\infty$. To this end, we replace $\frac{1}{p+P}$ by

$$\frac{1}{p+P} - \frac{1}{p} = -\frac{P}{p(p+P)}.$$

Since this is asymptotically equal to $-\frac{P}{p^2}$ as $p \to +\infty$, the regularized integral

$$J_{\rm reg}(P) := \int_1^\infty \left(\frac{1}{p+P} - \frac{1}{p}\right) dp$$

is convergent by power-counting. Mnemonically, we write

$$\boxed{J_{\rm reg}(P) = J(P) - J(0).} \tag{2.23}$$

However, this is not a rigorous relation, since $J(P) = +\infty$ and $J(0) = +\infty$.

(ii) Second example (two successive subtractions). The integral

$$J(P) = \int_1^\infty \frac{pdp}{p+P}, \qquad P > 0$$

is divergent. Since $\frac{p}{p+P} \simeq 1$ as $p \to +\infty$, we pass to the new integrand

$$g(p) := \frac{p}{p+P} - 1 = -\frac{P}{p+P}$$

by subtraction. However, since $g(p) \simeq -\frac{P}{p}$ as $p \to +\infty$, the integral $\int_1^\infty g(p)dp$ is still divergent. Continuing the subtraction method, we pass to the integrand

$$g(p) - \left(-\frac{P}{p}\right) = \frac{P^2}{(p+P)p}.$$

This leads us to the definition of the regularized integral

$$J_{\rm reg}(P) := \int_1^\infty \frac{P^2 dp}{(p+P)p}, \qquad P > 0$$

which is convergent by power-counting. Using analytic continuation, the function $P \mapsto J_{\rm reg}(P)$ makes sense for all complex numbers P in the subset $\mathbb{C}\backslash]-\infty, -1]$ of the complex plane. Mnemonically,

$$\boxed{J_{\rm reg}(P) = J(P) - J(0) - PJ'(0).} \tag{2.24}$$

In fact, this stands for the expression

$$J_{\rm reg}(P) = \int_1^\infty \left(\frac{p}{p+P} - 1 + \frac{P}{p}\right) dp.$$

2.2.2 The Choice of the Normalization Momentum

The divergent integral $J(P) := \int_1^\infty \frac{dp}{p+P}$ with $P > 0$ considered above can also be regularized by fixing the parameter $P_0 > 0$ and replacing (2.23) by the mnemonic relation

$$\boxed{\mathcal{J}_{\rm reg}(P) = J(P) - J(P_0).}$$

Rigorously, we define

$$\mathcal{J}_{\rm reg}(P) := \int_1^\infty \left(\frac{1}{p+P} - \frac{1}{p+P_0}\right) dp = (P_0 - P) \int_1^\infty \frac{dp}{(p+P)(p+P_0)}.$$

This integral is convergent by power-counting. Note that the regularized integral $\mathcal{J}_{\rm reg}(P)$ depends on the choice of the momentum P_0, which is called the normalization momentum.

> *In terms of physics, the choice of P_0 can be determined by the typical momentum of the physical experiment under consideration.*

For example, in particle accelerators the scattered particles have a fixed momentum, and hence a fixed energy. It turns out that the physics changes if the particle energy changes. This is the phenomenon of the running coupling constant in the Standard Model of particle physics.

> *Summarizing, regularization (and hence renormalization) methods produce additional parameters which have to be determined by physical experiments.*

The role of the renormalization group in physics. The change of the normalizing momentum is governed by a transformation which is described by the so-called renormalization group. In particular, this is crucial for studying the high-energy behavior of quantum field theories (see Chap. 3 of Vol. I on scale changes in physics). For example, quarks behave like free particles at very high energies. This is the so-called asymptotic freedom of quarks.[13]

Off-shell versus on-shell normalization. Recall that the 4-momentum $p = (p^0, p^1, p^2, p^3)$ of an elementary particle (in an inertial system) corresponds to the momentum vector $\mathbf{p} = p^1\mathbf{i} + p^2\mathbf{j} + p^3\mathbf{k}$ and the energy $E = cp^0$ where c denotes the velocity of light in a vacuum. If the particle has the rest mass m_0, then

$$\boxed{|p^0|^2 - |p^1|^2 - |p^2|^2 - |p^3|^2 = c^2 m_0^2.} \tag{2.25}$$

This corresponds to the energy relation $E^2 = (m_0c^2)^2 + c^2\mathbf{p}^2$. By definition, the 4-momentum p is on-shell (resp. off-shell) iff the mass-shell relation (2.25) is satisfied (resp. not satisfied).

2.2.3 The Method of Differentiating Parameter Integrals

> Improve the convergence behavior of a parameter-depending integral by differentiation with respect to the parameter.
>
> Folklore

[13] For the discovery of the asymptotic freedom of quarks, David Gross, David Politzer and Frank Wilczek were awarded the Nobel prize in physics in 2004.

Consider again the divergent integral

$$J(P) = \int_1^\infty \frac{dp}{p+P}, \qquad P > 0.$$

Formal differentiation with respect to the parameter P yields the convergent integral

$$J'(P) = -\int_1^\infty \frac{dp}{(p+P)^2} = -\frac{1}{1+P}, \qquad P > 0.$$

The differential equation

$$\mathsf{J}'(P) = -\frac{1}{1+P}, \qquad P > 0$$

has the general solution $\mathsf{J}(P) = C - \ln(1+P)$, with the arbitrary real constant C. Therefore, we define the regularized integral by setting

$$\boxed{\mathsf{J}_{\rm reg}(P) := C - \ln(1+P), \qquad P > 0.}$$

Note that the function J represents a *family* of functions which depends on the real parameter C.

In the general case, consider the divergent integral

$$J(P) = \int_1^\infty f(p,P)dp, \qquad P \in \Pi$$

where Π is an open interval. Let $n = 1, 2, \ldots$ The formal n-th derivative with respect to the parameter P looks like

$$J^{(n)}(P) = \int_1^\infty \frac{\partial^n f(p,P)}{\partial P^n}\, dp, \qquad P \in \Pi.$$

Suppose that the integrals are divergent for $n = 1, 2, \ldots, N-1$, whereas the integral $J^{N)}(P)$ is convergent for all $P \in \Pi$. Then we define the regularized integral $\mathsf{J}_{\rm reg}(P)$ by means of the differential equation

$$\mathsf{J}_{\rm reg}^{(N)}(P) = J^{(N)}(P), \qquad P \in \Pi. \tag{2.26}$$

Let $\mathsf{J} = \mathsf{J}(P)$ be a special solution of this differential equation. Then we use the general solution of (2.26) in order to define

$$\boxed{\mathsf{J}_{\rm reg}(P) := C_0 + C_1 P + \ldots + C_{N-1}P^{N-1} + \mathsf{J}(P), \qquad P \in \Pi.}$$

Here, $C_0, \ldots, C_{N-1}$ are arbitrary real parameters. In physics, these free parameters have to be determined by experiment.

2.2.4 The Method of Taylor Subtraction

Let us generalize (2.23) and (2.24). To this end, consider the divergent integral

$$J(P) = \int_1^\infty f(p,P)\, dp, \qquad P \in \Pi.$$

Choose the normalization point P_0 in the open interval Π. Let $n = 0, 1, 2, \ldots$ Mnemonically, we set

$$\boxed{\mathcal{J}_{\text{reg}}P) = J(P) - J(P_0) - (P - P_0)J'(P_0) - \ldots - \frac{(P - P_0)^n}{n!}J^{(n)}(P_0).}$$

Here, we assume that, for fixed n, the corresponding integral is convergent for all parameters $P \in \Pi$ and that n is the smallest number with this property. This is called the method of minimal (Taylor) subtraction. In terms of integrands, we get

$$\mathcal{J}_{\text{reg}}(P) = \int_1^\infty \left(f(p, P) - \sum_{k=0}^{n} \frac{(P - P_0)^k}{k!} \frac{\partial^k f(p, 0)}{\partial P^k} \right) dp.$$

2.2.5 Overlapping Divergences

> Overlapping divergences caused a lot of trouble in the history of renormalization theory.[14]
>
> Folklore

As a prototype for an integral with overlapping divergences, consider the divergent integral

$$J(P) = \int_1^\infty \int_1^\infty f(p, q, P)dpdq, \qquad P > 0$$

with the integrand

$$f(p, q, P) := \frac{p}{p + P} \cdot \frac{1}{p + q} \cdot \frac{q}{q + P}.$$

Fix the parameter $P > 0$. The point is that

- $\int_1^\infty f(p, q, P)dp = +\infty$ for all $q \geq 1$,
- and $\int_1^\infty f(p, q, P)dq = +\infty$ for all $p \geq 1$

by power-counting. This means that the integral is divergent if either of the integration variables p and q is fixed. We call this an overlapping divergence. In order to regularize the integral $J(P)$, we will use the following subtractions with respect to the parameter P.

(i) Minimal subtraction concerning the one-dimensional p-integral.
(ii) Minimal subtraction concerning the one-dimensional q-integral.
(iii) Minimal subtraction of the result from (i) and (ii) concerning the two-dimensional (p, q)-integral.

Ad (i). For fixed $q \geq 1$, we have $f(p, q, P) \simeq \frac{q}{q+P} \cdot \frac{1}{p}$ as $p \to +\infty$. Therefore, we replace $f(p, q, P)$ by

$$f(p, q, P) - \frac{q}{q + P} \cdot \frac{1}{p}.$$

Ad (ii). For fixed $p \geq 1$, we have $f(p, q, P) \simeq \frac{p}{p+P} \cdot \frac{1}{q}$ as $q \to +\infty$. Therefore, we pass over to

$$g(p, q, P) := f(p, q, P) - \frac{q}{(q + P)p} - \frac{p}{(p + P)q}.$$

[14] See the quotation on page 974.

This yields the integral

$$G(P) = \int_1^\infty \int_1^\infty g(p,q,P)dqdp, \qquad P > 0$$

with the integrand

$$g(p,q,P) = -\frac{p^2q^2 + pq^3 + p^3q + P(p^3 + p^2q + pq^2 + q^3)}{p(p+P)(p+q)(q+P)q}.$$

Power-counting shows that the integral $G(P)$ is still divergent. To show this, note that for all $p, q \geq 1$, we get

$$\frac{p^2q^2}{p(p+P)(p+q)(q+P)q} \geq \frac{pq}{(p+P)\cdot 2\cdot (q+P)}.$$

Furthermore, by power-counting,

$$\int_1^\infty \int_1^\infty \frac{pq\ dpdq}{2(p+P)(q+P)} = \int_1^\infty \frac{p\ dp}{2(p+P)} \int_1^\infty \frac{q\ dq}{(q+P)} = +\infty.$$

Ad (iii). Mnemonically, we define $J_{\rm reg}(P) := G(P) - G(0) - PG'(0)$ with $P > 0$. Rigorously, this gives the definition of the regularized integral:

$$J_{\rm reg}(P) := \int_1^\infty \int_1^\infty g_{\rm reg}(p,q,P)dpdq.$$

Here, the regularized integrand $g_{\rm reg}(p,q,P)$ is equal to

$$g(p,q,P) - g(p,q,0) - P\frac{\partial g(p,q,0)}{\partial P} = -\frac{P^2pq + P^3(p+q)}{p(p+P)(p+q)(q+P)q}.$$

Proposition 2.5 *The regularized integral $J_{\rm reg}(P)$is convergent.*

Proof. For example, consider the subintegral

$$A := \int_1^\infty \int_1^\infty \frac{pq\ dpdq}{p(p+P)(p+q)(q+P)q}.$$

From $(\sqrt{p} - \sqrt{q})^2 \geq 0$, we get $\sqrt{p}\sqrt{q} \leq \frac{1}{2}(p+q)$ for all $p, q \geq 0$. Hence

$$\frac{1}{p+q} \leq \frac{1}{2\sqrt{p}\sqrt{q}} \qquad \text{for all} \quad p, q > 0.$$

This implies

$$A \leq \frac{1}{2}\int_1^\infty \frac{dp}{(p+P)\sqrt{p}} \quad \int_1^\infty \frac{dq}{(q+P)\sqrt{q}}\ dq < \infty$$

by power-counting. Analogously, we obtain the convergence of the remaining subintegrals. □

The relation of this subtraction procedure and Zimmermann's forest formula will be discussed in Example 2 on page 985.

2.2.6 The Role of Counterterms

The method of subtractions changes the integrands by subtracting regularizing terms in order to enforce convergence of the integrals. Physicists try to connect the subtraction terms with additional terms in the Lagrangian of the quantum field theory under consideration. These additional terms of the Lagrangian are called counterterms. It is one of the tasks of renormalization theory

- to study the structure of the necessary subtraction terms and
- to show that the subtraction terms can be generated by appropriate counterterms of the classical Lagrangian (see Chap. 16).

The philosophy behind this approach is that the procedure of quantization adds quantum effects to the classical field theory. These quantum effects can be described by changing the classical Lagrangian by adding counterterms.

2.2.7 Euler's Gamma Function

Let us summarize some classical formulas which will be used for the dimensional regularization of divergent multi-dimensional integrals below. Recall that the factorial symbol $n!$ stands for the product $1 \cdot 2 \cdots n$. For example $3! = 6$. By convention, $0! := 1$. Hence

$$n! = n \cdot (n-1)!, \qquad n = 1, 2, \ldots$$

In 1729 it was Euler's aim to extrapolate the discrete values $n!$ to a smooth function $\Gamma :]0, \infty[\to \mathbb{R}$ with the property $\Gamma(1) = 1$ and

$$\Gamma(x+1) = x\Gamma(x) \qquad \text{for all} \quad x > 0.$$

Euler constructed such a function by means of the convergent integral

$$\Gamma(x) := \int_0^\infty t^{x-1} \mathrm{e}^{-t}\, dt \qquad \text{for all} \quad x > 0.$$

In 1811 Gauss showed that this function can be extended to a meromorphic function $\Gamma : \mathbb{C} \to \mathbb{C} \cup \{\infty\}$. He used the infinite product

$$\Gamma(z) = \lim_{n\to\infty} \frac{n! n^z}{z(z+1)\cdots(z+n)} \qquad \text{for all} \quad z \in \mathbb{C} \setminus \{0, -1, -2, \ldots\}.$$

The gamma function has the following properties:[15]

- $\Gamma(1) = 1$ and $\Gamma(z+1) = z\Gamma(z)$ for all $z \in \mathbb{C} \setminus \{0, -1, -2, \ldots\}$ (functional equation). Hence $\Gamma(n) = (n-1)!$ if $n = 1, 2, \ldots$ Explicitly,

$$\Gamma(z) = \frac{1}{z} + O(1), \qquad z \to 0. \tag{2.27}$$

[15] The proofs can be found in R. Remmert, Classical Topics in Complex Function Theory, Chap. 2, Springer, New York, 1998.

- More generally, the gamma function is a meromorphic function on the complex plane. It has precisely the poles $0, -1, -2, \ldots$. Each pole is simple and has the residue
$$\operatorname{res}_{-k}\Gamma = \frac{(-1)^k}{k!}, \qquad k = 0, 1, 2, \ldots$$
Therefore,
$$\Gamma(z-k) = \frac{(-1)^k}{k!(z+k)} + O(1), \qquad z \to 0, \quad k = 0, 1, 2, \ldots$$
We define the regularized value by
$$\Gamma(-k)_{\mathrm{reg}} := \lim_{z\to 0}\left(\Gamma(z-k) - \frac{(-1)^k}{k!(z+k)}\right), \qquad k = 0, 1, 2, \ldots$$
These limits are well defined.
- More precisely, in a neighborhood of the origin we have the Laurent series expansion
$$\boxed{\Gamma(z) = \frac{1}{z} - \gamma + \left(\frac{\gamma^2}{2} + \frac{\pi^2}{12}\right) z + O(z^2), \qquad z \to 0} \tag{2.28}$$
where $\gamma = 0.577\ldots$ is the Euler constant. Truncating the singularity at the point $z = 0$ and letting $z \to 0$, we obtain $\Gamma_{\mathrm{reg}}(0) = -\gamma$.
- In a neighborhood of the pole $z = -1$, we have the Laurent series expansion[16]
$$\Gamma(z-1) = -\frac{1}{z} + \gamma - 1 - \left(\frac{\gamma^2}{2} + \frac{\pi^2}{12} + \gamma\right) z + O(z^2), \tag{2.29}$$
as $z \to 0$. Therefore, $\Gamma_{\mathrm{reg}}(-1) = \gamma - 1$.
- The gamma function has no zeros. The reciprocal function $1/\Gamma$ is an entire function with the Weierstrass product representation
$$\frac{1}{\Gamma(z)} = z\mathrm{e}^{\gamma z}\prod_{k=1}^{\infty}\left(1 + \frac{z}{k}\right)\mathrm{e}^{-z/k} \qquad \text{for all} \quad z \in \mathbb{C}.$$
- $\Gamma(z) := \int_0^\infty t^{z-1}\mathrm{e}^{-t}dt$ for all complex numbers z with $\Re(z) > 0$.
- Euler's supplement: For all complex numbers z which are not integers, we have
$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin \pi z},$$
and $\Gamma(z)\Gamma(-z) = -\frac{\pi}{z \sin \pi z}$. In particular, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.
- Legendre's duplication formula:
$$\Gamma(2z) = \frac{2^{2z-1}}{\sqrt{\pi}}\,\Gamma(z)\Gamma(z + \tfrac{1}{2}).$$
This is true for all complex numbers z different from $0, -\frac{1}{2}, -1, -\frac{3}{2}, \ldots$
- Special values: $\Gamma(n + \frac{1}{2}) = \frac{(2n)!}{4^n n!}\sqrt{\pi}$ for all $n = 0, 1, 2, \ldots$

[16] For the proof, see Problem 2.11.

- The asymptotic Stirling formula for real arguments x:

$$\Gamma(x+1) \simeq \sqrt{2\pi x}\left(\frac{x}{e}\right)^x, \qquad x \to +\infty.$$

The gamma function is uniquely determined by its functional equation. More precisely, Wielandt proved the following theorem in 1939: Consider the half-plane $\mathcal{H} := \{z \in \mathbb{C} : \Re(z) > 0\}$. Suppose that the function $F : \mathcal{H} \to \mathbb{C}$ is holomorphic and satisfies the functional equation

$$F(1) = 1, \quad F(z+1) = zF(z) \qquad \text{for all} \quad z \in \mathcal{H}.$$

Moreover, let $|F(z)|$ be bounded on the strip $\{z \in \mathbb{C} : 1 \leq \Re(z) < 2\}$. Then $F = \Gamma$ on $\mathcal{H}$.

Let p and q be complex numbers with positive real part, $\Re(p), \Re(q) > 0$. Then the integral

$$B(p,q) := \int_0^1 x^{p-1}(1-x)^{q-1}\, dx \tag{2.30}$$

exists, and we have the famous Euler identity

$$\boxed{B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.} \tag{2.31}$$

The function B is called the Euler beta function.

Finally, we want to show that the gamma function allows us to compute the surface measure of the unit sphere $\mathbb{S}^{N-1}$ in $\mathbb{R}^N$. To fix the terminology, choose the dimension $N = 2, 3, \ldots$. For each point $p \in \mathbb{R}^N$, we set $|p| := \sqrt{(p^1)^2 + \ldots + (p^N)^2}$. Define the $(N-1)$-dimensional unit sphere $\mathbb{S}^{N-1} := \{p \in \mathbb{R}^N : |p| = 1\}$. Then the surface measure of $\mathbb{S}^{N-1}$ is given by the Jacobi formula

$$\boxed{\text{meas}(\mathbb{S}^{N-1}) = \frac{2\pi^{\frac{N}{2}}}{\Gamma\left(\frac{N}{2}\right)}.} \tag{2.32}$$

For example, we get $\text{meas}(\mathbb{S}^1) = 2\pi$ and $\text{meas}(\mathbb{S}^2) = 4\pi$ for the length of the unit circle and the surface measure of the 2-dimensional unit sphere, respectively. An elegant proof will be given in Problem 2.3. Moreover, for the volume of the N-dimensional ball $\mathbb{B}_R^N := \{x \in \mathbb{R}^N : |x| \leq R\}$ of radius R, we get

$$V(\mathbb{B}_R^N) = \text{meas}(\mathbb{S}^{N-1})\frac{R^N}{N}. \tag{2.33}$$

In fact, we have $V(\mathbb{B}_R^N) = \int_0^R \left(\int_{\mathbb{S}^{N-1}} dS\right) r^{N-1} dr$.

2.2.8 Integration Tricks

Let us summarize some integration tricks which are frequently applied in quantum field theory.

The sphere trick. Let $f : [0, \infty[\to \mathbb{C}$ be a continuous function. Fix the dimension $N = 2, 3, \ldots$. Moreover, suppose that $\sup_{0 \leq r < \infty} r^{N-1+\alpha}|f(r)| < \infty$ for fixed $\alpha > 1$. Then the following integral is absolutely convergent:

$$\boxed{\int_{\mathbb{R}^N} f(|p|)\, d^N p = \frac{2\pi^{\frac{N}{2}}}{\Gamma\left(\frac{N}{2}\right)} \int_0^\infty f(r) r^{N-1} dr.} \tag{2.34}$$

To prove this, set $r := |p|$, use spherical coordinates, and note that

$$\int_{\mathbb{R}^N} f(|p|)\, d^N p = \text{meas}(\mathbb{S}^{N-1}) \int_0^\infty f(r) r^{N-1} dr.$$

The Cauchy residue trick. We have

$$\boxed{\int_{-\infty}^{\infty} \frac{f(x)dx}{g(x)} = 2\pi \mathrm{i} \sum_{k=1}^{K} \text{res}_{z_k} \left(\frac{f}{g}\right).} \tag{2.35}$$

Here, we assume that f and g are polynomials such that the degree condition degree(g)-degree(f)≥ 2 is satisfied, and the function $\frac{f}{g}$ has no poles on the real axis. In equation (2.35), we sum over all poles $z_1, \ldots, z_K$ of the function $\frac{f}{g}$ on the upper half-plane. If there is no such pole, then the integral from (2.35) is equal to zero. For computations, it is convenient to know that the following hold: If the function $\frac{f}{g}$ has a pole of order $n = 1, 2, \ldots$ at the point z_k, then the residue of this pole is given by

$$\text{res}_{z_k} \left(\frac{f}{g}\right) = \lim_{z \to z_k} \frac{1}{(n-1)!} \frac{d^{n-1}}{dz^{n-1}} \left(\frac{(z-z_k)^n f(z)}{g(z)}\right).$$

The proof of (2.35) proceeds as in Problem 12.1 of Vol. I.

(i) First example. Let $a > 0$ and $\varepsilon > 0$. We want to show that

$$\int_{-\infty}^{\infty} \frac{dp}{p^2 - (a + \mathrm{i}\varepsilon)^2} = \frac{\pi \mathrm{i}}{a + \mathrm{i}\varepsilon}.$$

In fact, the equation $p^2 - (a + \mathrm{i}\varepsilon)^2 = 0$ has the simple zero $p_+ := a + \mathrm{i}\varepsilon$ on the upper half-plane, and the simple zero $p_- := -a - \mathrm{i}\varepsilon$ on the lower half-plane. Hence

$$\text{res}_{p_+} \frac{1}{p^2 - (a + \mathrm{i}\varepsilon)^2} = \lim_{p \to p_+} \frac{p - p_+}{(p - p_+)(p - p_-)} = \frac{1}{2p_+}.$$

This yields the claim. Similarly, we get

$$\int_{-\infty}^{\infty} \frac{dp}{p^2 - (a - \mathrm{i}\varepsilon)^2} = -\frac{\pi \mathrm{i}}{a - \mathrm{i}\varepsilon}.$$

This implies the limit

$$\lim_{\varepsilon \to +0} \int_{-\infty}^{\infty} \frac{dp}{p^2 - (a \pm \mathrm{i}\varepsilon)^2} = \pm \frac{\pi \mathrm{i}}{a}.$$

(ii) Second example. Analogously, we obtain

$$\int_{-\infty}^{\infty} \frac{dp}{p^2 + (a \pm \mathrm{i}\varepsilon)^2} = \frac{\pi}{a \pm \mathrm{i}\varepsilon}.$$

The Wick rotation trick. In addition to the assumptions formulated for (2.35) above, assume that the function $\frac{f}{g}$ has no poles on the closed first quadrant

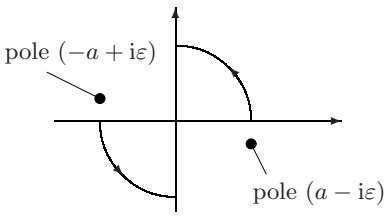


Fig. 2.1. Wick rotation

$\{x+\mathrm{i}y : x,y \geq 0\}$ and on the close third quadrant $\{x+\mathrm{i}y : x,y \leq 0\}$. Then the integral (2.35) is invariant under the transformation $x \mapsto \mathrm{i}x$. That is,

$$\boxed{\int_{-\infty}^{\infty} \frac{f(x)dx}{g(x)} = \int_{-\infty}^{\infty} \frac{f(\mathrm{i}x)\ \mathrm{i}dx}{g(\mathrm{i}x)}.}$$

Proof. Use the counter-clockwise rotation pictured in Fig. 2.1. Here, the real axis is rotated into the imaginary axis. By assumption, the rotating real axis does not cross poles of $\frac{f}{g}$. Hence

$$\int_{-\infty}^{\infty} \frac{f(x)dx}{g(x)} = \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{f(x)dx}{g(x)},$$

by the Cauchy theorem on the path-independence of integrals over holomorphic functions. Finally, set $x = \mathrm{i}y$ and replace the real variable y by x. □

For example, let $a > 0$ and $\varepsilon > 0$. Then

$$\int_{\mathbb{R}} \frac{dp}{p^2-(a-\mathrm{i}\varepsilon)^2} = \int_{\mathbb{R}} \frac{\mathrm{i}dp}{-p^2-(a-\mathrm{i}\varepsilon)^2} = -\mathrm{i}\int_{\mathbb{R}} \frac{dp}{p^2+(a-\mathrm{i}\varepsilon)^2} = -\frac{\pi\mathrm{i}}{a-\mathrm{i}\varepsilon}.$$

This coincides with the result obtained above. Physicists use this trick in order to reduce algebraic Feynman integrals on Minkowski space to the corresponding integrals on Euclidean space (see Sect. 2.2.12 on page 80). The goal of the following three tricks due to Feynman, Schwinger and Pauli–Villars is to simplify the structure of algebraic Feynman integrals.

The Feynman integration trick. For all nonzero real numbers a, b, we have the identity

$$\boxed{\frac{1}{ab} = \int_0^1 \frac{dx}{(ax+b(1-x))^2},} \tag{2.36}$$

by elementary integration. Furthermore, for all nonzero real numbers a, b, c, d, we obtain

$$\frac{1}{abc} = 2\int_0^1 dx \int_0^x \frac{dy}{(ay+b(x-y)+c(1-x))^3} \tag{2.37}$$

and

$$\frac{1}{abcd} = 6\int_0^1 dx \int_0^x dy \int_0^y \frac{dz}{(az+b(y-z)+c(x-y)+d(1-x))^4}.$$

The general case reads as follows.

Proposition 2.6 *Let $n = 2, 3, \ldots$ For all nonzero real numbers $a_1, \ldots, a_n$, we have*

$$\frac{1}{a_1 a_2 \cdots a_n} = (n-1)! \int_0^1 dx_1 \int_0^{x_1} dx_2 \cdots \int_0^{x_{n-2}} \frac{dx_{n-1}}{f(x_1, \ldots, x_{n-1})^n}$$

where $f(x_1, \ldots, x_{n-1}) := a_1 x_{n-1} + \sum_{k=2}^{n-1} a_k (x_{n-k} - x_{n-k+1}) + a_n (1 - x_1)$.

Proof. We will proceed by induction.

(I) The claim is true for $n = 2$.

(II) Suppose that the claim is true for the integer n where $n \geq 2$. Set

$$J(a_1, \ldots, a_{n+1}) := \int_0^1 dx_1 \int_0^{x_1} \cdots \int_0^{x_{n-1}} \frac{dx_n}{f(x_1, \ldots, x_{n+1})^{n+1}}.$$

Elementary integration yields

$$J(a_1, \ldots, a_{n+1}) = \frac{1}{a_1 - a_2} (J(a_2, a_3, a_4, \ldots, a_{n+1}) - J(a_1, a_3, a_4, \ldots, a_{n+1})).$$

Hence

$$J(a_1, \ldots, a_{n+1}) = \frac{1}{a_1 - a_2} \left(\frac{1}{a_2 a_3 \cdots a_{n+1}} - \frac{1}{a_1 a_3 \cdots a_{n+1}} \right) = \frac{1}{a_1 a_2 \cdots a_{n+1}}.$$

Thus, the claim is true for the integer $n + 1$. □

Differentiating the relation (2.36) with respect to the variable a, we obtain

$$\frac{1}{a^2 b} = 2 \int_0^1 \frac{x dx}{(ax + b(1-x))^3} \tag{2.38}$$

and

$$\frac{1}{a^3 b} = 3 \int_0^1 \frac{x^2 dx}{(ax + b(1-x))^4}. \tag{2.39}$$

These identities can be elegantly reformulated in terms of the Dirac delta function. Explicitly, we obtain

$$\boxed{\frac{1}{ab} = \int_0^1 \int_0^1 \frac{\delta(x + y - 1)\, dxdy}{(ax + by)^2}} \tag{2.40}$$

for all nonzero real numbers a, b.[17] For $n = 2, 3, \ldots$ and all nonzero real numbers $a_1, \ldots, a_n$, the general formula reads as

$$\frac{1}{a_1 a_2 \cdots a_n} = \int_0^1 \cdots \int_0^1 \frac{\delta(u_1 + u_2 + \ldots + u_n - 1)\, du_1 du_2 \cdots du_n}{(a_1 u_1 + a_2 u_2 + \ldots + a_n u_n)^n}.$$

This follows from Prop. 2.6 after integrating over u_n and substituting

[17] In fact, integrating over the variable y, we get

$$\int_0^1 \int_0^1 \frac{\delta(x + y - 1)\, dxdy}{(ax + by)^2} = \int_0^1 \frac{dx}{(ax + b(1-x))^2} = \frac{1}{ab}.$$

$$u_1 = x_{n-1}, \quad u_2 = x_{n-2} - x_{n-1}, \ldots, \quad u_{n-1} = x_1 - x_2.$$

Note that $u_1 + u_2 + \ldots + u_{n-1} = x_1$.

The Schwinger integration trick. For all real numbers $x > 0$ and all complex numbers z with $\Re(z) > 0$, we have

$$\boxed{\frac{1}{x^z} = \frac{1}{\Gamma(z)} \int_0^\infty y^{z-1} \mathrm{e}^{-xy} dy.} \tag{2.41}$$

This follows from applying the substitution $t = xy$ to $\Gamma(z) = \int_0^\infty t^{z-1} \mathrm{e}^{-t} dt$. Physicists call this the Schwinger integration trick.

The Pauli–Villars integration trick. For all real numbers P, m, M with $P \neq m^2$ and $P \neq M^2$, we have

$$\frac{1}{P - m^2} - \frac{1}{P - M^2} = \int_{M^2}^{m^2} \frac{du}{(u - P)^2}. \tag{2.42}$$

The Wick differentiation trick. Let $a > 0$ and $J \in \mathbb{R}$. Then

$$\boxed{\int_{-\infty}^{\infty} x \mathrm{e}^{-ax^2} dx = \frac{d}{dJ} \left(\int_{-\infty}^{\infty} \mathrm{e}^{-ax^2} \mathrm{e}^{Jx} dx \right)_{|J=0}.} \tag{2.43}$$

This is used in order to compute moments (e.g., correlation functions) in quantum field theory). In physics, J corresponds to an external source.

2.2.9 Dimensional Regularization via Analytic Continuation

Consider the integral

$$\int_0^R \frac{dr}{\sqrt{1 + r^2}} = \ln \left(\sqrt{1 + r^2} + r \right) \Big|_0^R = \ln \left(\sqrt{1 + R^2} + R \right).$$

Consequently, we get

$$\int_0^\infty \frac{dr}{\sqrt{1 + r^2}} = \lim_{R \to +\infty} \ln \left(\sqrt{1 + R^2} + R \right) = \infty. \tag{2.44}$$

That is, the integral is divergent of logarithmic type. Such slowly divergent integrals frequently occur in renormalization theory. In order to regularize the integral (2.44), we will use the following key formula[18]

$$\boxed{\int_0^\infty \frac{r^\alpha dr}{(a + br^\beta)^\gamma} = \left(\frac{a}{b} \right)^{\frac{\alpha+1}{\beta}} \frac{\Gamma\left(\frac{\alpha+1}{\beta} \right) \Gamma\left(\gamma - \frac{\alpha+1}{\beta} \right)}{a^\gamma \beta \Gamma(\gamma)}.} \tag{2.45}$$

[18] This formula follows from (2.30) and (2.31) by using a simple substitution (see Problem 2.4). Note that the two integrals

$$\int_0^1 r^\alpha dr \quad \text{and} \quad \int_1^\infty r^\beta dr$$

are finite if we have $\alpha > -1$ and $\beta < -1$, respectively.

In the classical sense, this so-called Liouville integral exists for all parameters

$$a, b, \beta, \gamma > 0, \quad \beta\gamma - \alpha > 1.$$

Using analytic continuation and the regularization of the gamma function, the integral (2.45) can be easily extended to a larger parameter domain. For example, we obtain the following two regularized integrals

$$\left[\int_0^\infty \frac{dr}{\sqrt{1+r^2}}\right]_{\text{reg}} := -\frac{\gamma}{2}, \tag{2.46}$$

and

$$\left[\int_0^\infty (1+r^2)^{\frac{1}{4}} dr\right]_{\text{reg}} := \frac{\Gamma(\frac{1}{2})\Gamma(-\frac{3}{4})}{2\Gamma(-\frac{1}{4})}. \tag{2.47}$$

In order to motivate the definition (2.46), choose the parameter $\varepsilon > 0$. Then we have the classical integral

$$\int_0^\infty \frac{dr}{(1+r^2)^{\frac{1}{2}+\varepsilon}} = \frac{\Gamma(\frac{1}{2})\Gamma(\varepsilon)}{2\Gamma(\frac{1}{2}+\varepsilon)}.$$

By (2.28), this is equal to

$$\frac{\Gamma(\frac{1}{2})}{2\Gamma(\frac{1}{2}+\varepsilon)}\left(\frac{1}{\varepsilon} - \gamma + O(\varepsilon)\right), \qquad \varepsilon \to 0.$$

Truncating the singular term $\frac{1}{\varepsilon}$ and letting $\varepsilon \to 0$, we get (2.46). Furthermore, the definition (2.47) is motivated by using the key equation (2.45) with the parameter values $a = b := 1, \alpha := 0, \beta := 2$, and $\gamma := -\frac{1}{4}$. Note that the values $\Gamma(\frac{1}{2}), \Gamma(-\frac{3}{4})$ and $\Gamma(-\frac{1}{4})$ are well-defined, since the meromorphic gamma function has only poles at the points $0, -1, -2, \ldots$

Dimensional regularization of integrals. Let $N = 2, 3, \ldots$ We want to study the integral

$$J_N(f) := \int_{\mathbb{R}^N} f(|p|)\, d^N p.$$

By the sphere trick (2.34) on page 70, we have the key formula

$$\boxed{J_N(f) = \frac{2\pi^{\frac{N}{2}}}{\Gamma\left(\frac{N}{2}\right)} \int_0^\infty f(r) r^{N-1} dr.} \tag{2.48}$$

The point is that the right-hand side of (2.48) is not only defined for the integer dimensions $N = 2, 3, \ldots$, but also for more general values of the parameter N. Observe that the interpolating gamma function is uniquely determined by Wielandt's uniqueness theorem considered above. Now the method of dimensional regularization proceeds as follows. We want to regularize the divergent integral $J_{N_0}(f)$ for fixed $N_0 = 2, 3, \ldots$

- To this end, we choose the parameter $N = N_0 - \varepsilon$, and we compute $J_{N_0-\varepsilon}(f)$ by (2.48).
- We truncate the singularity of the function $\varepsilon \mapsto J_{N_0-\varepsilon}$ at the point $\varepsilon = 0$. This way, we get the regularized expression $[J_{N_0-\varepsilon}(f)]_{\text{reg}}$.

- Finally, we define the regularized integral

$$\boxed{[J_{N_0}(f)]_{\mathrm{reg}} := \lim_{\varepsilon \to +0} [J_{N_0-\varepsilon}(f)]_{\mathrm{reg}}.}$$

Let us illustrate this with the integral

$$J_N := \int_{\mathbb{R}^N} \frac{d^N p}{(a^2 + |p|^2)^2}, \qquad N = 2, 3, 4$$

for all parameters $a > 0$. This integral behaves like $\int_1^\infty r^{N-5} dr$. Consequently, if $N = 2, 3$ (resp. $N = 4$), the integral is convergent (resp. divergent). We want to motivate the regularization

$$\left[\int_{\mathbb{R}^4} \frac{d^4 p}{(a^2 + |p|^2)^2} \right]_{\mathrm{reg}} := -\pi^2 \gamma. \tag{2.49}$$

To begin with, we set $N = 4 - 2\varepsilon$ where $\varepsilon > 0$. By the key formula (2.48),

$$J_{4-2\varepsilon} = \frac{2\pi^{2-\varepsilon}}{\Gamma(2-\varepsilon)} \int_0^\infty \frac{r^{3-2\varepsilon} dr}{(a^2 + r^2)^2}.$$

Using the Liouville integral (2.45), we obtain

$$J_{4-2\varepsilon} = \frac{2\pi^{2-\varepsilon}}{\Gamma(2-\varepsilon)} \cdot \frac{a^{4-2\varepsilon} \Gamma(2-\varepsilon)\Gamma(\varepsilon)}{2a^4 \Gamma(2)} = \frac{\pi^{2-\varepsilon}}{a^{2\varepsilon}} \cdot \Gamma(\varepsilon).$$

The Laurent series (2.28) of the gamma function at the pole $\varepsilon = 0$ tells us that

$$J_{4-2\varepsilon} = \frac{\pi^{2-\varepsilon}}{a^{2\varepsilon}} \left(\frac{1}{\varepsilon} - \gamma + O(\varepsilon) \right), \qquad \varepsilon \to 0.$$

Truncating the singular term $\frac{1}{\varepsilon}$, we get the regularization

$$[J_{4-2\varepsilon}]_{\mathrm{reg}} := \frac{\pi^{2-\varepsilon}}{a^{2\varepsilon}} \left(-\gamma + O(\varepsilon) \right), \qquad \varepsilon \to 0. \tag{2.50}$$

Finally, letting $\varepsilon \to +0$, we obtain (2.49).

Historical remarks. Dimensional regularization was introduced by Gerardus 't Hooft and Martinus Veltman in 1972.[19] They showed that, in contrast to the 1938 Fermi model, the electroweak Standard Model in particle physics is renormalizable. For this, 't Hooft and Veltman were awarded the Nobel prize in physics in 1999.

> *Dimensional regularization is the standard method used by physicists in modern renormalization theory.*[20]

[19] G. 't Hooft and M. Veltman, Regularization and renormalization of gauge fields, Nucl. Phys. **B44** (1972), 189–213. See also the monograph by M. Veltman, Diagrammatica: the Path to Feynman Diagrams, Cambridge University Press, 1995.

[20] Observe that dimensional regularization fails if the Dirac matrix γ_5 is involved, which corresponds to the appearance of chirality (processes with parity violation). In this case, one has to use other regularization methods like algebraic renormalization (see the quotation in Sect. 19.6 on page 1019).

2.2.10 Pauli–Villars Regularization

> Regularize divergent integrals by introducing additional ghost particles of large masses.
>
> Folklore

Prototype. We regularize the divergent integral

$$J = \int_1^\infty \frac{dp}{p}$$

by introducing the additional parameter $P_M > 0$ and by setting

$$J_{\rm reg}(P_M) := \int_1^\infty \left(\frac{1}{p} - \frac{1}{p+P_M}\right)\, dp = \int_1^\infty \frac{P_M dp}{p(p+P_M)}.$$

In terms of physics, $P_M = cM$ is the momentum of an additional (ghost) particle of mass M.[21] Explicitly,

$$\boxed{J_{\rm reg}(P_M) = \ln P_M.}$$

In fact, the Pauli–Villars regularized integral $J_{\rm reg}(P_M)$ is equal to the limit

$$\int_1^{P_{\max}} \left(\frac{1}{p} - \frac{1}{p+P_M}\right)\, dp = \ln P_M + \ln \frac{P_{\max}}{P_{\max}+P_M} \to \ln P_M,$$

as $P_{\max} \to +\infty$. Observe that

- the logarithmic divergence of the cut-off integral

$$\int_1^{P_{\max}} \frac{dp}{p} = \ln P_{\max}, \qquad P_{\max} \to +\infty,$$

- corresponds to the logarithmic divergence $J_{\rm reg}(P_M) = \ln P_M$ of the regularized integral as $P_M \to +\infty$.

Higher dimension. Let $m > 0, M > 0$. We want to regularize the divergent integral

$$J := \int_{\mathbb{R}^4} \frac{d^4p}{(|p|^2 + c^2m^2)^2},$$

in the sense of Pauli–Villars. To this end, we choose a parameter $M > 0$, and we introduce the convergent integral

$$J_{\rm reg}(M^2) := \int_{\mathbb{R}^4} \left(\frac{1}{(|p|^2 + c^2m^2)^2} - \frac{1}{(|p|^2 + c^2M^2)^2}\right)\, d^4p.$$

Proposition 2.7 $J_{\rm reg}(M^2) = \pi^2(\ln M^2 - \ln m^2)$.

Proof. Differentiation of $J_{\rm reg}(M^2)$ with respect to the parameter M^2 yields

$$J'_{\rm reg}(M^2) = 2c^2 \int_{\mathbb{R}^4} \frac{d^4p}{(|p|^2 + c^2M^2)^3} = \frac{\pi^2}{M^2},$$

[21] As usual, c denotes the velocity of light in a vacuum.

by Problem 2.10. This differentiation is admissible, since both $J_{\rm reg}(M^2)$ and the derivative $J'_{\rm reg}(M^2)$ are convergent integrals. Noting that $J_{\rm reg}(m^2) = 0$, we get $J_{\rm reg}(M^2) = \pi^2(\ln M^2 - \ln m^2)$. □

Comparison of the Pauli–Villars method with dimensional regularization. By (2.50), dimensional regularization of J yields

$$J_{4-2\varepsilon} = \frac{\pi^{2-\varepsilon}}{(cm)^{2\varepsilon}} \left(\frac{1}{\varepsilon} - \gamma + O(\varepsilon) \right), \qquad \varepsilon \to 0.$$

Comparing this with Prop. 2.7, we see that

- the logarithmic growth $J_{\rm reg}(M^2) = \ln M^2 - \ln m^2$ as $M \to +\infty$ of the Pauli–Villars regularization corresponds to
- the pole of the dimensional regularization $J_{4-2\varepsilon}$ at the point $\varepsilon = 0$.

The Pauli–Villars regularization preserves the relativistic invariance. However, the introduction of additional masses may destroy the gauge invariance.

2.2.11 Analytic Regularization

Basic idea. Fix $P_0 > 0$. Our goal is to regularize the divergent integral

$$J_{\frac{1}{2}}(P_0) = \int_0^\infty \frac{dp}{(p+P_0)^{1/2}}.$$

To this end, we start with the convergent integral

$$J_\lambda(P_0) := \int_0^\infty \frac{dp}{(p+P_0)^\lambda}, \qquad \lambda > 1.$$

Explicitly, we get[22]

$$J_\lambda(P_0) = \frac{P_0^{1-\lambda}}{\lambda - 1} \qquad \text{for all} \quad \lambda > 1.$$

Using analytic continuation, we define

$$\boxed{J_\lambda(P_0)_{|\rm reg} := \frac{P_0^{1-\lambda}}{\lambda - 1}}$$

for all complex numbers λ with $\lambda \neq 1$. In particular, this yields

$$J_{\frac{1}{2}}(P_0)_{|\rm reg} = -2P_0^{\frac{1}{2}}.$$

Example 1. Fix the parameter $m > 0$. We want to study the integral

$$J_\lambda(m) = \int_{-\infty}^\infty \frac{dx}{(x^2+m^2)^\lambda}.$$

Since $\frac{1}{(x^2+m^2)^\lambda} \simeq \frac{1}{x^{2\lambda}}$ as $|x| \to \infty$, the integral is convergent if $\lambda > \frac{1}{2}$, by power-counting. Using the Liouville integral formula (2.45) on page 73, we obtain

[22] In fact, $J_\lambda(P_0) = \lim_{P\to+\infty} \int_0^P \frac{dp}{(P+P_0)^\lambda} = \frac{1}{1-\lambda} \lim_{P\to+\infty} (P+P_0)^{1-\lambda} - P_0^{1-\lambda}$.

$$J_\lambda(m) := \frac{m^{1-2\lambda}\sqrt{\pi}\,\Gamma\left(\lambda-\frac{1}{2}\right)}{\Gamma(\lambda).}, \qquad \lambda > \tfrac{1}{2}. \tag{2.51}$$

The gamma function $z \mapsto \Gamma(z)$ has no zeros, and it has poles precisely at the points $z = 0, -1, -2, \ldots$ Using analytic continuation, we define

$$J_\lambda(m)_{\mathrm{reg}} := \frac{m^{1-2\lambda}\sqrt{\pi}\,\Gamma\left(\lambda-\frac{1}{2}\right)}{\Gamma(\lambda)}$$

for all complex numbers λ different from $\frac{1}{2} - k, k = 0, 1, 2, \ldots$ Here, we use the convention $m^{1-2\lambda} := \mathrm{e}^{(1-2\lambda)\ln m}$. In particular,

$$J_n(m)_{\mathrm{reg}} = \frac{m^{1-2n}\pi(2n-2)!}{4^{n-1}(n-1)!}, \qquad n = 1, 2, \ldots$$

Example 2. Fix both the parameter $m > 0$ and the regularization parameter $\varepsilon > 0$. Let us study the integral

$$\mathcal{J}_\lambda(m,\varepsilon) = \int_{-\infty}^{\infty} \frac{dx}{((m-\mathrm{i}\varepsilon)^2 - x^2)^\lambda}.$$

We introduce the regularization parameter $\varepsilon > 0$ in order to avoid singularities of the integrand on the real axis.[23] Let $\lambda > \frac{1}{2}$. The integrand of the convergent integral $\mathcal{J}_\lambda(m,\varepsilon)$ has singularities at the points $x_- := m - \mathrm{i}\varepsilon$ and $x_+ = -m + \mathrm{i}\varepsilon$. Set $x = \mathrm{i}y$. By Wick rotation, we get

$$\mathcal{J}_\lambda(m,\varepsilon) = \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{dx}{((m-\mathrm{i}\varepsilon)^2 - x^2)^\lambda} = \int_{-\infty}^{\infty} \frac{\mathrm{i}dy}{((m-\mathrm{i}\varepsilon)^2 + y^2)^\lambda}, \qquad \lambda > \tfrac{1}{2}$$

(see Fig. 2.1 on page 71). Using (2.51) and applying analytic continuation with respect to $m - \mathrm{i}\varepsilon$, we obtain

$$\mathcal{J}_\lambda(m,\varepsilon) = \frac{\mathrm{i}(m-\mathrm{i}\varepsilon)^{1-2\lambda}\sqrt{\pi}\,\Gamma\left(\lambda-\frac{1}{2}\right)}{\Gamma(\lambda)}, \qquad \lambda > \tfrac{1}{2}.$$

Here $(m-\mathrm{i}\varepsilon)^{1-2\lambda} = \mathrm{e}^{(1-2\lambda)\ln(m-\mathrm{i}\varepsilon)}$, where $\ln(m-\mathrm{i}\varepsilon)$ denotes the principal value of the logarithm (see page 480). Moreover, using analytic continuation with respect to λ, we define the regularized integral

$$\mathcal{J}_\lambda(m,\varepsilon)_{\mathrm{reg}} := \frac{\mathrm{i}(m-\mathrm{i}\varepsilon)^{1-2\lambda}\sqrt{\pi}\,\Gamma\left(\lambda-\frac{1}{2}\right)}{\Gamma(\lambda)}$$

for all complex numbers λ different from $\frac{1}{2} - k$ with $k = 0, 1, 2, \ldots$ In particular, since the gamma function Γ has poles at the points $0, -1, -2, \ldots$, we get

$$\mathcal{J}_\lambda(m,\varepsilon)_{\mathrm{reg}} = 0, \qquad \lambda = 0, -1, -2, \ldots$$

[23] For example, because of the decomposition

$$\frac{1}{m^2 - x^2} = \frac{1}{2m}\left(\frac{1}{x+m} - \frac{1}{x-m}\right),$$

the singularities at the points $x_+ = m$ and $x_- = -m$ are responsible for the non-existence of the integral $\int_{-\infty}^{\infty} \frac{dx}{m^2-x^2}$.

Adiabatic limit. In order to free ourselves from the choice of the small regularization parameter $\varepsilon > 0$, we consider the limit

$$\mathcal{J}_\lambda(m,0)_{\mathrm{reg}} := \lim_{\varepsilon\to+0} \mathcal{J}_\lambda(m,\varepsilon)_{\mathrm{reg}} = \frac{\mathrm{i}m^{1-2\lambda}\sqrt{\pi}\,\Gamma\left(\lambda-\frac{1}{2}\right)}{\Gamma(\lambda)}$$

for all complex numbers λ different from $\frac{1}{2}-k$ with $k=0,1,2,\ldots$

Critical values. For $\lambda=\frac{1}{2}-k$ with $k=0,1,2,\ldots$ and $\varepsilon>0$, we define

$$\mathcal{J}_{\frac{1}{2}-k}(m,\varepsilon) := \frac{\mathrm{i}(m-\mathrm{i}\varepsilon)^{2k}\sqrt{\pi}\,\Gamma_{\mathrm{reg}}(-k)}{\Gamma(\frac{1}{2}-k)}$$

together with

$$\mathcal{J}_{\frac{1}{2}-k}(m,0)_+ := \lim_{\varepsilon\to+0} \mathcal{J}_{\frac{1}{2}-k}(m,\varepsilon) = \frac{\mathrm{i}m^{2k}\sqrt{\pi}\,\Gamma_{\mathrm{reg}}(-k)}{\Gamma(\frac{1}{2}-k)}.$$

Recall that $\Gamma_{\mathrm{reg}}(0)=-\gamma$. Hence $\mathcal{J}_{\frac{1}{2}}(m,0)_+ = -\mathrm{i}\gamma$.

The Feynman integration trick. Let $a>0$ and $b>0$. Consider the integral

$$J=\int_{-\infty}^{\infty}\frac{dx}{(x^2+a)(x^2+b)}.$$

We want to show that $J=\frac{\pi}{\sqrt{ab}(\sqrt{a}+\sqrt{b})}$.

Proof. By the Feynman trick (2.36) on page 71,

$$J=\int_{-\infty}^{\infty}dx\int_0^1\frac{du}{(x^2+au+b(1-u))^2}.$$

Since the nonnegative integrand $(x,u)\mapsto\frac{1}{(x^2+au+b(1-u))^2}$ is integrable over the set $\mathbb{R}\times[0,1]$, the Fubini theorem tells us that the iterated integrations can be interchanged. Hence

$$J=\int_0^1 du\int_{-\infty}^{\infty}\frac{dx}{(x^2+au+b(1-u))^2}.$$

Using the Liouville integral (2.45) on page 73,

$$\int_{-\infty}^{\infty}\frac{dx}{(x^2+au+b(1-u))^2}=\frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{3}{2}\right)}{(au+b(1-u))^{3/2}}.$$

Therefore,

$$J=\frac{\pi}{2}\int_0^1\frac{du}{(au+b(1-u))^{3/2}}=\frac{\pi}{\sqrt{ab}\,(\sqrt{a}+\sqrt{b})}.$$

□

2.2.12 Application to Algebraic Feynman Integrals in Minkowski Space

> The important mathematical problem of evaluating (algebraic) Feynman integrals arises quite naturally in elementary particle physics when one treats various quantities (corresponding to Feynman diagrams) in the framework of perturbation theory.[24]
>
> Vladimir Smirnov, 2006

Fix the mass parameter $m > 0$ and the regularizing parameter $\varepsilon > 0$. Our goal is to compute the integral

$$J_\lambda(m,\varepsilon) = \int_{\mathbb{R}^4} \frac{d^4p}{(c^2(m-\mathrm{i}\varepsilon)^2 - p^2)^\lambda}. \tag{2.52}$$

Here, $p = (p^0, p^1, p^2, p^3)$, and $p^2 := |p^0|^2 - \sum_{j=1}^3 |p^j|^2$. Introducing the energy $E := cp^0$ and the momentum vector $\mathbf{p}$, we get

$$p^2 := \frac{E^2}{c^2} - \mathbf{p}^2 \tag{2.53}$$

where c denotes the velocity of light in a vacuum. To simplify notation, we set $c := 1$.

Wick rotation. Let $\lambda > 3$. By definition, the integral is to be understood as the following iterated integral:

$$J_\lambda(m,\varepsilon) := \int_{\mathbb{R}^3} d^3\mathbf{p} \int_{-\infty}^{\infty} \frac{dE}{((m-\mathrm{i}\varepsilon)^2 - E^2 + \mathbf{p}^2)^\lambda}.$$

By Wick rotation,[25]

$$\int_{-\infty}^{\infty} \frac{dE}{((m-\mathrm{i}\varepsilon)^2 - E^2 + \mathbf{p}^2)^\lambda} = \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{dE}{((m-\mathrm{i}\varepsilon)^2 - E^2 + \mathbf{p}^2)^\lambda}.$$

Replacing the variable E by $\mathrm{i}E$, this integral is equal to

$$\mathrm{i}\int_{-\infty}^{\infty} \frac{dE}{((m-\mathrm{i}\varepsilon)^2 + E^2 + \mathbf{p}^2)^\lambda}.$$

Hence

$$J_\lambda(m,\varepsilon) = \mathrm{i}\int_{\mathbb{R}^4} \frac{d^4p}{((m-\mathrm{i}\varepsilon)^2 + E^2 + \mathbf{p}^2)^\lambda}.$$

Using the sphere trick (2.34) on page 70, we get

$$J_\lambda(m,\varepsilon) = 2\mathrm{i}\pi^2 \int_0^\infty \frac{r^3 dr}{((m-\mathrm{i}\varepsilon)^2 + r^2)^\lambda}.$$

The Liouville integral (2.45) on page 73 tells us that[26]

$$J_\lambda(m,\varepsilon) = \mathrm{i}\pi^2 (m-\mathrm{i}\varepsilon)^{4-2\lambda} \frac{\Gamma(\lambda-2)}{\Gamma(\lambda)} = \frac{\mathrm{i}\pi^2 (m-\mathrm{i}\varepsilon)^{4-2\lambda}}{(\lambda-1)(\lambda-2)}, \quad \lambda > 3.$$

[24] V. Smirnov, Feynman Integral Calculus, Springer, Berlin, 2006.
[25] See Fig. 2.1 on page 71.
[26] Note that $\Gamma(\lambda) = (\lambda-1)\Gamma(\lambda-1) = (\lambda-1)(\lambda-2)\Gamma(\lambda-2)$.

Analytic continuation. For all complex numbers λ different from the singular exponents $\lambda = 1, 2$, analytic continuation with respect to λ yields

$$J_\lambda(m,\varepsilon) = \frac{\mathrm{i}\pi^2(m-\mathrm{i}\varepsilon)^{4-2\lambda}}{(\lambda-1)(\lambda-2)}.$$

In a neighborhood of $\lambda = 1$, we have $\frac{1}{(\lambda-1)(\lambda-2)} = -\frac{1}{\lambda-1} - 1 + O(\lambda-1)$ as $\lambda \to 1$. Therefore, we define the regularization

$$J_1(m,\varepsilon)_{\mathrm{reg}} := \lim_{\lambda\to 1}\left(J_\lambda(m,\varepsilon) + \frac{1}{\lambda-1}\right) = -\mathrm{i}\pi^2(m-\mathrm{i}\varepsilon)^2.$$

Similarly, since $\frac{1}{(\lambda-1)(\lambda-2)} = \frac{1}{\lambda-2} - 1 + O(\lambda-2)$ as $\lambda \to 2$, we define

$$J_2(m,\varepsilon)_{\mathrm{reg}} := \lim_{\lambda\to 2}\left(J_\lambda(m,\varepsilon) - \frac{1}{\lambda-2}\right) = -\mathrm{i}\pi^2.$$

Adiabatic limit. The limit $\varepsilon \to +0$ yields

$$J_1(m,0)_+ := \lim_{\varepsilon\to+0} J_1(m,\varepsilon)_{\mathrm{reg}} = -\mathrm{i}\pi^2 m^2,$$

and $J_2(m,0)_+ := \lim_{\varepsilon\to+0} J_2(m,\varepsilon)_{\mathrm{reg}} = -\mathrm{i}\pi^2$.

2.2.13 Distribution-Valued Meromorphic Functions

> In quantum field theory, Green's functions are closely related to distribution-valued meromorphic functions.
>
> Folklore

The following considerations will be used in the next section in order to study Newton's equation of motion in terms of the Fourier transform of tempered distributions. For the convenience of the reader, our approach will be chosen in such a way that it serves as a prototype for the study of Green's functions in quantum field theory later on. The essential tool is given by families of tempered distributions which analytically depend on a complex parameter λ. This approach dates back to a fundamental paper by Marcel Riesz in 1948.[27]

Basic idea. We want to study the integral

$$\int_0^\infty t^\lambda \varphi(t)dt \tag{2.54}$$

for all test functions $\varphi \in \mathcal{S}(\mathbb{R})$ and all complex numbers λ. If $\Re(\lambda) > -1$, then the integral is convergent. However, if $\Re(\lambda) \le -1$ and $\varphi(0) \neq 0$, then the integral is divergent. The idea is to use analytic continuation with respect to λ. For example, if $-2 < \Re(\lambda) < -1$, then we will show below that the analytic continuation of the integral looks like

$$\int_0^\infty t^\lambda\varphi(t)dt = \frac{\varphi(0)}{\lambda+1} + \frac{\varphi'(0)}{\lambda+2} + \int_0^1 t^\lambda(\varphi(t)-\varphi(0)-\varphi'(0)t)\,dt + \int_1^\infty t^\lambda\varphi(t)dt. \tag{2.55}$$

[27] See the footnote on page 92.

Using the terminology to be introduced below, relation (2.55) can be written as

$$t_+^\lambda = \frac{\delta}{\lambda+1} - \frac{\delta'}{\lambda+2} + (t_+^\lambda)_{\mathrm{reg},2}, \qquad -2 < \Re(\lambda) < -1.$$

To begin with, we define $t_+ := \theta(t)t$ for all $t \in \mathbb{R}$. Hence

$$t_+ = \begin{cases} t & \text{if} \quad t \geq 0, \\ 0 & \text{if} \quad t < 0. \end{cases}$$

The classical function $t \mapsto t_+$ corresponds to a tempered distribution denoted by

$$t_+(\varphi) := \int_{-\infty}^{\infty} t_+ \varphi(t) dt \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).$$

Observe the following crucial fact. If we set $f(t) := (t_+)^{-3/2}$, then the classical function f is not defined for arguments $t \leq 0$. However, the tempered distribution $t_+^{-3/2}$ to be introduced below is well defined. But it cannot be represented by a classical, locally integrable function.

The tempered distribution t_+^λ. For all complex numbers λ with positive real part, $\Re(\lambda) > 0$, we define the tempered distribution t_+^λ given by

$$t_+^\lambda(\varphi) := \int_{-\infty}^{\infty} t_+^\lambda \varphi(t) dt \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).$$

Hence $t_+^\lambda(\varphi) := \int_0^\infty t_+^\lambda \varphi(t) dt$ for all $\varphi \in \mathcal{S}(\mathbb{R})$ where $t_+^\lambda = \mathrm{e}^{\lambda \ln t}$ if $t > 0$.

Proposition 2.8 *Let $\varphi \in \mathcal{S}(\mathbb{R})$. The function $\lambda \mapsto t_+^\lambda(\varphi)$ defined on the open half-plane $\{\lambda \in \mathbb{C} : \Re(\lambda) > 0\}$ can be analytically extended to a holomorphic function on the punctured complex plane $\mathbb{C} \setminus \{-1, -2, \ldots\}$. Explicitly, choose $k = 1, 2, \ldots$ and the complex number λ with $\Re(\lambda) > -k$ and $\lambda \neq -1, -2, \ldots, -k+1$. Then*

$$t_+^\lambda = \sum_{r=1}^{k} \frac{(-1)^{r-1} \delta^{(r-1)}}{(\lambda + r)(r-1)!} + (t_+^\lambda)_{\mathrm{reg},k}.$$

Here, for all test functions $\varphi \in \mathcal{S}(\mathbb{R})$, the regular part is defined by

$$(t_+^\lambda)_{\mathrm{reg},k}(\varphi) := \int_0^1 t^\lambda \left(\varphi(t) - \sum_{r=0}^{k-1} \frac{\varphi^{(r)}(0)}{r!} \, t^r \right) dt + \int_1^\infty t^\lambda \varphi(t) \, dt.$$

We say that the map $\lambda \mapsto t_+^\lambda$ is a meromorphic function on the complex plane $\mathbb{C}$ with values in the space $\mathcal{S}'(\mathbb{R})$ of tempered distributions. This function has simple poles at the points $\lambda = -1, -2, \ldots$ with the residues

$$\mathrm{res}_{\lambda=-k}(t_+^\lambda) = \frac{(-1)^{k-1} \delta^{(k-1)}}{(k-1)!}, \qquad k = 1, 2, \ldots$$

Proof. Fix the test function $\varphi \in \mathcal{S}(\mathbb{R})$. We will proceed by induction.

(I) Let $k = 1$. We start with $t_+^\lambda(\varphi) = \int_0^\infty t^\lambda \varphi(t) \, dt$ for $\Re(\lambda) > 0$. Obviously, if $\Re(\lambda) > 0$, then $t_+^\lambda(\varphi)$ is equal to

$$y_+^\lambda(\varphi) := \int_0^1 t^\lambda(\varphi(t) - \varphi(0))\, dt + \int_1^\infty t^\lambda \varphi(t)\, dt + A$$

where $A := \varphi(0)\int_0^1 t^\lambda dt = \frac{1}{\lambda+1}$. The point is that the function $\lambda \mapsto y_+^\lambda(\varphi)$ is holomorphic for all complex numbers λ with $\Re(\lambda) > -1$. This follows from $\varphi(t) - \varphi(0) = O(t)$ as $t \to 0$. In fact, the Taylor expansion theorem tells us that

$$|t^\lambda(\varphi(t) - \varphi(0))| \le t^{\lambda+1} \sup_{0\le\tau\le t} |\varphi'(\tau)|, \qquad t \in [0,1].$$

Hence

$$y_+^\lambda(\varphi) = \int_0^1 t^\lambda(\varphi(t) - \varphi(0))\, dt + \int_1^\infty t^\lambda \varphi(t)\, dt + \frac{\varphi(0)}{\lambda+1}, \quad \Re(\lambda) > -1.$$

Thus, the function $\lambda \mapsto y_+^\lambda(\varphi)$ (defined for $\Re(\lambda) > -1$) is an analytic continuation of the function $\lambda \mapsto t_+^\lambda(\varphi)$ (defined for $\Re(\lambda) > 0$).

(II) Let $k = 2$. For all complex numbers λ with $\Re(\lambda) > -1$, the value $y_+^\lambda(\varphi)$ is equal to

$$z_+^\lambda(\varphi) := \int_0^1 t^\lambda(\varphi(t) - \varphi(0) - \varphi'(0)t)\, dt + \int_1^\infty t^\lambda \varphi(t)\, dt + B$$

where

$$B := \varphi'(0)\int_0^1 t^{\lambda+1}\, dt + \varphi(0)\int_0^1 t^\lambda dt = \frac{\varphi'(0)}{\lambda+2} + \frac{\varphi(0)}{\lambda+1}.$$

By Taylor expansion, $\varphi(t) - \varphi(0) - \varphi'(0)t = O(t^2)$ as $t \to 0$. Thus, the function $\lambda \mapsto z_+^\lambda(\varphi)$ is holomorphic on $\{\lambda \in \mathbb{C} : \Re(\lambda) > -2\} \setminus \{-1\}$. Consequently, the function $\lambda \mapsto z_+^\lambda(\varphi)$ (defined for $\Re(\lambda) > -2$ with $\lambda \neq -1$) is an analytic continuation of the function $\lambda \mapsto y_+^\lambda(\varphi)$ (defined for $\Re(\lambda) > -1$).

(III) An induction argument for $k = 1, 2, \ldots$ yields the desired result. □

Fundamental solution. Each tempered distribution $x \in \mathcal{S}'(\mathbb{R})$ which satisfies the equation

$$\frac{d^2x}{dt^2} = \delta \tag{2.56}$$

is called a tempered fundamental solution of the differential operator $\frac{d^2}{dt^2}$.[28]

Proposition 2.9 *The tempered distribution t_+ has the following properties:*

(i) $\ddot{t}_+ = \delta$.

(ii) $t_+ = 0$ *on* $]-\infty, 0[$.

(iii) *The general tempered fundamental solution of the differential operator* $\frac{d^2}{dt^2}$ *looks like*

$$x = t_+ + H$$

where H corresponds to the linear function $H(t) := a + bt$ for all $t \in \mathbb{R}$ with arbitrary complex numbers a, b as coefficients.

(iv) *The equation $\ddot{x} = \delta$, $x \in \mathcal{S}'(\mathbb{R})$ with $x = 0$ on $]-\infty, 0[$ has precisely one solution given by t_+.*

[28] Recall that we also write $\ddot{x}$ or x'' instead of $\frac{d^2x}{dt^2}$.

Because of (iv), t_+ is called the retarded tempered fundamental solution of the differential operator $\frac{d^2}{dt^2}$.

Proof. Ad (i). For all test functions $\varphi \in \mathcal{S}(\mathbb{R})$, integration by parts yields

$$\ddot{t}_+(\varphi) = t_+(\ddot{\varphi}) = \int_0^\infty t\ddot{\varphi}(t)dt = -\int_0^\infty \dot{\varphi}(t)dt = \varphi(0) = \delta(\varphi).$$

Ad (ii). If the test function $\varphi \in \mathcal{S}(\mathbb{R})$ vanishes outside $]-\infty, 0[$, then $t_+(\varphi) = \int_{-\infty}^0 t_+\varphi(t)dt = 0$.

Ad (iii). Let $H \in \mathcal{S}'(\mathbb{R})$ with $\ddot{H} = 0$. Note that sing supp $t_+ = \{0\}$ (see page 704 of Vol. I). By Theorem 8.16 on page 745, the differential operator $\frac{d^2}{dt^2}$ is hypoelliptic, that is, the tempered distribution H corresponds to a smooth function $t \mapsto H(t)$. From $\ddot{H}(t) \equiv 0$ it follows that $H(t) = a + bt$ for all $t \in \mathbb{R}$.

Finally, let $x \in \mathcal{S}'(\mathbb{R})$ with $\ddot{x} = \delta$. Set $H := x - t_+$. Then $\ddot{H} = 0$.

Ad (iv). If $a + bt = 0$ for all $t < 0$, then $a = b = 0$. □

According to the general theory of distributions, the following hold (see Sect. 11.7 of Vol. I). Let $F \in \mathcal{D}'(\mathbb{R})$ be a given distribution with compact support. Then the equation

$$\ddot{x} = F, \qquad x \in \mathcal{D}'(\mathbb{R}) \tag{2.57}$$

has the special solution $x = t_+ * F$ and the general solution

$$x = t_+ * F + H \tag{2.58}$$

where H corresponds to the linear function $H(t) = at + b$ for all $t \in \mathbb{R}$ with complex coefficients a, b. In particular, if $F : \mathbb{R} \to \mathbb{R}$ is continuous with compact support, then (2.58) looks like

$$x(t) = \int_{-\infty}^\infty (t-\tau)_+ F(\tau)d\tau + H(t) = \int_{-\infty}^t (t-\tau)F(\tau)d\tau + H(t) \tag{2.59}$$

for all $t \in \mathbb{R}$.

The tempered distribution t_-^λ. Parallel to $t_+ = \theta(t)t$, let us define the function $t_- := -\theta(-t)t$ for all $t \in \mathbb{R}$. Then

$$t_- = \begin{cases} 0 & \text{if} \quad t \geq 0, \\ |t| & \text{if} \quad t < 0. \end{cases}$$

The corresponding tempered distribution reads as

$$t_-(\varphi) = \int_{-\infty}^\infty t_-\varphi(t)dt = \int_{-\infty}^0 |t|\varphi(t)dt \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).$$

As for t_+, we get the following. The function $\lambda \mapsto t_-^\lambda(\varphi)$ defined on the open half-plane $\{\lambda \in \mathbb{C} : \Re(\lambda) > 0\}$ can be analytically extended to a holomorphic function on the punctured complex plane $\mathbb{C} \setminus \{-1, -2, \ldots\}$. Explicitly, choose $k = 1, 2, \ldots$ and the complex number λ with $\Re(\lambda) > -k$ and $\lambda \neq -1, -2, \ldots, -k+1$. Then

$$t_-^\lambda = \sum_{r=1}^k \frac{\delta^{(r-1)}}{(\lambda + r)(r-1)!} + (t_-^\lambda)_{\mathrm{reg},k}.$$

Here, for all test functions $\varphi \in \mathcal{S}(\mathbb{R})$, the regular part is defined by

$$(t_-^\lambda)_{\mathrm{reg},k}(\varphi) := \int_{-1}^{0} |t|^\lambda \left(\varphi(t) - \sum_{r=0}^{k-1} \frac{\varphi^{(r)}(0)}{r!}\, t^r \right) dt + \int_{-\infty}^{-1} |t|^\lambda \varphi(t)\, dt.$$

Parallel to the proof of Prop. 2.9, we obtain the following result.

Proposition 2.10 *The tempered distribution t_- has the following properties:*
(i) $\ddot{t}_- = \delta$.
(ii) $t_- = 0$ *on* $]0,\infty[$.
(iii) *The general tempered fundamental solution of the differential operator* $\frac{d^2}{dt^2}$ *looks like*

$$x = t_- + H$$

where H corresponds to the linear function $H(t) := a+bt$ for all $t \in \mathbb{R}$ with arbitrary complex numbers a, b as coefficients.
(iv) *The equation $\ddot{x} = \delta$, $x \in \mathcal{S}'(\mathbb{R})$ with $x = 0$ on $]0,\infty[$ has precisely one solution given by t_-.*

Because of (iv), t_- is called the advanced tempered fundamental solution of the differential operator $\frac{d^2}{dt^2}$.

The tempered distribution $|t|_*^\lambda$. For all complex numbers λ with the property $\lambda \neq -1, -3, -5, \ldots$, we define

$$\boxed{|t|_*^\lambda := t_+^\lambda + t_-^\lambda.} \tag{2.60}$$

Note that the common poles $-2, -4, \ldots$ of t_+^λ and t_-^λ cancel each other. Explicitly, for all complex numbers λ with $\Re(\lambda) > 0$, we have[29]

$$|t|_*^\lambda(\varphi) = \int_{-\infty}^{\infty} |t|^\lambda \varphi(t) dt \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).$$

The holomorphic function $\lambda \mapsto |t|_*^\lambda(\varphi)$ on the half-plane $\{\lambda \in \mathbb{C} : \Re(\lambda) > 0\}$ can be analytically continued to a holomorphic function on $\mathbb{C} \setminus \{-1, -3, \ldots\}$ which has simple poles at the points $-1, -3, \ldots$ with the residues

$$\mathrm{res}_{\lambda=-k}(|t|_*^\lambda) = \frac{2\delta^{(k-1)}}{(k-1)!}, \qquad k = 1, 3, 5, \ldots$$

Proposition 2.11 *The tempered distribution $\frac{1}{2}|t|_* = \frac{1}{2}(t_+ + t_-)$ is a fundamental solution of the differential operator* $\frac{d^2}{dt^2}$.

This follows from $\ddot{t}_+ = \delta$ and $\ddot{t}_- = \delta$. Naturally enough, $\frac{1}{2}|t|_*$ is called the retarded-advanced fundamental solution of $\frac{d^2}{dt^2}$.

The tempered distributions $(t + 0_+\mathrm{i})^\lambda$ and $(t - 0_+\mathrm{i})^\lambda$. Fix $\varepsilon > 0$. For all complex numbers λ and all real numbers t, we have[30]

$$(t \pm \varepsilon\mathrm{i})^\lambda = \mathrm{e}^{\lambda \ln|t \pm \varepsilon \mathrm{i}|}\, \mathrm{e}^{\lambda \mathrm{i} \arg(t \pm \varepsilon \mathrm{i})}.$$

Let $\lambda \in \mathbb{C}$. For any test function $\varphi \in \mathcal{S}(\mathbb{R})$, we define

[29] In this case, we also write $|t|^\lambda$ instead of $|t|_*^\lambda$.
[30] Recall that any complex number z can be uniquely represented by $z = |z|\mathrm{e}^{i \arg(z)}$ where $-\pi < \arg(z) \leq \pi$. For example, $\arg(-1) = \pi$.

$$(t+\varepsilon\mathrm{i})^\lambda(\varphi) := \int_{-\infty}^{\infty} (t+\varepsilon\mathrm{i})^\lambda \varphi(t)dt.$$

This yields the tempered distributions $(t+\varepsilon\mathrm{i})^\lambda$ and $(t-\varepsilon\mathrm{i})^\lambda$. The functions $\lambda \mapsto (t+\varepsilon\mathrm{i})^\lambda(\varphi)$ and $\lambda \mapsto (t-\varepsilon\mathrm{i})^\lambda(\varphi)$ are holomorphic functions on the complex plane $\mathbb{C}$.

Proposition 2.12 (i) *For any complex number λ, the limits*

$$(t+0_+\mathrm{i})^\lambda := \lim_{\varepsilon\to+0}(t+\varepsilon\mathrm{i})^\lambda$$

and $(t-0_+\mathrm{i})^\lambda := \lim_{\varepsilon\to+0}(t-\varepsilon\mathrm{i})^\lambda$ exist in the sense of tempered distributions.

(ii) *For any test function $\varphi \in \mathcal{S}(\mathbb{R})$, the functions $\lambda \mapsto (t+0_+\mathrm{i})^\lambda(\varphi)$ and $\lambda \mapsto (t-0_+\mathrm{i})^\lambda(\varphi)$ are holomorphic on the complex plane $\mathbb{C}$. Explicitly, for all $\lambda \in \mathbb{C} \setminus \{-1,-2,-3,\dots\}$,*

$$(t\pm 0_+\mathrm{i})^\lambda = t_+^\lambda + e^{\pm\mathrm{i}\lambda\pi}t_-^\lambda.$$

Moreover, if $k=1,2,\dots$, then we have the Sokhotski formula

$$(t\pm 0_+\mathrm{i})^{-k} = \lim_{\lambda\to-k}\left(t_+^\lambda + (-1)^k t_-^\lambda\right) \pm \frac{\mathrm{i}\pi(-1)^k}{(k-1)!}\,\delta^{(k-1)}.$$

Here, the limit is to be understood in the sense of tempered distributions.

Proof. Let $\Re(\lambda) > 0$. Then the following two classical limits

$$\lim_{\varepsilon\to+0}(t\pm\varepsilon\mathrm{i})^\lambda = \begin{cases} t^\lambda & \text{if} \quad t>0, \\ \mathrm{e}^{\pm\mathrm{i}\lambda\pi}|t|^\lambda & \text{if} \quad t<0 \end{cases}$$

exist. Hence $(t\pm 0_+\mathrm{i})^\lambda(\varphi) = t_+^\lambda(\varphi) + \mathrm{e}^{\pm\mathrm{i}\lambda\pi}t_-^\lambda(\varphi)$ for all $\varphi \in \mathcal{S}(\mathbb{R})$.

The proofs of the remaining statements can be found in Gelfand and Shilov (1964), Vol. 1, Chap. 1. □

The Fourier transform of tempered fundamental solutions. For the function $x \in \mathcal{S}(\mathbb{R})$, the Fourier transform $\mathcal{F}(x)$ from time t to energy E is given by[31]

$$\mathcal{F}(x)(E) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x(t)\mathrm{e}^{\mathrm{i}Et}dt, \qquad E\in\mathbb{R}.$$

By Sect. 11.3.4 of Vol. I, this can be extended to the Fourier transform of tempered distributions.

Proposition 2.13 (i) $\mathcal{F}(t_+) = -\frac{1}{\sqrt{2\pi}}(E+0_+\mathrm{i})^{-2}$.
(ii) $\mathcal{F}(t_-) = -\frac{1}{\sqrt{2\pi}}(E-0_+\mathrm{i})^{-2}$.
(iii) $\mathcal{F}(\frac{1}{2}|t|) = -\frac{1}{\sqrt{2\pi}}|E|_*^{-2}$.

Proof. We will use the method of the adiabatic limit. Fix $\varepsilon > 0$.

Ad (i). Define $t_{+,\varepsilon}(t) := t_+\mathrm{e}^{-\varepsilon t}$ for all $t\in\mathbb{R}$. Integration by parts yields

$$\mathcal{F}(t_{+,\varepsilon})(E) = \frac{1}{\sqrt{2\pi}}\int_0^\infty t\mathrm{e}^{-\varepsilon t}\mathrm{e}^{\mathrm{i}Et}\,dt = \int_0^\infty \frac{\mathrm{e}^{-\varepsilon t}\mathrm{e}^{\mathrm{i}Et}dt}{\sqrt{2\pi}(\varepsilon-\mathrm{i}E)} = -\frac{(E+\varepsilon\mathrm{i})^{-2}}{\sqrt{2\pi}}.$$

[31] We set $\hbar := 1$. Then $E=\omega$, that is, energy E and angular frequency ω coincide.

The factor $\mathrm{e}^{-\varepsilon t}$ enforces the convergence of the integral for all $E \in \mathbb{R}$. Letting $\varepsilon \to +0$, we get the claim. Note that the Fourier transform is sequentially continuous with respect of the convergence on the space $\mathcal{S}'(\mathbb{R})$ of tempered distributions.

Ad (ii). Set $t_{-,\varepsilon} := t_- \mathrm{e}^{\varepsilon t}$ for all $t \in \mathbb{R}$. Then

$$\mathcal{F}(t_{-,\varepsilon})(E) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} t\mathrm{e}^{\varepsilon t}\mathrm{e}^{\mathrm{i}Et}\, dt = -\frac{1}{\sqrt{2\pi}}(E - \varepsilon\mathrm{i})^{-2}.$$

Ad (iii). Recall that $|t|_* = t_+ + t_-$. By the Sokhotski formula on page 86,

$$(E + 0_+\mathrm{i})^{-2} + (E - 0_+\mathrm{i})^{-2} = 2(E_+^{-2} + E_-^{-2}) = 2|E|_*^{-2}. \qquad \square$$

2.2.14 Application to Newton's Equation of Motion

Consider the following initial-value problem

$$\boxed{m\ddot{x}(t) = F(t), \qquad t \in \mathbb{R}, \quad x(0) = a, \;\; \dot{x}(0) = v.} \tag{2.61}$$

This describes the motion of a particle of mass $m > 0$ on the real line with the initial position a and the initial velocity v at time $t = 0$. We want to discuss this classical problem by using a language which can be generalized to quantum field theory. To simplify notation, we set $m := 1$.

If $F : \mathbb{R} \to \mathbb{R}$ is continuous, then the unique solution of (2.61) reads as

$$x(t) = a + vt + \int_0^t (t - \tau)F(\tau)d\tau \qquad \text{for all} \quad t \in \mathbb{R}.$$

This solution formula can be written as

$$x(t) = a + vt + \int_{-\infty}^{\infty} G(t, \tau)F(\tau)d\tau, \qquad t \in \mathbb{R}, \tag{2.62}$$

with the Green's function

$$G(t, \tau) := \begin{cases} (t - \tau)_+ & \text{if} \quad \tau \geq 0, \\ (t - \tau)_- & \text{if} \quad \tau < 0. \end{cases}$$

Fundamental Solutions in the Time Space

Let $F \in \mathcal{D}'(\mathbb{R})$ be a distribution with compact support. If $g \in \mathcal{D}'(\mathbb{R})$ is a fundamental solution of $\frac{d^2}{dt^2}$, that is, $\ddot{g} = \delta$, then the distribution

$$x = g * F$$

is a solution of the Newtonian equation $\ddot{x} = F$, by the general theory of distributions (see Sect. 11.7 of Vol. I). In particular, suppose that the force function $F : \mathbb{R} \to F$ is continuous with compact support. This means that there exists a compact time interval $[t_1, t_2]$ such that

$$F(t) = 0 \qquad \text{for all} \quad t \notin [t_0, t_1].$$

Choosing the fundamental solutions $t_+, t_-, \frac{1}{2}|t|$, we get the following special solutions of the Newtonian equation for all $t \in \mathbb{R}$:

- Retarded solution $x_{\mathrm{ret}} = t_+ * F$:

$$x_{\mathrm{ret}}(t) = \int_{-\infty}^{\infty} (t-\tau)_+ F(\tau)d\tau = \int_{-\infty}^{t} (t-\tau)F(\tau)d\tau.$$

  The retarded solution is the unique solution of the initial-value problem[32]

$$m\ddot{x}(t) = F(t), \quad t \in \mathbb{R}, \quad x(t_1) = 0, \;\; \dot{x}(t_1) = 0.$$

  Alternatively, the retarded solution is the unique solution of the Newtonian equation $m\ddot{x}(t) = F(t), t \in \mathbb{R}$, which vanishes in a neighborhood of the point in time $t = -\infty$. This corresponds to the motion of a particle that rests in a neighborhood of $t = -\infty$. This motion is quite natural because the force vanishes near time $t = -\infty$ (remote past).
- Advanced solution $x_{\mathrm{adv}} = t_- * F$:

$$x_{\mathrm{adv}}(t) = \int_{-\infty}^{\infty} (t-\tau)_- F(\tau)d\tau = -\int_{t}^{\infty} (t-\tau)F(\tau)d\tau.$$

  The advanced solution is the unique solution of the initial-value problem

$$m\ddot{x}(t) = F(t), \quad t \in \mathbb{R}, \quad x(t_2) = 0, \;\; \dot{x}(t_2) = 0.$$

  Alternatively, the advanced solution is the unique solution of the Newtonian equation $m\ddot{x}(t) = F(t), t \in \mathbb{R}$, which vanishes in a neighborhood of the point in time $t = +\infty$. This corresponds to the motion of a particle that rests in a neighborhood of $t = +\infty$. This motion is quite natural because the force vanishes near time $t = +\infty$ (far future).
- Retarded-advanced solution $x_{\mathrm{ret/adv}}(t) = \frac{1}{2}|t| * F$:

$$x_{\mathrm{ret/adv}} = \tfrac{1}{2}(x_{\mathrm{ret}}(t) + x_{\mathrm{adv}}(t)) = \int_{-\infty}^{\infty} \tfrac{1}{2}|t-\tau| \; F(\tau)d\tau.$$

Note that the retarded solution $x_{\mathrm{ret}}(t)$ at time t only depends on the values of the force F on the time interval $]-\infty, t]$ in the past. This motivates the terminology 'retarded'.

The advanced solution $x_{\mathrm{adv}}(t)$ at time t only depends on the values of the force F on the time interval $]-\infty, t]$ in the future. This motivates the terminology 'advanced'. Moreover, the retarded-advanced solution is the arithmetical mean of the retarded and the advanced solution.

Furthermore, this motivates why $t_+, t_-, \frac{1}{2}|t|$ are called retarded, advanced, retarded-advanced fundamental solution of $\frac{d^2}{dt^2}$, respectively. The importance of retarded-advanced fundamental solutions in quantum electrodynamics was emphasized by Wheeler and Feynman in 1945.[33] The function

$$G^+(t,\tau) := (t-\tau)_+ \qquad \text{for all} \quad t, \tau \in \mathbb{R}$$

is called the retarded Green's function of the Newtonian equation $m\ddot{x} = F$ (with $m := 1$), and

$$G^-(t,\tau) := (t-\tau)_- \qquad \text{for all} \quad t, \tau \in \mathbb{R}$$

[32] Choose $m := 1$.

[33] A. Wheeler and R. Feynman, Interaction with the absorber as the mechanism of radiation, Rev. Mod. Phys. **17** (1945), 157–181. See the quotation on page 486.

is called the advanced Green's function. Fix $\tau \in \mathbb{R}$. Then, in the sense of tempered distributions, for all $t \in \mathbb{R}$ we get:

$$\frac{\partial^2}{\partial t^2} G^+(t,\tau) = \delta(t-\tau), \quad \frac{\partial^2}{\partial t^2} G^-(t,\tau) = \delta(t-\tau), \quad \frac{\partial^2}{\partial t^2} G(t,\tau) = \delta(t-\tau).$$

Furthermore, we have:

- $G^+(t,\tau) = 0$ for all $t \le \tau$.
- $G^-(t,\tau) = 0$ for all $t \ge \tau$.

The retarded and advanced Green's functions determine the Green's function. Explicitly, for all $t, \tau \in \mathbb{R}$,

$$G(t,\tau) = \begin{cases} G^+(t,\tau) & \text{if} \quad \tau \ge 0, \\ G^-(t,\tau) & \text{if} \quad \tau < 0. \end{cases}$$

The corresponding notions for the motion of a quantum particle on the real line will be considered in Sect. 8.5 on page 729.

Fundamental Solutions in the Energy Space

Our next goal is to compute the fundamental solutions t_+, t_- and $\frac{1}{2}|t|$ by using the Fourier transform.

The formal Fourier transform. The Fourier transform reads as[34]

$$\hat{x}(E) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}Et} x(t) dt, \qquad E \in \mathbb{R}$$

together with the inverse transform $x(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{x}(E) \mathrm{e}^{-\mathrm{i}Et} dE$ for all $t \in \mathbb{R}$. This implies

$$\frac{d}{dt} x(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (-\mathrm{i}E) \hat{x}(E) \mathrm{e}^{-\mathrm{i}Et} dt, \qquad t \in \mathbb{R}.$$

Thus, the differentiation operator $\frac{d}{dt}$ is transformed into multiplication by the factor $-\mathrm{i}E$.

As in most textbooks in physics, we will start by using the Fourier transform in a formal way. That is, we will use the usual rules for the classical Fourier transform without worrying about the existence of the classical Fourier transform. However, this will cause trouble. In a next step we will use the theory of tempered distributions in order to get the rigorous approach. This way we will understand why it is important to consider retarded (resp. advanced) fundamental solutions.

(i) First method (formal fundamental solution): Applying the formal Fourier transform to $\ddot{x} = F$, we get $-E^2 \hat{x}(E) = \hat{F}(E)$. Hence

$$\hat{x}(E) = -\frac{\hat{F}(E)}{E^2} \qquad \text{for all} \quad E \in \mathbb{R}.$$

If $\hat{F} \in \mathcal{S}(\mathbb{R})$ and $\hat{F}(E) = O(E^2)$ as $E \to +0$, then the tempered distribution corresponding to $\hat{x}$ reads as

[34] We write $\hat{x}$ instead of $\mathcal{F}(x)$.

$$\hat{x}(\varphi) = \int_{-\infty}^{\infty} \hat{x}(E)\varphi(E)dE \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}). \tag{2.63}$$

In the special case where $F = \delta$, we formally obtain

$$\hat{F}(E) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}Et}\delta(t)dt = \frac{1}{\sqrt{2\pi}}.$$

This yields the formal fundamental solution

$$\boxed{\hat{x}(E) = -\frac{1}{\sqrt{2\pi}\, E^2}, \qquad E \in \mathbb{R}} \tag{2.64}$$

of the differential operator $\frac{d^2}{dt^2}$. However, this is not a tempered distribution in the sense of (2.63). In fact, one has to add correction terms in order to get the rigorous expression (2.67) below.

(ii) Second method (adiabatic regularization): Suppose that $F(t) = 0$ for all $t < 0$, that is, the force F vanishes at all points in time $t < 0$. Fix $\varepsilon > 0$. Let $x = x(t)$ be a solution of $\ddot{x} = F$. Set $x_\varepsilon(t) := x(t)\mathrm{e}^{-\varepsilon t}$ and $F_\varepsilon(t) := F(t)\mathrm{e}^{-\varepsilon t}$ for all $t \in \mathbb{R}$. Then

$$\left(\frac{d}{dt} + \varepsilon\right)^2 x_\varepsilon(t) = F_\varepsilon(t) \qquad \text{for all} \quad t \in \mathbb{R}.$$

Formal Fourier transform yields $(-\mathrm{i}E + \varepsilon)^2 \hat{x}_\varepsilon(E) = \hat{F}_\varepsilon(E)$. Hence

$$\hat{x}_\varepsilon(E) = -\frac{\hat{F}_\varepsilon(E)}{(E + \mathrm{i}\varepsilon)^2} \qquad \text{for all} \quad E \in \mathbb{R}. \tag{2.65}$$

Choosing $F_\varepsilon := \delta$, we get the formal fundamental solution

$$\boxed{\hat{x}_\varepsilon(E) = -\frac{1}{\sqrt{2\pi}(E + \mathrm{i}\varepsilon)^2}, \qquad E \in \mathbb{R}} \tag{2.66}$$

of the differential operator $(\frac{d}{dt}+\varepsilon)^2$. Note that, in contrast to (2.64), $\hat{x}_\varepsilon$ is a tempered distribution if $\varepsilon > 0$.

The rigorous Fourier transform. (i) First method: According to Prop. 2.13 on page 86, the tempered distribution

$$-\frac{1}{\sqrt{2\pi}}\, |E|_*^{-2}$$

is the Fourier transform of the fundamental solution $\frac{1}{2}|t|$ of $\frac{d^2}{dt^2}$. This is the rigorous formulation of the formal expression (2.64). By (2.60), we get

$$\begin{aligned} -\frac{1}{\sqrt{2\pi}}\, |E|_*^{-2}(\varphi) = {} & \sqrt{\frac{2}{\pi}}\, \varphi(0) - \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} \frac{\varphi(E) - \varphi(0) - \varphi'(0)E}{E^2}\, dE \\ & - \frac{1}{\sqrt{2\pi}} \int_{1}^{\infty} \frac{\varphi(E)dE}{E^2} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-1} \frac{\varphi(E)dE}{E^2} \end{aligned} \tag{2.67}$$

for all test functions $\varphi \in \mathcal{S}(\mathbb{R})$. Note that this distribution does not correspond to a classical function, as in (2.63).

(ii) Second method: By Prop. 2.12 on page 86, the limit

$$-\frac{1}{\sqrt{2\pi}}\,(E+0_+\mathrm{i})^{-2} = -\lim_{\varepsilon\to+0}\frac{1}{\sqrt{2\pi}}\,(E+\varepsilon\mathrm{i})^{-2}$$

exists, in the sense of tempered distributions. By Prop. 2.13 on page 86, this limit is the Fourier transform of the retarded fundamental solution t_+ of $\frac{d^2}{dt^2}$. This is the rigorous formulation of the adiabatic limit to the expression (2.66).

The rigorous Fourier transform of the advanced fundamental solution t_- is given by

$$-\frac{1}{\sqrt{2\pi}}\,(E-0_+\mathrm{i})^{-2} = -\lim_{\varepsilon\to+0}\frac{1}{\sqrt{2\pi}}\,(E-\varepsilon\mathrm{i})^{-2}.$$

The rigorous Laplace transform. Suppose that $F\in\mathcal{S}(\mathbb{R})$ together with $F(t)=0$ for all $t<0$. Let $x=x(t)$ be the classical solution of (2.61) with the homogeneous initial condition $a=v=0$. Then the function $x=x(t)$ is at most of polynomial growth. Therefore, the integral

$$\hat{x}_\varepsilon(E) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x(t)\mathrm{e}^{-\varepsilon t}\mathrm{e}^{\mathrm{i}Et}dt$$

exists for all $E\in\mathbb{R},\varepsilon>0$. The damping factor $\mathrm{e}^{-\varepsilon t}$ enforces the convergence of the integral. Since $x(t)=0$ if $t<0$, this is equal to the Laplace transform

$$\sqrt{2\pi}\,\hat{x}_\varepsilon(E) = \int_0^{\infty} x(t)\mathrm{e}^{-\varepsilon t}\mathrm{e}^{\mathrm{i}Et}dt.$$

It follows from $\ddot{x}=F$ that

$$\int_0^{\infty}\big(\ddot{x}(t)-F(t)\big)\,\mathrm{e}^{-\varepsilon t}\mathrm{e}^{\mathrm{i}Et}dt = 0.$$

Since $x(0)=0$ and $\dot{x}(0)=0$, integration by parts yields

$$\int_0^{\infty}\big((\mathrm{i}E-\varepsilon)^2x(t)-F(t)\big)\mathrm{e}^{-\varepsilon t}\mathrm{e}^{\mathrm{i}Et}dt = 0.$$

Hence $(\mathrm{i}E-\varepsilon)^2\hat{x}_\varepsilon(E)-\hat{F}_\varepsilon(E)=0$. This implies

$$\hat{x}_\varepsilon(E) = -\frac{\hat{F}_\varepsilon(E)}{(E+\varepsilon\mathrm{i})^2},\qquad E\in\mathbb{R}$$

which coincides with (2.65).

Generalization to definite and indefinite quadratic forms. Let us fix the integers m and n with $1\le m\le n$. For the quadratic form

$$\mathcal{P}(x) := x_1^2+x_2^2+\ldots+x_m^2-x_{m+1}^2-\ldots-x_n^2,$$

we define

$$\mathcal{P}_+(x) := \begin{cases}\mathcal{P}(x) & \text{if}\quad P(x)\ge 0,\\ 0 & \text{if}\quad P(x)<0\end{cases}$$

together with

$$P_+^\lambda(\varphi) := \int_{\mathbb{R}^n}\mathcal{P}_+^\lambda(x)\varphi(x)d^nx.$$

This way it is possible to generalize the considerations above to higher dimensions. In particular, the special case where $n=4$ and $m=3$ corresponds to wave processes

in the 3-dimensional (x_1, x_2, x_3)-space with time t and $x_4 = ct$ (c denotes the velocity of light in a vacuum).

In a fundamental paper from 1948, Marcel Riesz used Riemann–Liouville integrals and its higher-dimensional variants in order to construct Green's functions and the solution of the initial-value problem (also called Cauchy problem) for wave equations.[35] This approach represents a far-reaching generalization of the simple considerations discussed above to quantum field theory. We will study this in a later volume. At this point, we refer to:

I. Gelfand and G. Shilov, Generalized Functions, Vol. 1, Academic Press, New York, 1964.

A. Komech, Linear Partial Differential Equations with Constant Coefficients. In: Yu. Egorov et al., Elements of the Modern Theory of Partial Differential Equations, Springer, New York, 1999, pp. 121–256.

N. Ortner and P. Wagner, Distribution-Valued Analytic Functions: Theory and Applications, Lecture Notes 37/2008, Max Planck Institute for Mathematics in the Sciences, Leipzig, 2008.
Internet: `http://www.mis.de/preprints/ln/lecturenote-3708`

2.2.15 Hints for Further Reading.

Many results on algebraic Feynman integrals can be found in

V. Smirnov, Feynman Integral Calculus, Springer, Berlin, 2006.

Furthermore, we recommend:

I. Gradshtein and I. Ryshik, Tables of Integrals, Series, and Products, Academic Press, New York, 1980.

A. Prudnikov, Yu. Brychkov, O. Manichev, Integrals and Series, Vols. 1–5, Gordon and Breach, New York, 1986.

L. Landau and E. Lifshitz, Course of Theoretical Physics, Vol. 4: Quantum Electrodynamics, Butterworth–Heinemann, Oxford, 1982 (Chap. XII).

N. Bogoliubov and D. Shirkov, Quantum Fields. Lectures given at the Moscow Lomonossov University, Benjamin, Reading, Massachusetts, 1983.

M. Veltman, Diagrammatica: the Path to Feynman Diagrams, Cambridge University Press, 1995.

V. Radanovic, Problem Book in Quantum Field Theory, Springer, New York, 2006.

A. Grozin, QED (Quantum Electrodynamics) and QCD (Quantum Chromodynamics): Practical Calculation and Renormalization of One-and Multi-Loop Feynman Diagrams, World Scientific Singapore, 2007.

W. McComb, Renormalization Methods: A Guide for Beginners, Oxford University Press, 2007.

[35] M. Riesz, The Riemann–Liouville integral and the Cauchy problem, Acta Math. **81** (1948), 1–223 (in French). See also C. Bär et al., Wave Equations on Lorentzian Manifolds and Quantization, European Mathematical Society, 2007.

2.3 Further Regularization Methods in Mathematics

In physical experiments, physicists measure finite numbers. In contrast to this, the theory frequently generates infinite expressions.

> *The main task is to extract relevant finite information from infinite expressions.*

In the history of mathematics, mathematicians frequently encountered this problem.[36] Let us discuss some important approaches.

2.3.1 Euler's Philosophy

One of the greatest masters in mathematics, Leonhard Euler, relied on formal analytic identities. He used the principle that the sum of a divergent series is the value of the function from which the series is derived. For example, consider the geometric series

$$1 + x + x^2 + \ldots = \frac{1}{1-x}. \tag{2.68}$$

According to Euler, we define

$$[1 + x + x^2 + \ldots]_{\mathrm{reg}} := \frac{1}{1-x} \qquad \text{for all} \quad x \in \mathbb{C} \setminus \{1\}.$$

We call this the regularized value of the series $1 + x + x^2 + \ldots$[37]. We have the following consistency theorem:

> *If the series* (2.68) *is convergent, then its sum coincides with the regularized value.*

However, this method is also able to assign finite values to a divergent series. For example, if we choose $x = -1$ and $x = 2$, then

$$[1 - 1 + 1 - 1 + \ldots]_{\mathrm{reg}} = \tfrac{1}{2},$$

and $[1 + 2 + 2^2 + 2^3 + \ldots]_{\mathrm{reg}} = -1$. In the case where $x = 1$, this simple regularization method fails. Then we have to use the more sophisticated zeta function regularization discussed on page 55. By this method, we assign the value $\zeta(0) = -\frac{1}{2}$ to the divergent series $1 + 1 + 1 + \ldots$

As a second example for Euler's philosophy, consider the formal power series expansion

$$f(x) = x - 1!x^2 + 2!x^3 - 3!x^4 + \ldots, \tag{2.69}$$

[36] For the history of this topic full of errors and pitfalls, we refer to M. Kline, Mathematical Thought from Ancient to Modern Times, Vol. 3, Sect. 47, Oxford University Press, 1972. We also refer to the following classic monographs:
G. Hardy, Divergent Series, Clarendon Press, Oxford, 1949.
K. Knopp, Theory and Applications of Infinite Series, Dover, New York, 1989.

[37] In modern language, the power series expansion (2.68) converges for all complex numbers x with $|x| < 1$. This yields the holomorphic function $f(x) := \frac{1}{1-x}$ on the open unit disc. Finally, the function $x \mapsto \frac{1}{1-x}$ on the pointed complex plane $\mathbb{C} \setminus \{1\}$ is the analytic continuation of f.

which is divergent for all nonzero complex numbers x. Formal differentiation yields

$$f'(x) = 1 - 2!x + 3!x^2 - 4!x^3 + \dots$$

Hence

$$x^2 f'(x) + f(x) = x, \qquad f(0) = 0.$$

Surprisingly enough, this differential equation has the rigorous solution

$$f(x) = x \int_0^\infty \frac{\mathrm{e}^{-t}}{1+xt}\, dt \qquad \text{for all} \quad x \geq 0. \tag{2.70}$$

Note that this integral exists for all $x \geq 0$. Therefore, Euler assigned the function f to the divergent series (2.69). In particular, choosing $x = 1$, we get the finite regularization

$$[2! - 3! + 4! - \dots]_{\mathrm{reg}} := \int_0^\infty \frac{\mathrm{e}^{-t}}{1+t}\, dt$$

of the divergent series $2! - 3! + 4! - \dots$

2.3.2 Adiabatic Regularization of Divergent Series

Suppose we are given the series

$$a_0 + a_1 + a_2 + \dots \tag{2.71}$$

with complex coefficients $a_0, a_1, a_2, \dots$ We assign the perturbed series

$$a_0 + a_1 x + a_2 x^2 + \dots, \tag{2.72}$$

and we assume that there exists a number $\varepsilon > 0$ such that the perturbed series (2.72) is convergent for all $x \in]1-\varepsilon, 1[$. Now we let the perturbation go to zero, that is, $x \to 1 - 0$. We define

$$\left[\sum_{k=0}^\infty a_k\right]_{\mathrm{reg}} := \lim_{x\to 1-0} \sum_{k=0}^\infty a_k x^k$$

if the limit exists. Physicists call this adiabatic regularization. The idea is to perturb a given physical system and to study the limit when the perturbation goes to zero. Mathematicians call this Abel summation. For example,

$$1 - 1 + 1 - \dots = \lim_{x\to 1-0} (1 - x + x^2 - \dots) = \lim_{x\to 1-0} \frac{1}{1+x} = \frac{1}{2}.$$

Abel proved the consistency of this method:

> *If the series $\sum_{k=0}^\infty a_k$ is convergent, then its sum is equal to the adiabatic regularization.*

2.3.3 Adiabatic Regularization of Oscillating Integrals

Fix the large angular frequency $\omega_{\max}$ and the large time interval $[-\frac{T}{2}, \frac{T}{2}]$. Let us consider the following three integrals:

(i) Almost white noise:

$$\delta_{\omega_{\max}}(t) := \frac{1}{2\pi} \int_{-\omega_{\max}}^{\omega_{\max}} \mathrm{e}^{\mathrm{i}\omega t} d\omega \qquad \text{for all} \quad \text{times} \ \ t \in \mathbb{R}.$$

(ii) White noise (Dirac's delta function):

$$\delta(t) := \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\omega t} d\omega \qquad \text{for all} \quad \text{times} \ \ t \in \mathbb{R}.$$

(iii) Time average of a harmonic oscillation with angular frequency ω:

$$m(\omega) := \lim_{T \to +\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \mathrm{e}^{\mathrm{i}\omega t} dt = \begin{cases} 1 & \text{if} \quad \omega = 0, \\ 0 & \text{if} \quad \omega \neq 0. \end{cases}$$

.

It follows from $\frac{1}{2\pi} \int_{-a}^{a} \mathrm{e}^{\mathrm{i}\omega t} d\omega = \frac{1}{2\pi i t} \cdot \mathrm{e}^{\mathrm{i}\omega t}|_{-a}^{a} = \frac{\sin at}{\pi t}$ for $t \neq 0$ that

$$\delta_{\omega_{\max}}(t) = \begin{cases} \frac{\sin \omega_{\max} t}{\pi t} & \text{if} \quad t \neq 0, \\ \omega_{\max} \cdot \sqrt{\frac{2}{\pi}} & \text{if} \quad t = 0. \end{cases}$$

The function $\delta_{\omega_{\max}}$ represents the superposition of harmonic waves with angular frequencies $\omega \in [-\omega_{\max}, \omega_{\max}]$, where each harmonic wave has the same amplitude. Physicists and engineers call this an approximation of white noise. From the physical point of view, it makes sense to consider the limit $\omega_{\max} \to +\infty$. If $t = 0$, then we get

$$\lim_{\omega_{\max} \to +\infty} \delta_{\omega_{\max}}(0) = +\infty.$$

However, if $t \neq 0$, then the limit $\lim_{\omega_{\max} \to +\infty} \delta_{\omega_{\max}}(t)$ does not exist because of strong oscillations. Therefore, it is necessary to regularize the situation. To this end, fix the regularization parameter $\varepsilon > 0$, and consider the following two integrals.

(a) Adiabatic Dirac delta function:

$$\delta_{\varepsilon}(t) := \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\omega t} \mathrm{e}^{-\varepsilon \omega^2} d\omega \qquad \text{for all} \quad t \in \mathbb{R}.$$

(b) Adiabatic time average of a harmonic oscillation:

$$m_{\varepsilon}(\omega) := \frac{\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\omega t} \mathrm{e}^{-\varepsilon t^2} dt}{\int_{-\infty}^{\infty} \mathrm{e}^{-\varepsilon t^2} dt} \qquad \text{for all} \quad \omega \in \mathbb{R}.$$

In order to compute these integrals, we use the key formula

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\omega t} \mathrm{e}^{-t^2/2} dt = \mathrm{e}^{-\omega^2/2},$$

telling us that the Gaussian function $t \mapsto \mathrm{e}^{-t^2/2}$ is invariant under the Fourier transformation. By rescaling, this implies

$$\delta_\varepsilon(t) = \frac{\mathrm{e}^{-t^2/4\varepsilon}}{\sqrt{4\pi\varepsilon}} \qquad \text{for all} \quad t \in \mathbb{R},$$

and

$$m_\varepsilon(\omega) = \mathrm{e}^{-\omega^2/4\varepsilon} \qquad \text{for all} \quad \omega \in \mathbb{R}.$$

Letting $\varepsilon \to +0$, we obtain the adiabatic limits

$$\lim_{\varepsilon \to +0} \delta_\varepsilon(t) = \begin{cases} +\infty & \text{if} \quad t = 0, \\ 0 & \text{if} \quad t \neq 0, \end{cases}$$

and $\lim_{\varepsilon \to +0} m_\varepsilon(\omega) = m(\omega)$ for all $\omega \in \mathbb{R}$. Note the crucial fact that

Adiabatic averaging cancels infinities.

This is the secret behind the incredible success of path integral methods in quantum physics.

The Feynman approach to quantum field theory via path integrals is based on formal adiabatic averages.

The basic ideas are studied in Chap. 7 of Vol. I, in terms of a finite-dimensional rigorous setting.

2.3.4 Regularization by Averaging

For a sequence (s_n) of complex numbers, Cauchy proved that

$$\lim_{n \to +\infty} \frac{s_0 + s_1 + \ldots + s_n}{n+1} = \lim_{n \to \infty} s_n$$

if the right-hand limit exists. This method can be successively applied to the sequence (s_n) of partial sums of an infinite series $\sum_{k=0}^{\infty} a_k$ with complex terms $a_0, a_1, \ldots$ Here, $s_n := a_0 + a_1 + \ldots a_n$. As usual, we define

$$\sum_{k=0}^{\infty} a_k := \lim_{n \to \infty} s_n$$

if the right-hand limit exists. More generally, we define the averaged regularization

$$\left[\sum_{k=0}^{\infty} a_k\right]_{\mathrm{reg,average}} := \lim_{n \to \infty} \frac{s_0 + s_1 + \ldots + s_n}{n+1}. \tag{2.73}$$

According to Cauchy, we have the following consistency theorem: If the series $\sum_{k=0}^{\infty} a_k$ is convergent, then its sum coincides with the regularization in the sense of (2.73). This summation method and its iterations were studied by Otto Hölder (1859–1937) in 1882. Mathematicians call this the Hölder summation. For example, consider the formula

$$\frac{1}{(1+x)^2} = 1 - 2x + 3x^2 - 4x^3 + \ldots,$$

which follows from $-\frac{1}{1+x} = -1 + x - x^2 + \ldots$ by formal differentiation. Setting $x = 1$, Euler assigned the value $\frac{1}{4}$ to the divergent series $1 - 2 + 3 - 4 + \ldots$. The Hölder summation of $1 - 2 + 3 - 4 + \ldots$ proceeds as follows. The partial sums are

$$s_0 = 1, \quad s_1 = -1, \quad s_2 = 2, \quad s_3 = -2, \dots$$

For the averaged partial sums $s_n^{(1)} := \frac{1}{n+1}(s_0 + ... + s_n)$, we get

$$s_0^{(1)} = 1, \quad s_1^{(1)} = 0, \quad s_2^{(1)} = \tfrac{2}{3}, \quad s_3^{(1)} = 0, \dots$$

The sequence $(s_n^{(1)})$ is not convergent. Averaging this sequence again, we get $s_n^{(2)} :== \frac{1}{n+1}(s_0^{(1)} + ... + s_n^{(1)})$ with

$$s_0^{(2)} = 1, \quad s_1^{(2)} = \tfrac{1}{2}, \quad s_2^{(2)} = \tfrac{5}{9}, \quad s_3^{(2)} = \tfrac{5}{12}, \dots$$

It can be shown that the sequence $(s_n^{(2)})$ converges to $\frac{1}{4}$, which coincides with the Euler sum of the divergent series $1 - 2 + 3 - 4 + \dots$

The reconstruction theorem for the Fourier series. In the 19th century, mathematicians discovered that the convergence problem for the Fourier series is highly sophisticated. This strongly influenced the development of analysis and functional analysis in the 20th century. Let us describe the basic ideas.[38] Suppose we are given the 2π-periodic continuous function $f : \mathbb{R} \to \mathbb{C}$. We want to approximate the function f by the simpler 2π-periodic functions $x \mapsto \mathrm{e}^{inx}$ with $n = 0, 1, 2, \dots$. Motivated by Gauss' method of least squares, we consider the following minimum problem

$$\int_{-\pi}^{\pi} \left| f(x) - \sum_{n=-N}^{N} a_n \mathrm{e}^{inx} \right|^2 dx = \min!$$

for the unknown complex coefficients $a_0, a_1, a_{-1}, a_2, a_{-2}, \dots$ This problem has the unique solution

$$a_n := \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \mathrm{e}^{-inx} dx, \qquad n = -N, -N+1, \dots, N$$

for $N = 0, 1, 2, \dots$ The complex numbers a_n are called the Fourier coefficients of the function f. Since the function f is continuous, we expect that the Fourier series

$$f(x) = \sum_{n=-\infty}^{\infty} a_n \mathrm{e}^{inx} \tag{2.74}$$

is convergent for all $x \in \mathbb{R}$. Unfortunately, this is not true, as was shown by Du Bois-Reymond in 1871. There arises the following question: How can we extract the information about the function f from the knowledge of its Fourier coefficients? There are two answers.

(i) Mean square convergence:

$$\lim_{N \to \infty} \int_{-\pi}^{\pi} |f(x) - \sum_{n=-N}^{N} a_n \mathrm{e}^{inx}|^2 dx = 0.$$

(ii) Averaging (Fejér 1904): $\left[\sum_{n=-\infty}^{\infty} a_n \mathrm{e}^{inx}\right]_{\text{reg,average}} = f(x)$ for all $x \in \mathbb{R}$.

[38] The proofs can be found in Zeidler (1995a), quoted on page 1049.

More generally, the mean square convergence (i) is true for all measurable functions $f :]-\pi, \pi[\to \mathbb{C}$ with $\int_{-\pi}^{\pi} |f(x)|^2 dx < \infty$ because of the fact that the family $\{e^{inx}/\sqrt{2\pi}\}_{n\in\mathbb{Z}}$ forms a complete orthonormal system in the Hilbert space $L_2(-\pi, \pi)$.

Tauberian theorems. Let us consider an arbitrary series $\sum_{n=0}^{\infty} a_n$ with complex terms $a_0, a_1, a_2, \ldots$ Then the following implications hold:

> *classical convergence* $\Rightarrow$ *regularization by averaging* $\Rightarrow$ *adiabatic regularization.*

In particular, this tells us that if the series can be regularized by averaging, then it can also be regularized in the adiabatic sense, and the regularized values are the same. Conversely, we have the following convergence theorem:

> *The series* $\sum_{n=0}^{\infty} a_n$ *is convergent if the series can be regularized in the adiabatic sense (or by averaging in the sense of* (2.73)*) and we have the relation* $|a_n| = O(\frac{1}{n})$ *as* $n \to +\infty$.

This was proved by Hardy and Littlewood in the 1910s by generalizing a weaker theorem proved by Tauber in 1890.[39] Combining the Tauberian theorem above with the Fejér theorem, we get the following basic theorem:

> *If the* 2π*-periodic function* $f : \mathbb{R} \to \mathbb{C}$ *is continuous and its Fourier coefficients satisfy the condition* $|a_n| = O(\frac{1}{n})$ *as* $n \to \infty$*, then the Fourier series converges to* f *on the real line.*

2.3.5 Borel Regularization

In 1899 Borel investigated the following summation method.[40] Define

$$\left[\sum_{n=0}^{\infty} a_n\right]_{\text{reg,Borel}} := \int_0^{\infty} dt\, e^{-t} \sum_{n=0}^{\infty} \frac{a_n}{n!}\, t^n \tag{2.75}$$

if the series $\sum_{n=0}^{\infty} \frac{a_n}{n!} t^n$ converges for all real numbers t and the integral exists. In this case, we say that the series $\sum_{n=0}^{\infty} a_n$ is Borel regularizable (or Borel summable). We have the following consistency theorem:

> *If the series* $\sum_{n=0}^{\infty} a_n$ *is convergent, then it is Borel regularizable and its sum coincides with the regularized value.*

For example, let $-\infty < \Re(z) < 1$. Set $z := x + y\mathrm{i}$ with $x, y \in \mathbb{R}$. Then

$$\begin{aligned}[1 + z + z^2 + \ldots]_{\text{reg,Borel}} &= \int_0^{\infty} dt\, e^{-t} \sum_{n=0}^{\infty} \frac{(zt)^n}{n!} \\ &= \int_0^{\infty} e^{-(1-z)t} dt = \int_0^{\infty} e^{-(1-x)t} e^{\mathrm{i}yt} dt = \frac{1}{1-z}.\end{aligned}$$

This shows that Borel summability is able to construct analytic continuation. For example, we have the following theorem:

---

[39] Numerous theorems of this type can be found in the monograph by J. Korevaar, Tauberian Theory: A Century of Developments, Springer, Berlin, 2004. In Sect. 6.5.1 of Vol. I, we have used a Tauberian theorem for the Laplace transformation in order to give an elegant proof of the prime number theorem.

[40] E. Borel, Mémoire sur les séries divergent, Ann. Sci. École Normale Superieur **16**(3) (1899), 9–136 (in French).

If the power series expansion $f(z) := \sum_{n=0}^{\infty} a_n z^n$ *has a finite positive radius of convergence and is Borel summable at the nonzero point* $z = 2z_0$, *then the function* f *has an analytic continuation to the open disc of radius* $|z_0|$ *centered at the point* z_0.

This represents Borel's circle method of analytic continuation. As another important example, let us mention the Nevalinna–Sokal theorem:[41] We are given the positive numbers R and r. Let $U_R(R)$ denote the open disc of radius R centered at the point R of the complex plane.

Suppose that the function $f : U_R(R) \to \mathbb{C}$ *is holomorphic with the expansion* $f(z) = \sum_{k=0}^{n-1} a_k z^k + R_n(z), n = 1, 2, \ldots$, *for all* $z \in U_R(R)$ *and the uniform remainder estimate*

$$\sup_{z \in U_R(R)} \sup_{n \in \mathbb{N}} \frac{r^n |R_n(z)|}{|z|^n n!} < \infty.$$

Then we have the absolutely convergent integral representation

$$f(z) = \frac{1}{z} \int_0^{\infty} e^{-t/z} F(t) dt \qquad \text{for all} \quad z \in U_R(R).$$

Here, we set[42] $F(t) := \sum_{n=0}^{\infty} \frac{a_n}{n!}\, t^n$.

Poincaré's asymptotic series. We now want to study a class of divergent series which is very useful for computing problems in mathematics and physics. Let $x > 0$, and let $a_0, a_1, \ldots$ be complex numbers. We write

$$f(x) \sim a_0 + \frac{a_1}{x} + \frac{a_2}{x^2} + \ldots, \qquad x \to +\infty \tag{2.76}$$

iff the remainder $|f(x) - (a_0 + \frac{a_1}{x} + \ldots + \frac{a_n}{x^n})|$ has the order of magnitude $o(\frac{1}{x^n})$ as $x \to +\infty$ for all indices $n = 0, 1, 2, \ldots$ In this case, we say that the function f has the asymptotic series expansion $a_0 + \frac{a_1}{x} + \frac{a_2}{x^2} + \ldots$ as $x \to +\infty$. This notion was used by Poincaré in order to study the solutions of linear singular ordinary differential equations of Fuchsian type and the perturbed motion of planets in celestial mechanics.[43]

Let us study a simple example. Consider the integral

$$f(x) := \int_x^{\infty} \frac{\mathrm{e}^{x-t}}{t}\, dt, \qquad x > 0.$$

We are interested in the behavior of this integral for large values of x. Repeated integration by parts yields

[41] A. Sokal, An improvement of Watson's theorem on Borel summability, J. Math. Phys. **21**(2) (1980), 261–263.

[42] In addition, we obtain that the function F is holomorphic on the open circle $U_r(0)$ (of radius r centered at the origin) and has an analytic continuation to the open neighborhood $\{t \in \mathbb{C} : \mathrm{dist}(t, \mathbb{R}_+) < r\}$ of the positive real axis.

[43] H. Poincaré, Sur les intégrales irregulières des équations linéaires, Acta Mathematica **8** (1886), 294–344 (in French). Analogously, we write the symbol

$$f(x) \sim a_0 + a_1(x-a) + a_2(x-a)^2 + \ldots, \qquad x \to a$$

iff the remainder $|f(x) - (a_0 + a_1(x-a) + \ldots + a_n(x-a)^n)|$ has the order of magnitude $o(|x-a|^n)$ as $x \to a$ for all indices $n = 0, 1, 2, \ldots$.

$$f(x) = \frac{1}{x} - \frac{1!}{x^2} + \frac{2!}{x^2} - \ldots + \frac{(-1)^n n!}{x^{n+1}} + R_{n+1}(x), \qquad x > 0$$

with the remainder $R_{n+1}(x) := (-1)^{n+1}(n+1)! \int_x^\infty \mathrm{e}^{x-t} t^{-n-2}\, dt$. Obviously, since $x - t \le 0$ implies $\mathrm{e}^{x-t} \le 1$, we have

$$|R_{n+1}(x)| \le (n+1)! \int_x^\infty \frac{dt}{t^{n+2}} \le \frac{n!}{x^{n+1}}. \tag{2.77}$$

Therefore, we have the asymptotic expansion

$$f(x) \sim \frac{1}{x} - \frac{1!}{x^2} + \frac{2!}{x^2} - \ldots, \qquad x \to +\infty.$$

Observe that this series is divergent for all $x > 0$. Nevertheless, this expansion is very useful for calculations. For example, if we take $x \ge 2n$, then it follows from (2.77) and the formula

$$n! = \sqrt{2\pi n} \left(\frac{n}{\mathrm{e}}\right)^n \mathrm{e}^{\vartheta(n)/12n}, \quad 0 < \vartheta(n) < 1, \ \ n = 1, 2, \ldots \tag{2.78}$$

due to Stirling (1692–1770) that

$$\left| f(x) - \left(\frac{1}{x} - \frac{1!}{x^2} + \frac{2!}{x^2} - \ldots + \frac{(-1)^n n!}{x^{n+1}} \right) \right| \le \frac{1}{2^{n-1} \mathrm{e}^{n-1} \sqrt{n}}$$

for all $n = 1, 2, \ldots$ Furthermore, let us mention that the Stirling formula (2.78) is closely related to the following famous asymptotic series

$$\ln \Gamma(x+1) - \left(x + \tfrac{1}{2}\right) \ln x + x - \ln \sqrt{2\pi} \sim \sum_{k=0}^{\infty} \frac{B_{2k}}{(2k-1)2k} \cdot \frac{1}{x^{2k-1}}, \quad x \to +\infty.$$

Here $B_2, B_4, \ldots$ denote the Bernoulli numbers. In particular, this implies

$$n! \sim n^n \mathrm{e}^{-n} \sqrt{2\pi n} \left(1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{51840n^3} + \ldots\right), \quad n \to \infty.$$

For example, the value $n!$ for large n appears in the statistical mechanics of gases where n is of the magnitude of Avogadro's number (also called Loschmidt's number).[44] The Bernoulli numbers were introduced by Jakob Bernoulli (1654–1705) in order to study combinatorial problems in the theory of probability.

2.3.6 Hadamard's Finite Part of Divergent Integrals

Let $f : \mathbb{R} \to \mathbb{C}$ be a smooth function. Choose $a > 0$. We define

$$\left[\int_0^a \frac{f(x)}{x}\, dx\right]_{\text{reg,finite}} := \int_0^a \frac{f(x) - f(0)}{x}\, dx. \tag{2.79}$$

This is called the finite part of the integral $\int_0^a \frac{f(x)}{x}\, dx$. In order to motivate this definition, note that

[44] Two gram of hydrogen contain $6.022 \cdot 10^{23}$ hydrogen atoms. This fundamental number was first approximately computed by Loschmidt (1821–1895) in 1866. Avogadro lived from 1776 until 1856.

$$\int_\varepsilon^a \frac{dx}{x} = \ln a - \ln \varepsilon, \qquad 0 < \varepsilon < a.$$

Hence $\int_0^a \frac{dx}{x} = \lim_{\varepsilon \to +0} \int_\varepsilon^a \frac{dx}{x} = +\infty$. However, by Taylor expansion,

$$f(x) - f(0) = O(x), \qquad x \to 0.$$

Therefore, the integral on the right-hand side of (2.79) exists. If $f(0) = 0$, then the finite part of the integral coincides with the integral itself. Similarly, for $n = 1, 2, \ldots$, we define

$$\left[\int_0^a \frac{f(x)}{x^{n+1}}\, dx\right]_{\text{reg,finite}} := \int_0^a \frac{f(x) - f(0) - \sum_{k=1}^n \frac{f^{(k)}(0)}{k!} x^k}{x^{n+1}}\, dx.$$

Hadamard used such regularized integrals in his monograph on hyperbolic partial differential equations (e.g., wave equations) in order to represent the solutions by integral formulas and to overcome the highly singular behavior of the corresponding Green's functions.[45]

2.3.7 Infinite-Dimensional Gaussian Integrals and the Zeta Function Regularization

We want to show that the passage from a finite to an infinite number of degrees of freedom may cause mathematical trouble. As a model, let us consider the Gaussian integral and its infinite-dimensional limit which is called a functional integral (or Feynman path integral). Let $\{\lambda_k\}$ be a sequence of positive numbers with $0 < \lambda_1 \le \lambda_2 \le \lambda_3 \le \ldots$, and let $\{J_k\}$ be a sequence of real numbers. Let us choose the dimension $N = 1, 2, \ldots$ By Sect. 7.23.3 of Vol. I, we have the formula

$$\int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\sum_{k=1}^N \lambda_k \varphi_k^2} \mathrm{e}^{\mathrm{i}\sum_{k=1}^N J_k \varphi_k} \prod_{k=1}^N \frac{d\varphi_k}{\sqrt{2\pi}} = \mathrm{e}^{\frac{1}{2}\zeta'_{N,\Lambda}(0)} \mathrm{e}^{-\frac{1}{2}\sum_{k=1}^N \lambda_k^{-1} J_k^2}$$

with the discrete zeta function

$$\zeta_{N,\Lambda}(s) := \frac{1}{\lambda_1^s} + \ldots + \frac{1}{\lambda_N^s} \qquad \text{for all} \quad s \in \mathbb{C}.$$

Furthermore, $\mathrm{e}^{-\zeta'_{N,\Lambda}(0)} = \prod_{k=1}^N \lambda_k$. Now we want to study the limit $N \to \infty$. Formally, we get

$$\lim_{N\to\infty} \int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\sum_{k=1}^N \lambda_k \varphi_k^2} \mathrm{e}^{\mathrm{i}\sum_{k=1}^N J_k \varphi_k} \prod_{k=1}^N \frac{d\varphi_k}{\sqrt{2\pi}} = \left(\prod_{k=1}^\infty \lambda_k\right)^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}\sum_{k=1}^\infty \lambda_k^{-1} J_k^2}.$$

As a rule, the product $\Pi_{k=1}^\infty \lambda_k$ is infinite. For example, this is true if we choose $\lambda_k = k$ for all $k = 1, 2, \ldots$ In this case, we have

$$\zeta_{N,\Lambda}(s) = \frac{1}{1^s} + \frac{1}{2^s} + \ldots + \frac{1}{N^s}.$$

The limit

[45] See the footnote on page 26.

$$\zeta(s) = \sum_{k=1}^{\infty} \frac{1}{k^s}$$

is divergent for all complex numbers s with $\Re(s) < 1$. Therefore, the value $\zeta(s)$ is meaningless near $s = 0$, in the usual sense. However, there exists an analytical extension of the Riemann zeta function to a meromorphic function ζ on $\mathbb{C} \setminus \{0\}$, and we will use this extension. In this sense, we have[46]

$$\zeta(0) = -\frac{1}{2} \quad \text{and} \quad \zeta'(0) = -\ln 2\pi.$$

We now choose the following definition:

$$\left[\int_{\mathbb{R}^\infty} \mathrm{e}^{-\frac{1}{2}\sum_{k=1}^{\infty} k\varphi_k^2} \mathrm{e}^{\mathrm{i}\sum_{k=1}^{\infty} J_k\varphi_k} \prod_{k=1}^{\infty} \frac{d\varphi_k}{\sqrt{2\pi}}\right]_{\mathrm{reg}} := \mathrm{e}^{\frac{1}{2}\zeta'(0)} \mathrm{e}^{-\frac{1}{2}\sum_{k=1}^{\infty} k^{-1}J_k^2}.$$

Here, $\mathrm{e}^{\frac{1}{2}\zeta'(0)} = \frac{1}{\sqrt{2\pi}}$, and we assume that the series $\sum_{k=1}^{\infty} k^{-1}J_k^2$ is convergent.

In the more general case where $0 < \lambda_1 \leq \lambda_2 \leq \ldots$, we introduce the Dirichlet series

$$\zeta_\Lambda(s) := \sum_{k=1}^{\infty} \mathrm{e}^{-s\ln\lambda_k}, \qquad \Re(s) > \sigma_0 \tag{2.80}$$

where $\sigma_0 := \limsup_{N\to+\infty} \frac{\ln N}{\ln \lambda_N}$, and we assume that σ_0 is finite. Then the series (2.80) is convergent (resp. divergent) for all the complex numbers s with $\Re(s) > \sigma_0$ (resp. $\Re(s) < \sigma_0$), and the function ζ_Λ has a singularity at the point $s = \sigma_0$. Now we assume that the function ζ_Λ can be analytically extended to a function ζ_Λ which is holomorphic in a neighborhood of $s = 0$. We also assume that the series $\sum_{k=1}^{\infty} \lambda_k^{-1} J_k^2$ is convergent. Then we define the regularized infinite-dimensional Gaussian integral by

$$\left[\int_{\mathbb{R}^\infty} \mathrm{e}^{-\frac{1}{2}\sum_{k=1}^{\infty} \lambda_k\varphi_k^2} \mathrm{e}^{\mathrm{i}\sum_{k=1}^{\infty} J_k\varphi_k} \prod_{k=1}^{\infty} \frac{d\varphi_k}{\sqrt{2\pi}}\right]_{\mathrm{reg}} := \mathrm{e}^{\frac{1}{2}\zeta_\Lambda'(0)} \mathrm{e}^{-\frac{1}{2}\sum_{k=1}^{\infty} \lambda_k^{-1}J_k^2}.$$

2.4 Trouble in Mathematics

> Formal manipulations in mathematics can lead to completely wrong results.
> Folklore

2.4.1 Interchanging Limits

In the mathematics and physics literature, one frequently encounters the interchange of limits. By considering three simple examples, we want to illustrate the crucial fact that the formal interchange of limits can lead to wrong results.

[46] For the proofs of all the following statements, see H. Edwards, Riemann's Zeta Function, Academic Press, New York, 1974, and T. Apostol, Introduction to Analytic Number Theory, Springer-Verlag, New York, 1986 (Dirichlet series).

(i) We want to show that

$$\boxed{\lim_{x\to 0}\lim_{n\to+\infty}\sum_{k=1}^{n}\frac{x^2}{(1+x^2)^k}\neq\lim_{n\to\infty}\lim_{x\to 0}\sum_{k=1}^{n}\frac{x^2}{(1+x^2)^k}.} \tag{2.81}$$

To this end, we consider the infinite series

$$f(x):=\sum_{k=1}^{\infty}\frac{x^2}{(1+x^2)^k},\qquad x\in\mathbb{R}.$$

If $x=0$, then $f(0)=0$. If $x\neq 0$, then the geometric series tells us that

$$f(x)=x^2\left(\frac{1}{1-\frac{1}{1+x^2}}-1\right)=1.$$

Consequently, we obtain

$$\lim_{x\to 0}\lim_{n\to+\infty}\sum_{k=1}^{n}\frac{x^2}{(1+x^2)^k}=\lim_{x\to 0}f(x)=1,$$

and $\lim_{n\to\infty}\lim_{x\to 0}\sum_{k=1}^{n}\frac{x^2}{(1+x^2)^k}=0$. This implies (2.81).

(ii) Let $x\in\mathbb{R}$. Set $f_n(x):=\frac{1}{n}\arctan(n^2x)$ for all positive integers n. Then $\lim_{n\to\infty}f_n(x)=0$. Moreover,

$$f_n'(x):=\frac{n}{1+n^4x^2}.$$

Hence $\lim_{n\to\infty}f_n'(0)=\infty$. Consequently,

$$\boxed{\frac{d}{dx}\lim_{n\to\infty}f_n(x)\neq\lim_{n\to\infty}\frac{d}{dx}f_n(x)\quad\text{at the point}\quad x=0.}$$

(iii) Set $f(x,y):=(2-xy)xy\cdot\mathrm{e}^{-xy}$. We want to show that

$$\boxed{\int_0^1 dy\int_0^\infty dx\,f(x,y)\neq\int_0^\infty dx\int_0^1 dy\,f(x,y).} \tag{2.82}$$

In fact, note that $\frac{d}{dz}\left(z^2\mathrm{e}^{-z}\right)=(2-z)z\mathrm{e}^{-z}$. If $y>0$, then

$$\int_0^\infty f(x,y)dx=\frac{1}{y}(xy)^2\mathrm{e}^{-xy}\Big|_0^\infty=0.$$

If $x>0$, then $\int_0^1 f(x,y)dy=\frac{1}{x}(xy)^2\mathrm{e}^{-xy}\Big|_0^1=x\mathrm{e}^{-x}$. Integration by parts yields

$$\int_0^\infty x\mathrm{e}^{-x}dx=-x\mathrm{e}^{-x}\Big|_0^\infty+\int_0^\infty\mathrm{e}^{-x}dx=1.$$

This proves the claim (2.82).

In the mathematical literature, one proves theorems which guarantee the interchange of limits.

Roughly speaking, one needs uniform convergence.

As an introduction, we recommend the textbook by V. Zorich, Analysis I, II, Springer, New York, 2003.

2.4.2 The Ambiguity of Regularization Methods

Unfortunately, it is possible that different regularization methods yield different results. As a simple example, consider the Euler series $1 - 1 + 1 - \dots$. The two functions

$$f_1(x) := \frac{1}{1+x} = 1 - x + x^2 - x^3 + \dots, \qquad |x| < 1$$

and

$$f_2(x) := \frac{1-x^2}{1-x^3} = \frac{1+x}{1+x+x^2} = 1 - x^2 + x^4 - \dots, \qquad |x| < 1$$

represent adiabatic regularization of the divergent series $1 - 1 + 1 - \dots$ As usual, define

$$[1 - 1 + 1 - 1 + \dots]_{j,\mathrm{reg}} := f_j(1), \qquad j = 1, 2.$$

Then we get the different values $f_1(1) = \frac{1}{2}$ and $f_2(1) = \frac{2}{3}$.

2.4.3 Pseudo-Convergence

In quantum field theory, physicists use the method of perturbation theory in order to compute physical quantities (e.g., the cross section of a scattering process) as formal power series expansions of the type

$$\boxed{\sigma(\kappa) = a_0 + a_1\kappa + a_2\kappa^2 + \dots, \qquad \kappa > 0} \tag{2.83}$$

where the positive parameter κ represents the so-called coupling constant, which measures the strength of interaction. There arises the following question:

> *What can we say about the mathematical meaning of* (2.83)*. Is this more than a formal power series expansion?*

In 1951 Dyson invented a heuristic physical argument in order to rule out the convergence of (2.83) in a neighborhood of the point $\kappa = 0$. Roughly speaking, he argued as follows: If the power series expansion (2.83) has a positive radius of convergence around the origin, then it also converges for small negative values of the coupling constant κ. However, such a negative coupling constant corresponds to repelling forces which destroy the physical system.

An example of pseudo-convergence. Suppose that the methods of formal perturbation theory produce the following quantity

$$\sigma(\kappa) = 1 + \kappa + \frac{\kappa^2}{2!} + \dots + \frac{\kappa^6}{6!} + \kappa^7 - 1!\kappa^8 + 2!\kappa^9 - 3!\kappa^{10} + \dots, \quad \kappa > 0.$$

If we compute the function f up to order six, then we get

$$\sigma(\kappa) = 1 + \kappa + \frac{\kappa^2}{2!} + \dots + \frac{\kappa^6}{6!}.$$

However, if we compute higher-order terms, then we encounter the divergent series $\kappa^7 - 1!\kappa^8 + 2!\kappa^9 - 3!\kappa^{10} + \dots$ We call this pseudo-convergence. According to (2.70) on page 94, a possible regularization reads as

$$f_{\mathrm{reg}}(\kappa) := 1 + \kappa + \frac{\kappa^2}{2!} + \dots \frac{\kappa^6}{6!} + \kappa^7 \int_0^\infty \frac{\mathrm{e}^{-t}}{1+\kappa t}\, dt \qquad \text{for all} \quad \kappa \geq 0.$$

The comparison with the results of physical experiments decides whether this regularization is meaningful from the physical point of view.

Asymptotic series and the Ritt theorem. An interesting mathematical result tells us that each formal power series expansion is the asymptotic series of a function f with respect to small positive values. Moreover, the function f is analytic on an appropriate open neighborhood of the positive real line, where the neighborhood has the shape of an angular sector. The precise result reads as follows. We are given the formal power series expansion

$$a_0 + a_1\kappa + a_2\kappa^2 + \dots$$

with complex coefficients $a_0, a_1, a_2, \dots$ We choose the angular sector

$$\mathcal{S} := \{z \in \mathbb{C} : z = r\mathrm{e}^{\mathrm{i}\varphi},\ -\gamma < \varphi < \gamma,\ r > 0\}$$

where γ is an arbitrary, but fixed number in the open interval $]0, \frac{\pi}{2}[$.

Then there exists an analytic function $\sigma : \mathcal{S} \to \mathbb{C}$ such that $\sigma(\kappa) \sim a_0 + a_1\kappa + a_2\kappa^2 + \dots$ as $\kappa \to 0$ on $\mathcal{S}$.

Explicitly, this means that for each index $n = 0, 1, 2, \dots$, we have

$$\sigma(\kappa) = a_0 + a_1\kappa + a_2\kappa^2 + \dots + a_n\kappa^n + R_n(\kappa)$$

where the remainder R_n has the property $\kappa^n R_n(\kappa) \to 0$ as $\kappa \to 0$ on $\mathcal{S}$.

The Borel theorem. Suppose that we are given the formal power series expansion

$$a_0 + a_1\kappa + a_2\kappa^2 + \dots$$

with real coefficients $a_0, a_1, a_2, \dots$. Let $r > 0$.

Then there exists a smooth function $\sigma :]-r, r[\to \mathbb{R}$ such that

$$\frac{d^n\sigma}{d\kappa^n}(0) = a_n \qquad \text{for all} \quad n = 0, 1, 2, \dots$$

In addition, the function σ is real-analytic.[47] *on the pointed interval $]-r, r] \setminus \{0\}$.*

The proof of the Borel theorem based on the proof of the Ritt theorem can be found in R. Remmert, Theory of Complex Functions, p. 300, Springer, New York, 1991.

2.4.4 Ill-Posed Problems

> Distinguish carefully between well-posed and ill-posed problems.
> Folklore

By using the simple example (2.88) below, we want to show that an uncritical use of the method of perturbation theory may lead to wrong results if the problem is ill-posed. This is a possible paradigm for quantum field theory. According to Hadamard (1865–1863), a mathematical problem is called well-posed iff it has a unique solution which depends continuously on the data of the problem. Otherwise, the problem is called ill-posed.

Roughly speaking, ill-posed problems refer to incomplete information.

[47] This means that the function σ can be represented as a local power series expansion in a sufficiently small neighborhood of each nonzero point in the interval $]-r, r[$.

For example, if we measure the gravitational field of earth by satellites, then the determination of the mass distribution of the earth by the measured data represents an ill-posed problem. Other important applications are computer tomography and inverse scattering problems in quantum mechanics (the determination of the potential by using scattering data).[48] As a prototype, let us consider the minimum problem

$$\boxed{||A(\alpha)\psi - f||^2 = \min!, \qquad \psi \in X} \tag{2.84}$$

which is motivated by the Gaussian method of least squares. Here, for any $\alpha \in]0,1[$, the operator $A(\alpha) : X \to X$ is linear on the real Hilbert space X of finite dimension $m = 1, 2, \ldots$ For fixed parameter $\alpha \in]0,1[$, we are given $f \in X$. We are looking for $\psi \in X$.

Well-posed problem. Let $L(X,X)$ denote the space of linear operators $B : X \to X$ equipped with the operator norm $||B|| := \max_{||\psi|| \leq 1} ||B\psi||$.

Proposition 2.14 *Suppose that the inverse operator $A(\alpha)^{-1} : X \to X$ exists for each parameter $\alpha \in]0,1[$, and $\alpha \mapsto A(\alpha)$ is a continuous map from $]0,1[$ into $L(X,X)$.*[49] *Then the minimum problem (2.84) is well-posed. For each parameter $\alpha \in]0,1[$, the unique solution is given by $\psi(\alpha) = A(\alpha)^{-1} f$.*

Proof. (I) Existence. Note that $||A(\alpha)\psi - f||^2 \geq 0$ for all $\psi \in X$, and we have $||A(\alpha)\psi(\alpha) - f||^2 = 0$.

(II) Uniqueness. $||A(\alpha)\psi - f||^2 = 0$ implies $A(\alpha)\psi - f = 0$, and hence $\psi = A(\alpha)^{-1} f$.

(III) Continuity. The continuity of the map $\alpha \to A(\alpha)$ from $]0,1[$ into $L(X,X)$ implies the continuity of the map $\alpha \mapsto A^{-1}(\alpha)$ from $]0,1[$ into $L(X,X)$ (see Zeidler (1995), Vol. 1, Sect. 1.23). Therefore, $\alpha_n \to \alpha$ and $f_n \to f$ on X as $n \to \infty$ imply $A(\alpha_n)^{-1} f_n \to A(\alpha)^{-1} f$ on X as $n \to \infty$, for the corresponding solutions. □

The generalized inverse operator. Let $A : X \to X$ be a linear operator on the real Hilbert space X. The trick is to introduce the operator

$$C := A^\dagger A.$$

Because of $C^\dagger = A^\dagger (A^\dagger)^\dagger = C$, the operator C is self-adjoint. Therefore, it possesses a complete orthonormal system $\psi_1, \ldots \psi_m$ of eigenvectors with the eigenvalues $\mu_1, \ldots, \mu_m$. Obviously, $\mu_k = \langle \psi_k | C\psi_k \rangle = \langle A\psi_k | A\psi_k \rangle \geq 0$. The values $\sigma_k := \sqrt{\mu_k}$ are called the singular values of the operator A.[50] Let us choose the indices in such a way that $\sigma_j > 0$ if $j = 1, \ldots, r$, and $\sigma_j = 0$ if $j = r+1, \ldots, m$. We will prove in Problem 2.5 that

$$A\psi = \sum_{j=1}^{r} \sigma_j |\varphi_j\rangle\langle\psi_j| \qquad \text{for all} \quad \psi \in X \tag{2.85}$$

where $\varphi_j := \sigma_j^{-1} A\psi_j$. In particular, we will show that $\psi_1, \ldots, \psi_r$ (resp. $\varphi_1, \ldots, \varphi_r$) is an orthonormal basis of $\ker(A)^\perp$ (resp. $\mathrm{im}(A)$). The operator

[48] We refer to the textbook by G. Ramm, Inverse Problems: Mathematical and Analytical Techniques in Engineering, Springer, New York, 2005.

[49] This is equivalent to the continuity of the matrix elements of $A(\alpha)$ with respect to a fixed basis.

[50] If the operator A is self-adjoint with the eigenvalues $\lambda_1, \ldots, \lambda_m$, then $\sigma_k = |\lambda_k|$ for all k.

$$\boxed{A_{\text{inv}} := \sum_{j=1}^{r} \sigma_j^{-1} |\psi_j\rangle\langle\varphi_j|} \tag{2.86}$$

is called the generalized inverse (or the Moore–Penrose inverse) of the operator A. This generalized inverse plays a crucial role in the theory of ill-posed problems. Consider again the minimum problem (2.84) for the fixed parameter $\alpha \in]0,1[$. To simplify notation, set $A := A(\alpha)$.

Proposition 2.15 *The general solution of the minimum problem* (2.84) *is given by the sum* $\psi = A_{\text{inv}} f + \varphi$ *where* φ *is an arbitrary solution of the equation* $A\varphi = 0$.

The proof will be given in Problem 2.7. This proof shows that $A_{\text{inv}} f$ is the unique solution of the modified minimum problem

$$||A\psi - f||^2 = \min!, \qquad \psi \in \ker(A)^{\perp}$$

where the symbol $\ker(A)^{\perp}$ denotes the orthogonal complement to the null space $\ker(A) := \{\psi \in X : A\psi = 0\}$ of the operator A. Since we have the orthogonality relation $||A_{\text{inv}} f + \varphi||^2 = ||A_{\text{inv}} f||^2 + ||\varphi||^2$ for all $\varphi \in \ker(A)$, the element $A_{\text{inv}} f$ is a solution of the problem

$$||A\psi - f||^2 = \min!, \qquad \psi \in X$$

which has the smallest norm among all the solutions. This solution is uniquely determined.

Note the following. Proposition 2.15 shows that the original minimum problem (2.84) has a unique solution iff $\ker(A) = 0$, that is, the operator is invertible. In this case, we have $A^{-1} = A_{\text{inv}}$; this means that the inverse operator coincides with the generalized inverse operator.

The Tikhonov regularization. Let us replace the original problem (2.84) by the regularized problem

$$||A\psi - f||^2 + \varepsilon||\psi||^2 = \min!, \qquad \psi \in X. \tag{2.87}$$

Let us introduce the so-called filter function $F_\varepsilon(\sigma) := \frac{\sigma^2}{\sigma^2+\varepsilon}$.

Proposition 2.16 *For each parameter* $\varepsilon > 0$, *the regularized problem* (2.87) *has the unique solution*

$$\psi(\varepsilon) := \sum_{j=1}^{r} F_\varepsilon(\sigma_j) \cdot \sigma_j^{-1} \langle\varphi_j|f\rangle \psi_j.$$

Since $F_\varepsilon(\sigma_j) \to 1$ *as* $\varepsilon \to +0$, *the solution* $\psi(\varepsilon)$ *goes to* $A_{\text{inv}} f$ *as* $\varepsilon \to +0$.

Recall that $A_{\text{inv}} f$ is a distinguished solution of the minimum problem (2.87) with $\varepsilon = 0$. The proof of Prop. 2.16 can be found in Problem 2.8. Fix $\varepsilon > 0$. For the filter function, we get

$$\lim_{\sigma \to +0} F_\varepsilon(\sigma) = 0, \qquad \lim_{\sigma \to \infty} F_\varepsilon(\sigma) = 1.$$

Therefore, the approximation of $A_{\text{inv}} f$ by $\psi(\varepsilon)$ critically depends on the behavior of f and on the magnitude of the singular values $\sigma_1, \ldots, \sigma_r$. For example, let $f = \varphi_1$. Noting that $\varphi_1, \ldots, \varphi_n$ forms an orthonormal system, we get

$$\psi(\varepsilon) = F_\varepsilon(\sigma_1) \cdot \sigma_1^{-1}\psi_1 = F_\varepsilon(\sigma_1) A_{\text{inv}} f.$$

If $\sigma_1 \gg \varepsilon$, then $F_\varepsilon(\sigma_1) \sim 1$, and hence $\psi(\varepsilon) \sim A_{\text{inv}} f$. In contrast to this, if $\sigma \ll \varepsilon$, then $F_\varepsilon(\sigma_1) \sim 0$, and $\psi(\varepsilon)$ strongly differs from $A_{\text{inv}} f$. The method of regularization plays a fundamental role in quantum field theory.

Special ill-posed problem. We want to study the problem (2.84) in the special case where

$$A(\alpha) := \begin{pmatrix} b(\alpha) & 0 \\ 0 & 1 \end{pmatrix}, \qquad b(\alpha) := \eta(1 + \alpha + \alpha^2 + \ldots + \alpha^{100}) - \alpha^{101}.$$

Let $0 < \alpha, \eta < 1$. Explicitly, the original problem (2.84) reads as

$$\boxed{(b(\alpha)\psi_1 - f_1)^2 + (\psi_2 - f_2)^2 = \text{min!}.} \tag{2.88}$$

If $b(\alpha) \neq 0$, then this problem has the unique solution

$$\psi_1 = \frac{f_1}{b(\alpha)}, \quad \psi_2 = f_2.$$

However, if $b(\alpha) = 0$, then (2.84) has the general solution

$$\psi_1 = \text{arbitrary real number}, \quad \psi_2 = f_2.$$

In this case, the problem is ill-posed.

The pitfalls of formal perturbation theory. Let $n = 1, 2, \ldots, 100$. Use the nth order approximation $b_n(\alpha) := \eta(1 + \alpha + \ldots + \alpha^n)$ of $b(\alpha)$, and replace the original problem (2.88) by the approximate problem

$$(b_n(\alpha)\psi_1 - f_1)^2 + (\psi_2 - f_2)^2 = \text{min!}.$$

This problem has the unique solution

$$\psi_1 = \frac{f_1}{b_n(\alpha)} = \frac{(1-\alpha)f_1}{\eta(1-\alpha^{n+1})} = \frac{f_1}{\eta}(1-\alpha)(1+\alpha^{n+1}+\ldots),$$

and $\psi_2 = f_2$. If we only consider terms up to order n, then the solution reads as

$$\psi_1 = \frac{f_1}{\eta}(1-\alpha), \quad \psi_2 = f_2. \tag{2.89}$$

Now consider the nth order case where $n = 101$. By Problem 2.9, there exists a number $\eta_0 > 0$ such that the nonlinear equation $b_{101}(\alpha) = 0$, that is,

$$\alpha = \eta^{1/101}(1 + \alpha + \ldots + \alpha^{100})^{1/101}, \qquad \alpha \in [0, 1]$$

has a unique solution $\alpha(\eta)$ for each parameter $\eta \in]0, \eta_0[$. If we choose the critical parameter $\alpha = \alpha(\eta)$, then the original problem (2.88) has the general solution $\psi_1 =$ arbitrary real number, $\psi_2 = f_2$. This solution differs drastically from (2.89).

This means that, in the present example, a catastrophe occurs in very high order of formal perturbation theory.

Therefore, if we use perturbation theory in a formal manner, we can never exclude such a singular behavior. This tells us that

The results obtained by formal perturbation theory have to be handled cautiously.

Formal perturbation theory and quantum field theory. Until now, all of the predictions made by quantum field theory are based on the method of formal perturbation theory. Surprisingly, for small coupling constants in quantum electrodynamics and electroweak interaction, the theoretical predictions coincide with the experimental results with very high accuracy. Physicists belief that this cannot happen by chance. There remains the task to create a mathematically rigorous theory which explains the great success of formal perturbation theory.

2.5 Mathemagics

> Euler truly did not sour his life with limiting value considerations, convergence and continuity criteria and he could not and did not wish to bother about the logical foundation of analysis, but rather he relied – only on occasion unsuccessfully – on his astonishing certitude of instinct and algorithmic power.[51]
>
> Emil Fellmann, 1975

> Seen statistically, Euler must have made a discovery every week... About 1911, Eneström published an almost complete (from today's viewpoint) list of works with 866 titles. Of the 72 volumes of Euler's *Collected Works* all but three have appeared as of today.[52] Euler's correspondence with nearly 300 colleagues is estimated to constitute 4500 to 5000 letters, of which perhaps a third appear to have been lost. These letters are to appear in 13 Volumes.
> Euler was not only one of the greatest mathematicians, but also in general one of the most creative human beings.
>
> Rüdiger Thiele, 1982

> In the entire history of mathematics, aside from the golden age of Greek mathematics, there has never been a better time than that of Leonhard Euler. It was his privilege to leave mathematics with a completely changed face, making it into a powerful machine that it is today.[53]
>
> Andreas Speiser, 1934

Pierre Cartier writes the following in his beautiful article *Mathemagics, A tribute to L. Euler and R. Feynman,* Séminaire Lotharingien **44**, 1–71 from the year 2000:[54]

[51] E. Fellmann, Leonhard Euler. In: Kindler Enzyklopädie. Die Großen der Weltgeschichte (The great people in the history of mankind), Vol. IV, pp. 495–531. Zürich, 1975. We also refer to E. Fellmann, Leonhard Euler, Birkhäuser, Basel 2007 (translated from German into English).

[52] L. Euler, Opera omnia, Vols. 1–72, Leipzig–Berlin, later Basel–Zürich, 1911ff. Edited by E. Fellmann.
Rüdiger Thiele, Leonhard Euler, Teubner, Leipzig, 1982 (in German) (reprinted with permission).

[53] A. Speiser, Leonhard Euler und die deutsche Philosophie (Euler and the German philosophy), Zürich, 1934 (in German).

[54] We also refer to V. Varadarajan, Euler through Time: A New Look at Old Themes, Amer. Math. Soc., Providence, Rhode Island, 2006.

> The implicit philosophical belief of the working mathematician is today the Hilbert–Bourbaki formalism. Ideally, one works within a *closed system*: the basic principles are clearly enunciated once for all, including (that is an addition of twentieth century science) the formal rules of logical reasoning clothed in mathematical form. The basic principles include precise definitions of all mathematical objects, and the coherence between the various branches of mathematical sciences is achieved through reduction to basic models in the universe of sets. A very important feature of the system is its *non-contradiction*; after Gödel (1906–1978), we have lost the initial hopes to establish the non-contradiction by a formal reasoning, but one can live with a corresponding belief in non-contradiction. The whole structure is certainly very appealing, but the illusion is that it is eternal, that it will function for ever according to the same principles. What history of mathematics teaches us is that the principles of mathematical deduction, and not simply the mathematical theories, have evolved over the centuries. In modern times, theories like General Topology or Lebesgue's Integration Theory represent an almost perfect model of precision, flexibility, and harmony, and their applications, for instance to probability theory, have been very successful. My thesis is:
>
> *There is another way of doing mathematics, equally successful, and the two methods should supplement each other and not fight.*
>
> This other way bears various names: symbolic method, operational calculus, operator theory... Euler was the first to use such methods in his extensive study of infinite series, convergent as well as divergent. The calculus of differences was developed by Boole (1815–1864) around 1860 in a symbolic way, then Heaviside (1850–1925) created his own symbolic calculus to deal with systems of differential equations in electric circuits. But the modern master was Feynman (1918–1988) who used his diagrams, his disentangling of operators, his path integrals...
>
> > *The method consists in stretching the formulas to their extreme consequences, resorting to some internal feeling of coherence and harmony.*
>
> There are obviously pitfalls in such methods, and only experience can tell you that for the Dirac delta function an expression like $x\delta(x)$ or $\delta'(x)$ is lawful, but not $\delta(x)/x$ or $\delta(x)^2$. Very often, these so-called symbolic methods have been substantiated by later rigorous developments, for instance, the Schwartz distribution theory gives a rigorous meaning to $\delta(x)$, but physicists used sophisticated formulas in "momentum space" long before Laurent Schwartz codified the Fourier transformation for distributions. The Feynman "sums over histories" have been immensely successful in many problems, coming from physics as well from mathematics, despite the lack of a comprehensive rigorous theory.

Newton (1643–1727), Leibniz (1646–1716), Euler (1707–1783) and their successors very successfully used infinitesimals, that is, quantities with the strange property

$$dx \neq 0 \qquad \text{and} \qquad (dx)^2 = 0. \tag{2.90}$$

Such quantities are still frequently used in the physics literature. Obviously, classical numbers dx do not have the property (2.90). Based on the notion of ultra-filters, we will show in Sect. 4.6 how to introduce rigorously such infinitesimals as *equivalence classes* of real numbers. We will embed this into the discussion of a general mathematical strategy called the strategy of equivalence classes.

Problems

2.1 *The Mittag-Leffler theorem.* Use the sketch of the proof given on page 58 in order to give a full proof. Hint: See Remmert (1998), p. 128.

2.2 *The partial fraction series of the cotangent function.* Use the sketch of the proof given on page 58 in order to give a full proof of the relation (2.16).
Hint: See Smirnov (1964), Vol. 3, Sect. 65.

2.3 *The surface measure of the* $(N-1)$*-dimensional unit sphere.* Prove formula (2.32) on page 69. Solution: We start with the Gaussian integral

$$J := (\sqrt{\pi})^N = \left(\int_{-\infty}^{\infty} \mathrm{e}^{-x^2} dx \right)^N = \int_{\mathbb{R}^N} \mathrm{e}^{-\sum_{j=1}^N x_j^2} d^N x.$$

Using spherical coordinates, we get

$$J = \int_0^\infty \left(\int_{\mathbb{S}^{N-1}} d\mu \right) \mathrm{e}^{-r^2} r^{N-1} dr.$$

Finally, the substitution $t = r^2$ yields

$$J = \frac{1}{2} \operatorname{meas}(\mathbb{S}^{N-1}) \int_0^\infty t^{\frac{N}{2}-1} \mathrm{e}^{-t} dt = \frac{1}{2} \operatorname{meas}(\mathbb{S}^{N-1}) \Gamma\left(\frac{N}{2}\right).$$

2.4 *Computation of Liouville integrals.* Prove the integral formula (2.45) on page 73. Solution: For example, consider the integral

$$J := 2 \int_0^\infty \frac{dr}{1+r^2}.$$

Using the substitution $x := 1/(1+r^2)$, we get

$$J = \int_0^1 x^{-1/2} (1-x)^{-1/2} dx = B(\tfrac{1}{2}, \tfrac{1}{2}) = \frac{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2})}{\Gamma(1)}.$$

In the general case, the proof proceeds analogously.

2.5 *Singular values of a linear operator.* Prove (2.85) on page 106.
Solution: Using the completeness relation $I = \sum_{k=1}^m |\psi_k\rangle\langle\psi_k|$, we get

$$A = \sum_{k=1}^m A|\psi_k\rangle\langle\psi_k|. \tag{2.91}$$

If $k = r+1, \ldots, m$, then $\sigma_k = 0$. Hence

$$\langle A\varphi | A\psi_k \rangle = \langle \varphi | A^\dagger A \psi_k \rangle = \sigma_k^2 \langle \varphi | \psi_k \rangle = 0 \qquad \text{for all} \quad \varphi \in X.$$

Therefore, $A\psi_k$ is perpendicular to the image $\operatorname{im}(A)$. Hence $A\psi_k = 0$. By (2.91), $A = \sum_{j=1}^r A|\psi_j\rangle\langle\psi_j|$. Finally, $A\psi_j = \sigma_j \varphi_j$. In addition, as preparation for an argument below, note that

$$\langle \varphi_j | \varphi_k \rangle = \sigma_j^{-2} \langle A\psi_j | A\psi_k \rangle = \sigma_j^{-2} \langle \psi_j | A^\dagger A \psi_k \rangle = \langle \psi_j | \psi_k \rangle = \delta_{jk}, \quad j,k = 1, \ldots, r.$$

Therefore, $\varphi_1, \ldots, \varphi_r$ is an orthonormal basis of the linear space $\operatorname{im}(A)$, and $\psi_1, \ldots, \psi_r$ is an orthonormal basis of the linear space $\ker(A)^\perp$.

2.6 *Quadratic variational problem.* Let $B : X \to X$ be a linear self-adjoint operator on the real finite-dimensional Hilbert space X. Suppose that there is a number $c > 0$ such that $\langle\psi|B\psi\rangle \geq c||\psi||^2$ for all $\psi \in X$. Show that, for given $f \in X$, the minimum problem

$$\frac{1}{2}\langle\psi|B\psi\rangle - \langle f|\psi\rangle = \min!, \qquad \psi \in X$$

has the unique solution $\psi = B^{-1}f$. (In particular, the inverse operator B^{-1} exists on X.) Hint: See Zeidler (1995), Vol. 2, Sect. 2.4.

2.7 *Proof of Proposition 2.15 on page 107.* Solution: The minimum problem

$$||f - f_0||^2 = \min!, \qquad f_0 \in \mathrm{im}(A)$$

has a unique solution, by Problem 2.6. Geometrically, f_0 is the orthogonal projection of f onto the linear space $\mathrm{im}(A)$. The equation

$$A\psi = f_0, \qquad \psi \in \ker(A)^\perp$$

has a unique solution ψ_0. Since $A\psi - f \in \mathrm{im}(A)$ and $f - f_0 \in \mathrm{im}(A)^\perp$, the orthogonal decomposition of $A\psi - f$ into the sum $(A\psi - f_0) + (f_0 - f)$ yields

$$||A\psi - f||^2 = ||A\psi - f_0||^2 + ||f_0 - f||^2.$$

Therefore, the original problem $||A\psi - f||^2 = \min!, \psi \in X$ is equivalent to the minimum problem

$$||A\psi - f_0||^2 = \min!, \qquad \psi \in X$$

which has the general solution $\psi = \psi_0 + \varphi$ with an arbitrary element φ in $\ker(A)$. By Problem 2.5, $\varphi_1, \ldots, \varphi_r$ is an orthonormal basis of $\mathrm{im}(A)$, and $\psi_1, \ldots, \psi_r$ is an orthonormal basis of $\ker(A)^\perp$. By Fourier expansion,

$$f_0 = \sum_{j=1}^{r} \langle\varphi_j|f_0\rangle\varphi_j = \sum_{j=1}^{r} \langle\varphi_j|f\rangle\varphi_j.$$

Similarly, $\psi_0 = \sum_{j=1}^{r} \langle\psi_j|\psi_0\rangle\psi_j$. From $A\psi_0 = f_0$ and $A\psi_j = \sigma_j\varphi_j$ we get

$$A\psi_0 = \sum_{j=1}^{r} \langle\psi_j|\psi_0\rangle\sigma_j\varphi_j = \sum_{j=1}^{r} \langle\varphi_j|f\rangle\varphi_j.$$

Hence $\langle\psi_j|\psi_0\rangle = \sigma_j^{-1}\langle\varphi_j|f\rangle, j = 1, \ldots, r$. By (2.86), this implies

$$\psi_0 = \sum_{j=1}^{r} \sigma_j^{-1}|\psi_j\rangle\langle\varphi_j|f\rangle = A_{\mathrm{inv}}f.$$

2.8 *Proof of Proposition 2.16 on page 107.* Solution: Note that $||A\psi - f||^2 + \varepsilon||\psi||^2$ is equal to

$$\langle A\psi - f|A\psi - f\rangle + \varepsilon\langle\psi|\psi\rangle = \langle\psi|(\varepsilon I + A^\dagger A)\psi\rangle - 2\langle A^\dagger f|\psi\rangle.$$

By Problem 2.6, the unique solution of the minimum problem (2.87) reads as

$$\psi(\varepsilon) = (\varepsilon I + A^\dagger A)^{-1}A^\dagger f.$$

Since $(\varepsilon I + A^\dagger A)\psi_j = (\varepsilon + \mu_j)\psi_j$, Fourier expansion yields

$$\psi(\varepsilon) = \sum_{k=1}^{m} (\varepsilon + \mu_k)^{-1} \langle \psi_k | A^\dagger f \rangle \psi_k.$$

Noting that $\langle \psi_k | A^\dagger f \rangle = \langle A\psi_k | f \rangle$ and that $A\psi_k = 0$ if $k \geq r+1$, we get

$$\psi(\varepsilon) = \sum_{j=1}^{r} (\varepsilon + \mu_j)^{-1} \sigma_j \langle \varphi_j | f \rangle \psi_j = \sum_{j=1}^{r} \frac{\sigma_j^2}{\sigma_j^2 + \varepsilon} \cdot \sigma_j^{-1} \langle \varphi_j | f \rangle \psi_j.$$

2.9 *A special nonlinear equation.* Suppose that the function $g : [0,1] \to \mathbb{R}$ is smooth. Then there exists a number $\eta_0 > 0$ such that, for each $\eta \in [0, \eta_0]$, the equation

$$\alpha = \eta g(\alpha), \qquad 0 \leq \alpha \leq 1$$

has a unique solution. This solution can be computed by the convergent iterative method $\alpha_{n+1} = \eta g(\alpha_n), n = 0, 1, \ldots$, with $\alpha_0 := 0$.
Solution: If $\alpha, \beta \in \mathbb{R}$, then $|f(\alpha) - f(\beta)| = |f'(\xi)(\alpha - \beta)| \leq \text{const } |\alpha - \beta|$. Now apply the Banach fixed point theorem to the interval $[0,1]$ (see Sect. 7.13 of Vol. I).

2.10 *Integration.* Let $a > 0$. Show that $\int_{\mathbb{R}^4} \frac{d^4p}{(|p|^2+a^2)^3} = \frac{\pi^2}{2a^2}$.
Solution: By the sphere trick (2.34), the integral is equal to

$$2\pi^2 \int_0^\infty \frac{r^3 dr}{(r^2 + a^2)^3}.$$

Setting $x = r^2$, we get

$$\pi^2 \int_0^\infty \frac{x dx}{(x + a^2)^3} = \pi^2 \int_0^\infty \left(\frac{1}{(x + a^2)^2} - \frac{a^2}{(x + a^2)^3} \right) dx = \frac{\pi^2}{2a^2}.$$

2.11 *Gamma function.* Prove the Laurent expansion (2.29) of the gamma function near the pole $z = -1$.
Solution: By the functional equation, $(z-1)\Gamma(z-1) = \Gamma(z)$. Hence

$$\Gamma(z-1) = -(1 + z + z^2 + O(z^3))\Gamma(z), \qquad z \to 0.$$

Finally, use the Laurent expansion (2.28) of $\Gamma(z)$ near the pole $z = 0$.

3. The Power of Combinatorics

> Hopf algebra is invading quantum field theory from both ends, both at the foundational level and the computational level... The approach from quantum theoretical first principle is still in its first infancy.[1]
>
> Héctor Figueroa and José Gracia-Bondia, 2005

In this series of monographs, we will show that:

> *There are highly complex mathematical structures behind the idea of the renormalization of quantum field theories.*

The combinatorial structure of Feynman diagrams lies at the heart of renormalization methods. In the standard Boguliubov–Parasiuk–Hepp–Zimmermann (BPHZ) approach, the regularization of algebraic Feynman integrals is carried out by an iterative method which was invented by Bogoliubov in the 1950s. It was shown by Zimmermann in 1969 that Bogoliubov's iterative method can be solved in a closed form called the Zimmermann forest formula. Finally, it was discovered by Kreimer in 1998 that Zimmermann's forest formula can be formulated by using the coinverse of an appropriate Hopf algebra for Feynman graphs. This will be thoroughly studied later on. In this chapter, we only want to discuss some basic ideas about Hopf algebras and Rota–Baxter algebras.

3.1 Algebras

> Products play a fundamental role in quantum field theory (e.g., normal products, time-ordered products, retarded products). They are used in order to construct correlation functions.
>
> Folklore

Algebras are linear spaces equipped with a distributive multiplication. Let us discuss this.

The algebra of smooth functions as a prototype. Fix $n = 1, 2, \ldots$ Recall that $\mathcal{E}(\mathbb{R}^N)$ denotes the set of all smooth complex-valued functions $f : \mathbb{R}^N \to \mathbb{C}$. We write $\mathcal{A}$ instead of $\mathcal{E}(\mathbb{R}^N)$. For all functions $f, g \in \mathcal{A}$ and all complex numbers α, β, we define the linear combination $\alpha f + \beta g$ and the product fg by setting for all $x \in \mathbb{R}^N$:

- $(\alpha f + \beta g)(x) := \alpha f(x) + \beta g(x)$,
- $(fg)(x) := f(x)g(x)$.

[1] H. Figueroa and J. Gracia-Bondia, Combinatorial Hopf algebras in quantum field theory I, Rev. Math. Phys. **17** (2005), 881–982.
Internet: `http://arXiv:hep-th/0408145`

In addition, we define the so-called unit element **1** by setting

- $\mathbf{1}(x) := 1$ for all $x \in \mathbb{R}^N$.

Then for all $f, g, h \in \mathcal{A}$ and all $\alpha, \beta \in \mathbb{C}$, the following hold:

(A1) Linearity: The set $\mathcal{A}$ is a complex linear space.
(A2) Consistency: $fg \in \mathcal{A}$, and $(\alpha f)g = f(\alpha g) = \alpha(fg)$.
(A3) Distributivity: $(\alpha f + \beta g)h = \alpha fh + \beta fh$, and

$$h(\alpha f + \beta g) = \alpha hf + \beta hg.$$

(A4) Associativity: $(fg)h = f(gh)$.
(A5) Commutativity: $fg = gf$.
(A6) Unitality: There exists precisely one element **1** in $\mathcal{A}$ such that

$$\mathbf{1}f = f\mathbf{1} = f \qquad \text{for all} \quad f \in \mathcal{A}.$$

Here, **1** is called the unit element of $\mathcal{A}$.

The definition of an algebra. As we will show later on, algebras play a fundamental role in the mathematical description of quantum processes. By definition, the set $\mathcal{A}$ is called a complex algebra iff for all $f, g \in \mathcal{A}$ and all complex numbers α and β, the linear combination $\alpha f + \beta g$ and the product fg are defined in such a way that the conditions (A1), (A2), and (A3) are always satisfied.[2] In addition, we use the following terminology.

- The algebra $\mathcal{A}$ is called associative iff condition (A4) is always satisfied.
- The algebra $\mathcal{A}$ is called commutative iff condition (A5) is always satisfied.
- The algebra $\mathcal{A}$ is called unital iff condition (A6) is satisfied.

For example, the space $\mathcal{D}(\mathbb{R}^N)$ of smooth test functions $f : \mathbb{R}^N \to \mathbb{C}$ with compact support is a complex algebra. This algebra is associative and commutative. The same is true for the spaces $\mathcal{E}(\mathbb{R}^N)$ and $\mathcal{S}(\mathbb{R}^N)$ of test functions.[3] In addition, the space $\mathcal{E}(\mathbb{R}^N)$ is unital.

For fixed $n = 2, 3, \ldots$, the set of complex $(n \times n)$-matrices forms a complex algebra which is associative, noncommutative, and unital. Here, the unit element is given by the unit matrix $I := \mathrm{diag}(1, 1, \ldots, 1)$.

A subset $\mathcal{B}$ of a complex algebra $\mathcal{A}$ is called a subalgebra iff it is an algebra with respect to the operations induced by $\mathcal{A}$. Explicitly, this means that if $f, g \in \mathcal{B}$ and $\alpha, \beta \in \mathbb{C}$, then $\alpha f + \beta g \in \mathcal{B}$ and $fg \in \mathcal{B}$.

Algebra morphism. Let $\mathcal{A}$ and $\mathcal{B}$ be algebras over $\mathbb{C}$. The map

$$\chi : \mathcal{A} \to \mathcal{B}$$

is called an algebra morphism iff it respects linear combinations and products, that is, the map χ is linear and we have $\chi(fg) = \chi(f)\chi(g)$ for all $f, g \in \mathcal{A}$. Bijective algebra morphisms are also called algebra isomorphisms.

The map $S : \mathcal{A} \to \mathcal{B}$ is called an algebra anti-morphism iff it is linear and we have $S(fg) = S(g)S(f)$ for all $f, g \in \mathcal{A}$.

Modification. Real algebras (also called algebras over $\mathbb{R}$) are defined analogously. We only replace the field $\mathbb{C}$ of complex numbers by the field $\mathbb{R}$ of real numbers.

Perspectives. In this series of monographs, algebras will be encountered quite often. Let us mention the following examples:

[2] Analogously, the definition of a real algebra is obtained by starting from a real linear space $\mathcal{A}$ and by replacing complex numbers by real numbers.

[3] The precise definition of the space $\mathcal{S}(\mathbb{R}^N)$ of smooth, rapidly decreasing functions $f : \mathbb{R}^N \to \mathbb{C}$ can be found in Sect. 10.3.3 of Vol. I.

- the Hopf algebra of linear differential operators (Sect. 3.3.2);
- Hopf algebras, formal power series expansions, and renormalization (Sect. 3.4);
- symmetries and Lie groups; linearized symmetries and Lie algebras (Vols. I–VI);
- the algebra of multilinear functionals (Sect. 3.2);
- the tensor algebra of a linear space (Vol. III);
- the algebra of symmetric multilinear functionals and the symmetric algebra of a linear space (Vol. III);
- the algebra of antisymmetric multilinear functionals and the Grassmann (or exterior) algebra of a linear space (Vol. III);
- the Clifford (or inner) algebras of a linear space equipped with a bilinear form (Vol. III);
- the enveloping algebra of a Lie algebra (Vol. III).

Concerning applications to physics, we mention the following:

- the convolution algebra and the Heaviside calculus for computing electric circuits in electrical engineering (Sect. 4.2);
- Lie algebras in classical mechanics based on Poisson brackets (Sect. 6.9.2);
- Lie super algebras and the supersymmetry of elementary particles and strings (Sect. 7.21 and Vols. III–VI);
- the $*$-algebra approach to physics (classical mechanics, statistical physics, quantum physics) (Sect. 7.17);
- $*$-algebras, C^*-algebras, and von Neumann algebras (Sect. 7.18 and Vol. IV);
- Clifford algebras and the Dirac equation for fermions (Vol. III);
- C^*-algebras and quantum information (Vol. IV);
- local nets of operator algebras and the algebraic approach to quantum field theory due to Haag and Kastler (Vols. IV–VI);
- operator algebras and spectral theory of observables in quantum physics (the Gelfand theory) (Vol. IV);
- the Gelfand–Naimark theorem and noncommutative geometry (Vol. IV);
- the Connes–Kreimer–Moscovici Hopf algebra and renormalization (Vol. IV);
- operator algebras and quantum gravity (Vol. VI).

Roughly speaking, products and hence algebras are everywhere.

3.2 The Algebra of Multilinear Functionals

Multilinear algebra studies all kinds of products.
Folklore

In what follows we want to study the elements of multilinear algebra, which play a crucial role in modern physics. The main tool are multilinear functionals. The tensor product $\otimes$ and the Grassmann product $\wedge$ correspond to special multilinear functionals which are called decomposable. A special role is played by symmetric and antisymmetric multilinear functionals, which is related to bosons and fermions in elementary particle physics, respectively.

In this section, the symbols X, Y, Z, X_α denote linear spaces over $\mathbb{K}$. Here, we choose $\mathbb{K} = \mathbb{R}$ and $\mathbb{K} = \mathbb{C}$; this corresponds to real and complex linear spaces, respectively. The index α runs in the nonempty index set $\mathcal{A}$. For the basic definitions concerning linear spaces, we refer to Sect. 7.3 of Vol. I. Recall that:

Two finite-dimensional real (resp. complex) linear spaces are linear isomorphic iff they have the same dimension.

This well-known theorem from the basic course in linear algebra essentially simplifies the theory of finite-dimensional linear spaces. It shows that a finite-dimensional linear space can be described by a single invariant, namely, its dimension. Note that an analogous theorem for infinite-dimensional linear spaces is not true. If $b_1, b_2, \ldots, b_n$ (resp. $c_1, \ldots, c_n$) is a basis of the linear space X (resp. Y), then the map $\chi : X \to Y$ given by

$$\chi\left(\sum_{k=1}^{n} \alpha_k b_k\right) := \sum_{k=1}^{n} \alpha_k c_k \qquad \text{for all} \quad \alpha_1, \ldots, \alpha_n \in \mathbb{K}$$

is a linear isomorphism. Furthermore, each isomorphism between n-dimensional linear spaces can be obtained this way.

The Cartesian product. By definition, the product set $X \times Y$ is given by

$$X \times Y := \{(x, y) : x \in X,\ y \in Y\}.$$

If $A : X \to X$ and $B : Y \to Y$ are linear operators, then the linear product operator $A \times B : X \times Y \to X \times Y$ is defined by

$$(A \times B)(x, y) := (Ax, By) \qquad \text{for all} \quad x \in X, y \in Y.$$

If $\mathcal{A}$ is a nonempty set, then the product $\prod_{\alpha\in\mathcal{A}} X_\alpha$ is defined to be the set of all tuples

$$(x_\alpha)_{\alpha\in\mathcal{A}} \qquad \text{where} \ \ x_\alpha \in X_\alpha \ \ \text{for all } \alpha \in \mathcal{A}.$$

Explicitly, the symbol (x_α) stands for a map $\alpha \mapsto x_\alpha$ from $\mathcal{A}$ into the union $\bigcup_{\alpha\in\mathcal{A}} X_\alpha$ with $x_\alpha \in X_\alpha$ for all $\alpha \in X_\alpha$.[4]

The direct sum $X \oplus Y$. The Cartesian product $X \times Y$ becomes a linear space over $\mathbb{K}$ if we introduce the linear combination

$$\lambda(x, y) + \mu(u, v) := (\lambda x + \mu u, \lambda y + \mu v), \quad x, u \in X,\ y, v \in Y,\ \lambda, \mu \in \mathbb{K}.$$

This linear space is denoted by $X \oplus Y$. If $A : X \to X$ and $B : Y \to Y$ are linear operators, then the linear operator $A \oplus B : X \oplus Y \to X \oplus Y$ is defined by

$$(A \oplus B)(x, y) := (Ax, By) \qquad \text{for all} \quad x \in X,\ y \in Y.$$

The product $\prod_{\alpha\in\mathcal{A}} X_\alpha$ becomes a linear space over $\mathbb{K}$ if we introduce the linear combinations

$$\lambda(u_\alpha) + \mu(v_\alpha) := (\lambda u_\alpha + \mu v_\alpha)$$

for all $(u_\alpha), (v_\alpha) \in \prod_{\alpha\in\mathcal{A}} X_\alpha$ and all $\lambda, \mu \in \mathbb{K}$. By definition, the direct sum

$$\bigoplus_{\alpha\in\mathcal{A}} X_\alpha$$

is a linear subspace of the product space $\prod_{\alpha\in\mathcal{A}} X_\alpha$ which consists precisely of all the tuples (x_α) with $x_\alpha \neq 0$ for at most a finite set of indices α.[5]

The coproduct $X \coprod Y$. We define

$$X \coprod Y := \{(1, x) : x \in X\} \cup \{(2, y) : y \in Y\}.$$

[4] If one wants to be sure that the Cartesian product $\Pi_{\alpha\in\mathcal{A}}$ is not empty for infinite index sets $\mathcal{A}$ of arbitrary cardinality, then one needs Zermelo's axiom of choice (see page 246).

[5] If the index set $\mathcal{A}$ is finite, then $\prod_{\alpha\in\mathcal{A}} X_\alpha$ and $\bigoplus_{\alpha\in\mathcal{A}} X_\alpha$ coincide.

If the spaces X and Y are disjoint, then we have $X \coprod Y \simeq X \cup Y$, in the sense of a bijection. Therefore, the coproduct is also called the disjoint union of X and Y. If $\mathcal{A}$ is an index set, then we define

$$\coprod_{\alpha \in \mathcal{A}} X_\alpha := \{(\alpha, x_\alpha) : \alpha \in \mathcal{A},\ x_\alpha \in X_\alpha\}.$$

The space $L(X,Y)$ of linear operators. The set of all linear operators $A : X \to Y$ forms a linear space over $\mathbb{K}$ which is denoted by $L(X,Y)$. In particular, the space $L(X,X)$ forms a complex algebra with respect to the usual linear combinations

$$\alpha A + \beta B$$

of linear operators $A, B : X \to X$ where $\alpha, \beta \in \mathbb{K}$. Furthermore, if $X = Y$, then $L(X,X)$ becomes an algebra over $\mathbb{K}$ with respect to the operator product AB.

The dual space X^d. Duality plays a fundamental role in quantum physics in order to describe quantum fields (e.g., in perturbative quantum field theory). The dual space X^d to the given linear space X is defined by

$$X^d := L(X, \mathbb{K}).$$

This is a linear space over $\mathbb{K}$. Let $A : X \to Y$ be a linear operator. Then the dual operator $A^d : Y^d \to X^d$ is defined by

$$(A^d f)(x) := f(A^d x) \qquad \text{for all} \quad x \in X.$$

This refers to all linear functionals $f \in Y^d$.

Cobasis. Let $b_1, \ldots, b_n$ be linearly independent elements of the linear space X. Define

$$b^k \left(\sum_{s=1}^n \alpha_s b_s \right) := \alpha_k \qquad \text{for all} \quad \alpha_1, \ldots, \alpha_n \in \mathbb{K},\ k = 1, \ldots, n.$$

In particular, $b^k(b_l) = \delta_l^k$ for $k, l = 1, \ldots, n$. We call $b^1, \ldots, b^n$ the dual system to $b_1, \ldots, b_n$.

Proposition 3.1 *The functionals $b^1, \ldots, b^n$ are linearly independent elements of the dual space X^d.*

Proof. Let $\sum_{k=1}^n \beta_k b^k = 0$ Applying this to b_l, we get $\sum_{k=1}^n \beta_k \delta_l^k = 0$. Hence $\beta_l = 0$ for $l = 1, \ldots, n$. □

If $b_1, \ldots, b_n$ is a basis of the linear space X, then $b^1, \ldots, b^n$ is a basis of the dual space X^d called the cobasis of X. Consequently, if X is a finite-dimensional linear space, then we have the linear isomorphism

$$X \simeq X^d.$$

Setting $X^{dd} := (X^d)^d$, we get the linear isomorphism $X^{dd} \simeq X$. In differential geometry, one writes dx^k instead of b^k.

For an arbitrary linear space X, we have

$$X \subseteq X^{dd}$$

where X is a linear subspace of X^{dd}. This is to be understood in the following sense. For fixed $x \in X$, define

$$F_x(f) := f(x) \qquad \text{for all} \quad f \in X^d.$$

The map $F_x : X^d \to \mathbb{K}$ is linear. Hence $F_x \in X^{dd}$.

Proposition 3.2 *The map $x \mapsto F_x$ is a linear injective morphism from X to X^{dd}.*

Therefore, the linear space X can be identified with a linear subspace of X^{dd}. The proof will be given in Problem 4.11 on page 259 based on Zorn's lemma.

The linear space $M_2(X,Y)$ of bilinear functionals. Let the symbol $M_2(X,Y;\mathbb{K})$ denote the space of all bilinear functionals

$$B : X \times Y \to \mathbb{K}.$$

Let $B, C \in M_2(X,Y;\mathbb{K})$ and $\beta, \gamma \in \mathbb{K}$. We define $\beta B + \gamma C$ by setting

$$(\beta B + \gamma C)(x,y) := \beta B(x,y) + \gamma C(x,y) \qquad \text{for all} \quad x \in X, y \in Y.$$

This way, $M_2(X,Y;\mathbb{K})$ becomes a linear space over $\mathbb{K}$. Let $f \in X^d$ and $g \in Y^d$. We define the tensor product $f \otimes g$ of the linear functionals f and g by setting

$$\boxed{(f \otimes g)(x,y) := f(x)g(y) \qquad \text{for all} \quad x \in X,\ y \in Y.}$$

Obviously, we have $(f \otimes g) \in M_2(X,Y;\mathbb{K})$. If X is a finite-dimensional space with the basis $b_1, \dots b_n$, then each $B \in M_2(X,Y;\mathbb{K})$ can be uniquely represented by

$$\boxed{B = \sum_{k,l=1}^{n} B(b_k, b_l) b^k \otimes b^l.} \tag{3.1}$$

This will be proved in Problem 3.1 on page 167. For infinite-dimensional linear spaces X and Y, the bilinear functional $B : X \otimes Y \to \mathbb{K}$ is called decomposable iff it can be represented by a finite sum of the form

$$B = \alpha_1 f_1 \otimes g_1 + \ldots + \alpha_n f_n \otimes g_n$$

where $f_k \in X^d, g_k \in Y^d, \alpha_k \in \mathbb{K}$, and $k = 1, 2, \dots$ The set of decomposable bilinear functionals forms a linear subspace of $M_2(X,Y;\mathbb{K})$, and it coincides with $M_2(X,Y;\mathbb{K})$ if X and Y are finite-dimensional spaces. Let $X = Y$. For $f, g \in X^d$, we define

$$\boxed{f \wedge g := f \otimes g - g \otimes f.}$$

The decomposable bilinear functionals

$$B = \alpha_1 f_1 \wedge g_1 + \ldots + \alpha_n f_n \wedge g_n$$

are antisymmetric, that is, $B(x,y) = -B(y,x)$ for all $x, y \in X$.

The algebra $M(X)$ of multilinear functionals. The symbol $M_k(X)$ denotes the set of all k-linear functionals

$$F : X \times \cdots \times X \to \mathbb{K}.$$

This means that the map $(x_1, \dots, x_k) \mapsto F(x_1, \dots, x_k)$ is linear in each argument. The space $M_k(X)$ becomes a linear space over $\mathbb{K}$ in a natural way by using linear combinations of k-linear functionals. If $B \in M_k(X)$ and $C \in M_l(X)$, then the product $B \otimes C$ is defined by

$$\boxed{(B \otimes C)(x_1, \dots x_k, x_{k+1}, \dots, x_{k+l}) := B(x_1, \dots, x_k) C(x_{k+1}, \dots, x_{k+l})}$$

for all $x_1, \dots, x_{k+l} \in X$. Obviously, $B \otimes C$ is an element of $M_{k+l}(X)$. In addition, we set

$$M_0(X) := \mathbb{K} \qquad \text{and} \qquad \alpha \otimes F = F \otimes \alpha := \alpha F$$

for all $\alpha \in \mathbb{K}$ and $F \in M_k(X)$. With respect to the product $\otimes$, the direct sum

$$M(X) := \bigoplus_{k=0}^{\infty} M_k(X)$$

becomes an associative algebra over $\mathbb{K}$, which is called the algebra of multilinear functionals on the linear space X. Explicitly, the elements of $M(X)$ are finite sums of the form

$$\alpha + F_1 + F_2 + \dots$$

where $\alpha \in \mathbb{K}$, and $F_k \in M_k(X)$ for $k = 1, 2 \dots$ In particular, the space $M_1(X)$ coincides with the dual space X^d. The algebra $M(X)$ over $\mathbb{K}$ possesses an additional structure, namely, it is graded. By definition, the elements of $M_k(X)$ are called homogeneous of degree k. This notion generalizes the degree of a polynomial.

The tensor product $X \otimes Y$ of two linear spaces. Let X and Y be linear spaces over $\mathbb{K}$ where $\mathbb{K} = \mathbb{R}, \mathbb{C}$. We want to construct a linear space $X \otimes Y$ equipped with a product $x \otimes y$ for $x \in X$ and $y \in Y$ such that the following properties are valid:

(P) Product property: For all $u, x \in X$, $v, y \in Y$, and $\alpha, \beta \in \mathbb{K}$, we have

$$\begin{aligned} (\alpha u + \beta x) \otimes y &= \alpha(u \otimes y) + \beta(x \otimes y), \\ x \otimes (\alpha v + \beta y) &= \alpha(x \otimes v) + \beta(x \otimes y). \end{aligned} \tag{3.2}$$

(B) Basis property: If $x_1, \dots, x_n$ and $y_1, \dots, y_m$ are linearly independent elements in the linear spaces X and Y, respectively, then all the elements

$$x_k \otimes y_l, \qquad k = 1, \dots, n, \;\; l = 1, \dots, m$$

are linearly independent in the tensor product $X \otimes Y$.

Property (B) excludes the trivial case where $x \otimes y = 0$ for all $x \in X, y \in Y$. Let us now realize these properties by a specific model based on bilinear functionals. For $x \in X$ and $y \in Y$, we define the tensor product $x \otimes y$ by setting

$$(x \otimes y)(f, g) := f(x)g(y) \qquad \text{for all} \quad f \in X^d, g \in Y^d.$$

This means that $x \otimes y : X^d \times Y^d \to \mathbb{K}$ is a bilinear functional on the product space $X^d \times Y^d$. By definition, the tensor product $X \otimes Y$ is the linear hull of all the products $x \otimes y$ with $x \in X$ and $y \in Y$. Explicitly, the elements of $X \otimes Y$ have the form

$$\alpha_1(x_1 \otimes y_1) + \dots + \alpha_n(x_n \otimes y_n), \qquad n = 1, 2, \dots \tag{3.3}$$

with $x_1, \dots, x_n \in X, y_1, \dots, y_n \in Y$, and $\alpha_1, \dots, \alpha_n \in \mathbb{K}$. The space $X \otimes Y$ is a linear subspace of the linear space of all bilinear functionals $B : X^d \times Y^d \to \mathbb{K}$.

Proposition 3.3 *The basis property* (B) *above is valid for the tensor product $X \otimes Y$.*

The proof will be given in Problem 4.14 on page 260. Observe that different expressions of the form (3.3) may describe the same element of the space $X \otimes Y$. The basis property (B) can be used in order to decide whether two expressions of the form (3.3) represent the same element of $X \otimes Y$ or not. For example, consider the equation

$$x \otimes y = u \otimes v + w \otimes z \tag{3.4}$$

where $x, u, w \in X$ and $y, v, z \in Y$. Set

$$\mathcal{X} := \operatorname{span}\{x, u, w\}, \qquad \mathcal{Y} := \operatorname{span}\{y, v, z\}.$$

Let $b_1, \ldots, b_r$ (resp. $c_1, \ldots, c_s$) be a basis of $\mathcal{X}$ (resp. $\mathcal{Y}$). Then, by the product property (P) above, we get

$$x \otimes y = \left(\sum_{j=1}^{r} \beta_j b_j \right) \otimes \left(\sum_{k=1}^{s} \gamma_k c_k \right) = \sum_{j,k} \beta_j \gamma_k (b_j \otimes c_k).$$

Similarly, we obtain

$$u \otimes v + w \otimes z = \sum_{j,k} \alpha_{jk} (b_j \otimes c_k).$$

By Prop. 3.3, equation (3.4) is valid iff $\beta_j \gamma_k = \alpha_{jk}$ for all indices j, k. The general case proceeds analogously. This argument shows that the following holds: If X and Y are finite-dimensional linear spaces over $\mathbb{K}$, then $X \otimes Y$ is also a finite-dimensional linear space over $\mathbb{K}$ with the dimension

$$\dim(X \otimes Y) = \dim X \cdot \dim Y.$$

On page 260, we will construct the tensor product $X \otimes Y$ in terms of equivalence classes. The two constructions yield isomorphic linear spaces. Important are only the properties (P), (B) above.

The tensor product $\mathcal{A} \otimes \mathcal{B}$ of two algebras. Let $\mathcal{A}$ and $\mathcal{B}$ be algebras over $\mathbb{K}$ with $\mathbb{K} = \mathbb{R}, \mathbb{C}$. Since $\mathcal{A}$ and $\mathcal{B}$ are linear spaces over $\mathbb{K}$, we have the tensor product $\mathcal{A} \otimes \mathcal{B}$ at hand. In addition, we define the product

$$(a \otimes b)(c \otimes d) := ac \otimes bd$$

for all $a, c \in \mathcal{A}$ and all $b, d \in \mathcal{B}$. In a natural way, this definition can be extended to expressions of the form (3.3).

Proposition 3.4 *The tensor product $\mathcal{A} \otimes \mathcal{B}$ is an algebra over $\mathbb{K}$.*

The proof will be given in Problem 4.17 on page 261. We have to show that the product on $\mathcal{A} \times \mathcal{B}$ does not depend on the choice of the representations (3.3). To this end, we will use the language of equivalence classes.

Tensor products will be studied in greater detail in Vol. III. In terms of physics, tensor products are used in order to describe composite particles. For example, the tensor product $\varphi \otimes \psi$ of the two states φ and ψ of single particles is the state of the composite particle. In terms of mathematics, tensor products are used in order to reduce multilinear functionals to linear operators on tensor products.

3.3 Fusion, Splitting, and Hopf Algebras

> In nature, one observes fusion and splitting of physical states. From the mathematical point of view, this corresponds to products and coproducts of Hopf algebras, respectively.
>
> Folklore

In this section, we will use tensor products in order to define Hopf algebras. In particular, the language of tensor products will tell us why coassociativity (CA) and counitality (CU) are dual concepts to associativity (A) and unitality (U) of algebras (see page 128).

3.3.1 The Bialgebra of Linear Differential Operators

For a beginner, the definition of a Hopf algebra seems to be rather involved. To help the reader, we start with the well-known algebra $\mathcal{LE}(\mathbb{R}^N)$ of linear differential operators with respect to N arguments and constant coefficients. This is the prototype of a Hopf algebra.

To begin with, fix the dimension $N = 1, 2, \ldots$ The points of the space $\mathbb{R}^N$ are denoted by $x = (x_1, \ldots, x_N)$. Recall that $\mathcal{E}(\mathbb{R}^N)$ denotes the space of smooth complex-valued functions

$$f : \mathbb{R}^N \to \mathbb{C}.$$

We introduce the partial derivative $\partial_j := \frac{\partial}{\partial x_j}$. For smooth functions f, we have $\partial_j \partial_k f = \partial_k \partial_j f$. This means that we have the commutativity property

$$\partial_j \partial_k = \partial_k \partial_j, \qquad j, k = 1, \ldots, N.$$

Furthermore, we set

$$\partial^\alpha := \partial_1^{\alpha_1} \partial_2^{\alpha_2} \cdots \partial_N^{\alpha_N}$$

where $\alpha_1, \ldots, \alpha_N$ are nonnegative integers. Define $|\alpha| := \alpha_1 + \ldots + \alpha_N$. By a linear differential operator with respect to N arguments and constant coefficients, we understand the symbol

$$\boxed{D := a_0 + \sum_{0<|\alpha|\leq m} a_\alpha \partial^\alpha, \qquad m = 1, 2, \ldots}$$

Here, the coefficients a_α are complex numbers. The positive integer m is called the order of the differential operator. We set $m := 0$ if $D := a_0$. We also introduce the symbol

$$\varepsilon(D) := a_0.$$

For example, if $N = 2$ we may choose

$$D := a_0 + a_1\partial_1 + a_2\partial_2 + a_{11}\partial_1\partial_1 + a_{12}\partial_1\partial_2 + a_{22}\partial_2\partial_2.$$

Obviously, if $f : \mathbb{R}^N \to \mathbb{C}$ is a smooth function, then so is Df. More precisely, the operator

$$D : \mathcal{E}(\mathbb{R}^N) \to \mathcal{E}(\mathbb{R}^N)$$

is linear. The set of all these linear operators is denoted by $\mathcal{LE}(\mathbb{R}^N)$. There exist the following operations:

- Linear combination: $\alpha D + \alpha' D'$.
- Product: DD'.
- Coproduct (Leibniz rule): $\Delta(D)(f, g) := D(fg)$.
- Coinverse (integration by parts): For all functions $f, g \in \mathcal{D}(\mathbb{R}^N)$, we have

$$\int_{\mathbb{R}^N} (Df)g \, d^N x = \int_{\mathbb{R}^N} fS(D)g \, d^N x.$$

This way, the set $\mathcal{LE}(\mathbb{R}^N)$ becomes a complex algebra, which is associative, commutative, and unital. The unit element $\mathbf{1}$ is the constant operator $D := 1$. We have

$$\varepsilon(\mathbf{1}) = 1.$$

Summarizing, we obtain the following:

The algebra $\mathcal{LE}(\mathbb{R}^N)$ of differential operators is a Hopf algebra.

This will be discussed in the sequel.[6] We will use the following terminology.

- The map $D \mapsto \Delta(D)$ is called the coproduct of the Hopf algebra $\mathcal{LE}(\mathbb{R}^N)$, and $\Delta(D)$ is called the coproduct applied to the differential operator D.
- The map $D \mapsto \varepsilon(D)$ is called the counit of the Hopf algebra $\mathcal{LE}(\mathbb{R}^N)$. The complex number $\varepsilon(D) = a_0$ is called the augmentation of the differential operator D.
- The map $D \mapsto S(D)$ is called the coinverse (or the antipode) of the Hopf algebra $\mathcal{LE}(\mathbb{R}^N)$. The differential operator $S(D)$ is called the coinverse of the differential operator D.

The explicit form of Δ and S will be given below, by using the Leibniz rule and integration by parts, respectively. In particular, we will see that the coproduct is a map of the form

$$\Delta : \mathcal{LE}(\mathbb{R}^N) \to \mathcal{LE}(\mathbb{R}^N) \otimes \mathcal{LE}(\mathbb{R}^N),$$

which describes a decomposition (splitting) of differential operators. In contrast to this, the product $(D, D') \mapsto DD'$ describes a fusion of the two differential operators D and D'.

The tensor product of linear differential operators. We are given the operators $D, D' \in \mathcal{LE}(\mathbb{R}^N)$. For all functions $f, g \in \mathcal{E}(\mathbb{R}^N)$, we define the tensor product $D \otimes D'$ by setting

$$(D \otimes D')(f, g) := (Df)(D'g).$$

The Leibniz rule. For smooth functions $f, g \in \mathcal{E}(\mathbb{R}^N)$, we have the product rule

$$\partial_j(fg) = (\partial_j f)g + f\partial_j g, \qquad j = 1, \dots, N, \tag{3.5}$$

which is also called the Leibniz rule. Iterated application yields

$$\partial_k\partial_j(fg) = (\partial_k\partial_j f)g + \partial_j f\partial_k g + \partial_k f\partial_j g + f\partial_k\partial_j g. \tag{3.6}$$

The binomic formula tells us that $(a+b)^n = a^n + \sum_{k=1}^{n-1} \binom{n}{k} a^{n-k}b^k + b^n$. Similarly, for $n = 2, 3, \dots$, we get

$$\partial_j^n(fg) = \partial_j^n f + \sum_{k=1}^{n-1} \binom{n}{k} (\partial_j^{n-k} f)\partial_j^k g + \partial_j^n g, \qquad j = 1, \dots, N.$$

The coproduct. Let $D \in \mathcal{LE}(\mathbb{R}^N)$. We define the coproduct $\Delta(D)$ of the linear differential operator D by setting

$$\boxed{\Delta(D)(f, g) := D(fg), \qquad \text{for all} \quad f, g \in \mathcal{E}(\mathbb{R}^N).} \tag{3.7}$$

Let us first consider some examples. If $D = \mathbf{1}$, then $\Delta(\mathbf{1})(f, g) = fg$. Hence

$$\Delta(\mathbf{1}) = \mathbf{1} \otimes \mathbf{1}.$$

By the Leibniz rule (3.5), $\Delta(\partial_j)(f, g) = (\partial_j \otimes \mathbf{1} + \mathbf{1} \otimes \partial_j)(f, g)$. Hence

[6] In 2004 Christian Brouder (Laboratoire de minéralogie-cristallographie, Paris) gave a fascinating lecture at the Max Planck Institute for Mathematics in the Sciences (Leipzig) entitled *Hopf algebras and quantum field theory*. I will follow this lecture, and I am very grateful to Christian Brouder for sending me the manuscript of his lecture.

$$\Delta(\partial_j) = \partial_j \otimes \mathbf{1} + \mathbf{1} \otimes \partial_j, \qquad j = 1, \ldots, N.$$

By (3.6), we get

$$\begin{aligned}\Delta(\partial_k\partial_j)(f,g) = {} & (\partial_k\partial_j \otimes \mathbf{1})(f,g) + (\partial_j \otimes \partial_k)(f,g) \\ & + (\partial_k \otimes \partial_j)(f,g) + (\mathbf{1} \otimes \partial_k\partial_j)(f,g).\end{aligned}$$

Hence

$$\Delta(\partial_k\partial_j) = \partial_k\partial_j \otimes \mathbf{1} + \partial_j \otimes \partial_k + \partial_k \otimes \partial_j + \mathbf{1} \otimes \partial_k\partial_j. \tag{3.8}$$

This corresponds to the four possible splittings of the product $\partial_k\partial_j$:

$$\partial_k\partial_j \otimes \mathbf{1}, \quad \partial_j \otimes \partial_k, \quad \partial_k \otimes \partial_j, \quad \mathbf{1} \otimes \partial_k\partial_j.$$

To simplify notation, in what follows we will write ΔD instead of $\Delta(D)$.

The Sweedler notation. Let $f, g \in \mathcal{E}(\mathbb{R}^N)$. In the general case, we obtain

$$(\Delta D)(f,g) = \sum_{k=1}^{K} D_{1,k} f D_{2,k} g \qquad \text{for all} \quad f, g \in \mathcal{E}(\mathbb{R}^N).$$

This means that $(\Delta D)(f,g) = \sum_{k=1}^{K} (D_{1,k} \otimes D_{2,k})(f,g)$. Hence

$$\Delta D = \sum_{k=1}^{K} D_{1,k} \otimes D_{2,k}.$$

To simplify notation, we briefly write

$$\boxed{\Delta D = \sum_{D} D_{(1)} \otimes D_{(2)}.} \tag{3.9}$$

This is the so-called Sweedler notation. In what follows, we set $\mathcal{A} := \mathcal{LE}(R^{\mathbb{N}})$. Then the coproduct represents a map of the type

$$\Delta : \mathcal{A} \to \mathcal{A} \otimes \mathcal{A},$$

which is an algebra morphism. In fact, the map Δ is linear, and we have

$$\Delta(D'D) = (\Delta D')(\Delta D) \qquad \text{for all} \quad D', D \in \mathcal{A}. \tag{3.10}$$

The proof will be given in Problem 3.5 on page 169.

Coassociativity. It follows from the associative law $f(gh) = (fg)h$ for functions $f, g, h \in \mathcal{E}(\mathbb{R}^N)$ and from the defining relation (3.7) that

$$(\Delta D)(f, gh) = (\Delta D)(fg, h).$$

This is called the coassociative law for the coproduct. We want to show that this is equivalent to the following relation:

$$\boxed{(\Delta \otimes \mathrm{id})\Delta = (\mathrm{id} \otimes \Delta)\Delta.} \tag{3.11}$$

To prove this, observe first that

$$(\Delta D)(f,gh) = D(f(gh)) = \sum_D D_{(1)} f D_{(2)}(gh)$$

and

$$D_{(2)}(gh) = \sum_{D_{(2)}} \left[D_{(2)}\right]_{(1)} g \left[D_{(2)}\right]_{(2)} h.$$

Similarly, we get

$$(\Delta D)((fg),h) = D((fg)h) = \sum_D D_{(1)}(fg) D_{(2)} h$$

and

$$D_{(1)}(fg) = \sum_{D_1} \left[D_{(1)}\right]_{(1)} f \left[D_{(1)}\right]_{(2)} g.$$

To simplify notation, we write $D_{(1)(2)}$ instead of $[D_{(1)}]_{(2)}$. It follows from

$$D(f(gh)) = D((fg)h)$$

that $\sum_D D_{(1)} f \sum_{D_{(2)}} D_{(2)(1)} g D_{(2)(2)} h$ is equal to

$$\sum_D \sum_{D_{(1)}} D_{(1)(1)} f D_{(1)(2)} g D_{(2)} h.$$

This means that $\left(\sum_D D_{(1)} \otimes \sum_{D_{(2)}} D_{(2)(1)} \otimes D_{(2)(2)}\right)(f,g,h)$ is equal to

$$\left(\sum_D \sum_{D_{(1)}} D_{(1)(1)} \otimes D_{(1)(2)} \otimes D_{(2)}\right)(f,g,h).$$

Hence

$$\sum_D D_{(1)} \otimes \sum_{D_{(2)}} D_{(2)(1)} \otimes D_{(2)(2)} = \sum_D \left(\sum_{D_{(1)}} D_{(1)(1)} \otimes D_{(1)(2)}\right) \otimes D_{(2)}.$$

This tells us that

$$\sum_D D_{(1)} \otimes \Delta D_{(2)} = \sum_D \Delta D_{(1)} \otimes D_{(2)}.$$

Recalling that $\Delta D = \sum_D D_{(1)} \otimes D_{(2)}$ and $(\mathrm{id} \otimes \Delta)(a \otimes b) = a \otimes \Delta b$, we obtain

$$(\mathrm{id} \otimes \Delta)\Delta D = (\Delta \otimes \mathrm{id})\Delta D.$$

This proves the claim (3.11).

The counit. For the differential operator $D = a_0 + \sum_{0<|\alpha|\le m} \partial^\alpha$, recall the definition $\varepsilon(D) := a_0$. The map

$$\varepsilon : \mathcal{A} \to \mathbb{C}$$

is an algebra morphism, that is, the map ε is linear and $\varepsilon(DD') = \varepsilon(D)\varepsilon(D')$ for all $D, D' \in \mathcal{A}$. We want to show that

$$\boxed{(\varepsilon \otimes \mathrm{id})\Delta = (\mathrm{id} \otimes \varepsilon)\Delta = \mathrm{id},} \tag{3.12}$$

where $\mathrm{id} : \mathcal{A} \to \mathcal{A}$ denotes the identical operator on the algebra $\mathcal{A}$. In fact, using the relation

$$\Delta D(1, g) = D(1g) = Dg$$

together with the Sweedler notation $\Delta D = \sum_D D_{(1)} \otimes D_{(2)}$, we get

$$Dg = \sum_D (D_{(1)} \otimes D_{(2)})(1, g) = \sum_D D_{(1)}(1) D_{(2)} g = \sum_D \varepsilon(D_{(1)}) D_{(2)} g.$$

Similarly, $Dg = \sum_D (D_{(1)} \otimes D_{(2)})(g, 1) = \sum_D D_{(1)} \varepsilon(D_{(2)}) g$. Hence

$$D = \sum_D \varepsilon(D_{(1)}) D_{(2)} = \sum_D D_{(1)} \varepsilon(D_{(2)}).$$

Furthermore,

$$(\varepsilon \otimes \mathrm{id}) \Delta D = (\varepsilon \otimes \mathrm{id}) \sum_D D_{(1)} \otimes D_{(2)} = \sum_D \varepsilon(D_{(1)}) D_{(2)} = D.$$

Similarly, $(\mathrm{id} \otimes \varepsilon) \Delta D = \sum_D D_{(1)} \varepsilon(D_{(2)}) = D$. This yields the claim (3.12).

The coinverse. For the differential operator $D = a_0 + \sum_{0<|\alpha|\le m} a_\alpha \partial_\alpha$, we define the coinverse $S(D)$ by setting

$$S(D) := a_0 + \sum_{0<|\alpha|\le m} (-1)^\alpha a_\alpha \partial^\alpha.$$

For example, $S(\mathbf{1}) = \mathbf{1}$, $S(\partial_j) = -\partial_j$, and $S(\partial_j \partial_k) = \partial_j \partial_k$ for $j, k = 1, \ldots, N$. Using the coinverse, the formula of integration by parts reads as

$$\int_{\mathbb{R}^N} (Df) g \, d^N x = \int_{\mathbb{R}^N} f S(D) g \, dx^N \qquad \text{for all} \quad f, g \in \mathcal{D}(\mathbb{R}^N).$$

Using the Sweedler notation $\Delta D = \sum_D D_{(1)} \otimes D_{(2)}$, we obtain

$$\sum_D S(D_{(1)}) D_{(2)} = \sum_D D_{(1)} S(D_{(2)}) = \varepsilon(D) \mathbf{1}. \tag{3.13}$$

The proof will be given in Problem 3.6 on page 170. Define

$$\eta(z) := z\mathbf{1} \qquad \text{for all} \quad z \in \mathbb{C}.$$

Introducing the map $\mu : \mathcal{A} \otimes \mathcal{A} \to \mathcal{A}$ by setting

$$\mu\Big(\sum_{j=1}^r f_j \otimes g_j\Big) := \sum_{j=1}^r f_j g_j,$$

the relation (3.13) can be written as

$$\boxed{\mu(S \otimes \mathrm{id}) \Delta = \mu(\mathrm{id} \otimes S) \Delta = \eta \varepsilon.} \tag{3.14}$$

In the sense of the general definition given next, the operator algebra $\mathcal{LE}(\mathbb{R}^N)$ is a Hopf algebra.

3.3.2 The Definition of Hopf Algebras

The canonical morphisms of an associative unital algebra. Let $\mathcal{A}$ be an associative unital complex algebra $\mathcal{A}$ with the unit element $\mathbf{1}$. For all $a, b \in \mathcal{A}$ and all complex numbers z, we define the following maps:

(i) Multiplication map: $\mu(a \otimes b) := ab$.
(ii) Unitality map: $\eta(z) := z\mathbf{1}$.
(iii) Identical map: $\mathrm{id}(a) := a$.

Using linear extension, we get the following three algebra morphisms:

$$\mu : \mathcal{A} \otimes \mathcal{A} \mapsto \mathcal{A}, \qquad \eta : \mathbb{C} \to \mathcal{A}, \qquad \mathrm{id} : \mathcal{A} \to \mathcal{A}.$$

These so-called three canonical morphisms of the algebra $\mathcal{A}$ have the following properties:

(A) Associativity: $\mu(\mu \otimes \mathrm{id}) = \mu(\mathrm{id} \otimes \mu)$.
(U) Unitality: $\mu(\eta \otimes \mathrm{id}) = \mu(\mathrm{id} \otimes \eta) = \mathrm{id}$.

Let us prove this. Relation (A) follows from the associative law $a(bc) = (ab)c$ for all $a, b, c \in \mathcal{A}$. In fact,

$$\mu(\mathrm{id} \otimes \mu)(a \otimes b \otimes c) = \mu(a \otimes \mu(b \otimes c)) = \mu(a \otimes bc) = a(bc).$$

Similarly, $\mu(\mu \otimes \mathrm{id})(a \otimes b \otimes c) = (ab)c$. This proves (A). Relation (U) follows from

$$\mu(\eta \otimes \mathrm{id})(1 \otimes a) = \mu(\eta(1) \otimes a) = \mu(\mathbf{1} \otimes a) = \mathbf{1}a = a.$$

Similarly, $\mu(\mathrm{id} \otimes \eta)(a \otimes 1) = a\mathbf{1} = a$. This yields (U) if we identify $a \otimes 1$ and $1 \otimes a$ with a. This corresponds to the isomorphisms $\mathcal{A} \otimes \mathbb{C} = \mathcal{A} = \mathbb{C} \otimes \mathcal{A}$.

Dualization and the definition of bialgebras. It is our goal to dualize the relations (A) and (U) above by using the replacement

$$\mu \Rightarrow \Delta, \qquad \eta \Rightarrow \varepsilon$$

and by commuting the factors. This way, we obtain the following two dual relations:

(CA) Coassociativity: $(\mathrm{id} \otimes \Delta)\Delta = (\Delta \otimes \mathrm{id})\Delta$.
(CU) Counitality: $(\mathrm{id} \otimes \varepsilon)\Delta = (\varepsilon \otimes \mathrm{id})\Delta = \mathrm{id}$.

Let $\mathcal{A}$ be an associative unital complex algebra. Such an algebra is called a complex bialgebra iff there exist two algebra morphisms

(i) $\Delta : \mathcal{A} \to \mathcal{A} \otimes \mathcal{A}$ (coproduct) and
(ii) $\varepsilon : \mathcal{A} \to \mathbb{C}$ (counit)

such that the conditions (CA) and (CU) are satisfied. The counitality map ε is also called the augmentation map.

The definition of Hopf algebras. The complex bialgebra $\mathcal{A}$ is called a Hopf algebra iff there exists a linear map $S : \mathcal{A} \to \mathcal{A}$ such that

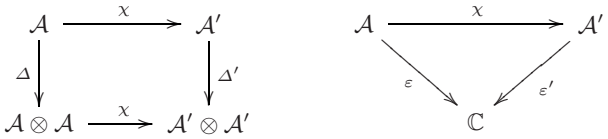
$$\mu(S \otimes \mathrm{id})\Delta = \mu(\mathrm{id} \otimes S)\Delta = \eta\varepsilon. \tag{3.15}$$

This condition looks strange at the first glance. However, we will show below that this is a very natural condition in terms of both Sweedler's notation and a convolution on the space of linear operators $L(\mathcal{A}, \mathcal{A})$ on the algebra $\mathcal{A}$.

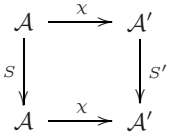
Hopf algebra morphism. Let $\mathcal{A}$ and $\mathcal{A}'$ be Hopf algebras. The map

$$\chi : \mathcal{A} \to \mathcal{A}'$$

is called a Hopf algebra morphism iff it is an algebra morphism and the following three diagrams are commutative:



(3.16)



(3.17)

The map $\chi : \mathcal{A} \to \mathcal{A}'$ is called a Hopf algebra isomorphism iff it is a bijective Hopf algebra morphism and the inverse map $\chi^{-1} : \mathcal{A}' \to \mathcal{A}$ is also a Hopf algebra morphism.[7]

The Sweedler notation. For carrying out concrete computations, one frequently uses the so-called Sweedler notation in bialgebras. Let $a \in \mathcal{A}$. The coproduct Δa is contained in the tensor product $\mathcal{A} \otimes \mathcal{A}$. Thus, there exist elements $a_1, \ldots, a_m, b_1, \ldots, b_m \in \mathcal{A}$ such that

$$\Delta a = \sum_{j=1}^{m} a_j \otimes b_j .$$

According to Sweedler, we write this as

$$\boxed{\Delta a = \sum_{a} a_{(1)} \otimes a_{(2)} .} \tag{3.18}$$

In this language, the basic relations of a Hopf algebra read as follows.

(i) Coassociativity: $\sum_a a_{(1)} \sum_{a_{(2)}} a_{(2)(1)} \otimes a_{(2)(2)}$ is equal to

$$\sum_{a} \left(\sum_{a_{(1)}} a_{(1)(1)} \otimes a_{(1)(2)} \right) a_{(2)} .$$

For this, we simply write $\sum_a a_{(1)} \otimes a_{(2)} \otimes a_{(3)}$. Intuitively, this reflects the regular behavior of splittings.

(ii) Counit: $\sum_a \varepsilon(a_{(1)}) a_{(2)} = \sum_a a_{(1)} \varepsilon(a_{(2)}) = a$.

(iii) Coinverse: $\sum_a a_{(1)} S a_{(2)} = \sum_a (S a_{(1)}) a_{(2)} = \varepsilon(a)\mathbf{1}$.

Condition (iii) refers to the product on the algebra $\mathcal{A}$. The coinverse is also called antipode.

As we will show later on, the coinverse lies at the heart of renormalization in quantum field theory.

[7] Bialgebra morphisms refer to (3.16), that is, the condition (3.17) on the coinverse drops out.

Commutative and cocommutative Hopf algebras. The Hopf algebra $\mathcal{A}$ is called commutative iff $ab = ba$ for all $a, b \in \mathcal{A}$. Moreover, $\mathcal{A}$ is called cocommutative iff the factors of the summands of the coproduct can be interchanged, that is,

$$\Delta a = \sum_{j=1}^{m} a_{1,j} \otimes a_{2,j} = \sum_{j=1}^{m} a_{2,j} \otimes a_{1,j} \qquad \text{for all} \quad a \in \mathcal{A}.$$

This means that $\sum_a a_{(1)} \otimes a_{(2)} = \sum_a a_{(2)} \otimes a_{(1)}$ in Sweedler's notation.

Convolution. Consider again an associative unital complex algebra $\mathcal{A}$ equipped with the additional structure of a bialgebra. We want to introduce a convolution on the space $L(\mathcal{A}, \mathcal{A})$ of linear operators $B : \mathcal{A} \to \mathcal{A}$. Let $B, C \in L(\mathcal{A}, \mathcal{A})$ be given. We define

$$(B * C)a := \sum_a B a_{(1)} C a_{(2)} \qquad \text{for all} \quad a \in \mathcal{A}.$$

Then $B * C \in L(\mathcal{A}, \mathcal{A})$. In other words, $B * C = \mu(B \otimes C)\Delta$. This means that the convolution $B * C$ is obtained by the following quite natural composition of maps:

$$\mathcal{A} \xrightarrow{\Delta} \mathcal{A} \otimes \mathcal{A} \xrightarrow{B \otimes C} \mathcal{A} \otimes \mathcal{A} \xrightarrow{\mu} \mathcal{A}.$$

Proposition 3.5 *The convolution on $L(\mathcal{A}, \mathcal{A})$ is associative and has the unit element $\eta\varepsilon$.*

Explicitly, this means that, for all $B, C, D \in L(\mathcal{A}, \mathcal{A})$, we have

$$(B * C) * D = B * (C * D), \qquad B * \eta\varepsilon = \eta\varepsilon * B = B.$$

The proof can be found in Problem 3.7 on page 170. Using this notion of convolution, the defining relation (3.15) of the coinverse S can be elegantly written as

$$\boxed{S * \mathrm{id} = \mathrm{id} * S = \eta\varepsilon}$$

where id is the identical map on $\mathcal{A}$. This means the following:

The complex bialgebra $\mathcal{A}$ is a Hopf algebra iff there exists a linear map $S : \mathcal{A} \to \mathcal{A}$ which is the two-sided inverse of the identical map on $\mathcal{A}$ for the convolution on $L(\mathcal{A}, \mathcal{A})$.

Historical remarks. Hopf algebras were studied first by Heinz Hopf (1894–1971) in 1941 in order to compute the cohomology of Lie groups and more general topological spaces.

H. Hopf, On the topology of group manifolds and its generalizations, Ann. Math. **42** (1941), 22–52 (in German).

This can be found in

E. Spanier, Algebraic Topology, Springer, New York, 1989.

In what follows, we will study the relation of Hopf algebras to both

- power series expansions and
- symmetry.

Roughly speaking, Hopf algebras are frequently used in order to carry out sophisticated computations for problems where a *nontrivial symmetry* is behind. Using the language of commutative diagrams, it turns out that a bialgebra can be understood best as a mathematical concept which combines the concept of algebra with its dual concept. One only has to reinverse the arrows in the commutative diagrams of an algebra. This will be thoroughly considered in Problem 3.4 on page 168. Hopf algebras are bialgebras which carry the additional structure of a coinverse. Typically, the coinverse comes from dualizing the inverse of a group structure (see Sect. 3.5.2).

3.4 Power Series Expansion and Hopf Algebras

3.4.1 The Importance of Cancellations

> The big surprise in renormalization theory is the appearance of unexpected huge cancellations in the lengthy computations.
>
> Folklore

It happens quite often in mathematics and physics that extremely complicated long expressions dramatically simplify by rearranging them as alternating sums and by cancelling the alternating terms. As the simplest example, let us mention the product

$$(1-x)(1+x+x^2+\ldots+x^{1000}) = (1+x+x^2+\ldots+x^{1000}) - (x+x^2+\ldots+x^{1001}).$$

Rearranging this, we get

$$1+x-x+x^2-x^2+\ldots+x^{1000}-x^{1000}-x^{1001}.$$

Thus, most of the terms cancel, and we finally get the simple expression

$$(1-x)(1+x+x^2+\ldots+x^{1000}) = 1-x^{1001}.$$

As a nontrivial example, let us mention that the elegant heat-kernel approach to the sophisticated Atiyah–Singer index theorem for elliptic differential operators on compact manifolds was discovered by Atiyah, Bott, and Patodi in 1973; they noticed completely unexpected cancellations in long formulas related to the spectral geometry on manifolds.[8] It is typical for topology that topological invariants are related to alternating sums. The prototype is the Euler characteristic (see Sect. 5.6.2 in Vol. I). Unexpected cancellations are also typical for quite lengthy computations in renormalization theory.[9] The experience of mathematicians and physicists shows that symmetries are behind cancellations. It turns out that the cancellations in renormalization theory can be based on Hopf algebras. As a prototype, we want to study the Faà di Bruno Hopf algebra related to the local diffeomorphism group in the complex plane. As a preparation for this, we need two classical formulas for the coefficients of power series expansions, namely,

- the Lagrange inversion formula (3.20) related to the famous Kepler equation (3.19) in celestial mechanics, and
- the Faà di Bruno composition formula (3.30).

Let us study these two formulas first.

[8] M. Atiyah, R. Bott, and V. Patodi, On the heat equation and the index theorem, Inventiones Math. **19** (1973), 279–330. See also P. Gilkey, Invariance Theory, the Heat Equation, and the Atiyah–Singer Index Theorem, CRC Press, Boca Raton, Florida, 1995.

[9] We refer to V. Rivasseau, From Perturbative to Constructive Renormalization, Princeton University Press, 1991. See also V. Rivasseau, An introduction to renormalization. In: B. Duplantier and V. Rivasseau, Poincaré-Seminar 2002: Vacuum Energy – Renormalization, pp. 213–240, Birkhäuser, Basel, 2003.

3.4.2 The Kepler Equation and the Lagrange Inversion Formula

The Kepler equation and the implicit function theorem. Let x, y be Cartesian coordinates in the plane, and denote time by the parameter t. According to Kepler (1571–1630), the motion $x = x(t), y = y(t)$ of a planet of mass m around the sun of mass m_{sun} is given by the equation

$$x = a\cos u - ea, \qquad y = b\sin u, \qquad t = \tau \cdot \sqrt{\frac{ma^3}{Gm_{\text{sun}}}}$$

with the so-called Kepler equation

$$\boxed{u = \tau + e\sin u.} \tag{3.19}$$

This equation relates the parameter τ to the Kepler parameter u. Here, we use the following notation: G gravitational constant, e eccentricity of the ellipse ($0 \le e < 1$), a major semi-axis, and b minor semi-axis. The difference $\tau := u - e\sin u$ was called the mean anomaly by Kepler.[10] The orbit of the planet is given by the following ellipse:

$$\frac{(x+ea)^2}{a^2} + \frac{y^2}{b^2} = 1.$$

The sun is located at the focal point $(x, y) = (-ea, 0)$. For the relation between the major and minor semi-axis, we have $b = a\sqrt{1-e^2}$.

Suppose we ware given the time t and we want to compute the position (x, y) of the planet at the time t. Then we have to solve the Kepler equation in order to get $u = u(\tau)$. Setting $F(u, \tau) := u - e\sin u - \tau$, the Kepler equation (3.19) can be written as

$$F(u, \tau) = 0.$$

Here, we have $F(0, 0) = 0$ and $F_u(0, 0) = 1 - e$. Since $F_u(0, 0) \neq 0$ and the function $(u, \tau) \mapsto F(u, \tau)$ is analytic on the product space $\mathbb{C} \times \mathbb{C}$, the implicit function theorem[11] tells us that there exists an open neighborhood U of the point $(0, 0)$ in the product space $\mathbb{C} \times \mathbb{C}$ such that the Kepler equation (3.19) has a unique solution curve $u = u(\tau)$. This curve is analytic. In other words, there exists a positive number τ_0 such that, for each complex number τ with $|\tau| < \tau_0$, equation (3.19) has the solution

$$u = b_0 + b_1\tau + b_2\tau^2 + \dots,$$

where the power series is convergent. This solution can be computed by the following iterative method

$$u_{n+1} = \tau + e\sin u_n, \qquad n = 0, 1, 2, \dots$$

with $u_0 := 0$. Then $u_1 = \tau$, $u_2 = \tau + e\sin\tau, \dots$ More elegantly, one can use the famous Lagrange inversion formula from 1771:

$$\boxed{u = \tau + e\,\sin\tau + \frac{e^2}{2!}\frac{d}{d\tau}\sin^2\tau + \dots + \frac{e^n}{n!}\frac{d^{n-1}}{d\tau^{n-1}}\sin^n\tau + \dots.} \tag{3.20}$$

Laplace (1749–1827) proved that this power series is convergent if $|\tau| < \tau_0$ with $\tau_0 := 0.6627$.[12] For computing the orbit of a planet, one needs the power series

[10] In the special case where $e = 0$, the orbit is a circle of radius $a = b$ and angular variable u.

[11] We refer to Zeidler (1986), Vol. 1, Sect. 4.7 (see the references on page 1049).

[12] For the proof, see the standard textbook on celestial mechanics by Y. Hagihara, Celestial Mechanics, Vol. I, Sect. 5.14, MIT Press, Cambridge, Massachusetts, 1970.

expansions of $\sin u$ and $\cos u$ with respect to τ. Lagrange showed that the following hold: If the function $u \mapsto f(u)$ is holomorphic in a neighborhood of the point $u = 0$, then

$$f(u(\tau)) = f(\tau) + ef'(\tau)\sin\tau + \frac{e^2}{2!}\frac{d}{d\tau}\left(f'(\tau)\sin^2\tau\right) + \dots$$
$$+\frac{e^n}{n!}\frac{d^{n-1}}{d\tau^{n-1}}\left(f'(\tau)\sin^{n-1}\tau\right) + \dots$$

for all complex numbers τ in a sufficiently small neighborhood of $\tau = 0$.

The Lagrange equation. Starting from the Kepler equation, Lagrange (1736–1813) studied in 1771 the more general equation[13]

$$\boxed{u = \tau + e\varphi(u).} \tag{3.21}$$

Here, we assume that τ and e are complex parameters which vary in a neighborhood of the origin in the Gaussian plane $\mathbb{C}$, and that the function $u \mapsto \varphi(u)$ is holomorphic in a neighborhood of the origin $u = 0$ in $\mathbb{C}$. By the implicit function theorem, equation (3.21) has a solution $u = u(\tau, e)$ which is unique for all points (τ, e) in a sufficiently small neighborhood V of the origin $(0, 0)$ in the product space $\mathbb{C} \times \mathbb{C}$. Moreover, the function $(\tau, e) \mapsto u(\tau, e)$ is holomorphic on V. Lagrange showed that

$$u(\tau, e) = \tau + e\varphi(\tau) + \sum_{n=2}^{\infty} \frac{e^n}{n!}\frac{d^{n-1}}{d\tau^{n-1}}\,\varphi(\tau)^n. \tag{3.22}$$

If the function $u \mapsto f(u)$ is holomorphic in an open neighborhood of the origin in $\mathbb{C}$, then we have

$$f(u(\tau, e)) = \tau + ef'(\tau)\varphi(\tau) + \sum_{n=2}^{\infty} \frac{e^n}{n!}\frac{d^{n-1}}{d\tau^{n-1}} f'(\tau)\varphi(\tau)^n. \tag{3.23}$$

The formulas (3.22) and (3.23) hold for all points (τ, e) in a sufficiently small open neighborhood of the origin $(0, 0)$ in $\mathbb{C} \times \mathbb{C}$. The proof will be given in Problem 3.12.

The inversion of a convergent power series. Suppose that the power series

$$f(x) = f_1 x + \frac{f_2}{2!}\,x^2 + \frac{f_3}{3!}\,x^3 + \dots \tag{3.24}$$

with complex coefficients $f_1, f_2, f_3, \dots$ is convergent in an open neighborhood of the origin $x = 0$ in the Gaussian plane $\mathbb{C}$. Assume that $f_1 \neq 0$. Then the map f is a local diffeomorphism at the origin. This means that the equation

$$y = f(x), \qquad x \in U(0)$$

can be uniquely solved in an open neighborhood $U(0)$ of the point $y = 0$ in $\mathbb{C}$, and the unique solution $x = f^{-1}(y)$ is holomorphic in an open neighborhood of $y = 0$ in $\mathbb{C}$. We write

$$f^{-1}(y) = g_1 + \frac{g_2}{2!}\,y^2 + +\frac{g_3}{3!}\,y^3 + \dots \tag{3.25}$$

[13] L. de Lagrange, Sur le problème de Kepler, Mémoires de l'Académie royale des Sciences et Belles-Lettres de Berlin **24** (1771). In: L. de Lagrange, Oeuvres (Collected Works), Vol. 3, pp. 113–138, Georg Olms, Hildesheim (Germany)/New York.

To simplify the following formulas, let us assume that $f_1 := 1$. Furthermore, we set

$$f(x) = x + bx^2 + cx^3 + dx^4 + \ldots, \quad f^{-1}(y) = Ay + By^2 + Cy^3 + Dy^4 + \ldots$$

The unknown coefficients $A, B, C, D, \ldots$ can be determined by using the substitution

$$y = f(x(y)) = (Ay + By^2 + Cy^3 + \ldots) + b(Ay + By^2 + Cy^3 + \ldots)^2 + \ldots$$

Then comparison of coefficients yields $A = 1$ and

$$B = -b, \qquad C = -c + 2b^2, \qquad D = -d + 5bc - 15b^3. \tag{3.26}$$

For higher-order terms, the formulas become more and more complex.

The challenge is to understand this complicated combinatorics.

For getting the classical solution of this problem, we will follow Lagrange. To this end, we set

$$\psi(x) := f_1 + \frac{f_2}{2!}x + \frac{f_3}{3!}x^2 + \ldots,$$

and we assume that $f_1 \neq 0$. Then we have to solve the equation $y = f(x)$ with $f(x) = x\psi(x)$, that is,

$$x = y\varphi(x), \qquad x \in U(0)$$

with $\varphi(x) := \frac{1}{\psi(x)}$. By (3.22), this yields the famous Lagrange inversion formula

$$\boxed{f^{-1}(y) = \frac{y}{f_1} + \sum_{n=2}^{\infty} \frac{y^n}{n!} \frac{d^{n-1}}{dx^{n-1}} \left(\frac{1}{\psi(x)^n} \right)_{|x=0}.} \tag{3.27}$$

This formula is valid for all complex numbers y in a sufficiently small open neighborhood of the origin $y = 0$ in the complex plane.

3.4.3 The Composition Formula for Power Series

> Francesco Faà di Bruno (1825–1888) (beatified in 1988) gave a formula equivalent to (3.30) below about a hundred and fifty years ago.[14]
>
> Héctor Figueroa and José Gracia–Bondia, 2005

The inversion of a formal power series. If the expression $f(x)$ from (3.24) only represents a formal power series, that it, the series is divergent for all $x \neq 0$, then we define the formal power series $f^{-1}(y)$ by the Lagrange formula (3.27), by using formal differentiation. Then we have the key relations

$$f^{-1}(f(x)) = x, \qquad f(f^{-1}(y)) = y$$

for the symbols x and y. This is to be understood in the sense of formal substitution. Our next goal is to understand the relation to combinatorics and Hopf algebras. We start with the two formal power series expansions

$$f(x) = f_1 x + \frac{f_2}{2!}\, x^2 + \frac{f_3}{3!}\, x^3 + \ldots = \sum_{n=1}^{\infty} \frac{f_n}{n!}\, x^n$$

[14] Faà di Bruno, Annali di Scienze Matematiche e Fisiche di Tortolini **6** (1855), 479; Quart. J. Pure Appl. Math. **1** (1857), 359. See also the footnote on page 115.

and

$$g(x) = g_1 x + \frac{g_2}{2!}\, x^2 + \frac{g_3}{3!}\, x^3 + \ldots = \sum_{n=1}^{\infty} \frac{g_n}{n!}\, x^n$$

with complex coefficients $f_1, f_2, f_3 \ldots$ and $g_1, g_2, g_3 \ldots$ Consider the composed map $h := f \circ g$, that is, $h(x) := f(g(x))$. We want to compute the coefficients $h_1, h_2, h_3, \ldots$ of the formal power series

$$h(x) = \sum_{n=1}^{\infty} \frac{h_n}{n!} x^n.$$

Comparison of coefficients. By substitution,

$$h(x) = \sum_{j=1}^{\infty} \frac{f_j}{j!} \left(\sum_{k=1}^{\infty} \frac{g_k}{k!}\, x^k \right)^j = \sum_{m=1}^{\infty} \frac{h_m}{m!}\, x^m.$$

Comparison of coefficients yields

$$\begin{aligned} h_1 &= g_1 f_1, \quad h_2 = g_2 f_1 + g_1^2 f_2, \\ h_3 &= g_3 f_1 + g_1^3 f_3 + 3 g_1 g_2 f_2, \\ h_4 &= g_4 f_1 + g_1^4 f_4 + 6 g_2 g_1^2 f_3 + (3 g_2^2 + 4 g_3 g_1) f_2, \end{aligned} \tag{3.28}$$

and so on. For computing an arbitrary coefficient h_k, one only needs to carry out a finite number of algebraic operations. However, the formulas become more and more complex. It is our goal to obtain a formula which is valid for all $h_1, h_2, h_3, \ldots$ This is the Faà di Bruno formula (3.30) below.

Generalized binomial coefficients, partitions, and the Bell polynomials. Let $n = 1, 2, \ldots$ and $k = 1, \ldots, n$. Moreover, let $\lambda := (\lambda_1, \ldots, \lambda_n)$ denote a tuple of nonnegative integers with

$$\lambda_1 + \lambda_2 + \ldots + \lambda_n = k, \qquad \lambda_1 + 2\lambda_2 + \ldots + n\lambda_n = n. \tag{3.29}$$

Define the generalized binomial coefficients

$$\binom{n}{\lambda;\ k} := \frac{n!}{\lambda_1! \lambda_2! \cdots \lambda_n! \cdot (1!)^{\lambda_1} (2!)^{\lambda_2} \cdots (n!)^{\lambda_n}}.$$

This coefficient allows a simple interpretation in terms of partitions. Suppose that in a partition of the set $\{1, 2, \ldots, n\}$ into k blocks there are λ_1 singletons, λ_2 two elements subsets, and so on. This means that (3.29) holds. Then the number of all the partitions of this type is equal to $\binom{n}{\lambda;\ k}$. The polynomial

$$B_{n,k}(x_1, x_2, \ldots, x_{n-k+1}) := \sum_{\lambda} \binom{n}{\lambda;\ k} x_1^{\lambda_1} x_2^{\lambda_2} \cdots x_n^{\lambda_n}$$

is called the Bell polynomial of type n, k. Here, the sum runs over all the tuples $\lambda = (\lambda_1, \ldots, \lambda_n)$ of nonnegative integers such that condition (3.29) is satisfied.[15] We are now ready to formulate the Faà di Bruno formula:

$$\boxed{h_n = \sum_{k=1}^{n} f_k B_{n,k}(g_1, g_2, \ldots, g_{n+1-k}).} \tag{3.30}$$

[15] John Bell (1883–1960).

The generating function for the Bell polynomials. For $n = 1, 2, \ldots$ and $k = 1, \ldots, n$, we obtain that the exponential function

$$\exp\left(z \sum_{m=1}^{\infty} x_m \frac{t^m}{m!}\right) \tag{3.31}$$

is equal to the sum

$$1 + \sum_{n=1}^{\infty} \frac{t^n}{n!} \sum_{k=1}^{n} z^k B_{n,k}(x_1, x_2, \ldots, x_{n+1-k}). \tag{3.32}$$

The function (3.31) is called the Faà di Bruno generating function of the Bell polynomials. Explicitly, $B_{n,1}(x_1, \ldots, x_n) = x_n, B_{n,n}(x_1) = x_1^n$. Furthermore, we have $B_{2,1}(x_1, x_2) = x_2, B_{2,2}(x_1) = x_1^2$, and

$$B_{3,1}(x_1, x_2, x_3) = x_3, \quad B_{3,2}(x_1, x_2) = 3x_1x_2, \quad B_{3,3}(x_1) = x_1^3,$$

as well as $B_{4,1} = x_4$, $B_{4,2} = 3x_2^2 + 4x_1x_3$, $B_{4,3} = 6x_1^2x_2$, $B_{4,4} = x_1^4$. Each Bell polynomial is homogeneous of degree k. For the proofs of these standard result in combinatorics, we refer to the monograph

L. Comtet, Advanced Combinatorics, Reidel, Dordrecht, 1974,

and to Section II.7 of the following beautiful survey article:

H. Figueroa and J. Gracia-Bondia, Combinatorial Hopf algebras in quantum field theory I, Rev. Math. Phys. **17** (2005), 881–982.
Internet: `http://arxiv.org/hep-th/0408145`

3.4.4 The Faà di Bruno Hopf Algebra for the Formal Diffeomorphism Group of the Complex Plane

> The Faà di Bruno Hopf algebras are of the same general type as the Kreimer–Connes–Moscovici Hopf algebras; they are in fact Hopf subalgebras of the Connes–Moscovici Hopf algebras. The latter appeared in connection with the index formula for transversally elliptic differential operators on a foliation.[16]
>
> Héctor Figueroa and José Gracia–Bondia, 2005

The formal diffeomorphism group. Let G be the set of al formal power series expansions of the form

$$f(x) = x + \sum_{k=2}^{\infty} \frac{f_k}{k!}\, x^k$$

with complex coefficients $f_2, f_3, \ldots$ With respect to composition $f \circ g$, the set G becomes a group called the (formal) local diffeomorphism group of the Gaussian plane $\mathbb{C}$ at the origin. The unit element $\mathbf{1}$ of G corresponds to the power series x.

The coordinate maps. We define the nth coordinate map $\chi_n : G \to \mathbb{C}$ by setting

[16] A. Connes and H. Moscovici, Hopf algebras, cyclic cohomology and the transverse index theorem, Commun. Math. Phys. **198** (1988), 199–246.
A. Connes and D. Kreimer, Hopf algebras, renormalization and noncommutative geometry, Commun. Math. Phys. **199** (1998), 203–242. See also the footnote on page 115.

$$\chi_n(f) := f_n, \qquad n = 1, 2, \ldots$$

This map sends the formal power series f to its nth coefficient f_n. For all indices $n, m = 1, 2$, all formal power series $f, g \in G$, and all complex numbers, we define the following operations:

(i) Linear combination: $(\alpha\chi_n + \beta\chi_m)(f) := \alpha\chi_n(f) + \beta\chi_m(f)$.
(ii) Product: $(\chi_n\chi_m)(f) := \chi_n(f)\chi_m(f)$.
(iii) Coproduct: $(\Delta\chi_n)(g, f) := \chi_n(f \circ g)$.
(iv) Counit: $\varepsilon(\chi_n) := \chi_n(\mathbf{1})$.
(v) Coinverse: $(S\chi_n)(f) := \chi_n(f^{-1})$.

Explicitly,

$$\varepsilon(\chi_n) = \begin{cases} 1 & \text{if} \quad n = 1, \\ 0 & \text{otherwise.} \end{cases}$$

By (3.28), special cases of the coproduct read as

$$\begin{aligned} &\Delta\chi_1 = \chi_1 \otimes \chi_1, \qquad \Delta\chi_2 = \chi_2 \otimes \chi_1 + \chi_1^2 \otimes \chi_2, \\ &\Delta\chi_3 = \chi_3 \otimes \chi_1 + \chi_1^3 \otimes \chi_3 + 3\chi_2\chi_1 \otimes \chi_2, \\ &\Delta\chi_4 = \chi_4 \otimes \chi_1 + \chi_1^4 \otimes \chi_4 + 6\chi_2\chi_1^2 \otimes \chi_3 + (3\chi_2^2 + 4\chi_3\chi_1) \otimes \chi_2. \end{aligned} \tag{3.33}$$

For example, $\Delta\chi_2(g, h) = \chi_2(g)\chi_1(f) + \chi_1(g)^2\chi_2(f) = g_2 f_1 + g_1^2 f_2$, which coincides with (3.28). More generally, it follows from Sect. 3.4.3 that

$$\boxed{\Delta\chi_n = \sum_{k=1}^{n} B_{n.k}(\chi_1, \chi_2, \ldots, \chi_{n+1-k}) \otimes \chi_k, \qquad n = 1, 2, \ldots} \tag{3.34}$$

For the coinverse, it follows from the inversion formulas (3.26) that $S\chi_1 = \chi_1$, and $S\chi_2 = -\chi_2$, as well as

$$S\chi_3 = -\chi_3 + 3\chi_2^2, \quad S\chi_4 = -\chi_4 + 10\chi_2\chi_3 - 15\chi_2^3. \tag{3.35}$$

In the general case, the coinverse $S\chi_n, n = 1, 2, \ldots$, is explicitly given by the Lagrange inversion formula for the power series coefficients of inverse functions. Different formulations will be considered in the next section (e.g., the generalized Zimmermann forest formula).

The Hopf algebra. Let the symbol $\mathbb{C}[\chi_1, \chi_2, \ldots]$ denote the complex algebra generated by the coordinate functions $\chi_1, \chi_2, \ldots$ Explicitly, the elements of this algebra are precisely all the polynomials

$$\sum_r \alpha_{\gamma_1\gamma_2\ldots\gamma_r} \chi_1^{\gamma_1}\chi_2^{\gamma_2} \cdots \chi_r^{\gamma_r}$$

of arbitrary order with respect to the variables $\chi_1, \chi_2, \ldots$ and complex coefficients $\alpha_{\ldots}$. This algebra contains the complex numbers as polynomials of degree zero. Equipped with the usual linear combinations and products of polynomials, the algebra $\mathbb{C}[\chi_1, \chi_2, \ldots]$ is a complex commutative algebra with the complex number 1 as unit element.[17] By multilinearity, we extend the coproduct Δ, the counit ε and the coinverse S to $\mathbb{C}[\chi_1, \chi_2, \ldots]$. For example, we define the coproduct of an algebra element

[17] This can be identified with the constant function χ_1.

$$\Delta\left(\sum \alpha_{\gamma_1\gamma_2\ldots\gamma_r}\chi_1^{\gamma_1}\chi_2^{\gamma_2}\cdots\chi_r^{\gamma_r}\right)$$

by the expression

$$\sum \alpha_{\gamma_1\gamma_2\ldots\gamma_r}(\Delta\chi_1)^{\gamma_1}(\Delta\chi_2)^{\gamma_2}\cdots(\Delta\chi_r)^{\gamma_r}.$$

The corresponding formulas for ε and S are obtained by replacing Δ by ε and S, respectively. Note that $\Delta\chi_n : G\times G \mapsto \mathbb{C}$ is a function of two variables $f, g \in G$, and the product $\Delta\chi_n\Delta\chi_m$ refers to the product of functions of two variables, that is $(\Delta\chi_n\Delta\chi_m)(g,f) := (\Delta\chi_n)(g,f)\cdot\Delta\chi_m(g,f)$.

Proposition 3.6 *The algebra* $\mathbb{C}[\chi_1,\chi_2,\ldots]$ *is a commutative Hopf algebra.*

The proof proceeds similarly to the proof given in Problem 3.10 on page 170.

3.4.5 The Generalized Zimmermann Forest Formula

> The coinverse (also called antipode) of Hopf algebras allows us to elegantly describe complicated inversion processes in mathematics and physics.
>
> Folklore

> It turned out that the whole iterative and intricate structure of renormalization theory could be mapped to the theory of Hopf algebras, with Zimmermann's forest formula for the counterterm coming along as antipode.[18]
>
> Dirk Kreimer, 1994

The iterative formula for the coinverse. The coinverse can be computed by the following iterative formula:

$$\boxed{S\chi_n = -\chi_n - \left(\sum a_{(1)}Sa_{(2)} - S\chi_n - \chi_n\right), \qquad n=2,3,\ldots} \tag{3.36}$$

where we have

$$\Delta\chi_n = \sum a_{(1)}\otimes a_{(2)},$$

in the Sweedler notation. Moreover, $S\chi_1 = \chi_1$, and $\chi_1 = 1$. Let us discuss formula (3.36). Observe first that this is a trivial identity. In fact, since $\varepsilon(\chi_n) = 0$ for $n = 2,3,\ldots$ and $\mu(\mathrm{id}\otimes S)\Delta = \eta\varepsilon$, we get

$$\mu(\mathrm{id}\otimes S)\Delta(\chi_n) = \eta\varepsilon(\chi_n) = 0, \quad n = 2,3,\ldots.$$

This implies

$$\mu(\mathrm{id}\otimes S)\sum a_{(1)}\otimes a_{(2)} = \mu\left(\sum a_{(1)}\otimes Sa_{(2)}\right) = \sum a_{(1)}Sa_{(2)} = 0.$$

Secondly, the right-hand side of (3.36) does not explicitly contain the term $S\chi_n$. This follows from $\Delta(\chi_n) = \chi_n\otimes\chi_1 + \chi_1^n\otimes\chi_n + \ldots$ and hence

$$\sum a_{(1)}Sa_{(2)} = \chi_nS\chi_1 + \chi_1^nS\chi_n + \ldots = \chi_n + S\chi_n + \ldots, \qquad n=2,3,\ldots$$

Let us use (3.36) in order to successively compute $S\chi_2, S\chi_3$, and $S\chi_4$.

[18] D. Kreimer, Knots and Feynman Diagrams, Cambridge University Press, 1994. See also Sect. 19.3 on page 990.

(i) $n = 2$. It follows from the coproduct $\Delta\chi_2 = \chi_2 \otimes \chi_1 + \chi_1^2 \otimes \chi_2$ that

$$S\chi_2 = -\chi_2 + (\chi_2 S\chi_1 + \chi_1^2 S\chi_2 - S\chi_2 - \chi_2) = -\chi_2.$$

(ii) $n = 3$. By $\Delta\chi_3 = \chi_3 \otimes \chi_1 + \chi_1^3 \otimes \chi_3 + 3\chi_2\chi_1 \otimes \chi_2$,

$$S\chi_3 = -\chi_3 - 3\chi_2\chi_1 S\chi_2 = -\chi_3 + 3\chi_2^2.$$

(iii) $n = 4$. By $\Delta\chi_4 = \chi_4 \otimes \chi_1 + \chi_1^4 \otimes \chi_4 + 6\chi_2\chi_1^2 \otimes \chi_3 + (3\chi_2^2 + 4\chi_3\chi_1) \otimes \chi_2$,

$$\begin{aligned} S\chi_4 &= -\chi_4 - 6\chi_2 S\chi_3 - (3\chi_2^2 + 4\chi_3)S\chi_2 \\ &= -\chi_4 - 6\chi_2(-\chi_3 + 3\chi_2^2) + (3\chi_2^2 + 4\chi_3)\chi_2 \\ &= -\chi_4 - 15\chi_2^3 + 10\chi_2\chi_3. \end{aligned}$$

Next we want to get global explicit formulas for the coinverse which are not based on an iterative process.

The alternating coinverse formula. We have $S\chi_1 = \chi_1$, $S\chi_2 = -\chi_2$, and

$$S\chi_n = -\chi_n + \sum_{j=2}^{n-1} (-1)^j B(\chi_1, \ldots, \chi_{n-1}), \qquad n = 3, 4, \ldots \tag{3.37}$$

Here, $B(\chi_1, \ldots, \chi_{n-1})$ is equal to

$$\sum_{0<k_{j-1}<\cdots<k_1<n} B_{n,k_1} B_{k_1,k_2} \cdots B_{k_{j-2},k_{j-1}} \chi_{k_{j-1}}.$$

To simplify notation, the arguments of the Bell polynomials are suppressed. For example, we get $S\chi_3 = -\chi_3 + B_{32}\chi_2 = -\chi_3 + 3\chi_2^2$. Furthermore,

$$\begin{aligned} S\chi_4 &= -\chi_4 + B_{4,2}\chi_2 + B_{4,3}\chi_3 - B_{4,3}B_{3,2}\chi_2 \\ &= -\chi_4 + 3\chi_2^3 + 4\chi_2\chi_3 + 6\chi_2\chi_3 - 18\chi_2^3 = -\chi_4 + 10\chi_2\chi_3 - 15\chi_2^3. \end{aligned}$$

Finally, let us compute

$$\begin{aligned} S\chi_5 &= -\chi_5 + (B_{5,2}\chi_2 + B_{5,3}\chi_3 + B_{5,4}\chi_4) \\ &\quad -(B_{5,4}B_{4,3}\chi_3 + B_{5,4}B_{4,2}\chi_2 + B_{5,3}B_{3,2}\chi_2) + B_{5,4}B_{4,3}B_{3,2}\chi_2 \\ &= -\chi_5 + 15\chi_2\chi_4 + 10\chi_3^2 + 25\chi_2^2\chi_3 - (130\chi_2^2\chi_3 + 75\chi_2^4) + 180\chi_2^4 \\ &= -\chi_5 + 15\chi_2\chi_4 + 10\chi_3^2 - 105\chi_2^2\chi_3 + 105\chi_2^4. \end{aligned}$$

The sophisticated Zimmermann type cancellation formula. The computations above are based on cancellations. Now we will formulate a formula for the coinverse which works in a much simpler way. For $n = 2, 3, \ldots$, this elegant formula reads as

$$\boxed{S\chi_n = \sum_{k=1}^{n-1} (-1)^k n(n+1)\cdots(n-1+k) B_{n-1,k}\left(\frac{\chi_2}{2}, \frac{\chi_3}{3}, \ldots\right).} \tag{3.38}$$

This is also called the generalized Zimmermann forest formula. For example,

$$S\chi_4 = -4B_{3,1} + 20B_{3,2} - 120B_{3,3} = -\chi_4 + 10\chi_2\chi_3 - 15\chi_2^3.$$

Remark. The iterative formula (3.36) represents a model for the Bogoliubov iterative formula in the Bogoliubov–Parasiuk–Hepp–Zimmermann (BPHZ) approach to the renormalization of quantum field theories. Similarly, the condensed coinverse formula (3.38) represents a model for the famous Zimmermann forest formula in the BPHZ approach. It is a general phenomenon observed in combinatorics that one frequently encounters alternating sums. As a rule, such formulas are not optimal, since huge cancellations occur.

> *It is the goal of combinatorics to get optimal formulas which contain a minimal number of cancellations.*

Such optimal formulas like (3.38) are not easy to get. The elementary proofs of the coinverse formulas (3.37) and (3.36) above can be found in Figueroa and Gracia-Bondia (2005) (see the reference on page 136). For the sophisticated proof of the condensed coinverse formula (3.38), we refer to:

> W. Schmitt, Antipodes and incidence Hopf coalgebras, J. Combinatorial Theory **A46** (1987), 264–290.
>
> M. Haimann and W. Schmitt, Incidence algebra antipodes and Lagrange inversion in one and several variables, J. Combinatorial Theory **A50** (1989), 172–189.

This proof is based on the modern theory of incidence Hopf algebras in combinatorics. The generalized Zimmermann forest formula (3.38) is a high-light in combinatorics.

3.4.6 The Logarithmic Function and Schur Polynomials

The basic formula. For the formal power series $1 + \sum_{n=1}^{\infty} c_n x^n$ with complex coefficients $c_1, c_2, \ldots$, we get the following formal power series

$$\boxed{\ln\left(1 + \sum_{n=1}^{\infty} c_n x^n\right) = \sum_{n=1}^{\infty} S_n(c_1, \ldots, c_n) x^n} \tag{3.39}$$

with the so-called Schur polynomials[19] $S_n(c_1, \ldots, c_n)$ defined by

$$\frac{1}{n!} \sum_{k=1}^{n} (-1)^{k-1} (k-1)! B_{n,k}(c_1, 2!c_2, 3!c_3, \ldots, (n+1-k)!c_{n+1-k}).$$

Conversely, we have

$$\exp\left(\sum_{n=1}^{\infty} S_n(c_1, \ldots, c_n) x^n\right) = 1 + \sum_{n=1}^{\infty} c_n x^n. \tag{3.40}$$

Set $g(x) := \sum_{n=1}^{\infty} \frac{g_n}{n!} x^n$ with $g_n := c_n n!$ and

$$f(y) := \ln(1+y) = \sum_{n=1}^{\infty} \frac{f_n}{n!}\, y^n$$

[19] Schur (1875–1941).

with $f_n := (-1)^{n-1}(n-1)!$. Then formula (3.39) follows from the Faà di Bruno composition formula (3.30) applied to $f \circ g$. □

Explicitly, the first Schur polynomials read as follows:

$$S_1(c_1) = c_1, \quad S_2(c_1, c_2) = c_2 - \frac{c_1^2}{2}, \quad S_3(c_1, c_2, c_3) = c_3 - c_1 c_2 + \frac{c_1^3}{3}.$$

3.4.7 Correlation Functions in Quantum Field Theory

> In the perturbative approach to quantum field theory, combinatorial formulas are used for computing correlation functions of interacting quantum fields by means of simpler correlation functions of free fields.
>
> Folklore

Let us briefly discuss the relation between the Faà di Bruno Hopf algebra and several types of correlation functions. Quantum field theories are described by correlation functions

$$C_n(x_1, x_2, \ldots, x_n), \qquad n = 2, 3, \ldots$$

which describe the correlations between the quantum field at the different space-time points $x_1, x_2, \ldots, x_n$. As a rule, the correlation functions have strong singularities if two space-time points coincide. In combinatorics, a basic tool is given by generating functions. Motivated by this, let us set

$$Z[J] := 1 + \sum_{n=1}^{\infty} \frac{\mathrm{i}^n}{n!\hbar^n} \int_{\mathbb{R}^{4n}} C_n(x_1, \ldots, x_n) J(x_1) \cdots J(x_n) d^4x_1 \cdots d^4x_n$$

and

$$Z_{\mathrm{red}}[J] := \sum_{n=1}^{\infty} \frac{\mathrm{i}^n}{n!\hbar^n} \int_{\mathbb{R}^{4n}} C_{n,\mathrm{red}}(x_1, \ldots, x_n) J(x_1) \cdots J(x_n) d^4x_1 \cdots d^4x_n.$$

Here, C_n (resp. $C_{n,\mathrm{red}}$) is called the n-correlation function (resp. the reduced n-correlation function). We assume that the functions C_n and $C_{n,\mathrm{red}}$ with $n = 1, 2, \ldots$ are symmetric with respect to all the arguments. The proof of the following proposition will be given in Problem 3.14 on page 173.

Proposition 3.7 *Let $n = 1, 2, \ldots$ Suppose that we have*

$$C_n(x_1, \ldots, x_n) = \sum_{k=1}^{n} \sum_{p \in \Pi_{n,k}} \left(\prod_p C_{\mathrm{red}} \right)(x_1, \ldots, x_n)$$

for all arguments $x_1, \ldots, x_n \in \mathbb{R}^4$. Then, $Z[J] = e^{Z_{\mathrm{red}}[J]}$, in the sense of a formal power series.

The symbol $\sum_{p \in \Pi_{n,k}} \left(\prod_p C_{\mathrm{red}} \right)(x_1, \ldots x_n)$ means that we sum over all possible partitions p of the set $\{1, \ldots, n\}$ into k blocks. For each partition p, we consider a product of reduced correlation functions whose arguments are given by p. For example, we have $C_1(x_1) = C_{1,\mathrm{red}}(x_1)$. Furthermore,

$$C_2(x_1, x_2) = C_{1,\mathrm{red}}(x_1) C_{1,\mathrm{red}}(x_2) + C_{2,\mathrm{red}}(x_1, x_2)$$

and

$$C_3(x_1,x_2,x_3) = C_{1,\mathrm{red}}(x_1)C_{1,\mathrm{red}}(x_2)C_{1,\mathrm{red}}(x_3) + C_{1,\mathrm{red}}(x_1)C_{2,\mathrm{red}}(x_2,x_3)$$
$$+C_{1,\mathrm{red}}(x_2)C_{2,\mathrm{red}}(x_1,x_3) + C_{1,\mathrm{red}}(x_3)C_{2,\mathrm{red}}(x_1,x_2) + C_{3,\mathrm{red}}(x_1,x_2,x_3).$$

If one knows the correlation functions $C_1, C_2, \ldots$, then the reduced correlation functions $C_{1,\mathrm{red}}, C_{2,\mathrm{red}}, \ldots$ can be computed by means of the basic formula

$$\boxed{Z_{\mathrm{red}}[J] = \ln Z[J].} \tag{3.41}$$

In particular, if $C_1 = 0$, then $C_{1,\mathrm{red}} = 0$ and

$$C_{2,\mathrm{red}}(x_1,x_2) = C_2(x_1,x_2). \tag{3.42}$$

Furthermore, $C_{3,\mathrm{red}} = C_3$ and

$$C_{4,\mathrm{red}}(x_1,x_2,x_3,x_4) = C_4(x_1,x_2,x_3,x_4) - C_2(x_1,x_2)C_2(x_3,x_4)$$
$$-C_2(x_1,x_3)C_2(x_2,x_4) - C_2(x_1,x_4)C_2(x_2,x_3) \tag{3.43}$$

for all $x_1, x_2, x_3, x_4 \in \mathbb{R}^4$. These formulas for the 2-correlation function and the 4-correlation function are encountered in the theory of interacting quantum fields. In the special case of free quantum fields, we have $C_{1,\mathrm{red}} = 0$ and $C_{n,\mathrm{red}} = 0$ for all $n = 3, 4, \ldots$ All the information about a free quantum field is contained in the 2-point correlation function C_2.

Physicists use this result in order to simplify the computation of correlation functions for interacting quantum fields. It turns out that the reduced correlation functions $C_{n,\mathrm{red}}$ correspond to connected Feynman graphs (also called Feynman diagrams), whereas the general correlation functions C_n correspond to all kinds of (connected or disconnected) Feynman graphs. Therefore, the functions C_n and $C_{n,\mathrm{red}}$ are also called the n-Green's function and the connected n-Green's function, respectively.

> *For interacting quantum fields, it is sufficient to investigate connected Feynman graphs. The effects related to disconnected Feynman graphs can be reduced to the corresponding effects coming from connected Feynman graphs.*

In other words, the reduced correlation functions form a (nonlinear) basis for general correlation functions. This is closely related to the theory of cumulants in probability theory to be studied in the next section.

It turns out that even the connected Feynman graphs can still have a certain redundancy. Therefore physicists pass over to Feynman graphs with a nontrivial topology (called one-particle irreducible graphs). This allows us to reduce the computation of n-correlation functions C_n to both irreducible vertex functions and 2-correlation functions.

Finally, let us discuss the relation to the Faà di Bruno Hopf algebra. To this end, set

$$X_n := \frac{\mathrm{i}^n}{\hbar^n} \int_{\mathbb{R}^{4n}} C_n(x_1,\ldots,x_n)J(x_1)\cdots J(x_n)d^4x_1\cdots d^4x_n.$$

Using the Faà di Bruno formula (3.31), we obtain that the exponential function $\exp(Z_{\mathrm{red}}[J]) = \exp\left(\sum_{n=1}^{\infty} \frac{1}{n!} \cdot X_n\right)$ is equal to

$$1 + \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{k=1}^{n} \sum_{\lambda \in \Pi_{n,k}} \frac{n!}{\lambda_1!\lambda_2!\cdots\lambda_n!(1!)^{\lambda_1}(2!)^{\lambda_2}\cdots(n!)^{\lambda_n}} \; X_1^{\lambda_1} X_2^{\lambda_2} \cdots X_n^{\lambda_n}.$$

In turn, Proposition 3.7 above tells us that this is equal to $Z[J]$.

3.4.8 Random Variables, Moments, and Cumulants

> Generating functions for the moments of random variables represent a basic tool in quantum field theory.
>
> Folklore

The following considerations are basic for the theory of random variables. Let X and Y be two independent random variables. Then, for the mean values and the mean fluctuations, we have the following additivity property

$$\overline{X+Y} = \overline{X} + \overline{Y}, \qquad (\Delta X + \Delta Y)^2 = (\Delta X)^2 + (\Delta Y)^2.$$

We want to generalize this to higher moments. The basic idea is to pass from the multiplicative family of moments to the additive family of cumulants by using the logarithmic function. In terms of algebra, the passage from cumulants to moments is given by Schur polynomials. In quantum field theory, the passage from moments to cumulants corresponds to the passage from correlation functions (i.e., Green's functions) to reduced correlation functions (i.e., connected Green's functions). In terms of Feynman diagrams, this corresponds to a passage from general Feynman graphs to connected Feynman graphs (see (3.41)). The proofs for the following statements can be found in Sect. II.12 of the textbook by A. Shiryaev, Probability, Springer, New York, 1996.

Random variable on the real line. Let $\varrho : \mathbb{R} \to \mathbb{R}$ be a nonnegative smooth function which rapidly decreases at infinity (i.e., $\varrho \in \mathcal{S}(\mathbb{R})$) and which satisfies the normalization condition $\int_{\mathbb{R}} \varrho(x)dx = 1$. The function ϱ generates a probability measure μ and a random variable X on the real line with the corresponding Lebesgue-Stieljes integral

$$\int_{\mathcal{B}} f d\mu = \int_{\mathcal{B}} f(x)\varrho(x)dx.$$

For each Borel subset $\mathcal{B}$ of the real line (e.g., intervals), the measure $\mu(\mathcal{B})$ is equal to $\mu(\mathcal{B}) = \int_{\mathcal{B}} d\mu$. Moreover, we get

$$p(X \in \mathcal{B}) = \mu(\mathcal{B}).$$

This means the following. Suppose that we measure the variable X in an experiment. Then the probability of finding the measured value in the set $\mathcal{B}$ is equal to $\mu(\mathcal{B})$. For a function $f : \mathbb{R} \to \mathbb{C}$, we define the mean value

$$\overline{f(X)} := \int_{\mathbb{R}} f(x)d\mu(x)$$

provided this integral exists.This is the case if the function f is smooth and the function $x \mapsto |f(x)|$ has at most polynomial growth at infinity.

Moments. The moments of the random variable X are defined by

$$m_k := \overline{X^k} = \int_{\mathbb{R}} x^k d\mu, \qquad k = 1, 2, \ldots$$

The characteristic function. The function

$$\varphi_X(\xi) := \overline{\mathrm{e}^{\mathrm{i}\xi x}} = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\xi x} d\mu(x), \qquad \xi \in \mathbb{R}$$

is called the characteristic function to the random variable X. In other words, this is the Fourier–Stieltjes transform with respect to the probability measure μ. Differentiation yields $\varphi_X'(0) = \mathrm{i}\int_{\mathbb{R}} x d\mu(x)$. Hence

$$\varphi_X^{(k)}(0) = \mathrm{i}^k m_k \qquad k = 1, 2, \dots$$

By Taylor expansion,

$$\varphi_X(\xi) := 1 + \sum_{k=1}^{n} \mathrm{i}^k m_k \frac{\xi^k}{k!} + o(\xi^n), \qquad \xi \to 0, \;\; n = 1, 2, \dots$$

This way, we obtain the formal power series expansion:

$$\varphi_X(\xi) := 1 + \sum_{k=1}^{\infty} \mathrm{i}^k m_k \frac{\xi^k}{k!}, \qquad \xi \to 0.$$

Therefore, the characteristic function φ_X is called the generating function for the moments of the random variable X. The Bochner theorem tells us the following:[20]

The continuous function $\chi : \mathbb{R} \to \mathbb{C}$ with $\chi(0) = 1$ is the characteristic function of a probability measure on the real line iff it is positive definite, that is,

$$\sum_{i,j=1}^{n} \chi(\xi_i - \xi_j)\eta_i \eta_j^{\dagger} \geq 0, \qquad n = 1, 2, \dots$$

for all real numbers $\xi_1, \dots, \xi_n$ and all complex numbers $\eta_1, \dots, \eta_n$.

This theorem is a jewel in harmonic analysis; it allows important generalizations concerning the spectral representation of kernels (in the sense of the theory of distributions).[21]

Cumulants. We pass from the generating function φ_X to the generating function

$$\ln \varphi_X(\xi) = \sum_{k=1}^{\infty} \mathrm{i}^k s_k \frac{\xi^k}{k!}. \tag{3.44}$$

This is a formal power series. The coefficients $s_1, s_2, \dots$ are called the cumulants of the random variable X. There exists a one-to-one relation between moments and cumulants. Explicitly, we get the rescaled Schur polynomials

$$m_1 = s_1, \qquad m_2 = s_2 + s_1^2, \qquad m_3 = s_3 + 3s_1 s_2 + s_1^3.$$

Conversely, we obtain $\mathrm{e}^{\ln \varphi_X(\xi)} = \varphi_X(\xi)$, and hence

$$s_1 = m_1 = \overline{X}, \quad s_2 = m_2 - m_1^2 = (\Delta X)^2, \quad s_3 = m_3 - 3m_1 m_2 + 2m_1^3.$$

The crucial additivity property of cumulants. Now consider two random variables X, Y on the real line which are independent, that is,

$$p(X \in \mathcal{B}, \;\; Y \in \mathcal{C}) = p(X \in \mathcal{B}) \cdot p(Y \in \mathcal{C})$$

[20] S. Bochner, Lectures on Fourier Integrals, Leipzig, 1932 (in German).

[21] See K. Maurin, Generalized Eigenfunction Expansions and Unitary Representations of Topological Groups, Sect. II.5, PWN, Warsaw, 1968.

for all Borel sets $\mathcal{B}, \mathcal{C}$ on the real line. In other words, the probability measure of the random vector (X, Y) on the plane is the product measure of the measures corresponding to X and Y. Then we have the product property

$$\varphi_{X+Y} = \varphi_X \varphi_Y$$

for the generating functions. This implies

$$\ln(\varphi_{X+Y}) = \ln \varphi_X + \ln \varphi_Y.$$

Consequently, for two independent random variables X and Y, the cumulants are additive quantities, that is,

$$s_{k,X+Y} = s_{k,X} + s_{k,Y}, \qquad k = 1, 2, \dots$$

In particular, we get $\overline{X+Y} = \overline{X} + \overline{Y}$, and $\{\Delta(X+Y)\}^2 = (\Delta X)^2 + (\Delta Y)^2$ if we choose $k = 1, 2$.

Solution of the moment problem. We want to reconstruct the probability measure from its moments. This is the moment problem.

Suppose that we are given a sequence $m_1, m_2, \dots$ of nonnegative numbers with

$$\limsup_{k\to\infty} \frac{m_k^{1/k}}{k} < \infty \qquad \text{and} \qquad \sum_{k=1}^{\infty} \frac{1}{m_{2k}^{1/2k}} = \infty.$$

Then there exists a uniquely determined probability measure μ on the real line such that the numbers $m_1, m_2, \dots$ are the moments of μ.

The moment problem played an important role in the development of measure theory and functional analysis. In particular, the Riesz–Markov representation theorem (about linear continuous functionals on spaces of real continuous functions defined on compact sets) and the Hahn–Banach theorem (on the extension of linear continuous functionals in Banach spaces) were proved for solving the moment problem. This fascinating history can be found in J. Dieudonné, History of Functional Analysis, 1900–1975, North-Holland, Amsterdam, 1983. We also refer to J. Shohat and J. Tamarkin, The Problem of Moments, New York, 1950, and P. Lax, Functional Analysis, Sect. 33.5, Wiley, New York, 2002.

Probability measures on the plane. Let $\varrho = \varrho(x, y)$ be a nonnegative smooth function on the plane with $\varrho \in \mathcal{S}(\mathbb{R}^2)$ and $\int_{\mathbb{R}^2} \varrho(x, y)dxdy = 1$. The corresponding moments of this probability measure are given by

$$m_{k,l} := \int_{\mathbb{R}^2} x^k y^l \varrho(x, y)dxdy, \qquad k, l = 0, 1, 2, \dots$$

with the generating function

$$\varphi(\xi, \eta) := \sum_{k,l=0}^{\infty} \mathrm{i}^{k+l} m_{k,l} \frac{\xi^k \eta^l}{k!l!}$$

given by $\varphi(\xi, \eta) := \overline{\mathrm{e}^{\mathrm{i}(\xi x+\eta y)}}$. Here, $m_{0,0} = 1$. Finally, set

$$\ln \varphi(\xi, \eta) = \sum_{k,l=0}^{\infty} \mathrm{i}^{k+l} s_{kl} \frac{\xi^k \eta^l}{k!l!}.$$

Here, $s_{0,0} = 0$. Then there exists a one-to-one correspondence between the moments m_{kl} and the cumulants s_{kl}. The same is true if we pass to probability measures on $\mathbb{R}^n$ with $n = 3, 4, \dots$ The power series expansions for φ and $\ln \varphi$ are to be understood as formal power series expansions.

3.5 Symmetry and Hopf Algebras

> Whoever understands symmetries can understand everything in this world.
> Folklore

Symmetries play a crucial role in mathematics and physics. As a rule, it is only possible to explicitly solve a mathematical or physical problem if a certain symmetry is available. In the history of sciences, mathematicians and physicists encountered more and more complicated symmetries.

3.5.1 The Strategy of Coordinatization in Mathematics and Physics

In what follows, we will show that appropriate complex-valued coordinate functions

$$\chi_j : G \to \mathbb{C}, \qquad j = 1, 2, \dots \tag{3.45}$$

on a group G generate a Hopf algebra denoted by $\mathcal{H}(G)$. This is the so-called coordinate algebra of G. In ancient times, mathematicians studied geometric objects. The 'coordinatization' of geometric objects arose in analytic geometry founded by Descartes (1596–1650).[22] The physicist Galilei (1564–1642) wrote:

> *Measure everything that is measurable, and make measurable everything that is not yet so.*

In terms of physics,

(i) the group G above describes abstract physical quantities, and
(ii) the coordinate functions $\chi_1, \chi_2, \dots$ correspond to measurements performed by an observer.

Typically, the mathematical structure concerning (ii) looks more complicated (e.g., Hopf algebras appear) than the mathematical structure concerning (i).[23] Furthermore, note the following:

> *The theory of Hopf algebras allows us to study symmetries which are not necessarily related to the classical theory of Lie groups and Lie algebras.*

For example, this leads to quantum groups. In the physics literature, roughly speaking, Hopf algebras are also called quantum groups. Sometimes, only special Hopf algebras are called quantum groups (e.g., one-parameter deformations of the enveloping algebras of semi-simple Lie algebras).

In (3.45), the translation of properties of the group G into properties of the coordinate functions corresponds to a general strategy used in mathematics and physics:

> *Investigate mathematical objects by studying families of maps defined on the objects.*

In topology, this leads to the crucial concept of cohomology. In physics, this means that we investigate the properties of the space-time by studying physical fields depending on space and time. For the convenience of the reader, let us summarize mathematical and physical topics which are closely related to the concept of Hopf algebra:

---

[22] The general strategy of 'coordinatization' in algebra and algebraic geometry is studied in I. Shafarevich, Algebra I, Springer, Berlin, 1990 (Encyclopedia of Mathematical Sciences).

[23] For example, see Sect. 3.5.3 on the coordinate Hopf algebra of an operator algebra.

- factorization of the scattering matrix and the Yang–Baxter equation,
- integrable models in statistical physics and the Yang–Baxter equation,
- solutions of the Yang–Baxter equation by means of Hopf algebras and the braid group,
- Artin's braid group and braid group statistics of particles,
- integrable models in quantum field theory and quantum groups,
- conformal field theory, Virasora algebras, affine Lie algebras (Kac–Moody algebras), Verma modules, vertex algebras, and operator products,
- vertex algebras, the completion of the classification of the finite simple groups by discovering the monster group, which acts on the monstrous moonshine algebra,
- vertex algebras and algebraic curves in algebraic geometry,
- complex function theory, Riemann surfaces, conformal field theory, and strings,
- fusion rules for Feynman diagrams in conformal field theory and the Verlinde formula,
- models in quantum gravitation, the Moyal product, and the Seiberg–Witten map,
- generalized differential calculi (with Leibniz rule) and noncommutative geometry,
- quantum groups and new topological invariants of knots and 3-dimensional manifolds due to Jones (related to von Neumann algebras), Vassiliev, and Kontsevich,
- Witten's topological quantum field theory and topological invariants of knots and 3-dimensional manifolds,
- number theory (lattices, modular forms, and zeta functions),
- Frobenius manifolds, quantum cohomology and moduli spaces.

In mathematics and physics, one frequently encounters the solution of problems by using iterative processes. The prototype of such processes is given by the equation

$$x - Ax = y$$

with the solution

$$x = (I - A)^{-1}y = (I + A + A^2 + \ldots)y$$

and the corresponding iterative method

$$x_{n+1} = Ax_n + y, \qquad n = 0, 1, 2, \ldots$$

with $x_0 := 0$. This yields

$$x_1 = y, \quad x_2 = y + Ay, \ldots, x_n = y + Ay + Ay^2 + \ldots A^{n-1}y.$$

However, there are more complicated problems where this simple method fails. As a rule of thumb, complicated iteration methods are governed by Hopf algebras. Two important examples are:

- the widely used implicit Runge–Kutta method for solving ordinary differential equations in numerical mathematics (the use of Butcher series with respect to the Hopf algebra of rooted trees for computing the coefficients of the higher-order methods),
- Feynman diagrams and renormalization.

We refer to:

J. Butcher, An algebraic theory of integration methods, Math. Comp. **26** (1972), 79–106.

C. Brouder, Runge–Kutta methods and renormalization, Eur. J. Phys. C **12** (2000), 521–534. Internet: `http://arxiv.org/hep-th/9904014`

Hints for further reading can be found in Sect. 17.4 of Vol. I.

3.5.2 The Coordinate Hopf Algebra of a Finite Group

Finite groups can equivalently described by incidence numbers.
Folklore

Let G be a finite group with the unit element e. Let $\mathcal{H}(G)$ denote the set of all complex-valued functions

$$\varphi : G \to \mathbb{C}$$

on the group G. For functions $\varphi, \psi \in \mathcal{H}(G)$, group elements $g, h \in G$, and complex numbers α, β, we define the following operations:

(i) Linear combination: $(\alpha\varphi + \beta\psi)(g) := \alpha\varphi(g) + \beta\psi(g)$.
(ii) Product: $(\varphi\psi)(g) := \varphi(g)\psi(g)$.
(iii) Unit element: $\mathbf{1}(g) := 1$.
(iv) Coproduct: $(\Delta\varphi)(g, h) := \varphi(gh)$.
(v) Counit: $\varepsilon(\varphi) := \varphi(e)$.
(vi) Coinverse: $(S\varphi)(g) := \varphi(g^{-1})$.

Concerning the definition of $\alpha\varphi + \beta\psi, \varphi\psi, \mathbf{1}, \Delta(\varphi), \varepsilon(\varphi)$, and $S(\varphi)$, note the following:

- The coproduct Δ sends complex-valued functions $g \mapsto \varphi(g)$ of one variable on the group G to functions $(g, h) \mapsto \varphi(gh)$ of two variables by using the group product gh.
- The counit ε sends complex-valued functions $g \mapsto \varphi(g)$ of one variable on the group G to complex numbers $\varphi(e)$ by using the unit element e of the group.
- The coinverse S sends complex-valued functions $g \mapsto \varphi(g)$ of one variable on G to functions $g \mapsto \varphi(g^{-1})$ of one variable by using the inverse g^{-1} of the group element g.

The maps Δ, S, and ε are linear and multiplicative. This means that, for all functions $\varphi, \psi : G \to \mathbb{C}$ and all complex numbers α, β, we have

$$\Delta(\alpha\varphi + \beta\psi) := \alpha\Delta\varphi + \beta\Delta\psi, \qquad \Delta(\varphi\psi) = \Delta\varphi\Delta\psi.$$

The same is true if we replace Δ by S (resp. ε).

Recall the following definition. For given functions $\varphi, \psi : G \to \mathbb{C}$, we define the tensor product $\varphi \otimes \psi$ by setting

$$(\varphi \otimes \psi)(g, h) := \varphi(g)\psi(h) \qquad \text{for all} \quad g, h \in G.$$

This is a function of the form $\varphi \otimes \psi : G \times G \to \mathbb{C}$. In other words, $\varphi \otimes \psi$ is a function of two variables. Note the following peculiarity which follows from the finiteness of the group G. The functions

$$\chi : G \times G \to \mathbb{C}$$

of two variables are in one-to-one correspondence to the elements of the tensor product $\mathcal{H}(G) \otimes \mathcal{H}(G)$. We will show this in (3.46) below by using a basis. In this sense, the coproduct is a map of the form $\Delta : \mathcal{H}(G) \to \mathcal{H}(G) \otimes \mathcal{H}(G)$.

Proposition 3.8 *The algebra $\mathcal{H}(G)$ is a commutative Hopf algebra. The dimension of $\mathcal{H}(G)$ is equal to the number of group elements. The Hopf algebra $\mathcal{H}(G)$ is cocommutative iff the group G is commutative.*

The proof will be given in Problem 3.11 on page 172.

Basis. Let us formulate the operations of the algebra $\mathcal{H}(G)$ in terms of a basis. Then the coproduct, the counit, and the coinverse can be expressed by using so-called incidence numbers. Let $g_1, g_2, \ldots, g_m$ denote the elements of the group G where g_1 is the unit element e. For $j = 1, \ldots, m$, we introduce the special coordinate functions $\chi_j : G \to \mathbb{C}$ by setting $(k = 1, \ldots, m)$:

$$\chi_j(g_k) := \begin{cases} 1 & \text{if} \quad g_k = g_j, \\ 0 & \text{if} \quad g_k \neq g_j. \end{cases}$$

Each function $\varphi : G \to \mathbb{C}$ can be uniquely represented by the formula

$$\varphi = \sum_{k=1}^{m} \alpha_k \chi_k$$

where $\alpha_k := \varphi(g_k)$ for $k = 1, \ldots, m$. Thus, the functions $\chi_1, \ldots, \chi_m$ form a basis of the linear space $\mathcal{H}(G)$. Furthermore, each function $\psi : G \times G \to \mathbb{C}$ of two variables, $\psi = \psi(g, h)$, can be uniquely represented as

$$\psi = \sum_{j,k=1}^{m} \alpha_{jk} \chi_j \otimes \chi_k \tag{3.46}$$

with $\alpha_{jk} := \psi(g_j, g_k)$. This follows from

$$(\chi_j \otimes \chi_k)(g, h) = \chi_j(g)\chi_k(h) \qquad \text{for all} \quad g, h \in G.$$

Consequently, the functions $\psi = \psi(g, h)$ of two variables can be identified with the elements of the tensor product $\mathcal{H}(G) \otimes \mathcal{H}(G)$.

Coproduct. For the coproduct $\Delta : \mathcal{H}(G) \to \mathcal{H}(G) \otimes \mathcal{H}(G)$, we get

$$\Delta\chi_j = \sum_{r,s=1}^{m} \beta_{jrs} \chi_r \otimes \chi_s, \qquad j = 1, \ldots, m$$

with the incidence numbers $\beta_{jrs} := \chi_j(g_r g_s)$. Explicitly,

$$\beta_{jrs} = \begin{cases} 1 & \text{if} \quad g_r g_s = g_j, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\beta_{jrs} = (\Delta\chi_j)(g_r, g_s)$. The linearity of the coproduct tells us that

$$\Delta\left(\sum_{j=1}^{m} \alpha_j \chi_j\right) = \sum_{j=1}^{m} \alpha_j \Delta\chi_j.$$

The counit. For the counit $\varepsilon : G \to \mathbb{C}$, we get the incidence numbers $\varepsilon(\chi_1), \ldots, \varepsilon(\chi_n)$. Explicitly,

$$\varepsilon(\chi_j) = \begin{cases} 1 & \text{if} \quad g_j = e, \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, $\varepsilon(\sum_{j=1}^{m} \alpha_j \chi_j) = \sum_{j=1}^{m} \alpha_j \varepsilon(\chi_j)$.

The coinverse. For the coinverse $S : \mathcal{H}(G) \to \mathcal{H}(G)$, we get

$$S\chi_j = \sum_{r=1}^{m} \gamma_{jr}\chi_r, \qquad j = 1, \ldots, m$$

with the incidence numbers $\gamma_{jr} := \chi_j(g_r^{-1})$. Explicitly,

$$\gamma_{jr} := \begin{cases} 1 & \text{if} \quad g_r^{-1} = g_j, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\gamma_{js} = (S\chi_j)(g_s)$. Furthermore, for all complex numbers $\alpha_1, \ldots, \alpha_m$, we obtain

$$S\left(\sum_{j=1}^{m} \alpha_j \chi_j\right) = \sum_{j=1}^{m} \alpha_j S\chi_j.$$

Summarizing, the Hopf algebra $\mathcal{H}(G)$ equivalently represents the finite group G in terms of incidence numbers for the group operations. This way, the Hopf algebra $\mathcal{H}(G)$ represents a 'coordinatization' of the group G.

3.5.3 The Coordinate Hopf Algebra of an Operator Group

Let X be a finite-dimensional complex Hilbert space with the complete orthonormal system $b_1, b_2, \ldots, b_n$. We consider a group $\mathcal{G}$ of linear operators

$$\mathsf{A} : X \to X$$

with the unit operator id. In terms of physics, the operators A in $\mathcal{G}$ represent physical quantities independent of any measurement by observers. In order to describe measurements, we have to assign real numbers to the operator A; these real numbers are called the coordinates of A. We will proceed in two steps. First we will assign a matrix A to the operator A. Then we will describe the matrix $A = (a_{jk})_{j,k=1,\ldots,n}$ by a family $\{\chi_{jk}\}_{j,k=1,\ldots,n}$ of coordinate functions. Moreover, we will pass to the polynomial algebra generated by the coordinate functions. Finally, this polynomial algebra can be equipped with the structure of a Hopf algebra.

(i) Matrix elements a_{jk} (Dirac calculus): We set

$$a_{jk} := \langle b_j | \mathsf{A} b_k \rangle, \qquad j, k = 1, 2, \ldots, n$$

with the corresponding matrix $A := (a_{jk})$. By matrix calculus (called Dirac calculus in physics), the map $\mathsf{m} : L(X, X) \to L(\mathbb{C}^n, \mathbb{C}^n)$ given by

$$\mathsf{m}(\mathsf{A}) := A \qquad \text{for all} \quad \mathsf{A} \in L(X, X)$$

is an algebra isomorphism. Explicitly, for all linear operators $\mathsf{A}, \mathsf{B} : X \to X$ and all complex numbers α, β, we have

$$\mathsf{m}(\alpha\mathsf{A} + \beta\mathsf{B}) = \alpha \cdot \mathsf{m}(\mathsf{A}) + \beta \cdot \mathsf{m}(\mathsf{B}), \qquad \mathsf{m}(\mathsf{AB}) = \mathsf{m}(\mathsf{A})\mathsf{m}(\mathsf{B}).$$

This way, the operator group $\mathcal{G}$ is isomorphically transformed into the matrix group G, which is a subgroup of the group $GL(n, \mathbb{C})$ of invertible complex $(n \times n)$-matrices with the unit matrix I as unit element.

(ii) Coordinate functions (Hopf algebra – quantum group): Let us make the additional assumption that $\det(A) = 1$ for all $A \in G$. That is, G is a subgroup of the Lie group $SL(n, \mathbb{C})$. Let $A \in G$ where $A = (a_{jk})$. The key definitions reads as follows:

$$\boxed{\chi_{jk}(A) := a_{jk}, \qquad j, k = 1, \dots n.}$$

This means that, for fixed indices j, k, the coordinate function $\chi_{jk} : G \to \mathbb{C}$ assigns to each matrix A in the group G its matrix element a_{jk}. Furthermore, for all $j, k = 1, \dots, n$ and all $A, B \in G$, we define the following:

- Coproduct: $(\Delta\chi_{jk})(A, B) := \chi_{jk}(AB)$.
- Counit: $\varepsilon(\chi_{jk}) := \chi_{jk}(I)$.
- Coinverse $(S\chi_{jk})(A) := \chi_{jk}(A^{-1})$.

Using the matrix product, we have

$$\left(\sum_{s=1}^{n} \chi_{js} \otimes \chi_{sk}\right)(A, B) = \sum_{s=1}^{n} \chi_{js}(A)\chi_{sk}(B) = \sum_{s=1}^{n} a_{js} b_{sk} = \chi_{jk}(AB).$$

Thus, for the coproduct we get

$$\boxed{\Delta\chi_{jk} = \sum_{s=1}^{n} \chi_{js} \otimes \chi_{sk}.}$$

Moreover, $\varepsilon(\chi_{jk}) = \delta_{jk}$. Since $\det(A) = 1$, the elements of the inverse matrix A^{-1} are polynomials in the elements of A. Thus, $S\chi_{jk}$ is a polynomial with respect to the functions $\chi_{11}, \dots, \chi_{nn}$. Let us now introduce the algebra $\mathbb{C}[\chi_{11}, \dots, \chi_{nn}]$ of all polynomials

$$\sum \alpha_{\gamma_{11} \dots \gamma_{nn}} (\chi_{11})^{\gamma_{11}} (\chi_{12})^{\gamma_{12}} \cdots (\chi_{nn})^{\gamma_{nn}}$$

in the variables $\chi_{11}, \dots, \chi_{nn}$ with complex coefficients $\alpha_{\dots}$. Here the exponents $\gamma_{11}, \dots$ are nonnegative integers. The operations of this algebra are the usual complex linear combinations and products. We extend the coproduct Δ above to the algebra $\mathbb{C}[\chi_{11}, \dots, \chi_{nn}]$ by multilinear extension, that is, we define

$$\Delta\left(\sum \alpha_{\gamma_{11} \dots \gamma_{nn}} (\chi_{11})^{\gamma_{11}} (\chi_{12})^{\gamma_{12}} \cdots (\chi_{nn})^{\gamma_{nn}}\right)$$

by the expression

$$\sum \alpha_{\gamma_{11} \dots \gamma_{nn}} (\Delta\chi_{11})^{\gamma_{11}} (\Delta\chi_{12})^{\gamma_{12}} \cdots (\Delta\chi_{nn})^{\gamma_{nn}}.$$

Analogous formulas are obtained by replacing Δ by ε (resp. S).

Proposition 3.9 *The algebra $\mathbb{C}[\chi_{11}, \dots, \chi_{nn}]$ is a commutative Hopf algebra. This Hopf algebra is cocommutative iff the operator group G is commutative.*

For the proof, we refer to Problem 3.11 on page 172. In modern mathematical physics, one studies deformations of Hopf algebras. In the present case, deformations of the Hopf algebra $\mathbb{C}[\chi_{11}, \dots, \chi_{nn}]$ can be described by deformations of the coordinate functions $\chi_{11}, \dots, \chi_{nn}$. The deformation of Hopf algebras is studied in the theory of quantum groups (see Vol. IV).

3.5.4 The Tannaka–Krein Duality for Compact Lie Groups

Try to dualize in mathematics as much as you can.
Folklore

We want to generalize the classical exponential function

$$\chi(x) := \mathrm{e}^{\mathrm{i}x}, \qquad x \in \mathbb{R}$$

in the setting of groups and C^*-algebras. The starting point is the functional equation

$$\chi(x+y) = \chi(x)\chi(y) \qquad \text{for all} \quad x, y \in \mathbb{R}.$$

This means that the map $\chi : \mathbb{R} \to U(1)$ is a group morphism from the additive group $\mathbb{R}$ of real numbers onto the multiplicative group $U(1)$ of the unit circle in the complex plane. Explicitly, $U(1) := \{z \in \mathbb{C} : |z| = 1\}$. The group $U(1)$ is the prototype of a compact Lie group. Our goal is to use the group $U(1)$ in order to study the structure of more general objects, namely, groups and C^*-algebras. By definition, the character of a group G is a group morphism $\chi : G \to U(1)$, that is,

$$\boxed{\chi(gh) = \chi(g)\chi(h) \qquad \text{for all} \quad g, h \in G.}$$

For example, the characters of the group $U(1)$ are given by the following maps $\chi_n : U(1) \to U(1)$, $n = 0, \pm 1, \pm 2, \ldots$, where

$$\chi_n(\mathrm{e}^{\mathrm{i}\varphi}) := \mathrm{e}^{\mathrm{i}n\varphi}, \qquad \varphi \in \mathbb{R}. \tag{3.47}$$

In terms of the angle variable φ, this corresponds to the map $\varphi \mapsto n\varphi$. The integer n is called the winding number (or the topological charge) of the map χ_n. In this connection, let us discuss the following dualities:

- de Rham duality for manifolds and cohomology,
- Pontryagin duality for commutative compact groups,
- Tannaka–Krein duality for noncommutative compact groups,
- Gelfand–Naimark duality for commutative C^*-algebras and noncommutative geometry (see Vol. IV).

This fits into the following general strategy in mathematics:

Study the structure of a mathematical object X by investigating maps

$$\chi : X \to Y$$

that live on the object X. From the physical point of view, the prototype for this strategy is given by physical fields χ that live on the space-time manifold X.

De Rham duality. Let X be a manifold, and let $\chi : X \to Y$ be a differential form on X. Using the Cartan derivative $d\chi$ of the differential form χ with the crucial property

$$d(d\chi) = 0,$$

that is, $d^2 = 0$, the topological properties of the manifold X can be studied by de Rham cohomology which lies at the heart of modern differential topology. This yields a powerful theory of topological invariants (called topological charges in physics) including the theory of characteristic classes and Chern numbers (see Chap. 5 of Vol. I). Cohomology also lies at the heart of both the methods of BRST-quantization and algebraic renormalization.

Pontryagin duality. Let G be a commutative compact topological group[24] (e.g., a finite group or the Lie group $U(1)$). Let G' be the group of characters

$$\chi : G \to U(1)$$

of the group G. This so-called dual group G' of G is always discrete (e.g., for the group $G = U(1)$, the dual group G' consists precisely of the maps χ_n introduced in (3.47) above, where $n = 0, \pm 1, \pm 2, \ldots$). Further dualization leads us to the bidual group G'' which consists of all the characters

$$\varrho : G' \to U(1)$$

of the dual group G'. For each group element $g \in G$, we define

$$\boxed{\varrho_g(\chi) := \chi(g) \qquad \text{for all} \quad \chi \in G'.}$$

In 1934, Pontryagin proved the following theorem:[25]

The map $g \mapsto \varrho_g$ is a group isomorphism from G onto the bidual group G''.

Tannaka–Krein duality. Let G be a compact Lie group with the unit element e. We do not assume that G is commutative (e.g., G is one of the groups $U(1), SU(2), SU(3)$ which arise in the Standard Model of particle physics). We want to generalize the Pontryagin duality to this more general noncommutative situation. The trick is to pass from complex numbers to real numbers and to use real Hopf algebras. The idea is to use the following objects:

- The orginal compact Lie group G.
- The dual real Hopf algebra $\mathcal{G}'$ of all the continuous functions $\chi : G \to \mathbb{R}$ of finite type.
- The dual group $\mathcal{G}''$ to the Hopf algebra $\mathcal{G}'$, which consists of all the algebra morphisms $\mu : \mathcal{G}' \to \mathbb{R}$ equipped with the convolution product.

Explicitly, we proceed as follows. First consider the real linear space $C(G, \mathbb{R})$ of continuous real-valued functions

$$\chi : G \to \mathbb{R}. \tag{3.48}$$

For each $g \in G$, we define the function

$$\chi_g(h) := \chi(hg) \qquad \text{for all} \quad h \in G.$$

The function χ is called of finite type iff the linear hull of the set $\{\chi_g : g \in G\}$ is finite-dimensional. By definition, the space $\mathcal{G}'$ consists of all the continuous functions (3.48) of finite type. Obviously, $\mathcal{G}'$ is a linear subspace of $C(G, \mathbb{R})$. Moreover, defining

$$\Delta\chi(g, h) := \chi(gh), \qquad \varepsilon(\chi) = \chi(e), \qquad S\chi(g) := \chi(g^{-1}), \qquad g, h \in G,$$

the real linear space $\mathcal{G}'$ becomes a real Hopf algebra. Furthermore, let $\mathcal{G}''$ denote the set of all the algebra morphisms $m : \mathcal{G}' \to \mathbb{R}$. The set $\mathcal{G}''$ becomes a group with respect to the convolution product

[24] This means that G is a separated topological space, and the map $(g, h) \mapsto gh^{-1}$ is a continuous map from $G \times G$ to G.

[25] L. Pontryagin, The theory of topological commutative groups, Ann. Math. **35** (1934), 361–388. See L. Pontryagin, Topological Groups, Gordon and Breach, 1966.

$$m * m' := \mu \circ (m \otimes m') \circ \Delta.$$

Explicitly, the map $m * m'$ is the composition of the maps

$$\mathcal{G}' \xrightarrow{\Delta} \mathcal{G}' \otimes \mathcal{G}' \xrightarrow{m \otimes m'} \mathcal{G}' \otimes \mathcal{G}' \xrightarrow{\mu} \mathcal{G}',$$

where the multiplication map μ sends the tensor product $\varphi \otimes \psi$ of functions $\varphi, \psi \in \mathcal{G}'$ to the product $\varphi\psi$. Finally, for each $g \in G$, we define the function $\varrho_g : \mathcal{G}' \to \mathbb{R}$ by setting

$$\boxed{\varrho_g(\chi) := \chi(g) \qquad \text{for all} \quad \chi \in \mathcal{G}'.}$$

Then, $\varrho_g \in \mathcal{G}''$. Tannaka and Krein created a duality theory which culminates in the following statement:[26]

The map $g \mapsto \varrho_g$ is a group isomorphism from G onto $\mathcal{G}''$.

3.6 Regularization and Rota–Baxter Algebras

It was noted long ago by Tricomi, and later independently by Cotlar, that the Hilbert transform

$$(Hf)(x) := \lim_{\varepsilon \to +0} \int_{|\xi| \geq \varepsilon} \frac{f(\xi)}{\xi - x}\, d\xi, \qquad x \in \mathbb{R},$$

operating on a suitable function algebra, satisfies the identity

$$(Hf)^2 = f^2 + 2H(fHf).$$

Later on, Glen Baxter was the first to point out that the evaluation of several functionals of sums of independent random variables depended on a purely algebraic study of a closely related identity,

$$Pf \cdot Pg = P(Pf \cdot g + f \cdot Pg - wfg).$$

Here, the fixed real parameter w is called the weight of the Rota–Baxter operator P.[27] The very same identity reappeared in the same guise in various estimates involving the iteration of the maximum function

$$x \mapsto \max(x, 0),$$

[26] T. Tannaka, On the duality theorem for noncommutative topological groups (in German), Tohoku math. **45** (1939), 1–12.
M. Krein, A principle of duality for compact groups and quadratic block algebras, Dokl. Akad. Nauk (N.S.) **69** (1949), 725–728 (in Russian).

[27] Idempotent Baxter operators (i. e., $P^2 = P$) have been known for a long time, often under the name of *Wiener–Hopf operators*. We refer to:
N. Wiener and E, Hopf, On a class of singular integral equations, Sitzber. Deutsch. Akad. Wiss Berlin, Kl. Math.-Phys.-Techn. (1931), pp. 696–706 (classic paper)(in German).
M. Krein, Integral equations on half line with kernel depending upon the difference of the arguments, Amer. Math. Soc. Transl. **22**(2) (1962), pp. 163–288.
I. Gochberg, One-Dimensional Singular Integral Equations, Vols. 1, 2, Birkhäuser, Basel, 1992.

> such as occur in the theory of almost-everywhere convergence... By looking at the problem in the rarified atmosphere of universal algebra, we were led to a systematic method for guessing and verifying identities for Baxter operators, based upon reducing all computations to identities between symmetric functions.[28]
>
> Gian-Carlo Rota and David Smith, 1972

It is our goal to generalize the classical operations of

- differentiation and integration of smooth functions, and
- regularization of singular functions

by using deformations. This will lead us to the notion of Rota–Baxter operators with weights. Hopf algebras and Rota–Baxter algebras play a crucial role in modern renormalization theory, as we will study later on.

Let $\mathcal{A}$ be a complex algebra. Consider an operator $P : \mathcal{A} \to \mathcal{A}$. One frequently encounters the following special cases.

(i) Linear map: The operator P is called linear iff we have

$$P(\alpha f + \beta g) = \alpha Pf + \beta Pg$$

for all $f, g \in \mathcal{A}$ and all complex numbers α, β.

(ii) Antilinear map: The operator P is called antilinear iff

$$P(\alpha f + \beta g) = \alpha^\dagger Pf + \beta^\dagger Pg$$

for all $f, g \in \mathcal{A}$ and all complex numbers α, β.

(iii) Multiplicative map: The operator P is called multiplicative iff for all $f, g \in \mathcal{A}$, we have the product property

$$P(fg) = Pf \cdot Pg.$$

The operator P is called an endomorphism of the algebra $\mathcal{A}$ iff it is linear and multiplicative. Bijective endomorphisms are called automorphisms.

(iv) Anti-multiplicative map: The operator P is called anti-multiplicative iff for all $f, g \in \mathcal{A}$,

$$P(fg) = Pg \cdot Pf.$$

The operator P is called an anti-endomorphism of the algebra $\mathcal{A}$ iff it is antilinear and anti-multiplicative. Bijective anti-endomorphisms are called anti-automorphisms.

(v) Derivation: The operator P is called a derivation iff it is linear, and for all $f, g \in \mathcal{A}$, we have the following Leibniz (product) rule:

$$P(fg) = Pf \cdot g + f \cdot Pg. \tag{3.49}$$

Derivations are also called infinitesimal endomorphisms (or generalized differential operators).

[28] G. Rota and D. Smith, Fluctuation theory and Baxter algebras, Istituto Nazionale di Alta Mathematica, Symposia matematica, Volume IX (1972), 179–201. This article is reprinted in J. Kung (Ed.), Gian-Carlo Rota on Combinatorics: Introductory Papers and Commentaries, pp. 481–201. Birkhäuser, Basel, 1995 (reprinted with permission).
G. Baxter, Combinatorial methods in fluctuation theory, Z. Wahrscheinlichkeitstheorie (Journal of Probability Theory) **1** (1963), 263–270.

(vi) Inverse derivation: The operator P is called an inverse derivation iff it is linear and for all $f, g \in \mathcal{A}$, we have the following rule:

$$Pf \cdot Pg = P(Pf \cdot g + f \cdot Pg). \tag{3.50}$$

As we will show below, this rule generalizes integration by parts. Inverse derivations are also called generalized integral operators.

(vii) Truncation operator: The operator P is called a truncation iff it is linear with the projector property $P^2 = P$ and with the truncation property, that is, for all $f, g \in \mathcal{A}$ we have

$$Pf \cdot Pg = P(Pf \cdot g + f \cdot Pg - fg). \tag{3.51}$$

The operator $R := I - P$ is called a regularization operator.

(viii) Rota–Baxter operator: Fix the real number w. The operator P is called a Rota–Baxter operator of weight w iff it is linear and for all $f, g \in \mathcal{A}$, we have the relation

$$\boxed{Pf \cdot Pg = P(Pf \cdot g + f \cdot Pg - wfg).} \tag{3.52}$$

Obviously, the inverse derivation (vi) (resp. the truncation operator (vii)) is a Rota–Baxter operator of weight $w = 0$ (resp. $w = 1$). If P is a Rota–Baxter operator of nonzero weight w, then $w^{-1}P$ (resp. $-w^{-1}P$) is a Rota–Baxter operator of weight 1 (resp. -1).

(ix) Rota–Baxter algebra: By definition, a Rota–Baxter algebra of weight w is a complex commutative unital algebra $\mathcal{A}$ equipped with a fixed Rota–Baxter operator of weight w.

Differentiation and integration. Set $\mathcal{A} := \mathcal{E}(\mathbb{R})$, that is, the complex algebra $\mathcal{A}$ consists of all the smooth functions $f : \mathbb{R} \to \mathbb{C}$.

- Fix $n = 1, 2, \ldots$ The operator

$$Pf := f^n \qquad \text{for all} \quad f \in \mathcal{A}$$

is multiplicative. This operator is linear iff $n = 1$. In this special case, the operator P is the trivial automorphism of the algebra $\mathcal{A}$.

- The operator $Pf := f^\dagger$ is anti-linear, anti-multiplicative, and bijective. Hence it is an anti-automorphism of $\mathcal{A}$. (Since $f^\dagger g^\dagger = g^\dagger f^\dagger$, the operator A is also multiplicative.)
- For fixed real number q, set

$$(Pf)(x) := f(qx) \qquad \text{for all} \quad x \in \mathbb{R}.$$

Then the operator $P : \mathcal{A} \to \mathcal{A}$ is an endomorphism.

- The operator $Pf := \frac{df}{dx}$ is a derivation of $\mathcal{A}$. In fact, the Leibniz rule of differentiation tells us that

$$\frac{d(fg)}{dx} = \frac{df}{dx} \cdot g + f \cdot \frac{dg}{dx}.$$

This is precisely relation (3.49).

- Define the integral operator $(Pf)(x) := \int_0^x f(u)du$ for all $f \in \mathcal{A}$. Choosing arbitrary functions $f, g \in \mathcal{A}$, we obtain

$$P(Pf \cdot g + f \cdot Pg) = Pf \cdot Pg.$$

This means that the operator $P : \mathcal{A} \to \mathcal{A}$ is a Rota–Baxter operator of weight $w = 0$. In fact, setting $F(x) := \int_0^x f(u)du$ and $G(x) := \int_0^x g(u)du$, and noting that $G(0) = F(0) = 0$, integration by parts yields

$$\int_0^x \big(F(u)g(u) + f(u)G(u)\big)\, du = \int_0^x \frac{d\{F(u)G(u)\}}{du}\, du = F(x)G(x).$$

3.6.1 Regularization of the Laurent Series

By a formal Laurent series with finite principal part, we understand a symbol of the form

$$f(z) = \frac{a_{-n}}{z^n} + \frac{a_{-n+1}}{z^{n-1}} + \ldots + \frac{a_{-1}}{z} + a_0 + a_1 z + a_2 z^2 + \ldots$$

where $a_{-n}, a_{-n+1}, \ldots$ are complex numbers and $n = 1, 2, \ldots$ The truncation operator $P : \mathcal{A} \to \mathcal{A}$ is defined by

$$\boxed{(Pf)(z) := \frac{a_{-n}}{z^n} + \frac{a_{-n+1}}{z^{n-1}} + \ldots + \frac{a_{-1}}{z}.}$$

That is, the truncation operator P assigns the principal part to the Laurent series. The operator $R := I - P$ is called the regularization operator. Explicitly,

$$(Rf)(z) := a_0 + a_1 z + a_2 z^2 + \ldots$$

That is, the regularization operator assigns the regular part to the Laurent series. The set of formal Laurent series with finite principal part forms a complex commutative unital algebra with respect to the usual linear combination and multiplication. For example,

$$2\left(\frac{1}{z} + z\right) + 3\left(\frac{2}{z^2} + z^2\right) = \frac{2}{z} + \frac{6}{z^2} + 2z + 3z^2,$$

and

$$\left(\frac{2}{z} + z\right)\left(\frac{3}{z^2} + z^2\right) = \frac{6}{z^3} + \frac{3}{z} + 2z + z^3.$$

Obviously, the truncation operator is linear and a projection operator, that is, $P^2 = P$.

Proposition 3.10 *The truncation operator $P : \mathcal{A} \to \mathcal{A}$ is a Rota–Baxter operator of weight $w = 1$.*

Consequently, the algebra of formal Laurent series with finite principal part is a Rota–Baxter algebra of weight $w = 1$ (with respect to the truncation operator P).
Proof. Let $f, g \in \mathcal{A}$. Since $P + R = I$, we have

$$\begin{aligned} Pf \cdot g + (f \cdot Pg - fg) &= Pf \cdot (Pg + Rg) - fRg \\ &= Pf \cdot (Pg + Rg) - (Pf + Rf)Rg = Pf \cdot Pg - Rf \cdot Rg. \end{aligned}$$

Summarizing, $Pf \cdot g + f \cdot Pg - fg = Pf \cdot Pg - Rf \cdot Rg$. Applying the operator P to this equation and noting that $Rg \cdot Rf$ is a power series expansion and hence $P(Rg \cdot Rf) = 0$, we get the desired relation

$$P(Pf \cdot g + fPg - fg) = Pf \cdot Pg.$$

□

3.6.2 Projection Operators

The preceding result can immediately be generalized to projection operators onto subalgebras of commutative algebras. This explains the importance of Rota–Baxter operators for algebraic structures. Let $\mathcal{A}$ be a commutative (real or complex) algebra. Suppose that we have the direct sum decomposition

$$\mathcal{A} := \mathcal{A}_+ \oplus \mathcal{A}_-$$

where $\mathcal{A}_+$ and $\mathcal{A}_-$ are subalgebras of $\mathcal{A}$. This means that for each $f \in \mathcal{A}$ we have the unique sum representation

$$f = f_+ + f_-, \qquad f_+ \in \mathcal{A}_+, \;\; f_- \in \mathcal{A}_-.$$

Now we set $Pf := f_-$ and $Rf := f_+$.

Proposition 3.11 *The projection operators $P : \mathcal{A} \to \mathcal{A}_-$ and $R : \mathcal{A} \to \mathcal{A}_+$ are Rota–Baxter operators of weight $w = 1$.*

The proof proceeds analogously to the proof of Prop. 3.10.

3.6.3 The q-Integral

Let $\mathcal{B}$ be the complex algebra of all polynomials $f : \mathbb{R} \to \mathbb{C}$ with $f(0) = 0$. Fix the parameter $q \in]-1, 1[$. The Jackson integral (or q-integral) is defined by

$$(q)\int_0^x f(x)dx := f(qx) + f(q^2x) + f(q^3x) + \ldots, \qquad x \in \mathbb{R}.$$

This series is always convergent. In particular, for $k = 1, 2, \ldots$, we get

$$(q)\int_0^x u^k du = \frac{x^k q^k}{1 - q^k}, \qquad x \in \mathbb{R}.$$

Let $n = 1, 2, \ldots$ By linearity,

$$\boxed{(q)\int_0^x \sum_{k=1}^n a_k u^k du = \sum_{k=1}^n a_k \frac{x^k q^k}{1 - q^k}.}$$

This definition looks strange. However, the point is that, after setting

$$(Pf)(x) := (q)\int_0^x f(u)du, \qquad u \in \mathbb{R}, \quad f \in \mathcal{B},$$

this notion of integral enjoys a formula of integration by parts, namely,

$$\boxed{P(Pf \cdot g + f \cdot Pg - wfg) = Pf \cdot Pg \qquad \text{for all} \quad f, g \in \mathcal{B}} \tag{3.53}$$

with $w = -1$. This is a deformation of the integration by parts formula

$$P(Pf \cdot g + f \cdot Pg) = Pf \cdot Pg$$

for classical integrals $(Pf)(x) := \int_0^x f(u)du$.

Proposition 3.12 *The q-integration operator $P : \mathcal{B} \to \mathcal{B}$ is a Rota–Baxter operator of weight $w = -1$.*

Proof. We have to show that (3.53) is valid. To see this, set $(Qf)(x) := f(qx)$ for all $x \in \mathbb{R}$. Then

$$Pf = Qf + Q^2 f + Q^3 f + \dots$$

Obviously, we have $Q(f \cdot g) = Qf \cdot Qg$. Hence $Q^2(f \cdot g) = Q^2 f \cdot Q^2 g$, and so on. Let $\varphi, \psi \in \mathcal{B}$. Using the commutativity property $\varphi\psi = \psi\varphi$, we obtain the following identity:

$$\begin{aligned} &Q\big((I-Q)\psi \cdot Q\varphi\big) + Q\big((I-Q)\varphi \cdot Q\psi\big) \\ &\quad = Q(\psi Q\varphi) - Q^2(\psi\varphi) + Q(\varphi Q\psi) - Q^2(\varphi\psi) \\ &\quad = Q(\varphi\psi) - Q^2(\varphi\psi) + Q\big(\psi Q\varphi + \varphi Q\psi - \varphi\psi - Q(\varphi\psi)\big) \\ &\quad = (I-Q)Q(\varphi\psi) - Q\big((I-Q)\varphi \cdot (I-Q)\psi\big). \end{aligned} \tag{3.54}$$

Setting $Sf := f + Pf = f + Qf + Q^2 f + \dots$, we get

$$S(I-Q)f = (I-Q)Sf = f \qquad \text{for all} \quad f \in \mathcal{B}.$$

Hence $S = (I-Q)^{-1}$ and $P = QS = SQ$. Multiplying the first and last members of (3.54) by S, we obtain

$$\begin{aligned} P\big((I-Q)\psi \cdot Q\varphi\big) + P\big((I-Q)\varphi \cdot Q\psi\big) &= Q\varphi \cdot Q\psi \\ &\quad - P\big((I-Q)\varphi \cdot (I-Q)\psi\big). \end{aligned} \tag{3.55}$$

Finally, set $f := (I-Q)\varphi$ and $g := (I-Q)\psi$. Then

$$Pf = QS(I-Q)\varphi = Q\varphi, \qquad Pg = Q\psi.$$

By (3.55), we get $P(gPf) + P(fPg) = Pf \cdot Pg - P(fg)$. This is the claim (3.53). □

Hints for further reading. The q-deformation of special polynomials and the q-integral were introduced by:

F. Jackson, On q-functions and a certain difference operator, Trans. Roy. Soc. Edinburgh **46** (1908), 253–281.

F. Jackson, q-Integration, Proc. Durham Phil. Soc. **7** (1927), 182–189.

In the last twenty years, physicists used this approach in order to study deformations of classical structures in order to get new approaches to the quantization of space and time. We refer to:

M. Majid, Foundations of Quantum Group Theory, Cambridge University Press, Cambridge, 1995.

M. Majid, A Quantum Groups Primer, Cambridge University Press, 2002.

J. Wess, Gauge theories on noncommutative space-time treated by the Seiberg–Witten method. In: U. Carow-Watamura et al. (Eds.) (2005), 179–192.

T. Kornwinder, Special functions and q-commuting variables, 1996. Internet: `http://www.q-alg/9608008`

H. Wachter, q-Integration on quantum spaces, Eur. Phys. J. C **32** (2004), 281–297. Internet: `http://www.hep-th/0206083`

A. Schmidt and H. Wachter, q-deformed quantum Lie algebras, J. Geometry and Physics, **56** (2006), 2289–2325.

H. Wachter, Towards a q-deformed quantum field theory, 24 pages. In: B. Fauser, J. Tolksdorf, and E. Zeidler (Eds.), Quantum Field Theory – Competitive Methods, Birkhäuser, Basel, 2008.

A. Schmidt, Towards a q-deformed supersymmetric field theory, 18 pages. In: B. Fauser, J. Tolksdorf, and E. Zeidler (Eds.) (2008) quoted above.

3.6.4 The Volterra–Spitzer Exponential Formula

The Volterra exponential formula. Consider the algebra $\mathcal{E}(\mathbb{R})$ of smooth functions $f : \mathbb{R} \to \mathbb{C}$ on the real line. Let $P : \mathcal{E}(\mathbb{R}) \to \mathcal{E}(\mathbb{R})$ be the integral operator

$$(Pf)(t) := \int_0^t f(u)du, \qquad t \in \mathbb{R},$$

and choose the fixed function $a \in \mathcal{E}(\mathbb{R})$. We will show below that, for all $t \in \mathbb{R}$, we have the Volterra exponential formula

$$\exp\left(\int_0^t a(u)du\right) = 1 + \int_0^t a(u)du + \int_0^t dv\, a(v) \int_0^v a(u)du + \int_0^t dw\, a(w) \int_0^w dv\, a(v) \int_0^v a(u)du + \ldots \tag{3.56}$$

Mnemonically, the right-hand side can be written as

$$1 + Pa + Pa \circ Pa + Pa \circ Pa \circ Pa + \ldots := 1 + Pa + P(aPa) + P(aP(aPa)) + \ldots$$

Obviously, for all $t \in \mathbb{R}$, there holds the simpler formula

$$\exp\left(\int_0^t a(u)du\right) = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\int_0^t a(u)du\right)^k. \tag{3.57}$$

The two formulas (3.56) and (3.57) are trivial special cases of the Dyson series which plays a fundamental role in quantum field theory (see Sect. 7.17.4 of Vol. I).

Proof. Let us prove (3.56). We start with the initial-value problem

$$\dot{f}(t) = a(t)f(t), \quad t \in \mathbb{R}, \quad f(0) = 1. \tag{3.58}$$

This is equivalent to the Volterra integral equation

$$f(t) = 1 + \int_0^t a(u)f(u)du, \qquad t \in \mathbb{R}. \tag{3.59}$$

In turn, this is equivalent to the operator equation

$$f = 1 + P(af). \tag{3.60}$$

The initial-value problem (3.56) has the unique solution $f(t) = \mathrm{e}^{\int_0^t a(u)du}$. In fact, differentiation yields $\dot{f}(t) = a(t)f(t)$. Furthermore, if

$$f = 1 + Pa + Pa \circ Pa + Pa \circ Pa \circ Pa + \ldots,$$

then $1 + P(af) = 1 + Pa \circ f = f$. Thus, f is at least a formal solution of (3.60). However, the argument from Sect. 7.17.3 of Vol. I shows that the infinite series $f = 1 + Pa + Pa \circ Pa + \ldots$ is uniformly convergent on each compact time interval. This proves (3.56). □

The Spitzer exponential formula. It is our goal to solve equation (3.60) in the case of Rota–Baxter algebras and to generalize the Volterra exponential formula (3.56) to the following Spitzer exponential formula:[29]

$$\boxed{\mathrm{e}^{P\left(\ln(1-az)^{-1}\right)} = \mathbf{1} + zPa + z^2 Pa \circ Pa + z^3 Pa \circ Pa \circ Pa + \ldots} \tag{3.61}$$

Let $\mathcal{A}$ be a Rota–Baxter algebra of weight $w = -1$ with respect to the Rota–Baxter operator $P : \mathcal{A} \to \mathcal{A}$, and let a be a fixed element of $\mathcal{A}$. Then the equation

$$f = \mathbf{1} + zP(af), \qquad f \in \mathcal{A} \tag{3.62}$$

has a solution given by (3.61). The solution of (3.62) and the Spitzer identity (3.61) are to be understood in the sense of formal power series expansions with respect to the variable z. If P is a Rota–Baxter operator of weight $w = 1$, then the statements above remain valid if we replace P by $-P$ and a by $-a$ in the Spitzer formula (3.61). The non-trivial proof can be found in:

> F. Spitzer, A combinatorial lemma and its application to probability theory, Trans. Amer. Math. Soc. **82** (1965), 323–339.

See also J. Kung (Ed.), Gian-Carlo Rota on Combinatorics, p. 490, Birkhäuser, Basel, 1995. A non-commutative version of Spitzer's identity can be found in

> K. Ebrahimi-Fard, Li Guo, and D. Kreimer, Spitzer's identity and the algebraic Birkhoff decomposition in perturbative quantum field theory, J. Phys. **A37** (2004), 11037–11052.

3.6.5 The Importance of the Exponential Function in Mathematics and Physics

> The exponential function is the most important function in mathematics.
> Folklore

The following exponential and logarithmic formulas play a crucial role:

- The Dyson series via time-ordering operator (see Sect. 7.17.4 of Vol. I).
- The Trotter exponential formula (see Sect. 8.3 of Vol. I).
- The Baker–Campbell–Hausdorff exponential formula for Lie algebras (see Sect. 8.4 of Vol. I).
- The Faà di Bruno exponential formula for Bell polynomials (see (3.32) on page 136).
- The logarithmic formula for Schur polynomials (see (3.39) on page 140).
- The logarithmic formula for reduced correlation functions (or connected Green's functions) (see (3.41) on page 142).
- The logarithmic formula for cumulants (see (3.44) on page 144).
- Group characters (see Sect. 3.5.4 on page 152).
- The Volterra–Spitzer formula (see (3.57) on page 160).
- The Spitzer formula for Rota–Baxter algebras (see (3.61) on page 161).
- The perturbation formula (see (3.70) on page 167).

[29] $P\left(\ln(1-az)^{-1}\right)$ stands for the formal power series $-\sum_{k=1}^{\infty} \frac{(-z)^k}{k!} P(a^k)$.

3.7 Partially Ordered Sets and Combinatorics

> Modern combinatorics was strongly influenced by the work of Gian-Carlo Rota (1933–1989) at the MIT (Massachusetts Institute of Technology) beginning in 1967.
>
> Folklore

3.7.1 Incidence Algebras and the Zeta Function

Let $\mathcal{P}$ be a partially ordered set. We define the closed interval

$$[x, z] := \{y \in \mathcal{P} : x \leq y \leq z\},$$

and the half-open interval

$$[x, z[:= \{y \in \mathcal{P} : x \leq y < z\}.$$

The set $\mathcal{P}$ is called locally finite iff each closed interval only contains a finite number of elements. The function

$$\zeta(x, y) := \begin{cases} 1 & \text{if} \quad x \leq y, \\ 0 & \text{otherwise} \end{cases}$$

is called the zeta function of $\mathcal{P}$, and the function

$$\delta(x, y) := \begin{cases} 1 & \text{if} \quad x = y, \\ 0 & \text{otherwise} \end{cases}$$

is called the Kronecker function of $\mathcal{P}$. In terms of physics, we may regard the elements $x, y, \ldots$ of $\mathcal{P}$ as events. The relation $x \leq y$ means that the event y is caused by the event x.

Let $\mathcal{P}$ be a locally finite partially ordered set. By definition, the set $\mathcal{A}(\mathcal{P})$ consists of all the functions

$$f : \mathcal{P} \times \mathcal{P} \to \mathbb{R}$$

with $f(x, y) = 0$ if $x \not\leq y$. For $f, g \in \mathcal{A}(\mathcal{P})$, we define the convolution

$$(f * g)(x, z) := \sum_{x \leq y \leq z} f(x, y) g(y, z).$$

Obviously, $f * \delta = \delta * f = f$.

> *The set $\mathcal{A}(\mathcal{P})$ is a real associative unital algebra with the convolution as product and the Kronecker function as unit element.*

The algebra $\mathcal{A}(\mathcal{P})$ is called the incidence algebra of $\mathcal{P}$. We have to prove the associative law for the convolution. In fact, $((f * g) * h)(x, z)$ is equal to

$$\sum_{x \leq y \leq z} (f * g)(x, y) h(y, z) = \sum_{x \leq y \leq z} \sum_{x \leq u \leq y} f(x, u) g(u, y) h(y, z)$$

$$= \sum_{x \leq u \leq z} \sum_{u \leq y \leq z} f(x, u) g(u, y) h(y, z) = (f * (g * h))(x, z).$$

3.7.2 The Möbius Function as an Inverse Function

Proposition 3.13 *There exists precisely one function $\mu \in \mathcal{A}(\mathcal{P})$ which is a two-sided inverse of the zeta function, that is, $\mu * \zeta = \zeta * \mu = \delta$.*

Proof. (I) Uniqueness. If $\mu' * \zeta = \zeta * \mu' = \delta$, then the associative law states $\mu = \mu * \delta = \mu * (\zeta * \mu') = (\mu * \zeta) * \mu' = \delta * \mu' = \mu'$.

(II) Existence. If $x \not\leq z$, then let $\mu(x, z) := 0$. If $x \leq z$, then we set $\mu(x, x) := 1$, and we use

$$\mu(x, z) := -\sum_{x \leq y < z} \mu(x, y) \tag{3.63}$$

as the basis for an inductive definition over the number of elements contained in the interval $[x, z]$. For example, if $[x, z]$ contains precisely two elements, then

$$\mu(x, z) = -\mu(x, x) = -1.$$

If $[x, z]$ contains precisely the three elements x, y, z, then $x < y < z$. Hence

$$\mu(x, z) = -\mu(x, x) - \mu(x, y) = -1 + 1 = 0.$$

Finally, let us prove the convolution formula

$$\sum_{x \leq y \leq z} \zeta(x, y)\mu(y, z) = \delta(x, z).$$

In fact, if $z = x$, then $1 = \zeta(x, x)\mu(x) = \delta(x, x)$. If $x < z$, then the definition of the Möbius function μ states

$$\sum_{x \leq y < z} \zeta(x, y)\mu(y, z) + \mu(x, z) = 0.$$

This is the claim, since $\zeta(x, z) = 1$ and $\delta(x, z) = 0$. □

The Möbius inversion formula. The following inversion formula represents the main property of the Möbius function.

Proposition 3.14 *Let $f : \mathcal{P} \to \mathbb{R}$ be a function on the locally finite partially ordered set $\mathcal{P}$. Suppose that there exists an element x_0 in $\mathcal{P}$ such that $f(x) = 0$ if $x_0 \not\leq x$. Define $F(x) := \sum_{y \leq x} f(y)$. Then*

$$f(x) = \sum_{y \leq x} F(y)\mu(y, x) \qquad \text{for all} \quad x \in \mathcal{P}. \tag{3.64}$$

Proof. Since $F(x) = \sum_{x_0 \leq y \leq x} f(y)$, the sum $F(x)$ is finite. By substitution,

$$\sum_{y \leq x} F(y)\mu(y, x) = \sum_{y \leq x} \sum_{z \leq y} f(z)\mu(y, x).$$

This is equal to $\sum_{y \leq x} \sum_{z} f(z)\zeta(z, y)\mu(y, x)$. Interchanging the order of summation, Prop. 3.13 tells us that this is equal to

$$\begin{aligned} \sum_{z} f(z) \sum_{y \leq x} \zeta(z, y)\mu(y, x) &= \sum_{z} f(z) \sum_{z \leq y \leq x} \zeta(z, y)\mu(y, x) \\ &= \sum_{z} f(z)\delta(z, x) = f(x). \end{aligned}$$

This identity proves the validity of the inversion formula (3.64). □

The discrete version of the fundamental theorem of calculus. Consider the continuous function $f : \mathbb{R} \to \mathbb{R}$. Set

$$F(x) := \int_0^x f(y)dy \qquad \text{for all} \quad x \in \mathbb{R}.$$

Then the fundamental theorem of calculus tells us that

$$f(x) = F'(x) \qquad \text{for all} \quad x \in \mathbb{R}.$$

For an arbitrary function $f : \{1, 2, \ldots\} \to \mathbb{R}$, the discrete version reads as

$$F(n) := \sum_{1 \le m \le n} f(m)$$

and

$$\boxed{f(n) = F(n) - F(n-1), \qquad n = 2, 3, \ldots, \quad f(1) = F(1).} \tag{3.65}$$

The proof is obvious. In order to see that this is a special case of the Möbius inversion formula, consider the set $\mathcal{P}$ of positive integers. This is a partially ordered set equipped with the usual $\le$-relation. The corresponding Möbius function is defined by $\mu(m, n) := 0$ if $m > n$. Moreover, $\mu(m, m) := 1$. For $n = m + 1, m + 2, \ldots$, we successively have

$$\mu(m, n) := -\sum_{m \le k < n} \mu(m, k).$$

Hence $\mu(m, m + 1) = -\mu(m, m) = -1$. This implies

$$\mu(m, m + 2) = -\mu(m, m) - \mu(m, m + 1) = -1 + 1 = 0.$$

Furthermore, $\mu(m, m+3) = -\mu(m, m)-\mu(m, m+1)-\mu(m, m+2) = -1+1+0 = 0$. Similarly, we get $\mu(m, m) := 1$, $\mu(m, m + 1) = -1$, and otherwise $\mu(m, n) = 0$. For the zeta function of $\mathcal{P}$, we get $\zeta(m, n) = 1$ if $m \le n$ and otherwise $\zeta(m, n) = 0$. Therefore,

$$F(n) = \sum_{1 \le m \le n} f(m)\zeta(m, n).$$

By Prop. 3.14, $f(n) = \sum_{1 \le m \le n} F(m)\mu(m, n)$. This coincides with (3.65).

In what follows, we will apply the Möbius inversion formula to the inclusion-exclusion principle in combinatorics, to the classical 1832 Möbius inversion formula in number theory, and to the Riemann zeta function.

3.7.3 The Inclusion–Exclusion Principle in Combinatorics

Let S_1 and S_2 be finite sets, and let $|S_j|$ denote the cardinality of the set S_j (i.e., the number of elements). Then

$$|S_1 \cup S_2| = |S_1| + |S_2| - |S_1 \cap S_2|.$$

This is the simplest case of the so-called inclusion-exclusion principle. In general, this principle states that the cardinality of the union of a finite family of finite sets can be computed as an alternating sum of cardinalities of intersections of those sets. Explicitly, let $n = 2, 3, \ldots$ If $S_1, \ldots S_n$ are finite sets, then

$$|S_1 \cup \cdots \cup S_n| = \sum_{1 \le i \le n} |S_i| - \sum_{1 \le i < j \le n} |S_1 \cap S_j| + \sum_{1 \le i < j < k \le n} |S_i \cap S_j \cap S_k|$$
$$- \ldots + (-1)^{n+1} \sum_{1 \le i_1 < \ldots < i_n \le n} |S_{i_1} \cap \cdots \cap S_{i_n}|. \tag{3.66}$$

Proof. We will use the Möbius inversion formula. To this end, let $\mathcal{P}$ be the collection of all the finite subsets of the set

$$\mathcal{E} := \{S_1, S_2, \ldots, S_n\},$$

including the empty subset. For example, $\emptyset$, $\{S_1\}$, $\{S_1, S_2\}, \ldots, \{S_1, \ldots, S_n\}$ are elements of the set $\mathcal{P}$, which becomes a partially ordered set with respect to the usual $\subseteq$-relation. For $\mathcal{A}, \mathcal{B} \in \mathcal{P}$, the Möbius function of $\mathcal{P}$ reads as

$$\mu(\mathcal{A}, \mathcal{B}) = (-1)^{|\mathcal{B} \setminus \mathcal{A}|}$$

if $\mathcal{A} \subseteq \mathcal{B}$. Otherwise $\mu(\mathcal{A}, \mathcal{B}) = 0$. Set $S := S_1 \cup \cdots \cup S_n$. For each element $\mathcal{A}$ of $\mathcal{P}$ different from $\emptyset$, define $f(\mathcal{A})$ as the cardinality of the set

$$\{x \in S : x \in S_j \text{ iff } S_j \in \mathcal{A}\}.$$

Furthermore, set $f(\emptyset) := 0$. For example, if $\mathcal{A} = \{S_1\}$, then

$$f(\mathcal{A}) = |S_1 \setminus \{S_2 \cup \cdots \cup S_n\}|.$$

Now introduce the function

$$F(\mathcal{A}) := \sum_{\mathcal{A} \subseteq \mathcal{B}} f(\mathcal{B}).$$

For example, if $\mathcal{E} := \{S_1, S_2\}$ and $\mathcal{A} := \emptyset$, then

$$F(\emptyset) = 0 + |S_1 \setminus S_2| + |S_2 \setminus S_1| + |S_1 \cap S_2| = |S_1 \cup S_2|.$$

The summands correspond to $\mathcal{B} = \emptyset, \{S_1\}, \{S_2\}, \{S_1, S_2\}$. If $\mathcal{A}$ is equal to $\{S_1\}$, then

$$F(\mathcal{A}) = |S_1 \setminus S_2| + |S_1 \cap S_2| = |S_1|.$$

If $\mathcal{A} = \{S_1, S_2\}$, then $F(\mathcal{A}) = |S_1 \cup S_2|$. In the general case where $n = 1, 2, \ldots$, we get $F(\emptyset) = |S|$. Moreover, if $\mathcal{A}$ is different from $\emptyset$, then

$$F(\mathcal{A}) = \bigcap_{S_j \in \mathcal{A}} S_j.$$

By Möbius inversion, $f(\mathcal{A}) = \sum_{\mathcal{A} \subseteq \mathcal{B}} F(\mathcal{B}) \mu(\mathcal{A}, \mathcal{B})$. Using $\mathcal{A} = \emptyset$, we obtain

$$f(\emptyset) = F(\emptyset) + \sum_{\mathcal{B} \neq \emptyset} (-1)^{|\mathcal{B}|} F(\mathcal{B}).$$

Since $f(\emptyset) = 0$, $F(\emptyset) = |S|$ and $F(\mathcal{B}) = |\bigcap_{S_j \in \mathcal{B}} S_j|$, this is the claim. □

3.7.4 Applications to Number Theory

The classical Möbius function. A positive integer n is called a prime number iff $n \geq 2$, and only the numbers 1 and n are divisors of n. Each positive integer $n \geq 2$ can be uniquely represented as

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}$$

where $p_1 < p_2 < \ldots < p_r$ are prime numbers and $\alpha_1, \alpha_2, \ldots, \alpha_r$ are positive integers. Let $n = 2, 3 \ldots$ By definition,

$$\mu(n) := (-1)^r \qquad \text{if } \alpha_1 = \alpha_2 = \ldots = \alpha_r = 1.$$

Otherwise $\mu(n) := 0$. Furthermore, $\mu(1) := 1$. For example, if p is a prime number, then $\mu(p) = -1$. If $n = pq$ is the product of two different prime numbers p and q, then $\mu(pq) = 1$. Moreover, if n is divisible by a square of a prime number, then $\mu(n) = 0$. For example $\mu(12) = 0$. Let $\mathcal{P} := \{1, 2, \ldots\}$. For $m, n \in \mathcal{P}$, we write

$$m \preceq n$$

iff m is a divisor of n. This way, $\mathcal{P}$ becomes a partially ordered set which is locally finite. For given function $f : \mathcal{P} \to \mathbb{R}$, we define

$$F(n) := \sum_{m \preceq n} f(m), \qquad n = 1, 2, \ldots$$

In 1832, Möbius (1790–1868) discovered the inversion formula

$$f(n) = \sum_{m \preceq n} F(m) \mu\left(\frac{n}{m}\right), \qquad n = 1, 2, \ldots \tag{3.67}$$

The relation to partially ordered sets. Formula (3.67) is a special case of the Möbius inversion formula (3.64) (due to Rota) on $\mathcal{P} = \{1, 2, \ldots\}$ with respect to $\preceq$. This follows from the formula

$$\mu(m, n) = \mu\left(\frac{n}{m}\right) \tag{3.68}$$

for all $m, n \in \mathcal{P}$ with $m \preceq n$. The proof of (3.68) will be given in Problem 3.15. For the zeta function on $\mathcal{P}$, we get $\zeta(m, n) := 1$ if $m \preceq n$. Otherwise $\zeta(m, n) = 0$.

The inverse of the Riemann zeta function. If s is a complex number with $\Re(s) > 1$, then the Riemann zeta function is given by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{\zeta(1, n)}{n^s}.$$

For the inverse function, we obtain

$$\frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\mu(1, n)}{n^s}.$$

The proof can be found in H. Edwards, Riemann's Zeta Function, Academic Press, New York, 1974.

3.8 Hints for Further Reading

As an introduction to combinatorics, we recommend the following monographs:

L. Comtet, Advanced Combinatorics, Reidel, Dordrecht, 1974.

A. Tucker, Applied Combinatorics, Wiley, New York, 1980.

M. Aigner, Combinatorial Theory, Springer, Berlin, 1997.

M. Petkovsek, H. Wilf, and D. Zeilberger, A=B, Peters, Wellesley, Massachusetts, 1996.
Internet: `http://www.cis.upenn.edu/ wilf/AequB.html`

R. Stanley, Enumerative Combinatorics, Cambridge University Press, 1997.

J. Gross and J. Yellen (Eds.), Handbook of Graph Theory, CRC Press, Boca Raton, Florida, 2004.

The contributions of Gian-Carlo Rota to modern combinatorics are collected in:

J. Kung (Ed.), Gian-Carlo Rota on Combinatorics: Introductory Papers and Commentaries, Birkhäuser, Basel, 1995.

A detailed discussion of the Kepler equation can be found in W. Neutsch and K. Scherer, Celestial Mechanics: An Introduction to Classical and Contemporary Methods, Wissenschaftsverlag, Mannheim, 1992. Finally, we recommend:

M. Sweedler, Hopf Algebras, Benjamin, New York, 1969.

P. Cartier, A primer of Hopf algebras. Preprint: l'Institut des Hautes Études Scientifiques (IHES), Bures-sur-Yvette (France), 2006, IHES/M/06/40. Internet: `http://www.cartier@ihes.fr`

Problems

3.1 *Bilinear forms.* Prove the identity (3.1) on page 120.
Solution: Let $b_1, \ldots, b_n$ be a basis of X. Since B is bilinear, we get

$$B\left(\sum_{k=1}^{n} \lambda_k b_k, \sum_{l=1}^{n} \mu_l b_l\right) = \sum_{k,l=1}^{n} \lambda_k \mu_l B(b_k, b_l).$$

It follows from the generalized orthogonality relation $b^r(b_s) = \delta^r_s$ that

$$B = \sum_{k,l=1}^{n} B(b_k, b_l) b^k \otimes b^l. \tag{3.69}$$

3.2 *Perturbation formula for the exponential function.* Let $A, B : X \to X$ be linear continuous operators on a complex Hilbert space X. Show that

$$\mathrm{e}^{z(A+B)} = \mathrm{e}^{zA} + \int_0^z \mathrm{e}^{(z-\zeta)A} B \mathrm{e}^{\zeta(A+B)} d\zeta \qquad \text{for all} \quad z \in \mathbb{C}. \tag{3.70}$$

Solution: Set $J(z) := \mathrm{e}^{-zA}\mathrm{e}^{z(A+B)}$. Differentiating the relation

$$\mathrm{e}^{z(A+B)} = \mathrm{e}^{zA} J(z)$$

with respect to the complex variable z, we get

$$(A+B)\mathrm{e}^{z(A+B)} = A\mathrm{e}^{zA}J(z) + \mathrm{e}^{zA}J'(z).$$

Hence $J'(z) = \mathrm{e}^{-zA}B\mathrm{e}^{zA}J(z)$. Since $J(0) = I$,

$$J(z) = I + \int_0^z \mathrm{e}^{-\zeta A}B\mathrm{e}^{\zeta A}J(\zeta)d\zeta.$$

Multiplying this by e^{zA}, we get the claim (3.70). Similarly, using the function $K(z) := \mathrm{e}^{z(A+B)}\mathrm{e}^{-zA}$, we get

$$\mathrm{e}^{z(A+B)} = \mathrm{e}^{zA} + \int_0^z \mathrm{e}^{\zeta(A+B)}B\mathrm{e}^{(z-\zeta)A}d\zeta \qquad \text{for all} \quad z \in \mathbb{C}. \tag{3.71}$$

3.3 *Special perturbation formula.* Let $t \in \mathbb{C}$. Use (3.70), in order to prove that

$$\frac{d\mathrm{e}^{A+tB}}{dt}|_{t=0} = \int_0^1 \mathrm{e}^{(1-\zeta)A}B\mathrm{e}^{\zeta A}d\zeta = \int_0^1 \mathrm{e}^{\zeta A}B\mathrm{e}^{(1-\zeta)A}d\zeta. \tag{3.72}$$

Solution: Set $z = 1$, and replace B by tB in (3.70).

3.4 *Hopf algebras and commutative diagrams.* Reformulate the defining relations (A), (U), (CA), (CU) of Hopf algebras on page 128 in the language of commutative diagrams.

Solution: We will use the following algebra morphisms

$$\mu : \mathcal{A}\otimes\mathcal{A} \to \mathcal{A}, \quad \Delta : \mathcal{A} \to \mathcal{A}\otimes\mathcal{A}, \quad \eta : \mathbb{C} \to \mathcal{A}, \quad \varepsilon : \mathcal{A} \to \mathbb{C}$$

and the linear map $S : \mathcal{A} \to \mathcal{A}$. Recall that $\mu(a\otimes b) = ab$ for all $a, b \in \mathcal{A}$. The commutativity of the diagrams

$$\begin{array}{ccc} \mathcal{A}\otimes\mathcal{A} & \xrightarrow{\mu} & \mathcal{A} \\ {\scriptstyle \mathrm{id}\otimes\mu}\uparrow & & \uparrow{\scriptstyle \mu} \\ \mathcal{A}\otimes\mathcal{A}\otimes\mathcal{A} & \xrightarrow{\mu\otimes\mathrm{id}} & \mathcal{A}\otimes\mathcal{A} \end{array} \tag{3.73}$$

and

$$\begin{array}{ccccc} \mathbb{C}\otimes\mathcal{A} & \xrightarrow{\eta\otimes\mathrm{id}} & \mathcal{A}\otimes\mathcal{A} & \xleftarrow{\mathrm{id}\otimes\eta} & \mathcal{A}\otimes\mathbb{C} \\ & {\scriptstyle \mathrm{id}}\searrow & \downarrow{\scriptstyle \mu} & \swarrow{\scriptstyle \mathrm{id}} & \\ & & \mathcal{A} & & \end{array} \tag{3.74}$$

is equivalent to (A) and (U) on page 128, respectively, that is,

$$\mu(\mu\otimes\mathrm{id}) = \mu(\mathrm{id}\otimes\mu) \quad \text{(associativity (A))}$$

and

$$\mu(\eta\otimes\mathrm{id}) = \mu(\mathrm{id}\otimes\eta) = \mathrm{id} \qquad \text{(unitality (U))},$$

respectively. Here, id is the identical map on $\mathcal{A}$, that is $\mathrm{id}(a) := a$ for all $a \in \mathcal{A}$. To simplify notation, we also denote the natural isomorphisms $i : \mathcal{A}\otimes\mathbb{C} \to \mathcal{A}$

and $j : \mathbb{C} \otimes \mathcal{A} \to \mathcal{A}$ (as well as their inverses) by the symbol id. In this sense, for example,

$$\mathrm{id}(a \otimes z) := za \qquad \text{for all} \quad a \in \mathcal{A}, z \in \mathbb{C}.$$

Let us now pass to the *dual concepts*. Reversing the arrows, we get the following two commutative diagrams

$$\begin{array}{ccc} \mathcal{A} \otimes \mathcal{A} & \xleftarrow{\Delta} & \mathcal{A} \\ {\scriptstyle \mathrm{id}\otimes\Delta}\downarrow & & \downarrow{\scriptstyle \Delta} \\ \mathcal{A} \otimes \mathcal{A} \otimes \mathcal{A} & \xleftarrow{\Delta\otimes\mathrm{id}} & \mathcal{A} \otimes \mathcal{A} \end{array} \tag{3.75}$$

and

$$\begin{array}{ccccc} \mathbb{C} \otimes \mathcal{A} & \xleftarrow{\varepsilon\otimes\mathrm{id}} & \mathcal{A} \otimes \mathcal{A} & \xrightarrow{\mathrm{id}\otimes\varepsilon} & \mathcal{A} \otimes \mathbb{C} \\ & {\scriptstyle \mathrm{id}}\nwarrow & \uparrow{\scriptstyle \Delta} & \nearrow{\scriptstyle \mathrm{id}} & \\ & & \mathcal{A} & & \end{array} \tag{3.76}$$

which tell us that

$$(\Delta \otimes \mathrm{id})\Delta = (\mathrm{id} \otimes \Delta)\Delta \qquad \text{(coassociativity (CA))}$$

and

$$(\varepsilon \otimes \mathrm{id})\Delta = (\mathrm{id} \otimes \varepsilon)\Delta = \mathrm{id} \qquad \text{(counitality (CU))},$$

respectively (see page 128). Finally, the commutative diagram

$$\begin{array}{ccccc} \mathcal{A} \otimes \mathcal{A} & \xleftarrow{\Delta} & \mathcal{A} & \xrightarrow{\Delta} & \mathcal{A} \otimes \mathcal{A} \\ {\scriptstyle \mathrm{id}\otimes S}\downarrow & & \uparrow{\scriptstyle \eta} & & \downarrow{\scriptstyle S\otimes\mathrm{id}} \\ & & \mathbb{C} & & \\ & & \uparrow{\scriptstyle \varepsilon} & & \\ \mathcal{A} \otimes \mathcal{A} & \xrightarrow{\mu} & \mathcal{A} & \xleftarrow{\mu} & \mathcal{A} \otimes \mathcal{A} \end{array} \tag{3.77}$$

is equivalent to the relation

$$\mu(\mathrm{id} \otimes S)\Delta = \mu(S \otimes \mathrm{id})\Delta = \eta\varepsilon \qquad \text{(coinverse)}.$$

3.5 *The coproduct as an algebra morphism.* Prove (3.10) on page 125. Solution: By the Sweedler notation, we get

$$(\Delta D')(\Delta D) = \left(\sum_{D'} D'_{(1)} \otimes D'_{(2)}\right)\left(\sum_{D} D_{(1)} \otimes D_{(2)}\right)$$

with the product $(D'_{(1)} \otimes D'_{(2)})(D_{(1)} \otimes D_{(2)}) = D'_{(1)}D_{(1)} \otimes D'_{(2)}D_{(2)}$ on $\mathcal{A} \otimes \mathcal{A}$. Applying the Leibniz rule to $\Delta(D'D)(f, g) = D'(D(fg))$, we get the claim $\Delta(D'D) = (\Delta D')(\Delta D)$.

3.6 *The coinverse of a differential operator.* Prove (3.13) on page 127.
Solution: Consider first $D := \partial_k\partial_j$. By (3.8) on page 125, the Sweedler notation $\Delta D = \sum_D D_{(1)} \otimes D_{(2)}$ corresponds to

$$\Delta(\partial_k\partial_j) = \partial_k\partial_j \otimes \mathbf{1} + \partial_j \otimes \partial_k + \partial_k \otimes \partial_j + \mathbf{1} \otimes \partial_k\partial_j.$$

Since $S(\mathbf{1}) = \mathbf{1}$, $S(\partial_j) = -\partial_j$, and $S(\partial_k\partial_j) = \partial_k\partial_j$, we get

$$\begin{aligned}\sum_D S(D_{(1)}) \otimes D_{(2)} &= S(\partial_k\partial_j) \otimes \mathbf{1} + S(\partial_j) \otimes \partial_k + S(\partial_k) \otimes \partial_j + S(\mathbf{1}) \otimes \partial_k\partial_j \\ &= \partial_k\partial_j \otimes \mathbf{1} - \partial_j \otimes \partial_k - \partial_k \otimes \partial_j + \mathbf{1} \otimes \partial_k\partial_j.\end{aligned}$$

Hence

$$\sum_D S(D_{(1)})D_{(2)} = \partial_k\partial_j - \partial_j\partial_k - \partial_k\partial_j + \partial_k\partial_j = 0.$$

Analogously, $\sum_D D_{(1)}S(D_{(2)}) = 0$. For $D := \partial_{k_1}\partial_{k_2}\cdots\partial_{k_m}$, proceed by induction.

3.7 *Properties of the convolution.* Prove Proposition 3.5 on page 130. Solution:
(I) Associativity. By definition of the convolution,

$$(B * C) * D = \mu((B * C) \otimes D)\Delta = \mu(\mu \otimes \mathrm{id})(B \otimes C \otimes D)(\Delta \otimes \mathrm{id})\Delta.$$

By coassociativity, $(\Delta \otimes \mathrm{id})\Delta = (\mathrm{id} \otimes \Delta)\Delta$. Finally, the associativity of the multiplication on $\mathcal{A}$ tells us that $(B * C) * D$ is equal to

$$\mu(\mathrm{id} \otimes \mu)(B \otimes C \otimes D))(\mathrm{id} \otimes \Delta)\Delta = \mu(B \otimes (C * D))\Delta = B * (C * D).$$

(II) Unit element. By definition of the convolution, $B * \eta\varepsilon = \mu(B \otimes \eta\varepsilon)\Delta$. This can be written as

$$B * \eta\varepsilon = \mu(\mathrm{id} \otimes \eta)(B \otimes \mathrm{id}_{\mathbb{C}})(\mathrm{id} \otimes \varepsilon)\Delta.$$

By unitality, $\mu(\mathrm{id}\otimes\eta)(B\otimes\mathrm{id}_{\mathbb{C}}) = B$. Moreover, counitality yields $(\mathrm{id}\otimes\varepsilon)\Delta = \mathrm{id}$. Hence $B * \eta\varepsilon = B$. Similarly, we get $\eta\varepsilon * B = B$.

3.8 *Uniqueness of the coinverse.* Show that the coinverse of a Hopf algebra is unique. Solution: Suppose that

$$\mathrm{id} * S = S * \mathrm{id} = \eta\varepsilon,$$

and $\mathrm{id} * S' = S' * \mathrm{id} = \eta\varepsilon$. By Prop. 3.5 on page 130,

$$S' * (\mathrm{id} * S) = S' * \eta\varepsilon = S'.$$

Similarly, $(S' * \mathrm{id}) * S = \eta\varepsilon * S = S$. By associativity of the convolution, $S' = S$.

3.9 *Anti-multiplicativity of the coinverse.* Show that the coinverse $S : \mathcal{A} \to \mathcal{A}$ of a Hopf algebra $\mathcal{A}$ is an algebra anti-morphism, that is, S is linear and we have $S(ab) = S(b)S(a)$ for all $a, b \in \mathcal{A}$.
Hint: See Lemma 1.26 of the monograph by J. Gracia-Bondia, J. Várilly, and H. Figueroa, Elements of Noncommutative Geometry, Birkhäuser, Boston, 2001.

3.10 *Hopf algebras and finite groups.* Prove Proposition 3.8 on page 148.
Solution: Set $\mathcal{A} := \mathcal{H}(G)$. For given functions $\varphi, \psi \in \mathcal{A}$, the tensor product $\varphi \otimes \psi$ denotes a function of two variables given by

$$(\varphi \otimes \psi)(g, h) := \varphi(g)\psi(h) \qquad \text{for all} \quad g, h \in G.$$

The tensor algebra $\mathcal{A} \otimes \mathcal{A}$ consists of all the finite sums

$$\varphi_1 \otimes \psi_1 + ... + \varphi_n \otimes \psi_n, \quad \varphi_j, \psi_j \in \mathcal{A}, \quad j = 1, ..., n, n = 1, 2, ...$$

Since the group G is finite, the tensor algebra $\mathcal{A} \otimes \mathcal{A}$ contains precisely all complex functions $\varphi : G \times G \to \mathbb{C}$ of two variables, $\varphi = \varphi(g, h)$, where $g, h \in G$. Similarly, $\mathcal{A} \otimes \mathcal{A} \otimes \mathcal{A}$ consists of all complex functions $\chi : G \times G \times G \to \mathbb{C}$ of three variables, $\chi = \chi(g, h, k)$, where $g, h, k \in G$. For all $\varphi, \psi \in \mathcal{A}$ and all $z \in \mathbb{C}$, we set

$$\mu(\varphi \otimes \psi) := \varphi\psi, \qquad \eta(z) := z\mathbf{1}, \qquad \mathrm{id}(\varphi) := \varphi.$$

We have to show the following:
(CA) $(\mathrm{id} \otimes \Delta)\Delta = (\Delta \otimes \mathrm{id})\Delta$ (coassociativity).
(CU) $(\varepsilon \otimes \mathrm{id})\Delta = (\mathrm{id} \otimes \varepsilon)\Delta = \mathrm{id}$ (counitality).
(S) $\mu(S \otimes \mathrm{id})\Delta = \mu(\mathrm{id} \otimes S)\Delta = \eta\varepsilon$ (coinverse).
These properties follow easily from the corresponding definitions of the maps Δ, ε, and S on page 148. Let us show this. In what follows, the symbols g, h, k, m are elements of the group G. The map Δ sends $\mathcal{A}$ to $\mathcal{A} \otimes \mathcal{A}$, that is, functions of one variable are transformed into functions of two variables. Explicitly, let

$$\psi := \Delta\varphi, \qquad \varphi \in \mathcal{A}.$$

Then $\psi(g, h) = \varphi(gh)$. If $\tau \in \mathcal{A} \otimes \mathcal{A}$, then the function $\sigma := \mu\tau$ is given by

$$\sigma(g) = \tau(g, g) \qquad \text{for all} \quad g \in G.$$

This follows from $\tau = \sum_{jk} \alpha_{jk}\chi_j \otimes \chi_k$ and $\sigma = \sum_{jk} \alpha_{jk}\chi_j\chi_k$ (see (3.46) on page 149).
Ad (CA). We will use $g(hk) = (gh)k$. The map $\mathrm{id} \otimes \Delta$ sends $\mathcal{A} \otimes \mathcal{A}$ to $\mathcal{A} \otimes \mathcal{A} \otimes \mathcal{A}$. Let $\chi := (\mathrm{id} \otimes \Delta)\psi$. Then

$$\chi(g, h, k) = \psi(g, hk) = \varphi(g(hk)).$$

Similarly, let $\varrho := (\Delta \otimes \mathrm{id})\psi$. Then

$$\varrho(g, h, k) = \psi(gh, k) = \varphi((gh)k).$$

Therefore, $\chi = \varrho$.
Ad (CU). We will use $eg = ge = g$. Let $\alpha := (\varepsilon \otimes \mathrm{id})\psi$. Then

$$\alpha(g) = \psi(e, g) = \varphi(eg).$$

Similarly, setting $\beta := (\mathrm{id} \otimes \varepsilon)\psi$, we get

$$\beta(g) = \psi(g, e) = \varphi(ge).$$

Consequently, $\alpha = \beta$.
Ad (S). We will use $g^{-1}g = gg^{-1} = e$. Let $\sigma := \mu\tau$ with $\tau := (S \otimes \mathrm{id})\psi$. Then

$$\sigma(g) = \tau(g, g) = \psi(g^{-1}, g) = \varphi(g^{-1}g) = \varphi(e).$$

Similarly, setting $\nu := \mu\lambda$ with $\lambda := (\mathrm{id} \otimes S)\psi$, we obtain

$$\nu(g) = \lambda(g, g) = \psi(g, g^{-1}) = \varphi(gg^{-1}) = \varphi(e).$$

Finally, let $\alpha := \eta\beta$ and $\beta := \varepsilon(\varphi)$. Then $\beta(g) = \varphi(e)$ and $\alpha(g) = \varphi(e)$. Therefore, $\sigma = \nu = \eta\varepsilon$.

3.11 *Hopf algebras and operator groups.* Prove Proposition 3.9 on page 151. Hint: Argue similarly to the solution of Problem 3.10.

3.12 *Proof of the Lagrange inversion formulas.* Prove the formulas (3.22) and (3.23) on page 133. Solution:
Ad (3.22). Using an induction argument, we show first that we have the following key relation

$$\frac{\partial^n u}{\partial e^n} = \frac{\partial^{n-1}}{\partial \tau^{n-1}} \left(\varphi(u)^n \frac{\partial u}{\partial \tau} \right), \qquad n = 1, 2, \ldots \tag{3.78}$$

Here, we briefly write u instead of $u(\tau, e)$.
(I) $n = 1$. Differentiating the Lagrange equation

$$u(\tau, e) = \tau + e\varphi(u(\tau, e)) \tag{3.79}$$

with respect to the variables τ and e, we get

$$u_\tau = 1 + e\varphi'(u)u_\tau, \qquad u_e = \varphi(u) + e\varphi'(u)u_e.$$

Multiplying this by u_e and u_τ, respectively, we get $u_e = \varphi(u)u_\tau$. This is the claim (3.78) with $n = 1$.
(II) $n \Rightarrow n + 1$. If the function F is smooth, then the chain rule tells us that

$$\frac{\partial}{\partial e} \left(F(u(\tau, e)) \frac{\partial u(\tau, e)}{\partial \tau} \right) = \frac{\partial}{\partial \tau} \left(F(u(\tau, e)\varphi(u(\tau, e)) \frac{\partial u(\tau, e)}{\partial e} \right). \tag{3.80}$$

Now suppose that equation (3.78) is true for n. Differentiating this with respect to the variable e and noting (3.80), we get

$$\frac{\partial^{n+1} u}{\partial e^{n+1}} = \frac{\partial^{n-1}}{\partial \tau^{n-1}} \frac{\partial}{\partial e} \left(\varphi(u)^n \frac{\partial u}{\partial \tau} \right) = \frac{\partial^n}{\partial \tau^n} \left(\varphi(u)^{n+1} \frac{\partial u}{\partial e} \right).$$

This proves (3.78). Finally, by Taylor expansion we get

$$u(\tau, e) = u(\tau, 0) + e \frac{\partial u(\tau, 0)}{\partial e} + \sum_{n=2}^{\infty} \frac{e^n}{n!} \frac{\partial^n u(\tau, 0)}{\partial e^n}.$$

By the Lagrange equation (3.79), we have $u(\tau, 0) = \tau$ and $u_\tau(\tau, 0) = 1$. Thus, it follows from the key relation (3.78) that

$$u(\tau, e) = u(\tau, 0) + e\varphi(\tau) + \sum_{n=2}^{\infty} \frac{e^n}{n!} \frac{d^{n-1}}{d\tau^{n-1}} \varphi(\tau)^n.$$

Ad (3.23). Use a similar argument.

3.13 *The first Bell polynomials.* Let x_1, x_2, x_3 be complex numbers. Compute the exponential function

$$f(t) := \mathrm{e}^{x_1 t + \frac{1}{2} x_2 t^2 + \frac{1}{6} x_3 t^3}$$

up to order three in order to compute the Bell polynomials $B_{3,j}, j = 1, 2, 3$.
Solution: By the addition theorem, $f(t) = \mathrm{e}^{x_1 t} \mathrm{e}^{\frac{1}{2} x_2 t^2} \mathrm{e}^{\frac{1}{6} x_3 t^3}$. Hence $f(t)$ is equal to

$$\begin{aligned} &(1 + x_1 t + \tfrac{1}{2} x_1^2 t^2 + \tfrac{1}{6} x_1^3 t^3 + \ldots)(1 + \tfrac{1}{2} x_2 t^2 + \ldots)(1 + \tfrac{1}{6} x_3 t^3 + \ldots) \\ &\qquad = 1 + x_1 t + \tfrac{1}{2}(x_2 + x_1^2) t^2 + \tfrac{1}{6}(x_3 + 3x_1 x_2 + x_1^3) t^3. \end{aligned}$$

Therefore, $B_{1,1} = x_1, B_{2,1} = x_2, B_{2,2} = x_1^2$, and

$$B_{3,1} = x_3, \quad B_{3,2} = 3x_1 x_2, \quad B_{3,3} = x_1^3.$$

3.14 *Reduced correlation functions.* Prove Proposition 3.7 on page 141.
Solution: As in Problem 3.13, the addition theorem for the exponential formula tells us that $\exp(Z_{\rm red}[J])$ is equal to the infinite product

$$\sum_{k_1=0}^{\infty} \frac{1}{k_1!} \left(\frac{\mathrm{i}}{\hbar} \int_{\mathbb{R}^4} C_{1,\rm red}(x_1) J(x_1) d^4x_1 \right)^{k_1} \times \sum_{k_2=0}^{\infty} \frac{1}{k_2!} \left(\frac{\mathrm{i}^2}{2!\hbar^2} \int_{\mathbb{R}^8} C_{2,\rm red}(x_1,x_2) J(x_1) J(x_2) d^4x_1 d^4x_2 \right)^{k_2} \times \cdots$$

We want to show that this is equal to $Z[J]$. To this end, we inspect the terms of order $m = 1, 2, \ldots$ with respect to J.
(I) Let $m = 1$. We obtain

$$\frac{\mathrm{i}}{\hbar} \int_{\mathbb{R}^4} C_{1,\rm red}(x_1) J(x_1) d^4x_1 .$$

By assumption, $C_1(x_1) = C_{1,\rm red}(x_1)$.
(II) Let $m = 2$. We have to choose the two products with $k_1 = 2, k_2 = 0$ and $k_1 = 0, k_2 = 2$. For $k_1 = 2$, we get

$$\left(\int_{\mathbb{R}^4} C_{1,\rm red}(x_1) J(x_1) d^4x_1 \right)^2 = \int_{\mathbb{R}^8} C_{1,\rm red}(x_1) J(x_1) C_{1,\rm red}(x_2) J(x_2) d^4x_1 d^4x_2,$$

up to the factor $\frac{\mathrm{i}^2}{2!\hbar^2}$. Adding the product for $k_1 = 0, k_2 = 2$, we get

$$\frac{\mathrm{i}^2}{2!\hbar^2} \int_{\mathbb{R}^8} \big(C_{1,\rm red}(x_1) C_{1,\rm red}(x_2) + C_{2,\rm red}(x_1,x_2) \big) J(x_1) J(x_2) d^4x_1 d^4x_2.$$

By assumption, $C_2(x_1,x_2) = C_{1,\rm red}(x_1) C_{1,\rm red}(x_2) + C_{2,\rm red}(x_1,x_2)$.
(III) For $m = 3, 4, \ldots$, the proof proceeds similarly.

3.15 *The classical Möbius function.* Prove (3.68) on page 166. Solution: Consider first $\mu(1,n)$ for $n = 2, 3, \ldots$ The basic trick is the binomial formula

$$0 = (1-1)^n = 1 - \binom{n}{1} + \binom{n}{2} + \ldots + (-1)^{n-1} \binom{n}{n-1} + (-1)^n . \quad (3.81)$$

Let $p, q, p_1, \ldots, p_r$ be pairwise different prime numbers. Note that $\mu(1,1) = 1$ and

$$\mu(1,n) = - \sum_{1 \preceq k \prec n} \mu(1,k).$$

(I) $\mu(1,p) = -1$. In fact, $\mu(1,p) = -\mu(1,1) = -1$. Moreover,

$$-\mu(1,pq) = \mu(1,1) - \mu(1,p) - \mu(1,q) = \left(1 - \binom{2}{1} + 1 \right) - 1 = -1.$$

By induction based on (3.81), the same argument shows that

$$\mu(p_1 p_2 \cdots p_r) = (-1)^r .$$

(II) $\mu(1,p^2)$=0. In fact, $\mu(1,p^2) = -\mu(1,1) - \mu(1,p) = -1 + 1 = 0$. Moreover, we have

$$\begin{aligned} -\mu(1,p^2q) &= \mu(1,1)+\mu(1,p)+\mu(1,q)+\mu(1,pq)+\mu(1,p^2) \\ &= \left(1-\binom{2}{1}+1\right)+0=0. \end{aligned}$$

By induction based on (3.81), the same argument shows that $\mu(1,n)=0$ if n contains the square of a prime number as divisor.

4. The Strategy of Equivalence Classes in Mathematics

> Quantum states are equivalence classes. Global physical fields are sections of bundles.
>
> Folklore

One of the main strategies in the sciences consists in using classifications. This means that we put single objects into classes. Instead of studying individual objects, we investigate the properties of classes. This is a simple, but extremely powerful general strategy. For example, the first systematic classification of plants and animals was developed by the Swedish biologist Carl von Linné (1707–1778). Baron de la Brède et de Montesquieu (1689–1755) said:

> Intelligence consists of this; that we recognize the similarity of different things and the difference between similar ones.

In terms of mathematics, this corresponds to introducing operations for equivalence classes. In this series of monographs, we will study the following topics:

(I) Simplifying mathematical theories or justifying formal approaches by introducing ideal elements:
- imaginary numbers (solution of algebraic equations, theory of analytic functions, Fourier transform, quantum theory, conformal field theory, string theory),
- infinite points in projective geometry (e.g., the compactification of algebraic curves, compact Riemann surfaces, elliptic and Abelian integrals),
- generalized derivatives and distributions (e.g., the Dirac delta function),
- Mikusiński's superfunctions (or operators) in electrical engineering (justification of the Heaviside calculus),
- Sato's hyperfunctions (generalized analytic functions, applications to partial differential equations),
- cardinal and ordinal numbers, transfinite induction (Cantor's structuring of the infinite),
- justification of Leibniz's infinitesimals in non-standard analysis via ultrafilters.

(II) Equivalence classes in algebra:
- the Gaussian ring of integers modulo a fixed integer, and the quadratic reciprocity law in number theory,
- the Gaussian ring, the Fermat–Euler theorem, and a modern coding algorithm,
- the construction of algebraic objects that satisfy prescribed relations (e.g., complex numbers, quaternions, octions, tensor algebras, Grassmann algebras, Clifford algebras, universal envelopping algebra of a Lie algebra, supersymmetric algebras, and quantum groups),
- quotient structures (e.g., quotien groups, quotient rings, quotient algebras, quotient fields),

- the quotient field over the convolution ring of continuous functions (Mikusiński's rigorous approach to the Heaviside calculus in electrical engineering),
- field extensions (e.g., algebraic equations and Galois theory, algebraic numbers, rational functions),
- central extensions of groups and Lie algebras (Bargmann's theorem on the projective representations of the Poincaré group, the Virasoro algebra and conformal quantum field theory),
- exact sequences, chain complexes, homology groups, and homological algebra,
- the cohomology of groups or Lie algebras,
- direct (resp. codirect) limits of sets, linear spaces, groups, function spaces, spaces of generalized functions (distributions), and topological spaces (e.g., Čech cohomology),[1]
- the strategy of generalized physical fields (fiber bundles, sheaves, Grothendieck's schemes in algebraic geometry, algebraic K-theory),
- the strategy of coordinatization (duality, Hopf algebras, quantum groups),
- the strategy of motives in number theory,
- categories and functors as a general tool in order to describe mathematical structures.

(III) Equivalence classes in analysis:

- Cantor's construction of real numbers as equivalence classes of Cauchy sequences, the completion of metric spaces, and Hensel's p-adic numbers,
- completion of a normed space to a Banach space: the Lebesgue integral, Sobolev spaces and the energetic approach to both the partial differential equations of mathematical physics and the calculus of variations (e.g., the justification of the Dirichlet principle in electrostatics),
- the completion of a pre-Hilbert space and the Gelfand–Naimark–Segal (GNS) construction for C^*-algebras in the algebraic approach to quantum theory.
- The Riemann–Hilbert problem and the Birkhoff decomposition (the Connes–Kreimer approach to renormalization and renormalization groups, the universal Connes–Kreimer–Moscovici Hopf algebra),
- the motivic Galois group in renormalization group theory due to Connes and Marcolli.

(IV) Equivalence classes in geometry:

- homogeneous spaces and transformation groups (orbit spaces and orbit types in gauge field theory),
- spaces of quantum states and projective geometry (e.g., the Hopf fibration and the electron spin),
- curvature and fiber bundles (the Standard Model in elementary particle physics),
- universal covering of a Lie group (e.g., the electron spin),
- universal covering of a Riemann surface; scattering of strings; global parametrization (uniformization) of Riemann surfaces and algebraic varieties, algebraic functions and Abelian integrals, Riemann's moduli space of Riemann surfaces,
- the Teichmüller space as a universal covering of Riemann's moduli space,
- Riemann's holonomy group of a differential equation (differential equations of Fuchsian type and special functions like Gauss' hypergeometric function),

[1] Direct and codirect limits are also called inductive and projective limits, respectively.

- holonomy group of a principal fiber bundle (the Ashtekar approach to quantum gravitation),
- lattices as quotient groups and special functions (periodic functions, elliptic functions, automorphic functions, modular functions, theta functions, Jacobi varieties),
- sheaf theory, cohomology with values in a sheaf, and the global construction of analytic functions and differentials, divisors and line bundles (the Cousin problems, Abelian integrals, the Riemann–Roch–Hirzebruch theorem),
- sheaves and Grothendieck's schemes in algebraic geometry and number theory (e.g., local rings, divisors),
- K-theory of operator algebras (noncommutative geometry and quantum field theory).
- Quantization of gauge theories via the Faddeev–Popov method (pseudomeasure on the orbit space induced by the gauge group, factorization of the Feynman functional integral, ghosts).

(V) Equivalence classes in topology:

- bundles and cocycles (physical fields and observers),
- topological quotient spaces (e.g., the topology of projective spaces or the spectrum of a quantum operator as the space of maximal ideals in the Gelfand theory of C^*-algebras),
- fundamental group and higher homotopy groups of a topological space,
- homology groups of a topological space (e.g., the Lefschetz fixed-point theorem),
- cohomology groups of a topological space (e.g., electrical circuits, topological charges of physical fields, de Rham cohomology, characteristic classes, Chern class, Thom class, Stiefel–Whitney class, existence of the 4-potential in Maxwell's theory of electromagnetism),
- dynamical systems and Floer homology (generalized Morse theory),
- dynamical systems and the Conley index as the homotopy type of a topological space (generalized Morse theory),
- Grothendieck's algebraic K-theory and the Riemann–Roch–Hirzebruch theorem,
- the Atiyah–Hirzebruch topological K-theory (generalized cohomology) of topological spaces (index of Fredholm operators, Atiyah–Singer index theorem, homotopy groups of the space of Fredholm operators of a Hilbert space, vector fields on spheres, string theory),
- Frobenius manifolds, moduli spaces, and quantum cohomology.

This impressive list of deep tools in mathematics and its relations to physics shows that:

Equivalence classes are everywhere in mathematics and physics.

The reader should note the following. One of the most important tools in modern mathematics are bundles, which globalize classical notions like linear spaces, Lie groups, and so on. We want to show that:

Bundles in mathematics are closely related to physical fields.

This helps to understand many deep relations between modern mathematics and modern physics.

Classification. Suppose that a given nonempty set X is decomposed into pairwise disjoint nonempty sets X_α:

$$X = \bigcup_{\alpha \in \mathcal{A}} X_\alpha .$$

Here, $\mathcal{A}$ denotes an index set (e.g., $\mathcal{A} := \{1, 2, \ldots, N\}$). For two elements x and y, we write

$$x \sim y$$

iff there exists an index α such that $x, y \in X_\alpha$. Obviously, for all $x, y, z \in X$, the following hold:

(R) Reflexivity: $x \sim x$.
(S) Symmetry: if $x \sim y$, then $y \sim x$.
(T) Transitivity: if $x \sim y$ and $y \sim z$, then $x \sim z$ (transitivity).

Equivalence relation. We are given the set X. Suppose that, for certain elements x and y of X, there exists a relation $x \sim y$ which has the properties (R), (S), and (T) above. Then the relation '$\sim$' is called an equivalence relation. Introduce the equivalence class $[x]$ of the element x by setting

$$\boxed{[x] := \{y \in X : \ x \sim y\}.}$$

The elements of $[x]$ are called the representatives of the equivalence class.

Proposition 4.1 *The equivalence classes yield a partition of the set X into pairwise disjoint subsets $[x]$.*

The set of these equivalence classes is denoted by the symbol

$$X/\sim \ := \{ \ [x] : \ x \in X\}.$$

This set is called the quotient space with respect to the equivalence relation '$\sim$'.
Proof. If $z \in [x]$ and $z \in [y]$, then $x \sim z$, $y \sim z$. Hence $x \sim y$, by symmetry and transitivity. This implies $[x] = [y]$. Conversely, if $[x] = [y]$, then $x \sim y$. Thus, either two equivalence classes coincide or they are disjoint. □

We will show on page 199 that quantum states are equivalence classes.

4.1 Equivalence Classes in Algebra

Let us discuss the basic ideas about quotient structures in algebra.

4.1.1 The Gaussian Quotient Ring and the Quadratic Reciprocity Law in Number Theory

> Gauss is supposed to have discovered a proof of the law of quadratic reciprocity in 1796 when he was nineteen... This law, which Gauss called the gem of arithmetic, is a basic result on congruences. After Gauss gave his six proofs, more than fifty others were given by later mathematicians.[2].
>
> Morris Kline, 1990

The ring $\mathbb{Z}$ of integers. Let R denote the set $\mathbb{Z}$ of integers $0, \pm 1, \pm 2, \ldots$ Then, for all $a, b, c \in R$, the following hold:

(R0) Consistency: $a + b, ab \in R$.

[2] M. Kline, Mathematical Thought from Ancient to Modern Times, Vol. 2, Oxford University Press, 1990.

(R1) Additivity: R is an additive group.[3]
(R2) Distributivity: $(a+b)c = ac+bc$ and $c(a+b) = ca+cb$.
(R3) Associativity: $(ab)c = a(bc)$.
(R4) Commutativity: $ab = ba$.
(R5) Unitality: There exists an element 1 (called the unit element of R) such that $1a = a1 = a$ for all $a \in R$.

A set R is called a ring iff there exist an addition $a+b$ and a multiplication ab for all $a, b \in R$ such that the properties (R0)–(R3) are satisfied. If, in addition, property (R4) is also met, then the ring is called commutative. Using this terminology, the ring $\mathbb{Z}$ of integers is a commutative ring with unit element 1.

- A ring R with unit element 1 is called a skew-field iff, for any given nonzero element a, there exists an element in R denoted by a^{-1} such that we have the relation[4] $aa^{-1} = a^{-1}a = 1$.
- A commutative skew-field is called a field.[5]

For example, the sets $\mathbb{Q}$ (rational numbers), $\mathbb{R}$ (real numbers), $\mathbb{C}$ (complex numbers) are fields, whereas the set $\mathbb{H}$ of quaternions is only a skew-field.[6]

A subset S of the ring R is called a subring iff it is a ring with respect to the operations on R. By definition, a two-sided ideal S of a ring R is a subring with

$$sr \in R \quad \text{and} \quad rs \in R \qquad \text{for all} \quad s \in S, r \in R.$$

In a commutative ring, two-sided ideals are briefly called ideals.[7] For a fixed integer m, we set

$$m\mathbb{Z} := \{mk : \ k \in \mathbb{Z}\}.$$

Obviously, $m\mathbb{Z}$ is an ideal of $\mathbb{Z}$. We call this the ideal generated by the integer m. We also briefly write (m) instead of $m\mathbb{Z}$.

By a ring morphism, we understand a map

$$\chi : R \to T \tag{4.1}$$

between the rings R and T which respects addition and multiplication, that is, for all $a, b \in R$, we have

$$\chi(a+b) = \chi(a) + \chi(b), \qquad \chi(ab) = \chi(a)\chi(b). \tag{4.2}$$

If R and T are fields, then the map (4.1) is called a field morphism iff (4.2) holds. Bijective field morphisms are called field isomorphisms. Analogously, we define morphisms and isomorphisms for skew-fields.

[3] See Sect. 7.5 of Vol. I.

[4] One easily shows that this inverse element a^{-1} is uniquely determined by a, and the nonzero elements of R form a group. A ring with unit element is also called a unital ring.

[5] For skew-fields and fields, we always exclude the trivial case $\{0\}$, that is, we assume that there exist two different elements, namely, the zero element 0 and the unit element 1.

[6] A quaternion is given by $ai + bj + ck + d$ where a, b, c, d are real numbers, and $i^2 = j^2 = k^2 = -1$, as well as $ij = -ji = k, jk = -kj = i, ki = -ik = j$.

[7] Ideals were introduced by Kummer (1810–1893) in order to prove a special case of Fermat's last theorem in number theory. The final proof of Fermat's last theorem was given by Wiles (born 1953) in 1994. Important contributions to the theory of ideals were made by Dedekind (1831–1916), as well as by von Neumann (1903–1957) and by Gelfand (born 1913) in the context of operator theory related to quantum mechanics (von Neumann algebras and Gelfand's C^*-algebras).

The Gaussian ring $\mathbb{Z}/\mathrm{mod}\ m$ of residue classes modulo m. Choose a fixed integer m. Let $x, y \in \mathbb{Z}$. Following Gauss we write

$$x \equiv y \mod m \qquad \text{iff} \qquad x - y \in (m).$$

This is an equivalence relation (also called congruence relation). The equivalence classes are denoted by $[x]$. They are also called residue classes. In other words, x is congruent to y modulo m iff the difference $x - y$ is divisible by the integer m. For example, if $m = 3$, then $2 \equiv 5 \mod 3$, and there are precisely three equivalence classes $[0], [1], [2]$, namely,

$$[0] = (m), \quad [1] = 1 + (m) = \{1, 1 \pm 3, 1 \pm 6, 1 \pm 9 \ldots\}, \quad [2] = 2 + (m).$$

In particular, the representatives of $[2]$ are $2, 5, 8, \ldots, -1, -4, -7 \ldots$ Addition and multiplication of residue classes are defined by

$$\boxed{[x] + [y] := [x + y], \qquad [x][y] = [xy].}$$

This definition does not depend on the choice of the representatives.[8] For example, if $m = 2$, then $[2]+[1] = [3] = [1]$, and $[2] = [4], [1] = [7]$, as well as $[4]+[7] = [11] = [1]$. The operations for the two elements of $\mathbb{Z}/\mathrm{mod}\ 2$ are given by

$$[1] + [1] = [0], \quad [0] + [1] = [1] + [0] = [1], \quad [0] + [0] = [0], \tag{4.3}$$

and $[0][1] = [1][0] = [0]$ and $[1][1] = [1]$. Thus, $\mathbb{Z}/\mathrm{mod}\ 2$ is a field (also briefly denoted by $\mathbb{Z}_2$). The same result can be obtained by computing with integers and by setting '$2 = 0$'. For example,

$$5 + 8 = 13 = 1 + 6 \cdot 2 = 1, \qquad 5 \cdot 8 = 40 = 20 \cdot 2 = 0.$$

This corresponds to $[5] + [8] = [13] = [1]$ and $[5][8] = [40] = [0]$. In the following proposition, assume that $m = 1, 2, 3, \ldots$.

Proposition 4.2 (i) *The quotient space $\mathbb{Z}/\mathrm{mod}\ m$ is a commutative ring with the unit element* $[1]$.

(ii) *The quotient ring $\mathbb{Z}/\mathrm{mod}\ m$ is a field iff m is a prime number.*

(iii) *The order of a finite field is a prime power.*

(iv) *Conversely, for any given prime power p^n, there is a unique finite field of order p^n (up to isomorphism).*

(v) *Each finite skew-field is a field.*

Finite fields are also called Galois fields. The proofs can be found in the standard textbooks on algebra. We refer to:

B. van der Waerden, Moderne Algebra, Vols. 1, 2, Springer, Berlin, 1930, 8th edition, 1993 (in German). English edition: Modern Algebra, Frederyck Ungar, New York, 1975.

S. Lang, Algebra, Springer, New York, 2002.

[8] This follows from the fact that the equivalence relation respects addition and multiplication. That is, if $x \equiv y \mod m$ and $u \equiv v \mod m$, then

$$x + u \equiv y + v \mod m \qquad \text{and} \qquad xu \equiv yv \mod m.$$

In the literature, one also writes $\mathbb{Z}_m$ instead of $\mathbb{Z}/\text{mod } m$. For example, the additive group $\mathbb{Z}_2$ from (4.3) is isomorphic to the multiplicative group $\{1, -1\}$. The isomorphism is given by

$$[0] \mapsto 1, \qquad [1] \mapsto -1.$$

The quadratic reciprocity law in number theory. In 1801, the young Gauss (1777–1855) published his *Disquisitiones arithmeticae* (investigations on arithmetic). This work founded modern number theory. In particular, Gauss proved the following so-called quadratic reciprocity law. This law was empirically discovered by Euler (1707–1783) and by Legendre (1752–1833). Gauss gave the first complete proof. The goal is to study the quadratic equation

$$\boxed{x^2 \equiv q \mod p} \tag{4.4}$$

and its dual equation $x^2 \equiv p \mod q$. If equation (4.4) has a solution, then we set $\left(\frac{q}{p}\right) := 1$. Otherwise, $\left(\frac{q}{p}\right) := -1$. This is called the Legendre symbol.

Theorem 4.3 *If p and q are prime numbers greater than two, then*

$$\left(\frac{q}{p}\right) = (-1)^{(p-1)(q-1)/4} \left(\frac{p}{q}\right).$$

In addition, $\left(\frac{p-1}{p}\right) = (-1)^{(p-1)/2}$ *and* $\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8}$.

Example. The equation

$$x^2 \equiv 4 \mod 5$$

has a solution, namely, $x = 2$. The equation

$$x^2 \equiv 2 \mod 3$$

has no solution. In fact, if $x = 0, 1, 2$, then $x^2 \equiv a \mod 3$ with $a = 0, 1, 1$. Using the quadratic reciprocity law, we have $\left(\frac{4}{5}\right) = 1$ and $\left(\frac{2}{3}\right) = -1$.

Hints for further reading. We refer to:

> Carl Friedrich Gauß, Disquisitiones arithmeticae, 1801 (in Latin).
> English edition: C. F. Gauss, Disquisitiones Arithmeticae, translated by A. Clarke, Yale University, New Haven, Connecticut, 1965.
> German edition: Untersuchungen über höhere Arithmetik (Investigations on higher arithmetic), Springer, Berlin, 1986.

The tremendous influence of this masterpiece on the development of modern mathematics is described in:

> C. Goldstein, N. Schappacher, and J. Schwermer, The Shaping of Arithmetic after Gauss' Disquisitiones Arithmeticae, Springer, Berlin, 2007.

In Chapter 5, we will discuss the enormous influence of another masterpiece of Gauss on the development of modern mathematics and physics, namely:

> Carl Friedrich Gauß, Disquisitiones generales circa superficies curvas (General theory of curved surfaces).
> In: C. F. Gauß, Werke (Collected Works), Vol. 5, pp. 217–256; 341–347, Göttingen 1863/1929 (in Latin).
> German edition, C. F. Gauß, Allgemeine Flächentheorie, Ostwalds Klassiker, Vol. 5, Leipzig, 1889.

This treatise founded differential geometry based on the crucial notion of curvature. Nowadays we know that the fundamental forces in the universe are described by curvature. We refer to:

P. Dombrowski, 150 years after Gauss' 'Disquisitiones generales circa superficies curvas', Astérisque **62** (1979).

G. Dunnington, Carl Friedrich Gauß, Titan of Science, New York, 1955.

4.1.2 Application of the Fermat–Euler Theorem in Coding Theory

More than 200 years the theorem of Fermat–Euler was considered to be only a result in pure mathematics. In 1977 however, Rivest, Shamir and Adleman published a bafflingly simple and yet extraordinarily secure code, which is based on the Fermat–Euler theorem.

The Euler function. The positive integers k and m are called relatively prime iff 1 is the only common divisor. For example, 3 and 5 are relatively prime, but 2 and 4 are not. Let $m = 1, 2, \ldots$ The Euler number $\varphi(m)$ tells us how many of the numbers $1, 2, \ldots, m$ are relatively prime to m. For example, $1, 3$ are relatively prime to 4, but 2, 4 are not. Hence $\varphi(4) = 2$. Furthermore,

$$\varphi(1) = \varphi(2) = 1, \quad \varphi(3) = 2, \quad \varphi(5) = 4, \quad \varphi(6) = 2.$$

If m is a prime number, then $\varphi(m) = m - 1$. Alternatively, $\varphi(m)$ is equal to the invertible elements of the ring $\mathbb{Z}/\text{mod } m$. The function $\varphi : \mathbb{N}^\times \to \mathbb{N}^\times$ is called the Euler function.

The Fermat–Euler theorem. This theorem due to Fermat (1601–1665) and Euler (1707–1783) tells us the following:

For positive integers a and m, which are assumed to be relatively prime, one has $a^{\varphi(m)} \equiv 1 \mod m$.

Fermat formulated this theorem in the special case where m is a prime number. Then $\varphi(m) = m - 1$. For example, $2^4 \equiv 1 \mod 5$. Let us briefly discuss the application of this theorem to the following sophisticated coding method.

Preparations by the operator. Step 1: Here two prime numbers p and q, roughly of the size 10^{100}, are chosen and kept *secrete*.

Step 2: One forms the product $m = pq$ and calculates $\varphi(m) = (p-1)(q-1)$.

Step 3: One chooses an additional positive integer s with $0 < s < \varphi(m)$.

Step 4: The person sending the message is given *publicly* the two numbers m and s.

Encoding the message. The message is simply encoded in a single positive integer n.[9] The person sending the message computes the number r by the following equation

$$n^s \equiv r \mod m.$$

The number r is the only information sent to the operator.

Decoding the message by the operator. Here one must reconstruct from the remainder r the original number n. The operator proceeds as follows.

Step 1: He uses the two numbers m and s in order to compute a positive integer t which satisfies the equation

[9] For example, one associates to every letter a two-digit number 10, 11, 12, ... and replaces in the message all occurrences of the letter by that two-digit number. Then forming the concatenation of all of these, one gets a big number n.

$$ts \equiv 1 \mod \varphi(m). \tag{4.5}$$

This equation always has a solution, since $\varphi(m)$ and s are relatively prime.

Step 2: He now just divides r^t by the number m. The remainder is the sought for number n which allows the operator to decode the message.

Justification of the procedure. The following fact is the key to this procedure.

Proposition 4.4 $r^t \equiv n \mod m$.

Proof. By the Fermat–Euler theorem, one has

$$\boxed{n^{\varphi(m)} \equiv 1 \mod m.}$$

By (4.5), there exists an integer k with $ts = 1 + k\varphi(m)$. It follows that

$$r^t \equiv n^{st} \equiv n^{1+k\varphi(m)} \equiv n \cdot n^{k\varphi(m)} \equiv n \mod m.$$

□

This proposition only determines the number $n + km$ for some natural number k. However, since the chosen number m is huge and the message is assumed to have a reasonable length, we always have $n < m$. Hence $k = 0$.

The security of this method. If an intruder wants to decode the message, he needs the number t, that is, $\varphi(m) = (p-1)(q-1)$.

> *To get this number he must determine the prime number decomposition of number m, which is known to him.*

The trick of this method is simply that because of the size of chosen for the prime numbers p and q, no computer is as yet able to determine the factors p and q in a reasonable amount of time. Since computers are becoming more and more powerful all the time, the security of this method is only guaranteed if one chooses new, larger numbers p and q from time to time. More sophisticated coding methods are based on the theory of elliptic curves in algebraic geometry.

The importance of quantum computers. The theory of quantum computers shows that there exist methods in order to factorize huge prime numbers in a reasonable amount of time. Shor's famous algorithm provides a quantum-computer method for factorizing an integer N in a number of steps which is polynomial (less than cubic) in the number of digits (i.e., $\ln N \cdot N^2$). Up to now, the experimental realization of quantum computers is missing and a challenge for the future. Problems of quantum information will be studied in Volume IV. We recommend:

> M. Nielsen and I. Chuang, Quantum Computation and Quantum Information, Cambridge University Press, 2001.
>
> D. Bouwmeester, A. Ekert, and A. Zeilinger (Eds.), The Physics of Quantum Information: Quantum Cryptography, Quantum Teleportation, Quantum Computation, Springer, Berlin, 2002.

Shor's fundamental paper is published on the Internet:

`http://arXiv.org/quant-ph/9508027`

For his important contributions to quantum computing, Peter Shor was awarded the Nevalinna prize in 1998.

4.1.3 Quotient Rings, Quotient Groups, and Quotient Fields

In a straightforward manner, let us generalize the Gaussian quotient ring $\mathbb{Z}/\mathrm{mod}\ m$. We recommend the reader to carry out the necessary proofs, which are easy consequences of the relevant definitions. [10]

Construction of quotient rings. Let J be a two-sided ideal of the ring R. For elements x, y of R, we write

$$x \sim y \qquad \text{iff} \qquad x - y \in J.$$

This is an equivalence relation. For the corresponding equivalence classes, we define an addition and a multiplication by setting

$$[x] + [z] := [x + z], \qquad [x][z] := [xz].$$

The point is that this definition does not depend on the choice of the representatives. This follows easily from the fact that J is a two-sided ideal. This way, the quotient space $R/\sim$ becomes a ring denoted by R/J. This ring is called the quotient ring of R modulo J. Set $\pi(x) := [x]$. The map

$$\pi : R \to R/J$$

is a surjective ring morphism. Such morphisms are also called ring epimorphisms.

The morphism theorem for rings. Suppose that we are given the ring morphism

$$f : R \to T \tag{4.6}$$

between the two rings R and T. We introduce the kernel of f by setting

$$\ker(f) := f^{-1}(0)$$

where 0 denotes the zero element of the ring T. Explicitly, the set $\ker(f)$ is equal to $\{r \in R : f(r) = 0\}$. We also introduce the image of the morphism f by setting $\mathrm{im}(f) := f(R)$. Finally, we set $f_*([r]) := f(r)$. This definition does not depend on the choice of the representative.

Theorem 4.5 *The map $f_* : R/\ker(f) \to \mathrm{im}(f)$ is a ring isomorphism.*

Simple rings. The ring morphism (4.6) is called trivial iff T is isomorphic to R or $T = \{0\}$. The ring R is called simple iff all the ring epimorphisms (4.6) are trivial. This is equivalent to the fact that the two-sided ideals of R are trivial (i.e., they are equal to $\{0\}$ or to R).

Quotient groups. If one replaces two-sided ideals by normal subgroups, then the construction of quotient rings can be immediately translated to the construction of quotient groups. Let S be a subgroup of the group G. Then S is called a normal subgroup of G iff $g^{-1}sg \in S$ for all $s \in S$ and all $g \in G$.

Construction of quotient groups. Let S be a normal subgroup of the group G. For elements x, y of G, we write

$$x \sim y \qquad \text{iff} \qquad xy^{-1} \in S.$$

This is an equivalence relation. For the corresponding equivalence classes, we define a multiplication by setting

[10] These proofs can be found in the two standard textbooks on algebra by van der Waerden (1930) and Lang (2002) quoted on page 180.

$$[x][y] := [xy].$$

The point is that this definition does not depend on the choice of the representatives. This follows easily from the fact that S is a normal subgroup. This way, the quotient set $G/\sim$ becomes a group denoted by G/S. This group is called the quotient group of G modulo S. Set $\pi(x) := [x]$. The map

$$\pi : G \to G/S$$

is a surjective group morphism. Such morphisms are also called group epimorphisms.

The morphism theorem for groups. Suppose that we are given the group morphism

$$f : G \to H \tag{4.7}$$

between the two groups G and H. We introduce the kernel of f by setting

$$\ker(f) := f^{-1}(1)$$

where 1 denotes the unit element of the group H. Explicitly, the set $\ker(f)$ is equal to $\{g \in G : f(g) = 1\}$. We also introduce the image of the morphism f by setting $\mathrm{im}(f) := f(G)$. Finally, we set $f_*([g]) := f(g)$. This definition does not depend on the choice of the representative.

Theorem 4.6 *The map $f_* : G/\ker(f) \to \mathrm{im}(f)$ is a group isomorphism.*

Simple groups. The group morphism (4.7) is called trivial iff H is isomorphic to G or $H = \{1\}$. The group G is called simple iff all the group epimorphisms (4.7) are trivial. This is equivalent to the fact that the normal subgroups of G are trivial (i.e., they are equal to $\{1\}$ or to G.)

The finite simple groups are completely classified. The mathematicians needed more than 100 years for this classification. Note that the full proof comprehends about 10 000 pages. We refer to:

D. Gorenstein, Classifying the finite simple groups, Bull. Amer. Math. Soc. **14** (1986), 1–98.

R. Solomon, A brief history of the classification of the finite simple groups. Bull. Amer. Math. Soc. **38**(3) (2001), 315–352.

The largest sporadic finite simple group has about 10^{64} elements. This group called the Monster was discovered by using methods rooted in quantum field theory (see Sect. 17.5 of Vol. I).

The permutation group. Fix the number $n = 1, 2, \ldots$ Let X be a finite set with n elements. The set of all bijective maps

$$\pi : X \to X$$

forms a group $\mathrm{Aut}(X)$ which is called the automorphism group of the set X. If $\pi, \sigma \in \mathrm{Aut}(X)$, then the product $\pi\sigma$ is given by the composition of maps, that is,

$$(\pi\sigma)(x) := \pi(\sigma(x)) \qquad \text{for all} \quad x \in X.$$

If we denote the elements of X by $1, 2, \ldots, n$, then the elements π of $\mathrm{Aut}(X)$ can be represented by the permutation

$$\begin{pmatrix} 1 & 2 & \ldots & n \\ x_1 & x_2 & \ldots & x_n \end{pmatrix}. \tag{4.8}$$

The $n!$ symbols (4.8) form a group called the symmetric group S_n. Obviously, we have the group isomorphism $\mathrm{Aut}(X) \simeq S_n$. By definition, a transposition τ is an element of S_n which transposes precisely two different elements. For example, the transposition between 1 and 3 reads as

$$\tau = \begin{pmatrix} 1\ 2\ 3\ 4 \dots\ n \\ 3\ 2\ 1\ 4 \dots\ n \end{pmatrix}.$$

The proof of the following proposition can be found in the standard textbook on algebra by Lang (2002), p. 31, quoted on page 180.

Proposition 4.7 *There exists precisely one group morphism*

$$\mathrm{sgn} : S_n \to \{1, -1\}$$

from the symmetric group S_n to the multiplicative group $\{1 - 1\}$ such that we have $\mathrm{sgn}(\tau) = -1$ for all transpositions τ.

The number $\mathrm{sgn}(\pi)$ is called the sign (or the parity) of the permutation π. Here, π is said to be even (resp. odd) iff $\mathrm{sgn}(\pi) = 1$ (resp. $= -1$). Each permutation $\pi \in S_n$ can be represented as a product of transpositions:

$$\pi = \tau_1 \tau_2 \cdots \tau_m.$$

Hence $\mathrm{sgn}(\pi) = \mathrm{sgn}(\tau_1)\mathrm{sgn}(\tau_2) \cdots \mathrm{sgn}(\tau_n) = (-1)^m$. The even permutations constitute the kernel of the map sgn: $S_n \to \{1, -1\}$, which is called the alternating group A_n. By the morphism theorem for groups, we have the isomorphism

$$S_n / A_n \simeq \{1, -1\}.$$

The group A_n of order $n!/2$ is simple if $n = 2, 3$, or $n \geq 5$. The group A_4 is not simple.

Quotient fields. The ring of integers $\mathbb{Z}$ is not a field, however, it can be extended to the field of rational numbers which can be represented by fractions $\frac{a}{b}$ of integers a, b where $b \neq 0$. We want to generalize this idea. In the next section, we will consider an elegant application to the Heaviside calculus in electrical engineering. In what follows all of the necessary computations resemble the computations known for rational numbers. However, our arguments only rely on the relations valid in rings and fields.

Zero divisors. In the Gaussian ring $\mathbb{Z}/\mathrm{mod}\ 6$, the decomposition $2 \cdot 3 = 6$ implies

$$[2][3] = [0].$$

The non-zero elements [2] and [3] are called zero divisors of the ring $\mathbb{Z}/\mathrm{mod}\ 6$. Similarly, let R be an arbitrary ring. An element a of R is called a zero divisor iff $a \neq 0$ and there exists a nonzero element of R such that $ab = 0$. Obviously, a field has never zero divisors. The Gaussian ring $\mathbb{Z}/\mathrm{mod}\ m$ has no zero divisors iff either $m = 1$ or m is a prime number.

Fractions. Let F be a field. For $a, b \in F$ with $a \neq 0$, we define

$$\boxed{\frac{a}{b} := ab^{-1}.}$$

It turns out that these fractions possess the usual properties known from rational numbers. In fact, for all $a, b, c, d \in F$ with $a \neq 0$ and $d \neq 0$, the following hold:[11]

[11] In (F4), we also assume that $c \neq 0$.

(F1) Equality: $\frac{a}{b} = \frac{c}{d}$ iff $ad = bc$.
(F2) Addition: $\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$.
(F3) Multiplication: $\left(\frac{a}{b}\right)\left(\frac{c}{d}\right) = \frac{ac}{cd}$.
(F4) Division: $\left(\frac{a}{b}\right)\left(\frac{c}{d}\right)^{-1} = \frac{ad}{bc}$.

Proof. In a field, we always have $bb^{-1} = 1$, and $(bc)^{-1} = c^{-1}b^{-1}$. Furthermore, note that $ac = ca$ because of commutativity.

Ad (F1). If $\frac{a}{b} = \frac{c}{d}$, then $ab^{-1} = d^{-1}c$. Multiplying this by b, we get $a = d^{-1}cb$. Hence $da = cb$. This argument can be reversed.

Ad (F2). We have to show that

$$(ad + bc)(bd)^{-1} = ab^{-1} + cd^{-1}.$$

In fact, this follows from $(ab^{-1} + cd^{-1})bd = ad + bc$.

Ad (F3), (F4). Argue similarly. □

The extension theorem. Let R be a commutative ring without zero divisors.[12] Then the following hold.

Theorem 4.8 (i) *The ring R can be extended to a field F. The elements of F have precisely the form ab^{-1} with $a, b \in R$ and $b \neq 0$.*

(ii) *The extension is universal, that is, each extension of R to a field $\mathcal{F}$ contains a subfield which is isomorphic to F.*

(iii) *If R is a real (resp. complex) algebra, then so is F.*

Proof. Ad (i). (I) Equivalence relation. Let $a, b, c, d \in R$ with $a \neq 0$ and $c \neq 0$. Our idea is to use symbols $\frac{a}{b}$ and to write

$$\frac{a}{b} \sim \frac{c}{d}$$

iff $ad = bc$. This is an equivalence relation. For example, $\frac{a}{b} \sim \frac{a}{b}$ follows from $ab = ba$. The equivalence classes are denoted by $[\frac{a}{b}]$.

(II) Multiplication of equivalence classes. We define

$$\left[\frac{a}{b}\right]\left[\frac{c}{d}\right] := \left[\frac{ac}{bd}\right].$$

The point is that this definition does not depend on the choice of the representatives of the equivalence classes. To see this, we have to show that

$$\left[\frac{a}{b}\right] = \left[\frac{a'}{b'}\right] \quad \text{and} \quad \left[\frac{c}{d}\right] = \left[\frac{c'}{d'}\right] \quad \text{imply} \quad \left[\frac{a}{b}\right]\left[\frac{c}{d}\right] = \left[\frac{a'}{b'}\right]\left[\frac{c'}{d'}\right].$$

In fact, if $ab' = ba', cd' = c'd$, then $ab'cd' = ba'c'd$. Hence $(ac)(b'd') = (bd)(a'c')$.

(III) Addition: We define

$$\left[\frac{a}{b}\right] + \left[\frac{c}{d}\right] := \left[\frac{ad + bc}{bd}\right].$$

Again one shows that this definition does not depend on the choice of the representatives. Here, we need that $b \neq 0$ and $c \neq 0$ imply $bc \neq 0$. This follows from our assumption that the ring R has no zero divisors. Furthermore, one checks that the

[12] We assume that R is not trivial, that is, it has at least one non-zero element. But, we do not assume that R has a unit element.

equivalence classes form a ring with the zero element $\mathbf{0} := [\frac{0}{b}]$ and the unit element $\mathbf{1} := [\frac{b}{b}]$.

(IV) Invertibility: If $a \neq 0$ and $b \neq 0$, then

$$\left[\frac{a}{b}\right]\left[\frac{b}{a}\right] = \left[\frac{ab}{ab}\right] = \mathbf{1}.$$

Consequently, the equivalence classes form a field F.

(V) Identification of the ring elements: For all nonzero elements b, b' of R, we have

$$\frac{ab}{b} \sim \frac{ab'}{b'},$$

since $(ab)b' = b(ab')$. Hence $\left[\frac{ab}{b}\right] = \left[\frac{ab'}{b'}\right]$. For each $a \in R$, define

$$\chi(a) := \left[\frac{ab}{b}\right]. \tag{4.9}$$

The map $\chi : R \to F$ does not depend on the choice of b. Thus, χ is well-defined. One shows easily that χ respects addition and multiplication. Thus, χ is a ring morphism. In addition, the map χ is injective. In fact, if

$$\left[\frac{ab}{b}\right] = \left[\frac{a'b}{b}\right],$$

then $ab^2 = a'b^2$, and hence $(a - a')b^2 = 0$. Since $b \neq 0$, and the ring R has no zero divisors, we get $b^2 \neq 0$, and hence $a - a' = 0$. Summarizing, the map $\chi : R \to \chi(R)$ is a ring isomorphism. If we identify the subring $\chi(R)$ of F with R, then the field F is an extension of R.

Ad (ii). Suppose that $\mathcal{F}$ is another field extension of R. Define

$$\varrho\Big(\Big[\frac{a}{b}\Big]\Big) := ab^{-1}.$$

This map does not depend on the representatives. One also shows that the map $\varrho : F \to \mathcal{F}$ is an injective ring morphism. Thus, F is isomorphic to $\varrho(\mathcal{F})$.

Ad (iii) Suppose that, say, R is a real algebra. For each real number α, we define

$$\alpha\left[\frac{a}{b}\right] := \left[\frac{\alpha a}{b}\right].$$

This map does not depend on the choice of the representatives. This way, the field F becomes a real algebra. □

4.1.4 Linear Quotient Spaces

Let X be a linear subspace of the linear space Z over $\mathbb{K}$. For $u, v \in Z$, we write

$$u \sim v \quad \text{iff} \quad u - v \in X.$$

This is an equivalence relation. The equivalence classes have the form

$$[u] = u + X = \{u + x : \ x \in X\}.$$

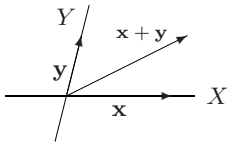


Fig. 4.1. Direct sum

Intuitively, if $Z = \mathbb{R}^3$ and X is a one-dimensional (resp. 2-dimensional) subspace of Z, then $u + X$ is a straight line (resp. plane) passing through the point u and being parallel to the straight line (resp. plane) X. For all $u, v \in Z$ and $\alpha, \beta \in \mathbb{K}$, we set

$$\alpha[u] + \beta[v] := [\alpha u + \beta v].$$

This definition does not depend on the choice of the representatives. This follows from $\alpha X + \beta X = \{\alpha x + \beta y : x, y \in X\} = X$, and hence

$$\alpha(u + X) + \beta(v + X) = (\alpha u + \beta v) + X.$$

This way, the quotient space $Z/\sim$ becomes a linear space over $\mathbb{K}$ which is denoted by Z/X. From the geometric point of view, the quotient space Z/X operates with straight lines or planes (i.e., linear manifolds).

Projection operator. Let X and Y be linear subspaces of the linear space Z over $\mathbb{K}$. Suppose that each element z of Z allows the following unique sum representation

$$\boxed{z = x + y, \qquad x \in X,\ y \in Y.}$$

Define

$$Pz := x \qquad \text{for all} \quad z \in Z.$$

The operator $P : Z \to X$ has the typical property that it is linear and $P^2 = P$ (Fig. 4.1). Precisely such operators are called projection operators. The operator

$$I - P : Z \to Y$$

is also a projection operator. In fact, the operator $I - P$ is linear, and $(I - P)^2$ is equal to $I - 2P + P^2 = I - P$. Hence $(I - P)^2 = I - P$. We have

$$Z = X \oplus Y, \qquad P(Z) = X,\ \ (I - P)(Z) = Y.$$

The mapping $y \mapsto [y]$ yields the following isomorphism.

Proposition 4.9 *The complementary linear subspace Y is linear isomorphic to the quotient space Z/X.*

Thus, for each linear subspace X of the linear space Z over $\mathbb{K}$, we get the direct sum decomposition

$$Z = X \oplus Z/X.$$

Note that if $Z = X \oplus Y$, then the quotient space Z/X is uniquely determined by Z and X, in contrast to the space Y. The dimension of the quotient space Z/X is called the codimension of the linear subspace X with respect to the linear space Z. We write

$$\operatorname{codim}(X) := \dim(Z/X).$$

The dimension of a linear space is an absolute invariant under linear isomorphisms. In contrast to this, the codimension of a linear space is only a relative invariant. If the dimension of Z is finite, then

$$\boxed{\operatorname{codim}(X) = \dim(Z) - \dim(X).}$$

For example, in the 3-dimensional space of our intuition, a plane has the dimension 2 and the codimension 1. The morphism theorem for linear morphisms (i.e., linear operators) $L : X \to Y$ between the two linear spaces X and Y over $\mathbb{K}$ reads as follows.

Theorem 4.10 *If $L : X \to Y$ is a linear morphism, then we have the linear isomorphism $X/\ker(L) \simeq \operatorname{im}(L)$.*

Here, the kernel $\ker(L) := L^{-1}(0)$ is a linear subspace of X. The proof proceeds as for rings on page 184.

4.1.5 Ideals and Quotient Algebras

Let A be an algebra over $\mathbb{K}$. For subsets B and C of A, we define

$$B + C := \{b + c : \ b \in B, \ c \in C\}, \qquad BC := \{bc : \ b \in B, \ c \in C\}.$$

Using this convenient notation, we obtain the following:

- The subset B of A is a linear subspace iff $\alpha B + \beta B = B$ for all $\alpha, \beta \in \mathbb{K}$.
- The subset J of A is called a two-sided ideal iff $JA = AJ = J$.

Let J be a two-sided ideal of A. Since the algebra A is a linear space, the quotient space A/J is a linear space. Additionally, A/J becomes an algebra over $\mathbb{K}$ if we introduce the following multiplication

$$[u][v] := [uv].$$

This definition does not depend on the choice of representatives. In fact,

$$(u + J)(v + J) = uv + (uJ + vJ + JJ) = uv + J.$$

The morphism theorem for algebra morphisms $\chi : A \to B$ between the two algebras A and B over $\mathbb{K}$ reads as follows.

Theorem 4.11 *If $\chi : A \to B$ is an algebra morphism, then we have the algebra isomorphism $A/\ker(\chi) \simeq \operatorname{im}(\chi)$.*

Here, the kernel $\ker(\chi) := \chi^{-1}(0)$ is a two-sided ideal of A. The proof proceeds as for rings on page 184.

Further constructions for linear spaces. In addition to the preceding material, there exist the following constructions for linear spaces X, Y, X_α:

- the inductive (or direct) limit: $\lim \operatorname{ind}_{\alpha \in \mathcal{A}} X_\alpha$ (Sect. 4.5.5);
- the projective (or inverse) limit: $\lim \operatorname{proj}_{\alpha \in \mathcal{A}} X_\alpha$ (Sect. 4.5.5);
- the K-ring $K(\mathrm{Vect}_{\mathbb{K}})$ generated by the semi-ring of finite-dimensional vector spaces over $\mathbb{K}$ (Sect. 4.4.9);
- the K-ring $K_{\mathbb{K}}(X)$ generated by the semi-ring of vector bundles of finite rank on the topological space X (Sect. 4.4.9);
- the tensor product $X \otimes Y$, and the tensor algebra $\bigotimes(X)$ (Vol. III);

- the algebra $A(X)$ of antisymmetric multilinear functionals, and the Grassmann (or exterior) algebra $\bigwedge(X)$ (Vol. III);
- the algebra $S(X)$ of symmetric multilinear functionals, and the symmetric algebra $\bigodot(X)$ (Vol. III);
- the Clifford (or inner) algebra $\bigvee(X)$ (Vol. III).

4.2 Superfunctions and the Heaviside Calculus in Electrical Engineering

> The historical experience of mathematicians shows that successful formal approaches invented by physicists can be rigorously justified, possibly, after large time delay. For example, this concerns Heaviside's calculus in electrical engineering (Laplace transform and Mikusiński's operational calculus), Dirac's delta function in quantum mechanics (Laurent Schwartz's theory of distributions), Dirac's operator calculus (Gelfand triplets and distributions), and Leibniz's infinitesimals in calculus (Robinson's non-standard analysis). There is no reason why there should not emerge a rigorous justification of quantum field theory in the future.
>
> Folklore

> In this book, the operators of the Heaviside calculus in electrical engineering are represented by (abstract) fractions of continuous functions.[13]
>
> Jan Mikusiński, 1959

The convolution algebra. Let R denote the set of all continuous functions

$$f : [0, \infty[\to \mathbb{C}.$$

For $f, g \in R$ and $\alpha, \beta \in \mathbb{C}$, we define the usual linear combination $\alpha f + \beta g$. As product $f * g$, we choose the convolution

$$(f * g)(t) := \int_0^\infty f(t)g(t-\tau)d\tau \qquad \text{for all} \quad t \geq 0.$$

For example, set $l(t) := 1$ for all $t \geq 1$. Let $f \in R$. Then

$$(l * f)(t) = \int_0^t f(\tau)d\tau \qquad \text{for all} \quad t \geq 0. \tag{4.10}$$

If the function $f : [0, \infty[\to \mathbb{R}$ is continuous and continuously differentiable, then the fundamental theorem of calculus tells us that

$$(l * f')(t) = \int_0^t f'(\tau)d\tau = f(t) - f(+0) \qquad \text{for all} \quad t \geq 0.$$

In terms of the convolution algebra R, this means $l * f' = f - f(+0)l$. Elementary properties of the convolution product tell us that R is a commutative ring and a complex algebra. The following theorem due to Titchmarsh (1899–1963) formulates a non-trivial property of the convolution product.

Theorem 4.12 *The convolution algebra R has no zero divisors.*

[13] J. Mikusiński, Operational Calculus, Pergamon Press, Oxford, 1959.

Explicitly, this means the following. Let $f, g : [0, \infty[\to \mathbb{C}$ be two continuous functions with $(f * g)(t) = 0$ for all $t \in \mathbb{R}$. Then at least one of the functions f or g vanishes identically. The elegant proof can be found in K. Yosida, Functional Analysis, Sect. VI.5, Springer, New York, 1995. The Titchmarsh theorem is one of the jewels in harmonic analysis.

The field of superfunctions. By Theorem 4.8 on page 187, the convolution algebra R can be extended to a field F which is uniquely determined (up to isomorphism). The field F is also a complex algebra. Here, F is called the Mikunsiński field. To simplify notation, we write fg instead of $f * g$. The elements of F are called superfunctions. In the sense of (F1)–(F4) on page 186, the elements of F are fractions of the form

$$\boxed{\frac{f}{g}, \qquad f, g \in R, \quad g \neq 0.}$$

Observe the crucial fact that fractions refer to the product on the field F, but not to the division of continuous functions in the usual sense. The unit element $\mathbf{1}$ of the field F is given by

$$\mathbf{1} = \frac{f}{f} \qquad \text{for all} \quad f \in R,\ f \neq 0.$$

Let us consider some examples of superfunctions.

(i) Integration: Choose the element l of R defined by $l(t) := 1$ for all $t \in \mathbb{R}$. For all $f \in R$, set $g := lf$ (in the sense of the multiplication on the field F). By (4.10), we have $g \in R$ and $g(t) = \int_0^t f(\tau) d\tau$ for all $t \geq 0$.

(ii) The unit element $\mathbf{1}$. Suppose that there exists a unit element in the convolution ring R. We denote this unit element by δ. It follows from $f * \delta = f$ for all $f \in R$ that

$$\int_0^t f(\tau)\delta(t - \tau) d\tau = f(t) \qquad \text{for all} \quad t \geq 0.$$

However, there is no continuous function $\delta : [0, \infty[\to \mathbb{C}$ which has this property. Thus, the convolution ring R has *no* unit element. But the Minkusiński field F has the unit element $\mathbf{1}$. This superfunction can be regarded as the rigorous version of the Dirac delta function in the Mikusiński setting. In particular, note that the element l of R is different from $\mathbf{1}$. We have $\mathbf{1} = \frac{l}{l}$.

(iii) Differentiation: Set $s := l^{-1}$. For all continuous and continuously differentiable functions $f : [0, \infty[\to \mathbb{C}$, we have the following key relation for the Heaviside calculus:

$$\boxed{f' = sf - f(+0)\mathbf{1}.} \tag{4.11}$$

In fact, it follows from $lf' = f - f(+0)l$ and $sl = \mathbf{1}$ that

$$f' = (sl)f' = sf - f(+0)\frac{l}{l} = sf - f(+0)\mathbf{1}.$$

(iv) Exponential function: Fix the complex number α. Set $h(t) := \mathrm{e}^{\alpha t}$ for all $t \geq 0$. Then, in terms of the Mikusiński field F,

$$h = \frac{\mathbf{1}}{s - \alpha\mathbf{1}}.$$

In fact, $h' = \alpha h$. Hence $\alpha h = sh - h(+0)\mathbf{1} = sh - \mathbf{1}$, by (iii). Thus, $(s - \alpha\mathbf{1})h = \mathbf{1}$.

Application to the Heaviside calculus. As a simple example, consider the initial-value problem

$$x'(t) - \alpha x(t) = f(t), \quad t \geq 0, \quad x(+0) = x_0. \tag{4.12}$$

We are given the continuous function $f : [0, \infty[\to \mathbb{C}$, and the complex numbers α, x_0. In the Mikusiński field F, equation (4.12) reads as

$$sx - x_0\mathbf{1} - \alpha x = f.$$

Hence $(s - \alpha\mathbf{1})x = x_0\mathbf{1} + f$. By (iv),

$$x = \frac{x_0\mathbf{1}}{s - \alpha\mathbf{1}} + \frac{f}{s - \alpha\mathbf{1}} = x_0 h + hf.$$

Recall that the product hf in the field F stands for the convolution product $h * f$. Therefore, the solution of our problem (4.12) is given by

$$x = x_0 h + h * f.$$

Explicitly, noting that $h * f = f * h$, the solution of (4.12) reads as

$$x(t) = \mathrm{e}^{\alpha t} x_0 + \int_0^t \mathrm{e}^{\alpha(t-\tau)} f(\tau)\, d\tau \qquad \text{for all} \quad t \geq 0.$$

This justifies rigorously the symbolic Heaviside method considered in (1.28) on page 30. The approach can be simplified by cancelling the (redundant) unit element $\mathbf{1}$ of the field F. Then

$$x = \frac{x_0}{s - \alpha} + \frac{f}{s - \alpha}.$$

Many applications in engineering can be found in the classical monograph by

J. Mikusiński, Operational Calculus, Pergamon Press, Oxford, 1959.

The Polish engineer Jan Mikusiński founded this approach in the 1950s. Roughly speaking, Mikusiński's method is equivalent to the use of the Laplace transform (see Sect. 1.31 on page 32). We also refer to: K. Yosida, Operational Calculus: A Theory of Hyperfunctions, Springer, New York, 1984. In 1959 Sato introduced a class of generalized functions which he called hyperfunctions:

M. Sato, Theory of hyperfunctions I, II, J. Fac. Sci. Univ. Tokyo **8** (1959), 139–193; (1969), 487–536.

P. Schapira, Mikio Sato – a visionary of mathematics, Notices Amer. Math. Soc. **54**(2) (2007), 243–245.

Roughly speaking, distributions in the sense of Laurent Schwartz are dual objects to smooth functions. Similarly, roughly speaking, Sato's hyperfunctions are dual objects to analytic functions. The general theory together with applications to partial differential equations is investigated in:

L. Hörmander, The Analysis of Linear Partial Differential Operators I, Chap. 9, Springer, New York, 1983.

We also refer to M. Sato, T. Miwa, and M. Jimbo, Holonomic quantum fields, Parts I–V, Publ. Res. Inst. Math. Sci. **14** (1978), 223–267; **15** (1979), 201–278; 577–629; 871–972; **16** (1980), 531–535.

4.3 Equivalence Classes in Geometry

Geometry is the invariant theory of groups of transformations.
Felix Klein
Erlangen program 1872

4.3.1 The Basic Idea of Geometry Epitomized by Klein's Erlangen Program

The geometry known in ancient times was Euclidean geometry, and it dominated mathematics for over 2000 years. The famous question as to the existence of non–Euclidean geometries led in the nineteenth century to the description of a series of different geometries. This being established, it was natural to consider the classification of possible geometries. Felix Klein (1849–1925) solved this problem and showed in 1872 with his Erlangen program that geometries can be conveniently classified by means of group theory. A geometry requires a group G of transformations. Every property or quantity remaining invariant under the action of the group G is a property of the associated geometry, which is therefore also referred to as a G-geometry. For example, the Euclidean geometry corresponds to invariants under the Euclidean group of motions, which consists of translations and rotations. In particular, the distance between two points is a property of Euclidean geometry. For the modern version of Klein's Erlangen program, we refer to:

R. Sharpe, Differential Geometry: Cartan's Generalization of Klein's Erlangen Program, Springer, New York, 1997.

4.3.2 Symmetry Spaces, Orbit Spaces, and Homogeneous Spaces

A G-space (X, G) describes a geometry on the space X which possesses the symmetry group G.

Folklore

In gauge theory, orbit spaces play a fundamental role. For example, in the Standard Model in particle physics, the orbits under the action of the gauge group represent the same physical state. Since the Feynman functional integral has to be taken over physical states, this integral becomes an integral over an orbit space of the gauge group. This is the idea of the method of Faddeev–Popov quantization in gauge theory (see Chap. 16 of Vol. I). At this point, we want to discuss same basic ideas of orbit spaces in geometry and theirs relations to quantum states and projective spaces. Let $n = 0, 1, 2, \ldots$ To fix the terminology, the symbol

$$\mathbb{S}^n := \{x \in \mathbb{R}^{n+1} : \ x_1^2 + x_2^2 + \ldots + x_{n+1}^2 = 1\}$$

denotes the n-dimensional unit sphere. Observe that this sphere is the boundary of the $(n+1)$-dimensional closed unit ball centered at the origin:

$$\mathbb{B}^{n+1} := \{x \in \mathbb{R}^{n+1} : \ x_1^2 + x_2^2 + \ldots + x_{n+1}^2 \leq 1\}.$$

Let $R > 0$, and let $a \in \mathbb{R}^{n+1}$, where $a = (a_1, \ldots, a_{n+1})$. The set

$$\mathbb{B}_R^{n+1}(a) := \{x \in \mathbb{R}^{n+1} : (x_1 - a_1)^2 + \ldots + (x_{n+1} - a_{n+1})^2 \leq R^2\}$$

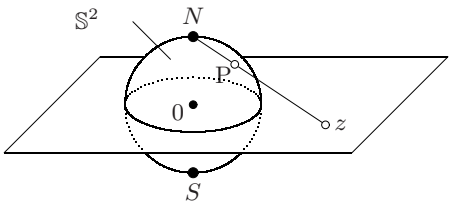


Fig. 4.2. Riemann sphere

is called the $(n+1)$-dimensional closed ball of radius R centered at the point a. The interior of $\mathbb{B}_R^{n+1}(a)$ is called the $(n+1)$-dimensional open ball of radius R centered at the point a. Explicitly,

$$\text{int } \mathbb{B}_R^{n+1}(a) = \{x \in \mathbb{R}^{n+1} : (x_1 - a_1)^2 + \ldots + (x_{n+1} - a_{n+1})^2 < R^2\}.$$

The boundary $\partial\mathbb{B}_R^{n+1}(a)$ of the closed ball $\mathbb{B}_R^{n+1}(a)$ is given by

$$\partial\mathbb{B}_R^{n+1}(a) = \{x \in \mathbb{R}^{n+1} : (x_1 - a_1)^2 + \ldots + (x_{n+1} - a_{n+1})^2 = R^2\}.$$

This is an n-dimensional sphere of radius R centered at the point a. We write $\mathbb{S}_R^n$ instead of $\partial\mathbb{B}_R^{n+1}(0)$. Recall that the set $\mathbb{C}$ of all complex numbers is called the Gaussian plane. If we add the point ∞, then we get the closed Gaussian plane

$$\overline{\mathbb{C}} := \mathbb{C} \cup \{\infty\}.$$

By stereographic projection, the closed Gaussian plane $\overline{\mathbb{C}}$ is bijective to the Riemann sphere $\mathbb{S}^2$ (Fig. 4.2). This stereographic projection preserves angles; thus, it is a conformal map. We will show in Sect. 4.3.5 on page 203 that the Riemann sphere is also bijective to the projective complex line $\mathbb{P}^1_{\mathbb{C}}$. The Riemann sphere $\mathbb{S}^2$ is an arcwise connected 1-dimensional complex manifold (i.e., it is a Riemann surface). To prove this, we have to introduce local complex coordinates. Note that the stereographic projection pictured in Fig. 4.2 sends the North Pole N (resp. the South Pole S) to the point ∞ (resp. to the origin 0) of the extended complex plane $\overline{\mathbb{C}}$.

- The points P of the punctured sphere $\mathbb{S}^2 \setminus \{N\}$ can be described by the local complex coordinates $z(P)$ via stereographic projection.
- Now consider the points P on the punctured sphere $\mathbb{S}^2 \setminus \{S\}$. If $P \neq N$, then stereographic projection assigns the complex number $z(P)$ to the point P. Finally, we assign the local complex coordinate $\zeta(P)$ to the point P by setting

$$\zeta(P) := \begin{cases} \frac{1}{z(P)} & \text{if} \quad P \neq N, \\ 0 & \text{if} \quad P = N. \end{cases}$$

- For a point $P \in \mathbb{S}^2 \setminus \{N, S\}$, the change of local coordinates from $z(P)$ to $\zeta(P)$ is described by the holomorphic map $z \mapsto \frac{1}{z}$ on $\mathbb{C} \setminus \{0\}$.

The space $\mathbb{C}/U(1)$ as the prototype of an orbit space. Let us consider the group $U(1) := \{e^{i\varphi} : \varphi \in \mathbb{R}\}$. Fix $g \in U(1)$, and define

$$\varrho(g)z := gz \qquad \text{for all} \quad z \in \mathbb{C}.$$

If $g, h \in U(1)$, then $\varrho(h)(\varrho(g)z) = \varrho(hg)z$ for all $z \in \mathbb{C}$. The map

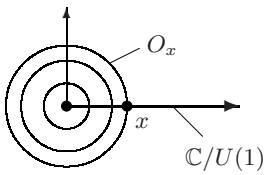


Fig. 4.3. Action of the rotation group on the plane

$$\varrho(g) : \mathbb{C} \to \mathbb{C}$$

represents a rotation of the Gaussian plane $\mathbb{C}$ about the origin $z = 0$. If $g = \mathrm{e}^{\mathrm{i}\varphi}$, then φ is the rotation angle. We write

$$w \sim z \mod U(1)$$

iff there exists an element g of the group $U(1)$ such that $w = \varrho(g)z$. This means that the point z can be moved to the point w by a rotation. This is an equivalence relation because of the group property of $U(1)$. The equivalence classes $[z]$ are called orbits of the action of the group $U(1)$ on the Gaussian plane $\mathbb{C}$. The set of these orbits is denoted by $\mathbb{C}/U(1)$, and we call this the orbit space. Intuitively, the orbits are concentric circles about the origin (Fig. 4.3). Obviously, the orbits decompose the Gaussian plane into pairwise disjoint sets. Let us reformulate this in the language of fiber bundles. To this end, we parametrize the orbit space $\mathbb{C}/U(1)$. Let O_x denote that orbit which intersects the real line at the point $x \geq 0$. Then

$$\mathbb{C} = \bigcup_{x \in [0,\infty[} O_x .$$

For each point $z \in \mathbb{C}$, there is precisely one orbit O_x which passes through z. We define $\pi(z) := x$. This way, we get the surjective map

$$\pi : \mathbb{C} \to [0, \infty[.$$

The pre-images of π are called fibers. Explicitly, $\pi^{-1}(x) = O_x$, that is, the fibers coincide with the orbits. The family of fibers

$$\{O_x\}_{x \in [0,\infty[}$$

is called an abstract fiber bundle. Obviously, $\mathbb{C}/U(1) \simeq [0, \infty[$, in the sense of a bijection. Thus, the orbit space $\mathbb{C}/U(1)$ can be identified with the interval $[0, \infty[$, which is not a manifold, but only a manifold with boundary.

Typically, orbit spaces are not manifolds, since they have singularities.

Klein spaces. Let X be a set, and let G be a group. The set of all bijections

$$A : X \to X$$

forms a group which is called the automorphism group of X. This group is denoted by $\mathrm{Aut}(X)$. The ordered pair

$$(X, G)$$

is called a Klein space (or G-space) iff there exists a group morphism

$$\varrho : G \to \mathrm{Aut}(X).$$

The map ϱ is called a representation of the group G by a transformation group on the space X. Explicitly, this means that, for all $g, h \in G$, the map $\varrho(g) : X \to X$ is a bijection with

$$\varrho(hg) = \varrho(h)\varrho(g).$$

The set $\{\varrho(g) : g \in G\}$ forms a subgroup of $\mathrm{Aut}(G)$ which is called a transformation group of X induced by the symmetry group G. We also say that the symmetry group G acts on the space X. For two points $x, y \in X$, we write

$$x \sim y \mod G$$

iff there exists an element g of the symmetry group G such that

$$y = \varrho(g)x.$$

This is an equivalence relation on the space X. The equivalence classes $[x]$ are called orbits of the action of the group G on the space X. The set of all orbits is denoted by

$$X/G.$$

This is also called the orbit space induced by the action of the group G on the space X. The stabilizer $\mathrm{stab}(x_0)$ of the point x_0 on the space X is defined by

$$\mathrm{stab}(x_0) := \{g \in G : \varrho(g)x_0 = x_0\}.$$

This is a subgroup of G. In other words, an element g of the symmetry group G belongs to the stabilizer $\mathrm{stab}(x_0)$ iff the point x_0 of the space X remains fixed under the action of g. The stabilizer $\mathrm{stab}(x_0)$ is called trivial iff it contains only the unit element of G. In geometry, one uses the following terminology:

- The action of the group G on the space X is called *transitive* iff there is only one orbit.
- The action of G on X is called *effective* iff the map $\varrho : G \to \mathrm{Aut}(X)$ is injective.
- The action of G on X is called *faithful* iff the map $\varrho : G \to \mathrm{Aut}(X)$ is bijective.
- The action of G on X is called *free* iff the stabilizer $\mathrm{stab}(x_0)$ is trivial for all points $x_0 \in X$.

In the case of Fig. 4.3, the stabilizer $\mathrm{stab}(x_0)$ is equal to $U(1)$ if $x_0 = 0$, and it is trivial if $x_0 \neq 0$. The action of the rotation group $U(1)$ on the Gaussian plane $\mathbb{C}$ is faithful, but neither transitive nor free. In contrast to this, for all $a \in \mathbb{C}$, the action of the translation group $z \mapsto z + a$ on $\mathbb{C}$ is transitive, faithful, and free.

Morphisms of Klein spaces. Let (X, G) and (Y, G) be Klein spaces with the corresponding representations

$$\varrho : G \to \mathrm{Aut}(X) \quad \text{and} \quad \sigma : G \to \mathrm{Aut}(Y)$$

of the group G on X and Y, respectively. By definition, a morphism (resp. isomorphism) from (X, G) to (Y, G) is a map (resp. a bijection) $\mu : X \to Y$ such that the following diagram is commutative for all elements g of the symmetry group G:

$$\begin{array}{ccc} X & \xrightarrow{\mu} & Y \\ {\scriptstyle \varrho(g)}\downarrow & & \downarrow{\scriptstyle \sigma(g)} \\ X & \xrightarrow{\mu} & Y. \end{array} \tag{4.13}$$

Homogeneous spaces. By definition, a homogeneous space X under the action of the symmetry group G is a Klein space (X, G) which has precisely one orbit (i.e., the action is transitive). Let us consider some examples.

- The unit circle $\mathbb{S}^1$ is a homogeneous space under the action of the rotation group in the plane.
- The 2-dimensional unit sphere $\mathbb{S}^2$ is a homogeneous space under the action of the rotation group. In fact, each given point of $\mathbb{S}^2$ can be moved to an arbitrary point of $\mathbb{S}^2$ by using a rotation.
- The Euclidean plane (resp. the 3-dimensional Euclidean space) is homogeneous under the action of the translation group.
- Let S be a subgroup of the group G. We want to show that the orbit space G/S is homogeneous under the action of the group G.
  To this end, we write $g \sim h$ iff $g^{-1}h \in S$. This is an equivalence relation. The corresponding equivalence classes $[g]$ form the quotient space $G/\sim$. Explicitly,

$$[g] = \{gs : s \in S\} \qquad \text{for all} \quad g \in G.$$

  Briefly, $[g] = gS$. Fix $s \in S$, and define the map $\varrho(s) : G \to G$ by setting

$$\varrho(s)g := sg \qquad \text{for all} \quad g \in G.$$

  This way, the subgroup S acts on the group G. The orbits of the Klein space (G, S) are precisely the equivalence classes $[g]$. Thus,

$$G/S = G/\sim .$$

  Now fix $h \in G$, and set

$$\chi(h)([g]) := [hg]$$

  for all equivalence classes $[g]$. The map $\chi(h)$ does not depend on the choice of the representatives. Therefore, the group G acts on G/S by means of the maps $\chi(h) : G/S \to G/S$. This way, we get the Klein space $(G/S, G)$. The action of G on G/S is transitive. In fact, for two equivalence classes $[g], [r]$, we have $\chi(rg^{-1})[g] = [r]$. Consequently, the Klein space $(G/S, G)$ has precisely one orbit, that is, the orbit space G/S is a homogeneous space.

Classification of Klein spaces. Let (X, G) be a Klein space. Then X is the disjoint union of the orbits O in X. For any orbit O, there is a Klein space (O, G), and O is a homogeneous space under the action of G. Therefore, it is sufficient to classify the Klein spaces (X, G) where the space X is homogeneous.

Theorem 4.13 *We are given the homogeneous space X with the symmetry group G. Then the following hold:*

(i) *If we fix a point x_0 in X, then the Klein space (X, G) is isomorphic to the Klein space $(G/\mathrm{stab}(x_0), G)$.*

(ii) *If the stabilizer $\mathrm{stab}(x_0)$ is a normal subgroup of the symmetry group G, then the quotient space $G/\mathrm{stab}(x_0)$ is a group.*

(iii) *If S and T are subgroups of G, then the Klein spaces $(G/S, G)$ and $(G/T, G)$ are isomorphic iff S is conjugate to T.*

Recall that, by definition, the subgroup S is conjugate to the subgroup T iff there exists an element g of G such that $S = \{gtg^{-1} : t \in T\}$. For the proof of the theorem, we refer to Problem 4.1.

4.3.3 The Space of Quantum States

> In contrast to the underlying complex Hilbert space, the space of quantum states has a nontrivial topological structure.
>
> Folklore

In the following sections of this chapter, the symbol $A \simeq B$ means that there exists a bijective map from the set A onto the set B. In other words, the sets A and B are equivalent, in the sense of set theory.

Complex Hilbert space. Let X be a complex Hilbert space. The symbol $X^\times$ denotes the nonzero elements of X.[14] For $\psi, \varphi \in X^\times$, we write

$$\psi \sim \varphi$$

iff there exists a number $\lambda \in \mathbb{C}^\times$ such that $\psi = \lambda\varphi$. Obviously, this is an equivalence relation. The corresponding equivalence classes

$$[\psi] = \{\lambda\psi \in X : \lambda \in \mathbb{C}^\times\}$$

are called quantum states (or rays). The set of all quantum states $[\psi]$ of the Hilbert space X is denoted by $\mathbb{P}(X)$. Thus,

$$\boxed{\mathbb{P}(X) = X^\times/\sim .}$$

Equivalently, this is the orbit space $X^\times/\mathbb{C}^\times$ of the Hilbert space X under the action $\psi \mapsto \lambda\psi$ of the multiplicative group $\mathbb{C}^\times$. The elements $\psi \in X^\times$ are called representatives of quantum states. The relation of $\mathbb{P}(X)$ to projective geometry will be studied below. Let $A : D(A) \to X$ be a self-adjoint operator on X (e.g., a complex self-adjoint $(n \times n)$-matrix acting on $X = \mathbb{C}^n$). For $\psi \in X^\times \cap D(A)$, we define

$$\overline{\psi} := \frac{\langle\psi|A\psi\rangle}{\langle\psi|\psi\rangle}.$$

This complex number only depends on the quantum state $[\psi]$. We call $\overline{\psi}$ the mean value of the observable A (e.g., energy) measured in the quantum state $[\psi]$. Let us introduce the following notions:

- $\mathbb{S}(X) := \{\psi \in X : ||\psi|| = 1\}$ (unit sphere of X);
- $\mathbb{B}(X) := \{\psi \in X : ||\psi|| \leq 1\}$ (closed unit ball in X);
- $\mathrm{int}\big(\mathbb{B}(X)\big) := \{\psi \in X : ||\psi|| < 1\}$ (open unit ball in X);[15]
- $\mathbb{S}(X)/\mathrm{U}(1)$ (the orbit space of the sphere $\mathbb{S}(X)$ under the action $\psi \mapsto \lambda\psi$ of the group $U(1)$ of complex numbers λ with $|\lambda| = 1$);
- $\mathbb{G}_m(X)$ is the space of all m-dimensional linear subspaces of the Hilbert space X (m-Grassmann space of X).[16]

We have the following bijections:

$$\boxed{\mathbb{P}(X) \simeq X^\times/\mathbb{C}^\times \simeq \mathbb{G}_1(X) \simeq \mathbb{S}(X)/U(1).} \tag{4.14}$$

In fact, the maps

$$[\psi] \mapsto [\psi] \cup \{0\}, \qquad [\psi] \mapsto [\psi] \cap \mathbb{S}(X)$$

[14] In particular, the symbol $\mathbb{R}^\times$ (resp. $\mathbb{C}^\times$) denotes the set of nonzero real (resp. complex) numbers.

[15] If $X := \mathbb{R}^n$, then we use the symbols $\mathbb{S}^n := \mathbb{S}(\mathbb{R}^n)$ and $\mathbb{B}^n(\mathbb{S}) := \mathbb{B}(\mathbb{R}^n)$.

[16] Grassmann (1809–1877).

yield the bijections $\mathbb{P}(X) \simeq \mathbb{G}_1(X)$ and $\mathbb{P}(X) \simeq \mathbb{S}(X)/U(1)$, respectively.

Real Hilbert space. Let X be a real Hilbert space. Then we introduce the same symbols as in the complex case above by using the following replacements:

$$\mathbb{C}^\times \Rightarrow \mathbb{R}^\times, \qquad U(1) \mapsto \{1, -1\}.$$

In particular, we obtain the following bijections:

$$\mathbb{P}(X) \simeq X^\times/\mathbb{R}^\times \simeq \mathbb{G}_1(X) \simeq \mathbb{S}(X)/\{1, -1\}.$$

Here, the Grassmann space $\mathbb{G}_1(X)$ is the set of all 1-dimensional linear subspaces of the real Hilbert space X, and $\mathbb{S}(X)/\{1, -1\}$ is the set of all antipodal pairs $(\psi, -\psi)$ of the unit sphere $\mathbb{S}(X)$.

4.3.4 Real Projective Spaces

> Algebraic geometry is undoubtedly the area of mathematics where the deviation is greatest between the intuitive ideas forming its starting point and the abstract and complex concepts at the foundation of modern research... The history of algebraic geometry has been divided into seven epochs:
> (i) 400 B.C.–1630 A.D.: Prehistory (theory of conics by Appolonius of Perga).
> (ii) 1630–1795: Exploration of plane curves (Descartes, Newton, Leibniz, Euler, Maclaurin, Bézout).
> (iii) 1795–1850: The golden age of projective geometry (Poncelet, Möbius, Plücker).
> (iv) 1850–1866: Riemann and conformal (birational) geometry.
> (v) 1866–1920: Development and chaos.
> (vi) 1920–1950: New structures in algebraic geometry (Poincaré, Élie Cartan, Hodge, de Rham, Lefschetz, Kähler, Weil, Kodaira).
> (vii) 1950ff: Sheaves (Leray, Henri Cartan, Serre), the Riemann–Roch–Hirzebruch theorem (Hirzebruch), and schemes (Grothendieck).
> The fourth epoch is without any doubt the most important of all in the history of algebraic geometry to this day. It is entirely stamped by the work of one man, one of the greatest mathematicians who ever lived, and also one of those who have had, most profoundly, the perception (or divination) of the essential unity of mathematics.[17]
>
> Jean Dieudonné, 1985

The real projective line $\mathbb{P}^1$. By definition, the points of $\mathbb{P}^1$ are the antipodal pairs $\{P_+, P_-\}$ of the unit sphere $\mathbb{S}^1$. To each point P on the real line $\mathbb{R}^1$, we assign the antipodal pair $\{P_+, P_-\}$ as pictured in Fig. 4.4. In particular, the origin $x = 0$ on the real line corresponds to $\{(0, 1), (0, -1)\}$. The equatorial pair $\{(1, 0), (-1, 0)\}$ represents the unique infinite point of the real line. We have the bijection

$$\boxed{\mathbb{P}^1 \simeq \mathbb{S}^1.}$$

[17] J. Dieudonné, History of Algebraic Geometry, Birkhäuser, Boston, 1985 (reprinted with permission).
See also K. Maurin, The Riemann Legacy: Riemannian Ideas in Mathematics and Physics of the 20th Century, Kluwer, Dordrecht. 1997.

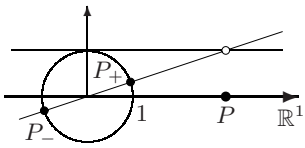


Fig. 4.4. Real projective line $\mathbb{P}^1$

In fact, $\mathbb{P}^1$ is bijective to the upper semi-circle where the two equatorial points $(1,0)$ and $(-1,0)$ are identified with each other. Gluing the two endpoints together, we get a circle. Consequently, the real projective line $\mathbb{P}^1$ can be equipped with such a topology that it is homeomorphic to the unit circle $\mathbb{S}^1$.

The real projective plane $\mathbb{P}^2$ and infinite points. In the Euclidean plane, two different straight lines do not always intersect. This is an imperfect situation, which prevents the formulation of an elegant duality between points and straight lines in Euclidean geometry.

> *In mathematics, one always tries to cure imperfect situations by introducing ideal elements.*

For example, the fact that the equation $x^2 + 1 = 0$ has no real solution led to the invention of the imaginary number $\sqrt{-1}$ in the 16th century. In geometry, one introduces infinite points. Let us restrict to the situation of a plane. By definition, an infinite point is the (non-oriented) direction of a straight line in the Euclidean plane. We say that two different straight lines intersect each other in an infinite point iff they have the same direction, that is, they are parallel. By definition, all the infinite points form the infinite straight line. The projective plane $\mathbb{P}^2$ is obtained from the Euclidean plane by adding the set of all infinite points. Furthermore, we add the infinite straight line to the set of all straight lines. For the projective plane, the following hold:

- Two different straight lines always uniquely determine one specific point (i.e., the intersection point).
- Two different points uniquely determine one specific straight line (i.e., the connecting straight line).

For example, a finite point P and an infinite point (i.e., a direction) P_∞ determine uniquely one specific straight line which passes through the point P and has the direction P_∞.

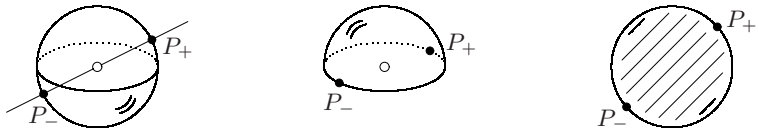
In analytical terms, we start with the two equations for the given straight lines:

$$y - x + 1 = 0, \qquad y - x + 2 = 0, \qquad (x, y) \in \mathbb{R}^2.$$

Using a fixed Cartesian coordinate system, we describe the Euclidean plane (resp. the 3-dimensional Euclidean space) by $\mathbb{R}^2$ (resp. $\mathbb{R}^3$). The basic trick of projective geometry is to pass from coordinates (x, y) in $\mathbb{R}^2$ to homogeneous coordinates (x, y, z) which are contained in $\mathbb{R}^3$. Explicitly, we use the replacement $x \Rightarrow x/z$, $y \Rightarrow y/z$ with the real nonzero number z. This yields the modified homogeneous equations

$$y - x + z = 0, \qquad y - x + 2z = 0, \qquad (x, y, z) \in \mathbb{R}^3, \quad (x, y, z) \neq (0, 0, 0).$$

The general solution (x, y, z) is given by the set $\{\lambda(1, 1, 0) : \lambda \in \mathbb{R}^\times\}$. By definition, this is a point of the projective plane $\mathbb{P}^2$. In general, for two nonzero tuples (x, y, z) and (x', y', z') in $\mathbb{R}^3$, we write



(a) (b) (c)

Fig. 4.5. The real projective plane $\mathbb{P}^2$ where P_+ is identified with P_-

$$(x, y, z) \sim (x', y', z')$$

iff there exists a nonzero real number λ with $(x, y, z) = \lambda(x', y', z')$. This is an equivalence relation. The equivalence classes

$$[(x, y, z)]$$

are called points of the projective plane $\mathbb{P}^2$. Each point of $\mathbb{P}^2$ has either the form $[(x, y, 1)]$ (finite point) or the form $[(x, y, 0)]$ with $x^2 + y^2 \neq 0$ (infinite point).

- The finite point $[(x, y, 1)]$ of $\mathbb{P}^2$ can be identified with the point (x, y) of $\mathbb{R}^2$.
- The infinite point $[(x, y, 0)]$ can be identified with a straight line in $\mathbb{R}^2$ which passes through the points $(0, 0)$ and (x, y).

Using the real Hilbert space $X := \mathbb{R}^3$ and the notation introduced in Sect. 4.3.3 on page 199, we have the following bijections:

$$\boxed{\mathbb{P}^2 \simeq (\mathbb{R}^3)^\times/\mathbb{R}^\times \simeq \mathbb{G}_1(\mathbb{R}^3) \simeq \mathbb{S}^2/\{1, -1\}.} \tag{4.15}$$

Let us discuss this. To begin with, note that the points of the real projective plane $\mathbb{P}^2$ can be identified with straight lines in $\mathbb{R}^3$ which pass through the origin. This yields the bijection $\mathbb{P}^2 \simeq \mathbb{G}_1(\mathbb{R}^3)$. Moreover, each of these straight lines intersects the 2-dimensional unit sphere $\mathbb{S}^2$ in precisely two points which are antipodal points (Fig. 4.5(a)). Therefore, the points of $\mathbb{P}^2$ can be identified with pairs of antipodal points of $\mathbb{S}^2$. This yields the bijection $\mathbb{P}^2 \simeq \mathbb{S}^2/\{1, -1\}$. Precisely the antipodal pairs of the equator of $\mathbb{S}^2$ correspond to infinite points of the real projective plane $\mathbb{P}^2$. In addition, we have the following bijections:

$$\mathbb{P}^2 \simeq (\mathbb{S}^2_+/\sim) \simeq (\mathbb{B}^1/\sim).$$

Here, $\mathbb{S}^2_+/\sim$ denotes the closed northern hemisphere of the sphere $\mathbb{S}^2$, where antipodal points P_+, P_- of the equator are identified with each other (Fig. 4.5(b)). Projection of $\mathbb{S}^2_+/\sim$ onto the Euclidean (x, y)-plane yields the set $\mathbb{B}^1/\sim$. Explicitly, $\mathbb{B}^1/\sim$ represents the closed unit disc where antipodal points P_+, P_- on the boundary are identified with each other (Fig. 4.5(c)).

Finally, we want to consider a hyperbola in the setting of projective geometry. Fix $a > 0$ and $b > 0$. Consider the hyperbola

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \qquad (x, y) \in \mathbb{R}^2, \tag{4.16}$$

with respect to Cartesian (x, y)-coordinates in the Euclidean plane. Using the replacement $x \Rightarrow x/z$, $y \Rightarrow y/z$ with the nonzero real number z, we get the equation of the hyperbola,

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = z^2, \qquad (x,y,z) \in \mathbb{R}^3, \quad (x,y,z) \neq (0,0,0), \tag{4.17}$$

with respect to homogenous coordinates (x, y, z). This equation has the solutions

$$(x,y,z) = (a, \pm b, 0).$$

Note that the two infinite points $[(a,b,0)]$ and $[(a,-b,0)]$ on the projective hyperbola (4.17) correspond to the two asymptotes $y = \frac{b}{a}x$ and $y = -\frac{b}{a}x$ of the Euclidean hyperbola (4.16), respectively.

The equation of the unit circle $x^2 + y^2 = 1$ passes over to the equation

$$x^2 + y^2 = z^2, \qquad (x,y,z) \in \mathbb{R}^3, \quad (x,y,z) \neq (0,0,0,0),$$

in homogeneous coordinates. Letting $z = 0$, the equation $x^2 + y^2 = 0$ implies $x = y = 0$. Consequently, the unit circle does not contain any infinite points, as expected.

Generalization. Let $n = 1, 2, \ldots$ Using the real Hilbert space $X := \mathbb{R}^{n+1}$ and the notation introduced in Sect. 4.3.3, set $\mathbb{P}^n := \mathbb{P}(\mathbb{R}^{n+1})$. Generalizing the relation (4.15), we get the following bijections:

$$\mathbb{P}^n \simeq (\mathbb{R}^{n+1})^\times / \mathbb{R}^\times \simeq \mathbb{G}_1(\mathbb{R}^{n+1}) \simeq \mathbb{S}(\mathbb{R}^{n+1})/\{1,-1\}.$$

In algebraic geometry, one always uses projective spaces in order to get elegant results of great generality for algebraic curves. As an example, we mention the theorem of Max Noether (1844–1921) on the genus of an irreducible algebraic curve. See the textbooks by:

G. Walker, Algebraic Curves, Princeton University Press, 1950.

G. Fischer, Plane algebraic curves, Vieweg, Braunschweig, 1994 (in German).

E. Brieskorn and H. Knörrer, Plane Algebraic Curves, Birkhäuser, Basel, 1981 (in German).

4.3.5 Complex Projective Spaces

The projective complex line $\mathbb{P}^1_{\mathbb{C}}$. For two nonzero points (w, z) and (w', z') in the space $\mathbb{C}^2$, we write

$$(w,z) \sim (w',z')$$

iff there exists a nonzero complex number λ such that $(w,z) = \lambda(w',z')$. This is an equivalence relation. The equivalence classes are denoted by $\mathbb{P}(\mathbb{C}^2)$ (or briefly by $\mathbb{P}^1_{\mathbb{C}}$). The set $\mathbb{P}^1_{\mathbb{C}}$ has the complex dimension one. Therefore, it is called the projective complex line. The group $U(1)$ acts on the unit sphere $\mathbb{S}(\mathbb{C}^2)$ by $(w,z) \mapsto \lambda(w,z)$ with $\lambda \in U(1)$. There exists the bijection

$$\mathbb{P}^1_{\mathbb{C}} \;\simeq\; \mathbb{S}(\mathbb{C}^2)/U(1)$$

between $\mathbb{P}^1_{\mathbb{C}}$ and the orbit space $\mathbb{S}(\mathbb{C}^2)/U(1)$. In 1931, Heinz Hopf discovered that there exists a bijective map

$$h : \mathbb{S}(\mathbb{C}^2)/U(1) \to \mathbb{S}^2$$

which is called the Hopf map (or the Hopf fibration). This map allows an elegant interpretation in terms of the electron spin (see Sect. 5.7.2 of Vol. I). In particular, the Hopf map tells us that the space of quantum states of a two-dimensional

Hilbert space is bijective to the 2-dimensional sphere. By the Hopf map, we have the bijection

$$\boxed{\mathbb{P}^1_{\mathbb{C}} \simeq \mathbb{S}^2.}$$

The n-dimensional complex projective space. For $n = 1, 2, \ldots$, choose the Hilbert space $X := \mathbb{C}^{n+1}$. By (4.14), we have the bijections

$$\mathbb{P}(\mathbb{C}^{n+1}) \simeq (\mathbb{C}^{n+1})^\times / \mathbb{C}^\times \simeq \mathbb{S}(\mathbb{C}^{n+1})/U(1).$$

The space $\mathbb{P}(\mathbb{C}^{n+1})$ is also briefly denoted by the symbol $\mathbb{P}^n_{\mathbb{C}}$, and it is called the n-dimensional complex projective space.

4.3.6 The Shape of the Universe

The recent WMAP (Wilkinson Microwave Anisotropy Probe) experiment of NASA measures the anisotropy of the radiation which was created in the universe 380 000 years after the Big Bang. Information about this experiment can be found on the NASA homepage:

`http://www.nasa.gov/home/`

`http://lambda.gsfc.nasa.gov`

Furthermore, we refer to:

R. Aurich, S. Lustig, F. Steiner, and H. Then (2004), Indications about the Shape of the Universe from the WMAP data, Phys. Rev. Lett. **94**, 021301. Internet:`http://arXiv.org/astro-ph/0412407`

R. Aurich and F. Steiner (2004), Quintessence and the curvature of the universe after WMAP, Int. J. Mod. Phys. D **13**, 123–136. Internet: `http://arXiv.org/astro-ph/0302264`

These papers are based on sophisticated mathematical methods from both spectral geometry and analytic number theory, and they make use of huge computer calculations. For example, the available WMAP data exclude a completely flat universe called quintessence. In fact, the WMAP data show that the curvature of our Universe is small, but the sign of the curvature is not known today. It is possible that our Universe is not simply connected, but it is a homogenous space X/G which has a complicated topological structure. In order to get better information about the true structure of our Universe, one needs more precise experimental data. Such improved experiments are planned for the near future by using the European Planck satellite. The papers quoted above can be viewed as cosmological versions of the following classical papers:

M. Kac, Can one hear the shape of a drum? Amer. Math. Monthly **73**(4), (1966), 1–23.

C. Gordon, D. Webb, and S. Volpert, You can't hear the shape of a drum, Bull. Amer. Math. Soc. **27**(1) (1992), 134–138.

C. Gordon and D. Webb, You can't here the shape of a drum, American Scientist **84** (1996), 46–55.

The last two papers tell us that the observed eigenfrequencies of a drum (i.e., the eigenvalues of the Laplacian on a compact Riemannian manifold) do not always uniquely determine the manifold, up to isometries.

Hints for further reading. The 1872 Klein Erlangen program on the relation between geometry and symmetry was substantially generalized to differential geometry by Élie Cartan (1859–1951) in the 1920s. We recommend the following monographs:

T. Ivey and J. Landsberg, Cartan for Beginners: Differential Geometry via Moving Frames and Exterior Differential Systems, Amer. Math. Soc., Providence, Rhode Island, 2003.

R. Sharpe, Differential Geometry: Cartan's Generalization of Klein's Erlangen Program, Springer, New York, 1997.

For the Ehresmann approach to modern differential geometry (based on symmetry groups, principal fiber bundles, and connections on the associated vector bundles), we refer to:

S. Kobayashi and K. Nomizu, Foundations of Differential Geometry, Vols. 1, 2, Wiley, New York, 1963.

J. Jost, Riemannian Geometry and Geometric Analysis, Springer, Berlin, 2008.

The relations to gauge theory and the Standard Model in elementary particle physics will be thoroughly studied in Vol. III.

4.4 Equivalence Classes in Topology

> Though I travelled many different roads, I always encountered analysis situs (topology).
>
> Henri Poincaré (1854–1912)

> Topology has its roots in geometric intuition, the theory of analytic functions, the theory of Abelian integrals over algebraic functions, and in physics.
>
> Folklore

4.4.1 Topological Quotient Spaces

In this Section, we need the notions 'topological space' and 'topology' introduced in Sect. 5.5 of Vol. I.

Quotient topology. Let X be a topological space. Suppose that we are given an equivalence relation $\sim$ on X. We have the canonical projection

$$\pi : X \to X/\sim \tag{4.18}$$

given by $\pi(x) := [x]$. We want to equip the quotient space $X/\sim$ of equivalence classes $[x]$ with a natural topology. To this end, we define:

> *A subset S of the quotient space $X/\sim$ is called open iff the set of all the corresponding representatives is open in the original space X.*

In other words, a subset S of $X/\sim$ is called open iff the set

$$\pi^{-1}(S) \text{ is an open subset of } X.$$

Then the map π from (4.18) has the obvious property that the pre-images of open sets are again open. Thus, π is continuous.

Examples. Topological quotient spaces are everywhere in mathematics and physics. Let us consider a few simple examples.

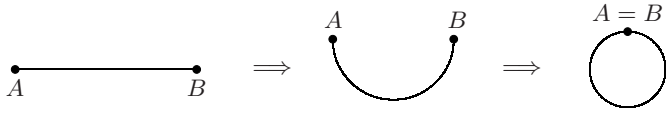


Fig. 4.6. The circle as the quotient space of an interval

(i) The unit circle as a quotient space of the unit interval: Consider the unit interval $I := [0, 1]$. We equip the set I with the subspace topology induced by the topology on the real line $\mathbb{R}$. Explicitly, a subset O' of I is called open iff there exists an open subset O of the real line such that

$$O' = O \cap I.$$

For example, the sets $[0, \varepsilon[$ and $]1 - \varepsilon, 1]$ are open in I if $0 < \varepsilon \leq \frac{1}{2}$. Now we want to identify the endpoints $A := \{0\}$ and $B := \{1\}$ of the interval I with each other (Fig. 4.6). Intuitively, we glue the points A and B together. This yields a circle. In terms of mathematics, we start with the decomposition

$$I = \{0, 1\} \cup \bigcup_{0<x<1} \{x\}$$

of the interval I into pairwise disjoint sets. These sets are the equivalence classes of the corresponding equivalence relation $\sim$. Explicitly, the equivalence classes are the following subsets off the interval I:

$$[0] = [1] := \{0, 1\} \quad \text{and} \quad [x] := \{x\} \qquad \text{for all} \quad x \in]0, 1[.$$

Choose $0 < \varepsilon \leq \frac{1}{2}$, and set $U_\varepsilon := [0, \varepsilon[\ \cup \ ]1 - \varepsilon, 1]$. Then, for example, the set

$$\mathcal{U}_\varepsilon := \{[x] : \ x \in U_\varepsilon\}$$

is an open neighborhood of the point $[0]$ in the quotient space $I/\sim$. The quotient space $I/\sim$ is homeomorphic to the unit circle $\mathbb{S}^1$.

(ii) The 2-dimensional sphere as a quotient space of the unit square: Consider the unit square

$$S := \{(x, y) \in \mathbb{R}^2 : 0 \leq x, y \leq 1\},$$

and identify all of the boundary points with each other. Intuitively, we glue all of the boundary points together (Fig. 4.7). The corresponding quotient space $S/\sim$ is homeomorphic to the 2-dimensional unit sphere $\mathbb{S}^2$.

(iii) The torus as the quotient space of a square: In contrast to example (ii), we only identify opposite boundary points of the unit square S with each other. The corresponding quotient space $S/\sim$ is homeomorphic to a 2-dimensional torus (Fig. 4.8).

(iv) The n-dimensional sphere as the quotient space of the n-dimensional cube: Let $n = 1, 2, \ldots$ Consider the n-dimensional unit cube

$$I^n := \{(x_1, \ldots, x_n) \in \mathbb{R}^n : \ 0 \leq x_1 \leq \ldots \leq x_n \leq 1\},$$

and identify all of the boundary points with each other. Then the corresponding quotient space $I^n/\sim$ is homeomorphic to the n-dimensional unit sphere $\mathbb{S}^n$.

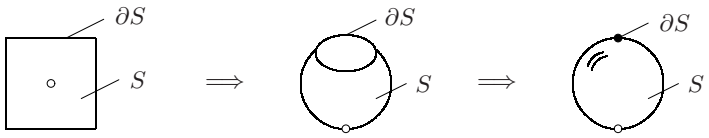


Fig. 4.7. The two-dimensional sphere as the quotient space of a square

(v) The real n-dimensional projective space as the quotient space of the n-dimensional sphere: Consider the sphere $\mathbb{S}^n$, $n = 1, 2, \ldots$, equipped with the equivalence relation

$$x \sim y \qquad \text{iff} \qquad x = \pm y.$$

The corresponding quotient space $\mathbb{S}^n/\{\pm I\}$ becomes a topological space with respect to the quotient topology. Since we have the bijection

$$\mathbb{S}^n/\{\pm I\} \simeq \mathbb{P}^n,$$

the n-dimensional real projective space $\mathbb{P}^n$ also becomes a topological space.

(vi) Orbit spaces: Example (v) is a special case of the following more general situation. Let X be a topological space, and suppose that the symmetry group G acts on X. By Sect. 4.3.2, there exists an equivalence relation on X such that the quotient space $X/\sim$ coincides with the orbit space X/G. Using the quotient topology, the orbit space X/G becomes a topological space.

(vii) The Gaussian plane factorized by a lattice: Consider the lattice

$$L := \{z \in \mathbb{C} : z = m + n\mathrm{i}, \quad m, n \in \mathbb{Z}\}.$$

This lattice is generated from the origin $z = 0$ by repeating the following translations: $z \mapsto z + 1$, $z \mapsto z + \mathrm{i}$, $z \mapsto z - 1$, and $z \mapsto z - \mathrm{i}$. For the points z, w in the complex plane $\mathbb{C}$, we write

$$z \sim w \quad \text{iff} \quad z - w \in L.$$

The corresponding quotient space $\mathbb{C}/L$ is homeomorphic to the quotient space $S/\sim$ from (iii) above, and hence it is homeomorphic to a 2-dimensional torus T. Let $f : \mathbb{C} \to \overline{\mathbb{C}}$ be a meromorphic function with the periods 1 and i, that is,

$$f(z + w) = f(z) \qquad \text{for all} \quad w \in L.$$

Then the function f can be considered as a function of the form $f : \mathbb{C}/L \to \overline{\mathbb{C}}$. In turn, this induces a function

$$f : T \to \overline{\mathbb{C}}$$

on the torus T. This was the basic idea of Riemann in order to reduce the study of elliptic functions (and of the corresponding Abelian integrals) to the investigation of the topological properties of a torus.

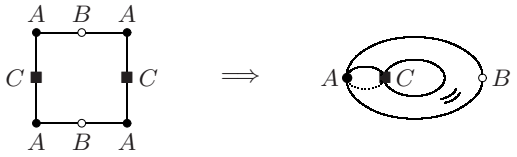


Fig. 4.8. The torus as the quotient space of a square

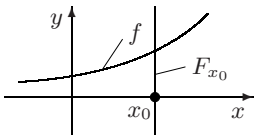


Fig. 4.9. Section of a fiber bundle

4.4.2 Physical Fields, Observers, Bundles, and Cocycles

> Physical fields can be described by bundles in mathematics. The change of the real values measured by different observers corresponds to cocyles.
> Folklore

We want to show how the modern mathematical language of bundles fits physics in a quite natural way. The point is that we have to model mathematically the following situation:

- Three observers measure the same physical effect.
- The transformation laws between the measured quantities are governed by the fact that physics has an invariant meaning; this means, that physics is independent of the specific observers.

This corresponds to the situation in geometry where invariant geometric objects (e.g., a sphere) are described by different local coordinate systems.

Prototype. Let $y = f(x)$ be a real function as pictured in Fig. 4.9. The subset

$$\text{graph}(f) := \{(x, f(x)) \in \mathbb{R}^2 : \ x \in \mathbb{R}\}$$

of $\mathbb{R}^2$ is called the graph of the map $f : \mathbb{R} \to \mathbb{R}$. We want to reformulate f as the section of a fiber bundle. To this end, we introduce the surjective map

$$\pi : \mathbb{R}^2 \to \mathbb{R}$$

given by $\pi(x, y) := x$. The map π is called the projection from the bundle space $\mathbb{R}^2$ onto the base space $\mathbb{R}$. The pre-image $\pi^{-1}(x)$ is called the fiber F_x at the point x of the base space $\mathbb{R}$. Explicitly,

$$F_x = \{(x, y) \in \mathbb{R}^2 : \ y \in \mathbb{R}\}.$$

Therefore, $F_x \simeq \mathbb{R}$ for all $x \in \mathbb{R}$, in the sense of a bijection. Here, $\mathbb{R}$ is called the typical fiber. If $x \neq x'$, then $F_x \neq F_{x'}$. Thus, we have the disjoint decomposition

$$\mathbb{R}^2 = \bigcup_{x \in \mathbb{R}} F_x$$

of the bundle space $\mathbb{R}^2$ into fibers. Set $s(x) := (x, f(x))$. The map

$$s : \mathbb{R} \to \mathbb{R}^2$$

has the property that $s(x) \in F_x$ for all $x \in \mathbb{R}$. This map is called a *section* of the bundle. This is the desired equivalent formulation of the given function f. The family $\{F_x\}_{x \in \mathbb{R}}$ is called a pre-bundle.

Unstructured fiber bundle. By definition, each surjective map

$$\boxed{\pi : B \to X} \tag{4.19}$$

is called an unstructured fiber bundle, in contrast to topological and smooth fiber bundles to be considered below. Here, we use the following terminology:

- the set B is called the bundle space,
- the set X is called the base space,
- the map π is called the bundle projection, and
- the pre-image $F_x := \pi^{-1}(x)$ is called the fiber over the base point x.

The map $s : X \to B$ is called a section iff the diagram

$$\begin{array}{ccc} & & B \\ & \overset{s}{\nearrow} & \big\downarrow \pi \\ X & \xrightarrow{\mathrm{id}} & X \end{array} \tag{4.20}$$

is commutative. Equivalently, $\pi \circ s = \mathrm{id}$ (i.e., the map s is a right inverse of the bundle projection π).

Bundle morphism. By definition, a morphism between the bundles

$$\pi : B \to X \qquad \text{and} \qquad \sigma : C \to Y$$

is a pair (f, g) of mappings such that the following diagram is commutative:

$$\begin{array}{ccc} B & \xrightarrow{f} & C \\ \big\downarrow \pi & & \big\downarrow \sigma \\ X & \xrightarrow{g} & Y. \end{array} \tag{4.21}$$

This implies that fibers are preserved. Explicitly, we have the induced maps

$$f : F_x \to F_{g(x)} \qquad \text{for all} \quad x \in X.$$

The bundle morphism (4.21) is called a bundle isomorphism iff the maps f and g are bijective.

Pre-bundles and physical fields. By definition, a pre-bundle is a family

$$\{\mathcal{F}_x\}_{x \in X}$$

of sets $\mathcal{F}_x$ (e.g., linear spaces, groups, rings) indexed by the set X. In terms of physics, the points x are events which are elements of the space-time X (i.e., x describes a position in space at a certain time). The points of the pre-fiber $\mathcal{F}_x$ describe additional degrees of freedom of the physical system at the event x. The family

$$\{\psi(x)\}_{x \in X}$$

is called a pre-section iff $\psi(x) \in \mathcal{F}_x$ for all $x \in X$. In terms of physics, this describes a physical field $x \mapsto \psi(x)$ which depends on space and time.

Proposition 4.14 *Each pre-bundle corresponds to a bundle. Here, the pre-fibers are bijectively equivalent to the fibers of the bundle, and the pre-sections correspond to sections of the bundle.*

Proof. We define the bundle space B as the coproduct

$$B := \coprod_{x \in X} \mathcal{F}_x$$

of the pre-fibers $\mathcal{F}_x$. Explicitly, $B = \{(x, y) : \ x \in X, \ y \in \mathcal{F}_x\}$.[18] The bundle projection

$$\pi : B \to X$$

is given by $\pi(x, y) := x$. The fiber F_x over the base point x is equal to

$$F_x = \pi^{-1}(x) = \{x\} \times \mathcal{F}_x.$$

Set $i_x(y) := (x, y)$. For each base point $x \in X$, the map

$$i_x : \mathcal{F}_x \to B$$

is injective with the image $i_x(\mathcal{F}_x) = F_x$. This induces the bijection

$$i_x : \mathcal{F}_x \to F_x$$

between the pre-fiber $\mathcal{F}_x$ and the fiber F_x. Set $s(x) := i_x(\psi(x))$. Then the map $s : X \to B$ is a section of the bundle. □

Operations with pre-bundles. Our general strategy reads as follows:

Operations between pre-fiber bundles are performed with respect to fibers.

For example, let $\{\mathcal{F}_x\}_{x\in X}$ and $\{\mathcal{G}_x\}_{x\in X}$ be pre-fiber bundles where all the fibers $\mathcal{F}_x$ and $\mathcal{G}_x$ are real finite-dimensional linear spaces. For linear spaces, we have the direct sum $\mathcal{F}_x \oplus \mathcal{G}_x$ and the tensor product $\mathcal{F}_x \otimes \mathcal{G}_x$ at hand.[19] This leads immediately to the operations

$$\{\mathcal{F}_x\}_{x\in X} \oplus \{\mathcal{G}_x\}_{x\in X} := \{\mathcal{F}_x \oplus \mathcal{G}_x\}_{x\in X}$$

and

$$\{\mathcal{F}_x\}_{x\in X} \otimes \{\mathcal{G}_x\}_{x\in X} := \{\mathcal{F}_x \otimes \mathcal{G}_x\}_{x\in X}.$$

We call $\{\mathcal{F}_x\}_{x\in X}$ a pre-subbundle of $\{\mathcal{G}_x\}_{x\in X}$ iff each fiber $\mathcal{F}_x$ is a linear subspace of the fiber $\mathcal{G}_x$. Finally, a morphism from $\{\mathcal{F}_x\}_{x\in X}$ to $\{\mathcal{G}_y\}_{y\in Y}$ is defined to be both a map $g : X \to Y$ and a family

$$\{L_x\}_{x\in X}$$

of linear maps $L_x : \mathcal{F}_x \to \mathcal{G}_y$ with $y = g(x)$ for all $x \in X$. This morphism is called an isomorphism iff all the maps L_x are linear isomorphisms and g is bijective.

Next it is our goal to equip fiber bundles $\pi : B \to X$ with an additional structure.

For example, π is a continuous or smooth map. This way, we will get topological and smooth fiber bundles, respectively.

Product bundle. Let X and F be sets. The simplest bundle is a product bundle which corresponds to the surjective map

$$\pi : X \times F \to X$$

where $\pi(x, y) := x$. The fiber over the base point $x \in X$ is given by

$$F_x := \pi^{-1}(x) = \{x\} \times F.$$

The set F is called the typical fiber. Product bundles are called trivial bundles. Now we want to construct nontrivial bundles.

Roughly speaking, the idea is to glue trivial bundles together.

[18] The coproduct is also called the disjoint union.
[19] For example, $\mathbb{R}^n \oplus \mathbb{R}^m = \mathbb{R}^{n+m}$ and $\mathbb{R}^n \otimes \mathbb{R}^m = \mathbb{R}^{nm}$.

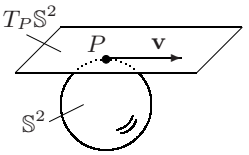


Fig. 4.10. The tangent bundle of a sphere

Vector bundle. Let $n = 1, 2, \ldots$ Consider the following situation.

(V1) Topological bundle: The map $\pi : B \to X$ is surjective and continuous where the bundle space B and the base space X are topological spaces.

(V2) Linear fibers: For each $x \in X$, the fiber $F_x := \pi^{-1}(x)$ carries the structure of a real n-dimensional linear space.

(V3) Local triviality: The bundle behaves locally like a trivial product bundle

$$U_\alpha \times \mathbb{R}^n.$$

More precisely, there exists a covering $\{U_\alpha\}$ of the base space X by open sets U_α. For each index α, we have a homeomorphism

$$\chi_\alpha : \bigcup_{x \in U_\alpha} F_x \to U_\alpha \times \mathbb{R}^n \tag{4.22}$$

which preserves the linear fibers. That is, all the maps

$$F_x \xrightarrow{\chi_\alpha} \{x\} \times \mathbb{R}^n \xrightarrow{p} \mathbb{R}^n$$

are linear isomorphisms from F_x onto $\mathbb{R}^n$. Here, $p(x, y) := y$.

If the conditions (V1)–(V3) are satisfied, then the bundle $\pi : B \to X$ is called a real vector bundle of rank n. A map $s : B \to X$ is called a section of the vector bundle iff the map s is continuous, and we have the commutative diagram (4.20).

By definition, a morphism (resp. isomorphism) from the vector bundle

$$\pi : B \to X$$

to the vector bundle $\sigma : C \to Y$ is a pair (f, g) of continuous maps (resp. homeomorphisms) $f : B \to C$ and $g : X \to Y$ such that the diagram (4.21) is commutative. Then the map f is fiber-preserving. In addition, we postulate that the linear structure of the fibers is respected. That is, all the induced fiber maps

$$f : F_x \to F_{g(x)}, \qquad x \in X$$

are linear morphisms (resp. linear isomorphisms).

Smooth vector bundle. The vector bundle introduced above is called smooth iff we replace the topological spaces (resp. continuous maps) by manifolds (resp. smooth maps). In other words, we replace the category of topological spaces by the category of manifolds.

Standard example (velocity field on a sphere and the tangent bundle). Consider the unit sphere $\mathbb{S}^2$ (Fig. 4.10). In order to describe a continuous velocity field

$$\mathbf{v} = \mathbf{v}(P)$$

on the sphere, we need the quite natural notion of the tangent bundle. Intuitively, think of an ocean which covers a planet; the velocity field describes the velocity vectors of the fluid particles. Note that the velocity vector $\mathbf{v}(P)$ is an element of the tangent space $T_P\mathbb{S}^2$ of the sphere $\mathbb{S}^2$ at the point P. It is our goal to show that:

A continuous velocity field on the sphere is a section of the tangent bundle of the sphere.

By definition, the tangent bundle $T\mathbb{S}^2$ of the sphere $\mathbb{S}^2$ consists of all the pairs

$$\{(P, \mathbf{v}) : \ P \in \mathbb{S}^2, \ \mathbf{v} \in T_P\mathbb{S}^2\}.$$

In other words, the tangent bundle $T\mathbb{S}^2$ is the set of all pairs $(P, \mathbf{v})$ where P is an arbitrary point of the sphere, and $\mathbf{v}$ is an arbitrary tangent vector of the sphere at the point P. Setting $\pi(P, \mathbf{v}) := P$, we get the vector bundle

$$\pi : T\mathbb{S}^2 \to \mathbb{S}^2.$$

The fiber F_P over the point $P \in \mathbb{S}^2$ is the tangent space $T_P\mathbb{S}^2$ at the point P.

In order to prove local triviality, we will introduce local coordinates in a natural way. Let $\{U_\alpha\}$ be a covering of the sphere $\mathbb{S}^2$ by nonempty open sets U_α which are different from $\mathbb{S}^2$. Fix the set U_α. Then there exist two continuous tangent vector fields $\mathbf{v}_1 = \mathbf{v}_1(P)$ and $\mathbf{v}_2 = \mathbf{v}_2(P)$ on U_α such that $\mathbf{v}_1(P), \mathbf{v}_2(P)$ is a basis of the tangent space $T_P\mathbb{S}^2$ at each point $P \in U_\alpha$. Therefore, each tangent vector $\mathbf{v}$ at the point $P \in U_\alpha$ can be represented in the form

$$\mathbf{v} = v^1(P)\mathbf{v}_1(P) + v^2(P)\mathbf{v}_2(P).$$

The tuple $v^1(P), v^2(P)$ of real numbers represents the local coordinates of the tangent vector $\mathbf{v}$ with respect to U_α. The map

$$\chi_\alpha : \bigcup_{P \in U_\alpha} T_P\mathbb{S}^2 \to U_\alpha \times \mathbb{R}^2$$

from (V3) above is given by $\chi_\alpha(P, \mathbf{v}) := \{P\} \times (v^1(P), v^2(P))$.

A section $s : \mathbb{S}^2 \to T\mathbb{S}^2$ of the tangent bundle is nothing else than a continuous tangent vector field on the sphere. Observe that it is not possible to introduce a global coordinate system on the sphere. In fact, Poincaré proved the crucial fact in 1885 that:

Every continuous tangent vector field on the sphere $\mathbb{S}^2$ vanishes at some point.

For the proof based on the mapping degree, we refer to Zeidler (1986), Vol. I, p. 558 (see the references on page 1049). Let $n = 1, 2, \ldots$ A sphere $\mathbb{S}^n$ is called parallelizable iff there exist n continuous tangent vector fields of the sphere $\mathbb{S}^2$ which are linearly independent at each point of the sphere. A fundamental theorem of topology tells us the following:[20]

Precisely the spheres $\mathbb{S}^1, \mathbb{S}^3, \mathbb{S}^7$ are parallelizable.

This is a special case of the theorem of Adams on page 233. Tangent bundles play a fundamental role in describing the mechanics of point particles by position and velocity. This will be studied in Chap. 6.

The sphere is not only a topological space, but a 2-dimensional real manifold. This means that it looks locally like an open subset of $\mathbb{R}^2$, and the change of local coordinates is given by diffeomorphisms (see Sect. 5.4 of Vol. I). Using local manifold coordinates for the points of the sphere, and describing the tangent vectors in terms of these local coordinates, the tangent bundle becomes a smooth vector bundle. The

[20] The proof based on K-theory can be found in F. Hirzebruch (1995) quoted on page 235.

sections of the smooth tangent bundle $T\mathbb{S}^2$ are precisely the smooth velocity fields $P \mapsto \mathbf{v}(P)$ on the sphere $\mathbb{S}^2$.

Bundle charts, observers, and transition maps. Let us discuss the maps (4.22) in terms of observers. We want to show that this is intimately related to the notion of cocycle which is of fundamental importance for topology. The map

$$\chi_\alpha : \bigcup_{x \in U_\alpha} F_x \to U_\alpha \times \mathbb{R}^n$$

is called a bundle chart. This map assigns to the point $P \in F_x$ on the bundle space B the local bundle coordinate

$$(x, y_\alpha(P)) \quad \text{with} \quad x \in U_\alpha,\ y_\alpha(P) \in \mathbb{R}^n.$$

In terms of physics, an observer O_α (corresponding to U_α and χ_α) measures the value $y_\alpha(P)$ at the space-time point x. Now suppose that

$$x \in U_\alpha \cap U_\beta,$$

and consider a second observer O_β (corresponding to U_β and χ_β). Then the point $P \in F_x$ of the bundle space B is also described by the local bundle coordinate

$$(x, y_\beta(P)) \quad \text{with} \quad x \in U_\beta,\ y_\beta(P) \in \mathbb{R}^n.$$

The change of local coordinates

$$\boxed{y_\beta(P) = T_{\beta\alpha}(x) y_\alpha(P)}$$

is given by the real invertible $(n \times n)$-matrix $T_{\beta\alpha}(x)$, that is, $T_{\beta\alpha}(x) \in GL(n, \mathbb{R})$. The map

$$T_{\beta\alpha} : U_\alpha \cup U_\beta \to GL(n, \mathbb{R})$$

is called the transition map of the bundle from the observer O_α to the observer O_β (corresponding to the pair (U_α, U_β) of open subsets of the base space X). The change of observers corresponds to the following commutative diagram

$$\begin{array}{ccc} O_\beta & \xleftarrow{T_{\beta\alpha}} & O_\alpha \\ {\scriptstyle T_{\gamma\beta}}\downarrow & \swarrow{\scriptstyle T_{\gamma\alpha}} & \\ O_\gamma & & \end{array} \tag{4.23}$$

For all indices α, β, γ, the following three conditions are satisfied:

(C1) $T_{\alpha\alpha}(x) = \mathrm{id}$ for all $x \in U_\alpha$.
(C2) $T_{\alpha\beta}(x) = T_{\beta\alpha}(x)^{-1}$ for all $x \in U_\alpha \cap U_\beta$.
(C3) $T_{\gamma\alpha}(x) = T_{\gamma\beta}(x) T_{\beta\alpha}(x)$ for all $x \in U_\alpha \cap U_\beta \cap U_\gamma$.

This allows the following quite natural interpretation in terms of observers. Condition (C2) tells us that $y_\alpha(P) = T_{\beta\alpha}(x)^{-1} y_\beta(P)$. Furthermore, condition (C3) corresponds to the commutativity of the diagram (4.23). Explicitly, if $x \in U_\alpha \cap U_\beta \cap U_\gamma$, then $y_\gamma(P) = T_{\gamma\alpha}(x) y_\alpha(P)$, and

$$y_\gamma(P) = T_{\gamma\beta}(x) y_\beta(P) = T_{\gamma\beta}(x) T_{\beta\alpha}(x) y_\alpha(P).$$

This implies (C3).

Observers, cocycles, and the construction of bundles. By definition, a cocycle on the topological space X with values in the group G consists of

- a covering $\{U_\alpha\}_{\alpha\in\mathcal{A}}$ of the space X by open subsets U_α and
- a family of continuous maps $T_{\beta\alpha} : U_\alpha \cap U_\beta \to$ such that the conditions (C1)–(C3) are satisfied for all indices $\alpha, \beta, \gamma \in \mathcal{A}$.

The proof of the following theorem shows that cocycles play a decisive role in describing all kind of physical fields by observers. In mathematics, it was discovered by Ĉech (1893–1960) in the 1930s that cocycles can be used in topology for constructing the so-called Ĉech cohomology. In modern mathematics, one uses a generalization of Ĉech cohomology on arbitrary topological spaces called cohomology with values in a pre-sheaf. This general approach dates back to Leray (1906–1998) in the late 1940s.

Theorem 4.15 *Suppose that we are given the topological space X and a cocycle $\{T_{\beta\alpha}\}$ on X with values in the group $GL(n,\mathbb{R})$, $n = 1, 2, \ldots$ Then there exists a vector bundle $\pi : B \to X$ with typical fiber $\mathbb{R}^n$ which corresponds to the given cocycle.*

Proof. (I) The bundle space B. The proof is strongly motivated by the physical picture of observers. We start with the coproduct

$$\mathcal{B} := \coprod_{\alpha\in\mathcal{A}} U_\alpha \times \mathbb{R}^n.$$

Explicitly, $\mathcal{B} := \{(\alpha, x; v_\alpha) : \ \alpha \in \mathcal{A}, x \in U_\alpha, v_\alpha \in \mathbb{R}^n\}$. For the elements of $\mathcal{B}$, we write

$$(\alpha, x, v_\alpha) \sim (\beta, y, v_\beta)$$

iff we have $x \in U_\alpha, y \in U_\beta, x = y$ and $v_\beta = T_{\beta\alpha} v_\alpha$. Because of the cocycle property, this is an equivalence relation on $\mathcal{B}$. By definition, the corresponding equivalence classes $[(\alpha, x, v_\alpha)]$ form the bundle space B. Briefly,

$$B := \mathcal{B}/\sim .$$

Naturally enough, the bundle projection $\pi : B \to X$ is defined by

$$\pi([\alpha, x, v_\alpha]) := x.$$

Intuitively, the equivalence class $[(\alpha, x, v_\alpha)]$ describes the physical field v at the event (space-time point) x by the values of the field measured by different observers.

(II) Topology of the bundle space B. We say that the subset O of B is open iff, for each point $[(\alpha, x, v_\alpha)]$ of the set O, there exist

- an open subset O_α of the base space X with $x \in O_\alpha \subseteq U_\alpha$ and
- an open subset V of the typical fiber $\mathbb{R}^n$ with $v_\alpha \in V$

such that

$$[(\alpha, y, w)] \subseteq O_\alpha \qquad \text{for all} \quad (y, w) \in O_\alpha \times V.$$

One checks in a straightforward manner that this definition of the open set O does not depend on the choice of the representatives in the bundle charts $U_\alpha \times \mathbb{R}^n$. This way, the bundle space B becomes a topological space, and the bundle projection $\pi : B \to X$ is continuous. □

The preceding proof tells us that, for $n = 1, 2, \ldots$, the following hold:

> *There exists a one-to-one correspondence between real vector bundles of rank n on the topological space X and cocycles on X with values in the group $GL(n,\mathbb{R})$.*

In terms of physics, this corresponds to the relation between the invariant formulation of physical fields and the coordinate formulation based on the transformation laws between different observers.

Operations with vector bundles. In Sect. 3.2, we have studied operations for linear spaces and linear operators. In a quite natural way, all of these operations can be translated to vector bundles. Since the fibers are linear spaces, the simple strategy is to perform the operations with respect to the fibers. This can be easily done

- by using the corresponding operations for the typical fibers,
- by applying the corresponding operations to the cocycles, and
- by constructing the bundle space via Theorem 4.15.

Let us discuss this. We are given the two real vector bundles V and W of finite rank with

- the bundle projections $\pi : B \to X$ and $\sigma : C \to X$,
- the fibers F_x and G_x,
- the typical fibers $\mathbb{R}^m$ and $\mathbb{R}^n$, and
- the cocycles $S_{\alpha\beta} : U_\alpha \cup U_\beta \to GL(m, \mathbb{R})$ and $T_{\alpha\beta} : U_\alpha \cup U_\beta \to GL(n, \mathbb{R})$, respectively.

First we assume that the two vector bundles V and W refer to the same open covering $\{U_\alpha\}$ of the base space X.

(i) Cartesian product: We want to construct the vector bundle $V \times W$ with the fibers $F_x \times G_x$ for all $x \in X$. To this end, we use the typical fiber $\mathbb{R}^m \times \mathbb{R}^n$ and the cocycle $S_{\alpha\beta} \times T_{\alpha\beta}$. Explicitly,

$$(S_{\alpha\beta} \times T_{\alpha\beta})(x)(v, w) = (S_{\alpha\beta}(x)v, T_{\alpha\beta}(x)w).$$

One checks easily that indeed the Cartesian product $S_{\alpha\beta} \times T_{\alpha\beta}$ of cocycles yields again a cocycle.

(ii) Direct sum: In order to construct the vector bundle $V \oplus W$ with the fibers $F_x \oplus G_x$ for all $x \in X$, we use the typical fiber $\mathbb{R}^m \oplus \mathbb{R}^n$ and the cocycle $S_{\alpha\beta} \oplus T_{\alpha\beta}$.

(iii) Tensor product: The tensor product $V \otimes W$ with the fibers $F_x \otimes G_x$ for all $x \in X$ is obtained by using the typical fiber $\mathbb{R}^m \otimes \mathbb{R}^n$ and the cocycle $S_{\alpha\beta} \otimes T_{\alpha\beta}$.

If the vector bundles V and W refer to the open coverings $\{U_\alpha\}$ and $\{V_\beta\}$ of the base space X, then we apply the construction above to the common covering consisting of all the intersections $U_\alpha \cap V_\beta$. Finally, it is not difficult to show that the following hold where the symbol $\simeq$ stands for bundle isomorphims: If $V \simeq \mathcal{V}$ and $W \simeq \mathcal{W}$, then

$$V \times W \simeq \mathcal{V} \times \mathcal{W}.$$

The same is true for $V \oplus W$ and $V \otimes W$. Therefore, the constructions above do not depend on the open coverings of the base space X, but only on the isomorphism classes of the vector bundles.

Generalizations. We have restricted ourselves to real vector bundles. The same arguments apply to complex vector bundles with the typical fiber $\mathbb{C}^n$. If we use the replacements

$$\begin{aligned} \textit{topological space} &\Rightarrow \textit{manifold}, \\ \textit{continuous map} &\Rightarrow \textit{smooth map}, \end{aligned} \tag{4.24}$$

then we get the notion of smooth vector bundles. The cocycles are then smooth maps. The same arguments apply to more general bundles (e.g., principal fiber bundles, where the fibers are Lie groups).

Modern differential geometry is based on the idea of parallel transport in bundles.

This allows us to introduce the notion of curvature. In fact, this represents the mathematics of the Standard Model in elementary physics for describing the fundamental forces in nature. The prototype is the Levi-Civita parallel transport of velocity fields on a sphere (see Vol. III).

4.4.3 Generalized Physical Fields and Sheaves

> During World War II, the French mathematician Jean Leray (1906–1998) was a prisoner of war from 1940 to 1945. He organized a university in his prison camp and himself gave a course on algebraic topology, a field he had become interested in connection with his collaboration with the Polish mathematician Juliusz Schauder (1899–1943) on applications of degree theory in nonlinear functional analysis.[21] Leray became dissatisfied both with the methods using triangulations and with those using inverse or direct limits (introduced by Čech (1893-1960) in the 1930s). In 1942 he published a series of four Notes in the *Comptes rendus* (of the French Academy of Sciences) outlining a new and original way of defining and studying cohomology ...
>
> In May 1946 Leray published two Notes in the *Comptes rendus* in which he introduced for the first time the notions of *sheaf, sheaf cohomology,* and of *spectral sequence*. In retrospect, it is difficult to exaggerate the importance of these concepts, which very rapidly became not only powerful tools in algebraic topology, but spread to many other parts of mathematics, some of which seem very remote from topology, such as algebraic geometry, number theory, and mathematical logic. These applications certainly went far beyond the wildest dreams of the inventor of these notions, and they undoubtedly rank at the same level in the history of mathematics as the methods invented by Poincaré (1854–1912) and Brouwer (1881–1966) in classical topology.[22]
>
> Jean Dieudonné, 1989

Let X be an arbitrary set, and let R be a ring. In order to study the structure of the set X, one can study the space of all the mappings

$$f : X \to R.$$

In terms of physics, the mapping f can be regarded as a physical field on the space X (e.g., a space-time manifold) with values in the ring R. For mathematics and

[21] In 1930 Schauder published the famous Schauder fixed-point theorem: *The fixed-point theorem in function spaces*, Studia Math. **2** (1930), 171–180 (in German). In 1936, Leray and Schauder created the fundamental Leray–Schauder degree theory for solving nonlinear operator equations (e.g., nonlinear partial differential equations and integral equations) in infinite-dimensional Banach spaces. This is thoroughly studied in Zeidler (1986), Vol. I (see the references on page 1049). In 1943, Juliusz Schauder and his wife were killed by the German occupation army in Poland.

[22] J. Dieudonné, A History of Algebraic and Differential Topology 1900–1960, Birkhäuser, Boston 1989 (reprinted with permission). See also the quotation on the history of algebraic geometry to be found on page 200.

physics, it is important to generalize this concept in the following way. Let X be a topological space with topology $\mathcal{T}$. Recall that $\mathcal{T}$ denotes the family of open subsets of X.

We assign to each set $U \in \mathcal{T}$ a ring $R(U)$.

This leads us to the concept of pre-sheaves and sheaves to be studied in later volumes.

Analytic functions and sheaves. The prototype of a sheaf is given by analytic functions on open subsets of the Gaussian plane $X := \mathbb{C}$. Let $\mathcal{T}(\mathbb{C})$ denote the family of open subsets of $\mathbb{C}$ (e.g., an open disc). We assign to each $U \in \mathcal{T}(\mathbb{C})$ the ring $R(U)$ of all holomorphic functions

$$f : U \to \mathbb{C}.$$

Suppose that we are given arbitrary open sets

$$U, V, W \in \mathcal{T}(\mathbb{C}) \quad \text{with} \quad W \subseteq V \subseteq U.$$

By definition, the map $r_{U,V} : R(U) \to R(V)$ assigns to each holomorphic function $f : U \to \mathbb{C}$ its restriction $r_{U,V}(f) : V \to \mathbb{C}$ to the set V. Then the following hold:

(S1) The following diagram is commutative:

$$\begin{array}{ccc} R(U) & \xrightarrow{r_{U,V}} & R(V) \\ & \searrow^{r_{U,W}} & \downarrow{\scriptstyle r_{V,W}} \\ & & R(W). \end{array} \tag{4.25}$$

All of the mappings $r_{U,V}, r_{V,W}, r_{U,W}$ are ring morphisms.

(S2) $r_{UU} = \mathrm{id}$.

Moreover, we have the following stronger properties which are closely related to analytic continuation. Let U be an open subset of X, and let $U_1, U_2, \ldots$ be a family of open subsets of X with

$$U = \bigcup_{j \in J} U_j$$

where the index set J is a finite or infinite set of positive integers.

(S3) The local-global principle: If $f, g \in R(U)$, then

$$r_{U,U_j}(f) = r_{U,U_j}(g) \quad \text{for all } j \in J \quad \text{implies} \quad f = g \text{ on } U.$$

(S4) The gluing condition: If $f_j \in R(U_j)$ for all $j \in J$ and we have

$$r_{U_j, U_j \cap U_k}(f_j) = r_{U_k, U_j \cap U_k}(f_k) \qquad \text{for all} \quad j, k \in J,$$

then there exists precisely one $f \in R(U)$ such that

$$r(U, U_j)(f) = f_j \qquad \text{for all} \quad j \in J.$$

The conditions (S1), (S2) (resp. (S1)–(S4)) are characteristic for pre-sheaves (resp. sheaves).

Germs. Consider a fixed point $x_0 \in X$. Let $\mathcal{T}(x_0)$ be the set of all the open subsets U of X which contain the point x_0. For $f \in R(U)$ and $g \in R(V)$ we write

$$f \sim g$$

iff f and g coincide on a sufficiently small neighborhood of the point x_0. More precisely, we assume that $U, V \in \mathcal{T}(x_0)$ and that there exists a set $W \in \mathcal{T}(x_0)$ with $W \subseteq U \cap V$ and the property

$$r_{U,W}(f) = r_{V,W}(g).$$

This is an equivalence relation. The equivalence classes are called germs of the pre-sheaf at the point x_0.

Example. In the special case of the sheaf of holomorphic functions on the Gaussian plane $\mathbb{C}$, two holomorphic functions

$$f : U \to \mathbb{C}$$

and $g : V \to \mathbb{C}$ belong to the same germ at the point z_0 iff they have the same power series expansion near the point z_0. In this case, the famous local-global principle of the theory of analytic functions tells us the crucial fact that the functions f and g have the same global analytic continuation. In particular, they have the same Riemann surface.

The pre-sheaf of smooth functions. Let X be a manifold (e.g., a real finite-dimensional linear space or a sphere). We assign to each open subset of X the ring $R(U)$ of smooth functions $f : U \to \mathbb{R}$. Then the conditions (S1) and (S2) above are satisfied. This way, we get the pre-sheaf $C^\infty_{\text{sheaf}}(X)$ of smooth functions of X. Two smooth functions

$$f : U \to \mathbb{C}$$

and $g : V \to \mathbb{C}$ belong to the same germ at the point $x_0 \in X$ iff they have the same values and the same derivatives of each order at the point x_0. Let $C^\infty_{x_0,\text{germ}}(X)$ denote the set of germs at the point x_0.

In contrast to germs of holomorphic functions, germs of smooth functions do not determine the global properties of smooth functions.

Historical remarks. It turns out that the language of pre-sheaves and sheaves is of fundamental importance for mathematics. In the 1950s, Jean Leray (1906–1998), Henri Cartan (born 1904), and Jean-Piere Serre (born 1926) showed that the theory of holomorphic functions of several variables can elegantly be formulated in terms of the cohomology of sheaves. In the 1950s and 1960s, Arthur Wightman and Res Jost emphasized the importance of the theory of holomorphic functions with several variables in axiomatic quantum field theory. In the 1960s, Alexandre Grothendieck (born 1928) based algebraic geometry and number theory on the new concept of schemes. Here, schemes are closely related to sheaf theory.

Hints for further reading. As an elementary introduction to sheaf theory and its applications, we refer to Chapter 19 of the handbook edited by Zeidler, Teubner-Taschenbuch der Mathematik II, 9th edition, Teubner, Wiesbaden, 2008 (in German) and to the following monographs:

K. Maurin, Methods of Hilbert Spaces, PWN, Warsaw, 1972.

K. Maurin, The Riemann Legacy: Riemannian Ideas in Mathematics and Physics of the 20th Century, Kluwer, Dordrecht, 1997.

R. Streater and A. Wightman, PCT, Spin, Statistics, and All That, Benjamin, New York, 1968.

For the relation between differential forms, de Rham cohomology, cohomology with values in a pre-sheaf (Čech cohomology), fiber bundles, and characteristic classes (e.g., Chern classes), we refer to:

R. Bott and L. Tu, Differential Forms in Algebraic Topology, Springer, New York, 1982.

As standard textbooks, we recommend:

O. Forster, Lectures on Riemann Surfaces, Springer, Berlin, 1981.

G. Bredon, Sheaf Theory, Springer, Berlin, 1997.

I. Shafarevich, Basic Algebraic Geometry, Vols. 1, 2, Springer, Berlin, 1994.

R. Hartshorne, Algebraic Geometry, Springer, New York, 1994.

D. Cox, J. Little, and D. O'Shea, Using Algebraic Geometry, Springer, New York, 1998.

Furthermore, we recommend the classic monograph:

F. Hirzebruch, Topological Methods in Algebraic Geometry, 3rd enlarged edition, Springer, New York, 1966.

4.4.4 Deformations, Mapping Classes, and Topological Charges

> Two continuous mappings are contained in the same mapping class iff they can be continuously deformed into each other. In important special cases, the space of mapping classes can be equipped with an additional group structure. This leads to Poincaré's fundamental group and the higher homotopy groups of topological spaces.
>
> Folklore

The space $[X, Y]$ of mapping classes. Let X and Y be topological spaces (e.g., subsets of $\mathbb{R}^n$). For two continuous maps $f, g : X \to Y$, we write

$$f \sim g$$

iff there exists a continuous map $H : X \times [0,1] \to Y$ with

$$H(x,0) = f(x) \quad \text{and} \quad H(x,1) = g(x) \qquad \text{for all} \quad x \in X.$$

This is an equivalence relation. The equivalence classes $[f]$ are called mapping classes. They form the space $[X, Y]$ of mapping classes from X to Y.

> *It is an important task of topology to describe the structure of mapping classes.*

This structure is only known for a collection of special cases. Mathematicians are looking for ever stronger tools in order to get new information about mapping classes. The crucial map

$$(x,t) \mapsto H(x,t)$$

is called a homotopy between the mappings f and g. Intuitively, the mappings f and g can be viewed as physical fields, and the homotopy H deforms

- the physical field f at the initial time $t = 0$
- into the physical field g at the final time $t = 1$.

The deformed value of $f(x)$ at time t is equal to $H(x,t)$. If $f \sim g$, then we also say that f is homotopically equivalent to g.

Examples. (i) The space $[\mathbb{S}^1, \mathbb{S}^1]$. For $m = 0, \pm 1, \pm 2, \ldots$, define the continuous map $\chi_m : \mathbb{S}^1 \to \mathbb{S}^1$ given by

$$\chi_m(\varphi) := m\varphi \qquad \text{for all} \quad \varphi \in \mathbb{R},$$

with respect to angle coordinates. Using complex numbers, the map χ_m corresponds to $\varphi \mapsto \mathrm{e}^{\mathrm{i}m\varphi}$. The map χ_m winds the original unit circle m times around itself. If

$m > 0$ (resp. $m < 0$), then we observe a clockwise (resp. counter-clockwise) winding. Each continuous map $f : \mathbb{S}^1 \to \mathbb{S}^1$ is homotopically equivalent precisely to one map χ_m. The number m is called the mapping degree $\deg(f)$ (or the winding number) of the map f.[23] Therefore, the mapping $f \mapsto \deg(f)$ yields the bijection

$$[\mathbb{S}^1, \mathbb{S}^1] \simeq \mathbb{Z}.$$

(ii) The space $[\mathbb{S}^2, \mathbb{S}^2]$. Example (i) can be generalized to higher-dimensional spheres. Using spherical coordinates, we define the continuous map $f : \mathbb{S}^2 \to \mathbb{S}^2$ by setting

$$\chi_m(\varphi, \vartheta) := (m\varphi, \vartheta) \qquad \text{for all} \quad \varphi \in \mathbb{R},\ \vartheta \in [-\tfrac{\pi}{2}, \tfrac{\pi}{2}].$$

Each continuous map $f : \mathbb{S}^2 \to \mathbb{S}^2$ is homotopically equivalent precisely to one map χ_m. The integer m is called the mapping degree $\deg(f)$ (or the winding number) of the map f. The mapping $f \mapsto \deg(f)$ yields the bijection

$$[\mathbb{S}^2, \mathbb{S}^2] \simeq \mathbb{Z}.$$

More generally, there holds the bijection

$$\boxed{[\mathbb{S}^n, \mathbb{S}^n] \simeq \mathbb{Z}, \qquad n = 1, 2, \dots} \tag{4.26}$$

This theorem is due to Heinz Hopf. In physics, the mapping degree $\deg(f)$ is called the topological charge of the physical field f. Analytic formulas for computing the mapping degree by means of the so-called Kronecker integral can be found in Sect. 5.7.3 of Vol. I.

(iii) We have the bijection

$$\boxed{[\mathbb{S}^3, \mathbb{S}^2] \simeq \mathbb{Z}.}$$

This famous theorem due to Heinz Hopf is closely related to both the electron spin in physics and the Hopf fibration of the 3-dimensional sphere in mathematics (see Sect. 5.7.2 of Vol. I).

(iv) The space $[\mathbb{S}^0, X]$: Two points x and y of a topological space X are called arcwise connected iff there exists a continuous map $c : [0, 1] \to X$ with

$$c(0) = x \qquad \text{and} \qquad c(1) = y.$$

Intuitively, the two points are connected by the continuous curve c. This is an equivalence relation. The equivalence classes are called the path components of the topological space X. Let $\mathbb{S}^0$ denote the 0-dimensional sphere (i.e., the point 0 on the real line). The topological space X is called arcwise connected iff it has precisely one path component.

> *There exists a bijection between the space of mapping classes* $[\mathbb{S}^0, X]$ *and the set of path components of the topological space* X.

For example, the interval $[0, 1]$ is arcwise connected, but the set

$$[0, 1] \cup [2, 3] \cup [4, 5]$$

[23] The general theory of mapping degree for mappings on finite-dimensional and infinite-dimensional Banach spaces is thoroughly studied in the author's monograph on fixed-point theory, Zeidler (1986), Vol. I (see the references on page 1049), including the Hopf theorem (4.26).

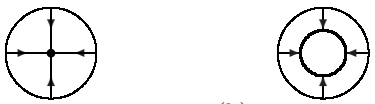


Fig. 4.11. Deformation retracts

has three path components. Furthermore, the space $\mathbb{R}^n$ is arcwise connected. Hence we have the bijection

$$[\mathbb{S}^0, \mathbb{R}^n] = \{0\}, \qquad n = 0, 1, 2, \ldots$$

The set $\mathbb{R}^2 \setminus \mathbb{S}^1$ has two path components (namely, the interior and the exterior of the unit circle). Therefore, we have the bijections

$$[\mathbb{S}^0, \mathbb{R}^2 \setminus \mathbb{S}^1] \simeq \{0, 1\} \simeq \mathbb{Z}_2.$$

Let C be a Jordan curve in $\mathbb{R}^2$ (i.e., C is a subset of $\mathbb{R}^2$ which is homeomorphic to $\mathbb{S}^1$). The classical Jordan curve theorem tells us that the set $\mathbb{R}^2 \setminus C$ has two path components.[24] Therefore, we have the bijection

$$[\mathbb{S}^0, \mathbb{R}^2 \setminus C] \simeq \mathbb{Z}_2.$$

Deformation retracts. Let Y be a subspace of the topological space X (i.e, Y is a subset of X equipped with the subspace topology – see page 243). By definition, Y is a retract of X iff there exists a continuous map

$$r : X \to Y$$

with $r(y) = y$ for all $y \in Y$. The map r is called a retraction. Furthermore, Y is called a deformation retract of X iff there exists a continuous map

$$H : X \times [0, 1] \to X$$

such that the following hold:

- $H(x, 0) = x$ for all $x \in X$,
- $H(x, 1) \in Y$ for all $x \in X$, and
- $H(y, 1) = y$ for all $y \in Y$.

This means that the identity map on X (at time $t = 0$) can be continuously deformed into a retraction $H(. , 1) : X \to Y$ (at time $t = 1$). For example,

- the midpoint of a disc is a deformation retract of the disc,
- and a fixed boundary circle of an annulus is a deformation retract of the annulus (Fig. 4.11).

Topologically equivalent topological spaces. Two topological spaces X and Y are called topologically equivalent iff there exists a homoemorphism

$$f : X \to Y.$$

Explicitly, this means that the map $f : X \to Y$ is continuous and bijective, and the inverse map $f^{-1} : Y \to X$ is also continuous. We also say that X and Y have the same topological type iff they are topologically equivalent.

[24] Camille Jordan (1838–1922).

Now let us introduce the weaker notion of homotopically equivalent topological spaces. Observe that topologically equivalent topological spaces are also homotopically equivalent, but the conserve is not always true. For example, the interval $[0,1]$ is topologically equivalent to any compact interval $[a,b]$ of positive length, but it is not topologically equivalent to the one-point interval $\{0\}$. However, the interval $[0,1]$ is homotopically equivalent to $\{0\}$.

Homotopically equivalent topological spaces. For two topological spaces, we write

$$X \sim Y$$

iff there exist two continuous maps $f : X \to Y$ and $g : Y \to X$ such that we have the homotopies

$$g \circ f \sim \mathrm{id}_X \qquad \text{and} \qquad f \circ g \sim \mathrm{id}_Y,$$

where id_X and id_Y are the identity maps on X and Y, respectively. In other words, the map $g \circ f$ (resp. $f \circ g$) can be continuously deformed into the identity map on X (resp. Y). We say that f and g are inverse up to homotopy. For topological spaces X, Y, Z, we have $X \sim X$. Moreover, $X \sim Y$ iff $Y \sim X$, and

$$X \sim Y, \quad Y \sim Z \quad \text{implies} \quad X \sim Z.$$

Consequently, $X \sim Y$ is an equivalence relation. We say that the topological spaces X and Y are homotopically equivalent (or the have the same homotopy type) iff $X \sim Y$. The following definitions will be frequently used:

- X is called *contractible* iff it is homotopically equivalent to a one-point space. Intuitively, this means that the space X can be continuously contracted into a point.
- X is called *arcwise connected* iff the space $[\mathbb{S}^0, X]$ of mapping classes consists precisely of one point. Intuitively, this means that two points of X can always be connected by a continuous curve.
- X is called *simply connected* iff the space $[\mathbb{S}^1, X]$ of mapping classes consists precisely of one point. Intuitively, this means that each closed continuous curve in X can be continuously contracted into a point.

Examples. (i) For $n = 1, 2, \ldots$, each convex set of $\mathbb{R}^n$ is contractible (e.g., an interval, a disc in $\mathbb{R}^2$, or a ball in $\mathbb{R}^n$).

(ii) A disc is arcwise connected and simply connected.

(iii) A circle (or an annulus) is arcwise connected, but not simply connected.

(iv) An annulus is homotopically equivalent to a circle (Fig. 4.11).

(v) If Y is a deformation retract of the topological space X, then Y and X are homotopically equivalent.

Cyclic groups. The following notions will be used in the next section. The set

$$\mathcal{Z} := \{\mathrm{e}^n : n = 0, \pm 1, \pm 2, \ldots\}$$

is a multiplicative group called the infinite cyclic group. We have the group isomorphism

$$\mathbb{Z} \simeq \mathcal{Z}$$

from the additive group $\mathbb{Z}$ onto the multiplicative group $\mathcal{Z}$. This isomorphism is given by the map $n \mapsto \mathrm{e}^n$. Let $m = 1, 2, \ldots$ The set

$$\mathcal{Z}_m := \{\mathrm{e}^{2\pi m \mathrm{i}/k} : k = 0, 1, \ldots, m-1\}$$

is a multiplicative group called the cyclic group of order m. We have the group isomorphism

$$\mathbb{Z}_m \simeq \mathcal{Z}_m$$

from the additive group $\mathbb{Z}_m$ onto the multiplicative group $\mathcal{Z}_m$. This isomorphism is given by the map

$$[k] \mapsto \mathrm{e}^{2\pi \mathrm{i} m/k}$$

where the equivalence class $[k]$ modulo m is equal to the set $k + m\mathbb{Z}$. Explicitly, we have $[k] := \{k + nm : n = 0, \pm 1, \pm 2, \ldots\}$. For example, we get

$$\mathcal{Z}_1 = \{1\}, \quad \mathcal{Z}_2 = \{1, -1\}, \quad \mathcal{Z}_3 = \{1, \mathrm{e}^{2\pi \mathrm{i}/3}, \mathrm{e}^{-2\pi \mathrm{i}/3}\}, \quad \mathcal{Z}_4 := \{1, \mathrm{i}, -1, -\mathrm{i}\}.$$

The direct product of two groups. If G and H are groups, then the Cartesian product

$$G \times H := \{(g, h) : \ g \in G, h \in H\}$$

becomes a group if we use the multiplication $(g, h)(g^+, h^+) := (gg^+, hh^+)$. The group $G \times H$ is called the direct product between the groups G and H.

4.4.5 Poincaré's Fundamental Group

Basic idea. Consider the unit circle $\mathbb{S}^1 := \{z \in \mathbb{C} : \ |z| = 1\}$. Let us recall the definition $\chi_n(\varphi) := \mathrm{e}^{\mathrm{i}n\varphi}$. The space of mapping classes $[\mathbb{S}^1, \mathbb{S}^1]$ can be equipped with a group structure by defining the multiplication

$$[\chi_m][\chi_k] := [\chi_{m+k}] \qquad \text{for all} \quad k, m \in \mathbb{Z}.$$

This group denoted by $\pi_1(\mathbb{S}^1)$ is Poincaré's fundamental group of the unit circle. We have

$$\pi_1(\mathbb{S}^1) = \mathcal{Z}.$$

Using the group isomorphism $\mathbb{Z} \simeq \mathcal{Z}$ given by $m \mapsto \chi_m$, we get

$$\pi_1(\mathbb{S}^1) \simeq \mathbb{Z}.$$

Here, $\pi_1(\mathbb{S}^1)$ is regarded as an additive group. Let us generalize this. It turns out that the fundamental group is a multiplicative group, which is not always commutative. Therefore, the fundamental group is not always isomorphic to an additive group.

The product of loops. Let X be an arbitrary topological space. Fix the point x_0 in X. Each continuous map

$$c : \mathbb{S}^1 \to X \qquad \text{with} \qquad c(1) = x_0 \tag{4.27}$$

is called a loop in X with the base point x_0. By definition, the set of all such loops forms the loop space $\Omega_{x_0} X$. Intuitively, the space $\Omega_{x_0} X$ consists of all the closed curves in the topological space X which pass through the point x_0. For two loops $c, d \in \Omega_{x_0} X$, we define the product $c \times d$ by using the angular coordinate $\varphi \in [0, 2\pi]$ and by setting

$$(c \times d)(\varphi) := \begin{cases} c(2\varphi) & \text{if} \quad 0 \le \varphi \le \pi, \\ d(2\varphi - 2\pi) & \text{if} \quad \pi \le \varphi \le 2\pi. \end{cases}$$

Then, $c \times d \in \Omega_{x_0} X$. Intuitively, this means that we glue the two curves together at the point x_0 (Fig. 4.12).

The fundamental group $\pi_1(X, x_0)$. Our next goal is to identify homotopic loops with each other. To this end, for $c, d \in \Omega_{x_0} X$, we write

$$c \sim d \mod x_0 \tag{4.28}$$

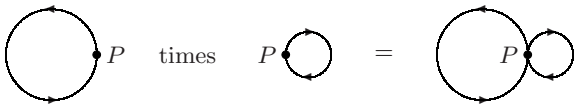


Fig. 4.12. Multiplication of two loops

iff the two curves are homotopic *modulo the point* x_0. Explicitly, this is defined in the following way: we assume that there exists a continuous map $H : \mathbb{S}^1 \times [0,1] \to X$ such that we have

$$H(\varphi, 0) = c(\varphi) \qquad \text{and} \qquad H(\varphi, 1) = d(\varphi) \qquad \text{for all} \quad \varphi \in [0, 2\pi]$$

together with $H(0,t) = H(2\pi, t) = x_0$ for all times $t \in [0,1]$. This is an equivalence relation. Intuitively, during the time interval $[0,1]$, the closed curve c is continuously deformed into the closed curve d in such a way that all of the deformed curves pass through the point x_0. By definition, the equivalence classes $[c]$ form the set $\pi_1(X, x_0)$. It is crucial that the product of loops is compatible with the equivalence relation (4.28). Explicitly, if $c \sim c^+ \mod x_0$ and $d \sim d^+ \mod x_0$, then

$$c \times c^+ \sim d \times d^+ \mod x_0.$$

Therefore, the set $\pi_1(X, x_0)$ of equivalence classes $[c]$ forms a group with respect to the following product:

$$[c][d] := [c \times d].$$

The unit element of this group is given by the equivalence class $[c]$ with $c(\varphi) := x_0$ for all $\varphi \in [0, 2\pi]$.

The group $\pi_1(X)$**.** If the group $\pi_1(X, x_0)$ does not depend on the choice of the base point x_0, up to group isomorphism, then we simply write $\pi_1(X)$ instead of $\pi_1(X, x_0)$. In particular, this is the case if the topological space X is arcwise connected.

> *The arcwise connected topological space X is simply connected iff the fundamental group $\pi_1(X)$ is trivial, that is, it consists of precisely one point.*

In this case, we briefly write $\pi_1(X) = 0$. In order to rigorously compute fundamental groups and higher homotopy groups, one uses exact sequences. This will be investigated in Vol. IV. Let us mention three examples.

(i) For the circle $\mathbb{S}^1$, we get $\pi_1(\mathbb{S}^1) = \mathcal{Z}$. Regarded as an additive group, we obtain $\pi_1(\mathbb{S}^1) = \mathbb{Z}$.
(ii) The 2-dimensional sphere $\mathbb{S}^2$ is simply connected, and hence $\pi_1(\mathbb{S}^2) = 0$.
(iii) The 2-dimensional torus $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$: Regarded as an additive group, we get $\pi_1(\mathbb{T}^2) = \mathbb{Z} \oplus \mathbb{Z}$.

Let us heuristically discuss the situation (iii). We choose two circles on the torus, namely, the equator e and a fixed meridian m. In Fig. 4.8 on page 207, the equator (resp. meridian) corresponds to the curve ABA (resp. ACA). Intuitively, the homotopy equivalence classes of loops on the torus have the form

$$[e]^r [m]^s, \qquad r, s \in \mathbb{Z}$$

together with $[e][m] = [m][e]$. In terms of an additive group, the equivalence classes have the form

$$r[e] + s[m], \qquad r, s \in \mathbb{Z}.$$

This yields $\pi_1(\mathbb{T}^2) = \mathbb{Z} \oplus \mathbb{Z}$.

The homotopy group $\pi_0(X)$. If the topological space X is arcwise connected, then we define

$$\pi_0(X) := \{0\}.$$

We briefly write $\pi_0(X) = 0$. If X has $m+1$ path components, $m = 0, 1, \ldots$, then we define

$$\pi_0(X) := \mathbb{Z}_m.$$

This is an additive group. Generally, the symbol $\pi_0(X)$ denotes the set of path components of the topological space X.

4.4.6 Loop Spaces and Higher Homotopy Groups

> The most intuitively evident topological invariant of a topological space is the number of connected pieces into which it falls. Over the past one hundred years or so we have come to realize that this primitive notion admits in some sense two higher-dimensional analogues. These are the *homotopy* and *cohomology groups* of the spaces in question[25]...
> By some divine justice the homotopy groups of a finite polyhedron or a manifold seem as difficult to compute as they are easy to define ... To this day not all the homotopy groups of say the 2-sphere have been computed.[26]
>
> Raoul Bott and Loring Tu, 1982

We will base our approach to higher homotopy groups on the study of loop spaces. Intuitively, loops are closed continuous curves. In terms of physics, the transport of physical information along loops is crucial, since it is possible that the initial information differs from the final information after one surrounding.

This gain (or loss) of physical information is caused by interactions.

Loops play a fundamental role in gauge theory (e.g., Wilson loops) and in the Ashtekar program for quantum gravitation (called loop gravity).[27]

Loop spaces. Consider the loop space $\Omega_{x_0}X$ introduced on page 223. The set $\Omega_{x_0}X$ becomes a topological space equipped with the compact-open topology. In order to introduce this topology, let C be a compact subset of $\mathbb{S}^1$, and let O be an open subset of X. Let $\mathcal{U}(C, O)$ denote the set of all the maps of the form (4.27) on page 223 with

$$c(C) \subseteq O.$$

All the sets $\mathcal{U}(C, O)$ form a subbasis of the compact-open topology on the loop space $\Omega_{x_0}X$. The definition of a subbasis of a topology can be found on page 241. In what follows, loop spaces are always equipped with the compact-open topology. Now, in order to get higher homotopy groups, our idea is to use the loop space of the loop space $\Omega_{x_0}X$, and so on.

The homotopy groups $\pi_k(X, x_0)$, $k \geq 2$. We define the second homotopy group of X by setting

$$\boxed{\pi_2(X, x_0) := \pi_1(\Omega_{x_0}X, c_0).}$$

[25] The basic relation between potentials in physics and cohomology are discussed in Sect. 16.8 of Vol. I.

[26] R. Bott and L. Tu, Differential Forms in Algebraic Topology, Springer, New York, 1982 (reprinted with permission).

[27] A. Ashtekar, M. Bojewald, and J. Lewandowski, Mathematical structure of loop quantum cosmology, Adv. Theor. Math. Phys. **7** (2003), 233–268.

Here, the symbol c_0 denotes the constant loop $c_0 : \mathbb{S}^1 \to \{x_0\}$. In the general case, we proceed by induction. For $k = 1, 2, \ldots$, we use the inductive definition

$$\boxed{\pi_k(X, x_0) := \pi_{k-1}(\Omega_{x_0} X, c_0), \qquad k = 2, 3, \ldots}$$

of the kth homotopy group of the topological space X.

The homotopy groups $\pi_k(X, x_0)$ with $k \geq 2$ are commutative.

The homotopy groups $\pi_k(X)$. If the groups $\pi_k(X, x_0), k = 1, 2, \ldots$, do not depend on the choice of the base point x_0, up to group isomorphism, then we simply write $\pi_k(X)$ instead of $\pi_k(X, x_0)$. In particular, this is the case if the topological space X is arcwise connected.

Mapping classes and homotopy groups. Let the topological space X be arcwise connected. Then there exist the bijections

$$[\mathbb{S}^k, X] \simeq \pi_k(X), \qquad k = 0, 1, 2, \ldots$$

Invariance of homotopy groups. The following result is crucial.

Theorem 4.16 *Homotopically equivalent topological spaces (e.g., homeomorphic spaces) have the same homotopy groups.*

This has the following consequence: Two topological spaces X and Y are *not* homotopically equivalent (and hence not homeomorphic) if there exists an index $k = 0, 1, 2, \ldots$ with $\pi_k(X) \neq \pi_k(Y)$.

Examples. The following homotopy groups are written as additive groups.

(i) Point: If the topological space X consists of precisely one point, then

$$\pi_k(X) = 0, \qquad k = 0, 1, 2, \ldots$$

A contractible space X has the same homotopy groups, since such a space is homotopically equivalent to a point.

(ii) Unit circle: $\pi_1(\mathbb{S}^1) = \mathbb{Z}$, and $\pi_k(\mathbb{S}^1) = 0$ if $k = 0$ or $k \geq 2$. An annulus or the surface of a cylinder have the same homotopy groups, since these topological spaces are homotopically equivalent to the unit circle. It follows from $\pi_1(\mathbb{S}^1) \neq 0$ that the unit circle is not contractible.

(iii) 2-dimensional sphere: $\pi_0(\mathbb{S}^2) = \pi_1(\mathbb{S}^2) = 0$ and $\pi_1(\mathbb{S}^2) = \pi_3(\mathbb{S}^2) = \mathbb{Z}$. Furthermore, $\pi_k(\mathbb{S}^2) = \mathbb{Z}_2$ if $k = 4, 5, 7, 8, 11$, and

$$\pi_6(\mathbb{S}^2) = \mathbb{Z}_{12}, \;\; \pi_9(\mathbb{S}^2) = \mathbb{Z}_3, \;\; \pi_{10}(\mathbb{S}^2) = \mathbb{Z}_{15}, \quad \pi_{22}(\mathbb{S}^2) = \mathbb{Z}_{132} \oplus \mathbb{Z}_2.$$

(iv) n-dimensional sphere: For $n = 1, 2, \ldots$, we have

$$\pi_n(\mathbb{S}^n) = \mathbb{Z}, \qquad \pi_k(\mathbb{S}^n) = 0, \quad k = 0, \ldots, n-1.$$

The product theorem. For the Cartesian product $X \times Y$ of two topological spaces X and Y, we get

$$\pi_k(X \times Y, (x_0, y_0)) = \pi_k(X, x_0) \times \pi_k(Y, y_0), \qquad k = 1, 2, \ldots,$$

in the sense of a direct group product.

Serre's finiteness theorem. In his 1951 thesis, Serre proved the following fundamental result: Let $m = 0, 1, \ldots$ and $n = 1, 2, \ldots$

With the only exceptions $\pi_n(\mathbb{S}^n) = \mathbb{Z}$ and $\pi_{4n-1}(\mathbb{S}^{2n})$, all the homotopy groups $\pi_m(\mathbb{S}^n)$ of spheres are finite.

The exceptional homotopy group $\pi_{4n-1}(\mathbb{S}^{2n})$ is the direct sum of $\mathbb{Z}$ with a finite group.[28] A list of known homotopy groups of spheres can be found in the Appendix to:

> C. Dodson and P. Parker, A Users' Guide to Algebraic Topology, Kluwer, Dordrecht, 1997.

Historical remark. The fundamental group dates back to Poincaré (1854–1912). The higher homotopy groups were introduced by Hurewicz (1904–1956) in the 1930s.

4.4.7 Homology, Cohomology, and Electrodynamics

The notions of homology and cohomology are deeply rooted in electrodynamics. This will be thoroughly studied in Vol. IV. In particular, we will show that the most simple approach to homology and cohomology groups is related to electrical circuits. This generalizes then to the Maxwell equations in electrodynamics by using differential forms and the de Rham cohomology.

4.4.8 Bott's Periodicity Theorem

> I was very fortunate to be the first to notice that the loop space of a Lie group is very easily attacked with Morse-theoretic methods.[29] It turns out if you look at the loop space rather than at the group, then the so-called diagram of the group on the universal covering of its maximal torus plays an essential role. So you can read off topological properties of the loop space much more easily from the diagram of the group than you can read off things about the group itself...
> During the period 1955–57 there was a controversy in homotopy theory. The question concerned the homotopy group of the unitary group in dimension 10. The homotopy theorists said it was the cyclic group $\mathbb{Z}_3$. The results of Borel and Hirzebruch predicted it to be 0. This contradiction intrigued me, and I thought I should be able to say something about it using Morse-theoretic techniques that Samelson and I had discovered. Finally I hit upon a very complicated method involving the exceptional Lie group G_2 to check the conundrum independently. My good friend Arnold Shapiro and I spent all weekend computing. At the end we came out of the side of Borel and Hirzebruch, so I was convinced that they were right. And if they were right the table of homotopy groups started to look periodic for a long stretch. In the odd dimensions they were $\mathbb{Z}$ up to nine dimensions, and in the even dimensions they were 0. So I thought, "Maybe they are periodic all the way." And fairly soon after I saw my old ideas would actually do the job. In this way the unitary group was then settled.
> Then I started to think about the orthogonal group, and that was much harder. But I do remember precisely when I suddenly saw how to deal

[28] Jean-Pierre Serre (born 1926) was awarded both the Fields medal in 1954 and the Abel prize in 2003.

[29] R. Bott, On torsion in Lie groups, Proc. Nat. Academy Sci. U.S.A. **40** (1954), 586–588.

> with it. That occurred after we had left the Institute in Princeton and were moving house. In a flash I saw how it all fitted together.[30]
>
> Raoul Bott, 2001

Let $U(n)$ be the group of complex unitary $(n \times n)$-matrices. It is crucial that the homotopy groups

$$\pi_k(U(n)), \qquad k = 0, 1, 2, \dots$$

are independent of n if the natural number n is sufficiently large. Explicitly, we need $0 \leq k < 2n$. In this case, we briefly write $\pi_k(U)$. These stable homotopy groups are completely known. In fact, the fundamental Bott periodicity theorem says that

$$\pi_0(U) = 0, \quad \pi_1(U) = \mathbb{Z}, \quad \pi_{k+2}(U) = \pi_k(U), \quad k = 0, 1, 2, \dots$$

This corresponds to the period 2.

Let $O(n)$ be the group of real orthogonal $(n \times n)$-matrices. If n is sufficiently large, that is, $0 \leq k < n - 1$, then the homotopy groups

$$\pi_k(O(n))$$

are independent of n. We briefly write $\pi_k(O)$. These stable homotopy groups are also completely known. In this case, the Bott periodicity theorem says that

$$\pi_{k+8}(O) = \pi_k(O), \qquad k = 0, 1, 2, \dots$$

together with $\pi_0(O) = \pi_1(O) = \mathbb{Z}_2$, and

$$\pi_j(O) = 0, \quad j = 2, 4, 5, 6, \quad \pi_m(O) = \mathbb{Z}, \quad m = 3, 7.$$

This corresponds to the period 8.

4.4.9 *K*-Theory

> K-theory was introduced by Grothendieck in his formulation of the Riemann–Roch–Hirzebruch theorem in 1958.[31] For each projective algebraic variety, Grothendieck constructed a group from the category of coherent algebraic sheaves, and showed that it had many nice properties. About 1960, Atiyah and Hirzebruch constructed a topological analog defined for any compact space X, a group $K(X)$ constructed from the category of vector bundles on X... Topological K-theory has become an important tool in topology. Using K-theory, Adams and Atiyah were able to give a simple proof that the only spheres which can be provided with H-space structures are the spheres of dimension 1, 3, and 7. Moreover, Adams was able to derive a substantial part of stable homotopy theory from K-theory. Further applications to analysis are found in the work of Atiyah–Singer on the index theorem, Quillen, Bass, and others. A key factor in these applications is the Bott periodicity for Lie groups.[32]
>
> Max Karoubi, 1978.

[30] Interview with Raoul Bott, Notices Amer. Math. Soc. **48**(4) (2001), 374–382 (reprinted with permission); R. Bott, The stable homotopy of the classical groups, Ann. of Math. **70** (1959), 313–337.

[31] Alexandre Grothendieck (born 1928) was awarded the Fields medal in 1966.

[32] M. Karoubi, K-Theory: An Introduction, Springer, Berlin, 1978 (reprinted with permission).

Semi-rings. Consider the set $\mathbb{N}$ of natural numbers $0, 1, 2, \ldots$ Set $S := \mathbb{N}$. Then, for all $m, n, s \in S$, the following hold.

(S0) Consistency: $m + n, mn \in S$.
(S1) Associativity: $(m + n) + s = m + (n + s)$ and $(mn)s = m(ns)$.
(S2) Zero element: There exists a uniquely determined element 0 in S with the property $m + 0 = m$ for all $m \in S$.
(S3) Commutativity: $m + n = n + m$ and $mn = nm$.
(S4) Regularity: It follows from $m, n \in S$ and $m + s = n + s$ for all $s \in S$ that $m = n$.

In general, a set S is called a semi-ring iff an addition $m+n$ and a multiplication mn exist for certain elements of S such that for all $m, n, s \in S$ the conditions (S0)–(S2) are satisfied. The semi-ring is called commutative (resp. regular) iff, in addition, condition (S3) (resp. (S4)) is satisfied.

The map $\chi : S \to S'$ from the semi-ring S into the semi-ring S' is called a semi-ring morphism iff it respects sums and products, that is,

$$\chi(m + n) = \chi(m) + \chi(n), \qquad \chi(mn) = \chi(m)\chi(n) \qquad \text{for all} \quad m, n \in S.$$

Bijective semi-ring morphisms are called semi-ring isomorphisms.

The universal ring extension $K(S)$ of a regular semi-ring S. The commutative semi-ring $\mathbb{N}$ of natural numbers $0, 1, 2, \ldots$ can be extended to the commutative ring $\mathbb{Z}$ of integers $0, \pm 1, \pm 2, \ldots$ We want to generalize this. In terms of integers, the idea of the following general construction is to identify, say, the integer -2 with the following family of pairs:

$$(0, 2), (2, 4), (4, 6), \ldots$$

Note that $-2 = 0 - 2 = 2 - 4 = \ldots$ Our approach will be based on the following commutative diagram:

$$\begin{array}{ccc} S & \xrightarrow{\ i\ } & K(S) \\ & \searrow^{\sigma} & \downarrow{\scriptstyle \varrho} \\ & & R \end{array} \tag{4.29}$$

Our goal is to extend the regular commutative semi-ring S to a ring $K(S)$ in a universal way. This means that all the ring morphisms $\varrho : K(S) \to R$ generate all possible semi-ring morphisms $\sigma : S \to R$ from the semi-ring S into the ring R.

Theorem 4.17 *Let S be a regular commutative semi-ring. Then:*

(i) *There exist a commutative ring $K(S)$ and an injective semi-ring morphism $i : S \to K(S)$.*

(ii) *Each extension of the semi-ring S to a commutative ring $\mathcal{R}$ contains a subring which is isomorphic to $K(S)$.*

(iii) *Each ring morphism $\varrho : K(S) \to R$ from the ring $K(S)$ into the arbitrary ring R generates the semi-ring morphism $\varrho \circ i : S \to R$ (i.e., the diagram (4.29) is commutative).*

(iv) *Conversely, each semi-ring morphism $\sigma : S \to R$ from the semi-ring S into the ring R can be obtained by* (iii).

Proof. We sketch the proof, and we recommend the reader to complete the proof by himself/herself.

Ad (i). (I) Equivalence classes. For $m, n, r, s \in S$, we consider the symbols (m, n) and (r, s).[33] We write $(m, n) \sim (r, s)$ iff

$$m + s = r + n.$$

This is an equivalence relation. Let us prove the transitivity. Here, we will need the regularity of S. Suppose that

$$[m, n] \sim [r, s] \quad \text{and} \quad [r, s] \sim [u, v].$$

Then $m + s = r + n$ and $r + v = u + s$. Adding this, we get

$$(m + v) + (s + r) = (u + n) + (s + r).$$

Since the semi-ring S is assumed to be regular, we obtain $m + v = u + n$. Hence

$$[m, n] \sim [u, v].$$

The corresponding equivalence classes $[(m, n)]$ form the space $K(S)$.

(II) Addition and multiplication of equivalence classes: We define

$$[(m, n)] + [(r, s)] := [(m + r, n + s)], \qquad [(m, n)][(r, s)] := [(mr + ns, ms + nr)].$$

These operations do not depend on the choice of the representatives. With respect to these operations, $K(S)$ becomes a commutative ring. For each m, define

$$i(m) := [(m, 0)].$$

The map $i : S \to K(S)$ is injective. In fact, if $i(m) = 0$, then $[(m, 0)] = [(0, 0]$. Hence $m + 0 = 0 + 0$. This implies $m = 0$.

Ad (ii). If the ring R is an extension of the semi-ring S, then we define the map $\varrho : K(S) \to R$ given by

$$\varrho([(m, n)]) := m - n.$$

This definition does not depend on the choice of the representatives (m, n). The map ϱ is injective. Therefore, we can identify the image $\varrho(S)$ with S.

Ad (iii). This is obvious.

Ad (iv). Define $\varrho([(m, 0)]) := \sigma(m)$. □

Example. Let S be the set of the linear spaces $\mathbb{R}^0, \mathbb{R}^1, \mathbb{R}^2, \ldots$ Naturally enough, $\mathbb{R}^0 := \{0\}$. For all $m, n \in \mathbb{N}$, we define the direct sum $\oplus$ and the tensor product $\otimes$ by setting

$$\mathbb{R}^m \oplus \mathbb{R}^n := \mathbb{R}^{m+n}, \qquad \mathbb{R}^n \otimes \mathbb{R}^n := \mathbb{R}^{mn}.$$

The set S is a semi-ring which is isomorphic to the semi-ring $\mathbb{N}$, and the ring $K(S)$ is isomorphic to the ring $\mathbb{Z}$ of integers. Define the fiber dimension map $j : K(S) \to \mathbb{Z}$ by setting

$$j : ([\mathbb{R}^m, \mathbb{R}^n)]) := m - n. \tag{4.30}$$

This map does not depend on the choice of the representatives. In fact, if

$$[(\mathbb{R}^m, \mathbb{R}^n)] = [(\mathbb{R}^r, \mathbb{R}^s)],$$

then $\mathbb{R}^m \oplus \mathbb{R}^s = \mathbb{R}^r \oplus \mathbb{R}^s$. Hence $m + s = r + n$. This implies $m - n = r - s$.

[33] Intuitively, if m and n are natural numbers, then (m, n) corresponds to the integer $m - n$. We use equivalence classes, since different symbols (m, n) and (r, s) may yield the same integer. Note that $m - n = r - s$ iff $m + s = r + n$.

Topological K-theory represents a highly non-trivial generalization of this simple example.

Let us discuss this.

The universal Grothendieck ring $K(S)$ for general commutative semi-rings. Each ring is regular. Therefore, if a ring $\mathcal{R}$ contains a subset S which is a semi-ring, then S is regular. Consequently, if the given semi-ring S is not regular, then it cannot be extended to a ring. The following theorem yields the optimal generalization of Theorem 4.17.

Theorem 4.18 *For each commutative semi-ring S, the following universal statements hold:*

(i) *There exist a ring $K(S)$ and a semi-ring morphism $i : S \to K(S)$.*

(ii) *Each ring morphism $\varrho : K(S) \to R$ from the ring $K(S)$ into the arbitrary ring R generates the semi-ring morphism $\varrho \circ i : S \to R$ (i.e., the diagram (4.29) above is commutative).*

(iii) *Conversely, each semi-ring morphism $\sigma : S \to R$ from the semi-ring S into the ring R can be obtained by* (ii).

(iv) *If the semi-ring S is regular, then the map $i : S \to K(S)$ is injective.*

Proof. We have to modify the proof of Theorem 4.17 in the following way. For $m, n, r, s \in S$, we write $(m, n) \sim (r, s)$ iff there exists an element p of S such that

$$m + n + p = r + s + p.$$

This modification is needed in order to guarantee the transitivity of the equivalence relation in the case where the semi-ring S is not regular. □

Prototype. Choose $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. Let $\mathrm{Vect}_{\mathbb{K}}$ denote the semi-ring of all finite-dimensional linear spaces X over $\mathbb{K}$ with respect to the direct sum $\oplus$ and the tensor product $\otimes$. Here, isomorphic linear spaces are identified with each other. One shows as above that the corresponding K-ring $K(\mathrm{Vect}_{\mathbb{K}})$ is isomorphic to the ring $\mathbb{Z}$.

Application to vector bundles on a compact manifold. Let M be a finite-dimensional compact manifold (e.g., a sphere). Choose $\mathbb{K}^n$ with $n = 0, 1, 2, \ldots$. Here, $\mathbb{K} = \mathbb{R}$ (real case) or ($\mathbb{K} = \mathbb{C}$) (complex case). Let $\mathrm{Vect}_{\mathbb{K}}(M)$ be the space of vector bundles over M with typical fiber $\mathbb{K}^n$, where $n = 0, 1, 2, \ldots$ More precisely, two isomorphic vector bundles are identified with each other. If $V, W \in \mathrm{Vect}_{\mathbb{K}}(M)$, then the direct sum and the tensor product

$$V \oplus W, \qquad V \otimes W$$

are well-defined by Sect. 4.4.2. This way, the space $\mathrm{Vect}_{\mathbb{K}}(M)$ becomes a semi-ring. The Grothendieck ring $K(\mathrm{Vect}_{\mathbb{K}}(M))$ is called the K-ring of the manifold M. To simplify notation, we write $K_{\mathbb{K}}(M))$ instead of $K(\mathrm{Vect}_{\mathbb{K}}(M))$.

This way, we assign to each compact manifold M the ring $K_{\mathbb{K}}(M)$.

It turns out that this is a sophisticated tool in modern topology. Each element of $K_{\mathbb{K}}(M)$ is an equivalence class $[(V, W)]$ where V and W are vector bundles (up to bundle isomorphisms) with typical fibers $\mathbb{K}^m$ and $\mathbb{K}^n$, respectively. We define the map $j : K_{\mathbb{K}}(M) \to \mathbb{Z}$ by setting

$$j([V, W]) := m - n.$$

Similarly as in (4.30), this definition does not depend on the choice of the representatives. For an arcwise connected manifold M, we define the reduced K-ring

$$\mathcal{K}(M) := j^{-1}(0).$$

It turns out that the map $j : K_{\mathbb{R}}(M) \to \mathbb{Z}$ is a ring epimorphism. If we introduce the equivalence relation $A \sim B$ iff $A - B \in \mathcal{K}(M)$, then the K-ring $K_{\mathbb{R}}(M)$ is decomposed in the equivalence classes $[A]$. Each of these classes is bijective to the reduced K-ring $\mathcal{K}(X)$. Furthermore,

$$K_{\mathbb{R}}(M)/\mathcal{K}(M) \simeq \mathbb{Z}.$$

That is, the quotient ring of the ring $K_{\mathbb{R}}(M)$ modulo the reduced ring $\mathcal{K}(M)$ is isomorphic to the ring of integers. The following deep theorem tells us that the additive group of the ring $\mathcal{K}(\mathbb{S}^n)$ is known for all spheres $\mathbb{S}^n$, $n = 1, 2, \ldots$

Theorem 4.19 *The additive groups $\mathcal{K}(\mathbb{S}^n)$ are given by*

$$\mathcal{K}(\mathbb{S}^1) = \mathcal{K}(\mathbb{S}^2) = \mathbb{Z}/\text{mod } 2, \qquad \mathcal{K}(\mathbb{S}^4) = \mathcal{K}(\mathbb{S}^8) = \mathbb{Z},$$

and $\mathcal{K}(\mathbb{S}^n) = 0$ if $n = 3, 5, 6, 7$.

Moreover, we have the periodicity property $\mathcal{K}(\mathbb{S}^n) = \mathcal{K}(\mathbb{S}^{n+8})$ for all dimensions $n = 1, 2, \ldots$

The following result is a special case of the Atiyah–Jänich theorem which will be considered on page 235. Let $F(\mathcal{H})$ denote the space of all linear continuous Fredholm operators $A : \mathcal{H} \to \mathcal{H}$ on the complex infinite-dimensional separable Hilbert space $\mathcal{H}$.

Theorem 4.20 *For $n = 1, 2, \ldots$, we have*

$$K_{\mathbb{C}}(\mathbb{S}^n) \simeq \pi_n(F(\mathcal{H})), \qquad n = 1, 2, \ldots,$$

in the sense of a bijection.

In particular, the K-ring $K_{\mathbb{C}}(\mathbb{S}^1)$ of the unit circle $\mathbb{S}^1$ is bijective to the fundamental group $\pi_1(F(\mathcal{H}))$ of the space of linear continuous Fredholm operators on the complex infinite-dimensional separable Hilbert space $\mathcal{H}$.

This shows that the sophisticated K-ring $K_{\mathbb{C}}(\mathbb{S}^1)$ of the unit circle sees the topological structure of the space $F(\mathcal{H})$ of linear continuous Fredholm operators on the Hilbert space $\mathcal{H}$. The index of linear continuous Fredholm operators is the crucial quantity in the formulation of the Atiyah–Singer index theorem. In fact, it turns out that K-theory can be used in order to prove the Atiyah–Singer theorem about the index of elliptic partial differential equations (or more general pseudo-differential equations) on compact Riemannian manifolds. The prototype of this result can be found in Sect. 5.6.9 of Vol. I (see the hints for further reading on page 235).

The theorem of Adams about vector fields on spheres. Choose the number $n = 1, 2, 3, \ldots$ By definition, the number $\text{Span}(\mathbb{S}^n)$ is equal to the maximal number of continuous tangent vector fields on the n-dimensional sphere $\mathbb{S}^n$ which are linearly independent at every point of the sphere. Let $k(n)$ be the number of factors 2 in the prime number decomposition of the even natural number $n + 1$ (e.g., $48 = 2^3 \cdot 3^2$ implies $k(48) = 3$). Dividing $k(n)$ by 4, we get

$$k(n) = 4m(n) + r(n)$$

where $r(n) = 0, 1, 2, 3$ and $k(n) = 0, 1, 2, \ldots$ (e.g., $m(48) = 0$ and $r(48) = 3$).

Theorem 4.21 *If n is even, then* Span $(\mathbb{S}^n) = 0$. *If n is odd, then*

$$\text{Span}(\mathbb{S}^n) = 8m(n) + 2^{r(n)} - 1.$$

Using K-theory, this famous theorem was proven by J. Adams, Vector fields on spheres, Ann. of Math. **75** (1962), 603–632. For example,

$$\text{Span}(\mathbb{S}^1) = 1,\ \text{Span}(\mathbb{S}^3) = 3,\ \text{Span}(\mathbb{S}^5) = 1,\ \text{Span}(\mathbb{S}^7) = 7.$$

Furthermore, $\text{Span}(\mathbb{S}^{4k+1}) = 1$ and $(\mathbb{S}^{8k+3}) = 3$ if $k = 0, 1, 2, \ldots$ Finally, $\text{Span}(\mathbb{S}^n) = n$ iff $n = 1, 3, 7$. In terms of physics, the theorem of Adams tells us an important information about the velocity fields on n-dimensional spheres.

The theorem of Kervaire and Milnor on division algebras. A real finite-dimensional nonzero algebra $\mathcal{A}$ is called a division algebra iff for given elements a, b of $\mathcal{A}$ with $a \neq 0$, the two equations

$$ax = b, \qquad ay = b$$

have unique solutions x and y in $\mathcal{A}$. Examples for division algebras are $\mathbb{R}$ (the field of real numbers), $\mathbb{C}$ (the field of complex numbers), $\mathbb{H}$ (the skew-field of quaternions, $\mathbb{O}$) (the non-associative algebra of octonions).[34] The dimensions of these division algebras are 1, 2, 4, 8, respectively.

Theorem 4.22 *The dimension of an arbitrary division algebra is 1, 2, 4, 8.*

This theorem was independently proven by Kervaire and Milnor in 1958 by using Bott's periodicity theorem from 1958.[35] A proof based on K-theory can be found in F. Hirzebruch (1995) quoted on page 235. Interestingly enough, so far only topological proofs are known for this purely algebraic theorem.

4.4.10 Application to Fredholm Operators

> In the winter semester 1900/01 Holmgren, who had come from Uppsala (Sweden) to study under Hilbert in Göttingen (Germany), held a lecture in Hilbert's seminar on Fredholm's work on linear integral equations which had been published the previous year. This was a decisive day in Hilbert's life. He took up Fredholm's discovery with great zeal, and combined it with his variational method.
>
> Otto Blumenthal, 1932[36]

To explain the basic idea of Fredholm operators, consider the linear system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1, \\ a_{21}x_1 + a_{22}x_2 &= b_2. \end{aligned} \tag{4.31}$$

We are given the complex numbers a_{ij}, b_i with $i, j = 1, 2$. We are looking for the complex numbers x_1, x_2. The homogeneous dual problem is given by

[34] See H. Ebbinghaus et al., Numbers, Springer, New York, 1995.

[35] J. Milnor, Some consequences of a theorem of Bott, Ann. of Math. **68** (1958), 444–449.

[36] O. Blumenthal, Hilbert's biography. In: D. Hilbert, Collected Works, Vol. 3, pp. 388–429, Springer, Berlin, 1932 (in German). Hilbert's friend Otto Blumenthal died in the German concentration camp Theresienstadt (Terežin) in 1944.

$$a_{11}y_1 + a_{21}y_2 = 0,$$
$$a_{12}y_1 + a_{22}y_2 = 0. \tag{4.32}$$

The classical Fredholm alternative reads as follows:

Equation (4.31) *has a solution iff* $b_1y_1 + b_2y_2 = 0$ *for all solutions* y_1, y_2 *of the dual equation* (4.32).
Uniqueness implies existence (i.e., if we know that equation (4.31) *has at most one solution, then there exists a unique solution).*

The theory of linear Fredholm operators generalizes the Fredholm alternative to infinite dimensions.

Linear Fredholm Operators. Let $\mathcal{H}$ be a complex separable Hilbert space, and let $L(\mathcal{H},\mathcal{H})$ be the space of all linear and continuous operators $A : \mathcal{H} \to \mathcal{H}$. Here, $L(\mathcal{H},\mathcal{H})$ is a complex Banach space equipped with the operator norm

$$||A|| := \sup_{||x||\leq 1} ||Ax||.$$

The operator $A \in L(\mathcal{H},\mathcal{H})$ is called a Fredholm operator iff the two conditions $\dim\ker(A) < \infty$ and $\operatorname{codim}\operatorname{im}(A) < \infty$ are fulfilled. Fredholm operators are distinguished by the fact that the solution set of the equation (4.34) below has finite dimension, and we have only a finite number of linearly independent solvability conditions. The finite number

$$\boxed{\operatorname{ind}(A) := \dim\ker(A) - \operatorname{codim}\operatorname{im}(A)} \tag{4.33}$$

is called the index of the operator A. If $\dim\mathcal{H} < \infty$, then $\operatorname{ind}(A) = 0$. For example, set $\mathcal{H} := \mathbb{C}^2$. Then the operator $A : \mathcal{H} \to \mathcal{H}$ given by the equation (4.31) has the index zero. The set of all Fredholm operators in $L(\mathcal{H},\mathcal{H})$ is denoted by $F(\mathcal{H})$. Consider the equation

$$\boxed{Ax = b, \qquad x \in \mathcal{H}.} \tag{4.34}$$

We assume that $A \in F(\mathcal{H})$. Let $b \in \mathcal{H}$ be given. Suppose that x is a solution of (4.34). Then

$$(A^d f)(x) = f(Ax) = f(b).$$

This implies that $f(b) = 0$ for all $f \in \mathcal{H}^d$ with $A^d f = 0$. A linear Fredholm operator A has the important property that this simple necessary solvability condition for equation (4.34) is also sufficient for the existence of a solution. The following hold:

The equation (4.34) *has a solution iff* $f(b) = 0$ *for all linear continuous functionals* $f \in \mathcal{H}^d$ *with* $A^d f = 0$. *The solution set of* (4.34) *is a linear manifold of dimension* $\dim\ker(A)$. *The number*

$$\dim(A^d) = \dim\ker(A) - \operatorname{ind}(A)$$

is equal to the number of linearly independent solvability conditions for (4.34). *In particular, if* $\operatorname{ind}(A) = 0$, *then equation* (4.34) *has the crucial property that uniqueness implies existence.*

Stability of the index of a Fredholm operator. The importance of the index of a Fredholm operator relies on the fact that it is invariant under sufficiently small perturbations with respect to the operator norm.

The set $F(\mathcal{H})$ of Fredholm operators is open in $L(\mathcal{H},\mathcal{H})$, and the index is locally constant. Hence the index is constant on each component (i.e., maximal connected subset) of $F(\mathcal{H})$.

The Atiyah–Jänich theorem about the homotopic classification of families of linear Fredholm operators. Let M be a compact connected separated topological space (e.g., a sphere). We consider continuous maps of the form

$$\mu : M \to F(\mathcal{H}).$$

More precisely, let us consider the space $[M, F(\mathcal{H})]$ of mapping classes (with respect to homotopies) introduced on page 219.

Theorem 4.23 *If $\dim \mathcal{H} = \infty$, then we have $[M, F(\mathcal{H})] \simeq K_{\mathbb{C}}(M)$, in the sense of a bijection.*

This important theorem due to Atiyah and Jänich relates the K-ring $K_{\mathbb{C}}(M)$ of a compact connected separated topological space M to the homotopic structure of families of linear Fredholm operators parameterized by M. If we replace the set of Fredholm operators $F(\mathcal{H})$ by the set $GL(\mathcal{H})$ of invertible linear operators on $L(\mathcal{H},\mathcal{H})$, then the following theorem due to Kuipers shows that the situation is much simpler.

If $\dim \mathcal{H} = \infty$, then the space $[M, GL(\mathcal{H})]$ consists precisely of one point.

The proofs can be found in B. Booss and D. Bleecker, Topology and Analysis, Springer, New York, 1985.

4.4.11 Hints for Further Reading

K-theory. The theory of linear and nonlinear Fredholm operators and its applications to differential and integral equations can be found in Zeidler (1995a), (1995b) (see the references on page 1049). As an introduction to K-theory and its applications to the structure of the vector fields on spheres, we recommend:

F. Hirzebruch, Division algebras and topology. In: H. Ebbinghaus et al. (Eds.), Numbers, pp. 281–301, Springer, New York, 1995.

Furthermore, we recommend:

B. Booss and D. Bleecker, Topology and Analysis, Springer, New York, 1985 (Fredholm operators, the Atiyah–Singer index theorem, and topological gauge theory).

M. Atiyah, K-Theory, Benjamin, New York, 1967.

H. Bass, Algebraic K-Theory, Benjamin, 1968.

N. Karoubi, K-Theory: An Introduction, Springer, Berlin, 1978 (topological K-theory).

P. Gilkey, Invariance Theory, the Heat Equation, and the Atiyah–Singer Index Theorem, CRC Press, Boca Raton, Florida, 1995.

Introduction to topology:

S. Matveev, Lectures on Algebraic Topology, European Mathematical Society, 2006.

V. Guillemin and A. Pollack, Differential Topology, Prentice Hall, Englewood Cliffs, New Jersey, 1974.

G. Naber, Topological Methods in Euclidean Spaces, Cambridge University Press, 1980.

R. Bott and L. Tu, Differential Forms in Algebraic Topology, Springer, New York, 1982.

G. Bredon, Topology and Geometry, Springer, New York, 1993.

We also refer to Chapter 18 of the handbook edited by Zeidler, Teubner-Taschenbuch der Mathematik II, 9th edition, Teubner, Wiesbaden, 2008 (in German), and to the following two surveys:

C. Dodson and P. Parker, A Users' Guide to Algebraic Topology, Kluwer, Dordrecht, 1997.

S. Novikov, Topology I: General Survey (Encyclopedia of Mathematical Sciences), Springer, New York, 1996.

Topology and Physics:

A. Schwarz, Topology for Physicists, Springer, Berlin, 1994.

A. Schwarz, Quantum Field Theory and Topology, Springer, Berlin, 1993.

K. Marathe and G. Martucci, The Mathematical Foundations of Gauge Theories, North-Holland, Amsterdam, 1992.

G. Naber, Topology, Geometry, and Gauge Fields, Springer, New York, 1997.

G. Naber, Space-Time and Singularities, Cambridge University Press, 1988.

R. Hwa and V. Teplitz, Homology and Feynman Diagrams, Benjamin, 1966.

Standard references in topology:

A. Hatcher, Algebraic Topology, Cambridge University Press, 2002.
Internet: `http://www.math.cornell.edu/∼hatcher`

A. Hatcher, Spectral Sequences in Algebraic Topology, 2003ff.
Internet: `http://www.math.cornell.edu/∼hatcher`

A. Hatcher, Vector Bundles and K-Theory, 2003ff.
Internet: `http://www.math.cornell.edu/∼hatcher`

E. Spanier, Algebraic Topology, Springer, New York, 1989.

S. Hu, Homotopy Theory, Academic Press, New York, 1959.

F. Hirzebruch, Topological Methods in Algebraic Geometry, Springer, New York, 1966.

D. Husemoller, Fiber Bundles, Springer, New York, 1994.

G. Bredon, Sheaf Theory, Springer, Berlin, 1997.

A. Pressley and G. Segal (1986), Loop Groups, Clarendon Press, Oxford.

H. Baues, Homotopy Type and Homology, Oxford University Press, 1996.

W. Lück, Algebraische Topologie, Vieweg, Wiesbaden, 2005 (in German).

Perspectives

In the following volumes of this monograph, we will study important applications of topology to physics.

In quantum physics, one encounters quantum numbers.

This means that crucial quantum phenomena are described by discrete numbers. In mathematics, one knows two important methods in order to classify mathematical structures by discrete numbers. This concerns

- symmetry (e.g., the representation theory of groups and algebras), and
- qualitative behavior (topological invariants).

In Vol. IV on quantum mathematics, we will show that these two approaches are crucial for quantum physics.

4.5 The Strategy of Partial Ordering

Partially ordered sets occur quite often in mathematics.
Folklore

For natural numbers a, b, c, we have the following ordering relations:

(P1) Reflexivity: $a \leq a$;
(P2) Symmetry: If $a \leq b$ and $b \leq a$, then $a = b$.
(P3) Transitivity: $a \leq b$ and $b \leq c$ imply $a \leq c$.

In general, a set $\mathcal{N}$ is called partially ordered iff there is a $\leq$-relation defined on certain elements a, b of $\mathcal{N}$ such that (P1)–(P3) hold for all $a, b, c \in \mathcal{N}$. We write $a < c$ iff $a \leq b$ and $a \neq b$. Partially ordered sets are also called posets. In addition, a set $\mathcal{N}$ is called totally ordered iff the following hold:

(O1) $\mathcal{N}$ is partially ordered;
(O2) we have $a \leq b$ or $b \leq a$ for all $a, b \in \mathcal{N}$.

For example, both the set of natural numbers and the set of real numbers are totally ordered. For all $(a, b), (c, d) \in \mathbb{R}^2$, we define

$$(a, b) \leq (c, d) \qquad \text{iff} \qquad a \leq b \ \text{ and } \ c \leq d.$$

This way, $\mathbb{R}^2$ becomes a partially ordered set, which is not totally ordered. In fact, we have neither $(1, 2) \leq (2, 1)$ nor $(2, 1) \leq (1, 2)$. Furthermore, for a poset $\mathcal{N}$ and its subsets $\mathcal{S}$, we introduce the following terminology:

- Maximal element of $\mathcal{N}$: An element m of $\mathcal{N}$ is called maximal iff there is no element a in $\mathcal{N}$ with $m < a$. Similarly, an element m of $\mathcal{N}$ is called minimal iff there is no element a in $\mathcal{N}$ with $a < m$.
- Greatest element of $\mathcal{N}$: An element g of $\mathcal{N}$ is called greatest iff $a \leq g$ for all $a \in \mathcal{N}$.[37] Similarly, an element s of $\mathcal{N}$ is called smallest iff $s \leq a$ for all $a \in \mathcal{N}$.
- Upper bound of the subset $\mathcal{S}$: The element $b \in \mathcal{N}$ is called an upper (resp. lower) bound of the subset $\mathcal{S}$ of $\mathcal{N}$ iff $a \leq b$ (resp. $b \leq a$) for all $a \in \mathcal{S}$.
- Supremum $\sup(\mathcal{S})$ of the subset $\mathcal{S}$: The supremum of $\mathcal{S}$ is the smallest upper bound of $\mathcal{S}$. We denote this by $\sup(\mathcal{S})$.
- Infimum $\inf(\mathcal{S})$ of the subset $\mathcal{S}$: The infimum of $\mathcal{S}$ is the greatest lower bound of $\mathcal{S}$. We denote this by $\inf(\mathcal{S})$. If $\sup(\mathcal{S})$ (resp. $\inf(\mathcal{S})$) exists, then it is unique.
- Well-ordering: The set $\mathcal{N}$ is well-ordered iff it is totally ordered and each nonempty subset of $\mathcal{N}$ has a smallest element.
- Lattice: The set $\mathcal{N}$ is called a lattice iff it is partially ordered and the supremum $\sup(\{a, b\})$ and the infimum $\inf(\{a, b\})$ exist for each two-point subset $\{a, b\}$ of the set $\mathcal{N}$.

[37] By symmetry, the greatest element is uniquely determined. In contrast to this, maximal elements are not always unique.

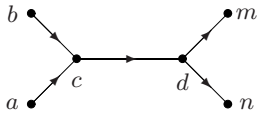


Fig. 4.13. Feynman diagram as a partially ordered set

- Complete lattice: The set $\mathcal{N}$ is called a complete lattice iff the supremum $\sup(\mathcal{S})$ and the infimum $\inf(S)$ exist for each nonempty subset $\mathcal{S}$ of $\mathcal{N}$.

For the theory of lattices, we refer to:

G. Birkhoff, Lattice Theory, Amer. Math. Soc., New York, 1968.

Examples. Each interval of the real line is an ordered set. The interval $]0,1]$ has the point $x = 1$ as greatest element, but this interval has no smallest element. As a subset of the real line, the interval $]0,1]$ has the point $x = 1$ as supremum and the point $x = 0$ as infimum. The real line is a lattice, but not a complete lattice, since the set $\mathbb{R}$ has neither a supremum nor an infimum.

4.5.1 Feynman Diagrams

Scattering processes between elementary particles can be represented by special graphs called Feynman diagrams. This is a basic tool in elementary particle physics and quantum field theory.[38] Fig. 4.13 shows the prototype of a Feynman diagram. We have vertices and oriented connecting lines. The basic idea is to write

$$a < c$$

if an arrow points from the vertex a to the vertex c (which is different from the vertex a). In addition, we assume that this can be extended to a partially ordered set by adding the reflexivity property (P1) above and by demanding the validity of the transitivity property (P3) above. For example, Fig. 4.13 describes a set of vertices a, b, c, d, m, n with the basic relations

$$a < c, \quad b < c, \quad c < d, \quad d < m, \quad d < n.$$

The other $\leq$-relations are obtained from this by using (P1) and (P3). For example, we write $a \leq a$, and

$$a < d, \quad a < m, \quad a < n.$$

Summarizing, all of the $\leq$-relations of the set $\mathcal{N} = \{a, b, c, d, m, n\}$ corresponding to Fig. 4.13 are given by the following list:

- $a \leq a, a \leq c, a \leq d, a \leq m, a \leq n$;
- $b \leq b, b \leq c, b \leq d, b \leq m, b \leq n$;
- $c \leq c, c \leq d, c \leq m, c \leq n$;
- $d \leq d, d \leq m, d \leq n$;
- $m \leq m, n \leq n$.

[38] The history of Feynman diagrams can be found in D. Kaiser, Drawing Theories Apart: The Dispersion of Feynman Diagrams in Postwar Physics, University of Chicago Press, 2005.

One checks easily that $\mathcal{N}$ is indeed a partially ordered set. Intuitively, for two vertices s and r the relation $s \leq r$ means that either $s = r$ or there exists a chain of oriented arrows which points from the vertex s to the vertex r. In Fig. 4.13, the six vertices a, b, c, d, m, n do not correspond to a totally ordered set, since we have neither $a \leq b$ nor $b \leq a$. But the three vertices a, c, d correspond to a totally ordered set.

The Feynman diagrams used in quantum electrodynamics will be discussed in Chap. 14. In contrast to the prototype depicted in Fig. 4.13, we will have different types of connecting lines corresponding to electrons, positrons, and photons in quantum electrodynamics. Moreover, we have to distinguish between internal lines and external lines. To explain this, let us distinguish between the internal vertices c, d and the external vertices a, b, m, n. Then the line cd is called an internal line and ac, bc, dm, dn are called external lines. Now Fig. 4.13 describes the scattering of the two incoming particles ac and bc resulting in the two outgoing particles dm and dn. The external lines describe particles observed in scattering processes, whereas the internal lines describe quantum fluctuations of the vacuum (ground state) which strongly influence the scattering process.

Traditionally, for the sake of simplicity, physicists do not draw the external vertices. We will do the same. Thus, using the convention of physicists, the diagram depicted in Fig. 4.13 describes the five internal lines ac, bc, cd, dm, dn which correspond to five virtual particles living in the vacuum. Much material about mathematical graph theory can be found in J. Gross and J. Yellen (Eds.), Handbook of Graph Theory, CRC Press, Boca Raton, Florida, 2004.

4.5.2 The Abstract Entropy Principle in Thermodynamics

Our goal is an assertion of the form

$$\boxed{s \leq a \qquad \text{implies} \qquad S(s) = S(a).} \tag{4.35}$$

Theorem 4.24 *Let $S : X \to [-\infty, \infty[$ be a monotone increasing function on the nonempty partially ordered set X, that is, $a \leq b$ implies $S(a) \leq S(b)$. In addition, suppose that each monotone increasing sequence in X has an upper bound. Then there exists an element s of X with the stability property (4.35) for all $a \in X$.*

This theorem due to Brézis and Browder allows the following physical interpretation:[39] We regard the elements a of X as states of a many-particle system in physics. The function S assigns to the state a the entropy $S(a)$. The ordering relation $a \leq b$ means that the state a may pass to the state b at later time. Thus, a monotone increasing sequence $a_1 \leq a_2 \leq \dots$ corresponds to a possible time development of the physical system with the corresponding increasing entropy values

$$S(a_1) \leq S(a_2) \leq \dots,$$

by the second law of thermodynamics. Theorem 4.24 tells us the existence of a stable equilibrium state s in the following sense: If the system is in the state s, then the entropy can no longer increase.

The proof of Theorem 4.24 can be found in Zeidler (1986), Vol. III, p. 163 (see the references on page 1049).

[39] H. Brézis and F. Browder, A general ordering principle in nonlinear functional analysis, Advances in Math. **21** (1976), 355–364.

4.5.3 Convergence of Generalized Sequences

> In general topological spaces, classical sequences are not enough for the characterization of closed sets by convergence.
>
> Folklore

Directed index set. In classical analysis, a sequence (x_n) is equipped with an index $n = 0, 1, 2, \ldots$ which lies in the set $\mathbb{N} = \{0, 1, 2, \ldots\}$ of natural numbers. In modern analysis (e.g., the theory of von Neumann algebras), one needs sequences (x_ν) where the index ν lies in a more general directed set $\mathcal{N}$. By definition, a set $\mathcal{N}$ is called directed (or a generalized index set) iff the following hold:

(D1) $\mathcal{N}$ is partially ordered.
(D2) If $a, b \in \mathcal{N}$, then there exists an element d in $\mathcal{N}$ such that $a \leq d$ and $b \leq d$.

For example, both the set of natural numbers and the set of real numbers are directed. Moreover, let X be an arbitrary set. Then the family $\{U(x)\}$ of all the subsets $U(x)$ of X which contain the given point x forms a directed set by the convention

$$U(x) \leq V(x) \quad \text{iff} \quad V(x) \subseteq U(x).$$

Note that $U(x), V(x) \leq U(x) \cap V(x)$. We say that the family $\{U(x)\}$ is directed by inverse inclusion.

Generalized sequence. Let X be a separated topological space, and let $\mathcal{N}$ be a directed set.[40] By definition, a generalized sequence (x_ν) in the topological space X is a map which assigns to each index ν in $\mathcal{N}$ an element x_ν in X. Obviously, every classical sequence (x_n) with natural numbers n as indices is a generalized sequence. For a generalized sequence (x_ν), we write

$$\lim_{\nu \to \infty} x_\nu = x \tag{4.36}$$

iff for each open neighborhood $U(x)$ of the point x, there exists an index ν_0 such that $x_\nu \in U(x)$ for all indices ν with $\nu_0 \leq \nu$. The following statements hold:

(a) Fix the point x in X. Consider the sequence $(x_{U(x)})$, where the index $U(x)$ is an arbitrary open neighborhood of the point x, and $x_{U(x)}$ is some point of the set $U(x)$. Then the point x is the limit of the sequence $(x_{U(x)})$.
(b) A subset C of the topological space X is closed iff for each generalized sequence (x_ν) in C it follows from (4.36) that $x \in C$.
(c) In particular, if X is a metric space (e.g., a subset of $\mathbb{R}^m$ or of a Hilbert space), then statement (b) remains valid if we only use classical sequences.
(d) The generalized limit is unique (in a separated space).

For the proofs, see

> J. Kelley, General Topology, van Nostrand, New York, 1955.

This monograph contains a systematic theory of generalized sequences. A summary (including the theory of metric and topological spaces, Hilbert spaces, Banach spaces, and locally convex spaces) can be found in Zeidler (1986), Vol. I, Appendix (see the references on page 1049). Generalized sequences are important for characterizing von Neumann algebras in quantum mechanics (see Sect. 7.18 on page 654). Generalized sequences are also called Moore–Smith sequences or nets.

[40] The topological space X is called separated (or a Hausdorff space) iff for any two different points x and y in X, there exist disjoint open neighborhoods $U(x)$ and $V(y)$. Most topological spaces have this property. Hausdorff (1868–1942) wrote the first textbook on set theory, and the theory of topological and metric spaces in 1914: F. Hausdorff, Foundation of Set Theory, Teubner, Leipzig (in German). In 1942 Hausdorff and his family committed suicide in order to escape the deportation into a German concentration camp.

4.5.4 Inductive and Projective Topologies

The reader who is not familiar with the concept of topological space should look at Sect. 5.5 of Vol. I. The prototype of a topological space is the real line $\mathbb{R}$. The terminology of set theory (e.g., images and pre-images of maps) is explained in the Appendix to Vol. I.

Images and pre-images. Let $f : X \to Y$ be a map. Then, for all subsets A and B of X and all subsets U and V of Y, the following hold:

- $f(A \cup B) = f(A) \cup f(B)$;
- $f(A \cap B) \subseteq f(A) \cap f(B)$;
- $f^{-1}(U \cup V) = f^{-1}(U) \cup f^{-1}(V)$;
- $f^{-1}(U \cap V) = f^{-1}(U) \cap f^{-1}(V)$.

We will use this in order to construct topologies which change arbitrary maps into continuous ones.

Topological spaces and continuous maps. Let X be a set. A family $\mathcal{T}$ of subsets of X is called a topology on X iff the following hold:

(T1) $X \in \mathcal{T}$ and $\emptyset \in \mathcal{T}$;
(T2) $\mathcal{T}$ is invariant under forming *finite* intersections (i.e., if $X_1, \ldots, X_n \in \mathcal{T}$, then $\cap_{k=1}^{n} X_k \in \mathcal{T}$); and
(T3) $\mathcal{T}$ is invariant under forming *arbitrary* unions (i.e., if $X_\alpha \in \mathcal{T}$ for all $\alpha \in \mathcal{A}$, then $\cup_{\alpha \in \mathcal{A}} X_\alpha \in \mathcal{T}$).

The set X equipped with a fixed topology $\mathcal{T}$ is called a topological space. The subsets of X belonging to $\mathcal{T}$ are called $\mathcal{T}$-open (or briefly open).

A map $f : X \to Y$ between two topological spaces X and Y is continuous iff the pre-images of open sets are open.

Basis of a topology. Let X be a topological space with the topology $\mathcal{T}$. A family $\mathcal{B}$ of open subsets of X is called a basis of the topology $\mathcal{T}$ iff any open subset of X is the union of sets from $\mathcal{B}$.

Subbasis of a topology. Again let X be a topological space with the topology $\mathcal{T}$. A family $\mathcal{S}$ of open subsets of the topological space X is called a subbasis of the topology $\mathcal{T}$ iff all the finite intersections of the sets from $\mathcal{S}$ form a basis of the topology $\mathcal{T}$. We also say that the family $\mathcal{S}$ generates the topology $\mathcal{T}$ iff it forms a subbasis of $\mathcal{T}$.

As an example, consider the real line $\mathbb{R}$ equipped with the usual topology. Then the family of all open intervals together with the empty set form a basis of the topology on $\mathbb{R}$. Moreover, the family of all open intervals of infinite length forms a subbasis of the topology on $\mathbb{R}$. This follows from the fact that each finite open interval can be represented as the intersection of two open intervals of infinite length.

Comparison of topologies. Let $\mathcal{T}$ and $\mathcal{S}$ be two topologies on the set X. Suppose that

$$\mathcal{T} \subseteq \mathcal{S}.$$

Then the topology $\mathcal{T}$ is called weaker than the topology $\mathcal{S}$, and $\mathcal{S}$ is called stronger than $\mathcal{T}$. Let $\mathcal{F}$ be a family of subsets of X with $X \in \mathcal{F}$ and $\emptyset \in \mathcal{F}$.

There exists a uniquely determined weakest topology $\mathcal{T}$ on X with the property $\mathcal{F} \subseteq \mathcal{T}$.

Explicitly, a subset Y of X is contained in $\mathcal{T}$ iff it is the union of finite intersections of sets from $\mathcal{F}$. In other words, the family $\mathcal{F}$ is a subbasis of the topology $\mathcal{T}$ (i.e., the family $\mathcal{F}$ generates the topology $\mathcal{T}$).

In addition, the family $\mathcal{F}$ is a basis of the topology $\mathcal{T}$ iff it is invariant under finite intersections.

Inductive topology. We are given the family of maps

$$i_\alpha : X_\alpha \to Y, \qquad \alpha \in \mathcal{A} \tag{4.37}$$

on the topological spaces X_α. We want to introduce a nontrivial topology on the *target space* Y such that all the maps i_α are continuous.[41]

There exists a uniquely determined strongest topology on the target space Y such that all the maps (4.37) are continuous.

This topology is called the inductive topology on the target space Y with respect to the family $\{i_\alpha\}_{\alpha\in\mathcal{A}}$ of maps. Explicitly, a subbasis $\mathcal{S}$ of the inductive topology is given in the following way. By definition, the subset O of the target space Y belongs to $\mathcal{S}$ iff the pre-image $i_\alpha^{-1}(O)$ is open in X_α for all indices α. Let us consider two typical examples.

(i) The coproduct topology: By definition, the coproduct $\coprod_{\alpha\in\mathcal{A}} X_\alpha$ of the family $\{X_\alpha\}_{\alpha\in\mathcal{A}}$ of topological spaces X_α consists of all the pairs (α, x), where $\alpha \in \mathcal{A}$ and $x \in X_\alpha$. The canonical maps

$$i_\alpha : X_\alpha \to \coprod_{\beta\in\mathcal{A}} X_\beta, \qquad \alpha \in \mathcal{A}$$

are given by $i_\alpha(x) := (\alpha, x)$. The inductive topology with respect to $\{i_\alpha\}$ is called the coproduct topology on $\coprod_{\alpha\in\mathcal{A}} X_\alpha$. Naturally enough, the map

$$f : \coprod_{\alpha\in\mathcal{A}} X_\alpha \to Z$$

is continuous iff all the composite maps $f \circ i_\alpha : X_\alpha \to Z,\ \alpha \in \mathcal{A}$, are continuous.

(ii) Quotient topology: Let $\sim$ be an equivalence relation on the topological space X. Consider the canonical map

$$\pi : X \to X/\sim$$

given by $\pi(x) := [x]$. The quotient topology on $X/\sim$ is the inductive topology with respect to the map π. This is the strongest topology on $X/\sim$ such that the canonical map π is continuous. Explicitly, a subset S of $X/\sim$ is open iff the pre-image $\pi^{-1}(S)$ is open in the initial space X. Let Z be a topological space. The map

$$f : X/\sim \ \to \ Z$$

is continuous iff the composite map $f \circ \pi : X \to Z$ is continuous.

Now we want to study the dual situation.

Projective topology. We are given the family of maps

$$\pi_\alpha : Y \to X_\alpha, \qquad \alpha \in \mathcal{A}, \tag{4.38}$$

where X_α is a topological space for each index α. We want to introduce a nontrivial topology on the *initial space* Y such that all the maps π_α are continuous.[42]

[41] If only the set Y and the empty set are open on Y, then we obtain the weakest topology on the target space Y such that all the maps i_α are continuous. We are not interested in this trivial case.

[42] If all the subsets of Y are open, then we obtain the strongest topology on the initial space Y such that all the maps π_α are continuous. We are not interested in this trivial case.

There exists a uniquely determined weakest topology on the initial space Y such that all the maps (4.38) are continuous.

This topology is called the projective topology on the initial space Y with respect to the family $\{\pi_\alpha\}_{\alpha\in\mathcal{A}}$ of maps. Explicitly, the projective topology is generated by all the sets $\pi_\alpha^{-1}(O)$ where O is an open set in X_α, and α is an arbitrary index. Let us consider two typical examples.

(i) The product topology: By definition, the Cartesian product $\prod_{\alpha\in\mathcal{A}} X_\alpha$ consists of all the tuples $(x_\alpha)_{\alpha\in\mathcal{A}}$. The canonical map

$$\pi_\alpha : \prod_{\beta\in\mathcal{A}} X_\beta \to X_\alpha, \qquad \alpha \in \mathcal{A}$$

is given by $\pi_\alpha((x_\beta)) = x_\alpha$. The projective topology with respect to $\{\pi_\alpha\}_{\alpha\in\mathcal{A}}$ is called the product topology on $\prod_{\alpha\in\mathcal{A}} X_\alpha$. Let Z be a topological space. Naturally enough, the map

$$f : Z \to \prod_{\beta\in\mathcal{A}} X_\beta$$

is continuous iff all the composite maps $\pi_\alpha \circ f : Z \to X_\alpha, \alpha \in \mathcal{A}$, are continuous.

(ii) The subspace topology: Let Y be a subset of the topological space X. The canonical map

$$j : Y \to X$$

is given by $j(x) := x$. By definition, the subspace topology on Y is the projective topology with respect to the map j. This is the weakest topology on Y such that the canonical map j is continuous. Explicitly, a subset S of Y is open iff there exists an open set O in X such that $S = O \cup Y$. Let Z be a topological space. The map

$$f : Z \to Y$$

is continuous iff the composite map $j \circ f : Z \to X$ is continuous.

4.5.5 Inductive and Projective Limits

The notion of limit for sequences in a topological space can be generalized to the limit of mathematical structures (e.g., linear spaces, groups, topological spaces).

Folklore

The prototype of an inductive limit is the union

$$\bigcup_{n=0}^{\infty} X_n = \lim_{n\to\infty} \text{ind}\, X_n \tag{4.39}$$

of a sequence $X_0 \subseteq X_1 \subseteq X_2 \subseteq \ldots$ of linear spaces $X_0, X_1, \ldots$ over $\mathbb{K}$ with the additional property that X_k is a linear subspace of X_{k+1} for all $k = 0, 1, \ldots$ Obviously, $\cup_{n=0}^{\infty} X_n$ is again a linear space over $\mathbb{K}$. The prototype of a projective limit is the intersection

$$\bigcap_{n=0}^{\infty} X_n = \lim_{n\to\infty} \text{proj}\, X_n \tag{4.40}$$

of a sequence $X_0 \supseteq X_1 \supseteq X_2 \supseteq \ldots$ of linear spaces $X_0, X_1, \ldots$ over $\mathbb{K}$ with the additional property that X_{k+1} is a linear subspace of X_k for all $k = 0, 1, \ldots$ Obviously, $\cap_{n=0}^{\infty} X_n$ is again a linear space over $\mathbb{K}$. We want to generalize this.

- The inductive limit will be a quotient space of the coproduct.
- The projective limit will be a subspace of the Cartesian product.

The inductive (resp. projective) limit is also called the direct (resp. inverse) limit. These limits were introduced in topology in about 1930. Let $\mathcal{A}$ be a directed set (e.g., $\mathcal{A} = \mathbb{N}$).

Inductive limit. Our starting point is the commutative diagram

$$\begin{array}{ccc} X_\alpha & \xrightarrow{\pi_{\alpha\beta}} & X_\beta \\ & \searrow^{\pi_{\alpha\gamma}} & \downarrow{\scriptstyle \pi_{\beta\gamma}} \\ & & X_\gamma \end{array} \tag{4.41}$$

for all $\alpha, \beta, \gamma \in \mathcal{A}$ with $\alpha \leq \beta \leq \gamma$. In the special case of (4.39), we choose $\pi_{\alpha\beta}(x) := x$ for all $\alpha, \beta = 0, 1, 2, \ldots$ with $\alpha \leq \beta$.

(i) Linear spaces: We are given the commutative diagrams (4.41) where $X_\alpha, X_\beta, X_\gamma$ are linear spaces over $\mathbb{K}$, and $\pi_{\alpha\beta}, \pi_{\beta\gamma}, \pi_{\beta\gamma}$ are linear morphisms. Furthermore, we assume that $\pi_{\alpha\alpha} = \mathrm{id}$ for all $\alpha \in \mathcal{A}$. For the elements $(\alpha, x), (\beta, y)$ of the coproduct $\coprod_{\alpha\in\mathcal{A}} X_\alpha$, we write

$$(\alpha, x) \sim (\beta, y)$$

iff there exists an index γ such that $\alpha, \beta \leq \gamma$, and $\pi_{\alpha\gamma}(x) = \pi_{\beta\gamma}(y)$. This is an equivalence relation on $\coprod_{\alpha\in\mathcal{A}} X_\alpha$. We write

$$\liminf_{\alpha\in\mathcal{A}} X_\alpha := \left(\coprod_{\alpha\in\mathcal{A}} X_\alpha \right) / \sim,$$

and call this the inductive limit of the diagram (4.41). This inductive limit is a linear space over $\mathbb{K}$. The canonical maps

$$i_\alpha : X_\alpha \to \liminf_{\alpha\in\mathcal{A}} X_\alpha$$

are defined by $i_\alpha(x) := [(\alpha, x)]$. Let Z be a linear space over $\mathbb{K}$. The map

$$f : \liminf_{\alpha\in\mathcal{A}} X_\alpha \to Z$$

is linear iff all the maps $f \circ i_\alpha : X_\alpha \to Z, \alpha \in \mathcal{A}$, are linear.

(ii) Topological spaces: The situation (i) can be immediately generalized to the category of topological spaces by replacing linear spaces (resp. linear morphisms) by topological spaces (resp. continuous maps).

Now let us investigate the dual situation.

Projective limit. We start with the following commutative diagrams

$$\begin{array}{ccc} X_\alpha & \xleftarrow{\pi_{\alpha\beta}} & X_\beta \\ & \nwarrow^{\pi_{\alpha\gamma}} & \uparrow{\scriptstyle \pi_{\beta\gamma}} \\ & & X_\gamma \end{array} \tag{4.42}$$

for all α, β, γ with $\alpha \leq \beta \leq \gamma$. In the special case (4.40), we choose the maps $\pi_{\alpha\beta}(x) := x$ for all $\alpha, \beta = 0, 1, 2, \ldots$ with $\alpha \leq \beta$.

(a) Linear spaces. We are given the commutative diagrams (4.42) where $X_\alpha, X_\beta, X_\gamma$ are linear spaces over $\mathbb{K}$, and $\pi_{\alpha\beta}, \pi_{\beta\gamma}, \pi_{\beta\gamma}$ are linear morphisms. Furthermore, we assume that $\pi_{\alpha\alpha} = \mathrm{id}$ for all $\alpha \in \mathcal{A}$. By definition, an element (x_α) of the product set $\prod_{\alpha\in\mathcal{A}} X_\alpha$ is called a thread iff $\beta \leq \gamma$ always implies $x_\beta = \pi_{\beta\gamma}(x_\gamma)$. The set of all threads is denoted by

$$\lim_{\alpha\in\mathcal{A}} \mathrm{proj}\, X_\alpha$$

and called the projective limit of the diagram (4.42). This is a linear subspace of $\prod_{\alpha\in\mathcal{A}} X_\alpha$. The canonical projections

$$\pi_\alpha : \prod_{\beta\in\mathcal{A}} X_\beta \to X_\alpha, \quad \alpha \in \mathcal{A}$$

are defined by $\pi_\alpha((x_\beta)) := x_\alpha$. Let Z be a linear space over $\mathbb{K}$. The map

$$f : Z \to \lim_{\alpha\in\mathcal{A}} \mathrm{proj}\, X_\alpha$$

is linear iff all the maps $\pi_\alpha \circ f : Z \to X_\alpha, \alpha \in \mathcal{A}$, are linear.

(b) Topological spaces: The situation (a) can be immediately generalized to the category of topological spaces by replacing linear spaces (resp. linear morphisms) by topological spaces (resp. continuous maps).

4.5.6 Classes, Sets, and Non-Sets

> A traveller who refuses to pass over a bridge until he has personally tested the soundness of every part of it is not likely to go far; something must be risked, even in mathematics.
>
> Horace Lamb (1849–1934)

> On the 7th of December 1873, the theory of sets left behind forever its age of innocence, for on that day Georg Cantor (1845–1918) proved that the set of real numbers is uncountable... In 1901 Bertrand Russel (1872–1970) discovered the inconsistency of the comprehension axiom (Russel's antinomy)... In 1908 Ernst Zermelo (1871–1953) proposed a system of axioms for set theory which heralded a new approach...
>
> By a theorem of mathematical logic, proved in 1931 by Kurt Gödel (1906–1978), the consistency of the set-theoretic axiom systems cannot be proved without methodological means beyond those they represent.[43]
>
> Heinz-Dieter Ebbinghaus, 1995.

> No one should ever drive us from the paradise which Cantor (1845–1918) created for us.[44]
>
> David Hilbert (1862–1943)

The *theory of sets* was created by Georg Cantor. This was motivated by the question about the uniqueness of the trigonometric Fourier expansion in the case where the convergence of the Fourier series is violated in an infinite number of points.[45]

[43] H. Ebbinghaus, Set theory and mathematics. In: H. Ebbinghaus et al. (Eds.), Numbers, Springer, New York, 1995, pp. 355–379 (reprinted with permission).

[44] D. Hilbert, On the infinite, Math. Ann. **95** (1926), 161–190 (in German).

[45] W. Purkert and H. Ilgauds, Georg Cantor, Teubner, Leipzig, 1985 (in German). I. Grattan-Guiness, Towards a biography of Georg Cantor, Annals of Science **27** (1971), 345–391.

This way, Cantor was led to the study of the structure of infinite sets. In about 1900 mathematicians were shocked by discovering contradictions in set theory. For example, in 1901 Russel showed that the notion 'set of all sets' is contradictory. A detailed discussion of this antinomy (and the different axiomatic ways of avoiding this difficulty due to Zermelo–Fraenkel and von Neumann–Bernays–Gödel) can be found in the article by Ebbinghaus (1995) quoted above.

Von Neumann's idea of classes. In 1925, the young John von Neumann (1903–1957) showed how to avoid such contradictions. He introduced the notion of class. By definition, a class is a collection of objects.

- A class is called a set iff it can be the element of some class.
- A class is called a non-set iff it can never be an element of a class.

Thus, sets can be gathered into collections. Non-sets, intuitively speaking, are so large that they cannot be gathered into larger collections. For example, the class **U** of all sets is a non-set. This class is called the universe. Moreover, the class of all linear spaces (or groups, rings, fields, Hilbert spaces, topological spaces) is a non-set. Intuitively, the class of linear spaces cannot be a set, since there exist linear spaces of arbitrarily high cardinality. The rigorous approach to classes has to be based on axioms. A detailed study can be found in:

> D. Klaua, General Set Theory, Akademie-Verlag, Berlin, 1964, Vols. 1, 2 (in German).

We also recommend the article on set theory by Ebbinghaus (1995) quoted above and the following books:

> A. Fraenkel, Y. Bar-Hillel, and A. Lévy, Foundations of Set Theory, North-Holland, Amsterdam, 1973.
>
> K. Devlin, The Joy of Sets: Fundamentals of Contemporary Set Theory, Springer, New York, 1997.

The axioms are the starting point; they formulate formal terms like 'class', 'set', and 'non-set' together with formal relations and operations between them. Pierre Cartier writes:

> A very important feature of an axiomatic system in mathematics is its *non-contradiction*; after Gödel's work in 1931, we have lost the initial hopes to establish the non-contradiction of mathematics by a formal reasoning, but one can live with a corresponding belief in non-contradiction.[46]

The outstanding philosopher and mathematician Bertrand Russel (1872–1970) remarked:

> Thus mathematics may defined as the subject in which we never know what we are talking about, nor whether what we are saying is true.

Zermelo's axiom of choice. One of the cornerstones in the foundations of class theory is the sophisticated axiom of choice, which reads as follows for the universe U:

[46] B. Russel and A. Whithead, Principia Mathematica, Vols. 1–3, Cambridge University Press, 1910.
K. Gödel, On formally undecidable theorems of the *Principia Mathematica* and related systems, Monatshefte Math. und Physik **38** (1931), 173–198 (in German).
Yu. Manin, A Course in Mathematical Logic, Springer, New York, 1977.
H. Beckert, On the epistemology of the infinite. Abhandlungen der Sächsischen Akademie der Wissenschaften zu Leipzig, Mathematisch-naturwissenschaftliche Klasse, Bd. **59**, Heft 3 (2001). Hirzel, Stuttgart/Leipzig (in German).

There exists a map $C : \mathrm{U} \to \mathrm{U}$ such that the image $C(S)$ of any nonempty set S is equal to some one-point set $\{s\}$ where s is an element of S.

Intuitively, the map C selects precisely one element s from each nonempty set S. Therefore, C is called a selection map. The axiom of choice first formulated by Zermelo (1871–1953) in 1908 seems to be trivial, but it is not trivial at all. One of the surprising results of set theory is the fact that the axiom of choice is independent of the remaining axioms of set theory. That is, one can postulate either the validity or the non-validity of the axiom of choice.[47]

In the present series of monographs, we will use classes, and we will always assume that the axiom of choice is valid.

Most of mathematicians use this convention. As a consequence, we will have the lemma of Zorn at hand, as a general existence principle in mathematics. A typical application of Zorn's lemma will be considered in Problem 4.10 on page 259. From a pragmatic point of view, the lemma of Zorn is an extremely useful tool in mathematics. Therefore, we do not want to loose this gadget. The axiom of choice has the following consequences which are very far from common sense:

(i) The existence of non-measurable sets: There exist an infinite number of subsets of the real line which are not measurable in the sense of Lebesgue. In other words, the elementary pre-measure for intervals on the real line cannot be extended to a translation-invariant measure on the real line which allows us to assign a measure to all subsets of the real line.
(ii) Cantor's well-ordering principle: Any set can be well-ordered.
(iii) Obviously, the set of real numbers is not well-ordered with respect to the usual ordering. For example, the open interval $]0, 1[$ has no smallest element. However, by (ii), there exists another (highly abstract) total ordering on the real line for which the real line is well-ordered.

For the proof of (i), see E. Stein and R. Shakarchi, Princeton Lectures in Analysis III: Measure Theory, p. 24, Princeton University Press, 2003. The proof of (ii) can be based on Zorn's lemma (see N. Dunford and J. Schwartz, Linear Operators, Vol. I, Sect. I.2.9, Wiley, New York, 1958).

4.5.7 The Fixed-Point Theorem of Bourbaki–Kneser

Theorem 4.25 *The map $f : \mathcal{N} \to \mathcal{N}$ on a nonempty partially ordered set $\mathcal{N}$ has a fixed-point (i.e., the equation $f(x) = x$ has a solution) if the following two conditions are satisfied:*
(i) *$x \leq f(x)$ for all $x \in \mathcal{N}$.*
(ii) *Each nonempty totally ordered subset of $\mathcal{N}$ has a supremum.*

For the map $f : [0, 1] \to [0, 1]$, the simple intuitive meaning of this fixed-point theorem is pictured in Fig. 4.14. This innocently looking theorem is one of the cornerstones of set theory. The fairly sophisticated proof of this theorem can be found in Zeidler (1986), Vol. I, Sect. 11.8 (see the references on page 1049). Observe that the proof does not use the axiom of choice.

[47] K. Gödel, The consistency of the axiom of choice and the generalized continuum hypothesis, Proc. Nat. Acad. Sci. U.S.A. **24**(1938), 556–557.
P. Cohen, Set Theory and the Continuum Hypothesis, Benjamin, New York, 1966. See also K. Devlin, The Joy of Sets, Springer, New York, 1997.

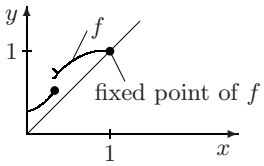


Fig. 4.14. The Bourbaki–Kneser fixed-point theorem

4.5.8 Zorn's Lemma

Theorem 4.26 *A nonempty partially ordered set contains a maximal element if each nonempty, totally ordered subset has an upper bound.*

This is one of the most important existence principles in mathematics. The proof of this theorem follows from the Bourbaki–Kneser theorem and the axiom of choice. This proof can be found in Zeidler (1986), Vol. I, p. 511. Next we want to show how non-standard analysis can be based on Zorn's lemma by constructing ultrafilters.

4.6 Leibniz's Infinitesimals and Non-Standard Analysis

> Extend the field $\mathbb{R}$ of classical real numbers to the field $*\mathbb{R}$ of generalized real numbers in order to get infinitesimal numbers and infinite numbers.
>
> Folklore

Non-standard analysis was introduced by Abraham Robinson (1918–1974) in 1960. The goal of non-standard analysis is to rigorously justify the elegant Leibniz calculus of infinitesimals. In particular, we will prove that a classical real function $f : \mathbb{R} \to \mathbb{R}$ has the classical derivative $f'(x)$ at the point x iff the difference

$$\boxed{\frac{df(x)}{dx} - f'(x)}$$

is an infinitesimal number for all nonzero infinitesimal numbers dx. Here, we use the Leibniz differential

$$df(x) := f(x + dx) - f(x).$$

The point is that the differential quotient $\frac{df(x)}{dx}$ is indeed a quotient, in the strict algebraic sense, between the generalized numbers $df(x)$ and dx. In addition, we have to show that the expression $f(x + dx)$ makes sense. To this end, we will uniquely extend the classical function $f : \mathbb{R} \to \mathbb{R}$ to a function

$$f : *\mathbb{R} \to *\mathbb{R},$$

where $*\mathbb{R}$ contains the non-standard numbers $x + dx$. The point is that the totally ordered field $*\mathbb{R}$ is an extension of the classical totally ordered field $\mathbb{R}$.

Main ideas. Let us first summarize the key properties. The rigorous justification will be considered below.

- Generalized numbers: The classical set $\mathbb{R}$ of real numbers can be extended to a larger set $*\mathbb{R}$ whose elements are called generalized numbers.

- Operations: Sums $a+b$, products ab and inequalities $a \leq b$ for real numbers can be extended to generalized numbers by preserving the usual rules. In particular, for each nonzero generalized number a, there exists a unique generalized number a^{-1} such that
$$aa^{-1} = a^{-1}a = 1.$$
Moreover, if a and b are generalized numbers, then
$$a \leq b \qquad \text{or} \qquad b \leq a.$$
For all generalized numbers a, we define the modulus by setting
$$|a| := \begin{cases} a & \text{if} \quad a > 0, \\ 0 & \text{if} \quad a \leq 0. \end{cases}$$
- Infinite numbers: In contrast to the set $\mathbb{R}$ of real numbers, the extended set $*\mathbb{R}$ contains both nonzero infinitesimal numbers and infinite numbers. By definition, a generalized number ω is called a positive (resp. negative) infinite number iff
$$\boxed{r \leq \omega \quad \text{for all real numbers } r}$$
(resp. $\omega \leq r$ for all real numbers r). Naturally enough, a generalized number a is called *finite* iff it is not an infinite number. This is equivalent to saying that there exist real numbers r and s such that $r \leq a \leq s$.
- Infinitesimal numbers: A generalized number a is called an infinitesimal number (or briefly an infinitesimal) iff
$$\boxed{|a| < \varepsilon \qquad \text{for all real numbers } \varepsilon > 0.}$$
Trivially, the real number $a = 0$ is an infinitesimal. Nonzero infinitesimals are not contained in $\mathbb{R}$, but in the extended set $*\mathbb{R}$.
- The sum $\alpha + \beta$ and the product $\alpha\beta$ of infinitesimals α and β are again infinitesimals.
- The inverse α^{-1} of a positive (resp. negative) infinitesimal is a positive (resp. negative) infinite number.
- The product $a\alpha$ of a real number a with an infinitesimal α is an infinitesimal.[48]

Summarizing, for a generalized number a, only the following three alternatives are possible:

(i) a is a real number;
(ii) a is an infinite number;
(iii) $a = r + \alpha$, where r is a real number, and α is a nonzero infinitesimal.

In case (iii), r is called the standard part of the generalized number a. The standard part is uniquely determined by a. Precisely the generalized numbers from (ii) and (iii) are called non-standard numbers.

Now we are going to rigorously justify the notion of generalized numbers. To this end, we need the following concepts:

- totally ordered field,
- ultrafilter,
- ultra-cofinite set.

[48] Further rules can be found in Problem 4.8 on page 258.

As an essential ingredient, we will use Zorn's lemma (based on the axiom of choice) in order to ensure the existence of ultrafilters. Let us discuss this. A totally ordered field X is defined to be a field which has the following additional properties for all $a, b \in X$:

- X is a totally ordered set.
- $a \leq b$ and $c \leq d$ imply $a + c \leq b + d$.
- $a \leq b$ and $0 \leq c \leq d$ imply $ac \leq bd$.

For example, the field $\mathbb{R}$ of real numbers is a totally ordered field. Our goal is to extend $\mathbb{R}$ to the totally ordered field $*\mathbb{R}$.

4.6.1 Filters and Ultrafilters

Filters. By a filter $\mathcal{F}$ of the nonempty set X, we understand a nonempty family of subsets of X such that the following hold.

- If $A, B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}$, and the intersection $A \cap B$ is not empty.
- If $A \in \mathcal{F}$ and $A \subseteq B \subseteq X$, then $B \in \mathcal{F}$.

For example, let x be a given point on the real line $X := \mathbb{R}$. Then the set of all open intervals J with $x \in J$ forms a filter of the real line.

Ultrafilters. A filter $\mathcal{U}$ of X is called an ultrafilter iff it is maximal with respect to inclusion. Explicitly, there is no filter $\mathcal{F}$ of X with $\mathcal{U} \subseteq \mathcal{F}$ and $\mathcal{F} \neq \mathcal{U}$.

Proposition 4.27 *Each filter of a nonempty set X is contained in some ultrafilter of X.*

Proof. Let $\mathcal{F}(X)$ denote the set of all filters of X. Observe that $\mathcal{F}(X)$ is a partially ordered set with respect to inclusion "$\subseteq$". Let the family $\{F_\alpha\}$ of filters F_α be a totally ordered subset of $\mathcal{F}(X)$, that is, we have $F_\alpha \subseteq F_\beta$ or $F_\alpha \subseteq F_\beta$ for all indices α, β. Define the union

$$\mathcal{B} := \bigcup_\alpha F_\alpha$$

of the filters F_α. Then, the family $\mathcal{B}$ of subsets of X is a filter of X. In addition, we have $F_\alpha \subseteq \mathcal{B}$ for all indices α. Thus, $\mathcal{B}$ is an upper bound of the subset $\{F_\alpha\}$ of the partially ordered set $\mathcal{F}(X)$. By the Zorn lemma on page 248, the set $\mathcal{F}(X)$ contains a maximal element $\mathcal{U}$. This is the desired ultrafilter. □

Proposition 4.28 *Let $\mathcal{U}$ be an ultrafilter of the partially ordered set X. Then, for each nonempty subset A of X, we have either $A \in \mathcal{U}$ or $(X \setminus A) \in \mathcal{U}$.*

The proof of this complementarity principle for ultrafilters will be given in Problem 4.5 on page 258.

The role played by high abstraction. Ultrafilters are highly abstract objects far away from our intuition. We only know the existence of such objects. But we will see in the next section that ultrafilters are extremely useful in order to give the Leibniz infinitesimals a sound mathematical basis.

This kind of high abstraction is typical for parts of modern mathematics, like the theory of motives in both modern number theory and modern algebraic geometry. For example, we refer to the survey article by

> Y. André, Une introduction aux motives: motifs pures, motifs mixtes, périodes, Panoramas et Sythèses **17** (2004) (in French).

There are hints in recent research papers that possibly the rigorous justification of quantum field theory has to be based on highly abstract objects. For example, this concerns the motivic Galois group introduced by Connes and Marcolli, in the framework of the Tannakian category. Roughly speaking, the motivic Galois group contains the information about all possible one-dimensional renormalization groups in quantum field theory. For details, we refer to the recent monograph by

> A. Connes and M. Marcolli, Noncommutative Geometry, Quantum Fields, and Motives, Amer. Math. Soc., Providence, Rhode Island, 2008.
> Internet: `http://www.math.fsu.edu/~marcolli/bookjune4.pdf`

4.6.2 The Full-Rigged Real Line

> Non-standard analysis adds infinitesimals and infinite numbers to the classical real line.
>
> Folklore

Ultra-cofinite sets of natural numbers. Let $\mathbb{N} := \{0, 1, 2, \ldots\}$ denote the set of natural numbers. A subset of $\mathbb{N}$ is called cofinite iff its complement is finite. For example, the sets $\mathbb{N}$ and $\{2, 3, 4, \ldots\}$ are cofinite. The family of cofinite sets forms a filter $\mathcal{F}$ of $\mathbb{N}$, which is called the Fréchet filter. By Prop. 4.27, there exists an ultrafilter $\mathcal{U}$ which contains the Fréchet filter $\mathcal{F}$. Explicitly, the family $\mathcal{U}$ of subsets of natural numbers has the following four properties.

(i) Intersection property: If $A, B \in \mathcal{U}$, then $A \cap B \in \mathcal{U}$, and the intersection $A \cap B$ is not empty.
(ii) Extension property: If $A \in \mathcal{U}$ and $A \subseteq B \subseteq \mathbb{N}$, then $B \in \mathcal{U}$.
(iii) Cofinite sets: Each cofinite set of $\mathbb{N}$ is contained in $\mathcal{U}$.
(iv) Complementarity: For each nonempty subset S of $\mathbb{N}$, one has either $S \in \mathcal{U}$ or $(\mathbb{N} \setminus S) \in \mathcal{U}$.

The set $\mathcal{U}$ is not uniquely determined. To show this, consider the two sets $\mathbb{N}_{\text{even}}$ and $\mathbb{N}_{\text{odd}}$ of even and odd natural numbers, respectively. These two sets are not cofinite. Suppose that the family $\mathcal{U}$ satisfies the conditions (i)–(iv). By (iv), $\mathcal{U}$ contains either $\mathbb{N}_{\text{even}}$ or $\mathbb{N}_{\text{odd}}$. Thus, $\mathcal{U}$ is different from the Fréchet filter. Furthermore, we will show in Problem 4.6 that there exist two families $\mathcal{U}_{\text{even}}$ and $\mathcal{U}_{\text{odd}}$ which satisfy the conditions (i)–(iv) and contain the set $\mathbb{N}_{\text{even}}$ and $\mathbb{N}_{\text{odd}}$, respectively.

> *In what follows, we will fix a family* $\mathcal{U}$ *which has the properties* (i)–(iv) *above.*

The elements of $\mathcal{U}$ are called *ultra-cofinite* subsets of $\mathbb{N}$. In particular, cofinite sets of natural numbers are always ultra-cofinite. By the complementarity principle for ultrafilters (Prop. 4.28), finite sets of natural numbers are never ultra-cofinite.

The construction of generalized numbers. Let us consider sequences (a_n) of real numbers, $n = 1, 2, \ldots$. We write

$$a_n = b_n \qquad \text{a.e.} \quad \text{(almost everywhere)}$$

iff there is a cofinite set C such that $a_n = b_n$ for all $n \in C$. More generally, we use the notation

$$a_n = b_n \qquad \text{u.a.e.} \quad \text{(ultra-almost everywhere)}$$

iff there is an ultra-cofinite set U such that $a_n = b_n$ for all $n \in U$. Since cofinite sets are ultra-cofinite,

$$a_n = b_n \quad \text{a.e.} \qquad \text{implies} \qquad a_n = b_n \quad \text{u.a.e.}$$

Let us now pass to the key definition of non-standard analysis. For two sequences (a_n) and (b_n) of real numbers, we write

$$(a_n) \sim (b_n) \qquad \text{iff} \qquad a_n = b_n \quad \text{u.a.e.}$$

The intersection property (i) of $\mathcal{U}$ guarantees that this is an equivalence relation. Define

$$\boxed{*\mathbb{R} := \text{set of all equivalence classes } [(a_n)].}$$

The equivalence classes $[(a_n)]$ are called generalized numbers. We equip the set $*\mathbb{R}$ of generalized numbers with a sum, a product and a $\leq$-relation by setting

$$[(a_n)] + [(b_n)] := [a_n + b_n], \quad [(a_n)][(b_n)] := [a_n b_n],$$

and

$$[(a_n)] \leq [(b_n)] \quad \text{iff} \quad a_n \leq b_n \quad \text{u.a.e.}$$

We have to show that this definition does not depend on the choice of the representatives. For example, this is true for sums. In fact,

$$a_n = c_n \quad \text{and} \quad b_n = d_n \quad \text{u.a.e.} \qquad \text{imply} \quad a_n + b_n = c_n + d_n \quad \text{u.a.e.},$$

by the intersection property (i) of $\mathcal{U}$. Hence

$$a_n \sim c_n \qquad \text{and} \quad b_n \sim d_n \qquad \text{imply} \quad (a_n + b_n) \sim (c_n + d_n).$$

The same argument applies to products. For inequalities, observe that if

$$a_n \leq b_n, \quad a_n = c_n \quad b_n = d_n \quad \text{u.a.e.},$$

then $c_n \leq d_n$ u.a.e.

Theorem 4.29 *The set $*\mathbb{R}$ is a totally ordered field.*

Proof. From $a_n + b_n = b_n + a_n$ we obtain the commutative law

$$[(a_n)] + [(b_n)] = [(b_n)] + [(a_n)].$$

Similarly, one obtains all of the other laws. Let us only consider two facts that are not completely obvious.

(I) Ordering. Let us show that

$$[(a_n)] \leq [(b_n)] \qquad \text{or} \qquad [(b_n)] \leq [(a_n)].$$

In fact, if $a_n \leq b_n$ u.a.e., then $[(a_n)] \leq [(b_n)]$. Suppose now that $a_n \leq b_n$ u.a.e. is not true. By the complementarity property of $\mathcal{U}$, $a_n > b_n$ u.a.e. Hence $[(b_n)] \leq [(a_n)]$.

(II) Inverse element. Let $[(a_n)] \neq 0$. By (I), $[(a_n)] > 0$ or $[(a_n)] < 0$. Suppose that $[(a_n)] > 0$, i.e., $a_n > 0$ u.a.e. Hence there exists a set $A \in \mathcal{U}$ such that $a_n > 0$ for all $n \in A$. Define

$$b_n := \begin{cases} a_n^{-1} & \text{if} \quad n \in A, \\ 0 & \text{if} \quad n \notin A. \end{cases}$$

Then, $a_n b_n = 1$ for all $n \in A$. In other words, $a_n b_n = 1$ u.a.e., and hence $[(a_n)][(b_n)] = [(1)]$. □

Observe that the arguments (I) and (II) above do not work if we restrict ourselves to cofinite index sets. For example, suppose that we have

$$a_n > b_n \qquad \text{for all} \quad n = 2, 4, 6, \ldots$$

and $a_n < b_n$ for all $n = 1, 3, 5, \ldots$ Then the relation $a_n \leq b_n$ a.e. is not satisfied. But this does not imply that $a_n > b_n$ a.e. This explains why we use the more complicated concept of ultra-cofinite sets. A special role is played by equivalence classes $[(r)]$ of constant sequences (r).

We can identify real numbers r with generalized numbers $[(r)]$.

This way, we regard $*\mathbb{R}$ as an extension of $\mathbb{R}$. In particular, each real number is also a generalized number. In what follows, we will frequently write r instead of $[(r)]$. To make this identification precise, define the map

$$\varphi(r) := [(r)] \qquad \text{for all} \quad r \in \mathbb{R}.$$

Obviously, from $[(r)] + [(s)] = [(r + s)]$ and $[(r)][(s)] = [(rs)]$ we get

$$\varphi(r + s) = \varphi(r) + \varphi(s) \qquad \text{and} \qquad \varphi(r)\varphi(s) = \varphi(rs).$$

This tells us that the map φ is an isomorphism from $\mathbb{R}$ onto $\varphi(\mathbb{R})$ which also preserves the $\leq$-relation. In particular,

$$s \leq [(a_n)] \leq r$$

means that $s \leq a_n \leq r$ u.a.e.

Example 4.30 *Let $a_n := \frac{1}{n}$. Obviously, $-\varepsilon < \frac{1}{n} < \varepsilon$ a.e., and hence*

$$-\varepsilon < [(\tfrac{1}{n})] < \varepsilon \qquad \textit{for all real numbers} \quad \varepsilon > 0.$$

Thus, the generalized number $[(\frac{1}{n})]$ is a nonzero infinitesimal.

Let $a_n := n$. Obviously, $r \leq n$ a.e., and hence

$$r \leq [(n)] \qquad \textit{for all real numbers} \quad r.$$

Thus, $[(n)]$ is a positive infinite number.

Similarly, we obtain the following:

- $([(a_n)]$ is an infinitesimal if $a_n \to 0$ as $n \to \infty$, and
- $[(a_n)]$ is a positive infinite number if $a_n \to +\infty$ as $n \to \infty$.

The standard part of finite generalized numbers. For generalized numbers a, b, we write

$$\boxed{a \approx b}$$

iff the difference $a - b$ is an infinitesimal. In this case, we also say that a and b are infinitesimal neighbors.

Proposition 4.31 *For each finite generalized number a, there exists a unique real number r such that $a \approx r$.*

Proof. Uniqueness. If $a \approx r$ and $a \approx s$, then the real number $r-s$ is an infinitesimal, and hence $r = s$.

Existence. Set $a := [(a_n)]$. Since the generalized number a is finite, there exist real numbers r and s such that

$$s \leq [(a_n)] \leq r.$$

Hence $s \leq a_n \leq r$ u.a.e. Construct the set $A := \{r \in \mathbb{R} : a_n \leq r \text{ u.a.e }\}$. Since this set of real numbers is bounded from below, there exists the infimum, $\inf(A)$. For each real number $\varepsilon > 0$,

$$|a_n - \inf(A)| < \varepsilon \quad \text{u.a.e.}$$

This implies the desired relation $[(a_n)] \approx \inf(A)$. □

Extension of functions. Naturally enough, each real function $f : \mathbb{R} \to \mathbb{R}$ can be uniquely extended to a function $f : *\mathbb{R} \to *\mathbb{R}$ by letting

$$f([(a_n)]) := [f(a_n)].$$

This definition does not depend on the choice of the representative, since $a_n = b_n$ u.a.e. implies $f(a_n) = f(b_n)$ u.a.e.

Theorem 4.32 (convergence). *Let r and x_0 be real numbers. For each real function $f : \mathbb{R} \to \mathbb{R}$, the following two statements are equivalent:*

(i) $\lim_{x \to x_0} f(x) = r$.
(ii) $f(x_0 + dx) \approx r$ *for all infinitesimals* $dx \neq 0$.

Proof. (i) ⇒ (ii). For each real number $\varepsilon > 0$, there exists a real number $\delta(\varepsilon) > 0$ such that

$$0 < |x - x_0| < \delta(\varepsilon) \qquad \text{implies} \qquad |f(x) - r| < \varepsilon.$$

Let $dx := [(a_n)]$. Choose $\varepsilon > 0$. From $0 < |a_n| < \delta(\varepsilon)$ u.a.e we get

$$|f(x_0 + a_n) - r| < \varepsilon \quad \text{u.a.e.}$$

This is true for all $\varepsilon > 0$. Hence the number

$$f([(x_0 + a_n)]) - r = [(f(x_0 + a_n) - r)]$$

is an infinitesimal.

(ii) ⇒ (i) (indirect argument). Suppose that $\lim_{x \to x_0} f(x) = r$ is not true. Then there exist a real number $\varepsilon > 0$ and a real sequence (a_n) such that

$$\lim_{n \to \infty} a_n = x_0, \quad a_n \neq x_0 \qquad \text{and} \qquad |f(a_n) - r| \geq \varepsilon \qquad \text{for all} \quad n \in \mathbb{N}.$$

Set $dx := [(a_n)] - x_0$. Then, dx is a nonzero infinitesimal. From the assumption (ii) we get $f(x_0 + dx) \approx r$. Hence $|f(a_n) - r| < \varepsilon$ u.a.e. This is a contradiction. □

Corollary 4.33 *Let $r \in \mathbb{R}$. For each real function $f : \mathbb{R} \to \mathbb{R}$, the following two statements are equivalent:*

(i) $\lim_{x \to +\infty} f(x) = r$.
(ii) $f(\omega) \approx r$ *for all infinite numbers* $\omega > 0$.

This follows analogously to the proof of the preceding theorem.

Leibniz differentials and the differential quotient. We are given a real function $f : \mathbb{R} \to \mathbb{R}$. Define the Leibniz differential by the quite natural relation $df(x) := f(x + dx) - f(x)$.

Theorem 4.34 *The following two statements are equivalent:*

(i) *The classical derivative* $f'(x)$ *exists.*
(ii) *For all infinitesimals* $dx \neq 0$, *we have* $\frac{df(x)}{dx} \approx f'(x)$.

Proof. Let $dx = [(h_n)]$. Introduce the function $g : \mathbb{R} \to \mathbb{R}$ by letting

$$g(h) := \frac{f(x+h) - f(x)}{h} \qquad \text{if} \quad h \neq 0,$$

and $g(0) := 0$. According to Theorem 4.32 on convergence, the following two statements are equivalent:

(i) $\lim_{h \to 0} g(h) = f'(x)$.
(ii) $g(dx) \approx f'(x)$ for all infinitesimals $dx \neq 0$.

Finally, observe that

$$\frac{df(x)}{dx} = \frac{f(x+dx) - f(x)}{dx} = \frac{[f(x+h_n) - f(x)]}{[(h_n)]} = [g(h_n)] = g(dx).$$

This finishes the proof. □

Example 4.35 (product rule). *Consider two functions* $f, g : \mathbb{R} \to \mathbb{R}$ *which have the derivatives* $f'(x)$ *and* $g'(x)$ *at the point* x. *Then, the product function* fg *has the deriviative* $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$.

In fact, for all infinitesimals $dx \neq 0$, we obtain

$$\begin{aligned} \frac{d(fg)(x)}{dx} &= \frac{f(x+dx)g(x+dx) - f(x)g(x)}{dx} \\ &= \frac{f(x+dx) - f(x)}{dx} g(x+dx) + \frac{g(x+dx) - g(x)}{dx} f(x). \end{aligned}$$

Since $g(x+dx) \approx g(x) + g'(x)dx$, we get

$$\frac{d(fg)(x)}{dx} \approx f'(x)g(x) + g'(x)f(x) + f'(x)g'(x)dx.$$

The product of an infinitesimal with a real number is again an infinitesimal. Therefore, $f'(x)g'(x)dx$ is an infinitesimal. Hence

$$\frac{d(fg)(x)}{dx} \approx f'(x)g(x) + f(x)g'(x).$$

Finally, use Theorem 4.34. □

Infinite sums. Consider the sum

$$S(n) := \sum_{j=1}^{n} c_j, \qquad n = 1, 2, \ldots$$

Extend the values $S(n)$ to a piecewise constant function $S : \mathbb{R} \to \mathbb{R}$ by letting

$$S(x) := \begin{cases} S(n) & \text{if} \quad n \leq x < n+1, \quad n = 1, 2, \ldots, \\ 0 & \text{if} \quad x < 1. \end{cases}$$

As we have shown above, the function S can be uniquely extended to a function $S : {}^*\mathbb{R} \to {}^*\mathbb{R}$. Let $\omega > 0$ be a positive infinite number. By convention, the symbol

$$\sum_{j=1}^{\omega} c_j$$

stands for $S(\omega)$. Let r be a real number. In terms of classical convergence,

$$\sum_{j=1}^{\infty} c_j = r \qquad \text{iff} \qquad \lim_{x \to +\infty} S(x) = r.$$

Thus, by Corollary 4.33, the following two statements are equivalent:

(i) $\sum_{j=1}^{\infty} c_j = r$.
(ii) $\sum_{j=1}^{\omega} c_j \approx r$ for all infinite numbers $\omega > 0$.

Integrals. Let $f : [a, b] \to \mathbb{R}$ be a continuous function on the compact interval $[a, b]$. Define the sum

$$S(n) := \sum_{j=1}^{n} f(x_j)\Delta x, \qquad n = 1, 2, \ldots,$$

where $\Delta x := (b-a)/n$ and $x_j := a + j\Delta x$. Construct $S(x)$ as above. It follows from classical calculus that

$$\lim_{x \to +\infty} S(x) = \int_a^b f(x)dx.$$

Thus, in terms of non-standard analysis,

$$S(\omega) \approx \int_a^b f(x)dx$$

for all infinite numbers $\omega > 0$. Mnemonically, we write

$$\boxed{\sum_{j=1}^{\omega} f(x + jdx)dx \approx \int_a^b f(x)dx}$$

for all infinite numbers $\omega > 0$, where $dx := (b-a)/\omega$.

Incompleteness of the extended totally ordered field ${}^*\mathbb{R}$. By definition, a completely ordered set S is a totally ordered set with the following additional property: If a subset of S has an upper bound, then it also has a supremum.

Proposition 4.36 *The set of infinitesimals has an upper bound, but no supremum in* ${}^*\mathbb{R}$.

The proof will be given in Problem 4.9. Thus, in contrast to the completely ordered set $\mathbb{R}$, the extended set ${}^*\mathbb{R}$ is not completely ordered.

Non-standard mathematics. The idea of non-standard analysis can be extended to all branches of mathematics. This leads to the so-called non-standard mathematics. A fundamental result of non-standard mathematics tells us that:

> *For each non-standard proof of a mathematical theorem on standard objects, there exists a proof within standard mathematics.*

Hints for further reading. As an introduction to non-standard mathematics and its applications, we recommend:

A. Prestel, Non-standard analysis. In: H. Ebbinghaus et al. (Eds.), Numbers, Springer, Berlin, 1995, pp. 305–327.

Furthermore, we refer to:

C. Edwards, The Historical Development of the Calculus, Springer, New York, 1979.

A. Robinson, Non-Standard Analysis, North–Holland, London, 1966.

E. Keisler, Elementary Calculus, Prindle, Boston, 1976.

E. Nelson, Internal set theory: a new approach to non-standard analysis, Bull. Amer. Math. Soc. **83** (1977), 1165–1198.

D. Landers and L. Rogge, Nichtstandardanalysis, Springer, Berlin, 1994 (in German).

K. Potthoff, Introduction to the Theory of Mathematical Models and its Applications, Wissenschaftliche Buchgesellschaft, Darmstadt, 1981 (in German).

Typical applications concern stochastic processes (e.g., Brownian motion). The basic idea is to consider Brownian motion as a random walk with respect to infintesimal time intervals: We refer to:

S. Albeverio, J. Fenstad, and R. Høegh-Krohn, Non-Standard Methods in Stochastic Analysis and Mathematical Physics, Academic Press, New York, 1986.

E. Nelson, Dynamical Theories of Brownian Motion, Princeton University Press, 1967.

E. Nelson, Quantum Fluctuations, Princeton University Press, 1985.

The theory of topological spaces in terms of filters can be found in:

N. Bourbaki, The Elements of Mathematics, Vol. II, General Topology, Springer, New York, 1990.

For partially ordered sets and lattices, we refer to:

G. Birkhoff, Lattice Theory, Amer. Math. Soc., New York, 1968.

Concerning the philosophical aspects of mathematics and physics, we recommend:

I. Kobzarev and Yu. Manin, Elementary Particles: Mathematics, Physics, and Philosophy, Kluwer, Dordrecht, 1989.

Yu. Manin, A Course in Mathematical Logic, Springer, New York, 1977.

B. Russel, Introduction to Mathematical Philosophy, Dover, New York, 1993.

A. Tarski, An Introduction to Logic and to the Methodology of the Deductive Sciences, Oxford University Press, 1994.

K. Popper, A realistic view of logic, physics, and history. In: Physics, Logic, and History, edited by W. Yourgrau and A. Breck, Plenum, New York, 1974.

S. Chandrasekhar, Truth and Beauty: Aesthetics and Motivations in Science, Chicago University Press, Chicago, Illinois, 1990.

Tian Yu Cao, Conceptual Developments of 20th Century Field Theories, Cambridge University Press, 1998.

Tian Yu Cao (Ed.), Conceptual Foundations of Quantum Field Theory, Cambridge University Press, 1999 (with contributions by leading physicists).

The Cambridge Dictionary of Philosophy, edited by R. Audi, Cambridge University Press, 2005.

Problems

4.1 *The structure of homogeneous spaces.* Prove Theorem 4.13 on page 198.
Hint: See B. Simon, Representations of Finite and Compact Groups, p. 4, Amer. Math. Soc., Providence, Rhode Island, 1996.

4.2 *Power sets.* Denote by 2^S the family of all subsets of the given set S. Show that 2^S becomes a totally ordered set with respect to the relation $A \subseteq B$, and S is the only maximal element.

4.3 *The Fréchet filter.* Show that the family of all cofinite subsets of the set of natural numbers is a filter.

4.4 *Extension of filters.* Let A be a subset of X that does not belong to the filter $\mathcal{F}$ of X. Show that there exists a filter $\mathcal{G}$ of X such that $\mathcal{F} \subseteq \mathcal{G}$ and $(X \setminus A) \in \mathcal{G}$.
Solution: Observe first that if $F \in \mathcal{F}$, then

$$F \cap (X \setminus A) \neq \emptyset.$$

Otherwise, $F \subseteq A$, and hence $A \in \mathcal{F}$. This is a contradiction to our assumption $A \notin \mathcal{F}$. Consider now the family $\mathcal{G}$ of all subsets C of X which have the following property: $F \cap (X \setminus A) \subseteq C$ for some $F \in \mathcal{F}$. Then $\mathcal{G}$ is the desired filter.

4.5 *The complementarity principle for ultrafilters.* Prove Proposition 4.28 on page 250. Solution: If $A \in \mathcal{U}$, then there exists a filter $\mathcal{G}$ such that $\mathcal{U} \subseteq \mathcal{G}$ and $(X \setminus A) \in \mathcal{G}$, by Problem 4.4. Since $\mathcal{U}$ is maximal, $\mathcal{U} = \mathcal{G}$. Hence $(X \setminus A) \in \mathcal{U}$.

4.6 *Ultra-cofinite sets of natural numbers.* Show the following:
(i) There exists an ultrafilter $\mathcal{U}_{\text{even}}$ of the set $\mathbb{N}$ of natural numbers that contains the cofinite sets and the set $\mathbb{N}_{\text{even}}$ of even natural numbers.
(ii) There exists an ultrafilter $\mathcal{U}_{\text{odd}}$ of the set $\mathbb{N}$ of natural numbers that contains the cofinite sets and the set $\mathbb{N}_{\text{odd}}$ of odd natural numbers.
(iii) $\mathbb{N}_{\text{even}} \notin \mathcal{U}_{\text{odd}}$ and $\mathbb{N}_{\text{odd}} \notin \mathcal{U}_{\text{even}}$.
Solution: Ad (i), (iii). The set $\mathbb{N}_{\text{even}}$ is not cofinite. By Problem 4.4, there exists a filter $\mathcal{G}$ which contains the cofinite sets and the set $\mathbb{N}_{\text{even}}$. The filter $\mathcal{G}$ is contained in some ultrafilter which will be denoted by $\mathcal{U}_{\text{even}}$. Obviously, $\mathbb{N}_{\text{even}} \in \mathcal{U}_{\text{even}}$. By the complementarity principle for ultrafilters on page 250, the complement $\mathbb{N}_{\text{odd}}$ to the set $\mathbb{N}_{\text{even}}$ is not contained in the ultrafilter $\mathcal{U}_{\text{even}}$.
Ad (ii), (iii). Argue similarly.

4.7 *Non-standard analysis.* Show that $*\mathbb{R}$ is a totally ordered field by completing the proof of Theorem 4.29 on page 252.

4.8 *Computational rules.* Show that the following hold true:

- The sum $\alpha+\beta$ and the product $\alpha\beta$ of infinitesimals is again an infinitesimal.
- The product $a\alpha$ of a real number a with an infinitesimal α is an infinitesimal.
- The sum $\omega + \mu$ of infinite numbers of the same sign is an infinite number.
- The product $\omega\mu$ of infinite numbers is an infinite number.
- The product $a\omega$ of a nonzero real number a with an infinite number ω is an infinite number.

Hint: If $[(a_n)]$ and $[(b_n)]$ are infinitesimals, then $|a_n|, |b_n| < \varepsilon$ u.a.e, and hence $|a_n b_n| < \varepsilon^2$ for all $\varepsilon > 0$. Thus, the product $[(a_n)][(b_n)]$ is an infinitesimal. Furthermore, note that

$$[(n^{-1})][(rn)] = r, \quad [(n^{-2})][(n)] = [(n^{-1})], \quad [(n^{-1})][(n^2)] = [(n)].$$

This shows that products between infinitesimals and infinite numbers can be real numbers, infinitesimals, or infinite numbers.

4.9 *Incomplete ordering of* $*\mathbb{R}$. Prove Prop. 4.36 on page 256.
Solution: The set S of infinitesimals does not have a supremum. Otherwise, we would have

$$\alpha \leq \sup(S) \leq r \qquad \text{for all real numbers } r \text{ and all infinitesimals } \alpha.$$

Thus, $\sup(S)$ is a nonzero infinitesimal. Since $2 \cdot \sup(S)$ is also an infinitesimal, we obtain a contradiction.

4.10 *A typical application of Zorn's lemma.* Let X be a linear subspace of the linear space Z over $\mathbb{K} = \mathbb{R}, \mathbb{C}$. Show that there exists a linear subspace Y of Z such that $Z = X \oplus Y$.
Solution: If $X = Z$, then choose $Y := \{0\}$. If $X \neq Z$ choose an element $z \in Z \setminus X$. Then, $z \neq 0$, and we have

$$X \oplus \operatorname{span}(z) \subseteq Z.$$

In the case where the dimension of Z is finite, the proof follows easily by using an induction argument with respect to the dimension of Z. In the case where the space Z has an infinite dimension of arbitrary cardinality, one has to use transfinite induction. However, the easiest proof is obtained by applying Zorn's lemma. To this end, consider the family $\mathcal{F}$ of all linear subspaces L of Z with $L \cap X = \{0\}$. With respect to inclusion $\subseteq$, the set $\mathcal{F}$ is partially ordered. Each nonempty, totally ordered subset of $\mathcal{F}$ has an upper bound, namely, the union of the corresponding sets. By Zorn's lemma, the set $\mathcal{F}$ has a maximal element Y, which is the desired linear subspace. Otherwise, there would exist an element z of Z with $z \notin X \oplus Y$, and the linear subspace $\operatorname{span}(Y, z)$ would contradict the maximality of Y.

4.11 *Proof of Proposition 3.2 on page 120.* Solution:
(I) Auxiliary result. Let $u \in X$ with $u \neq 0$. By Problem 4.10, there exists a linear subspace Y of X such that $X = \operatorname{span}(u) \oplus Y$. Define $f(\alpha u + y) := \alpha$ for all $\alpha \in \mathbb{K}$, $y \in Y$. Then $F_u(f) = f(u) = 1$.
(II) To prove the claim, assume that $F_x = F_z$. Then $F_{x-z} = F_x - F_z = 0$. We have to show that $x = y$. In fact, assume that $x - z \neq 0$. By (I), $F_{x-z} \neq 0$. This is a contradiction.

4.12 *Construction of linear functionals.* Let X be a linear subspace of the linear space Z over $\mathbb{K} = \mathbb{R}, \mathbb{C}$. Show that any linear functional $F : X \to \mathbb{K}$ defined on the subspace X can be extended to a linear functional $F : Z \to \mathbb{K}$ defined on the total space Z.
Solution: Choose a linear subspace Y of Z such that $Z = X \oplus Y$, and set $F(x + y) := F(x)$ for all $x \in X, y \in Y$.

4.13 *Construction of a biorthogonal system.* Let Z be a linear space over $\mathbb{K} = \mathbb{R}, \mathbb{C}$. Let $b_1, \ldots, b_n$ be linearly independent elements of Z. Show that there exist elements $b^1, \ldots, b^n$ of the dual space Z^d such that

$$b^k(b_l) = \delta_l^k \qquad k, l = 1, \ldots, n.$$

Solution: Let $X := \operatorname{span}\{b_1, \ldots, b_n\}$. Define

$$F(\alpha_1 b_1 + \ldots + \alpha_n b_n) := \alpha_1 \qquad \text{for all} \quad \alpha_1, \ldots, \alpha_n \in \mathbb{K}.$$

This is a linear functional on X. By Problem 4.12, this functional can be extended to a linear functional on Z, which we call b^1. Analogously, we get $b^2, \ldots, b^n$.

4.14 *Proof of Proposition 3.3 on page 121.* Solution: By Problem 4.13, we construct linear functionals $x^j \in X^d$ and $y^k \in Y^d$ such that

$$x^j(x_r) = \delta^j_r, \qquad y^k(y_s) = \delta^k_s, \qquad j, r = 1, \ldots, n,\ k, s = 1, \ldots, m.$$

Hence

$$(x_r \otimes y_s)(x^j, y^k) = x^j(x_r) y^k(y_s) = \delta^j_r \delta^k_s.$$

Now suppose that

$$\sum_{r,s} \alpha_{rs} x_r \otimes y_s = 0.$$

Applying this to (x^j, y^k), we get $\alpha_{jk} = 0$ for all indices j, k.

4.15 *The tensor product of linear spaces.* Let X and Y be linear spaces over $\mathbb{K}$ where $\mathbb{K} = \mathbb{R}, \mathbb{C}$. We want to construct the tensor product $X \otimes Y$ by using equivalence classes. We start with the set Z of all formal finite sums of the form

$$\alpha_1(x_1, y_1) + \ldots + \alpha_n(x_n, y_n), \qquad n = 1, 2, \ldots \tag{4.43}$$

where $x_1, \ldots, x_n \in X, y_1, \ldots, y_n \in Y$, and $\alpha_1, \ldots, \alpha_n \in \mathbb{K}$. In a natural sense, the set Z becomes a linear space. Let L be the smallest linear subspace of Z which contains all the sums

$$(\alpha u + \beta x, y) - \alpha(u, y) - \beta(x, y), \qquad (x, \alpha v + \beta y) - \alpha(x, v) - \beta(x, y)$$

where $u, x \in X$, $v, y \in Y$, and $\alpha, \beta \in \mathbb{K}$. Consider the factor space Z/L. For the equivalence class $[(x, y)]$ in Z/L, we write

$$x \otimes y := [(x, y)].$$

The linear hull of all the elements $x \otimes y$ is called the tensor product $X \otimes Y$. Explicitly, the elements of $X \otimes Y$ have the form

$$\alpha_1(x_1 \otimes y_1) + \ldots + \alpha_n(x_n \otimes y_n), \qquad n = 1, 2, \ldots \tag{4.44}$$

Show that the construction of L implies the typical product properties:

- $(\alpha u + \beta x) \otimes y = \alpha(u \otimes y) + \beta(x \otimes y)$,
- $x \otimes (\alpha v + \beta y) = \alpha(x \otimes v) + \beta(x \otimes y)$

for all $u, x \in X, v, y \in Y$, and all $\alpha, \beta \in \mathbb{K}$.

4.16 *The basis property of the tensor product.* Let $b_1, \ldots, b_n$ (resp. $c_1, \ldots, c_m$) be linearly independent elements of X (resp. Y). Show that the elements $b_j \otimes c_k$ with $j = 1, \ldots, n$ and $k = 1, \ldots, m$ are linearly independent.
Solution: Choose the element (4.43) and define

$$\{\alpha_1(x_1, y_1) + \ldots + \alpha_n(x_n, y_n)\}(f, g) := \alpha_1 f(x_1) g(y_1) + \ldots + \alpha_n f(x_n) g(y_n)$$

for all $f \in X^d$ and $g \in Y^d$. This is a bilinear functional on $X^d \times Y^d$. For the elements of the space L, this bilinear functional vanishes identically. For example,

$$\{(x, v + y) - (x, v) - (x, y)\}(f, g) = f(x)g(v + y) - f(x)g(v) - f(x)g(y) = 0$$

because of $g(v+y)=g(v)+g(y)$. Thus, the definition

$$[\alpha_1(x_1,y_1)+\ldots+\alpha_n(x_n,y_n)](f,g):=\alpha_1 f(x_1)g(y_1)+\ldots+\alpha_n f(x_n)g(y_n)$$

for all $f\in X^d,\ g\in Y^d$ does not depend on the choice of the representatives. We write

$$(x\otimes y)(f,g):=[(x,y)](f,g)=f(x)g(y)\qquad\text{for all}\quad f\in X^d,\ g\in Y^d.$$

Similarly, $\{\alpha_1(x_1\otimes y_1)+\ldots+\alpha_n(x_n,\otimes y_n)\}(f,g)$ is equal to

$$[\alpha_1(x_1,y_1)+\ldots+\alpha_n(x_n,y_n)](f,g).$$

Now argue as in the proof of Problem 4.14.
Remark on some ismorphism. Both the tensor product constructed above and the tensor product constructed on page 121 satisfy the properties (P) and (B) formulated on page 121. Therefore, the two linear spaces are isomorphic in a natural way, by writing the elements

$$\alpha_1(x_1\otimes y_1)+\ldots+\alpha_n(x_n\otimes y_n) \tag{4.45}$$

in the form $\sum_{j,k}\alpha_{jk}b_j\otimes b_k$ (see page 121). Because of the basis property (B), the complex coefficients α_{jk} are uniquely determined by the expression (4.45).

4.17 *The tensor product of algebras.* Let X and Y be algebras over $\mathbb{K}$. Define the product

$$(u,v)(x,y):=(ux,vy),\qquad u,x\in X,\ v,y\in Y. \tag{4.46}$$

Show that the tensor product $X\otimes Y$ becomes an algebra by (4.46).
Solution: First of all, the set Z introduced in Problem 4.15 becomes an algebra with respect to the product (4.46). Moreover, L is an ideal of Z. To see this, note that

$$(w,z)\{(x,v+y)-(x,v)-(x,y)\}=(wx,z(v+y))-(wx,zv)-(wx,zy).$$

The right-hand side is equal to $(wx,zv+zy)-(wx,zv)-(wx,zy)$, which lives in L. Similar arguments show that if $\zeta\in Z$ and $l\in L$, then $\zeta l, l\zeta\in L$. Thus, L is indeed an ideal of Z. Consequently, the quotient space Z/L is a quotient algebra, and $X\otimes Y$ is a subalgebra of Z/L. This proves Prop. 3.4.
Simplification. First suppose that X and Y are linear spaces over $\mathbb{K}$. Computations concerning the tensor product $X\otimes Y$ can be simply performed as follows. Consider all the expressions (4.43) and add the following relations

$$(\alpha u+\beta x,y)-\alpha(u,y)-\beta(x,y)=0,\quad (x,\alpha v+\beta y)-\alpha(x,v)-\beta(x,y)=0.$$

Finally, replace (x,y) by $x\otimes y$. If X and Y are algebras over $\mathbb{K}$, then we add the product $(u,v)(x,y)=(ux,vy)$.

5. Geometrical Optics

In order to understand classical mechanics and quantum mechanics, one has to understand geometrical optics.
The Huygens principle is the first general principle in the history of physics, which describes the propagation of physical effects.

Folklore

The philosophers of antiquity speculated about the nature of light, being familiar with burning glasses, with the rectilinear propagation of light, and with refraction and reflection. The first systematic writings on optics of which we shave any definite knowledge are due to Greek philosophers and mathematicians (Empedocles (490–430 B.C.) and Euclid (360–290 B.C.).[1]

Max Born and Emil Wolf, 1959

In 1636, the year in which Harvard College was founded, Réne Descartes (1596–1650) was putting his last hand to his *Discourse sur la méthode de bien conduire sa raison* which contained among others his geometry and also his dioptrics. In 1637, this book came into the hands of Pierre de Fermat (1601–1665). In 1657 Fermat received from the physician of King Louis XIV and of Mazarin, Cureau de la Chambre, in his time a very reputed man who was also a physicist of note, a treatise about optics. In the letter in which he acknowledged the receipt of this book, he stated for the first time his idea that the law of refraction might be deduced from the minimum principle of shortest time, just like the Greek engineer Heron of Alexandria (100 A.D.) had done for the reflection of light... In a letter dated January 1, 1662, he announces to Cureau de la Chambre that he found to his amazement that his principle was yielding a new demonstration of Descartes' refraction law ...

Huygens (1629–1695) was working at the Paris Academy from 1666 until 1681. What Huygens did for optics is exactly the same thing which was done two hundred years later by Maxwell (1831–1879) for electricity. Huygens replaced Fermat's long distance principle of least time by a 'contact principle'. The light waves of Huygens were exactly as hypothetical as the atoms of Democritos (400 B.C.) ...
The theory of Huygens was shortly afterwards killed by Newton (1643–1727) who showed that it was inconsistent with the propagation of longitudinal waves and the possibility of the existence of transversal waves

[1] M. Born and E. Wolf, Principles of Optics, Cambridge University Press, 1959. This is the classic textbook on optics. The seventh (expanded) edition was published in 1999.

had not been devised at yet.[2] In consequence the influence of Huygens was delayed for one hundred and twenty-five years if we consider the progress in optics and was lost altogether for the progress of the Calculus of Variations ...
The very first solution for the brachistochrone problem which Johann Bernoulli (1667–1748) found in 1696 contains the demonstration of the crucial fact that the minimum is really attained for the cycloid ... Bernoulli's method, in which something of the field theory of Weierstrass (1815–1897) appears for the first time did not attract the attention even of Bernoulli's contemporaries and remained completely ignored for nearly two hundred years. These two pages which I discovered by chance more than thirty years ago, have had a very decisive influence on the work I myself did in the Calculus of Variations. I succeeded gradually in simplifying the exposition of this theory and came finally to the point where I found to my astonishment that the method to which I have directed through long and hard work was contained, at least in principle, in the admirable book *Traité de la luminiére* (Treatise on light) of Huygens from the year 1690.[3]

5.1 Ariadne's Thread in Geometrical Optics

The following three equations describe light rays $y = y(x), x \in \mathbb{R}$ and wave fronts $S(x, y) = \text{const}$ of light in the plane.

(i) The Euler–Lagrange equation of light rays:

$$\boxed{\frac{d}{dx}\left(\frac{n(x,y(x))\cdot y'(x)}{\sqrt{1+y'(x)^2}}\right) = n_y(x,y)\sqrt{1+y'(x)^2}, \qquad x \in \mathbb{R}.} \tag{5.1}$$

The smooth function $n : \mathbb{R}^2 \to \mathbb{R}$ with $\inf_{(x,y)\in\mathbb{R}^2} n(x, y) > 0$ is the refraction index. This means that a light ray passing through the point (x_0, y_0) has the velocity

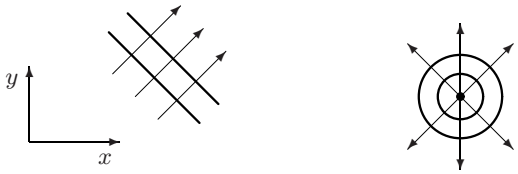
$$\frac{c}{n(x_0, y_0)}$$

at (x_0, y_0). Here, c is the velocity of the light in a vacuum. The crucial function $L : \mathbb{R}^3 \to \mathbb{R}$ given by

$$L(x, y, y') := \frac{n(x, y)}{c}\sqrt{1 + y'^2}$$

[2] According to Maxwell's theory, light corresponds to *transversal* electromagnetic waves. This means that the oscillating electric and magnetic fields are perpendicular to the direction of propagation of light.

[3] C. Carathéodory, Beginning of research in the calculus of variations, Osiris, Vol. III (1937), pp. 224–240. This lecture was given by Carathéodory (1873–1950) at the tercentenary celebration of the Harvard University, Cambridge, Massachusetts, U.S.A. The complete article is reprinted in the collected works by Carathéodory (1954), Vol. II, pp. 93–107 (see the references on page 1052).
Constantin Carathéodory, 1936



(a) Linear wave front (b) Circular wave front

Fig. 5.1. Light rays and wave fronts

is called the Lagrangian. In terms of the Lagrangian, the Euler–Lagrange equation (5.1) looks like

$$\frac{d}{dx}L_{y'}(x, y(x), y'(x)) = L_y(x, y(x), y'(x)), \qquad x \in \mathbb{R}. \tag{5.2}$$

The real number $y'(x)$ is called the slope of the light ray $y = y(x)$ at the point x. We also introduce the co-slope at x:

$$p(x) := \frac{y'(x) \cdot n(x, y)}{c\sqrt{1 + y'(x)^2}}, \qquad x \in \mathbb{R}. \tag{5.3}$$

Conversely, we have

$$y'(x) = \frac{p(x)}{\sqrt{\frac{n(x,y)^2}{c^2} - p(x)^2}}, \qquad x \in \mathbb{R}.$$

(ii) The Hamilton canonical equations of light rays:

$$\boxed{y'(x) = H_p(x, y(x), p(x)), \quad p'(x) = -H_y(x, y(x), p(x)), \quad x \in \mathbb{R}.} \tag{5.4}$$

Here, $H(x, y, p) := -\sqrt{\frac{n(x,y)^2}{c^2} - p^2}$. The function H is called the Hamiltonian of geometrical optics. An elementary computation shows that the Euler–Lagrange equation (5.1) is equivalent to (5.4). The crucial point is that a symplectic geometry is behind (5.4). This will be discussed in Sect. 6.10.

(iii) The Hamilton–Jacobi equation (or eikonal equation):

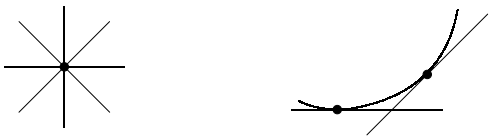
$$\boxed{S_x(x, y)^2 + S_y(x, y)^2 = \frac{n(x, y)^2}{c^2}, \qquad (x, y) \in \mathbb{R}^2.} \tag{5.5}$$

The function $S : \mathbb{R}^2 \to$ is called the eikonal. In terms of the Hamiltonian H, the eikonal equation can be written as

$$S_x + H(x, y, S_y) = 0.$$

The three equations (5.1), (5.4), and (5.5) reflect the duality between light rays and wave fronts, as we will show below (Fig. 5.1).

Prototypes of solutions. Let us consider the special case of a vacuum: $n(x, y) \equiv 1$.



(a) focal point (b) envelope (caustic)

Fig. 5.2. Singular points of light rays

(a) Linear waves: Equations (5.1) and (5.5) have the solutions

$$y(x) := \text{const}, \qquad S(x,y) := \frac{x}{c}, \qquad x, y \in \mathbb{R}.$$

This corresponds to the straight-line light rays $y(x) = \text{const}$ and to the orthogonal straight-line wave fronts $S(x,y) = \text{const}$. If the light travels from the point (x_0, y_0) to the point (x_1, y_1), then it needs the time

$$S(x_1, y_0) - S(x_0, y_0) = \frac{x_1 - x_0}{c}.$$

(b) Circular waves: Equations (5.1) and (5.5) also have the solutions

$$y(x) := vx, \qquad S(x,y) := \frac{\sqrt{x^2+y^2}}{c}, \qquad x, y \in \mathbb{R}$$

with fixed real parameter $v \neq 0$. This corresponds to the straight-line light rays $y(x) = vx$ and to the circular wave fronts $S(x,y) = \text{const}$. If the light travels from the point $(0,0)$ to the point (x_1, y_1), then it needs the time

$$S(x_1, y_1) - S(0,0) = \frac{\sqrt{x_1^2 + y_1^2}}{c}.$$

Relation to point mechanics. Let us use the following substitution:

$$x \Rightarrow t, \quad y \Rightarrow q.$$

Then the light ray $y = y(x)$ passes over to the motion $q = q(t)$ of a mass point on the real line. Moreover, the slope $y'(x)$ passes over to the velocity (time derivative) $\dot{q}(t)$. This way, it is possible to pass from geometrical optics to classical mechanics, by using the appropriate Lagrangian L. In particular, the eikonal S having the physical dimension of time passes over to the action S having the physical dimension of energy times time.

Regular and singular behavior of light rays. In the regular situation, two different points are connected by a uniquely determined light ray. If the situation is not regular (i.e., singular), light rays may intersect. We distinguish between

- isolated intersection points (focal points) and
- envelopes of light rays (caustics).

In real life, the light intensity of focal points and caustics is much larger than that of regular configurations. In the second half of the 20th century, Arnold and his school in Moscow thoroughly studied the classification of singular behavior in geometrical optics. This is closely related to the sophisticated theory of singularities in algebraic geometry (Fig. 5.2).

Gravitational lenses. Suppose that the light coming from a distant quasar Q in the universe passes through a galaxy. According to Einstein's theory of general relativity, the gravitational forces related to the galaxy cause refraction of light. This means that the galaxy acts like a lens called gravitational lens. Astronomers observe the same caustic effects as for lenses in laboratories on earth. For example, there exists one quasar which has four different images in the Hubble Space Telescope. This is the so-called Einstein cross.

Perspectives. Geometrical optics is closely related to the following mathematical subjects:

- duality between systems of first-order ordinary differential equations (e.g., light rays) and first-order partial differential equations (e.g., wave fronts) in analysis,
- symplectic geometry,
- contact geometry,
- light rays and non-Euclidean geometry (geodesics, Riemannian geometry and curvature, Kähler geometry).

It was discovered by Hamilton (1805–1865) that geometrical optics and point mechanics can be described by the same mathematics.

Light in Maxwell's theory of electrodynamics. In 1873 Maxwell (1831–1879) published his *Treatise on Electricity and Magnetism*. For a fairly large class of materials (without external charges and external currents), in an inertial system, the Maxwell equations look like

$$\mathbf{curl}\,\mathbf{E} = -\mathbf{B}_t, \qquad \mathbf{curl}\,\mathbf{H} = \mathbf{D}_t, \quad \operatorname{div}\mathbf{D} = 0, \quad \operatorname{div}\mathbf{B} = 0 \tag{5.6}$$

together with the constitutive laws $\mathbf{D} = \varepsilon\mathbf{E}$ and $\mathbf{B} = \mu\mathbf{H}$.

Here, we use a Cartesian (x, y, z)-coordinate system, and $\mathbf{E}, \mathbf{D}, \mathbf{B}, \mathbf{H}, \varepsilon, \mu$ depend on the point (x, y, z). Moreover, ε (resp. μ) are material functions called the dielectricity (resp. magnetic permeability) function.[4] The velocity of light in the material at the point (x, y, z) is given by

$$c_{\text{matter}}(x, y, z) = \frac{c}{n(x, y, z)} = \frac{1}{\sqrt{\varepsilon(x, y, z)\mu(x, y, z)}},$$

where c is the velocity of light in a vacuum, and $n(x, y, z)$ is the refraction index at the point (x, y, z). The refraction index is caused by the interaction between electromagnetic waves and the atoms of the material.

In Maxwell's theory, light corresponds to electromagnetic waves.

Light rays are the streamlines of the energy flow caused by the electric field $\mathbf{E}$ and the magnetic field $\mathbf{B}$. More precisely, the velocity vector field of the energy flow is given by

$$\mathbf{v}(x, y, z, t) := \frac{\mathbf{E}(x, y, z, t) \times \mathbf{H}(x, y, z, t)}{\eta(x, y, z, t)},$$

where $\eta = \frac{1}{2}(\mathbf{ED} + \mathbf{BH})$ is the electromagnetic energy density. Thus, the light rays $\mathbf{x} = \mathbf{x}(t)$ are the solutions of the differential equation

$$\dot{\mathbf{x}}(t) = \mathbf{v}(\mathbf{x}(t), t), \qquad t \in \mathbb{R}.$$

Now to the point. Geometrical optics represents the short-wavelength limit

$$\lambda \to 0$$

[4] In a vacuum, we have $\varepsilon(x, y, z) \equiv \varepsilon_0$ and $\mu(x, y, z) \equiv \mu_0$.

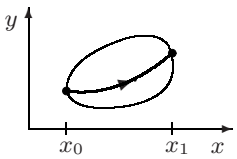


Fig. 5.3. Fermat's principle

of the Maxwell theory, where λ denotes the wavelength. To this end, one makes the ansatz

$$\mathbf{E}(x,y,z,t) = \mathbf{E}_0(x,y,z)\,\mathrm{e}^{\mathrm{i}\omega S(x,y,z)}\cdot\mathrm{e}^{-\mathrm{i}\omega t} + O(\lambda),$$
$$\mathbf{B}(x,y,z,t) = \mathbf{B}_0(x,y,z)\,\mathrm{e}^{\mathrm{i}\omega S(x,y,z)}\cdot\mathrm{e}^{-\mathrm{i}\omega t} + O(\lambda), \quad \lambda\to 0,$$

where $\omega := 2\pi c/\lambda$. More precisely, one has to take the real part of $\mathbf{E}$ and $\mathbf{B}$. The eikonal function S satisfies the first order partial differential equation

$$S_x^2 + S_y^2 + S_z^2 = \frac{n^2}{c^2}$$

with the refraction index $n = n(x,y,z)$. For the amplitude $\mathbf{E}_0$, we get the so-called transport equation $2\,\mathbf{grad}\,\mathbf{E}_0\cdot\mathbf{grad}\,S + \mathbf{E_0}\Delta S = 0$.

5.2 Fermat's Principle of Least Time

Fermat's principle reads as follows:

Light rays move in such a way that the travelling time is minimal.

The variational problem. Let the refraction index $n:\mathbb{R}^2\to\mathbb{R}$ be given as on page 264. We are given the two points (x_0,y_0) and (x_1,y_1) in $\mathbb{R}^2$ with $x_0<x_1$. We are looking for a smooth curve $y=y(x), x_0\le x\le x_1$, which is a solution of the following minimum problem:

$$\boxed{\int_{x_0}^{x_1}\frac{n(x,y(x))}{c}\sqrt{1+y'(x)^2}\,dx = \min!, \quad y(x_0)=y_0, y(x_1)=y_1.} \tag{5.7}$$

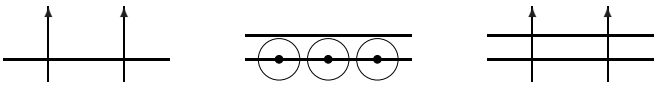
Here, the integral $\int_{x_0}^{x_1}\frac{n(x,y(x))}{c}\sqrt{1+y'(x)^2}\,dx$ represents the time needed by the light ray for passing from the point (x_0,y_0) to (x_1,y_1). In order to motivate (5.7), let us divide the light ray curve $y=y(x)$ into small pieces of length Δs. For the velocity of light at the point $(x,y(x))$, we get

$$\lim_{\Delta t\to 0}\frac{\Delta s}{\Delta t} = \frac{c}{n(x,y(x))}.$$

Approximately, the time needed by the light ray is equal to

$$\sum\Delta t = \sum\frac{n}{c}\Delta s = \sum\frac{n}{c}\sqrt{1+y'^2}\,\Delta x.$$

Letting $\Delta t\to 0$, we obtain (5.7). In what follows, we will use the Lagrangian L from (5.2). Then the Fermat principle (5.7) looks like



(a) (b) (c)

Fig. 5.4. Huygens' principle for linear wave fronts

$$\int_{x_0}^{x_1} L(x, y(x), y'(x))\, dx = \min!, \quad y(x_0) = y_0, y(x_1) = y_1.$$

Necessary condition for light rays. The following theorem is basic.

Theorem 5.1 *Any smooth solution $y = y(x)$ of Fermat's minimum problem (5.7) satisfies the Euler–Lagrange equation*

$$\frac{d}{dx} L_{y'}(x, y(x), y'(x)) = L_y(x, y(x), y'(x)), \quad x_0 \le x \le x_1, \tag{5.8}$$

which coincides with (5.1).

Proof. Fix the function $h \in \mathcal{D}(]x_0, x_1[)$ and the real parameter ε. Set

$$J(\varepsilon) := \int_{x_0}^{x_1} L(x, y(x) + \varepsilon h(x), y'(x) + \varepsilon h'(x))\, dx.$$

This means that we replace the solution curve $x \to y(x)$ by the perturbed curve

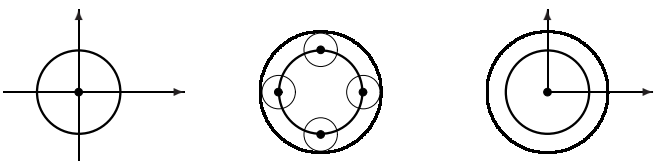
$$x \mapsto y(x) + \varepsilon h(x).$$

Since $h(x_0) = h(x_1) = 0$, the perturbed curve also passes through the two points (x_0, y_0) and (x_1, y_1) (Fig. 5.3). Consequently, the function J has a minimum at the point $\varepsilon = 0$. Hence $J'(0) = 0$. This implies

$$\int_{x_0}^{x_1} \{L_y(x, y(x), y'(x)) \cdot h(x) + L_{y'}(x, y(x), y'(x)) \cdot h'(x))\}\, dx = 0.$$

Integration by parts implies

$$\int_{x_0}^{x_1} \{L_y(x, y(x), y'(x)) - \frac{d}{dx} L_{y'}(x, y(x), y'(x))\} \cdot h(x) dx = 0$$



(a) (b) (c)

Fig. 5.5. Huygens' principle for circular wave fronts in a vacuum

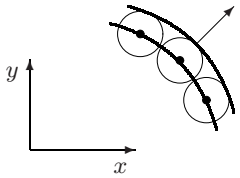


Fig. 5.6. Huygens' general principle

for all test functions $h \in \mathcal{D}(]x_0, x_1[)$. By the variational lemma (see Sect. 10.4.1 of Vol. I), we get (5.8). □

Sufficient condition for light rays. The main problem is to decide whether a given solution $y = y(x)$ of the Euler–Lagrange equation (5.8) is a solution of the minimum problem (5.7). This will be studied in Sect. 5.4.

5.3 Huygens' Principle on Wave Fronts

Huygens' principle reads as follows:

Wave fronts are the envelopes of elementary waves.

Intuitively, this means the following. Consider a wave front W_t at fixed time t. Each point of W_t originates a family of elementary wave fronts. For any later time $t+\Delta t$, the envelope of the elementary waves at time $t + \Delta t$ forms the wave front at time $t + \Delta t$.

For example, in a vacuum with refraction index $n \equiv 1$, the elementary wave fronts are circles (Figs. 5.4 & 5.5). In the general case of an arbitrary refraction index n, elementary waves are deformed circles (Fig. 5.6).

Heuristic motivation of Huygens' principle. Let us investigate the regular situation, that is, two different points P and Q in the plane can always be connected by precisely one light ray. By definition, the optical distance

$$d(P, Q)$$

between the two points P and Q is the time needed by a light ray for passing from P to Q. We will essentially use geometric arguments based on the optical distance. Consider the situation pictured in Fig. 5.7.

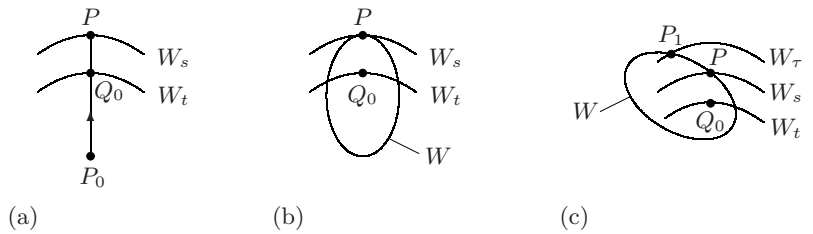


Fig. 5.7. Motivation of Huygens' principle

- Fix the point P_0. Consider the wave front $W_t := \{P \in \mathbb{R}^2 : d(P_0, P) = t\}$ in the plane.
- For $0 < t < s$, the corresponding wave fronts W_t, W_s cannot intersect each other. In fact, suppose that $Q \in W_t \cap W_s$. Then $d(P_0, Q) = t$ and $d(P_0, Q) = s$ with $t \neq s$, a contradiction.
- Choose a point $P \in W_s$. The light ray connecting the point P_0 with P intersects the wave front W_t in the point Q_0 (Fig. 5.7(a)). For the optical distance, we get $d(P_0, P) = s$ and $d(P_0, Q_0) = t$. Hence $d(Q_0, P) = s - t$.
- Consider the elementary wave front $W := \{Q \in \mathbb{R}^2 : d(Q_0, Q) = s - t\}$ originated at the point Q_0.
- We claim that P is a contact point of the wave front W_s and the elementary wave front W, as pictured in Fig. 5.7(b). To prove this, suppose that the claim is not true. Then P is a proper intersection point between W_s and W. By Fig. 5.7(c), there exists a point $P_1 \in W_\tau \cap W$ with $\tau > s$. Naturally enough, we assume that the triangle inequality is valid for the optical distance. Considering the triangle $P_0Q_0P_1$, we get
$$d(P_0, P_1) \leq d(P_0, Q_0) + d(Q_0, P_1).$$
Hence $\tau \leq t + (s - t) = s$, a contradiction.

Short-range forces in nature. We distinguish between

- short-range forces and
- long-range forces.

The Huygens principle corresponds to a short-range force. In contrast to this, consider the Newtonian equation

$$m\ddot{x}(t) = F(t), \qquad t \in \mathbb{R}, \quad x(0) = x_0, \quad \dot{x}(0) = v_0$$

for the motion of a point with mass $m > 0$ on the real line. We are given the smooth force $F : \mathbb{R} \to \mathbb{R}$ and both the initial position x_0 and the initial velocity v_0 at time $t = 0$. This problem has the unique solution

$$x(t) = x_0 + v_0 t + \frac{1}{m} \int_0^t (t - \tau) F(\tau) d\tau, \qquad t \in \mathbb{R}.$$

If we change the initial position x_0 at time $t = 0$, then the position $x(t)$ of the particle changes immediately at each time $t > 0$. Thus the force F has infinite range. (See also the discussion on Faraday's locality principle in physics on page 13).

5.4 Carathéodory's Royal Road to Geometrical Optics

> Once a day the Ptolemean king asked Euclid (ca. 360–290 B.C.) for showing him an easy approach to mathematics. Euclid answered that there is no 'royal road' to mathematics. The same anecdote is also told about Alexander the Great (356–323 B.C.) and Aristotle (384–322 B.C.).
>
> Folklore

> Constantin Carathéodory (1873–1950), a student of Hilbert (1862–1943) and the leading expert in the calculus of variations, instinctively felt that there must be a simple 'royal road' to the calculus of variations. He was obsessed by the question "Why extremals play such a distinguished role?" They are being used and they function fabulously, but how can one reason

out and show their necessity and simplicity? Carathéodory's wonderful paper *Die Methode der geodätischen Äquidistanten*, Acta Mathematica **47** (1925), 199–230, was the result of his investigations.[5]

Krystzof Maurin, 1997

Carathéodory's fundamental equation reads as

$$\boxed{\begin{aligned} S_x(x,y) &= L(x,y,v(x,y)) - v(x,y)L_{y'}(x,y,v(x,y)),\\ S_y(x,y) &= L_{y'}(x,y,v(x,y)), \qquad (x,y)\in\mathbb{R}. \end{aligned}} \tag{5.9}$$

We are given the smooth function $L:\mathbb{R}^3\to\mathbb{R}$,

$$L(x,y,v) := \frac{n(x,y)v(x,y)}{c\sqrt{1+v^2}}, \qquad (x,y,v)\in\mathbb{R}^3,$$

together with the smooth function $n:\mathbb{R}^2\to\mathbb{R}$, where $\inf_{(x,y)\in\mathbb{R}^2} n(x,y) > 0$. We are looking for smooth functions

- $v = v(x,y)$ (slope function) and
- $S = S(x,y)$ (eikonal)

from $\mathbb{R}^2$ to $\mathbb{R}$ which satisfy (5.9). The smooth solutions $y = y(x), x\in\mathbb{R}$ of the so-called slope equation

$$y'(x) = v(x,y(x)), \qquad x\in\mathbb{R} \tag{5.10}$$

are called the trajectories of (5.9). The set of all the solutions of (5.10) is called a field of trajectories. Note that two different trajectories of the field do not intersect each other, since the solution of the initial-value problem for the slope equation (5.10) is unique.[6] Again let us consider Fermat's minimum problem

$$\int_{x_0}^{x_1} \frac{n(x,y(x))}{c}\,\sqrt{1+y'(x)^2}\,dx = \min!, \quad y(x_0)=y_0, y(x_1)=y_1. \tag{5.11}$$

The following main theorem in geometrical optics tells us that the trajectories of the field are solutions of (5.11), that is, they are light rays. Furthermore, the equations $S(x,y) = \text{const}$ describe wave fronts.

Theorem 5.2 *Let $v, S:\mathbb{R}^2\to\mathbb{R}$ be a smooth solution of Carathéodory's fundamental equation (5.9). Fix the points (x_0,y_0) and (x_1,y_1) in $\mathbb{R}^2$. Then the following hold.*

(i) *Let $y = y(x), x_0\le x\le x_1$ be a trajectory which corresponds to the slope equation (5.10) and satisfies the boundary condition $y(x_0)=y_0, y(x_1)=y_1$. Such a trajectory is a solution of the minimum problem (5.11).*

(ii) *The difference $S(x_1,y_1)-S(x_0,y_0)$ is the optical distance between the points (x_0,y_0) and (x_1,y_1). In other words,*

$$S(x_1,y_1) - S(x_0,y_0) = \int_{x_0}^{x_1} \frac{n(x,y(x))}{c}\,\sqrt{1+y'(x)^2}\,dx.$$

[5] K. Maurin, Riemann's Legacy: Riemann's Ideas in Mathematics and Physics of the 20th Century, Kluwer, Dordrecht, 1997 (reprinted with permission).

[6] We assume that the solutions of the slope equation exist for all $x\in\mathbb{R}$. Otherwise the following results are only valid locally. Fig. 5.8 shows a field of trajectories on the open set U.

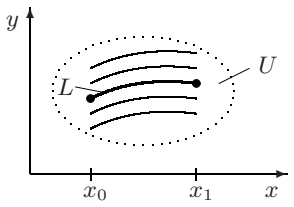


Fig. 5.8. Field of light rays

Proof. (I) Differentiating the Lagrangian $L(x,y,y') = \frac{n(x,y)}{c}\sqrt{1+y'^2}$, we get

$$L_{y'y'}(x,y,y') = \frac{n(x,y)}{c(\sqrt{1+y'^2})^3} \geq 0 \qquad \text{for all} \quad x,y,y' \in \mathbb{R}.$$

(II) Let us introduce the Weierstrass excess function

$$\mathcal{E}(x,y,v,w) := L(x,y,w) - L(x,y,v) - (w-v)L_{y'}(x,y,v).$$

By Taylor expansion, $\mathcal{E}(x,y,v,w) = L_{y'y'}(x,y,v_*) \geq 0$ for all $x,y,v,w \in \mathbb{R}$.

(III) The integral trick. Let $y = \eta(x)$ be a smooth function with $\eta(x_0) = y_0$ and $\eta(x_1) = y_1$. Then the integral

$$\int_{x_0}^{x_1} \Big(S_x(x,\eta(x)) + S_y(x,\eta(x))\eta'(x) \Big) dx = \int_{x_0}^{x_1} \frac{dS(x,\eta(x))}{dx}\, dx$$

is equal to the difference $S(x_1,y_1) - S(x_0,y_0)$.

(IV) Using Carathéodory's equation (5.9) and setting $y := \eta(x)$, we get the key relation

$$L(x,\eta(x),\eta'(x)) = S_x(x,\eta(x)) + S_y(x,\eta(x))\eta'(x) + \mathcal{E}(x,\eta(x),v(x,\eta(x)),\eta'(x)).$$

By (II) and (III),

$$\int_{x_0}^{x_1} L(x,\eta(x),\eta'(x))\, dx \geq S(x_1,y_1) - S(x_0,y_0).$$

(V) Choosing $\eta(x) := y(x)$, it follows from $y'(x) = v(x,y(x))$ that

$$\mathcal{E}(x,y(x),v(x,y(x)),y'(x)) \equiv 0.$$

Hence

$$\int_{x_0}^{x_1} L(x,y(x),y'(x))\, dx = S(x_1,y_1) - S(x_0,y_0).$$

□

Example (vacuum). Let $n(x,y) \equiv 1$. Then the functions

$$v(x,y) := 0, \qquad S(x,y) = x \qquad \text{for all} \quad (x,y) \in \mathbb{R}^2$$

are solutions of Carathéodory's equation (5.9). This corresponds to the straight-line light rays $y(x) \equiv \text{const}$ and the wave fronts $x = \text{const}$.

Hilbert's invariant integral. Let v, S be a smooth solution of Carathéodory's equation (5.9). Then, for any smooth function $y = \eta(x)$, the line integral

$$\int_{(x_0,y_0)}^{(x_1,y_1)} \{L(x,y,v(x,y)) - v(x,y)L_{y'}(x,y,v(x,y))\}\,dx + L_{y'}(x,y,v(x,y))\,dy \tag{5.12}$$

along the curve $x = x, y = \eta(x)$ is equal to the difference $S(x_1,y_1) - S(x_0,y_0)$. That is, the line integral does not depend on the path. This follows as in step (III) of the proof of the theorem above, noting that $dy = \eta'(x)dx$. The integral (5.12) is called Hilbert's invariant integral. In fact, Carathéodory's 'royal road' to geometrical optics and to the calculus of variations was strongly motivated by Hilbert's invariant integral, which is closely related to the Poincaré–Cartan integral invariant (see page 423).

5.5 The Duality between Light Rays and Wave Fronts

Legendre transformation. Recall that $L(x,y,v) := \frac{n(x,y)}{c}\sqrt{1+v^2}$. In order to simplify the approach considered above, we will use the Legendre transformation $(x,y,v) \mapsto (x,y,p)$ and $L(x,y,v) \mapsto H(x,y,p)$ by setting

$$p := L_v(x,y,v), \qquad H := vp - L.$$

Explicitly, we get

$$p = \frac{n(x,y)v}{c\sqrt{1+v^2}}, \quad v = \frac{p}{c\sqrt{\frac{n(x,y)^2}{c^2} - p^2}}, \quad H(x,y,p) = -\sqrt{\frac{n(x,y)^2}{c^2} - p^2}.$$

The Hamilton–Jacobi equation for the wave fronts. Introducing the co-slope function

$$p(x,y) := \frac{n(x,y)v(x,y)}{c\sqrt{1+v(x,y)^2}}, \tag{5.13}$$

Carathéodory's fundamental equation (5.9) passes over to

$$S_y(x,y) = p(x,y), \qquad S_x(x,y) = -H(x,y,p(x,y)).$$

This implies the Hamilton–Jacobi equation

$$S_x(x,y) + H(x,y,S_y(x,y)) = 0. \tag{5.14}$$

The Hamilton canonical equation for the light rays. As we have shown on page 265, the Legendre transformation sends the Euler–Lagrange equation for light rays to the Hamilton canonical equation (5.4) on page 265.

Contact transformation. The Legendre transformation satisfies the differential relation

$$dL - pdv = vdp - dH,$$

which is typical for contact transformations (see Sect. 5.7). In fact, by the product rule for differentials, $dH = d(pv) - dL = (dp)v + pdv - dL$.

The classical maximum principle. Define

$$\mathcal{H}(x,y,v,p) := pv - L(x,y,v) \qquad \text{for all} \quad (x,y,v) \in \mathbb{R}^3,$$

where $L(x,y,v) := n(x,y)(1+v^2)^\gamma$ for all $(x,y,v) \in \mathbb{R}^3$. Suppose first that $\gamma > \frac{1}{2}$. Then, for any fixed $(x,y,p) \in \mathbb{R}^3$, we define

$$H(x,y,p) := \max_{v\in\mathbb{R}} \mathcal{H}(x,y,v,p).$$

Note that the function $v \mapsto \mathcal{H}(x,y,v,p)$ is strictly concave and goes to $-\infty$ as $v \to \pm\infty$. Thus, the global maximum problem

$$\mathcal{H}(x,y,v,p) = \max!, \qquad v \in \mathbb{R}$$

has a unique solution given by $\mathcal{H}_v(x,y,v,p) = 0$. Hence

$$p = L_v(x,y,v) = 2\gamma n(x,y)v(1+v^2)^{\gamma-1}.$$

The map $v \mapsto p$ is invertible. Hence

$$H(x,y,p) = \mathcal{H}(x,y,v(p),p). \tag{5.15}$$

This corresponds to the Legendre transformation. In geometrical optics, we encounter the case where $\gamma = \frac{1}{2}$. Then we have to consider the local problem

$$\mathcal{H}(x,y,v,p) = \text{critical!}, \qquad v \in \mathbb{R}.$$

This yields $\mathcal{H}_v(x,y,v,p) = 0$. Again this implies (5.15).

In the sense of duality in general optimization theory (in the framework of nonlinear functional analysis), the Hamiltonian H is the conjugate function to the Lagrangian L (see Zeidler (1986), Vol. III, quoted on page 1049).

5.5.1 From Wave Fronts to Light Rays

Suppose that we have a smooth solution $S = S(x,y), (x,y) \in \mathbb{R}$, of the Hamilton–Jacobi equation (5.14).[7] Our goal is to construct a solution of Carathéodory's fundamental equation (5.9) on page 272. To this end, we first define

$$p(x,y) := S_y(x,y), \qquad (x,y) \in \mathbb{R}^2.$$

Then we use the inverse Legendre transformation in order to get

$$v(x,y) := \frac{n(x,y)p(x,y)}{c\sqrt{\frac{n(x,y)^2}{c^2} - p(x,y)^2}}, \qquad (x,y) \in \mathbb{R}^2.$$

This yields the solution S, v of (5.9). Light rays $y = y(x)$ are obtained by using the slope equation $y'(x) = v(x,y(x))$.

[7] If $S = S(x,y)$ is only a local solution, then the following argument is locally valid.

5.5.2 From Light Rays to Wave Fronts

Fix the point $(x_0, y_0) \in \mathbb{R}^2$. Then the initial-value problem

$$S_x(x,y) + H(x,y,S_y(x,y)) = 0, \quad S(x_0,y) = 0 \tag{5.16}$$

has always a smooth solution $S = S(x,y)$ in a sufficiently small neighborhood of the point (x_0, y_0). This is the special case of a general result for first-order partial differential equations due to Cauchy (1789–1857).[8] Let us sketch the proof. We use Cauchy's characteristic system, which is an extension of the Hamilton canonical equations:

$$\begin{aligned} y'(x) &= H_p(x, y(x), p(x)), & y(x_0) &= \eta, \\ p'(x) &= -H_y(x, y(x), p(x)), & p(x_0) &= 0, \\ \sigma'(x) &= p(x)y'(x) - H(x, y(x), p(x)), & \sigma(x_0) &= 0. \end{aligned} \tag{5.17}$$

This system can be solved by first solving the Hamilton canonical equations for getting $y = y(x), p = p(x)$ and then computing

$$\sigma(x) = \int_{x_0}^{x} \{p(x)y'(x) - H(x, y(x), p(x)\}dx.$$

Varying the parameter η, we get the solutions $y = y(x,\eta), \sigma = \sigma(x,\eta)$ of (5.17) depending on η. Now the solution of (5.16) is obtained by setting

$$S(x, y(x,\eta)) := \sigma(x,\eta).$$

The solutions $y = y(x), p = p(x), \sigma = \sigma(x)$ of (5.17) are called characteristics of (5.16). The projections of the characteristics from the (x,p,σ)-space onto the (y,p)-space are solutions of the Hamilton–Jacobi equations. Finally, the projections $y = y(x)$ are light rays starting at the wave front $x_0 = \textit{constant}$ in the (x,y)-plane. This shows that

> *In order to understand geometrical optics best, one has to pass to higher dimensions.*

The relation to Lagrangian submanifolds will be studied in Sect. 6.13 on page 419.

5.6 The Jacobi Approach to Focal Points

Jacobi's sufficient criterium for a minimum. Let $y = y(x), x \in \mathbb{R}$, be a smooth solution of the Euler–Lagrange equation (5.1). The corresponding Jacobi equation reads as

$$\boxed{\frac{d}{dx}\big(\alpha(x)h'(x)\big) + \beta(x)h'(x) + \gamma h(x) = \lambda h(x), \quad h(x_0) = h(x_1) = 0,} \tag{5.18}$$

where the coefficients look like

$$\alpha(x) := L_{y'y'}(P), \quad \beta(x) := L_{yy'}(P), \quad \gamma(x) := \beta'(x) - L_{yy}(P)$$

[8] See Zeidler (2004), p. 523, quoted on page 1049. We also refer to C. Carathéodory, Calculus of Variations and Partial Differential Equations of First Order, Chelsea, New York, 1982.

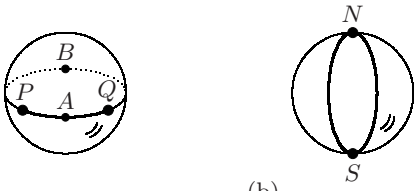


Fig. 5.9. Geodesics on earth

with $P := (x, y(x), y'(x))$. Explicitly, for $L(x, y, y') = \frac{n(x,y)}{c}\sqrt{1+y'^2}$, we get

$$\alpha(x) = \frac{n(x, y(x))}{c(\sqrt{1+y'(x)^2})^3}, \quad \beta(x) = \frac{n_y(x, y(x)) \cdot y'(x)}{c\sqrt{1+y'(x)^2}},$$

and $\gamma(x) = \beta'(x) - \frac{n_{yy}(x,y)}{c}\sqrt{1+y'(x)^2}$.

Theorem 5.3 *If the smallest eigenvalue of* (5.18) *is positive, then the given curve* $y = y(x)$ *is a solution of Fermat's principle of least time* (5.7).

The proof can be found in Zeidler (1986), Vol. III, p. 205, quoted on page 1049. This proof generalizes the proof given on page 376 for the harmonic oscillator. Since

$$\min_{x_0 \le x \le x_1} \alpha(x) > 0,$$

the Jacobi eigenvalue problem (5.18) corresponds to a regular Sturm–Liouville problem.[9] Such problems are special cases of the functional-analytic Hilbert–Schmidt theory for compact self-adjoint operators in Hilbert space (see Zeidler (1995a), Chap. 4, quoted on page 1049). The Sturm–Liouville problem (5.18) is called singular iff the finite interval $[x_0, x_1]$ is replaced by an infinite interval, or the function $\alpha = \alpha(x)$ vanishes at least at one point of the interval. Typically, such singular problems arise in quantum mechanics for investigating the spectrum of atoms or molecules (e.g., the hydrogen atom). This will be studied in Vol. III. In terms of functional analysis, singular Sturm–Liouville problems are treated by von Neumann's functional-analytic theory of unbounded self-adjoint operators in Hilbert space (based on the spectral family).

Fermat's principle of critical time and extremals. By definition, the variational problem

$$\int_{x_0}^{x_1} L(x, y(x), y'(x))\, dx = \text{critical!}, \qquad y(x_0) = y_0,\ y(x_1) = y_1 \tag{5.19}$$

means that $J'(0) = 0$ where

$$J(\varepsilon) := \int_{x_0}^{x_1} L(x, y(x) + \varepsilon h(x), y'(x) + \varepsilon h'(x))\, dx$$

for any $h \in \mathcal{D}(]x_0, x_1[)$. By convention, the solutions $y = y(x)$ of the Euler–Lagrange equation (5.1) are called extremals.

[9] Sturm (1803–1855), Jacobi (1804–1851), Liouville (1809–1882).

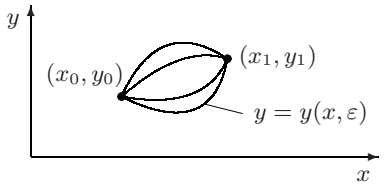


Fig. 5.10. Focal points

Proposition 5.4 *The smooth trajectory $y = y(x)$ is a solution of Fermat's principle of critical time* (5.19) *iff it is an extremal.*

Proof. Use the critical condition $J'(0) = 0$ and the same argument as in the proof of Theorem 5.1. □

Intuitively, extremals are light rays which run locally in minimal time, but not globally. To explain this, as a paradigm, replace light rays by curves of minimal length on the surface of earth. Choose two points P and Q on the equator as pictured in Fig. 5.9(a).

The segment PAQ of the equator is a curve of minimal length. However, the segment PBQ of the equator is only locally a curve of minimal length, but not globally. The North Pole N and the South Pole S are the prototypes of focal points, that is, the uniqueness of the connecting curve of minimal length is violated (Fig. 5.9(b)).

Focal points and Jacobi fields. Fix the points (x_0, y_0) and (x_1, y_1). Consider the family

$$y = y(x, \varepsilon), \quad x_0 \le x \le x_1$$

of smooth extremals which pass through the focal points (x_0, y_0) and (x_1, y_1) for all parameters $\varepsilon \in]-\varepsilon_0, \varepsilon_0[$ (Fig. 5.10).

The function $y = h(x)$ defined by

$$h(x) := y_\varepsilon(x, 0), \qquad x_0 \le x \le x_1$$

is called the Jacobi field of the family of extremals. By Taylor expansion,

$$y(x, \varepsilon) = y(x) + \varepsilon h(x) + O(\varepsilon^2), \qquad \varepsilon \to 0, \qquad x_0 \le x \le x_1,$$

where we set $y(x) := y(x, 0)$. Thus the Jacobi field describes the first-order deviation of the family of extremals from the special extremal $x \mapsto y(x)$. If $h(x) \equiv 0$, then the first-order deviation vanishes. In particular, if there is a unique extremal $y = y(x)$ which connects the point (x_0, y_0) with the point (x_1, y_1), then a nontrivial Jacobi field does not exist.

Theorem 5.5 *The Jacobi field $h = h(x)$ satisfies the Jacobi equation*

$$\frac{d}{dx}\big(\alpha(x)h'(x)\big) + \beta(x)h'(x) + \gamma h(x) = 0, \quad h(x_0) = h(x_1) = 0. \qquad . \quad (5.20)$$

Proof. Differentiate the equation for the extremals,

$$\frac{d}{dx} L_{y'}(x, y(x, \varepsilon), y_x(x, \varepsilon)) = L_y(x, y(x, \varepsilon), y_x(x, \varepsilon)),$$

with respect to the parameter ε. □

The equation (5.20) is a special case of the Jacobi eigenvalue problem (5.18) with $\lambda = 0$.

By definition, the point (x_1, y_1) is conjugate to the point (x_0, y_0) iff the Jacobi equation (5.20) has a nontrivial solution, $h(x) \not\equiv 0$. In particular, the situation considered in Theorem 5.3 corresponds to points (x_0, y_0) and (x_1, y_1) which are not conjugate. Finally, consider a sphere of radius r as pictured in Fig. 5.9. It can be shown that the points P and Q on the equator are conjugate iff their distance is equal to πr. In particular, North Pole and South Pole are conjugate points.

5.7 Lie's Contact Geometry

> Contact geometry of dimension $2n + 1$ $(n = 1, 2, \ldots)$ is based on contact elements which are pairs (P, Π) of a point P and a $2n$-dimensional plane Π through the point P. Contact geometry is closely related to envelopes of curves and surfaces (e.g., this concerns Huygens' principle in geometrical optics).
> Contact transformations provide the general setting for transforming ordinary and partial differential equations in such a way that solutions of the original equation pass over to solutions of the transformed equation.
>
> Folklore

> Contact geometry does for geometric optics and the theory of wave propagation what symplectic geometry does for mechanics. Both these geometries and their isomorphisms were conceived by a single man – Sophus Lie (1842–1899). The thesis that Lie is the father of both geometries is firmly founded...
> Lie regarded the theory of groups of contact transformations (founded in 1871) as his greatest discovery and achievement.[10]
>
> Krystzof Maurin, 1997

5.7.1 Basic Ideas

Invariance of solutions of differential equations. In order to simplify the solution of the differential equation

$$F(x, y(x), y'(x)) = 0, \qquad x \in \mathbb{R} \tag{5.21}$$

one can use the point transformation $\xi = a(x, y), \eta = b(x, y)$. This yields

$$G(\xi, \eta(\xi), \eta'(\xi)) = 0, \qquad \xi \in \mathbb{R}. \tag{5.22}$$

However, it turns out that it is frequently convenient to use the more general transformation

$$\xi = A(x, y, y'), \qquad \eta = B(x, y, y'), \qquad \eta' = C(x, y, y') \tag{5.23}$$

from $\mathbb{R}^3$ to $\mathbb{R}^3$. Quite naturally, we postulate:

> *Solutions of the differential equation* (5.21) *are transformed into solutions of the differential equation* (5.22).

[10] K. Maurin, Riemann's Legacy: Riemann's Ideas in Mathematics and Physics of the 20th Century, Kluwer, Dordrecht, 1997 (reprinted with permission).

This restricts the possible transformations of the form (5.23). It turns out that contact transformations are the right setting. To discuss this, let us use the language of differentials. We write (5.21) and (5.22) as

$$\boxed{F(x,y,y')=0, \qquad dy-y'dx=0} \tag{5.24}$$

and

$$G(\xi,\eta,\eta')=0, \qquad d\eta-\eta' d\xi=0. \tag{5.25}$$

Here, x, y, y' and ξ, η, η' are regarded as independent variables which are constrained by the equations (5.24) and (5.25), respectively. The diffeomorphism (5.23) from $\mathbb{R}^3$ onto $\mathbb{R}^3$ is called a contact transformation with respect to the contact forms $dy - y'dx$ and $d\eta - \eta' d\xi$ iff, for any point (x, y, y'), there exists a nonzero real number $\varrho(x, y, y')$ such that

$$\boxed{d\eta-\eta' d\xi=\varrho(x,y,y')(dy-y'dx)} \tag{5.26}$$

for all $(x, y, y') \in \mathbb{R}^3$.[11] Observe that $dy-y'dx = 0$ implies $d\eta-\eta' d\xi = 0$, by (5.26).

Prototype of a contact transformation. The Legendre transformation

$$\xi=y', \quad \eta=xy'-y, \quad \eta'=x \tag{5.27}$$

with the inverse transformation $x = \eta', y = \xi\eta' - \eta, y' = \xi$ represents a contact transformation. In fact,

$$d\eta-\eta' d\xi=dx\cdot y'+xdy'-dy-xdy'=-(dy-y'dx).$$

The Legendre transformation sends solutions of the original differential equation (5.21) to solutions of (5.22).

Example. Consider the Clairaut differential equation[12]

$$y-xy'=g(y'). \tag{5.28}$$

By Legendre transformation, we get $-\eta = g(\xi)$. The inverse Legendre transformation yields the parameterized solution

$$x=-g'(\xi), \quad y=-\xi g'(\xi)+g(\xi), \qquad \xi\in\mathbb{R}$$

of (5.28). Hence $y = \xi x+g(\xi), x \in \mathbb{R}$. This is a family of straight lines parameterized by the real parameter ξ.

In particular, let $g(y') := -\frac{1}{2}y'^2$. Then the Clairaut differential equation (5.28) has the family of straight lines

$$y=x\xi-\tfrac{1}{2}\xi^2, \qquad x\in\mathbb{R}, \tag{5.29}$$

as solutions where ξ is a parameter. Writing this family as $F(x, y, \xi) = 0$, the envelope of this family is obtained from the equation

[11] Expressing the differentials by partial derivatives, we get

$$B_x dx+B_y dy+B_{y'}dy'-C(A_x dx+A_y dy+A_{y'}dy')=\varrho(dy-y'dx).$$

Comparing the coefficients of dx, dy, dy', we get the following system of first-order partial differential equations: $B_x - CA_x = -\varrho y'$, $B_y - CA_y = \varrho$, $B_{y'} - CA_{y'} = 0$.

[12] Clairaut (1713–1765), Legendre (1752–1832).

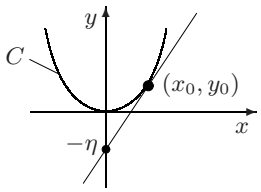


Fig. 5.11. Legendre transformation

$$F_\xi(x, y, \xi) = 0, \qquad F(x, y, \xi) = 0, \tag{5.30}$$

by elementary differential geometry. Hence $x - \xi = 0$. This yields the parabola

$$y(x) = \tfrac{1}{2}x^2, \qquad x \in \mathbb{R}. \tag{5.31}$$

Summarizing, the parabola (5.31) and the family (5.28) of its tangent lines form solutions of the differential equation $y - xy' = -\frac{1}{2}y'^2$ (Fig. 5.11).

Geometric interpretation of the Legendre transformation. A smooth curve $C : y = (x)$ can be described in the following two ways, which are dual to each other:

(i) The curve C is a set of points $\{(x, y(x)) : x \in \mathbb{R}\}$.
(ii) The curve C is the envelope of the family of its tangent lines.

The Legendre transformation passes from (i) to (ii). To explain this, consider the smooth curve $C : y = y(x)$ pictured in Fig. 5.11. Fix $x_0 \in \mathbb{R}$, and set

$$y_0 := f(x_0), \quad y_0' := f'(x_0).$$

The tangent line of the curve C at the point (x_0, y_0) is given by the equation $y = y_0'(x - x_0) + y_0$. Hence

$$y = y_0'x + y_0 - y_0'x_0.$$

This tangent line can be characterized by the two coordinates

$$\xi := y_0', \qquad \eta := y_0'x_0 - y_0.$$

Thus, the equation of the tangent line reads as

$$y - \xi x + \eta = 0, \qquad x \in \mathbb{R}.$$

The intuitive meaning of the tangent line coordinates (ξ, η) is the following:

- the symbol ξ represents the slope of the tangent line, and
- $(0, -\eta)$ represents the intersection point of the tangent line and the y-axis.

In the (ξ, η)-space, the family of tangent lines is given by an equation of the form

$$\eta = g(\xi).$$

Let us compute $g'(\xi)$ by a general argument. For the envelope C of the family of tangent lines $y - \xi x + g(\xi) = 0$, we get the system

$$-x + g'(\xi) = 0, \qquad y - \xi x + g(\xi) = 0,$$

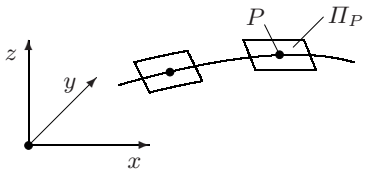


Fig. 5.12. Strip

with $x = x_0, y = y_0$, by (5.30). Hence $x_0 = g'(\xi)$. Set $\eta' := g'(\xi)$. Summarizing, we get

$$\xi := y_0', \qquad \eta := y_0' x_0 - y_0, \qquad \eta' = x_0 \qquad \text{for all} \quad (x_0, y_0) \in C.$$

Replacing (x_0, y_0, y_0') by $(x, y, y') \in \mathbb{R}^3$, we obtain the Legendre transformation $\xi := y', \eta := y'x - y, \eta' = x$ for all $(x, y, y') \in \mathbb{R}^3$.

Properties of the contact form. Consider the contact form

$$\boxed{\alpha := dz - ydx}$$

on the manifold $\mathbb{R}^3 := \{(x, y, z) : \ x, y, z \in \mathbb{R}\}$. Fix the point $P = (x_0, y_0, z_0)$ in $\mathbb{R}^3$. Let

$$x = x(t), \ \ y = y(t), \ z = z(t), \quad -1 < t < 1$$

be a smooth curve C on $\mathbb{R}^3$ which passes through the point P at time $t = 0$. The tangent vector (i.e., the velocity vector)

$$v := (\dot{x}(0), \dot{y}(0), \dot{z}(0))$$

of the trajectory C at the point P_0 is called admissible iff $\alpha_P(v) = 0$. Since

$$dx(v) = \dot{x}(0), \quad dy(v) = \dot{y}(0), \quad dz(v) = \dot{z}(0),$$

this means that

$$\dot{z}(0) - y_0 \dot{x}(0) = 0. \tag{5.32}$$

The set Π_P of all admissible velocity vectors at the point P forms a plane through the point P which is given by the equation

$$\Pi_P : z - z_0 - y_0(x - x_0) = 0 \qquad \text{for all} \quad (x, y, z) \in \mathbb{R}^3,$$

by (5.32). The pair (P, Π_P) is called the contact element at the point P (with respect to the contact form α). By definition, a strip is a family of contact elements (planes) along a curve (Fig. 5.12).

As Cauchy (1789–1857) and Lie (1842–1899) observed, the initial-value problem for general first-order partial equations can be elegantly handled by constructing the solution with the help of strips (see M. Giaquinta and S. Hildebrandt, Calculus of Variations, Vol. 2, Chap. 10, Springer, Berlin, 1995). This is one of the essential sources of contact geometry.

(i) For the derivative of the contact form α, we get[13]

$$d\alpha = d(dz) - dy \wedge dx = -dy \wedge dx = dx \wedge dy.$$

This is the so-called volume form on the (x, y)-plane.

(ii) We have $\alpha \wedge d\alpha = dz \wedge dx \wedge dy = dx \wedge dy \wedge dz$. This is the volume form on $\mathbb{R}^3$.

[13] Note that $dx \wedge dy = -dy \wedge dx$ and $dx \wedge dx = 0$. Similar relations are obtained by the cyclic permutation $x \Rightarrow y \Rightarrow z \Rightarrow x$. Moreover, $d(dx) = d(dy) = d(dz) = 0$,

5.7.2 Contact Manifolds and Contact Transformations

Contact form. Fix $n = 1, 2, \ldots$ Let $\mathcal{M}$ be a $(2n+1)$-dimensional real manifold. This is called a contact manifold iff there exists a smooth differential 1-form α on $\mathcal{M}$ with[14]

$$\alpha \wedge (d\alpha)^n \neq \quad \text{on } \mathcal{M}.$$

The form α is called the contact form.

Contact elements. The tangent vector v at the point P of $\mathcal{M}$ is called admissible iff

$$\alpha_P(v) = 0.$$

The set Π_P of all admissible tangent vectors at the point P forms a linear $2n$-dimensional subspace of the tangent space $T_P\mathcal{M}$ of $\mathcal{M}$ at the point P. The pair (P, Π_P) is called the contact element at the point P.

Example. Let $n = 1, 2, \ldots$. The $(2n+1)$-dimensional manifold

$$\mathbb{R}^{2n+1} = \{(q^1, \ldots, q^n, p_1, \ldots, p_n, z) : \ q^j, p_j, z \in \mathbb{R},\ j = 1, \ldots, n\}$$

is a contact manifold with respect to the contact form

$$\boxed{\alpha := dz - \sum_{j=1}^{n} p_j dq^j.}$$

In fact, $d\alpha = d(dz) - \sum_{j=1}^n dp_j \wedge dq^j = \sum_{j=1}^n dq^j \wedge dp_j$, and

$$\alpha \wedge (d\alpha)^n = dz \wedge dq^1 \wedge \cdots \wedge dq^n \wedge dp_1 \wedge \cdots \wedge dp_n.$$

This is called the volume form on $\mathbb{R}^{2n+1}$. The differential 2-form

$$\omega := \sum_{j=1}^{n} dq^j \wedge dp_j$$

is called the symplectic form on the $2n$-dimensional manifold

$$\mathbb{R}^{2n} = \{(q^1, \ldots, q^n, p_1, \ldots, p_n) : \ q^j, p_j \in \mathbb{R},\ j = 1, \ldots, n\}.$$

Contact transformation. Let $\mathcal{M}, \alpha$ and $\mathcal{N}, \beta$ be $(2n+1)$-dimensional contact manifolds.

A diffeomorphism $f : \mathcal{M} \to \mathcal{N}$ is called a contact transformation iff it sends contact elements to contact elements.

Explicitly, this means the following. The diffeomorphism f sends smooth curves passing through the point P to smooth curves passing through the image point $f(P)$. Differentiating the curves with respect to the curve parameter, we get the linearization $f'(P)$ which sends tangent vectors at P to tangent vectors at $f(P)$. The diffeomorphism f is a contact transformation iff

$$\alpha_P(v) = 0 \quad \text{implies} \quad \beta_{f(P)}(f'(P)v) = 0$$

for all tangent vectors v at the point P and all points $P \in \mathcal{M}$.

---

and

$$d(adx + bdy + cdz) = da \wedge dx + db \wedge dy + dc \wedge dz.$$

The theory of differential forms will be thoroughly studied in Vol. III. As an introduction to differential forms, we recommend the two textbooks by V. Zorich, Analysis, Vol. II, Springer, Berlin, 2003, and by I. Agricola and T. Friedrich, Global Analysis: Differential Forms in Analysis, Geometry and Physics, Amer. Math. Soc., Providence, Rhode Island, 2002.

[14] The symbol $(d\alpha)^n$ denotes the n-fold wedge product $d\alpha \wedge \cdots \wedge d\alpha$.

5.7.3 Applications to Geometrical Optics

Consider the manifold $\mathbb{R}^5_L := \{(x, y, y', p, L) : \ x, y, y', p, L \in \mathbb{R}\}$. Fixing (x, y) in $\mathbb{R}^2$, we obtain the 3-dimensional submanifold

$$\mathbb{R}^3_L(x, y) := \{(x, y, y', p, L) : \ y', p, L \in \mathbb{R}\},$$

which is a contact manifold with respect to the contact form

$$dL - pdy'.$$

Replacing the coordinate L by the symbol H, we obtain the manifold $\mathbb{R}^5_H$ and its submanifold $\mathbb{R}^3_H(x, y)$, which is a 3-dimensional contact manifold with respect to the contact form $dH - y'dp$. We claim that:

The Legendre transformation $H = y'p - L$ is a contact transformation from $\mathbb{R}^3_L(x, y)$ onto $\mathbb{R}^3_H(x, y)$.

Proof. Consider a curve

$$C : y' = y'(\tau), \quad p = p(\tau), \quad L = L(\tau)$$

which passes through the point $P \in \mathbb{R}^3_L(x, y)$ at time $\tau = 0$. This curve has the tangent vector $v = (\dot{y}'(0), \dot{p}(0), \dot{L}(0))$ at the point P, where the dot denotes the derivative with respect to the curve parameter τ. The Legendre transformation sends the point P, the curve C, and the tangent vector v to P_*, C_* and v_*, respectively. Explicitly,

$$C_* : y' = y'(\tau), \quad p = p(\tau), \quad H = H(\tau) = y'(\tau)p(\tau) - L(\tau).$$

Hence $v_* = (\dot{y}'(0), \dot{p}(0), \dot{H}(0))$ with

$$\dot{H}(0) = \dot{y}'(0)p(\tau) + y'(0)\dot{p}(0)) - \dot{L}(0),$$

by the product rule of calculus. This yields[15]

$$\boxed{(dH - y'(0)dp)(v_*) = -(dL - p(0)dy')(v).} \tag{5.33}$$

Thus, $(dL - p(0)dy')(v) = 0$ implies $(dH - y'(0)dp)(v_*) = 0$. This tells us that the Legendre transformation sends admissible tangent vectors to admissible tangent vectors. In other words, it sends the contact element at the point P to the contact element at the image point P_*. Consequently, the Legendre transformation is a contact transformation. □

Mnemonic approach. It follows from $H = py' - L$ that we obtain

$$dH = dp \cdot y' + pdy' - dL.$$

Hence

$$dL - pdy' = -(dH - y'dp). \tag{5.34}$$

[15] Note that $-(dL - p(0)dy')(v) = -\dot{L}(0) + p(0)\dot{y}'(0)$ and

$$(dH - y'(0)dp)(v_*) = \dot{H}(0) - y(0)\dot{p}(0) = -\dot{L}(0) + p(0)\dot{y}'(0).$$

Therefore, $dL - pdy' = 0$ implies $dH - y'dp = 0$, and we are done.

Observe that the mnemonic argument works since Leibniz's product rule $d(y'p) = dy' \cdot p + y'dp$ for differentials corresponds to the product rule

$$\frac{d(y'(\tau)p(\tau))}{d\tau} = \frac{dy'(\tau)}{d\tau} \cdot p(\tau) + y'(\tau)\frac{dp(\tau)}{d\tau}$$

for derivatives, which we have used in order to obtain the key relation (5.33) in our proof above. For practical computations, physicists and mathematicians frequently use the mnemonic argument. We will do this, too (see Problem 5.1).

Perspective. One can rigorously show that Fermat's principle of least time is equivalent to Huygens' principle for the propagation of wave fronts. To this end, one has to use one-parameter families of contact transformations in an appropriate higher-dimensional space. This can be found in Giaquinta and Hildebrandt (1995), Vol. 2, Sect. 3.4. In a general setting, we will study this in Vol. III by considering the Cartan–Kähler theorem for systems of differential forms and its applications to the following topics:

- differential geometry (e.g., the main theorem on the construction of surfaces by means of the Gaussian invariants and its generalization to higher dimensions),
- the theory of Lie groups (the reconstruction of a Lie group from its Lie algebra),
- geometrical optics, and
- thermodynamics.

5.7.4 Equilibrium Thermodynamics and Legendre Submanifolds

> Thermodynamical systems in equilibrium states (e.g., chemical substances) can be described by the Gibbs contact form and its integral manifolds.
>
> Folklore

> Every mathematician knows that it is impossible to understand any elementary course in thermodynamics. The reason is that thermodynamics is based – as Gibbs (1839–1903) has explicitly proclaimed – on the rather complicated mathematical theory of contact geometry.[16]
>
> Vladimir Arnold, 1990

Consider a chemical substance in thermodynamical equilibrium. The physics of this substance is governed by the Gibbs contact form

$$\boxed{\Gamma := dE - TdS + PdV - \mu dN.} \tag{5.35}$$

Here, the symbols have the following physical meaning:

- T (absolute temperature),
- V (volume), P (pressure), N (particle number),
- E (inner energy), S (entropy), μ (chemical potential).

Physicists also introduce the following quantities:

- $F := E - TS$ (free energy),
- $\Omega := E - TS - \mu N$ (statistical potential or Gibbs potential),
- $H := E + PV$ (enthalpy),

[16] V. Arnold, Contact geometry: the geometrical method of Gibbs' thermodynamics, Proceedings of the Gibbs Symposium, Yale University, Connecticut (U.S.A), 1989, pp. 163–179, Amer. Math. Soc., Providence, Rhode Island, 1990.

- $G = F + PV$ (free enthalpy).

In physics, one has frequently to change the independent thermodynamical variables.

The following approach based on the Gibbs contact form is very flexible with respect to changing variables.

The fundamental Gibbs equation. The 7-dimensional real manifold

$$\mathbb{R}^7_{\text{Gibbs}} := \{(E, T, S, P, V, \mu, N) \in \mathbb{R}^7\}$$

is called the Gibbs manifold. This is a contact manifold with respect to the Gibbs contact form (5.35). A submanifold $\mathcal{M}$ of the Gibbs manifold $\mathbb{R}^7_{\text{Gibbs}}$ is called an *integral manifold* of the Gibbs contact form Γ iff the fundamental Gibbs equation

$$\boxed{\Gamma = 0 \qquad \text{on } \mathcal{M}} \tag{5.36}$$

is satisfied. This means that

$$\Gamma_{\mathcal{P}}(v) = 0 \tag{5.37}$$

for all tangent vectors v at the point $\mathcal{P}$ and all points $\mathcal{P}$ in $\mathcal{M}$. The maximal dimension of an integral manifold for Γ is equal to three. Integral manifolds of maximal dimension are called *Legendre submanifolds*.

Thermodynamical processes. Let us first consider one-dimensional integral manifolds of the Gibbs contact form Γ. Consider the smooth curve

$$\mathcal{C}: \quad t \mapsto (E(t), T(t), S(t), P(t), V(t), \mu(t), N(t)) \tag{5.38}$$

on the open time interval $\mathcal{U}$. By definition, the curve $\mathcal{C}$ describes a thermodynamical process iff it is an one-dimensional integral manifold of Γ. The velocity vector looks like

$$v(t) = (\dot{E}(t), \dot{T}(t), \dot{S}(t), \dot{P}(t), \dot{V}(t), \dot{\mu}(t), \dot{N}(t)).$$

In order to get a submanifold $\mathcal{C}$ of the Gibbs manifold, we have to assume that $v(t) \neq 0$ for all $t \in \mathcal{U}$. Note that dE is a linear functional at each point of the Gibbs manifold with $dE(v(t)) = \dot{E}(t)$, as well as $dS(v(t)) = \dot{S}(t)$ and so on. The curve $\mathcal{C}$ satisfies equation (5.37) iff

$$\boxed{\dot{E}(t) - T(t)\dot{S}(t) + P(t)\dot{V}(t) - \mu(t)\dot{N}(t) = 0, \quad t \in \mathcal{U}.} \tag{5.39}$$

Suppose that we are given the smooth energy function

$$E = E(S, V, N) \qquad \text{for all} \quad (S, V, N) \in \mathcal{V},$$

where $\mathcal{V}$ is a nonempty open subset of $\mathbb{R}^3$. Choose $E(t) := E(S(t), V(t), N(t))$. Fix the point $\mathcal{P}$ and suppose that the curve $\mathcal{C}$ passes through the point $\mathcal{P}$ at the fixed time t. By the chain rule, it follows from (5.39) that

$$E_S(\mathcal{P})\dot{S}(t) + E_V(\mathcal{P})\dot{V}(t) + E_N(\mathcal{P})\dot{N}(t) = T(t)\dot{S}(t) - P(t)\dot{V}(t) + \mu(t)\dot{N}(t).$$

Since the curve $\mathcal{C}$ can be chosen in such a way that the derivatives $\dot{E}(t), \dot{S}(t)$ and $\dot{N}(t)$ attain arbitrary values, comparison of coefficients gives

$$\boxed{T = E_S(\mathcal{P}), \quad P = -E_V(\mathcal{P}), \quad \mu = E_N(\mathcal{P}), \qquad \mathcal{P} \in \mathcal{V}.} \tag{5.40}$$

In terms of physics, this equation tells us the crucial fact that the energy function $E = E(S, V, N)$ allows us to compute all the remaining physical quantities T, P, μ by merely using differentiation. This motivates why the function $E = E(S, V, N)$ is called a *thermodynamical potential* by physicists.

Reversibility and heat production. A process realized in nature is called *reversible* iff the process obtained by time reflection can also be realized in nature. Otherwise the process is called irreversible. Biological processes are irreversible. For example, the life process of a human being cannot be reversed in time. Applying time reflection $t \mapsto -t$ to the process $\mathcal{C}$ from (5.38), we get

$$\mathcal{C}_- : \quad t \mapsto (E(-t), T(-t), S(-t), P(-t), V(-t), \mu(-t), N(-t)).$$

If the process equation (5.39) is satisfied for $\mathcal{C}$, then it is also satisfied for $\mathcal{C}_-$. Thus, the process $\mathcal{C}$ is reversible. The process $\mathcal{C}$ is called quasi-stationary. In terms of physics, this means that the process runs very slowly in such a way that, approximately, the chemical substance is in a thermodynamical equilibrium state at each point in time. This represents an idealization of real thermodynamical processes. For quasi-stationary processes, the integral

$$Q([t_0, t_1]) = \int_{t_0}^{t_1} T(t)\dot{S}(t)\, dt$$

is the amount of heat energy produced by the process during the time interval $[t_0, t_1]$. More general thermodynamical processes including irreversible processes will be discussed in Sect. 7.17.9 in the setting of the second law in thermodynamics. It turns out that entropy measures the complexity and the stored information of many-particle systems in nature.

Coordinate transformation on the Gibbs manifold and the statistical potential Ω. By setting

$$\Omega := E - ST - \mu N$$

with the inverse transformation $E = \Omega + ST + \mu T$, let us introduce the new coordinates $(\Omega, T, S, P, V, \mu, N)$ on the Gibbs manifold $\mathbb{R}^7_{\text{Gibbs}}$. The process (5.38) yields $\Omega(t) = E(t) - S(t)T(t) - \mu(t)N(t)$. Differentiating this with respect to time t, the product rule tells us that

$$\dot{E}(t) = \dot{\Omega}(t) + \dot{S}(t)T(t) + S(t)\dot{T}(t) + \dot{\mu}(t)N(t) + \mu(t)\dot{N}(t).$$

From the process equation (5.39), we obtain

$$\boxed{\dot{\Omega}(t) = -S(t)\dot{T}(t) - P(t)\dot{V}(t) - N(t)\dot{\mu}(t), \quad t \in \mathcal{U}.} \tag{5.41}$$

Suppose that we are given the smooth function

$$\Omega = \Omega(T, V, \mu), \qquad (T, V, \mu) \in \mathcal{W}$$

where $\mathcal{W}$ is a nonempty open subset of $\mathbb{R}^3$. As above, equation (5.41) implies

$$S = -\Omega_T, \quad P = -\Omega_V, \quad N = -\Omega_\mu. \tag{5.42}$$

In addition, $E = \Omega - ST - \mu N$. The statistical potential Ω introduced by Gibbs is crucial in thermodynamics. We will show in Sect. 7.17.3 on page 638 that the statistical potential Ω can be constructed by means of the method of statistical physics. Then the other thermodynamical quantities can be obtained from (5.42).

Mnemonic approach. Dividing the equation $\Gamma = 0$, that is,

$$dE = TdS - PdV + \mu dN \tag{5.43}$$

by the symbol dt, we get the process equation

$$\frac{dE}{dt} = T\frac{dS}{dt} - P\frac{dV}{dt} + \mu\frac{dN}{dt}.$$

This is (5.39). Let $E = E(S, V, N)$. From (5.43), we get

$$E_S dS + E_V dV + E_N dN = TdS - PdV + \mu dN.$$

Comparing the coefficients of dS, dV, dN, we obtain

$$E_S = T, \quad E_V = -P, \quad E_N = \mu.$$

Passing to $\Omega = E - ST - \mu N$, we get $d\Omega = dE - dS \cdot T - SdT - d\mu \cdot N + \mu dN$. Using (5.43),

$$d\Omega = -SdT - PdV - Nd\mu.$$

If $\Omega = \Omega(T, V, \mu)$, then

$$\Omega_T dT + \Omega_V dV + \Omega_\mu d\mu = -SdT - PdV - Nd\mu.$$

Hence $\Omega_T = -S, \Omega_V = -P, \Omega_\mu = -N$. The rigorous justification of this mnemonic approach can be based on Problem 5.1.

Thermodynamical potentials. Now we want to study Legendre submanifolds of the Gibbs contact form Γ. Such submanifolds describe the physics of chemical substances in terms of thermodynamical potentials.

(i) Basic variables S, V, N (inner energy E as thermodynamical potential): Set $\mathcal{P} := (S, V, N)$. Let

$$\mathcal{P} \mapsto (E, T, S, P, V, \mu, N) \tag{5.44}$$

be a smooth immersion from the nonempty open subset $\mathcal{V}$ of $\mathbb{R}^3$ into the Gibbs manifold $\mathbb{R}^7_{\text{Gibbs}}$ such that

$$\boxed{T = E_S, \quad P = -E_V, \quad \mu = E_N \qquad \text{on } \mathcal{V}.}$$

Then the image of this map is a Legendre submanifold of the Gibbs contact form Γ. The immersion property guarantees that the image of the map (5.44) is a 3-dimensional submanifold of the Gibbs manifold.[17]

[17] The smooth map $(x, y) \mapsto (u(x, y), v(x, y), w(x, y))$ from an open subset $\mathcal{U}$ of $\mathbb{R}^2$ into $\mathbb{R}^3$ is an immersion iff the linearization of this map is injective at each point $\mathcal{P} = (x_0, y_0)$. That is, the matrix of the first-order partial derivatives

$$\begin{pmatrix} u_x(\mathcal{P}) & v_x(\mathcal{P}), & w_x(\mathcal{P}) \\ u_y(\mathcal{P}) & v_y(\mathcal{P}) & w_y(\mathcal{P}) \end{pmatrix} \tag{5.45}$$

has maximal rank for all $\mathcal{P} \in \mathcal{U}$ (i.e., rank = 2 in the present case).

An analogous result holds for general maps from $(x_1, \ldots, x_m)$ to $(u_1(x_1, \ldots, x_m), \ldots, u_n(x_1, \ldots, x_m))$, where $1 \leq m < n$ (see Zeidler (1986), Vol. IV, p. 551ff, quoted on page 1049).

(ii) Basic variables: T, V, μ (statistical potential $\Omega := E - ST - \mu N$ as thermodynamical potential): Set $\mathcal{P} := (T, V, \mu)$. Let

$$\mathcal{P} \mapsto (\Omega, T, S, P, V, \mu, N)$$

be a smooth immersion from the nonempty open subset $\mathcal{V}$ of $\mathbb{R}^3$ into the Gibbs manifold $\mathbb{R}^7_{\text{Gibbs}}$ (equipped with new coordinates) such that

$$\boxed{S = -\Omega_T, \quad P = -\Omega_V, \quad N = -\Omega_\mu \quad \text{on } \mathcal{V}.}$$

Then the image of this map is a Legendre submanifold of the contact form Γ which is equal to $d\Omega + SdT + PdV + Nd\mu$.

(iii) Basic variables: T, V, N (free energy $F := E - TS$ as thermodynamical potential): Set $\mathcal{P} := (T, V, N)$. Let

$$\mathcal{P} \mapsto (F, T, S, P, V, \mu, N)$$

be a smooth immersion from the nonempty open subset $\mathcal{V}$ of $\mathbb{R}^3$ into the Gibbs manifold $\mathbb{R}^7_{\text{Gibbs}}$ (equipped with new coordinates) such that

$$\boxed{S = -F_T, \quad P = -F_V, \quad \mu = F_N \quad \text{on } \mathcal{V}.}$$

Then the image of this map is a Legendre submanifold of the contact form Γ which is equal to $dF + SdT + PdV - \mu dN$.

(iv) Basic variables: T, P, N (free enthalpy $G := E - ST + PV$ as thermodynamical potential). Set $\mathcal{P} := (T, P, N)$. Let

$$\mathcal{P} \mapsto (G, T, S, P, V, \mu, N)$$

be a smooth immersion from the nonempty open subset $\mathcal{V}$ of $\mathbb{R}^3$ into the Gibbs manifold $\mathbb{R}^7_{\text{Gibbs}}$ (equipped with new coordinates) such that

$$\boxed{S = -G_T, \quad V = -G_P, \quad \mu = G_N \quad \text{on } \mathcal{V}.}$$

Then the image of this map is a Legendre submanifold of the contact form Γ which is equal to $dG + SdT + VdP - \mu dN$.

5.8 Light Rays and Non-Euclidean Geometry

Geometry is the essential of the beauty of the world.[18]
Johannes Kepler, Harmonies of the World, 1619

In humbleness we have to admit that if "number" is a product of our imagination, "space" has a reality outside of our imagination, to which we *a priori* cannot assign its laws.

Carl Friedrich Gauss in a letter to Bessel

[18] The original Latin version in Kepler's treatise *Harmonices mundi* reads as follows: *Geometria est archetypus pulcheritudinis mundi.*

An axiomatic approach to Euclidean geometry was formulated by the Greek mathematician Euclid (about 360 B.C.–290 B.C.) in his famous *Elements*. More than 2000 years later, three mathematicians proved independently that there exist logically consistent non-Euclidean geometries for which Euclid's parallel axiom does not hold, namely, Gauss (1777–1855) in 1792ff, Lobachevsky (1792–1856) in 1826ff, and Bólayi, Jr. (1802–1860) in 1832.[19] Morris Kline writes in his history of mathematics:[20]

> Gauss said in a letter to Bessel of January 27, 1829, that he probably would never publish his findings in non-Euclidean geometry, because he feared ridicule, or, as he put it, the clamor of Boeotians, a figurative reference to a dull-witted Greek tribe.

The reader should notice that in Gauss' time the intellectual world at large was dominated by the preeminent German philosopher Immanuel Kant (1742–1804) who claimed in his 1781 major work *Critique of Pure Reason* that Euclidean geometry is fixed *a priori* in all the human brains. Gauss did not share Kant's opinion. The development of Einstein's theory of general relativity proved that Gauss was right.

Models of non-Euclidean geometry (based on subsets of the Euclidean plane equipped with a non-Euclidean metric) were given by Beltrami (1835–1900) in 1868, Klein (1842–1925) in 1871, and Poincaré (1854–1912) in 1902. The Poincaré model was very influential because of his close relation to elementary geometric and physical intuition.[21] We will study this model and its relations to modern mathematics and physics in Sect. 5.10. The modern axiomatic approach to geometry was published by Hilbert in his 1899 monograph.[22]

The fascinating history of geometry is described by C. Scriba and P. Schreiber, 5000 Years of Geometry, Springer, Heidelberg, 2003 (in German). We also refer to the monographs by F. Klein, Development of Mathematics in the 19th Century, Math. Sci. Press, New York, 1979, and by E. Scholz, The History of the Idea of Manifold, Birkhäuser, Basel, 1980 (in German).

---

[19] The Hungarian mathematician Bólyai Farkas (1775–1856), a friend of Gauss, published his textbook on geometry in Latin entitled *Tentamen* (Attempt at explanation) in 1832. His son, Bólyai Janos, wrote an Appendix to this textbook, where he represented his new hyperbolic non-Euclidean geometry.
N. Lobachevsky, The Elements of Geometry, Kasan, 1829 (in Russian); Geometrical Investigations on the Theory of Parallel Lines, Berlin, 1840 (in German).
We recommend the following two collections of fundamental papers, which reflect a fascinating part of the history of both mathematics and physics:
H. Reichhardt (Ed.), Gauss and the Beginnings of Non-Euclidean Geometry (with articles by Bólayi, Lobachevsky, and Klein), Teubner, Leipzig, 1985 (in German).
J. Böhm and H. Reichhardt (Eds.), Gauss' Surface Theory, Riemannian Manifolds, and the Minkowski Space-Time (with papers by Gauss, Riemann, and Minkowski), Teubner, Leipzig, 1984 (in German).

[20] M. Kline, Mathematics Thought from Ancient to Modern Times, Oxford University Press, New York, Vol. 3, 1972.

[21] H. Poincaré, La science et l'hypothèse, Flamarion, Paris, 1902 (in French).

[22] D. Hilbert, Foundations of Geometry, Teubner, Leipzig, 1899 (in German). The 12th edition appeared in 1977.

5.8.1 Linear Symplectic Spaces

Symplectic forms. Let Y be a real linear space. By a symplectic form κ on Y, we understand a bilinear map $\kappa : Y \times Y \to \mathbb{R}$ which has the following two properties:

- $\kappa(v, w) = -\kappa(w, v)$ for all $v, w \in Y$ (antisymmetry).
- If $\kappa(v_0, w) = 0$ for all $w \in Y$ and fixed $v_0 \in Y$, then $v_0 = 0$ (non-degeneracy).

The real linear space Y equipped with a symplectic form κ is called a symplectic space.

Symplectic morphism. Let Y, Z be real, symplectic spaces with the symplectic forms κ, σ, respectively. By definition, a symplectic morphism is a linear map $A : Y \to Z$ with

$$\sigma(Av, Aw) = \kappa(v, w) \qquad \text{for all} \quad v, w \in Y.$$

If, in addition, this map A is bijective, then it is called a symplectic isomorphism.

Almost complex space. The space Y is called almost complex iff it is a real linear space and there exists a linear operator $J : Y \to Y$ with $J^2 = -I$. Here, I denotes the unit operator.

Example 1 (the space $\mathbb{C}_\mathbb{R}$). Let $z = x + \mathrm{i}y$ and $\zeta = \xi + \mathrm{i}\eta$ be complex numbers, that is, $x, y, \xi, \eta \in \mathbb{R}$. The set $\mathbb{C}$ of all complex numbers is a one-dimensional complex Hilbert space equipped with the inner product

$$\langle z|\zeta\rangle := z^\dagger \zeta.$$

Explicitly,

$$\boxed{\langle z|\zeta\rangle = (x - \mathrm{i}y)(\xi + \mathrm{i}\eta) = (x\xi + y\eta) + \mathrm{i}(x\eta - y\xi).} \tag{5.46}$$

Let $z \in \mathbb{C}$. Restricting the multiplication αz to real numbers α, the set $\mathbb{C}$ becomes a linear space denoted by $\mathbb{C}_\mathbb{R}$.

- $\mathbb{C}_R$ is a real 2-dimensional linear space. Setting

$$\langle z|\zeta\rangle_\mathbb{R} := \Re(\langle z|\zeta\rangle) \ \text{ and } \ \kappa(z, \zeta) := \Im(\langle z|\zeta\rangle) \qquad \text{for all} \quad z, \zeta \in \mathbb{C}_\mathbb{R},$$

  we get the key decomposition

$$\boxed{\langle z|\zeta\rangle = \langle z|\zeta\rangle_\mathbb{R} + \mathrm{i}\kappa(z, \zeta) \qquad \text{for all} \quad z, \zeta \in \mathbb{C}_\mathbb{R}.}$$

  Explicitly, $\langle z|\zeta\rangle_\mathbb{R} = x\xi + y\eta$ and $\kappa(z, \zeta) = x\eta - y\xi$. The following are met:
- $\mathbb{C}_\mathbb{R}$ is a real 2-dimensional Hilbert space equipped with the inner product $\langle z|\zeta\rangle_\mathbb{R}$. The elements 1 and i form an orthonormal basis of $\mathbb{C}_\mathbb{R}$. The real space $\mathbb{C}_\mathbb{R}$ is called the realification of the complex space $\mathbb{C}$.
- $\mathbb{C}_\mathbb{R}$ is a symplectic space equipped with the symplectic form κ. The two complex numbers $z = x + \mathrm{i}y$ and $\zeta = \xi + \mathrm{i}\zeta$ span a parallelogram whose area is equal to $\kappa(z, \zeta)$ (Fig. 5.13).
- $\mathbb{C}_\mathbb{R}$ is an almost complex space equipped with the operator $Jz := \mathrm{i}z$ for all $z \in \mathbb{C}_\mathbb{R}$.

For all $z, \zeta \in \mathbb{C}_\mathbb{R}$, the following hold:

- $\langle Jz|J\zeta\rangle_\mathbb{R} = \langle z|\zeta\rangle_\mathbb{R}$ (i.e., J is unitary on $\mathbb{C}_\mathbb{R}$).
- $\kappa(z, \zeta) = \langle Jz|\zeta\rangle_\mathbb{R}$ and $\kappa(Jz, J\zeta) = \kappa(z, \zeta)$ (J is symplectic on $\mathbb{C}_\mathbb{R}$).

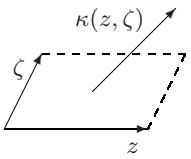


Fig. 5.13. Kähler form

This follows from the fact that $\langle .|. \rangle$ is an inner product on $\mathbb{C}$. For example,

$$\langle Jz|J\zeta\rangle_{\mathbb{R}} = \Re(\langle \mathrm{i}z|\mathrm{i}\zeta\rangle) = \Re(-\mathrm{i}^2\langle z|\zeta\rangle) = \Re(\langle z|\zeta\rangle) = \langle z|\zeta\rangle_{\mathbb{R}}.$$

Furthermore, $\kappa(Jz, J\zeta) = \Im(\langle Jz|J\zeta\rangle) = \Im(\langle z|\zeta\rangle) = \kappa(z,\zeta)$.

Example 2 (the space $\mathbb{R}^2$). For real numbers x, y, ξ, η, set

$$v := \begin{pmatrix} x \\ y \end{pmatrix}, \quad w := \begin{pmatrix} \xi \\ \eta \end{pmatrix}, \quad \mathcal{J} := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Note that $\mathcal{J}^2 = -I$. The space $\mathbb{R}^2$ consists of all matrices of the form v, w above. Define $\langle v|w\rangle_{\mathbb{R}} := v^d w$ and $\sigma(v,w) := v^d \mathcal{J} w$. Explicitly, we obtain $\langle v|w\rangle_{\mathbb{R}} = x\xi + y\eta$ and

$$\sigma(v,w) = \det(v,w) = \begin{vmatrix} x & \xi \\ y & \eta \end{vmatrix} = x\eta - y\xi.$$

The following hold:

- $\mathbb{R}^2$ is a real 2-dimensional linear space.
- $\mathbb{R}^2$ is a real Hilbert space equipped with the inner product $\langle .|. \rangle_{\mathbb{R}}$.
- $\mathbb{R}^2$ is a symplectic space equipped with the symplectic form σ.
- $\mathbb{R}^2$ is an almost complex space equipped with the operator $\mathcal{J}$.

The two examples above are intimately related to each other. In fact, the map

$$x + \mathrm{i}y \mapsto (x, y)$$

yields a linear isomorphism from the real space $\mathbb{C}_{\mathbb{R}}$ onto the real space $\mathbb{R}^2$. This map preserves both the Hilbert space structure and the symplectic structure (i.e., the map is unitary, and it is a symplectic isomorphism).

Matrix groups. The set of all real (2×2)-matrices

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \qquad \alpha, \beta, \gamma, \delta \in \mathbb{R}$$

with $\det A \neq 0$ (i.e., $\alpha\delta - \beta\gamma \neq 0$) forms a group with respect to matrix multiplication. More precisely, this is a real 4-dimensional Lie group denoted by $GL(2,\mathbb{R})$ (general real linear group). The subgroup of $GL(2,\mathbb{R})$ which consists of all real (2×2)-matrices A with $\det A = 1$ (i.e., $\alpha\delta - \beta\gamma = 1$) forms a real 3-dimensional Lie group which is denoted by $SL(2,\mathbb{R})$ (special real linear group).

Proposition 5.6 *For a linear operator $A : \mathbb{R}^2 \to \mathbb{R}^2$, the following statements are equivalent:*

(i) *A is a symplectic isomorphism.*
(ii) $A^d \mathcal{J} A = \mathcal{J}$.
(iii) $A \in SL(2, \mathbb{R})$ (i.e., $\det A = 1$).

Proof. (i) $\Leftrightarrow$ (ii): Use $\sigma(Av, Aw) = (Av)^d \mathcal{J} Aw = v^d (A^d \mathcal{J} A) w$ together with $\sigma(v, w) = v^d \mathcal{J} w$.

(ii)$\Leftrightarrow$ (iii): Explicitly, $A^d \mathcal{J} A = \mathcal{J}$ means that

$$\begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

This is equivalent to $\alpha\delta - \beta\gamma = 1$. □

Symplectic geometry. The set of symplectic isomorphisms forms the group $SL(2, \mathbb{R})$ which is called the symplectic group. This is a Lie group. Precisely the invariants of $SL(2, \mathbb{R})$ are properties of the symplectic geometry on $\mathbb{R}^2$. The most important symplectic invariant is the area

$$\sigma(v, w) = \det(v, w) = \begin{vmatrix} x & \xi \\ y & \eta \end{vmatrix}$$

spanned by the vectors v and w. In fact, if $A \in SL(2, \mathbb{R})$, then $\sigma(Av, Aw)$ is equal to $\sigma(v, w)$. That is, $\det(Av, Aw) = \det(v, w)$ for all $A \in SL(2, \mathbb{R})$. The symplectic group $SL(2, \mathbb{R})$ and its fundamental applications are studied in the monograph by S. Lang, $SL(2, \mathbb{R})$, Addison–Wesley, Reading, Massachusetts, 1975. In a straightforward way, the preceding examples can be generalized to complex Hilbert spaces, as we will show in the next section.

The Lie algebra $sl(2, \mathbb{R})$. The relation between Lie groups and Lie algebras will be thoroughly studied in Vol. III. We will show that:

> *The theory of Lie groups and their Lie algebras represents a far-reaching generalization of the exponential function and the logarithmic function.*

At this point, we only make the following remark with a view to hyperbolic non-Euclidean geometry below. The set of all real (2×2)-matrices

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \qquad \alpha, \beta, \gamma, \delta \in \mathbb{R} \tag{5.47}$$

with $\operatorname{tr}(A) = 0$ (i.e., $\alpha + \delta = 0$) forms a real 3-dimensional Lie algebra with respect to the Lie product $[A, B] := AB - BA$. This Lie algebra is denoted by $sl(2, \mathbb{R})$. Setting

$$B := \mathrm{e}^A, \qquad A \in sl(2, \mathbb{R}),$$

we get a map $A \mapsto B$ from the Lie algebra $sl(2, \mathbb{R})$ into the Lie group $SL(2, \mathbb{R})$. This so-called exponential map is a diffeomorphism from a sufficiently small open neighborhood of the origin $A = 0$ in $sl(2, \mathbb{R})$ onto a sufficiently small open neighborhood of the unit element I in $SL(2, \mathbb{R})$. Observe that the inverse map $A = \ln B$ yields a parametrization of the real 3-dimensional manifold $SL(2, \mathbb{R})$ near the unit element I. The Lie algebra $sl(2, \mathbb{R})$ represents the parameter space.

On page 317, we will introduce the symmetry group $Sym(\mathcal{H}_{\mathbb{C}})$ of the hyperbolic geometry on the hyperbolic plane. There exists a group epimorphism

$$\varrho : SL(2,\mathbb{R}) \to Sym(\mathcal{H}_{\mathbb{R}})$$

with the kernel $N = \{I, -I\}$. Thus, we obtain the group isomorphism

$$\boxed{Sym(\mathcal{H}_{\mathbb{R}}) \simeq SL(2,\mathbb{R})/N.} \tag{5.48}$$

In a sufficiently small neighborhood of the unit element I, the groups $Sym(\mathcal{H}_{\mathbb{R}})$ and $SL(2,\mathbb{R})$ are locally isomorphic. Hence the Lie algebra of $Sym(\mathcal{H}_{\mathbb{R}})$ is equal to the Lie algebra of $SL(2,\mathbb{R})$. Thus, $Sym(\mathcal{H}_{\mathbb{R}})$ has the Lie algebra $sl(2,\mathbb{R})$. This fact will critically be used in Sect. 5.11 in order to display Ariadne's thread to gauge theory in terms of hyperbolic light-ray geometry.

The Lie group $SL(2,\mathbb{C})$ and Einstein's theory of special relativity. The symbol $SL(2,\mathbb{C})$ denotes the set of all complex (2×2)-matrices

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \qquad \alpha, \beta, \gamma, \delta \in \mathbb{C}$$

with $\det A = 1$, that is, $\alpha\delta - \beta\gamma = 1$. For complex (2×2)-matrices, we have

$$A \in SL(2,\mathbb{C}) \qquad \text{iff} \qquad A^d \mathcal{J} A = \mathcal{J}.$$

Therefore, the group $SL(2,\mathbb{C})$ is the symmetry group of the symplectic geometry on the complex linear space $\mathbb{C}^2$.

The group $SL(2,\mathbb{C})$ plays the key role in special relativity.

More precisely, the change of inertial systems under preservation of the orientation of both space and time is described by the proper Lorentz group $SO^+(3,1)$. The simply connected, real, 6-dimensional Lie group $SL(2,\mathbb{C})$ is the universal covering group of $SO^+(3,1)$. Explicitly, we have the Lie group isomorphism

$$\boxed{SO^+(3,1) \simeq SL(2,\mathbb{C})/N} \tag{5.49}$$

with the normal subgroup $N := \{I, -I\}$ of $SL(2,\mathbb{C})$. From the mathematical point of view, the Lie group epimorphism

$$\chi : SL(2,\mathbb{C}) \to SO^+(3,1)$$

with the kernel N is responsible for the appearance of the electron spin (intrinsic angular momentum of the electron) in special relativity. We will thoroughly study this in Vol. III. As an introduction, we recommend the classics by B. van der Waerden, Group Theory and Quantum Mechanics, Springer, New York, 1974.

Comparing the group isomorphisms (5.48) and (5.49) with each other, one discovers that Einstein's theory of special relativity represents a quite natural generalization of non-Euclidean hyperbolic geometry, in terms of the Lie groups $SL(2,\mathbb{R})$ and $SL(2,\mathbb{C})$.

The group $SU(n)$. Let $n = 1, 2, \ldots$ Recall the following terminology.

- $U(n)$: A complex $(n \times n)$-matrix A is called unitary iff $AA^\dagger = A^\dagger A = I$. These matrices form the group $U(n)$ which is a real n^2-dimensional Lie group.
- $U(1)$: In particular, the real 1-dimensional Lie group $U(1)$ consists of all complex numbers z with $|z| = 1$.
- $SU(n)$: The matrices A in $U(n)$ with $\det A = 1$ form a real Lie group of dimension $n^2 - 1$. In particular, $SU(2)$ is a subgroup of $SL(2,\mathbb{C})$.

It is quite remarkable that the low-dimensional Lie groups

$$U(1),\ SU(2),\ SU(3),\ SL(2,\mathbb{C})$$

play a fundamental role in describing elementary particles in the setting of the Standard Model.[23]

In contrast to this, the symmetry group for describing gravitation in the setting of Einstein's theory of general relativity is the complicated infinite-dimensional diffeomorphism group of the real 4-dimensional space-time manifold, which is a pseudo-Riemannian manifold. In the search for a unified theory for the four fundamental forces in nature, one has to combine the finite-dimensional Lie groups describing elementary particles with infinite-dimensional groups related to gravitation. This is an open problem.

5.8.2 The Kähler Form of a Complex Hilbert Space

> The splitting of the inner product of a complex Hilbert space into real part and complex part corresponds to a splitting of the unitary geometry in Hilbert spaces into symplectic geometry and Kähler geometry. These geometries are fundamental for both classical and modern physics.
>
> Folklore

This section serves as a preparation for the theory of Kähler mannifolds to be considered in the next section in the framework of non-Euclidean geometry.

From the complex Hilbert space X to the real Hilbert space $X_{\mathbb{R}}$. Let X be a complex Hilbert space with the inner product $\langle .|. \rangle$. For all $v, w \in X$, define

$$\boxed{\langle v|w\rangle_{\mathbb{R}} := \Re(\langle v|w\rangle), \qquad \kappa(v,w) := \Im(\langle v|w\rangle).}$$

Here, $\kappa : X \times X \to \mathbb{R}$ is called the Kähler form of the complex Hilbert space X. Alternatively, observing that $\langle v|w\rangle^\dagger = \langle w|v\rangle$, we get

$$\langle v|w\rangle_{\mathbb{R}} = \frac{1}{2}\big(\ \langle v|w\rangle + \langle w|v\rangle\ \big), \qquad \kappa(v,w) = \frac{1}{2\mathrm{i}}\big(\ \langle v|w\rangle - \langle w|v\rangle\ \big)$$

for all $v, w \in X$. The following hold: Let $v, w \in X$. Considering only real linear combinations $\alpha v + \beta w$ with $\alpha, \beta \in \mathbb{R}$, the set X becomes a real linear space denoted by $X_{\mathbb{R}}$. Then:

- $X_{\mathbb{R}}$ is a real Hilbert space equipped with the inner product $\langle .|. \rangle_{\mathbb{R}}$. The real Hilbert space $X_{\mathbb{R}}$ is called the realification of the original complex Hilbert space X.
- $X_{\mathbb{R}}$ is a symplectic space equipped with the symplectic form κ.

Proof. Note that $\langle v|v\rangle$ is real for all $v \in X$. Therefore, $\langle v|v\rangle_{\mathbb{R}} = 0$ implies $\langle v|v\rangle = 0$. Hence $v = 0$. Furthermore, it follows from $\langle w|v\rangle^\dagger = \langle v|w\rangle$ that

$$\kappa(v,w) = \Im(\langle v|w\rangle) = -\Im(\langle w|v\rangle) = -\kappa(w,v).$$

Finally, suppose that $\kappa(v_0, v) = 0$ for all $v \in X_{\mathbb{R}}$. Choose $v := \mathrm{i}v_0$. Then $\Im(\langle v_0|\mathrm{i}v_0\rangle) = 0$. Thus, $\Im(\mathrm{i}\langle v_0|v_0\rangle) = 0$. This implies $\langle v_0|v_0\rangle_{\mathbb{R}} = 0$, and hence $v_0 = 0$. □

Define the operator $Jv := \mathrm{i}v$ for all $v \in X$. Then, $J^2 = -I$. For all $v, w \in X_{\mathbb{R}}$, we get:

[23] Eugene Wigner (1902–1995) wrote a famous essay *On the unreasonable effectiveness of mathematics in the natural sciences*, Comm. Pure Appl. Math. **13** (1960), 1–14.

- $\langle Jv|Jw\rangle_{\mathbb{R}} = \langle v|w\rangle_{\mathbb{R}}$ and
- $\langle Jv|w\rangle_{\mathbb{R}} = -\langle v|Jw\rangle_{\mathbb{R}}$.

Thus, the operator $J : X_{\mathbb{R}} \to X_{\mathbb{R}}$ is unitary and skew-adjoint. In fact,

$$\langle Jv|Jw\rangle_{\mathbb{R}} = \langle \mathrm{i}v|\mathrm{i}w\rangle_{\mathbb{R}} = \Re(-\mathrm{i}^2\langle v|w\rangle) = \langle v|w\rangle_{\mathbb{R}}.$$

Moreover, $\langle Jv|w\rangle_{\mathbb{R}} = \Re(-\mathrm{i}\langle v|w\rangle) = -\Re(\langle v|\mathrm{i}w\rangle) = -\langle v|Jw\rangle_{\mathbb{R}}$. This tells us that $J^\dagger = -J$.

From the real Hilbert space Y to the complex Hilbert space $Y_{\mathbb{C}}$. Let Y be a real Hilbert space with the inner product $\langle .|.\rangle_{\mathbb{R}}$, and let Y be almost complex, that is, there exists a linear operator $\mathcal{J} : Y \to Y$ with

$$\boxed{\mathcal{J}^2 = -I.}$$

Let $v \in Y$. Define the complex multiplication

$$(\alpha + \beta\mathrm{i})v := \alpha v + \beta\mathcal{J}v \qquad \text{for all} \quad \alpha, \beta \in \mathbb{R}.$$

This way, the space Y becomes a complex linear space denoted by $Y_{\mathbb{C}}$. One checks this by direct computation. For example,

$$\mathrm{i}(\mathrm{i}v) = \mathcal{J}(\mathcal{J}v) = \mathcal{J}^2 v = -v \qquad \text{for all} \quad v \in Y.$$

Similarly, one gets $a(bv) = (ab)v$ for all $a, b \in \mathbb{C}$. The complex space $Y_{\mathbb{C}}$ is called a complexification of the real space Y.

Proposition 5.7 *Let $\mathcal{J} : Y \to Y$ be a unitary operator on the real Hilbert space Y with $\mathcal{J}^2 = -I$. Then:*

(i) *The real linear space Y becomes a symplectic space equipped with the symplectic form $\kappa(v, w) := \langle \mathcal{J}v|w\rangle_{\mathbb{R}}$ for all $v, w \in Y$.*

(ii) *The complex linear space $Y_{\mathbb{C}}$ becomes a complex Hilbert space equipped with the inner product $\langle v|w\rangle := \langle v|w\rangle_{\mathbb{R}} + \mathrm{i}\kappa(v, w)$ for all $v, w \in Y$.*

Proof. Ad (i). Since $\mathcal{J}$ is unitary, it follows from $\mathcal{J}^{-1} = \mathcal{J}^\dagger$ and $\mathcal{J}^2 = -I$ that $\mathcal{J}^\dagger = -\mathcal{J}$. Hence

$$\kappa(v, w) = \langle \mathcal{J}v|w\rangle_{\mathbb{R}} = -\langle v|\mathcal{J}w\rangle_{\mathbb{R}} = -\kappa(w, v).$$

Suppose that $\kappa(v_0, w) = 0$ for all $w \in Y$. Choose $w := \mathcal{J}v_0$. Then we obtain

$$\kappa(v_0, \mathcal{J}v_0) = \langle \mathcal{J}v_0|\mathcal{J}v_0\rangle_{\mathbb{R}} = \langle v_0|v_0\rangle_{\mathbb{R}} = 0.$$

Hence $v_0 = 0$.

Ad (ii). $\langle w|v\rangle = \langle w|v\rangle_{\mathbb{R}} + \mathrm{i}\kappa(w, v) = \langle v|w\rangle_{\mathbb{R}} - \mathrm{i}\kappa(v, w) = \langle v|w\rangle^\dagger$. Furthermore, let us show that $\langle v|\mathrm{i}w\rangle = \mathrm{i}\langle v|w\rangle$. In fact, $\langle v|\mathrm{i}w\rangle = \langle v|\mathcal{J}w\rangle$. This is equal to

$$\langle v|\mathcal{J}w\rangle_{\mathbb{R}} + \mathrm{i}\kappa(v, \mathcal{J}w) = \kappa(w, v) + \mathrm{i}\langle v|w\rangle_{\mathbb{R}} = \mathrm{i}(\langle v|w\rangle_{\mathbb{R}} + \mathrm{i}\kappa(v, w)) = \mathrm{i}\langle v|w\rangle.$$

Finally, $\langle v|v\rangle = \langle v|v\rangle_{\mathbb{R}}$. Therefore, $\langle v|v\rangle > 0$ if $v \neq 0$. □

5.8.3 The Refraction Index and Geodesics

The Poincaré model in non–Euclidean geometry is obtained by replacing straight lines in Euclidean geometry by light rays in the upper half-plane equipped with an appropriate refraction index.

- The light rays are called hyperbolic straight lines.
- The angles between the light rays are called hyperbolic angles.

In contrast to Euclidean geometry, the sum of angles of a hyperbolic triangle is not equal to π, but less than π. The Poincaré model represents an extremely elegant formulation of Lobachevsky's sophisticated work on non-Euclidean geometry. Formulated in the right mathematical language due to Élie Cartan (i.e., based on both the connection form and the curvature form), the Poincaré model is the *prototype* of both the Einstein theory of general relativity on gravitation and the Standard Model in elementary particle physics. In the following sections, we want to study the Poincaré model and its relations to numerous important subjects in modern mathematics and physics. In greater detail, we will investigate this in Vol. III on gauge theory.[24]

Regard the open upper half-plane $\mathcal{H}_{\mathbb{R}} := \{(x, y) \in \mathbb{R}^2 : \ y > 0\}$ as an optical medium with the refraction index $n(x, y)$ at the point (x, y). We assume that the function $(x, y) \mapsto n(x, y)$ is smooth and that $n(x, y) > 0$ for all $(x, y) \in \mathbb{R}$. We also define

$$\mathcal{H}_{\mathbb{C}} := \{z \in \mathbb{C} : \ \Im(z) > 0\}.$$

The map $(x, y) \mapsto x + \mathrm{i}y$ is a natural bijection from $\mathcal{H}_{\mathbb{R}}$ onto $\mathcal{H}_{\mathbb{C}}$. We will use the following notation:

- $C : x = x(\tau), y = y(\tau), \tau_0 \le \tau \le \tau_1$ (smooth curve),
- $Q(\tau) := (x(\tau), y(\tau), \dot{x}(\tau), \dot{y}(\tau))$,
- $\mathcal{L}(x, y, \dot{x}, \dot{y}) := n(x, y)^2(\dot{x}^2 + \dot{y}^2)$ (energetic Lagrangian),
- $L(x, y, \dot{x}, \dot{y}) := \sqrt{\mathcal{L}(x, y, \dot{x}, \dot{y})}$ (metric Lagrangian),
- $l(C) := \int_{\tau_0}^{\tau_1} L(Q(\tau))\, d\tau$ (length of the curve C),
- $E(C) := \int_{\tau_0}^{\tau_1} \mathcal{L}(Q(\tau))\, d\tau$ (geodesic energy of the curve C).

Here, $[\tau_0, \tau_1]$ is a compact parameter interval.

Fermat's principle of minimal arc length. The integral

$$T(C) := \int_{\tau_0}^{\tau_1} \frac{n(x(\tau), y(\tau))}{c} \sqrt{\dot{x}(\tau)^2 + \dot{y}(\tau)^2}\, d\tau$$

is equal to the time needed by a light ray to pass from the point $(x(\tau_0), y(\tau_0))$ to the point $(x(\tau_1), y(\tau_1))$ along the curve C. We define $l(C) := c \cdot T(C)$. Then, $l(C)$ has the physical dimension of length, and we call $l(C)$ the arc length of the curve C (with respect to the refraction index $n = n(x, y)$). In particular, the arc length between the points $(x(\tau_0), y(\tau_0))$ and $(x(\tau), y(\tau))$ is given by

[24] Fermat (1601–1665), Huygens (1629–1695), Euler (1707–1783), Lagrange (1736–1813), Gauss (1777–1855), Lobachevsky (1792–1856), Riemann (1826–1866), Beltrami (1835–1900), Lie (1842–1899), Klein (1849–1925), Poincaré (1854–1912), Élie Cartan (1859–1951), Hilbert (1862–1943), Minkowski (1864–1909), Einstein (1879–1955), Weyl (1885–1955), Hodge (1903–1975), de Rham (1903–1990), Kähler (1906–2000), Chern (1911–2004).

$$s(\tau) = \int_{\tau_0}^{\tau} n(x(\sigma), y(\sigma)) \sqrt{\dot{x}(\sigma)^2 + \dot{y}(\sigma)^2}\, d\sigma, \qquad \tau_0 \le \tau \le \tau_1.$$

Differentiating this with respect to the parameter τ, we get

$$\left(\frac{ds(\tau)}{d\tau}\right)^2 = n(x(\tau), y(\tau))^2 \left(\left(\frac{dx(\tau)}{d\tau}\right)^2 + \left(\frac{dy(\tau)}{d\tau}\right)^2\right).$$

Mnemonically, we write

$$\boxed{ds^2 = n(x,y)^2(dx^2 + dy^2).}$$

This differs from the Euclidean expression $ds^2 = dx^2 + dy^2$ by the conformal factor $n(x,y)^2$.

The Euler–Lagrange equations. Suppose that we are given the two points (x_0, y_0) and (x_1, y_1) on the upper half-plane $\mathcal{H}_{\mathbb{R}}$. Fermat's problem of least arc length reads as

$$l(C) = \min!, \qquad (x(\tau_k), y(\tau_k)) = (x_k, y_k), \quad k = 0, 1. \tag{5.50}$$

Using the notation introduced above, this means:

$$\boxed{\int_{\tau_0}^{\tau_1} L(Q(\tau))\, d\tau = \min!, \qquad (x(\tau_k), y(\tau_k)) = (x_k, y_k), \quad k = 0, 1.} \tag{5.51}$$

Here, the initial point (x_0, y_0) and the final point (x_1, y_1) are fixed. The smooth solutions of (5.51) satisfy the following Euler–Lagrange equations:

$$\frac{d}{d\tau} L_{\dot{x}}(Q(\tau)) = L_x(Q(\tau)), \quad \frac{d}{d\tau} L_{\dot{y}}(Q(\tau)) = L_y(Q(\tau)).$$

This follows as in the proof of Theorem 5.1 on page 269. Hence

$$\frac{d}{d\tau}\left(\frac{\mathcal{L}_{\dot{x}}}{\sqrt{\mathcal{L}}}\right) = \frac{\mathcal{L}_x}{\sqrt{\mathcal{L}}}, \qquad \frac{d}{d\tau}\left(\frac{\mathcal{L}_{\dot{y}}}{\sqrt{\mathcal{L}}}\right) = \frac{\mathcal{L}_y}{\sqrt{\mathcal{L}}}. \tag{5.52}$$

The failure of the Legendre transformation for the metric Lagrangian. We set

$$p := L_{\dot{x}}, \quad q := L_{\dot{y}}, \quad H := \dot{x}p + \dot{y}q - L.$$

Obviously,

$$\mathcal{L}_{\dot{x}}\dot{x} + \mathcal{L}_{\dot{y}}\dot{y} = 2\mathcal{L}. \tag{5.53}$$

This yields

$$H = \frac{\mathcal{L}_{\dot{x}}\dot{x} + \mathcal{L}_{\dot{y}}\dot{y} - 2\mathcal{L}}{2\sqrt{\mathcal{L}}} \equiv 0.$$

Thus, the Legendre transformation does not work. The reason for this degeneracy is the homogeneity of the Lagrangian (with respect to $\dot{x}$ and $\dot{y}$) which leads to the Euler identity (5.53). This homogeneity of the Lagrangian allows us the rescaling of the curve parameter without changing the integral $\int_{\tau_0}^{\tau_1} L d\tau$.

One possibility to overcome the trouble is to fix the parameter $\tau := x$. This was successfully done in Sect. 5.5. Now we want to use a general geometric method by choosing the curve parameter τ as arc length.

5.8.4 The Trick of Gauge Fixing

> Many variational problems in mathematics and physics possess additional degrees of freedom called gauge degrees of freedom (e.g., the choice of parametrization or the choice of potentials). This concerns problems in geometry (e.g., geodesics and minimal surfaces), in quantum electrodynamics in terms of the 4-potential, in the Standard Model of elementary particle physics, and in string theory based on the conformal invariance of the action functional.[25] The strategy for handling such problems is to fix a special gauge. However, the introduction of additional constraints has to be done with care. Moreover, one has to ensure that physical statements do not depend on the choice of the gauge fixing. In quantum field theory this is closely related to the Ward identities, the more general Slavnov–Taylor identities, and the cohomological BRST-symmetry theory due to Becchi, Rouet, Stora and Tyutin.
>
> Folklore

Replace the Euler–Lagrange equations (5.52) by the new system

$$\frac{d}{d\tau}\mathcal{L}_{\dot{x}}(Q(\tau)) = \mathcal{L}_x(Q(\tau)), \qquad \frac{d}{d\tau}\mathcal{L}_{\dot{y}}(Q(\tau)) = \mathcal{L}_y(Q(\tau)) \tag{5.54}$$

$$\mathcal{L}(Q(\tau)) = 1, \qquad \tau_0 \le \tau \le \tau_1. \tag{5.55}$$

Here, we add the constraint (5.55), which tells us that

$$\dot{s}(\tau) = \sqrt{n(x(\tau), y(\tau))^2(\dot{x}(\tau)^2 + \dot{y}(\tau)^2)} = \sqrt{\mathcal{L}(Q(\tau))} \equiv 1.$$

Consequently, the curve parameter τ equals arc length. Thus, we obtain the following equivalence of equations.

Proposition 5.8 *Each smooth solution of* (5.54), (5.55) *is a solution of the Euler–Lagrange equations* (5.52) *parameterized by arc length.*

Conversely, each solution of the Euler–Lagrange equations (5.52) *parameterized by arc length is a solution of* (5.54), (5.55).

The basic trick is that the equation (5.54) has the function $\mathcal{L}$ as conserved quantity, as we will show below via Legendre transformation. This means that for each solution of (5.54) with the initial condition $L(Q(\tau_0)) = 1$, we get $Q(\tau) \equiv 1$. That is, the constraint equation (5.55) is automatically satisfied.

5.8.5 Geodesic Flow

First let us investigate the so-called geodesic flow equations:

$$\frac{d}{d\tau}\mathcal{L}_{\dot{x}}(Q(\tau)) = \mathcal{L}_x(Q(\tau)), \qquad \frac{d}{d\tau}\mathcal{L}_{\dot{y}}(Q(\tau)) = \mathcal{L}_y(Q(\tau)). \tag{5.56}$$

The precise relation to Fermat's minimum principle will be studied below. This will motivate the following basic definition.

Geodesics. The solutions $x = x(\tau), y = y(\tau)$ of (5.56) in the open upper half-plane $\mathcal{H}_{\mathbb{R}}$ with the real curve parameter τ are called geodesics with respect to the metric $ds^2 = n(x, y)^2(dx^2 + dy^2)$.

[25] This will be studied in Vol. III.

Intuitively, geodesics are light rays in the optical medium $\mathcal{H}_{\mathbb{R}}$ with the refraction index $n(x,y)$ at the point (x,y).

Explicitly, the equation (5.56) for geodesics reads as[26]

$$\begin{aligned} \ddot{x} + \frac{n_x}{n}\,\dot{x}^2 + \frac{2n_y}{n}\,\dot{x}\dot{y} - \frac{n_x}{n}\,\dot{y}^2 = 0, \\ \ddot{y} - \frac{n_y}{n}\,\dot{x}^2 + \frac{2n_x}{n}\,\dot{x}\dot{y} + \frac{n_y}{n}\,\dot{y}^2 = 0. \end{aligned} \tag{5.57}$$

This is a second-order system of ordinary differential equations for the geodesics. We will show in Vol. III that Einstein's equation of motion for light rays (and celestial bodies) has the same structure. Setting $z(\tau) = x(\tau) + \mathrm{i}y(\tau)$ and noting that $\partial_z n := \frac{1}{2}(n_x - \mathrm{i}n_y)$, the equation for the geodesics can elegantly be written as

$$\boxed{\dot{w} = -\frac{2\partial_z n(z)}{n(z)}\,\dot{z}^2, \quad \dot{w} = z.} \tag{5.58}$$

This is the equation for a dynamical system called the geodesic flow. Explicitly, one has to replace $\dot{z}$ and $n(z)$ by $\dot{z}(\tau)$ and $n(z(\tau))$, respectively, and so on.

5.8.6 Hamilton's Duality Trick and Cogeodesic Flow

> Mathematicians and physicists enjoy Hamiltonian systems, since such dynamical systems allow the application of methods from symplectic geometry. In particular, the Hamiltonian function of a Hamiltonian system is always a conserved quantity.
>
> Folklore

The cogeodesic flow equations read as follows:

$$\dot{p} = -\mathcal{H}_x, \;\; \dot{q} = -\mathcal{H}_y, \;\; \dot{x} = \mathcal{H}_p, \;\; \dot{y} = H_q. \tag{5.59}$$

Here, we use the Hamiltonian function

$$\mathcal{H}(x,y,p,q) := \frac{p^2}{4n(x,y)^2} + \frac{q^2}{4n(x,y)^2},$$

which is dual to the energetic Lagrangian $\mathcal{L}(x,y,\dot{x},\dot{y}) = n(x,y)^2(\dot{x}+\dot{y}^2)$. We are looking for smooth trajectories $\tau \mapsto (x(\tau), y(\tau), p(\tau), q(\tau))$ which satisfy equation (5.59). Setting $z = x + \mathrm{i}y$ and $u = p + \mathrm{i}q$, the cogeodesic flow equation can be written as

$$\boxed{\dot{u} = -\partial_{\bar{z}}\mathcal{H}, \qquad \dot{z} = \partial_{\bar{u}}\mathcal{H}.} \tag{5.60}$$

---

[26] Note that $\mathcal{L} = n^2(\dot{x}^2 + \dot{y}^2)$, and hence $\mathcal{L}_{\dot{x}} = 2n^2\dot{x}, \mathcal{L}_x = 2nn_x(\dot{x}^2+\dot{y}^2)$. This implies

$$\frac{d}{d\tau}\mathcal{L}_{\dot{x}} = (4nn_x\dot{x} + 4nn_y\dot{y})\dot{x} + 2n^2\ddot{x}.$$

Theorem 5.9 (i) *The geodesic flow equation* (5.56) *is equivalent to the cogeodesic flow equation* (5.59) *via Legendre transformation with respect to the energetic Lagrangian* $\mathcal{L}$.

(ii) *The solutions of* (5.59) *are constant along the Hamiltonian function* $\mathcal{H}$ *(i.e.,* $\mathcal{H}$ *is a conserved quantity).*

(iii) *The energetic Lagrangian* $\mathcal{L}$ *is a conserved quantity of the geodesic flow equation* (5.56).

(iv) *Each solution of the geodesic flow equation* (5.56) *which satisfies the initial condition* $\mathcal{L}(Q(\tau_0)) = 1$ *is a solution of the Euler–Lagrange equations* (5.52) *parameterized by arc length.*

Proof. Ad (i). The Legendre transformation with respect to the energetic Lagrangian $\mathcal{L}$ is given by

$$p := \mathcal{L}_{\dot{x}}, \quad q := \mathcal{L}_{\dot{y}}, \quad \mathcal{H} = p\dot{x} + \dot{y}q - \mathcal{L}.$$

This way, the Euler–Lagrange equations (5.56) with respect to $\mathcal{L}$ pass over to the Hamiltonian canonical equations (5.59) with respect to $\mathcal{H}$. Explicitly,

$$p = 2n(x,y)^2\dot{x}, \quad q = 2n(x,y)^2\dot{y}, \quad \mathcal{H} = n(x,y)^2(\dot{x}^2 + \dot{y}^2).$$

Ad (ii). For a solution of (5.59), we get

$$\begin{aligned}\frac{d}{d\tau}\mathcal{H}(x(\tau), y(\tau), p(\tau), q(\tau)) &= \mathcal{H}_x\dot{x} + \mathcal{H}_y\dot{y} + \mathcal{H}_p\dot{p} + \mathcal{H}_q\dot{q}\\ &= -\dot{p}\dot{x} - \dot{q}\dot{y} + \dot{x}\dot{p} + \dot{y}\dot{q} \equiv 0.\end{aligned}$$

Ad (iii). This follows from (ii) by Legendre transformation.

Ad (iv). By (iii), the solutions of (5.59) satisfy (5.54), (5.55), and hence they satisfy (5.52). □

5.8.7 The Principle of Minimal Geodesic Energy

Parallel to the Fermat problem (5.51) of least arc length, we formulate the problem of least geodesic energy:

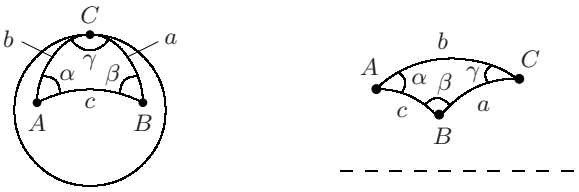
$$\boxed{\int_{\tau_0}^{\tau} \mathcal{L}(Q(\tau)) = \min!, \qquad (x(\tau_k), y(\tau_k)) = (x_k, y_k), \quad k = 0, 1.} \tag{5.61}$$

Here, the initial point (x_0, y_0) and the final point (x_1, y_1) are fixed. Observe that the metric Lagrangian $L = \sqrt{\mathcal{L}}$ has been replaced by the energetic Lagrangian $\mathcal{L}$. There holds the following equivalence principle.

Theorem 5.10 (i) *The smooth solutions of the energetic minimum problem* (5.61) *satisfy the geodesic flow equations* (5.56).

(ii) *Each smooth solution of the energetic minimum problem* (5.61) *is also a solution of the Fermat minimum problem* (5.51).

(iii) *Conversely, each solution of the Fermat minimum problem* (5.51) *parameterized by arc length is a solution of the energetic minimum problem* (5.61).



(a) spherical (b) hyperbolic

Fig. 5.14. Geodesic triangles

Proof. Ad (i). Equation (5.56) is the Euler–Lagrange equation to (5.61).

Ad (ii). By the Schwarz inequality,

$$\int_{\tau_0}^{\tau_1} \sqrt{\mathcal{L}(Q(\tau))}\, d\tau \le (\tau_1 - \tau_0)^{1/2} \left(\int_{\tau_0}^{\tau_1} \mathcal{L}(Q(\tau))\, d\tau \right)^{1/2}.$$

Equality holds iff $\mathcal{L}(Q(\tau)) \equiv \text{const}$, by Problem 5.4. If $\tau \mapsto (x(\tau), y(\tau))$ is a smooth solution of (5.61), then it is parameterized by arc length. Hence $\mathcal{L}(Q(\tau)) \equiv 1$.

Ad (iii). Argue as in (ii). □

5.9 Spherical Geometry

Spherical geometry is the geometry on the surface of earth.
Folklore

Geodesics. Consider the 2-dimensional sphere

$$\mathbb{S}_r^2 := \{(x, y, \zeta) \in \mathbb{R}^3 : \; x^2 + y^2 + \zeta^2 = r^2\}$$

of radius r. The equator and the meridian circles passing through both North Pole and South Pole are called geodesic lines of the sphere. More precisely, a curve on the sphere is called a geodesic line (or a spherical straight line) iff it is the intersection between the sphere $\mathbb{S}_r^2$ and a plane which passes through the origin $(0, 0, 0)$. Connected subarcs of geodesic lines are called geodesics.[27] If a geodesic connects the two points A and B on the sphere $\mathbb{S}_r^2$ and the length of the geodesic is less or equal to πr, then it is a curve of shortest length. If an airplane wants to fly from point A to point B by using a route of shortest distance, then it has to fly along a geodesic of shortest length (see Fig. 5.9 on page 277).

Geodesic triangles. Consider a triangle ABC as pictured in Fig. 5.14(a). The sides of the triangle are geodesics of shortest length. For the sum of the angles α, β, γ, we have

$$\boxed{\alpha + \beta + \gamma = \pi + K\mathsf{A}.} \tag{5.62}$$

Here, $K = 1/r^2$ is the Gaussian curvature of the sphere $\mathbb{S}_r^2$, and A is the spherical area of the triangle. In contrast to this, for a triangle in Euclidean geometry, we

[27] The term *geodesic* was introduced by Liouville (1809–1882) in 1850 and was taken from geodesy.

have $\alpha+\beta+\gamma=\pi$. Formula (5.62) allows us to measure the radius of earth, by measuring the angles and the surface area of a triangle on earth.

We will show below that equation (5.62) remains valid in hyperbolic geometry, if we set $K=-1$ and A denotes the hyperbolic area of the hyperbolic triangle (Fig 5.14(b)). Formally, this is obtained by starting with $K=1/r^2$. Now replace the radius r of the sphere by the imaginary radius $\mathrm{i}r$. Finally, set $r=1$. This way, we get $K=-1$.

The Pythagorean theorem. For the spherical geodesic triangle ABC with geodesic sides of shortest length, and angles α,β,γ together with the special choice $\gamma:=\frac{\pi}{2}$ (Fig. 5.14), we get the spherical Pythagorean theorem

$$\cos\left(\frac{a}{r}\right)=\cos\left(\frac{b}{r}\right)\cos\left(\frac{c}{r}\right). \tag{5.63}$$

Here, a,b,c is the length of the sides BC,CA,AB, respectively.

Now consider the geodesic triangle ABC in hyperbolic geometry pictured in Fig. 5.14(b). As we will show below, in hyperbolic geometry, geodesics are semicircles centered at the x-axis. Formally, replacing r by $\mathrm{i}r$ and setting $r=1$ and noting that $\cos(\mathrm{i}\varphi)=\cosh\varphi$, we get the hyperbolic Pythagorean theorem

$$\cosh a=\cosh b\cosh c.$$

This formula can be rigorously justified.

Finally, let us study the Euclidean limit case $r\to+\infty$. Since we have the approximation formula $\cos\varphi=1-\frac{1}{2}\varphi^2+O(\varphi^4)$ as $\varphi\to+\infty$, it follows from (5.63) that

$$a^2+b^2=c^2+O\left(\frac{1}{r^2}\right),\qquad r\to+\infty.$$

Letting $r\to+\infty$, we get

$$a^2+b^2=c^2.$$

This is the classical Pythagorean theorem in Euclidean geometry.

5.9.1 The Global Gauss–Bonnet Theorem

There exists a deep relation between curvature and topology. The prototype is the famous Gauss–Bonnet formula

$$\boxed{\int_{\mathbb{S}^2_r}\frac{K}{2\pi}\,d\mathsf{A}=\chi(\mathbb{S}^2_r),} \tag{5.64}$$

where $\chi(\mathbb{S}^2_r)=2$ is the Euler characteristic of the sphere $\mathbb{S}^2_r$. In order to discuss this formula, consider first the triangulation of the unit sphere (i.e., $r=1$) pictured in Fig. 5.15(a). Here, we have four geodesic triangles. For each triangle, we get

$$\alpha+\beta+\gamma=\pi+\mathsf{A} \tag{5.65}$$

with $\alpha=\beta=\frac{\pi}{2},\gamma=\pi$, and $\mathsf{A}=\frac{1}{4}\,\mathrm{meas}(\mathbb{S}^2)=\pi$. The Euler characteristic of this triangulation is given by the following Euler polyhedron formula

$$\boxed{\chi=T-S+V.} \tag{5.66}$$

Here, T (resp. S and V) is the number of triangles (resp. sides of triangles and vertices). Explicitly, $T=4,S=6,V=4$. Hence $\chi=2$. The Euler characteristic χ

(a) $T - S + V = 4 - 6 + 4 = 2$ (b) $T - S + V = 8 - 12 + 6 = 2$

Fig. 5.15. The global Gauss–Bonnet theorem

does not depend on the choice of the triangulation. Consider, for example, the triangulation pictured in Fig. 5.15(b). Then we have the numbers $T = 8, S = 12, V = 6$. Again $T - S + V = 2$. The point is that the most important topological invariant – the Euler characteristic – can be obtained in terms of analysis by integrating over a density $\varrho(x, y) := \frac{K(x,y)}{2\pi}$. The integral (5.64) is called a topological charge in physics.

The Gauss–Bonnet formula relates analysis and differential geometry (*i.e., curvature*) *to topology.*

For a long time, it was a famous open problem to find the generalization of the Gauss–Bonnet theorem to higher dimension. This problem was solved by Chern (1911–2004) in 1945.[28] Motivated by this problem, Chern created the theory of characteristic classes. Let us explain this for the sphere. If we use the volume form $\upsilon(\mathbb{S}^2_r)$ with

$$\int_{\mathbb{S}^2_r} \upsilon(\mathbb{S}^2_r) = \mathrm{meas}(\mathbb{S}^2_r) = 4\pi r^2,$$

then the form $\gamma := \frac{K}{2\pi}\upsilon$ is called the Chern form, and we have

$$\int_{\mathbb{S}^r_2} \gamma = \chi(\mathbb{S}^2_r).$$

If μ is a 1-form on $\mathbb{S}^2_r$, then

$$\int_{\mathbb{S}^2_r} d\mu = 0,$$

by the Poincaré–Stokes integral theorem for differential forms. Hence

$$\boxed{\int_{\mathbb{S}^2_r} \gamma + d\mu = \chi(\mathbb{S}^2_r).} \tag{5.67}$$

This shows the relation to de Rham cohomology, as we will discuss next.

[28] S. Chern, A simple intrinsic proof of the Gauss–Bonnet formula for closed Riemannian manifolds, Ann. of Math. **45** (1945), 747–752.
S. Chern, Characteristic classes of Hermitean manifolds, Ann. Math. **47** (1946), 85–121.

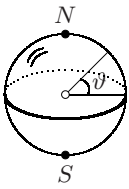


Fig. 5.16. Geographic latitude ϑ

5.9.2 De Rham Cohomology and the Chern Class of the Sphere

> Cohomology is deeply routed in the following topics: Gauss' surface theory, the Kirchhoff–Weyl theory of electrical networks, and Maxwell's theory of electromagnetism. Cohomology lies at the heart of both modern differential topology and modern quantum field theory (the BRST approach).[29]
>
> Folklore

Recall de Rham cohomology from Sect. 16.8.2 of Vol. I. Two smooth differential 2-forms ω and σ on $\mathbb{S}^2_r$ are called equivalent, $\omega \sim \sigma$, iff there exists a smooth 1-form on $\mathbb{S}^2_r$ such that

$$\omega = \sigma + d\mu \quad \text{on } \mathbb{S}^2_r.$$

The equivalence class $[\omega]$ corresponding to ω is called the de Rham cohomology class of ω. All the de Rham cohomology classes of 2-forms on $\mathbb{S}^2_r$ form a real 1-dimensional linear space called the second cohomology group $H^2(\mathbb{S}^2_r)$ of the sphere (in the sense of de Rham). By (5.67), the Euler characteristic of $\mathbb{S}^2_r$ only depends on the cohomology class $[\gamma]$ of the Chern form γ. This cohomology class is called the first Chern class of the sphere $\mathbb{S}^2_r$. We write $c_1(\mathbb{S}^2_r) = [\gamma]$. The Chern class $[\gamma]$ is a generator of the second cohomology group $H^2(\mathbb{S}^2_r)$ of the sphere.

Representing de Rham cohomology classes by harmonic forms (Hodge theory). Riemann based his approach to complex analysis on the Laplacian. In the 1930s, Hodge generalized Riemann's theory in a substantial way. We will study this in Vol. IV on quantum mathematics. At this point, we would like to sketch the following. To begin with, let us introduce spherical coordinates, namely,

- geographic length $\varphi : -\pi < \varphi \leq \pi$, and
- geographic latitude ϑ: $-\frac{\pi}{2} \leq \vartheta \leq \frac{\pi}{2}$.

In particular, the points on the equator, the North Pole, the South Pole have the geographic latitude $\vartheta = 0, \frac{\pi}{2}, -\frac{\pi}{2}$, respectively (Fig. 5.16).[30] Mnemonically, the Riemannian metric on $\mathbb{S}^2_r$ is given by

$$ds^2 = r^2 \cos^2 \vartheta \cdot d\varphi^2 + r^2 d\vartheta^2,$$

and the volume form reads as $\upsilon = r^2 \cos^2 \vartheta \; d\varphi \wedge d\vartheta$. Every smooth curve on the sphere, $C : \varphi = \varphi(\tau), \vartheta = \vartheta(\tau), \tau_0 \leq \tau \leq \tau_1$, has the length

$$l(C) = \int_{\tau_0}^{\tau_1} \dot{s}(\tau) \, d\tau = \int_{\tau_0}^{\tau_1} \sqrt{r^2 \cos^2 \vartheta(\tau) \cdot \dot{\varphi}(\tau)^2 + r^2 \dot{\vartheta}(\tau)^2} \, d\tau.$$

[29] Both the classical roots and the modern extensions will be thoroughly studied in Vol. III on gauge theory and in Vol. IV on quantum mathematics.

[30] To avoid technicalities, we exclude the North Pole and the South Pole. The full approach has to be based on this coordinate system and two local coordinate systems at both the north and the south pole.

Every open subset $\mathcal{U}$ of the sphere $\mathbb{S}_r^2$ has the surface area

$$\text{meas}(\mathcal{U}) = \int_{\mathcal{U}} \upsilon = \int_{\mathcal{U}} r^2 \cos\vartheta \; d\varphi d\vartheta.$$

Let $C^\infty(\mathbb{S}_r^2)$ denote the set of all smooth functions $f : \mathbb{S}_r^2 \to \mathbb{R}$. This is a real linear space. For $f, g \in \mathbb{C}^\infty(\mathbb{S}_r^2)$, the Cartan derivative works as follows:

- $df = f_\varphi d\varphi + f_\vartheta d\vartheta$,
- $d(f d\varphi + g d\vartheta) = df \wedge d\varphi + dg \wedge d\vartheta$,
- $d(f d\varphi \wedge d\vartheta) = 0$.

The Hodge star operator is defined as follows:

- $*1 := \upsilon$, and $*\upsilon = 1$,
- $*f = f\upsilon$, and $*(f\upsilon) = f$,
- $*\big(f\cos\vartheta \; d\varphi + g \; d\vartheta\big) = -g\cos\vartheta \; d\varphi + f d\vartheta$.

Let $\Omega^p(\mathbb{S}_r^2)$ denote the set of all p-forms on the sphere $\mathbb{S}_r^2$ with smooth coefficients. Introducing the inner product

$$\boxed{\langle \omega | \mu \rangle := \int_{\mathbb{S}_r^2} \omega \wedge *\mu, \qquad \text{for all} \quad \omega, \mu \in \Omega^p(\mathbb{S}_r^2),}$$

the space $\Omega^p(\mathbb{S}_r^2)$ becomes a real pre-Hilbert space for $p = 0, 1, 2$. In particular, for $p = 0$, we have $\Omega^0(\mathbb{S}_r^2) = C^\infty(\mathbb{S}_r^2)$, and

$$\langle f | g \rangle = \int_{\mathbb{S}_r^2} f \wedge *g = \int_{\mathbb{S}_r^2} fg \cdot \upsilon = \int_{\mathbb{S}_r^2} fg \cdot r^2 \cos\vartheta \; d\varphi d\vartheta.$$

This is the usual inner product on the sphere $\mathbb{S}_r^2$ with respect to the surface measure. We define the operator

$$d^* := - * d * .$$

Then, $\langle \omega | d\mu \rangle = \langle d^* \omega | \mu \rangle$ for all $\omega, \mu \in \Omega^p(\mathbb{S}_r^2)$ with $p = 0, 1, 2$. That is, the operator d^* is the adjoint operator to the Cartan operator d. Now to the point. The operator

$$\boxed{\Delta := dd^* + d^* d} \tag{5.68}$$

is called the Laplacian (or the Laplace–Beltrami operator) of the sphere $\mathbb{S}_r^2$. For example, if $f \in C^\infty(\mathbb{S}_r^2)$, then

$$\Delta f = -\frac{f_{\varphi\varphi}}{r^2 \cos^2\vartheta} - \frac{f_{\vartheta\vartheta}}{r^2} - \frac{\tan\vartheta \cdot f_\vartheta}{r^2}.$$

This is the usual Laplacian written in spherical coordinates with constant radius r. Let $\omega \in \Omega^p(\mathbb{S}_r^2)$ for fixed $p = 0, 1, 2$. Then the following two statements are equivalent:

(i) $\Delta\omega = 0$ on $\mathbb{S}_r^2$ (Laplace equation on the sphere).
(ii) $d\omega = 0$ and $d^*\omega = 0$ on $\mathbb{S}_r^2$ (Yang–Mills equation on the sphere).

The p-form ω is called harmonic iff (i) is satisfied. All the possible harmonic forms on the sphere are given as follows:

- harmonic 0-forms: $\omega \equiv a$ with $a \in \mathbb{R}$,
- harmonic 1-forms: $\omega \equiv 0$,
- harmonic 2-forms: $\omega = a(*1)$ with $a \in \mathbb{R}$.

The main result of Hodge theory is the following one:[31]

> *For a compact Riemannian manifold $\mathcal{M}$, every de Rham cohomology class contains precisely one harmonic form.*

Consequently, the de Rham cohomology class $H^q(\mathcal{M})$ is isomorphic to the real linear space of all harmonic q-forms. The dimension

$$\beta_q := \dim H^q(\mathcal{M}), \qquad q = 0, 1, \ldots, n$$

is called the q-th *Betti number* of the n-dimensional manifold $\mathcal{M}$. Moreover, the alternating sum

$$\boxed{\chi(\mathcal{M}) = \beta_0 - \beta_1 + \beta_2 - \ldots + (-1)^n \beta_n}$$

is called the *Euler characteristic* of $\mathcal{M}$. Note the following:

- The de Rham cohomology works for all compact manifolds. We do not need any additional mathematical structure.
- In contrast to this, Hodge theory can only be applied to compact Riemannian manifolds. We need the additional metric structure in order to introduce the Hodge star operator, and hence the Laplacian.

In particular, for the sphere, we get $H^0(\mathbb{S}^2_r) = H^2(\mathbb{S}^2_r) = \mathbb{R}$ and $H^1(\mathbb{S}^2_r) = 0$. Hence

$$\beta_0 = \beta_2 = 1, \quad \beta_1 = 0.$$

The relation $\beta_0 = \beta_2$ is called the Poincaré duality of the sphere. For the Euler characteristic of the sphere, we get

$$\chi(\mathbb{S}^2_r) = \beta_0 - \beta_1 + \beta_2 = 1 - 1 + 1 = 2.$$

In terms of topology, the following holds:

- $\beta_0 = 1$ tells us that the sphere $\mathbb{S}^2_r$ is path-connected, that is, two points on the sphere can always be connected by a continuous path.
- $\beta_1 = 0$ tells us that the sphere is simply connected. That is, every closed continuous curve on the sphere can be continuously contracted into a point.
- $\beta_2 = 1$ tells us that the sphere is orientable.

Note the following:

- Cohomology: The definition of the Euler characteristic above in terms of differential forms, displays the analytic character of this quantity. This corresponds to cohomology in topology.
- Homology: The Euler characteristic can also be characterized by the Euler polyhedron formula (5.66) in terms of triangulations of the basic manifolds. The Euler polyhedron formula is the starting point for homology in topology.

Surprisingly enough, the two completely different approaches yield the same quantity. This is the deep duality between homology and cohomology. In terms of physics, this means that there exists a deep relation between the geometric structure of the space-time manifold (homology) and the physical fields that live on the space-time manifold (cohomology).

[31] For the proof, see J. Jost, Riemannian Geometry and Geometric Analysis, Sect. 2.2, 5th edition, Springer, Berlin, 2008.

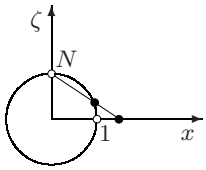


Fig. 5.17. Stereographic projection

5.9.3 The Beltrami Model

Beltrami's model of elliptic geometry. Choose the radius $r = 1$ of the sphere. The stereographic projection pictured in Fig. 5.17 maps the unit sphere (minus North Pole) $\mathbb{S}^2 \setminus \{N\}$ onto the complex plane $\mathbb{C}$ in a bijective manner. Explicitly, the point (ξ, η, ζ) on the unit sphere corresponds to the point $z = x + \mathrm{i}y$ in the complex plane $\mathbb{C}$ given by

$$z = \frac{\xi + \mathrm{i}\eta}{1 - \zeta}.$$

In particular, the South Pole goes to the origin $z = 0$, and the equator is mapped onto the unit circle. For a given smooth curve on the unit sphere,

$$\xi = \xi(\tau), \quad \eta = \eta(\tau), \quad \zeta = \zeta(\tau), \qquad \tau_0 \le \tau \le \tau_1,$$

we get the following transformation formulas:

$$\dot{\xi}(\tau)^2 + \dot{\eta}(\tau)^2 + \dot{\zeta}(\tau)^2 = \frac{\dot{x}(\tau)^2 + \dot{y}(\tau)^2}{(x(\tau)^2 + y(\tau)^2 + 1)^2}.$$

This allows us to transplant the metric from the unit sphere to the complex plane $\mathbb{C}$. Explicitly,

$$ds^2 = \frac{dx^2 + dy^2}{1 + x^2 + y^2}. \tag{5.69}$$

By definition, the elliptic Beltrami model $\mathcal{B}_{\mathrm{ellip}}$ consists of the open unit disc

$$\mathcal{B} := \{(x, y) \in \mathbb{R}^2 : \ x^2 + y^2 < 1\}$$

equipped with the Riemannian metric (5.69).

Beltrami's hyperbolic model. Using the rescaling $x \mapsto rx$ and $y \mapsto ry$ and setting $r := \mathrm{i}$, the Riemannian metric (5.69) passes over to the metric

$$ds^2 = \frac{dx^2 + dy^2}{1 - x^2 - y^2} \tag{5.70}$$

on the open unit disc $\mathcal{B}$. This yields the hyperbolic Beltrami model $\mathcal{B}_{\mathrm{hyp}}$. Published in 1868, this was the first model in the history of mathematics which globally realized Lobachevsky's hyperbolic geometry on a subset of the Euclidean plane equipped with a non-Euclidean Riemannian metric. We will show in Problem 5.6 that this model is equivalent to the 1881 Poincaré model on the open upper half-plane to be considered in the next section. Next let us discuss the different local and global properties of the hyperbolic Beltrami model.

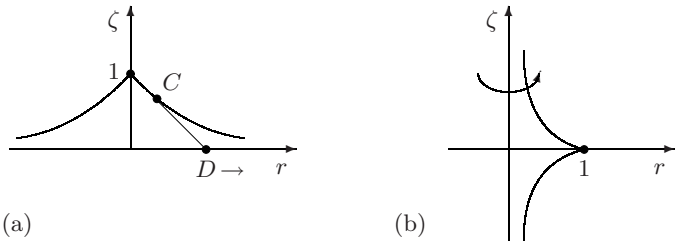


Fig. 5.18. Pseudosphere

The pseudosphere of Gauss. Consider the curve $r = r(\zeta)$ pictured in Fig. 5.18(a). This so-called tractrix was first studied by Leibniz (1646–1716). The tractrix has the following characteristic property: If we consider the tangent to the curve through a point C, which intersects the ζ-axis at the point D, then the length of the segment CD is constant. The name tractrix comes from the fact that the curve arises when a dog at the point D pulls a cart at the point C in direction of the ζ-axis. This is a situation which earlier often actually occurred in mining.[32] The pseudosphere $\mathbb{S}_{\text{pseudo}}$ is obtained through rotation of the tractrix about the ζ-axis in a Cartesian (ξ, η, ζ)-coordinate system (Fig. 5.18(b)). Introducing polar coordinates φ, r we obtain

$$\xi = r\cos\varphi, \quad \eta = r\sin\varphi, \quad \zeta = \pm\int_1^r \frac{\sqrt{1-\varrho^2}}{\varrho}\, d\varrho.$$

Naturally, the curves $\varphi = \text{const}$ and $r = \text{const}$ are called meridians and circles of latitude. The element of arc is

$$ds^2 = d\xi^2 + d\eta^2 + d\zeta^2 = r^2 d\varphi^2 + \frac{dr^2}{r^2}.$$

Letting $x = \varphi, y = 1/r$ with $-\pi < \varphi < \pi, 0 < r < 1$ we obtain

$$ds^2 = \frac{dx^2 + dy^2}{y^2}, \qquad -\pi < x < \pi, \;\; y > 1.$$

The Gaussian curvature. Using local real coordinates u, v, Gauss describes the local metric properties of a smooth surface by the first fundamental form:

$$ds^2 = E(u,v)du^2 + 2F(u,v)dudv + G(u,v)dv^2.$$

Mnemonically, this means that the curve length of a given smooth curve on the surface, $C : u = u(\tau), v = v(\tau), \tau_0 \le \tau \le \tau_1$, is equal to

$$l(C) = \int_{\tau_0}^{\tau_1} \dot{s}(\tau)d\tau$$

with $\dot{s}^2 = E(P)\dot{u}^2 + 2F(P)\dot{u}\dot{v} + G(P)\dot{v}^2\, d\tau$, and $\dot{s} \ge 0$. Here, $P, \dot{s}, \dot{u}, \dot{v}$ depend on the curve parameter τ. To introduce the Gaussian curvature $K(P)$ of the surface at the point P, we choose a Cartesian (x, y, ζ)-coordinate system at the point, where the (x, y)-plane is the tangent plane of the surface at the point P (Fig. 5.19).

[32] The Latin word *trahere* means *to pull*.

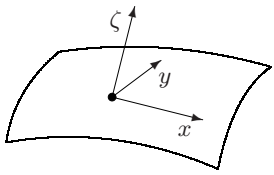


Fig. 5.19. Local coordinate system

Using Taylor expansion, the local equation of the surface near the point P with the local coordinates $(0,0)$ looks like

$$\zeta(x,y) = \tfrac{1}{2}(ax^2 + by^2) + \dots$$

The dots stand for terms of higher order (i.e., $o(x^2+y^2)$ as $x^2+y^2 \to 0$). The principal axis theorem for quadratic forms tells us that this normal form is always possible, after a suitable rotation of the $(x,y,\zeta$-coordinate system about the ζ-axis. Now we define the Gaussian curvature $K(P)$ of the surface at the point P by setting

$$\boxed{K(P) := ab.} \tag{5.71}$$

This means that curvature depends on the local quadratic terms of the Taylor expansion. In particular, for a sphere of radius r, say, near the south pole, we get

$$\zeta = r\left(1 - \sqrt{1 - \frac{x^2+y^2}{r^2}}\right) = \frac{1}{2}\cdot\frac{x^2+y^2}{r} + \dots$$

Thus, $a = b = \frac{1}{r}$ and hence $K = 1/r^2$.

The theorema egregium of Gauss – a gem of mathematics and physics. The Gaussian curvature is defined by using the surrounding space. Gauss posed the following fundamental question:

> *Is it possible to compute the Gaussian curvature without using the surrounding space, but only by using the functions E, F, G of the intrinsic metric?*

After tedious computations, Gauss found the answer. The theorema egregium tells us that the Gaussian curvature K only depends on the metric functions E, F, G and their partial derivatives up to second order. Explicitly, there holds the following universal formula:[33]

$$K = \frac{1}{D}\begin{vmatrix} \left(-\frac{1}{2}G_{uu} + F_{uv} - \frac{1}{2}E_{vv}\right) & \frac{1}{2}E_u & \left(F_u - \frac{1}{2}E_v\right) \\ F_v - \frac{1}{2}G_u & E & F \\ \frac{1}{2}G_v & F & G \end{vmatrix} - \frac{1}{D}\begin{vmatrix} 0 & \frac{1}{2}E_v & \frac{1}{2}G_u \\ \frac{1}{2}E_v & E & F \\ \frac{1}{2}G_u & F & G \end{vmatrix}. \tag{5.72}$$

On February 19, 1826 Gauss wrote to Olbers:

[33] Here, we set $D := (EG - F^2)^2$. The proof of this formula can be found in J. Stoker, Differential Geometry, Sect. VI.7, Wiley, New York, 1989.

> I hardly know any period of my life, where I earned so little real gain for truly exhausting work, as during this winter. I found many, many beautiful things, but my work on other things has been unsuccessful for months.

Finally, on October 8, 1827, Gauss presented the general surface theory. The title of his paper was *Disquisitiones circa superficies curvas* (Investigations about curved surfaces) (in Latin). The most important result of this masterpiece of mathematics is the theorema egregium. Gauss proved that

> *Curvature is an intrinsic property of a 2-dimensional manifold which is equipped with the notion of arc length.*

In particular, for the pseudosphere with pseudoradius $r = 1$, the theorema egregium yields $K \equiv -1$ (outside the singular equator). Riemann generalized Gauss' work to higher dimensions. Erich Worbs writes:[34]

> In 1854 Bernhard Riemann (1826–1866) presented three topics for his inaugural lecture in Göttingen. Gauss (1777–1855), in recollection of his own struggle with Euclid's parallel axiom, chose – in breaking with tradition – the third one: *On the hypotheses which lie at the foundation of geometry.* In his lecture Riemann presented the fundamentals of a geometry for the n-dimensional curved metric space (nowadays called Riemannian geometry). This must have made an extremely deep impression on Gauss, who at that time was already very weak. Later, on his way home, he spoke with unusual excitement to Wilhelm Weber about the depth of the presentation.

Riemann considered all the 2-dimensional submanifolds of an n-dimensional Riemannian manifold and computed their Gauss curvature. This family of Gauss curvatures forms the Riemann curvature tensor. Einstein used the Riemann curvature tensor in order to intrinsically describe gravitation as the curvature of the 4-dimensional space-time manifold. In the Standard Model of elementary particle physics, strong interaction and electroweak interaction are intrinsically represented by the curvature of a principal fiber bundle with the structure group $U(1) \times SU(2) \times SU(3)$ (see Vol. III). Here, the structure group describes the possible gauge transformations.

Hilbert's no-go theorem for the embedding of the global hyperbolic Beltrami manifold into the 3-dimensional Euclidean space. Hilbert proved in 1901 that the hyperbolic Beltrami manifold $\mathcal{B}_{\text{hyp}}$ (and hence also the hyperbolic Poincaré plane $\mathcal{H}_{\mathbb{R}}$) cannot be isometrically embedded into the space $\mathbb{R}^3$. The precise definition of an isometric embedding will be given on page 312. Intuitively, Hilbert showed that the hyperbolic Beltrami model cannot be realized on a submanifold of $\mathbb{R}^3$ in such a way that the arc length of curves is preserved.[35] Note that the pseudosphere is not a submanifold of $\mathbb{R}^3$, since there are no tangent planes at the points of the equator. The hyperbolic Beltrami model can be locally realized on the pseudosphere, but not globally. For a long time, it was an open problem whether Riemannian manifolds can be isometrically embedded into a Euclidean space $\mathbb{R}^m$ of sufficiently high dimension m. This problem was positively solved by John Nash in 1956.

The Nash embedding theorem and the sophisticated regularization method (hard implicit function theorem). Let $n = 1, 2, \ldots$ There are the following two crucial embedding theorems:

[34] E. Worbs, Carl Friedrich Gauß: Ein Lebensbild (A Biography), Koehler & Amelang, Leipzig (in German).

[35] The proof can be found in S. Novikov and T. Taimanov, Geometric Structures and Fields, Sect. 4.3.4, Amer. Math. Soc., Providence, Rhode Island, 2006.

(a) regular curve (b) singular curve

Fig. 5.20. Curves

(W) *Every paracompact real n-dimensional manifold can be embedded into the Euclidean space* $\mathbb{R}^{2n+1}$ (Whitney's embedding theorem from 1936).[36]

(N) *Every paracompact real n-dimensional Riemannian manifold can be isometrically embedded into some Euclidean space* $\mathbb{R}^m$*, where* $m = O(n^3)$ *as* $n \to \infty$ (Nash's embedding theorem from 1956).

Here, we use the following terminology. In this series of monographs, manifolds are always assumed to be smooth (see Sect. 5.4 of Vol. I). Moreover, a manifold is called paracompact iff it possesses a finite or at most countable system of local coordinate systems (i.e., charts). Compact manifolds are paracompact. But, the converse is not true. For example, the real line $\mathbb{R}$ (or the space $\mathbb{R}^m$ with $m \geq 1$) is paracompact, but not compact. The map $\chi : \mathcal{M} \to \mathcal{N}$ from the manifold $\mathcal{M}$ into the manifold $\mathcal{N}$ is called an *embedding* if and only if

- the image $\chi(\mathcal{M})$ is a submanifold of $\mathcal{N}$, and
- the map $\chi : \mathcal{M} \to \chi(\mathcal{M})$ is a diffeomorphism.

We say that the manifold $\mathcal{M}$ can be embedded into the manifold $\mathcal{N}$ iff there exists an embedding $\chi : \mathcal{M} \to \mathcal{N}$. This definition excludes certain pathologies (e.g., self-intersections; see Fig. 5.20(b)).

A real manifold is called *Riemannian* (resp. pseudo-Riemannian) iff each tangent space is equipped with the structure of a real Hilbert space (resp. indefinite Hilbert space), and the inner product depends smoothly on the point (with respect to local coordinates). Riemannian and pseudo-Riemannian manifolds will be thoroughly studied in Vol. III (e.g., the 4-dimensional space-time in general relativity is pseudo-Riemannian).

The map $\chi : \mathcal{M} \to \mathcal{N}$ between two Riemannian manifolds $\mathcal{M}$ and $\mathcal{N}$ is called an *isometry* iff it is a length-preserving diffeomorphism from $\mathcal{M}$ onto $\mathcal{N}$. The map $\chi : \mathcal{M} \to \mathcal{N}$ is called an isometric embedding iff it is a length-preserving embedding. Then the map χ is an isometry from $\mathcal{M}$ onto the image $\chi(\mathcal{M})$.

In order to formulate a typical theorem on embeddings, fix $1 \leq m \leq n$. Let $\mathcal{U}$ be a nonempty open subset of $\mathbb{R}^m$.

The smooth, injective map $F : \mathcal{U} \to \mathbb{R}^n$ *is an embedding iff the linearization*

$$F'(\tau) = \left(\frac{\partial F^j(\tau)}{\partial \tau^k} \right), \qquad j = 1, \ldots, n, \ k = 1, \ldots, m$$

[36] For the proof, we refer to Zeidler (1986), Vol. IV, p. 588 (see the references on page 1049). The reader might wonder why the dimension $2n + 1$ appears. The point is that one needs additional information about the tangent spaces in the proof. Therefore, one has to use the tangent bundle $T\mathcal{M}$ of the manifold $\mathcal{M}$; the tangent bundle consists of all the pairs (x, v) where x is an arbitrary point of the manifold, and v is an arbitrary tangent vector at the point $x \in \mathcal{M}$. As Whitney showed, one needs an extra dimension in order to realize the $2n$-dimensional tangent bundle $T\mathcal{M}$ in $\mathbb{R}^{2n+1}$.

has the rank=m for all parameter values $\tau \in \mathcal{U}$, and the inverse mapping $F^{-1} : F(\mathcal{U}) \to \mathcal{U}$ is continuous.

Then we say that the real m-dimensional submanifold $F(\mathcal{U})$ of the target space $\mathbb{R}^n$ is an embedded manifold of $\mathbb{R}^n$. To illustrate this standard theorem in the theory of manifolds, let us consider two simple examples:

(i) Curves in $\mathbb{R}^3$: Set $F(t) := (x(t), y(t), z(t))$. Let $\mathcal{U}$ be an open interval of the real line $\mathbb{R}$. The smooth injective map $F : \mathcal{U} \to \mathbb{R}^3$ is an embedding iff

$$\dot{x}(t)^2 + \dot{y}(t)^2 + \dot{z}(t)^2 \neq 0 \qquad \text{for all} \quad t \in \mathcal{U},$$

and the inverse map $F^{-1} : F(\mathcal{U}) \to \mathcal{U}$ is continuous. Then the image $F(\mathcal{U})$ is called an embedded curve of $\mathbb{R}^3$ (Fig. 5.20(a)).
In particular, choose $x(t) :\equiv t, z(t) :\equiv 0$. Then $F(x) = (x, y(x), 0)$. Here, $F(x)$ lives in the (x, y)-plane. The smooth map

$$F : \mathcal{U} \to \mathbb{R}^3,$$

corresponding to the plane curve $y = y(x)$, is an embedding iff $y'(x) \neq 0$ for all $x \in \mathcal{U}$.

(ii) Surfaces in $\mathbb{R}^3$: Set $F(x, y) := (f(x, y), g(x, y), z(x, y))$. Let $\mathcal{U}$ be a nonempty open subset of $\mathbb{R}^2$. The smooth injective map $F : \mathcal{U} \to \mathbb{R}^3$ is an embedding iff the linearization

$$F'(x, y) := \begin{pmatrix} f_x(x, y) & g_x(x, y) & z_x(x, y) \\ f_y(x, y) & g_y(x, y) & z_y(x, y) \end{pmatrix}$$

has rank =2 for all $(x, y) \in \mathcal{U}$, and $F^{-1} : F(\mathcal{U}) \to \mathcal{U}$ is continuous. Then the set $F(\mathcal{U})$ is called an embedded surface of $\mathbb{R}^3$.
In particular, choose $f(x, y) :\equiv x, g(x, y) :\equiv y$. Hence $F(x, y) = (x, y, z(x, y))$. Then the smooth map

$$F : \mathcal{U} \to \mathbb{R}^3,$$

corresponding to the surface equation $z = z(x, y)$, is an embedding iff

$$z_x(x, y)^2 + z_y(x, y)^2 \neq 0 \qquad \text{for all} \quad (x, y) \in \mathcal{U}.$$

This condition means that the surface has the tangent plane

$$(x - x_0)z_x(x_0, y_0) + z_y(x_0, x_0)(y - y_0) - z(x_0, y_0) = 0$$

at every point $(x_0, y_0, z(x_0, y_0))$ with $(x_0, y_0) \in \mathcal{U}$.

In order to prove his embedding theorem, John Nash had to solve a nonlinear system (S) of partial differential equations. Unfortunately, the classical implicit function theorem failed. The corresponding iterative method for (S) has the nasty property that the functions loose smoothness in each iterative step. To cure this defect, Nash invented a refined iteration technique:

The idea is to regularize the iterated functions in each step and to use the fast convergence of the Newton method in order to compensate the effects caused by the modifications.

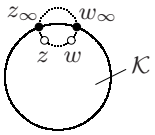


Fig. 5.21. Klein's model of hyperbolic geometry

This approach was further developed by Jürgen Moser (1928–1999) in 1966. Nowadays there exists a well-established mathematical tool called the *hard implicit function theorem*, which allows many interesting applications to sophisticated problems arising in celestial mechanics and the theory of partial differential equations (see the papers by Moser and Hamilton, and the monograph by Krantz and Parker quoted on page 352). There also exist interesting relations to problems arising in quantum field theory (see the papers by Bricmont, Gawędzki, and Kupiainen quoted on page 353).[37]

The elegant method of gauge fixing. In 1989, Matthias Günther showed that the proof of the Nash embedding theorem can be simplified substantially (see the paper quoted on page 352). The basic idea is to add appropriate constraints to the underdetermined system (S) of nonlinear first-order partial differential equations in order to enforce the application of the classical implicit function theorem.

Klein's projective model of hyperbolic geometry. In 1871, the young Felix Klein (1849–1925) published the following model. Consider the open unit disc $\mathcal{K} := \{z \in \mathbb{C} : |z| < 1\}$. By definition, "straight lines" are subsets of $\mathcal{K}$ that correspond to circles, which orthogonally intersect the unit circle (Fig. 5.21). The points of the unit circle are the "points of infinity." For given two different points $z, w \in \mathcal{K}$, there exists a uniquely determined "straight line" which passes through z and w. The "distance" between the points z and w is defined to be the positive number

$$d(z,w) := \ln\left(\frac{z-z_\infty}{z-w_\infty} : \frac{w-z_\infty}{w-w_\infty}\right).$$

It turns out that this model based on the cross ratio is equivalent to Lobachevsky's hyperbolic geometry. Note that the cross ratio is the most important invariant in projective geometry. In contrast to the Beltrami model, the elementary Klein model does not use the theory of Riemannian manifolds.

5.10 The Poincaré Model of Hyperbolic Geometry

> The Poincaré model is the most elegant formulation of hyperbolic non-Euclidean geometry.
>
> Folklore

Elements of the following model were introduced by Henri Poincaré (1854–1912) in 1881. After a long discussion with Felix Klein, Poincaré presented the final form of this model in his famous monograph "La science et l'hypothèse," Paris, 1902. In

[37] For the creation of the theory of non-cooperative games based on Nash-equilibrium states (Ann. of Math. **54** (1951), 286–295), John Nash (born 1928) was awarded the Nobel prize in economics in 1994. His biography can be found in the bestseller by S. Nasar, *A Beautiful Mind*, Simon and Schuster, New York, 1998, which was the basis for a Hollywood movie.

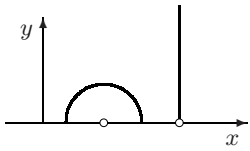


Fig. 5.22. Hyperbolic geodesics (light rays)

what follows, we will specify the situation from Sect. 5.8.3 on page 297 by choosing the special refraction index

$$\boxed{n(x,y) := \frac{1}{y}, \qquad x \in \mathbb{R},\ y > 0.} \tag{5.73}$$

Mnemonically, this corresponds to the metric

$$ds^2 = \frac{dx^2 + dy^2}{y^2}. \tag{5.74}$$

Setting $dz = dx + \mathrm{i}dz$ and $d\bar{z} = dx - \mathrm{i}dy$, mnemonically, we get[38]

$$ds^2 = \frac{dz d\bar{z}}{y^2}. \tag{5.75}$$

Hyperbolic plane. The upper half-plane $\mathcal{H}_{\mathbb{R}}$ equipped with the metric (5.74) is called the real hyperbolic plane. Similarly, $\mathcal{H}_{\mathbb{C}}$ equipped with the metric (5.75) is called the complex hyperbolic plane. Note that:

- The real hyperbolic plane $\mathcal{H}_{\mathbb{R}}$ is a real 2-dimensional manifold.
- The complex hyperbolic plane $\mathcal{H}_{\mathbb{C}}$ is a complex 1-dimensional manifold.

The geometry corresponding to the refraction index (5.73) is called hyperbolic geometry.

Hyperbolic geodesics. By (5.57) on page 300, the equation for geodesics corresponding to the refraction index $n(x,y) = \frac{1}{y}$ reads as

$$\ddot{x} = \frac{2\dot{x}\dot{y}}{y}, \qquad \ddot{y} = \frac{\dot{y}^2 - \dot{x}^2}{y}. \tag{5.76}$$

Proposition 5.11 *The geodesics of the hyperbolic plane are obtained as follows: Consider all circles and straight lines of the complex plane which orthogonally intersect the x-axis, and restrict them to the open upper half-plane.*

The points of the x-axis are called the "points of infinity" of the hyperbolic plane. Maximal geodesics are called straight lines (Fig. 5.22).

Proof. (I) Let $y \equiv \text{const}$. The second equation of (5.76) yields $x \equiv \text{const}$. This is a point, but not a curve. Thus, the curve $y \equiv \text{const}$ is not a geodesic.

(II) We are looking for a geodesic of the form $x = (y)$. Instead of using the system (5.76), it is easier to use the variational problem

[38] Rigorously, we have to use the metric tensors $\frac{dx\otimes dx + dy\otimes dy}{y^2}$ and $\frac{dz\otimes d\bar{z}}{y^2}$.

(a) Euclidean geometry (b) hyperbolic geometry

Fig. 5.23. Parallel lines

$$\int_{y_0}^{y_1} \frac{\sqrt{x'(y)^2+1}}{y}\,dy = \min!, \quad x(y_0)=x_0,\ x(y_1)=x_1.$$

The Euler–Lagrange equation reads as $\frac{d}{dy}\left(\frac{x'(y)}{y\sqrt{x'(y)^2+1}}\right)=0$. Hence

$$\frac{x'(y)}{y\sqrt{x'(y)^2+1}} \equiv \text{const} = x_*.$$

This yields $x \equiv \text{const}$ or $x = x_* \pm \sqrt{R^2-y^2}$ with the constant $R>0$. □

Geodesics. For an arbitrary smooth curve, $x=x(y), y_0 \le y \le y_1$, the hyperbolic arc length satisfies the following inequality:

$$\int_{y_0}^{y_1} \frac{\sqrt{x'(y)^2+1}}{y}\,dy \ge \int_{y_0}^{y_1} \frac{dy}{y} = \ln y_1 - \ln y_0.$$

This goes to $+\infty$ if $y_0 \to +0$. Therefore:

The hyperbolic arc length of hyperbolic straight lines is infinite.

Hyperbolic parallels and the violation of Euclid's parallel axiom. In Euclidean geometry, for any given point P_0 outside a given straight line $\mathcal{G}$, there exists precisely one straight line which does not intersect $\mathcal{G}$. This is called a parallel straight line to $\mathcal{G}$.

In the hyperbolic geometry of the real hyperbolic plane, geodesic lines are also called *hyperbolic straight lines*. Then, for any given point P_0 outside a given hyperbolic straight line $\mathcal{G}$, there exist infinitely many hyperbolic straight lines through the point P_0 which do not intersect $\mathcal{G}$. In other words, there exist infinitely many hyperbolic parallel straight lines to $\mathcal{G}$ which pass through the point P_0 (Fig. 5.23(b)).

Consequently, the hyperbolic geometry of the real hyperbolic plane represents a non-Euclidean geometry.

The real (resp. and complex) hyperbolic plane is equipped with the following mathematical structures:

- complete metric space,
- Riemannian manifold,
- non-compact Riemann surface,
- Kähler manifold, and
- symplectic manifold.

In what follows we want to discuss this.

The hyperbolic plane as a complete metric space with respect to geodesic distance. We are given the two points $P, Q \in \mathcal{H}_{\mathbb{C}}$. Let the curve

$$C : z = z(\tau), \quad \tau_0 \leq \tau \leq \tau_1$$

be a geodesic which connects the point P with Q (see Fig. 5.22 on page 315). Using the so-called geodesic distance defined by

$$d(P, Q) := l(C),$$

where $l(C)$ is the hyperbolic arc length of the geodesic C, the complex hyperbolic plane becomes a metric space. This means that, for arbitrary points $P, Q, S \in \mathcal{H}_{\mathbb{C}}$, the following hold:

(M1) $d(P, Q) \geq 0$ and $d(P, Q) = 0$ iff $P = Q$,
[M2) $d(P, Q) = d(Q, P)$,
(M3) $d(P, Q) \leq d(P, S) + d(S, Q)$.

Moreover, the complex hyperbolic plane is a complete metric space. By definition, this means that each Cauchy sequence is convergent.[39]

The symmetry group of the hyperbolic geometry. Consider the Möbius transformation[40]

$$w = \frac{\alpha z + \beta}{\gamma z + \delta}$$

with real coefficients $\alpha, \beta, \gamma, \delta$ and $\alpha\delta - \beta\gamma \neq 0$. Each of these transformations represents a conformal diffeomorphism from the complex hyperbolic plane $\mathcal{H}_{\mathbb{C}}$ onto itself. The set of all these transformations

$$T : \mathcal{H}_{\mathbb{C}} \to \mathcal{H}_{\mathbb{C}}$$

forms a group $Sym(\mathcal{H}_{\mathbb{C}})$, which is called the symmetry group of the complex hyperbolic plane $\mathcal{H}_{\mathbb{C}}$. By definition, precisely the invariants under this group are properties of hyperbolic geometry on $\mathcal{H}_{\mathbb{C}}$. Examples are: arc length, angle, surface area, geodesic line, geodesic distance, and Gaussian curvature.

The action of the real symplectic group $SL(2, \mathbb{R})$ on the complex hyperbolic plane. The map

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \mapsto \frac{\alpha z + \beta}{\gamma z + \delta} \tag{5.77}$$

is a group morphism from the group $SL(2, \mathbb{R})$ onto the group $Sym(\mathcal{H}_{\mathbb{C}})$. Precisely the unit matrix I and $-I$ are mapped to the identical Möbius transformation with $\alpha = \delta = 1$ and $\beta = \gamma = 0$ Consequently, we have the group isomorphism

$$Sym(\mathcal{H}_{\mathbb{C}}) \simeq SL(2, \mathbb{R})/\{\pm I\}.$$

Next we want to study the curvature properties of the hyperbolic plane. This can be done most elegantly by using the notion of Kähler manifold.

[39] A sequence (P_n) of points in $\mathcal{H}_{\mathbb{C}}$ is called Cauchy iff, for any $\varepsilon > 0$, there exists an index n_0 such that $d(P_n, P_m) < \varepsilon$ for all $n, m \geq n_0$. Completeness means that, for any Cauchy sequence (P_n), there exists a point P in the complex hyperbolic plane such that $\lim_{n \to +\infty} d(P_n, P) = 0$.

[40] Möbius (1790–1868).

5.10.1 Kähler Geometry and the Gaussian Curvature

> Many important manifolds arising in mathematics and physics (e.g., Riemann surfaces) can be equipped with an Hermitean Kähler metric.
>
> Folklore

> Erich Kähler's seminal article *On a remarkable Hermitean metric*[41] has grown into a domain in itself. What is even more striking concerning this article is that more or less every half page, Erich Kähler opens a new path that has later turned out to be crucial for the development of the subject.
>
> Jean-Piere Bourguignon, 2004

There are the following three fundamental geometries used in physics:

(i) symplectic geometry (geometrical optics and classical mechanics),
(ii) Riemannian und pseudo-Riemannian geometry (general theory of relativity on gravitation),
(iii) unitary geometry and Hilbert spaces (quantum theory).

Kähler geometry combines these three essential mathematical structures with each other. In fact, Kähler geometry plays a fundamental role in complex geometry, algebraic geometry, and string theory. We will study this in later volumes. We refer to the hints for further reading on Kähler geometry to be found on page 350. At this point, we will only discuss a few basic ideas.

Our goal is to use the complexification $\mathcal{H}_{\mathbb{C}}$ of the real 2-dimensional hyperbolic plane $\mathcal{H}_{\mathbb{R}}$ in order to simplify computations, by using methods from complex analysis. To this end, we introduce the following notation:

- $z := x + \mathrm{i}y$ and $\bar{z} := x - \mathrm{i}y$ (variables),
- $dz := dx + \mathrm{i}dy$ and $d\bar{z} := dx - \mathrm{i}dy$ (differentials),
- $\partial_x = \frac{\partial}{\partial x}$ and $\partial_y = \frac{\partial}{\partial y}$ (partial derivatives),
- $\partial_z := \partial_x - \mathrm{i}\partial_y$ and $\partial_{\bar{z}} := \partial_x + \mathrm{i}\partial_y$,
- $\partial := \partial_z \cdot dz$ and $\bar{\partial} := \partial_{\bar{z}} \cdot d\bar{z}$.

In particular, for the smooth function $U : \mathcal{H}_{\mathbb{C}} \to \mathbb{C}$, we get the differential forms $\bar{\partial}U(z) = \partial_{\bar{z}}U(z)\, d\bar{z}$, and

$$\partial\bar{\partial}U(z) = \partial_z\partial_{\bar{z}}U(z)\, dz \wedge d\bar{z}, \qquad z \in \mathcal{H}_{\mathbb{C}}. \tag{5.78}$$

The function U is called the Kähler potential of the 2-form located on the right-hand side of (5.78).

The Riemann surface. By definition, complex one-dimensional connected manifolds are called Riemann surfaces. The complex hyperbolic plane $\mathcal{H}_{\mathbb{C}}$ is a (non-compact) Riemann surface.[42]

Tangent space. Consider the complex hyperbolic plane $\mathcal{H}_{\mathbb{C}}$. Fix the point $P := z_0$. Let $z = z(\tau), \tau_0 \le \tau \le \tau_1$, be a smooth curve on $\mathcal{H}_{\mathbb{C}}$ through the point P, that is, say, $z(0) = z_0$ and $0 \in]\tau_0, \tau_1[$. Then the derivative $v := \dot{z}(0)$ is called a tangent vector of $\mathcal{H}_{\mathbb{C}}$ at the point P. The set of all possible tangent vectors v at the point P is called the tangent space $T_P\mathcal{H}_{\mathbb{C}}$ of $\mathcal{H}_{\mathbb{C}}$ at the point P. For all $v, w \in T_P\mathcal{H}_{\mathbb{C}}$, we will use the following definitions:

---

[41] E. Kähler, Über eine bemerkenswerte Hermitesche Metrik, Abhandlungen des Mathematischen Seminars der Universität Hamburg **9** (1932/33), 173–186.
J. Bourguignon, The unabated vitality of Kählerian geometry. In: E. Kähler, Collected Works, de Gruyter, Berlin, 2004, pp. 737–766.

[42] The relation of the complex hyperbolic plane to the general theory of Riemann surfaces is studied in J. Jost, Compact Riemann Surfaces: An Introduction to Contemporary Mathematics, 3rd edn., Springer, Berlin, 2006.

- $dx_P(v) := \Re(v)$ and $dy_P(v) := \Im(v)$,
- $dz_P(v) := v$ and $d\bar{z}_P(v) := \bar{v}$,
- $(dz \otimes d\bar{z})_P(v, w) := dz_P(v) d\bar{z}_P(w) = v\bar{w}$,
- $(d\bar{z} \otimes dz)_P(v, w) = d\bar{z}_P(v) dz_P(w) := \bar{v}w$.

Moreover, $dz \wedge d\bar{z} := dz \otimes d\bar{z} - d\bar{z} \otimes dz$. Hence

- $(dz \wedge d\bar{z})_P(v, w) = v\bar{w} - \bar{v}w$ for all $v, w \in T_P\mathcal{H}_{\mathbb{C}}$.

The two fundamental bilinear tensor fields on the complex hyperbolic plane. Let us introduce

- the complex-valued Hermitean metric tensor field $h := n(z)^2\, dz \otimes d\bar{z}$ and
- the real-valued Kähler 2-form $\kappa := \frac{\mathrm{i}}{2} n(z)^2\, dz \wedge d\bar{z}$

on the complex manifold $\mathcal{H}_{\mathbb{C}}$. Here, $z \to n(z)$ is a given smooth function with positive values on $\mathcal{H}_{\mathbb{C}}$.[43] Trivially, we have

$$d\kappa = 0,$$

since $d\kappa = \frac{\mathrm{i}}{2} d(n(z)^2) \wedge dz \wedge d\bar{z}$, and wedge products with three factors of the form dx, dy vanish identically. However, in complex dimensions > 1, the condition $d\kappa = 0$ represents a crucial restriction on the structure of Kähler manifolds. The tensor field h is called Hermitean, since the map $(v, w) \mapsto h(v, w)$ represents an Hermitean form. That is, the following conditions are satisfied for all points $P \in \mathcal{H}_{\mathbb{C}}$ and all tangent vectors $u, v, w \in T_P\mathcal{H}_{\mathbb{C}}$, as well as all complex numbers α, β:

- $h_P(v, w) \in \mathbb{C}$ and $h_P(v, w)^\dagger = h_P(w, v)$,
- $h_P(\alpha u + \beta v, w) = \alpha h_P(u, w) + \beta h_P(v, w)$.

Define

$$\boxed{\langle v|w\rangle_P := h_P(v, w)^\dagger \qquad \text{for all} \quad v, w \in T_P\mathcal{H}_{\mathbb{C}}}$$

and all points $P \in \mathcal{H}_{\mathbb{C}}$.

> *This way, the tangent space $T_P\mathbb{C}$ becomes a one-dimensional complex Hilbert space equipped with the inner product $\langle .|. \rangle$.*

In particular, we have $\langle w|\alpha u + \beta v\rangle_P = \alpha\langle w|u\rangle_P + \beta\langle w|v\rangle_P$ for all complex numbers α, β.[44] Now we are able to apply the theory of complex Hilbert spaces. According to Sect. 5.8.2, for all tangent vectors $v, w \in T_P\mathcal{H}_{\mathbb{C}}$, we define:[45]

$$\langle v|w\rangle_{\mathbb{R}} := \Re(\langle v|w\rangle) \quad \text{and} \quad \kappa_P(v, w) := \Im(\langle v|w\rangle).$$

Note that the definition of κ coincides with the Kähler form introduced above. If we regard the tangent space $T_P\mathcal{H}_{\mathbb{C}} = \mathbb{C}$ as a real 2-dimensional space $\mathbb{C}_{\mathbb{R}}$, this space becomes a real Hilbert space equipped with the inner product $\langle v|w\rangle_{\mathbb{R}}$. This corresponds to the Riemannian structure on the real manifold $\mathcal{H}_{\mathbb{R}}$, where the tangent spaces $T_P\mathcal{H}_{\mathbb{R}}$ are isomorphic to $\mathbb{R}^2$, as real spaces.

[43] Explicitly, for all $v, w \in \mathbb{C}$, we have

$$h_P(v, w) = n(z)^2 v\bar{w}, \qquad \kappa_P(v, w) = \tfrac{\mathrm{i}}{2} n(z)^2 (v\bar{w} - w\bar{v}).$$

Note that, in the present case, the tangent space $T_P\mathcal{H}_{\mathbb{C}}$ coincides with $\mathbb{C}$. However, we are going to use a language which prepares the generalization to arbitrary Kähler manifolds on page 323.

[44] This is the convention used by physicists for inner products (in order to fit the Dirac calculus). In this series of monographs, we adopt this convention.

[45] All of the expressions depend on the choice of the point P in $\mathcal{H}_{\mathbb{C}}$. To simplify notation, we suppress the index P.

Hermitean forms are named after Charles Hermite (1822–1901) who studied the principal axis transformation for complex self-adjoint matrices. To this end, he had to replace real quadratic forms by complex-valued expressions which are called Hermitean forms nowadays. Henri Poincaré (1854–1912), one of the greatest mathematicians of all times, was a student of Hermite at the Ecole Polytechnique in Paris.

Hyperbolic geometry. In order to get the hyperbolic geometry, we choose the refraction index function

$$\boxed{n(z) := \frac{1}{y} \qquad \text{for all} \quad z \in \mathcal{H}_{\mathbb{C}}.}$$

Gauss based his theory of 2-dimensional surfaces on two symmetric bilinear tensor fields (called the first and second fundamental form). Élie Cartan noticed that it is very convenient to use antisymmetric bilinear forms, too. In modern differential geometry, Cartan's antisymmetric differential forms play a fundamental role (e.g., the curvature form to be considered below).

Hyperbolic arc length. Let $C : z = z(\tau), \tau_0 \le \tau \le \tau_1$, be a smooth curve on $\mathcal{H}_{\mathbb{C}}$. The hyperbolic arc length of C is equal to

$$l(C) = \int_{\tau_0}^{\tau_1} ||\dot{z}(\tau)||\, d\tau = \int_{\tau_0}^{\tau_1} \frac{\sqrt{\dot{x}(\tau)^2 + \dot{y}(\tau)^2}}{y(\tau)}\, d\tau.$$

Hyperbolic angle. Let $z = z_j(\tau), \tau_0 \le \tau \le \tau_1$, $j = 1, 2$, be two smooth curves passing through the point P, that is, say, $z_j(0) = z_0$ for $j = 1, 2$, and $0 \in]\tau_0, \tau_1[$. By definition, the hyperbolic angle α between the two curves at the intersection point P is defined by[46]

$$\cos\alpha = \frac{\Re\langle \dot{z}_1 | \dot{z}_2 \rangle}{||\dot{z}_1|| \cdot ||\dot{z}_2||} = \frac{\dot{x}_1\dot{x}_2 + \dot{y}_1\dot{y}_2}{\sqrt{\dot{x}_1^2 + \dot{y}_1^2} \cdot \sqrt{\dot{x}_2^2 + \dot{y}_2^2}}.$$

This coincides with the angle α defined in Euclidean geometry.

Hyperbolic area measure. For the Kähler form, we get

$$\kappa = \tfrac{\mathrm{i}}{2} n(z)^2 dz \wedge d\bar{z} = \tfrac{\mathrm{i}}{2} n(z)^2 (dx + \mathrm{i}dy) \wedge (dx - \mathrm{i}dy) = n(z)^2 dx \wedge dy$$

with $n(z) = \frac{1}{y}$. Let $\mathcal{U}$ be a bounded open subset of $\mathcal{H}_{\mathbb{C}}$. The hyperbolic area of the set $\mathcal{U}$ is defined by

$$\mathrm{meas}(\mathcal{U}) := \int_{\mathcal{U}} \kappa.$$

Hence $\mathrm{meas}(\mathcal{U}) = \int_{\mathcal{U}} \frac{1}{y^2}\, dx \wedge dy = \int_{\mathcal{U}} \frac{1}{y^2}\, dxdy$.

Orientation. Let v and w be two tangent vectors at the point P. Then:

- v and w are linearly independent iff $\kappa_P(v, w) \neq 0$.
- The ordered pair (v, w) is positively oriented iff $\kappa_P(v, w) > 0$.

Kähler form and Kähler potential. Again fix the point P which corresponds to $z \in \mathcal{H}_{\mathbb{C}}$. Setting $U(x, y) := -2\mathrm{i} \ln y$, we obtain

$$\kappa_P = \partial_z \partial_{\bar{z}} U(z)\, dz \wedge d\bar{z}.$$

[46] Here, $\dot{z}_j, \dot{x}_j, \dot{y}_j$ stand for $\dot{z}_j(0), \dot{x}_j(0), \dot{y}_j(0)$, respectively.

That is, the function U is the Kähler potential of κ.[47] The use of potentials is crucial in physics for simplifying the computation of forces and physical fields. For example, in electrodynamics, the four-potential simplifies the computation of solutions of the Maxwell equations. Similarly, the Kähler potential simplifies computations to be performed in Kähler geometry.

Hyperbolic Gaussian curvature. By definition, the Gaussian curvature $K(P)$ of the complex hyperbolic plane $\mathcal{H}_{\mathbb{C}}$ at the point P is equal to

$$K(P) := -\frac{4}{n(z)^2}\partial_z\partial_{\bar{z}} \ln n(z). \tag{5.79}$$

This definition refers to a general smooth refraction index $n : \mathcal{H}_{\mathbb{C}} \to \mathbb{R}$ with $n(z) > 0$ for all $z \in \mathcal{H}_{\mathbb{R}}$. In the special case of hyperbolic geometry, we have $n(x + \mathrm{i}y) := y$. Then

$$\boxed{K \equiv -1.} \tag{5.80}$$

Thus the hyperbolic plane has a negative constant Gauss curvature.

The Laplacian. The differential operator

$$\Delta := -\frac{4}{n(z)^2}\frac{\partial^2}{\partial z \partial \bar{z}} = -\frac{1}{n(x,y)^2}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)$$

is called the Laplacian (or the Laplace–Beltrami operator) related to the refraction index $n(x, y)$). Formula (5.79) reads as

$$K(P) = \Delta \ln n(z).$$

The Gauss–Bonnet theorem on hyperbolic triangles. Consider a hyperbolic geodesic triangle as pictured in Fig. 5.14(b) on page 302. Then

$$\boxed{\alpha + \beta + \gamma = \pi + KA} \tag{5.81}$$

with $K = -1$. For general refraction index $n = n(x, y)$, we obtain

$$\alpha + \beta + \gamma = \pi + \int_{\mathcal{T}} K(x,y)n(x,y)^2 dxdy.$$

Here, we integrate over the triangle set $\mathcal{T}$. If the Gaussian curvature is constant, then we get (5.81). In particular, in the case of Euclidean geometry we have (5.81) with $K \equiv 0$.

Sartorius von Waltershausen, a good friend of Gauss, reported that Gauss tried to check the light-ray geometry of our universe by measuring the sum $\alpha + \beta + \gamma$ of the angles α, β, γ of a huge triangle established by three light rays running between the tops of three mountains in Germany: Hoher Hagen, Brocken, and Inselsberg. The distance between the mountains is about 150 km. Gauss measured

$$\alpha + \beta + \gamma = \pi,$$

within the limits of accuracy. Nowadays we know that the sum $\alpha + \beta + \gamma$ is different from π for sufficiently large light-ray triangles, by Einstein's theory of general relativity.

[47] Explicitly, $\partial_z\partial_{\bar{z}}U(y) = \frac{1}{4}\frac{\partial^2}{\partial y^2}U(y) = \frac{\mathrm{i}}{2y^2}$.

The first Chern class of the complex hyperbolic plane. The differential form

$$\gamma_P := \frac{\mathrm{i}K(z)n(z)^2}{4\pi}\ dz \wedge d\bar{z} = \frac{K(P)}{2\pi}\ dx \wedge dy$$

is called the Chern form of both the complex manifold $\mathcal{H}_{\mathbb{C}}$ and the real manifold $\mathcal{H}_{\mathbb{R}}$ (with respect to the refraction index n). In the special case where $n(x+\mathrm{i}y) := \frac{1}{y}$ for all $z \in \mathcal{H}_{\mathbb{C}}$, we get the Chern form of both the complex and the real hyperbolic plane. Here, $K \equiv -1$. The corresponding cohomology class $[\gamma]$ is called the first Chern class of both $\mathcal{H}_{\mathbb{C}}$ and $\mathcal{H}_{\mathbb{R}}$.

> *The crucial point is that the first Chern class $[\gamma]$ does not depend on the choice of the refraction index $n = n(x, y)$. The first Chern class reflects a topological property of the real hyperbolic plane $\mathcal{H}_{\mathbb{R}}$, namely, the orientability of this manifold. The same is true for the complex hyperbolic plane $\mathcal{H}_{\mathbb{C}}$.*

The first Chern class of the real hyperbolic plane is the prototype of the modern topological theory of characteristic classes created by Chern in about 1945. Note that characteristic classes are important topological invariants, which are frequently encountered in modern mathematics and physics. As an introduction, we refer to J. Milnor and Stasheff, Characteristic Classes, Princeton University Press, 1974, as well as to R. Bott and L. Tu, Differential Forms in Algebraic Topology, Springer, New York, 1982.

Next we want to introduce the general notion of a Kähler manifold. For the convenience of the reader, we start with local Kähler manifolds.

Local Kähler manifold. Fix $n = 1, 2, \ldots$ In what follows, we will use the Einstein convention, that is, we sum over equal upper and lower indices from 1 to n. The elements of the space $\mathbb{C}^n$ are the points $z = (z^1, \ldots, z^n)$ with $z^k \in \mathbb{C}^n$ for all $k = 1, \ldots, n$. Let $\mathcal{U}$ be a nonempty open subset of $\mathbb{C}^n$. The two key formulas read as follows:

$$h_z := h_{r\bar{s}}(z)\ dz^r \otimes d\bar{z}^{\bar{s}}, \qquad \kappa_z := \tfrac{\mathrm{i}}{2} h_{r\bar{s}}(z)\ dz^r \wedge d\bar{z}^{\bar{s}}. \tag{5.82}$$

Explicitly, for all $v, w \in \mathbb{C}^n$, we have

$$h(v, w) = v^r h_{r\bar{s}} \bar{w}^{\bar{s}}, \qquad \kappa(v, w) = \tfrac{\mathrm{i}}{2}(v^r h_{r\bar{s}} \bar{w}^{\bar{s}} - w^r h_{r\bar{s}} \bar{v}^{\bar{s}}).$$

By definition, the open set $\mathcal{U}$ is a local Kähler manifold with respect to h, κ iff the following hold:

- All the functions $h_{r\bar{s}} : \mathcal{U} \to \mathbb{C}$ are smooth.
- The complex $(n \times n)$-matrix $(h_{r\bar{s}})$ is self-adjoint (also called Hermitean). That is, $h_{r\bar{s}}(z)^\dagger = h_{s\bar{r}}(z)$ for all $r, s = 1, \ldots, n$.
- For any $z \in \mathcal{U}$, all the eigenvalues of the matrix $(h_{r\bar{s}}(z))$ are positive.
- $d\kappa = 0$ on $\mathcal{U}$.

If we define, $\langle v|w\rangle_z^\dagger := h_z(v, w)^\dagger$ for all $v, w \in \mathbb{C}^n$, then this is an inner product on $\mathbb{C}^n$. Moreover,

$$\kappa(v, w) = \Im(\langle v|w\rangle) \qquad \text{for all} \quad v, w \in \mathbb{C}^n.$$

We also define $\langle v|w\rangle_{\mathbb{R}} := \Re(\langle v|w\rangle)$ for all $v, w \in \mathbb{C}^n$. We have the following key definitions:

- $R := \Delta \ln \det(h_{r\bar{s}})$ (scalar curvature of $\mathcal{U}$),
- $R_{r\bar{s}}\ dz^r \wedge d\bar{z}^s := -\partial\bar{\partial} \ln \det(h_{r\bar{s}})$ (Ricci form),
- $\gamma := \frac{\mathrm{i}}{2\pi} R_{r\bar{s}}\ dz^r \wedge d\bar{z}^{\bar{s}}$ (Chern form),
- $[\gamma] \in H^2(\mathcal{U})$ (first Chern class of $\mathcal{U}$),

- $v := dz^1 \wedge d\bar{z}^1 \wedge dz^2 \wedge d\bar{z}^2 \wedge \cdots \wedge dz^n \wedge d\bar{z}^n$ (volume form of $\mathcal{U}$).

Here, we set $\partial := \frac{\partial}{\partial z^k}\, dz^k$ and $\bar{\partial} := \frac{\partial}{\partial \bar{z}^k}\, d\bar{z}^k$. The operator

$$\Delta := -(\det(h_{r\bar{s}}))^{-1/2}\partial\bar{\partial}$$

is called the Laplace–Beltrami operator. For $n = 1$, the Gaussian curvature is given by $K = R/2$.

Global Kähler manifold. Let $\mathcal{M}$ be a complex n-dimensional manifold equipped with a smooth complex-valued 2-tensor field h on $\mathcal{M}$. Set

$$\langle v|w\rangle_P := h_P(v,w)^\dagger \qquad \text{for all} \quad v, w \in T_P\mathcal{M}.$$

Then, $\mathcal{M}$ is called a Kähler manifold iff the following hold:

- Each tangent space $T_P\mathcal{M}$ becomes a complex Hilbert space equipped with the inner product $\langle .|.\rangle$.
- Define $\kappa_P(v,w) := \Im(\langle v|w\rangle_P)$ for all $v, w \in T_P\mathcal{M}$.
- $d\kappa = 0$ on $\mathcal{M}$.

In other words, h and κ look like (5.82) in local coordinates. The condition $d\kappa = 0$ guarantees that the equivalence class $[\kappa]$ is an element of the second (de Rham) cohomology group $H^2(\mathcal{M})$ of the Kähler manifold $\mathcal{M}$.

5.10.2 Kähler–Einstein Geometry

A Kähler manifold $\mathcal{M}$ is called a Kähler–Einstein manifold iff the Ricci form is proportional to the Kähler form, that is there exists a complex constant λ such that

$$\boxed{R_{j\bar{k}} dz^j \wedge d\bar{z}^{\bar{k}} = \lambda h_{j\bar{k}} dz^j \wedge d\bar{z}^{\bar{k}} \qquad \text{on } \mathcal{M}.}$$

These manifolds are related to Einstein's equations for curved space-time manifolds with vanishing external energy-matter field. This models a universe where only gravitation is acting (see (5.96) on page 330).

5.10.3 Symplectic Geometry

The real 2-dimensional manifold $\mathcal{H}_\mathbb{R}$ becomes a symplectic manifold equipped with the symplectic form $\sigma := \frac{1}{y^2}\, dx \wedge dy$. The map $F : \mathcal{H}_\mathbb{R} \to \mathcal{H}_\mathbb{R}$ is called a symplectic isomorphism iff

$$\boxed{F^*\sigma = \sigma \qquad \text{on } \mathcal{H}_\mathbb{R}.}$$

Recall that, for a diffeomorpism $F : \mathbb{R}^2 \to \mathbb{R}^2$ of the form

$$\xi = \xi(x,y), \qquad \eta = \eta(x,y), \qquad (x,y) \in \mathbb{R}^2,$$

the linearization $F'(P) : \mathbb{R}^2 \to \mathbb{R}^2$ at the point $P = (x,y)$ is given by the linear map

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \xi_x(x,y) & \xi_y(x,y) \\ \eta_x(x,y) & \eta_y(x,y) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}.$$

The determinant of $F'(P)$ is called the Jacobian of F at the point P. Synonymously, we will use the following symbols:

$$\det F'(P) = \begin{vmatrix} \xi_x(x,y) & \xi_y(x,y) \\ \eta_x(x,y) & \eta_y(x,y) \end{vmatrix} = \frac{\partial(\xi(x,y),\eta(x,y))}{\partial(x,y)}.$$

Proposition 5.12 (i) *The diffeomorphism $F : \mathbb{R}^2 \to \mathbb{R}^2$ is a symplectic isomorphism with respect to the symplectic form $dx \wedge dy$ iff $F'(P) \in SL(2,\mathbb{R})$ for all points $P \in \mathbb{R}^2$, that is, F preserves both the Euclidean area measure and the orientation.*

(ii) *The diffeomorphism $F : \mathcal{H}_{\mathbb{R}} \to \mathcal{H}_{\mathbb{R}}$ is a symplectic isomorphism with respect to the symplectic form σ iff*

$$\frac{1}{\eta(x,y)^2} \cdot \frac{\partial(\xi(x,y),\eta(x,y))}{\partial(x,y)} = \frac{1}{y^2} \qquad \text{for all} \quad (x,y) \in \mathcal{H}^2_{\mathbb{R}},$$

that is, F preserves both the hyperbolic area measure and the orientation of the real hyperbolic plane $\mathcal{H}_{\mathbb{R}}$.

Explicitly, $F'(P) \in SL(2,\mathbb{R})$ means that $\frac{\partial(\xi(x,y),\eta(x,y))}{\partial(x,y)} = 1$. Concerning (ii), for any bounded open subset of $\mathcal{U}$, we get

$$\int_{F(\mathcal{U})} \frac{d\xi d\eta}{\eta^2} = \int_{\mathcal{U}} \frac{1}{\eta^2(x,y)} \frac{\partial(\xi(x,y),\eta(x,y))}{\partial(x,y)}\, dxdy = \int_{\mathcal{U}} \frac{dxdy}{y^2}.$$

This tells us that

- the hyperbolic area measure meas$(F(\mathcal{U}))$ of the transformed set $F(\mathcal{U})$
- equals the hyperbolic area measure meas$(\mathcal{U})$ of the original set $\mathcal{U}$.

Using the elegant language of differential forms, one briefly writes

$$\boxed{\int_{F(\mathcal{U})} \sigma = \int_{\mathcal{U}} F^*\sigma.} \tag{5.83}$$

The proof of Prop. 5.12 will be given in Problem 5.3.

5.10.4 Riemannian Geometry

> The doubts about what we may believe about the geometry of physical space, raised by the work of Gauss, Lobachevsky and Bólayi, stimulated one of the major creations of the 19th century, Riemannian geometry. The creator was Bernhard Riemann (1826–1866) – the deepest philosopher of geometry.[48]
>
> Morris Kline, 1972

> It was Riemann, who probably more than anyone else, enriched mathematics with new ideas. These ideas display an unusual degree of vitality and impulse the whole of mathematics as well as many branches of physics.[49]
>
> Krysztof Maurin, 1997

[48] M. Kline, Mathematical Thought from Ancient to Modern Times, Vols. 1–3, Oxford University Press, New York, 1972.

[49] K. Maurin, The Riemann Legacy: Riemannian Ideas in Mathematics and Physics of the 20th Century, Kluwer, Dordrecht, 1997.

Hyperbolic straight lines (i.e., semi-circles centered at the x-axis or straight lines parallel to the y-axis) are curved compared with Euclidean straight lines. Intuitively, this is responsible for the nontrivial curvature of the real hyperbolic plane. The main task of differential geometry is to measure this curvature of the manifold. This was done by Gauss, Riemann, and Élie Cartan in different settings.[50]

Metric properties of the real hyperbolic plane. For all points P of the real hyperbolic plane $\mathcal{H}_{\mathbb{R}}$ and all tangent vectors $v, w \in \mathcal{H}_{\mathbb{R}}$, we set:

- $g_P := \frac{dx \otimes dx + dy \otimes dy}{y^2}$ (metric tensor of $\mathcal{H}_{\mathbb{R}}$ at the point P),
- $\langle v|w\rangle_P := g_P(v,w)$ (inner product on the tangent space $T_P\mathcal{H}_{\mathbb{R}}$ of the real hyperbolic plane at the point P),
- $||v|| := \sqrt{\langle v|v\rangle_P}$ (length of the tangent vector v at the point P),
- $v_P := \frac{dx \wedge dy}{y^2}$ (volume form of $\mathcal{H}_{\mathbb{R}}$ at the point P).

These quantities can be used in order to introduce the following fundamental notions for the real hyperbolic plane.

(i) Length $l(C)$ of a smooth curve C: $x = x(\tau), y = y(\tau), \tau_0 \le \tau \le \tau_1$: We define

$$\boxed{l(C) := \int_{\tau_0}^{\tau_1} ||v(\tau)||\, d\tau,} \tag{5.84}$$

where $v(\tau) := (\dot{x}(\tau), \dot{y}(\tau))$ is the tangent vector of the curve C at the point P corresponding to the curve parameter τ.

(ii) Angle α between two curves C and C_* at the intersection point P:

$$\boxed{\cos\alpha := \frac{\langle v|v_*\rangle_P}{||v||_P \cdot ||v_*||_P}.} \tag{5.85}$$

Here, v (resp. v_*) are tangent vectors of the smooth curve C (resp. C_*) at the intersection point P.

(iii) Area of an open bounded subset $\mathcal{U}$ of $\mathcal{H}_{\mathbb{R}}$:

$$\boxed{\mathrm{meas}(\mathcal{U}) := \int_{\mathcal{U}} v.} \tag{5.86}$$

This means that $\mathrm{meas}(\mathcal{U}) = \int_{\mathcal{U}} \frac{dxdy}{y^2}$.

Geodesics (generalized straight lines). For introducing curvature, our starting point is the equation for geodesics. Set

$$v(\tau) := \begin{pmatrix} \dot{x}(\tau) \\ \dot{y}(\tau) \end{pmatrix}.$$

By (5.57) on page 300, the equation for a geodesics $x = x(\tau), y = y(\tau)$ can be written as

[50] Gauss (1777–1855), Betti (1823–1892), Riemann (1826–1866), Christoffel (1829–1900), Ricci-Curbastro (1853–1925), Bianchi (1856–1928), Élie Cartan (1869–1951), Levi-Civita (1873–1941), Einstein (1879–1955), Weyl (1885–1955), Ehresmann (1905–1979), Yang (born 1922).

$$\boxed{\dot{v} + (\dot{x}\mathcal{A}_1(x,y) + \dot{y}\mathcal{A}_2(x,y)) \cdot v = 0,} \tag{5.87}$$

by introducing the so-called connection matrices

$$\mathcal{A}_1 := \begin{pmatrix} \varrho_x & \varrho_y \\ -\varrho_y & \varrho_x \end{pmatrix}, \quad \mathcal{A}_2 := \begin{pmatrix} \varrho_y & -\varrho_x \\ \varrho_x & \varrho_y \end{pmatrix}.$$

Here, we set $\varrho(x,y) := \ln n(x,y)$.

Curvature properties of the real hyperbolic plane. In what follows we will use the Einstein convention, that is, we will sum over equal upper and lower indices from 1 to 2. We set $x^1 := x, x^2 := y$, and

$$\partial_j := \frac{\partial}{\partial x^j}, \qquad j = 1, 2.$$

In Euclidean geometry, the equation for a geodesic (i.e., a straight line) reads as $\frac{dv(\tau)}{d\tau} \equiv 0$. This motivates us to write the equation (5.87) of a geodesic on the hyperbolic plane in the following form:

$$\boxed{\frac{Dv(\tau)}{d\tau} \equiv 0 \qquad \text{along } \gamma = \gamma(\tau).} \tag{5.88}$$

To this end, we introduce the following notation:

- $\nabla_i := \partial_i + \mathcal{A}_i$ (Ricci's covariant partial derivative),
- $\frac{D}{d\tau} := \dot{x}^i(\tau)\nabla_i$ (covariant directional derivative),
- $\mathcal{A} := \mathcal{A}_i dx^i$ (Cartan's connection 1-form). In terms of Cartan's connection 1-form $\mathcal{A}$, the equation (5.87) of geodesics reads as

$$\boxed{\dot{v}(\tau) + \mathcal{A}_{\gamma(\tau)}(\dot{\gamma}(\tau)) \cdot v(\tau) = 0, \qquad \tau_0 \le \tau \le \tau_1.} \tag{5.89}$$

  Here, we write the curve C as $\gamma(\tau) := (x(\tau), y(\tau))$, and we set $v(\tau) := \dot{\gamma}(\tau)$.
- $\mathcal{F} := \frac{1}{2}\mathcal{F}_{jk}\, dx^j \wedge dx^k$ (Cartan's curvature 2-form). Here,

$$\mathcal{F}_{jk} := \nabla_j \nabla_k - \nabla_k \nabla_j, \qquad j, k = 1, 2.$$

  Using the Lie bracket $[\mathcal{A}, \mathcal{B}]_- := \mathcal{A}\mathcal{B} - \mathcal{B}\mathcal{A}$, we get the fundamental structural equation

$$\boxed{\mathcal{F}_{jk} = \partial_j \mathcal{A}_k - \partial_k \mathcal{A}_j + [\mathcal{A}_j, \mathcal{A}_k]_-, \quad j, k = 1, 2.}$$

  Note that $\mathcal{F}_{jk} = -\mathcal{F}_{kj}$ for $j, k = 1, 2$.
- $\mathcal{R}_P(a, b, u, v) := -\langle \mathcal{F}(a, b)u | v \rangle_P$ for all tangent vectors $a, b, u, v \in T_P\mathcal{H}_{\mathbb{R}}$ at the point P (Riemann's curvature tensor).[51]

[51] Unfortunately, the definition of the Riemann curvature tensor is not unique in the literature; there appear modified definitions based on permutations of the arguments a, b, u, v. Our definition is chosen in such a way that the Gaussian curvature and the scalar curvature have the same sign. Our terminology coincides with the terminology used in the following two modern textbooks on Riemannian geometry:
J. Jost, Riemannian Geometry and Geometric Analysis, 5th edition, Springer, Berlin, 2008.
S. Novikov and T. Taimanov, Geometric Structures and Fields, Amer. Math. Soc., Providence, Rhode Island.

The Riemann curvature tensor has the following symmetries for all arguments $a, b, u, v \in T\mathcal{H}_{\mathbb{R}}$ and all points $p \in \mathcal{H}_{\mathbb{R}}$:

- $\mathcal{R}_P(a,b,u,v) = -\mathcal{R}_P(b,a,u,v)$,
- $\mathcal{R}_P(a,b,u,v) = -\mathcal{R}_P(a,b,v,u)$,
- $\mathcal{R}_P(a,b,u,v) = \mathcal{R}_P(u,v,a,b)$.

That is, $\mathcal{R}_P(a,b,u,v)$ is antisymmetric in the first and second pairs of indices, and it is symmetric relative to permutations of these pairs of indices.

Components. Introducing the notation for the matrix elements, we write

$$\mathcal{A}_i = (\Gamma_{is}^r) = \begin{pmatrix} \Gamma_{i1}^1 & \Gamma_{i2}^1 \\ \Gamma_{i1}^2 & \Gamma_{i2}^2 \end{pmatrix}, \qquad \mathcal{F}_{jk} = (R_{sjk}^r) = \begin{pmatrix} R_{1jk}^1 & R_{2jk}^1 \\ R_{1jk}^2 & R_{2jk}^2 \end{pmatrix}. \tag{5.90}$$

Here, r (resp. s) is the row (resp. column) index. For the inner product on the tangent space $T_P\mathcal{H}_{\mathbb{R}}$, we have

$$\langle v|w\rangle_P = v^i g_{ij} w^j$$

with $g_{ij}(P) := n(P)^2\delta_{ij}$,and $g^{ij}(P) := n(P)^{-2}\delta_{ij}$ for $i,j = 1,2$. Note that

$$g^{is}g_{sj} = \delta_j^i \qquad \text{for all} \quad i,j = 1,2.$$

Here, we use the notation $\delta_j^i := \delta_{ij}$ (Kronecker symbol). That is, $\delta_{11} = \delta_{22} = 1$ and $\delta_{12} = \delta_{21} = 0$. The equation (5.87) for the geodesics reads as

$$\ddot{x}^k + \dot{x}^i \Gamma_{ij}^k \dot{x}^j = 0, \qquad k = 1,2. \tag{5.91}$$

Finally, setting $R_{ijkl} := g_{ir}R_{jkl}^r$, we get

$$\boxed{\mathcal{R}_P(a,b,c,d) = -R_{ijkl}(P)\cdot a^i b^j u^k v^l.}$$

Thus, the real numbers $R_{ijkl}(P)$ are the (negative) components of the Riemann curvature tensor $\mathcal{R}_P$ at the point P. Moreover, the real numbers $\Gamma_{ij}^k(P)$ are called the Christoffel symbols at the point P. For a smooth real-valued function U on the hyperbolic plane $\mathcal{H}_{\mathbb{R}}$, the Laplacian looks like

$$\Delta U = -g^{ij}\nabla_i\nabla_j U.$$

The Ricci tensor is defined by

$$\mathrm{Ric}_P(u,v) := u^j R_{jl} v^l \qquad \text{for all} \quad u,v \in T_P\mathcal{H}_{\mathbb{R}},$$

where $R_{jl} := g^{ik}R_{ijkl}$. Finally, the scalar curvature is defined by

$$\boxed{R(P) := g^{jl}R_{jl}(P).}$$

Explicitly, we get $\mathcal{F}_{21} = -\mathcal{F}_{12}, \mathcal{F}_{11} \equiv \mathcal{F}_{22} \equiv 0$, and

$$\mathcal{F}_{12} = \begin{pmatrix} 0 & -\varrho_{xx} - \varrho_{yy} \\ \varrho_{xx} + \varrho_{yy} & 0 \end{pmatrix}.$$

Hence $\mathcal{F} = \mathcal{F}_{12}\, dx^1 \wedge dx^2$. Using (5.72), an explicit computation shows that

$$\boxed{R_{ijkl} = K(g_{ik}g_{jl} - g_{jk}g_{il}), \qquad i,j,k,l = 1,2,} \tag{5.92}$$

where K is the Gaussian curvature. Moreover,[52]

$$\boxed{R_{ij} = Kg_{ij}, \qquad R = 2K.}$$

This implies $\mathrm{Ric}_P = K(P)g_P$ for all points P of $\mathcal{H}_{\mathbb{R}}$. A real n-dimensional Riemannian manifold $\mathcal{M}$ is called an *Einstein manifold* iff there is a real constant λ such that

$$\mathrm{Ric}_P = \lambda g_P \qquad \text{for all} \quad P \in \mathcal{M}.$$

Therefore, a real 2-dimensional Riemannian manifold is an Einstein manifold iff it has constant Gaussian curvature (e.g., the 2-dimensional sphere and the hyperbolic plane).[53] By (5.92), from the symmetry relation $g_{ij} = g_{ji}$ it follows that

- $R_{ijkl} = -R_{jikl}$ (antisymmetry with respect to i,j),
- $R_{ijkl} = -R_{ijlk}$ (antisymmetry with respect to the first pair of indices),
- $R_{ijkl} = -R_{ijlk}$ (antisymmetry with respect to the second pair of indices),
- $R_{ijkl} = R_{klij}$ (symmetry between the first pair and the second pair of indices),
- $R_{i[jkl]} = 0$ (the antisymmetrization with respect to the last three indices vanishes – first Bianchi identity). [54]
- $\partial_{[s}R_{ijk]l} = 0$ (antisymmetrization with respect to the first three indices vanishes identically – second Bianchi identity).

Hence $R_{1212} = -R_{2112} = -R_{1221} = R_{2121}$. The remaining components R_{ijkl} vanish identically, since $R_{iikl} = R_{ijkk} = 0$. Therefore, only the component $R_{1212} = K(g_{11}g_{22} - g_{12}g_{21})$ is essential.[55]

The theorema egregium of Gauss. Consider the energetic minimum problem

$$\int_{\tau_0}^{\tau_1} \dot{x}^i(\tau) g_{ij}(P(\tau)) \dot{x}^j(\tau)\, d\tau = \min!$$

with the fixed end points $(x^1(\tau_0), x^2(\tau_0))$ and $(x^1(\tau_1), x^2(\tau_1))$. Every smooth solution of this minimum problem satisfies the Euler–Lagrange equation (5.91) with the Christoffel symbols[56]

$$\Gamma^k_{ij} = \tfrac{1}{2} g^{ks} \left(\partial_i g_{sj} + \partial_j g_{is} - \partial_s g_{ij}\right).$$

Mnemonically, observe the position of the summation index s. In terms of matrix elements, the fundamental structural equation

$$\boxed{\mathcal{F}_{kl} = \partial_k \mathcal{A}_l - \partial_l \mathcal{A}_k + \mathcal{A}_k \mathcal{A}_l - \mathcal{A}_l \mathcal{A}_k, \qquad j,k = 1,2} \tag{5.93}$$

for the Cartan curvature 2-form $\mathcal{F} = \frac{1}{2}\mathcal{F}_{ij} dx^i \wedge dx^j$ reads as

---

[52] Note that $R_{jl} = R_{ijkl}g^{ik} = K(2g_{jl} - \delta^k_l g_{jk}) = Kg_{jl}$.

[53] We refer to A. Besse, Einstein Manifolds, Springer, New York, 1987.

[54] Since $R_{ijkl} = -R_{ijlk}$, we get $R_{i[jkl]} = 2(R_{ijkl} + R_{iklj} + R_{ikjl})$.

[55] The situation completely changes in higher dimensions. Consider a real n-dimensional Riemannian manifold, $n = 2, 3, \ldots$ Then, the Riemann curvature tensor has $n^2(n^2-1)/12$ essential components. In particular, if $n = 4$, then we get 20 essential components. This is the case of Einstein's theory of general relativity. As we will show in Vol. III, the scalar curvature R and its sign are crucial for our universe, but not the 20 components of the Riemann curvature tensor.

[56] We will prove this in Problem 5.7.

$$R^i_{jkl} = \partial_k \Gamma^i_{lj} - \partial_l \Gamma^i_{kj} + \Gamma^i_{ks}\Gamma^s_{lj} - \Gamma^i_{ls}\Gamma^s_{kj}. \tag{5.94}$$

This looks awkwardly compared with the elegant equation (5.93). In particular, we have

$$\boxed{K = \frac{R_{1212}}{\det(g_{ij})}.} \tag{5.95}$$

This is the theorema egregium of Gauss in terms of the Riemann curvature tensor. This theorem tells us that the Gaussian curvature can be computed by knowing the components g_{ij} of the metric tensor and their partial derivatives up to order 2. These quantities can be intrinsically measured on the manifold without using any embedding of the manifold into a higher-dimensional surrounding space.

Cartan's local curvature forms. We define:

- $\omega^k_j := \Gamma^k_{ij}\, dx^i$ (local connection 1-form),
- $\Theta^k := \frac{1}{2}(\Gamma^k_{ij} - \Gamma^k_{ji})\, dx^i \wedge dx^j$ (local torsion 2-form),
- $\Omega^i_j := \frac{1}{2} R^i_{jkl}\, dx^k \wedge dx^l$ (local curvature 2-form).

Since $\Gamma^k_{ij} = \Gamma^k_{ji}$, we get $\Theta^k = 0$, i.e., the connection is torsion-free. For $i, j = 1, 2$, it follows from (5.94) that

- $\Theta^i = \omega^i_s \wedge dx^s = 0$ (torsion-free connection), and
- $\Omega^i_j = d\omega^i_j + \omega^i_s \wedge \omega^s_j$ (structural equation).

Using $dd\omega = 0$ and the product rule $d(\omega \wedge \mu) = d\omega \wedge \mu - \omega \wedge d\mu$ for 1-forms ω and μ, we get the following:

- $d\Theta^i = d\omega^i_s \wedge dx^s = 0$ (first Bianchi identity), and
- $d\Omega^j_i = d\omega^i_s \wedge \omega^s_j - \omega^i_s \wedge d\omega^s_j = 0$ (second Bianchi identity).

This implies

- $R_{[ijk]l} = 0$ (first Bianchi identity), and
- $\partial_{[s}\mathcal{F}_{kl]} = 0$ (second Bianchi identity).

One can show that this is equivalent to the second Bianchi identity given above.[57] In Vol. III we will use Cartan's structural equations in order to compute elegantly the Schwarzschild solution of the Einstein equations in general relativity. This solution describes the gravitational field of both the sun and black holes.

Killing the indices and the construction of invariants. The invariance properties of the expressions above under changing local coordinates will be studied in Vol. III based on a general approach. At this point let us only mention that the theory will be designed in such a way that

Expressions without any free indices are invariants.

For example, $\mathcal{A}, \mathcal{F}, \mathcal{R}, \mathrm{Ric}, R$, and the Gaussian curvature K are invariants, that is, they do not depend on the choice of the local coordinate system. Hence they possess a geometric meaning.

We briefly say that invariants are generated by killing the indices.

[57] In the present 2-dimensional case, the argument is trivial. However, it can be applied to higher dimensions. Then, the argument is crucial.

This is a special case of the index principle to be studied in Vol. III.

Symmetry. Recall that $\varrho(x,y) := \ln n(x,y)$, where $n = n(x,y)$ is the refraction index at the point. The equation (5.87) for the geodesics can be written as

$$\ddot{x}^k + (\dot{x}^1, \dot{x}^2)\Gamma^k(P)\begin{pmatrix}\dot{x}^1\\ \dot{x}^2\end{pmatrix} = 0, \qquad k = 1,2$$

with the symmetric matrices

$$\Gamma^1 := (\Gamma^1_{ij}) = \begin{pmatrix}\varrho_x & \varrho_y\\ \varrho_y & -\varrho_x\end{pmatrix}, \quad \Gamma^2 = (\Gamma^2_{ij}) = \begin{pmatrix}-\varrho_y & \varrho_x\\ \varrho_x & \varrho_y\end{pmatrix}.$$

In addition, the matrices Γ^1, Γ^2 are traceless. Thus, they are contained in the Lie algebra $sl(2,\mathbb{R})$ of the symmetry group $Sym(2,\mathbb{R})$ of the hyperbolic plane. This means that the equations for the geodesics reflect the symmetry of the hyperbolic plane on an infinitesimal level.

Example 1 (Euclidean geometry). Let $n \equiv 1$. Then the covariant partial derivative (resp. the covariant directional derivative) coincides with the classical partial derivative (resp. directional derivative), that is, $\nabla_j = \partial_j, j = 1,2$ and $\frac{D}{d\tau} = \frac{d}{d\tau}$. The Christoffel symbols vanish identically. Hence $\mathcal{A}, \mathcal{F}, \mathcal{R} \equiv 0$, and $K \equiv 0$. That is, all the curvature quantities vanish identically.

Example 2 (hyperbolic geometry). Let $n(x,y) := \frac{1}{y}$ for all $x \in \mathbb{R}, y > 0$. Hence $\varrho(x,y) = -\ln y$, and we have $\mathcal{A} = \mathcal{A}_1 dx + \mathcal{A}_2 dy$ with the connection matrices

$$\mathcal{A}_1 := \begin{pmatrix}0 & -\frac{1}{y}\\ \frac{1}{y} & 0\end{pmatrix}, \quad \mathcal{A}_2 := \begin{pmatrix}-\frac{1}{y} & 0\\ 0 & -\frac{1}{y}\end{pmatrix}.$$

Moreover, we get $\mathcal{F} = \mathcal{F}_{12} dx^1 \wedge dx^2$ with

$$\mathcal{F}_{12} = \begin{pmatrix}R^1_{112} & R^1_{212}\\ R^2_{112} & R^1_{212}\end{pmatrix} = \begin{pmatrix}0 & -\frac{1}{y^2}\\ \frac{1}{y^2} & 0\end{pmatrix}.$$

For the metric tensor, we get

$$g_{11} = g_{22} = \frac{1}{y^2}, \quad g_{12} = g_{21} = 0, \quad g^{11} = g^{22} = y^2, g^{12} = g^{21} = 0.$$

Hence $R_{1212} = g_{1s}R^s_{212} = g_{11}R^1_{212} = -y^{-4}$. By the theorema egregium (5.95),

$$K \equiv -1, \qquad R \equiv -2, \quad R_{kl} \equiv -g_{kl}, \qquad k,l = 1,2.$$

The Einstein equations in general relativity. These fundamental equations read as

$$\boxed{\mathrm{Ric}(g) - \tfrac{1}{2}Rg + \Lambda g = \kappa \mathsf{T}.} \tag{5.96}$$

Here, g is the metric tensor of a 4-dimensional pseudo-Riemannian space-time manifold with signature $(1,-1,-1,-1)$. The Ricci tensor Ric(g) is responsible for gravitation in the universe. The geodesics describe the motion of both celestial bodies

and light rays.[58] Here, $\kappa := 8\pi G/c^4$ where G is Newton's gravitational constant, and c is the velocity of light in a vacuum. Moreover, Λ is the so-called cosmological constant, which acts as a negative pressure if $\Lambda > 0$. This constant is responsible for the accelerated expansion of our universe observed by astronomers. The energy-momentum tensor T describes the distribution of matter and energy in the universe. If $\mathsf{T} \equiv 0$, then we get $\mathrm{Ric}(g) = (\frac{1}{2}R - \Lambda)g$. If g is a solution of this equation, then $R = 4\Lambda$. Hence

$$\mathrm{Ric}(g) = \Lambda g.$$

The Einstein equations will be studied in Vol. III. We refer to Zeidler (1986), Vol. IV, and to Grøn and Hervik, Einstein's Theory of General Relativity: with Modern Applications in Cosmology, Springer, New York, 2007.

Historical remarks. In 1954, the young physicists Yang and Mills tried to generalize the Maxwell equations. They used an idea of Hermann Weyl published in 1929.[59]

- Weyl formulated the Maxwell equations as a gauge theory based on the commutative Lie group $U(1)$.
- The goal of Yang and Mills was to replace the commutative group $U(1)$ by the noncommutative Lie group $SU(2)$.

To this end, they replaced the relation

$$\mathcal{F}_{jk} = \partial_j \mathcal{A}_k - \partial_k \mathcal{A}_j, \qquad j = 1, 2, 3, 4$$

between the electromagnetic field tensor $\{\mathcal{F}_{ij}\}$ and the 4-potential $\{\mathcal{A}_j\}$ by a modified relation of the type

$$\mathcal{F}_{jk} = \partial_j \mathcal{A}_k - \partial_k \mathcal{A}_j + \mathcal{A}_j \mathcal{A}_k - \mathcal{A}_k \mathcal{A}_j, \qquad j, k = 1, 2, 3, 4, \tag{5.98}$$

where the complex 2×2 matrices $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4$ are elements of the Lie algebra $su(2)$ of the symmetry group $SU(2)$. Dianzhou Zhang asked Professor Yang in an interview:[60] An interesting question is whether you understood in 1954 the tremendous importance of your joint paper with Mills on noncommutative gauge theory. Yang answered:

> No. In the 1950s we felt our work was elegant. I realized its importance in the 1960s and its great importance to physics in the 1970s. Its relationship to deep mathematics became only clear to me after 1974.

In the 1960s and the early 1970s, equations of the type (5.98) above were used in order to formulate the *Standard Model* in elementary particle physics. Here, the symmetry group $U(1) \times SU(2) \times SU(3)$ is used.

---

[58] Using components, we have $ds^2 = g_{\mu\nu} dx^\mu dx^\nu$ and

$$R_{\mu\nu} - \tfrac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = \kappa T_{\mu\nu}, \qquad \mu, \nu = 0, 1, 2, 3. \tag{5.97}$$

We sum over equal upper and lower indices from 0 to 3. Multiplying equation (5.97) by $g^{\mu\nu}$, we get $R - 2R + 4\Lambda = \kappa g^{\mu\nu} T_{\mu\nu}$. Hence $R = 4\Lambda - \kappa \operatorname{tr}(\mathsf{T})$.

[59] H. Weyl, Elektron and Gravitation, Z. Phys. **56** (1929), 330–352 (in German). C. Yang, C. and R. Mills, Conservation of isotopic spin and isotopic spin invariance, Phys. Rev. **96** (1954), 191–195.

[60] D. Zhang, C.N. Yang and contemporary mathematics: An interview, Mathematical Intelligencer **15**(4) (1993), 13–21. In 1957, Chen Ning Yang (born 1922) and Tsung-Dao Lee (born 1926) were awarded the Nobel prize in physics for their penetrating investigation of the so-called parity laws which has led to important discoveries regarding the elementary particles.

- *The crucial field tensor* $\{\mathcal{F}_{ij}\}$ *describes all the 12 particles (i.e., the photon, 8 gluons, and 3 vector bosons) which are responsible for the interaction*
- *between the 12 basic particles (i.e., 6 quarks and 6 leptons – the electron, 3 neutrinos, the muon, and the tau) and their antiparticles.*

The typical difficulty of any gauge theory is the fact that the interacting particles are massless at the very beginning. In contrast to this, the three vector bosons observed in nature are quite heavy – their masses equal about 100 proton masses. One needs an additional field – the Higgs field – in order to generate the masses of the vector bosons. This so-called Higgs mechanism will be thoroughly studied in Vol. III.

In the late 1960s, Yang discovered the relation between Yang–Mills theory and Riemannian geometry. He reports in the same interview as quoted above:

> In the late 1960s, I began a new formulation of gauge fields, through the approach of non-integrable phase factors. It happend that one semester, I was teaching general relativity, and I noticed that the formula (5.93) above in gauge theory and the formula (5.94) above in Riemannian geometry are not just similar – they are, in fact *the same* if one makes the right identification of symbols.

Yang continues:

> With an appreciation of the geometrical meaning of gauge theory, I consulted Jim Simons, a distinguished geometer, who was then the chairman of the Mathematics Department at Stony Brook (Long Island, New York). He said gauge theory must be related to connections on fiber bundles. I then tried to understand fiber bundle theory from such books as Steenrod's *The Topology of Fiber Bundles*, Princeton University Press, 1951, but I learned nothing. The language of modern mathematics is too cold and abstract for a physicist.[61]

In 1975, Wu and Yang wrote a paper about global gauge theory. In this paper, they published a quite interesting dictionary about the completely different terminology of mathematicians and physicists concerning the same topic.[62] For example:

$$\begin{aligned}
\text{connection in mathematics} &\Leftrightarrow \text{potential in physics},\\
\text{curvature in mathematics} &\Leftrightarrow \text{field tensor (interaction) in physics},\\
\text{structural equation} &\Leftrightarrow \text{field tensor} - \text{potential relation},\\
\text{change of bundle coordinates} &\Leftrightarrow \text{gauge transformation},\\
\text{structure group} &\Leftrightarrow \text{gauge group}.^{63}
\end{aligned}$$

In a long historical process, mathematicians tried to understand curvature, but physicists studied the forces acting in the universe. Nowadays we know that mathematicians and physicists did the same from an abstract mathematical point of view. Observe that

[61] Based on Élie Cartan's fundamental papers, the final form of modern differential geometry (to be presented in Vol. III) was created by Charles Ehresmann, Les connexions infinitésimales dans un espace fibré differentiable, Colloque de Topologie, Bruxelles, 1950, pp. 29–55. The first monograph on modern differential geometry was written by S. Kobayashi and K. Nomizu, Foundations of Differential Geometry, Vols. 1, 2, Wiley, New York, 1963.

[62] T. Wu and C. Yang, Concept of non-integrable phase factors and global formulation of gauge fields, Phys. Rev. **D12** (1975), 3845–3857.

The Cartan curvature 2-form is more fundamental than the Riemann curvature tensor.

In fact, since $\mathcal{R}_P(a,b,u,v) := -\langle \mathcal{F}(a,b)u|v\rangle_P$ for all $a,b,u,v \in T_P\mathcal{H}_{\mathbb{R}}$, the Riemann curvature tensor only works if the tangent space of the manifold is equipped with the additional structure of a Hilbert space (or an indefinite Hilbert space, as in general relativity). However, the notion of curvature in the sense of Cartan is independent of such an additional structure. In terms of mathematics, the notion of curvature can be introduced without using a metric, one only needs what is called a connection. In terms of physics, this corresponds to the transport of information.

5.11 Ariadne's Thread in Gauge Theory

The prototype of a gauge transformation in physics is given by the formula

$$\psi_+(x,y,z,t) = \mathrm{e}^{\mathrm{i}\alpha(x,y,z,t)}\psi(x,y,z,t), \quad (x,y,z,t) \in \mathbb{R}^4. \tag{5.99}$$

Here $\psi : \mathbb{R}^4 \to \mathbb{C}$ is a complex-valued function depending on both Cartesian space coordinates $(x,y,z) \in \mathbb{R}^3$ and the time coordinate $t \in \mathbb{R}$. The real-valued phase function $\alpha : \mathbb{R}^4 \to \mathbb{R}$ depends on space and time. The phase factor $\mathrm{e}^{i\alpha}$ is contained in the group $U(1)$ of all complex numbers ζ with $|\zeta| = 1$. Linearization at the unit element $\zeta = 1$ of the group $U(1)$ yields

$$\mathrm{e}^{\mathrm{i}\alpha} = 1 + \mathrm{i}\alpha + o(\alpha), \qquad \alpha \to 0.$$

The purely imaginary numbers $\mathrm{i}\alpha$ form the Lie algebra $u(1)$ of the Lie group $U(1)$. From the physical point of view, suppose that the function ψ is a solution of the Schrödinger equation for a quantum particle moving in $\mathbb{R}^3$, which satisfies the normalization condition $\int_{\mathbb{R}^3} |\psi(x,y,z,t)|^2 \, dxdydz = 1$ for all times $t \in \mathbb{R}$. The real number

$$p_t(\mathcal{U}) := \int_{\mathcal{U}} |\psi(x,y,z,t)|^2 \, dxdydz$$

equals the probability of finding the quantum particle in the open subset $\mathcal{U}$ of $\mathbb{R}^3$ at time t. Observe the crucial fact that the probability $p_t(\mathcal{U})$ is invariant under the gauge transformation (5.99). This follows from $|\mathrm{e}^{\mathrm{i}\alpha}| = 1$, and hence $|\psi_+(x,y,z,t)| = |\psi(x,y,z,t)|$.

In terms of physics, roughly speaking, gauge theory describes additional internal degrees of freedom of physical systems which do not affect the physics.

For example, the choice of the potential does not influence the forces in classical Newtonian mechanics.

From the mathematical point of view, gauge theory studies invariants under gauge transformations.

Only such gauge invariants can be observed in physical experiments. Therefore, gauge theory is part of the theory of invariants. In order to explain the basic ideas of gauge theory, let us return to the hyperbolic plane $\mathcal{H}_{\mathbb{R}}$.

5.11.1 Parallel Transport of Physical Information – the Key to Modern Physics

> In modern physics, the Huygens principle is replaced by the parallel transport of physical information.[64]
>
> Folklore

Again set $x^1 = x, x^2 := y$. Consider a smooth function $\psi : \mathcal{H}_{\mathbb{R}} \to \mathbb{R}^2$. We write

$$\psi(x^1, x^1) = \begin{pmatrix} \psi^1(x^1, x^2) \\ \psi^2(x^1, x^2) \end{pmatrix}. \tag{5.100}$$

Intuitively, we regard $\psi(x^1, x^2)$ as the value of a physical field at the point (x^1, x^2) of the hyperbolic plane $\mathcal{H}_{\mathbb{R}}$. Let $GL(2, \mathbb{R})$ denote the Lie group of all real invertible (2×2)-matrices.

The equation of parallel transport. Let

$$C : x^1 = x^1(\tau), \quad x^2 = x^2(\tau), \qquad \tau_0 \le \tau \le \tau_1$$

be a smooth curve on the hyperbolic plane $\mathcal{H}_{\mathbb{R}}$. Set $\gamma(\tau) := (x^1(\tau), x^2(\tau))$. Furthermore, let $\psi(\tau) := \psi(\gamma(\tau))$. By definition, the field ψ is parallel along the curve C iff the differential equation

$$\boxed{\dot{\psi}(\tau) + \dot{x}^i(\tau)\mathcal{A}_i(\gamma(\tau)) \cdot \psi(\tau) = 0} \tag{5.101}$$

is satisfied for all $\tau \in [\tau_0, \tau_1]$. This definition is motivated by (5.57) on page 300.[65] In particular, as a consequence of our definition, the tangent vector field

$$\tau \mapsto (\dot{x}^1(\tau), \dot{x}^2(\tau))$$

of a geodesic C is parallel along the curve C itself. This generalizes the elementary intuitive fact that, in Euclidean geometry, the tangent vector of a straight line C is parallel along the straight line C itself. The operator $\mathit{\Pi}_\tau$ of parallel transport is defined by

$$\mathit{\Pi}_\tau \psi_0 := \psi(\tau), \qquad \tau \ge 0,$$

where $\tau \mapsto \psi(\tau)$ is the unique solution of the differential equation (5.101) with the initial condition $\psi(\tau_0) := \psi_0$.

Parallel transport and covariant directional derivative. The following theorem shows that the covariant directional derivative along the curve C is related to parallel transport in a quite natural way.

Theorem 5.13 *For any $\tau \in]\tau_0, \tau_1[$, we have*

$$\frac{D\psi(\tau)}{d\tau} = \frac{d\mathit{\Pi}_{-\sigma}\psi(\tau + \sigma)}{d\sigma}_{|\sigma=0}. \tag{5.102}$$

Proof. We have to distinguish between

[64] The importance of parallel transport in differential geometry was emphasized in a classical paper by T. Levi-Civita, Nozione di parallelismo in una varieta qualunque e consequente specificazione geometrica della curvature Riemanniana, Rend. Palermo **42** (1917), 73–205 (in Italian).

[65] Using components, equation (5.101) reads as $\dot{\psi}^k + \dot{x}^i \mathit{\Gamma}^k_{ij} \psi^j = 0, k = 1, 2$.

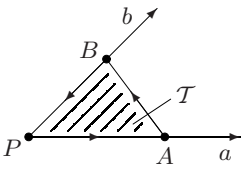


Fig. 5.24. Parallel transport and curvature

- the given physical field $\psi(\tau) = \psi(x^1(\tau), x^2(\tau))$ and
- the parallel transport $\psi_* = \psi_*(\tau)$ which satisfies equation (5.101).

Using (5.101), Taylor expansion at the point $\tau + \sigma$ yields

$$\psi_*(\tau+\sigma) - \psi_*(\tau) = -\dot{x}^i(\tau+\sigma)\mathcal{A}_i(x^1(\tau+\sigma), x^2(\tau+\sigma))\ \psi_*(\tau+\sigma) + o(\sigma),$$

as $\sigma \to 0$. Assuming that $\psi_*(\tau+\sigma) = \psi(\tau+\sigma)$, we claim that

$$\frac{D\psi(\tau)}{d\tau} = \lim_{\sigma\to 0} \frac{\psi_*(\tau) - \psi(\tau)}{\sigma}.$$

That is, we compare

- the value $\psi(\tau)$ of the given physical field with
- the value obtained from $\psi(\tau+\sigma)$ by backward parallel transport.

The trick is to use the decomposition

$$\frac{\psi_*(\tau) - \psi(\tau)}{\sigma} = \frac{\psi_*(\tau) - \psi_*(\tau+\sigma)}{\sigma} + \frac{\psi(\tau+\sigma) - \psi(\tau)}{\sigma}.$$

Letting $\sigma \to 0$, we get $\dot{x}^i(\tau)\mathcal{A}_i(x^1(\tau), x^2(\tau))\ \psi(\tau) + \frac{d\psi(\tau)}{d\tau}$. This coincides with $\frac{D\psi(\tau)}{d\tau}$ defined above. □

Parallel transport and curvature. Consider the triangle $\mathcal{T}$ pictured in Fig. 5.24. The triangle is spanned by the vectors τa and τb with $\tau > 0$. We are given the field $\psi(P)$ at the point P in the hyperbolic plane $\mathcal{H}_{\mathbb{R}}$. We carry out a clockwise parallel transport of the field $\psi(P)$ along the boundary $\partial\mathcal{T} = \overline{PABP}$ of the triangle $\mathcal{T}$. We are starting at the initial point P and we are finishing at the final point P after turning once around the triangle. The result of this parallel transport is denoted by $\Pi_{\partial\mathcal{T}}\psi(P)$. The following theorem tells us that, roughly speaking, curvature measures the difference $\psi(P) - \Pi_{\partial\mathcal{T}}\psi(P)$. This means that

Curvature measures the path-dependence of the parallel transport.

In particular, if this difference vanishes, then also the curvature vanishes at the point P. Let $\mathrm{meas}(\mathcal{T}) = \frac{1}{2}\det(\tau a, \tau b)$ be the Euclidean area measure of the triangle $\mathcal{T}$. Now we contract the triangle $\mathcal{T}$ to the point P by letting $\tau \to 0$.

Theorem 5.14 *We have*

$$\mathcal{F}_P(a,b)\psi(P) = \det(a,b)\ \lim_{\mathcal{T}\to P} \frac{\psi(P) - \Pi_{\partial\mathcal{T}}\psi(P)}{\mathrm{meas}(\mathcal{T})}.$$

The proof based on the Poincaré–Stokes integral theorem will be given in Problem 5.8.

Cartan's covariant differential. For the physical field ψ, the classical Cartan calculus yields the following:

- $d\psi = \partial_k \psi \cdot dx^k$,
- $d^2\psi = d(\partial_j \psi) \wedge dx^k = \partial_j \partial_k \psi \cdot dx^j \wedge dx^k \equiv 0$ (Poincaré identity).

Now we replace the partial derivative ∂_k by the covariant partial derivative ∇_k. This way, we get the following definitions:

- $D\psi := \nabla_k \psi \cdot dx^k$ (covariant differential),
- $D^2\psi := D(\nabla_k \psi) \wedge dx^k = \nabla_j \nabla_k \psi \cdot dx^j \wedge dx^k$,
- $D^3\psi := D(\nabla_j \nabla_k \psi) \wedge dx^j \wedge dx^k = \nabla_i \nabla_j \nabla_k \psi \cdot dx^i \wedge dx^j \wedge dx^k \equiv 0$ (Bianchi identity).

Let us discuss this. For any real linear space X, the space $L(X, X)$ of linear operators $A : X \to X$ becomes a real Lie algebra with respect to the Lie bracket $[A, B] := AB - BA$ for all $A, B \in L(X, X)$. In particular, by cyclic permutation, we have the Jacobi identity

$$[\,[A, B], C] + [\,[B, C], A] + [\,[C, A], B] = 0.$$

Explicitly, $(AB - BA)C - C(AB - BA) + (BC - CB)A - A(BC - CB) + (CA - AC)B - B(CA - AC) = 0$. Now we apply this to the covariant partial differential operator ∇_k. Hence

$$\Big([\,[\nabla_i, \nabla_j], \nabla_k] + [\,[\nabla_j, \nabla_k], \nabla_i] + [\,[\nabla_k, \nabla_i], \nabla_j]\Big)\psi = 0. \tag{5.103}$$

Since $dx^j \wedge dx^j = -dx^k \wedge dx^j$, we get

$$D^2\psi = \tfrac{1}{2}(\nabla_j \nabla_k - \nabla_k \nabla_l)\psi \cdot dx^j \wedge dx^k = \tfrac{1}{2}[\nabla_j, \nabla_k]\ \psi \cdot dx^j \wedge dx^k.$$

Therefore, we have the structural equation[66]

$$\boxed{\mathcal{F}\psi = D^2\psi,}$$

where $\mathcal{F}\psi = \frac{1}{2}\mathcal{F}_{jk}\psi \cdot\ dx^j \wedge dx^k$ with

$$\mathcal{F}_{jk}\psi = [\nabla_j, \nabla_k]\psi, \qquad j, k = 1, 2.$$

This tells us that the Cartan curvature 2-form measures the structure of the Lie algebra generated by the covariant partial derivatives. From (5.103) we obtain the Bianchi identity $D^3\psi = 0$, which is equivalent to the relation

$$\boxed{\nabla_{[r}\mathcal{F}_{kl]} = 0.} \tag{5.104}$$

This means that antisymmetrization with respect to the indices r, k, l always vanishes identically.[67] Alternatively, $D\psi, D^2\psi$, and $D^3\psi = 0$ can be written as follows:

- $D\psi = d\psi + \mathcal{A}\psi$ (covariante differential),
- $\mathcal{F} = d\mathcal{A} + \mathcal{A} \wedge \mathcal{A}$ (structural equation),
- $d\mathcal{F} + \mathcal{A} \wedge \mathcal{F} - \mathcal{F} \wedge \mathcal{A} = 0$ (Bianchi identity).

For a detailed discussion, we refer to Vol. III. There we will also show that the velocity vector fields on a manifold form a Lie algebra generated by the Lie derivative, and that there exists a close relation between this Lie algebra and curvature.

---

[66] Mnemonically, we write $\mathcal{F} = D^2$.

[67] For 2-dimensional manifolds, the Bianchi identity is trivial, since 3-forms vanish identically on a 2-dimensional manifold. However, the argument above also applies to n-dimensional manifolds, $n = 3, 4, \ldots$ Then we we get nontrivial results (see Vol. III).

5.11.2 The Phase Equation and Fiber Bundles

> In a natural way, the global mathematical description of physical fields is based on the language of bundles.
>
> Folklore

The main trick due to Élie Cartan is to reduce the parallel transport of a physical field to a dynamical system in a principal fiber bundle (Cartan's method of moving frame).[68] To this end, we will replace the differential equation of parallel transport by the phase equation (5.106) below. First let us introduce some bundles.

Tangent bundle. By definition, the tangent bundle $T\mathcal{H}_{\mathbb{R}}$ of the hyperbolic plane $\mathcal{H}_{\mathbb{R}}$ consists of all the pairs (P, v), where P is a point of the hyperbolic plane, and v is a tangent vector at the point P. That is,

$$T\mathcal{H}_{\mathbb{R}} := \{(P, v) : P \in \mathcal{H}_{\mathbb{R}},\ v \in T_P\mathcal{H}_{\mathbb{R}}\}.$$

This is a real 4-dimensional manifold.

Cotangent bundle. By definition, the cotangent space $T_P^*\mathcal{H}\mathbb{R}$ of the hyperbolic plane at the point P is the dual space to the tangent space $T_P\mathcal{H}_{\mathbb{R}}$. That is, the elements of $T_P^*\mathcal{H}_{\mathbb{R}}$ are linear functionals of the form $\omega : T_P\mathcal{H}_{\mathbb{R}} \to \mathbb{R}$.

Principal fiber bundle. The product set

$$\mathcal{P} := \mathcal{H}_{\mathbb{R}} \times GL(2, \mathbb{R}) = \{(P, G) : P \in \mathcal{H}_{\mathbb{R}},\ G \in GL(2, \mathbb{R})\}$$

is called the (trivial) principal fiber bundle over $\mathcal{H}_{\mathbb{R}}$ with the group $GL(2, \mathbb{R})$ as typical fiber.

The fundamental phase equation. Let $C : P = P(\tau), \tau_0 \le \tau \le \tau_1$ be a smooth curve on the hyperbolic plane. Here, the point P corresponds to the coordinates (x^1, x^2). The main idea is to replace the differential equation

$$\dot{\psi}(\tau) + \dot{x}^i(\tau)\mathcal{A}_i(P(\tau)) \cdot \psi(\tau) = 0, \quad \tau_0 \le \tau \le \tau_1, \quad \psi(\tau_0) = \psi_0 \tag{5.105}$$

of parallel transport along the curve by the phase equation

$$\boxed{\dot{G}(\tau) + \dot{x}^i(\tau)\mathcal{A}_i(P(\tau))\, G(\tau) = 0, \qquad \tau_0 \le \tau \le \tau_1} \tag{5.106}$$

along the curve C.

Proposition 5.15 *If $G = G(\tau)$ is a smooth solution of the phase equation (5.106), then the function*

$$\psi(\tau) := G(\tau)\psi_0, \qquad \tau_0 \le \tau \le \tau_1$$

is a solution of the equation (5.105) of parallel transport.

The proof follows by differentiation. This way the parallel transport along the curve C corresponds to a curve

$$\boxed{(P, G) = (P(\tau), G(\tau)), \qquad \tau_0 \le \tau \le \tau_1}$$

on the principal fiber bundle $\mathcal{P}$.

Associated vector bundle. The (trivial) bundle

$$\mathcal{V} := \{(P, \psi) :\ P \in \mathcal{H}_{\mathbb{R}},\ \psi \in \mathbb{R}^2\}$$

is called the associated vector bundle to the principal fiber bundle $\mathcal{P}$ with the typical fiber $\mathbb{R}^2$. The parallel transport $\psi = \psi(\tau)$ of the physical field corresponds to the curve

$$(P, \psi) = (P(\tau), \psi(\tau)), \qquad \tau_0 \le \tau \le \tau_1$$

in the vector bundle $\mathcal{V}$.

[68] The intuitive background will be studied in Vol. III.

5.11.3 Gauge Transformations and Gauge-Invariant Differential Forms

Gauge transformations. Let P denote a point in the hyperbolic plane $\mathcal{H}_{\mathbb{R}}$ with the coordinates (x^1, x^1). The transformation

$$\boxed{\psi_+(P) = \mathcal{G}(P)\psi(P)} \tag{5.107}$$

is called a gauge transformation. More precisely, we assume that the transformation matrix $\mathcal{G}(P)$ is contained in the group $GL(2,\mathbb{R})$ for all points P in $\mathcal{H}_{\mathbb{R}}$ and that the map $\mathcal{G} : \mathcal{H}_{\mathbb{R}} \to GL(2,\mathbb{R})$ is smooth. We add the transformation law

$$(P, G) \mapsto (P, G_+) \qquad G_+ = \mathcal{G}(P)G$$

for the coordinates of the principal fiber bundle $\mathcal{P} = \mathcal{H}_{\mathbb{R}} \times GL(2,\mathbb{R})$. The standard example of the gauge transformation of a velocity field will be investigated below. We postulate:

Parallel transport is compatible with gauge transformations.

In order to reach this goal, we introduce the following transformed quantities for all $i, j = 1, 2$:

- $\mathcal{A}_j^+ := \mathcal{G}\mathcal{A}_j\mathcal{G}^{-1} - \mathcal{G}^{-1}\partial_j\mathcal{G}$ (connection matrices),[69]
- $\nabla_j^+ := \partial_j + \mathcal{A}_j^+$ (covariant partial derivative),
- $\frac{D^+}{d\tau} := \frac{d}{d\tau} + \dot{x}^i\mathcal{A}_i^+$ (covariant directional derivative),
- $\mathcal{A}^+ = \mathcal{A}_i^+ dx^i$ (connection 1-form),
- $\mathcal{F}_{ij}^+ := \nabla_i^+\nabla_j^+ - \nabla_j^+\nabla_i^+$ (curvature matrices),
- $\mathcal{F}^+ := \frac{1}{2}\mathcal{F}_{ij}^+ \, dx^i \wedge dx^j$ (curvature 2-form).

Theorem 5.16 *If $\tau \mapsto \psi(\tau)$ is parallel along the curve C, then so is ψ^+ with respect to the transformed connection matrices.*

Proof. From $I = \mathcal{G}\mathcal{G}^{-1}$, we get $0 = \partial_j(\mathcal{G}\mathcal{G}^{-1}) = \partial_j\mathcal{G} \cdot \mathcal{G}^{-1} + \mathcal{G}\partial_j(\mathcal{G}^{-1})$. Hence

$$\partial_j\mathcal{G} \cdot \mathcal{G}^{-1} = -\mathcal{G}\partial_j(\mathcal{G}^{-1}) = \mathcal{G} \cdot \mathcal{G}^{-2}\partial_j\mathcal{G} = \mathcal{G}^{-1}\partial_j\mathcal{G}.$$

Therefore,

$$\mathcal{A}_j^+ = \mathcal{G}\mathcal{A}_j\mathcal{G}^{-1} - \partial_j\mathcal{G} \cdot \mathcal{G}^{-1}. \tag{5.108}$$

Differentiating $\psi^+(\tau) = \mathcal{G}(P(\tau)) \cdot \psi(\tau)$ with respect to the curve parameter τ, we get

$$\dot{\psi}^+ = (\dot{x}^i\partial_i\mathcal{G})\psi + \mathcal{G}\dot{\psi}.$$

Substituting $\psi = \mathcal{G}^{-1}\psi^+$ and $\dot{\psi} = -\dot{x}^i\mathcal{A}_i\psi$, we obtain $\dot{\psi}^+ + \dot{x}^i\mathcal{A}_i^+\psi^+ = 0$. □

We have the following transformation laws:

- $\nabla_j^+\psi^+(P) = \mathcal{G}(P) \cdot \nabla_j\psi(P)$,
- $\frac{D^+\psi^+(P)}{d\tau} = \mathcal{G}(P)\frac{D\psi(P)}{d\tau}$,

[69] Explicitly, $\mathcal{A}_j^+(P) := \mathcal{G}(P)\mathcal{A}_j(P)\mathcal{G}(P)^{-1} - \mathcal{G}(P)^{-1}\partial_j\mathcal{G}(P)$.

- $\mathcal{A}_P^+ = \mathcal{G}(P)\mathcal{A}_P\mathcal{G}(P)^{-1}$,
- $\mathcal{F}_{ij}^+(P) = \mathcal{G}(P)\mathcal{F}_{ij}(P)\mathcal{G}(P)^{-1}$,
- $\mathcal{F}_P^+ = \mathcal{G}(P)\mathcal{F}_P\mathcal{G}(P)^{-1}$.

This shows the crucial fact that:

Both the covariant partial derivative and the covariant directional derivative are transformed like the physical field ψ itself.

For the covariant partial derivative, by (5.108), this follows from

$$\begin{aligned}\nabla_j^+\psi^+ &= \nabla_j^+(\mathcal{G}\psi) = \partial_j(\mathcal{G}\psi) + \mathcal{A}_j^+(\mathcal{G}\psi)\\ &= \partial_j\mathcal{G}\cdot\psi + \mathcal{G}\partial_j\psi + \mathcal{G}\mathcal{A}_j\psi - \partial_j\mathcal{G}\cdot\psi = \mathcal{G}(\partial_j + \mathcal{A}_j)\psi = \mathcal{G}\nabla_j\psi.\end{aligned}$$

A similar argument yields the transformation law for the covariant directional derivative. For the curvature components $\mathcal{F}_{ij}$, observe that

$$\nabla_i^+\nabla_j^+\psi^+ = \nabla_i^+(\mathcal{G}\nabla_j\psi) = \mathcal{G}(\nabla_i\nabla_j)\psi = \mathcal{G}(\nabla_i\nabla_j)\mathcal{G}^{-1}\psi^+.$$

Similarly,

$$\mathcal{F}_{ij}^+\psi^+ = (\nabla_i^+\nabla_j^+ - \nabla_j^+\nabla_i^+)\psi^+ = \mathcal{G}(\nabla_i\nabla_j - \nabla_j\nabla_i)\mathcal{G}^{-1}\psi^+ = (\mathcal{G}\mathcal{F}_{ij}\mathcal{G}^{-1})\psi^+.$$

The Maurer–Cartan form. Let us consider the principal fiber bundle $\mathcal{P} := \mathcal{H}_{\mathbb{R}} \times GL(2,\mathbb{R})$. The 1-form

$$\mathsf{M}_{(P,G)} := G^{-1}dG \qquad \text{for all} \quad (P,G) \in \mathcal{P}$$

is called the Maurer–Cartan form on $\mathcal{P}$.[70] Explicitly, $\mathsf{M}_{(P,G)} = G^{-1}\partial_j G\, dx^j$. Consider a smooth curve $C : x^j = x^j(\tau), \tau_0 \le \tau \le \tau_1, j = 1,2$, on the base manifold $\mathcal{H}_{\mathbb{R}}$. Define $G(\tau) := G(x^1(\tau), x^2(\tau))$. Then $\dot{G}(\tau) = \dot{x}^j(\tau)\partial_j G(P(\tau))$. Hence, for all $\tau \in [\tau_0, \tau_1]$,

$$\mathsf{M}_{(P(\tau),G(\tau))}(\dot{P}(\tau), \dot{G}(\tau)) = G(\tau)^{-1}\dot{x}^j(\tau)\partial_j G(P(\tau)) = G^{-1}(\tau)\dot{G}(\tau).$$

Fix $\mathcal{G} \in GL(2,\mathbb{R})$, and set $f_{\mathcal{G}}(P,G) := (P, \mathcal{G}G)$ for all $(P,G) \in \mathcal{P}$.

Proposition 5.17 *For any $\mathcal{G} \in GL(2,\mathbb{R})$, the Maurer–Cartan form M is invariant under the gauge transformation $f_{\mathcal{G}} : \mathcal{P} \to \mathcal{P}$.*

Proof. We have to show that $f_{\mathcal{G}}^*\mathsf{M} = \mathsf{M}$. In fact, the linearization

$$f'_{\mathcal{G}}(P,G) : T_{(P,G)}\mathcal{P} \to T_{(P,\mathcal{G}G)}\mathcal{P}$$

is given by $f'_{\mathcal{G}}(v,V) = (v, \mathcal{G}V)$ for all $(v,V) \in T_P\mathcal{H}_{\mathbb{R}} \times T_G GL(2,\mathbb{R})$.[71] Hence

$$(f_{\mathcal{G}}^*\mathsf{M})_G(v,V) = \mathsf{M}_{(P,\mathcal{G}G)}(v, \mathcal{G}V) = (\mathcal{G}G)^{-1}\mathcal{G}V = G^{-1}V = \mathsf{M}_G(v,V). \qquad \square$$

Proposition 5.18 *$d\mathsf{M} + \mathsf{M} \wedge \mathsf{M} = 0$ on $\mathcal{P}$.*

[70] Maurer (1859–1927) worked on the theory of invariants at the University of Tübingen (Germany).

[71] To see this, note that differentiation of $H(\tau) := \mathcal{G}G(\tau)$ yields $\dot{H}(0) = \mathcal{G}\dot{G}(0)$.

This is called the Maurer–Cartan structural equation of the Lie group $GL(2,\mathbb{R})$. As we will show in Vol. III, intuitively, this equation tells us that the curvature of the Lie group $GL(2,\mathbb{R})$ (with respect to left-translation on the group) vanishes identically, and M is the corresponding connection 1-form. This is true for all Lie groups.

Proof. It follows from $GG^{-1} = I$ that $dG \cdot G^{-1} + GdG^{-1} = 0$. Hence

$$d(G^{-1}dG) = dG^{-1} \wedge dG = (dG^{-1})G \wedge G^{-1}dG = -G^{-1}dG \wedge G^{-1}dG. \qquad \square$$

Gauge-invariant differential forms. For any point (P,G) of the principal fiber bundle $\mathcal{P} = \mathcal{H}_{\mathbb{R}} \times GL(2,\mathbb{R})$, we define:

- $\mathsf{A}_{(P,G)} := G^{-1}dG + G^{-1}\mathcal{A}_P G$ (global connection 1-form on $\mathcal{P}$).

Proposition 5.19 *The global connection form* A *on the principal fiber bundle* $\mathcal{P}$ *is invariant under the gauge transformation* $(P,G) \mapsto (P,G_+)$. *Explicitly, this means that,* $\mathsf{A}^+_{(P,G_+)} = \mathsf{A}_{(P,G)}$.

Proof. By Prop. 5.17, $G_+^{-1}dG_+ = G^{-1}dG$. Hence

$$\begin{aligned}\mathsf{A}^+_{(P,G_+)} &= G_+^{-1}dG^+ + G_+^{-1}\mathcal{A}_P^+ G_+ \\ &= G^{-1}dG + (\mathcal{G}G)^{-1}\mathcal{G}\mathcal{A}_P\mathcal{G}^{-1}(\mathcal{G}G) = G^{-1}dG + G^{-1}\mathcal{A}_P G = \mathsf{A}_{(P,G)}.\end{aligned}$$

$\square$

The global phase equation. Using the global connection 1-form A, the equation

$$\boxed{\mathsf{A}_{(P(\tau),G(\tau))}(\dot{P}(\tau),\dot{G}(\tau)) = 0, \qquad \tau_0 \le \tau \le \tau_1} \tag{5.109}$$

is equivalent to the phase equation (5.106). In fact, equation (5.109) means that $G(\tau)^{-1}\dot{G}(\tau) + G(\tau)^{-1}\mathcal{A}_{P(\tau)}(\dot{P}(\tau))G(\tau) = 0$. This is equivalent to

$$\dot{G}(\tau) + \mathcal{A}_{P(\tau)}(\dot{P}(\tau))G(\tau) = 0.$$

Explicitly, $\dot{G}(\tau) + \dot{x}^j(\tau)\mathcal{A}_j(P(\tau))G(\tau) = 0$.

Cartan's global structural equation. We define the global curvature form F on the principal fiber bundle $\mathcal{P} = \mathcal{H}_{\mathbb{R}} \times GL(2,\mathbb{R})$ by setting

$$\boxed{\mathsf{F} := d\mathsf{A} + \mathsf{A} \wedge \mathsf{A}.} \tag{5.110}$$

Sections of the principal fiber bundle and gauge transformations. Let $s : \mathcal{H}_{\mathbb{R}} \to \mathcal{P}$ be a section of the principal fiber bundle $\mathcal{P} := \mathcal{H}_{\mathbb{R}} \times GL(2,\mathbb{R})$. By definition, this is a smooth map of the form

$$s(P) := (P, G(P)) \qquad \text{for all} \quad P \in \mathcal{H}_{\mathbb{R}}.$$

In Vol. III, we will develop a calculus for differential forms on principal fiber bundles with values in a Lie algebra. In this setting, we will show in Vol. III that

$$\boxed{s^*\mathsf{F} = \mathcal{F}^+.}$$

Explicitly, $(s^*\mathsf{F})_P = G_+^{-1}(P)\mathcal{F}_P G_+$ where $G_+(P) = G(P)G$. Thus, the pull-back $s^*\mathsf{F}$ of the global curvature form F is precisely the gauge transformation of the Cartan curvature 2-form $\mathcal{F}$. Similarly, the pull-back $s^*\mathsf{A}$ of the global connection 1-form A yields the gauge-transformation of the phase equation which implies the equation of parallel transport for physical fields. Summarizing, we get the following:

The global curvature form F (*defined on the principal fiber bundle* $\mathcal{P} = \mathcal{H}_{\mathbb{R}} \times GL(2,\mathbb{R})$ *with values in the Lie algebra* $gl(2,\mathbb{R})$) *represents an invariant mathematical object which carries all the information on the gauge transformations of the Cartan curvature 2-form on the base manifold* $\mathcal{H}_{\mathbb{R}}$.

Standard example (velocity fields). Let

$$x_+^j = \chi^j(x^1, x^2), \qquad j = 1, 2 \tag{5.111}$$

be a diffeomorphism from the hyperbolic plane $\mathcal{H}_{\mathbb{R}}$ onto itself. This represents a change of coordinates on the hyperbolic plane. Regard the parameter τ as time. In terms of physics, the curve

$$C: \quad x^i = x^i(\tau), \qquad \tau_0 \le \tau \le \tau_1, \;\; i = 1, 2$$

describes the motion of a point on the hyperbolic plane with the velocity vector

$$v(\tau) := \begin{pmatrix} \dot{x}^1(\tau) \\ \dot{x}^2(\tau) \end{pmatrix}$$

at time τ. After the coordinate change (5.111), the curve looks like

$$C: \;\; x_+^i(\tau) = \chi^i(x^1(\tau), x^2(\tau)), \qquad \tau_0 \le \tau \le \tau_1, \;\; i = 1, 2.$$

Differentiation with respect to time yields the transformation law of the velocities:

$$\begin{pmatrix} \dot{x}_+^1(\tau) \\ \dot{x}_+^2(\tau) \end{pmatrix} = \begin{pmatrix} \partial_1 \chi^1(P(\tau)) & \partial_2 \chi^1(P(\tau)) \\ \partial_1 \chi^2(P(\tau)) & \partial_2 \chi^2(P(\tau)) \end{pmatrix} \begin{pmatrix} \dot{x}^1(\tau) \\ \dot{x}^2(\tau) \end{pmatrix}.$$

This is a gauge transformation of the form $v_+ = G(P(\tau))\, v$. Observe that, in this example, a change of coordinates generates a transformation law for the velocity vectors in the tangent space, which represents a gauge transformation.[72]

5.11.4 Perspectives

The approach above was chosen in order to display the historical development from Gauss' theorema egregium to Cartan's structural equation. In modern differential geometry, the axiomatic presentation reverses the historical order.

- The starting point is the fundamental phase equation, as a differential equation on a principal fiber bundle $\mathcal{P}$.
- More precisely, one starts with a velocity field on $\mathcal{P}$ which splits into horizontal and vertical velocity vectors, with respect to the fibers.
- This velocity vector field generates the dynamical system of parallel transport. The corresponding ordinary differential equation coincides with the equation of parallel transport, which is based on the connection 1-form A on $\mathcal{P}$.
- Projection onto horizontal vector fields is used in order to replace the differential $d\omega$ by the covariant differential $D\omega$ of differential forms ω on the principal fiber bundle $\mathcal{P}$.

[72] The relation to the classical tensor calculus will be studied in Vol. III. We refer to J. Schouten, Ricci Calculus, Springer, Berlin, 1954.

(a) $g = 1$ (b) $g = 2$

Fig. 5.25. Oriented surface of genus g without boundary

- This yields the curvature form

$$\boxed{\mathsf{F} := D\mathsf{A},} \tag{5.112}$$

and the Bianchi identity $D\mathsf{F} = 0$. It turns out that (5.112) is equivalent to Cartan's structural equation (5.110).

The extremely elegant formula (5.112) generalizes both the theorema egregium of Gauss and the relation between the electromagnetic field tensor F and the 4-potential A in Maxwell's theory of electromagnetism. Moreover, formula (5.112) lies at the heart of both the Standard Model in elementary physics and Einstein's theory of general relativity. In terms of mathematics, this approach effectively describes all kinds of differential geometries:

- The typical Lie group of the principal fiber bundle $\mathcal{P}$ is the symmetry group of the geometry under consideration.
- Physicists are interested in the study of partial differential equations for physical fields. In this general setting, physical fields ψ are sections of a vector bundle $\mathcal{V}$ associated to the principal fiber bundle $\mathcal{P}$.
- There exists a natural way of transplanting the parallel transport from $\mathcal{P}$ to $\mathcal{V}$.
- The parallel transport on $\mathcal{V}$ generates the corresponding covariant differentiation for physical fields, which is used in order to formulate the basic partial differential equations for the physical field under consideration.

The point is that one has to replace product bundles by general fiber bundles. Observe that:

> *General fiber bundles are obtained by gluing local product bundles together, with the aid of a cocycle.*

In particular, the choice of arbitrary Lie groups allows us to take all kind of symmetries into account. In 1872 Felix Klein (1849–1925) formulated his Erlangen program:

> *Geometry is the invariant theory of transformation groups (symmetry groups).*

Sophus Lie (1842–1899), Élie Cartan (1859–1951), and their successors realized this program in differential geometry by investigating Lie groups and the invariant calculus of differential forms. This is a fascinating chapter in the history of mathematics and physics. We will thoroughly study this in Vol. III. We also refer to the standard textbook by S. Kobayashi and K. Nomizu, Foundations of Differential Geometry, Vols. 1, 2, Wiley, New York, 1963.

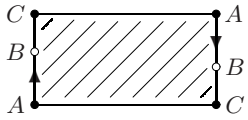


Fig. 5.26. Möbius strip

5.12 Classification of Two-Dimensional Compact Manifolds

> The notion of manifold is of fundamental importance for both mathematics and physics. There arises the problem of classifying manifolds in terms of topology. This has been one of the most important research topics in topology in the last 150 years.
>
> Folklore

In 1863 Möbius classified the compact orientable 2-manifolds.[73] Some examples are pictured in Fig. 5.25.[74] In 1865 Möbius published a strange surface called the Möbius strip nowadays. This twisted surface is obtained from a rectangle by gluing together two opposite sides in a twisted manner (Fig. 5.26). If we walk along this surface, then we reach both sides of the rectangle. Therefore, the Möbius strip is a one-sided surface, and hence it is not possible to define an orientation. Intuitively, a surface is oriented iff the movement of a small oriented circle along a closed curve never changes the orientation of the circle. The main theorem on classical topological surface theory reads as follows:[75]

> *The compact 2-manifolds $\mathcal{M}$ and $\mathcal{N}$ (without or with boundary) are homeomorphic iff the following three conditions hold:*
> (i) *$\mathcal{M}$ and $\mathcal{N}$ have the same genus g,*
> (ii) *$\mathcal{M}$ and $\mathcal{N}$ have the same number of boundaries, and*
> (iii) *both $\mathcal{M}$ and $\mathcal{N}$ are either orientable or non-orientable.*

The genus attains the values $g = 0, 1, 2, \ldots$ Let us first consider a few typical examples. Recall that the Euler characteristic of the 2-manifold $\mathcal{M}$ is given by

$$\chi(\mathcal{M}) = \beta_0 - \beta_1 + \beta_2,$$

[73] Möbius (1790–1868) was director of the observatory in Leipzig from 1820 until his death. Gauss recommended him for this position.

[74] Let $n = 1, 2, \ldots$ By definition, an n-manifold is a real n-dimensional arcwise connected manifold without boundary. Similarly, an n-manifold with boundary is a real n-dimensional arcwise connected manifold with boundary (see Sects. 5.4 and 10.4.2 of Vol. I). An n-dimensional sphere $\mathbb{S}^n_r$ of radius $r > 0$ is described by the equation $x_1^2 + \ldots + x_{n+1}^2 = r^2$ in $\mathbb{R}^{n+1}$. The Betti numbers of an n-sphere are given by $\beta_0 = \beta_n = 1$ and $\beta_k = 0$ if $k = 1, 2, \ldots, n-1$. This yields the Euler characteristic $\chi(\mathbb{S}^n_r) = \sum_{k=0}^{n} (-1)^k \beta_k = 1 + (-1)^n$.

[75] In 1907 the first complete proof of this theorem was given by M. Dehn and P. Heegard, Analysis Situs (Topology), Enzyklopädie der mathematischen Wissenschaften, Vol. III, Nr. 3, pp. 153–220, Teubner, Leipzig, 1907 (in German). We refer to the classic textbooks by Levi (1929) and Rinow (1975), Sect. XII.2, and to the nicely written textbook by Kinsey (1993).

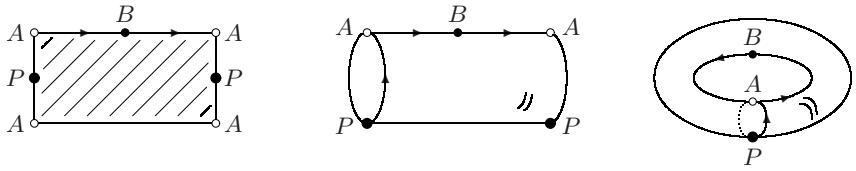


Fig. 5.27. The torus $\mathbb{T}^2$

where $\beta_0, \beta_1, \beta_2$ are the Betti numbers of $\mathcal{M}$. Since the 2-manifold $\mathcal{M}$ is arcwise connected, we always have $\beta_0 = 1$. Moreover, we have $\beta_2 = 1$ (resp. $\beta_2 = 0$) iff $\mathcal{M}$ is orientable (resp. non-orientable).

- The 2-sphere $\mathbb{S}^2_r$ of radius $r > 0$: This surface has the genus $g = 0$, the Betti numbers
$$\beta_0 = \beta_2 = 1, \ \ \beta_1 = 0$$
and the Euler characteristic $\chi(\mathbb{S}^2_r) = 2$. The fundamental group is trivial, $\pi_1(\mathbb{S}^2_r) = 0$, that is, the sphere $\mathbb{S}^2_r$ is simply connected.
- The 2-dimensional torus $\mathbb{T}^2$: This surface can be obtained by identifying the opposite boundary points of a rectangle (Fig. 5.27). Alternatively, the torus is homeomorphic to a sphere with one handle attached to it (Fig. 5.25(a)). This surface has the genus $g = 1$, the Betti numbers
$$\beta_0 = \beta_2 = 2, \ \ \beta_1 = 2,$$
and the Euler characteristic $\chi(\mathbb{T}^2) = 0$. For the additive fundamental group, we get
$$\pi_1(\mathbb{T}^2) = \mathbb{Z} \oplus \mathbb{Z}.$$
This reflects the fact that there exist two different types of closed curves on the torus which cannot be contracted to one point. For example, in Fig. 5.27(c) this concerns the equator ABA and the meridian circle PAP. Consequently, the torus $\mathbb{T}^2$ is not simply connected.
- The real 2-dimensional projective space $\mathbb{P}^2$: This space is obtained from the closed unit disc by identifying diametrically opposed boundary points of the unit disc with each other (Fig. 5.28). The topological space $\mathbb{P}^2$ is compact, arcwise connected, non-orientable, and it has the Betti numbers
$$\beta_0 = 1, \ \ \beta_1 = \beta_2 = 0.$$
This yields the Euler characteristic $\chi(\mathbb{P}^2) = 1$. The genus of $\mathbb{P}^2$ is given by $g = 1$, as we will discuss below. The additive fundamental group of $\mathbb{P}^2$ is given by

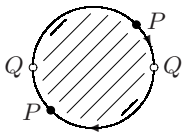


Fig. 5.28. The projective space $\mathbb{P}^2$

(a) $g = 0,\ r = 2$ (b) $g = 1,\ r = 1$

Fig. 5.29. Oriented surface of genus g with r boundaries

$$\pi_1(\mathbb{P}^2) = \mathbb{Z}_2.$$

This group consists of the two elements $\{0, 1\}$ with $1+1 = 0$. The element "1" of $\pi_1(\mathbb{P}^2)$ corresponds to the boundary curve PQP (Fig. 5.28). This curve cannot be continuously contracted to a point within $\mathbb{P}^2$. Thus, the projective space $\mathbb{P}^2$ is not simply connected. In contrast to this, taking the identification of diametrically opposed boundary points into account, the curve $PQPQP$ can be continuously contracted to the center of the unit disc. This corresponds to "1+1=0."

Let us now study the general case. We are given the compact 2-manifold $\mathcal{M}$ of genus g. Let $r = 0, 1, 2, \ldots$ be the number of boundaries. Then there exists a manifold $\mathcal{N}$ which is homeomorphic to $\mathcal{M}$ and which represents one of the following normal forms:[76]

(I) Suppose that $\mathcal{M}$ is orientable.

(I-1) Let $r = 0$. Then the normal form $\mathcal{N}$ is obtained from the unit sphere $\mathbb{S}^2$ by taking $2g$ open discs away and by attaching g handles. The genus is also equal to the number of 'holes' (Fig. 5.25).
- Betti numbers of $\mathcal{M}$: $\beta_0 = \beta_2 = 1,\ \beta_1 = 2g$.
- Euler characteristic of $\mathcal{M}$: $\chi = 2 - 2g$.
- Additive fundamental group of $\mathcal{M}$: $\pi_1 = \mathbb{Z} \oplus \ldots \oplus \mathbb{Z}$ ($2g$ summands).

(I-2) Let $r > 0$. The normal form $\mathcal{N}$ is obtained from (I-1) by taking r open discs away (Fig. 5.29).
- Betti numbers of $\mathcal{M}$: $\beta_0 = \beta_2 = 1$ and $\beta_1 = 2g + r - 1$.
- Euler characteristic of $\mathcal{M}$: $\chi = 3 - 2g - r$.

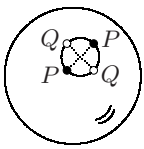
(II) Suppose that $\mathcal{M}$ is non-orientable.

(II-1) Let $r = 0$. The normal form $\mathcal{N}$ is obtained from the unit sphere $\mathbb{S}^2$ by taking g open discs away, and by identifying diametral points of the boundary circles with each other. The number g is the genus of $\mathcal{M}$. For example, the surface pictured in Fig. 5.30(a) is homeomorphic to the projective space $\mathbb{P}^2$.
- Betti numbers of $\mathcal{M}$: $\beta_0 = 1,\ \beta_1 = g - 1, \beta_2 = 0$. Here, $g = 1, 2, \ldots$
- Euler characteristic of $\mathcal{M}$: $\chi = 2 - g$.

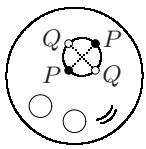
(II-2) Let $r > 0$. The normal form is obtained from (II-1) by taking r open discs away. Again the number g is called the genus of $\mathcal{M}$.
- Betti numbers of $\mathcal{M}$: $\beta_1 = 1,\ \beta_1 = g + r - 1,\ \beta_2 = 0$.
- Euler characteristic of $\mathcal{M}$: $\chi = 2 - g - r$.

In particular, the Möbius strip corresponds to $g = 1$ and $r = 1$. Hence $\beta_1 = 1$ and $\chi = 0$.

[76] Note that the Betti numbers, the Euler characteristic, and the fundamental groups of $\mathcal{M}$ and $\mathcal{N}$ coincide.



(a) $g = 1,\ r = 0$



(b) $g = 1,\ r = 2$

Fig. 5.30. Non-orientable surface of genus g with r boundaries

5.13 The Poincaré Conjecture and the Ricci Flow

> The study of mathematics, like the Nil, begins in minuteness, but ends in magnificence.
>
> C.C. Colton, 1820

> Friedman's 1982 proof of the 4-dimensional Poincaré hypothesis was an extraordinary tour de force. His methods were so sharp that as to actually provide a complete classification of all compact simply connected topological 4-manifolds, yielding many previously unknown examples of such manifolds.[77]
>
> John Milnor, 1986

The Poincaré conjecture for 3-manifolds was one of the seven Millenium Prize Problems announced by the Clay Mathematics Institute in Cambridge, Massachusetts, in the year 2000 (see Sect. 1.7 of Vol. I).

The topological characterization of the 2-sphere. The following result was already known in the second half of the 19th century:

> *A compact simply connected 2-manifold is homeomorphic to a 2-sphere.*

This homeomorphism can be chosen as a diffeomorphism.

The topological characterization of the 3-sphere. The famous Poincaré conjecture claims the following:

> *A compact simply connected* 3*-manifold is diffeomorphic to a* 3*-sphere.*

This conjecture was proven by Grigori Perelman (born 1966) in 2003. He invented an ingenious approach for solving this outstanding problem. The main idea comes from physics. Here, Perelman uses the physical picture of the flow of fluid particles in order to deform the original 3-manifold into the final 3-sphere. More precisely, he applies the so-called Ricci flow on manifolds which was thoroughly studied in the pioneering papers by Richard Hamilton in the 1980s and 1990s. The Ricci flow is governed by the partial differential equation

$$\boxed{\frac{\partial g(t)}{\partial t} = -2\mathrm{Ric}(g(t)), \qquad t \geq 0.} \tag{5.113}$$

Here, the parameter t can be regarded as time. This equation describes the time-deformation of the metric tensor g of a Riemannian manifold $\mathcal{M}$. Equation (5.113)

[77] J. Milnor, The work of M. H. Freedman. In: M. Atiyah and D. Iagolnitzer (Eds.), Fields Medallists' Lectures, World Scientific, Singapore, 2003, pp. 405–406.

is of the type of a diffusion process (or heat conduction process). The main difficulty is that there may appear singularities during the time evolution. One has to control the possible types of singularities, and one has to regularize and renormalize them (see the hints for further reading on page 351).

The generalization of the idea of the flow of fluid particles is also crucial for quantum physics. In Sect. 7.11.5 we will show that the Feynman path integral corresponds to a diffusion process in imaginary time (the Feynman–Kac formula). Moreover, the modern approach to renormalization in quantum field theory is based on the flow generated by the renormalization group (see Chap. 3 of Vol. I).

The homotopy type of a topological space. Recall from Vol. I the following terminology. Let X and Y be topological spaces. Let id_X denote the identity map on X. The two continuous maps $f, g : X \to X$ are called homotopic iff there exists a continuous map $H : X \times [0,1] \to X$ such that

$$H(x,0) = f(x) \qquad \text{and} \qquad H(x,1) = g(x) \qquad \text{for all} \quad x \in X.$$

We write $f \sim g$. The map H is called a homotopy.

By definition, the space X has the same homotopy type as the space Y iff there exist continuous maps $F : X \to Y$ and $G : Y \to X$ such that

$$GF \sim \mathrm{id}_X \qquad \text{and} \qquad FG \sim \mathrm{id}_Y.$$

That is, the composite maps $GF : X \to X$ and $FG : Y \to Y$ are homotopic to the corresponding identity maps. We also say that the space X is homotopy equivalent to the space Y. For example, a topological space is called contractible iff it has the same homotopy type as a single point (see Sect. 5.5 of Vol. I)..

By definition, the topological space X has the same topological type as the topological space Y iff X is homeomorphic to Y. Explicitly, this means that there exists a bijective continuous map $F : X \to Y$ such that the inverse map $F^{-1} : Y \to X$ is also continuous. Setting $G := F^{-1}$, we get $GF = \mathrm{id}_X$ and $FG = \mathrm{id}_Y$. Thus, if X and Y have the same topological type, then they also have the same homotopy type. However, the converse is not always true.

The topological characterization of the n-sphere. The generalized Poincaré conjecture reads as follows:

> *If an n-manifold has the same homotopy type as an n-sphere, then it has actually the same topological type as an n-sphere.*

Nowadays we know that this statement is true for all dimensions $n = 1, 2, \ldots$ For $n \geq 5$, the proof was given by Smale, and independently by Stallings and Zeeman and by Wallace in 1960-61. For $n = 4$, Freedman gave the proof in 1982. Perelman settled the most difficult case $n = 3$ in 2002. For their seminal results, Smale, Freedman and Perelman were awarded the Fields medal in 1966, 1986, and 2006, respectively. Perelman refused the award. Another formulation of the Poincaré conjecture reads as follows:

> *If an n-manifold has the same fundamental group and the same homology as the n-sphere, then it is actually homeomorphic to the n-sphere.*[78]

This is true for all dimensions $n = 1, 2, \ldots$ The positive answer to the Poincaré conjecture tells us the following highly nontrivial result:

> *Algebraic topology is able to detect spheres in all dimensions.*

[78] By definition, two topological spaces X and Y have the same homology iff they have the same homology groups with integer coefficients. For an n-sphere, these homology groups are given by $H_0 = H_n = \mathbb{Z}$ and $H_k = 0$ if $k = 1, 2, \ldots, n-1$ (see Vol. IV).

5.14 A Glance at Modern Optimization Theory

In the 1950s, modern control theory was founded by generalizing the duality between wave fronts and light rays.

(i) Dynamic programming (generalized wave fronts): Bellman created dynamic programming by generalizing the Hamilton–Jacobi partial differential equation to the Bellman functional equation. In geometrical optics, the eikonal function S measures the time needed for the propagation of light. In dynamic programming, the function S measures the quantity to be optimized (e.g., the costs of a production process).

(ii) Pontryagin's maximum principle (generalized light rays): Pontryagin generalized the Hamilton canonical equations for light rays (and the maximum principle in geometrical optics) to the computation of optimal trajectories in modern technology. Let us discuss some important examples.

- For the return of a spaceship from moon to earth, one has to compute a trajectory such that the heating of the spaceship remains minimal. Of interest in the optimal solution is the fact that the spaceship penetrates the earth's atmosphere rather deeply (from 120 km altitude to 50 km) and then it climbs again to the altitude of 75 km. On the other hand, the velocity falls almost monotonically. The computation (performed by Roland Bulirsch for the NASA in the 1960s) can be found in Stoer and Bulirsch, Introduction to Numerical Analysis, Springer, New York, 1993.
- For the moon landing, one needs a feed-back control program which guarantees minimal fuel consumption of the moon landing ferry. Here, the braking process is controlled by measuring the distance between the ferry and the moon surface.
- For the flight to Mars, one needs a trajectory of the spaceship which again guarantees minimal fuel consumption, by taking the gravitational forces of the planets in our solar system into account.

The proof for the validity of Pontryagin's maximum principle is highly sophisticated. For a detailed study of (i) and (ii) in the setting of nonlinear functional analysis, we refer to Zeidler (1986), Vol. III, quoted on page 353.

5.15 Hints for Further Reading

Standard textbooks:

M. Born and E. Wolf, Principles of Optics, 7th edn., Cambridge University Press, 1999 (classical optics).

L. Mandel and E. Wolf, Optical Coherence and Quantum Optics, Cambridge University Press, 1995 (modern laser optics).

S. Rodriguez, The Scattering of Light by Matter, Scuola Normale Superiore, Pisa, 2001.

Classic monograph on geometrical optics:

C. Carathéodory, Geometrische Optik, Springer, Berlin, 1937 (in German).

Introduction:

H. Römer, Theoretical Optics: An Introduction, Wiley, Weinheim, 2005.

G. Scharf, From Electrostatics to Optics, Springer, Berlin, 1994.

V. Arnold, Huygens and Borrow, Newton and Hooke: Pioneers in Mathematical Analysis and Catastrophe Theory: From Evolvents to Quasi-Crystals, Birkhäuser, Basel, 1990.

Caustics and singularities:

V. Arnold, Singularities of Caustics and Wave Fronts, Kluwer, Dordrecht, 1991.

Yu. Kravtsov and Yu. Orlov, Caustic, Catastrophes and Wave Fields, Springer, Berlin, 1999.

V. Arnold, S. Gusein-Zade, and A. Varchenko, Singularities of Differentiable Maps, Vols. 1, 2, Birkhäuser, Basel, 1985.

W. Ebeling, Functions of Several Complex Variables, and Their Singularities, Amer. Math. Soc., Providence, Rhode Island, 2007.

Galaxies as gravitational lenses for the light coming from quasars:

J. Ehlers, E. Falco, and P. Schneider, Gravitational Lenses, Springer, New York, 1992.

Light in modern cosmology:

S. Weinberg, Cosmology, Oxford University Press, 2008.

Viscosity solutions of the Hamilton–Jacobi equation including singularities (e.g., caustics):[79]

P. Lions, Generalized Solutions of Hamilton–Jacobi Equations, Pitman, Boston, 1982.

The Huygens principle in both geometrical optics and the calculus of variations:

K. Maurin, The Riemann Legacy, Kluwer, Dordrecht, 1997.

V. Arnold, Lectures on Partial Differential Equations, Springer, Berlin, 2004.

M. Giaquinta and S. Hildebrandt, Calculus of Variations, Vols. 1, 2, Springer, Berlin, 1995.

The Huygens principle for wave equations:

P. Günther, Huygens' Principle and Hyperbolic Differential Equations, Academic Press, San Diego, 1988.

P. Günther, Huygens' principle and Hadamard's conjecture, Math. Intelligencer **13** (1991), 56–63.

M. Radzikowski, Micro-local approach to the Hadamard condition in quantum field theory on curved space-time, Commun. Math. Phys. **179** (1996), 529–553.

R. Brunetti, K. Fredenhagen, and M. Köhler, The micro-local spectrum condition and the Wick polynomials of free fields, Commun. Math. Phys. **180** (1996), 633–652.

R. Brunetti and K. Fredenhagen, Micro-local analysis and interacting quantum field theories: renormalization on physical backgrounds, Commun. Math. Phys. **208** (2000), 623–661.

[79] For his fundamental contributions to the modern theory of nonlinear partial differential equations, Piere-Louis Lions was awarded the Fields medal in 1994.

C. Bär, N. Ginoux, and F. Pfäffle, Wave Equations on Lorentzian Manifolds and Quantization, European Mathematical Society, 2007.

The modern version of Euclid's *Elements*:

D. Hilbert, Foundations of Geometry, Teubner, Leipzig, 1899 (12th edn. 1977) (in German).

Geometry in physics:

K. Maurin, The Riemann Legacy: Riemannian Ideas in Mathematics and Physics of the 20th Century, Kluwer, Dordrecht, 1997.

S. Novikov and T. Taimanov, Geometric Structures and Fields, Amer. Math. Soc., Providence, Rhode Island, 2006.

J. Jost, Geometry and Physics, Springer, Berlin, 2008.

C. Scriba and P. Schreiber, 5000 Years Geometry: History, Cultures, People, Springer, Berlin, 2005 (in German).

Geometry of 2-dimensional surfaces:

J. Stoker, Differential Geometry, Wiley, New York, 1989.

J. Jost, J. Eschenburg, Differential Geometry and Minimal Surfaces, Springer, Berlin, 2nd enlarged edn., 2007 (in German).

J. Jost, Compact Riemann Surfaces: An Introduction to Contemporary Mathematics, 3rd edn., Springer, Berlin, 2006.

Contact geometry:

C. Carathéodory, Calculus of Variations and Partial Differential Equations of First Order, Chelsea, New York, 1982.

V. Arnold, Contact Geometry and Wave Propagation, L'Enseignement Mathématique, Université de Genève, 1989.

V. Arnold, Contact geometry: the geometrical method of Gibbs's thermodynamics, Proceedings of the Gibbs Symposium, Yale University, Connecticut (U.S.A), 1989, pp. 163–179, Amer. Math. Soc., Providence, Rhode Island, 1990.
M. Giaquinta and S. Hildebrandt, Calculus of Variations, Vol. 2, Springer, Berlin, 1995.

Symplectic geometry:

C. Siegel, Symplectic Geometry, Academic Press, New York, 1964.

S. Lang, $SL(2,\mathbb{R})$, Addison–Wesley, Reading, Massachusetts, 1975.

V. Guillemin, S. Sternberg, Geometric Asymptotics, Amer. Math. Soc., Providence, Rhode Island, 1989.

V. Guillemin, S. Sternberg, Symplectic Techniques in Physics, Cambridge University Press, 1990.

H. Hofer and E. Zehnder, Symplectic Invariants and Hamiltonian Dynamics, Birkhäuser, Basel, 1994.

Kähler geometry:

J. Jost, Nonlinear Methods in Riemannian and Kählerian Geometry, Birkhäuser, Basel, 1991.

J. Jost, Riemannian Geometry and Geometric Analysis, Springer, Berlin, 2005.

W. Ballmann, Lectures on Kähler Manifolds, European Mathematical Society, 2006.

A. Moroianu, Lectures on Kähler Geometry, Cambridge University Press, 2007.

J. Bourguignon, The unabated vitality of Kählerian geometry, pp. 737–766. In: E. Kähler, Mathematical Works, de Gruyter, Berlin, 2004.

K. Becker, M. Becker, and J. Schwarz, String Theory and M-Theory (Master theory), Cambridge University Press, 2007.

Einstein geometry:

A. Besse, Einstein Manifolds, Springer, New York, 1987.

Ø. Grøn and S. Hervik, Einstein's Theory of General Relativity: with Modern Applications in Cosmology, Springer, New York, 2007.

Poissonian geometry:

Marsden, J., Ratiu, T. (1999), Introduction to Mechanics and Symmetry, Springer, New York.

J. Marsden, T. Ratiu (Eds.), The Breadth of Symplectic and Poissonian Geometry, Birkhäuser, Boston, 2004.

S. Waldmann, Poissonian Geometry and Deformation Quantization, Springer, Berlin, 2007 (in German).

Gauge theory:

S. Kobayashi and K. Nomizu, Foundations of Differential Geometry, Vols. 1, 2, Wiley, New York, 1963.

K. Marathe and G. Martucci, The Mathematical Foundations of Gauge Theories, North-Holland, Amsterdam, 1992.

G. Naber, Topology, Geometry, and Gauge Fields, Springer, New York, 1997.

S. Novikov and T. Taimanov, Geometric Structures and Fields, Amer. Math. Soc., Providence, Rhode Island, 2006.

C. Burgess and G. Moore, The Standard Model: A Primer, Cambridge University Press, 2007.

L. Brown et al. (Eds.), The Rise of the Standard Model, Cambridge University Press, 1995.

Topology of surfaces:

C. Kinsey, Topology of Surfaces, Springer, New York, 1993.

F. Levi, Geometrical Configurations: with an Introduction to the Combinatorial Topology of Surfaces, Leipzig, 1929 (in German).

W. Rinow, Textbook on Topology, Deutscher Verlag der Wissenschaften, Berlin, 1975 (in German).

Ricci flow and the proof of the Poincaré conjecture:

S. Smale, The generalized Poincaré conjecture in dimension greater than four, Ann. Math. **74**(2) (1961), 391–406.

J. Milnor, The work of M. H. Freedman. In: M. Atiyah and D. Iagolnitzer (Eds.), Fields Medallists' Lectures, World Scientific, Singapore, 2003, pp. 405–406.

P. Topping, Lectures on the Ricci Flow, Cambridge University Press, 2006.

H. Cao and X. Zhu, A complete proof of the Poincaré and geometrization conjectures – Application of the Hamilton–Perelman theory of Ricci flow, Asian J. Math. **10**(2) (2006), 195–492.

R. Hamilton, The formation of singularities in the Ricci flow, pp. 1–118. In: H. Cao, B. Chow, S. Chu, and S. Yau (Eds.), Collected Papers on Ricci Flow, International Press, Boston, 2003.

H. Cao, S. Yau, and X. Zhu (Eds.), Structure of Three-dimensional Space: The Poincaré and Geometrization Conjectures, International Press, Boston, 2006.

J. Morgan, G. Tian, Ricci Flow and the Poincaré Conjecture, Amer. Math. Soc., Providence, Rhode Island/Clay Mathematics Institute, Cambridge, Massachusetts, 2007.

The renormalization group flow and renormalization in statistical physics and quantum field theory:

W. McComb, Renormalization Methods: A Guide for Beginners, Oxford University Press, 2007.

Classical papers in non–Euclidean geometry:

C. Gauss, Disquisitiones generales circa superficies curvas (Investigations about curved surfaces), 1827. In: C. Gauss, Collected Works, Vol. 5, pp. 217–256; 341–347, Göttingen, 1863ff (in Latin). German translation: C. Gauss, Allgemeine Flächentheorie, Ostwalds Klassiker, Vol. 5, Leipzig, 1889.

B. Riemann, On the hypotheses which lie at the foundation of geometry, Göttinger Abhandlungen **13** (1854), 272–287 (in German).

E. Beltrami, Fundamental theory of spaces of constant curvature, Ann. di Mat. **2**(2) (1868), 232–255 (in Italian).

D. Hilbert, On manifolds of constant Gaussian curvature, Trans. Amer. Math. Soc. **2** (1901), 87–99 (in German).

H. Poincaré, La science et l'hypothèse, Flamiron, Paris, 1902 (in French).

J. Nash, The embedding problem for Riemannian manifolds, Ann. of Math. **63** (1956), 20–63.

M. Günther, On the Nash embedding theorem, Math. Nachrichten **144** (1989), 165–187 (in German).

P. Dombrowski, 150 years after Gauss' *Disquisitiones generales circa superficies curvas*, Astérisque **62** (1979).

R. Sharpe, Differential Geometry: Cartan's Generalization of Klein's Erlangen Program, Springer, New York, 1997.

T. Ivey and J. Landsberg, Cartan for Beginners: Differential Geometry via Moving Frames and Exterior Differential Systems, Amer. Math. Soc., Providence, Rhode Island, 2003.

S. Nasar, A Beautiful Mind: A Biography of John Forbes Nash, Jr., Simon and Schuster, New York, 1998.

The hard implicite function theorem:

J. Moser, A rapidly convergent iteration method and nonlinear partial differential equations I, II, Ann. Scuola Norm. Sup. Pisa **20** (1966), 226–315, 449–535.

R. Hamilton, The inverse function theorem of Nash and Moser, Bull. Amer. Math. Soc. **7** (1982), 65–222.

S. Krantz and H. Parks, The Implicit Function Theorem: History, Theory, and Applications, Birkhäuser, Boston, 2002 (classical and hard implicit function theorem).

J. Bricmont and A. Kupiainen, Renormalizing of partial differential equations, pp. 83–115. In: V. Rivasseau (Ed.), Constructive Physics, Springer, Berlin, 1995.

J. Bricmont, K. Gawędzki, and A. Kupiainen, Kolmogorov–Arnold–Moser (KAM) theorem and quantum field theory, Commun. Math. Phys. **201**(3) (1999), 699–727.

The relations between the calculus of variations and modern optimization theory based on the unifying point of view of nonlinear functional analysis are thoroughly studied in:

E. Zeidler, Nonlinear Functional Analysis and its Applications, Vol. III: Variational Methods and Optimization, Springer, New York, 1986.

We also refer to the following survey article:

R. Klötzler, Commentaries on optimal control and duality, pp. 307–359. In: Appendix to C. Carathéodory, Calculus of Variations and First-Order Partial Differential Equations, Part II, Teubner, Leipzig, 1994 (in German).

Problems

5.1 *Leibniz's product rule for differentials.* Let $x, h \in \mathbb{R}^n$ with $n = 1, 2, \dots$ We write $x := (x^1, \dots, x^n)$ and $h = (h^1, \dots, h^n)$. Choose $i = 1, \dots n$. The basic definition is given by

$$\boxed{dx^i(h) = h^i, \qquad h \in \mathbb{R}^n.} \tag{5.114}$$

This means that dx^i is a linear functional on $\mathbb{R}^n$. For each smooth function $f : \mathbb{R}^n \to \mathbb{R}$, we define[80]

$$\boxed{df(x) := \frac{\partial f(x)}{\partial x^i}\, dx^i.} \tag{5.115}$$

We sum over $i = 1, \dots n$. Explicitly, we get the linear functional $df(x) : \mathbb{R}^n \to \mathbb{R}$ with

$$df_x(h) = \frac{\partial f(x)}{\partial x^i}\, dx^i(h) = \frac{\partial f(x)}{\partial x^i}\, h^i.$$

Prove that, for given smooth functions $f, g : \mathbb{R}^n \to \mathbb{R}$, we have the Leibniz product rule

$$\boxed{d(fg) = (df)g + f(dg).} \tag{5.116}$$

Solution: $d(fg) = (fg)_{x^i} dx^i = f_{x^i} g dx^i + f g_{x^i} dx^i$.

[80] To streamline notation, we sometimes write df_x instead of $df(x)$.

5.2 *Leibniz's chain rule and the pull-back of differentials.* To the differential

$$df = \frac{\partial f}{\partial x^i}\, dx^i$$

we want to apply the transformation formula $x^i = x^i(u^1, \ldots, u^m), i = 1, \ldots, n$.

(i) Elegant formal argument due to Leibniz: It follows from $dx^i = \frac{\partial x^i}{\partial u^j}\, du^j$ that

$$df = \frac{\partial f}{\partial x^i} \frac{\partial x^i}{\partial u^j}\, du^j. \tag{5.117}$$

We sum over $i = 1, \ldots, n$ and $j = 1, \ldots, m$. This corresponds to the chain rule

$$\frac{\partial f}{\partial u^j} = \frac{\partial f}{\partial x^i} \frac{\partial x^i}{\partial u^j}, \qquad j = 1, \ldots, m$$

where we sum over $i = 1, \ldots, n$.

(ii) Rigorous argument: We write $x^i = F^i(u^1, \ldots, u^m)$, and we define the pull-back F^*df of the functional df by setting

$$F^*df := df \circ F'. \tag{5.118}$$

This means that we use the linearized map $F'(u) : \mathbb{R}^n \to \mathbb{R}^n$ in order to transform the linear functional df into the linear functional F^*df. Explicitly, this means that

$$(F^*df)_u(k) = df_x(F'(u)k) \qquad \text{for all} \quad k \in \mathbb{R}^m,$$

where $x = F(u)$. Show that (ii) corresponds to (i) because of the quite natural relation

$$\boxed{F^*(df) = d(f \circ F).} \tag{5.119}$$

between the differential $d(f \circ F)$ of the transformed function $f \circ F$ and the pull-back F^*df of the original differential df.

Solution: Set $h := F'(u)k$. Then

$$h^i = \frac{\partial F^i(u)}{\partial u^j}\, k^j, \qquad i = 1, \ldots, n$$

where we sum over $j = 1, \ldots, m$. Then

$$df_x(F'(u)k) = \frac{\partial f(x)}{\partial x^i} \frac{\partial F^i(u)}{\partial u^j}\, k^j$$

where $x = x(u)$. Set $g := f \circ F$. Then $g(u) = f(F(u))$. By the chain rule,

$$\frac{\partial g(u)}{\partial u^j} = \frac{\partial f(x)}{\partial x^i} \frac{\partial F^i(u)}{\partial u^j}.$$

Hence $dg_u(k) = \frac{\partial g(u)}{\partial u^j} k^j = df_x(F'(u)k)$. This finishes the proof.

In practical computation, we recommend the reader to use the elegant Leibniz formula (i) *above.*

Our proof justifies this.

The language of physicists. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function. Set $\Delta x := h$ and $\Delta f(x) := f(x + \Delta x) - f(x)$. By Taylor expansion,

$$\Delta f(x) = f'(x)\Delta x + o(\Delta x), \qquad \Delta x \to 0.$$

Physicist frequently write df and dx instead of Δf and Δx, respectively, and they cancel the remainder $o(\Delta x)$. This is convenient for practical purposes. However, from the mathematical point o view, one has to distinguish between

- the linear functional $df(x) : \mathbb{R}^n \to \mathbb{R}$ and
- the real number $\Delta f(x)$.

5.3 *Proof of Proposition 5.12 on page 324.* Solution: Ad (i). Let $\omega := dx \wedge dy$. Consider the diffeomorphism $F : \mathbb{R}^2 \to \mathbb{R}^2$ given by $\xi = \xi(x, y), \eta = \eta(x, y)$. We want to determine the map F in such a way that

$$F^*\omega = \omega \qquad \text{on } \mathbb{R}^2.$$

To this end, we will use the Leibniz method discussed in Problem 5.2. From

$$d\xi = \xi_x dx + \xi_y dy, \qquad d\eta = \eta_x dx + \eta_y dy,$$

we get $d\xi \wedge d\eta = (\xi_x \eta_y - \xi_y \eta_x)\, dx \wedge dy$. Therefore, the equation

$$d\xi \wedge d\eta = dx \wedge dy \qquad \text{on } \mathbb{R}^2$$

is equivalent to $\xi_x \eta_y - \xi_y \eta_x = 1$ on $\mathbb{R}^2$.
Ad (ii). Use the same argument.

5.4 *The real Schwarz inequality.* Let $f, g : [a, b] \to \mathbb{R}$ be two continuous functions on the compact interval $[a, b]$. Show that

$$\left| \int_a^b f(x)g(x)dx \right| \le \left(\int_a^b |f(x)|^2 dx \right)^{1/2} \left(\int_a^b |g(x)|^2 dx \right)^{1/2}.$$

Equality holds iff f and g are linearly dependent, that is, there exist real numbers α, β with $\alpha^2 + \beta^2 \neq 0$ and

$$\alpha f(x) + \beta g(x) = 0 \qquad \text{for all} \quad x \in [a, b].$$

Solution: Set $\langle f|g\rangle := \int_a^b f(x)g(x)dx$, and $||f||^2 = \sqrt{\langle f|f\rangle}$. By homogeneity, it is sufficient to prove the claim for $||g|| = 1$. Define $F(t) := ||f + tg||^2$ for all $t \in \mathbb{R}$. Then

$$F(t) = \langle f + tg|f + tg\rangle = \langle f|f\rangle + 2t\langle f|g\rangle + t^2.$$

The equation

$$F(t) = 0, \qquad t \in \mathbb{R}$$

has the zeros $t_\pm = -\langle f|g\rangle \pm \sqrt{\langle f|g\rangle^2 - ||f||^2}$. Note that $F(t) \ge 0$ for all $t \in \mathbb{R}$.

- If $F(t) > 0$ for all $t \in \mathbb{R}$, then $\langle f|g\rangle^2 - ||f||^2 < 0$.
- If $F(t) = 0$ for some t, then $\langle f|g\rangle^2 - ||f||^2 = 0$.

5.5 *The complex Schwarz inequality.* Let $f, g : [a, b] \to \mathbb{C}$ be two continuous functions on the compact interval $[a, b]$. Show that

$$\left| \int_a^b f(x)^\dagger g(x)dx \right| \le \left(\int_a^b |f(x)|^2 dx \right)^{1/2} \left(\int_a^b |g(x)|^2 dx \right)^{1/2}.$$

Equality holds iff f and g are linearly dependent, that is, there exist complex numbers α, β with $|\alpha|^2 + |\beta|^2 \neq 0$ and

$$\alpha f(x) + \beta g(x) = 0 \qquad \text{for all} \quad x \in [a, b].$$

Solution: Set $\langle f|g\rangle := \int_a^b f(x)^\dagger g(x)dx$. Replace g by $\mathrm{e}^{\mathrm{i}\alpha}g$, and choose the angle α in such a way that $\langle f|g\rangle$ is real. Observe that $||\mathrm{e}^{\mathrm{i}\alpha}g|| = ||g||$. Now use the same argument as above.
Historical remark. The Schwarz inequality is named after Amandus Schwarz (1843–1921) who published this inequality in 1884. However, this inequality was obtained much earlier by Cauchy (1789–1857) in 1821 (for sequences of numbers) and by Bunyakovskii (1804–1889) in 1859 (for integrals). Therefore, the inequality is also called the Cauchy–Bunyakovskii inequality.

5.6 *The Beltrami model in non-Euclidean geometry.* Set $z := x + \mathrm{i}y$. Let $\mathrm{int}(\mathbb{B}^1)$ denote the open unit disc $\{z \in \mathbb{C} : \ |z| < 1\}$ in the complex plane $\mathbb{C}$. for $\varepsilon = 1$ and $\varepsilon = -1$, study the following two metrics

$$ds^2 = \frac{4(dx^2 + dy^2)}{(1 + \varepsilon|z|^2)^2} \qquad \text{on} \ \ \mathrm{int}(\mathbb{B}^1).$$

For $\varepsilon = -1$ (resp. $\varepsilon = 1$), this metric represents the hyperbolic (resp. elliptic) Beltrami model. Let $\varepsilon = -1$. Show that the function

$$w = \left(\frac{z+1}{z-1}\right)^2, \qquad z \in \mathrm{int}(\mathbb{B}^1)$$

maps conformally the open unit disc onto the open upper-half plane $\mathcal{H}_\mathbb{C}$. Show that the inverse map transforms the hyperbolic metric on $\mathcal{H}_\mathbb{C}$ into the metric above with $\varepsilon = -1$. Prove that the hyperbolic straight lines on $\mathcal{H}_\mathbb{C}$ become circles on $\mathrm{int}(\mathbb{B}^1)$ which orthogonally intersect the unit circle. The points on the unit circle are regarded as "points of infinity." Use the corresponding Kähler metric in order to compute the Gaussian curvature $K \equiv 1$ (resp. $K \equiv -1$) if $\varepsilon = 1$ (resp. $\varepsilon = -1$).

5.7 *Geodesics and the Christoffel symbols.* Fix $n = 1, 2, \ldots$ Let $\mathcal{U}$ be a nonempty set of $\mathbb{R}^n$. Set $x = (x^1, \ldots, x^n)$. We are given a family of smooth functions

$$g_{ij}(x) : \mathbb{R}^n \to \mathbb{R}, \qquad i, j = 1, \ldots, n$$

with $g_{ij}(x) = g_{ji}(x)$ for all points $x \in \mathcal{U}$ and all indices $i, j = 1, \ldots, n$. Suppose that the symmetric $(n \times n)$-matrix $\mathcal{G} := (g_{ij}(x))$ is invertible for all $x \in \mathcal{U}$. Denote the entries of the inverse matrix $\mathcal{G}(x)^{-1}$ by $g^{ij}(x)$. Now consider the energetic variational problem

$$\int_{\tau_0}^{\tau_1} g_{ij}(x(\tau))\dot{x}^i(\tau)\dot{x}^j(\tau)\, d\tau = \text{critical!} \tag{5.120}$$

with fixed end points $x(\tau_0)$ and $x(\tau_1)$. We are looking for a smooth function $x : [\tau_0, \tau_1] \to \mathcal{U}$ on the compact interval $[\tau_0, \tau_1]$. In what follows we sum over equal upper and lower indices from 1 to n. The meaning of variational problems concerning critical values is explained on page 805 of Vol. I.
Show that every smooth solution of the variational problem satisfies the Euler–Lagrange equation

$$\ddot{x}^k(\tau) + \Gamma^k_{ij}(x(\tau))\dot{x}^i(\tau)\dot{x}^j(\tau) = 0, \qquad \tau_0 \leq \tau \leq \tau_1, \ \ k = 1, \ldots, n$$

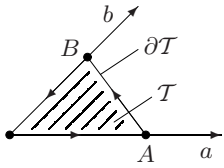


Fig. 5.31. Special parallel transport

with the Christoffel symbols.

$$\Gamma_{ij}^k := \tfrac{1}{2} g^{ks}(\partial_i g_{sj} + \partial_j g_{is} - \partial_s g_{ij}).$$

Solution: Set $L(x, \dot{x}) := g_{ij}(x)\dot{x}^i\dot{x}^j$. The solutions of the variational problem are solutions of the following Euler–Lagrange equations:

$$\frac{d}{d\tau} L_{\dot{x}^s} - L_{x^s} = 0, \qquad s = 1, \dots, n.$$

Since $L_{\dot{x}^s} = g_{sj}\dot{x}^j + g_{is}\dot{x}^i$ and $L_{x^s} = (\partial_s g_{ij})\dot{x}^i\dot{x}^j$, we get

$$g_{sj}\ddot{x}^j + g_{is}\ddot{x}^i + (\partial_r g_{sj})\dot{x}^r\dot{x}^j + (\partial_r g_{is})\dot{x}^r\dot{x}^i - (\partial_s g_{ij})\dot{x}^i\dot{x}^j = 0.$$

Because of the symmetry $g_{ij} = g_{ji}$, we obtain

$$2g_{sj}\ddot{x}^j + (\partial_i g_{sj} + \partial_j g_{is} - \partial_s g_{ij})\dot{x}^i\dot{x}^j = 0.$$

Since $g^{ks}g_{sj} = \delta_j^k$ (Kronecker symbol), we get

$$\ddot{x}^k + \tfrac{1}{2} g^{ks}(\partial_i g_{sj} + \partial_j g_{is} - \partial_s g_{ij})\ \dot{x}^i\dot{x}^j = 0.$$

5.8 *Proof of Theorem 5.14 on page 335.* Solution: (I) Special case: To display the simple idea of the proof, let us first consider the special case pictured in Fig. 5.31. The differential equation for the clockwise parallel transport along the boundary $\partial\mathcal{T}$ of the triangle $\mathcal{T}$ reads as

$$\dot{\psi}(\tau) = -\dot{x}^j(\tau)\mathcal{A}_j(P(\tau))\psi(\tau), \quad 0 \le \tau \le \tau_1, \quad \psi(0) = \psi_0. \tag{5.121}$$

Hence

$$\Pi_{\partial\mathcal{T}}\psi_0 - \psi_0 = \int_0^{\tau_1} \dot{\psi}(\tau)d\tau = -\int_{\partial\mathcal{T}} dx^j \mathcal{A}_j(P)\psi(P).$$

The basic trick of the proof is to extend the values of ψ on the boundary $\partial\mathcal{T}$ to the triangle $\mathcal{T}$ in a smooth way. Set

$$\omega := dx^j \mathcal{A}_j \psi.$$

By the Poincaré–Stokes integral theorem, we have

$$\int_{\partial\mathcal{T}} \omega = \int_{\mathcal{T}} d\omega.$$

See Sect. 10.4.2 of Vol. I. Hence

$$\int_{\partial\mathcal{T}} \omega = \int_{\mathcal{T}} \partial_k(\mathcal{A}_j\psi)\ dx^k \wedge dx^j = \int_{\partial\mathcal{T}} (\partial_1(\mathcal{A}_2\psi) - \partial_2(\mathcal{A}_1\psi))\ dx^1 dx^2.$$

Set

$$\mathcal{B} := \lim_{\tau \to +0} \frac{\int_{\partial \mathcal{T}} (\partial_1(\mathcal{A}_2\psi) - \partial_2(\mathcal{A}_1\psi))\, dx^1 dx^2}{\mathrm{meas}(\mathcal{T})}.$$

Noting that $\partial_k(\mathcal{A}_j\psi) = \partial_k\mathcal{A}_j \cdot \psi + \mathcal{A}_j\partial_k\psi$, we get

$$\mathcal{B} = \partial_1\mathcal{A}_2(P)\psi_0 - \partial_2\mathcal{A}_1(P)\psi_0 + \mathcal{A}_2(P)\partial_1\psi(P) - \mathcal{A}_2(P)\partial_1\psi(P).$$

By (5.121), $\partial_1\psi(P) = -\mathcal{A}_2(P)\psi_0$ and $\psi_2(P) = \mathcal{A}_2(P)\psi_0$. Hence

$$\mathcal{B} = \{\partial_1\mathcal{A}_2(P) - \partial_2\mathcal{A}_1(P) + \mathcal{A}_1(P)\mathcal{A}_2(P) - \mathcal{A}_2(P)\mathcal{A}_1(P)\}\psi_0 = \mathcal{F}_{12}(P)\psi_0.$$

Finally, by Fig. 5.31, $\det(a, b) = a^1b^2$. Hence

$$\mathcal{F}_P(a, b) = \mathcal{F}_{12}(dx^1 \wedge dx^2)(a, b) = \mathcal{F}_{12}(P)(a^1b^2 - a^2b^1) = \mathcal{F}_{12}(P)a^1b^2.$$

(II) General case: Proceed similarly as in (I).

6. The Principle of Critical Action and the Harmonic Oscillator – Ariadne's Thread in Classical Mechanics

Since the divine plan is the most perfect thing there is, there can be no doubt that all actions in the universe can be determined by the calculus of the minima and maxima from the corresponding causes.

Leonhard Euler

The history of the principle of least action has often been described. Yet the matter is still controversial, and there seems to be no general agreement who invented the principle, Leibniz (1646–1717), Euler (1707–1783), or Maupertuis (1698–1759)... We mention that the first mathematical treatment of the action principle was given by Euler in the *Additamentum* of his *Methodus inveniendi.*[1]

Mariano Giaquinta and Stefan Hildebrandt, 1996

By generalizing the method of Euler in the calculus of variations, Lagrange (1736–1813) discovered, how one can write, in a single line, the basic equation for all problems in analytic mechanics.

Carl Gustav Jakob Jacobi (1804–1851)

When we quantize a classical theory, wave packets behave like particles... A wave packet might decay into two wave packets. When two wave packets come near to each other, they scatter and perhaps produce more wave packets. This naturally suggests the physics of particles can be described in these terms...
Quantum field theory grew out of essentially these sorts of physical ideas. It struck me as limiting that even after some 75 years, the whole subject of quantum field theory remains rooted in this harmonic paradigm, to use a dreadfully pretentious word. We have not been able to get away from the basic notions of oscillations and wave packets. Indeed, string theory, the heir to quantum field theory, is still firmly founded on this harmonic

[1] L. Euler, Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes, sive solutio problematis isoperimetrici lattissimo sensu accepti (A method for finding curves which have a minimal or maximal property or solutions of the generalized isoperimetric problem), Bousquet, Lausannae et Genevae 1744 (see Euler, Opera omnia, Ser. I, Vol. 24, quoted on page 1053).
M. Giaquinta and S. Hildebrandt, Calculus of Variations, Vols. 1, 2, Springer, Berlin, 1996 (reprinted with permission). We recommend reading this standard textbook, which also contains many illuminating historical comments.

> paradigm. Surely, a brilliant young physicist, perhaps a reader of this book, will take us beyond.[2]
>
> Anthony Zee, 2003

The aim of this and the following chapter is to explain the basic physical and mathematical ideas of classical mechanics and quantum mechanics by considering the so-called harmonic oscillator. In all fields of physics, one encounters oscillating systems. Let us mention the following examples:

- electromagnetic waves and light (photons);
- laser beams (coherent states);
- oscillating molecules in a gas or a liquid;
- sound waves (phonons);
- oscillations of a crystal lattice (phonons);
- oscillations of a string (e.g., a violin string);
- waves in a plasma (plasmons);
- matter waves of elementary particles (e.g., electrons);
- gravitational waves (gravitons).

The harmonic oscillator represents the simplest oscillating system. The quantization of the harmonic oscillator is the basis of quantum mechanics, quantum field theory, and condensed matter physics.

System of physical units. In this chapter, we will use the international system of units, SI (see the Appendix of Vol. I).

6.1 Prototypes of Extremal Problems

> The calculus of variations has its roots in extremal problems for real-valued functions.
>
> Folklore

The one-dimensional problem. Let $f : J \to \mathbb{R}$ be a smooth function on the open interval J. Consider the minimum problem

$$\boxed{f(x) = \min!, \qquad x \in J.} \tag{6.1}$$

Let us recall some standard results from classical calculus.

(i) Necessary condition for a local minimum: If x_0 is a solution of (6.1), then $f'(x_0) = 0$ and $f''(x_0) \geq 0$.

(ii) Sufficient condition for a local minimum: If $f'(x_0) = 0$ and $f''(x_0) > 0$, then the function f has a local minimum at the point x_0. This means that there exists a sufficiently small positive number ε such that $f(x) \geq f(x_0)$ for all $x \in]x_0 - \varepsilon, x_0 + \varepsilon[$.

(iii) Sufficient condition for a global minimum: If $f'(x_0) = 0, f''(x_0) > 0$, and $f''(x) \geq 0$ on J, then the function f is convex on J, and the minimum problem (6.1) has the unique solution x_0.

[2] A. Zee, Quantum Field Theory in a Nutshell, Princeton University Press, 2003 (reprinted with permission).

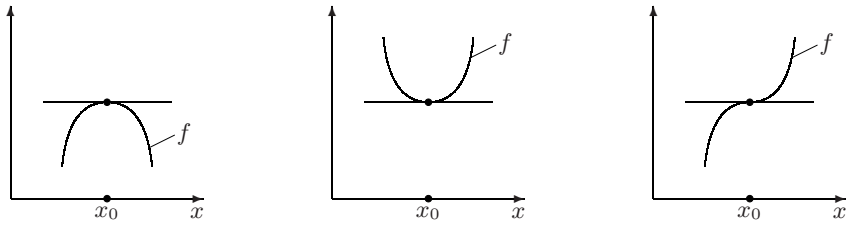


Fig. 6.1. Critical points

Now consider the more general problem

$$f(x) = \text{critical!}, \qquad x_0 \in J. \tag{6.2}$$

By definition, the point x_0 is a solution of (6.2) iff $f'(x_0) = 0$. We say that x_0 is a critical point of f, and the function is critical (or stationary) at the point x_0. Intuitively, the function f is critical at the point x_0 iff the tangent line of the graph of the function f at the point x_0 is horizontal (Fig. 6.1). For example, the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) := x^2$ has a global minimum at the point $x_0 = 0$. In contrast to this, the function $f(x) := x^3$ has the unique critical point $x_0 = 0$, but no minimal point. The function $f(x) := \sin x$ has precisely the critical points $x_0 = \pm\pi(n + \frac{1}{2})$ with $n = 0, 1, 2, \ldots$

Quadratic minimum problem. Set $f(x, y) := \alpha x^2 + \beta y^2$, and assume that $\beta > \alpha > 0$. Then the free minimum problem

$$f(x, y) = \min!, \qquad (x, y) \in \mathbb{R}^2 \tag{6.3}$$

has the unique solution $(x_0, y_0) = (0, 0)$. The constrained minimum problem

$$f(x, y) = \min!, \qquad x^2 + y^2 = 1, \qquad (x, y) \in \mathbb{R}^2 \tag{6.4}$$

has the unique solution $(x_0, y_0) = (1, 0)$ with $f(x_0, y_0) = \alpha$. Consider now the more general quadratic function

$$f(x, y) := ax^2 + 2bxy + cy^2, \tag{6.5}$$

where a, b, c are real numbers. Using the matrix $A := \begin{pmatrix} a & b \\ b & c \end{pmatrix}$, this can be written as $f(x, y) = (x, y)A \begin{pmatrix} x \\ y \end{pmatrix}$. Assume that the characteristic equation

$$\det(\lambda I - A) = 0, \qquad \lambda \in \mathbb{R},$$

that is, $\lambda^2 - (a + c)\lambda + ac - b^2 = 0$, has two solutions α and β with $\beta > \alpha > 0$.

Then the free minimum problem (6.3) *has the unique solution* $(0, 0)$*, and the constrained minimum problem* (6.4) *has a solution* (x_0, y_0) *with the minimal value* $f(x_0, y_0) = \alpha$.

In fact, it follows from the principal axis theorem for symmetric matrices A that the problems for the quadratic function f from (6.5) can be reduced to the corresponding problems for the simpler function $f(x, y) = \alpha x^2 + \beta y^2$, by using a rotation

(orthogonal transformation). The numbers α and β are the eigenvalues of the matrix A. It can be shown that the unique solution (x_0, y_0) of (6.4) is precisely the normalized eigenvector of the matrix A which belongs to the smallest eigenvalue α.

The Morse index. Let $\alpha \neq 0$ and $\beta \neq 0$. By definition, the Morse index $i_f(0,0)$ of the quadratic form f from (6.5) at the point $(0,0)$ is equal to the number of negative eigenvalues of the matrix A. Then the following hold:

- If $i_f(0,0) = 0$ (i.e., $\alpha > 0$ and $\beta > 0$), then f has a global minimum at the point $(0,0)$.
- If $i_f(0,0) = 2$ (i.e., $\alpha < 0$ and $\beta < 0$), then f has a global maximum at the point $(0,0)$.
- If $i_f(0,0) = 1$ (i.e., $\alpha\beta < 0$), then f has a saddle point at $(0,0)$.

For example, the function $f(x,y) := 2x^2 - y^2$ has a saddle point at $(0,0)$, whereas $f(x,y) := -x^2 - y^2$ has a global maximum at $(0,0)$.

First and second variation. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a smooth function. For fixed $(h,k) \in \mathbb{R}^2$, we set

$$\chi(\sigma) := f(x_0 + \sigma h, y_0 + \sigma k), \qquad \sigma \in \mathbb{R}.$$

We define

$$\boxed{\delta f(x_0, y_0; h, k) := \chi'(0),} \tag{6.6}$$

and call this the first variation of the function f at the point (x_0, y_0) in direction of (h,k). Similarly, we define the second variation of f by setting

$$\delta^2 f(x_0, y_0; h, k) := \chi''(0).$$

It follows from the Taylor theorem that

$$\chi(\sigma) = \chi(0) + \chi'(0)\sigma + \tfrac{1}{2}\chi''(0)\sigma^2 + \dots$$

Hence

$$f(x_0 + \sigma h, y_0 + \sigma k) = f(x_0, y_0) + \delta f(x_0, y_0; h, k)\sigma + \tfrac{1}{2}\delta^2 f(x_0, y_0; h, k)\sigma^2 + \dots$$

Explicitly, in terms of partial derivatives,

$$\begin{aligned} \delta f(x_0, y_0; h, k) &= f_x(x_0, y_0)h + f_y(x_0, y_0)k, \\ \delta^2 f(x_0, y_0; h, k) &= f_{xx}(x_0, y_0)h^2 + 2f_{xy}(x_0, y_0)hk + f_{yy}(x_0, y_0)k^2. \end{aligned}$$

Physicists use the notation $\delta x := \sigma h$ and $\delta y := \sigma k$. Then

$$f(x_0 + \delta x, y_0 + \delta y) = f(x_0, y_0) + \delta f(x_0, y_0; \delta x, \delta y) + \tfrac{1}{2}\delta^2 f(x_0, y_0; \delta x, \delta y) + \dots$$

If we introduce the two matrices $f'(x_0, y_0) := (f_x(x_0, y_0), f_y(x_0, y_0))$, and

$$f''(x_0, y_0) := \begin{pmatrix} f_{xx}(x_0, y_0) & f_{xy}(x_0, y_0) \\ f_{xy}(x_0, y_0) & f_{yy}(x_0, y_0) \end{pmatrix},$$

then $\delta f(x_0, y_0; \delta x, \delta y) = f'(x_0, y_0) \begin{pmatrix} \delta x \\ \delta y \end{pmatrix}$, and

$$\delta^2 f(x_0, y_0; \delta x, \delta y) = (\delta x, \delta y) f''(x_0, y_0) \begin{pmatrix} \delta x \\ \delta y \end{pmatrix}.$$

Here, $f'(x_0, y_0)$ (resp. $f''(x_0, y_0)$) is called the first (resp. second) Fréchet derivative[3] of the map $f : \mathbb{R}^2 \to \mathbb{R}$. Furthermore, the matrix $f''(x_0, y_0)$ (resp. the determinant $\det f''(x_0, y_0)$) is called the Hessian (resp. the Jacobian) of the map f at the point (x_0, y_0).

Observe that the use of the function χ allows us to reduce extremal problems for functions of several real variables to the inspection of functions of one real variable.

The same trick can be applied to variational problems.

This will be shown in Sect. 6.5.

Critical points. By definition, the function f has the critical point (x_0, y_0) iff the function χ has the critical point $\sigma = 0$, for all choices $h, k \in \mathbb{R}$. This means that

$$\delta f(x_0, y_0; h, k) = 0 \qquad \text{for all} \quad (h, k) \in \mathbb{R}^2.$$

Equivalently, $f'(x_0, y_0) = 0$, that is, $f_x(x_0, y_0) = f_y(x_0, y_0) = 0$.

Nonlinear minimum problem. Again let $f : J \to \mathbb{R}$ be a smooth function, where J is a nonempty open subset of $\mathbb{R}^2$. For the minimum problem

$$\boxed{f(x, y) = \min!, \qquad (x, y) \in J} \tag{6.7}$$

the following hold:

(i) Necessary condition for a local minimum. If (x_0, y_0) is a solution of (6.7), then $f'(x_0, y_0) = 0$.[4] This is the prototype of the Euler–Lagrange equation in the calculus of variations.

(ii) Sufficient condition for a local minimum. Suppose that $f'(x_0, y_0) = 0$, and suppose that Jacobi's accessory minimum problem

$$\delta^2 f(x_0, y_0; h, k) = \min!, \qquad h^2 + k^2 = 1, \quad (h, k) \in \mathbb{R}^2$$

has a positive minimal value.[5] Then f has a local minimum at the point (x_0, y_0). This is the prototype of Jacobi's eigenvalue method in the calculus of variations.

The Morse index. Suppose that $f'(x_0, y_0) = 0$ together with

$$\det f''(x_0, y_0) \neq 0.$$

The Morse index $i_f(x_0, y_0)$ of the map $f : J \to \mathbb{R}$ at the point (x_0, y_0) is equal to the number of negative eigenvalues of the matrix $f''(x_0, y_0)$, by definition.

- If $i_f(x_0, y_0) = 0$, then f has a local minimum at the point (x_0, y_0).
- If $i_f(x_0, y_0) = 2$, then f has a local maximum at the point (x_0, y_0).
- If $i_f(x_0, y_0) = 1$, then f has a saddle point at (x_0, y_0).

Constrained minimum problem and Lagrange multipliers. We are given the smooth functions $f, g : \mathbb{R}^2 \to \mathbb{R}$. We consider the minimum problem

$$f(x, y) = \min!, \quad g(x, y) = 0, \quad (x, y) \in \mathbb{R}^2. \tag{6.8}$$

[3] Fréchet (1878–1973).

[4] To prove this, note that, for fixed $h, k \in \mathbb{R}$, the function χ has a minimum at the point $\sigma = 0$. Hence $\chi'(0) = 0$. This implies $\delta f(x_0, y_0; h, k) = 0$ for all $h, k \in \mathbb{R}$. Therefore, $f'(x_0, y_0) = 0$.

[5] This is equivalent to the fact that the eigenvalues of $f''(x_0, y_0)$ are positive. In turn, this is equivalent to $f_{xx}(x_0, y_0) > 0$ together with $\det f''(x_0, y_0) > 0$.

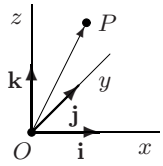


Fig. 6.2. Cartesian coordinate system

- Necessary condition for a constrained local minimum: If (x_0, y_0) is a local solution of (6.8) and $g'(x_0, y_0) \neq 0$, then there exists a real number λ such that

$$\boxed{f'(x_0, y_0) - \lambda g'(x_0, y_0) = 0.} \tag{6.9}$$

  The number λ is called a Lagrange multiplier.
- Sufficient condition for a constrained local minimum: Suppose that the point (x_0, y_0) satisfies the side condition $g(x_0, y_0) = 0$. Furthermore, suppose that there exists a real number λ such that the condition (6.9) holds and the definiteness condition

$$\delta^2 f(x_0, y_0; h, k) - \lambda \delta^2 g(x_0, y_0; h, k) > 0$$

  is satisfied for all $(h, k) \in \mathbb{R}^2$ with $h^2 + k^2 > 0$ and $\delta g(x_0, y_0; h, k) = 0$. Then the point (x_0, y_0) is a local solution of (6.8).

For example, let $f(x, y) := ax^2 + 2bxy + cy^2$ with real numbers a, b, c, and let $g(x, y) := x^2 + y^2$. Then each local solution (x_0, y_0) of (6.8) satisfies the eigenvalue equation (6.9), that is, $f_x(x_0, y_0) = \lambda g_x(x_0, y_0)$ and $f_y(x_0, y_0) = \lambda g_y(x_0, y_0)$. Explicitly,

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \lambda \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}.$$

Therefore, the Lagrange multiplier λ is an eigenvalue of the matrix A. All the proofs can be found in Zeidler (1986), Vol. III (see the references on page 1049).

6.2 The Motion of a Particle

The Euclidean manifold $\mathbb{E}^3$. We start with the motion

$$Q = Q(t), \qquad t \in \mathbb{R}$$

of a particle (e.g., the motion of a planet around the sun) in the 3-dimensional space of our intuition. Here, $Q(t)$ is the position of the particle at time t. The set of all positions Q is called the Euclidean manifold $\mathbb{E}^3$. In what follows, it is convenient to use the language of both the vector calculus and the theory of manifolds. First fix a point O as origin (e.g., the position of the sun). Then the motion can also be described by the vector equation

$$\mathbf{q} = \mathbf{q}(t), \qquad t \in \mathbb{R}.$$

Here, the position vector $\mathbf{q}(t)$ points from the origin O to the point $Q(t)$. We also write $\mathbf{q}(t) = \overrightarrow{OQ(t)}$. Let $T_O\mathbb{E}^3$ denote the space of all vectors with initial point at

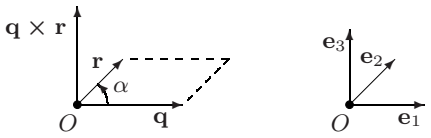


Fig. 6.3. Vector product

the origin O. This is a 3-dimensional real Hilbert space equipped with the inner product

$$\langle \mathbf{q}|\mathbf{r}\rangle_O := \mathbf{qr} \qquad \text{for all} \quad \mathbf{q}, \mathbf{r} \in T_O\mathbb{E}^3.$$

Recall that $\mathbf{qr} := |\mathbf{q}| \cdot |\mathbf{r}| \cdot \cos\alpha$. Here, $|\mathbf{q}|$ denotes the length of the vector $\mathbf{q}$, and α is the angle between the two vectors $\mathbf{q}$ and $\mathbf{r}$. The angle α is chosen in such a way that $0 \leq \alpha \leq \pi$ (Fig. 6.3). Similarly, let $T_Q\mathbb{E}^3$ denote the space of all vectors with the initial point Q. The space $T_Q\mathbb{E}^3$ is a real 3-dimensional Hilbert space equipped with the inner product

$$\langle \mathbf{v}|\mathbf{w}\rangle_Q := \mathbf{vw} \qquad \text{for all} \quad \mathbf{v}, \mathbf{w} \in T_Q\mathbb{E}^3.$$

At time t, the particle has the velocity vector

$$\dot{\mathbf{q}}(t) = \lim_{\Delta t \to 0} \frac{\mathbf{q}(t + \Delta t) - \mathbf{q}(t)}{\Delta t}$$

whose initial point coincides with the position $Q(t)$ of the particle at time t, that is, $\dot{\mathbf{q}}(t) \in T_{Q(t)}\mathbb{E}^3$ (Fig. 6.4). The space $T_Q\mathbb{E}^3$ consists of all the possible velocity vectors at the point Q. In the terminology of the theory of manifolds, the space $T_Q\mathbb{E}^3$ is called the tangent space of the Euclidean manifold $\mathbb{E}^3$ at the point Q. To simplify notation, we denote the 3-dimensional real Hilbert space $T_O\mathbb{E}^3$ by the symbol E^3. We then have the Hilbert space isomorphism

$$E^3 \simeq T_Q\mathbb{E}^3 \qquad \text{for all} \quad Q \in \mathbb{E}^3.$$

The Lie algebra of the Hilbert space $T_Q\mathbb{E}^3$. Let $\mathbf{q}, \mathbf{r} \in T_Q\mathbb{E}^3$. Recall that, by definition, the vector product

$$\mathbf{q} \times \mathbf{r}$$

is a vector of length $|\mathbf{q}| \cdot |\mathbf{r}| \cdot \sin\alpha$, and the three vectors $\mathbf{q}, \mathbf{r}, \mathbf{q} \times \mathbf{r}$ form a right-handed orthogonal system. The angle α between the vectors $\mathbf{q}$ and $\mathbf{r}$ is given as pictured in Fig. 6.3, that is, $0 \leq \alpha \leq \pi$. Moreover, $\mathbf{q} \times \mathbf{r} \in T_Q\mathbb{E}^3$. The difference between right-handed and left-handed systems of the three basis vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$

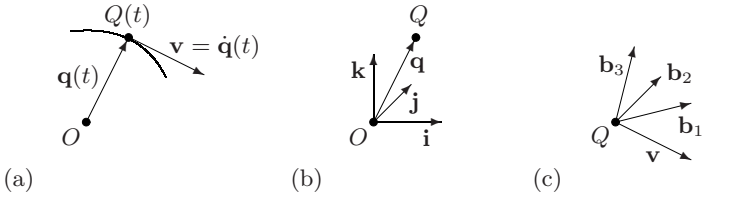


Fig. 6.4. Motion of a particle

(a) Right-handed (b) Left-handed

Fig. 6.5. Orientation

is pictured in Fig. 6.5. For all vectors $\mathbf{q}, \mathbf{p}, \mathbf{r} \in T_Q\mathbb{E}^3$ and all real numbers a, b, the vector product has the following properties:

(i) $\mathbf{q} \times \mathbf{r} \in T_Q\mathbb{E}^3$ (consistency).
(ii) $\mathbf{q} \times \mathbf{r} = -\mathbf{r} \times \mathbf{q}$ (anticommutative law).
(iii) $(a\mathbf{q} + b\mathbf{p}) \times \mathbf{r} = a(\mathbf{q} \times \mathbf{r}) + b(\mathbf{p} \times \mathbf{r})$ (distributive law).
(iv) $\mathbf{q} \times (\mathbf{p} \times \mathbf{r}) + \mathbf{p} \times (\mathbf{r} \times \mathbf{q}) + \mathbf{r} \times (\mathbf{q} \times \mathbf{p}) = 0$ (Jacobi identity).

Note that the vector product is *not* associative. The missing associativity is replaced by the Jacobi identity, which is based on cyclic permutations. Thus, with respect to the vector product $\mathbf{q} \times \mathbf{r}$, the linear real linear 3-dimensional space $T_Q\mathbb{E}^3$ becomes a real Lie algebra. We will show in Sect. 6.12 that this Lie algebra is isomorphic to the Lie algebra $su(E^3)$ of infinitesimal rotations.

6.3 Newtonian Mechanics

> The rise of modern science was accompanied with the replacement of authorities or traditions by causes in explaining phenomena. One of the ultimate goals of science is to understand the world, and this is approached by scientific explanation, that is, by finding out causes for various phenomena. According to Aristotle, however, there are different kinds of cause: material, formal, efficient, and final causes. Before the rise of modern science, teleological explanation based on the notion of final cause was a dominant mode of explanation. With the revival of Neoplatonism, Archimedianism and atomism in the Renaissance, there began a transformation in basic assumptions of scientific explanation. Copernicus, Kepler, Galileo, and Descartes, for example, believed that the underlying truth and universal harmony of the world can be perfectly represented by simple and exact mathematical expressions. The mathematization of nature led to a certain degree of popularity of formal cause. But the most popular and powerful conception of causality, in fighting, against the teleological explanation, was a mechanical one based on the notion of efficient cause. Different from final and formal causes, the idea of efficient cause focuses on how the cause is transmitted to the effect, that is, on the mode of transmission. According to the classical mechanical view, causality can be reduced to the laws of motion of bodies in space and time...
>
> Tian Yu Cao, 1998
> *Conceptual Developments of 20th Century Field Theories* [6]

[6] Cambridge University Press, Cambridge, United Kingdom, 1998 (reprinted with permission).

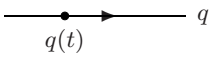


Fig. 6.6. Motion on the real line

> So that we may say now that the door is opened, for the first time, to a new method fraught with numerous and wonderful results which in future years will command the attention of other minds.
>
> Galileo Galilei (1564–1642)

> *Lex prima*: A stationary body will remain motionless, and a moving body will continue to move in the same direction with unchanging speed unless it is acted on by some force.
> *Lex secunda*: The time-rate-of-change of the momentum of a body is proportional to the force.
> *Lex tertia*: If any body exerts a force on another object, then the second object also exerts an equal and opposite force on the first.
> It remains that, from the same principles, I now demonstrate the frame of the System of the World.
>
> Isaac Newton (1643–1727)
> *Philosophiae Naturalis Principia Mathematica,* London, 1687[7]

> Who, by a vigor of mind almost divine, the motions and figures of the planets, the paths of comets, and the tides of the sea first demonstrated.[8]
>
> Newton's Epitaph, Westminster Abbey, London

> When one considers all that Newton achieved, and the cultural and scientific environment in which he achieved it, there is reason to regard him as the greatest scientist – and perhaps the greatest genius – that ever lived.
>
> Anthony Philip French

The motion $q = q(t)$ of a classical point particle of mass m on the real line is described by the Newtonian equation "time-derivative of momentum equals force,"

$$\boxed{\dot{p}(t) = F(q(t)),} \tag{6.10}$$

together with the initial condition $q(t_0) = q_0$ and $\dot{q}(t_0) = v_0$ (Fig. 6.6). Here, the position q_0 and the velocity v_0 of the particle are prescribed at the initial time t_0. It turns out that many problems in physics can be essentially simplified by using potentials. To explain this in the present situation, use the force function F in order to define

$$W := \int_{q_0}^{q} F(x)dx,$$

and $U(q) := U(0) - W$. Hence

$$\boxed{F(q) = -U'(q).}$$

[7] See the footnote on page 12.

[8] Newton's grave in Westminster Abbey is framed by five smaller gravestones with famous names: Michael Faraday (1791–1867), George Green (1793–1841), Sir William Thomson (Lord Kelvin of Largs) (1824–1907), James Clerk Maxwell (1831–1879) and Paul Dirac (1902–1984).

The function U is called a potential of the force F. Note that the normalization constant $U(0)$ of the potential can be chosen arbitrarily. A change of $U(0)$ represents the simplest *gauge transformation* in physics. We will show later on that gauge transformations play a fundamental role in modern elementary particle physics (gauge field theory). Using the trajectory $q = q(t)$, we also define

$$\boxed{E(t) := \tfrac{1}{2}m\dot{q}(t)^2 + U(q(t)).}$$

In what follows, we will use the following terminology:

$q(t)$	position of the particle at time t,
$\dot{q}(t)$	velocity at time t,
$\ddot{q}(t)$	acceleration at time t,
$p(t) := m\dot{q}(t)$	momentum at time t,
$F(q)$	force acting at the point q,
W	work done by the force if the particle moves from the initial point q_0 to the final point q,
$U(q)$	potential energy of the particle at position q,
$\frac{1}{2}m\dot{q}(t)^2$	kinetic energy of the particle at time t,
$E(t)$	total energy of the particle at time t.

Constant mass. Let us first consider mass points of constant positive mass m. The Newtonian equation of motion reads then as

$$\boxed{m\ddot{q}(t) = F(q(t)), \quad t \in \mathbb{R}, \quad q(t_0) = q_0,\ \dot{q}(t_0) = v_0.} \tag{6.11}$$

This corresponds to "force equals mass times acceleration". For a smooth force $F : \mathbb{R} \to \mathbb{R}$, it follows from the general theory of ordinary differential equations that there exists a maximal open time interval J containing the initial time t_0 such that the initial-value problem (6.11) has a unique solution $q = q(t)$ on J. Such a solution is called a trajectory of the particle. For example, the initial-value problem

$$\ddot{x}(t) = 2x(t)\dot{x}(t), \qquad x(0) = 0, \ \ \dot{x}(0) = 1$$

has the unique solution $x(t) = \tan t, -\frac{\pi}{2} < t < \frac{\pi}{2}$, which blows up as time t goes to $\frac{\pi}{2}$ from the left. This example shows that the trajectories do not always exist for all times.

Theorem 6.1 *The total energy E of the particle is constant during the motion.*

Proof. Differentiation with respect to time t yields

$$\dot{E}(t) = m\ddot{q}(t)\dot{q}(t) + U'(q(t))\dot{q}(t) = \big(m\ddot{q}(t) - F(q(t))\big)\dot{q}(t) = 0.$$

Hence $E(t)$ is constant on the time interval J. □

Example 6.2 (The falling stone). *Consider the constant force $F := -mg$. Then, the unique solution of the equation of motion (6.11) is given by*

$$q(t) = q_0 + v_0 t - \tfrac{1}{2}gt^2 \qquad \textit{for all times} \quad t \in \mathbb{R}. \tag{6.12}$$

This law of falling bodies was first discovered by Galilei in about 1600. If the initial velocity is negative, $v_0 < 0$, equation (6.12) describes a falling stone on earth of mass m. Here, $q(t)$ denotes the height of the stone at time t, and $g = 9.81\ \mathrm{m/s^2}$ is the so-called acceleration constant on earth. The potential energy of the stone is given by

$$U(q) = mgq,$$

provided we use the normalization condition $U(0) := 0$. If the stone is falling down from the height $q_0 > 0$ to the ground $q = 0$, then the stone looses the potential energy mgq_0. The stone possesses the constant total energy

$$E = \tfrac{1}{2}mv_0^2 + mgq_0$$

during the motion. If the stone hits the ground, then the total (mechanical) energy E is converted into heat energy.

Loss of mass. If the mass $m(t)$ changes in time, then we get the momentum $p(t) = m(t)q(t)$, and the Newtonian equation of motion reads as

$$\boxed{\frac{d}{dt}\big(m(t)q(t)\big) = F(q(t)).} \tag{6.13}$$

This models a rocket loosing mass by burning fuel. Einstein discovered that the mass of each moving body depends on its velocity,

$$m(t) = \frac{m_0}{\sqrt{1 - \dot{q}(t)^2/c^2}}.$$

Here, the symbols m_0 and c denote the rest mass of the particle and the speed of light in a vacuum, respectively. Equation (6.13) passes then over to the basic equation of relativistic motion. In this chapter, we will only consider point particles having time-independent constant positive mass m (non-relativistic motion).

The harmonic oscillator. The equation

$$\boxed{\ddot{q} + \omega^2 q = 0,} \tag{6.14}$$

along with the initial condition $q(0) = q_0$, $\dot{q}(0) = v_0$, describes the motion of a so-called harmonic oscillator. We are given the positive constant ω. The unique solution of this initial-value problem reads as

$$\boxed{q = q_0 \cos \omega t + \frac{v_0}{\omega} \sin \omega t,} \tag{6.15}$$

for all real times t. This motion has the time period T, where

$$T = \frac{2\pi}{\omega}.$$

The quantities ω and $\nu := \omega/2\pi$ are called angular frequency and frequency, respectively. Hence $T = 1/\nu$.

The harmonic oscillator represents the simplest oscillating system. Let us motivate this. Consider a point of constant positive mass m under the action of the repulsive force

$$F(q) = -\alpha q.$$

Here, α is a positive constant. Introducing the constant $\omega := \alpha/m$, equation (6.14) corresponds to the Newtonian equation, $m\ddot{q} = -\alpha q$. The universality of the harmonic oscillator follows from the fact that each smooth force F, with $F(0) = 0$ and the symmetry property $F(-q) = -F(q)$ for all real q, allows the Taylor expansion

$$F(q) = -\alpha q - \beta q^3 + ... \qquad \text{for small} \quad q \in \mathbb{R}.$$

If the force F is repulsive, then $q > 0$ implies $F(q) < 0$. Hence $\alpha > 0$ and $\beta \geq 0$. The first approximation reads as $F(q) = -\alpha q$. If we set $\alpha := m\omega^2$, then we obtain a force which precisely corresponds to the harmonic oscillator. This force has the potential

$$U = \tfrac{1}{2} m\omega^2 q^2.$$

For the motion of the harmonic oscillator, energy conservation means that the function (6.15) satisfies the condition

$$\boxed{\tfrac{1}{2} m\dot{q}(t)^2 + \tfrac{1}{2} m\omega^2 q(t)^2 = E,}$$

for all times t, along with the constant total energy $E := \frac{1}{2} m v_0^2 + \frac{1}{2} m\omega^2 q_0^2$.

Anharmonic oscillator. The second approximation of the force F reads as

$$F(q) = -\alpha q - \beta q^3$$

with the positive constant β. We set $\beta = m\kappa$, and call κ the coupling constant. Now the equation of motion reads as

$$\boxed{\ddot{q}(t) + \omega^2 q(t) = -\kappa q(t)^3.} \tag{6.16}$$

The potential is given by $U(q) := \frac{1}{2} m\omega^2 q^2 + \frac{1}{4} m\kappa q^4$ with the constant total energy

$$E := \tfrac{1}{2} m v_0^2 + \tfrac{1}{2} m\omega^2 q_0^2 + \tfrac{1}{4} m\kappa q_0^4.$$

6.4 A Glance at the History of the Calculus of Variations

> Bees – by virtue of a certain geometrical forethought – know that the hexagon is greater than the square and the triangle and will hold more money for the same expenditure of material.
>
> Pappus of Alexandria, 300 B.C.

> Every process in nature will occur in the shortest possible way.
>
> Leonardo da Vinci (1452–1519)

> A light ray between two points needs the shortest possible time.
>
> Pierre de Fermat (1601–1665)

> Johann Bernoulli, professor of mathematics, greets the most sophisticated mathematicians in the world. Experience shows that noble intellectuals are driven to work for the pursuit of the knowledge by nothing more than being confronted with difficult and useful problems.
> Six months ago, in the June edition of the Leipzig *Acta eruditorum* (journal of scientists), I presented such a problem. The allotted six-month deadline has now gone by, but no trace of a solution has appeared. Only the famous

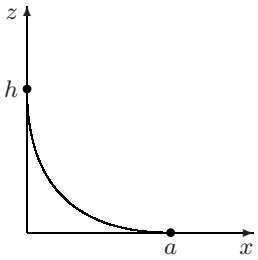


Fig. 6.7. Johann Bernoulli's brachistochrone

Leibniz informed me that he had unravelled the knot of this brilliant and outstanding problem, and he kindly asked me to extend the deadline until next Easter. I agreed to this honorable request... I will repeat the problem here once more.
Two points, at different distances from the ground, and not in a vertical line, should be connected by such a curve that a body under the influence of gravitational forces passes in the shortest possible way from the upper to the lower point (Fig. 6.7).[9]

Johann Bernoulli, January 1697

This paper solves my brother's problem, to whom I will set other problems in return.

Jakob Bernoulli, May 1697

The Euler *Calculus of Variations* (Methodus inveniendi) from the year 1744 is one of the most beautiful mathematical works that has ever been written.

Constantin Carathéodory (1873–1950)

Read Euler, he is the master of us all.

Marquise de Pierre Simon Laplace (1749–1824)

One needs to have delved but little into the principles of differential calculus to know the method of how to determine the greatest and least ordinates of curves. But there are maxima or minima problems of a higher order, which in fact depend on the same method, which however cannot be subjected to this method. These are the problems where it is a matter of finding the curves themselves.
The first problem of this type, which the geometers solved, is that of the brachistochrone or the curve of fastest fall which Johann Bernoulli proposed toward the end of the preceding century. One attained this only in special ways, and it was only some time later and on the occasion of the investigations concerning isoperimetric problems that the great geometer

[9] We are given the initial point $(0, h)$ and the final point $(a, 0)$ of the unknown curve. The solution is then the arc of a cycloid,

$$x = C(p - \sin p), \quad z = h + C(\cos p - 1), \qquad 0 \le p \le p_0,$$

where the constant C is determined by the final point. The proof can be found in Zeidler (1995b), p. 132 (see the references on page 1049).

> of whom we just spoke and his extraordinary brother Jakob Bernoulli gave some rules in order to solve several other problems of this type.
> But since these rules were not of sufficient generality, the famous Euler undertook to refer all investigations of this type to a general method. [10] But even as sophisticated and fruitful as his method is, one must nevertheless confess that it is not sufficiently simple... Now here one finds a method which requires only a simple use of the principles of differential and integral calculus.
>
> Joseph Louis Lagrange, 1762

> As I see, your analytic solution of the isoperimetric problem contains all that one can wish for in this situation. I am very happy that this theory which I have treated since the first attempts almost alone, has been brought precisely by you to the highest degree of perfection.
> The importance of the situation has occasioned me with the help of your new insights to myself conceive of an analytic solution, but which I shall not make known before you have published your deliberations, in order not to deprive you of the least part of the fame due you.
>
> Euler, in a letter to the young Lagrange[11]

6.5 Lagrangian Mechanics

> The mathematician is perfect only in so far as he is a perfect being, in so far as he perceives the beauty of truth; only then will his work be thorough, transparent, comprehensive, pure, clear, attractive, and even elegant. All this is necessary in order to resemble Lagrange.
>
> Johann Wolfgang von Goethe (1749–1832)
> *Wilhelm Meisters Wanderjahre*

We start with the variational problem

$$\boxed{\int_{t_0}^{t_1} L(q(t), \dot{q}(t), t)\, dt = \text{critical !}} \tag{6.17}$$

along with the boundary condition $q(t_0) = q_0$ and $q(t_1) = q_1$. Here, the finite interval $[t_0, t_1]$ and the boundary positions q_0 and q_1 are given. The function

$$L := \text{kinetic energy minus potential energy} \tag{6.18}$$

[10] Euler used an involved difference method. He could merely apply his method to one-dimensional variational problems.

[11] Euler was born in Basel (Switzerland) in 1707, and he studied at the University of Basel. In 1727 he moved to the newly founded Academy in St. Petersburg (Russia). From 1741 until 1766 Euler worked at the Berlin Academy. In 1766, he moved back to Saint Petersburg where he died in 1783.
The Prussian king Frederyck the Great (1712–1786) appointed Lagrange as Euler's successor at the Berlin Academy. In 1786 Lagrange returned from Berlin to Paris. In 1788 he published his famous treatise "Méchanique analitique" about the foundations of celestial mechanics. The title changed to "Mécanique analytique" in 1813, after the reform of the French orthography initiated by the French Academy.

is called the Lagrangian, and

$$S[q] := \int_{t_0}^{t_1} L(q(t), \dot{q}(t), t)\, dt$$

is called the action of the motion along the given trajectory $q = q(t)$ during the fixed time interval $[t_0, t_1]$. The action S has the physical dimension of energy times time.

6.5.1 The Harmonic Oscillator

As an example, let us consider the special case of the harmonic oscillator where

$$L(q, \dot{q}) := \tfrac{1}{2} m\dot{q}^2 - \tfrac{1}{2} m\omega^2 q^2.$$

The constants m and ω are assumed to be positive. The following theorem tells us that the Euler–Lagrange equation to the variational problem (6.17) coincides with the Newtonian equation of motion for the harmonic oscillator.

Proposition 6.3 *Each smooth solution of* (6.17) *satisfies the Euler–Lagrange equation* $\ddot{q} + \omega^2 q = 0$.

Proof. To simplify notation, let us set $m = \omega := 1$. The following argument is typical for all kind of variational problems. Recall that $\mathcal{D}(t_0, t_1)$ denotes the set of all smooth functions $h :]t_0, t_1[\to \mathbb{R}$, which have compact support, that is, each function h vanishes in some neighborhood of the boundary points. In particular, $h(t_0) = h(t_1) = 0$. The functions h are called test functions.

(I) Minimum problem. Replace first "critical!" by "min!". Suppose that the function $q = q(t)$ is a solution of (6.17). The basic idea due to Lagrange consists in reducing the original variational problem to a simpler problem for a real function χ. To this end, fix a function $h \in \mathcal{D}(t_0, t_1)$, and consider the following family of functions,

$$r(t) := q(t) + \sigma h(t), \qquad t_0 \le t \le t_1,$$

which depends on the real parameter σ. Since $h(t_0) = h(t_1) = 0$, the function $r = r(t)$ satisfies the boundary condition $r(t_0) = q_0,\ r(t_1) = q_1$ from (6.17). Set $\chi(\sigma) := \int_{t_0}^{t_1} L(r(t), \dot{r}(t))\, dt$. Explicitly,

$$\chi(\sigma) := \int_{t_0}^{t_1} \tfrac{1}{2}\{(\dot{q}(t) + \sigma\dot{h}(t))^2 - (q(t) + \sigma h(t))^2\}\, dt,$$

for all $\sigma \in \mathbb{R}$. If $\sigma = 0$, then $r = q$. Since q is a solution of the minimum problem (6.17) and the function r is admissible (i.e., r satisfies the boundary condition of the variational problem), we get

$$\chi(\sigma) \ge \chi(0) \qquad \text{for all} \quad \sigma \in \mathbb{R}.$$

Thus, the real function χ has a minimum at the point $\sigma = 0$. Hence we get the key relation

$$\boxed{\chi'(0) = 0.} \tag{6.19}$$

Differentiation with respect to the parameter σ yields

$$\chi'(0) = \int_{t_0}^{t_1} (\dot{q}\dot{h} - qh)dt.$$

Integration by parts implies the so-called variational equation[12]

$$\chi'(0) = \int_{t_0}^{t_1} \big(-\ddot{q}(t) - q(t)\big)h(t)\ dt = 0. \tag{6.20}$$

This is true for all test functions $h \in \mathcal{D}(t_0, t_1)$. Therefore, the variational lemma (see Sect. 7.20.2 of Vol. I) tells us that $\ddot{q}(t) + q(t) = 0$ on $[t_0, t_1]$.

(II) Critical point. By definition, problem (6.17) means that each of the functions χ introduced above is critical at the point $\sigma = 0$. This implies (6.19). We now proceed as in (I). □

First variation. Motivated by Sect. 6.1, we set $\delta S(q; h) := \chi'(0)$. Explicitly,

$$\delta S(q, h) = \int_{t_0}^{t_1} \big(\dot{q}(t)\dot{h}(t) - q(t)h(t)\big)\ dt \qquad \text{for all} \quad h \in \mathcal{D}(t_0, t_1).$$

This is called the first variation of the action functional S at the point q in direction of h.

First functional derivative. After integrating by parts, we obtain

$$\delta S(q, h) = \frac{\delta S[q]}{\delta q}(h) \qquad \text{for all} \quad h \in \mathcal{D}(t_0, t_1) \tag{6.21}$$

with $\frac{\delta S[q]}{\delta q}(h) := \int_{t_0}^{t_1} \frac{\delta S[q]}{\delta q(t)}\ h(t)dt$ and

$$\frac{\delta S[q]}{\delta q(t)} := -\ddot{q}(t) - q(t) \qquad \text{for all} \quad t \in [t_0, t_1].$$

Here, the functional $h \mapsto \frac{\delta S[q]}{\delta q}(h)$ is called the functional derivative of the action functional S at the point q, and the function $t \mapsto \frac{\delta S[q]}{\delta q(t)}$ is called the local functional derivative of the action functional S at the point q and at time t. The Euler–Lagrange equation can elegantly be written as

$$\boxed{\frac{\delta S[q]}{\delta q(t)} = 0, \qquad t_0 \le t \le t_1.} \tag{6.22}$$

Second variation. Define $\delta^2 S(q; h) := \chi''(0)$. Explicitly,

$$\delta^2 S(q; h) = \int_{t_0}^{t_1} \big(\dot{h}(t)^2 - h(t)^2\big)\ dt.$$

This is called the second variation of the action functional S at the point q in direction of h. Integration by parts yields

$$\delta^2 S(q; h) = -\int_{t_0}^{t_1} \Big(\ddot{h}(t) + h(t)\Big)h(t)\ dt \qquad \text{for all} \quad h \in \mathcal{D}(t_0, t_1).$$

It follows from $\chi(1) = \chi(0) + \chi'(0) + \frac{1}{2}\chi''(0)$ that

[12] Note that the boundary term $\dot{q}(t_1)h(t_1) - \dot{q}(t_0)h(t_0)$ vanishes, since the function h vanishes at the boundary.

$$S[q+h] = S[q] + \delta S(q,h) + \tfrac{1}{2}\delta^2 S(q;h)$$

for all smooth functions $q : [t_0, t_1] \to \mathbb{R}$, and all test functions $h \in \mathcal{D}(t_0, t_1)$. Physicists write $\delta q := h$. If q is a critical point of S (i.e., q is a solution of the Euler–Lagrange equation), then $\delta S(q;h) = 0$. Hence

$$S[q+h] = S[q] + \tfrac{1}{2}\delta^2 S(q;h) \qquad \text{for all} \quad h \in \mathcal{D}(t_0, t_1). \tag{6.23}$$

Consequently, q is a minimum of S if

$$\delta^2 S(q;h) \geq 0$$

for all test functions $h \in \mathcal{D}(t_0, t_1)$. In the next section, we will use this observation in order to prove the Jacobi eigenvalue criterion for the existence of a minimum.

Second functional derivative. Naturally enough, we define

$$\frac{\delta^2 S[q]}{\delta q^2}(h,k) := \frac{d}{d\sigma}\frac{\delta S[q+\sigma k]}{\delta q}(h)_{|\sigma=0} \qquad \text{for all} \quad h,k \in \mathcal{D}(t_0,t_1).$$

Explicitly, after integrating by parts,

$$\frac{\delta^2 S[q]}{\delta q^2}(h,k) = \int_{t_0}^{t_1} \Big(\dot{k}(t)\dot{h}(t) - k(t)h(t)\Big)\, dt = \int_{t_0}^{t_1} \Big(-\ddot{k}(t) + k(t)\Big)h(t)\, dt$$

for all $h, k \in \mathcal{D}(t_0, t_1)$. The bilinear functional $(h,k) \mapsto \frac{\delta^2 S[q]}{\delta q^2}(h,k)$ is called the second functional derivative of the action functional S at the point q. In particular, for the second variation we get

$$\delta^2 S(q;h) = \frac{\delta^2 S[q]}{\delta q^2}(h,h) \qquad \text{for all} \quad h \in \mathcal{D}(t_0,t_1).$$

6.5.2 The Euler–Lagrange Equation

Suppose that the Lagrangian L is smooth. The same argument as in the proof of Prop. 6.3 yields the following fundamental result due to Lagrange.

Theorem 6.4 *Each smooth solution of the variational problem* (6.17) *satisfies the Euler–Lagrange equation*

$$\boxed{\frac{d}{dt}L_{\dot{q}} = L_q.} \tag{6.24}$$

Explicitly, this means

$$\frac{d}{dt}L_{\dot{q}}(q(t), \dot{q}(t), t) = L_q(q(t), \dot{q}(t), t).$$

If the trajectory $q = q(t)$ has several components, $q = (q_1, ..., q_M)$, then we have to write down the Euler–Lagrange equation for each component. Explicitly,

$$\frac{d}{dt}L_{\dot{q}_m} = L_{q_m}, \qquad m = 1, ..., M. \tag{6.25}$$

Moreover, Lagrange's elegant argument can be immediately generalized to multi-dimensional integrals. Examples can be found in the problem section to Chap. 14 of Vol. I. As an introduction to the calculus of variations and optimization theory, we recommend the author's textbook Zeidler (1986), Vol. III (see the references on page 1049).

6.5.3 Jacobi's Accessory Eigenvalue Problem

> Returning to the concepts of maximum and minimum, it is a nuisance that there reigns such confusion in these words. One says that an expression attains a maximum or a minimum if one simply wishes to say that it is critical (or extremal) and hence its first variation vanishes, also in the case when neither a minimum nor a maximum occurs.
>
> Carl Gustav Jacobi (1804–1851)

Let us study the principle of least action for the harmonic oscillator

$$S[q] := \int_{t_0}^{t_1} \tfrac{1}{2}\Big(\dot{q}^2(t) - \omega^2 q^2(t)\Big)dt = \min! \tag{6.26}$$

together with the boundary condition $q(t_0) = q_0, q(t_1) = q_1$. Here, we set $m := 1$. The corresponding Euler–Lagrange equation reads as

$$\ddot{q}(t) + \omega^2 q(t) = 0, \quad t \in [t_0, t_1], \quad q(t_0) = q_0,\ q(t_1) = q_1. \tag{6.27}$$

Following Jacobi, we add the accessory eigenvalue problem

$$-\ddot{h}(t) - \omega^2 h(t) = \lambda h(t), \quad t \in [t_0, t_1], \quad h(t_0) = h(t_1) = 0. \tag{6.28}$$

The eigensolutions are given by

$$h_n(t) := \sin\frac{n\pi(t-t_0)}{t_1 - t_0}, \qquad \lambda_n := \frac{n^2\pi^2}{(t_1-t_0)^2} - \omega^2, \qquad n = 1, 2, \dots \tag{6.29}$$

Proposition 6.5 *Let $q = q(t)$ be a solution of the Euler–Lagrange equation (6.27), and let the smallest eigenvalue λ_1 of (6.28) be positive. Then the function q is a solution of the minimum problem (6.26).*

For example, choose $t_0 := 0, t_1 := \pi, q_1 = q_0 + 1$, and $\omega := \frac{1}{2}$. Then, we have $\lambda_1 > 0$, and hence the function $q = q_0 + \sin\omega t$ is a solution of (6.26).

Proof. Let us choose the real Hilbert space $X := L_2(t_0, t_1)$ with the inner product

$$\langle h|k\rangle := \int_{t_0}^{t_1} h(t)k(t)\,dt.$$

We will use the operator $A : D(A) \to X$ defined by

$$Ah := -\ddot{h} - \omega^2 h \qquad \text{for all} \quad h \in X.$$

Here, $D(A)$ is defined to be the set of all twice continuously differentiable functions $h : [t_0, t_1] \to \mathbb{R}$ with $h(t_0) = h(t_1) = 0$. A classical result tells us[13] that the eigenfunctions $h_1, h_2, \dots$ of the operator A form a complete orthonormal system in the Hilbert space X, that is,

$$h = \sum_{n=1}^{\infty} \langle h_n|h\rangle h_n \qquad \text{for all} \quad h \in X.$$

[13] We refer to Zeidler (1995a), Sect. 4.5 (see the references on page 1049).

Integration by parts shows that $\langle Ah|k\rangle = \langle h|Ak\rangle$ for all $h, k \in D(A)$, that is, the operator A is symmetric. Hence $\langle h_n|Ah\rangle = \langle Ah_n|h\rangle = \lambda_n\langle h_n|h\rangle$. This implies

$$Ah = \sum_{n=1}^{\infty}\langle h_n|Ah\rangle h_n = \sum_{n=1}^{\infty}\lambda_n\langle h_n|h\rangle h_n \qquad \text{for all} \quad h \in D(A).$$

Therefore, $\langle Ah|h\rangle = \sum_{n=1}^{\infty}\lambda_n\langle h_n|h\rangle^2$. This yields the key relation

$$\boxed{\delta^2 S(q;h) = \sum_{n=1}^{\infty}\lambda_n\langle h_n|h\rangle^2 \qquad \text{for all} \quad h \in D(A).} \tag{6.30}$$

By (6.23),

$$S[q+h] = S[q] + \tfrac{1}{2}\delta^2 S(q;h) \qquad \text{for all} \quad h \in D(A).$$

If $\lambda_1 > 0$, then $\delta^2 S(q;h) \geq 0$. Hence $S[q+h] \geq S[q]$ for all $h \in D(A)$. □

6.5.4 The Morse Index

> The Morse index describes the global behavior of the action functional of the harmonic oscillator with respect to arbitrary time intervals. This global behavior is governed by the appearance of focal points of the trajectories of the harmonic oscillator, which correspond to focal points in geometric optics.
>
> Folklore

We are given the time interval $[t_0, t_1]$. Suppose that all of the eigenvalues $\lambda_1, \lambda_2, \ldots$ introduced in (6.28) are different from zero. Motivated by Sect. 6.1 and the key relation (6.30), the Morse index $\mu(t_0, t_1)$ of the second variation $h \mapsto \delta^2 S(q;h)$ on the time interval $[t_0, t_1]$ is defined to be the number of negative eigenvalues $\lambda_1, \lambda_2, \ldots$[14] Explicitly, for $n = 0, 1, 2, \ldots$, we obtain

$$\boxed{\mu(t_0, t_1) = n \qquad \text{for all} \quad t_1 \in \left]t_0 + \frac{n\pi}{\omega},\ t_0 + \frac{(n+1)\pi}{\omega}\right[.} \tag{6.31}$$

To discuss this, consider first the case where

$$t_0 < t_1 < t_0 + \frac{\pi}{\omega}.$$

Then $0 < \lambda_1 < \lambda_2 < \ldots$ Thus, all the eigenvalues $\lambda_1, \lambda_2, \ldots$ are positive. Hence

$$\mu(t_0, t_1) = 0.$$

Suppose that the function $q = q(t)$ is a solution of the Euler–Lagrange equation (6.27). By Prop. 6.5, the function $q = q(t)$ is a solution of the minimum problem (6.26).

Consider now the case where $t_0 + \frac{n\pi}{\omega} < t_1 < \frac{(n+1)\pi}{\omega}$ with $n = 1, 2, \ldots$ The we have $\lambda_1 < 0$. Hence

$$S[q+h_1] = S[q] + \tfrac{1}{2}\delta^2 S(q;h_1) = S[q] + \tfrac{1}{2}\lambda_1 < S[q].$$

[14] Morse (1892–1977).

In this case, the function $q = q(t)$ is not a solution of the minimum problem (6.26) on the time interval $[t_0, t_1]$.

Observe that, by (6.29), the Morse index jumps at the critical time points

$$t_{\text{crit}} := t_0 + \frac{n\pi}{\omega}, \qquad n = 1, 2, \dots$$

The Morse index plays a crucial role for the harmonic oscillator in quantum mechanics, where it is responsible for the jumps of the Feynman propagator (see Sect. 7.9.4 on page 576). In quantum mechanics, the Morse index is also called the Maslov index.

6.5.5 The Anharmonic Oscillator

Let $U : \mathbb{R} \to \mathbb{R}$ be a given smooth function. By definition, an anharmonic oscillator has the potential

$$U(q) := \tfrac{1}{2} m\omega^2 q^2 + \tfrac{1}{4} m\kappa q^4$$

with the positive constants m (mass), ω (angular frequency), and κ (coupling constant). The action functional to the potential U is given by

$$S[q] := \int_{t_0}^{t_1} \Big(\tfrac{1}{2} m\dot{q}^2(t) - U\big(q(t)\big) \Big) dt.$$

Setting $\chi(\sigma) := S[q + \sigma h]$ with the real parameter σ, we get the nth variation $\delta^n S(q, h) := \chi^{(n)}(0)$ where $n = 1, 2, \dots$

In what follows the function $q : [t_0, t_1] \to \mathbb{R}$ is assumed to be smooth.

Choose the test function $h \in \mathcal{D}(t_0, t_1)$. Explicitly, after integrating by parts, we get the first variation

$$\delta S(q; h) = \frac{\delta S[q]}{\delta q}(h) := \int_{t_0}^{t_1} \frac{\delta S[q]}{\delta q(t)} \cdot h(t) dt$$

with the local first functional derivative $\frac{\delta S[q]}{\delta q(t)} := -m\ddot{q}(t) - U'(q(t))$, and the second variation

$$\begin{aligned} \delta^2 S(q; h) &= \int_{t_0}^{t_1} \Big(m\dot{h}(t)^2 - U''\big(q(t)\big) h(t)^2 \Big) dt \\ &= \int_{t_0}^{t_1} \Big(- m\ddot{h}(t) - U''\big(q(t)\big) h(t) \Big) h(t) dt. \end{aligned}$$

This is a quadratic form with respect to h. Introducing the second functional derivative $\frac{\delta^2 S[q]}{\delta q^2}(h, k) := \frac{d}{d\sigma} \frac{\delta S[q + \sigma k]}{\delta q}(h)_{|\sigma = 0}$, we obtain

$$\frac{\delta^2 S[q]}{\delta q^2}(h, k) = \int_{t_0}^{t_1} \Big(- m\ddot{k}(t) - U''\big(q(t)\big) k(t) \Big) h(t) dt$$

for all $k, h \in \mathcal{D}(t_0, t_1)$. Furthermore, $\delta^2 S(q; h) = \frac{\delta^2 S[q]}{\delta q^2}(h, h)$. It follows from the Taylor expansion $\chi(1) = \chi(0) + \chi'(0) + \frac{1}{2}\chi''(\vartheta)$ with $0 < \vartheta < 1$ that

$$S[q + h] = S[q] + \delta S(q; h) + \delta^2 S(q + \vartheta h; h).$$

The principle of critical action versus the principle of least action. Following Jacobi, let us compare the principle of critical action

$$S[q] = \text{critical!}, \qquad q(t_0) = q_0,\ q(t_1) = q_1 \tag{6.32}$$

with the principle of least action

$$S[q] = \min!, \qquad q(t_0) = q_0,\ q(t_1) = q_1. \tag{6.33}$$

We say that the function $q : [t_0, t_1] \to \mathbb{R}$ is a local minimum of the action functional S (i.e., a local solution of (6.33)) iff it satisfies the boundary condition and, for each test function $h \in \mathcal{D}(t_0, t_1)$, there exists a positive number σ_0 such that

$$S[q + \sigma h] \geq S[q] \qquad \text{for all} \quad \sigma \in [-\sigma_0, \sigma_0].$$

The accessory eigenvalue problem reads as

$$-m\ddot{h}(t) - U'\big(q(t)\big)h(t) = \lambda h(t), \quad t \in [t_0, t_1], \quad h(t_0) = h(t_1) = 0. \tag{6.34}$$

This classical boundary-eigenvalue problem has a complete orthonormal system $h_1, h_2, \ldots$ of eigenfunctions in the real Hilbert space $L_2(t_0, t_1)$ with the corresponding simple eigenvalues $\lambda_1 < \lambda_2 < \ldots$ Moreover,

$$\delta S^2(q; h) = \sum_{n=1}^{\infty} \lambda_n \langle h_n | h \rangle^2 h_n,$$

where $\langle h_n | h \rangle = \int_{t_0}^{t_1} h_n(t)h(t)dt$. Suppose that all of the eigenvalues are different from zero. By definition, the Morse index $i_S(q)$ is equal to the number of negative eigenvalues. The smallest eigenvalue λ_1 is the minimal value of the constrained quadratic minimum problem

$$\delta^2 S(q; h) = \min, \qquad \int_{t_0}^{t_1} h(t)^2 dt = 1, \qquad h \in C_0^2[t_0, t_1],$$

where $C_0^2[t_0, t_1]$ denotes the set of all twice continuously differentiable functions $h : [t_0, t_1] \to \mathbb{R}$ with $h(t_0) = h(t_1) = 0$.

(i) Necessary condition for a local minimum: If q is a solution of (6.32) (e.g., q is a local solution of (6.33)), then q is a solution of the Euler–Lagrange equation

$$\frac{\delta S[q]}{\delta q(t)} = 0, \quad t_0 \leq t \leq t_1, \quad q(t_0) = q_0, \ \ q(t_1) = q_1, \tag{6.35}$$

that is, $m\ddot{q}(t) = -U'(q(t))$.

(ii) Sufficient condition for a local minimum: Conversely, suppose that the function q is a solution of (6.35) and the Morse index $i_s(q)$ is equal to zero. Then the function q is a local minimum of the action functional S. If the Morse index $i_S(q)$ is positive, then q is not a local minimum of the action functional S.

The linearized Euler–Lagrange equation and Jacobi fields. Choose the test function $h \in \mathcal{D}(t_0, t_1)$, and replace the Euler–Lagrange equation

$$m\ddot{q} + U'(q) = 0$$

by $m(\ddot{q}+\sigma\ddot{h})+U'(q+\sigma h) = 0$. Differentiation with respect to the real parameter σ at $\sigma = 0$ yields the so-called linearized Euler–Lagrange equation (or Jacobi equation)

$$m\ddot{h}(t) + U'(q(t))h(t) = 0, \quad t \in [t_0, t_1], \quad h(t_0) = h(t_1) = 0. \tag{6.36}$$

The solutions h of this equation are called Jacobi fields. By definition, the point t_0 is conjugate to the point t_1 iff problem (6.36) has a nontrivial solution (i.e., there exists a nontrivial Jacobi field). Equivalently, the accessory eigenvalue problem (6.34) has the eigenvalue $\lambda = 0$.

If the interval $[t_0, t_1]$ contains an interior point which is conjugate to t_0, then each solution q of the Euler–Lagrange equation (6.35) is not a local minimum of the action functional S.

For example, in the case of the harmonic oscillator, the conjugate points to t_0 are $t_{n,\mathrm{conj}} = t_0 + n\pi/\omega$ with $n = 1, 2, \ldots$ If $t_1 < t_{1,\mathrm{conj}}$, then the Morse index i_S is equal to zero. If we steadily increase the interval $[t_0, t_1]$ for fixed t_0, then the Morse index jumps each point $t_{1,\mathrm{conj}}, t_{2,\mathrm{conj}}, \ldots$ The jump is equal to one.

We will show in Sect. 7.10 that the appearance of conjugate points complicates the computation of the Feynman path integral for the quantized harmonic oscillator.

In turn, this complicates the most important approximation method in quantum mechanics – the WKB method – which studies the singular limit $\hbar \to 0$ (i.e., the Planck action quantum goes to zero). In geometric optics, conjugate points correspond to focal points which represent singularities of the light ray configuration.

The classical results quoted above are special cases of a more general functional-analytic result whose proof can be found in Zeidler (1986), Vol. III, p. 201 (see the references on page 1049). Roughly speaking, the following hold:

The principle of least action is valid for sufficiently small time intervals.

6.5.6 The Ginzburg–Landau Potential and the Higgs Potential

Let us study the prototype of a phase transition by considering the Ginzburg–Landau potential

$$\boxed{U(q) := (q^2 - a^2)^2, \qquad q \in \mathbb{R}}$$

for fixed positive parameter a. The principle of critical action

$$\int_{t_0}^{t_1} \Big(\tfrac{1}{2} m \dot{q}(t)^2 - U\big(q(t)\big) \Big) dt = \text{critical!}$$

together with the boundary condition $q(t_0) = q_0$, $q(t_1) = q_1$ yields the equation of motion

$$\boxed{m\ddot{q}(t) = F(q(t)) \qquad t \in \mathbb{R}} \tag{6.37}$$

with the force $F(q) = -U'(q) = 4(a^2 - q^2)q$. The energy

$$E(t) = \tfrac{1}{2} m \dot{q}^2(t) + U(q(t))$$

is constant along the trajectories $q = q(t)$ (solutions of (6.37)). Let us discuss the qualitative behavior of the motion of a particle governed by (6.37). Since $U(q) \geq 0$ for all $q \in \mathbb{R}$, the energy E of the particle is always nonnegative.

- If $E = 0$, then $\dot{q}(t) = 0$ for all times $t \in \mathbb{R}$. Hence the particle rests either at the point $q = a$ or at the point $q = -a$.

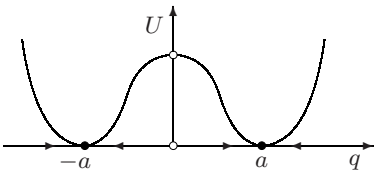


Fig. 6.8. The Ginzburg–Landau potential

- For fixed energy $E \geq 0$, the particle is only located at positions which belong to the set
$$\{q \in \mathbb{R} : E - U(q) \geq 0\}.$$
For example, if $0 \leq E < U(0)$, then the particle never passes the origin $q = 0$. In contrast to this, if $E > U(0)$, then the particle is able to pass the origin (Fig. 6.8).

Taylor expansion of the force near the point q_0 yields

$$F(q) = -U'(q_0) - U''(q_0)(q - q_0) + o(q - q_0), \quad q \to q_0,$$

where $U'(q) = 4(q^2 - a^2)q$ and $U''(q) = 12q^2 - 4a^2$. The critical equation

$$\boxed{U'(q_0) = 0}$$

has the solutions $q_0 = \pm a, 0$. The force vanishes precisely at these equilibrium points. Since $U''(\pm a) > 0$ and $U''(0) < 0$,

- the force $F(q) = -U''(a)(q - a) + \dots$ is attracting near the equilibrium point $q = a$,
- the force $F(q) = -U''(-a)(q + a) + \dots$ is attracting near the equilibrium point $q = -a$, and
- the force $F(q) = -U''(0)q + \dots$ is repelling near the equilibrium point $q = 0$.

Consequently, the Ginzburg–Landau potential describes the motion of a particle that has the two stable equilibrium points $q = \pm a$ and one unstable equilibrium point $q = 0$. A passage from $q = a$ to $q = -a$ models a phase transition.

Superconductivity. The Ginzburg–Landau potential was used by Ginzburg and Landau in 1950 in order to study phase transitions in superconductivity.[15]

The Higgs particle. In the Standard Model of elementary particle physics, Weinberg introduced the field of the so-called Higgs particle in order to generate the large masses of the vector bosons $W^{\pm}, Z^0$. The potential of the Higgs field is of the Landau–Ginzburg type. We refer to:

P. Higgs, Broken symmetry and the masses of gauge bosons, Phys. Rev. Lett. **13** (1964), 508–509.

S. Weinberg, A model of leptons, Phys. Rev. Lett. **19** (1967), 1264–1266.

This will be thoroughly studied in Vol. III on gauge field theory.

[15] B. Ginzburg and L. Landau, On the theory of superconductivity, J. Experimental and Theoretical Physics **20** (1950), 1064–1082 (in Russian).

6.5.7 Damped Oscillations, Stability, and Energy Dissipation

The damped oscillator equation

$$m\ddot{q}(t) = -\alpha q(t) - \beta\dot{q}(t), \qquad t \in \mathbb{R}, \qquad q(0) = q_0,\ \dot{q}(0) = v_0 \tag{6.38}$$

with the positive constants α, β describes the motion of a particle on the real line under the influence of the repulsive force $-\alpha q$ and the friction force $-\beta\dot{q}$. Typically, friction forces depend on the velocity of the particle. Letting $\omega := \sqrt{\alpha/m}$ and $\gamma := \beta/2m$ we get the equation

$$\ddot{q}(t) + 2\gamma\dot{q}(t) + \omega^2 q(t) = 0, \qquad t \in \mathbb{R}, \quad q(0) = q_0,\ \dot{q}(0) = v_0.$$

If the friction force is sufficiently small, $0 < \gamma < \omega$, then the unique motion is given by

$$q = q_0 \mathrm{e}^{-\gamma t} \cos\left(t\sqrt{\omega^2 - \gamma^2}\right) + \frac{v_0 + \gamma q_0}{\sqrt{\omega^2 - \gamma^2}}\, \mathrm{e}^{-\gamma t} \sin\left(t\sqrt{\omega^2 - \gamma^2}\right).$$

This corresponds to damped oscillations.

Asymptotic stability. As time goes to plus infinity, the motion passes to rest at the origin, independently of the initial position and the initial velocity:

$$\lim_{t \to +\infty} q(t) = 0.$$

We call this asymptotic stability of the motion.

Irreversibility. Obviously, the nontrivial process $q = q(t)$ is irreversible; that is, if $q = q(t)$ is a nontrivial solution of the equation of motion (6.38), then the time-reversed process $q = q(-t)$ is not a solution of (6.38). Such a process would correspond to oscillations with increasing amplitude as time goes to plus infinity. For the mechanical energy

$$E(t) = \tfrac{1}{2} m\dot{q}(t)^2 + \tfrac{1}{2} m\omega^2 q(t)^2$$

of the motion $q = q(t)$, we get $\dot{E}(t) = m\ddot{q}\dot{q} + m\omega^2 q\dot{q}$, and $m\ddot{q} = -m\omega^2 q - \beta\dot{q}$. Hence

$$\dot{E}(t) = -\beta\dot{q}(t)^2 \leq 0 \qquad \text{for all times} \quad t \geq 0.$$

If the initial velocity does not vanish, $v_0 \neq 0$, then energy conservation is violated because of $\dot{E}(0) < 0$. From the physical point of view, the mechanical energy is partly converted into heat. Such processes are called dissipative. By the first law of thermodynamics, the sum of mechanical energy and heat energy is conserved.

6.5.8 Resonance and Small Divisors

We consider an harmonic oscillator with angular eigenfrequency ω under the influence of an external periodic force $F(t) := \sin \alpha t$ with the angular frequency α. This corresponds to the following equation of motion:

$$m\ddot{q}(t) + m\omega^2 q(t) = \sin \alpha t, \qquad t \in \mathbb{R} \qquad q(0) = q_0,\ \dot{q}(0) = v_0. \tag{6.39}$$

To simplify notation, we normalize the mass by setting $m := 1$.

(N) Non-resonance case $\alpha \neq \omega$ (i.e., the angular frequency α of the external force is different from the angular eigenfrequency ω of the harmonic oscillator): The unique solution of (6.39) reads as

$$q(t) = q_0 \cos \omega t + \frac{v_0}{\omega} \sin \omega t + \frac{\sin \alpha t + \sin \omega t}{2(\alpha + \omega)\omega} - \frac{\sin \alpha t - \sin \omega t}{2(\alpha - \omega)\omega}, \qquad t \in \mathbb{R}.$$

If the difference $\alpha - \omega$ is small, then $\alpha - \omega$ is called a *small divisor*.

(R) Resonance case $\alpha = \omega$ (i.e., the angular frequency α of the external force is equal to the angular eigenfrequency ω of the harmonic oscillator): The unique solution of (6.39) reads as

$$q(t) = q_0 \cos \omega t + \frac{v_0}{\omega} \sin \omega t + \frac{\sin \omega t}{2\omega^2} - \frac{t}{2\omega} \cos \omega t, \qquad t \in \mathbb{R}.$$

The last term $t \cos \omega t$ corresponds to an oscillation of angular frequency ω. The point is that this term is dangerous as its amplitude grows without bound as time t goes to plus infinity. In real life, such resonance effects can lead to the destruction of structures (e.g., vibrations of a bridge induced by wind or traffic, or vibrations of a building caused by an earthquake). The occurrence of the dangerous resonance term $t \cos \omega t$ is understandable, when one realizes that the resonance solution (R) can be derived from the non-resonance solution (N) by passing to the limit $\alpha \to \omega$. In fact, the classical Bernoulli–de l'Hospital rule for limits tells us that

$$\lim_{\alpha \to \omega} \frac{\sin \alpha t - \sin \omega t}{\alpha - \omega} = \lim_{\alpha \to \omega} t \cos \alpha t = t \cos \omega t.$$

Observe the following:

Many complicated phenomena in nature are caused by resonance effects (e.g., the chaotic motion of some asteroids in celestial mechanics or internal resonances of quantum fields).

In terms of mathematics, resonances are related to singular situations which correspond to small (or zero) divisors.

6.6 Symmetry and Conservation Laws

Newton and his successors noticed that there exist conservation laws that simplify the integration of the equations of motion. For example, this concerns the conservation of the following quantities: energy, momentum, angular momentum, Runge–Lenz vector. In 1918, Emmy Noether (1882–1935) proved a general theorem which shows that

The symmetries of the Lagrangian are responsible for conservation laws.

For example, invariance of the Lagrangian under time translations leads to conservation of energy. More general, we will show in Sec. 6.6.2 that smooth continuous symmetries of the action integral imply conservation laws.

6.6.1 The Symmetries of the Harmonic Oscillator

Let $q = q(t)$ be a solution of the harmonic oscillator equation

$$\ddot{q}(t) + \omega^2 q(t) = 0, \qquad \text{for all} \quad t \in \mathbb{R}. \tag{6.40}$$

This equation has the following symmetries.

(i) Time translation: $q = q(t + t_0)$ is a solution of (6.40) for each fixed $t_0 \in \mathbb{R}$.
(ii) Time reflection: $q = q(-t)$ is a solution of (6.40).
(iii) Spatial reflection: $q = -q(t)$ is a solution of (6.40).
(iv) Rescaling of the position: $q = \alpha q(t)$ is a solution of (6.40) for each fixed real number $\alpha > 0$.

Summarizing, we say that the harmonic oscillator equation is invariant under time translations, time reflections, spatial reflections, and rescaling of position.

6.6.2 The Noether Theorem

The Noether theorem is one of the most important and most beautiful theorems in mathematical physics. To begin with, consider the Euler–Lagrange equation

$$\boxed{\frac{d}{dt} L_{\dot{q}}(q(t), \dot{q}(t), t) = L_q(q(t), \dot{q}(t), t) \qquad \text{for all} \quad t \in \mathbb{R}.} \tag{6.41}$$

Suppose that the Lagrangian $L : \mathbb{R}^3 \to \mathbb{R}$ is smooth. To simplify the argument, we assume that all the solutions $q = q(t)$ of the equation of motion (6.41) are smooth, and they exist for all times $t \in \mathbb{R}$. For such a solution, we define the momentum

$$p(t) := L_{\dot{q}}(q(t), \dot{q}(t), t),$$

and the energy

$$E(t) := p(t)\dot{q}(t) - L(q(t), \dot{q}(t), t).$$

The Euler–Lagrange equation (6.41) is then equivalent to the momentum equation:

$$\dot{p}(t) = L_q(q(t), \dot{q}(t), t) \qquad \text{for all} \quad t \in \mathbb{R}. \tag{6.42}$$

For the time derivative of the energy function, we get the energy equation:

$$\dot{E}(t) = -L_t(q(t), \dot{q}(t), t) \qquad \text{for all} \quad t \in \mathbb{R}. \tag{6.43}$$

This follows from $\dot{E} = \dot{p}\dot{q} + p\ddot{q} - L_q\dot{q} - L_{\dot{q}}\ddot{q} - L_t$, and $p = L_{\dot{q}}$, $\dot{p} = L_q$. From (6.42) and (6.43), we obtain immediately the following two statements:

(i) Conservation of momentum: If the Lagrangian $L = L(\dot{q}, t)$ does not depend on position q, then $p(t) = \text{const}$ for all times $t \in \mathbb{R}$.
(ii) Conservation of energy: If the Lagrangian $L = L(q, \dot{q})$ does not depend on time t, then $E(t) = \text{const}$ for all times $t \in \mathbb{R}$.

For example, choose the Lagrangian $L := \frac{1}{2}m\dot{q}^2 - U(q, t)$. The corresponding Euler–Lagrange equation

$$m\ddot{q}(t) = -U_q(q(t), t)$$

describes the motion of a particle on the real line under the influence of the (time-dependent) force $-U_q(q, t)$ with the potential U. Momentum and energy are given by $p(t) := m\dot{q}(t)$ and $E(t) := \frac{1}{2}m\dot{q}(t)^2 + U(q(t), t)$, respectively. If the potential $U = U(q)$ does not depend on time, then we have the energy conservation

$$\tfrac{1}{2}m\dot{q}(t)^2 + U(q(t)) = \text{const} = E_0 \qquad \text{for all} \quad t \in \mathbb{R}.$$

Moreover, if the potential U vanishes identically, $U = 0$, then we get the momentum conservation

$$m\dot{q}(t) = \text{const} = p_0 \qquad \text{for all} \quad t \in \mathbb{R}.$$

Setting $v_0 := p_0/m$, the general solution is then given by the free motion

$$q = q_0 + v_0 t \qquad \text{for all} \quad t \in \mathbb{R}$$

of constant velocity v_0.

Special case of the Noether theorem. We want to show that momentum conservation is a consequence of the invariance of the action integral

$$S_\Omega[q] := \int_\Omega L(q(t), \dot{q}(t), t)dt$$

under spatial translations. To explain this, let us introduce the transformation $T_\varepsilon : \mathbb{R}^2 \to \mathbb{R}^2$ by setting

$$T_\varepsilon(t, q) := (t, q + \varepsilon) \qquad \text{for all} \quad (t, q) \in \mathbb{R}^2. \tag{6.44}$$

In particular, this transformation sends each trajectory q to a new trajectory $T_\varepsilon q$. Explicitly, $(T_\varepsilon q)(t) := q(t) + \varepsilon$ for all times $t \in \mathbb{R}$. Now we postulate the invariance property

$$\boxed{S_{T_\varepsilon \Omega}[T_\varepsilon q] = S[q]} \tag{6.45}$$

of the action integral. We assume that this relation is valid for all smooth functions $q : \mathbb{R} \to \mathbb{R}$, all compact intervals $\Omega := [t_0, t_1]$, and all real parameters ε in some open neighborhood of $\varepsilon = 0$.

Proposition 6.6 (i) *The invariance of the action integral under spatial translations implies conservation of momentum for each solution of the Euler–Lagrange equation.*

(ii) *The invariance of the action integral under time translations implies conservation of energy for each solution of the Euler–Lagrange equation.*

Proof. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. If the interval Ω contracts to the point t_0, then

$$\lim_{\Omega \to t_0} \frac{1}{\text{meas}(\Omega)} \int_\Omega f(t)dt = f(t_0).$$

Consequently, if $\int_\Omega f(t)dt = 0$ for all compact intervals Ω, then $f(t) = 0$ for all $t \in \mathbb{R}$.

Ad (i). (I) General trajectory $q = q(t)$. The transformation T_ε from (6.44) sends the trajectory $t \mapsto q(t)$ to $t \mapsto q(t) + \varepsilon$ and leaves the time interval $[t_0, t_1]$ invariant. Thus, the invariance condition (6.45) tells us that

$$\int_{t_0}^{t_1} L(q(t) + \varepsilon, \tfrac{d}{dt}(q(t) + \varepsilon), t)dt = \int_{t_0}^{t_1} L(q(t), \dot{q}(t), t)dt$$

for all $\varepsilon \in \mathbb{R}$. Differentiation with respect to the parameter ε at $\varepsilon = 0$ yields

$$\int_{t_0}^{t_1} L_q(q(t), \dot{q}(t), t)dt = 0$$

for all intervals $[t_0, t_1]$. Hence $L_q(q(t), \dot{q}(t), t) = 0$ for all $t \in \mathbb{R}$.

(II) Motion $q = q(t)$. If we choose a solution $q = q(t)$ of the Euler–Lagrange equation (6.41), then we obtain the momentum equation,

$$\dot{p}(t) = L_q(q(t), \dot{q}(t), t),$$

by (6.42). Consequently, $\dot{p}(t) = 0$. This implies the desired momentum conservation, namely, $p(t) = \text{const}$ for all times $t \in \mathbb{R}$.

Ad (ii). (I) General trajectory $q = q(t)$. Now we use the time translation

$$\tau := t + \varepsilon.$$

Set $T_\varepsilon(t, q) := (t + \varepsilon, q)$ for all $(t, q) \in \mathbb{R}^2$. For fixed parameter ε, the trajectory

$$t \mapsto q(t)$$

is transformed into the trajectory $\tau \mapsto q(\tau, \varepsilon)$ with $q(\tau, \varepsilon) := q(t)$. Moreover, the time interval $[t_0, t_1]$ is transformed into $[t_0 + \varepsilon, t_1 + \varepsilon]$. The invariance condition (6.45) reads as

$$\int_{t_0+\varepsilon}^{t_1+\varepsilon} L(q(\tau, \varepsilon), q_\tau(\tau, \varepsilon), \tau) d\tau = \int_{t_0}^{t_1} L(q(t), \dot{q}(t), t) dt$$

for all $\varepsilon \in \mathbb{R}$. Using the transformation $\tau = t + \varepsilon$, we get

$$\int_{t_0}^{t_1} L(q(t), \dot{q}(t), t + \varepsilon) dt = \int_{t_0}^{t_1} L(q(t), \dot{q}(t), t) dt.$$

Differentiation with respect to the real parameter ε at $\varepsilon = 0$ yields

$$\int_{t_0}^{t_1} L_t(q(t), \dot{q}(t), t) dt = 0,$$

and hence $L_t(q(t), \dot{q}(t), t) = 0$.

(II) Motion $q = q(t)$. For a solution $q = q(t)$ of the Euler-Lagrange equation (6.41), we have the energy equation,

$$\dot{E}(t) = -L_t(q(t), \dot{q}(t), t),$$

by (6.43). Hence $\dot{E}(t) = 0$. This implies energy conservation, $E(t) = \text{const}$ for all $t \in \mathbb{R}$. □

General case of the Noether theorem. Let us now study a smooth change

$$\boxed{\tau = \tau(t, \varepsilon), \quad Q = Q(t, q, \varepsilon), \qquad t, q \in \mathbb{R}} \tag{6.46}$$

from time t and position q to time τ and position Q, respectively. We assume that this transformation depends on the small real parameter ε in such a way that the value $\varepsilon = 0$ corresponds to the identical transformation, that is,

$$t = \tau(t, 0), \quad q = Q(t, q, 0), \qquad t, q \in \mathbb{R}.$$

We also assume that, for each value of ε in some open neighborhood of $\varepsilon = 0$, the time transformation $t \mapsto \tau(t, \varepsilon)$ is a diffeomorphism from $\mathbb{R}$ onto itself. Naturally enough, for fixed parameter ε, the given trajectory $q = q(t)$ is transformed into the new trajectory $q = q(\tau, \varepsilon)$ given by

$$\tau = \tau(t,\varepsilon), \quad q(\tau(t,\varepsilon),\varepsilon) = Q(t,q(t),\varepsilon), \qquad t \in \mathbb{R}.$$

The linearization of the transformation of trajectories with respect to the small parameter ε reads as

$$\boxed{\tau = t + \delta t + o(\varepsilon), \qquad q(\tau,\varepsilon) = q(t) + \delta q(t) + o(\varepsilon), \qquad \varepsilon \to 0.} \tag{6.47}$$

Here, we set $\delta t := \varepsilon\tau_\varepsilon(t,0)$ and

$$\delta q(t) := \varepsilon \frac{\partial}{\partial \varepsilon} q(\tau(t,\varepsilon),\varepsilon)_{|\varepsilon=0}.$$

In fact, Taylor expansion with respect to the small parameter ε yields

$$\tau(t,\varepsilon) = \tau(t,0) + \tau_\varepsilon(t,0)\varepsilon + o(\varepsilon), \qquad \varepsilon \to 0.$$

Hence $\tau = t + \delta t + o(\varepsilon)$. Similarly, we get $q(\tau(t,\varepsilon),\varepsilon) = q(t) + \delta q(t) + o(\varepsilon)$. To streamline the notation, let us denote the smooth transformation (6.46) by

$$T_\varepsilon : \mathbb{R}^2 \to \mathbb{R}^2,$$

that is, $T_\varepsilon(t,q) = (\tau, Q)$. In 1918, the following famous theorem was proven by Emmy Noether.[16]

Theorem 6.7 *Suppose that the action integral* $S_\Omega[q] := \int_\Omega L(q(t),\dot q(t),t)dt$ *has the invariance property*

$$\boxed{S_{T_\varepsilon\Omega}[T_\varepsilon q] = S_\Omega[q]} \tag{6.48}$$

for all smooth trajectories $q = q(t)$ *on* $\mathbb{R}$*, all compact time intervals* Ω*, and all real parameters* ε *in some open neighborhood of* $\varepsilon = 0$*. Then each smooth solution* $q = q(t)$ *of the Euler-Lagrange equation*

$$\frac{d}{dt} L_{\dot q}(q(t),\dot q(t),t) = L_q(q(t),\dot q(t),t), \qquad \textit{for all} \quad t \in \mathbb{R}$$

satisfies the conservation law

$$\frac{d}{dt}\left(L\delta t + (\delta q - \dot q \delta t) L_{\dot q}\right) = 0 \qquad \textit{for all} \quad t \in \mathbb{R}. \tag{6.49}$$

Before proving this, let us make some comments.

Local symmetry condition. The conservation law (6.49) reads explicitly as

$$\frac{d}{dt}\big(L(P)\delta t + (\delta q(t) - \dot q(t)\delta t) L_{\dot q}(P)\big) = 0 \qquad \text{for all} \quad t \in \mathbb{R},$$

where $P := (q(t),\dot q(t),t)$. This holds for all $\varepsilon \in \mathbb{R}$. Note that δt and δq linearly depend on ε. The invariance condition (6.48) for the action integral can be replaced by the following local symmetry condition

$$\boxed{L\big(q(\tau,\varepsilon), q_\tau(\tau,\varepsilon),\tau)\big)_{|\tau=\tau(t,\varepsilon)} \cdot \tau_t(t,\varepsilon) = L\big(q(t),\dot q(t),t\big).} \tag{6.50}$$

We assume that this is true for all times $t \in \mathbb{R}$, all smooth trajectories $q = q(t)$ on $\mathbb{R}$, and all real parameters ε in some open neighborhood of $\varepsilon = 0$. The proof to be given below shows that

[16] E. Noether, Invariant variational problems, Göttinger Nachrichten, Math.-phys. Klasse 1918, 235–257 (in German).

The Noether theorem remains valid if we use trajectories $q = q(t)$ with n degrees of freedom, that is, we have $q(t) \in \mathbb{R}^n$ for all $t \in \mathbb{R}$, where[17] $n = 1, 2, \ldots$

The point is that we will carry out the proof below in such a way that it can be generalized immediately to variational problems for multi-dimensional integrals. This will be considered in Vol. III in connection with the investigation of physical field theories (e.g., the theory of general relativity, the Standard Model in elementary particle physics, and the string theory). In particular, we will show that the energy-momentum tensor for physical fields follows from the relativistic invariance of the action integral together with the Noether theorem.

Example. For the motion $q = q(t)$ on the real line, the conservation law (6.49) reads as

$$\frac{d}{dt}\big(p(t)\delta q(t) - E(t)\delta t\big) = 0 \qquad \text{for all} \quad t \in \mathbb{R}. \tag{6.51}$$

(i) Conservation of momentum: For the translation of position,

$$\tau = t, \quad Q = q + \varepsilon,$$

we obtain $\delta t = 0$ and $\delta q = \varepsilon$, by (6.47). Then $q(\tau, \varepsilon) = q(t) + \varepsilon$. If the Lagrangian $L = L(q, \dot{q}, t)$ does not depend on position q, then

$$L(q(t) + \varepsilon, \dot{q}(t), t) = L(q(t), \dot{q}(t), t) \qquad \text{for all} \quad t, \varepsilon \in \mathbb{R}.$$

This is the local symmetry condition (6.50). By (6.51), $\varepsilon\dot{p}(t) = 0$ for all t, ε. Hence $p(t) = \text{const}$ for all $t \in \mathbb{R}$.

(ii) Conservation of energy: For the time translation

$$\tau = t + \varepsilon, \quad Q = q,$$

we get $\delta t = \varepsilon$ and $\delta q = 0$, by (6.47). Then $q(\tau, \varepsilon) = q(t)$. If the given Lagrangian $L = L(q, \dot{q}, t)$ does not depend on time t, then

$$L(q(t), \dot{q}(t), t + \varepsilon) = L(q(t), \dot{q}(t), t) \qquad \text{for all} \quad t, \varepsilon \in \mathbb{R}.$$

This is the local symmetry condition (6.50). By (6.51), $\varepsilon\dot{E}(t) = 0$ for all t, ε. Hence $E(t) = \text{const}$ for all $t \in \mathbb{R}$.

Proof of Theorem 6.7. In what follows, the primed quantity q' denotes the derivative of the function $q = q(t)$ with respect to time t.

(I) General trajectories. The *global* symmetry condition (6.48) reads explicitly as

$$\boxed{\int_{T_\varepsilon \Omega} L(q(\tau, \varepsilon), q_\tau(\tau, \varepsilon), \tau) d\tau = \int_{\Omega} L(q(t), q'(t), t) dt.}$$

Using the transformation $\tau = \tau(t, \varepsilon)$, we get

$$\int_{\Omega} L(q(\tau, \varepsilon), q_\tau(\tau, \varepsilon), \tau)_{|_{\tau = \tau(t,\varepsilon)}} \cdot \tau_t(t, \varepsilon) dt = \int_{\Omega} L(q(t), q'(t), t) dt,$$

by the substitution rule for integrals. Contracting the interval Ω to the point t, we obtain the *local* symmetry condition

[17] Here, $\dot{q}(t) L_{\dot{q}}(P) = \sum_{k=1}^{n} \dot{q}^k(t) \frac{\partial L}{\partial \dot{q}^k}(P)$.

$$L(q(\tau,\varepsilon),q_\tau(\tau,\varepsilon),\tau)_{|_{\tau=\tau(t,\varepsilon)}}\cdot\tau_t(t,\varepsilon)=L(q(t),q'(t),t), \tag{6.52}$$

which coincides with (6.50). Since $\tau(t,0)=t$ and $q(t,0)=q(t)$, we get

$$\tau_t(t,0)=1,\quad q_\tau(t,0)=q'(t),\quad q_{\tau\tau}(t,0)=q''(t).$$

Differentiating the local symmetry condition (6.52) with respect to the real parameter ε at $\varepsilon=0$, the chain rule yields

$$\begin{aligned}&L_q(P)\big(q_\tau(t,0)\tau_\varepsilon(t,0)+q_\varepsilon(t,0)\big)\tau_t(t,0)\\&\quad+L_{q'}(P)\big(q_{\tau\tau}(t,0)\tau_\varepsilon(t,0)+q_{\varepsilon\tau}(t,0)+L_t(P)\tau_\varepsilon(t,0)\big)\tau_t(t,0)\\&\quad+L(P)\tau_{\varepsilon t}(t,0)=0.\end{aligned}$$

Here, we set $P:=\big(q(t),q'(t),t\big)$. Introducing $\alpha(t):=\varepsilon q_\varepsilon(t,0)$, we obtain

$$L_q(P)(q'\delta t+\alpha)+L_{q'}(P)(q''\delta t+\alpha')+L_t(P)\delta t+L(P)(\delta t)'=0.$$

Setting $L'(t):=\frac{d}{dt}L(P)$, the chain rule tells us that

$$L'=L_qq'+L_{q'}q''+L_t.$$

This implies the following equation

$$L'\delta t+L_q\alpha+L_{q'}\alpha'+L(\delta t)'=0. \tag{6.53}$$

Explicitly, $L'(P)\delta t+L_q(P)\alpha(t)+L_{q'}(P)\alpha'(t)+L(P)(\delta t)'(t)=0$. To compute the function $\alpha=\alpha(t)$, differentiate $q(\tau(t,\varepsilon),\varepsilon)$ with respect to the parameter ε at $\varepsilon=0$. Then

$$\delta q(t)=\varepsilon q_\tau(\tau(t,0),0)\tau_\varepsilon(t,0)+\varepsilon q_\varepsilon(\tau(t,0),0).$$

Noting that $\tau(t,0)=t$, we get

$$\delta q(t)=q'(t)\delta t+\alpha(t).$$

Now the key relation (6.53) reads as

$$\boxed{L'\delta t+L(\delta t)'+L_q(\delta q-q'\delta t)+L_{q'}(\delta q-q'\delta t)'=0.} \tag{6.54}$$

This is valid for each trajectory $q=q(t)$.

(II) Motion. Now consider the special case where the trajectory $q=q(t)$ satisfies the Euler-Lagrange equation. Then

$$(L_{q'}(P))'=L_q(P).$$

Therefore, using the product rule, the key relation (6.54) passes over to the equation

$$(L\delta t)'+\big(L_{q'}\left(\delta q-q'\delta t\right)\big)'=0.$$

This is the desired conservation law (6.49). □

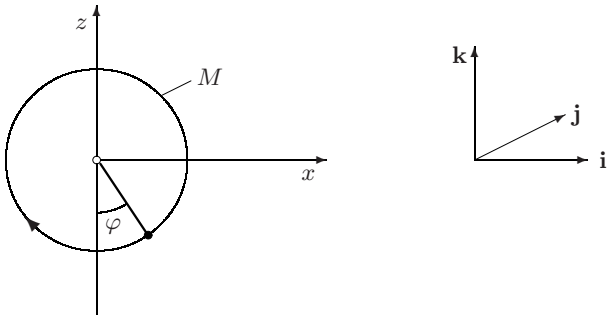


Fig. 6.9. Motion of a pendulum

6.7 The Pendulum and Dynamical Systems

> Henri Poincaré (1854–1912) originated not only new theories, but completely new branches of mathematics like the theory of dynamical systems, differential topology, and algebraic topology. His ideas are so great, his way of thinking of and looking at mathematical reality has been so widely accepted that to us, his descendants, it seems strange that people have thought differently, for example, that dynamical systems should be considered on manifolds, and not only on $\mathbb{R}^n$ – because indeed a plane pendulum is a motion on the circle, and a spherical pendulum is a motion on a sphere.[18]
>
> Krysztof Maurin, 1999

Motions of mass point systems are frequently governed by constraints (e.g., the oscillations of molecules). Whereas the free motion of N mass points in Euclidean space has $3N$ degrees of freedom, constrained motions possess less than $3N$ degrees of freedom. The Lagrangian approach to mechanics allows an elegant reduction to the true number of degrees of freedom by considering the motion with respect to appropriately chosen local coordinates. This corresponds to the theory of dynamical systems on manifolds. In order to illustrate this, let us consider the motion $\mathbf{x} = \mathbf{x}(t)$ of a pendulum of mass m and length l.

In a right-handed Cartesian coordinate system, we set $\mathbf{x} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ (Fig. 6.9). The gravitational force

$$\mathbf{F} = -mg\mathbf{k}$$

acts on a particle of mass m. Here, $g = 9.81\text{m/s}^2$ is the acceleration of gravity. Letting $U(\mathbf{x}) := mgz$, we get $F = -\mathbf{grad}\ U$. Therefore, the force $\mathbf{F}$ has the potential U.

6.7.1 The Equation of Motion

For the Lagrangian, we get

$$L := \text{kinetic energy minus potential energy} = \tfrac{1}{2}m\dot{\mathbf{x}}^2 - U(\mathbf{x}).$$

[18] K. Maurin, The Riemann Legacy, Kluwer, Dordrecht, 1999 (reprinted with permission).

The principle of critical action reads as

$$\int_{t_0}^{t_1} \Big(\tfrac{1}{2} m\dot{\mathbf{x}}(t)^2 - mgz(t)\Big)dt = \text{critical !}, \quad \mathbf{x}(t_0) = \mathbf{x}_0,\ \mathbf{x}(t_1) = \mathbf{x}_1,$$

together with the side condition $\mathbf{x}(t)^2 = l^2$ for all times $t \in [t_0, t_1]$. The side condition tells us that the motion proceeds on a circle of radius l. In order to *eliminate* the side condition, we make the ansatz

$$\mathbf{x}(t) = l(\sin\varphi(t)\ \mathbf{i} - \cos\varphi(t)\ \mathbf{k}),$$

where φ denotes the elongation angle (Fig. 6.9). For the velocity vector, we then get $\dot{\mathbf{x}}(t) = l(\cos\varphi(t)\ \mathbf{i} + \sin\varphi(t)\ \mathbf{k})\dot{\varphi}(t)$. Therefore, the principle of critical action passes over to

$$\int_{t_0}^{t_1} \Big(\tfrac{1}{2} ml^2\dot{\varphi}^2(t) + mgl\cos\varphi(t)\Big)dt = \text{critical!}, \quad \varphi(t_0) = \varphi_0,\ \varphi(t_1) = \varphi_1.$$

Each smooth solution $\varphi = \varphi(t)$ of the action principle satisfies the Euler–Lagrange equation $\frac{d}{dt}L_{\dot{\varphi}} = L_{\varphi}$ with the Lagrangian $L := \frac{1}{2}ml^2\dot{\varphi}^2 + mgl\cos\varphi$. This yields the nonlinear equation of motion for the pendulum,

$$\boxed{\ddot{\varphi} + \omega^2 \sin\varphi = 0,} \tag{6.55}$$

together with the angular frequency parameter $\omega := \sqrt{g/l}$. Equation (6.55) is also called the one-dimensional sin-Gordon equation. Let us study the situation where the pendulum rests at the initial time $t_0 := 0$. Then we have to add the initial condition

$$\varphi(0) = \varphi_0, \quad \dot{\varphi}(0) = 0.$$

We assume that $-\pi < \varphi_0 < \pi$.

6.7.2 Elliptic Integrals and Elliptic Functions

The linearized pendulum and the classical sinus function. Let us first consider the special case where the elongation angle φ is small. Using the approximation $\sin\varphi = \varphi + ...$, we obtain the harmonic oscillator equation, $\ddot{\varphi} + \omega^2\varphi = 0$, which has the solution

$$\boxed{\varphi = \varphi_0 \sin\left(\sqrt{\frac{g}{l}}\ t + \frac{\pi}{2}\right), \qquad t \in \mathbb{R}, \qquad T = 2\pi\sqrt{\frac{l}{g}}\ ,} \tag{6.56}$$

where T is the time period of the pendulum.

The nonlinear pendulum and the Jacobi sinus function. Now let us pass over to the general case.

> *From the practical point of view, the main task consists in computing the time period T of the pendulum.*

It turns out that this problem leads to the theory of elliptic integrals and elliptic functions created by Euler, Gauss, Legendre, Jacobi, Riemann, and Weierstrass in the 18th and 19th century. We will show that

$$\boxed{T = 4\sqrt{\frac{l}{g}}\, K(k), \qquad k := \sin\frac{\varphi_0}{2}.}$$

To this end, let us start with Legendre's normal form of the elliptic integral of the first kind,

$$F(\psi;k) := \int_0^{\psi} \frac{d\chi}{\sqrt{1-k^2\sin^2\chi}}\,.$$

For $\psi = \frac{\pi}{2}$, Legendre introduced the symbol $K(k) := F(\frac{\pi}{2}, k)$. The inverse function to $x = F(\psi;k)$ is called the amplitudinis function, $\psi = \mathrm{am}\,(x;k)$. Moreover, the function

$$y = \sin \mathrm{am}(x;k)$$

is called the *sinus amplitudinis function*. This Jacobian sinus function is also denoted by $y = \mathrm{sn}(x;k)$. Analytic continuation of this function yields a meromorphic function on the complex plane, which is double-periodic (i.e., elliptic) with the real period $4K(k)$ and the imaginary period $2K(\sqrt{1-k^2})\mathrm{i}$. In the limit case $k = 0$, we get the classical sinus function,

$$\sin \mathrm{am}(x;0) = \sin x \qquad \text{for all} \quad x \in \mathbb{C}.$$

Using Newton's binomial formula $(1+z)^\alpha = 1 + \alpha z + ...$ for $\alpha = -\frac{1}{2}$ and Walli's integral formula[19]

$$\int_0^{\pi/2} \sin^{2n}\psi\, d\psi = \frac{1\cdot 3\cdot 5\cdots(2n-1)}{2\cdot 4\cdot 6\cdots 2n}\cdot\frac{\pi}{2},$$

we obtain the convergent power series expansion

$$K(k) = \frac{\pi}{2}\Big(1 + 4k^2 + \Big(\frac{1\cdot 3}{2\cdot 4}\Big)^2 k^4 + ...\Big), \qquad 0 \le k < 1.$$

Proposition 6.8 *The unique solution $\varphi = \varphi(t)$ of the nonlinear pendulum equation (6.55) together with the initial condition $\varphi(0) = \varphi_0$, $\dot\varphi(0) = 0$ reads as*

$$\sin\frac{\varphi(t)}{2} = k \sin\ \mathrm{am}\left(\sqrt{\frac{g}{l}}\, t + K(k); k\right), \qquad t \in \mathbb{R}, \tag{6.57}$$

where $k = \sin\frac{\varphi_0}{2}$. This is a periodic motion of time period $T = 4\sqrt{l/g}\, K(k)$ between the elongation angles φ_0 and $-\varphi_0$ at time $t = 0$ and $t = T/2$, respectively.

For small angles φ_0, we obtain the classical first-order approximation formula (6.56). The second-order approximation of the time period reads as

$$T = 2\pi\sqrt{\frac{l}{g}}\left(1 + \frac{\varphi_0^2}{16}\right) + O(\varphi_0^4), \qquad \text{as} \quad \varphi_0 \to 0.$$

[19] Wallis (1616–1703).

Proof. (I) Energy conservation. Recall that $\omega := \sqrt{g/l}$. If $\varphi = \varphi(t)$ is a solution of (6.55), then

$$\tfrac{1}{2}ml^2\dot{\varphi}(t)^2 - mgl\cos\varphi(t) = \text{const} = E \qquad \text{for all} \quad t \in \mathbb{R}. \tag{6.58}$$

In fact, differentiation with respect to time t yields $ml^2\ddot{\varphi} + mgl\sin\varphi = 0$. Here, $E = \frac{1}{2}m\dot{\mathbf{x}}^2 + U(\mathbf{x})$ equals the total energy. Equation (6.58) describes conservation of energy. The initial condition $\varphi(0) = \varphi_0$, $\dot{\varphi}(0) = 0$ tells us that $E = -mgl\cos\varphi_0$. Using the identity $\cos\varphi = 1 - 2\sin^2\frac{\varphi}{2}$, from (6.58) we obtain

$$\left(\frac{d\varphi}{dt}\right)^2 = 4\omega^2\left(\sin^2\frac{\varphi_0}{2} - \sin^2\frac{\varphi}{2}\right).$$

Since $\dot{\varphi}(t)^2 \geq 0$, the motion of the pendulum satisfies $-\varphi_0 \leq \varphi(t) \leq \varphi_0$ for all times t. Integrating $dt = ...d\varphi$, we get

$$t = \frac{1}{2\omega}\int_0^{\varphi} \frac{d\chi}{\sqrt{\sin^2\frac{\varphi_0}{2} - \sin^2\frac{\chi}{2}}} - \text{const.} \tag{6.59}$$

The constant follows from the initial condition $t(\varphi_0) = 0$.

(II) Legendre's normal form. We want to transform the elliptic integral (6.59) into $F(\psi; k)$. To this end, we introduce the new parameter ψ by letting

$$\sin\frac{\varphi}{2} = k\sin\psi, \qquad -\frac{\pi}{2} \leq \psi \leq \frac{\pi}{2}.$$

In particular, we assume that $\psi = \pi/2$ corresponds to $\varphi = \varphi_0$. Note that

$$\sqrt{\sin^2\frac{\varphi_0}{2} - \sin^2\frac{\varphi}{2}} = k\sqrt{1 - \sin^2\psi} = k\cos\psi,$$

and $\cos\frac{\varphi}{2} = \sqrt{1 - \sin^2\frac{\varphi}{2}} = \sqrt{1 - k^2\sin^2\psi}$. It follows from

$$\frac{1}{2}\cos\frac{\varphi}{2}\,d\varphi = k\cos\psi\,d\psi$$

that

$$\omega t = \int_0^{\psi} \frac{d\psi}{\sqrt{1 - k^2\sin^2\psi}} - \text{const.}$$

By the initial condition $t(\frac{\pi}{2}) = 0$, the constant is equal to $F(\frac{\pi}{2}; k) = K(k)$. Therefore, $\psi = \operatorname{am}(\omega t + K(k); k)$, and hence we obtain the desired motion (6.57) of the pendulum.

(III) Approximation. If the maximal angle φ_0 is small, then the parameter k is also small. In first order, as $k \to 0$,

$$\sin\frac{\varphi}{2} = k\sin\operatorname{am}(\omega t + K(0); 0) + o(k) = k\sin\left(\omega t + \frac{\pi}{2}\right) + o(k).$$

Hence $\sin\frac{\varphi}{2} = k\cos\omega t$. If $t = 0$, then $\sin\frac{\varphi_0}{2} = k$. Finally, from $\sin\varphi = \varphi + ...$ we obtain the approximative solution $\varphi = \varphi_0\cos\omega t$, which coincides with (6.56). □

The Jacobi theta functions. Fix a complex number τ with positive imaginary part, that is, $\Im(\tau) > 0$. For each complex number z and each parameter $q := \mathrm{e}^{\mathrm{i}\pi\tau}$, Jacobi introduced the following functions:

$$\vartheta_0(z;q) := \sum_{n=-\infty}^{\infty} (-1)^n q^{n^2} \mathrm{e}^{2\pi \mathrm{i} n z}$$

and $\vartheta_1(z;q) := 2\sum_{n=0}^{\infty}(-1)^n q^{(n+\frac{1}{2})^2} \sin(2n+1)\pi z$, as well as

$$\vartheta_2(z;q) := 2\sum_{n=0}^{\infty} q^{(n+\frac{1}{2})^2} \cos(2n+1)\pi z, \quad \vartheta_3(z;q) := 1 + 2\sum_{n=1}^{\infty} q^{n^2} \cos 2n\pi z.$$

These functions are entire functions of period one with respect to the variable z. Now choose

$$\tau := \frac{K(\sqrt{1-k^2})}{K(k)}\,\mathrm{i}, \qquad q := \mathrm{e}^{\mathrm{i}\pi\tau}.$$

For the Jacobi sinus function, we get

$$\mathrm{sn}(z;q) = \sin \mathrm{am}(z;k) = 2K(k)\,\frac{\vartheta_0(0;q)\;\vartheta_1\left(\frac{z}{2K(k)}\,;q\right)}{\vartheta_1'(0,;q)\;\vartheta_0\left(\frac{z}{2K(k)}\,;q\right)}\,, \qquad z \in \mathbb{C}.$$

The function $z \mapsto \mathrm{sn}(z;q)$ has poles precisely at the points

$$2mK(k) + (2n+1)K\left(\sqrt{1-k^2}\right)\mathrm{i},$$

where m and n are arbitrary integers. Fix k with $0 \le k < 1$. Observe that

$$x = \int_0^{\mathrm{sn}\,(x;k)} \frac{dy}{\sqrt{(1-y^2)(1-k^2y^2)}} \qquad \text{for all} \quad x \in]-1,1[.$$

In the special case where $k = 0$, we have $\mathrm{sn}(x;0) = \sin x$. The map $x \mapsto \mathrm{sn}(x;k)$ is one-to-one on the interval $]-1,1[$. The values of $\mathrm{sn}(z;k)$ for arbitrary complex numbers are then obtained by analytic continuation.

Gauss' lemniscatic sinus function. Jakob Bernoulli (1654–1705) introduced the curve

$$(x^2+y^2)^2 - (x^2-y^2)^2 = 0, \qquad (x,y) \in \mathbb{R}^2 \tag{6.60}$$

in the Leipzig *Acta eruditorum* (journal of scientists) in 1694. This is an algebraic curve of fourth order called the lemniscate. This curve is obtained in the following way. Start with the hyperbola

$$x^2 - y^2 = 1.$$

Using polar coordinates $x = r\cos\varphi, y = r\sin\varphi$, $x^2 - y^2 = r^2\cos 2\varphi = 1$. Hence

$$r = \frac{1}{\sqrt{\cos 2\varphi}}, \qquad -\tfrac{\pi}{4} < \varphi < \tfrac{\pi}{4}, \quad \tfrac{3\pi}{4} < \varphi < \tfrac{5\pi}{4}.$$

By inversion with respect to unit circle, $r \mapsto \frac{1}{r}$, we obtain

$$r = \sqrt{\cos 2\varphi}, \qquad -\tfrac{\pi}{4} \le \varphi \le \tfrac{\pi}{4}, \quad \tfrac{3\pi}{4} \le \varphi \le \tfrac{5\pi}{4}.$$

This is the lemniscate (6.60) (Fig. 6.10). In about 1800, the young Gauss studied the arc length

$$s(r) = \int_0^r \frac{d\varrho}{\sqrt{1-\varrho^4}} \tag{6.61}$$

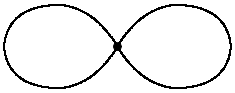


Fig. 6.10. Jakob Bernoulli's lemniscate

of the lemniscate. The inverse function, $r = \mathrm{sl}(s)$, was called the lemniscatic sinus function by Gauss. Using analytic continuation, Gauss observed that the function $r = \mathrm{sl}(s)$ has both a real and a purely imaginary period, namely, l, and $l\mathrm{i}$. Explicitly,

$$\mathrm{sl}(z + l) = \mathrm{sl}\, z \qquad \mathrm{sl}(z + l\mathrm{i}) = \mathrm{sl}\, z \qquad \text{for all} \quad z \in \mathbb{C}.$$

This way, for the first time, Gauss discovered the existence of double-periodic functions, which are also called elliptic functions. Explicitly, the period is the total length

$$l = 4 \int_0^1 \frac{d\varrho}{\sqrt{1 - \varrho^4}}$$

of the lemniscate. Moreover, by numerical experiments, Gauss discovered that

$$l = \frac{2\pi}{M(1, \sqrt{2})}.$$

Furthermore, for the Legendre elliptic integral, we have

$$K(k) = \int_0^{\pi/2} \frac{d\chi}{\sqrt{1 - k^2 \sin^2 \chi}} = \frac{1}{2M(1, \sqrt{1 - k^2})}, \quad 0 < k < 1.$$

Here, $M(a_0, b_0)$ denotes the Gaussian arithmetic-geometric mean of the positive real numbers a_0 and b_0. By definition, $M(a_0, b_0)$ is the limit of the rapidly convergent iterative method

$$a_{n+1} = \tfrac{1}{2}(a_n + b_n), \qquad b_{n+1} = \sqrt{a_n b_n}, \qquad n = 0, 1, 2, ...$$

That is, $M(a_0, b_0) = \lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n$.

Bernoulli's lemniscate was one of the starting points of algebraic geometry.

The essential step was made by Riemann (1826–1866) who noticed that algebraic curves in the plane $\mathbb{R}^2$ can be understood best by extending them to complex values. This way we obtain a curve in the complex space $\mathbb{C}^2$ which can be described by a compact Riemann surface of real dimension two (and complex dimension one). For the theory of elliptic functions, we refer to:

A. Hurwitz and R. Courant, Lectures on Complex Function Theory and Elliptic Integrals, Springer, Berlin, 1964 (in German).

V. Armitage, Elliptic Functions, Cambridge University Press, 2006.

Hints for further reading on plane algebraic curves can be found on page 203.

6.7.3 The Phase Space of the Pendulum and Bundles

> It was a great achievement of Poincaré (1854–1912) to show that the proper domain of analytic dynamics is the cotangent bundle TM^d of the position space M – this discovery was so fundamental that nowadays it seems to be natural and obvious.[20]
>
> Krysztof Maurin, 1996

> The categories of differentiable manifolds and vector bundles provide a useful context for the mathematics needed in mechanics, especially the new topological and qualitative results.[21]
>
> Ralph Abraham and Jerrold Marsden, 1978

> Too often in the physical sciences, the space of states is postulated to be a linear space when the basic problem is essentially nonlinear; this confuses the mathematical development.[22]
>
> Stephen Smale, 1980

This section should help the reader to understand the intuitive roots of the language of bundle theory in modern mathematics and physics. In mechanics, one has to distinguish the following crucial notions:

- position space M;
- state space TM (tangent bundle of the manifold M);
- phase space TM^d (cotangent bundle of M).

Let us discuss this for the prototype of a mechanical system with nontrivial topology – the circular pendulum.

> *Our strategy is to introduce only such quantities which have a geometrical meaning, that is, they are independent of the choice of local coordinates.*

The position space. The position of the circular pendulum is described by a point P of the circle $\mathbb{S}^1_R$ of radius R. We set

$$\boxed{M := \mathbb{S}^1_R.}$$

This is the set of all positions. Naturally enough, the circle M is called the position space of the pendulum.

The tangent space. By definition, the symbol

$$T_P M$$

denotes the tangent space of the circle M at the point P. In terms of physics, the vector $\mathbf{v}$ is contained in the tangent space $T_P M$ iff there exists a motion $\mathbf{x} = \mathbf{x}(t)$ on the circle M such that, at time $t = 0$, the particle is at the point P and has the velocity vector

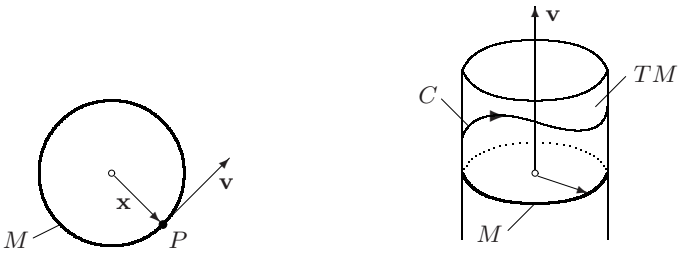
$$\dot{\mathbf{x}}(0) = \mathbf{v}.$$

The motion of the particle on the circle M is uniquely determined by knowing the tuple $(P, \mathbf{v})$ at the initial time $t = 0$. Here, P is the initial position, and $\mathbf{v}$ is the initial velocity vector (Fig. 6.11(a)). Let $\mathbf{x} = \mathbf{x}(t)$ be the counter-clockwise motion

[20] K. Maurin, The Riemann Legacy, Kluwer, Dordrecht, 1997.

[21] R. Abraham and J. Marsden, Foundations of Mechanics, Addison-Wesley, Reading, Massachusetts, 1978.

[22] S. Smale, The Mathematics of Time: Essays on Dynamical Systems, Economic Processes, and Related Topics, Springer, New York, 1980.



(a) position space M (b) tangent bundle TM

Fig. 6.11. The tangent bundle of the pendulum

of a particle on the circle M with angular velocity one. Set $\mathbf{e} := \dot{\mathbf{x}}(t)$. This way, we obtain a unit tangent vector $\mathbf{e}_P$ at each point P of the circle. Moreover, the velocity vectors $\mathbf{v} \in T_P M$ can be uniquely represented as

$$\mathbf{v} = v \cdot \mathbf{e}_P, \qquad v \in \mathbb{R}.$$

The real number v is called the coordinate of the velocity vector $\mathbf{v}$ at the point P.

The state space (tangent bundle). All the tuples $(P, \mathbf{v})$ are called states of the particle. The set of all possible states

$$\boxed{TM := \{(P, \mathbf{v}) : \quad P \in M, \ \ \mathbf{v} \in T_P M\}}$$

is called the state space of the pendulum. In mathematics, TM is called the tangent bundle of the circle M. For fixed P, the set

$$\mathcal{F}_P := \{(P, \mathbf{v}) : \mathbf{v} \in T_P M\}$$

is called a fiber of TM with respect to the base point P. This fiber can be identified with the tangent space $T_P M$ at the point P. The Lagrangian of the pendulum can be written as

$$L(P, \mathbf{v}) = \tfrac{1}{2} m \mathbf{v}^2 - U(P)$$

where $\frac{1}{2} m \mathbf{v}^2$ is the kinetic energy, and $U(P)$ is the potential energy of the pendulum at the point P of the circle M. Consequently, the Lagrangian

$$\boxed{L : TM \to \mathbb{R}}$$

is a real-valued function on the state space TM (tangent bundle of the position space M). In order to get an intuitive geometric interpretation of the tangent bundle TM, let us assign to the vector $\mathbf{v}$ the coordinate v. The map

$$(P, \mathbf{v}) \mapsto (P, v)$$

is a bijection from TM onto the product set $M \times \mathbb{R}$. In terms of geometry, the product set can be regarded as a cylindrical surface with the circle M as equator. The fibers $\{(P, v) : v \in \mathbb{R}\}$ of the cylindrical surface are generating straight lines perpendicular to the equator.

The motion $\mathbf{x} = \mathbf{x}(t)$ of the circular pendulum corresponds to a curve C in the tangent bundle TM (Fig. 6.11(b)).

From the point of view of the theory of bundles, product bundles like $M \times \mathbb{R}$ are trivial. For general mechanical systems, as a rule, the tangent bundle is not a trivial product, but it is obtained by gluing together local products.

This gluing procedure generates nontrivial topological properties.

The tangent bundle TM can be synonymously written as the coproduct (or disjoint union) of all the tangent spaces TM_P of the base manifold M:

$$TM = \coprod_{P \in M} T_P M.$$

One also writes $TM = \{T_P M\}_{P \in M}$ (bundle of the tangent spaces).

Duality and the costate space (phase space). Duality plays a crucial role in mathematics in order to substantially simplify investigations. Let us dualize the concept of the tangent bundle. To begin with, define the differential form $d\varphi_P$ at the point P by setting

$$d\varphi_P(\mathbf{v}) = v \qquad \text{for all} \quad \mathbf{v} \in T_P M.$$

In particular, $d\varphi_P(\mathbf{e}_P) = 1$. In terms of physics, the linear functional

$$d\varphi_P : T_P M \to \mathbb{R}$$

describes a special measurement process, which assigns to the velocity vector $\mathbf{v}$ the real number v (velocity component). By definition, the dual tangent space $(T_P M)^d$ consists of all linear functionals $F : T_P M \to \mathbb{R}$. The elements of $(T_P M)^d$ are called cotangent vectors (or covelocity vectors), and $(T_P M)^d$ itself is called the cotangent space of the circle M at the point P. Each cotangent vector $F \in (T_P M)^d$ can be represented by

$$F = p \cdot d\varphi_P.$$

The real number p is called the coordinate of F at the point P. Then

$$F(\mathbf{v}) = F(v\mathbf{e}_P) = v \cdot d\varphi_P(\mathbf{e}_P) = vp.$$

The tuple (P, F) is called a costate. The set of costates

$$\boxed{TM^d := \{(P, F) : \quad P \in M, \;\; F \in (T_P M)^d\}}$$

is called the costate space (or the phase space) of the pendulum. In mathematics, TM is called the cotangent bundle of the circle M. Assigning the coordinate p to the covector F, the map

$$(P, F) \mapsto (P, p)$$

is a bijection from TM^d onto the product set $M \times \mathbb{R}$. Using the notion of disjoint union, the cotangent bundle can be synonymously written as

$$TM^d = \coprod_{P \in M} (T_P M)^d.$$

One also writes $TM^d = \{(T_P M)^d\}_{P \in M}$ (bundle of the dual tangent spaces). As we will show later on, there exists a quite natural diffeomorphism

$$\boxed{\mathcal{L} : TM \to TM^d}$$

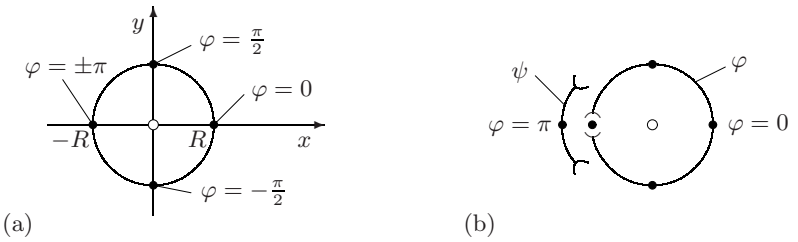


Fig. 6.12. Local coordinates of a circle

from the tangent bundle TM onto the cotangent bundle TM^d of the position space (manifold) M. This diffeomorphism $\mathcal{L}$ is called Legendre transformation, which is a contact transformation in the sense of Sophus Lie.

Local coordinates. By definition, an n-dimensional manifold looks locally like an open set in $\mathbb{R}^n$; the points in $\mathbb{R}^n$ represent local coordinates of the manifold.[23] In this setting, the angle variable φ is not an admissible coordinate for the global circle. In fact, it is impossible to map homeomorphically the circle M onto an open subset Ω of the real line $\mathbb{R}$, since the first Betti numbers are different. Explicitly, we have $\beta_1(M) = 1$ and $\beta_1(\Omega) = 0$. In fact, for the circle, we need at least two local coordinate systems with coordinates denoted by φ and ψ (Fig. 6.12(b)). For example,

- φ varies in the open interval $]-\pi, \pi[$;
- ψ varies in the open interval, say, $]\pi - \frac{\pi}{4}, \pi + \frac{\pi}{4}[$;
- the point $(x, y) = (R, 0)$ on the circle has the local coordinate $\varphi = 0$, and the point $(-R, 0)$ on the circle has the local coordinate $\psi = \pi$.

It is the goal of the analysis on manifolds to introduce such quantities which do not depend on the choice of local coordinates. Such properties possess an invariant geometric (or physical) meaning. This will be thoroughly studied in Vol. III. The general theory of invariants is a vivid branch of mathematics with many applications to physics; it was created by Cayley (1821–1895), and further developed by Weyl (1885–1955). We refer to:

H. Weyl, The Classical Groups: Their Invariants and Representations, Princeton University Press, 1946.

P. Olver, Classical Invariant Theory, Cambridge University Press, 1999.

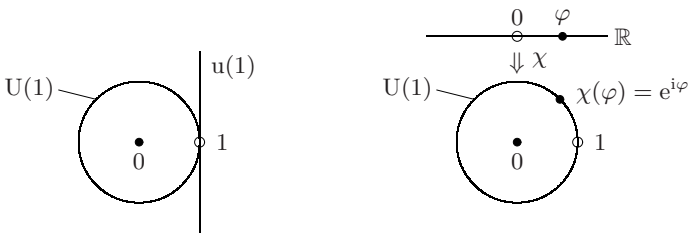
The covering space of the position space. The true topological character of the global angle variable φ becomes clear if we consider the surjective map $\chi : \mathbb{R} \to M$ given by

$$\boxed{\chi(\varphi) := R\mathrm{e}^{\mathrm{i}\varphi} \qquad \text{for all} \quad \varphi \in \mathbb{R}.} \tag{6.62}$$

This is called a covering map of the circle M, and the real line $\mathbb{R}$ is called a covering space of M (Fig. 6.13). Since the real line is simply connected, it is called a universal covering space of the circle M. The map χ is smooth and has the period 2π.

A deep mathematical result tells us that all Lie groups and all compact Riemann surfaces have universal covering spaces.

[23] The precise definition of manifolds and diffeomorphisms, as well as topological spaces and homeomorphisms can be found in Sects. 5.4ff of Vol. I.



(a) tangent Lie algebra $u(1)$ (b) universal covering group $\mathbb{R}$

Fig. 6.13. The Lie group U(1)

Equivalently, this tells us that there exist global parametrizations of algebraic curves, by the famous 1907 uniformization theorem due to Koebe (1882–1945) and Poincaré (1854–1912). For example, elliptic functions are global parametrizations of elliptic curves. As an introduction to uniformization, we recommend:

L. Pontryagin, Topological Groups, Gordon and Breach, New York, 1966.

J. Jost, Compact Riemann Surfaces: An Introduction to Contemporary Mathematics, 3rd edition, Springer, Berlin, 2006.

R. Narasimhan, Compact Riemann Surfaces, 2nd edition, Birkhäuser, Basel, 1996.

The position space as a Lie group. Choose $R = 1$, that is, the pendulum has the length one. Then the position space M can be identified with the Lie group

$$U(1) := \{z \in \mathbb{C} : |z| = 1\}.$$

Note that the elements of $U(1)$ can be represented by $z = \mathrm{e}^{\varphi \mathrm{i}}$ with $\varphi \in \mathbb{R}$.

The additive (simply connected) Lie group $\mathbb{R}$ is called the *universal covering group* of the Lie group $U(1)$ (Fig 6.13). In a Cartesian (x, y)-coordinate system with the right-handed orthonormal basis $\mathbf{i}, \mathbf{j}$, the tangent space of the unit circle $U(1)$ at the unit element 1 is given by

$$T_1 U(1) = \{v\mathbf{j} : \quad v \in \mathbb{R}\}.$$

By definition, this tangent space is called the Lie algebra $u(1)$ of the Lie group $U(1)$. In terms of complex numbers, this can be written as

$$u(1) = \{v\mathrm{i} : \quad v \in \mathbb{R}\}.$$

This is the space of purely imaginary numbers. The passage

$$U(1) \Rightarrow u(1) \tag{6.63}$$

from the Lie group $U(1)$ to the Lie algebra $u(1)$ is given by the Taylor expansion

$$\boxed{\mathrm{e}^{\varphi \mathrm{i}} = 1 + \varphi \mathrm{i} + o(\varphi), \qquad \varphi \to 0.}$$

Therefore, the passage (6.63) from the Lie group to the Lie algebra is nothing else than an application of the linearization principle to Lie groups.

First-order differential operators. Let $C^\infty(M)$ be the space of all smooth 2π-periodic functions $f : \mathbb{R} \to \mathbb{R}$. These functions can be regarded as smooth functions on the circle M. By definition, a first-order differential operator

$$D : C^\infty(M) \to C^\infty(M)$$

on the circle M has the form

$$D := v(\varphi)\frac{d}{d\varphi}, \qquad \varphi \in \mathbb{R}$$

with the fixed function $v \in C^\infty(M)$. Explicitly, for all $f \in C^\infty(M)$, we get

$$(Df)(\varphi) = v(\varphi)\frac{df(\varphi)}{d\varphi}, \qquad \varphi \in \mathbb{R}.$$

Furthermore, let $\mathbf{Vec}(M)$ denote the space of all smooth velocity vector fields $\mathbf{v}$ on the circle M. Explicitly, the velocity vector field $\mathbf{v}$ is given by

$$v(\varphi)\mathbf{e}_{P(\varphi)} \qquad \text{for all} \quad \varphi \in \mathbb{R},$$

where $v \in C^\infty(M)$. Define

$$\boxed{\mathbf{v}[f] := Df \qquad \text{for all} \quad f \in C^\infty(M).} \tag{6.64}$$

This way, we obtain a one-to-one map

$$\mathbf{v} \mapsto D$$

between the smooth velocity vector fields $\mathbf{v}$ on the circle M and the first-order differential operators D on M (with smooth coefficients). The identification $\mathbf{v} \mapsto D$ between velocity vector fields and first-order differential operators dates back to Sophus Lie (1842–1899); it is frequently used in the modern mathematical literature on finite-dimensional manifolds. However, such an identification does not exist on infinite-dimensional manifolds. Therefore, in this series of monographs, we will use an approach to velocity vectors on manifolds which fits both the finite-dimensional case and the infinite-dimensional case.

Sections and physical fields. We want to show that the modern notion of the section of a bundle comprehends

- vector fields on M (sections of the tangent bundle TM),
- first-order differential operators (sections of the tangent bundle TM),
- differential forms on M (sections of the cotangent bundle TM^d).

The map

$$\boxed{s : M \to TM} \tag{6.65}$$

given by $s(P) := \mathbf{v}(P)$ with $\mathbf{v}(P) \in T_P M$ is called a section of the tangent bundle TM. In terms of physics, this is a velocity field

$$P \mapsto \mathbf{v}(P)$$

on the circle M. We will show in Vol. III that physical fields are sections of vector bundles or sections of more general fiber bundles (vector fields, tensor fields, spinor fields, and so on). If we identify the velocity vector field $\mathbf{v} = \mathbf{v}(P)$ with the

differential operator D, then the section (6.65) equivalently describes a first-order differential operator $P \to v(P)\frac{d}{\varphi}$ on the circle. Dually, the map

$$\boxed{s : M \to TM^d} \tag{6.66}$$

given by $s(P) := F_P$ with $F_P \in T_P M$ is called a section of the cotangent bundle TM^d. This is a differential form $P \mapsto p(P)d\varphi_P$ on the circle M.

The notion of section generalizes the notion of function in classical analysis.

In fact, the real function $f : \mathbb{R} \to \mathbb{R}$ can be identified with the map

$$s : \mathbb{R} \to T\mathbb{R}$$

given by $s(x) := (x, f(x))$. Here, the tangent bundle of the real line $\mathbb{R}$ is defined by

$$T\mathbb{R} := \{(x, v) : x \in \mathbb{R}, \ \ v \in \mathbb{R}\}.$$

In other words, $T\mathbb{R} = \mathbb{R}^2$. The map s is a section of the tangent bundle $T\mathbb{R}$ of the real line.

Perspective. For a general mechanical system of n degrees of freedom with $n = 1, 2, \ldots$, the following hold:

- the position space M is a real n-dimensional manifold,
- the state space TM (tangent bundle of M) is a $2n$-dimensional manifold, and
- the phase space TM^d (cotangent bundle) is a $2n$-dimensional symplectic manifold.

The Lagrangian formulation of mechanics is based on the Lagrangian

$$L : TM \to \mathbb{R},$$

whereas the dual Hamiltonian formulation of mechanics is based on the Hamiltonian

$$H : TM^d \to \mathbb{R},$$

which represents the energy function of the mechanical system, as a rule. In contrast to the cotangent bundle TM^d, the tangent bundle TM is not always a symplectic manifold. Therefore, the Hamiltonian formulation of mechanics has advantages over the Lagrangian formulation. In statistical mechanics, the Hamiltonian approach will allow us to use the volume measure of the phase space in order to construct the key probability measure (see Sect. 7.17.5 on page 645). Summarizing, the language of manifolds allows us to study the global aspects of the motion of mechanical systems.

The complexity of a mechanical system is reflected by the complex topology of the tangent bundle and the cotangent bundle.

6.8 Hamiltonian Mechanics

The Hamiltonian approach to mechanics is centered at the concept of energy. In this setting, the following two important results are quite natural:

- Conservation of energy, and
- conservation of phase volume.

The latter property is crucial for classical statistical physics (Gibbs measure). From the geometric point of view, the passage from Lagrangian mechanics to Hamiltonian mechanics corresponds to a passage

- from the tangent bundle TM of the position space M (position, velocity vector)
- to the dual cotangent bundle TM^d of M (position, differential form).

This transformation is called the Legendre transformation which is a contact transformation in the sense of Lie. The cotangent bundle TM^d always carries a natural symplectic structure. Note the following:

- Lagrangian mechanics corresponds to Riemannian geometry (the metric is given by the kinetic energy).
- Hamiltonian mechanics corresponds to symplectic geometry.

One of the great problems of mathematics and physics in the 19th century consisted in solving the N-body problem in celestial mechanics.

The aim was to investigate the stability of our solar system.

To simplify considerations, Jacobi and his successors used transformations of time and position which preserve the form of the canonical equations.

Such transformations are called canonical transformations.

It turns out that canonical transformations coincide with symplectic transformations. That is, canonical transformations preserve the symplectic structure, and hence they present the symmetry transformations of symplectic geometry. When creating his mechanics, Hamilton (1805–1865) was motivated by an analogy between mechanics and Huygens' geometrical optics created in the 17th century:

- The trajectories $q = q(t), p = p(t)$ of a particle in mechanics correspond to light rays (solutions of the canonical ordinary differential equations), and
- the action function $S = S(q, t)$ in mechanics corresponds to the eikonal function in geometric optics, which determines the wave fronts (solutions of the Hamilton–Jacobi partial differential equation).
- The duality between trajectories and wave fronts was fully established by Carathéodory in the framework of his 'royal road to the calculus of variations' in 1925. This is based on the fundamental notion of geodesic fields.

The relation between geometric optics and wave optics plays a fundamental role for understanding Schrödinger's quantum mechanics which he discovered in 1926. In the 1950s, optimal control theory was invented independently by Bellman and Pontryagin. This theory allows many applications in technology (e.g., moon landing and return of a spaceship to earth):

- Pontryagin's theory generalizes Hamilton's canonical equations (Pontryagin's maximum principle and light rays), whereas
- Bellman's theory of dynamic programming generalizes the Hamilton–Jacobi partial equation (the Hamilton–Jacobi–Bellman equation and wave fronts).

In the Bellman approach, the action function S corresponds to the cost function. A detailed study can be found in Zeidler (1986), Vol. III (see the references on page 1049). In what follows, let us study the main ideas on an elementary level by considering the harmonic oscillator as a prototype.

6.8.1 The Canonical Equation

> In Hamiltonian mechanics, we pass from position and velocity to position and momentum.
>
> Folklore

The Newtonian equation $m\ddot{q} = -m\omega^2 q$ for the motion of the harmonic oscillator can be written as a first order system

$$\dot{q} = \frac{p}{m}, \qquad \dot{p} = -m\omega^2 q. \tag{6.67}$$

Here, the variable p is called momentum. Introducing the so-called Hamiltonian

$$H(q,p) := \frac{p^2}{2m} + \frac{m\omega^2 q^2}{2},$$

equation (6.67) is identical with the so-called canonical equation

$$\boxed{\dot{q} = H_p, \qquad \dot{p} = -H_q.} \tag{6.68}$$

Explicitly, for all times $t \in \mathbb{R}$,

$$\dot{q}(t) = H_p(q(t), p(t)), \qquad \dot{p}(t) = -H_q(q(t), p(t)).$$

Note that H represents the energy function of the harmonic oscillator. In terms of the Lagrangian

$$L = \frac{m\dot{q}^2}{2} - \frac{m\omega^2 q^2}{2},$$

we obtain

$$p = L_{\dot{q}}, \qquad H = p\dot{q} - L.$$

The transformation from the variables $q, \dot{q}$ to q, p and from the Lagrangian

$$L = L(q, \dot{q})$$

to the Hamiltonian $H = H(q,p)$ is called Legendre transformation.

6.8.2 The Hamiltonian Flow

The Hamiltonian formulation (6.68) possesses the advantage over the Newtonian formulation that we can apply the well-elaborated methods of the theory of dynamical systems to mechanics. Let us consider this. The unique solution of the initial-value problem

$$q(0) = q_0, \quad p(0) = p_0$$

to the canonical equation (6.67) is given by

$$q(t) := q_0 \cos \omega t + \frac{p_0}{m\omega} \sin \omega t, \tag{6.69}$$

$$p(t) := -m\omega q_0 \sin \omega t + p_0 \cos \omega t.$$

Define

$$\boxed{F_t(q_0, p_0) := (q(t), p(t)), \qquad t \in \mathbb{R}.}$$

This way, we get a flow on the phase space $\{(q,p) \in \mathbb{R}^2\}$ which is called the Hamiltonian flow corresponding to H (Fig. 6.14). The trajectories are ellipses. The flow transports the point (q_0, p_0) at time $t = 0$ to the point $(q(t), p(t))$ at time t. The flow can be interpreted as

- the motion of fluid particles, or
- the propagation of light rays.[24]

[24] In the case of light rays, the symbol t denotes a curve parameter, but not time.

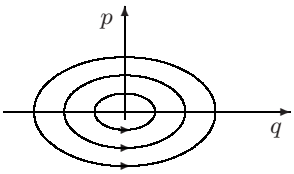


Fig. 6.14. Hamiltonian flow on phase space

Proposition 6.9 *The Hamiltonian flow* $\{F_t\}$ *has the following two properties.*

(i) *The energy H(q(t),p(t)) is constant along each trajectory.*
(ii) *The flow is volume preserving.*

Proof. Ad (i). This follows from energy conservation.

Ad (ii). The flow transports the domain G in phase space at time $t = 0$ to the domain G_t at time t. Observe

$$\text{meas } G_t = \int_{G_t} dqdp = \int_G D(t)dq_0dp_0 = \int_G dq_0dp_0 = \text{meas } G.$$

Setting $\alpha := q_0, \beta := p_0$, observe that (6.69) implies

$$D(t) := \frac{\partial(q(t), p(t))}{\partial(\alpha, \beta)} = \begin{pmatrix} \frac{\partial q(t)}{\partial \alpha} & \frac{\partial p(t)}{\partial \alpha} \\ \frac{\partial q(t)}{\partial \beta} & \frac{\partial p(t)}{\partial \beta} \end{pmatrix} = \begin{pmatrix} \cos \omega t & -m\omega \sin \omega t \\ \frac{1}{m\omega} \sin \omega t & \cos \omega t \end{pmatrix} = 1.$$

□

6.8.3 The Hamilton–Jacobi Partial Differential Equation

The action. The quantity

$$S[q] := \int_{t_0}^{t_1} L(q(t), \dot{q}(t))\, dt$$

is called the action along the trajectory $q = q(t)$, $a \le t \le b$. The action is the most fundamental quantity in physics. Let us study the action along trajectories of the harmonic oscillator. For given real values T, Q, consider the following boundary-value problem for the Newtonian equation

$$m\ddot{q}(t) + \omega^2 q(t) = 0, \qquad t \in \mathbb{R}, \tag{6.70}$$

$$q(0) = 0, \quad q(T) = Q.$$

To simplify notation, let $m = \omega = 1$. If $0 < T < \pi$, problem (6.70) has the unique solution

$$q = v \sin t, \qquad v := \frac{Q}{\sin T}. \tag{6.71}$$

Varying the parameters T, Q, we obtain a family of trajectories through the origin (Fig. 6.15(a)). The action along the trajectory (6.71) is given by

$$S(T, Q) := \int_0^T \tfrac{1}{2}(\dot{q}(t)^2 - q(t)^2)\, dt = \tfrac{1}{2}Q^2 \cot T.$$

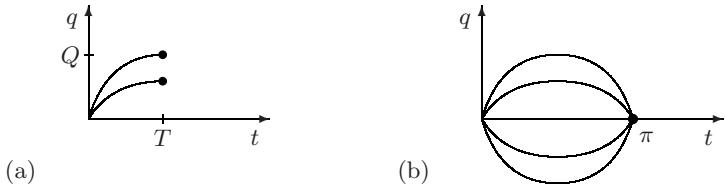


Fig. 6.15. Trajectories and focal points

Introducing the momentum $p(T) = \dot{q}(T) = v \cos T$ and the energy

$$H(q(T), p(T)) = \tfrac{1}{2}\left(p(T)^2 + q(T)^2\right) = \tfrac{1}{2}v^2,$$

we obtain

$$S_Q(T, Q) = p(T) \quad \text{and} \quad S_T(T, Q) = -H(q(T), p(T)).$$

This implies the so-called *Hamilton–Jacobi partial differential equation*

$$\boxed{S_T(T, Q) + H(Q, S_Q(T, Q)) = 0.} \tag{6.72}$$

Focal points. If $Q = 0$ and $T = \pi$, the trajectory from (6.70), (6.71) is not uniquely determined by the boundary condition. Each function $q(t) := v \sin t,\ t \in \mathbb{R}$ satisfies the condition $q(\pi) = 0$. In this critical case, a focal point occurs (Fig. 6.15(b)). In fact, the action $S(T, Q) = \frac{1}{2}Q^2 \cot T$ is singular at the point $T = \pi$. Such singularities are responsible for typical difficulties which arise in the theory of the Hamilton–Jacobi equation. In geometric optics, such singularities correspond to focal points and caustics of lenses, which are caused by the intersection and the envelops of families of light rays, respectively.

6.9 Poissonian Mechanics

The Poissonian formulation of mechanics has the following two advantages:

- Conservation laws can be expressed in terms of Poisson brackets.[25]
- Quantum mechanics can be obtained from classical mechanics by replacing the Poisson bracket $\{A, B\}$ with the commutator

$$\frac{1}{\mathrm{i}\hbar} \cdot [A, B]_-$$

where the Lie bracket $[A, B]_-$ is equal to $AB - BA$. This was implicitly discovered by Heisenberg in 1925. The general quantization rule was formulated by Dirac in 1926.

[25] Poisson (1781–1840).

6.9.1 Poisson Brackets and the Equation of Motion

Consider the space $C^\infty(\mathbb{R}^2)$ of smooth real functions $F = F(q,p)$. Define the Poisson bracket

$$\{A,B\} := A_q B_p - B_q A_p,$$

where $A_q = \partial A/\partial q$, and so on. In particular,

$$\{q,p\} = 1, \quad \{q,q\} = \{p,p\} = 0.$$

Let $q = q(t), p = p(t)$ be the motion of a particle which satisfies the canonical equations

$$\dot{p} = -H_q, \quad \dot{q} = H_p$$

for a given Hamiltonian $H = H(q,p)$. Then

$$\boxed{\dot{F} = \{F,H\}} \tag{6.73}$$

for each $F \in C^\infty(\mathbb{R}^2)$. Explicitly,

$$\frac{dF}{dt}(q(t),p(t)) = \{F,H\}(q(t),p(t)).$$

In fact, $\dfrac{dF}{dt} = F_q\dot{q} + F_p\dot{p} = F_q H_p - F_p H_q = \{F,H\}$.

6.9.2 Conservation Laws

Proposition 6.10 *If $\{F,H\} \equiv 0$, then F is a conserved quantity.*

Proof. By (6.73), $\frac{d}{dt}F(q(t),p(t)) = 0$. Hence $F(q(t),p(t)) = \text{const}$ for all t. □

For example, the Hamiltonian H is a conserved quantity, since obviously $\{H,H\} = 0$. This corresponds to conservation of energy.

The space $C^\infty(\mathbb{R}^2)$ forms a real Lie algebra with respect to the Poisson brackets.

Proof. For all functions $A, B, C \in C^\infty(\mathbb{R}^2)$ and all real numbers α and β, we have to show that:

(i) $\{A,B\} = -\{B,A\}$,
(ii) $\{\alpha A + \beta B, C\} = \alpha\{A,C\} + \beta\{B,C\}$,
(iii) $\{A,\{B,C\}\} + \{B,\{C,A\}\} + \{C,\{A,B\}\} = 0$ (Jacobi identity).

For example, $\{A,B\} = A_qB_p - B_qA_p = -\{B,A\}$. Similarly, we get the other claims. □

If A and B are conserved quantities, then so is $\{A,B\}$.

Proof. Set $C := \{A,B\}$. By the Jacobi identity, $\{H,A\} = \{H,B\} = 0$ implies $\{H,C\} = \{H,\{A,B\}\} = -\{A,\{B,H\}\} - \{B,\{H,A\}\} = 0$. □

6.10 Symplectic Geometry

A deeper understanding of classical mechanics is based on symplectic geometry. By definition, a quantity belongs to symplectic geometry iff it is invariant under symplectic transformations. Let us explain the basic ideas.

6.10.1 The Canonical Equations

Set $\mathbf{q}(t) := (q(t), p(t))$ and $\dot{\mathbf{q}}(t) := (\dot{q}(t), \dot{p}(t))$. We want to consider the motion $q = q(t), p = p(t)$ of a particle which satisfies the canonical equations

$$\dot{p}(t) = -H_q(\mathbf{q}(t)), \quad \dot{q}(t) = H_p(\mathbf{q}(t)). \tag{6.74}$$

Using the language of differential forms, this equation can elegantly be written as

$$\boxed{\dot{\mathbf{q}} \rfloor \omega = dH} \tag{6.75}$$

along the trajectory $\mathbf{q} = \mathbf{q}(t)$, where

$$\omega := dq \wedge dp$$

is called the symplectic form of $\mathbb{R}^2$. To show this, recall the definitions

$$dq(\mathbf{w}) := a \quad \text{and} \quad dp(\mathbf{w}) := b$$

for all $\mathbf{w} = (a, b)$ in $\mathbb{R}^2$. Furthermore,

$$(dq \wedge dp)(\mathbf{v}, \mathbf{w}) := dq(\mathbf{v})dp(\mathbf{w}) - dq(\mathbf{w})dp(\mathbf{v})$$

for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$. Finally, the inner product $\mathbf{v} \rfloor \omega$ is defined by

$$(\mathbf{v} \rfloor \omega)(\mathbf{w}) := \omega(\mathbf{v}, \mathbf{w}).$$

It follows from $dH = H_q(\mathbf{q}(t))\, dq + H_p(\mathbf{q}(t))\, dp$ that

$$dH(\mathbf{w}) = H_q(\mathbf{q}(t))\, a + H_p(\mathbf{q}(t))\, b.$$

On the other hand,

$$(\dot{\mathbf{q}} \rfloor \omega)(\mathbf{w}) := (dq \wedge dp)(\dot{\mathbf{q}}, \mathbf{w}) = dq(\dot{\mathbf{q}})dp(\mathbf{w}) - dq(\mathbf{w})dp(\dot{\mathbf{q}}).$$

Thus, the equation

$$(\dot{\mathbf{q}} \rfloor \omega)(\mathbf{w}) = dH(\mathbf{w})$$

for all $\mathbf{w} = (a, b)$ in $\mathbb{R}^2$ is equivalent to

$$\dot{q}(t)b - \dot{p}(t)a = H_q(\mathbf{q}(t))\, a + H_p(\mathbf{q}(t))\, b \qquad \text{for all} \quad a, b \in \mathbb{R}.$$

In turn, this is equivalent to the canonical equations (6.74).

Proposition 6.11 *For all $A, B \in C^\infty(\mathbb{R}^2)$,*

$$dA \wedge dB = \{A, B\}\, dq \wedge dp.$$

Proof. Note that $dq \wedge dq = dp \wedge dp = 0$. Therefore, it follows from

$$dA = A_q dq + A_p dp \quad \text{and} \quad dB = B_q dq + B_p dp$$

that $dA \wedge dB = (A_q B_p - B_q A_p)\, dq \wedge dp$. This completes the proof. □

6.10.2 Symplectic Transformations

Consider a diffeomorphism

$$Q = Q(q,p), \quad P = P(q,p), \qquad (q,p) \in \Omega \tag{6.76}$$

from an open set Ω in $\mathbb{R}^2$ onto another open set in $\mathbb{R}^2$. Such a diffeomorphism is called a symplectic transformation iff

$$dQ \wedge dP = dq \wedge dp \qquad \text{on } \Omega.$$

This is equivalent to the condition

$$\frac{\partial(Q,P)}{\partial(q,p)} \equiv 1 \qquad \text{on } \Omega.$$

That is, each symplectic transformation is volume preserving. In fact, this follows from

$$dQ \wedge dP = \{Q,P\}\, dq \wedge dp$$

along with $\{Q,P\} = \partial(Q,P)/\partial(q,p)$.

Proposition 6.12 *Poisson brackets are invariant under symplectic transformations.*

Proof. Define the transformed functions

$$\mathcal{A}(Q,P) := A(q,p) \quad \text{and} \quad \mathcal{B}(Q,P) := B(q,p),$$

where (q,p) and (Q,P) are related to each other through (6.76). Then

$$dA \wedge dB = \{A,B\}\, dq \wedge dp \quad \text{and} \quad d\mathcal{A} \wedge d\mathcal{B} = \{\mathcal{A},\mathcal{B}\}\, dQ \wedge dP.$$

Since the $\wedge$-product is invariant under diffeomorphisms,

$$dA \wedge dB = d\mathcal{A} \wedge d\mathcal{B}.$$

Therefore, $dQ \wedge dP = dq \wedge dp$ implies $\{A,B\}(q,p) = \{\mathcal{A},\mathcal{B}\}(Q,P)$. □

Proposition 6.13 *Symplectic transformations preserve the structure of the Hamiltonian canonical equations.*

In classical mechanics, such transformations are called canonical transformations. They were introduced by Jacobi in order to simplify the solution of complicated problems in celestial mechanics.[26]

Proof. We are given the Hamiltonian $H = H(q,p)$. The canonical equations

$$\dot{q} = H_p, \quad \dot{p} = -H_p$$

can be written as

$$\dot{q} = \{q,H\}, \quad \dot{p} = \{p,H\}.$$

For the transformed motion $Q(t) := Q(q(t),p(t))$, $P(t) := P(q(t),p(t))$, we get

[26] The classical theory of canonical transformations can be found in C. Carathéodory, Calculus of Variations and Partial Differential Equations of First Order, Chap. 6, Chelsea, New York, 1982.

$$\dot{Q} = Q_q \dot{q} + Q_p \dot{p} = Q_q H_p - Q_p H_q.$$

Hence

$$\dot{Q} = \{Q, H\}, \quad \dot{P} = \{P, H\}.$$

Define the transformed Hamiltonian $\mathcal{H}(Q, P) := H(q, p)$. By Prop. 6.12 above, we obtain

$$\dot{Q} = \{Q, \mathcal{H}\}, \quad \dot{P} = \{P, \mathcal{H}\}.$$

Hence $\dot{Q} = \mathcal{H}_P, \dot{P} = -\mathcal{H}_Q$. □

Now consider the trajectories

$$q = q(t; Q, P), \quad p = p(t; Q, P) \tag{6.77}$$

of the canonical equations with respect to the Hamiltonian $H = H(q, p)$ together with the initial conditions $q(0) = Q, p(0) = P$. Define

$$\boxed{F_t(Q, P) := (q(t), p(t)).}$$

This way, for each time t, we get a transformation $(Q, P) \mapsto (q(t), p(t))$.

Theorem 6.14 *The flow operator F_t is symplectic for each time t.*

Proof. We have to show that

$$\frac{\partial(q, p)}{\partial(Q, P)} = 1 \quad \text{for all} \quad t,$$

where q and p are given through (6.77). This is identical with

$$\{p, q\}(Q, P) = 1 \quad \text{for all} \quad t.$$

At the initial time $t = 0$, we have $q(0; Q, P) = Q$ and $p(0; Q, P) = P$. This implies

$$\{q, p\}(Q, P) = 1 \quad \text{for} \quad t = 0.$$

Therefore, it remains to show that

$$\frac{d}{dt}\{q, p\} = 0 \quad \text{for all} \quad t.$$

To this end, consider the canonical equations

$$q_t = \{q, H\}, \quad p_t = \{p, H\}.$$

By the product rule of calculus, we obtain

$$\frac{d}{dt}\{q, p\} = \{q_t, p\} + \{q, p_t\} = \{\{q, H\}, p\} + \{q, \{p, H\}\}.$$

Thus, the Jacobi identity yields $\frac{d}{dt}\{q, p\} = \{\{q, p\}, H\} = 0$, since $\{q, p\} = 1$. □

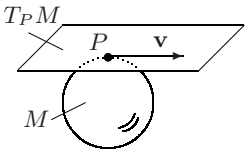


Fig. 6.16. Motion of a particle on a sphere $M = \mathbb{S}^2_R$

6.11 The Spherical Pendulum

> Two-dimensional spheres are the simplest curved surfaces. They serve as prototypes for the geometry and analysis of manifolds.
>
> Folklore

We want to study the motion of a point of mass m on a sphere $\mathbb{S}^2_R$ of radius R under the influence of the gravitational force. This is called a spherical pendulum (Fig. 6.16). In a Cartesian (x, y, z)-coordinate system, the position of a point of the sphere is described by the vector

$$\mathbf{q} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \tag{6.78}$$

with initial point at the center of the ball. The equation of the sphere reads as

$$x^2 + y^2 + z^2 = R^2.$$

The point $(0, 0, R)$ (resp. $(0, 0, -R)$) is called the North Pole (resp. South Pole).

6.11.1 The Gaussian Principle of Critical Constraint

The Newtonian equation of motion for the spherical pendulum reads as

$$m\ddot{\mathbf{q}}(t) = -mg\mathbf{k} + \mathbf{F}_c(\mathbf{q}(t), \dot{\mathbf{q}}(t)), \qquad t \in \mathbb{R}. \tag{6.79}$$

Here, the vector $-gm\mathbf{k}$ describes the gravitational force on the surface of earth. This force acts in direction of the negative z-axis. The additional constraining force $\mathbf{F}_c$ keeps the particle on the sphere. In order to compute the constraining force, let us use the most general principle in classical mechanics, namely, the Gaussian principle of critical constraint (see Sect. 7.28 of Vol. I). We have to solve the following extremal principle:

$$(m\ddot{\mathbf{q}} - \mathbf{F})^2 = \text{critical!} \tag{6.80}$$

with the gravitational force $\mathbf{F} = -gm\mathbf{k}$ and the constraint $\mathbf{q}^2 = R^2$. As we have shown on page 492 of Vol. I, this leads to the constraining force

$$\mathbf{F}_c(\mathbf{q}, \dot{\mathbf{q}}) = \lambda(\mathbf{q}, \dot{\mathbf{q}}) \cdot \mathbf{q}$$

which is a normal force depending on position and velocity. Explicitly,

$$\lambda(\mathbf{q}, \dot{\mathbf{q}}) := \frac{mg \cdot \mathbf{q}\mathbf{k} - \dot{\mathbf{q}}^2}{R^2}.$$

Using this, the equation of motion (6.79) looks rather complicated. It is our goal to simplify the approach.

In order to eliminate the constraining force $\mathbf{F}_c$*, we will use the principle of critical action in terms of spherical coordinates.*

The gravitational force on the surface of earth has the potential $U(z) := mgz$. In fact, the gravitational force $-mg\mathbf{k}$ is equal to $-U'(z)\mathbf{k}$. Therefore, the Lagrangian, $L :=$ *kinetic energy minus potential energy*, reads as

$$L(\mathbf{q}, \dot{\mathbf{q}}) := \tfrac{1}{2}m\dot{\mathbf{q}}^2 - mgz. \tag{6.81}$$

This Lagrangian is basic for the following approach.

6.11.2 The Lagrangian Approach

In mechanics it is important to use the appropriate coordinates.
Folklore

Spherical coordinates. Let φ be the geographic length of the sphere, and let ϑ be the geographic latitude with $-\pi \le \varphi \le \pi$ and $-\frac{\pi}{2} \le \vartheta \le \frac{\pi}{2}$. The equator is described by the equation $\vartheta = 0$ (see Fig. 5.16 on page 305). For the relation between Cartesian coordinates x, y, z and spherical coordinates, we get

$$x = R\cos\varphi\cos\vartheta, \quad y = R\sin\varphi\cos\vartheta, \quad z = R\sin\vartheta. \tag{6.82}$$

The coordinate line $\vartheta = \text{const}$ is a parallel line of latitude; it has the tangent vector

$$\mathbf{b}_1(P) := \frac{\partial \mathbf{q}}{\partial \varphi}(P) = R(-\sin\varphi\cos\vartheta\ \mathbf{i} + \cos\varphi\cos\vartheta\ \mathbf{j})$$

at the point $P = (\varphi, \vartheta)$, by (6.78). Similarly, the meridian $\varphi = \text{const}$ has the tangent vector

$$\mathbf{b}_2(P) := \frac{\partial \mathbf{q}}{\partial \vartheta}(P) = R(-\cos\varphi\sin\vartheta\ \mathbf{i} - \sin\varphi\sin\vartheta\ \mathbf{j} + \cos\vartheta\ \mathbf{k})$$

at the point P. The two vectors $\mathbf{b}_1, \mathbf{b}_2$ form an orthogonal basis of the tangent plane $T_P\mathbb{S}^2_R$ of the sphere at the point P.[27] The exterior unit normal vector $\mathbf{N}$ of the sphere at the point P is given by

$$\mathbf{N} := \mathbf{e}_1 \times \mathbf{e}_2, \quad \text{where } \mathbf{e}_j := \frac{\mathbf{b}_j(P)}{|\mathbf{b}_j(P)|}, \quad j = 1, 2.$$

The three vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{N}$ form a right-handed orthonormal system at the point P of the sphere $\mathbb{S}^2_R$.

Velocity vector (tangent vector). The motion of a point on the sphere is described by the position vector

$$\mathbf{q}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}, \qquad t \in \mathbb{R}.$$

Differentiating the functions

$$x(t) = R\cos\varphi(t)\sin\vartheta(t), \quad y(t) = R\sin\varphi(t)\cos\vartheta(t), \quad z(t) = R\sin\vartheta(t)$$

with respect to time t, we obtain

[27] The point P has to be different from the North Pole and the South Pole, since the spherical coordinates are singular at the two poles. This can be cured by introducing new local coordinates near the two poles (e.g., the coordinates (x, y)). To simplify the approach, we restrict ourselves to spherical coordinates.

$$\dot{x}(t) = -R\dot{\varphi}(t)\sin\varphi(t)\cos\vartheta(t) - R\dot{\vartheta}(t)\cos\varphi(t)\sin\vartheta(t).$$

Furthermore, $\dot{z}(t) = R\dot{\vartheta}(t)\cos\vartheta(t)$, and

$$\dot{y}(t) = R\dot{\varphi}(t)\cos\varphi(t)\cos\vartheta(t) - R\dot{\vartheta}(t)\sin\varphi(t)\sin\vartheta(t).$$

Since $\cos^2\alpha + \sin^2\alpha = 1$, we get

$$|\dot{\mathbf{q}}(t)|^2 = \dot{x}(t)^2 + \dot{y}(t)^2 + \dot{z}(t)^2 = R^2(\dot{\varphi}(t)^2\cos^2\vartheta(t) + \dot{\vartheta}(t)^2).$$

The transformed Lagrangian. In terms of spherical coordinates, the Lagrangian reads as

$$L(\varphi,\vartheta,\dot{\varphi},\dot{\vartheta}) = \tfrac{1}{2}mR^2(\dot{\varphi}^2\cos^2\vartheta + \dot{\vartheta}^2) - mgR\sin\vartheta.$$

The principle of critical action postulates that

$$\int_{t_0}^{t_1} L(\varphi(t),\vartheta(t),\dot{\varphi}(t),\dot{\vartheta}(t))\,dt = \text{critical!} \tag{6.83}$$

with $(\varphi(t_0),\vartheta(t_0)) = (\varphi_0,\vartheta_0)$ and $(\varphi(t_1),\vartheta(t_1)) = (\varphi_1,\vartheta_1)$ (boundary condition).

The Euler–Lagrange equation. Each smooth solution of (6.83) satisfies $\frac{d}{dt}L_{\dot{\varphi}} = L_{\varphi}$ and $\frac{d}{dt}L_{\dot{\vartheta}} = L_{\vartheta}$. Explicitly, the Euler–Lagrange equation for the motion of the spherical pendulum looks like

$$\boxed{\frac{d}{dt}\left(\dot{\varphi}(t)\cos^2\vartheta(t)\right) = 0, \qquad \ddot{\vartheta}(t) + \omega^2\cos\vartheta(t) = 0} \tag{6.84}$$

with $\omega := \sqrt{\frac{g}{R}}$. Setting $\vartheta = -\frac{\pi}{2} + \alpha$, the second equation passes over to

$$\ddot{\alpha} + \omega^2\sin\alpha = 0.$$

This is the equation of a circular pendulum. The first equation of (6.84) can be written as

$$\ddot{\varphi}(t) - 2\dot{\varphi}(t)\dot{\vartheta}(t)\sin\vartheta(t) = 0. \tag{6.85}$$

Geodesics. If the gravitational force vanishes (i.e., $g = 0$), then the spherical pendulum is called free. The trajectories of the free pendulum are called geodesics of the sphere. A piece of the equator ($\varphi(t) = t, \vartheta(t) = 0$) is always a geodesic. By definition, a great circle of the sphere is obtained by the intersection between the sphere and a plane passing through the center of the sphere. After a rotation, if necessary, we can assume that the great circle is the equator. Consequently, pieces of great circles are geodesics.

For the motion $\mathbf{q} = \mathbf{q}(t)$ of the free pendulum, it follows from equation (6.79) that the tangential component of the acceleration vector $\ddot{\mathbf{q}}(t)$ vanishes. According to Levi-Civita, we say that the vector $\ddot{\mathbf{q}}(t)$ is parallel along the trajectory. As we will show in Vol. III on gauge theory, the parallel transport of physical quantities is crucial for modern mathematics and physics. This corresponds to the transport of physical information.

6.11.3 The Hamiltonian Approach

Introduce the generalized momenta

$$p_\varphi := L_{\dot\varphi} = mR^2\dot\varphi\cos^2\vartheta, \qquad p_\vartheta := L_{\dot\vartheta} = mR^2\dot\vartheta,$$

and the Hamiltonian, $H := p_\varphi\dot\varphi + p_\vartheta\dot\vartheta - L$. That is,

$$H(\varphi, p_\varphi, \vartheta, p_\vartheta) = \frac{p_\varphi^2}{2mR^2\cos^2\vartheta} + \frac{p_\vartheta^2}{2mR^2} + mgR\sin\vartheta.$$

Then the canonical equations read as

$$\dot p_\varphi = -H_\varphi = 0, \quad \dot\varphi = H_{p_\varphi}, \quad \dot p_\vartheta = -H_\vartheta, \quad \dot\vartheta = H_{p_\vartheta}.$$

In particular, p_φ (z-component of the angular momentum) and H (energy) are conserved. Set

$$z_0 := R\sin\vartheta(0), \qquad p_\varphi := mR^2\dot\varphi(0)\cos^2\vartheta(0),$$
$$E := \frac{mR^2}{2}\big(\dot\varphi(0)^2\cos^2\vartheta(0) + \dot\vartheta(0)^2 + mgR\sin\vartheta(0)\big),$$

and $V(z) := 2\left(\frac{E}{m} - gz\right)(R^2 - z^2) - \left(\frac{p_\varphi}{m}\right)^2$.

Proposition 6.15 *For given data $\varphi(0), \dot\varphi(0), \vartheta(0), \dot\vartheta(0)$ with $\vartheta(0) \in]-\frac{\pi}{2}, \frac{\pi}{2}[$, there exists a unique motion $\varphi = \varphi(t), \vartheta = \vartheta(t)$ of the spherical pendulum given by the following elliptic integrals:*

$$t(z) = R\int_{z_0}^{z} \frac{d\zeta}{\sqrt{V(\zeta)}}, \qquad \varphi(z) = \varphi(0) + \frac{p_\varphi R}{m}\int_{z_0}^{z} \frac{d\zeta}{(R^2-\zeta^2)\sqrt{V(\zeta)}}.$$

Moreover, $z(t) = R\sin\vartheta(t)$.

Proof. We have the two conservation laws

$$mR^2\dot\varphi(t)\cos^2\vartheta(t) = p_\varphi, \qquad \frac{p_\varphi^2}{2mR^2\cos^2\vartheta(t)} + \frac{mR^2\dot\vartheta(t)^2}{2} + mgR\sin\vartheta(t) = E,$$

where p_φ and E (energy) are constants. The substitution $z = R\sin\vartheta$ yields

$$\dot\varphi = \frac{p_\varphi}{m(R^2-z^2)}, \qquad \dot\vartheta = \frac{d\vartheta}{dz}\cdot\frac{dz}{dt} = \frac{1}{R\cos\vartheta}\,\frac{dz}{dt} = \frac{1}{\sqrt{R^2-z^2}}\,\frac{dz}{dt}.$$

Therefore, by the energy conservation law, we obtain

$$\frac{mR^2}{R^2-z^2}\left(\frac{dz}{dt}\right)^2 + \frac{p_\varphi^2}{m(R^2-z^2)} + 2mgz = 2E.$$

Hence

$$R^2\left(\frac{dz}{dt}\right)^2 = V(z).$$

This implies

$$\frac{dt}{dz} = \frac{R}{\sqrt{V(z)}}.$$

Moreover,

$$\frac{d\varphi}{dt} = \frac{d\varphi}{dz}\cdot\frac{dz}{dt} = \frac{d\varphi}{dz}\cdot\frac{1}{R}\sqrt{V(z)} = \frac{p_\varphi}{m(R^2-z^2)}.$$

Finally, integrating this over z, we obtain $t = t(z)$ and $\varphi = \varphi(z)$. □

6.11.4 Geodesics of Shortest Length

Arc length. Let $\mathbf{q} = \mathbf{q}(t),\ t_0 \le t \le t_1$, be a curve on the sphere $\mathbb{S}^2_R$. By definition, the arc length of this curve between the points $\mathbf{q}_0$ and $\mathbf{q}(t)$ is given by

$$s(t) := \int_{t_0}^{t} |\dot{\mathbf{q}}(\tau)|\, d\tau = \int_{t_0}^{t} R\sqrt{\dot\varphi(\tau)^2 \cos^2\vartheta(\tau) + \dot\vartheta(\tau)^2}\, d\tau.$$

Differentiating this with respect to time t, we get

$$\dot s(t) = R\sqrt{\dot\varphi(t)^2 \cos^2\vartheta(t) + \dot\vartheta(t)^2}.$$

Hence

$$\left(\frac{ds(t)}{dt}\right)^2 = R^2\cos^2\vartheta(t)\left(\frac{d\varphi(t)}{dt}\right)^2 + R^2\left(\frac{d\vartheta(t)}{dt}\right)^2.$$

Mnemonically, we write

$$\boxed{ds^2 = R^2\cos^2\vartheta \cdot d\varphi^2 + R^2 d\vartheta^2.} \tag{6.86}$$

Curves of shortest length. Now we are looking for a smooth curve

$$\mathbf{q} = \mathbf{q}(t), \quad t_0 \le t \le t_1$$

on the sphere $\mathbb{S}^2_R$ which connects the two point $\mathbf{q}_0$ and $\mathbf{q}_1$, and which has minimal length. This is the optimal route for an aircraft which is flying from the city $\mathbf{q}_0$ to the city $\mathbf{q}_1$. We have to solve the variational problem

$$\int_{t_0}^{t_1} R\sqrt{\dot\varphi(t)^2 \cos^2\vartheta(t) + \dot\vartheta(t)^2}\, dt = \min! \tag{6.87}$$

with the side condition $(\varphi(t_0), \vartheta(t_0)) = (\varphi_0, \vartheta_0)$ and $(\varphi(t_1), \vartheta(t_1)) = (\varphi_1, \vartheta_1)$.

Theorem 6.16 *If the arc length s is chosen as parameter, then every solution of the variational problem (6.87) satisfies the following system of equations:*

$$\ddot\varphi(s) - 2\dot\varphi(s)\dot\vartheta(s)\sin 2\vartheta(s) = 0, \qquad \ddot\vartheta(s) = 0. \tag{6.88}$$

Proof. The variational problem (6.87) can be written as

$$\int_{t_0}^{t_1} \mathcal{L}\, dt = \min!$$

with $\mathcal{L} := \sqrt{L}$ where L is the Lagrangian of the spherical pendulum (6.83) with mass $m = 2$ and vanishing gravitational force (i.e., $g = 0$). The Euler–Lagrange equations

$$\frac{d}{ds}\mathcal{L}_{\dot\varphi} = \mathcal{L}_\varphi = 0, \qquad \frac{d}{ds}\mathcal{L}_{\dot\vartheta} = \mathcal{L}_\vartheta = 0$$

read as $\frac{d}{ds}\left(\frac{L_{\dot\varphi}}{2\sqrt{L}}\right) = 0$ and $\frac{d}{ds}\left(\frac{L_{\dot\vartheta}}{2\sqrt{L}}\right) = 0$. The variational integral is invariant under reparametrizations. Therefore, we can choose the arc length as parameter (i.e., $t = s$). Then $L = 1$. Hence

$$\frac{d}{ds}L_{\dot\varphi} = 0, \qquad \frac{d}{ds}L_{\dot\vartheta} = 0.$$

By (6.84), we obtain the claim (6.88). □

Suppose we are given two points $\mathbf{q}_0$ and $\mathbf{q}_1$ on the sphere $\mathbb{S}^2_R$. After a rotation, if necessary, we always can assume that the two points lie on the equator of the sphere. Then there are two arcs of the equator which connect the two given points. The smaller arc is a geodesic of shortest length between $\mathbf{q}_0$ and $\mathbf{q}_1$. Observe that two points do not always uniquely determine the connecting geodesic. For example, if the given points $\mathbf{q}_0$ and $\mathbf{q}_1$ represent the North Pole and the South Pole, respectively, then each meridian is a connecting geodesic (see Fig. 5.9 on page 277).

6.12 The Lie Group $SU(E^3)$ of Rotations

Invariance under rotations leads to conservation of angular momentum.
Folklore

For the motion $\mathbf{q} = \mathbf{q}(t)$ of a particle in the 3-dimensional space, the time-dependent vector $\mathbf{p}(t) := m\dot{\mathbf{q}}(t)$ (mass times velocity) is called the momentum of the particle at time t. The angular momentum at time t is defined by

$$\boxed{\mathbf{a}(t) := \mathbf{q}(t) \times \mathbf{p}(t).} \tag{6.89}$$

In this section, we study the motion $\mathbf{q} = \mathbf{q}(t)$ of a particle given by the Newtonian equation

$$m\ddot{\mathbf{q}}(t) = -U'(\mathbf{q}(t)), \qquad t \in \mathbb{R} \tag{6.90}$$

under the assumption that the potential $U = U(\mathbf{q})$ only depends on the distance $|\mathbf{q}|$. In other words, the potential U is invariant under rotations about the origin.

6.12.1 Conservation of Angular Momentum

Proposition 6.17 *The angular momentum is conserved for the motion* (6.90).

Proof. The trick is that $\mathbf{b} \times \mathbf{b} = 0$ holds for all vectors $\mathbf{b}$. Let us use Cartesian (x, y, z)-coordinates. Set $r := \sqrt{x^2 + y^2 + z^2}$. Then $U(\mathbf{q}) = V(r)$. It follows from

$$\frac{\partial}{\partial x} V(r) = V'(r) \frac{\partial r}{\partial x} = V'(r) \frac{x}{r}$$

and analogous formulas for y and z that

$$U'(\mathbf{q}) = V'(r) \frac{\mathbf{q}}{r}.$$

Let $\mathbf{q} = \mathbf{q}(t)$ be a solution of (6.90). By the product rule,

$$\frac{d}{dt}\mathbf{a}(t) = \dot{\mathbf{q}}(t) \times m\dot{\mathbf{q}}(t) + \mathbf{q}(t) \times m\ddot{\mathbf{q}}(t) = -\mathbf{q}(t) \times V'(r(t)) \frac{\mathbf{q}(t)}{r(t)} = 0.$$

□

Next we want to discuss the relation between the symmetry group of rotations and angular momentum.

The Lie group $SU(E^3)$ of rotations. In the 3-dimensional space, all the rotations about the origin form a Lie group denoted by $SU(E^3)$. Each such rotation $\mathbf{q} \mapsto \mathbf{q}^+$ can be represented by the Euler formula

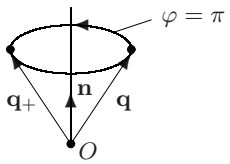


Fig. 6.17. Rotation

$$\mathbf{q}^+ = \mathbf{q}\cos\varphi + (\mathbf{n}\times\mathbf{q})\sin\varphi + (\mathbf{qn})\mathbf{n}(1-\cos\varphi). \tag{6.91}$$

This formula describes the counterclockwise rotation of the position vector $\mathbf{q}$ at the origin about the axis unit vector $\mathbf{n}$ with the rotation angle φ (Fig. 6.17). We write $\mathbf{q}^+ = R_{\mathbf{n},\varphi}\mathbf{q}$.

The Lie algebra $su(3)$ of infinitesimal rotations. Linearization of (6.91) with respect to the angle φ yields

$$\mathbf{q}^+ = \mathbf{q} + (\varphi\mathbf{n}\times\mathbf{q}). \tag{6.92}$$

We call the transformation $\mathcal{R}_\mathbf{b}$ given by $\mathcal{R}_\mathbf{b}\mathbf{q} := \mathbf{b}\times\mathbf{q}$ an infinitesimal rotation. Using the the usual linear combination of maps and the Lie product

$$[\mathcal{R}_\mathbf{a}, \mathcal{R}_\mathbf{b}]_- := \mathcal{R}_\mathbf{a}\mathcal{R}_\mathbf{b} - \mathcal{R}_\mathbf{b}\mathcal{R}_\mathbf{a},$$

the following hold:

The set of infinitesimal rotations forms a real 3-dimensional Lie algebra denoted by $su(E^3)$.

To prove this, we have to show that $\mathcal{R}_\mathbf{a}, \mathcal{R}_\mathbf{b} \in su(E^3)$ implies

$$[\mathcal{R}_\mathbf{a}, \mathcal{R}_\mathbf{b}]_- \in su(E^3).$$

This follows from the general theory of Lie groups. However, in order to display the geometric meaning behind the Lie algebra property of $su(E^3)$, let us use the following argument based on the well-known geometric properties of the vector product. In fact, it follows from the cyclic Jacobi identity for the vector product

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) + \mathbf{b}\times(\mathbf{c}\times\mathbf{a}) + \mathbf{c}\times(\mathbf{a}\times\mathbf{b}) = 0$$

and from $\mathbf{a}\times\mathbf{b} = -\mathbf{b}\times\mathbf{a}$ that

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) - \mathbf{b}\times(\mathbf{a}\times\mathbf{c}) = (\mathbf{a}\times\mathbf{b})\times\mathbf{c}.$$

Hence

$$\boxed{\mathcal{R}_\mathbf{a}\mathcal{R}_\mathbf{b}\mathbf{c} - \mathcal{R}_\mathbf{b}\mathcal{R}_\mathbf{a}\mathbf{c} = \mathcal{R}_{\mathbf{a}\times\mathbf{b}}\mathbf{c}.} \tag{6.93}$$

In addition, we have the exponential formula $R_{\mathbf{n},\varphi} = \mathrm{e}^{\varphi\mathcal{R}_\mathbf{n}}$.

Conservation of angular momentum and the Noether theorem. We want to show that the conservation of angular momentum is a consequence of the Noether theorem with respect to the rotation group. To this end, suppose that $U(\mathbf{q}) = V(|\mathbf{q}|)$. Then the Lagrangian

$$L(\mathbf{q},\dot{\mathbf{q}}) = \tfrac{1}{2}m\dot{\mathbf{q}}^2 - U(\mathbf{q}) \tag{6.94}$$

is invariant under rotations, and hence the action

$$S := \int_{t_0}^{t_1} L(\mathbf{q}(t), \dot{\mathbf{q}}(t))\, dt$$

is invariant under rotations. Each rotation $\mathbf{q}^+ = R_{\mathbf{n},\varphi}\mathbf{q}$ about the origin can be represented by the Euler formula (6.91). Hence

$$\delta\mathbf{q} := \varepsilon \frac{d}{d\varepsilon} R_{\mathbf{n},\varepsilon}\mathbf{q}_{|\varepsilon=0} = \varepsilon(\mathbf{n} \times \mathbf{q}).$$

Applying the Noether theorem (6.49) on page 387 to the three components of the vector function $t \mapsto \mathbf{q}(t)$, we get

$$\boxed{\frac{d}{d\tau}(L\delta t - (\delta\mathbf{q} - \dot{\mathbf{q}}\delta t)L_{\dot{\mathbf{q}}}) = 0} \tag{6.95}$$

for each solution $t \mapsto \mathbf{q}(t)$ of (6.90). Here, $\delta t = 0$. Therefore, $\delta\mathbf{q} \cdot L_{\dot{\mathbf{q}}}$ is conserved. Consequently,

$$\delta\mathbf{q}(t) \cdot \mathbf{p}(t) = \varepsilon(\mathbf{n} \times \mathbf{q}(t))\, \mathbf{p}(t) = \varepsilon\mathbf{n}(\mathbf{q}(t) \times \mathbf{p}(t)) = \text{const}, \quad t \in \mathbb{R}$$

for all parameters $\varepsilon > 0$ and all unit vectors $\mathbf{n}$. Hence $\mathbf{q}(t) \times \mathbf{p}(t) = \text{const}$.

Conservation of angular momentum and Poisson brackets. Now we want to describe an alternative approach for proving the conservation of angular momentum. This will be based on the use of Poisson brackets. In fact, this approach will correspond to Lie's momentum map to be defined in (6.97) below.

The Lagrangian L from (6.94) yields the momentum $\mathbf{p} = L_{\dot{\mathbf{q}}} = m\dot{\mathbf{q}}$ and the Hamiltonian

$$H(\mathbf{q}, \mathbf{p}) := \mathbf{p}\dot{\mathbf{q}} - L(\mathbf{q}, \dot{\mathbf{q}}) = \frac{\mathbf{p}^2}{2m} + V(|\mathbf{q}|).$$

This implies the partial derivatives $H_{\mathbf{p}} = \mathbf{p}/m$ and $H_{\mathbf{q}} = V'(|\mathbf{q}|)\frac{\mathbf{q}}{|\mathbf{q}|}$.

Let $\mathbf{a}(\mathbf{q}, \mathbf{p}) := \mathbf{q} \times \mathbf{p}$. For fixed vector $\mathbf{n}$, set $A(\mathbf{q}, \mathbf{p}) := \mathbf{a}(\mathbf{q}, \mathbf{p})\mathbf{n}$, that is,

$$A(\mathbf{q}, \mathbf{p}) := (\mathbf{q} \times \mathbf{p})\mathbf{n}.$$

This way, the vector-valued function $\mathbf{a} = \mathbf{a}(\mathbf{q}, \mathbf{p})$ is replaced by the real-valued function $A = A(\mathbf{q}, \mathbf{p})$. To simplify notation, we do not indicate that the function A depends on the vector $\mathbf{n}$. Our goal is to show that, for the Poisson bracket, we have

$$\boxed{\{A, H\} = 0.} \tag{6.96}$$

Then the function A is conserved. Hence

$$\frac{d}{dt}(\mathbf{q}(t) \times \mathbf{p}(t))\, \mathbf{n} = 0 \qquad \text{for all} \quad t.$$

This is true for all vectors $\mathbf{n}$. Consequently, we get

$$\frac{d}{dt}(\mathbf{q}(t) \times \mathbf{p}(t)) = 0 \qquad \text{for all} \quad t.$$

This tells us that the angular momentum is conserved.

It remains to prove (6.96). Using cyclic permutation, we obtain the well-known vector identity $(\mathbf{a} \times \mathbf{b})\mathbf{c} = (\mathbf{b} \times \mathbf{c})\mathbf{a}$. Hence $A(\mathbf{q}, \mathbf{p}) = (\mathbf{p} \times \mathbf{n})\mathbf{q}$. Then

$$\frac{\partial A}{\partial \mathbf{q}}(\mathbf{q},\mathbf{p})\mathbf{h} = \frac{d}{d\sigma}A(\mathbf{q}+\sigma\mathbf{h},\mathbf{p})_{|\sigma=0} = (\mathbf{p}\times\mathbf{n})\mathbf{h}.$$

Analogously, $\frac{\partial A}{\partial \mathbf{p}}(\mathbf{q},\mathbf{p})\mathbf{h} = (\mathbf{n}\times\mathbf{q})\mathbf{h}$. Finally, note that $(\mathbf{a}\times\mathbf{b})\mathbf{c} = 0$ iff the three vectors $\mathbf{a},\mathbf{b},\mathbf{c}$ are not linearly independent. Since $H_{\mathbf{p}}$ is parallel to $\mathbf{p}$, and $H_{\mathbf{q}}$ is parallel to $\mathbf{q}$, we get

$$\{A,H\} = A_{\mathbf{q}}H_{\mathbf{p}} - A_{\mathbf{p}}H_{\mathbf{q}} = (\mathbf{p}\times\mathbf{n})H_{\mathbf{p}} - (\mathbf{n}\times\mathbf{q})H_{\mathbf{p}} = 0.$$

6.12.2 Lie's Momentum Map

For all vectors $\mathbf{q},\mathbf{p}$ and all vectors $\mathbf{n}$, define

$$\boxed{M(\mathbf{q},\mathbf{p})\mathcal{R}_{\mathbf{n}} := (\mathbf{q}\times\mathbf{p})\mathbf{n}.}$$

This is Lie's momentum map related to the Lie algebra $su(E^3)$. More precisely, this is a map of the form

$$M : (T\mathbb{E}^3)^d \to su(E^3)^d. \tag{6.97}$$

Let us briefly discuss this. If we fix the tuple $(\mathbf{q},\mathbf{p})$, then the map

$$\mathcal{R}_{\mathbf{n}} \mapsto M(\mathbf{q},\mathbf{p})\mathcal{R}_{\mathbf{n}} \tag{6.98}$$

assigns to each infinitesimal rotation $\mathcal{R}_{\mathbf{n}}$ a real number. This map is a linear functional on the Lie algebra $su(E^3)$. Hence the map (6.98) is an element of the dual space $su(E^3)^d$ to the Lie algebra $su(E^3)$. Consequently, the map

$$(\mathbf{q},\mathbf{p}) \mapsto M(\mathbf{q},\mathbf{p})$$

assigns to each point $(\mathbf{q},\mathbf{p})$ of the phase space an element of the dual space $su(E^3)^d$. Finally, the points $(\mathbf{q},\mathbf{p})$ of the phase space form the cotangent bundle $(T\mathbb{E}^3)^d$ of the Euclidean manifold $\mathbb{E}^3$. This yields the map (6.97).

6.13 Carathéodory's Royal Road to the Calculus of Variations

6.13.1 The Fundamental Equation

Field of trajectories. Fix $n = 1,2,\ldots$ Caratheodory's fundamental equation reads as

$$\boxed{\begin{aligned} S_t(q,t) &= L(q,v(q,t),t) - v(q,t)L_{\dot q}(q,v(q,t),t),\\ S_q(q,t) &= L_{\dot q}(q,v(q,t),t), \qquad (q,t)\in\mathbb{R}^{n+1}. \end{aligned}} \tag{6.99}$$

We are given the smooth Lagrangian $L : \mathbb{R}^{2n+1} \to \mathbb{R}$, where L depends on the variables $q,\dot q \in \mathbb{R}^n$ and $t \in \mathbb{R}$. We set $q = (q^1,\ldots,q^n)$, as well as $L_q = (L_{q^1},\ldots,L_{q^n})$.

(H) *We assume that all the eigenvalues of the matrix $L_{\dot q\dot q}(q,\dot q,t)$ of the second-order partial derivatives of L (with respect to $\dot q^1,\ldots,\dot q^n$) are positive for any fixed argument $(q,\dot q,t)\in\mathbb{R}^{2n+1}$.*

This means that the function $\dot{q} \mapsto L(q, \dot{q}, t)$ is strictly convex on $\mathbb{R}^n$ for any fixed $(q, t) \in \mathbb{R}^{n+1}$. We are looking for smooth functions $v : \mathbb{R}^{n+1} \to \mathbb{R}$ and

$$S : \mathbb{R}^{n+1} \to \mathbb{R},$$

which are called velocity field and action function, respectively. The smooth solutions $q = q(t), t \in \mathbb{R}$ of the so-called velocity equation

$$\dot{q}(t) = v(q(t), t), \qquad x \in \mathbb{R} \tag{6.100}$$

are called trajectories of (6.99). The set of all the solutions of (6.100) is called a field of trajectories. Note that different trajectories of the field do not intersect, since the solution of the initial-value problem for the velocity equation (6.100) is unique. This generalizes the situation pictured in Fig. 5.8 on page 273.[28] In addition, let us consider the principle of least action:

$$\int_{t_0}^{t_1} L(t, q(t), \dot{q}(t))\, dt = \min!, \quad q(t_0) = q_0, q(t_1) = q_1. \tag{6.101}$$

Theorem 6.18 *Let v, S be a smooth solution of Carathéodory's fundamental equation (6.99). Fix the points (t_0, q_0) and (t_1, q_1) in $\mathbb{R}^{n+1}$. Then the following hold:*

(i) *Let $q = q_*(t), t_0 \le t \le t_1$, be a trajectory of (6.99) which satisfies the boundary condition $q_*(t_0) = q_0$ and $q_*(t_1) = q_1$. Then q_* is a solution of the minimum problem (6.101).*

(ii) *The difference $S(q_1, t_1) - S(t_0, q_0)$ is equal to the integral*

$$\int_{t_0}^{t_1} L(q_*(t), \dot{q}_*(t), t)\, dt.$$

The proof proceeds analogously to the proof of Theorem 5.2 on page 272.

Extremals. By definition, precisely the solutions $q = q(t)$ of the Euler–Lagrange equation

$$\frac{d}{dt} L_{\dot{q}}(q(t), \dot{q}(t), t) = L_q(q(t), \dot{q}(t), t)$$

are called extremals of the Lagrangian L. In particular, the solutions of the minimum problem (6.101) are extremals. Theorem 6.18 shows that

The trajectories of (6.99) are extremals.

This statement remains true if the convexity assumption (H) above is not satisfied. Then, for the proof, one has to use the Carathéodory equation (6.99) together with the integrability condition $S_{tq} = S_{qt}$. Therefore, the field of trajectories related to (6.100) is also called a field of extremals.

Legendre transformation. We define the co-velocity (or momentum)

$$p := L_{\dot{q}}(q, \dot{q}, t).$$

Because of assumption (H) above, this equation is locally invertible, by the implicit function theorem. To simplify the formulation, we assume that the Legendre transformation $(q, \dot{q}, t) \mapsto (q, p, t)$ is a diffeomorphism from $\mathbb{R}^{2n+1}$ onto $\mathbb{R}^{2n+1}$. This yields the smooth Hamiltonian $H : \mathbb{R}^{2n+1} \to \mathbb{R}$ given by

$$H(q, p, t) := p\dot{q}(q, p, t) - L(q, \dot{q}(q, p, t), t) \qquad \text{for all} \quad (q, p, t) \in \mathbb{R}^{2n+1}.$$

The proof of the following theorem proceeds as on page 274.

[28] We assume that the solutions of the velocity equation exist for all times $t \in \mathbb{R}$. Otherwise, the following results are only valid locally.

Theorem 6.19 *If v, S is a smooth solution of Carthéodory's fundamental equation (6.99), then the action function S is a solution of the Hamilton–Jacobi equation $S_t(q,t) + H(q, S_q(q,t), t) = 0$.*

6.13.2 Lagrangian Submanifolds in Symplectic Geometry

In geometrical optics, one wants to construct wave fronts by means of families of light rays. The point is that only *special families* of light rays allow this construction. This is intimately related to the notion of *Lagrange brackets*, which were introduced by Lagrange in the 18th century in order to simplify computations in celestial mechanics in the framework of perturbation theory.[29] The point is that the Lagrange brackets of a family of light rays are constant in time along the light rays (i.e., they are first integrals of the Hamilton canonical equations). In modern symplectic geometry, the Lagrange brackets are reformulated as *Lagrangian submanifolds* of a symplectic manifold. Replacing light rays by trajectories of particles in classical mechanics, we will study

- the construction of a solution of the Hamilton–Jacobi partial differential equation
- by the help of suitable families of trajectories which satisfy the Hamilton canonical system of ordinary differential equations.

The following general approach contains geometrical optics in the 3-dimensional (x, y, z)-space as a special case. To this end, we consider the light rays

$$y = y(x), \quad z = z(x),$$

and we choose $x := t, y := q^1, z := q^2, y' := \dot{q}^1, z' := \dot{q}^2$, as well as

$$L(x, y, z, y', z') := \frac{n(x,y,z)}{c}\sqrt{1 + y'^2 + z'^2}.$$

The Lagrange brackets. Consider the smooth map

$$q = q(u), \quad p = p(u), \qquad u \in U \tag{6.102}$$

from the nonempty open subset U of $\mathbb{R}^n$ into $\mathbb{R}^{2n}$. The target space $\mathbb{R}^{2n}$ is a symplectic manifold equipped with the symplectic 2-form

$$\omega := \sum_{k=1}^{n} dq^k \wedge dp_k.$$

Since $dq^k = \sum_{r=1}^{n} q^k_{u^r} du^r$ and $dp_k = \sum_{s=1}^{n} (p_k)_{u^s} du^s$, it follows from the antisymmetry relation $du^r \wedge du^s = -du^s \wedge du^r$ that

$$\omega = \sum_{r<s} [u^r, u^s]\, du^r \wedge du^s.$$

Here, we introduce the so-called Lagrange brackets

[29] We have shown in Sect. 17.17.4 of Vol. I that the fundamental Dyson series in quantum electrodynamics is closely related to the method of the variation of the parameter. Lagrange invented this method in order to compute the secular perturbations of the orbits of planets.

$$[u^r, u^s](u) := \sum_{k=1}^{n} \frac{\partial q^k(u)}{\partial u^r}\frac{\partial p_k(u)}{\partial u^s} - \frac{\partial q^k(u)}{\partial u^s}\frac{\partial p_k(u)}{\partial u^r}.$$

Lagrangian submanifolds. Let (6.102) be an immersion. Then the image set is an n-dimensional submanifold of the symplectic manifold $\mathbb{R}^{2n}$. This submanifold is called a Lagrangian submanifold iff all the Lagrange brackets vanish identically on the parameter space U.

The general definition in symplectic geometry reads as follows:

- Fix the number $n = 1, 2, \ldots$ Let $\mathcal{M}$ be a $2n$-dimensional symplectic manifold with the symplectic 2-form ω, that is, $d\omega = 0$ on $\mathcal{M}$ and ω is non-degenerated.[30]
- An n-dimensional isotropic submanifold $\mathcal{L}$ of $\mathcal{M}$ is called a Lagrangian submanifold. That is, for any point P on the submanifold $\mathcal{L}$, we have

$$\omega_P(v, w) = 0 \qquad \text{for all} \quad v, w \in T_P\mathcal{L}.$$

The main theorem on Lagrangian manifolds in the calculus of variations. We want to construct a local solution of the Hamilton–Jacobi partial differential equation by using a suitable n-parameter family of trajectories of the Hamilton canonical system of ordinary differential equations. To this end, fix the point $(u_0, t_0) \in \mathbb{R}^{n+1}$. Consider the smooth functions

$$q = q(t, u), \quad p = p(t, u), \qquad u \in J,\ u \in U, \tag{6.103}$$

where J is an open interval, and U is an arcwise connected open subset of $\mathbb{R}^{n+1}$ with $(t_0, u_0) \in J \times U$. Suppose that the following assumptions are met:

(H1) For any fixed parameter $u \in U$, the trajectories from (6.103) are solutions of the Hamilton canonical equations $\dot{p} = -H_q, \dot{q} = H_p$ on the time interval J.

(H2) $\det q_u(u, t) \neq 0$ for all $(u, t) \in U \times J$.

(H3) $[u^j, u^k](t_0, u) = 0$ for all $u \in U$ and all $j, k = 1, \ldots, n$.

By hypothesis (H1), for any time $t \in J$, the map (6.103) is an immersion, and hence its image is a submanifold of the (q, p)-phase space $\mathbb{R}^{2n}$. Moreover, by the crucial hypothesis (H3), the image of the map (6.103) at the initial time t_0 is a Lagrangian submanifold of the phase space $\mathbb{R}^{2n}$. By (H2), the implicit function theorem tells us that the equation $q = q(t, u)$ is uniquely solvable in a sufficiently small open neighborhood of the point (t_0, u_0). This yields the smooth function $u = u(t, q)$ in a sufficiently small open neighborhood of the point (t_0, q_0) with $q_0 := q(t_0, u_0)$. Now to the point. For any fixed parameter $u \in U$, compute the curve integral

$$S_*(t_1, u_1) := \int_{(t_0, u_0)}^{(t_1, u_1)} (p\dot{q} - H(q, p, t))dt + qq_u\, du. \tag{6.104}$$

Here, we replace the symbols q and p by (6.103), and we choose a fixed smooth curve $t = t(\tau), u = u(\tau), \tau_0 \leq \tau \leq \tau_1$, which connects the points (t_0, u_0) and (t_1, u_1) in the set $J \times U$. Explicitly,

$$qq_u\, du = \sum_{k=1}^{n} qp_{u^k} du^k.$$

The proof of the following theorem will show that the integral (6.104) does not depend on the choice of the path.

[30] Explicitly, for any point $P \in \mathcal{M}$, it follows from $\omega_P(v_0, v) = 0$ for all tangent vectors $v \in T_P\mathcal{M}$ and fixed $v_0 \in T_P\mathcal{M}$ that $v_0 = 0$. Here, $T_P\mathcal{M}$ denotes the tangent space of the manifold $\mathcal{M}$ at the point P.

The integral (6.104) *is called the Poincaré–Cartan integral invariant.*

Finally, let us introduce the desired function S by setting

$$S(q,t) := S_*(t, u(t,q)).$$

Theorem 6.20 *For any fixed time $t \in J$, the image of the map* (6.103) *is a Lagrangian submanifold of the phase space $\mathbb{R}^{2n}$.*

The function $S = S(q,t)$ is a smooth solution of the Hamilton–Jacobi equation $S_t(q,t) + H(q, S_q(q,t), t) = 0$ on a sufficiently small open neighborhood of the point (q_0, t_0).

The proof of this theorem can be found in Zeidler (1986), Vol. IV, p. 85 (see the references on page 1049). The basic idea is to show that the Lagrange brackets of (6.103) are constant in time. Thus, by (H3), the Lagrange brackets vanish identically on $J \times U$. This implies that the integral (6.104) is path-independent. Finally, differentiating (6.104) with respect to both the time t_1 and the parameter u_1, we get the Hamilton–Jacobi differential equation for the function S.

6.13.3 The Initial-Value Problem for the Hamilton–Jacobi Equation

Fix the point (q_0, t_0) in $\mathbb{R}^{n+1}, n = 1, 2, \ldots$ We are given the smooth real-valued function S_0 on an open neighborhood of the point (q_0, t_0) in $\mathbb{R}^{n+1}$.

Theorem 6.21 *The initial-value problem for the Hamilton–Jacobi partial differential equation,*

$$S_t(q,t) + H(q, S_q(q,t), t) = 0, \qquad S(q, t_0) = S_0(q),$$

has a unique smooth local solution at the point (q_0, t_0).

The full proof can be found in Zeidler (1986), Vol. IV, p. 88 (see the references on page 1049). The basic idea reads as follows. Fix the parameter $u \in \mathbb{R}^n$. Consider the Hamilton canonical system

$$\dot{q}(t) = H_p(q(t), p(t), t), \qquad \dot{p}(t) = -H_q(q(t), p(t), t)$$

with the initial condition $q(t_0, u) = S_0(u)$, $p(t_0, u) = 0$. This yields the family $q = q(t,u), p = p(t,u)$. Trivially, all the Lagrange brackets $[u^j, u^k](t_0, u)$ vanish at the initial time t_0. By Theorem 6.20, we obtain the desired solution. Uniqueness follows by reduction to the unique solvability of the initial-value problem for the Hamilton canonical ordinary differential equations.

6.13.4 Solution of Carathéodory's Fundamental Equation

Consider the solution $S = S(q,t)$ of the Hamilton–Jacobi equation from Theorem 6.21.

Theorem 6.22 *There exists a smooth velocity field $v = v(q,t)$ such that S, v is a smooth local solution of Carathéodory's fundamental equation* (6.99) *at the point (q_0, t_0).*

Proof. Set $p(q,t) := S_q(q,t)$. Local inversion of the Legendre transformation $p(q,t) = L_{\dot{q}}(q, v(q,t), t)$ yields the desired function $v = v(q,t)$ (compare the proof given in Sect. 5.5.1 on page 275). □

6.14 Hints for Further Reading

As an introduction to classical mechanics, we recommend:

F. Scheck, Mechanics: From Newton's Law to Deterministic Chaos, Springer, Berlin, 2000.

V. Arnold, Mathematical Theory of Classical Mechanics, Springer, Berlin, 1978.

W. Thirring, Classical Mathematical Physics: Dynamical Systems and Fields, Springer, New York, 1997.

H. Goldstein, Classical Mechanics, Addison-Wesley, Reading Massachusetts, 2002.

Standard textbook on classical mechanics:

R. Abraham and J. Marsden, Foundations of Mechanics, Addison-Wesley, Reading, Massachusetts, 1978.

Furthermore, we recommend:

J. Marsden and T. Ratiu, Introduction to Mechanics and Symmetry, Springer, New York, 1999.

V. Arnold, Geometrical Methods in the Theory of Ordinary Differential Equations, Springer, New York, 1983.

V. Arnold et al. (Eds.), Dynamical Systems, Vols. 1–10, Encyclopedia of Mathematical Sciences, Springer, Berlin, 1987–2003.

For monographs on the fascinating field of celestial mechanics, we refer to:

C. Siegel and J. Moser, Lectures on Celestial Mechanics, Springer, Berlin, 1971.

W. Neutsch and K. Scherer, Celestial Mechanics: An Introduction to Classical and Contemporary Methods, Wissenschaftsverlag, Mannheim, 1992.

D. Boccaletti and G. Pucacco, Theory of Orbits. Vol 1: Integrable Systems and Non-Perturbative Methods, Vol. 2: Perturbative and Geometrical Methods, Springer, Berlin, 1996.

Y. Hagihara, Celestial Mechanics, Vols. 1–6, MIT Press, Cambridge, Massachusetts, and Japan Society for the Promotion of Sciences, Tokyo, 1970–1976.

Applications to the modern mathematical theory in material sciences can be found in:

S. Müller, Variational Models for Microstructure and Phase Transitions. In: F. Bethuel, S. Hildebrandt, and M. Struwe (Eds.), Calculus of Variations and Geometric Evolution Problems, Cetraro 1996, pp. 85-210, Springer, Berlin, 1999.
Internet: `http://www.mis.mpg.de/preprints/ln/lecturenote-0298`

B. Ben Belgacem, S. Conti, A. DeSimone, and S. Müller, Energy scaling of compressed elastic films – three-dimensional elasticity and reduced theories, Arch. Rat. Mech. Anal. **164** (2002), 1–37.

G. Dolzmann, Variational Methods for Crystalline Microstructure: Analysis and Computation, Springer, Berlin, 2003.

G. Friesecke, R. James, and S. Müller, A hierarchy of plate models derived from nonlinear elasticity by gamma-convergence, Arch. Rat. Mech. Anal. **180** (2006), 183–236.

F. Schuricht, A new mathematical foundation for contact interactions in continuum physics, Arch. Rat. Mech. Anal. **184** (2007), 495–551.

There are close relations between the theory of microstructures and the renormalization techniques in quantum field theory. In both cases one has to master highly singular limits with respect to the relevant scales.

Problems

Much material on the Euler–Lagrange equations for important variational problems in physics can be found in the problem section to Chap. 14 of Vol. I.

6.1 *Trajectory of a stone thrown.* Compute the trajectory of a stone of mass m which has the initial position $\mathbf{q}_0 = 0$ and the initial velocity $\mathbf{v}$ at time $t = 0$. Solution: The Newtonian equation of motion $m\ddot{\mathbf{q}}(t) = -mg\mathbf{k}$ has the solution $\mathbf{q} = \mathbf{v}t - \frac{1}{2}gt^2\mathbf{k}$. This is a parabola if $\mathbf{v}$ is not parallel to $\mathbf{k}$. If $\mathbf{v} = -v\mathbf{k}$, then $z(t) = -vt - \frac{1}{2}gt^2$.

6.2 *Further problems.* Numerous exercises with solutions can be found in the textbook by F. Scheck, Mechanics: From Newton's Law to Deterministic Chaos, Springer, Berlin, 1999, pp. 417–516.

7. Quantization of the Harmonic Oscillator – Ariadne's Thread in Quantization

> Whoever understands the quantization of the harmonic oscillator can understand everything in quantum physics.
>
> Folklore

> Almost all of physics now relies upon quantum physics. This theory was discovered around the beginning of this century. Since then, it has known a progress with no analogue in the history of science, finally reaching a status of universal applicability.
> The radical novelty of quantum mechanics almost immediately brought a conflict with the previously admitted corpus of classical physics, and this went as far as rejecting the age-old representation of physical reality by visual intuition and common sense. The abstract formalism of the theory had almost no direct counterpart in the ordinary features around us, as, for instance, nobody will ever see a wave function when looking at a car or a chair. An ever-present randomness also came to contradict classical determinism.[1]
>
> Roland Omnès, 1994

> Quantum mechanics deserves the interest of mathematicians not only because it is a very important physical theory, which governs all microphysics, that is, the physical phenomena at the microscopic scale of 10^{-10}m, but also because it turned out to be at the root of important developments of modern mathematics.[2]
>
> Franco Strocchi, 2005

In this chapter, we will study the following quantization methods:

- Heisenberg quantization (matrix mechanics; creation and annihilation operators),
- Schrödinger quantization (wave mechanics; the Schrödinger partial differential equation),
- Feynman quantization (integral representation of the wave function by means of the propagator kernel, the formal Feynman path integral, the rigorous infinite-dimensional Gaussian integral, and the rigorous Wiener path integral),
- Weyl quantization (deformation of Poisson structures),

[1] From the Preface to R. Omnès, The Interpretation of Quantum Mechanics, Princeton University Press, Princeton, New Jersey, 1994. Reprinted by permission of Princeton University Press. We recommend this monograph as an introduction to the philosophical interpretation of quantum mechanics.

[2] F. Strocchi, An Introduction to the Mathematical Structure of Quantum Mechanics: A Short Course for Mathematicians, Lecture Notes, Scuola Normale, Pisa (Italy). Reprinted by permission of World Scientific Publishing Co. Pte. Ltd. Singapore, 2005.

- Weyl quantization functor from symplectic linear spaces to C^*-algebras,
- Bargmann quantization (holomorphic quantization),
- supersymmetric quantization (fermions and bosons).

We will choose the presentation of the material in such a way that the reader is well prepared for the generalizations to quantum field theory to be considered later on.

Formally self-adjoint operators. The operator $A : D(A) \to X$ on the complex Hilbert space X is called formally self-adjoint iff the operator is linear, the domain of definition $D(A)$ is a linear dense subspace of the Hilbert space X, and we have the symmetry condition

$$\langle \chi | A\varphi \rangle = \langle A\chi | \varphi \rangle \qquad \text{for all} \quad \chi, \psi \in D(A).$$

Formally self-adjoint operators are also called symmetric operators. The following two observations are crucial for quantum mechanics:

- If the complex number λ is an eigenvalue of A, that is, there exists a nonzero element $\varphi \in D(A)$ such that $A\varphi = \lambda\varphi$, then λ is a real number. This follows from $\lambda = \langle \varphi | A\varphi \rangle = \langle A\varphi | \varphi \rangle = \lambda^\dagger$.
- If λ_1 and λ_2 are two different eigenvalues of the operator A with eigenvectors φ_1 and φ_2, then φ_1 is orthogonal to φ_2. This follows from

$$(\lambda_1 - \lambda_2)\langle \varphi_1 | \varphi_2 \rangle = \langle A\varphi_1 | \varphi_2 \rangle - \langle \varphi_1 | A\varphi_2 \rangle = 0.$$

In quantum mechanics, formally self-adjoint operators represent formal observables.

For a deeper mathematical analysis, we need self-adjoint operators, which are called observables in quantum mechanics.

Each self-adjoint operator is formally self-adjoint. But, the converse is not true. For the convenience of the reader, on page 683 we summarize basic material from functional analysis which will be frequently encountered in this chapter. This concerns the following notions: formally adjoint operator, adjoint operator, self-adjoint operator, essentially self-adjoint operator, closed operator, and the closure of a formally self-adjoint operator. The reader, who is not familiar with this material, should have a look at page 683. Observe that, as a rule, in the physics literature one does not distinguish between formally self-adjoint operators and self-adjoint operators. Peter Lax writes:[3]

> The theory of self-adjoint operators was created by John von Neumann to fashion a framework for quantum mechanics. The operators in Schrödinger's theory from 1926 that are associated with atoms and molecules are partial differential operators whose coefficients are singular at certain points; these singularities correspond to the unbounded growth of the force between two electrons that approach each other... I recall in the summer of 1951 the excitement and elation of von Neumann when he learned that Kato (born 1917) has proved the self-adjointness of the Schrödinger operator associated with the helium atom.[4]

[3] P. Lax, Functional Analysis, Wiley, New York, 2003 (reprinted with permission). This is the best modern textbook on functional analysis, written by a master of this field who works at the Courant Institute in New York City. For his fundamental contributions to the theory of partial differential equations in mathematical physics (e.g., scattering theory, solitons, and shock waves), Peter Lax (born 1926) was awarded the Abel prize in 2005.

[4] J. von Neumann, General spectral theory of Hermitean operators, Math. Ann. **102** (1929), 49–131 (in German).

> And what do the physicists think of these matters? In the 1960s Friedrichs[5] met Heisenberg and used the occasion to express to him the deep gratitude of the community of mathematicians for having created quantum mechanics, which gave birth to the beautiful theory of operators in Hilbert space. Heisenberg allowed that this was so; Friedrichs then added that the mathematicians have, in some measure, returned the favor. Heisenberg looked noncommittal, so Friedrichs pointed out that it was a mathematician, von Neumann, who clarified the difference between a self-adjoint operator and one that is merely symmetric."What's the difference," said Heisenberg.

As a rule of thumb, a formally self-adjoint (also called symmetric) differential operator can be extended to a self-adjoint operator if we add appropriate boundary conditions. The situation is not dramatic for physicists, since physics dictates the 'right' boundary conditions in regular situations. However, one has to be careful. In Problem 7.19, we will consider a formally self-adjoint differential operator which cannot be extended to a self-adjoint operator.

> *The point is that self-adjoint operators possess a spectral family which allows us to construct both the probability measure for physical observables and the functions of observables (e.g., the propagator for the quantum dynamics).*

In general terms, this is not possible for merely formally self-adjoint operators. The following proposition displays the difference between formally self-adjoint and self-adjoint operators.

Proposition 7.1 *The linear, densely defined operator $A : D(A) \to X$ on the complex Hilbert space X is self-adjoint iff it is formally self-adjoint and it always follows from*

$$\langle \psi | A\varphi \rangle = \langle \chi | \varphi \rangle$$

for fixed $\psi, \chi \in X$ and all $\varphi \in D(A)$ that $\psi \in D(A)$.

Therefore, the domain of definition $D(A)$ of the operator A plays a critical role. The proof will be given in Problem 7.7.

Unitary operators. As we will see later on, for the quantum dynamics, unitary operators play the decisive role. Recall that the operator $U : X \to X$ is called unitary iff it is linear, bijective, and it preserves the inner product, that is,

$$\langle U\chi | U\varphi \rangle = \langle \chi | \varphi \rangle \qquad \text{for all} \quad \chi, \varphi \in X.$$

This implies $||U\varphi|| = ||\varphi||$ for all $\varphi \in X$. Hence

$$||U|| := \sup_{||\varphi|| \leq 1} ||U\varphi|| = 1$$

if we exclude the trivial case $X = \{0\}$.

The shortcoming of the language of matrices noticed by von Neumann. Let $A : D(A) \to X$ and $B : D(B) \to X$ be linear, densely defined, formally

J. von Neumann, Mathematical Foundations of Quantum Mechanics (in German), Springer, Berlin, 1932. English edition: Princeton University Press, 1955.
T. Kato, Fundamental properties of the Hamiltonian operators of Schrödinger type, Trans. Amer. Math. Soc. **70** (1951), 195–211.

[5] Schrödinger (1887–1961), Heisenberg (1901–1976), Friedrichs (1902–1982), von Neumann (1903–1957), Kato (born 1917).

self-adjoint operators on the infinite-dimensional Hilbert space X. Let $\varphi_0, \varphi_1, \varphi_2, \ldots$ be a complete orthonormal system in X with $\varphi_k \in D(A)$ for all k. Set

$$a_{jk} := \langle \varphi_j | A\varphi_k \rangle \qquad j, k = 0, 1, 2, \ldots$$

The way, we assign to the operator A the infinite matrix (a_{jk}). Similarly, for the operator B, we define

$$b_{jk} := \langle \varphi_j | B\varphi_k \rangle \qquad j, k = 0, 1, 2, \ldots$$

Suppose that the operator B is a proper extension of the operator A. Then

$$a_{jk} = b_{jk} \qquad \text{for all} \quad j, k = 0, 1, 2, \ldots,$$

but $A \neq B$. Thus, the matrix (a_{jk}) does not completely reflect the properties of the operator A. In particular, the matrix (a_{jk}) does not see the crucial domain of definition $D(A)$ of the operator A. Jean Dieudonné writes:[6]

> Von Neumann took pains, in a special paper, to investigate how Hermitean (i.e., formally self-adjoint) operators might be represented by infinite matrices (to which many mathematicians and even more physicists were sentimentally attached) ... Von Neumann showed in great detail how the lack of "one-to-oneness" in the correspondence of matrices and operators led to their weirdest pathology, convincing once for all the analysts that matrices were a totally inadequate tool in spectral theory.

7.1 Complete Orthonormal Systems

> A complete orthonormal system of eigenstates of an observable (e.g., the energy operator) cannot be extended to a larger orthonormal system of eigenstates.
>
> Folklore

Basic question. Let $H : D(H) \to X$ be a formally self-adjoint operator on the infinite-dimensional separable complex Hilbert space X. Physicists have invented algebraic methods for computing eigensolutions of the form

$$H\varphi_n = E_n\varphi_n, \qquad n = 0, 1, 2, \ldots \tag{7.1}$$

The idea is to apply so-called ladder operators which are based on the use of commutation relations (related to Lie algebras or super Lie algebras). We will encounter this method several times. In terms of physics, the operator H describes the energy of the quantum system under consideration. Here, the real numbers $E_0, E_1, E_2, \ldots$ are the energy values, and $\varphi_0, \varphi_1, \varphi_2, \ldots$ are the corresponding energy eigenstates. Suppose that $\varphi_0, \varphi_1, \varphi_2, \ldots$ is an orthonormal system, that is,

$$\langle \varphi_k | \varphi_n \rangle = \delta_{kn}, \qquad k, n = 0, 1, 2, \ldots$$

There arises the following crucial question.

[6] J. Dieudonné, History of Functional Analysis, 1900–1975, North-Holland, Amsterdam, 1983 (reprinted with permission).
J. von Neumann, On the theory of unbounded matrices, J. reine und angew. Mathematik **161** (1929), 208–236 (in German).

Is the system of the computed energy eigenvalues $E_0, E_1, E_2 \ldots$ complete?

The following theorem gives us the answer in terms of analysis.

Theorem 7.2 *If the orthonormal system $\varphi_0, \varphi_1, \ldots$ is complete in the Hilbert space X, then there are no other energy eigenvalues than $E_0, E_1, E_2, \ldots$, and the system $\varphi_0, \varphi_1, \varphi_2, \ldots$ cannot be extended to a larger orthonormal system of eigenstates.*

Before giving the proof, we need some analytical tools.

Completeness. By definition, the orthonormal system $\varphi_0, \varphi_1, \varphi_2 \ldots$ is complete iff, for any $\varphi \in X$, the Fourier series

$$\boxed{\varphi = \sum_{n=0}^{\infty} \langle \varphi_n | \varphi \rangle \varphi_n}$$

is convergent in X, that is, $\lim_{N\to\infty} ||\varphi - \sum_{n=0}^{N} \langle \varphi_n|\varphi\rangle \varphi_n|| = 0$. The proof of the following proposition can be found in Zeidler (1995a), Chap. 3 (see the references on page 1049).

Proposition 7.3 *Let $\varphi_0, \varphi_1, \varphi_2 \ldots$ be an orthonormal system in the infinite-dimensional separable complex Hilbert space X. Then the following statements are equivalent.*

(i) *The system $\varphi_0, \varphi_1, \varphi_2, \ldots$ is complete.*

(ii) *For all $\varphi, \psi \in X$, we have the convergent series*

$$\langle \psi | \varphi \rangle = \sum_{n=0}^{\infty} \langle \psi | \varphi_n \rangle \langle \varphi_n | \varphi \rangle, \tag{7.2}$$

which is called the Parseval equation.

(iii) $I = \sum_{n=0}^{\infty} \varphi_n \otimes \varphi_n$ *(completeness relation).*[7]

(iv) *For all $\varphi \in X$, we have the convergent series $||\varphi||^2 = \sum_{n=0}^{\infty} |\langle \varphi_n|\varphi\rangle|^2$.*

(v) *Let $\varphi \in X$. If all the Fourier coefficients of φ vanish, that is, we have $\langle \varphi_n|\varphi\rangle = 0$ for all n, then $\varphi = 0$.*

(vi) *The linear hull of the set $\{\varphi_0, \varphi_1, \varphi_2, \ldots\}$ is dense in the Hilbert space X. Explicitly, for any $\varphi \in X$ and any number $\varepsilon > 0$, there exist complex numbers $a_0, \ldots, a_n$ such that $||\varphi - (a_1\varphi_1 + \ldots + a_n\varphi_n)|| < \varepsilon$.*

Proof of Theorem 7.2. Suppose that $H\varphi = E\varphi$ with $\varphi \neq 0$ and that the eigenvalue E is different from $E_0, E_1, E_2, \ldots$. Since the eigenvectors for different eigenvalues are orthogonal to each other, we get $\langle \varphi_n|\varphi\rangle = 0$ for all indices n. By Prop. 7.3(v), $\varphi = 0$. This is a contradiction. □

The Dirac calculus. According to Dirac, we write equation (7.1) as

$$H|E_n\rangle = E_n|E_n\rangle, \qquad n = 0, 1, 2, \ldots$$

Moreover, the completeness relation from Prop. 7.3(iii) reads as

$$\boxed{I = \sum_{n=0}^{\infty} |\varphi_n\rangle\langle\varphi_n|.} \tag{7.3}$$

[7] This means that $\varphi = \lim_{N\to\infty} \sum_{n=0}^{N} (\varphi_n \otimes \varphi_n)\varphi$ for all $\varphi \in X$. Here, we use the convention $(\varphi_n \otimes \varphi_n)\varphi := \varphi_n \langle \varphi_n|\varphi\rangle$.

Mnemonically, from (7.3) we obtain $|\varphi\rangle = \sum_{n=0}^{\infty} |\varphi_n\rangle\langle\varphi_n|\varphi\rangle$ and

$$\langle\psi|\varphi\rangle = \langle\psi| \cdot |\varphi\rangle = \langle\psi| \cdot I|\varphi\rangle = \sum_{n=0}^{\infty} \langle\psi|\varphi_n\rangle\langle\varphi_n|\varphi\rangle.$$

This coincides with the Fourier series expansion $\varphi = \sum_{n=0}^{\infty} \langle\varphi_n|\varphi\rangle\varphi_n$ and the Parseval equation (7.2).

The following investigations serve as a preparation for the quantization of the harmonic oscillator in the sections to follow.

7.2 Bosonic Creation and Annihilation Operators

> Whoever understands creation and annihilation operators can understand everything in quantum physics.
>
> Folklore

The Hilbert space $L_2(\mathbb{R})$. We consider the space $L_2(\mathbb{R})$ of complex-valued (measurable) functions $\psi : \mathbb{R} \to \mathbb{C}$ with $\int_{-\infty}^{\infty} |\psi(x)|^2 dx < \infty$. This becomes a complex Hilbert space equipped with the inner product

$$\langle\varphi|\psi\rangle := \int_{-\infty}^{\infty} \varphi(x)^\dagger \psi(x) dx \qquad \text{for all} \quad \varphi, \psi \in L_2(\mathbb{R}).$$

Moreover, $||\psi|| := \sqrt{\langle\psi|\psi\rangle}$. The precise definition of $L_2(\mathbb{R})$ can be found in Vol. I, Sect. 10.2.4. Recall that the Hilbert space $L_2(\mathbb{R})$ is infinite-dimensional and separable. For example, the complex-valued function ψ on the real line is contained in $L_2(\mathbb{R})$ if we have the growth restriction at infinity,

$$|\psi(x)| \leq \frac{\text{const}}{1 + |x|} \qquad \text{for all} \quad x \in \mathbb{R},$$

and ψ is either continuous or discontinuous in a reasonable way (e.g., ψ is continuous up to a finite or a countable subset of the real line). Furthermore, we will use the space $\mathcal{S}(\mathbb{R})$ of smooth functions $\psi : \mathbb{R} \to \mathbb{C}$ which rapidly decrease at infinity (e.g., $\psi(x) := \mathrm{e}^{-x^2}$). The space $\mathcal{S}(\mathbb{R})$ is a linear subspace of the Hilbert space $L_2(\mathbb{R})$. Moreover, $\mathcal{S}(\mathbb{R})$ is dense in $L_2(\mathbb{R})$. The precise definition of $\mathcal{S}(\mathbb{R})$ can be found in Vol. I, Sect. 2.7.4.

The operators a and $a^\dagger$. Fix the positive number x_0. Let us study the operator

$$\boxed{a := \frac{1}{\sqrt{2}} \left(\frac{x}{x_0} + x_0 \frac{d}{dx} \right).}$$

More precisely, for each function $\psi \in \mathcal{S}(\mathbb{R})$, we define

$$(a\psi)(x) := \frac{1}{\sqrt{2}} \left(\frac{x\psi(x)}{x_0} + x_0 \frac{d\psi(x)}{dx} \right) \qquad \text{for all} \quad x \in \mathbb{R}. \tag{7.4}$$

This way, we get the operator $a : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$. We also define the operator $a^\dagger : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ by setting

$$\boxed{a^\dagger := \frac{1}{\sqrt{2}} \left(\frac{x}{x_0} - x_0 \frac{d}{dx} \right).} \tag{7.5}$$

Explicitly, for each function $\psi \in \mathcal{S}(\mathbb{R})$, we set[8]

$$(a^\dagger \psi)(x) := \frac{1}{\sqrt{2}} \left(\frac{x\psi(x)}{x_0} - x_0 \frac{d\psi(x)}{dx} \right) \qquad \text{for all} \quad x \in \mathbb{R}.$$

The operators a and $a^\dagger$ have the following properties:

(i) The operator $a^\dagger : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ is the formally adjoint operator to the operator $a : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ on the Hilbert space $L_2(\mathbb{R})$.[9] This means that

$$\langle \varphi | a\psi \rangle = \langle a^\dagger \varphi | \psi \rangle \qquad \text{for all} \quad \varphi, \psi \in \mathcal{S}(\mathbb{R}).$$

(ii) We have the commutation relation

$$\boxed{[a, a^\dagger]_- = I}$$

where I denotes the identity operator on the Hilbert space $L_2(\mathbb{R})$. Recall that $[A, B]_- := AB - BA$.

(iii) Set $\varphi_0(x) := c_0 \mathrm{e}^{-x^2/2x_0^2}$ with the normalization constant $c_0 := \frac{1}{\sqrt{x_0 \sqrt{\pi}}}$. Then $a\varphi_0 = 0$.

(iv) The operator $N : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ given by $N := a^\dagger a$ is formally self-adjoint, and it has the eigensolutions

$$\boxed{N\varphi_n = n\varphi_n, \qquad n = 0, 1, 2, \ldots}$$

where we set

$$\boxed{\varphi_n := \frac{(a^\dagger)^n}{\sqrt{n!}}\, \varphi_0.} \tag{7.6}$$

(v) For $n = 0, 1, 2, \ldots$, we have

$$a^\dagger \varphi_n = \sqrt{n+1}\ \ \varphi_{n+1}, \qquad a\varphi_{n+1} = \sqrt{n+1}\ \ \varphi_n.$$

Because of these relations, the operators a and $a^\dagger$ are called ladder operators.[10]

(vi) The functions $\varphi_0, \varphi_1, \ldots$ form a complete orthonormal system of the complex Hilbert space $L_2(\mathbb{R})$. This means that

$$\langle \varphi_n | \varphi_m \rangle = \int_{-\infty}^{\infty} \varphi_n(x)^\dagger \varphi_m(x)\, dx = \delta_{nm}, \qquad n, m = 0, 1, 2, \ldots$$

[8] In applications to the harmonic oscillator later on, the quantity x has the physical dimension of length. We introduce the typical length scale x_0 in order to guarantee that the operators a and $a^\dagger$ are dimensionless.

[9] In functional analysis, one has to distinguish between the formally adjoint operator $a^\dagger : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ and the adjoint operator $a^* : D(a^*) \to L_2(\mathbb{R})$ which is an extension of $a^\dagger$, that is, $\mathcal{S}(\mathbb{R}) \subseteq D(a^*) \subseteq L_2(\mathbb{R})$ and $a^*\varphi = a^\dagger \varphi$ for all $\varphi \in \mathcal{S}(\mathbb{R})$ (see Problem 7.4).

[10] Ladder operators are frequently used in the theory of Lie algebras and in quantum physics in order to compute eigenvectors and eigenvalues. Many examples can be found in H. Green, Matrix Mechanics, Noordhoff, Groningen, 1965, and in Shi-Hai Dong, Factorization Method in Quantum Mechanics, Springer, Dordrecht, 2007 (including supersymmetry). We will encounter this several times later on.

Moreover, for each function ψ in the complex Hilbert space $L_2(\mathbb{R})$, the Fourier series

$$\psi = \sum_{n=0}^{\infty} \langle \varphi_n | \psi \rangle \varphi_n$$

is convergent in $L_2(\mathbb{R})$. Explicitly,

$$\lim_{k \to \infty} ||\psi - \sum_{n=0}^{k} \langle \varphi_n | \psi \rangle \varphi_n|| = 0.$$

Recall that $||f||^2 = \langle f | f \rangle = \int_{-\infty}^{\infty} |f(x)|^2 \, dx$.

(vii) The matrix elements a_{mn} of the operator a with respect to the basis $\varphi_0, \varphi_1, \ldots$ are defined by

$$a_{mn} := \langle \varphi_m | a \varphi_n \rangle, \qquad m, n = 0, 1, 2, \ldots$$

Explicitly, $a_{mn} = \sqrt{n} \;\; \delta_{m,n-1}$. Therefore,

$$(a_{mn}) = \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & 0 \ldots \\ 0 & 0 & \sqrt{2} & 0 & 0 \ldots \\ 0 & 0 & 0 & \sqrt{3} & 0 \ldots \\ \vdots & & & & \end{pmatrix}.$$

Similarly, we introduce the matrix elements $(a^\dagger)_{mn}$ of the operator $a^\dagger$ by setting

$$(a^\dagger)_{mn} := \langle \varphi_m | a^\dagger \varphi_n \rangle, \qquad m, n = 0, 1, 2, \ldots$$

Then $(a^\dagger)_{mn} = a^\dagger_{nm}$. Thus, the matrix to the operator $a^\dagger$ is the adjoint matrix to the matrix (a_{mn}).

Let us prove these statements. To simplify notation, we set $x_0 := 1$.

Ad (i). For all functions $\varphi, \psi \in \mathcal{S}(\mathbb{R})$, integration by parts yields

$$\int_{-\infty}^{\infty} \varphi(x)^\dagger \left(x + \frac{d}{dx} \right) \psi(x) dx = \int_{-\infty}^{\infty} \left(x - \frac{d}{dx} \right) \varphi(x)^\dagger \cdot \psi(x) dx.$$

Hence $\langle \varphi | a \psi \rangle = \langle a^\dagger \varphi | \psi \rangle$.

Ad (ii). Obviously, $2aa^\dagger \psi = (x + \frac{d}{dx})(x - \frac{d}{dx})\psi = x^2 \psi + \psi - \psi''$. Similarly,

$$2a^\dagger a \psi = \left(x - \frac{d}{dx} \right) \left(x + \frac{d}{dx} \right) \psi = x^2 \psi - \psi - \psi''.$$

Hence $(aa^\dagger - a^\dagger a)\psi = \psi$.

Ad (iii). Note that $\sqrt{2} \, a \mathrm{e}^{-x^2/2} = (x + \frac{d}{dx}) \mathrm{e}^{-x^2/2} = 0$.

Ad (iv). For all $\varphi, \psi \in \mathcal{S}(\mathbb{R})$,

$$\langle \varphi | a^\dagger a \psi \rangle = \langle a \varphi | a \psi \rangle = \langle a^\dagger a \varphi | \psi \rangle.$$

Hence $\langle \varphi | N \psi \rangle = \langle N \varphi | \psi \rangle$. Thus, the operator N is formally self-adjoint. We now proceed by induction. Obviously, $N\varphi_0 = a^\dagger (a \varphi_0) = 0$. Suppose that $N \varphi_n = n \varphi_n$. Then, by (ii),

$$N(a^\dagger \varphi_n) = a^\dagger a a^\dagger \varphi_n = a^\dagger (a^\dagger a + I) \varphi_n.$$

This implies

$$N(a^\dagger\varphi_n) = a^\dagger(N+I)\varphi_n = (n+1)a^\dagger\varphi_n.$$

Thus, $N\varphi_{n+1} = (n+1)\varphi_{n+1}$.

Ad (v). By definition of the state φ_n,

$$a^\dagger\varphi_n = \frac{(a^\dagger)^{n+1}}{\sqrt{n!}}\,\varphi_0 = \sqrt{n+1}\,\frac{(a^\dagger)^{n+1}}{\sqrt{(n+1)!}}\,\varphi_0 = \sqrt{n+1}\,\varphi_{n+1}.$$

Moreover, by (ii) and (iv),

$$\sqrt{n+1}\,a\varphi_{n+1} = aa^\dagger\varphi_n = (a^\dagger a + I)\varphi_n = (n+1)\varphi_n.$$

Ad (vi). We first show that the functions $\varphi_0, \varphi_1, ...$ form an orthonormal system. In fact, by the Gaussian integral,

$$\langle\varphi_0|\varphi_0\rangle = \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-x^2}}{\sqrt{\pi}}\,dx = 1.$$

We now proceed by induction. Suppose that $\langle\varphi_n|\varphi_n\rangle = 1$. Then

$$(n+1)\langle\varphi_{n+1}|\varphi_{n+1}\rangle = \langle a^\dagger\varphi_n|a^\dagger\varphi_n\rangle = \langle\varphi_n|aa^\dagger\varphi_n\rangle = \langle\varphi_n|(N+I)\varphi_n\rangle.$$

By (iv), this is equal to $(n+1)\langle\varphi_n|\varphi_n\rangle$. Hence $\langle\varphi_{n+1}|\varphi_{n+1}\rangle = 1$.

Since the operator N is formally self-adjoint, eigenvectors of N to different eigenvalues are orthogonal to each other. Explicitly, it follows from

$$n\langle\varphi_n|\varphi_m\rangle = \langle N\varphi_n|\varphi_m\rangle = \langle\varphi_n|N\varphi_m\rangle = m\langle\varphi_n|\varphi_m\rangle$$

that $\langle\varphi_n|\varphi_m\rangle = 0$ if $n \neq m$. Finally, we will show below that the functions $\varphi_0, \varphi_1, ...$ coincide with the Hermite functions which form a complete orthonormal system in $L_2(\mathbb{R})$.

Ad (vii). By (v),

$$\langle\varphi_m|a\varphi_n\rangle = \sqrt{n}\langle\varphi_m|\varphi_{n-1}\rangle = \sqrt{n}\,\delta_{m,n-1}.$$

Moreover, $(a^\dagger)_{mn} = \langle\varphi_m|a^\dagger\varphi_n\rangle = \langle a\varphi_m|\varphi_n\rangle = (a_{nm})^\dagger$. □

Physical interpretation. In quantum field theory, the results above allow the following physical interpretation.

- The function φ_n represents a normalized n-particle state.
- Since $N\varphi_n = n\varphi_n$ and the state φ_n consists of n particles, the operator N is called the particle number operator.
- Since $N\varphi_0 = 0$, the state φ_0 is called the (normalized) vacuum state; there are no particles in the state φ_0.
- By (v) above, the operator $a^\dagger$ sends the n-particle state φ_n to the $(n+1)$-particle state φ_{n+1}. Naturally enough, the operator $a^\dagger$ is called the particle creation operator. In particular, the n-particle state

$$\varphi_n = \frac{(a^\dagger)^n}{\sqrt{n!}}\,\varphi_0$$

is obtained from the vacuum state φ_0 by an n-fold application of the particle creation operator a.[11]

[11] For the vacuum state φ_0, physicists also use the notation $|0\rangle$.

- Similarly, by (v) above, the operator a sends the $(n+1)$-particle state φ_{n+1} to the n-particle state φ_n. Therefore, the operator a is called the particle annihilation operator.

The position operator Q and the momentum operator P. We set

$$Q := \frac{x_0}{\sqrt{2}}(a^\dagger + a), \qquad P := \frac{\mathrm{i}\hbar}{x_0\sqrt{2}}(a^\dagger - a).$$

This way, we obtain the two linear operators $Q, P : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ along with the commutation relation

$$\boxed{[Q, P]_- = \mathrm{i}\hbar I.}$$

This follows from $[a, a^\dagger]_- = I$. In fact,

$$[Q, P]_- = \tfrac{1}{2}[a^\dagger + a, \mathrm{i}\hbar(a^\dagger - a)]_-.$$

Hence $2[Q, P]_- = \mathrm{i}\hbar[a, a^\dagger]_- - \mathrm{i}\hbar[a^\dagger, a]_- = 2\mathrm{i}\hbar[a, a^\dagger]_- = 2\mathrm{i}\hbar I$. Explicitly, for all functions $\psi \in \mathcal{S}(\mathbb{R})$ and all $x \in \mathbb{R}$,

$$(Q\psi)(x) = x\psi(x), \qquad (P\psi)(x) = -\mathrm{i}\hbar\, \frac{d\psi(x)}{dx}.$$

Hence $P = -\mathrm{i}\hbar\frac{d}{dx}$. The operators Q, P are formally self-adjoint, that is,

$$\langle\varphi|Q\psi\rangle = \langle Q\varphi|\psi\rangle, \qquad \langle\varphi|P\psi\rangle = \langle P\varphi|\psi\rangle$$

for all functions $\varphi, \psi \in \mathcal{S}(\mathbb{R})$. In fact,

$$\langle\varphi|Q\psi\rangle = \int_{-\infty}^{\infty} \varphi(x)^\dagger x\psi(x)\, dx = \int_{-\infty}^{\infty} (x\varphi(x))^\dagger \psi(x)\, dx = \langle Q\varphi|\psi\rangle.$$

Furthermore, noting that $(\mathrm{i}\varphi(x))^\dagger = -\mathrm{i}\varphi(x)^\dagger$, integration by parts yields

$$\langle\varphi|P\psi\rangle = \int_{-\infty}^{\infty} \varphi(x)^\dagger(-\mathrm{i}\hbar\psi'(x))dx = \int_{-\infty}^{\infty} (-\mathrm{i}\hbar\varphi'(x))^\dagger \psi(x)\, dx = \langle P\varphi|\psi\rangle.$$

The Hermite functions. To simplify notation, we set $x_0 := 1$. We will show that the functions $\varphi_0, \varphi_1, \ldots$ introduced above coincide with the classical Hermite functions.[12] To this end, for $n = 0, 1, 2, \ldots$, we introduce the Hermite polynomials

$$H_n(x) := (-1)^n \mathrm{e}^{x^2}\, \frac{d^n \mathrm{e}^{-x^2}}{dx^n} \tag{7.7}$$

along with the Hermite functions

$$\psi_n(x) := \frac{\mathrm{e}^{-x^2/2} H_n(x)}{\sqrt{2^n n! \sqrt{\pi}}}, \qquad x \in \mathbb{R}. \tag{7.8}$$

Explicitly, $H_0(x) = 1, H_1(x) = 2x$, and $H_2(x) = 4x^2 - 2$. For $n = 0, 1, 2, \ldots$, the following hold:

[12] Hermite (1822–1901).

(a) For all complex numbers t and x,

$$\mathrm{e}^{-t^2+2xt} = \sum_{n=0}^{\infty} H_n(x)\frac{t^n}{n!}.$$

Therefore, the function $(t,x) \mapsto \mathrm{e}^{-t^2+2xt}$ is called the generating function of the Hermite polynomials.

(b) The polynomial H_n of nth degree has precisely n real zeros. These zeros are simple.

(c) First recursive formula:

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x), \qquad x \in \mathbb{R}.$$

(d) $H_{2n+1}(0) = 0$, and $H_{2n}(0) = (-1)^n \cdot 2^n \cdot 1 \cdot 3 \cdot 5 \cdots (2n-1)$.

(e) $H_n(x) = 2^n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + 1$ for all $x \in \mathbb{R}$.

(f) Second recursive formula:

$$H_n(x) = H_n(0) + 2n\int_0^x H_{n-1}(y)dy, \qquad x \in \mathbb{R}.$$

(g) The Hermite functions $\psi_0, \psi_1, \ldots$ form a complete orthonormal system in the complex Hilbert space $L_2(\mathbb{R})$.

(h) $a^\dagger \psi_n = \sqrt{n+1}\ \psi_n$ for $n = 0, 1, 2, \ldots$

(j) $\psi_n = \varphi_n$ for $n = 0, 1, 2 \ldots$

(k) $x^2\psi_n(x) - \psi_n''(x) = (2n+1)\psi_n(x)$ for all $x \in \mathbb{R}$.

Let us prove this.

Ad (a). By the Cauchy formula,

$$f^{(n)}(x) = \frac{n!}{2\pi\mathrm{i}}\int_C \frac{f(z)}{(z-x)^{n+1}}\,dz, \qquad x \in \mathbb{C}.$$

Here, we assume that the function f is holomorphic on the complex plane $\mathbb{C}$. Moreover, C is a counter-clockwise oriented circle centered at the point x. Hence

$$(-1)^n \mathrm{e}^{-x^2} H_n(x) = \frac{n!}{2\pi\mathrm{i}}\int_C \frac{\mathrm{e}^{-z^2}}{(z-x)^{n+1}}\,dz.$$

Substituting $z = t + x$,

$$H_n(x) = \frac{n!}{2\pi\mathrm{i}}\int_{C_0} \frac{\mathrm{e}^{-t^2+2tx}}{t^{n+1}}\,dt.$$

Here, the circle C_0 is centered at the origin. Using again the Cauchy formula along with Taylor expansion, we get the claim (a).

Ad (b). The proof will be given in Problem 7.26.

Ad (c). Differentiate relation (a) with respect to t, and use comparison of coefficients.

Ad (d). Use an induction argument based on (c).

Ad (e). Use the definition (7.7) of H_n along with an induction argument.

Ad (f). Differentiate relation (a) by x, and use comparison of coefficients. Then, $H_n' = 2nH_{n-1}$.

Ad (g). The proof can be found in Zeidler (1995a), p. 210 (see the references on page 1049).

Ad (h). Use the definition of ψ_n and the relation $\sqrt{2}\, a^\dagger = x - \frac{d}{dx}$.

Ad (j). Obviously, $\varphi_0 = \psi_0$. By (h), both ψ_1 and φ_1 are generated from φ_0 the same way. Hence $\varphi_1 = \psi_1$. Similarly, $\varphi_2 = \psi_2$, and so on.

Ad (k). This follows from $a^\dagger a \varphi_n = n\varphi_n$ together with $\varphi_n = \psi_n$ and

$$a^\dagger a \psi_n = \frac{1}{2}\left(x - \frac{d}{dx}\right)\left(x + \frac{d}{dx}\right)\psi_n.$$

□

The normal product. Let $n = 1, 2, \ldots$. Again choose $x_0 := 1$. Consider

$$Q^n = \frac{1}{\sqrt{2^n}}\,(a + a^\dagger)^n = \frac{1}{\sqrt{2^n}}(a + a^\dagger)\cdots(a + a^\dagger).$$

This is a polynomial with respect to a and $a^\dagger$. By definition, the normal product $: Q^n :$ is obtained from Q^n by rearranging the factors in such a way that $a^\dagger$ (resp. a) stands left (resp. right). Explicitly, by the binomial formula,

$$\boxed{: Q^n := \frac{1}{\sqrt{2^n}} \sum_{k=0}^{n} \binom{n}{k} (a^\dagger)^k a^{n-k}.}$$

We get the key relation

$$\langle \varphi_0 | : Q^n : \varphi_0 \rangle = 0, \qquad n = 1, 2, \ldots,$$

telling us that the vacuum expectation value of the normal product is equal to zero. This follows from $a\varphi_0 = 0$, which implies $\langle \varphi_0 | \ldots a\varphi_0 \rangle = 0$ together with $\langle \varphi_0 | a^\dagger \ldots \rangle = \langle a\varphi_0 | \ldots \rangle = 0$. Finally, we set $: Q^0 := I$ if $n = 0$.

For example, $Q^2 = \frac{1}{2}(a + a^\dagger)(a + a^\dagger)$ is equal to $\frac{1}{2}(a^2 + aa^\dagger + a^\dagger a + (a^\dagger)^2)$. Hence

$$: Q^2 := \tfrac{1}{2}a^2 + a^\dagger a + \tfrac{1}{2}(a^\dagger)^2.$$

This implies $: Q^2 : \psi = (x^2 - \frac{1}{2})\psi$. Hence $: Q^2 := x^2 - \frac{1}{2}$. It turns out that $Q^n = x^n + \ldots$ is a polynomial of degree n. Explicitly,

$$\boxed{: Q^n := \frac{H_n(x)}{2^n}, \qquad n = 0, 1, 2, \ldots.}$$

For the proof, we refer to Problem 7.27.

Coherent states. For each complex number α, we define

$$\boxed{\varphi_\alpha := \mathrm{e}^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}}\, \varphi_n.} \tag{7.9}$$

By the Parseval equation,

$$||\varphi_\alpha||^2 = \mathrm{e}^{-|\alpha|^2} \sum_{n=0}^{\infty} \frac{|\alpha|^{2n}}{n!} = 1 \qquad \text{for all} \quad \alpha \in \mathbb{C}.$$

Therefore, the infinite series (7.9) is convergent in the Hilbert space $L_2(\mathbb{R})$. On page 478, we will prove that

$$a\varphi_\alpha = \alpha\varphi_\alpha \qquad \text{for all} \quad \alpha \in \mathbb{C}. \tag{7.10}$$

This tells us that the so-called coherent state φ_α is an eigenstate of the annihilation operator a. There exists a continuous family $\{\varphi_\alpha\}_{\alpha\in\mathbb{C}}$ of eigenstates of the operator a. In terms of physics, the coherent state φ_α is the superposition of states $\varphi_0, \varphi_1, \varphi_2, \ldots$ with the fixed particle number $0, 1, 2, \ldots$, respectively, and it is stable under particle annihilation, by (7.10).

Coherent states are frequently used as a nice tool for studying special physical situations in quantum optics, quantum statistics, and quantum field theory (e.g., the mathematical modelling of laser beams).

A finite family of bosonic creation and annihilation operators. The normal product and the following considerations are crucial for quantum field theory. Let $n = 1, 2, ..$ On the complex Hilbert space $L_2(\mathbb{R}^n)$ equipped with the inner product[13]

$$\langle\varphi|\psi\rangle := \int_{\mathbb{R}^n} \varphi(x)^\dagger \psi(x) dx$$

for all $\varphi, \psi \in L_2(\mathbb{R}^n)$, we define the operators

$$a_j, a_j^\dagger : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n), \qquad j = 1, ..., n$$

given by

$$a_j := \frac{1}{\sqrt{2}} \left(x_j + \frac{\partial}{\partial x_j} \right), \qquad a_j^\dagger := \frac{1}{\sqrt{2}} \left(x_j - \frac{\partial}{\partial x_j} \right).$$

Explicitly, for all functions $\psi \in \mathcal{S}(\mathbb{R}^n)$,

$$(a_j\psi)(x) := \frac{1}{\sqrt{2}} \left(x_j \psi(x) + \frac{\partial \psi(x)}{\partial x_j} \right), \qquad x \in \mathbb{R}^n.$$

For all functions $\varphi, \psi \in \mathcal{S}(\mathbb{R}^n)$, we have

$$\langle\varphi|a_j\psi\rangle = \langle a_j^\dagger \varphi|\psi\rangle, \qquad j = 1, ..., n,$$

that is, the operator $a_j^\dagger$ is the formally adjoint operator to the operator a_j on $\mathcal{S}(\mathbb{R}^n)$. For $j, k = 1, ..., n$, we have the following commutation relations

$$\boxed{[a_j, a_k^\dagger]_- = \delta_{jk} I,} \tag{7.11}$$

and

$$\boxed{[a_j, a_k]_- = [a_j^\dagger, a_k^\dagger]_- = 0.} \tag{7.12}$$

A special role is played by the state

$$\varphi_0(x) := c_0 \mathrm{e}^{-x^2}, \qquad x \in \mathbb{R}^n$$

with $x^2 := x_1^2 + ... + x_n^2$ and the normalization constant $c_0 := \pi^{-n/4}$. Then

$$\langle\varphi_0|\varphi_0\rangle = \int_{\mathbb{R}^n} \frac{\mathrm{e}^{-\frac{1}{2}x_1^2 - ... - \frac{1}{2}x_n^2}}{(\sqrt{\pi})^n} \, dx_1 \cdots dx_n = \left(\int_{\mathbb{R}} \frac{\mathrm{e}^{-\frac{1}{2}y^2}}{\sqrt{\pi}} \, dy \right)^n = 1.$$

[13] The definition of the spaces $\mathcal{S}(\mathbb{R}^n)$ and $L_2(\mathbb{R}^n)$ can be found in Vol. I, Sects. 2.7.4 and 10.2.4, respectively.

The operator $N : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ given by

$$N := \sum_{j=1}^{n} a_j^\dagger a_j$$

has the eigensolutions

$$\boxed{N|k_1 k_2 \dots k_n\rangle = (k_1 + k_2 + ... + k_n)|k_1 k_2 ... k_n\rangle} \tag{7.13}$$

with $k_1, k_2, \dots, k_n = 0, 1, 2, \dots$ Here, we set

$$|k_1 k_2 \dots k_n\rangle := \frac{(a_1^\dagger)^{k_1}}{\sqrt{k_1!}} \frac{(a^\dagger)^{k_2}}{\sqrt{k_2!}} \cdots \frac{(a^\dagger)^{k_n}}{\sqrt{k_n!}} \, \varphi_0.$$

The system of states $|k_1 k_2 \dots k_n\rangle$ forms a complete orthonormal system in the complex Hilbert space $L_2(\mathbb{R}^n)$. The operator N is formally self-adjoint, that is,

$$\langle \varphi | N\psi \rangle = \langle N\varphi | \psi \rangle \qquad \text{for all} \quad \varphi, \psi \in \mathcal{S}(\mathbb{R}^n).$$

The proofs for the claims above proceed analogously as for the operators a and $a^\dagger$. We use the following terminology. There are n types of elementary particles called bosons.

- The state $|k_1 k_2 \dots k_n\rangle$ corresponds to k_1 bosons of type 1, k_2 bosons of type 2,..., and k_n bosons of type n.
- The operator $a_j^\dagger$ is called the creation operator for bosons of type j.
- The operator a_j is called the annihilation operator for bosons of type j.
- The operator N is called the particle number operator.
- Since $N\varphi_0 = 0$, the state φ_0 is called the (normalized) vacuum state. Instead of φ_0, physicists also write $|0\rangle$.

7.3 Heisenberg's Quantum Mechanics

> Quantum mechanics was born on December 14, 1900, when Max Planck delivered his famous lecture before the German Physical Society in Berlin which was printed afterwards under the title "On the law of energy distribution in the normal spectrum." In this paper, Planck assumed that the emission and absorption of radiation always takes place in discrete portions of energy or *energy quanta* $h\nu$, where ν is the frequency of the emitted or absorbed radiation. Starting with this assumption, Planck arrived at his famous formula
>
> $$\varrho = \frac{\alpha \nu^3}{\mathrm{e}^{h\nu/kT} - 1}$$
>
> for the energy density ϱ of black-body radiation at temperature T.[14]
>
> Barthel Leendert van der Waerden, 1967

[14] B. van der Waerden, Sources of Quantum Mechanics, North-Holland, Amsterdam, 1967 (reprinted with permission).

> The present paper seeks to establish a basis for theoretical quantum mechanics founded exclusively upon relationships between quantities which in principle are observable.[15]
>
> Werner Heisenberg, 1925

> The recently published theoretical approach of Heisenberg is here developed into a systematic theory of quantum mechanics with the aid of mathematical *matrix theory*. After a brief survey of the latter, the mechanical equations of motions are derived from a variational principle and it is shown that using Heisenberg's quantum condition, the principle of energy conservation and Bohr's frequency condition follow from the mechanical equations. Using the anharmonic oscillator as example, the question of uniqueness of the solution and of the significance of the phases of the partial vibrations is raised. The paper concludes with an attempt to incorporate electromagnetic field laws into the new theory.[16]
>
> Max Born and Pascal Jordan, 1925

There exist three different, but equivalent approaches to quantum mechanics, namely,

(i) Heisenberg's particle quantization from the year 1925 and its refinement by Born, Dirac, and Jordan in 1926,
(ii) Schrödinger's wave quantization from 1926, and
(iii) Feynman's statistics over classical paths via path integral from 1942.

In what follows we will thoroughly discuss these three approaches in terms of the harmonic oscillator. Let us start with (i).

The classical harmonic oscillator. Recall that the differential equation

$$\ddot{q}(t) + \omega^2 q(t) = 0, \qquad t \in \mathbb{R} \tag{7.14}$$

describes the motion $q = q(t)$ of a point of mass m on the real line which oscillates with the positive angular frequency ω. We add the initial condition $q(0) = q_0$ and $\dot{q}(0) = v_0$. Let us introduce the momentum $p := m\dot{q}$ and the Hamiltonian

$$H(q,p) := \frac{p^2}{2m} + \frac{m\omega^2 q^2}{2}$$

which represents the energy of the particle. Recall that the equation of motion (7.14) is equivalent to the canonical equations $\dot{p} = -H_q, \dot{q} = H_q$. Explicitly,

$$\dot{p}(t) = -m\omega^2 q(t), \qquad m\dot{q}(t) = p(t), \qquad t \in \mathbb{R},$$

along with the initial conditions $q(0) = q_0$ and $p(0) = p_0$. Note that $p_0 = mv_0$ where v_0 is the initial velocity of the particle. Let us introduce the typical length scale

$$x_0 := \sqrt{\frac{\hbar}{m\omega}}$$

which can be formed by using the parameters m, ω and $\hbar$. Let a be an arbitrary complex number. The general solution of (7.14) is given by

[15] W. Heisenberg, Quantum-theoretical re-interpretation of kinematic and mechanical relations, Z. Physik **33** (1925), 879–893 (in German).
[16] M. Born and P. Jordan, On Quantum Mechanics, Z. Physik **34** (1925), 858–888 (in German).

$$\boxed{q(t) = \frac{x_0}{\sqrt{2}}\,(a^\dagger \mathrm{e}^{\mathrm{i}\omega t} + a\mathrm{e}^{-\mathrm{i}\omega}), \qquad t \in \mathbb{R}.} \tag{7.15}$$

For the momentum, we get

$$\boxed{p(t) = m\dot{q}(t) = \frac{\mathrm{i}\hbar}{x_0\sqrt{2}}\,(a^\dagger \mathrm{e}^{\mathrm{i}\omega t} - a\mathrm{e}^{-\mathrm{i}\omega}), \qquad t \in \mathbb{R}.}$$

Letting $t = 0$, we obtain

$$a = \frac{1}{\sqrt{2}}\left(\frac{q(0)}{x_0} + \frac{\mathrm{i}x_0 p(0)}{\hbar}\right)$$

for the relation between the Fourier coefficient a and the real initial values $q(0)$ and $p(0)$. Hence, for the conjugate complex Fourier coefficient,

$$a^\dagger = \frac{1}{\sqrt{2}}\left(\frac{q(0)}{x_0} - \frac{\mathrm{i}x_0 p(0)}{\hbar}\right).$$

For the Hamiltonian,

$$H(q(t), p(t)) = \hbar\omega(a^\dagger a + \tfrac{1}{2}), \qquad t \in \mathbb{R}.$$

This expression does not depend on time t which reflects conservation of energy for the motion of the harmonic oscillator. Note that

$$q(t)^\dagger = q(t), \qquad p(t)^\dagger = p(t) \qquad \text{for all} \quad t \in \mathbb{R},$$

and that $a, a^\dagger$ are dimensionless. In quantum mechanics, this classical reality condition will be replaced by the formal self-adjointness of the operators $q(t)$ and $p(t)$.

The classical uncertainty relation. The motion $q = q(t)$ has the time period $T = 2\pi/\omega$. Let us now study the time means of the classical motion. For a T-periodic function $f : \mathbb{R} \to \mathbb{R}$, we define the mean value

$$\bar{f} = \frac{1}{T}\int_{-T/2}^{T/2} f(t)dt,$$

and the mean fluctuation Δf by

$$(\Delta f)^2 = \overline{(f - \bar{f})^2} = \frac{1}{T}\int_{-T/2}^{T/2} (f(t) - \bar{f})^2 dt.$$

To simplify computations, let us restrict ourselves to the special case where the initial velocity of the particle vanishes, $p_0 = 0$. Then we get the energy $E = m\omega^2 q(0)^2/2$, along with[17]

$$\bar{q} = \bar{p} = 0, \qquad \Delta p = m\omega\Delta q, \qquad \Delta q = \sqrt{\frac{E}{m\omega^2}}.$$

This implies the so-called classical uncertainty relation:

$$\boxed{\Delta q \Delta p = \frac{E}{\omega}.} \tag{7.16}$$

[17] Note that $\int_{-T/2}^{T/2} \mathrm{e}^{\mathrm{i}k\omega t}dt = \int_{-T/2}^{T/2} \mathrm{e}^{\mathrm{i}2\pi kt/T}dt = 0$ if $k = 1, 2, \ldots$

Poisson brackets. In order to quantize the classical harmonic oscillator, it is convenient to write the classical equation of motion in terms of Poisson brackets. Recall that

$$\{A(q,p), B(q,p)\} := \frac{\partial A(q,p)}{\partial q}\frac{\partial B(q,p)}{\partial p} - \frac{\partial B(q,p)}{\partial q}\frac{\partial A(q,p)}{\partial p}.$$

For example, $\{q,p\} := 1$, $\{q,H\} = H_p = p/m$, and $\{p,H\} = -H_q = -m\omega^2 q$. Thus, for all times $t \in \mathbb{R}$, the equations of motion for the harmonic oscillator read as

$$\dot{q}(t) = \{q(t), H(q(t),p(t))\}, \qquad \dot{p}(t) = \{p(t), H(q(t)), p(t)\}, \tag{7.17}$$

together with $\{q(t), p(t)\} = 1$.

7.3.1 Heisenberg's Equation of Motion

> In a recent paper, Heisenberg puts forward a new theory which suggests that it is not the equations of classical mechanics that are in any way at fault, but that the mathematical operations by which physical results are deduced from them require modification. All the information supplied by the classical theory can thus be made use of in the new theory ... We make the fundamental assumption that the difference between the Heisenberg products is equal to $\mathrm{i}\hbar$ times their Poison bracket
>
> $$xy - yx = \mathrm{i}\hbar\{x,y\}. \tag{7.18}$$
>
> It seems reasonable to take (7.18) as constituting the general quantum conditions.[18]
>
> Paul Dirac, 1925

The general quantization principle. We are looking for a simple principle which allows us to pass from classical mechanics to quantum mechanics. This principle reads as follows:

- position $q(t)$ and momentum $p(t)$ of the particle at time t become operators,
- and Poisson brackets are replaced by Lie brackets,

$$\{A(q,p), B(q,p)\} \quad \Rightarrow \quad \frac{1}{\mathrm{i}\hbar}\,[A(q,p), B(q,p)]_-.$$

Recall that $[A,B]_- := AB - BA$. Using this quantization principle, the classical equation of motion (7.17) passes over to the equation of motion for the quantum harmonic oscillator

$$\boxed{\begin{aligned} \mathrm{i}\hbar\dot{q}(t) &= [q(t), H(q(t),p(t))]_-, \\ \mathrm{i}\hbar\dot{p}(t) &= [p(t), H(q(t),p(t))]_- \end{aligned}} \tag{7.19}$$

together with

[18] P. Dirac, The fundamental equations of quantum mechanics, Proc. Royal Soc. London Ser. A **109** (1925), no. 752, 642–653.
A far-reaching generalization of Dirac's principle to the quantization of general Poisson structures was proven by Kontsevich. In 1998, he was awarded the Fields medal for this (see the papers by Kontsevich (2003) and by Cattaneo and Felder (2000) quoted on page 676).

$$\boxed{[q(t), p(t)]_- = \mathrm{i}\hbar I.} \tag{7.20}$$

The latter equation is called the Heisenberg–Born–Jordan commutation relation.

The method of Fourier quantization. In order to solve the equations of motion (7.19), (7.20), we use the classical solution formula

$$\begin{aligned} q(t) &= \frac{x_0}{\sqrt{2}}\,(a^\dagger \mathrm{e}^{\mathrm{i}\omega t} + a\mathrm{e}^{-\mathrm{i}\omega t}),\\ p(t) &= m\dot{q}(t) = \frac{\mathrm{i}\hbar}{x_0\sqrt{2}}\,(a^\dagger \mathrm{e}^{\mathrm{i}\omega t} - a\mathrm{e}^{-\mathrm{i}\omega t}) \end{aligned} \tag{7.21}$$

for all times $t \in \mathbb{R}$. But we replace the classical Fourier coefficients a and $a^\dagger$ by operators a and $a^\dagger$ which satisfy the commutation relation

$$\boxed{[a, a^\dagger]_- = I.}$$

These operators can be found in Sect. 7.2. Let us check that indeed we obtain a solution. First of all note that

$$\begin{aligned} [q(t), p(t)]_- &= \tfrac{1}{2}\mathrm{i}\hbar[a^\dagger \mathrm{e}^{\mathrm{i}\omega t} + a\mathrm{e}^{-\mathrm{i}\omega t}, a^\dagger \mathrm{e}^{\mathrm{i}\omega t} - a\mathrm{e}^{-\mathrm{i}\omega t}]_-\\ &= \tfrac{1}{2}\mathrm{i}\hbar([a, a^\dagger]_- - [a^\dagger, a]_-) = \mathrm{i}\hbar[a, a^\dagger]_- = \mathrm{i}\hbar I. \end{aligned}$$

As in the classical case, one checks easily that

$$m\dot{q}(t) = p(t), \qquad \dot{p}(t) = -m\omega^2 q(t).$$

Moreover, it follows from $[q, p]_- = \mathrm{i}\hbar$ that

$$[q, p^2]_- = ([q, p]_-)p + p[q, p]_- = 2\mathrm{i}\hbar p.$$

Similarly, for $n = 1, 2, ...,$

$$[q, p^n]_- = \mathrm{i}\hbar n p^{n-1}, \qquad [p, q^n]_- = -\mathrm{i}\hbar n q^{n-1},$$

by induction. Hence

$$2m[q(t), H(q(t), p(t))]_- = [q(t), p(t)^2]_- = 2\mathrm{i}\hbar p(t) = 2m\mathrm{i}\hbar\dot{q}(t).$$

This is the first equation of motion. Similarly, we get the second equation of motion

$$[p(t), H(q(t), p(t))]_- = \tfrac{1}{2}[p(t), m\omega^2 q^2(t)]_- = -\mathrm{i}\hbar m\omega^2 q(t) = \mathrm{i}\hbar\dot{p}(t).$$

For the Hamiltonian, it follows from $[a, a^\dagger]_- = I$ that

$$H(q(t), p(t)) = \hbar\omega(a^\dagger a + \tfrac{1}{2}). \tag{7.22}$$

Matrix elements. Let us use the results from Sect. 7.2. Recall that the states

$$\varphi_n := \frac{(a^\dagger)^n}{\sqrt{n!}}\,\varphi_0, \qquad n = 0, 1, 2, ...$$

form a complete orthonormal system of the complex Hilbert space $L_2(\mathbb{R})$. In addition, $\varphi_n \in \mathcal{S}(\mathbb{R})$ for all n. For the physical interpretation of Heisenberg's quantum

mechanics, infinite-dimensional matrices play a crucial role. Let us discuss this. We assign to each linear operator $A : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ the matrix elements

$$A_{mn} := \langle \varphi_m | A\varphi_n \rangle, \qquad m, n = 0, 1, 2, \dots$$

For two linear formally self-adjoint operators $A, B : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$, we get the product rule

$$\boxed{(AB)_{mn} = \sum_{k=0}^{\infty} A_{mk} B_{kn}, \qquad m, n = 0, 1, 2, \dots} \tag{7.23}$$

In fact, by the Parseval equation (7.2), this follows from

$$\langle \varphi_m | AB\varphi_n \rangle = \langle A\varphi_m | B\varphi_n \rangle = \sum_{k=0}^{\infty} \langle A\varphi_m | \varphi_k \rangle \langle \varphi_k | B\varphi_n \rangle$$

along with $\langle A\varphi_m | \varphi_k \rangle = \langle \varphi_m | A\varphi_k \rangle$.

Examples. Let us now compute the matrix elements of $H, q(t)$, and $p(t)$. It follows from $N\varphi_n = n\varphi_n$ that

$$H\varphi_n = \hbar\omega(N + \tfrac{1}{2}I)\varphi_n = \hbar\omega(n + \tfrac{1}{2})\varphi_n.$$

Hence $H_{mn} = \langle \varphi_m | H\varphi_n \rangle = E_n \langle \varphi_m | \varphi_n \rangle = E_n \delta_{nm}$ with $E_n = \hbar\omega(n + \frac{1}{2})$. This yields the diagonal matrix

$$(H_{mn}) = \begin{pmatrix} E_0 & 0 & 0 & 0 & \dots \\ 0 & E_1 & 0 & 0 & \dots \\ \vdots & & & & \end{pmatrix}.$$

It follows from Sect. 7.2 that $a_{kn} = \sqrt{n}\, \delta_{k,n-1}$. Thus, by (7.21),

$$q_{kn}(t) = \frac{x_0}{\sqrt{2}} \left(a_{nk} \mathrm{e}^{\mathrm{i}\omega t} + a_{kn} \mathrm{e}^{-\mathrm{i}\omega t} \right). \tag{7.24}$$

This way, we get the self-adjoint matrix

$$(q_{kn}(t)) = \frac{x_0}{\sqrt{2}} \begin{pmatrix} 0 & \sqrt{1}\, \mathrm{e}^{-\mathrm{i}\omega t} & 0 & 0 & \dots \\ \sqrt{1}\, \mathrm{e}^{\mathrm{i}\omega t} & 0 & \sqrt{2}\, \mathrm{e}^{-\mathrm{i}\omega t} & 0 & \dots \\ 0 & \sqrt{2}\, \mathrm{e}^{\mathrm{i}\omega t} & 0 & 0 & \dots \\ \vdots & & & & \end{pmatrix}$$

for all times $t \in \mathbb{R}$. Similarly,

$$p_{kn}(t) = m\dot{q}_{kn}(t), \qquad k, n = 0, 1, 2, \dots$$

By the product rule (7.23), for the square of the position matrix (q_{kn}) we get

$$(q_{kn})^2 = \frac{x_0^2}{2} \begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ 0 & 3 & 0 & 0 & \dots \\ 0 & 0 & 5 & 0 & \dots \\ \vdots & & & & \end{pmatrix}. \tag{7.25}$$

Similarly,

$$(p_{kn})^2 = \frac{\hbar^2}{2x_0^2} \begin{pmatrix} 1\ 0\ 0\ 0\ ... \\ 0\ 3\ 0\ 0\ ... \\ 0\ 0\ 5\ 0\ ... \\ \vdots \end{pmatrix}.$$

7.3.2 Heisenberg's Uncertainty Inequality for the Harmonic Oscillator

In order to discuss the physical meaning of the matrices introduced above, we will use the following terminology:

- The elements ψ of the complex Hilbert space $L_2(\mathbb{R})$ normalized by the condition $\langle\psi|\psi\rangle = 1$ are called normalized states of the quantum harmonic oscillator,
- whereas the linear, formally self-adjoint operators $A : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ are called formal observables.

Two normalized states ψ and φ are called equivalent iff

$$\varphi = \mathrm{e}^{\mathrm{i}\alpha}\psi$$

for some real number α. We say that φ and ψ differ by phase. Consider some normalized state ψ and some formal observable A. The number

$$\bar{A} := \langle\psi|A\psi\rangle$$

is interpreted as the mean value of the observable A measured in the state ψ.[19] Moreover, the nonnegative number ΔA given by

$$(\Delta A)^2 := \langle\psi|(A - \bar{A})^2\psi\rangle$$

is interpreted as the fluctuation of the measured mean value $\bar{A}$. Let us choose $n = 0, 1, 2, \ldots$ For the state φ_n of the quantum harmonic oscillator, we get the following measured values for all times $t \in \mathbb{R}$.

(i) Energy: $\bar{E} = E_n = \hbar\omega(n + \frac{1}{2})$ and $\Delta E = 0$.

(ii) Position: $\bar{q}(t) = q_{nn}(t) = 0$ and $\Delta q(t) = x_0\sqrt{n + \frac{1}{2}}$.

(iii) Momentum: $\bar{p}(t) = p_{nn}(t) = 0$ and $\Delta p(t) = \frac{\hbar}{x_0}\sqrt{n + \frac{1}{2}}$.

(iv) Heisenberg's uncertainty inequality:

$$\boxed{\Delta q(t)\Delta p(t) \geq \frac{\hbar}{2}.}$$

Let us prove this.

Ad (i). For the energy, it follows from the eigensolution $H\varphi_n = E_n\varphi_n$ that

$$\bar{E} = \langle\varphi_n|H\varphi_n\rangle = E_n\langle\varphi_n|\varphi_n\rangle = E_n,$$

and $\Delta E = ||(H - E_n I)\varphi_n|| = 0$.

[19] Since the operator A is formally self-adjoint, the number $\bar{A}$ is real. Furthermore, note that $\langle\psi|(A - \bar{A})^2\psi\rangle = \langle(A - \bar{A})\psi|(A - \bar{A})\psi\rangle = ||(A - \bar{A}I)\psi||^2 \geq 0$.

Ad (ii). Note that

$$(\Delta q)^2 = \langle \varphi_n | q(t)^2 \varphi_n \rangle.$$

Therefore, $(\Delta q)^2$ is the nth diagonal element of the product matrix $(q_{kn})^2$ which can be found in (7.25). Analogously, we get (iii). The uncertainty inequality is an immediate consequence of (ii) and (iii). □

The famous Heisenberg uncertainty inequality for the quantum harmonic oscillator tells us that the state φ_n has the sharp energy E_n, but it is impossible to measure sharply both position and momentum of the quantum particle at the same time. Thus, there exists a substantial difference between classical particles and quantum particles.

It is impossible to speak of the trajectory of a quantum particle.

7.3.3 Quantization of Energy

> I have the best of reasons for being an admirer of Werner Heisenberg. He and I were young research students at the same time, about the same age, working on the same problem. Heisenberg succeeded where I failed... Heisenberg - a graduate student of Sommerfeld - was working from the experimental basis, using the results of spectroscopy, which by 1925 had accumulated an enormous amount of data[20]...
>
> Paul Dirac, 1968

The measured spectrum of an atom or a molecule is characterized by two quantities, namely,

- the wave length λ_{nm} of the emitted spectral lines (where $n, m = 0, 1, 2, \ldots$ with $n > m$), and
- the intensity of the spectral lines.

In Bohr's and Sommerfeld's semi-classical approach to the spectra of atoms and molecules from the years 1913 and 1916, respectively, the spectral lines correspond to photons which are emitted by jumps of an electron from one orbit of the atom or molecule to another orbit. If $E_0 < E_1 < E_2 < \ldots$ are the (discrete) energies of the electron corresponding to the different orbits, then a jump of the electron from the higher energy level E_n to the lower energy level E_m produces the emission of one photon of energy $E_n - E_m$. According to Einstein's light quanta hypothesis from 1905, this yields the frequency

$$\nu_{nm} = \frac{E_n - E_m}{h}, \qquad n > m \tag{7.26}$$

of the emitted photon, and hence the wave length $\lambda_{nm} = c/\nu_{nm}$ of the corresponding spectral line is obtained. The intensity of the spectral lines depends on the transition probabilities for the jumps of the electrons. In 1925 it was Heisenberg's philosophy to base his new quantum mechanics only on quantities which can be measured in physical experiments, namely,

- the energies $E_0, E_1, \ldots$ of bound states and

[20] In: A. Salam (Ed.), From a Life of Physics. Evening Lectures at the International Center for Theoretical Physics, Trieste (Italy), with outstanding contributions by Abdus Salam, Hans Bethe, Paul Dirac, Werner Heisenberg, Eugene Wigner, Oscar Klein, and Eugen Lifshitz, International Atomic Energy Agency, Vienna, Austria, 1968.
A. Sommerfeld, Atomic Structure and Spectral Lines, Methuen, London, 1923.

- the transition probabilities for changing bound states.[21]

Explicitly, Heisenberg replaced the trajectory $q = q(t), t \in \mathbb{R}$ of a particle in classical mechanics by the following family $(q_{nm}(t))$ of functions

$$q_{nm}(t) = q_{nm}(0)\mathrm{e}^{\mathrm{i}\omega_{nm}t}, \qquad n, m = 0, 1, 2, \ldots$$

where $\omega_{nm} = 2\pi\nu_{nm}$, and the frequencies ν_{nm} are given by (7.26). It follows from (7.26) that

$$\nu_{nk} + \nu_{km} = \nu_{nm}, \qquad n < k < m.$$

In physics, this is called the Ritz combination principle for frequencies.[22] In terms of mathematics, this tells us that the family $\{\nu_{nm}\}$ of frequencies represents a cocycle generated by the family $\{E_n\}$ of energies. Thus, this approach is based on a simple variant of cohomology.[23] In order to compute the intensities of spectral lines, Heisenberg was looking for a suitable quadratic expression in the amplitudes $q_{nm}(0)$. Using physical arguments and analogies with the product formula for Fourier series expansions, Heisenberg invented the composition rule

$$(q^2(0))_{nm} := \sum_{k=0}^{\infty} q_{nk}(0)q_{km}(0) \tag{7.27}$$

for defining the square $(q_{nm}(0))^2$ of the scheme $(q_{nm}(0))$. Applying this to the harmonic oscillator (and the anharmonic oscillator as a perturbed harmonic oscillator), Heisenberg obtained the energies

$$E_n = \omega\hbar(n + \tfrac{1}{2}), \qquad n = 0, 1, 2, \ldots$$

for the quantized harmonic oscillator.

After getting Heisenberg's manuscript, Born (1882–1970) noticed that the composition rule (7.27) resembled the product for matrices $q(t) = (q_{nm}(t))$, which he learned as a student in the mathematics course. He guessed the validity of the rule

$$qp - pq = \mathrm{i}\hbar. \tag{7.28}$$

But he was only able to verify this for the diagonal elements. After a few days of joint work with his pupil Pascal Jordan (1902–1980), Born finished a joint paper with Jordan on the new quantum mechanics including the commutation rule (7.28); nowadays this is called the Heisenberg–Born–Jordan commutation rule (or briefly the Heisenberg commutation rule). At that time, Heisenberg was not in Göttingen, but on the island Helgoland (North Sea) in order to cure a severe attack of hay fever. After coming back to Göttingen, Heisenberg wrote together with Born and Jordan a fundamental paper on the principles of quantum mechanics. The English translation of the following three papers can be found in van der Waerden (1968):

[21] Heisenberg's thinking was strongly influenced by the Greek philosopher Plato (428–347 B.C.). Nowadays one uses the Latin version 'Plato'. The correct Greek name is 'Platon'. Plato's Academy in Athens had unparalleled importance for Greek thought. The greatest philosophers, mathematicians, and astronomers worked there. For example, Aristotle (384–322 B.C.) studied there. In 529 A.D., the Academy was closed by the Roman emperor Justitian.

[22] Ritz (1878–1909) worked in Göttingen.

[23] The importance of cohomology for classical and quantum physics will be studied in Vol. IV on quantum mathematics.

W. Heisenberg, Quantum-theoretical re-interpretation of kinematics and mechanical relations), Z. Physik **33** (1925), 879–893.

M. Born, P. Jordan, On quantum mechanics, Z. Physik **35** (1925), 858–888.

M. Born, W. Heisenberg, and P. Jordan, On quantum mechanics II, Z. Physik **36** (1926), 557–523.

At the same time, Dirac formulated his general approach to quantum mechanics:

P. Dirac, The fundamental equations of quantum mechanics, Proc. Royal Soc. London Ser. A **109** (1926), no. 752, 642–653.

Heisenberg, himself, pointed out the following at the *Trieste Evening Lectures* in 1968:

> It turned out that one could replace the quantum conditions of Bohr's theory by a formula which was essentially equivalent to the sum-rule in spectroscopy by Thomas and Kuhn... I was however not able to get a neat mathematical scheme out of it. Very soon afterwards both Born and Jordan in Göttingen and Dirac in Cambridge were able to invent a perfectly closed mathematical scheme: Dirac with very ingenious new methods on abstract noncommutative q-numbers (i.e., quantum-theoretical numbers), and Born and Jordan with more conventional methods of matrices.

7.3.4 The Transition Probabilities

Let us discuss the meaning of the entries q_{kn} of the position matrix on page 445. Suppose that the quantum particle is an electron of electric charge $-e$ and mass m. Let ε_0 and c be the electric field constant and the velocity of light of a vacuum, respectively. Furthermore, let h be the Planck action quantum, and set $\hbar := h/2\pi$.[24] According to Heisenberg, the real number

$$\boxed{\gamma_{kn} := \frac{\omega_{kn}^3 e^2 (t_2 - t_1)}{3\pi\varepsilon_0 \hbar c^3} \, |q_{kn}(0)|^2, \qquad n, k = 0, 1, 2, \ldots, n \neq k} \tag{7.29}$$

is the transition probability for the quantum particle to pass from the state φ_k to the state φ_n during the time interval $[t_1, t_2]$. Here, $\omega_{kn} := (E_k - E_n)/\hbar$. This will be motivated below. Note that $\gamma_{kn} = \gamma_{nk}$. Explicitly,

$$\gamma_{kn} := \frac{\omega^2 e^2 (t_2 - t_1)}{6\pi\varepsilon_0 c^3 m} \, (n\delta_{k,n-1} + k\delta_{n,k-1}).$$

This means the following.

- Forbidden spectral lines: The transition of the quantum particle from the state φ_n of energy E_n to the state φ_k of energy E_k is forbidden, i.e., $\gamma_{kn} = 0$, if the energy difference $E_n - E_k$ is equal to $\pm 2\hbar\omega, \pm 3\hbar\omega, \ldots$
- Emission of radiation: The transition probability from the energy E_{n+1} to the energy E_n during the time interval $[t_1, t_2]$ is equal to

$$\gamma_{n+1,n} = \frac{\omega^2 e^2 (t_2 - t_1)}{6\pi\varepsilon_0 c^3 m} \, (n + 1), \qquad n = 0, 1, 2, \ldots \tag{7.30}$$

  In this case, a photon of energy $E = \hbar\omega$ is emitted. The meaning of transition probability is the following. Suppose that we have $\mathcal{N}$ oscillating electrons in the

[24] The numerical values can be found on page 949 of Vol. I.

state φ_n. Then the number of electrons which jump to the state φ_{n+1} during the time interval $[t_1, t-2]$ is equal to $\mathcal{N}\gamma_{n,n+1}$. Then the emitted mean energy E, which passes through a sufficiently large sphere during the time interval $[t_1, t_2]$, is equal to

$$E = \mathcal{N}\gamma_{n+1,n} \cdot \hbar\omega.$$

This quantity determines the intensity of the emitted spectral line.

- Absorption of radiation: The transition probability from the energy E_n to the energy E_{n+1} during the time interval $[t_1, t_2]$ is equal to

$$\gamma_{n,n+1} = \gamma_{n+1,n}, \qquad n = 0, 1, 2, ...$$

In this case, a photon of energy $E_{n+1} - E_n = \hbar\omega$ is absorbed.

Motivation of the transition probability. We want to motivate formula (7.29).

Step 1: Classical particle. Let $q = q(t)$ describe the motion of a classical particle of mass m and electric charge $-e$ on the real line. This particle emits the mean electromagnetic energy $\mathcal{E}$ through a sufficiently large sphere during the time interval $[t_1, t_2]$. Explicitly,

$$\mathcal{E} = \frac{e^2(t_2 - t_1)}{6\pi\varepsilon_0 c^3} \operatorname{mean}(\ddot{q}^2(t))$$

(see Landau and Lifshitz (1982), Sect. 67). We assume that the smooth motion of the particle has the time period T. Then we have the Fourier expansion

$$q(t) = \sum_{r=-\infty}^{\infty} q_r \mathrm{e}^{\mathrm{i}\omega_r t}, \qquad t \in \mathbb{R}$$

with the angular frequency $\omega := 2\pi/T$ and $\omega_r := r\omega$. Since the function $t \mapsto q(t)$ is real, we get $q_r(t)^\dagger = q_{-r}(t)$ for all $r = 0, \pm 1, \pm 2, \ldots$ Hence

$$\ddot{q}^2(t) = \sum_{r,s=-\infty}^{\infty} \omega_r^2 q_r \omega_s^2 q_s \mathrm{e}^{\mathrm{i}(\omega_r+\omega_s)t}.$$

Since $\operatorname{mean}\left(\mathrm{e}^{\mathrm{i}(\omega_r+\omega_s)t}\right) = \frac{1}{T}\int_0^T \mathrm{e}^{\mathrm{i}(\omega_r+\omega_s)t}dt = \delta_{0,r+s}$, we get

$$\operatorname{mean}(\ddot{q}^2(t)) = \sum_{r=-\infty}^{\infty} \omega_r^4 q_r q_{-r} = 2\sum_{r=1}^{\infty} \omega_r^4 |q_r|^2.$$

This yields $\mathcal{E} = \sum_{r=1}^{\infty} \mathcal{E}_r$ with

$$\mathcal{E}_r := \frac{e^2(t_2 - t_1)}{3\pi\varepsilon_0 c^3} \cdot \omega_r^4 |q_r|^2.$$

Step 2: Quantum particle. In 1925 Heisenberg postulated that, for the harmonic oscillator, the passage from the classical particle to the quantum particle corresponds to the two replacements

(i) $\omega_r \Rightarrow \omega_{kn} := (E_k - E_n)/\hbar$, and
(ii) $q_r \Rightarrow q_{kn}(0)$.

Let $k > n$. If the quantum particle jumps from the energy level E_k to the lower energy level E_n, then a photon of energy $E_k - E_n = \hbar\omega_{kn}$ is emitted. Using the replacements (i) and (ii) above, we get $\mathcal{E} = \sum_{k\geq 1}\sum_{n=0}^{k-1}\mathcal{E}_{kn}$ with

$$\mathcal{E}_{kn} := \frac{e^2(t_2 - t_1)}{3\pi\varepsilon_0 c^3}\cdot\omega_{kn}^4|q_{kn}(0)|^2.$$

By definition, the real number

$$\gamma_{kn} := \frac{\mathcal{E}_{kn}}{\hbar\omega_{kn}}, \qquad k > n$$

is the transition probability for a passage of the quantum particle from the energy level E_k to the lower energy level E_n during the time interval $[t_1, t_2]$. From (7.24) we get $|q_{kn}(0)|^2 = \frac{\hbar}{2m\omega}\, k\delta_{n,k-1}$. Hence $\gamma_{kn} = 0$ for the choice $k = n+2, n+3, \ldots$ Moreover,

$$\gamma_{n+1,n} = \frac{\mathcal{E}_{n+1,n}}{\hbar\omega} = \frac{e^2(t_2 - t_1)}{6\pi\varepsilon_0 c^3 m}\cdot\omega^2(n+1), \qquad n = 0, 1, 2, \ldots$$

This motivates the claim (7.30).

7.3.5 The Wightman Functions

> Both the Wightman functions and the correlation functions of the quantized harmonic oscillator are the prototypes of general constructions used in quantum field theory.
>
> Folklore

As we have shown, the motion of the quantum particle corresponding to the quantized harmonic oscillator is described by the time-depending operator function

$$q(t) = \frac{x_0}{\sqrt{2}}\,(a^\dagger \mathrm{e}^{\mathrm{i}\omega t} + a\mathrm{e}^{-\mathrm{i}\omega t}), \qquad t \in \mathbb{R} \tag{7.31}$$

with the initial condition $q(0) = Q$ and $p(0) = P$. Using this, we define the n-point Wightman function of the quantized harmonic oscillator by setting

$$\boxed{W_n(t_1, t_2, \ldots, t_n) := \langle 0|q(t_1)q(t_2)\cdots q(t_n)|0\rangle} \tag{7.32}$$

for all times $t_1, t_2, \ldots, t_n \in \mathbb{R}$. This is the vacuum expectation value of the operator product $q(t_1)q(t_2)\cdots q(t_n)$. In contrast to the operator function (7.31), the Wightman functions are classical complex-valued functions. It turns out that

The Wightman functions know all about the quantized harmonic oscillator.

Using the Wightman functions, we avoid the use of operator theory in Hilbert space. This is the main idea behind the introduction of the Wightman functions.

Proposition 7.4 (i) $W_2(t,s) = \frac{x_0^2}{2}\cdot e^{-\mathrm{i}\omega(t-s)}$ *for all* $t, s \in \mathbb{R}$.

(ii) $W_n \equiv 0$ *if* n *is odd. For example,* $W_1 \equiv 0$ *and* $W_3 \equiv 0$.

(iii) $W_4(t_1,t_2,t_3,t_4) = W_2(t_1,t_2)W_2(t_3,t_4) + 2W_2(t_1,t_3)W_2(t_2,t_4)$ *for all time points* $t_1, t_2, t_3, t_4 \in \mathbb{R}$.

(iv) $W_n(t_1,t_2,\ldots,t_n)^\dagger = W(t_n,\ldots,t_2,t_1)$ *for all times* $t_1, t_2, \ldots t_n$ *and all positive integers* n.

Proof. We will systematically use the orthonormal system $\varphi_0, \varphi_1, \ldots$ introduced on page 433 together with $a\varphi_0 = 0$, $a^\dagger\varphi_0 = \varphi_1$ and

$$a\varphi_n = \sqrt{n}\,\varphi_{n-1}, \quad a^\dagger\varphi_n = \sqrt{n+1}\,\varphi_{n+1}, \qquad n = 1, 2, \ldots$$

Recall that the vacuum state φ_0 is also denoted by $|0\rangle$. The computation of vacuum expectation values becomes extremely simple when using the intuitive meaning of the operator a (resp. $a^\dagger$) as a particle creation (resp. annihilation) operator. Let us explain this by considering a few typical examples. First let us show that most of the vacuum expectation values vanish.

- The state $a^\dagger a^\dagger\varphi_0$ contains two particles. Hence

$$\langle\varphi_0|a^\dagger a^\dagger\varphi_0\rangle = \text{const}\cdot\langle\varphi_0|\varphi_2\rangle = 0,$$

  by orthogonality.
- The state $aa^\dagger a^\dagger\varphi_0$ contains one particle. Hence

$$\langle\varphi_0|aa^\dagger a^\dagger\varphi_0\rangle = \text{const}\cdot\langle\varphi_0|\varphi_1\rangle = 0.$$

- $Aa\varphi_0 = 0$ for arbitrary expressions A, since $a\varphi_0 = 0$.
- Analogously, $aaaa^\dagger a^\dagger\varphi_0 = 0$. In fact,

$$aaaa^\dagger a^\dagger\varphi_0 = a(aaa^\dagger a^\dagger)\varphi_0 = \text{const}\cdot a\varphi_0 = 0.$$

  Formally, the state $aaaa^\dagger a^\dagger\varphi_0$ contains "2 minus 3" particles. In general, states with a 'negative' number of particles are equal to zero.

Therefore, it only remains to compute vacuum expectation values $\langle\varphi_0|A\varphi_0\rangle$ where the state $A\varphi_0$ contains no particle.

> *This means that A is a product of creation and annihilation operators where the number of creation operators equals the number of annihilation operators.*

The following examples will be used below.

- The state $aa^\dagger\varphi_0$ contains no particle. Here,

$$aa^\dagger\varphi_0 = a\varphi_1 = \varphi_0. \tag{7.33}$$

  Hence $\langle\varphi_0|aa^\dagger\varphi_0\rangle = \varphi_0|\varphi_0\rangle = 1$.
- The state $aaa^\dagger a^\dagger\varphi_0$ contains no particle. Explicitly,

$$aaa^\dagger a^\dagger\varphi_0 = aaa^\dagger\varphi_1 = \sqrt{2}\,aa\varphi_2 = 2a\varphi_1 = 2\varphi_0. \tag{7.34}$$

  Hence $\langle\varphi_0|aaa^\dagger a^\dagger\varphi_0\rangle = 2$.
- Similarly,

$$aa^\dagger aa^\dagger\varphi_0 = aa^\dagger a\varphi_1 = aa^\dagger\varphi_0 = a\varphi_1 = \varphi_0. \tag{7.35}$$

  Hence $\langle aa^\dagger aa^\dagger\varphi_0\rangle = 1$.

Ad (i). To simplify notation, set

$$a_j := \frac{x_0 \mathrm{e}^{-\mathrm{i}\omega t_j}}{\sqrt{2}}\, a, \qquad a_j^\dagger := \frac{x_0 \mathrm{e}^{\mathrm{i}\omega t_j}}{\sqrt{2}}\, a^\dagger.$$

We have $W_2(t_1, t_2) = \langle \varphi_0 | A \varphi_0 \rangle$ with the state

$$A\varphi_0 = (a_1^\dagger + a_1)(a_2^\dagger + a_2)\varphi_0.$$

Only the state $a_1 a_2^\dagger \varphi_0$ gives a non-vanishing contribution to the Wightman function W_2. By (7.33), $W_2(t_1, t_2)$ is equal to

$$\langle \varphi_0 | a_1 a_2^\dagger \varphi_0 \rangle = \frac{x_0^2}{2} \cdot \mathrm{e}^{-\mathrm{i}\omega t_1} \mathrm{e}^{\mathrm{i}\omega t_2} \langle \varphi_0 | a a^\dagger \varphi_0 \rangle = \frac{x_0^2}{2} \cdot \mathrm{e}^{-\mathrm{i}\omega(t_1 - t_2)}.$$

Ad (ii). First note that $\langle \varphi_0 | (a^\dagger + a)\varphi_0 \rangle = \langle \varphi_0 | \varphi_1 \rangle = 0$. The state

$$A\varphi_0 := (a_1^\dagger + a_1)(a_2^\dagger + a_2)(a_3^\dagger + a_3)\varphi_0$$

is the sum of particle states with an odd number of particles. Hence we obtain $\langle \varphi_0 | A \varphi_0 \rangle = 0$, by orthogonality. The same is true for an odd number of factors $(a_j^\dagger + a_j)$.

Ad (iii). We have $W_4(t_1, t_2, t_3, t_4) = \langle \varphi_0 | A \varphi_0 \rangle$ with the state

$$A\varphi_0 := (a_1^\dagger + a_1)(a_2^\dagger + a_2)(a_3^\dagger + a_3)(a_4^\dagger + a_4) = a_1 a_2 a_3^\dagger a_4^\dagger + a_1 a_2^\dagger a_3 a_4^\dagger + \ldots$$

The dots denote terms whose contribution to W_4 vanishes. By (7.34) and (7.35), $W_4(t_1, t_2, t_3, t_4)$ is equal to

$$\begin{aligned} \langle \varphi_0 | a_1 a_2 a_3^\dagger a_4^\dagger \varphi_0 \rangle + \langle \varphi_0 | a_1 a_2^\dagger a_3 a_4^\dagger \varphi_0 \rangle &= 2 W_2(t_1, t_3) W_2(t_2, t_4) \\ &\quad + W_2(t_1, t_2) W_2(t_3, t_4). \end{aligned}$$

Ad (iv). Since the operator $Q(t)$ is formally self-adjoint,

$$\langle \varphi_0 | Q(s) Q(t) \varphi_0 \rangle = \langle Q(t) Q(s) \varphi_0 | \varphi_0 \rangle = \langle \varphi_0 | Q(t) Q(s) \varphi_0 \rangle^\dagger.$$

Hence $W_2(s, t) = W_2(t, s)^\dagger$. The general case proceeds analogously. □

Similar arguments for computing vacuum expectation values via creation and annihilation operators are frequently used in quantum field theory.

Theorem 7.5 (i) *Equation of motion: For any $s \in \mathbb{R}$, the 2-point Wightman function $t \mapsto W_2(t, s)$ satisfies the classical equation of motion for the harmonic oscillator, that is,*

$$\frac{\partial^2 W_2(t, s)}{\partial t^2} + \omega^2 W_2(t, s) = 0, \qquad t \in \mathbb{R}.$$

(ii) *Reconstruction property: For all times $t, s \in \mathbb{R}$,*

$$q(t - s) = \frac{\sqrt{2}}{x_0} \left(W_2(t, s) a + W_2(s, t) a^\dagger \right). \tag{7.36}$$

Relation (7.5) tells us that the knowledge of the 2-point Wightman function W_2 allows us to reconstruct the quantum dynamics of the harmonic oscillator.
Proof. Note that $\ddot{q}(t) + \omega^2 q(t) = 0$, and hence

$$\frac{\partial^2 W(t,s)}{\partial^2 t} + \omega^2 W(t,s) = \langle \varphi_0 | (\ddot{q}(t) + \omega^2 q(t)) q(s) \varphi_0 \rangle = 0.$$

□

Perspectives. In 1956 Wightman showed that it is possible to base quantum field theory on the investigation of the vacuum expectation values of the products of quantum fields. These vacuum expectation values are called Wightman functions. The crucial point is that the Wightman functions are highly singular objects in quantum field theory. In fact, they are generalized functions.[25] However, they are also boundary values of holomorphic functions of several complex variables. This simplifies the mathematical theory. Using a similar construction as in the proof of the Gelfand–Naimark–Segal (GNS) representation theorem for C^*-algebras in Hilbert spaces, Wightman proved a reconstruction theorem which shows that the quantum field (as a Hilbert-space valued distribution) can be reconstructed from its Wightman distributions. Basic papers are:

A. Wightman, Quantum field theories in terms of vacuum expectation values, Phys. Rev. **101** (1956), 860–866.

R. Jost, A remark on the CPT-theorem, Helv. Phys. Acta **30** (1957), 409–416 (in German).

F. Dyson, Integral representations of causal commutators, Phys. Rev. **110**(6) (1958), 1460–1464.

A. Wightman, Quantum field theory and analytic functions of several complex variables, J. Indian Math. Soc. **24** (1960), 625–677.

H. Borchers, On the structure of the algebra of field operators, Nuovo Cimento **24** (1962), 214–236.

A. Uhlmann, Über die Definition der Quantenfelder nach Wightman und Haag (On the definition of quantum fields according to Wightman and Haag), Wissenschaftliche Zeitschrift der Karl-Marx-Universität Leipzig **11**(1962), 213–217 (in German).

A. Wightman and L. Gårding, Fields as operator-valued distributions in relativistic quantum theory, Arkiv för Fysik **28** (1964), 129–189.

R. Haag and D. Kastler, An algebraic approach to quantum field theory, J. Math. Phys. **5** (1964), 848–861.

K. Hepp, On the connection between the LSZ formalism and the Wightman field theory, Commun. Math. Phys. **1** (1965)(2), 95–111.

H. Araki and R. Haag, Collision cross sections in terms of local observables, Commun. Math. Phys. **4**(2) (1967), 7–91.

O. Steinmann, A rigorous formulation of LSZ field theory, Commun. Math. Phys. **10** (1968), 245–268.

R. Seiler, Quantum theory of particles with spin zero and one half in external fields, Commun. Math. Phys. **25** (1972), 127–151.

H. Epstein and V. Glaser, The role of locality in perturbation theory, Ann. Inst. Poincaré A **19**(3) (1973), 211–295.

[25] See Sect. 15.6 of Vol. I.

K. Osterwalder and R. Schrader, Axioms for Euclidean Green's functions I, II, Commun. Math. Phys. **31** (1973), 83–112; **42** (1975), 281–305.

D. Buchholz, The physical state space of quantum electrodynamics, Commun. Math. Phys. **85** (1982), 49–71.

J. Glimm and A. Jaffe, Quantum Field Theory and Statistical Mechanics: Expositions, Birkhäuser, Boston, 1985.

D. Buchholz, On quantum fields that generate local algebras, J. Math. Phys. **31** (1990), 1839–1846.

D. Buchholz, M. Porrmann, and U. Stein (1991), Dirac versus Wigner: towards a universal particle concept in local quantum field theory, Phys. Lett. **267 B**(39 (1991), 377–381.

J. Fröhlich, Non-Perturbative Quantum Field Theory: Mathematical Aspects and Applications, Selected Papers, World Scientific, Singapore, 1992.

D. Buchholz and R. Verch, Scaling algebras and renormalization group in algebraic quantum field theory, Rev. Math. Phys. **7** (1995), 1195–2040.

S. Doplicher, K. Fredenhagen, and J. Roberts, The structure of space-time at the Planck scale and quantum fields, Commun. Math. Phys. **172** (1995), 187–220.

As an introduction to axiomatic quantum field theory, we recommend the following monographs:

N. Bogoliubov et al., Introduction to Axiomatic Quantum Field Theory, Benjamin, Reading, Massachusetts, 1975.

R. Haag, Local Quantum Physics: Fields, Particles, Algebras, Springer, Berlin, 1996.

H. Araki, Mathematical Theory of Quantum Fields, Oxford University Press, New York, 1999.

C. Bär, N. Ginoux, and F. Pfäffle, Wave Equations on Lorentzian Manifolds and Quantization, European Mathematical Society, 2007.

We also recommend:

R. Streater and R. Wightman, PCT, Spin, Statistics, and All That, Benjamin, New York, 1968.

M. Reed and B. Simon, Methods of Modern Mathematical Physics. Vol. 2 (the mathematical structure of Wightman distributions), Vol. 3 (the Haag–Ruelle scattering theory), Academic Press, New York, 1972.

B. Simon, The $P(\varphi)_2$-Euclidean Quantum Field Theory, Princeton University Press, 1974 (constructive quantum field theory for a special nontrivial model in a 2-dimensional space-time).

J. Glimm and A. Jaffe, Mathematical Methods of Quantum Physics, Springer, New York, 1981 (constructive quantum field theory based on the use of functional integrals).

N. Bogoliubov et al., General Principles of Quantum Field Theory, Kluwer, Dordrecht, 1990.

In recent years, Klaus Fredenhagen (Hamburg University) has written a series of important papers together with his collaborators. The idea is to combine the operator-algebra methods of axiomatic quantum field theory (due to Gårding–Wightman and Haag–Kastler) with the methods of perturbation theory, by using formal power series expansions. We refer to:

M. Dütsch and K. Fredenhagen, A local perturbative construction of observables in gauge theories: The example of QED (quantum electrodynamics), Commun. Math. Phys. **203** (1999), 71–105.

R. Brunetti and K. Fredenhagen, Micro-local analysis and interacting quantum field theories: renormalization on physical backgrounds, Commun. Math. Phys. **208** (2000), 623–661.

M. Dütsch and K. Fredenhagen, Algebraic quantum field theory, perturbation theory, and the loop expansion, Commun. Math. Phys. **219**(1) (2001), 5–30.

M. Dütsch and K. Fredenhagen, The master Ward identity and the generalized Schwinger–Dyson equation in classical field theory, Commun. Math. Phys. **243** (2003), 275–314.

R. Brunetti, K. Fredenhagen, and R. Verch, The generally covariant locality principle – a new paradigm for local quantum field theory, Commun. Math. Phys. **237** (2003), 31–68.

R. Brunetti and K. Fredenhagen, Towards a background-independent formulation of perturbative quantum gravity, pp. 151–157. In: B. Fauser, J. Tolksdorf, and E, Zeidler (Eds.), Quantum Gravity: Mathematical Models and Experimental Bounds, Birkhäuser, Basel, 2006.

K. Fredenhagen, K. Rehren, and E. Seiler, Quantum field theory: where we are. Lecture Notes in Physics **721** (2007), 61–87
Internet 2006: `http://arxiv.org/hep-th/0603155`

We also recommend the lectures given by Klaus Fredenhagen at Hamburg University. These lectures are available on the Internet:

`http://unith.desy.de/research/aqft/lecture-notes`

Furthermore, we recommend the lectures on quantum field theory given by Arthur Jaffe at Harvard University:

A. Jaffe, Introduction to Quantum Field Theory. Lecture Notes, partially available at: `www.rathurjaffe.com/Assets/pdf/IntroQFT.pdf`

7.3.6 The Correlation Functions

In contrast to the Wightman functions, the correlation functions reflect causality.

Folklore

Parallel to (7.32), we now define the n-point correlation function (also called the n-point Green's function) by setting

$$\boxed{C_n(t_1, t_2, \ldots, t_n) := \langle 0 | \mathcal{T}(q(t_1)q(t_2)\cdots q(t_n))|0\rangle} \tag{7.37}$$

for all times $t_1, t_2, \ldots, t_n \in \mathbb{R}$. Here, the symbol $\mathcal{T}$ denotes the time-ordering operator, that is, we define

$$\mathcal{T}(q(t_1)q(t_2)\cdots q(t_n)) := q(t_{\pi(1)})q(t_{\pi(2)})\cdots q(t_{\pi(n)})$$

where the permutation π of the indices $1, 2, \ldots, n$ is chosen in such a way that $t_{\pi(1)} \geq t_{\pi(2)} \geq \ldots \geq t_{\pi(n)}$. For example, using the slightly modified Heaviside function θ_*, we obtain[26]

[26] We set $\theta_*(t) := 1$ if $t > 0$, $\theta_*(t) := 0$ if $t < 0$, and $\theta_*(0) := \frac{1}{2}$.

$$C_2(t,s) = \theta_*(t-s)W_2(t,s) + \theta_*(s-t)W_2(s,t)) = \frac{x_0^2}{2} \cdot \mathrm{e}^{-\mathrm{i}\omega|t-s|} \tag{7.38}$$

for all $t, s \in \mathbb{R}$. This relates the 2-point correlation function C_2 to the 2-point Wightman function W_2 by taking causality into account. In particular, we have $C_2(t,s) = W_2(t,s)$ if $t \geq s$.

Theorem 7.6 *For any $s \in \mathbb{R}$, the 2-point correlation function $t \mapsto C_2(t,s)$ satisfies the inhomogeneous classical equation of motion for the harmonic oscillator, that is,*

$$\frac{\partial^2 C_2(t,s)}{\partial t^2} + \omega^2 C_2(t,s) = \frac{\hbar}{m\mathrm{i}} \cdot \delta(t-s), \qquad t \in \mathbb{R}, \tag{7.39}$$

in the sense of tempered distributions on the real line.

This theorem tells us that the function $F(t) := \frac{m\mathrm{i}}{\hbar} \cdot C_2(t,0)$ satisfies the differential equation

$$\ddot{F}(t) + \omega^2 F(t) = \delta(t), \qquad t \in \mathbb{R}.$$

In terms of mathematics, the function F is a fundamental solution of the differential operator $\frac{d^2}{dt^2} + \omega^2$, in the sense of tempered distributions (see Sect. 11.7 of Vol. I).

The language of mathematicians. In order to prove Theorem 7.6, we will use the theory of generalized functions (distributions) introduced in Chap. 11 of Vol. I. Let $\psi \in \mathcal{S}(\mathbb{R})$. Integrating by parts twice, we get

$$\begin{aligned}\int_s^\infty \mathrm{e}^{-\mathrm{i}\omega(t-s)}\ddot{\psi}(t)dt &= -\dot{\psi}(s) + \int_s^\infty \mathrm{i}\omega\mathrm{e}^{-\mathrm{i}\omega(t-s)}\dot{\psi}(t)dt \\ &= -\dot{\psi}(s) - \mathrm{i}\omega\psi(s) - \omega^2 \int_s^\infty \mathrm{e}^{-\mathrm{i}\omega(t-s)}\psi(t)dt.\end{aligned}$$

Similarly,

$$\int_{-\infty}^s \mathrm{e}^{-\mathrm{i}\omega(s-t)}\ddot{\psi}(t)dt = \dot{\psi}(s) - \mathrm{i}\omega\psi(s) - \omega^2 \int_{-\infty}^s \mathrm{e}^{-\mathrm{i}\omega(s-t)}\psi(t)dt.$$

Hence

$$\int_{-\infty}^\infty \mathrm{e}^{-\mathrm{i}\omega|t-s|}\ \ddot{\psi}(t)dt = -2\mathrm{i}\omega\psi(s) - \omega^2 \int_{-\infty}^\infty \mathrm{e}^{-\mathrm{i}\omega|t-s|}\ \psi(t)dt.$$

In terms of distribution theory, this is equivalent to

$$\frac{\partial^2 \mathrm{e}^{-\mathrm{i}\omega|t-s|}}{\partial t^2} + \omega^2 \mathrm{e}^{-\mathrm{i}\omega|t-s|} = -2\mathrm{i}\omega\delta(t-s), \qquad t \in \mathbb{R}.$$

□

The language of physicists. We want to show how to obtain the claim of Theorem 7.6 by using Dirac's delta function in a formal setting.[27] For fixed $s \in \mathbb{R}$, consider

$$C(t) := \theta_*(t-s)W(t) + \theta_*(s-t)Z(t), \qquad t \in \mathbb{R}.$$

Differentiating this with respect to time t by means of the product rule and noting that $\dot{\theta}_*(t) = \delta(t)$, we get

$$\dot{C}(t) = \delta(t-s)W(t) - \delta(s-t)Z(t) + \theta_*(t-s)\dot{W}(t) + \theta_*(s-t)\dot{Z}(t).$$

[27] Both the formal Dirac calculus and its relations to the rigorous theory are thoroughly investigated in Sect. 11.2ff of Vol. I.

Using $\delta(t-s)=\delta(s-t)$ and $\delta(t)=0$ if $t\neq 0$, we obtain

$$\dot{C}(t)=\delta(t-s)(W(s)-Z(s))+\theta_*(t-s)\dot{W}(t)+\theta_*(s-t)\dot{Z}(t).$$

Hence

$$\begin{aligned}\ddot{C}(t)=\dot{\delta}(t-s)(W(s)-Z(s))+\delta(t-s)(\dot{W}(s)-\dot{Z}(s))\\ +\theta_*(t-s)\ddot{W}(t)+\theta_*(s-t)\ddot{Z}(t).\end{aligned}$$

Choosing $C(t):=C_2(t)$ and

$$W(t):=W_2(t,s)=\frac{x_0^2}{2}\,\mathrm{e}^{-\mathrm{i}\omega(t-s)}$$

together with $Z(t):=W_2(s,t)$, we get the differential equation (7.39) above.

The physical meaning of correlation functions for the harmonic oscillator. Let $\varphi\in L_2(\mathbb{R})$ with $\langle\varphi|\varphi\rangle=1$. We regard φ as a physical state of the quantized harmonic oscillator on the real line. The operator function $q=q(t),\ t\in\mathbb{R}$ from (7.31) on page 451 describes the motion of the quantum particle. According to the general approach introduced in Sect. 7.9 of Vol. I, we assign to the state φ the following real numbers:

(i) Mean position of the particle in the state φ at time t: $\bar{q}(t):=\langle\varphi|q(t)|\varphi\rangle$.
(ii) Mean fluctuation of the particle position at time t:

$$\Delta q(t):=\sqrt{\langle\varphi|(q(t)-\bar{q}(t))^2\varphi\rangle}.$$

(iii) Correlation coefficient: For $t,s\in\mathbb{R}$, we define

$$\gamma(t,s):=\frac{\overline{(q(t)-\bar{q}(t))(q(s)-\bar{q}(s))}}{\Delta q(t)\Delta q(s)}.$$

By the Schwarz inequality, $|\gamma(t,s)|\leq 1$. If $|\gamma(t,s)|=1$ (resp. $\gamma(t,s)=0$), then the position of the particle in the state φ at time t is strongly correlated (resp. not correlated) to the position in the state φ at time s.
(iv) Causal correlation coefficient:

$$\gamma_{\mathrm{causal}}(t,s):=\gamma(t,s)\qquad\text{if } t\geq s.$$

Furthermore, $\gamma_{\mathrm{causal}}(t,s):=\gamma(s,t)$ if $s\geq t$.
(v) Transition amplitude: Let $\varphi,\psi\in L_2(\mathbb{R})$ with $\langle\varphi|\varphi\rangle=\langle\psi|\psi\rangle=1$. The complex number $\langle\psi|q(t)\varphi\rangle$ is called the transition amplitude (for the position) from the state φ to the state ψ at time t.

To illustrate this, consider the ground state φ_0 of the harmonic oscillator. Then $W_2(t,s)=\frac{\hbar}{2m\omega}\mathrm{e}^{-\mathrm{i}\omega(t-s)}$. Thus, in the ground state, we have:

- Mean position $\bar{q}(t)=0$.
- Mean fluctuation: $\Delta q(t)=\sqrt{\langle\varphi_0|q(t)q(t)\varphi_0\rangle}=\sqrt{W_2(t,t)}=\sqrt{\frac{\hbar}{2m\omega}}$.
- Correlation coefficient:

$$\gamma(t,s)=\frac{W_2(t,s)}{\sqrt{W_2(t,t)}\sqrt{W_2(s,s)}}=\mathrm{e}^{-\mathrm{i}\omega(t-s)},\quad t\geq s,$$

and $\gamma_{\mathrm{causal}}(t,s)=\mathrm{e}^{-\mathrm{i}\omega|t-s|}$. Hence $|\gamma(t,s)|=1$. This means that, in the ground state, the position of the quantum particle at time t is strongly correlated to the position at time s.

- Transition amplitude from the state φ_0 to the state φ_n:

$$\langle\varphi_1|q(t)\varphi_0\rangle = \mathrm{e}^{\mathrm{i}\omega t}, \quad \langle\varphi_n|q(t)\varphi_0\rangle = 0, \qquad n = 2, 3, 2, \ldots$$

By (7.29), the transition probability γ_{n0} for passing from the state φ_0 to the state φ_n during the time interval $[t_1, t_2]$ is proportional to $|\langle\varphi_n|q(0)\varphi_0\rangle|^2$. Explicitly, $\gamma_{10} = \frac{\omega^2 e^2 (t_2 - t_1)}{6\pi\varepsilon_0 c^3 m}$ and $\gamma_{n0} = 0$ if $n = 2, 3, \ldots$

7.4 Schrödinger's Quantum Mechanics

> In particular, I would like to mention that I was mainly inspired by the thoughtful dissertation of Mr. Louis de Broglie (Paris, 1924). The main difference here lies in the following. De Broglie thinks of travelling waves, while, in the case of the atom, we are led to standing waves... I am most thankful to Hermann Weyl with regard to the mathematical treatment of the equation of the hydrogen atom.[28]
>
> Erwin Schrödinger, 1926

7.4.1 The Schrödinger Equation

In 1926 Schrödinger invented wave quantum mechanics based on a wave function $\psi = \psi(x, t)$. The Schrödinger equation for the motion of a quantum particle of mass m on the real line is given by

$$\boxed{\mathrm{i}\hbar\psi_t = -\frac{\hbar^2}{2m}\psi_{xx} + U\psi.} \tag{7.40}$$

Explicitly, the Schrödinger equation reads as

$$\mathrm{i}\hbar\frac{\partial\psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\psi(x,t)}{\partial^2 x} + U(x)\psi(x,t).$$

Schrödinger's quantization. The Schrödinger equation (7.40) is obtained by applying Schrödinger quantization to the classical energy equation

$$\boxed{E = \frac{p^2}{2m} + U.} \tag{7.41}$$

This means that we replace the classical momentum p and the classical energy E by differential operators. Explicitly,

$$E \Rightarrow \mathrm{i}\hbar\frac{\partial}{\partial t}, \qquad p \Rightarrow -\mathrm{i}\hbar\frac{\partial}{\partial x}.$$

From (7.41) we get

$$\boxed{\mathrm{i}\hbar\frac{\partial}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + U.}$$

[28] E.Schrödinger, Quantization as an eigenvalue problem (in German), Ann. Phys. **9** (1926), 361–376. See also E. Schrödinger, Collected Papers on Wave Mechanics, Blackie, London, 1928.

Applying this to the function ψ, we obtain the one-dimensional Schrödinger equation (7.40). Schrödinger generalized this in a straightforward manner to three dimensions, and he computed the spectrum of the hydrogen atom.

The physical interpretation of the wave function ψ. If the potential U vanishes, $U \equiv 0$, then the function

$$\psi_0(x,t) := C\mathrm{e}^{-\mathrm{i}E(p)t/\hbar}\mathrm{e}^{\mathrm{i}px/\hbar}$$

is a solution of the Schrödinger equation (7.40). Here, C is a fixed complex number, p is a fixed real number, and $E(p) := \frac{p^2}{2m}$. The function ψ_0 corresponds to a stream of freely moving electrons on the real line with momentum p and energy $E(p)$. There arises the following question:

What is the physical meaning of the function $\psi = \psi(x,t)$ in the general case?

Interestingly enough, Schrödinger did not know the answer when publishing his paper in 1926. The answer was found by Born a few months later.

By applying the Schrödinger equation to scattering processes, Born discovered the random character of quantum processes.

According to Born, we have to distinguish the following two cases:

(i) Single quantum particle: Suppose that $0 < \int_{\mathbb{R}} |\psi(x)|^2 dx < \infty$. Then, the value

$$\varrho(x,t) := \frac{|\psi(x,t)|^2}{\int_{\mathbb{R}} |\psi(x,t)|^2 dx}$$

represents the particle probability density at position x at time t. That is, the value

$$\int_J \varrho(x,t)dx$$

is equal to the probability of finding the particle in the interval J at time t. Naturally enough, $\int_{\mathbb{R}} \varrho(x,t)dx = 1$. If we measure the position x of the quantum particle, then the mean position $\bar{x}$ and the fluctuation Δx of the position at time t are given by

$$\boxed{\bar{x}(t) = \int_{\mathbb{R}} x\varrho(x,t)dx}$$

and

$$\boxed{(\Delta x)^2 = \overline{(x-\bar{x})^2} = \int_{\mathbb{R}} (x-\bar{x})^2 \varrho(x,t)dx.}$$

By definition, Δx is non-negative. In the theory of probability, a fundamental inequality due to Chebyshev (1821–1894) tells us that

$$P(\bar{x} - r\Delta x \le x \le \bar{x} + r\Delta x) \ge 1 - \frac{1}{r^2}$$

for all $r > 0$. In particular, choose $r = 4$. Then this inequality tells us that the probability of measuring the position x of the quantum particle in the interval $[\bar{x} - 4\Delta x, \bar{x} + 4\Delta x]$ is larger than $1 - \frac{1}{16} = 0.93$.

(ii) Stream of quantum particles: Suppose that $\int_{\mathbb{R}} |\psi(x,t))|^2 dx = \infty$. Then, the function ψ corresponds to a stream of particles on the real line with the particle density

$$\varrho(x,t) := |\psi(x,t)|^2, \qquad x \in \mathbb{R}, \quad t \in \mathbb{R},$$

and the current density vector

$$\mathbf{J}(x,t) = \mathcal{J}(x,t)\mathbf{e}, \qquad x \in \mathbb{R}, \quad t \in \mathbb{R}$$

at the point x at time t. Here, the unit vector $\mathbf{e}$ points in direction of the positive x-axis, and we define

$$\mathcal{J} := \frac{\mathrm{i}\hbar}{2m}(\psi\psi_x^\dagger - \psi^\dagger\psi_x).$$

This definition is motivated by the fact that each smooth solution ψ of the Schrödinger equation (7.40) satisfies the following conservation law[29]

$$\boxed{\varrho_t + \operatorname{div}\mathbf{J} = 0.} \tag{7.42}$$

Explicitly, $\operatorname{div}\mathbf{J} = \mathcal{J}_x$. For $a < b$, this implies the relation

$$\int_a^b \varrho(x,t)dx = \mathcal{J}(a,t) - \mathcal{J}(b,t)$$

which describes the change of the particle number on the interval $[a,b]$ by the particle stream. For example, the function

$$\psi_0(x,t) = C\mathrm{e}^{-\mathrm{i}E(p)t/\hbar}\mathrm{e}^{\mathrm{i}px/\hbar}$$

corresponds to a stream of quantum particles with the constant particle density $\varrho(x,t) = |C|^2$, the velocity $v = p/m$, and the current density vector

$$\mathbf{J} = v\varrho\mathbf{e}.$$

There exist fascinating long-term developments in mathematics. In his books "Geometry" and "Algebra" from 1550 and 1572, respectively, Bombielli (1526–1572) systematically used the symbol $\sqrt{-1}$ in order to solve algebraic equations of third and fourth order. Almost 400 years later, the physicist Schrödinger used the number $\mathrm{i} = \sqrt{-1}$ in order to formulate the basic equations of quantum mechanics. We are going to show that the use of complex numbers is substantial for quantum physics. Freeman Dyson writes in his foreword to Odifreddi's book:[30]

> One of the most profound jokes of nature is the square root of -1 that the physicist Erwin Schrödinger put into his wave equation in 1926 ... The Schrödinger equation describes correctly everything we know about the behavior of atoms. It is the basis of all of chemistry and most of physics. And that square root of -1 means that nature works with complex numbers. This discovery came as a complete surprise, to Schrödinger as well as to everybody else. According to Schrödinger, his fourteen-year-old girlfriend Itha Junger said to him at the time: "Hey, you never even thought when

[29] In fact, $\varrho_t = (\psi\psi^\dagger)_t = \psi_t\psi^\dagger + \psi\psi_t^\dagger$. By (7.40), $\varrho_t = -\mathcal{J}_x$.

[30] P. Odifreddi, The Mathematical Century: The 30 Greatest Problems of the Last 100 Years, Princeton University Press, Princeton, New Jersey, 2004. Reprinted by permission of Princeton University Press.

you began that so much sensible stuff would come out of it." All through the nineteenth century, mathematicians from Abel to Riemann and Weierstrass had been creating a magnificent theory of functions of complex variables. They had discovered that the theory of functions became far deeper and more powerful if it was extended from real to complex numbers. But they always thought of complex numbers as an artificial construction, invented by human mathematicians as a useful and elegant abstraction from real life. It never entered their heads that they had invented was in fact the ground on which atoms move. They never imagined that nature had got there first.

In what follows, we want to show that the notion of Hilbert space is an appropriate setting for describing quantum mechanics in terms of mathematics. Originally, the special Hilbert space l^2 (as an infinite-dimensional variant of $\mathbb{R}^n$) was introduced by Hilbert in the beginning of the 20th century in order to study eigenvalue problems for integral equations.

7.4.2 States, Observables, and Measurements

The Hilbert space approach. In 1926, the young Hungarian mathematician von Neumann Janos came to Göttingen as Hilbert's assistant.[31] In Göttingen, von Neumann learned about the new quantum mechanics of physicists. It was his goal to give quantum mechanics a rigorous mathematical basis. As a mathematical framework, he used the notion of Hilbert space. For example, in the present case of the motion of a quantum particle on the real line, we choose the Hilbert space $L_2(\mathbb{R})$ with the inner product

$$\boxed{\langle\psi|\chi\rangle = \int_{\mathbb{R}} \psi(x)^\dagger \chi(x) dx \qquad \text{for all} \quad \psi, \chi \in L_2(\mathbb{R}),}$$

and the norm $||\psi|| := \sqrt{\langle\psi|\psi\rangle}$. The general terminology reads as follows.

(S) States: Each nonzero element ψ of $L_2(\mathbb{R})$ is called a state. In terms of physics, this describes a state of a single quantum particle on the real line. Two nonzero elements ψ, χ of $L_2(\mathbb{R})$ represent equivalent states iff there exists a nonzero complex number μ with

$$\psi = \mu\chi.$$

In terms of physics, equivalent states represent the same physical state of the particle. The state ψ is called normalized iff $||\psi|| = 1$.

(O) Observables: The linear, formally self-adjoint operators

$$A : D(A) \subseteq X \to X$$

are called formal observables. Explicitly, this means that the domain of definition $D(A)$ is a linear subspace of X. Moreover, for all $\psi, \chi \in D(A)$ and all complex numbers α, β, we have

[31] Von Neumann (1903–1957) was born in Budapest (Hungary). He studied mathematics and chemistry in Berlin, Budapest, and Zurich. The German (resp. English) translation of the Hungarian name 'Janos' is Johann (resp. John). Von Neumann was an extraordinarily gifted mathematician. He was known for his ability to understand mathematical subjects and to solve mathematical problems extremely fast. In 1933, von Neumann got a professorship at the newly founded Institute for Advanced Study in Princeton, New Jersey (U.S.A.).

$$A(\alpha\psi + \beta\chi) = \alpha A\psi + \beta A\chi$$

together with the symmetry condition $\langle\psi|A\chi\rangle = \langle A\psi|\chi\rangle$.[32]

(M) Measurements: If we measure the formal observable A in the normalized state ψ, then we get the mean value

$$\boxed{\bar{A} := \langle\psi|A\psi\rangle,}$$

and the mean fluctuation[33]

$$\boxed{\Delta A := ||(A - \bar{A}I)\psi||.}$$

(C) Correlation coefficient: Let $A, B : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ be two formal observables. The correlation coefficient between A and B in the state ψ is defined by

$$\gamma := \frac{\mathrm{Cov}(A, B)}{\Delta A \cdot \Delta B}$$

together with the covariance

$$\mathrm{Cov}(A, B) := \overline{(A - \bar{A}I)(B - \bar{B}I)} = \langle\psi|(A - \bar{A}I)(B - \bar{B}I)\psi\rangle.$$

Hence $\mathrm{Cov}(A, B) = \langle(A - \bar{A}I)\psi|(B - \bar{B}I)\psi\rangle$.

By the Schwarz inequality, $|\gamma| \leq 1$.

- If $\gamma = 0$, then there is no correlation between the formal observables A and B. In other words, A and B are independent formal observables.
- If $|\gamma| = 1$, then the correlation between A and B is large. That is, the formal observable A depends strongly on the formal observable B.

Proposition 7.7 *The mean value is a real number.*

This is a consequence of $\langle\psi|A\psi\rangle^\dagger = \langle A\psi|\psi\rangle = \langle\psi|A\psi\rangle$. □

The following result underlines the importance of eigenvalue problems in quantum mechanics.

Proposition 7.8 *Suppose that the normalized state ψ is an eigenvector of the formal observable A with eigenvalue λ,*

$$A\psi = \lambda\psi.$$

Then, the measurement of A in the state ψ yields $\bar{A} = \lambda$ and $\Delta A = 0$.

[32] For a deeper mathematical analysis, von Neumann introduced the stronger notion of an observable. By definition, an observable is an essentially self-adjoint operator (see Vol. I, p. 677).

[33] Explicitly, $(\Delta A)^2 = \langle A\psi - \bar{A}\psi|A\psi - \bar{A}\psi\rangle$. If $A\psi \in D(A)$, then

$$(\Delta A)^2 = \langle\psi|(A - \bar{A}I)^2\psi\rangle = \overline{(A - \bar{A}I)^2}.$$

In this case, we say that λ is a sharp value of the formal observable A. For the proof, $\langle\psi|A\psi\rangle = \lambda\langle\psi|\psi\rangle = \lambda$, and $A\psi - \bar{A}\psi = A\psi - \lambda\psi = 0$. □

Examples. The operators $Q, P, H : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ are defined by

$$(Q\psi)(x) := x\psi(x), \quad (P\psi)(x) = -\mathrm{i}\hbar\psi'(x), \qquad x \in \mathbb{R},$$

for all functions $\psi \in \mathcal{S}(\mathbb{R})$. We call Q and P the position operator and the momentum operator, respectively. Moreover, we introduce the energy operator (Hamiltonian)

$$H := \frac{P^2}{2m} + U,$$

where we assume that $U \in \mathcal{S}(\mathbb{R})$. Then the fundamental operator equation

$$\mathrm{i}\hbar\dot{\psi} = H\psi$$

coincides with the Schrödinger equation (7.40).

Proposition 7.9 *The operators $Q, P, H : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ are formally self-adjoint on the Hilbert space $L_2(\mathbb{R})$, and there holds the commutation relation*

$$QP - PQ = \mathrm{i}\hbar I \qquad \text{on } \mathcal{S}(\mathbb{R}). \tag{7.43}$$

Proof. The formal self-adjointness of Q and P together with (7.43) are proved on page 436. Let $\psi \in \mathcal{S}(\mathbb{R})$. The formal self-adjointness of H follows from

$$\langle\psi|P^2\psi\rangle = \langle P\psi|P\psi\rangle = \langle P^2\psi|\psi\rangle.$$

Hence $\langle\psi|H\psi\rangle = \langle H\psi|\psi\rangle$. □

7.4.3 The Free Motion of a Quantum Particle

The classical motion of a particle of mass m on the real line is governed by the Hamiltonian $H := \frac{p^2}{2m}$ together with the canonical equations

$$\dot{q} = H_p = \frac{p}{m}, \qquad \dot{p} = -H_q = 0.$$

For given initial position $q(0) = q_0$ and initial velocity $\dot{q}(0) = v$, the unique solution reads as $q(t) = q_0 + vt$ for all times $t \in \mathbb{R}$ with the total energy

$$E(p) := \frac{p^2}{2m} = \frac{mv^2}{2}.$$

The free motion of a quantum particle on the real line is governed by the Hamiltonian operator

$$\boxed{H := \frac{P^2}{2m}.} \tag{7.44}$$

Recall that $P = -\mathrm{i}\hbar\frac{d}{dx}$, and hence

$$H = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2}.$$

At this point, we regard the operators P and H as differential operators which act on smooth functions (or on generalized functions).[34] For the functional-analytic approach to quantum mechanics, it is important to appropriately specify the domain of definition of the operators under consideration. This will be discussed below. For fixed nonzero complex number C, define the functions

$$\varphi_p(x) := C\mathrm{e}^{\mathrm{i}px/\hbar}, \qquad \psi_p(x,t) = \varphi_p(x)\mathrm{e}^{-\mathrm{i}tE(p))/\hbar}, \qquad x,t \in \mathbb{R}.$$

Then the function ψ_p satisfies the Schrödinger equation

$$\mathrm{i}\hbar\dot{\psi}_p = H\psi_p.$$

Moreover, for all parameters $p \in \mathbb{R}$, we have

$$P\varphi_p = p\varphi_p, \qquad H\varphi_p = E(p)\varphi_p.$$

These equations remain valid if we replace φ_p by ψ_p. From the physical point of view, the function ψ_p describes a homogeneous stream of quantum particles (e.g., electrons) with particle density $\varrho = |C|^2$ and velocity v. Note that the functions φ_p and $x \mapsto \psi_p(x,t)$ do not live in the Hilbert space $L_2(\mathbb{R})$.

Let $\varphi, \chi \in \mathcal{S}(\mathbb{R})$. Normalizing the function φ_p above by $C := \frac{1}{\sqrt{2\pi\hbar}}$, we get the Fourier transform

$$\hat{\varphi}(p) = \int_{\mathbb{R}} \varphi_p(x)^\dagger \varphi(x)dx, \qquad p \in \mathbb{R}$$

together with the inverse transform

$$\varphi(x) = \int_{\mathbb{R}} \varphi_p(x)\hat{\varphi}(p)dp, \qquad x \in \mathbb{R}.$$

The operator $\mathcal{F} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ is bijective (see Vol. I, p. 87). We write $\hat{\varphi} = \mathcal{F}\varphi$. This Fourier transform can be uniquely extended to a unitary operator of the form $\mathcal{F} : L_2(\mathbb{R}) \to L_2(\mathbb{R})$,s that is, we have

$$\langle\varphi|\chi\rangle = \langle\hat{\varphi}|\hat{\chi}\rangle, \qquad \text{for all} \quad \varphi,\chi \in L_2(\mathbb{R}),$$

which is called the Parseval equation of the Fourier transform.

The quantum dynamics of a freely moving particle. Let us now study the three operators

- $P : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ (momentum operator),
- $Q : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ position operator), and
- $H : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ (Hamiltonian).

These operators are formally self-adjoint on the Hilbert space $L_2(\mathbb{R})$. In the Fourier space, the operators P and H correspond to the following multiplication operators

$$(\hat{P}\hat{\varphi})(p) = p\hat{\varphi}(p), \qquad (\hat{H}\hat{\varphi})(p) = E(p)\hat{\varphi}(p), \qquad p \in \mathbb{R}.$$

This holds for all $\varphi \in \mathcal{S}(\mathbb{R})$, and hence for all $\hat{\varphi} \in \mathcal{S}(\mathbb{R})$. For given $\varphi_0 \in \mathcal{S}(\mathbb{R})$, the quantum dynamics

$$\boxed{\psi(t) = \mathrm{e}^{-\mathrm{i}Ht/\hbar}\varphi_0, \qquad t \in \mathbb{R}}$$

is given in the Fourier space by the equation

[34] The Schwartz $\mathcal{S}'(\mathbb{R})$ of tempered distributions and the Schwartz space $\mathcal{D}'(\mathbb{R})$ of distributions are investigated in Sect. 11.3 of Vol. I. Here, $\mathcal{S}'(\mathbb{R}) \subset \mathcal{D}'(\mathbb{R})$.

$$\hat{\psi}(p,t) = \mathrm{e}^{-\mathrm{i}E(p)t/\hbar}\hat{\varphi}_0(p), \qquad p \in \mathbb{R}$$

for each time $t \in \mathbb{R}$. Transforming this back to the original Hilbert space $L_2(\mathbb{R})$ by using the Fourier transform, we get the quantum dynamics

$$\boxed{\mathrm{e}^{-\mathrm{i}tH_0/\hbar}\varphi_0 = \mathcal{F}^{-1}\hat{\psi}(t) \qquad \text{for all} \quad t \in \mathbb{R}.} \tag{7.45}$$

We have $\psi(t) \in \mathcal{S}(\mathbb{R})$ for all times $t \in \mathbb{R}$, and this function satisfies the Schrödinger equation for all times.[35]

The full quantum dynamics. Consider equation (7.45). Observe the following peculiarity. The right-hand side of (7.45) is well-defined for initial states $\varphi_0 \in L_2(\mathbb{R})$ if we do not use the classical Fourier transform, but the extended Fourier transform $\mathcal{F} : L_2(\mathbb{R}) \to L_2(\mathbb{R})$. In this sense, we understand the dynamics

$$\psi(t) = \mathrm{e}^{-\mathrm{i}tH/\hbar}\varphi_0, \qquad t \in \mathbb{R}$$

for all initial states $\varphi_0 \in L_2(\mathbb{R})$. In terms of functional analysis, for any fixed time t, the operator $\mathrm{e}^{-\mathrm{i}tH/\hbar} : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ is unitary. Therefore, $\mathrm{e}^{-\mathrm{i}tH/\hbar}\varphi_0$ makes sense for all $\varphi_0 \in L_2(\mathbb{R})$. In this general setting, the function

$$\psi : [0, +\infty[\to L_2(\mathbb{R})$$

is continuous, but not necessarily differentiable. Therefore, it can be regarded as a generalized solution of the Schrödinger equation $\mathrm{i}\hbar\dot{\psi}(t) = H\psi(t), t \in \mathbb{R}$.

Measurement of observables. Suppose that we are given a normalized state $\varphi \in \mathcal{S}(\mathbb{R})$, that is,

$$||\varphi||^2 = \int_{\mathbb{R}} |\varphi(x)|^2 dx = 1.$$

By the Parseval equation,

$$||\hat{\varphi}||^2 = \int_{\mathbb{R}} |\hat{\varphi}(p)|^2 dp = ||\varphi||^2 = 1.$$

Let us now measure the position, the momentum, and the energy of a quantum particle on the real line where the particle is in the normalized state $\varphi \in \mathcal{S}(\mathbb{R})$.

(i) Measurement of position: For the mean value $\bar{x}$ and the mean fluctuation $\Delta x \geq 0$ of the particle position, we get

$$\bar{x} = \langle \varphi | Q\varphi \rangle = \int_{\mathbb{R}} x|\varphi(x)|^2 dx$$

and

$$(\Delta x)^2 = \langle \varphi | (Q - \bar{x}I)^2 \varphi \rangle = \int_{\mathbb{R}} (x - \bar{x})^2 |\varphi(x)|^2 dx.$$

The number $\int_J |\varphi(x)|^2 dx$ is the probability for measuring the particle position in the interval J.

[35] Fix $t \in \mathbb{R}$. The symbol $\psi(t)$ (resp. $\hat{\psi}(t)$) stands for the function $x \mapsto \psi(x,t)$ (resp. $p \mapsto \hat{\psi}(p,t)$) on $\mathbb{R}$.

(ii) Measurement of momentum: For the mean value $\bar{p}$ and the mean fluctuation Δp of the particle momentum, we get

$$\bar{p} = \langle \varphi | P\varphi \rangle = \langle \hat{\varphi} | \hat{P}\hat{\varphi} \rangle = \int_{\mathbb{R}} p \, |\hat{\varphi}(p)|^2 dp$$

and

$$\Delta p = \langle \varphi | (P - \bar{p}I)^2 \varphi \rangle = \int_{\mathbb{R}} (p - \bar{p})^2 |\hat{\varphi}(p)|^2 dp.$$

The number $\int_J |\hat{\varphi}(p)|^2 dp$ is the probability for measuring the particle momentum in the interval J.

(iii) Measurement of energy: Suppose we are given a measuring instrument which analyzes the energy of freely moving particles. The measured energy corresponds to the observable H. For the mean value $\bar{E}$ and the mean fluctuation ΔE of the energy in the normalized state $\varphi \in \mathcal{S}(\mathbb{R})$, we get

$$\bar{E} = \langle \varphi | H\varphi \rangle = \langle \hat{\varphi} | \hat{H}\hat{\varphi} \rangle = \int_{\mathbb{R}} E(p) |\hat{\varphi}(p)|^2 dp$$

and

$$\Delta E = \langle \varphi | (H - \bar{E}I)^2 \varphi \rangle = \int_{\mathbb{R}} (E(p) - \bar{E})^2 |\hat{\varphi}(p)|^2 dp.$$

The number

$$\int_{E(p) \in J} |\hat{\varphi}(p)|^2 dp$$

is the probability for measuring the particle energy in the given energy interval J. Recall that $E(p) = p^2/2m$. Fix the positive real number E. Then we have $E(p) \leq E$ iff $|p|^2 \leq 2mE$. Thus, the number

$$\int_{|p| \leq \sqrt{2mE}} |\hat{\varphi}(p)|^2 dp$$

is equal to the probability for measuring the energy $E(p)$ of the particle in the interval $[0, E]$.

The full functional-analytic approach to the free quantum particle will be studied in Sect. 7.6.4 on page 509.

7.4.4 The Harmonic Oscillator

Let us quantize the classical harmonic oscillator in the sense of Schrödinger's quantum mechanics. We will see that we obtain the same results as in Heisenberg's version of quantum mechanics. In Sect. 7.4.5, we will explain why Schrödinger's quantum mechanics is equivalent to Heisenberg's quantum mechanics. Choosing $\varphi \in \mathcal{S}(\mathbb{R})$, recall the definition of the position operator Q and the momentum operator P,

$$(Q\varphi)(x) := x\varphi(x), \qquad (P\varphi)(x) := -\mathrm{i}\hbar\varphi'(x) \qquad \text{for all} \quad x \in \mathbb{R}.$$

Quantization means that we replace the classical Hamiltonian function

$$H(q,p) = \frac{p^2}{2m} + \frac{m\omega^2 q^2}{2}$$

by the Hamiltonian operator

$$\boxed{H := \frac{P^2}{2m} + \frac{m\omega^2 Q^2}{2}.}$$

The Schrödinger equation for the wave function $\psi = \psi(x,t),\ x,t \in \mathbb{R}$, reads as

$$\boxed{\mathrm{i}\hbar\dot{\psi} = H\psi} \tag{7.46}$$

along with the prescribed initial condition $\psi(x,0) = \psi_0(x)$ for all $x \in \mathbb{R}$. Explicitly,

$$\mathrm{i}\hbar\psi_t(x,t) = -\frac{\hbar^2}{2m}\psi_{xx}(x,t) + \frac{m\omega^2 x^2}{2}\,\psi(x,t), \qquad x,t \in \mathbb{R}.$$

We are going to show that

The Hamiltonian operator H knows all about the quantized harmonic oscillator.

This is a typical feature for all quantum systems. Making the classic Fourier ansatz

$$\psi(x,t) := \varphi(x)\mathrm{e}^{-\mathrm{i}Et/\hbar}, \qquad x,t \in \mathbb{R},$$

we get the stationary Schrödinger equation

$$\boxed{E\varphi = H\varphi} \tag{7.47}$$

for the time-independent function φ. Explicitly,

$$E\varphi(x) = -\frac{\hbar^2}{2m}\,\varphi''(x) + \frac{m\omega^2}{2}\,x^2\varphi(x), \qquad x \in \mathbb{R}.$$

Again let us use the typical length $x_0 := \sqrt{\frac{\hbar}{\omega m}}$.

The eigensolutions of the Hamiltonian. Our mathematical investigation of the quantized harmonic oscillator will be based on the eigensolutions of the Hamiltonian. Motivated by Sect. 7.2, the basic trick is to introduce the two operators $a, a^\dagger : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ by letting

$$a := \frac{1}{\sqrt{2}}\left(\frac{Q}{x_0} + \frac{\mathrm{i}x_0 P}{\hbar}\right), \qquad a^\dagger := \frac{1}{\sqrt{2}}\left(\frac{Q}{x_0} - \frac{\mathrm{i}x_0 P}{\hbar}\right). \tag{7.48}$$

This forces the crucial factorization

$$H = \hbar\omega(a^\dagger a + \tfrac{1}{2})$$

of the Hamiltonian operator. Starting from the Gaussian probability density,

$$\varrho(x) := \frac{\mathrm{e}^{-x^2/2\sigma^2}}{\sigma\sqrt{2\pi}},$$

with the mean value $\bar{x} = 0$ and the mean fluctuation $\sigma := \frac{x_0}{\sqrt{2}}$, we define

$$\varphi_0(x) := \sqrt{\varrho(x)} \qquad \text{for all} \quad x \in \mathbb{R}.$$

The following theorem is basic for quantum physics.

Theorem 7.10 *The Hamiltonian H of the quantized harmonic oscillator has the eigensolutions $H\varphi_n = E_n\varphi_n$, $n = 0, 1, 2, , ...$ with the energy eigenvalues*

$$\boxed{E_n := \hbar\omega(n + \tfrac{1}{2})} \tag{7.49}$$

and the eigenstates

$$\varphi_n := \frac{(a^\dagger)^n}{\sqrt{n!}}\ \varphi_0.$$

The system $\varphi_0, \varphi_1, ...$ forms a complete orthonormal system in the Hilbert space $L_2(\mathbb{R})$.

Proof. To simplify notation, let $x_0 = 1$ by the rescaling $x \mapsto x/x_0$. The proof follows then from Sect. 7.2 on page 432. □

Explicitly, for all $x \in \mathbb{R}$ and $n = 0, 1, 2, ...$, we have

$$\varphi_n(x) = \frac{1}{\sqrt{2^n n! x_0 \sqrt{\pi}}}\ H_n\left(\frac{x}{x_0}\right)\ \exp\left\{-\frac{1}{2}\left(\frac{x}{x_0}\right)^2\right\}.$$

Mnemonically, physicists write $|E_n\rangle$ instead of φ_n.

Corollary 7.11 *For $n = 0, 1, 2, ...$,*
(i) $\bar{x} = \langle\varphi_n|Q\varphi_n\rangle = 0$;
(ii) $(\Delta x)^2 = \langle\varphi_n|(Q - \bar{x}I)^2\varphi_n\rangle = x_0^2(n + \frac{1}{2})$;
(iii) $\bar{p} = \langle\varphi_n|P\varphi_n\rangle = 0$;
(iv) $(\Delta p)^2 = \langle\varphi_n|(P - \bar{p}I)^2\varphi_n\rangle = \frac{\hbar^2}{x_0^2}(n + \frac{1}{2})$.

Proof. Let $x_0 = 1$ by the rescaling $x \mapsto x/x_0$.

Ad (i), (iii). Note that the Hermite functions φ_n are odd or even by (7.8). Hence

$$\int_{\mathbb{R}} x|\varphi_n(x)|^2 dx = 0, \qquad \int_{\mathbb{R}} \varphi_n(x)^\dagger \varphi_n'(x) dx = 0.$$

Ad (ii). Let $n = 0, 1, 2, ...$ By Sect. 7.2,

$$a\varphi_{n+1} = \sqrt{n+1}\ \varphi_n, \qquad a^\dagger\varphi_n = \sqrt{n+1}\ \varphi_{n+1}, \qquad a^\dagger a\varphi_n = n\varphi_n.$$

From $2\langle\varphi_n|Q^2\varphi_n\rangle = \langle\varphi_n|(a + a^\dagger)^2\varphi_n\rangle$ we get

$$2\langle\varphi_n|Q^2\varphi_n\rangle = \langle\varphi_n|(a^2 + aa^\dagger + a^\dagger a + a^\dagger a^\dagger)\varphi_n\rangle = 2n + 1.$$

In fact, because of $\langle\varphi_{n+1}|\varphi_{n-1}\rangle = 0$, we obtain

$$\langle\varphi_n|a^2\varphi_n\rangle = \langle a^\dagger\varphi_n|a\varphi_n\rangle = 0.$$

Moreover, $\langle\varphi_n|aa^\dagger\varphi_n\rangle = \langle a^\dagger\varphi_n|a^\dagger\varphi_n\rangle = n + 1$.

Ad (iv). Similarly, $2\langle\varphi_n|P^2\varphi_n\rangle = -\hbar^2\langle\varphi_n|(a - a^\dagger)^2\varphi_n\rangle = \hbar^2(2n + 1)$. □

Physical interpretation. Let us discuss some physical consequences.

(i) Ground state: The state

$$\psi(x,t) := \mathrm{e}^{-\mathrm{i}E_0 t/\hbar}\varphi_0(x), \qquad t,x \in \mathbb{R}$$

represents the lowest-energy state of the harmonic oscillator called ground state (or vacuum state). The sharp energy of the ground state equals $E_0 = \hbar/2$. For the mean position $\bar{x}$ and the mean fluctuation Δx of the particle position in the ground state, it follows from Corollary 7.11 that

$$\bar{x} = 0, \qquad \Delta x = \sigma = \frac{x_0}{\sqrt{2}}.$$

For the mean momentum $\bar{p}$ and the mean fluctuation Δp of the particle momentum in the ground state, we get $\bar{p} = 0$ and $\Delta x \Delta p = \frac{\hbar}{2}$.

(ii) The uncertainty inequality: In the normalized state

$$\psi(x,t) := \mathrm{e}^{-\mathrm{i}E_n t/\hbar}\varphi_n(x), \qquad n = 0,1,\ldots,$$

the particle has the sharp energy $E_n = \hbar\omega(n+\frac{1}{2})$, and

$$\bar{x} = 0, \qquad \Delta x = x_0\sqrt{n+\tfrac{1}{2}}$$

as well as

$$\bar{p} = 0, \qquad \Delta x \Delta p = \frac{E_n}{\omega} = \hbar\left(n+\tfrac{1}{2}\right).$$

From this we get

$$\boxed{\Delta x \Delta p \geq \frac{\hbar}{2}.}$$

In 1927 Heisenberg discovered that this inequality is the special case of a fundamental law in nature called the uncertainty of position and momentum (see Sect. 7.4.6 on page 475).

(iii) Measurement of energy: The energy states $\varphi_0, \varphi_1, \ldots$ form a complete orthonormal system in the Hilbert space $L_2(\mathbb{R})$.[36] This means that we have the orthogonality relation

$$\langle\varphi_n|\varphi_m\rangle = \int_{\mathbb{R}} \varphi_n(x)^\dagger \varphi_m(x)dx = \delta_{nm}, \qquad n,m = 0,1,2,\ldots.$$

Completeness means that for each $\chi \in L_2(\mathbb{R})$, the Fourier series

$$\chi = \sum_{n=0}^{\infty} \langle\varphi_n|\chi\rangle\varphi_n$$

converges in the Hilbert space $L_2(\mathbb{R})$. In other words,

$$\lim_{N\to+\infty} \int_{\mathbb{R}} \left|\chi(x) - \sum_{n=0}^{N} \langle\varphi_n|\chi\rangle\varphi_n(x)\right|^2 dx = 0.$$

Moreover, for given complex numbers a_n, the series $\sum_{n=0}^{\infty} a_n\varphi_n$ converges in $L_2(\mathbb{R})$ iff

[36] The properties of complete orthonormal systems in Hilbert spaces are thoroughly studied in Zeidler (1995a), Sect. 3.1 (see the references on page 1049).

$$\sum_{n=0}^{\infty} |a_n|^2 < \infty. \tag{7.50}$$

In addition, for all $\chi, \varphi \in L_2(\mathbb{R})$, we have the Parseval equation

$$\langle \chi|\varphi\rangle = \sum_{n=0}^{\infty} \langle \chi|\varphi_n\rangle\langle\varphi_n|\varphi\rangle. \tag{7.51}$$

Suppose now that $\langle\chi|\chi\rangle = 1$. Then

$$\sum_{n=0}^{\infty} |\langle\chi|\varphi_n\rangle|^2 = 1.$$

This motivates the following definition. If the particle is in the normalized state χ, then the number

$$\boxed{\sum_{E_n \in J} |\langle\varphi_n|\chi\rangle|^2}$$

is equal to the probability of measuring the energy value E of the particle in the interval J. In particular, choosing the open interval $J :=]-\infty, E[$, we obtain the energy distribution function

$$\mathsf{F}(E) := \sum_{E_n < E} |\langle\varphi_n|\chi\rangle|^2. \tag{7.52}$$

In particular, in the state χ we measure the mean energy

$$\bar{E} = \int_{\mathbb{R}} E \, d\mathsf{F}(E) = \sum_{n=0}^{\infty} E_n |\langle\varphi_n|\chi\rangle|^2$$

and the mean energy fluctuation

$$(\Delta E)^2 = \int_{\mathbb{R}} (E - \bar{E})^2 d\mathsf{F}(E) = \sum_{n=0}^{\infty} (E_n - \bar{E})^2 |\langle\varphi_n|\chi\rangle|^2.$$

(iv) Self-adjoint extension of the formally self-adjoint Hamiltonian H: Let us define an operator $H : D(H) \subseteq L_2(\mathbb{R}) \to L_2(\mathbb{R})$ by setting

$$H\psi := \sum_{n=0}^{\infty} E_n \langle\varphi_n|\psi\rangle\varphi_n.$$

Naturally enough, an element $\psi \in L_2(\mathbb{R})$ belongs to the domain of definition, $D(H)$, of the operator H iff the infinite series converges. This means that $\sum_{n=0}^{\infty} E_n^2 |\langle\varphi_n|\psi\rangle|^2 < \infty$. The operator $H : D(H) \to L_2(\mathbb{R})$ is an extension of the operator $H : \mathcal{S}(\mathbb{R}) \to L_2(\mathbb{R})$. In fact, if $\psi \in \mathcal{S}(\mathbb{R})$, then we obtain $\langle\varphi_n|H\psi\rangle = \langle H\varphi_n|\psi\rangle = E_n\langle\varphi_n|\psi\rangle$. Hence

$$H\psi = \sum_{n=0}^{\infty} \langle\varphi_n|H\psi\rangle\varphi_n = \sum_{n=0}^{\infty} E_n\langle\varphi_n|\psi\rangle\varphi_n.$$

The quantum dynamics of the harmonic oscillator. For each initial state ψ_0 in $L_2(\mathbb{R})$ and each time $t \in \mathbb{R}$, we define

$$\mathrm{e}^{-\mathrm{i}Ht/\hbar}\psi_0 := \sum_{n=0}^{\infty} \mathrm{e}^{-\mathrm{i}E_n t/\hbar}\langle\varphi_n|\psi_0\rangle\varphi_n.$$

This series is convergent because of

$$\sum_{n=0}^{\infty} |\mathrm{e}^{-\mathrm{i}E_n t/\hbar}\langle\varphi_n|\psi_0\rangle|^2 = \sum_{n=0}^{\infty} |\langle\varphi_n|\psi_0\rangle|^2 = ||\psi_0||^2 < \infty.$$

By definition, the equation

$$\boxed{\psi(t) = \mathrm{e}^{-\mathrm{i}Ht/\hbar}\psi_0 \qquad \text{for all} \quad t \in \mathbb{R}}$$

describes the dynamics of the quantum harmonic oscillator on the real line. The following theorem motivates this definition.

Theorem 7.12 *For each time $t \in \mathbb{R}$, the operator $e^{-iHt/\hbar} : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ is unitary.*

For given initial value $\psi_0 \in D(H)$, the function $\psi(t) := e^{-\mathrm{i}Ht/\hbar}\psi_0$ satisfies the Schrödinger equation $\mathrm{i}\hbar\dot{\psi}(t) = H\psi(t)$ for all times $t \in \mathbb{R}$.

Proof. For each $\psi_0 \in L_2(\mathbb{R})$ and all $t, s \in \mathbb{R}$,

$$\mathrm{e}^{-\mathrm{i}Hs/\hbar}(\mathrm{e}^{-\mathrm{i}\hbar Ht/\hbar}\psi_0) = \sum_{n=0}^{\infty} \mathrm{e}^{-\mathrm{i}E_n s/\hbar}\mathrm{e}^{-\mathrm{i}E_n t/\hbar}\langle\varphi_n|\psi_0\rangle\varphi_n = \mathrm{e}^{-\mathrm{i}H(t+s)/\hbar}\psi_0.$$

Choosing $s = -t$, this implies

$$\mathrm{e}^{\mathrm{i}Ht/\hbar}(\mathrm{e}^{-\mathrm{i}Ht/\hbar}\psi_0) = \psi_0.$$

Thus, the operator $\mathrm{e}^{\mathrm{i}tH/\hbar}$ is the inverse operator to the operator $\mathrm{e}^{-\mathrm{i}tH/\hbar}$ on the Hilbert space $L_2(\mathbb{R})$. Moreover, because of $|\mathrm{e}^{-\mathrm{i}E_n t/\hbar}| = 1$ it follows from the Parseval equation that

$$||\mathrm{e}^{-\mathrm{i}tH/\hbar}\psi_0||^2 = \sum_{n=0}^{\infty} |\mathrm{e}^{-\mathrm{i}tE_n/\hbar}|^2|\langle\varphi_n|\psi_0\rangle|^2 = ||\psi_0||^2.$$

Therefore, the operator $\mathrm{e}^{-\mathrm{i}Ht/\hbar} : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ is bijective and norm preserving, i.e., this operator is unitary. In particular, if the initial state ψ_0 is normalized, then so is $\mathrm{e}^{-\mathrm{i}Ht/\hbar}\psi_0$ for each time $t \in \mathbb{R}$.

Choose now $\psi_0 \in D(H)$. Formal differentiation with respect to time t yields

$$\mathrm{i}\hbar\dot{\psi}(t) = \sum_{n=0}^{\infty} E_n \mathrm{e}^{-\mathrm{i}E_n t/\hbar}\langle\varphi_n|\psi_0\rangle\varphi_n.$$

To justify this formal differentiation, it is sufficient to use the following majorant series[37]

[37] We refer to Zeidler (1995a), Sect. 5.8 (see the references on page 1049).

$$\sum_{n=0}^{\infty} |E_n \mathrm{e}^{-\mathrm{i}E_n t/\hbar}\langle\varphi_n|\psi_0\rangle|^2 \le \sum_{n=0}^{\infty} |E_n\langle\varphi_n|\psi_0\rangle|^2 = ||H\psi_0||^2 < \infty. \qquad \square$$

Transition probabilities. Let ψ_0 and ψ_1 be two normalized states in the Hilbert space $L_2(\mathbb{R})$. By definition, the real number

$$\boxed{\tau := |\langle\psi_1|\mathrm{e}^{-\mathrm{i}Ht/\hbar}\psi_0\rangle|^2}$$

represents the transition probability from the initial state ψ_0 to the final state ψ_1 during the time interval $[0, t]$. In order to motivate this definition, observe that

- $0 \le \tau \le 1$;
- $\tau = 1$ for the final state $\psi_1 := \mathrm{e}^{-\mathrm{i}tH/\hbar}\psi_0$;
- $\sum_{n=0}^{\infty} \tau_n = 1$ if τ_n corresponds to the final energy state φ_n, i.e.,

$$\tau_n := |\langle\varphi_n|\mathrm{e}^{-\mathrm{i}Ht/\hbar}\psi_0\rangle|^2.$$

In fact, it follows from the Schwarz inequality that

$$\tau \le ||\psi_1|| \cdot ||\mathrm{e}^{-\mathrm{i}tH/\hbar}\psi_0|| = ||\psi_1|| \cdot ||\psi_0|| = 1.$$

Moreover, $\langle \mathrm{e}^{-\mathrm{i}Ht/\hbar}\psi_0|\mathrm{e}^{-\mathrm{i}Ht/\hbar}\psi_0\rangle = \langle\psi_0|\psi_0\rangle = 1$. Finally,

$$\sum_{n=0}^{\infty} |\langle\varphi_n|\mathrm{e}^{-\mathrm{i}Ht/\hbar}\psi_0\rangle|^2 = ||\mathrm{e}^{-\mathrm{i}Ht/\hbar}\psi_0||^2 = ||\psi_0||^2 = 1.$$

7.4.5 The Passage to the Heisenberg Picture

Using the harmonic oscillator, we want to discuss in which sense the Heisenberg approach to quantum mechanics is equivalent to the Schrödinger approach.

Formal approach. The basic transformation from the Schrödinger picture to the Heisenberg picture reads as

$$\boxed{\psi(t) \mapsto \psi(0), \qquad A \mapsto A(t) := \mathrm{e}^{\mathrm{i}Ht/\hbar} A \mathrm{e}^{-\mathrm{i}Ht/\hbar}}$$

for all times $t \in \mathbb{R}$.

(S) Schrödinger picture: In this setting, the states $\psi(t)$ of the quantum harmonic oscillator on the real line are elements of the Hilbert space $L_2(\mathbb{R})$ which depend on time t,

$$\psi(t) = \mathrm{e}^{-\mathrm{i}Ht/\hbar}\psi(0), \qquad t \in \mathbb{R}.$$

The formal observables are formally self-adjoint operators

$$A : D(A) \subseteq L_2(\mathbb{R}) \to L_2(\mathbb{R})$$

which do not depend on time t. Differentiating the state function $t \mapsto \psi(t)$ with respect to time t, we get the Schrödinger equation

$$\boxed{\mathrm{i}\hbar\dot{\psi}(t) = H\psi(t), \qquad t \in \mathbb{R}.}$$

(H) Heisenberg picture: Here, the states $\psi(0)$ of the quantum harmonic oscillator are elements of the Hilbert space $L_2(\mathbb{R})$ which do not depend on time t. The formal observables $A(t)$ are operators on the Hilbert space $L_2(\mathbb{R})$ which depend on time t. Differentiating the function $t \mapsto A(t)$ with respect to time, we get the Heisenberg equation

$$\boxed{\mathrm{i}\hbar\dot{A}(t) = A(t)H - HA(t), \qquad t \in \mathbb{R}.}$$

From the physical point of view, we are interested in measurements of quantities in physical experiments. The point is that both the Schrödinger picture and the Heisenberg picture yield the same mean values. Explicitly,

$$\boxed{\bar{A}(t) = \langle\psi(t)|A\psi(t)\rangle = \langle\psi(0)|A(t)\psi(0)\rangle.}$$

Rigorous approach. Let us start with the Schrödinger picture for the quantum harmonic oscillator on the real line. Consider the self-adjoint Hamiltonian $H : D(H) \subseteq L_2(\mathbb{R}) \to L_2(\mathbb{R})$ introduced in Sect. 7.4.4. Explicitly,

$$H\varphi = \sum_{n=0}^{\infty} E_n \langle\varphi_n|\varphi\rangle\varphi_n.$$

Here, we have $\varphi \in D(H)$ iff this series is convergent in the Hilbert space $L_2(\mathbb{R})$. Define

$$D_0(H) := \mathrm{span}\{\varphi_0, \varphi_1, \varphi_2, ...\},$$

i.e., $D_0(H)$ is the set of finite linear combinations of the eigenfunctions $\varphi_0, \varphi_1, ...$ with complex coefficients.

Theorem 7.13 *Let $A : \mathcal{S}(\mathbb{R}) \to L_2(\mathbb{R})$ be a formally self-adjoint operator which maps $D_0(H)$ into itself. Then, for each $\varphi \in D_0(H)$ and all times $t \in \mathbb{R}$, the expression*

$$A(t)\varphi := e^{\mathrm{i}Ht/\hbar} A e^{-\mathrm{i}Ht/\hbar}\varphi$$

is well-defined, and we have the differential equation

$$i\hbar\frac{d}{dt}(A(t)\varphi) = (A(t)H - HA(t))\varphi.$$

Proof. All of the expressions are well-defined, since they refer to finite linear combinations of the eigenfunctions $\varphi_0, \varphi_1, ...$ Note that $\mathrm{e}^{-\mathrm{i}Ht/\hbar}\varphi_n$ is equal to $\mathrm{e}^{-\mathrm{i}E_n t/\hbar}\varphi_n$, and we have $A\varphi_n \in D_0(H)$ for all n. □

Example. The transformation of the formal observables

$$Q, P : \mathcal{S}(\mathbb{R}) \to L_2(\mathbb{R})$$

from the Schrödinger picture to the Heisenberg picture yields

$$Q(t)\varphi = \frac{x_0}{\sqrt{2}}\,(a^\dagger \mathrm{e}^{\mathrm{i}\omega t} + a\mathrm{e}^{-\mathrm{i}\omega t})\varphi$$

and

$$P(t)\varphi = \frac{\mathrm{i}\hbar}{x_0\sqrt{2}}\,(a^\dagger \mathrm{e}^{\mathrm{i}\omega t} - a\mathrm{e}^{-\mathrm{i}\omega t})\varphi$$

for all $\varphi \in D_0(H)$ and all times $t \in \mathbb{R}$.

Proof. To simplify notation, let $x_0 = \hbar = 1$. It follows from the basic relations $\mathrm{e}^{\mathrm{i}Ht}\varphi_n = \mathrm{e}^{\mathrm{i}E_n t}\varphi_n$ and $a\varphi_n = \sqrt{n}\,\varphi_{n-1}$ that

$$\mathrm{e}^{\mathrm{i}Ht}a\varphi_n = \mathrm{e}^{\mathrm{i}E_{n-1}t}a\varphi_n.$$

Noting that $E_n = \omega(n + \frac{1}{2})$,

$$\mathrm{e}^{\mathrm{i}Ht}a\mathrm{e}^{-\mathrm{i}Ht}\varphi_n = \mathrm{e}^{\mathrm{i}E_{n-1}t}\mathrm{e}^{-\mathrm{i}E_n t}a\varphi_n = \mathrm{e}^{-\mathrm{i}\omega t}a\varphi_n.$$

Similarly, $a^\dagger\varphi_n = \sqrt{n+1}\,\varphi_{n+1}$ implies

$$\mathrm{e}^{\mathrm{i}Ht}a^\dagger\mathrm{e}^{-\mathrm{i}Ht}\varphi_n = \mathrm{e}^{\mathrm{i}E_{n+1}t}\mathrm{e}^{\mathrm{i}E_n t}a^\dagger\varphi_n = \mathrm{e}^{\mathrm{i}\omega t}a^\dagger\varphi_n.$$

Summarizing,

$$\mathrm{e}^{\mathrm{i}Ht}Q\mathrm{e}^{-\mathrm{i}Ht}\varphi_n = \frac{1}{\sqrt{2\omega}}\,\mathrm{e}^{\mathrm{i}Ht}(a^\dagger + a)\mathrm{e}^{-\mathrm{i}Ht}\varphi_n = Q(t)\varphi_n.$$

The proof for P proceeds similarly. □

7.4.6 Heisenberg's Uncertainty Principle

In 1927 Heisenberg discovered that there exists a deep difference between classical mechanics and quantum mechanics.[38] He derived the following fundamental result in quantum physics:

> *The classical notion of the trajectory of a particle, which has a precise position and a precise velocity at the same time, is not meaningful anymore in quantum mechanics.*

Explicitly, for the operators $Q, P : \mathcal{S}(\mathbb{R}) \to L_2(\mathbb{R})$ called position operator Q and momentum operator P, we have the Heisenberg commutation relation

$$\boxed{(QP - PQ)\varphi = \mathrm{i}\hbar\varphi \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).} \tag{7.53}$$

Let $\varphi \in \mathcal{S}(\mathbb{R})$ be a normalized state in the Hilbert space $L_2(\mathbb{R})$. We claim that

$$\boxed{\Delta x \Delta p \geq \frac{\hbar}{2}.} \tag{7.54}$$

This means that it is impossible to measure precisely the position and the momentum of the quantum particle in the state φ at the same time. The uncertainty inequality (7.54) follows from (7.53) as a special case of Theorem 10.4 on page 524 of Vol. I.

[38] W. Heisenberg, The intuitive meaning of kinematics in quantum mechanics, Z. Physik **43** (1927), 172–199 (in German).

7.4.7 Unstable Quantum States and the Energy-Time Uncertainty Relation

> In particle accelerators, many particles are unstable; such so-called resonances only live a very short time.
>
> Folklore

We are going to show that wave packets are unstable in quantum mechanics. There exists a fundamental inequality between the life-time of the wave packet and its mean energy fluctuation which is called the energy–time uncertainty relation.

Wave packets and the Fourier transformation. Let $E(p) := p^2/2m$ denote the energy of a freely moving classical particle on the real line with mass m, momentum $p \in \mathbb{R}$, and velocity $v = p/m$. For each nonzero complex number C, the standing plane wave

$$\psi(x,t) := Ce^{-\mathrm{i}tE(p)/\hbar}\, e^{\mathrm{i}px/\hbar}, \qquad x,t \in \mathbb{R}$$

describes a stream of particles with mass m, momentum p, velocity $v = p/m$, energy $E(p)$, and particle density $\varrho = |C|^2$. Since $|\psi(x,t)|^2 = |C|^2$, the wave function ψ does not live in the Hilbert space $L_2(\mathbb{R})$. However, using the superposition

$$\psi(x,t) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} A(p) e^{-itE(p)/\hbar}\, e^{\mathrm{i}px/\hbar} dp \tag{7.55}$$

of standing plane waves with different momenta, we can construct so-called wave packets which live in the Hilbert space $L_2(\mathbb{R})$ if the amplitude function $A = A(p)$ lives in the space $\mathcal{S}(\mathbb{R})$. The Fourier transformation yields

$$A(p)e^{-iE(p)t/\hbar} = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \psi(x,t) \mathrm{e}^{-\mathrm{i}xp/\hbar} dx.$$

Let us consider a typical example. Choose the Gauss distribution

$$A(p) = \frac{1}{\sqrt{\Delta p \sqrt{2\pi}}}\ \exp\left(-\frac{(p-\bar{p})^2}{4(\Delta p)^2}\right), \tag{7.56}$$

where the real numbers $\bar{p}$ and $\Delta p > 0$ are given. In order to understand the physics of wave packets, let us introduce the following quantities

$$\Delta x_0 := \frac{\hbar}{2\Delta p}, \qquad \Delta E := \frac{(\Delta p)^2}{2m}, \qquad \Delta t := \frac{\hbar}{2\Delta E},$$

and $\bar{v} := \bar{p}/m$, $\bar{x} := \bar{v}t$, , $\bar{E} := \bar{p}^2/2m$, as well as

$$\Delta x = \Delta x_0 \sqrt{1 + \left(\frac{t}{\Delta t}\right)^2}. \tag{7.57}$$

The following proposition summarizes the properties of the wave packet.

Proposition 7.14 *The absolute value of the wave function ψ from (7.55), (7.56) is a Gauss function,*

$$|\psi(x,t)|^2 = \frac{1}{\Delta x\sqrt{2\pi}} \exp\left(-\frac{(x-\bar{x})^2}{2(\Delta x)^2}\right).$$

The mean values and mean fluctuations of the position operator Q and the momentum operator P in the state ψ at time t are $\bar{x}, \Delta x, \bar{p}, \Delta p$, respectively.

This follows by using classical formulas for Gauss–Fresnel integrals. □

This result allows the following physical interpretation. The wave function ψ lives in the Hilbert space $L_2(\mathbb{R})$. It represents a particle with mean momentum $\bar{p}$, mean energy $\bar{E}$, mean fluctuation of momentum Δp and mean fluctuation of energy ΔE. Moreover, the mean position $\bar{x} = \bar{v}t$ of the particle moves with the velocity $\bar{v} = \bar{p}/m$ called the group velocity of the wave packet. It is quite remarkable that

The wave packet is unstable.

In fact, by (7.57), the mean fluctuations Δx of the position of the particle go to infinity as time goes to infinity, that is, the particle is spread over the whole real line after a sufficiently long time. The lifetime of the particle can be measured by the quantity Δt. According to (7.57), the position fluctuations Δx increase in the time interval $[0, \Delta t]$ by the factor $\sqrt{2}$.

The energy-time uncertainty principle. The equation

$$\Delta p \Delta x = \frac{\hbar}{2}$$

for the ground state of a harmonic oscillator represents a special case of the general momentum-position uncertainty inequality $\Delta p \Delta x \geq \frac{\hbar}{2}$. It shows that the Heisenberg uncertainty inequality cannot be improved. Furthermore, we have the equation

$$\Delta E \Delta t = \frac{\hbar}{2}$$

for the Gaussian wave packet. In general, physicists assume that for all unstable particles, there holds the energy-time uncertainty inequality

$$\boxed{\Delta E \Delta t \geq \frac{\hbar}{2}} \tag{7.58}$$

for the lifetime Δt of the particle and its energy fluctuation ΔE. In high-energy particle accelerators, physicists observe frequently so-called resonances. These are unstable particles of mass Δm which decay after the time Δt. By Einstein's mass-energy equivalence, we have

$$\Delta E = c^2 \Delta m$$

where c denotes the speed of light in a vacuum. From (7.58) we get the following fundamental inequality in particle physics

$$\boxed{\Delta m \Delta t \geq \frac{\hbar}{2c^2}}$$

between the mass Δm of a resonance and its lifetime Δt.

The energy-time uncertainty principle is motivated by Einstein's theory of special relativity. Let us explain this. In special relativity, an event corresponds to a four-vector

$$(x, y, z, ct)$$

in Minkowski space. This is a combination of space and time. Similarly, there exists a combination of momentum (p_x, p_y, p_z) and energy E described by the four-vector

$$(p_x, p_y, p_z, \frac{E}{c}).$$

The momentum-energy uncertainty principle yields

$$\Delta p_x \Delta x \geq \frac{\hbar}{2}, \quad \Delta p_y \Delta y \geq \frac{\hbar}{2}, \quad \Delta p_z \Delta z \geq \frac{\hbar}{2}.$$

Postulating complete relativistic symmetry in nature, we can replace p_x and x by E/c and ct, respectively. This yields (7.58).

The energy-time uncertainty inequality represents one of the basic principles of modern physics. Physicists call the ground state of our world the vacuum. This ground state cannot be observed in a straight-forward way. However, there exist quantum fluctuations of the vacuum which can be observed as physical effects; for example, this concerns the fine structure of the energy spectrum of the hydrogen atom, the anomalous magnetic moment of the electron, and the vaporization of black holes in the universe. To understand this, one needs the methods of quantum field theory.

7.4.8 Schrödinger's Coherent States

There arises the following question: Is it possible to construct a stable time-dependent wave packet by the superposition of time-dependent eigenstates of the quantum harmonic oscillator? The positive answer was found by Schrödinger in 1926.[39] For each complex number $\alpha = |\alpha| \mathrm{e}^{\mathrm{i}\delta}$, we define the coherent state

$$\boxed{\psi_\alpha(x,t) := \mathrm{e}^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \mathrm{e}^{-\mathrm{i}E_n t/\hbar} \varphi_n(x) \, \frac{\alpha^n}{\sqrt{n!}}, \qquad x, t \in \mathbb{R}}$$

where the pair $\varphi_n, E_n = \hbar\omega(n + \frac{1}{2})$ is the nth eigensolution of the Hamiltonian for the quantum harmonic oscillator. For each $\alpha \in \mathbb{C}$, the function ψ_α possesses the following properties:

(i) Schrödinger equation: ψ_α is a solution of the time-dependent Schrödinger equation for the harmonic oscillator.

(ii) Normalization: $x \mapsto \psi_\alpha(x,t)$ is a normalized state in the Hilbert space $L_2(\mathbb{R})$ for each time $t \in \mathbb{R}$.

(iii) Mean position: $\bar{x}(t) = \langle \psi_\alpha(t) | Q \psi_\alpha \rangle = \sqrt{2}\, x_0 |\alpha| \cos(\omega t - \delta)$ for all times $t \in \mathbb{R}$. Recall that $x_0 := \sqrt{\hbar/m\omega}$.

(iv) Probability density: For all $x, t \in \mathbb{R}$,

$$|\psi_\alpha(x,t)|^2 = \frac{1}{\sigma\sqrt{2\pi}} \, \mathrm{e}^{-(x-\bar{x}(t))^2/2\sigma^2}.$$

This is a Gaussian distribution where the mean value $\bar{x}(t)$ oscillates with the angular frequency ω, and the time-independent mean fluctuation is given by $\sigma = x_0/\sqrt{2}$.

(v) Eigenvectors of the annihilation operator: $a\psi_\alpha(x,0) = \alpha\psi_\alpha(x,0)$ for all $x \in \mathbb{R}$.

Let us prove this. Explicitly, for all $x, t \in \mathbb{R}$,

$$\psi_\alpha(x,t) = \frac{1}{\sqrt{x_0 2^n \sqrt{\pi}}} \, \mathrm{e}^{-|\alpha|^2/2} \, \mathrm{e}^{-x^2/2x_0^2} \, \mathrm{e}^{-\mathrm{i}\omega t/2} \sum_{n=0}^{\infty} \frac{\alpha^n \mathrm{e}^{-\mathrm{i}n\omega t}}{n!} \, H_n\left(\frac{x}{x_0}\right).$$

The generating function for the Hermite polynomials reads as

[39] E. Schrödinger, The continuous passage from micromechanics to macromechanics, Naturwissenschaften **44** (1926), 664–666 (in German).

$$A\mathrm{e}^{-\xi^2+2\xi\eta} = A\sum_{n=0}^{\infty} \frac{\xi^n}{n!}\, H_n(\eta).$$

Choosing the quantities

$$\xi := \frac{\alpha \mathrm{e}^{-\mathrm{i}\omega t}}{\sqrt{2}}, \quad \eta := \frac{x}{x_0}, \quad A := \frac{1}{\sqrt{x_0\sqrt{\pi}}}\, \mathrm{e}^{-|\alpha|^2/2}\, \mathrm{e}^{-x^2/2x_0^2}\, \mathrm{e}^{-\mathrm{i}\omega t/2},$$

we get $\psi_\alpha(x,t) = A\mathrm{e}^{-\xi^2+2\xi\eta}$. The claims follow now easily by using standard calculus formulas along with $\mathrm{e}^{\mathrm{i}z} = \cos z + \mathrm{i}\sin z$. For (v), note that $a\varphi_0 = 0$ and $a\varphi_n = \sqrt{n}\,\varphi_{n-1}$ if $n = 1, 2, \ldots$ □

In the 1960s, coherent states were used in laser optics for the representation of coherent light waves. As a standard textbook on coherent states and laser optics, we recommend the monograph by L. Mandel and E. Wolf, Optical Coherence and Quantum Optics, Cambridge University Press, 1995.

7.5 Feynman's Quantum Mechanics

> It is a curious historical fact that quantum mechanics began with two quite different mathematical formulations: the differential equation of Schrödinger, and the matrix algebra of Heisenberg. The two, apparently dissimilar approaches, were proved to be mathematically equivalent. These two points of view were destined to complement one another and to be ultimately synthesized in Dirac's transformation theory.
>
> This paper will describe what is essentially a third formulation of nonrelativistic quantum theory. This formulation was suggested by some of Dirac's remarks concerning the relation of classical action to quantum mechanics. A probability amplitude is associated with an entire motion of a particle as a function of time, rather than simply with a position of the particle at a particular time.
>
> The formulation is mathematically equivalent to the more usual formulations. There are, therefore, no fundamentally new results. However, there is a pleasure in recognizing old things from a new point of view. Also, there are problems for which the new point of view offers a distinct advantage.[40]
>
> Richard Feynman, 1948

> The calculations that I did for Hans Bethe, using the Schrödinger equation, took me several months of work and several hundred sheets of paper. Dick Feynman (1918–1988) could get the same answer, calculating on a blackboard, in half an hour.[41]
>
> Freeman Dyson, 1979

Convention. Let z be a nonzero complex number with

$$z = |z|\mathrm{e}^{\mathrm{i}\varphi}, \qquad -\pi < \varphi < \pi,$$

that is, we exclude the non-positive real values, $z \le 0$. In the following sections, $\sqrt{z}$ denotes the principal value of the square root defined by

[40] R. Feynman, Space-time approach to nonrelativistic quantum mechanics, Phys. Rev. **20** (1948), 367–387.

[41] F. Dyson, Disturbing the Universe, Harper & Row, New York, 1979.

$$\sqrt{z} := \sqrt{|z|}\, \mathrm{e}^{\mathrm{i}\varphi/2}. \tag{7.59}$$

For example, $\sqrt{\mathrm{i}} = \mathrm{e}^{\mathrm{i}\pi/4}$. If we use the principal values, then the function

$$z \mapsto \sqrt{z} \tag{7.60}$$

is holomorphic on the set $\mathbb{C}\backslash]-\infty, 0]$ (the complex plane cut along the negative real axis). Thus, analytic continuation of the function $f(x) := \sqrt{x}, x > 0$ yields the function (7.60). This fact will be frequently used in what follows. The idea is to pass from time t to imaginary time $\mathrm{i}t$ and to use analytic continuation in order to translate well-known results from diffusion processes to quantum processes. This is called the Euclidean strategy in quantum physics. The following golden rule holds:

> *Apply analytic continuation only to such quantities that you can measure in physical experiments.*

Analytic continuation of functions depending on energy plays a crucial role in studying the following subjects:

- scattering processes,
- the energies energies of bound states, and
- the energies of unstable particles having finite lifetime (called resonances).

For this, we refer to Sect. 8.3.5 on page 713. In terms of the double-sheeted Riemann surface $\mathcal{R}$ of the multi-valued square-root function (used by physicists in quantum physics), the principal value of $\sqrt{z}$ in the open upper (resp. lower) half-plane corresponds to the first (resp. second sheet) of $\mathcal{R}$ (see Fig. 8.6 on page 714).

Similarly, as for the square root, the value $\ln z := \ln|z| + \mathrm{i}\varphi$ is called the principal value of the logarithm, where the argument φ of the square root is uniquely defined as above by the condition $-\pi < \varphi < \pi$. The function $z \mapsto \ln z$ is holomorphic on $\mathbb{C}\backslash]-\infty, 0]$.

7.5.1 Main Ideas

The basic idea of Feynman's approach to quantum mechanics is

- to describe the time-evolution of a quantum system by an integral formula, which is equivalent to the Schrödinger differential equation,
- and to represent the kernel $\mathcal{K}(x, t; y, t_0)$ of the integral formula by a path integral.

From the physical point of view, Feynman emphasized that

> *The description of quantum particles becomes easier if we use probability amplitudes as basic quantities, but not transition probabilities.*

The reason is that, in contrast to transition probabilities, probability amplitudes satisfy a simple composition rule (also called product rule) which is at the heart of Feynman's approach to quantum theory. In terms of finite-dimensional Hilbert spaces, the following hold:

- Feynman's probability amplitudes are precisely the complex-valued Fourier coefficients $c_1, c_2, \ldots, c_n$ of a state vector.
- Feynman's composition rule for probability amplitudes coincides with the Parseval equation (7.81) for Fourier coefficients in mathematics.[42]

[42] Parseval des Chénes (1755–1836), Fourier (1768–1830), Dirac (1902–1984), von Neumann (1903–1957), Laurent Schwartz (1915–2004), Feynman (1918–1988), Gelfand (born 1913).

- The transition probabilities correspond to the quadratic quantities

$$|c_1|^2, |c_2|^2, \ldots, |c_n|^2,$$

which do *not* linearly depend on the corresponding state vector, in contrast to the Fourier coefficients $c_1, c_2, \ldots, c_n$.

In infinite-dimensional Hilbert spaces, one has to replace Fourier series by Fourier integrals and their generalizations (e.g., Fourier–Stieltjes integrals). In physics, this corresponds to the formal Dirac calculus. In terms of mathematics, one has to use von Neumann's spectral theory for self-adjoint operators and the more general Gelfand theory of generalized eigenfunctions based on Laurent Schwartz's language of distributions (generalized functions).

Feynman's integral formula. According to Schrödinger, the motion of a quantum particle of mass $m > 0$ on the real line is described by the differential equation

$$\mathrm{i}\hbar\psi_t(x,t) = -\frac{\hbar^2}{2m}\psi_{xx}(x,t) + U(x)\psi(x,t), \quad \psi(x,t_0) = \psi_0(x), \tag{7.61}$$

for all positions $x \in \mathbb{R}$ and all times $t > t_0$. Feynman used the fact that the solution of this initial-value problem can be represented by the integral formula

$$\boxed{\psi(x,t) = \int_{\mathbb{R}} \mathcal{K}(x,t;x_0,t_0)\psi_0(x_0)dx_0, \qquad x \in \mathbb{R},\ t > t_0.} \tag{7.62}$$

The main task is to compute the kernel $\mathcal{K}$, which is called the (retarded) Feynman propagator kernel. There exist two different methods.[43]

(i) The Fourier method: Following Fourier's approach to the heat conduction equation, one can use eigenfunction expansions (e.g., Fourier series or Fourier integrals) in order to get the kernel $\mathcal{K}$. For the heat kernel, we will discuss this in (7.77) below.[44]

(ii) The path integral method: In his 1942 Princeton dissertation, Feynman (1918–1988) invented the path integral representation

$$\boxed{\mathcal{K}(x,t;x_0,t_0) = \int_{C\{t_0,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\ \mathcal{D}q.} \tag{7.63}$$

Here, we sum over all possible classical paths $q : [t_0, t] \to \mathbb{R}$ on the real line with fixed endpoints: $q(t_0) = x_0$ and $q(t) = x$. The symbol $S[q]$ denotes the classical action of the path $q = q(\tau), t_0 < \tau < t$.

According to Feynman, the passage from classical mechanics to quantum mechanics corresponds to a statistics over all possible classical paths where the statistical weight $e^{\mathrm{i}S[q]/\hbar}$ depends on the classical action.

[43] In terms of finite-dimensional Hilbert spaces, the two methods are thoroughly investigated in Volume I. For the Fourier method (resp. the Feynman path integral method), see formula (7.82) on page 421 of Vol. I (resp. formula (7.78) on page 417 of Vol. I).

[44] J. Fourier, La théorie de la chaleur (heat theory), Paris, 1822. Interestingly enough, Fourier (1768–1830) was obsessed with heat, keeping his rooms extremely hot.

This is a highly intuitive interpretation of the quantization of classical processes. Let us discuss the intuitive background.

Causality and the product rule for the Feynman propagator. The Feynman propagator kernel satisfies the following product rule:

$$\mathcal{K}(x,t;x_0,t_0) = \int_{\mathbb{R}} \mathcal{K}(x,t;y,\tau)\mathcal{K}(y,\tau;x_0,t_0)\, dy, \qquad t > \tau > t_0. \tag{7.64}$$

It follows from (7.62) that this relation reflects causality. To explain this, choose $t_0 < \tau < t$. We start with a wave function $\psi = \psi(x_0, t_0)$ at the initial time t_0. For the wave function at the intermediate time τ and at the final time t, we get

$$\psi(y,\tau) = \int_{\mathbb{R}} \mathcal{K}(y,\tau;x_0,t_0)\psi(x_0,t_0)dx_0 \tag{7.65}$$

and

$$\psi(x,t) = \int_{\mathbb{R}} \mathcal{K}(x,t;y,\tau)\psi(y,\tau)dy, \tag{7.66}$$

respectively. By causality, we expect that $\psi(x,t)$ at the final time t can also be generated by the wave function at the initial time t_0, that is,

$$\psi(x,t) = \int_{\mathbb{R}} \mathcal{K}(x,t;x_0,t_0)\psi(x_0,t_0)dx_0. \tag{7.67}$$

Now the product formula (7.64) tells us that indeed the composition of the two formulas (7.65) and (7.66) yields (7.67).

The infinitesimal Feynman propagator kernel. In order to obtain his path integral, Feynman used the causality condition (7.64) and the following magic approximation formula:

$$\boxed{\mathcal{K}(x+\Delta x, t+\Delta t; x, t) = \mathrm{e}^{\mathrm{i}\Delta S/\hbar} \cdot \mathcal{K}_{\mathrm{fluct}}(t+\Delta t; t).} \tag{7.68}$$

This is an approximation formula for small position differences Δx and small time differences Δt. Explicitly, we use

- the classical action difference

$$\Delta S := \left(\frac{m}{2}\left(\frac{\Delta x}{\Delta t}\right)^2 - U(x)\right)\Delta t$$

  with the discrete velocity $\frac{\Delta x}{\Delta t}$ and the discrete energy $\Delta E := \Delta S/\Delta t$, and
- the infinitesimal quantum fluctuation term

$$\mathcal{K}_{\mathrm{fluct}}(t+\Delta t; t) := \sqrt{\frac{m}{2\pi\hbar \mathrm{i}\Delta t}}.$$

Here, ΔS is an approximation of the classical action

$$S[q] := \int_t^{t+\Delta t} \left\{\frac{m}{2}\,\dot{q}(\tau)^2 - U(q(\tau))\right\}\, d\tau$$

for a classical trajectory $q = q(\tau)$ which connects the two points x and $x+\Delta x$, that is, $q(t) = x$ and $q(t+\Delta t) = x + \Delta x$. Here, the symbol m denotes the mass of the particle on the real line. The magic formula (7.68) tells us that

The passage from classical mechanics to quantum mechanics is obtained by adding quantum fluctuations.

The magic formula (7.68) combines the infinitesimal strategy due to Newton (1643–1727) and Leibniz (1646–1616) with the principle of least action due to Leibniz, Maupertuis (1698–1759) and Euler (1707–1783). Introducing the (complex) characteristic length[45]

$$l := \frac{1}{\mathcal{K}_{\text{fluct}}(t+\Delta t;t)} = \sqrt{\frac{2\pi\hbar \mathrm{i}\Delta t}{m}},$$

the magic Feynman formula (7.68) reads as

$$\boxed{\mathcal{K}(x+\Delta x, t+\Delta t; x, t) = \frac{\mathrm{e}^{\mathrm{i}\Delta S/\hbar}}{l}.}$$

This reflects the fact that the Feynman propagator kernel $\mathcal{K}$ has the physical dimension $(\text{length})^{-1}$ for the motion of a quantum particle on the real line.

The global Feynman propagator kernel. Combining the causality principle (7.64) with the magic formula (7.68) for the infinitesimal propagator kernel, Feynman arrived at the following global kernel formula:

$$\boxed{\mathcal{K}(x,t;x_0,t_0) = \lim_{N\to\infty} \frac{1}{l} \int_{\mathbb{R}^{N-1}} \mathrm{e}^{\mathrm{i}\sum \Delta S/\hbar} \ \frac{dq_1}{l} \cdots \frac{dq_{N-1}}{l}} \tag{7.69}$$

with the discretized action

$$\sum \Delta S := \sum_{n=0}^{N-1} \left\{ \frac{m}{2} \left(\frac{q_{n+1}-q_n}{\Delta t} \right)^2 - U(q_n) \right\} \Delta t.$$

Here, we add the boundary conditions: $q_0 := x_0$ and $q_N := x$. The crucial Feynman formula (7.69) tells us that the global Feynman propagator kernel $\mathcal{K}(x,t;x_0,t_0)$ is obtained by summing over all possible time-ordered products of infinitesimal Feynman propagator kernels. This is a special case of the following general principle in natural philosophy:

In nature, interactions are obtained by the superposition of all possible infinitesimal interactions taking causality into account.

Introducing the path-integral notation, we briefly write

$$\int_{C\{t_0,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar} \mathcal{D}q := \lim_{N\to\infty} \frac{1}{l} \int_{\mathbb{R}^{N-1}} \mathrm{e}^{\mathrm{i}\sum \Delta S/\hbar} \ \frac{dq_1}{l} \cdots \frac{dq_{N-1}}{l}. \tag{7.70}$$

Physicists use the following two methods for computing path integrals:

(i) the limit formula (7.70) and
(ii) infinite-dimensional Gaussian integrals.

Method (i) corresponds to an approximation of continuous paths by polygons. Method (ii) generalizes the finite-dimensional formula

$$\int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle} \ \mathrm{e}^{\langle b|x\rangle} \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}} = \frac{\mathrm{e}^{\langle b|A^{-1}b\rangle}}{\sqrt{\det A}}$$

[45] The square root is to be understood as principal value: $l = \mathrm{e}^{\mathrm{i}\pi/4}\sqrt{\frac{2\pi\hbar\Delta t}{m}}$.

to infinite dimensions. In this context, one has to define the determinant $\det A$ of an infinite-dimensional operator A by generalizing the finite-dimensional formula

$$\det A = \prod_{n=1}^{N} \lambda_n$$

for the eigenvalues $\lambda_1, \ldots, \lambda_N$ of the operator A. Here, we will use the analytic continuation of the Riemann zeta function and its generalization to elliptic differential operators on Riemannian manifolds (see Sect. 7.9). Summarizing, we will get the following key formula:

$$\boxed{\mathcal{K}(x,t;x_0,t_0) = \int_{C\{t_0,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar} \mathcal{D}q = \mathcal{N} \int_{C\{t_0,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}_G q} \tag{7.71}$$

which is basic for modern physics. This formuula tells us that the Feynman path integral differs from the normalized infinite-dimensional Gaussian integral by a normalization factor $\mathcal{N}$. Fortunately enough, the explicit knowledge of the normalization factor $\mathcal{N}$ is not necessary in many applications to quantum field theory. In terms of mathematics, formula (7.71) connects different subjects of mathematics with each other: spectral theory of elliptic differential operators on Riemannian manifolds, harmonic analysis, analytic number theory, distributions and pseudo-differential operators, Fourier integral operators, random walks and stochastic processes (Brownian motion), topological quantum field theory (topological invariants of knots, manifolds and algebraic varieties). This concerns the following mathematical branches: analysis, differential geometry, algebraic topology, algebraic geometry, and theory of probability.

The innocently looking formula (7.71) *emphasizes the unity of mathematics.*

The WKB (Wentzel, Kramers, Brioullin) method. The passage from Maxwell's wave optics to geometric optics corresponds to the limit $\lambda \to 0$ (i.e., the wavelength λ goes to zero). Similarly, the passage from quantum mechanics to classical mechanics corresponds to the limit

$$\hbar \to 0$$

called the classical limit. More precisely, this means that quantum effects occur if the quotient $\hbar/S_{\mathrm{daily}}$ is sufficiently small. Here, S_{daily} is the action of processes in daily life. Explicitly, $\hbar \sim 10^{-34}$Js and $S_{\mathrm{daily}} \sim 1$Js. Shortly after Schrödinger's publication of his wave mechanics in 1926, Wentzel, Kramers, and Brioullin independently investigated the limit $\hbar \to 0$ parallel to geometric optics.[46] In terms of the Feynman path integral, the refined WKB method yields the following elegant key formula

$$\boxed{\mathcal{K}(x,t;x_0,t_0) = \mathrm{e}^{\mathrm{i}S[q_{\mathrm{class}}]/\hbar}\, \mathcal{K}_{\mathrm{fluct}}(x,t;x_0,t_0)} \tag{7.72}$$

[46] G. Wentzel, A generalization of the quantum condition in wave mechanics, Z. Physik **38** (1926), 518–529 (in German).
H. Kramers, Wave mechanics and half-integer quantization, Z. Physik **39** (1927), 828–840 (in German).
M. Brioullin, La méchanique ondulatoire de Schrödinger; une méthode générale de résolution par approximations successives, Comptes Rendus Acad. Sci. (Paris) **183** (1926), 24–44 (in French).

where $S[q_{\text{class}}]$ is the action along the classical path with the boundary condition $q_{\text{class}}(t_0) = x_0$ and $q_{\text{class}}(t) = x$. The factor $\mathcal{K}_{\text{fluct}}$ describes quantum fluctuations (see Sect. 7.10 on page 580).

Diffusion processes and the Euclidean strategy in quantum mechanics. The diffusion equation

$$\frac{\partial \psi(x,t)}{\partial t} = \kappa \psi_{xx} - V(x), \quad \psi(x,t_0) = \psi_0(x) \tag{7.73}$$

for all $x \in \mathbb{R}$ and all $t > t_0$ describes the diffusion of particles on the real line, where $\psi(x,t)$ denotes the particle density at the position x at time t, and $\kappa > 0$ is the diffusion constant. Using the replacement

$$t \Rightarrow \mathrm{i}t, \tag{7.74}$$

and setting $\kappa := \hbar/2m$, $U(x) := -\hbar V(x)$, the diffusion equation (7.73) passes over to the Schrödinger equation (7.61).[47] We expect that, by the replacement (7.74), each result on the classical diffusion equation (7.73) generates a result in quantum mechanics. This is called the Euclidean strategy. For example, let $V(x) \equiv 0$. We will show below that the classical diffusion kernel

$$\mathcal{P}(x,t;x_0,t_0) = \sqrt{\frac{m}{2\pi\hbar(t-t_0)}} \cdot \mathrm{e}^{-m(x-x_0)^2/2\hbar(t-t_0)} \tag{7.75}$$

passes over to the Feynman propagator kernel $\mathcal{K}(x,t;x_0,t_0) := \mathcal{P}(x,\mathrm{i}t;x_0,\mathrm{i}t_0)$. Explicitly,

$$\mathcal{K}(x,t;x_0,t_0) = \sqrt{\frac{m}{2\pi\mathrm{i}\hbar(t-t_0)}} \cdot \mathrm{e}^{\mathrm{i}m(x-x_0)^2/2\hbar(t-t_0)} \tag{7.76}$$

for all positions $x, x_0 \in \mathbb{R}$ and all times $t > t_0$.

Brownian motion. In 1905 Einstein studied the Brownian motion of tiny particles suspended in a liquid. This was the beginning of the theory of stochastic processes, which was developed as a mathematical theory by Wiener and Kolmogorov in the early 1920s and in the early 1930s, respectively.[48] Comparing the Schrödinger equation (7.61) with the diffusion equation (7.73), we arrive at the following intuitive interpretation of quantum mechanics emphasized by Feynman:

> *The motion of a quantum particle on the real line can be regarded as Brownian motion (i.e., a random walk) in imaginary time.*

This formal analogy motivated Mark Kac in 1949 to prove the famous Feynman–Kac formula[49] which represents the diffusion kernel (7.75) as a path integral, in rigorous mathematical terms see Sect. 7.11.5 on page 588.

Historical remarks on Feynman's discovery. The following quotation is taken from the comprehensive handbook on Feynman path integrals in quantum mechanics written by Christian Grosche and Frank Steiner:[50]

[47] Alternatively, if we regard $\psi(x,t)$ as the temperature at the point x at time t, then the equation (7.73) describes the heat conduction on the real line.

[48] Robert Brown (1773–1858), Einstein (1879–1955), Schrödinger (1887–1961), Wiener (1894–1964), Kolmogorov (1903–1987), Mark Kac (1914–1984), Feynman (1918–1988).

[49] M. Kac, On distributions of certain Wiener functionals, Trans. Amer. Math. Soc. **65** (1949), 1–13.

[50] C. Grosche and F. Steiner, Handbook of Feynman Path Integrals, Springer, Berlin, 1998 (reprinted with permission).

Feynman was working as a research assistant at Princeton during 1940–41. In the course of his graduate studies he discovered with Wheeler an action principle using half advanced and half retarded potentials.[51] The problem was the infinite self-energy of the electron, and it turned out that the new "action principle" could deal successfully with the infinity arising in the application of classical electrodynamics.

The problem then became one of applying this action principle to quantum mechanics in such a way that classical mechanics could arise naturally as a special case of quantum mechanics when the Planck quantum of action h was allowed to go to zero.

Feynman searched for any ideas which might have been previously worked out in connecting quantum-mechanical behavior with such classical ideas as the Lagrangian and Hamilton's action integral ... At a Princeton beer party Richard Feynman learned from Herbert Jehle, a former student of Schrödinger in Berlin, who had newly arrived from Europe, of Dirac's paper.[52] Dirac showed that

$$\langle q(t)|q(t_0)\rangle \qquad \text{corresponds to} \qquad \mathrm{e}^{\frac{\mathrm{i}}{\hbar}\int_{t_0}^{t}\mathcal{L}dt},$$

where $\mathcal{L}$ is the Lagrangian. The natural question that then arose was what Dirac had meant by the phrase "corresponds to." Feynman found that Dirac's statement actually means "proportionally to", that is,

$$\mathcal{K}(x+\Delta x, t+\Delta t; x, t) = \mathrm{const}(\Delta t)\cdot \mathrm{e}^{\mathrm{i}\Delta S/\hbar}.$$

Based on this result and the causality composition law (7.64) in the limit $N \to \infty$, Feynman interpreted the multiple-integral construction (7.70) as an "integral over all paths" and wrote this down in his Ph. D. thesis presented to the Faculty of Princeton University on May 4, 1942.[53] During the war Feynman worked at Los Alamos (New Mexico), and after the war his primary direction of work was towards quantum electrodynamics. So it happened that a complete theory of the *path integral approach to quantum mechanics* was worked out only in 1947. Feynman submitted his paper to the *Physical Review*, but the editors rejected it! Thus he rewrote it and sent it to *Reviews of Modern Physics*, where it finally appeared in spring 1948 under the title "Space-time approach to non-relativistic quantum mechanics."[54] Feynman's paper is one of the most beautiful and most influential papers in physics written during the last fifty years.[55]

[51] J. Wheeler and R. Feynman, Interaction with the absorber as the mechanism of radiation, Rev. Mod. Phys. **17** (1945), 157–181.

[52] P. Dirac, The Lagrangian in quantum mechanics, Soviet Union Journal of Physics (in German). Reprinted in J. Schwinger (Ed.) (1958), pp. 312–320.

[53] R. Feynman, The principle of least action in quantum mechanics, Ph.D. thesis, Princeton, New Jersey, 1942.

[54] Rev. Mod. Phys. **20** (1948), 367–387.

[55] Feynman's approach to quantum mechanics has a forerunner. In 1924 Wentzel published a paper where one can find the basic formulae and their interpretation as they were adopted twenty years later by Feynman. In fact, Wentzel's paper was published before the fundamental papers by Heisenberg (1925) and Schrödinger (1926). See G. Wentzel, Zur Quantenoptik (On quantum optics), Z. Physik **22** (1924), 193–199. This is discussed in: S. Antoci and D. Liebscher, The third way to quantum mechanics is the forgotten first, Annales de Fondation Louis de Broglie **21** (1996), 349–368 (see also S. Antoci and D. Liebscher, Wentzel's path integrals, Int. J. Math. Phys. **37** (1998), 531–535).

7.5.2 The Diffusion Kernel and the Euclidean Strategy in Quantum Physics

Formal motivation of the diffusion kernel. In order to discuss the basic idea of the Euclidean strategy in quantum mechanics, let us start with considering the classical diffusion equation

$$\psi_t(x,t) = \kappa\psi_{xx}(x,t), \qquad x \in \mathbb{R}, t > t_0, \quad \psi(x,t_0) = \psi(x) \tag{7.77}$$

where $\kappa := \hbar/2m$. We want to obtain the kernel $\mathcal{P}$ from (7.75), by using the Fourier method in a formal way. We start with the following two conditions

(C1) $P_t(x,t) = \kappa P_{xx}(x,t), x \in \mathbb{R}, t > 0$, and
(C2) $\lim_{t\to+0} P(x,t) = \delta(x)$, $x \in \mathbb{R}$.

Taking the existence of P for granted, set $\mathcal{P}(x,t;x_0,t_0) := P(x-x_0;t-t_0)$. We want to show that the function

$$\psi(x,t) := \int_{\mathbb{R}} \mathcal{P}(x,t;x_0,t_0)\psi_0(x_0)dx_0, \quad x \in \mathbb{R}, t > t_0$$

is a solution of (7.77). In fact, it follows from (C1) that

$$\psi_t(x,t) - \kappa\psi_{xx}(x,t) = \int_{\mathbb{R}} (\mathcal{P}_t - \mathcal{P}_{xx})\psi_0(x_0)dx_0 = 0, \quad x \in \mathbb{R}, t > 0.$$

By (C2), $\lim_{t\to t_0+0} \psi(x,t) = \int_{\mathbb{R}} \lim_{t\to t_0+0} \mathcal{P}(x,t;x_0,t_0)\psi_0(x_0)dx_0$, and hence

$$\lim_{t\to t_0+0} \psi(x,t) = \int_{\mathbb{R}} \delta(x-x_0)\psi_0(x_0)dx_0 = \psi_0(x).$$

It remains to determine the function P. Let $p \mapsto \hat{P}(p,t)$ be the Fourier transform of $x \mapsto P(x,t)$. By (C1) and (C2),

$$\hat{P}_t(p,t) = -\kappa p^2 \hat{P}(p,t), \quad t > 0, \qquad \hat{P}(p,0) = \frac{1}{\sqrt{2\pi}}.$$

Hence $\hat{P}(p,t) = \frac{1}{\sqrt{2\pi}}\, \mathrm{e}^{-\kappa p^2 t}$. By Fourier transform,

$$P(x,t) = \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}px}\, \mathrm{e}^{-\kappa p^2 t}\, dp, \qquad x \in \mathbb{R}, t > 0.$$

Hence $P(x,t) = \frac{1}{\sqrt{4\pi\kappa t}}\, \mathrm{e}^{-x^2/4\kappa t}$ (see the Gaussian integral (7.182) on page 560). This finishes the classical motivation for the diffusion kernel (7.75).

The classical existence theorem for the diffusion equation. The proof of the following standard result in the theory of partial differential equations can be found in H. Triebel, Higher Analysis, Sect. 42, Barth, Leipzig, 1989.

Theorem 7.15 *We are given the initial function $\psi_0 \in \mathcal{D}(\mathbb{R})$. Choose the kernel $\mathcal{P}$ as in (7.75). Then the function*

$$\psi(x,t) := \int_{\mathbb{R}} \mathcal{P}(x,t;x_0,t_0)\psi_0(x_0)\, dx_0, \quad x \in \mathbb{R}, t > t_0 \tag{7.78}$$

is a classical solution of the diffusion equation (7.77). In addition, we have the initial condition $\lim_{t\to t_0+0} \psi(x,t) = \psi_0(x)$ for all $x \in \mathbb{R}$.

The classical existence theorem for the free quantum particle on the real line. Consider the Schrödinger equation

$$\mathrm{i}\hbar\psi_t(x,t) = -\frac{\hbar^2}{2m}\psi_{xx}(x,t), \quad x \in \mathbb{R}, t > t_0, \quad \psi(x,t_0) = \psi_0(x) \tag{7.79}$$

for the motion of a free quantum particle of mass m on the real line. Let D denote the set of all Gaussian functions $\mathrm{e}^{-\beta(x-\alpha)^2}$, $x \in \mathbb{R}$ with real parameter α and positive parameter β. The complex linear hull, span D, is a dense subset of the Hilbert space $L_2(\mathbb{R})$.

Theorem 7.16 *We are given the initial function $\psi_0 \in \operatorname{span} D$. Choose the kernel $\mathcal{K}$ as in* (7.75)*. Then the function*

$$\psi(x,t) := \int_{\mathbb{R}} \mathcal{K}(x,t;x_0,t_0)\psi_0(x_0)\, dx_0, \qquad x \in \mathbb{R}, t > t_0 \tag{7.80}$$

is a classical solution of the Schrödinger equation (7.79)*. In addition, we have the initial condition $\lim_{t\to t_0+0} \psi(x,t) = \psi_0(x)$, in the sense of the convergence on the Hilbert space $L_2(\mathbb{R})$.*

The proof can be found in Zeidler (1995a), Sect. 5.22.2 (see the references on page 1049).

Formal perspectives. In the next sections, we will study the following topics in a formal manner:

- Propagator theory via the formal Dirac calculus (Sect. 7.5.3).
- Formal motivation of the definition of the Feynman path integral (Sect. 7.7.6).

Rigorous perspectives. Furthermore, we will rigorously investigate the following mathematical topics:

- Von Neumann's operator calculus and the functional-analytic approach to both the Feynman propagator and the Euclidean Feynman propagator (Sect. 7.6.3).
- Functional-analytic theory of the motion of a free quantum particle on the real line (Sect. 7.6.4).
- Functional-analytic theory of the motion of a harmonic oscillator on the real line and the Maslov index (Sect. 7.6.7).
- The Euclidean Feynman propagator and von Neumann's density matrix in quantum statistics (Sect. 7.6.8).
- Computation of the Feynman path integral for both the free quantum particle and the quantized harmonic oscillator (Sects. 7.7.3 and 7.7.4).
- The relation between infinite-dimensional Gaussian integrals and the Feynman propagator kernel including applications to the free quantum particle and the quantized harmonic oscillator (Sect. 7.9).
- The semi-classical WKB method (Sect. 7.10).
- Brownian motion, the Wiener integral, and the Feynman–Kac formula for diffusion processes (Sect. 7.11).

7.5.3 Probability Amplitudes and the Formal Propagator Theory

> Feynman's approach to quantum theory can be understood best by using Dirac's formal calculus; this can be generalized straightforward to quantum field theory.
>
> Folklore

The Parseval equation. Let $\varphi_1, \ldots, \varphi_N$ be an orthonormal basis of the complex N-dimensional Hilbert space Y. This means that the orthonormality condition

$$\langle \varphi_k | \varphi_l \rangle = \delta_{kl}, \qquad k, l = 1, \ldots, N$$

is satisfied. The basis property tells us that, for all $\varphi, \psi \in Y$, we have

(F) the Fourier expansion $|\psi\rangle = \sum_{k=1}^{N} |\varphi_k\rangle\langle\varphi_k|\psi\rangle$,
(C) the completeness relation $I = \sum_{k=1}^{N} |\varphi_k\rangle\langle\varphi_k|$, and[56]
(P) the Parseval equation

$$\langle \psi | \varphi \rangle = \sum_{k=1}^{n} \langle \psi | \varphi_k \rangle \langle \varphi_k | \varphi \rangle. \tag{7.81}$$

These classical properties of Fourier expansions are discussed in Sect. 7.10 of Vol. I. The complex numbers $c_1 := \langle \psi | \varphi_1 \rangle, \ldots, c_N := \langle \psi | \varphi_N \rangle$ are called the Fourier coefficients. Suppose that $||\psi|| = 1$. By the Parseval equation,

$$||\psi||^2 = \sum_{k=1}^{N} |c_k|^2 = 1.$$

If ψ is the state of a quantum particle, then $|c_k|^2$ is the probability for observing the particle in the state φ_k; the Fourier coefficients $c_1, \ldots, c_N$ are called the probability amplitudes of the particle state ψ.

The Schrödinger equation. Consider again the Schrödinger equation

$$\mathrm{i}\hbar\psi_t = -\frac{\hbar^2}{2m}\frac{\partial^2 \psi}{\partial x^2} + U\psi \tag{7.82}$$

for the motion of a quantum particle on the real line. Here, $m > 0$ is the mass of the particle. We assume that the smooth potential function $U : \mathbb{R} \to \mathbb{R}$ has compact support, that is, $U \in \mathcal{D}(\mathbb{R})$. In terms of physics, the potential U describes the force acting on the quantum particle. If $U \equiv 0$, then the quantum particle is called free. Set

$$H_0\varphi := -\frac{\hbar^2}{2m}\frac{\partial^2 \varphi}{\partial x^2} + U\varphi \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).$$

Then the operator $H_0 : \mathcal{D}(\mathbb{R}) \to L_2(\mathbb{R})$ is essentially self-adjoint on the Hilbert space $L_2(\mathbb{R})$. Let $H : W_2^2(\mathbb{R}) \to L_2(\mathbb{R})$ be the self-adjoint extension of H_0. Then the Schrödinger equation reads as

$$\mathrm{i}\hbar\dot{\psi}(t) = H\psi(t), \qquad t > t_0, \qquad \psi(t_0) = \psi_0$$

with the unique solution $\psi(t) = \mathrm{e}^{-\mathrm{i}H(t-t_0)/\hbar}\psi_0$ (see Theorem 7.25 on page 507).

The formal Dirac calculus. It is our goal to study the Schrödinger equation (7.82) by means of the formal Dirac calculus on the real line.[57] In particular, we will use

- the orthonormality condition $\langle x | x_0 \rangle = \delta(x - x_0)$ for all $x, x_0 \in \mathbb{R}$, and

[56] In mathematics, one also writes $\psi = \sum_{k=1}^{N} \langle \varphi_k | \psi \rangle \varphi_k$ and $I = \sum_{k=1}^{N} \varphi_k \otimes \varphi_k$.
[57] This formal calculus is thoroughly discussed in Sect. 11.2.5 of Vol. I. The rigorous justification of the Dirac calculus can be found in Sect. 12.2 of Vol. I.

- the completeness relation

$$\boxed{I = \int_{\mathbb{R}} |x\rangle\langle x| \, dx,} \tag{7.83}$$

where I denotes the unit operator. Using the trivial identity $\langle\psi|\varphi\rangle = \langle\psi|I\varphi\rangle$ and the completeness relation (7.83), we formally get the Parseval equation

$$\langle\psi|\varphi\rangle = \int_{\mathbb{R}} \langle\psi|x\rangle\langle x|\varphi\rangle dx. \tag{7.84}$$

This elegant formal argument is called Dirac's substitution trick.[58]

Formal operator kernel. The operator equation $\varphi = A\psi$ is equivalent to the integral relation

$$\langle x|\varphi\rangle = \int_{\mathbb{R}} \langle x|A|x_0\rangle\langle x_0|\psi\rangle \, dx_0, \qquad x \in \mathbb{R},$$

by using the completeness relation. Setting $\mathcal{A}(x, x_0) := \langle x|A|x_0\rangle$ for all positions $x, x_0 \in \mathbb{R}$, we get

$$\varphi(x) = \int_{\mathbb{R}} \mathcal{A}(x, x_0)\psi(x_0)dx_0, \qquad x, x_0 \in \mathbb{R}.$$

The function $(x, x_0) \mapsto \mathcal{A}(x, x_0)$ is called the kernel of the operator $\mathcal{A}$. In rigorous terms, this is not always a classical function. For example, the identical operator $A = I$ has the kernel

$$\mathcal{A}(x, x_0) = \langle x|x_0\rangle = \delta(x - x_0).$$

If we choose the Hamiltonian H, then the stationary Schrödinger equation

$$-\frac{\hbar^2}{2m}\psi''(x) + U(x)\psi(x) = \varphi(x), \qquad x \in \mathbb{R} \tag{7.85}$$

means that $\varphi = H\psi$. Formally, this is equivalent to the integral relation

$$\langle x|\varphi\rangle = \int_{\mathbb{R}} \langle x|H|x_0\rangle\langle x_0|\psi\rangle dx_0, \qquad x \in \mathbb{R}, \tag{7.86}$$

by using the completeness relation (7.83). Now we want to study the kernels $\mathcal{K}$ and $\mathcal{G}$ to the Feynman propagator $\mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}$ and the negative resolvent operator $(H - \mathcal{E}I)^{-1}$, respectively. Here, $\mathcal{K}$ and $\mathcal{G}$ is called the Feynman propagator kernel and the energetic Green's function, respectively.

> *In terms of modern mathematics, the Dirac calculus is a forerunner of the theory of pseudo-differential operators, where differential operators and integral operators are treated on equal footing.*

[58] Writing $\langle x|\varphi\rangle = \varphi(x)$ and $\langle\psi|x\rangle = \langle x|\psi\rangle^\dagger = \psi(x)^\dagger$, equation (7.84) reads as

$$\langle\psi|\varphi\rangle = \int_{\mathbb{R}} \psi(x)^\dagger\varphi(x)dx.$$

This is the inner product on the Hilbert space $L_2(\mathbb{R})$.

We refer to the treatise by L. Hörmander, The Analysis of Linear Partial Differential Operators, Vols. 1–4, Springer, New York, 1983.

The key formulas. The Feynman approach to quantum physics is based on the following formal arguments.

(i) The Feynman propagator kernel $\mathcal{K}$: For all positions $x, x_0 \in \mathbb{R}$ and all times $t > t_0$, we define the Feynman propagator kernel

$$\boxed{\mathcal{K}(x,t;x_0,t_0) := \langle x|\mathrm{e}^{-\mathrm{i}H(t-t_0)/\hbar}|x_0\rangle.} \tag{7.87}$$

- Integral representation for the dynamics of the quantum particle: For the solution $\psi(t) = \mathrm{e}^{-\mathrm{i}H(t-t_0)/\hbar}\psi_0$ of the Schrödinger equation (7.82), we have

$$\boxed{\psi(x,t) = \int_{\mathbb{R}} \mathcal{K}(x,t;x_0,t_0)\psi(x_0,t_0)dx_0, \quad x \in \mathbb{R},\ t > t_0.} \tag{7.88}$$

- Schrödinger equation for the Feynman propagator kernel: For all positions $x, x_0 \in \mathbb{R}$ and all times $t > t_0$,

$$\mathrm{i}\hbar\mathcal{K}_t(x,t;x_0,t_0) = -\frac{\hbar^2}{2m}\mathcal{K}_{xx}(x,t;x_0,t_0) + U(x)\mathcal{K}(x,t;x_0,t_0).$$

- Singularity at the initial time t_0 :

$$\lim_{t\to t_0+0} \mathcal{K}(x,t;x_0,t_0) = \delta(x-x_0), \qquad x, x_0 \in \mathbb{R}. \tag{7.89}$$

- Causality relation: For all positions $x, x_0 \in \mathbb{R}$ and all times $t > \tau > t_0$,

$$\mathcal{K}(x,t;x_0,t_0) = \int_{\mathbb{R}} \mathcal{K}(x,t;y,\tau)\mathcal{K}(y,\tau;x_0,t_0)\ dy. \tag{7.90}$$

This is the product rule for the Feynman propagator kernel.

Formal proof. By the completeness relation $\int_{\mathbb{R}} |x_0\rangle\langle x_0|\ dx_0 = I$,

$$\langle x|\psi\rangle = \langle x|\mathrm{e}^{-\mathrm{i}H(t-t_0)/\hbar}|\psi_0\rangle = \int_{\mathbb{R}} \langle x|\mathrm{e}^{-\mathrm{i}H(t-t_0)/\hbar}|x_0\rangle\langle x_0|\psi_0\rangle dx_0.$$

This is (7.88). The differential equation for $\mathcal{K}$ follows from the fact that the two expressions

$$\mathrm{i}\hbar\psi_t(x,t) = \int_{\mathbb{R}} \mathrm{i}\hbar\mathcal{K}_t(x,t;x_0,t_0)\psi_0(x_0)\ dx_0,$$
$$H\psi(x,t) = \int_{\mathbb{R}} H\mathcal{K}(x,t;x_0,t_0)\psi_0(x_0)\ dx_0$$

are equal to each other for all initial functions ψ_0. Hence $\mathrm{i}\hbar\mathcal{K}_t = H\mathcal{K}$. Furthermore,

$$\lim_{t\to t_0+0}\langle x|\mathrm{e}^{-\mathrm{i}H(t-t_0)}|x_0\rangle = \langle x|x_0\rangle = \delta(x-x_0).$$

From the group property $\mathrm{e}^{u+v} = \mathrm{e}^u\mathrm{e}^v,\ u, v \in \mathbb{C}$ of the exponential function, we get

$$e^{-iH(t-t_0)/\hbar} = e^{-iH(t-\tau)\hbar} e^{-iH(\tau-t_0)/\hbar}, \qquad t_0 < \tau < t.$$

This implies

$$\langle x|e^{-iH(t-t_0)/\hbar}|x_0\rangle = \int_{\mathbb{R}} \langle x|e^{-iH(t-\tau)/\hbar}|y\rangle\langle y|e^{-iH(\tau-t_0)/\hbar}|x_0\rangle\, dy, \quad (7.91)$$

which is the causality relation (7.90). □

(ii) The resolvent kernel $\mathcal{R}$: Let $\varrho(H)$ be the resolvent set of the Hamiltonian H on the Hilbert space $L_2(\mathbb{R})$. By definition, the complex number $\mathcal{E}$ is contained in $\varrho(H)$ iff the inverse operator

$$(\mathcal{E}I - H)^{-1} : L_2(\mathbb{R}) \to L_2(\mathbb{R})$$

exists, and it is continuous. This operator is called the resolvent[59] of the Hamiltonian H at the point $\mathcal{E}$. We briefly write $R_\mathcal{E} := (\mathcal{E}I - H)^{-1}$. The complement $\sigma(E) := \mathbb{C} \setminus \varrho(H)$ is called the spectrum of H.

The spectrum $\sigma(H)$ is a closed subset of the real line; the complementary resolvent set $\varrho(H)$ is an open subset of the complex plane.

The points E in the spectrum $\sigma(H)$ are the energy values of the quantum particle described by the Hamiltonian H. For all positions $x, x_0 \in \mathbb{R}$ and all complex numbers $\mathcal{E} \in \varrho(H)$, we define the resolvent kernel

$$\boxed{\mathcal{R}(x, x_0; \mathcal{E}) := \langle x|(\mathcal{E}I - H)^{-1}|x_0\rangle.}$$

This kernel has the following properties.

- Integral representation of the resolvent: For each given complex number $\mathcal{E} \in \varrho(H)$, the equation

$$(\mathcal{E}I - H)\psi = \chi$$

has the unique solution $\psi = (\mathcal{E}I - H)^{-1}\chi$. This is equivalent to the integral relation

$$\psi(x) = \int_{\mathbb{R}} \mathcal{R}(x, x_0; \mathcal{E})\chi(x_0)dx_0, \qquad x \in \mathbb{R}. \quad (7.92)$$

This follows from

$$\langle x|\psi\rangle = \langle x|(\mathcal{E}I - H)^{-1}\varphi\rangle = \int_{\mathbb{R}} \langle x|(\mathcal{E}I - H)^{-1}|x_0\rangle\langle x_0|\varphi\rangle\, dx_0.$$

- The resolvent equation: For all $\mathcal{E}, \mathcal{E}' \in \varrho(H)$, we have Hilbert's resolvent equation $R_\mathcal{E} - R_{\mathcal{E}'} = (\mathcal{E}' - \mathcal{E})R_{\mathcal{E}'}R_\mathcal{E}$. This implies

$$\mathcal{R}(x, x_0; \mathcal{E}) - \mathcal{R}(x, x_0; \mathcal{E}') = (\mathcal{E}' - \mathcal{E}) \int_{\mathbb{R}} \mathcal{R}(x, y; \mathcal{E}')\mathcal{R}(y, x_0; \mathcal{E})\, dy.$$

In fact, Hilbert's resolvent equation implies

$$\langle x|R_\mathcal{E} x_0\rangle - \langle x|R_{\mathcal{E}'} x_0\rangle = (\mathcal{E}' - \mathcal{E}) \int_{\mathbb{R}} \langle x|R_{\mathcal{E}'} y\rangle\langle y|R_\mathcal{E} x_0\rangle\, dy.$$

[59] Physicists frequently use the negative resolvent operator $-(\mathcal{E}I - H)^{-1}$ which is equal to $(H - \mathcal{E}I)^{-1}$.

(iii) The energetic Green's function $\mathcal{G}$: For all positions $x, x_0 \in \mathbb{R}$ and all complex numbers $\mathcal{E} \in \varrho(H)$, we define

$$\boxed{\mathcal{G}(x, x_0; \mathcal{E}) := -\mathcal{R}(x, x_0; \mathcal{E}).}$$

For each complex number $\mathcal{E} \in \varrho(H)$, the inhomogeneous stationary equation $(H - \mathcal{E}I)\psi = \varphi$, that is,

$$-\frac{\hbar^2}{2m}\psi''(x) + (U(x) - \mathcal{E})\psi(x) = \varphi(x), \qquad x \in \mathbb{R},$$

has the solution $\psi = -(\mathcal{E}I - H)^{-1}\varphi$. By (7.92),

$$\psi(x) = \int_{\mathbb{R}} \mathcal{G}(x, y; \mathcal{E})\varphi(y)dy, \qquad x \in \mathbb{R}.$$

Choosing $\varphi(x) := \delta(x - x_0)$, we obtain $\psi(x) = \mathcal{G}(x, x_0; \mathcal{E})$. This implies that, for all $\mathcal{E} \in \varrho(H)$, we get

$$-\frac{\hbar^2}{2m}\mathcal{G}_{xx}(x, x_0; \mathcal{E}) + (U(x) - \mathcal{E})\mathcal{G}(x, x_0; \mathcal{E}) = \delta(x - x_0), \qquad x, x_0 \in \mathbb{R}.$$

Therefore, the function $(x, x_0) \mapsto \mathcal{G}(x, x_0; \mathcal{E})$ is called the energetic Green's function (or the energetic 2-point function) with respect to the complex number $\mathcal{E} \notin \sigma(H)$. Now let us show that the energetic Green's function has singularities at the spectral points $\mathcal{E} \in \sigma(H)$, which correspond to the physical energy values of the quantum particle described by the Hamiltonian H.

(iv) The energetic Fourier transform: Let $\{|E_k\rangle\}_{k\in\mathcal{N}}$ be the complete orthonormal system of (generalized) eigenstates of the Hamiltonian H with the index set $\mathcal{N}$. That is, we have

- the (generalized) eigenvalue equation $H|E_k\rangle = E_k|E_k\rangle$,
- the completeness relation $\int_{\mathcal{N}} |E_k\rangle\langle E_k|\, d\mu(k) = I$, and
- the orthonormality relation $\langle E_k|E_l\rangle = \delta_\mu(k - l)$ for all $k, l \in \mathcal{N}$.

Here, μ is a measure on the set $\mathcal{N}$. This measure is called the energy measure of the Hamiltonian H. The Dirac delta function δ_μ with respect to the measure μ has the characteristic property that[60]

$$\int_{\mathcal{N}} \delta_\mu(k - k_0) f(k)\, d\mu(k) = f(k_0).$$

Thus, the Dirac delta function δ_μ generalizes the Kronecker symbol. Now let us assign to each energy state $|E_k\rangle$ the so-called energy function $\chi_k(x) := \langle x|E_k\rangle$ for all $x \in \mathbb{R}$.

- The Fourier–Stieltjes transform:

$$\boxed{\hat{\psi}(k) = \int_{\mathbb{R}} \chi_k(x)^\dagger \psi(x)dx, \qquad k \in \mathcal{N}.} \tag{7.93}$$

- The inverse Fourier–Stieltjes transform:

$$\boxed{\psi(x) = \int_{\mathcal{N}} \chi_k(x)\hat{\psi}(k)d\mu(k), \qquad x \in \mathbb{R}.} \tag{7.94}$$

[60] Mnemonically, this follows from $\int_{\mathcal{N}} \langle E_{k_0}|E_k\rangle\langle E_k|f\rangle d\mu(k) = \langle E_{k_0}|f\rangle$, by using the completeness relation.

- The stationary Schrödinger equation: For all indices $k \in \mathcal{N}$,

$$-\frac{\hbar^2}{2m}\chi_k''(x) + U(x)\chi_k(x) = E_k\chi_k(x), \qquad x \in \mathbb{R}. \tag{7.95}$$

  This tells us that the function χ_k is an eigenfunction corresponding to the energy eigenvalue E_k.
- The function $\psi_k(x,t) := \mathrm{e}^{-\mathrm{i}E_k t/\hbar}\chi_k(x)$ satisfies the instationary Schrödinger equation:

$$\mathrm{i}\hbar\frac{\partial\psi_k(x,t)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\psi_k(x,t)}{\partial x^2} + U(x)\psi_k(x,t) = E_k\psi(x,t), \quad x,t \in \mathbb{R}.$$

Formal proof. Ad (7.93). By the completeness relation,

$$\langle E_k|\psi\rangle = \int_R \langle E_k|x\rangle\langle x|\psi\rangle\, dx.$$

Ad (7.94). Similarly, $\langle x|\psi\rangle = \int_{\mathcal{N}}\langle x|E_k\rangle\langle E_k|\psi\rangle d\mu(k)$.
Ad (7.95). From $H|E_k\rangle = E_k|E_k\rangle$, we get

$$\langle x|E_k\rangle = \langle x|H|E_k\rangle = \int_{\mathbb{R}}\langle x|H|x_0\rangle\langle x_0|E_k\rangle dx_0.$$

Now use the formal equivalence between (7.85) and (7.86). □

(v) The energetic representation of the Feynman propagator kernel: For all positions $x, x_0 \in \mathbb{R}$ and all times $t > t_0$, we have

$$\boxed{\mathcal{K}(x,t;x_0,t_0) = \int_{\mathcal{N}} \mathrm{e}^{-\mathrm{i}E_k(t-t_0)/\hbar}\chi_k(x)\chi_k(x_0)^\dagger d\mu(k)} \tag{7.96}$$

and

$$\boxed{\mathcal{G}(x,x_0;E+\mathrm{i}\varepsilon) = \int_{\mathcal{N}} \frac{\chi_k(x)\chi_k(x_0)^\dagger}{E_k - E - \mathrm{i}\varepsilon}\, d\mu(k).} \tag{7.97}$$

Formal proof. Ad (7.96). To simplify notation, we set $\hbar := 1$ and $t_0 := 0$. By the completeness relation,

$$\langle x|\mathrm{e}^{-\mathrm{i}tH}|x_0\rangle = \int_{\mathcal{N}}\langle x|E_k\rangle\langle E_k|\mathrm{e}^{-\mathrm{i}tH}|x_0\rangle d\mu(k).$$

Moreover, $\mathrm{e}^{-\mathrm{i}tH}|E_k\rangle = \mathrm{e}^{-\mathrm{i}tE_k}|E_k\rangle$. Hence

$$\langle E_k|\mathrm{e}^{-\mathrm{i}tH}|x_0\rangle = \langle x_0|\mathrm{e}^{\mathrm{i}tH}|E_k\rangle^\dagger = \mathrm{e}^{-\mathrm{i}tE_k t}\langle x_0|E_k\rangle^\dagger = \mathrm{e}^{-\mathrm{i}tE_k t}\chi_k(x_0)^\dagger.$$

Ad (7.97). Replace $\mathrm{e}^{-\mathrm{i}tH}$ by $(H - (E+\mathrm{i}\varepsilon)I)^{-1}$. □

(vi) The passage from time to energy: For all positions $x, x_0 \in \mathbb{R}$, all times $t > t_0$, all energies $E \in \mathbb{R}$, and all energy damping parameters $\varepsilon > 0$, the following transformation formulas are valid.

- The Fourier–Laplace transform of the Feynman propagator kernel:

$$\boxed{\mathcal{G}(x,x_0;E+\mathrm{i}\varepsilon) = \frac{\mathrm{i}}{\hbar}\int_{t_0}^{\infty} \mathrm{e}^{\mathrm{i}(E+\mathrm{i}\varepsilon)(t-t_0)/\hbar}\, \mathcal{K}(x,t;x_0,t_0)dt.}$$

- The Fourier–Laplace transform of the energetic Green's function:

$$\boxed{\mathcal{K}(x,t;x_0,t_0) = \frac{1}{2\pi\mathrm{i}} \cdot PV \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}(E+\mathrm{i}\varepsilon)(t-t_0)/\hbar}\, \mathcal{G}(x,x_0;E+\mathrm{i}\varepsilon)\, dE.}$$

Recall that the symbol $PV\int_{-\infty}^{\infty}\ldots$ stands for the limit $\lim_{R\to+\infty}\int_{-R}^{R}\ldots$ (principal value of the integral).

Formal proof. This follows immediately from (7.96) and (7.97) combined with the two classical formulas

$$\frac{\mathrm{i}}{\hbar}\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}(E+\mathrm{i}\varepsilon)(t-t_0)/\hbar}\mathrm{e}^{-\mathrm{i}E_k(t-t_0)/\hbar}\theta(t-t_0)dt = \frac{1}{E_k - E - \mathrm{i}\varepsilon}$$

and

$$\theta(t-t_0)\mathrm{e}^{-\mathrm{i}E_k(t-t_0)/\hbar} = \frac{1}{2\pi\mathrm{i}} \cdot PV \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-\mathrm{i}(E+\mathrm{i}\varepsilon)(t-t_0)/\hbar}}{E_k - E - \mathrm{i}\varepsilon}\, dE,$$

which are valid for the following quantities: all energies $E, E_k \in \mathbb{R}$, all times $t, t_0 \in \mathbb{R}$ with $t \neq t_0$, and all damping parameters $\varepsilon > 0$. The proof of the latter two formulas can be found in Problem 7.35. □

The preceding formal propagator theory is very convenient from the mnemonical point of view. Our next goal is to show how this formal approach can be translated into a rigorous approach. To this end, we will use

- the von Neumann operator calculus in Hilbert spaces,
- tempered distributions, Gelfand triplets, and the theory of generalized eigenfunctions, and
- tempered distributions and the Schwartz kernel theorem.

We will apply this to:

- the free quantum particle (Sect. 7.6.4),
- the harmonic oscillator (Sect. 7.6.7), and
- ideal gases (Sect. 7.6.8).

7.6 Von Neumann's Rigorous Approach

> Rigorous propagator theory is based on von Neumann's operator calculus for functions of self-adjoint operators.
>
> Folklore

As a preparation for the rigorous propagator theory to be considered in the next section, let us summarize von Neumann's operator calculus. In this section, we consider an arbitrary complex separable Hilbert space X of finite or infinite dimension. The inner product on X is denoted by $\langle\psi|\varphi\rangle$ for all $\varphi, \psi \in X$. For fixed initial time t_0, the given function $\psi : [t_0, \infty[\to X$ with values in the Hilbert space X is called continuously differentiable iff the following is met:

- For all $t > t_0$, the derivative $\dot{\psi}(t) := \lim_{h\to 0} h^{-1}(\psi(t+h) - \psi(t))$ exists (in the sense of the convergence on the Hilbert space X).
- The function $t \mapsto \psi(t)$ is continuous on the closed interval $[0, \infty[$.
- The function $t \mapsto \dot{\psi}(t)$ is continuous on the open interval $]t_0, \infty[$, and the limit $\lim_{t\to t_0+0} \dot{\psi}(t)$ exists.

It is our goal to construct continuously differentiable solutions of the Schrödinger equation $\mathrm{i}\hbar\dot{\psi} = H\psi$ in the form $\psi(t) = \mathrm{e}^{-\mathrm{i}tH/\hbar}\psi_0$. To this end, we need the construction of the operator $\mathrm{e}^{-\mathrm{i}tH/\hbar}$.

7.6.1 The Prototype of the Operator Calculus

The basic idea is to use a complete orthonormal system $\varphi_0, \varphi_1, \ldots$ in the infinite-dimensional Hilbert space X.[61] The two key formulas are given by the series expansions

$$\boxed{H\varphi = \sum_{k=0}^{\infty} E_k \cdot \langle \varphi_k | \varphi \rangle \varphi_k \qquad \text{for all} \quad \varphi \in D(H)} \tag{7.98}$$

and

$$\boxed{\mathsf{F}(H)\varphi = \sum_{k=0}^{\infty} F(E_k) \cdot \langle \varphi_k | \varphi \rangle \varphi_k \qquad \text{for all} \quad \varphi \in D.} \tag{7.99}$$

To discuss this, observe first that

- the infinite series $\sum_{k=0}^{\infty} a_k \varphi_k$ with complex numbers $a_0, a_1, a_2, \ldots$ is convergent iff
- $\sum_{k=0}^{\infty} |a_k|^2 < \infty$.

In particular, the completeness of $\varphi_0, \varphi_1, \ldots$ guarantees that

$$\varphi = \sum_{k=0}^{\infty} \langle \varphi | \varphi_k \rangle \, \varphi_k \qquad \text{for all} \quad \varphi \in X.$$

(i) The operator H: We are given the real numbers $E_0, E_1, \ldots$ We define

$$H\varphi_k := E_k \varphi_k, \qquad k = 0, 1, \ldots$$

In a natural way, we want to extend the operator H to a linear subspace $D(H)$ of X. To this end, we define

$$D(H) := \{\varphi \in X : \sum_{k=0}^{\infty} |E_k|^2 |\langle \varphi | \varphi_k \rangle|^2 < \infty\}.$$

In other words, we have $\varphi \in D(H)$ iff the infinite series from (7.98) is convergent in X. Now, for all $\varphi \in D(H)$, we define $H\varphi$ by the convergent series (7.98). In particular, $\varphi_k \in D(H)$ for all k.

The operator $H : D(H) \to X$ is self-adjoint.

The spectrum $\sigma(H)$ of H is the closure of the set $\{E_0, E_1, \ldots\}$. The resolvent set $\varrho(H)$ of the operator H is the largest open subset of the complex plane which does not contain the energy values $E_0, E_1, \ldots$

(ii) The operator $\mathsf{F}(H) : D \to X$: We are given the function $F : \mathbb{R} \to \mathbb{C}$. Let D be the set of all elements φ of X such that the series (7.99) is convergent. Explicitly,

$$D := \{\varphi \in X : \sum_{k=0}^{\infty} |F(E_k)|^2 |\langle \varphi_k | \varphi \rangle|^2 < \infty\}.$$

Finally, for any $\varphi \in D$, define $\mathsf{F}(H)\varphi$ by the convergent series (7.99). The operator $\mathsf{F}(H) : D \to X$ is self-adjoint if the function F is real-valued.

[61] If the Hilbert space X is finite-dimensional with dimension $N = 1, 2, \ldots$, then all of the following formulas remain valid if we replace the symbol $\sum_{k=0}^{\infty}$ by $\sum_{k=0}^{N-1}$.

(iii) The spectral family $\{\mathsf{E}_\lambda(H)\}_{\lambda\in\mathbb{R}}$ of the self-adjoint operator H. Fix the real number λ and consider the function

$$e_\lambda(E) := \begin{cases} 1 & \text{if} \quad E < \lambda, \\ 0 & \text{if} \quad E \geq \lambda. \end{cases} \tag{7.100}$$

In other words, e_λ is the characteristic function of the open interval $]-\infty, \lambda[$. Define

$$\mathsf{E}_\lambda(H)\varphi := \sum_{k=1}^{\infty} e_\lambda(E_k)\langle\varphi_k|\varphi\rangle\varphi_k.$$

This series is convergent for all $\varphi \in X$. The operator $\mathsf{E}_\lambda(H) : X \to X$ is the orthogonal projection onto the closed linear subspace spanned by all the eigenvectors φ_k with $E_k \in]-\infty, \lambda[$.

(iv) The propagator $\mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}$: Let $t, t_0 \in \mathbb{R}$. Since $|\mathrm{e}^{-\mathrm{i}(t-t_0)/\hbar}| \leq 1$, the operator

$$\mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}\varphi := \sum_{k=0}^{\infty} \mathrm{e}^{-\mathrm{i}(t-t_0)E_k/\hbar}\langle\varphi_k|\varphi\rangle\varphi_k$$

is defined for all $\varphi \in X$. In addition, the operator $\mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar} : X \to X$ is unitary. For given $\psi_0 \in D(H)$, set

$$\psi(t) := \mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}\psi_0 \qquad \text{for all} \quad t \in \mathbb{R}.$$

Then the function $\psi : \mathbb{R} \to X$ is continuously differentiable, and it is a solution of the Schrödinger equation.

$$\mathrm{i}\hbar\dot{\psi}(t) = H\psi(t), \qquad t \in \mathbb{R}, \qquad \psi(t_0) = \psi_0.$$

Proof. First use formal differentiation. This yields

$$\mathrm{i}\hbar\dot{\psi}(t) = \sum_{k=0}^{\infty} E_k \mathrm{e}^{-\mathrm{i}(t-t_0)E_k/\hbar}\langle\varphi_k|\varphi\rangle\varphi_k = H\psi(t).$$

Since we have the convergent majorant series

$$\sum_{k=0}^{\infty} |E_k \mathrm{e}^{-\mathrm{i}(t-t_0)E_k/\hbar}\langle\varphi_k|\varphi\rangle|^2 \leq \sum_{k=0}^{\infty} |E_k|^2 |\langle\varphi_k|\varphi\rangle|^2 < \infty,$$

the formal differentiation can be rigorously justified in the same way as for classical infinite series (see Sect. 5.8, Zeidler (1995a), quoted on page 1049). □

(v) The Euclidean propagator $\mathrm{e}^{-(t-t_0)H/\hbar}$: Suppose that $E_k \geq 0$ for all k. Fix the real number t_0. Let $t \geq t_0$. Since $0 \leq \mathrm{e}^{-(t-t_0)E_k/\hbar} \leq 1$, the operator

$$\mathrm{e}^{-(t-t_0)H/\hbar}\varphi := \sum_{k=0}^{\infty} \mathrm{e}^{-(t-t_0)E_k/\hbar}\langle\varphi_k|\varphi\rangle\varphi_k$$

is defined for all $\varphi \in X$. We have $||\mathrm{e}^{-(t-t_0)H/\hbar}\varphi|| \leq ||\varphi||$ for all $\varphi \in X$, that is, the operator $\mathrm{e}^{-(t-t_0)H/\hbar} : X \to X$ is non-expansive.[62] For given $\psi_0 \in D(H)$, set

[62] Note that $||\mathrm{e}^{-(t-t_0)H/\hbar}\varphi||^2 = \sum_{k=0}^{\infty} |\mathrm{e}^{-(t-t_0)E_k/\hbar}\langle\varphi_k|\varphi\rangle|^2$. Thus, for all $t \geq t_0$,

$$||\mathrm{e}^{-(t-t_0)H/\hbar}\varphi||^2 \leq \sum_{k=0}^{\infty} |\langle\varphi_k|\varphi\rangle|^2 = ||\varphi||^2.$$

$$\psi(t) := \mathrm{e}^{-(t-t_0)H/\hbar}\psi_0 \qquad \text{for all} \quad t \geq t_0.$$

Then the function $\psi : \mathbb{R} \to X$ is continuously differentiable on $[t_0, \infty[$, and it is a solution of the Euclidean Schrödinger equation

$$\hbar\dot{\psi}(t) = -H\psi(t), \qquad t \geq t_0, \qquad \psi(t_0) = \psi_0.$$

(vi) The resolvent $(\mathcal{E}I - H)^{-1}$: Let $\mathcal{E}$ be a non-real complex number. The series

$$R_{\mathcal{E}}\varphi := \sum_{k=0}^{\infty} \frac{\langle\varphi_k|\varphi\rangle}{\mathcal{E} - E_k}\,\varphi_k$$

is convergent for all $\varphi \in X$. This follows from

$$\left|\frac{1}{\mathcal{E} - E_k}\right|^2 = \frac{1}{(\Im\mathcal{E})^2 + (E_k - \Re\mathcal{E})^2} \leq \frac{1}{(\Im\mathcal{E})^2}.$$

Hence $||R_{\mathcal{E}}||^2 \leq \mathrm{const}(\mathcal{E})\cdot||\varphi||^2$. Thus, the operator $R_{\mathcal{E}}$ is linear and continuous. In addition, it can be easily shown that $R_{\mathcal{E}} = (\mathcal{E}I - H)^{-1}$.

(vii) The Fourier–Laplace transform of the propagator from time to energy: The integral

$$\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\mathcal{E}t} f(t)dt$$

does not exist (in the classical sense) if $\mathcal{E}$ is a real number and $f(t) \equiv 1$. However, if we choose both the complex energy $\mathcal{E} := E + \mathrm{i}\varepsilon$ (with $\varepsilon > 0$) and the truncation function $f(t) := \theta(t - t_0)$, then the integral[63]

$$\int_{t_0}^{\infty} \mathrm{e}^{\mathrm{i}Et}\mathrm{e}^{-\varepsilon t}dt$$

exists because of the damping factor $\mathrm{e}^{-\varepsilon t}$. This is the basic idea behind the use of both truncated propagators and complex energies in quantum physics. In order to explain this, choose the linear self-adjoint operator $H : D(H) \to X$ as in (i) above. Let t, t_0 be arbitrary real time parameters, and let $\mathcal{E}$ be a non-real complex parameter called energy. It is convenient to introduce the following operators, which we will frequently encounter in this treatise:

- $P(t, t_0) := \mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}$ (propagator),
- $P^+(t, t_0) := \theta(t - t_0)P(t, t_0)$ (retarded propagator or Feynman propagator),
- $P^-(t, t_0) := -\theta(t_0 - t)P(t, t_0)$ (advanced propagator),
- $G(\mathcal{E}) := (H - \mathcal{E}I)^{-1}$ (Green's operator),[64]
- $G^+(\mathcal{E}) := G(\mathcal{E})$ if $\Im(\mathcal{E}) > 0$ (retarded!Green's operator),
- $G^-(\mathcal{E}) := G(\mathcal{E})$ if $\Im(\mathcal{E}) < 0$ (advanced!Green's operator).

Proposition 7.17 *Let $t, t_0 \in \mathbb{R}$ and $\varphi, \chi \in X$. Then:*
(i) *For all energies $\mathcal{E}$ in the open upper half-plane (i.e., $\Im(\mathcal{E}) > 0$), we have the Fourier–Laplace transformation*

[63] Recall that $\theta(t - t_0) = 1$ if $t \geq t_0$ and $\theta(t - t_0) = 0$ if $t < t_0$ (Heaviside function).
[64] Since the operator $G(\mathcal{E})$ depends on the choice of the complex energy $\mathcal{E}$, we also call it the energetic Green's operator.

$$\langle \chi | G^+(\mathcal{E})\varphi \rangle = \frac{\mathrm{i}}{\hbar} \int_{\mathbb{R}} e^{\mathrm{i}\mathcal{E}(t-t_0)/\hbar} \langle \chi | P^+(t,t_0)\varphi \rangle \, dt \tag{7.101}$$

together with the inverse transformation

$$\langle \chi | P^+(t,t_0)\varphi \rangle = \frac{1}{2\pi \mathrm{i}} \cdot PV \int_{\mathbb{R}} e^{-\mathrm{i}\mathcal{E}(t-t_0)/\hbar} \langle \chi | G^+(\mathcal{E})\varphi \rangle \, d\Re(\mathcal{E})$$

where we assume that $t \neq t_0$.
(ii) *For all energies* $\mathcal{E}$ *in the open lower half-plane (i.e.,* $\Im(\mathcal{E}) < 0$*), we have the Fourier–Laplace transformation*

$$\langle \chi | G^-(\mathcal{E})\varphi \rangle = \frac{\mathrm{i}}{\hbar} \int_{\mathbb{R}} e^{\mathrm{i}\mathcal{E}(t-t_0)/\hbar} \langle \chi | P^-(t,t_0)\varphi \rangle \, dt \tag{7.102}$$

together with the inverse transformation

$$\langle \chi | P^-(t,t_0)\varphi \rangle = \frac{1}{2\pi \mathrm{i}} \cdot PV \int_{\mathbb{R}} e^{-\mathrm{i}\mathcal{E}(t-t_0)/\hbar} \langle \chi | G^-(\mathcal{E})\varphi \rangle \, d\Re(\mathcal{E})$$

where we assume that $t \neq t_0$.

Complete proofs for this prototype of operator calculus including the statements above can be found in Zeidler (1995a), Chap. 5 (see the references on page 1049). For the proof of Prop. 7.17 above, we refer to Problem 7.36. The Fourier–Laplace transform is also briefly called the Laplace transform.[65]

Interestingly enough, both retarded (i.e., causal) propagators and advanced (i.e., non-causal) propagators play a crucial role in quantum field theory.

From the mathematical point of view, the reason is that the relevant perturbation theory depends on quantities which are constructed by using both retarded and advanced propagators. Physicists interpret this by saying that

- the interaction between elementary particles is governed by virtual particles (which are graphically represented by the internal lines of the Feynman diagrams), and
- the virtual particles violate basic laws of physics (e.g., the relation between energy and momentum or causality).

7.6.2 The General Operator Calculus

> The observation which comes closest to an explanation of the mathematical concepts cropping up in physics which I know is Einstein's statement that the only physical theories which we are willing to accept are the beautiful ones. It stands to argue that the concepts of mathematics, which invite the exercise of much a wit, have the quality of beauty.[66]
>
> Eugene Wigner, 1959

[65] Laplace (1749–1827), Fourier (1768–1830).

[66] E. Wigner, The unreasonable effectiveness of mathematics in the natural sciences, Richard Courant Lecture in Mathematical Sciences delivered at New York University, May 11, 1959. In: E. Wigner, Philosophical Reflections and Syntheses. Annotated by G. Emch. Edited by J. Mehra and A. Wightman, Springer, New York, 1995, pp. 534–549.

Let X be a complex separable finite-dimensional or infinite-dimensional Hilbert space. We make the following assumption:

(A) *The linear operator $H : D(H) \to X$ is self-adjoint.*

This includes tacitly that the domain of definition $D(H)$ is a linear dense subspace of X. Von Neumann's famous spectral theorem tells us the following.

Theorem 7.18 *For each pair $\varphi \in D(H), \chi \in X$, there exists a (complex-valued) measure $\mu_{\chi,\varphi}$ on the real line such that*

$$\langle \chi | H\varphi \rangle = \int_{\mathbb{R}} E \cdot d\mu_{\chi,\varphi}(E).$$

We have $\varphi \in D(H)$ iff $\int_{\mathbb{R}} |E|^2 d\mu_{\varphi,\varphi}(E) < \infty$.

Furthermore, if $||\varphi|| = 1$, then $\mu_{\varphi,\varphi}$ is a classical probability measure, that is, $\int_{\mathbb{R}} d\mu_{\varphi,\varphi} = 1$. In order to get a physical interpretation, assume that the operator H is the Hamiltonian of a quantum system. Let φ be a unit vector in the Hilbert space X, that is, $||\varphi|| = 1$, and let Ω be an interval on the real line. Then the real number

$$\int_{\Omega} d\mu_{\varphi,\varphi}(E)$$

is the probability of finding the quantum system in the state φ. Moreover,

$$\bar{E} := \int_{\mathbb{R}} E \cdot d\mu_{\varphi,\varphi}(E)$$

is the mean energy value measured in the state φ of the quantum system. Note that this spectral theorem depends on the self-adjointness of the operator H, but it fails for formally self-adjoint operators which are not self-adjoint. Therefore, as it was discovered by von Neumann in 1929, the full probabilistic interpretation of observables in quantum mechanics is only valid for self-adjoint operators.

Let $F : \mathbb{R} \to \mathbb{C}$ be a continuous function (or, more generally, a piecewise continuous and bounded function like the Heaviside function). Let D be the set of all elements φ in X with $\int_{\mathbb{R}} |F(E)|^2 d\mu_{\varphi,\varphi}(E) < \infty$. The von Neumann operator calculus is based on the following fact.

Theorem 7.19 *There exists a uniquely determined self-adjoint operator denoted by $\mathsf{F}(H) : D \to X$ such that, for all $\varphi \in D$, $\chi \in X$, there holds the key relation $\langle \chi | \mathsf{F}(H)\varphi \rangle = \int_{\mathbb{R}} F(E) \cdot d\mu_{\chi,\varphi}(E)$.*

For example, if $F(E) \equiv 1$, then $\mathsf{F}(H) = I$ (unit operator), and for all χ, φ in X we get $\langle \chi | \varphi \rangle = \int_{\mathbb{R}} d\mu_{\chi,\varphi}$.

Sketch of the proof. An elegant short proof of Theorems 7.18 and 7.19 can be found in I. Sigal, Scattering Theory for Many-Body Quantum Mechanical Systems: Rigorous Results, Springer, New York, 1983. In the spirit of the Dirac calculus, the idea of Sigal's proof is to use the regularized (rescaled) resolvent

$$\delta_\varepsilon(EI - H) := \frac{1}{2\pi} \cdot (H - (E + \mathrm{i}\varepsilon)I)^{-1}, \qquad E \in \mathbb{R},\ \varepsilon > 0$$

with the typical property

$$w - \lim_{E_0 \to +\infty} \int_{-E_0}^{E_0} \delta_\varepsilon(EI - H) dE = I, \qquad \varepsilon > 0. \tag{7.103}$$

This justifies the designation as a (regularized) operator delta function. Note that we use the weak limit in (7.103).[67]

Step 1: The operator $\mathsf{F}(H)$ in the regular case: Let $F \to \mathbb{C}$ be a smooth function with compact support, that is, $F \in \mathcal{D}(\mathbb{R})$. We use the key formula

$$\mathsf{F}_\varepsilon(H) := w - \lim_{E_0 \to +\infty} \int_{-E_0}^{E_0} \delta_\varepsilon(EI - H) F(E) dE$$

and the limit formula

$$\mathsf{F}(H) := w - \lim_{\varepsilon \to +0} \mathsf{F}_\varepsilon(H)$$

in order to introduce the operator $\mathsf{F}(H)$ on the Hilbert space X. It can be shown that the limits exist.

Step 2: The spectral family $\{\mathsf{E}_\lambda\}_{\lambda \in \mathbb{R}}$ of the operator H: We extend the definition of the operator $\mathsf{F}(H)$ to more general (discontinuous) bounded functions $F : \mathbb{R} \to \mathbb{C}$ which are the pointwise limit

$$F(E) = \lim_{n \to \infty} F_n(E), \qquad E \in \mathbb{R}$$

of an increasing sequence (F_n) of nonnegative functions $F_n \in \mathcal{D}(\mathbb{R})$. In particular, choosing the characteristic function e_λ of the open interval $]-\infty, \lambda[$, we get the operator $\mathsf{E}_\lambda(H)$.

Step 3: The spectral measure μ: For given $\varphi \in X$ with $||\varphi|| = 1$, we define the probability measure $\mu_{\varphi,\varphi}$ on the real line by setting

$$\int_{]-\infty,\lambda[} d\mu_{\varphi,\varphi}(E) := \langle \varphi | \mathsf{E}_\lambda \varphi \rangle.$$

This is the measure of the open interval $]-\infty, \lambda[$; the function $\lambda \mapsto \langle \varphi | \mathsf{E}_\lambda \varphi \rangle$ represents the distribution function of the measure $\mu_{\varphi,\varphi}$, in terms of the theory of probability. More generally, for given $\varphi, \chi \in X$, we construct the (complex-valued) measure $\mu_{\chi,\varphi}$ on the real line by setting

$$\int_{]-\infty,\lambda[} d\mu_{\chi,\varphi}(E) = \langle \chi | \mathsf{E}_\lambda \varphi \rangle. \tag{7.105}$$

The spectral family of H has the following properties for all real numbers λ, λ_0 and all $\varphi \in X$:

(S1) The operator $\mathsf{E}_\lambda : X \to X$ is an orthogonal projection (i.e., the operator E_λ is linear, continuous, self-adjoint, and $\mathsf{E}_\lambda^2 = \mathsf{E}_\lambda$).
(S2) The function $\lambda \mapsto \langle \varphi | \mathsf{E}_\lambda \varphi \rangle$ is nondecreasing on the real line.
(S3) $\lim_{\lambda \to -\infty} \mathsf{E}_\lambda \varphi = 0$ and $\lim_{\lambda \to \infty} \mathsf{E}_\lambda \varphi = \varphi$.
(S4) $\lim_{\lambda \to \lambda_0 - 0} \mathsf{E}_\lambda \varphi = \mathsf{E}_{\lambda_0} \varphi$. □

[67] Recall that, by definition, the weak limit

$$w - \lim_{n \to \infty} \psi_n = \psi \tag{7.104}$$

exists on the Hilbert space X iff $\lim_{n \to \infty} \langle \varphi | \psi_n \rangle = \langle \varphi | \psi \rangle$ for all $\varphi \in X$. In particular, let $\varphi_1, \varphi_2, \ldots$ be a complete orthonormal system in X. Then the weak convergence (7.104) is equivalent to the boundedness of the sequence (ψ_n) and the convergence of all the Fourier coefficients, that is, $\lim_{n \to \infty} \langle \varphi_k | \psi_n \rangle = \langle \varphi_k | \psi \rangle$ for all k.

The spectral family $\{\mathsf{E}_\lambda\}$ of H is also called the spectral resolution of H.

Corollary 7.20 *For any self-adjoint operator $H : D(H) \to X$, there exists precisely one spectral family $\{E_\lambda\}$ with the properties* (S1)–(S4) *such that Theorem* 7.18 *holds with* (7.105)*. Explictly, the spectral family is given by the limit*

$$\langle\psi|\mathsf{E}_\lambda\varphi\rangle = \lim_{\delta\to+0}\lim_{\varepsilon\to+0}\int_{-\infty}^{\lambda+\delta}\langle\psi|(R_{s-\mathrm{i}\varepsilon} - R_{s+\mathrm{i}\varepsilon})\varphi\rangle ds$$

for all $\psi, \varphi \in X$. Here, $R_\mu := (\mu I - H)^{-1}$ is the resolvent of H.

In terms of physics, the spectral family of the observable H uniquely determines the probability measure of H. The proof of the Corollary can be found in K. Jörgens and F. Rellich, Eigenvalue Problems for Ordinary Differential Equations, p. 113, Springer, Berlin, 1976 (in German). For other proofs of the crucial spectral theorem, we refer to the following monographs:

E. Nelson, Topics in Dynamics: Flows, Princeton University Press, 1969.

K. Maurin, Methods of Hilbert Spaces, Polish Scientific Publishers, Warsaw, 1972.

M. Reed and B. Simon, Methods of Modern Mathematical Physics I: Functional Analysis, Academic Press, New York, 1972.

F. Riesz and B. Nagy, Functional Analysis, Frederyck Ungar, New York, 1978.

F. Berezin and M. Shubin, The Schrödinger Equation, Kluwer, Dordrecht, 1991.

K. Yosida, Functional Analysis, Springer, New York, 1995.

P. Lax, Functional Analysis, Wiley, New York, 2002.

Von Neumann's generalized Fourier transform. Alternatively, von Neumann's spectral theorem above can be obtained from von Neumann's diagonalization theorem:

> *Each linear self-adjoint operator is unitarily equivalent to a multiplication operator $\hat{f}(\lambda) \mapsto \lambda\hat{f}(\lambda)$ on an appropriate function space.*

This represents a far-reaching generalization of the classical Fourier transformation $f \mapsto \hat{f}$. The precise formulation can be found in Sect. 12.2.3 of Vol. I in the setting of the rigorous justification of the Dirac calculus.

Gelfand's theory of C^*-algebras. It was discovered by Gelfand in the 1940s that one can use the theory of C^* algebras in order to construct von Neumann's operator calculus (see the monographs Maurin (1972) and Yosida (1995) quoted above). Note that C^*-algebras play a fundamental role in quantum mechanics, quantum field theory, statistical physics, the Standard Model in particle physics, quantum gravity, and quantum information. The point is that C^*-algebras allow us to describe states and observables in a general setting. We will thoroughly study this in Vol. IV on quantum mathematics (see also Sect. 7.16.3ff for the definition of C^*-algebras together with the construction of the Weyl quantization functor).

The Rellich–Kato perturbation theorem. The operator

$$H + C : D(H) \to X$$

is self-adjoint if the following conditions are satisfied:

- The operator $H : D(H) \to X$ is self-adjoint.

- The perturbation $C : D(C) \to X$ is linear and symmetric, and the domain of definition $D(C)$ contains the set $D(H)$.
- There are fixed real numbers $0 \le a < 1$ and $b \ge 0$ such that

$$||C\varphi|| \le a||H\varphi|| + b||\varphi|| \qquad \text{for all} \quad \varphi \in D(H).$$

In particular, the assumptions are satisfied if the operator $C : X \to X$ is linear, symmetric, and continuous. The proof can be found in Zeidler (1995a), p. 417 (see the references on page 1049). In 1951, this criterion was used by Kato in order to prove that the Hamiltonian operators of molecules are essentially self-adjoint.

Classification of the spectrum. As we will discuss below, every self-adjoint operator $H : D(H) \to X$ generates a unique decomposition

$$\boxed{X = X_{\text{bound}} \oplus X_{\text{scatt}} \oplus X_{\text{sing}}} \tag{7.106}$$

of the Hilbert space X into pairwise orthogonal closed linear subspaces. It turns out that, in terms of quantum mechanics,

- the elements of X_{bound} correspond to bound states of the quantum system,
- and the elements of X_{scatt} correspond to scattering states.

The elements of X_{sing} are called singular states. In regular situations, the singular space X_{sing} is trivial, that is, it only consists of the zero element.[68]

(i) Bound states: The element φ of X is called an eigenstate of the Hamiltonian H iff there exists a real number E such that

$$\boxed{H\varphi = E\varphi, \qquad \varphi \neq 0.}$$

The number E is called the eigenvalue to the eigenstate φ.[69] By definition, the space X_{bound} is the closed linear hull of the eigenstates of H. The eigenstates of H form a complete orthonormal system of X_{bound}.

(ii) Classification of states by means of the spectral measure: Let the nonzero state $\varphi \in X$ be given. Consider the spectral measure μ_φ on the real line.[70] Then:

- $\varphi \in X_{\text{bound}}$ iff μ_φ is a point measure, that is, there exists a finite or countable set $\Omega = \{x_1, x_2, \ldots\}$ such that $\mu_\varphi(\{x_k\}) > 0$ for all k and $\mu_\varphi(\mathbb{R} \setminus \Omega) = 0$.
- $\varphi \in X_{\text{scatt}}$ iff the measure μ_φ has a density, that is, there exists a nonnegative integrable function $\varrho : \mathbb{R} \to \mathbb{R}$ such that $\mu_\varphi(\Omega) = \int_\Omega \varrho(x)dx$ for all intervals Ω.[71]
- $\varphi \in X_{\text{sing}}$ iff the measure μ_φ is singular, that is, there exists a set Ω of Lebesgue measure zero such that $\mu_\varphi(\Omega) > 0$ and $\mu_\varphi(\mathbb{R} \setminus \Omega) = 0$.

The operator H maps each of the three Hilbert spaces X_{bound}, X_{scatt} and X_{sing} into itself.

---

[68] The importance of both the absolutely continuous spectrum and the subspace X_{scatt} for the functional-analytic scattering theory will be discussed in Sect. 9 on page 747.

[69] On page 526 we will introduce eigencostates (or generalized eigenfunctions). Such costates do not always live in the infinite-dimensional Hilbert space X, but in an extension of X. Observe that each eigenstate is an eigencostate, but the converse is not always true. The eigenvalues of eigencostates are called generalized eigenvalues.

[70] To simplify notation, we write μ_φ instead of $\mu_{\varphi,\varphi}$.

[71] Equivalently, the monotone function $\lambda \mapsto \langle\varphi|\mathsf{E}_\lambda\varphi\rangle$ is differentiable almost everywhere on $\mathbb{R}$, and the first derivative is integrable over $\mathbb{R}$.

- The spectrum of the restriction $H : D(H) \cap X_{\rm bound} \to X_{\rm bound}$ is called the pure point spectrum $\sigma_{pp}(H)$.
- The spectrum of the restriction $H : D(H) \cap X_{\rm scatt} \to X_{\rm scatt}$ is called the absolutely continuous spectrum $\sigma_{\rm ac}(H)$.
- The spectrum of the restriction $H : D(H) \cap X_{\rm sing} \to X_{\rm sing}$ is called the singular spectrum $\sigma_{\rm sing}(H)$.

We have the following representation of the spectrum of the operator H:

$$\boxed{\sigma(H) = \sigma_{pp}(H) \cup \sigma_{ac}(H) \cup \sigma_{\rm sing}(H).}$$

The union $\sigma_{\rm c}(H) := \sigma_{\rm ac}(H) \cup \sigma_{\rm sing}(H)$ of the disjoint sets $\sigma_{ac}(H)$ and $\sigma_{\rm sing}(H)$ is called the continuous spectrum of H.

Recall that $\sigma(H)$ is a closed subset of the real line, and the open complement $\varrho(H) := \mathbb{C} \setminus \sigma(H)$ is the resolvent set of H. We have $E \in \varrho(H)$ iff the inverse operator $(EI - H)^{-1} : X \to X$ (i.e., the resolvent) exists as a linear continuous operator. We say that $\sigma_{pp}(H)$ is empty iff $X_{\rm bound} = \{0\}$. An analogous terminology will be used for $\sigma_{\rm ac}(H)$ and $\sigma_{\rm sing}(H)$.

The discrete spectrum. By definition, the discrete spectrum $\sigma_{\rm disc}$ of the operator H is the set of all eigenvalues of finite multiplicity which are isolated points of the spectrum $\sigma(H)$.

The Weyl stability theorem for the essential spectrum. By definition, the essential spectrum $\sigma_{\rm ess}(H)$ of the operator H is the complement to the discrete spectrum. That is, we have the disjoint decomposition

$$\sigma(H) = \sigma_{\rm disc}(H) \cup \sigma_{\rm ess}(H).$$

Explicitly, the essential spectrum contains precisely the following points:

- the eigenvalues of infinite multiplicity,
- the accumulation points of the set of eigenvalues,
- the points of the continuous spectrum.

Weyl proved that we have $E \in \sigma_{\rm ess}(H)$ iff there exists a sequence (φ_n) in the domain of definition $D(H)$ with

- $\lim_{n\to\infty} ||H\varphi_n - E\varphi_n|| = 0$;
- $||\varphi_n|| = 1$ for all n and $w - \lim_{n\to\infty} \varphi_n = 0$;
- the sequence (φ_n) has no convergent subsequence.

Such sequences are called Weyl sequences. The following theorem tells us that the essential spectrum of the self-adjoint operator H is invariant under compact perturbations. The linear operator $C : X \to X$ is called compact iff it is continuous and each sequence $(C\varphi_n)$ contains a convergent subsequence provided (φ_n) is bounded.

Theorem 7.21 *Let $H : D(H) \to X$ be a self-adjoint operator, and let C be a linear compact self-adjoint operator. Then the operator $H + C$ is self-adjoint and $\sigma_{\rm ess}(H + C) = \sigma_{\rm ess}(H)$.*

A variant of this theorem was proven by Weyl in 1909.[72]

Characterization of the spectrum by means of the spectral family. Let $H : D(H) \to X$ be a linear self-adjoint operator on the complex Hilbert space X. Set $P_{\lambda_0}\psi := \lim_{\lambda\to\lambda_0+0}(\mathsf{E}_\lambda - \mathsf{E}_{\lambda_0})\psi$ for all $\psi \in X$.

[72] H. Weyl, On the completely continuous difference of two bounded quadratic forms, Rend. Circ. Mat. Palermo **27** (1909), 373–392 (in German).

Theorem 7.22 (i) *The real number λ_0 is not contained in the spectrum $\sigma(H)$ of the operator H iff the spectral family $\{\mathsf{E}_\lambda\}_{\lambda\in\mathbb{R}}$ is constant in some open neighborhood of the point λ_0.*

(ii) *The real number λ_0 is an eigenvalue of H iff the spectral family jumps at the point λ_0. That is, $P_{\lambda_0} \neq 0$. The operator $P_{\lambda_0} : X \to X$ is the orthogonal projection operator onto the eigenspace of H to the eigenvector λ_0.*

(iii) *The real number λ_0 is contained in the essential spectrum $\sigma_{\rm ess}(H)$ iff* $\dim(\mathsf{E}_{\lambda_0+\varepsilon} - \mathsf{E}_{\lambda_0-\varepsilon})(X) = \infty$ *for all $\varepsilon > 0$.*

A comprehensive summary of spectral theory, measure theory, and other tools of modern analysis together with applications can be found in the Appendix to Zeidler, Nonlinear Functional Analysis and its Applications, Vol. IIB, Springer, New York, 1986. We also refer to Reed and Simon, Methods of Modern Mathematical Physics, Vols. 1–4, Academic Press, New York, 1972–1979.

7.6.3 Rigorous Propagator Theory

> The function $\psi(t) = \mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}\psi(t_0)$, for all times $t \in \mathbb{R}$, describes the dynamics of a quantum system corresponding to the self-adjoint Hamiltonian H.
>
> Folklore

It is our goal to translate the formal propagator theory from Sect. 7.5.3 into a rigorous mathematical setting.

Quantum Dynamics

The abstract Schrödinger equation. Consider the initial-value problem

$$\mathrm{i}\hbar\dot{\psi}(t) = H\psi(t), \quad t > t_0, \quad \psi(t_0) = \psi_0. \tag{7.107}$$

This is the basic equation in quantum mechanics.

Theorem 7.23 *Let $H : D(H) \to X$ be a linear self-adjoint operator on the complex separable Hilbert space X. For given initial state $\psi_0 \in D(H)$, the Schrödinger equation (7.107) has a unique, continuously differentiable solution $\psi : [t_0, \infty[\to \mathbb{R}$. Explicitly,*

$$\psi(t) := \mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}\psi_0, \qquad t \geq t_0. \tag{7.108}$$

The operator $e^{-\mathrm{i}(t-t_0)H/\hbar} : X \to X$ is unitary for all times $t \in X$.

The proof can be found in H. Triebel, Higher Analysis, Sect. 22, Barth, Leipzig, 1989.

Generalized solution. For given initial value $\psi_0 \in X$, the function $\psi = \psi(t)$ is well-defined by (7.108). This function is continuous on $[0, \infty[$. In contrast to this, if $\psi_0 \notin D(H)$, then as a rule, it is not true that the derivative $\dot{\psi}(t)$ exists. Therefore, we call $\psi(t) = \mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}\psi_0$ with $\psi_0 \in X$ a generalized solution of the Schrödinger equation (7.107). This solution is defined for all times $t \in \mathbb{R}$.

One-parameter unitary groups. By definition, a one-parameter unitary group on the Hilbert space X is a family $\{U(t)\}_{t\in\mathbb{R}}$ of operators with the following properties:

- $U(t) : X \to X$ is unitary for all times $t \in \mathbb{R}$.
- $U(t+s) = U(t)U(s)$ for all $t, s \in \mathbb{R}$, and $U(0) = I$.

Such a group is called strongly continuous iff the function $t \mapsto U(t)\varphi_0$ is continuous on the real line for all $\varphi_0 \in X$. The following classical result was proven by Stone (1903–1989) in 1932. [73]

Theorem 7.24 *Let X be a complex separable Hilbert space.*

(i) *If $\{U(t)\}_{t\in\mathbb{R}}$ is a strongly continuous, one-parameter unitary group on X, then there exists a unique self-adjoint operator $H : D(H) \to X$ such that*

$$U(t) = e^{-itH/\hbar}\varphi_0 \qquad \text{for all} \quad t \in \mathbb{R}. \tag{7.109}$$

We have $H\varphi_0 = \lim_{t\to 0} \frac{U(t)\varphi_0 - \varphi_0}{t}$. This limit exists precisely iff $\varphi_0 \in D(H)$. The operator H is called the generator of the one-parameter unitary group.

(ii) *Conversely, if $H : D(H) \to X$ is a self-adjoint operator, then formula (7.109) defines a strongly continuous, one-parameter unitary group on X.*

The proof can be found in Zeidler, Nonlinear Functional Analysis and its Applications, Vol. II/A, Sect. 19.21, Springer, New York, 1986.

The Feynman propagator. Let $t, t_0 \in \mathbb{R}$. In terms of Theorem 7.23, the unitary operator

$$P(t, t_0) := \mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}$$

on the Hilbert space X is called the propagator at time t (generated by the Hamiltonian H with respect to the initial time t_0). The truncated operator[74]

$$P^+(t, t_0) := P(t, t_0)\theta(t - t_0), \qquad t \in \mathbb{R}$$

is called the retarded propagator (or the Feynman propagator) at time t (with respect to the initial time t_0.) Obviously, $P(t_0, t_0) = I$. We get

$$\boxed{P(t, t_0) = P(t, \tau)P(\tau, t_0) \qquad \text{for all} \quad t, \tau, t_0 \in \mathbb{R}.}$$

This so-called reversible propagator equation (or group equation) follows from

$$U(t - \tau)U(\tau - t_0) = U(t - \tau + \tau - t_0) = U(t - t_0),$$

which is the consequence of the fact that $\{U(t)\}_{t\in\mathbb{R}}$ forms a group.

Euclidean Quantum Dynamics

The Euclidean Schrödinger equation. Consider the initial-value problem

$$\dot{\psi}(t) = -H\psi(t), \quad t > t_0, \quad \psi(t_0) = \psi_0. \tag{7.110}$$

We assume that the linear self-adjoint operator $H : D(H) \to X$ is nonnegative, that is, $\langle\varphi|H\varphi\rangle \geq 0$ for all $\varphi \in D(H)$. Observe that both the diffusion equation and the heat conduction equation are of this type. Since diffusion is an irreversible process, we expect that the initial condition ψ_0 does not uniquely determine the state $\psi(t)$ in the past $t < t_0$. Mathematically, this is reflected by the fact that the solution (7.111) below is not defined for $t < t_0$.

[73] M. Stone, On one-parameter unitary groups in Hilbert space, Ann. Math. **33** (1932), 643–648.

[74] Recall that $\theta(t-t_0) := 1$ if $t \geq t_0$, and $\theta(t-t_0) := 0$ if $t < t_0$ (Heaviside function).

Theorem 7.25 *For given $\psi_0 \in D(H)$, the Euclidean Schrödinger equation (7.110) has a unique, continuously differentiable solution $\psi : [t_0, \infty[\to \mathbb{R}$. This solution is given by*

$$\psi(t) = e^{-(t-t_0)H}\psi_0, \qquad t \geq t_0. \tag{7.111}$$

The operator family $\{e^{-tH}\}_{t\geq 0}$ forms a non-expansive semigroup, that is, the linear self-adjoint operators $e^{-tH} : X \to X$ satisfy

$$e^{-tH}e^{-sH} = e^{-(t+s)H} \qquad \text{for all} \quad t, s \geq 0,$$

as well as $e^{-tH}{}_{|t=0} = I$, and $\sup_{t\geq 0} ||e^{-tH}|| \leq 1$.

The proof can be found in H. Triebel, Higher Analysis, Sect. 22, Barth, Leipzig, 1989. In order to understand the specifics of the Euclidean quantum dynamics, suppose that the nonnegative self-adjoint operator $H : D(H) \to X$ has a complete orthonormal system $\varphi_0, \varphi_1, \varphi_2, \ldots$ of eigenvectors with $H\varphi_k = E_k\varphi_k$ for all k. Then $E_k = E_k\langle\varphi_k|\varphi_k\rangle = \langle\varphi_k|H\varphi_k\rangle \geq 0$ for all k. For $\psi_0 \in X$, the Parseval equation tells us that $||\psi_0||^2 = \sum_{k=1}^{\infty} |\langle\varphi_k|\psi\rangle|^2$. The series

$$\mathrm{e}^{-tH}\psi_0 = \sum_{k=1}^{\infty} \mathrm{e}^{-E_k t}\langle\varphi_k|\psi\rangle\varphi_k \tag{7.112}$$

is convergent iff $\sum_{k=0}^{\infty} \mathrm{e}^{-2tE_k}|\langle\varphi_k|\psi\rangle|^2 < \infty$. This is true if $t \geq 0$ because of $0 \leq \mathrm{e}^{-E_k t} \leq 1$. However, if $t < 0$, then the convergence of (7.112) can be violated. This reflects the irreversibility of diffusion and heat conduction processes.

The Euclidean propagator. Let $t \geq t_0$. The operator

$$P(t, t_0) := \mathrm{e}^{-(t-t_0)H}$$

is non-expansive on the Hilbert space X, that is $\sup_{t\geq t_0} ||\mathrm{e}^{-(t-t_0)H}|| \leq 1$. This operator is called the Euclidean propagator at time t (generated by the Hamiltonian H with respect to the initial time t_0). Obviously, $P(t_0, t_0) = I$. Furthermore, we have

$$\boxed{P(t, t_0) = P(t, \tau)P(\tau, t_0) \qquad \text{for all} \quad t \geq \tau \geq t_0.}$$

This so-called irreversible propagator equation (or semi-group equation) follows from $\mathrm{e}^{-(t-\tau)H}\mathrm{e}^{-(\tau-t_0)H} = \mathrm{e}^{-(t-\tau+\tau-t_0)H} = \mathrm{e}^{-(t-t_0)H}$, by Theorem 7.25.

Historical remarks. In the 19th century, mathematicians and physicists (e.g., Gauss, Green, Fourier, Riemann and Maxwell) discovered that one can use integral formulas of the type

$$u(x) = \int G(x, y)f(y)dy$$

in order to represent the solutions u of partial differential equations of the form $Lu = f$ which appear in hydrodynamics, gas-dynamics, elasticity, heat conduction, diffusion, and electrodynamics. The integral kernel G is called the Green's function.

Functional analysis was founded by Hilbert in the early 1910s in order to generalize Fredholm's theory of integral equations. At this time, differential equations were reduced to integral equations with Green's functions as integral kernels. In von Neumann's approach to quantum mechanics in the late 1920s, differential operators were regarded as independent mathematical objects, namely, as self-adjoint operators in a Hilbert space. In contrast to this, in his monograph

Principles of Quantum Mechanics,

Clarendon Press, Oxford, 1930, Dirac used his calculus in order to construct (generalized) integral kernels like the Dirac delta function. In the preface to his monograph

Mathematical Foundations of Quantum Mechanics,

Springer, Berlin 1932, von Neumann pointed out that he did not use Dirac's method because of lack of mathematical rigor.

In the 1940s, Feynman was strongly influenced by Dirac's approach. The Feynman propagators are nothing other than special Green's functions. In the 1950s, the two approaches due to Dirac and von Neumann were combined with each other by Gelfand; he used Laurent Schwartz's theory of generalized functions founded in the 1940s and Grothendieck's theory of nuclear spaces. As a typical example, we will consider the free quantum particle in Sect. 7.6.4. In the 1960s, the theory of pseudo-differential operators was created by Kohn and Nirenberg; this approach represents a further generalization of the theory of operator kernels. In quantum mechanics, this is related to the Weyl calculus introduced in the late 1920s by Hermann Weyl (see Sect. 7.12 on Weyl quantization).

Rigorous Operator Kernel

The operator kernel knows all about the operator.
Folklore

Let $N = 1, 2, \ldots$, and let D be a dense subset of $L_2(\mathbb{R}^N)$. The linear continuous operator $A : L_2(\mathbb{R}^N) \to L_2(\mathbb{R}^N)$ is said to have a continuous kernel iff there exists a continuous function $\mathcal{A} : \mathbb{R}^{2N} \to \mathbb{C}$ such that[75]

$$\langle \chi | A\varphi \rangle = \int_{\mathbb{R}^{2N}} \chi(x)^\dagger \mathcal{A}(x,y)\varphi(y) dx^N dy^N \tag{7.113}$$

for all $\varphi, \chi \in D$. This kernel is unique. In fact, if $\mathcal{A}$ and $\mathcal{B}$ are two continuous kernels corresponding to the operator A, then

$$\int_{\mathbb{R}^{2N}} (\chi(x)\varphi(y)^\dagger)^\dagger (\mathcal{A}(x,y) - \mathcal{B}(x,y)) dx^N dy^N = 0$$

for all $\varphi, \chi \in D$. Since the set of functions $(x,y) \mapsto \chi(x)\varphi(y)^\dagger$ with $\varphi, \chi \in D$ is dense in the complex Hilbert space $L_2(\mathbb{R}^{2N})$, we obtain the desired result $\mathcal{A}(x,y) = \mathcal{B}(x,y)$ on $\mathbb{R}^{2N}$.

More generally, if relation (7.113) is true for a function $\mathcal{A} \in L_2(\mathbb{R}^{2N})$, then this function is uniquely determined (as an element of the Hilbert space $L_2(\mathbb{R}^{2N})$ by the operator A. The function $\mathcal{A}$ is called the L_2-kernel of the operator A. Equation (7.113) generalizes the matrix equation

$$\chi^\dagger A\varphi = \sum_{j,k=1}^{n} \chi_j^\dagger A_{jk} \varphi_k.$$

[75] In classical mathematics, one uses $(A\varphi)(x) = \int_{\mathbb{R}^N} \mathcal{A}(x,y)\varphi(y)dy^N$. This is equivalent to (7.113). However, the bilinear formulation (7.113) is crucial for defining the notion of kernel for generalized functions.

Therefore, the kernel $(x, y) \mapsto \mathcal{A}(x, y)$ can be regarded as a continuous version of the complex $(n \times n)$-matrix (A_{jk}). The kernel $\mathcal{A}$ is called self-adjoint iff

$$\boxed{\mathcal{A}(x, y)^\dagger = \mathcal{A}(y, x) \qquad \text{for all} \quad x, y \in \mathbb{R}^N.}$$

This generalizes self-adjoint matrices. In 1904 Hilbert discovered the importance of self-adjoint integral kernels for both

- the spectral theory of integral operators and
- the Fourier expansions to regular boundary-value problems for second-order ordinary differential equations (i.e., the regular Sturm–Liouville problems).

In 1910, Weyl generalized this to singular Sturm–Liouville problems which are typical for computing the spectra of atoms and molecules in quantum mechanics.[76]

7.6.4 The Free Quantum Particle as a Paradigm of Functional Analysis

> Extend the pre-Hamiltonian to a self-adjoint operator on an appropriate Hilbert space X of quantum states, and use costates related to a Gelfand triplet with respect to X.
>
> The golden rule

The modern theory of differential and integral equations is based on functional analysis, which was created by Hilbert (1862–1943) in the beginning of the 20th century.[77] The development of functional analysis was strongly influenced by the questions arising in quantum mechanics and quantum field theory. In this section, we want to study thoroughly how the motion of a free quantum particle on the real line is related to fundamental notions in functional analysis.

This is Ariadne's thread in functional analysis.

This way, the formal considerations from Sect. 7.5.3 will obtain a sound basis for the free quantum particle.

The main idea of the modern strategy in mathematics and physics consists in describing differential operators and integral operators by abstract operators related to generalized integral kernels.

(i) In the language used by physicists, this concerns the Dirac calculus based on Dirac's delta function and Green's functions (also called Feynman propagators).
(ii) In the language used by mathematicians this is closely related to:
- Lebesgue's passage from the Riemann integral to the Lebesgue integral based on measure theory in about 1900;
- von Neumann's passage from formally self-adjoint operators to self-adjoint operators and his generalization of the classical Fourier transform via spectral theory in the late 1920s;
- Laurent Schwartz's theory of generalized functions including the kernel theorem in the 1940s;

[76] Weyl used methods on singular integral equations. These methods were developed in Weyl's Ph.D. thesis advised by Hilbert in Göttingen in 1908.

[77] As an introduction, we recommend P. Lax, Functional Analysis, Wiley, New York, 2002, and E. Zeidler, Applied Functional Analysis, Vols. 1, 2, Springer, New York, 1995.

- the generalization of von Neumann's spectral theory by Gelfand and Kostyuchenko in 1955 (based on quantum costates as generalized functions and the corresponding Gelfand triplets);
- the extension of the Gelfand–Kostyuchenko approach to general nuclear spaces by Maurin in 1959.[78]

Tempered Distributions

> In order to translate the very elegant, but formal Dirac calculus into mathematics, one has to leave the Hilbert space of states used by von Neumann in about 1930. Folklore

In what follows, we will use

- the space $\mathcal{S}(\mathbb{R})$ of smooth test functions $\varphi : \mathbb{R} \to \mathbb{C}$ which decrease rapidly at infinity,
- and the space $\mathcal{S}'(\mathbb{R})$ of tempered distributions introduced on page 615 of Vol. I.

Our basic tools will be

- the Fourier transform and
- the language of tempered distributions, and Gelfand triplets.

The main idea of our functional-analytic approach to the free quantum particle on the real line is to study the three energy operators

$$H_{\text{pre}} \subseteq H_{\text{free}} \subseteq H_{\text{pre}}^d.$$

Here, we start with $H_{\text{pre}}\varphi := -\frac{\hbar^2}{2m}\frac{d^2\varphi}{dx^2}$ for all $\varphi \in \mathcal{S}(\mathbb{R})$. This is the one-dimensional Laplacian. We first extend the (self-dual and formally self-adjoint) pre-Hamiltonian $H_{\text{pre}} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ on the space of test functions $\mathcal{S}(\mathbb{R})$ to the dual Hamiltonian

$$H_{\text{pre}}^d : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$$

on the space of tempered distributions. The restriction of the operator H_{pre}^d to the Hilbert space $L_2(\mathbb{R})$ yields the self-adjoint Hamiltonian

$$H_{\text{free}} : D(H_{\text{free}}) \to L_2(\mathbb{R})$$

used by von Neumann. Here, $\mathcal{S}(\mathbb{R}) \subseteq D(H_{\text{free}}) \subseteq L_2(\mathbb{R})$ where the domain $D(H_{\text{free}})$ of the free Hamiltonian H_{free} is the Sobolev space $W_2^2(\mathbb{R})$. In general, Sobolev spaces play a crucial role in the modern theory of linear and nonlinear partial differential equations. We recommend:

> L. Evans, Partial Differential Equations, Amer. Math. Soc., Providence, Rhode Island, 1998.
>
> Yu. Egorov and M. Shubin, Foundations of the Classical Theory of Partial Differential Equations, Springer, New York, 1998.

[78] I. Gelfand and A. Kostyuchenkov, On the expansion in eigenfunctions of differential operators and other operators, Doklady Akad. Nauk **103** (1955), 349–352 (in Russian).
K. Maurin, General eigenfunction expansion and the spectral representation of general kernels: a generalization of distribution theory to Lie groups, Bull. Acad. Sci. Polon. Sér. math. astr. et phys. **7** (1959), 471–479 (in German).

Yu. Egorov, A. Komech, and M. Shubin, Elements of the Modern Theory of Partial Differential Equations, Springer, New York, 1999.

P. Lax, Hyperbolic Partial Differential Equations, Courant Institute, New York, 2007.

R. Dautray and J. Lions, Mathematical Analysis and Numerical Methods for Science and Technology, Vols. 1–6, Springer, New York, 1988.

H. Triebel, Theory of Function Spaces, Birkhäuser, Basel, 1992.

We also refer to the author's monographs: Zeidler (1986), Vols. 1–4, and Zeidler (1995a), (1995b) (see the references on page 1049).

The Schrödinger Equation

The instationary Schrödinger equation. The motion of a free quantum particle of mass $m > 0$ on the real line is governed by the following initial-value problem

$$\boxed{\mathrm{i}\hbar\psi_t(x,t) = -\frac{\hbar^2}{2m}\psi_{xx}(x,t), \quad x,t \in \mathbb{R}, \quad \psi(0,x) = \psi_0(x).} \tag{7.114}$$

Here, the wave function ψ_0 is given at the initial time $t = 0$.

The stationary Schrödinger equation. Using the classical Fourier ansatz $\psi(x,t) := \mathrm{e}^{-\mathrm{i}tE/\hbar}\varphi(x)$, equation (7.114) implies the eigenvalue problem

$$\boxed{-\frac{\hbar^2}{2m}\,\varphi''(x) = E\varphi(x), \qquad x \in \mathbb{R}.} \tag{7.115}$$

We are looking for a nonzero function φ and a complex number E.

The Weyl lemma tells us that each solution of (7.115), in the sense of distributions, is a classical smooth function.[79]

Explicitly, all the solutions of (7.115) are given by

$$\varphi_p(x) := \frac{\mathrm{e}^{\mathrm{i}px/\hbar}}{\sqrt{2\pi\hbar}}, \qquad x \in \mathbb{R}$$

with the energy $E(p) := \frac{p^2}{2m}$. Here, p is an arbitrary real number. For any $p \in \mathbb{R}$, we have

$$-\mathrm{i}\hbar\frac{d\varphi_p}{dx} = p\varphi_p.$$

The normalization factor of φ_p is chosen in such a way that we obtain the Parseval equation (7.118) below.

The wave number. To simplify notation, physicists introduce the wave number $k := p/\hbar$, which has the physical dimension of inverse length. Furthermore, for fixed $k \in \mathbb{R}$, let

$$\chi_k(x) := \frac{\mathrm{e}^{\mathrm{i}kx}}{\sqrt{2\pi}} \qquad \text{for all} \quad x \in \mathbb{R}.$$

[79] H. Weyl, The method of orthogonal projection in potential theory, Duke Math. J. **7** (1940), 414–444. An elementary proof of the Weyl lemma for the Laplacian can be found in Zeidler (1986), Vol. IIA, p. 78 (see the references on page 1049). This is the origin of Hörmander's theory of hypoelliptic differential operators (see Sect. 8.6.3).

Then $-\frac{\hbar^2}{2m}\chi_k'' = E_k\chi_k$ for all $k \in \mathbb{R}$ with the energy

$$E_k = \frac{\hbar^2 k^2}{2m}.$$

Hence $|k| = \frac{\sqrt{2mE_k}}{\hbar}$.

Particle stream. If $k > 0$ (resp. $k < 0$), then the function

$$\psi(x,t) := \mathrm{e}^{-\mathrm{i}tE_k/\hbar}\chi_k(x), \qquad x, t \in \mathbb{R}$$

describes a homogeneous stream of free particles which moves from left to right (resp. right to left). The particles of the stream have the momentum $p = \hbar k$, the velocity

$$v = \frac{\hbar k}{m},$$

and the particle density $\varrho = |\chi_k|^2 = \frac{1}{2\pi}$ (see Sect. 7.4.1 on page 459).

The main trouble. The plane-wave functions χ_k possess a well-defined physical meaning, but they do not live in the Hilbert space $L_2(\mathbb{R})$, since $|\chi_k(x)| = \text{const}$ and hence $\int_{\mathbb{R}} |\chi_k(x)|^2 dx = \infty$.

Thus, the Hilbert space setting is not enough for studying quantum mechanics.

In order to overcome this difficulty, one has to introduce the concept of costates and eigencostates (generalized eigenfunctions). This will be done below. Before studying the Schrödinger equation (7.114) and its Hamiltonian H_{free}, we will investigate Gelfand triplets, the extended Fourier transform, Sobolev spaces, the position operator, and the momentum operator.

The Extended Fourier Transform

We want to study the operators $\mathcal{F}_{\text{pre}} \subseteq \mathcal{F} \subseteq \mathcal{F}_{\text{pre}}^d$, where $\mathcal{F} : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ is unitary (i.e., $\mathcal{F}$ is a Hilbert space isomorphism). This is the key property of the Fourier transform. As we will show below, in terms of physics the Fourier transform describes the duality between position and momentum.

The classical Fourier transform. Recall that $\chi_k(x) := \frac{\mathrm{e}^{\mathrm{i}kx}}{\sqrt{2\pi}}$ for all $x \in \mathbb{R}$ and all wave numbers $k \in \mathbb{R}$. In terms of the function χ_k, the Fourier transform $\hat{\varphi}$ of the test function $\varphi \in \mathcal{S}(\mathbb{R})$ reads as

$$\boxed{\hat{\varphi}(k) = \int_{\mathbb{R}} \chi_k(x)^\dagger \varphi(x)\, dx, \qquad \text{for all} \quad k \in \mathbb{R}.} \tag{7.116}$$

The inverse Fourier transform is given by

$$\boxed{\varphi(x) := \int_{\mathbb{R}} \chi_k(x)\hat{\varphi}(k)\, dk \qquad \text{for all} \quad x \in \mathbb{R}.} \tag{7.117}$$

Here, the function φ is represented as a superposition of plane waves χ_k. For all test functions $\psi, \varphi \in \mathcal{S}(\mathbb{R})$, we have the crucial Parseval equation

$$\int_{\mathbb{R}} \psi(x)^\dagger \varphi(x)\, dx = \int_{\mathbb{R}} \hat{\psi}(k)^\dagger \hat{\varphi}(k)\, dk, \tag{7.118}$$

which shows that the Fourier transform respects the inner product on the Hilbert space $L_2(\mathbb{R})$. Setting $(\mathcal{F}_{\rm pre}\varphi)(k) := \hat{\varphi}(k)$ for all $k \in \mathbb{R}$, we obtain the operator

$$\mathcal{F}_{\rm pre} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$$

called the classical Fourier transform (or the pre-Fourier transform). This operator is linear, bijective, and sequentially continuous (see Vol. I, p. 614). Moreover, for all $\varphi, \psi \in \mathbb{R}$, we have the following two relations:

(U) $\langle\psi|\varphi\rangle = \langle\mathcal{F}_{\rm pre}\psi|\mathcal{F}_{\rm pre}\varphi\rangle$ (pre-unitary), and

(S) $\int_{\mathbb{R}} \psi(x) \cdot (\mathcal{F}_{\rm pre}\varphi)(x)\, dx = \int_{\mathbb{R}} (\mathcal{F}_{\rm pre}\psi)(x) \cdot \varphi(x)\, dx$ (self-duality).

Relation (U) coincides with the Parseval equation (7.118), whereas relation (S) follows from interchanging integration. Explicitly,

$$\int_{\mathbb{R}} \psi(x) \left(\int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}kx} \varphi(k) dk \right) dx = \int_{\mathbb{R}} \varphi(k) \left(\int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}kx} \psi(x) dx \right) dk.$$

Finally, use the replacement $k \Leftrightarrow x$.

The Gelfand triplet. It is crucial to leave the Hilbert space $L_2(\mathbb{R})$ and to use the extension $\mathcal{S}'(\mathbb{R})$ of $L_2(\mathbb{R})$ by considering the functions in $L_2(\mathbb{R})$ as tempered distributions. To this end, we introduce the Gelfand triplet (also called the rigged Hilbert space $L_2(\mathbb{R})$):

$$\boxed{\mathcal{S}(\mathbb{R}) \subseteq L_2(\mathbb{R}) \subseteq \mathcal{S}'(\mathbb{R}).}$$

The elements of $L_2(\mathbb{R})$ (resp. $\mathcal{S}'(\mathbb{R})$) are called states (resp. costates). Recall that the inner product on the complex separable Hilbert space $L_2(\mathbb{R})$ is given by

$$\langle\psi|\varphi\rangle = \int_{\mathbb{R}} \psi(x)^\dagger \varphi(x) dx \qquad \text{for all} \quad \varphi, \psi \in L_2(\mathbb{R}).$$

The linear space $\mathcal{S}(\mathbb{R})$ of test functions is dense in $L_2(\mathbb{R})$. For any given function $\psi \in L_2(\mathbb{R})$, we define

$$T_\psi(\varphi) := \int_{\mathbb{R}} \psi(x)\varphi(x) dx \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).$$

Then, T_ψ is a tempered distribution. The map $\psi \mapsto T_\psi$ is an injective map from $L_2(\mathbb{R})$ into $\mathcal{S}'(\mathbb{R})$. Therefore, we may identify ψ with T_ψ. This will frequently be done in the future, by using the symbol ψ instead of T_ψ. In addition, if $\psi \in L_2(\mathbb{R})$, then we define the costate $\langle\psi|$ by setting

$$\langle\psi|(\varphi) := \int_{\mathbb{R}} \psi(x)^\dagger \varphi(x) dx \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).$$

Here, the costate $\langle\psi|$ is a tempered distribution. Obviously, $\langle\psi|(\varphi) = \langle\psi|\varphi\rangle$. Finally, for $k \in \mathbb{R}$, let us define the costate $\langle k|$ by setting

$$\langle k|(\varphi) := \int_{\mathbb{R}} \chi_k^\dagger(x) \varphi(x) dx \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).$$

Motivated by the Dirac calculus, we will write $\langle k|\varphi\rangle$ instead of $\langle k|(\varphi)$. Let $\varphi \in \mathcal{S}(\mathbb{R})$. The relation to the Fourier transform is given by

$$\langle k|\varphi\rangle = \hat{\varphi}(k) \qquad \text{for all} \quad k \in \mathbb{R}.$$

The extended Fourier transform. Recall that $\mathcal{F}_{\rm pre}\varphi := \hat{\varphi}$ for all test functions $\varphi \in \mathcal{S}(\mathbb{R})$. For any tempered distribution $T \in \mathcal{S}'(\mathbb{R})$, define

$$(\mathcal{F}^d_{\rm pre}T)(\varphi) := T(\mathcal{F}_{\rm pre}\varphi) \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).$$

The operator $\mathcal{F}^d_{\rm pre} : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ is linear and bijective. Next we want to show that

$$\mathcal{F}^d_{\rm pre}\psi = \mathcal{F}_{\rm pre}\psi \qquad \text{for all} \quad \psi \in \mathcal{S}(\mathbb{R}). \tag{7.119}$$

Hence $\mathcal{F}_{\rm pre} \subseteq \mathcal{F}^d_{\rm pre}$. For the proof, fix $\psi \in \mathcal{S}(\mathbb{R})$. By the self-duality of the Fourier transform considered on page 513,

$$\mathcal{F}^d_{\rm pre}T_\psi = T_{\mathcal{F}_{\rm pre}\psi}.$$

Thus, identifying ψ with T_ψ, we get the claim (7.119). Our key definition reads as

$$\mathcal{F}\psi := \mathcal{F}^d_{\rm pre}\psi \qquad \text{for all} \quad \psi \in L_2(\mathbb{R}).$$

In other words, the operator $\mathcal{F}$ is the restriction of the operator $\mathcal{F}^d_{\rm rm}$ to the Hilbert space $L_2(\mathbb{R})$. The Plancherel theorem tells us that the operator

$$\mathcal{F} : L_2(\mathbb{R}) \to L_2(\mathbb{R})$$

is unitary. That is, we have the Parseval equation $\langle \mathcal{F}\psi | \mathcal{F}\varphi \rangle = \langle \psi | \varphi \rangle$ for all functions $\psi, \varphi \in L_2(\mathbb{R})$. Explicitly,

$$\boxed{(\mathcal{F}\psi)(k) = \lim_{R\to+\infty} \frac{1}{\sqrt{2\pi}} \int_{-R}^{R} \mathrm{e}^{-\mathrm{i}kx}\psi(x)dx \qquad \text{for all} \quad k \in \mathbb{R}.}$$

The convergence is to be understood in the sense of the Hilbert space $L_2(\mathbb{R})$.

Simplifying notation. Motivated by $\mathcal{F}_{\rm pre} \subseteq \mathcal{F} \subseteq \mathcal{F}^d_{\rm pre}$, we write $\mathcal{F}$ instead of $\mathcal{F}^d_{\rm pre}$ (and $\mathcal{F}_{\rm pre}$). This way, we get the extended Fourier transform

$$\mathcal{F} : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$$

with $(\mathcal{F}T)(\varphi) = T(\mathcal{F}\varphi)$ for all $T \in \mathcal{S}'(\mathbb{R})$ and all $\varphi \in \mathcal{S}(\mathbb{R})$.

The Sobolev space $W_2^m(\mathbb{R})$. Let $m = 1, 2, \ldots$ By definition,

$$W_2^m(\mathbb{R}) := \{\varphi \in L_2(\mathbb{R}) : \varphi^{(j)} \in L_2(\mathbb{R}), j = 1, \ldots, m\}. \tag{7.120}$$

Here, the function φ and its jth derivatives $\varphi^{(j)}$, $j = 1, 2, \ldots$, are to be understood in the sense of tempered distributions (see (7.121)). Thus, $W_2^m(\mathbb{R}) \subseteq \mathcal{S}'(\mathbb{R})$. The space $W_2^m(\mathbb{R})$ becomes a complex separable Hilbert space equipped with the inner product

$$\boxed{\langle \psi | \varphi \rangle := \sum_{j=0}^{m} \int_{\mathbb{R}} \psi^{(j)}(x)^\dagger \varphi^{(j)}(x)dx.}$$

In 1936, spaces of this type were introduced by Sobolev (1885–1967) in order to study singular solutions of wave equations. The Fourier transform allows the following useful characterization of Sobolev spaces. Let $m = 1, 2, \ldots$

Proposition 7.26 *$\psi \in W_2^m(\mathbb{R})$ iff $\psi \in L_2(\mathbb{R})$ and $\int_{\mathbb{R}} |k|^{2m} |\hat{\psi}(k)|^2 dp < \infty$.*

Costates and Dual Operators

The theory of distributions is based on duality. Costates are dual states.
Folklore

Our goal is to construct eigencostates for the following observables: position, momentum, and energy of a free particle. The following investigations serve as preparation for this. Fix the state $\psi \in L_2(\mathbb{R})$. There are two possibilities for assigning a costate to ψ, namely,

- T_ψ (i.e., $T_\psi(\varphi) := \int_{\mathbb{R}} \psi(x)\varphi(x)dx$) for all $\varphi \in \mathcal{S}(\mathbb{R})$) , and
- $T_{\psi^\dagger}$ (i.e., $T_{\psi^\dagger}(\varphi) := \int_{\mathbb{R}} \psi(x)^\dagger \varphi(x)dx$ for all $\varphi \in \mathcal{S}(\mathbb{R})$).

The map $\psi \mapsto T_\psi$ (resp. $\psi \mapsto T_{\psi^\dagger}$) is injective and linear (resp. antilinear). According to Dirac, we set

$$\boxed{\langle\psi| := T_{\psi^\dagger}.}$$

Moreover, we write $|\psi\rangle$ instead of ψ. In particular, for all $\varphi \in \mathcal{S}(\mathbb{R})$,

$$\langle\psi|(\varphi) = T_{\psi^\dagger}(\varphi) = \int_{\mathbb{R}} \psi(x)^\dagger \varphi(x)dx = \langle\psi|\varphi\rangle.$$

Dual operators. In what follows, duality plays the crucial role. Let us assume that

(H) *The linear operator $A : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ is sequentially continuous.*

This means that $\lim_{n\to\infty} \varphi_n = \varphi$ in $\mathcal{S}(\mathbb{R})$ implies $\lim_{n\to\infty} A\varphi_n = A\varphi$ in $\mathcal{S}(\mathbb{R})$ (see Vol. I, p. 537). We want to construct the dual operator

$$A^d : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R}).$$

To this end, choose $T \in \mathcal{S}'(\mathbb{R})$, and define

$$\boxed{(A^dT)(\varphi) := T(A\varphi) \qquad \text{for all} \quad \varphi \in \mathbb{R}.}$$

Then $A^dT \in \mathcal{S}'(\mathbb{R})$. In fact, $\lim_{n\to\infty} \varphi_n = \varphi$ in $\mathcal{S}(\mathbb{R})$ implies

$$\lim_{n\to\infty} (A^dT)(\varphi_n) = \lim_{n\to\infty} T(A\varphi_n) = T(\lim_{n\to\infty} A\varphi_n) = T(A\varphi) = (A^dT)(\varphi).$$

Obviously, the operator A^d is linear.

Formally self-adjoint operators and pre-observables. Suppose that there exists a formally adjoint operator $A^\dagger : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ to the operator A from (H) above (see Problem 7.4). Then

$$\boxed{A^d\langle\psi| = \langle A^\dagger\psi| \qquad \text{for all} \quad \psi \in \mathcal{S}(\mathbb{R}).}$$

Indeed, for all $\varphi \in \mathcal{S}(\mathbb{R})$, we obtain

$$(A^d\langle\psi|)(\varphi) = \langle\psi|(A\varphi) = \langle\psi|A\varphi\rangle = \langle A^\dagger\psi|\varphi\rangle = \langle A^\dagger\psi|(\varphi).$$

In particular, if the operator A is formally self-adjoint (i.e., $A = A^\dagger$), then we obtain $A^d\langle\psi| = \langle A\psi|$ for all $\psi \in \mathcal{S}(\mathbb{R})$.

Self-dual operators and the Fourier transform. The operator A from (H) above is called self-dual iff

$$\int_{\mathbb{R}} \psi(x) \cdot (A\varphi)(x)dx = \int_{\mathbb{R}} (A\psi)(x) \cdot \varphi(x)dx \qquad \text{for all} \quad \varphi, \psi \in \mathcal{S}(\mathbb{R}).$$

Then $(A^d T_\psi) = T_{A\psi}$ for all $\psi \in \mathcal{S}(\mathbb{R})$. Identifying ψ with $T\psi$, we obtain

$$A^d\psi = A\psi \qquad \text{for all} \quad \psi \in \mathcal{S}(\mathbb{R}).$$

Hence $A \subseteq A^d$. To simplify notation, we frequently denote the dual operator A^d by the symbol

$$A : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R}),$$

and we regard this as an extension of the operator $A : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$. A typical example is the Fourier transform considered on page 512.

Antiself-dual operators and the derivative operator. The operator A from (H) above is called antiself-dual iff

$$\int_{\mathbb{R}} \psi(x) \cdot (A\varphi)(x)dx = -\int_{\mathbb{R}} (A\psi)(x) \cdot \varphi(x)dx \qquad \text{for all} \quad \varphi, \psi \in \mathcal{S}(\mathbb{R}).$$

Then $-A^d T_\psi = T_{A\psi}$ for all $\psi \in \mathcal{S}(\mathbb{R})$. Identifying ψ with $T\psi$, we obtain

$$-A^d\psi = A\psi \qquad \text{for all} \quad \psi \in \mathcal{S}(\mathbb{R}).$$

Hence $A \subseteq (-A^d)$. To simplify notation, we frequently denote the operator $-A^d$ by the symbol

$$A : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R}),$$

and we regard this as an extension of $A : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$. As a typical example, let us consider the derivative operator $A := \frac{d}{dx}$. Integration by parts shows that this operator is antiself-dual.[80] This way, we obtain the extension

$$\frac{d}{dx} : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R}).$$

Let $T \in \mathcal{S}'(\mathbb{R})$. Then $(\frac{d}{dx}T)(\varphi) = T(-\frac{d\varphi}{dx})$ for all $\varphi \in \mathcal{S}(\mathbb{R})$. This is the usual definition of the derivative of a tempered distribution. More generally, let $T \in \mathcal{S}'(\mathbb{R})$. The nth derivative of T is defined by

$$\boxed{\left(\frac{d^n T}{dx^n}\right)(\varphi) := (-1)^n T\left(\frac{d^n \varphi}{dx^n}\right), \quad n = 1, 2, \dots} \tag{7.121}$$

for all test functions $\varphi \in \mathcal{S}(\mathbb{R})$. This definition is based on the fact that the operator $\frac{d^n}{dx^n} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ is self-dual (resp. antiself-dual) if n is even (resp odd).

Each tempered distribution has derivatives of arbitrary order, which are again tempered distributions.

[80] For $n = 1, 2, \dots$ and all $\varphi, \psi \in \mathcal{S}(\mathbb{R})$, integration by parts yields

$$\int_{\mathbb{R}} \frac{d^n \psi(x)}{dx^n}\, \varphi(x)\, dx = (-1)^n \int_{\mathbb{R}} \psi(x)\, \frac{d^n \varphi(x)}{dx^n}\, dx.$$

Eigencostates

> For quantum mechanics, it is crucial to replace eigenvectors by eigencostates.
>
> Folklore

Let $A : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ be a linear operator, and let $\{T_\gamma\}_{\gamma\in\Gamma}$ be a system of nonzero tempered distributions $T_\gamma \in \mathcal{S}'(\mathbb{R})$ with

$$A^d T_\gamma = \lambda_\gamma T_\gamma \qquad \text{for all} \quad \gamma \in \Gamma, \tag{7.122}$$

where $\lambda_\gamma \in \mathbb{C}$ for all $\gamma \in \Gamma$. Explicitly, this means that

$$\boxed{T_\gamma(A\varphi) = \lambda_\gamma T(\varphi) \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}),\ \gamma \in \Gamma.}$$

Then all the distributions T_γ are called eigencostates (or generalized eigenfunctions) of the operator A. The system $\{T_\gamma\}$ is called complete iff, for any given test function $\varphi \in \mathcal{S}(\mathbb{R})$,

$$T_\gamma(\varphi) = 0 \ \text{ for all}\, \gamma \in \Gamma \qquad \text{implies} \qquad \varphi = 0.$$

In addition, if there exists a measure μ on the index set Γ with the generalized Parseval equation

$$\boxed{\int_{\mathbb{R}} \psi(x)^\dagger \varphi(x) dx = \int_\Gamma T_\gamma(\psi)^\dagger T_\gamma(\varphi)\, d\mu(\gamma)} \tag{7.123}$$

for all $\varphi, \psi \in \mathcal{S}(\mathbb{R})$, then the system $\{T_\gamma\}_{\gamma\in\Gamma}$ is called a complete orthonormal system of eigencostates of the operator A. Obviously, the latter property is stronger than completeness. In fact, if $T_\gamma(\varphi) = 0$ for all γ, then $\langle\varphi|\varphi\rangle = 0$, and hence $\varphi = 0$. The complex numbers $T_\gamma(\varphi)$ are called the generalized Fourier coefficients of the test function $\varphi \in \mathcal{S}(\mathbb{R})$. The function

$$\gamma \mapsto T_\gamma(\varphi)$$

is called the generalized Fourier transform of the function $\varphi \in \mathcal{S}(\mathbb{R})$ with respect to the operator A.

The Dirac calculus. It turns out that the Dirac calculus represents a very elegant method in order to formulate quantum mechanics and quantum field theory in a very elegant way. For $\varphi \in \mathcal{S}(\mathbb{R})$, we use the following notation:

- $T_\gamma \Rightarrow \langle\gamma|$,
- $T_\gamma(\varphi) \Rightarrow \langle\gamma|\varphi\rangle$, and
- $\langle\varphi|\gamma\rangle := \langle\gamma|\varphi\rangle^\dagger$.

Then, the generalized Parseval equation (7.123) reads as

$$\langle\psi|\varphi\rangle = \int_\Gamma \langle\psi|\gamma\rangle\langle\gamma|\varphi\rangle\, d\mu(\gamma) \qquad \text{for all} \quad \varphi, \psi \in \mathcal{S}(\mathbb{R}). \tag{7.124}$$

Mnemonically, in order to obtain (7.124) we write $\langle\psi|\varphi\rangle = \langle\psi| \cdot I \cdot |\varphi\rangle$ together with

$$\boxed{I = \int_\Gamma |\gamma\rangle\langle\gamma|\, d\mu(\gamma).}$$

This is Dirac's formal completeness relation.

The Position Operator

We want to study the following three operators $Q_{\rm pre} \subseteq Q \subseteq Q^d_{\rm pre}$.

- Let $\varphi \in \mathcal{S}(\mathbb{R})$. The pre-position operator $Q_{\rm pre} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ is defined by $(Q_{\rm pre}\varphi)(x) := x\varphi(x)$ for all $x \in \mathbb{R}$. The operator $Q_{\rm pre}$ is formally self-adjoint and self-dual. [81]
- Let $T \in \mathcal{S}'(\mathbb{R})$. The dual position operator $Q^d_{\rm pre} : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ is defined by $(Q^d_{\rm pre}T)(\varphi) := T(Q_{\rm pre}\varphi)$ for all $\varphi \in \mathcal{S}(\mathbb{R})$. This means that

$$(Q^d_{\rm pre}T)(\varphi) := T(Q_{\rm pre}\varphi) \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).$$

- The operator $Q : D(Q) \to L_2(\mathbb{R})$ is the restriction of $Q^d_{\rm pre}$ to $L_2(\mathbb{R})$. Explicitly, we set

$$D(Q) := \{\varphi \in L_2(\mathbb{R}) : \int_{\mathbb{R}} |x\varphi(x)|^2 dx < \infty\},$$

and $(Q\varphi)(x) := x\varphi(x)$ for all $x \in \mathbb{R}$ and all $\varphi \in D(Q)$.

The spectral family of the position operator. Fix $\lambda \in \mathbb{R}$, and choose $\varphi \in L_2(\mathbb{R})$. Define the operator $\mathsf{E}_\lambda : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ by setting

$$\boxed{(\mathsf{E}_\lambda\varphi)(x) := e_\lambda(x)\varphi(x) \qquad \text{for all} \quad x \in \mathbb{R},} \tag{7.125}$$

where e_λ is the characteristic function of the open interval $]-\infty, \lambda[$ (see (7.100) on page 497).

Proposition 7.27 *The operator family $\{\mathsf{E}_\lambda\}_{\lambda\in\mathbb{R}}$ is the spectral family of the self-adjoint position operator $Q : D(Q) \to L_2(\mathbb{R})$.*

Proof. The self-adjointness of Q will be proved in Problem 7.15. For all functions $\varphi, \psi \in L_2(\mathbb{R})$,

$$\langle\psi|\mathsf{E}_\lambda\varphi\rangle = \int_{-\infty}^{\infty} \psi(x)^\dagger e_\lambda(x)\varphi(x)dx = \int_{-\infty}^{\lambda} \psi(x)^\dagger\varphi(x)dx.$$

Hence $\frac{d}{d\lambda}\langle\psi|\mathsf{E}_\lambda\varphi\rangle = \psi(\lambda)^\dagger\varphi(\lambda)$. This implies $d\langle\psi|\mathsf{E}_\lambda\varphi\rangle = \psi(\lambda)^\dagger\varphi(\lambda)d\lambda$. Therefore,

$$\langle\psi|Q\varphi\rangle = \int_{-\infty}^{\infty} \psi(x)^\dagger x\varphi(x) = \int_{-\infty}^{\infty} \lambda \cdot d\langle\psi|\mathsf{E}_\lambda\varphi\rangle.$$

Finally, one checks easily that the conditions (S1)–(S4) for a spectral family (formulated on page 502) are satisfied. By the uniqueness statement from Corollary 7.20 on page 502, $\{\mathsf{E}_\lambda\}_{\lambda\in\mathbb{R}}$ is the spectral family of Q. □

Let the function $f : \mathbb{R} \to \mathbb{C}$ be measurable (e.g., piecewise continuous) and bounded on all compact intervals. Define

$$D(f(Q)) := \{\varphi \in L_2(\mathbb{R}) : \int_{\mathbb{R}} |f(x)|^2|\varphi(x)|^2 dx < \infty\}.$$

[81] In fact, for all $\varphi, \psi \in \mathcal{S}(\mathbb{R})$, we have $\int_{\mathbb{R}} \psi(x)^\dagger \cdot x\varphi(x)dx = \int_{\mathbb{R}} (x\psi(x))^\dagger\varphi(x)dx$ and

$$\int_{\mathbb{R}} \psi(x) \cdot x\varphi(x)\, dx = \int_{\mathbb{R}} x\psi(x) \cdot \varphi(x)\, dx.$$

For all $\varphi \in D(f(Q))$ and all $\psi \in L_2(\mathbb{R})$, set

$$\langle \psi | f(Q)\varphi \rangle := \int_{\mathbb{R}} f(\lambda) \cdot d\langle \psi | \mathsf{E}_\lambda \varphi \rangle = \int_{\mathbb{R}} \psi(x)^\dagger f(x)\varphi(x)dx.$$

This way, we uniquely obtain the linear operator $f(Q) : D(f(Q)) \to L_2(\mathbb{R})$. This operator is self-adjoint (resp. continuous on $L_2(\mathbb{R})$) if the function f is real-valued (resp. bounded on $\mathbb{R}$).

Measurement of position. Let $\psi \in L_2(\mathbb{R})$ with $\int_{\mathbb{R}} |\psi(x)|^2 dx = 1$. According to the general approach, the spectral family of the observable Q uniquely determines the measurements of Q in the normalized state ψ.

- Distribution function F: The probability of measuring the observable Q in the open interval $]-\infty, \lambda[$ is given by

$$\mathsf{F}(\lambda) := \langle \psi | \mathsf{E}_\lambda \psi \rangle = \int_{-\infty}^{\lambda} |\psi(x)|^2 dx.$$

  This is the probability of measuring the position of the particle in the interval $]-\infty, \lambda[$.
- The probability for measuring the position of the particle in the interval $[x_0, x_1]$ is equal to $\int_{[x_0,x_1]} d\mathsf{F}(\lambda) = \int_{x_0}^{x_1} |\psi(x)|^2 dx$.
- Mean position of the particle: $\bar{x} = \int_{\mathbb{R}} x \, d\mathsf{F}(x) = \int_{\mathbb{R}} x|\psi(x)|^2 dx$.
- Square of the position fluctuation:

$$(\Delta x)^2 = \int_{\mathbb{R}} (x - \bar{x})^2 \, d\mathsf{F}(x) = \int_{\mathbb{R}} (x - \bar{x})^2 |\psi(x)|^2 dx.$$

The complete orthonormal system of eigencostates of the position operator.

Proposition 7.28 (i) *The operator $Q : D(Q) \to L_2(\mathbb{R})$ has no eigenvectors in the Hilbert space $L_2(\mathbb{R})$.*
(ii) *For the spectrum, $\sigma(Q) = \sigma_{\mathrm{ess}}(Q) =]-\infty, \infty[$.*
(iii) *$X_{\mathrm{scatt}} = L_2(\mathbb{R})$, and $\sigma_{\mathrm{ac}}(Q) = \sigma(Q)$.*

Proof. Ad (i). Suppose that $Q\psi = \lambda\psi$, where $\psi \in L_2(\mathbb{R})$ and $\lambda \in \mathbb{R}$. Then we obtain $(x - \lambda)\psi(x) = 0$ for almost all $x \in \mathbb{R}$. Hence $\psi(x) = 0$ for almost all $x \in \mathbb{R}$. Thus, $\psi = 0$ in $L_2(\mathbb{R})$.

Ad (ii). Use Theorem 7.22 on page 505 and (7.125).

Ad (iii). For any $\varphi \in L_2(\mathbb{R})$, the function $\lambda \mapsto \langle \varphi | \mathsf{E}_\lambda \varphi \rangle$ is differentiable almost everywhere on $\mathbb{R}$, and the first derivative is integrable over $\mathbb{R}$. Thus, $\varphi \in X_{\mathrm{scatt}}$ (see page 503). □

Fix $x \in \mathbb{R}$. Let us consider the Dirac delta distribution $\delta_x \in \mathcal{S}'(\mathbb{R})$ defined by $\delta_x(\varphi) := \varphi(x)$ for all $\varphi \in \mathcal{S}(\mathbb{R})$.

Proposition 7.29 *The system $\{\delta_x\}_{x\in\mathbb{R}}$ represents a complete orthonormal system of eigencostates of the position operator Q_{pre}.*

Proof. Let $\varphi, \psi \in \mathcal{S}(\mathbb{R})$. For any parameter $x \in \mathbb{R}$,

$$\boxed{Q^d_{\mathrm{pre}}\delta_x = x\delta_x.}$$

In fact, $\delta_x(Q_{\mathrm{pre}}\varphi) = x\varphi(x) = x\delta_x(\varphi)$. Furthermore, we have the generalized Parseval equation $\langle \psi | \varphi \rangle = \int_{\mathbb{R}} \psi(x)^\dagger \varphi(x)dx = \int_{\mathbb{R}} \delta_x(\psi)^\dagger \delta_x(\varphi)dx$. □

In the setting of the Dirac calculus, physicists write $\langle x|$ instead of δ_x. Then

$$\langle\psi|\varphi\rangle = \int_{\mathbb{R}} \langle\psi|x\rangle\langle x|\varphi\rangle \, dx \qquad \text{for all} \quad \varphi, \psi \in \mathcal{S}(\mathbb{R}).$$

Mnemonically, this remains true for all $\psi, \varphi \in L_2(\mathbb{R})$. Dirac's formal completeness relation reads as

$$\boxed{I = \int_{\mathbb{R}} |x\rangle\langle x| \, dx.}$$

The relation between eigencostates and the spectral family. Set

$$\psi_0(x) := \mathrm{e}^{-x^2/2} \qquad \text{for all} \quad x \in \mathbb{R}.$$

Then $\psi_0 \in \mathcal{S}(\mathbb{R})$. This function generates the (not normalized) Gaussian measure

$$\mu(J) := \int_J |\psi_0(x)|^2 dx = \int_J \mathrm{e}^{-x^2} \, dx$$

for all intervals J on the real line. Fix $\lambda \in \mathbb{R}$. For all test functions $\varphi \in \mathcal{S}(\mathbb{R})$, define

$$\boxed{T_\lambda(\varphi) := \frac{d\langle\psi_0|\mathsf{E}_\lambda\varphi\rangle}{d\langle\psi_0|\mathsf{E}_\lambda\psi_0\rangle}.}$$

Proposition 7.30 *The family $\{T_x\}_{x\in\mathbb{R}}$ of tempered distributions with*

$$T_x = \frac{\delta_x}{\psi_0(x)}$$

represents a complete orthonormal system of eigencostates of the position operator Q_{pre}. Using the Gaussian measure $d\mu(x) = \psi_0(x)^2 dx$, we have the generalized Parseval equation

$$\int_{\mathbb{R}} \psi(x)^\dagger \varphi(x) dx = \int_{\mathbb{R}} T_x(\psi)^\dagger T_x(\varphi) \, d\mu(x) \qquad \textit{for all} \quad \varphi, \psi \in \mathcal{S}(\mathbb{R}).$$

Proof. By the proof of Prop. 7.27, $d\langle\psi_0|\mathsf{E}_\lambda\varphi\rangle = \psi_0(\lambda)\varphi(\lambda)d\lambda$. Hence

$$T_\lambda(\varphi) = \frac{\psi_0(\lambda)\varphi(\lambda)}{\psi_0(\lambda)^2} = \frac{\varphi(\lambda)}{\psi_0(\lambda)}.$$

Finally, use $\delta_x(\varphi) = \varphi(x)$. □

The square Q^2 of the position operator. By von Neumann's functional calculus, the self-adjoint operator $Q^2 : D(Q^2) \to L_2(\mathbb{R})$ has the domain of definition

$$D(Q^2) = \{\psi \in L_2(\mathbb{R}) : \int_{\mathbb{R}} x^4 |\psi(x)|^2 dx < \infty\}.$$

For $\lambda \in \mathbb{R}$, we get $(\lambda I - Q^2)\psi(x) = f(x)$. If $\lambda < 0$ and $f \in L_2(\mathbb{R})$ then the function

$$(\lambda I - Q^2)^{-1} f(x) = \frac{f(x)}{\lambda - x^2}, \qquad x \in \mathbb{R}$$

is contained in $L_2(\mathbb{R})$. If $\lambda \geq 0$, this is not the case for special choice of f. Hence the spectrum of Q^2 is equal to $[0, \infty[$. Let us compute the spectral family of Q^2. For all $\varphi, \psi \in L_2(\mathbb{R})$,

$$\langle\psi|Q^2\varphi\rangle = \int_{-\infty}^{\infty} \psi(x)^\dagger x^2 \varphi(x) dx.$$

Setting $\lambda = x^2$, we get $\langle\psi|Q^2\varphi\rangle = \int_0^\infty \lambda \varrho_{\psi,\varphi}(\lambda) d\lambda$ with the spectral density

$$\varrho_{\psi,\varphi}(\lambda) := \frac{1}{2\sqrt{\lambda}} \Big(\psi(\sqrt{\lambda})^\dagger \varphi(\sqrt{\lambda}) + \psi(-\sqrt{\lambda})^\dagger \varphi(-\sqrt{\lambda})\Big).$$

Thus, we get $\langle\psi|\mathsf{E}_{\lambda_0}(Q^2)\varphi\rangle = \int_0^\infty e_{\lambda_0}(E)\varrho_{\psi,\varphi}(\lambda)d\lambda$ for all $\lambda_0 \in \mathbb{R}$. The definition of the function e_λ can be found in (7.100) on page 497. In particular, $\mathsf{E}_{\lambda_0} = 0$ if $\lambda_0 \leq 0$.

Proposition 7.31 (i) *The operator Q^2 has no eigenvectors in the Hilbert space $L_2(\mathbb{R})$.*

(ii) *For the spectrum $\sigma(Q^2) = \sigma_{\rm ess}(Q^2) = \sigma_{\rm ac}(Q^2) = [0, \infty[$.*

Proof. Ad (i). Use the same argument as for the operator Q above.

Ad (ii). Use the spectral family together with Theorem 7.22 on page 505. □

The Momentum Operator

We want to study the following three operators $P_{\rm pre} \subseteq P \subseteq (-P^d_{\rm pre})$.

- Let $\varphi \in \mathcal{S}(\mathbb{R})$. The pre-momentum operator $P_{\rm pre} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ is defined by $(P_{\rm pre}\varphi)(x) := -\mathrm{i}\hbar \frac{d}{dx}\varphi(x)$ for all $x \in \mathbb{R}$. The operator $P_{\rm pre}$ is formally self-adjoint and antiself-dual.[82]
- Let $T \in \mathcal{S}'(\mathbb{R})$. The dual momentum operator $P^d_{\rm pre} : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ is defined by $(P^d_{\rm pre}T)(\varphi) := T(P_{\rm pre}\varphi)$ for all $\varphi \in \mathcal{S}(\mathbb{R})$. In the sense of tempered distributions, we have
$$P^d_{\rm pre} = \mathrm{i}\hbar \frac{d}{dx}.$$
This follows from $\mathrm{i}\hbar\frac{dT}{dx}(\varphi) = -\mathrm{i}\hbar T(\varphi') = T(P_{\rm pre}\varphi)$ for all $\varphi \in \mathcal{S}(\mathbb{R})$.
- The operator $P : D(P) \to L_2(\mathbb{R})$ is the natural extension of the operator $P_{\rm pre}$. Explicitly, we set $D(P) := \{\varphi \in L_2(\mathbb{R}) : \varphi' \in L_2(\mathbb{R})\}$, and
$$P\varphi := -\mathrm{i}\hbar \frac{d\varphi}{dx} \qquad \text{for all} \quad \varphi \in D(P).$$
Here, the derivative is to be understood in the sense of tempered distributions. In other words, $D(P) = W^1_2(\mathbb{R})$.

The Fourier transform, and the duality between position and momentum. Choose $\chi := P_{\rm pre}\varphi$ where $\varphi \in \mathcal{S}(\mathbb{R})$. For the Fourier transform, we get $\hat{\chi}(k) = \hbar k \hat{\varphi}(k)$ for all $k \in \mathbb{R}$. Thus, the operator $\hbar^{-1}P_{\rm pre}$ corresponds to the multiplication operator $Q_{\rm pre}$ in the Fourier space. This means that the following diagram is commutative:

[82] In fact, for all $\varphi, \psi \in \mathcal{S}(\mathbb{R})$, we have $\int_{\mathbb{R}} \psi^\dagger(x)(-\mathrm{i}\varphi'(x))dx = \int_{\mathbb{R}}(-\mathrm{i}\psi'(x))^\dagger\varphi(x)dx$ and
$$\int_{\mathbb{R}} \psi(x)(-\mathrm{i}\varphi'(x))dx = -\int_{\mathbb{R}}(-\mathrm{i}\psi'(x))\varphi(x)dx.$$

$$\begin{array}{ccc} \mathcal{S}(\mathbb{R}) & \xrightarrow{\hbar^{-1}P_{\mathrm{pre}}} & \mathcal{S}(\mathbb{R}) \\ \mathcal{F}\downarrow & & \downarrow\mathcal{F} \\ \mathcal{S}(\mathbb{R}) & \xrightarrow{Q_{\mathrm{pre}}} & \mathcal{S}(\mathbb{R}). \end{array}$$

Passing to the extended unitary Fourier transform $\mathcal{F} : L_2(\mathbb{R}) \to L_2(\mathbb{R})$, we obtain the following commutative diagram:

$$\begin{array}{ccc} D(P) & \xrightarrow{\hbar^{-1}P} & L_2(\mathbb{R}) \\ \mathcal{F}\downarrow & & \downarrow\mathcal{F} \\ D(Q) & \xrightarrow{Q} & L_2(\mathbb{R}). \end{array}$$

Since the operator $Q : D(Q) \to L_2(\mathbb{R})$ is self-adjoint and the property of self-adjointness is invariant under unitary transformations, the position operator $P : D(P) \to L_2(\mathbb{R})$ is self-adjoint (see Problem 7.14).

The spectral family of the wave number operator. Recall that the momentum p corresponds to the wave number $k = \hbar^{-1}p$. Therefore, the operator $K := \hbar^{-1}P$ is called the wave number operator. Since the spectral family of a self-adjoint operator is invariant under unitary transformations, we obtain the spectral family $\{\mathsf{E}_\lambda\}_{\lambda\in\mathbb{R}}$ of the wave number operator K from the spectral family $\{\mathsf{E}_\lambda(Q)\}_{\lambda\in\mathbb{R}}$ of the position operator Q in the Fourier space. Explicitly, $\mathsf{E}_\lambda = \mathcal{F}^{-1}\mathsf{E}_\lambda(Q)\mathcal{F}$ for all $\lambda \in \mathbb{R}$. This means that, for all functions $\varphi, \psi \in L_2(\mathbb{R})$ and all real numbers λ, we get

$$\boxed{\langle\psi|\mathsf{E}_\lambda\varphi\rangle = \int_{-\infty}^{\lambda} \hat{\psi}(k)^\dagger\hat{\varphi}(k)dk.}$$

Proposition 7.32 *The operator family $\{\mathsf{E}_\lambda\}_{\lambda\in\mathbb{R}}$ is the spectral family of the self-adjoint wave number operator $\hbar^{-1}P : D(P) \to L_2(\mathbb{R})$.*

Let the function $f : \mathbb{R} \to \mathbb{C}$ be measurable (e.g., piecewise continuous) and bounded on all compact intervals. Define

$$D(f(K)) := \{\varphi \in L_2(\mathbb{R}) : \int_{\mathbb{R}} |f(k)|^2|\hat{\varphi}(k)|^2dk < \infty\}.$$

For all $\varphi \in D(f(K))$ and all $\psi \in L_2(\mathbb{R})$, set

$$\langle\psi|f(K)\varphi\rangle := \int_{\mathbb{R}} f(\lambda)\cdot d\langle\psi|\mathsf{E}_\lambda\varphi\rangle = \int_{\mathbb{R}} f(k)\hat{\psi}(k)^\dagger\hat{\varphi}(k)dk.$$

This way, we obtain the linear operator $f(K) : D(f(K)) \to L_2(\mathbb{R})$. This operator is self-adjoint (resp. continuous on $L_2(\mathbb{R})$) if the function f is real-valued (resp. bounded on $\mathbb{R}$).

Measurement of the wave number. Let $\psi \in L_2(\mathbb{R})$ with the normalization condition $\int_{\mathbb{R}} |\psi(x)|^2dx = 1$. According to the general approach, the spectral family of the observable $K = \hbar^{-1}P$ uniquely determines the measurements of the wave number $k = \hbar^{-1}p$ in the normalized state ψ.

- Distribution function F: The probability of measuring the wave number observable K in the open interval $]-\infty, \lambda[$ is given by

$$\mathsf{F}(\lambda) := \langle\psi|\mathsf{E}_\lambda\psi\rangle = \int_{-\infty}^{\lambda} |\hat{\psi}(k)|^2 dk.$$

  This is the probability of measuring the wave number $k = \hbar^{-1}p$ of the particle in the open interval $]-\infty, \lambda[$.
- The probability of measuring the wave number of the particle in the interval $[k_0, k_1]$ is equal to

$$\int_{[k_0,k_1]} d\mathsf{F}(k) = \int_{k_0}^{k_1} |\hat{\psi}(k)|^2 dk.$$

- Mean wave number of the particle: $\bar{k} = \int_{\mathbb{R}} k \, d\mathsf{F}(k) = \int_{\mathbb{R}} k|\hat{\psi}(k)|^2 dk$.
- Square of the wave number fluctuation:

$$(\Delta k)^2 = \int_{\mathbb{R}} (k-\bar{k})^2 \, d\mathsf{F}(k) = \int_{\mathbb{R}} (k-\bar{k})^2 |\hat{\psi}(k)|^2 dk.$$

Moreover, we get the mean momentum $\bar{p} = \hbar\bar{k}$ and the mean momentum fluctuation $\Delta p = \hbar\Delta k$.

The complete orthonormal system of eigencostates of the momentum operator.

Proposition 7.33 (i) *The operator $P : D(P) \to L_2(\mathbb{R})$ has no eigenvectors in the Hilbert space $L_2(\mathbb{R})$.*
(ii) *For the spectrum, $\sigma(P) = \sigma_{\rm ess}(P) =]-\infty, \infty[$.*
(iii) *$X_{\rm scatt} = L_2(\mathbb{R})$, and $\sigma_{\rm ac}(P) = \sigma(P)$.*

This follows from Prop. 7.28 on page 519 and from the fact that the wave number operator $\hbar^{-1}P$ is unitarily equivalent to the position operator Q.

Proposition 7.34 *The system $\{\langle k|\}_{k\in\mathbb{R}}$ represents a complete orthonormal system of eigencostates of the momentum operator $P_{\rm pre}$.*

Proof. Let $\varphi, \psi \in \mathcal{S}(\mathbb{R})$. For any parameter $k \in \mathbb{R}$,

$$\boxed{P^d_{\rm pre}\langle k| = \hbar k \, \langle k|.}$$

In fact, using $P_{\rm pre}\chi_k = \hbar k \chi_k$, we get

$$\langle k|P_{\rm pre}\varphi\rangle = \int_{\mathbb{R}} \chi_k^\dagger P_{\rm pre}\varphi \, dx = \int_{\mathbb{R}} (P_{\rm pre}\chi_k)^\dagger \varphi dx = \hbar k \int_{\mathbb{R}} \chi_k^\dagger \varphi dx = \hbar k \, \langle k|\varphi\rangle.$$

Furthermore, we have the generalized Parseval equation

$$\langle\psi|\varphi\rangle = \int_{\mathbb{R}} \hat{\psi}(k)^\dagger \hat{\varphi}(k) dk = \int_{\mathbb{R}} \langle\psi|k\rangle\langle k|\varphi\rangle \, dk.$$

Thus, $\langle k|\varphi\rangle = 0$ for all $k \in \mathbb{R}$ implies $\langle\varphi|\varphi\rangle = 0$, and hence $\varphi = 0$. □

Dirac's formal completeness relation reads as

$$\boxed{I = \int_{\mathbb{R}} |k\rangle\langle k| \, dk.}$$

The relation between eigencostates and the spectral family. Set

$$\psi_0(x) := \sqrt{\hbar}\, \mathrm{e}^{-x^2\hbar^2/2} \qquad \text{for all} \quad x \in \mathbb{R}.$$

Then $\psi_0 \in \mathcal{S}(\mathbb{R})$, and $\hat{\psi}_0(k) = \mathrm{e}^{-k^2/2}$. This function generates the (not normalized) Gaussian measure

$$\mu(J) := \int_J \hat{\psi}_0(k)^2 dk = \int_J \mathrm{e}^{-k^2}\, dk$$

for all intervals J on the real line. Fix $\lambda \in \mathbb{R}$. For all test functions $\varphi \in \mathcal{S}(\mathbb{R})$, define

$$\boxed{T_\lambda(\varphi) := \frac{d\langle\psi_0|\mathsf{E}_\lambda\varphi\rangle}{d\langle\psi_0|\mathsf{E}_\lambda\psi_0\rangle}.}$$

Proposition 7.35 *The family $\{T_k\}_{k\in\mathbb{R}}$ of tempered distributions with*

$$T_k = \frac{\langle k|}{\hat{\psi}_0(k)}$$

represents a complete orthonormal system of eigencostates of the wave number operator $\hbar^{-1}P_{\mathrm{pre}}$. Using the Gaussian measure $d\mu(k) = |\psi_0(k)|^2 dk$, we have the generalized Parseval equation

$$\int_{\mathbb{R}} \psi(x)^\dagger \varphi(x) dx = \int_{\mathbb{R}} T_k(\psi)^\dagger T_k(\varphi)\, d\mu(k) \qquad \textit{for all} \quad \varphi, \psi \in \mathcal{S}(\mathbb{R}).$$

Proof. By the proof of Prop. 7.34, $d\langle\psi_0|\mathsf{E}_\lambda\varphi\rangle = \hat{\psi}_0(\lambda)\hat{\varphi}(\lambda)d\lambda$. Hence

$$T_\lambda(\varphi) = \frac{\hat{\psi}_0(\lambda)\hat{\varphi}(\lambda)}{\hat{\psi}_0(\lambda)^2} = \frac{\hat{\varphi}(\lambda)}{\hat{\psi}_0(\lambda)}.$$

Finally, use the Parseval equation for the Fourier transform. □

7.6.5 The Free Hamiltonian

> The free Hamiltonian is a paradigm for general Hamiltonians in quantum mechanics and quantum field theory.
>
> Folklore

The functional-analytic approach to quantum dynamics is based on the study of the energy operator (also called the Hamiltonian). In this section, we want to investigate thoroughly the Hamiltonian H_{free} of the free quantum particle on the real line, which is called the free Hamiltonian. The two key operator equations are the instationary Schrödinger equation

$$\boxed{\mathrm{i}\hbar\dot{\psi}(t) = H_{\mathrm{free}}\psi(t), \quad t > t_0, \quad \psi(t_0) = \psi_0} \tag{7.126}$$

with the solution $\psi(t) = \mathrm{e}^{-\mathrm{i}(t-t_0)H_{\mathrm{free}}/\hbar}\psi_0$ (the Feynman propagator) and the inhomogeneous stationary Schrödinger equation

$$\boxed{H_{\mathrm{free}}\varphi = \mathcal{E}\varphi + f} \tag{7.127}$$

with the solution $\varphi = (H_{\mathrm{free}} - \mathcal{E}I)^{-1}f$ (the energetic Green's operator). Here, we have to assume that the complex energy $\mathcal{E}$ is not contained in the spectrum $\sigma(H_{\mathrm{free}})$ of the free Hamiltonian. We will show that:

- The Feynman propagator kernel $\mathcal{K}$ describes the solution of the initial-value problem for the instationary Schrödinger equation (7.126),

$$\mathrm{i}\hbar\psi_t(x,t) = -\frac{\hbar^2}{2m}\psi_{xx}(x,t), \qquad \psi(t_0,x) = \psi_0(x),$$

by means of the integral formula

$$\psi(x,t) = \int_{\mathbb{R}} \mathcal{K}(x,t;x_0,t_0)\psi_0(x_0)dx_0, \quad t > t_0, x \in \mathbb{R}.$$

- The energetic Green's function $\mathcal{G}$ describes the solution of the inhomogeneous stationary Schrödinger equation (7.126),

$$-\frac{\hbar^2}{2m}\varphi''(x) = \mathcal{E}\varphi(x) + f(x), \qquad x \in \mathbb{R},\ \mathcal{E} \in \mathbb{C},$$

by means of the integral formula

$$\varphi(x) = \int_{\mathbb{R}} \mathcal{G}(x,x_0;\mathcal{E})f(x_0)dx_0, \qquad x \in \mathbb{R},\ \mathcal{E} \in \mathbb{C} \setminus \sigma(H_{\text{free}})$$

where $\sigma(H_{\text{free}}) = [0,\infty[$.

The energetic Green's function carries the information on the energy spectrum of the particle.

The Feynman propagator kernel $\mathcal{K}$ and the energetic Green's function $\mathcal{G}$ are related to each other by the Laplace transform.

This corresponds to

- the duality between energy and time, and
- the duality between causality and analyticity,

which is crucial for both quantum mechanics and quantum field theory.

Using the results on the momentum operators $P_{\text{pre}} \subseteq P \subseteq P^d_{\text{pre}}$ obtained on page 521, we want to study the energy operators $H_{\text{pre}} \subseteq H_{\text{free}} \subseteq H^d_{\text{pre}}$.

- The pre-Hamiltonian $H_{\text{pre}} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ is defined by

$$H_{\text{pre}} := \frac{P^2_{\text{pre}}}{2m}.$$

Explicitly, $H_{\text{pre}}\varphi = -\frac{\hbar^2}{2m}\ \varphi''$ for all $\varphi \in \mathcal{S}(\mathbb{R})$. The operator H_{pre} is formally self-adjoint and self-dual.

- The operator $H^d_{\text{pre}} : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ is defined by $H^d_{\text{pre}} := \frac{(P^d_{\text{pre}})^2}{2m}$. For any tempered distribution $T \in \mathcal{S}'(\mathbb{R})$,

$$H^d_{\text{pre}}T = -\frac{\hbar}{2m}\ \frac{d^2T}{dx^2}.$$

- By von Neumann's functional calculus, the operator $P^2 : D(P^2) \to L_2(\mathbb{R})$ is self-adjoint, and $D(P^2) = \{\psi \in L_2(\mathbb{R}) : \int_{\mathbb{R}} |k^2\hat{\psi}(k)|^2 dk < \infty\}$. By Prop. 7.26, $D(P^2) = W^2_2(\mathbb{R})$.
- We define the self-adjoint free Hamiltonian $H_{\text{free}} : D(H_{\text{free}}) \to L_2(\mathbb{R})$ by setting $H_{\text{free}} := \frac{P^2}{2m}$. Hence $D(H_{\text{free}}) = W^2_2(\mathbb{R})$.

Eigencostates. Recall that $\langle k|\varphi\rangle = \hat{\varphi}(k)$ for all $\varphi \in \mathcal{S}(\mathbb{R})$, where $\hat{\varphi}$ is the Fourier transform of φ. Moreover, following Dirac, we set $\langle \varphi|k\rangle := \langle k|\varphi\rangle^\dagger$. Recall that $E_k := \frac{\hbar^2 k^2}{2m}$ is the energy of a classical free particle on the real line which has the momentum $p = \hbar k$.

Proposition 7.36 *The system $\{\langle k|\}_{k\in\mathbb{R}}$ is a complete orthonormal system of eigencostates of the energy operator H_{pre}. Explicitly,*

(a) $H^d_{\mathrm{pre}}\langle k| = E_k\langle k|$ *for all wave numbers* $k \in \mathbb{R}$.

(b) $\langle\psi|\varphi\rangle = \int_{\mathbb{R}} \langle\psi|k\rangle\langle k|\varphi\rangle\, dk$ *for all* $\psi, \varphi \in L_2(\mathbb{R})$.

Proof. Since $P^d_{\mathrm{pre}}\langle k| = \hbar k\langle k|$, we get $H^d_{\mathrm{pre}}\langle k| = \frac{(P^d_{\mathrm{pre}})^2}{2m}\langle k| = \frac{(\hbar k)^2}{2m}\langle k|$. This is (a). Claim (b) coincides with the Parseval equation for the Fourier transform. □

In terms of distribution theory, the costate $\langle k|$ corresponds to the function $\chi_k^\dagger(x) = \frac{\mathrm{e}^{-\mathrm{i}kx}}{\sqrt{2\pi}}$ for all $x \in \mathbb{R}$. Passing from k to $-k$, claim (a) is equivalent to

$$-\frac{\hbar^2}{2m}\frac{d^2\chi_k}{dx^2} = \frac{\hbar^2 k^2}{2m}\cdot\chi_k \qquad \text{for all} \quad k \in \mathbb{R}.$$

The elements of the Hilbert space $L_2(\mathbb{R})$ correspond to states of a single particle. The function χ_k is not a state, but it describes a particle stream, as discussed on page 512.

The spectrum of the free Hamiltonian H_{free} acting in the Hilbert space X of states. We have $X_{\mathrm{scatt}} = L_2(\mathbb{R})$ and

$$\boxed{\sigma(H_{\mathrm{free}}) = \sigma_{\mathrm{ac}}(H_{\mathrm{free}}) = \sigma_{\mathrm{ess}}(H_{\mathrm{free}}) = [0, +\infty[.}$$

That is, the spectrum of the free Hamiltonian H_{free} contains all the energy values $E \geq 0$. The spectrum coincides with both the absolutely continuous spectrum and the essential spectrum. The pure point spectrum is empty, that is, there is no state of the free quantum particle on the real line which has a sharp energy. In other words, there are no bound states. In addition, the singular spectrum is empty. The resolvent set of the operator H_{free} is given by $\varrho(H_{\mathrm{free}}) = \mathbb{C} \setminus [0, +\infty[$.

The proof follows from the corresponding properties of the operator Q^2 and the fact that the operator $\hbar^{-2}P^2$ is unitarily equivalent to Q^2, by Fourier transform (see page 520).

The quantum dynamics: We will use Theorem 7.23 together with the Stone theorem on page 505ff. Set $P(t, t_0) := \mathrm{e}^{-\mathrm{i}(t-t_0)H_{\mathrm{free}}/\hbar}$. For all times $t, t_0 \in \mathbb{R}$, the operator

$$P(t, t_0) : L_2(\mathbb{R}) \to L_2(\mathbb{R})$$

is unitary. For each given initial state $\psi_0 \in L_2(\mathbb{R})$ at time t_0, we set

$$\boxed{\psi(t) := P(t, t_0)\psi_0, \qquad t \in \mathbb{R}.}$$

The function $t \mapsto \psi(t)$ describes the motion of the free quantum particle on the real line with the initial condition $\psi(t_0) = \psi_0$. If $\psi_0 \in D(H_{\mathrm{free}})$ (e.g., we choose $\psi_0 \in \mathcal{S}(\mathbb{R})$), then the function $\psi : [0, \infty[\to L_2(\mathbb{R})$ is continuously differentiable, and we have the Schrödinger equation

$$\mathrm{i}\hbar\dot{\psi}(t) = H_{\mathrm{free}}\psi(t), \quad t \in \mathbb{R}, \quad \psi(t_0) = \psi_0.$$

The operator $P(t,t_0)$ is called the propagator of the free quantum particle at time t (with respect to the initial time t_0). In terms of the unitary Fourier transform $\mathcal{F} : L_2(\mathbb{R}) \to L_2(\mathbb{R})$, the propagator $P(t,t_0)$ corresponds to the multiplication with the function $k \mapsto \mathrm{e}^{-\mathrm{i}(t-t_0)E_k/\hbar}$ in the Fourier space. This means that, for all $\psi_0 \in L_2(\mathbb{R})$, we get

$$P(t,t_0)\psi_0 = \mathcal{F}^{-1} M \mathcal{F} \psi_0, \qquad t, t_0 \in \mathbb{R}$$

with the multiplication operator $(M\hat{\psi}_0)(k) := \mathrm{e}^{-\mathrm{i}(t-t_0)E_k/\hbar}\hat{\psi}_0(k)$ for all wave numbers $k \in \mathbb{R}$.

The spectral measure of the free Hamiltonian H_{free}. Let the function $F : [0,\infty[\to \mathbb{C}$ be continuous (or piecewise continuous) and bounded. Then, for all $\chi, \varphi \in \mathcal{S}(\mathbb{R})$,

$$\boxed{\langle \chi | \mathsf{F}(H_{\mathrm{free}})\varphi \rangle = \int_0^\infty F(E)\, \varrho_{\chi,\varphi}(E) dE} \tag{7.128}$$

with the smooth density function

$$\varrho_{\chi,\varphi}(E) := \sqrt{\frac{m}{2\hbar^2 E}} \Big(\hat{\chi}(k)^\dagger \hat{\varphi}(k) + \hat{\chi}(-k)^\dagger \hat{\varphi}(-k) \Big), \quad E > 0.$$

Here, $k := \sqrt{2mE}/\hbar$. Moreover, $\hat{\chi}$ (resp. $\hat{\varphi}$) is the Fourier transform of χ (resp. φ) from (7.116). Formula (7.128) can be uniquely extended to all $\chi, \varphi \in L_2(\mathbb{R})$. The operator $\mathsf{F}(H_{\mathrm{free}}) : X \to X$ is linear and continuous. Formula (7.128) remains valid if we replace the function F by its complex-conjugate function $F^\dagger$ and the operator $\mathsf{F}(H_{\mathrm{free}})$ by its adjoint operator $\mathsf{F}(H_{\mathrm{free}})^\dagger$, respectively. If the function F is real-valued, then the operator $\mathsf{F}(H_{\mathrm{free}})$ is self-adjoint. Furthermore,

$$\langle \chi | H_{\mathrm{free}} \varphi \rangle = \int_0^\infty E \varrho_{\chi,\varphi}(E) dE \qquad \text{for all} \quad \chi, \varphi \in \mathcal{S}(\mathbb{R}).$$

Proof. We have $\langle \chi | \mathsf{F}(H)\varphi \rangle = \int_{-\infty}^\infty F\left(\frac{\hbar^2 k^2}{2m}\right) \hat{\chi}^\dagger(k)\hat{\varphi}(k) dk$. This is equal to

$$\int_0^\infty F\left(\frac{\hbar^2 k^2}{2m}\right) \Big(\hat{\chi}^\dagger(k)\hat{\varphi}(k) + \hat{\chi}^\dagger(-k)\hat{\varphi}(-k) \Big)\, dk = \int_0^\infty F(E) \varrho_{\chi,\varphi}(E) dE.$$

□

The spectral family of the free Hamiltonian H_{free}. Let $\lambda \in \mathbb{R}$. Choosing the characteristic function e_λ of the interval $]-\infty, \lambda[$ (see (7.100) on page 497), we get

$$\langle \chi | \mathsf{E}_\lambda(H_{\mathrm{free}})\varphi \rangle = \int_0^\infty e_\lambda(E) \varrho_{\chi,\varphi}(E) dE \qquad \text{for all} \quad \chi, \varphi \in \mathcal{S}(\mathbb{R}).$$

In particular, if $\lambda \leq 0$, then $\mathsf{E}_\lambda(H_{\mathrm{free}}) = 0$.

Measurements of the energy. Let $\varphi \in \mathcal{S}(\mathbb{R})$ be a normalized state in the Hilbert space $L_2(\mathbb{R})$ (i.e., $\int_{\mathbb{R}} |\varphi(x)|^2 dx = 1$). This state describes a free quantum particle on the real line. Let $0 \leq E_0 < E_1 \leq \infty$. Then:

- Probability of measuring the energy of the particle in the interval$[E_0, E_1]$:

$$\int_{E_0}^{E_1} \varrho_{\varphi,\varphi}(E) dE.$$

- Mean energy of the particle: $\bar{E} = \int_0^\infty E \varrho_{\varphi,\varphi}(E) dE$.
- Square of the energy fluctuation: $(\Delta E)^2 = \int_0^\infty (E - \bar{E})^2 \varrho_{\varphi,\varphi}(E) dE$.

The Feynman Propagator Kernel

For all positions $x, x_0 \in \mathbb{R}$ and times $t > t_0$, define

$$\mathcal{K}(x,t;x_0,t_0) := \sqrt{\frac{m}{2\pi i\hbar(t-t_0)}} \cdot \mathrm{e}^{im(x-x_0)^2/2\hbar(t-t_0)}.$$

Let $\psi_0 \in \mathcal{S}(\mathbb{R})$. Then we have the following integral representation of the quantum dynamics:

$$\boxed{\big(P(t,t_0)\psi_0\big)(x) = \int_{\mathbb{R}} \mathcal{K}(x,t;x_0,t_0)\psi_0(x_0)dx_0, \qquad x \in \mathbb{R},\ \ t > t_0.}$$

This is the *key formula* for solving the initial-value problem for the instationary Schrödinger equation (7.126) on page 524. For all $\chi, \varphi \in \mathcal{S}(\mathbb{R})$, we obtain the kernel formula

$$\langle \chi | P(t-t_0)\varphi\rangle = \int_{\mathbb{R}^2} \chi(x)^\dagger \mathcal{K}(x,t;x_0,t_0)\varphi(x_0)dxdx_0, \qquad t > t_0.$$

For $t > t_0$, the function $(x,y) \mapsto \mathcal{K}(x,t;y,t_0)$ is called the Feynman propagator kernel of the free quantum particle.

The Euclidean Propagator Kernel

Set $P_{\mathrm{Euclid}}(t,t_0) := \mathrm{e}^{-(t-t_0)H_{\mathrm{free}}/\hbar}$. The operator

$$P_{\mathrm{Euclid}}(t,t_0) : L_2(\mathbb{R}) \to L_2(\mathbb{R}), \qquad t \geq t_0$$

is linear, continuous, and nonexpansive, that is, $||P_{\mathrm{Euclid}}(t,t_0)|| \leq 1$ for all $t \geq t_0$. For each given initial state $\psi_0 \in L_2(\mathbb{R})$ at time t_0, we set

$$\psi(t) := P_{\mathrm{Euclid}}(t,t_0)\psi_0, \qquad t \geq t_0.$$

If $\psi_0 \in \mathcal{S}(\mathbb{R})$, then the function $\psi : [0,\infty[\to L_2(\mathbb{R})$ is continuously differentiable, and we have the Euclidean Schrödinger equation

$$\hbar\dot{\psi}(t) = -H_{\mathrm{free}}\psi(t), \quad t > t_0, \quad \psi(t_0) = \psi_0. \tag{7.129}$$

The operator $P_{\mathrm{Euclid}}(t,t_0)$ is called the Euclidean propagator of the free quantum particle at time t (with respect to the initial time t_0). In terms of the unitary Fourier transform $\mathcal{F} : L_2(\mathbb{R}) \to L_2(\mathbb{R})$, the Euclidean propagator $P(t,t_0)$ corresponds to the multiplication with the function $k \mapsto \mathrm{e}^{-(t-t_0)E_k/\hbar}$ in the Fourier space. This means that, for all initial states $\psi_0 \in L_2(\mathbb{R})$, we get

$$P_{\mathrm{Euclid}}(t,t_0)\psi_0 = \mathcal{F}^{-1}M\mathcal{F}\psi_0, \qquad t \geq t_0$$

with the multiplication operator $(M\hat{\psi}_0)(k) := \mathrm{e}^{-(t-t_0)E_k/\hbar}\hat{\psi}_0(k)$ for all $k \in \mathbb{R}$. For all positions $x, x_0 \in \mathbb{R}$ and all times $t > t_0$, define

$$\mathcal{P}(x,t;x_0,t_0) = \sqrt{\frac{m}{2\pi\hbar(t-t_0)}} \cdot \mathrm{e}^{-m(x-x_0)^2/2\hbar(t-t_0)}.$$

Then we have the following integral representation:

$$\boxed{\big(P_{\text{Euclid}}(t,t_0)\psi_0\big)(x) = \int_{\mathbb{R}} \mathcal{P}(x,t;x_0,t_0)\psi_0(x_0)dx_0, \qquad x\in\mathbb{R},\ \ t>t_0.}$$

This is the key formula for solving the initial-value problem for the Euclidean Schrödinger equation (7.129). For all $\chi,\varphi\in L_2(\mathbb{R})$, we obtain the kernel formula

$$\langle\chi|P_{\text{Euclid}}(t,t_0)\varphi\rangle = \int_{\mathbb{R}^2} \chi(x)^\dagger \mathcal{P}(x,t;x_0,t_0)\varphi(x_0)dxdx_0, \qquad t>t_0.$$

For $t>t_0$, the function $(x,x_0)\mapsto \mathcal{P}(x,t;x_0,t_0)$ is called the Euclidean propagator kernel of the free quantum particle.

The Energetic Green's Function

The inhomogeneous stationary Schrödinger equation. Consider the inhomogeneous equation.

$$\boxed{-\frac{\hbar^2}{2m}\varphi''(x) = \mathcal{E}\varphi(x) + f(x), \qquad x\in\mathbb{R},} \tag{7.130}$$

which passes over to the stationary Schrödinger equation (7.115) if $f(x)\equiv 0$. Equation (7.130) corresponds to the operator equation

$$H_{\text{free}}\varphi - \mathcal{E}\varphi = f, \qquad \varphi\in D(H_{\text{free}}). \tag{7.131}$$

We want to solve this equation. Let $\mathcal{E}\in\varrho(H_{\text{free}})$ (i.e., $\mathcal{E}\in\mathbb{C}\setminus[0,\infty[$). Then the resolvent

$$(\mathcal{E}I - H_{\text{free}})^{-1} : L_2(\mathbb{R}) \to L_2(\mathbb{R})$$

exists as a linear continuous operator. For given $f\in L_2(\mathbb{R})$, the equation (7.131) has the unique solution

$$\varphi = (H_{\text{free}} - \mathcal{E}I)^{-1} f.$$

Von Neumann's operator calculus tells us that for all $\chi, f\in L_2(\mathbb{R})$, we have

$$\langle\chi|(H_{\text{free}} - \mathcal{E}I)^{-1}f\rangle = \int_0^\infty \frac{\varrho_{\chi,f}(E)}{E-\mathcal{E}}\,dE.$$

The retarded Green's function. Our goal is to represent the solution of the inhomogeneous Schrödinger equation (7.130) by an integral formula. To this end, we introduce the function

$$\mathcal{G}^+(x,y;\mathcal{E}) := \frac{\mathrm{i}m\cdot \mathrm{e}^{\mathrm{i}k|x-y|}}{\hbar^2 k}, \qquad x,y\in\mathbb{R}. \tag{7.132}$$

Here, $k := \sqrt{2m\mathcal{E}}/\hbar$. We assume that $\Im(\mathcal{E})>0$. The square root is to be understood as principal value. This choice of the complex energy $\mathcal{E}$ guarantees that the function $\mathcal{G}^+$ decays exponentially as $|x-y|\to\infty$.

Proposition 7.37 *Let $\Im(\mathcal{E})>0$. For given $f\in\mathcal{S}(\mathbb{R})$, the unique solution of the inhomogeneous Schrödinger equation (7.130) reads as*

$$\varphi(x) = \int_{\mathbb{R}} \mathcal{G}^+(x,y;\mathcal{E})f(y)dy, \qquad x\in\mathbb{R}. \tag{7.133}$$

The proof will be given in Sect. 8.5.2 on page 731. By Prop. 7.37, we get

$$\langle\chi|(H_{\text{free}} - \mathcal{E}I)^{-1}\varphi\rangle = \int_{\mathbb{R}^2} \chi(x)^\dagger \mathcal{G}^+(x,y;\mathcal{E})\varphi(y)dxdy, \qquad \Im(\mathcal{E}) > 0$$

for all $\chi, \varphi \in \mathcal{S}(\mathbb{R})$. Therefore the function $(x,y) \mapsto \mathcal{G}^+(x,y;\mathcal{E})$ is the kernel of the (negative) resolvent $(H_{\text{free}} - \mathcal{E}I)^{-1}$; this kernel is called the retarded (energetic) Green's function of the Hamiltonian H_{free}. Note that, for fixed $y \in \mathbb{R}$, the retarded Green's function behaves like

- $\mathrm{e}^{\mathrm{i}kx}$ as $x \to +\infty$, and
- $\mathrm{e}^{-\mathrm{i}kx}$ as $x \to -\infty$ where $k > 0$.

This corresponds to outgoing waves at infinity, $x = \pm\infty$.

The advanced Green's function. Now we pass from the positive wave number k to the negative wave number $-k$, that is, we change outgoing waves into ingoing waves at infinity. To this end, define

$$\mathcal{G}^-(x,y;\mathcal{E}) := -\frac{\mathrm{i}m \cdot \mathrm{e}^{-\mathrm{i}k|x-y|}}{\hbar^2 k}, \qquad x,y \in \mathbb{R}. \tag{7.134}$$

Here, $k := -\sqrt{2m\mathcal{E}}/\hbar$. We assume that $\Im(\mathcal{E}) < 0$. The square root is to be understood as principal value. This choice of the complex energy $\mathcal{E}$ guarantees that the function $\mathcal{G}^-$ decays exponentially as $|x-y| \to \infty$.

Proposition 7.38 *Let $\Im(\mathcal{E}) < 0$. For given $f \in \mathcal{S}(\mathbb{R})$, the unique solution of the inhomogeneous Schrödinger equation (7.130) reads as*

$$\varphi(x) = \int_{\mathbb{R}} \mathcal{G}^-(x,y;\mathcal{E})f(y)dy, \qquad x \in \mathbb{R}.$$

Thus, for all $\chi, \varphi \in \mathcal{S}(\mathbb{R})$ we obtain

$$\langle\chi|(H_{\text{free}} - \mathcal{E}I)^{-1}\varphi\rangle = \int_{\mathbb{R}^2} \chi(x)^\dagger \mathcal{G}^-(x,y;\mathcal{E})\varphi(y)dxdy.$$

This means that the function $(x,y) \mapsto \mathcal{G}^-(x,y;\mathcal{E})$ is the kernel of the (negative) resolvent $(H_{\text{free}} - \mathcal{E}I)^{-1}$; this kernel is called the advanced (energetic) Green's function of the Hamiltonian H_{free}. Note that, for fixed $y \in \mathbb{R}$, the advanced Green's function behaves like

- $\mathrm{e}^{-\mathrm{i}kx}$ as $x \to +\infty$ and
- $\mathrm{e}^{\mathrm{i}kx}$ as $x \to -\infty$ where $k > 0$.

This corresponds to incoming waves at infinity, $x = \pm\infty$.

The Fourier–Laplace transform of the Feynman propagator kernel. Fix the initial-time t_0. Then, for all times $t > t_0$, all positions $x, y \in \mathbb{R}$, and all complex energies $\mathcal{E}$ in the open upper half-pane (i.e., $\Im(\mathcal{E}) > 0$), we have

$$\mathcal{G}^+(x,y;\mathcal{E}) := \frac{\mathrm{i}}{\hbar}\int_{t_0}^{\infty} \mathrm{e}^{\mathrm{i}\mathcal{E}(t-t_0)/\hbar}\, \mathcal{K}(x,t;y,t_0)\, dt$$

together with the inverse formula

$$\mathcal{K}(x,t;y,t_0) = \frac{1}{2\pi\mathrm{i}} \cdot PV \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\mathcal{E}(t-t_0)/\hbar}\, \mathcal{G}^+(x,y;\mathcal{E})\, d\Re(\mathcal{E}).$$

The global energetic Green's function. The retarded Green's function is holomorphic in the open upper half-plane. This function can be analytically continued to a global analytic function on a double-sheeted Riemann surface. This global Green's function is given by

$$\mathcal{G}_{\text{global}}(x,y;\mathcal{E}) = \frac{\mathrm{i}m \cdot \mathrm{e}^{\mathrm{i}k(\mathcal{E})|x-y|}}{\hbar^2 k(\mathcal{E})}$$

where $k(\mathcal{E}) := \frac{\sqrt{2m}}{\hbar} \cdot \sqrt{\mathcal{E}}$. Here, the function $\mathcal{E} \mapsto k(\mathcal{E})$ has to be regarded as a global analytic function defined on the Riemann surface $\mathcal{R}$ of the square-root function $\sqrt{\ } : \mathcal{R} \to \mathbb{C}$. This Riemann surface will be studied in Sect. 8.3.5 on page 713. In terms of $\mathcal{R}$, the retarded (resp. advanced) Green's function is defined on the open upper (resp. lower) half-plane of the first sheet of the Riemann surface $\mathcal{R}$. The two functions jump along the positive real axis (see Fig. 8.6 on page 714).

Perturbation of the Free Quantum Dynamics

If the motion of the free particle on the real line is perturbed by the potential U, then we get the perturbed Schrödinger equation

$$\boxed{\mathrm{i}\hbar\psi_t(x,t) = -\frac{\hbar^2}{2m}\psi_{xx}(x,t) + U(x)\psi(x,t), \quad x \in \mathbb{R},\ t > t_0,\ \ \psi(x,t_0) = \psi_0.} \tag{7.135}$$

This is the prototype of a quantum system under interaction. Let us introduce the Hamiltonian

$$H\varphi := -\frac{\hbar^2}{2m}\frac{d^2\varphi}{dx^2} + U\varphi \qquad \text{for all} \quad \varphi \in W_2^2(\mathbb{R}).$$

In other words, $H = H_{\text{free}} + U$.

Theorem 7.39 *If the function $U : \mathbb{R} \to \mathbb{R}$ is smooth and has compact support, then the Hamiltonian $H : W_2^2(\mathbb{R}) \to L_2(\mathbb{R})$ is self-adjoint.*

Proof. Let $x \in \mathbb{R}$. Define the operator $C : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ by setting

$$(C\varphi)(x) := U(x)\varphi(x) \qquad \text{for all} \quad \varphi \in L_2(\mathbb{R}).$$

Then $||C\varphi|| \le \text{const} \cdot ||\varphi||$ for all $\varphi \in L_2(\mathbb{R})$. In fact,

$$\langle U\varphi | U\varphi \rangle = \int_{\mathbb{R}} \varphi(x)^\dagger U(x)^2 \varphi(x) dx \le \text{const} \int_{\mathbb{R}} |\varphi(x)|^2 dx.$$

Since the operator $H_{\text{free}} : W_2^2(\mathbb{R}) \to L_2(\mathbb{R})$ is self-adjoint, it follows from the Rellich–Kato perturbation theorem on page 502, that the perturbed operator $H = H_{\text{free}} + C$ is also self-adjoint on $W_2^2(\mathbb{R})$. □

A detailed study of equation (7.135) can be found in Chap. 8. This concerns the relation between scattering processes and bound states.

The Beauty of Harmonic Analysis

The motion of a free quantum particle is governed by the Fourier transform. Let us explain the relation to the translation group on the real line. For each $a \in \mathbb{R}$, the transformation

$$T_a x := x + a \qquad \text{for all} \quad x \in \mathbb{R}$$

represents a translation of the real line. For each smooth function $\psi : \mathbb{R} \to \mathbb{C}$, we define the operator

$$(\mathcal{T}_a \psi)(x) := \psi(T_a^{-1} x).$$

Explicitly, $\mathcal{T}_a \psi(x) = \psi(x - a)$. The operator D defined by

$$D\psi(x) := \lim_{a \to 0} \frac{\mathcal{T}_a \psi(x) - \psi(x)}{a} = -\psi'(x) \qquad \text{for all} \quad x \in \mathbb{R}$$

is called the infinitesimal translation. By Taylor expansion,

$$\mathcal{T}_a \psi(x) = \psi(x) + D\psi(x) + \tfrac{1}{2} D^2 \psi(x) + \tfrac{1}{3!} D^3 \psi(x) + \dots$$

The Fourier transform is related to the eigenfunctions $\chi_k(x) := \frac{\mathrm{e}^{\mathrm{i}kx}}{\sqrt{2\pi}}$ of the infinitesimal operator D. Explicitly,

$$\mathrm{i}\hbar D \chi_k = \hbar k \chi_k, \qquad k \in \mathbb{R}.$$

Note that $\mathrm{i}\hbar D$ corresponds to the momentum operator on the real line. If we replace the translation group by another Lie group, then we get a generalization of the preceding situation which leads to

- more general infinitesimal transformations (differential operators),
- more general eigenfunctions (special functions of mathematical physics),
- and a generalization of the Fourier transform.

This is the subject of a beautiful branch in mathematics called harmonic analysis, which will be encountered quite often in this treatise. In the 20th century, the protagonist of harmonic analysis was Hermann Weyl (1885–1955). We recommend:

> G. Mackey, The Scope and History of Commutative and Noncommutative Harmonic Analysis, Amer. Math. Soc., Providence, Rhode Island, 1992.
>
> G. Mackey, Induced Representations of Groups and Quantum Mechanics, Benjamin, New York, 1968.
>
> G. Mackey, Unitary Group Representations in Physics, Probability, and Number Theory, Benjamin, Reading, Massachusetts, 1978.

7.6.6 The Rescaled Fourier Transform

> The rescaled Fourier transform fits best the duality between position and momentum of quantum particles in the setting of the Dirac calculus.
>
> Folklore

Introducing the function $\varphi_p(x) := \frac{\mathrm{e}^{\mathrm{i}px/\hbar}}{\sqrt{2\pi\hbar}}$ for all $x \in \mathbb{R}$, we obtain the key relation

$$-\mathrm{i}\hbar \frac{d\varphi_p}{dx} = p\varphi_p \qquad \text{for all} \quad p \in \mathbb{R}.$$

That is, the function φ_p is a generalized eigenfunction of the momentum operator with the momentum p as eigenvalue. The normalization is dictated by the Parseval

equation (7.138) below. Let $\varphi, \psi \in \mathcal{S}(\mathbb{R})$. The rescaled Fourier transform is given by the following two formulas

$$\tilde{\varphi}(p) = \int_{\mathbb{R}} \varphi_p^\dagger(x)\varphi(x)dx \qquad \text{for all} \quad p \in \mathbb{R} \tag{7.136}$$

and

$$\varphi(x) = \int_{\mathbb{R}} \varphi_p(x)\tilde{\varphi}(p)dp \qquad \text{for all} \quad x \in \mathbb{R} \tag{7.137}$$

together with the Parseval equation

$$\int_{\mathbb{R}} \psi(x)^\dagger \varphi(x)dx = \int_{\mathbb{R}} \tilde{\psi}(p)^\dagger \tilde{\varphi}(p)\, dp. \tag{7.138}$$

The classical Fourier transform is obtained by choosing $\hbar := 1$. Setting $\mathcal{F}_\hbar \varphi := \tilde{\varphi}$, we obtain the linear, bijective, sequentially continuous operator

$$\mathcal{F}_\hbar : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$$

which is called the rescaled Fourier transform. As in Sect. 7.6.4, this operator can be extended to a linear bijective operator

$$\mathcal{F}_\hbar : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$$

such that the restriction $\mathcal{F}_\hbar : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ is unitary. The commutative diagram

$$\begin{array}{ccc} D(P) & \xrightarrow{P} & L_2(\mathbb{R}) \\ {\scriptstyle \mathcal{F}_\hbar}\downarrow & & \downarrow{\scriptstyle \mathcal{F}_\hbar} \\ D(Q) & \xrightarrow{Q} & L_2(\mathbb{R}) \end{array}$$

tells us that the momentum operator P and the position operator Q are unitarily equivalent. According to Dirac, for fixed momentum $p \in \mathbb{R}$, we introduce the momentum costate $\langle p|$ by setting

$$\boxed{\langle p|(\varphi) := \int_{\mathbb{R}} \varphi_p^\dagger(x)\varphi(x)dx, \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).}$$

Mnemonically, we write this as $\langle p|\varphi\rangle$. Replacing the wave number costate $\langle k|$ from Sect. 7.6.4 by the momentum costate $\langle p|$, we get the following formulas of the Dirac calculus:

- $\langle p|\varphi\rangle = \tilde{\varphi}(p)$,
- $I = \int_{\mathbb{R}} |p\rangle\langle p|\, dp$,
- $P_{\text{pre}}^d \langle p| = p\, \langle p|$,
- $H_{\text{pre}}^d \langle p| = E(p)\, \langle p|$ with the energy value $E(p) := \frac{p^2}{2m}$.

The system $\{\langle p|\}_{p\in\mathbb{R}}$ forms a complete orthonormal system of costates for both the momentum operator P_{pre} and the free Hamiltonian H_{pre}. Adding the mnemonical formulas

- $\langle x|\varphi\rangle = \varphi(x)$ and $\langle x|p\rangle = \varphi_p(x)$,
- $I = \int_{\mathbb{R}} |x\rangle\langle x|\, dx$,

as well as $\langle a|b\rangle^\dagger = \langle b|a\rangle$, we automatically obtain

$$\boxed{\langle p|\varphi\rangle = \int_{\mathbb{R}} \langle p|x\rangle\langle x|\varphi\rangle \, dx, \qquad \langle x|\varphi\rangle = \int_R \langle x|p\rangle\langle p|\varphi\rangle \, dp}$$

which is the rescaled Fourier transform (7.136), (7.137) above. Similarly, the Parseval equation (7.138) above is obtained by

$$\boxed{\langle \psi|\varphi\rangle = \int_{\mathbb{R}} \langle \psi|x\rangle\langle x|\varphi\rangle \, dx = \int_{\mathbb{R}} \langle \psi|p\rangle\langle p|\varphi\rangle \, dp.}$$

This shows that the rescaled Fourier transform is nothing else than a change from the position coordinate x to the momentum coordinate p which respects "inner products."

Note that, as a rule, physicists use the wave number costates $\langle k|$ in scattering theory, and the momentum costates $\langle p|$ in the Feynman path integral approach. We will follow this convention.

7.6.7 The Quantized Harmonic Oscillator and the Maslov Index

> The global behavior of the quantized harmonic oscillator is governed by the Morse indices (also called Maslov indices) of the classical harmonic oscillator.
>
> Folklore

Let us continue the study of the quantized harmonic oscillator on the real line started in Sect. 7.4.4 on page 467. The initial-value problem for the corresponding Schrödinger equation reads as

$$\mathrm{i}\hbar\psi_t(x,t) = -\frac{\hbar^2}{2m}\psi_{xx}(x,t) + \frac{m\omega^2 x^2}{2}\,\psi(x,t), \quad \psi(x,t_0) = \psi_0(x) \tag{7.139}$$

for all position coordinates $x \in \mathbb{R}$ and all times $t > t_0$. Let us introduce the pre-Hamiltonian $H_{\mathrm{pre}} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ by setting

$$\big(H_{\mathrm{pre}}\varphi\big)(x) := -\frac{\hbar^2}{2m}\frac{d^2\varphi(x)}{dx^2} + \frac{m\omega^2 x^2}{2}\,\varphi(x), \qquad x \in \mathbb{R}.$$

By Sect. 7.4.4, the equation $H_{\mathrm{pre}}\varphi = E\varphi$ has the eigensolutions (φ_n, E_n) with the energy eigenvalues $E_n = \hbar\omega(n + \frac{1}{2})$ and the eigenfunctions

$$\varphi_n(x) = \frac{1}{\sqrt{2^n n! x_0 \sqrt{\pi}}}\; H_n\left(\frac{x}{x_0}\right)\; \exp\left\{-\frac{1}{2}\left(\frac{x}{x_0}\right)^2\right\}, \qquad n = 0, 1, 2, \ldots,$$

where $x_0 := \sqrt{\frac{\hbar}{m\omega}}$. Here, $H_0, H_1, H_2, \ldots$ are the Hermite polynomials introduced on page 436. Furthermore, $\varphi_n \in \mathcal{S}(\mathbb{R})$ for all n. We will use the Hilbert space $L_2(\mathbb{R})$ with the inner product

$$\langle \chi|\varphi\rangle := \int_{\mathbb{R}} \chi^\dagger(x)\varphi(x)dx, \qquad \chi, \varphi \in L_2(\mathbb{R}).$$

For introducing operator kernels, we will also use the Hilbert space $L_2(\mathbb{R}^2)$ equipped with the inner product

$$\langle \mathcal{A}|\mathcal{B}\rangle_{L_2(\mathbb{R}^2)} := \int_{\mathbb{R}^2} \mathcal{A}(x,y)^\dagger \mathcal{B}(x,y)dxdy, \qquad \mathcal{A}, \mathcal{B} \in L_2(\mathbb{R}^2).$$

(i) The self-adjoint Hamiltonian H: The point is that the eigenfunctions $\varphi_0, \varphi_1, \ldots$ form a complete orthonormal system in the Hilbert space. The pre-Hamiltonian H_{pre} can be extended to the self-adjoint operator $H : D(H) \to L_2(\mathbb{R})$ given by

$$H\varphi := \sum_{n=0}^{\infty} E_n \langle \varphi_n | \varphi \rangle \varphi_n .$$

Here, $\varphi \in D(H)$ iff this series is convergent in the Hilbert space $L_2(\mathbb{R})$, that is, $\sum_{n=0}^{\infty} E_n^2 |\langle \varphi_n | \varphi \rangle|^2 < \infty$. The operator H is called the Hamiltonian of the quantized harmonic oscillator.

(ii) The spectrum of the Hamiltonian H: The spectrum $\sigma(H)$ consists of the energy values $E_0, E_1, E_2, \ldots$ of the quantized harmonic oscillator. This is a pure point spectrum; the absolutely continuous spectrum, the essential spectrum, and the singular spectrum of H are empty.

(iii) The kernel theorem: Let $\lambda_0, \lambda_1, \ldots$ be complex numbers. Consider the operator $A : D(A) \to L_2(\mathbb{R})$ given by

$$A\varphi = \sum_{n=0}^{\infty} \lambda_n \langle \varphi_n | \varphi \rangle \varphi_n . \tag{7.140}$$

We assume that the domain of definition $D(A)$ consists of all the functions $\varphi \in L_2(\mathbb{R})$ for which the series on the right-hand side of (7.140) is convergent in $L_2(\mathbb{R})$, that is, $\varphi \in D(A)$ iff $\sum_{k=0}^{\infty} |\lambda_n \langle \varphi_n | \varphi \rangle|^2 < \infty$.

Theorem 7.40 (a) *Hilbert–Schmidt operator with $L_2(\mathbb{R}^2)$-kernel: If*

$$\sum_{n=0}^{\infty} |\lambda_n|^2 < \infty,$$

then the operator $A : X \to X$ defined by (7.140) is linear, continuous, and compact. The series

$$\mathcal{A}(x,y) := \sum_{n=0}^{\infty} \lambda_n \varphi_n(x) \varphi_n(y)^{\dagger}, \qquad (x,y) \in \mathbb{R}^2 \tag{7.141}$$

is convergent in the Hilbert space $L_2(\mathbb{R}^2)$, and the operator A has the $L_2(\mathbb{R}^2)$-kernel $\mathcal{A}$. That is, for all $\varphi, \chi \in L_2(\mathbb{R})$, we have

$$(A\varphi)(x) = \int_{\mathbb{R}} \mathcal{A}(x,y) \varphi(y) dy, \qquad x \in \mathbb{R},$$

together with the bilinear form

$$\langle \chi | A\varphi \rangle = \int_{\mathbb{R}^2} \chi(x)^{\dagger} \mathcal{A}(x,y) \varphi(y) dx dy. \tag{7.142}$$

If all the numbers $\lambda_0, \lambda_1, \ldots$ are real, then the operator A is self-adjoint.

(b) *Trace-class operator: If $\sum_{n=0}^{\infty} |\lambda_n| < \infty$, then* (i) *is valid. The operator $A : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ is called a trace class (or nuclear) operator; its trace is given by* $\operatorname{tr}(A) = \sum_{n=0}^{\infty} \lambda_n$.[83]

[83] The general definition of Hilbert–Schmidt operators and trace-class operators will be given in Sect. 7.16.4 on page 629.

(c) *The Schwartz kernel T: If the condition* $\sup_n |\lambda_n| < \infty$ *is satisfied, then the operator* $A : X \to X$ *is linear and continuous. There exists a uniquely determined tempered distribution* $T \in \mathcal{S}'(\mathbb{R}^2)$ *such that*

$$\langle \chi | A\varphi \rangle = T(\chi^\dagger \otimes \varphi) \qquad \text{for all} \quad \chi, \varphi \in \mathcal{S}(\mathbb{R}).$$

More precisely, there exist a continuous function $\mathcal{A} : \mathbb{R}^2 \to \mathbb{C}$ *of polynomial growth and nonnegative integers* r *and* s *such that*

$$T(\chi^\dagger \otimes \varphi) = \int_{\mathbb{R}^2} \chi^{(r)}(x)^\dagger \mathcal{A}(x,y)\varphi^{(s)}(y)dxdy \quad \text{for all} \ \ \chi, \varphi \in \mathcal{S}(\mathbb{R}).$$

Proof. Ad (a). Since the functions $\varphi_0, \varphi_1, \ldots$ form a complete orthonormal system in the Hilbert space $L_2(\mathbb{R})$, the tensor products

$$(\varphi_k^\dagger \otimes \varphi_l)(x,y) := \varphi_k(x)^\dagger \varphi(y), \qquad (x,y) \in \mathbb{R}^2, \quad k,l = 0,1,\ldots$$

represent a complete orthonormal system in the Hilbert space $L_2(\mathbb{R}^2)$ (see Zeidler (1995a), p. 224). Consequently, the series (7.141) is convergent in $L_2(\mathbb{R}^2)$ iff $\sum_{n=0}^\infty |\lambda_n|^2 < \infty$. The remaining claims are standard results in functional analysis (see Zeidler (1995a), Sect. 4.4).
Ad (b). If $\sum_{n=0}^\infty |\lambda_n| < \infty$, the $\lim_{n\to\infty} \lambda_n = 0$. Consequently, there exists a natural number n_0 such that $\sum_{n=n_0}^\infty |\lambda_n|^2 \le \sum_{n=n_0}^\infty |\lambda_n|$.
Ad (c). This is the Schwartz kernel theorem. The proof can be found in I. Gelfand and N. Vilenkin, Generalized Functions, Vol. 4, Sect. I.1.3, Academic Press, New York, 1964. □

(iv) The resolvent and the energetic Green's function of the Hamiltonian H: Let the complex number $\mathcal{E}$ be different from all the eigenvalues $E_0, E_1, \ldots$ Introduce $G(\mathcal{E}) := (H - \mathcal{E}I)^{-1}$. Then the energetic Green's operator

$$G(\mathcal{E}) : L_2(\mathbb{R}) \to L_2(\mathbb{R})$$

is linear and continuous. Explicitly,

$$G(\mathcal{E})\varphi = \sum_{n=0}^\infty \frac{\langle \varphi_n | \varphi \rangle}{E_n - \mathcal{E}} \, \varphi_n, \qquad \varphi \in L_2(\mathbb{R}).$$

The operator $G(\mathcal{E})$ has an $L_2(\mathbb{R}^2)$-kernel called the energetic Green's function of the quantized harmonic oscillator. Explicitly,

$$\mathcal{G}(x,y;\mathcal{E}) = \sum_{n=0}^\infty \frac{\varphi_n(x)\varphi_n(y)^\dagger}{E_n - \mathcal{E}}, \qquad x,y \in \mathbb{R}.$$

This series is convergent in $L_2(\mathbb{R}^2)$. For all $x \in \mathbb{R}$, we have

$$(G(\mathcal{E})\varphi)(x) = \int_{\mathbb{R}} \mathcal{G}(x,y;\mathcal{E})\varphi(y)dy.$$

The operator $R(\mathcal{E}) := -G(\mathcal{E})$ is called the resolvent of H.

(v) The Euclidean propagator kernel: Let $t > t_0$. Set $\beta := (t-t_0)/\hbar$. Since the series $\sum_{n=0}^{\infty} \mathrm{e}^{-\beta E_n}$ is convergent, it follows from Theorem 7.40(ii) that the Euclidean propagator $P_{\mathrm{Euclid}}(t, t_0) := \mathrm{e}^{-\beta H}$ is a trace-class operator on $L_2(\mathbb{R})$, and it has an $L_2(\mathbb{R}^2)$-kernel given by the series

$$\mathcal{P}(x,t;y,t_0) := \sum_{n=0}^{\infty} \mathrm{e}^{-\beta E_n} \varphi_n(x) \varphi_n(y)^\dagger,$$

which is convergent in the Hilbert space $L_2(\mathbb{R}^2)$.

Proposition 7.41 *For all positions $x, y \in \mathbb{R}$ and all times $t > 0$, the Euclidean propagator kernel reads as*

$$\mathcal{P}(x,t;y,0) = \frac{1}{x_0\sqrt{2\pi \sinh \omega t}} \exp\left\{ -\frac{(x^2+y^2)\cosh \omega t - 2xy}{2x_0^2 \sinh \omega t} \right\}.$$

For $t > t_0$, we get $\mathcal{P}(x,t;y,t_0) = \mathcal{P}(x,t-t_0;y,0)$.

Proof. This is the classical Mehler formula for Hermite polynomials which can be found in A. Erdéley et al. (Eds.), Higher Transcendental Functions, Vol. III, McGraw-Hill, New York, 2006. Explicitly, the Mehler formula reads as

$$\begin{aligned} &\frac{1}{\sqrt{1-z^2}} \exp\left\{ -\frac{1}{2(1-z^2)} [(x^2+y^2)(1+z^2) - 4xyz] \right\} \\ &\qquad = \exp\left(-\frac{x^2}{2} - \frac{y^2}{2} \right) \sum_{n=0}^{\infty} \frac{z^n}{2^n n!} H_n(x) H_n(y) \end{aligned} \tag{7.143}$$

for all $x, y \in \mathbb{R}$ and all complex numbers z with $|z| < 1$. □

We will see in Sect. 7.6.8 that the Euclidean propagator of a single harmonic oscillator governs the thermodynamics of an ideal gas if we set $\beta := 1/kT$ where T is the temperature and k is the Boltzmann constant.

(vi) The generalized Feynman propagator kernel and the Maslov indices: We want to show that analytic continuation of the Euclidean propagator kernel yields the function

$$\boxed{\mathcal{K}(x,t;y,0) := \frac{\mathrm{e}^{-\mathrm{i}\pi/4}\, \mathrm{e}^{-\mathrm{i}\pi\mu(0,t)/2}}{x_0\sqrt{2\pi|\sin \omega t|}} \exp\left(\mathrm{i} \frac{(x^2+y^2)\cos \omega t - 2xy}{2x_0^2 \sin \omega t} \right).} \tag{7.144}$$

This so-called Feynman–Souriau formula is valid for both

- all positions $x, y \in \mathbb{R}$ and
- all non-critical times $t \in]t_{n,\mathrm{crit}}, t_{n+1,\mathrm{crit}}[$ with $n = 0, 1, 2, ...$

Here, the critical times are given by $t_{n,\mathrm{crit}} := \frac{n\pi}{\omega}$. The Maslov index is defined by

$$\mu(0,t) := n \qquad \text{for all} \quad t \in]t_{n,\mathrm{crit}},\ t_{n+1,\mathrm{crit}}[. \tag{7.145}$$

For all $t > t_0$, we set $\mathcal{K}(x,t;y,t_0) := \mathcal{K}(x,t-t_0;y,0)$. The function $\mathcal{K}$ is called the generalized Feynman propagator kernel (or briefly the Feynman propagator kernel) of the quantized harmonic oscillator. The additional factors

$$\mathrm{e}^{-\mathrm{i}\pi/4}\, \mathrm{e}^{-\mathrm{i}\pi\mu(0,t)/2} \tag{7.146}$$

appearing in (7.144) are called the critical Maslov phase factors. In terms of mathematics, in the following proof we will show that these phase factors are obtained in a natural way by means of analytic continuation. In terms of physics, we will show below that the Maslov phase factors are closely related to causality.

Proof. To simplify notation, we set $\omega := 1$. In order to find the analytic continuation, we replace the real variable t by the complex variable z. This way, using Prop. 7.41 we get

$$\mathcal{P}(x,z;y,0) = \frac{1}{x_0\sqrt{2\pi\sinh z}} \exp\left\{-\frac{(x^2+y^2)\cosh z - 2xy}{2x_0^2\sinh z}\right\}.$$

Now set $z := \mathrm{i}t$. Then $\sinh z = \mathrm{i}\sin t$ and $\cosh z = \cos t$ for all $t \in \mathbb{C}$. Suppose that

$$t_{n,\mathrm{crit}} < t < t_{n+1,\mathrm{crit}}, \qquad n = 0, 1, 2, \ldots$$

Then $\sin t = (-1)^n |\sin t|$. Considering the square-root function on its Riemann surface (see Fig. 8.6 on page 714), we obtain

$$\sqrt{\mathrm{i}\sin t} = \sqrt{(-1)^n \mathrm{i}|\sin t|} = \sqrt{\mathrm{e}^{\mathrm{i}n\pi}\mathrm{e}^{\mathrm{i}\pi/2}|\sin t|} = \mathrm{e}^{\mathrm{i}n\pi/2}\mathrm{e}^{\mathrm{i}\pi/4}\sqrt{|\sin t|}.$$

This yields the claim (7.144). □

Focal points and the Morse index (Maslov index). We want to show that the singularities of the Feynman propagator kernel $\mathcal{K}(x,t;y,t_0)$ are related to the Morse indices of focal points in classical mechanics. To this end, consider a harmonic oscillator of mass $m > 0$ and angular frequency $\omega > 0$ on the real line. The classical equation of motion reads as

$$m\ddot{q}(\tau) + \omega^2 q(\tau) = 0, \qquad \tau \in \mathbb{R}, \qquad q(0) = q_0, \;\; \dot{q}(0) = q_1$$

with the characteristic length $x_0 := \sqrt{\hbar/m\omega}$. In Sect. 6.5.4, we have introduced the crucial Morse (or Maslov) index which coincides with (7.145) above. Explicitly, the critical points in time are characterized by the fact that the boundary value problem

$$\ddot{q}(t) + \omega^2 q(t) = 0, \qquad 0 < t < t_{n,\mathrm{crit}}, \qquad q(0) = q(t_{n,\mathrm{crit}}) = 0$$

has not only the trivial solution $q(t) \equiv 0$, but also a nontrivial solution, namely, $q(t) := \sin \omega t$. Observe that the function $\mathcal{P}$ has singularities precisely at the critical points in time, since $\sin \omega t_{n,\mathrm{crit}} = 0$. Moreover, the Morse index $\mu(0,t)$ jumps at the critical points in time.

The Feynman propagator kernel $\mathcal{K}(x,t;y,t_0)$ of the quantized harmonic oscillator contains information about the global behavior of the classical harmonic oscillator.

This phenomenon is typical for the quantization of classical dynamical systems.[84]

Causality and the motivation of the Maslov phase factors. Using the Dirac delta function in a formal way, we want to motivate formula (7.146) above in terms of physics. To simplify notation, let us use the convention $\omega = \hbar = m := 1$. Hence $x_0 = 1$. The starting point is the product formula (7.90) for the propagator kernel, that is,

[84] See M. Gutzwiller, Chaos in Classical and Quantum Mechanics, Springer, New York, 1990.

$$\mathcal{K}(x,t;y,0)=\int_{\mathbb{R}}\mathcal{K}(x,t-\tau;z,0)\mathcal{K}(z,\tau;y,0)\,dz \tag{7.147}$$

which is based on the causality relation $\mathrm{e}^{-\mathrm{i}tH}=\mathrm{e}^{-\mathrm{i}(t-\tau)H}\mathrm{e}^{-\mathrm{i}\tau H}$.

(I) Consider the first critical time interval $0<t<t_{1,\mathrm{crit}}$ with $t_{1,\mathrm{crit}}=\pi$. Then , then analytic continuation of the Euclidean propagator $\mathcal{P}$ from Prop. 7.41 yields the regular Feynman propagator kernel

$$\mathcal{K}(x,t;y,0)=\frac{\mathrm{e}^{-\mathrm{i}\pi/4}}{\sqrt{2\pi\sin t}}\;\exp\left(\mathrm{i}\frac{(x^2+y^2)\cos t-2xy}{2\sin t}\right),\quad 0<t<\pi.$$

Let us now study the limit $t\to\pi-0$. If $t=\frac{\pi}{2}$, then

$$\mathcal{K}\left(x,\frac{\pi}{2};y,0\right)=\frac{\mathrm{e}^{-\mathrm{i}\pi/4}\mathrm{e}^{-\mathrm{i}xy}}{\sqrt{2\pi}}.$$

By the product rule (7.147), we get

$$\begin{aligned}\lim_{t\to\pi-0}\mathcal{K}(x,t;y,0):&=\int_{\mathbb{R}}\mathcal{K}\left(x,\frac{\pi}{2};z,0\right)\mathcal{K}\left(z,\frac{\pi}{2};y,0\right)dz\\&=\mathrm{e}^{-\mathrm{i}\pi/2}\cdot\frac{1}{2\pi}\int_{\mathbb{R}}\mathrm{e}^{-\mathrm{i}(x+y)z}dz=\mathrm{e}^{-\mathrm{i}\pi/2}\delta(x+y).\end{aligned}$$

(II) Now consider the second critical time interval $\pi<t<2\pi$. We want to define the propagator kernel on the interval $]\pi,2\pi[$ in such a way that

$$\lim_{t\to\pi+0}\mathcal{K}(x,t;y,0)=\lim_{t\to\pi-0}\mathcal{K}(x,t;y,0)=\mathrm{e}^{-\mathrm{i}\pi}\delta(x+y).$$

The appropriate definition looks like

$$\mathcal{K}(x,t;y,0):=\frac{\mathrm{e}^{-\mathrm{i}\pi/4}\mathrm{e}^{-\mathrm{i}\pi/2}}{\sqrt{2\pi|\sin t|}}\;\exp\left(\mathrm{i}\frac{(x^2+y^2)\cos t-2xy}{2\sin t}\right),\qquad \pi<t<2\pi.$$

To see this, set $t:=\pi+\tau$. Using $\sin(\pi+\tau)=-\sin\tau$ together with $\lim_{\tau\to0}\frac{\sin\tau}{\tau}=1$, we obtain

$$\lim_{\tau\to+0}\mathcal{K}(x,\pi+\tau;y,0)=\mathrm{e}^{-\mathrm{i}\pi/2}\lim_{\tau\to+0}\frac{\mathrm{e}^{-\mathrm{i}\pi/4}\mathrm{e}^{\mathrm{i}(x+y)^2/2\tau}}{\sqrt{2\pi\tau}}=\mathrm{e}^{-\mathrm{i}\pi/2}\delta(x+y).$$

The latter limit follows from

$$\lim_{\tau\to+0}\mathcal{K}_{\mathrm{free}}(z,\tau;0,0)=\delta(z)$$

for the propagator kernel $\mathcal{K}_{\mathrm{free}}(z,\tau;0,0)=\mathrm{e}^{-\mathrm{i}\pi/4}\cdot\frac{\mathrm{e}^{\mathrm{i}z^2/2\tau}}{\sqrt{2\pi\tau}}$ of a free quantum particle on the real line.

(III) Similarly, we extend the definition of the propagator kernel $\mathcal{K}$ to the other critical time intervals. □

Using the theory of distributions, the formal argument above can be reformulated in terms of rigorous mathematics.

7.6.8 Ideal Gases and von Neumann's Density Operator

> The statistical physics of the multi-particle system of N harmonic oscillators is governed by the Euclidean propagator of a single harmonic oscillator.
>
> Folklore

We want to explain the following fundamental principle in physics:

> *In order to pass from quantum mechanics to statistical physics, apply the replacement*
>
> $$\frac{\mathrm{i}t}{\hbar} \mapsto \frac{1}{kT}.$$
>
> *Here, we use the following notation: t time, T absolute temperature, h Planck's quantum of action, $\hbar = h/2\pi$, and k Boltzmann constant.*

It turns out that the computation methods in statistical physics are frequently easier to handle than the corresponding methods in quantum mechanics. The reason is that, for $T > 0$ and $t > 0$, the integral

$$\int_0^\infty \mathrm{e}^{-E/kT} dE$$

is well-defined whereas the oscillating integral

$$\int_0^\infty \mathrm{e}^{-\mathrm{i}Et/\hbar} dE$$

does not exist. The Euclidean trick in physics is to start with imaginary time $t = -\mathrm{i}\tau$. Then $\mathrm{i}t = \tau$ is real. At the end of the computation, one performs an analytic continuation to real time t, if possible. Fortunately enough, this trick works well in many cases.

A gas of quantum particles on the real line. The following situation is the prototype of quantum statistics. Consider a large fixed number of N identical quantum particles (bosons) on the real line which are harmonic oscillators of mass m and fixed angular frequency $\omega > 0$. To simplify notation, physicists introduce the quantity

$$\boxed{\beta := \frac{1}{kT}}$$

in statistical physics. Here, T is the absolute temperature of the gas, and k is the universal Boltzmann constant. The physical dimension of kT is energy. For studying the physics of the gas, the following two quantities

$$x_0 := \sqrt{\frac{\hbar}{m\omega}}, \qquad \beta\hbar\omega = \frac{\hbar\omega}{kT}$$

are important. Here, x_0 has the physical dimension of length, and $\beta\hbar\omega$ is dimensionless. It is our aim to compute the following physical quantities of the gas at the temperature $T > 0$.

(i) Total energy of the gas:

$$\boxed{E = N\bar{E} = N\hbar\omega\left(\frac{1}{2} + \frac{1}{\mathrm{e}^{\beta\hbar\omega} - 1}\right).}$$

(ii) Relative energy fluctuations:

$$\frac{\Delta E}{E} = \frac{\Delta \bar{E}}{\bar{E}\sqrt{N}} = \frac{1}{\sqrt{N}\cosh\frac{\beta\hbar\omega}{2}}.$$

For large particle number N, the relative energy fluctuations are small, as expected by experience for gases in daily life.

(iii) Mass density of the gas:

$$\mu(x,T) = Nm\varrho(x,T) = \frac{Nm}{x_0}\sqrt{\frac{\tanh\frac{\beta\omega\hbar}{2}}{\pi}}\ \exp\left\{\frac{x^2(1-\cosh\beta\hbar\omega)}{x_0^2\sinh\beta\hbar\omega}\right\}.$$

Here, the density function $\varrho(x,T) := \langle x|\varrho(T)|x\rangle$ is related to von Neumann's density operator $\varrho(T)$. The derivative of energy with respect to temperature,

$$C(T) = E_T(T,N),$$

is called the heat capacity of the gas. A small change ΔT of temperature produces the following amount of heat,

$$\Delta Q = C(T)\Delta T.$$

The heat capacity can be measured in physical experiments. We will compute below the mean energy $\bar{E}$ and the mean energy fluctuation $\Delta\bar{E}$ of one particle. For the total energy, this yields $E = N\bar{E}$. Moreover, we assume that the single particles behave independently. Then, by the theory of probability, the total energy dispersion is additive,

$$(\Delta E)^2 = (\Delta\bar{E})^2 + \ldots + (\Delta\bar{E})^2 = N(\Delta\bar{E})^2.$$

Hence $\Delta E/E = \Delta\bar{E}/\bar{E}\sqrt{N}$.

Bose–Einstein condensation. To understand the physics of our gas, let us consider the two important special cases of high temperature and low temperature.

(H) For high temperature T (i.e., β is small), we get up to terms of lower order:[85]

$$E = NkT, \qquad \frac{\Delta E}{E} = \frac{1}{\sqrt{N}}, \qquad \mu(x,T) = \frac{Nm}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2\sigma^2}.$$

The mass density function μ is a Gaussian distribution with mean fluctuation $\sigma := x_0/\sqrt{\beta\hbar\omega}$. The energy law, $E = N\cdot kT$, is a special case of the classical Boltzmann law of energy equipartition. This law tells us that, for many-particle systems at high temperature, each degree of freedom contributes the amount of mean energy kT to the total energy of the system. For the heat capacity of the gas, we get $C = Nk$.

(L) For low temperature T, we obtain:

$$\lim_{T\to+0} E = \tfrac{1}{2}\hbar\omega N, \qquad \lim_{T\to+0}\frac{\Delta E}{E} = 0.$$

As expected, the particle energy is equal to the ground state energy of the harmonic oscillator. Physicists say that the excited energy states are frozen at low temperatures. This crucial phenomenon is called Bose-Einstein condensation.[86]

[85] Note that $\sinh\beta\hbar\omega = \beta\hbar\omega + O(\beta^2)$ and $\cosh\beta\hbar\omega = 1 + \beta^2\hbar^2\omega^2 + O(\beta^4)$ as $\beta\to 0$.

[86] In 2001, Eric Cornell, Wolfgang Ketterle, and Carl Wieman were awarded the Nobel prize in physics for the experimental achievement of Bose–Einstein condensation in dilute gases of alkali atoms, and for fundamental studies of the properties of the condensates.

Note that the behavior of the gas at low temperatures is governed by typical quantum effects.

The partition function. The possible energies of the gas particles are given by

$$E_n = \hbar\omega(n + \tfrac{1}{2}), \qquad n = 0, 1, 2, , ...$$

By statistical physics, the physical properties of this many-particle system follow from the partition function

$$Z(\beta) := \sum_{n=0}^{\infty} \mathrm{e}^{-\beta E_n}.$$

Recall that $\beta := 1/kT$. For a single particle, the probability of having the energy E_n is equal to

$$p_n := \frac{\mathrm{e}^{-\beta E_n}}{Z(\beta)}.$$

This yields the mean energy $\bar{E}$ and the mean energy fluctuation $\Delta\bar{E} \geq 0$ of a single particle, namely,

$$\bar{E} = \sum_{n=0}^{\infty} E_n p_n, \qquad (\Delta\bar{E})^2 = \sum_{n=0}^{\infty} (E_n - \bar{E})^2 p_n.$$

We claim that

$$\boxed{\bar{E} = \hbar\omega \left(\frac{1}{2} + \frac{1}{\mathrm{e}^{\beta\hbar\omega} - 1} \right), \qquad \frac{\Delta\bar{E}}{\bar{E}} = \frac{1}{\cosh \frac{\beta\hbar\omega}{2}}.} \tag{7.148}$$

Proof. By the geometric series $1 + q + q^2 + ... = \frac{1}{1-q}$ for $|q| < 1$, we get

$$Z(\beta) = \sum_{n=0}^{\infty} \mathrm{e}^{-\beta E_n} = \frac{\mathrm{e}^{-\beta\hbar\omega/2}}{1 - \mathrm{e}^{-\beta\hbar\omega}} = \frac{1}{2\sinh \frac{\beta\hbar\omega}{2}}.$$

Observe now that

$$\bar{E} = -\frac{Z'(\beta)}{Z(\beta)}, \qquad \overline{E^2} = \frac{Z''(\beta)}{Z(\beta)}, \qquad (\Delta E)^2 = \overline{E^2} - \bar{E}^2.$$

This yields the claim (7.148) after an elementary computation. □

The Wick trick (source trick). Alternatively, define the modified partition function

$$\mathcal{Z}(\beta, J) := \sum_{n=0}^{\infty} \mathrm{e}^{-E_n(\beta - J)} = \frac{1}{2\sinh \frac{(\beta - J)\omega\hbar}{2}}$$

where J is an additional small real parameter. Then $\mathcal{Z}(\beta, 0) = Z(\beta)$, and

$$\bar{E} = \frac{\mathcal{Z}_J(\beta, 0)}{\mathcal{Z}(\beta, 0)}, \qquad \overline{E^2} = \frac{\mathcal{Z}_{JJ}(\beta, 0)}{\mathcal{Z}(\beta, 0)}.$$

Tricks of this kind frequently appear while computing path integrals in quantum field theory; those tricks are also closely related to the Wick theorem in quantum field theory published in 1950. Therefore, we will briefly speak of the Wick trick. Behind this trick, there is the following general strategy in physics which was introduced by Schwinger: Add some source term to the physical system, and study

the change of the physical system under a change of the source J (see Chap. 14 of Vol. I).

Von Neumann's density operator. Let $H : D(H) \to L_2(\mathbb{R})$ be the self-adjoint Hamiltonian operator of the quantum harmonic oscillator on the real line,

$$H = \frac{P^2}{2m} + \frac{m\omega^2 Q^2}{2}.$$

Let $\varphi_0, \varphi_1, ...$ be the eigensolutions of H with

$$H\varphi_n = E_n\varphi_n, \qquad n = 0, 1, 2, ...$$

For any state $\varphi \in L_2(\mathbb{R})$ and any temperature $T > 0$, define

$$\mathrm{e}^{-\beta H}\varphi := \sum_{n=0}^{\infty} \mathrm{e}^{-\beta E_n}\langle\varphi_n|\varphi\rangle\varphi_n. \tag{7.149}$$

Note that

$$||\mathrm{e}^{-\beta H}\varphi||^2 = \sum_{n=0}^{\infty} |\mathrm{e}^{-\beta E_n}\langle\varphi_n|\varphi\rangle|^2 \leq \sum_{n=0}^{\infty} |\langle\varphi_n|\varphi\rangle|^2 = ||\varphi||^2.$$

Therefore, the operator $\mathrm{e}^{-\beta H} : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ is linear and continuous. For the trace, we get

$$\mathrm{tr}\,\mathrm{e}^{-\beta H} = \sum_{n=0}^{\infty}\langle\varphi_n|\mathrm{e}^{-\beta H}\varphi_n\rangle = \sum_{n=0}^{\infty}\mathrm{e}^{-\beta E_n}.$$

This is precisely the partition function Z. Therefore, the operator $\mathrm{e}^{-\beta H}$ is of trace class. In order to pass to the language of physicists, denote the vector φ_n by $|E_n\rangle$. Mnemonically, we write

$$\mathrm{e}^{-\beta H} = \sum_{n=0}^{\infty}\mathrm{e}^{-\beta E_n}|E_n\rangle\langle E_n|.$$

In fact, this implies $\mathrm{e}^{-\beta H}|\varphi\rangle = \sum_{n=0}^{\infty}\mathrm{e}^{-\beta E_n}|E_n\rangle\langle E_n|\varphi\rangle$ which coincides with (7.149). If $\chi_0, \chi_1, ...$ is an arbitrary complete orthonormal system in the Hilbert space $L_2(\mathbb{R})$, then

$$\mathrm{tr}\,\mathrm{e}^{-\beta H} = \sum_{n=0}^{\infty}\langle\chi_n|\mathrm{e}^{-\beta H}\chi_n\rangle = \sum_{n=0}^{\infty}\mathrm{e}^{-\beta E_n}\langle\chi_n|E_n\rangle\langle E_n|\chi_n\rangle.$$

The relation between the propagator $P(t,0) := \mathrm{e}^{-\mathrm{i}Ht/\hbar}$ and the operator $\mathrm{e}^{-\beta H}$ is given by

$$\mathrm{e}^{-\beta H} = P\left(-\mathrm{i}\beta\hbar, 0\right).$$

Now to the point. The linear bounded operator $\varrho : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ defined by

$$\boxed{\varrho := \frac{\mathrm{e}^{-\beta H}}{\mathrm{tr}\,\mathrm{e}^{-\beta H}}}$$

is called the density operator for our many-particle system of quantum harmonic oscillators on the real line. Explicitly,

$$\varrho = \sum_{n=0}^{\infty} p_n |E_n\rangle\langle E_n|$$

where $p_n = \mathrm{e}^{-\beta E_n}/Z(\beta)$. The real numbers

$$\varrho_{ij} := \langle \chi_i | \varrho \chi_j \rangle, \qquad i,j = 0,1,2,...$$

are called the entries of the density matrix with respect to the complete orthonormal system $\chi_0, \chi_1, ...$ For the mean energy value $\bar{E}$ and the mean energy fluctuation $\Delta\bar{E}$ of a particle, we get

$$\boxed{\bar{E} = \mathrm{tr}(\varrho H), \qquad (\Delta E)^2 = \mathrm{tr}(\varrho(H-\bar{E})^2).}$$

In fact, since $\varrho|E_n\rangle = p_n|E_n\rangle$ for all n,

$$\mathrm{tr}(\varrho H) = \sum_{n=0}^{\infty} \langle E_n|\varrho H|E_n\rangle = \sum_{n=0}^{\infty} p_n E_n \langle E_n|E_n\rangle = \sum_{n=0}^{\infty} p_n E_n = \bar{E}.$$

A similar argument applies to $\Delta\bar{E}$. Using the language of physicists, define[87]

$$\varrho(x,T) := \frac{\langle x|\mathrm{e}^{-\beta H}|x\rangle}{\mathrm{tr}\,\mathrm{e}^{-\beta H}}.$$

Since $\varphi_n(x) = \langle x|E_n\rangle$,

$$\varrho(x,T) = \sum_{n=0} p_n \langle x|E_n\rangle\langle E_n|x\rangle = \sum_{n=0}^{\infty} p_n |\varphi_n(x)|^2.$$

Recall that the function $x \mapsto |\varphi_n(x)|^2$ is the particle density of the nth energy state of the harmonic oscillator. Moreover,

$$\int_{\mathbb{R}} \varrho(x,T)dx = \sum_{n=0}^{\infty} p_n = 1.$$

Therefore, it is reasonable to regard $\varrho(x,T)$ as the (normalized) particle density of the gas at the point x at the temperature T.

Semiclassical quantum statistics and the Dirac calculus (formal approach). We want to explain how the Dirac calculus allows us to formally pass from the density operator ϱ to the semiclassical Gibbs statistics for high temperatures. Let $A : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ be a linear continuous operator of trace class. For a complete orthonormal system $\chi_0, \chi_1, ...$ of the complex Hilbert space $L_2(\mathbb{R})$, we get

$$\boxed{\mathrm{tr}\, A = \sum_{n=0}^{\infty} \langle \chi_n | A\chi_n\rangle.} \tag{7.150}$$

The point is that this number is finite, and it does not depend on the choice of the complete orthonormal system $\chi_0, \chi_1, ...$ The trick of the Dirac calculus is to formally extend the trace formula (7.150) to complete orthonormal systems of generalized eigenfunctions. For example, using the system $\{\langle x|\}_{x\in\mathbb{R}}$, we get

[87] See the formal Dirac calculus on page 596 of Vol. I.

$$\boxed{\operatorname{tr} A = \int_{\mathbb{R}} \langle x|A|x\rangle dx.} \tag{7.151}$$

Applying this formal approach, we are going to show that for high temperatures T, we obtain the following approximative formulas.[88]

(i) Mean value of energy:

$$\boxed{\bar{E} = \int_{\mathbb{R}^2} H(x,p)\varrho(x,p;T)\frac{dxdp}{h}.}$$

(ii) Mean energy fluctuation: $(\Delta\bar{E})^2 = \int_{\mathbb{R}^2}(H(x,p)-\bar{E})^2\varrho(x,p;T)\frac{dxdp}{h}$.

Here, $H(x,p) := \frac{p^2}{2m} + \frac{m\omega^2x^2}{2}$. For the density function in the phase space,

$$\boxed{\varrho(x,p;T) := \frac{\mathrm{e}^{-\beta H(x,p)}}{\int_{\mathbb{R}^2}\mathrm{e}^{-\beta H(x,p)}\frac{dxdp}{h}}.} \tag{7.152}$$

For a given function $A = A(x,p)$, the mean value $\bar{A}$ is defined by

$$\bar{A} = \int_{\mathbb{R}^2} A(x,p)\varrho(x,p;T)\frac{dxdp}{h}.$$

If the function $A = A(x)$ only depends on the position variable x, then

$$\bar{A} = \int_{\mathbb{R}} A(x)\varrho(x,T)dx$$

where we define

$$\varrho(x,T) := \int_{\mathbb{R}} \varrho(x,p;T)\frac{dp}{h}.$$

Let us prove (i) and (ii) above in a formal way. To begin with, observe that $\langle x|P^2|p\rangle = p^2\langle x|p\rangle$ and

$$\langle x|Q^2|p\rangle = \langle Q^2x|p\rangle = x^2\langle x|p\rangle.$$

Hence

$$\langle x|H|p\rangle = \langle x|\frac{P^2}{2m} + \frac{m\omega^2Q^2}{2}|p\rangle = H(x,p)\langle x|p\rangle.$$

Up to terms of order $O(\beta^2)$ as $\beta \to 0$, we get

$$\langle x|\mathrm{e}^{-\beta H}|p\rangle = \langle x|\,(I-\beta H)\,|p\rangle.$$

Hence

$$\langle x|\mathrm{e}^{-\beta H}|p\rangle = (1-\beta H(x,p))\,\langle x|p\rangle = \mathrm{e}^{-\beta H(x,p)}\langle x|p\rangle.$$

For the trace, we obtain $\operatorname{tr}\mathrm{e}^{-\beta H} = \int_{\mathbb{R}}\langle x|\mathrm{e}^{-\beta H}|x\rangle dx$. Using Dirac's completeness relation $\int_{\mathbb{R}}|p\rangle\langle p|\,dp = I$, we obtain

$$\operatorname{tr}\mathrm{e}^{-\beta H} = \int_{\mathbb{R}^2}\langle x|\mathrm{e}^{-\beta H}|p\rangle\langle p|x\rangle dxdp.$$

[88] Note that $dxdp/h$ and $\varrho(x,p;T)$ are dimensionless.

Since $\langle x|p\rangle = \mathrm{e}^{\mathrm{i}px}/\sqrt{h}$ and $\langle p|x\rangle = \langle x|p\rangle^\dagger$, we get

$$\operatorname{tr}\mathrm{e}^{-\beta H} = \int_{\mathbb{R}^2} \mathrm{e}^{-\beta H(x,p)}\,\frac{dxdp}{h}.$$

Summarizing, from $\varrho = \mathrm{e}^{-\beta H}/\operatorname{tr}\mathrm{e}^{-\beta H}$ it follows that

$$\langle x|\varrho|p\rangle = \varrho(x,p;T)\langle x|p\rangle$$

where $\varrho(x,p;T)$ is defined by (7.152). Finally,

$$\bar{E} = \operatorname{tr}(\varrho H) = \int_{\mathbb{R}} \langle x|\varrho H|x\rangle dx = \int_{\mathbb{R}^2} \langle x|\varrho|p\rangle\langle p|H|x\rangle dxdp.$$

Hence $\bar{E} = \int_{\mathbb{R}^2} \varrho(x,p;T)H(x,p)\frac{dxdp}{h}$. Similarly, we argue for $\Delta\bar{E}$.

Rigorous justification. To begin with, observe that the formal Dirac calculus tells us that

$$\langle x|\mathrm{e}^{-\beta H}|x\rangle = \sum_{n=0}^{\infty} \mathrm{e}^{-\beta E_n}\langle x|E_n\rangle\langle E_n|x\rangle = \sum_{n=0}^{\infty} \mathrm{e}^{-\beta E_n}|\langle x|E_n\rangle|^2,$$

and $\operatorname{tr}\mathrm{e}^{-\beta/kT} = \int_{\mathbb{R}}\langle x|\mathrm{e}^{-\beta H}|x\rangle dx$. In order to obtain a rigorous formulation, let us write this as

$$\boxed{\operatorname{tr}\mathrm{e}^{-\beta H} = \lim_{m\to\infty}\int_{\mathbb{R}}\sum_{n=0}^{m}\mathrm{e}^{-\beta E_n}|\varphi_n(x)|^2 dx.} \tag{7.153}$$

Proposition 7.42 *The trace formula* (7.153) *holds. Explicitly, the trace* $\operatorname{tr}\, e^{-\beta H}$ *is the partition function of the quantum harmonic oscillator.*

Proof. The trace class operator $\mathrm{e}^{-\beta H}$ has the complete orthonormal system of eigenvectors $\varphi_0, \varphi_1, \ldots$ with $\mathrm{e}^{-\beta H}\varphi_n = \mathrm{e}^{-\beta E_n}\varphi_n$ for all n. The trace is the sum of the eigenvalues. Hence

$$\operatorname{tr}\mathrm{e}^{-\beta H} = \sum_{n=0}^{\infty}\mathrm{e}^{-\beta E_n}.$$

On the other hand, it follows from $||\varphi_n||^2 = \int_{\mathbb{R}}|\varphi_n(x)|^2 dx = 1$ that

$$\lim_{m\to\infty}\int_{\mathbb{R}}\sum_{n=0}^{m}\mathrm{e}^{-\beta E_n}|\varphi_n(x)|^2 dx = \lim_{m\to\infty}\sum_{n=0}^{m}\mathrm{e}^{-\beta E_n}.$$

□

Introduce the kernel to the operator $\mathrm{e}^{-\beta H}$ by setting

$$\mathcal{P}(x,y;T) = \langle x|\mathrm{e}^{-\beta H}|y\rangle = \sum_{n=0}^{\infty}\mathrm{e}^{-\beta E_n}\langle x|E_n\rangle\langle E_n|y\rangle,$$

in the language of the Dirac calculus. This means that we define

$$\boxed{\mathcal{P}(x,y;T) := \sum_{n=0}^{\infty}\mathrm{e}^{-\beta E_n}\varphi_n(x)\varphi_n(y)^\dagger.}$$

Recall that $\beta := 1/kT$ and $x_0 := \sqrt{\hbar/m\omega}$. The Mehler formula (7.143) tells us the following.

Proposition 7.43 *For all positions $x, y \in \mathbb{R}$ and all temperatures $T > 0$, we get*

$$\mathcal{P}(x, y; T) = \frac{1}{x_0\sqrt{2\pi \sinh \beta\hbar\omega}} \exp\left\{-\frac{(x^2 + y^2)\cosh \beta\hbar\omega - 2xy}{2x_0^2 \sinh \beta\hbar\omega}\right\}.$$

Moreover, for the partition function of the quantum harmonic oscillator, we have the trace formula

$$\int_{\mathbb{R}} \mathcal{P}(x, x; T)dx = \sum_{n=0}^{\infty} e^{-E_n/kT} = \operatorname{tr} e^{-H/kT}.$$

For the density function $\varrho(x, T) := \langle x|\varrho|x\rangle$, this implies

$$\varrho(x, T) = \frac{\mathcal{P}(x, x; T)}{Z(\beta)} = \frac{1}{x_0}\sqrt{\frac{\tanh \frac{\beta\omega\hbar}{2}}{\pi}} \exp\left\{\frac{x^2(1 - \cosh \beta\hbar\omega)}{x_0^2 \sinh \beta\hbar\omega}\right\}.$$

Von Neumann's equation of motion for general density operators. We are given real numbers $p_0, p_1, ...$ with $0 \le p_n \le 1$ and $p_0 + p_1 + ... = 1$. Choose a complete orthonormal system $\langle 0|, \langle 1|, ...$ in the Hilbert space $L_2(\mathbb{R})$. Define

$$\varrho_0 := \sum_{n=0}^{\infty} p_n |n\rangle\langle n|.$$

Moreover, for each time $t \in \mathbb{R}$, we define

$$\boxed{\varrho(t) := \mathrm{e}^{\mathrm{i}Ht/\hbar}\varrho_0\mathrm{e}^{-\mathrm{i}Ht/\hbar}.}$$

This is the equation of motion for an arbitrary density operator in the Hilbert space $L_2(\mathbb{R})$. This equation corresponds to the time-dependence of observables in the Heisenberg picture. For an observable $A : D(A) \to L_2(\mathbb{R})$, we define the mean value

$$\bar{A}(t) := \operatorname{tr}(\varrho(t)A), \qquad t \in \mathbb{R}$$

if this trace exists. In the special case where $\varrho_0 = \sum_{n=0}^{\infty} p_n |E_n\rangle\langle E_n|$, we obtain $\varrho(t) = \varrho_0$ for all times $t \in \mathbb{R}$.

7.7 The Feynman Path Integral

> The history of mathematics shows that every well-working formal calculus used in physics can be rigorously justified once a day, by finding the appropriate rigorous tools.
>
> Folklore

7.7.1 The Basic Strategy

In Chap. 7 of Vol. I, we studied discrete path integrals in a rigorous setting for N degrees of freedom. In this section, we will study the limit $N \to \infty$. Our plan is the following one:

(i) We start with the definition of the Feynman path integral (7.156) below as a limit in position space, where $N \to \infty$.
(ii) We rigorously show that this limit exists (in a generalized sense) in the two special cases of
- the free quantum particle on the real line (Sect. 7.7.3) and
- the harmonic oscillator (Sect. 7.7.4).

It turns out that these limits coincide with the propagator kernel introduced in Sects. 7.6.4 and 7.6.7 by using the rigorous method of Fourier analysis combined with analytic continuation.
(iii) This brings us to the formulation of the *propagator hypothesis* saying that the Feynman path integral always represents the Feynman propagator kernel of the Schrödinger equation. We motivate this propagator hypothesis by using the Dirac calculus in a formal sense (Sect. 7.7.6).
(iv) In Sect. 7.8, we will rigorously study finite-dimensional Gaussian integrals with N degrees of freedom. Motivated by this, in Sect. 7.9 we will give the definition of normalized infinite-dimensional Gaussian integrals by using the spectral theory of quadratic forms and the determinant of infinite-dimensional operators based on the analytic continuation of the corresponding zeta function.
(v) For the free quantum particle and the harmonic oscillator, we rigorously show that the normalized infinite-dimensional Gaussian integral represents the Feynman propagator kernel, up to a normalization factor (Sects. 7.9.3 and 7.9.4).
(vi) This brings us to the *spectral hypothesis* saying that the Feynman path integral can be computed by means of infinite-dimensional Gaussian integrals, up to a normalization factor. This is the basic method successfully used by physicists in quantum field theory. Fortunately enough, the normalization factor does not play any role, as a rule, since it drops out by considering quotients of path integrals.

The following remarks are in order:

- The concrete calculations performed by physicists show that the propagator hypothesis above is right in quantum mechanics.[89] Many concrete examples can be found in the following two standard references:

  C. Grosche and F. Steiner, Handbook of Feynman Path Integrals, Springer, Berlin, 1998 (950 references).

  H. Kleinert, Path Integrals in Quantum Mechanics, Statistics, and Polymer Physics, World Scientific, River Edge, New York, 2004.
- In Sect. 7.11, we will study the rigorous Wiener path integral for Brownian motion together with Cameron's no-go theorem for the Feynman path integral.
- In Sect. 7.12, we will investigate the relation between the Weyl calculus and the Feynman path integral (method of pseudo-differential operators).

Detailed hints to both the mathematical and physical literature concerning the Feynman path integral can be found in Sect. 7.22 on page 667.

The creation of a comprehensive rigorous mathematical theory for Feynman path integrals (also called functional integrals) in quantum field theory is a challenge for the mathematics of the future.

[89] However, observe the following peculiarity: If caustics appear in classical mechanics, then one has to handle carefully the Maslov indices in quantum mechanics, as in the case of the harmonic oscillator in Sect. 7.6.7.

7.7.2 The Basic Definition

Let us consider the Schrödinger equation

$$\mathrm{i}\hbar\psi_t(x,t) = -\frac{\hbar^2}{2m}\psi_{xx}(x,t) + U(x)\psi(x,t), \qquad x \in \mathbb{R},\ t > s \tag{7.154}$$

together with the corresponding classical action

$$S[q] := \int_s^t \left\{ \tfrac{1}{2} m\dot{q}(\tau)^2 - U\big(q(\tau)\big) \right\} d\tau. \tag{7.155}$$

We assume that the potential $U : \mathbb{R} \to \mathbb{R}$ is smooth. Choose $N = 1, 2, \ldots$, and divide the time interval $[s,t]$ into the subintervals $t_0 = s < t_1 < \ldots < t_{N-1} < t_N = t$, where

$$t_j := s + j\Delta t, \quad j = 0, 1, \ldots, N, \quad \Delta t := \frac{t-s}{N}.$$

Fix the positions $x, y \in \mathbb{R}$ on the real line. Let the symbol $C\{s,t\}$ denote the set of all continuous functions $q : [s,t] \to \mathbb{R}$ with the boundary condition

$$q(s) := x, \qquad q(t) := y.$$

For each path $q \in C\{s,t\}$, we set $q_j := q(t_j)$, where $j = 0, 1, \ldots, N$. By definition, the discrete action of this path reads as

$$\boxed{S_N := \sum_{n=0}^{N-1} \left\{ \frac{m}{2} \left(\frac{q_{n+1} - q_n}{\Delta t} \right)^2 - U(q_n) \right\} \Delta t.}$$

Finally, let us introduce the characteristic length $l := \sqrt{\frac{2\pi\hbar \mathrm{i}\Delta t}{m}}$. Here, the square root is to be understood in the sense of the principal value.

Basic definition. Our definition of the Feynman path integral reads as

$$\boxed{\int_{C\{s,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}q := \lim_{N\to\infty} \frac{1}{l} \int_{\mathbb{R}^{N-1}} \mathrm{e}^{\mathrm{i}S_N/\hbar}\, \frac{dq_1}{l} \cdots \frac{dq_{N-1}}{l}.} \tag{7.156}$$

Since the boundary values $q_0 = y$ and $q_N = x$ are fixed, the integrals on the right-hand side of (7.156) are well-defined $(N-1)$-dimensional integrals of the real variables $q_1, \ldots, q_{N-1}$. We assume that the limit $N \to \infty$ exists.

Intuitive interpretation. We regard $\int_{C\{s,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar} \mathcal{D}q$ as an integral over the paths in the space $C\{s,t\}$. The definition (7.156) will be motivated in great detail in Sect. 7.7.6 on page 555. The path integral depends on x, t, y, s. We write

$$K(x,t;y,s) = \int_{C\{s,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}q.$$

In the following two sections, we will show that, for the free quantum particle and the harmonic oscillator on the real line, the function K is nothing else than the Feynman propagator kernel $\mathcal{K}$.

7.7.3 Application to the Free Quantum Particle

Let us consider the Schrödinger equation (7.154) above with vanishing potential, $U(x) \equiv 0$. The corresponding classical action reads as

$$S[q] := \int_s^t \tfrac{1}{2} m \dot{q}(\tau)^2 d\tau.$$

In Sect. 7.5.1, we have computed the corresponding Feynman propagator kernel

$$\mathcal{K}(x,t;y,s) = \sqrt{\frac{m}{2\pi\hbar \mathrm{i}(t-s)}}\ \mathrm{e}^{\mathrm{i}m(x-y)^2/2\hbar(t-s)} \tag{7.157}$$

for a freely moving quantum particle on the real line (see Theorem 7.16 on page 488).

Proposition 7.44 *For the free quantum particle, the Feynman path integral coincides with the Feynman propagator kernel, that is, we have*

$$\mathcal{K}(x,t;y,s) = \int_{C\{s,t\}} e^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}q \tag{7.158}$$

for all positions $x, y \in \mathbb{R}$ and all times $t > s$.

In the following proof, we will use a slight modification (7.160) of the original definition (7.156) of the path integral. In terms of physics, we separate the classical contribution from the quantum fluctuations. In terms of mathematics, we pass to homogeneous boundary conditions.

Proof. To simplify notation, set $\hbar := 1$ and $s := 0$.

(I) The classical trajectory. The action of a classical free particle of mass m on the real line is given by

$$S[q] := \int_0^t \tfrac{1}{2} m \dot{q}(\tau)^2\, d\tau.$$

Recall that the boundary-value problem

$$m\ddot{q}(\tau) = 0, \quad 0 < \tau < t, \quad q(0) = y,\ q(t) = x$$

corresponds to the motion of the particle with given endpoints. The unique solution is $q_{\mathrm{class}}(\tau) = y + \frac{\tau}{t}(x-y)$ with the classical action

$$S[q_{\mathrm{class}}] = \int_0^t \tfrac{1}{2} m \dot{q}_{\mathrm{class}}(\tau)^2 d\tau = \frac{m(x-y)^2}{2t}.$$

(II) Decomposition of trajectories. In order to study perturbations of the classical trajectory, we consider the trajectories

$$q(\tau) = q_{\mathrm{class}}(\tau) + r(\tau), \qquad \tau \in [0,t]$$

where $r \in C_0^2[0,t]$, that is, the function $r : [0,t] \to \mathbb{R}$ is twice continuously differentiable and satisfies the boundary condition $r(0) = r(t) = 0$. Then

$$S[q] = S[q_{\mathrm{class}}] + S[r]. \tag{7.159}$$

In fact, integration by parts yields

$$\int_0^t \dot{q}_{\text{class}}(\tau)\dot{r}(\tau)d\tau = -\int_0^t \ddot{q}_{\text{class}}(\tau)r(\tau)d\tau = 0,$$

since q_{class} satisfies the classical equation of motion, $m\ddot{q}_{\text{class}}(\tau) = 0$. Motivated by (7.159), let us slightly modify the definition (7.156) of the path integral by setting

$$\boxed{\int_{C\{0,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}q := \mathrm{e}^{\mathrm{i}S[q_{\text{class}}]/\hbar} \lim_{N\to\infty} \frac{1}{l} \int_{\mathbb{R}^{N-1}} \mathrm{e}^{\mathrm{i}S_N/\hbar}\, \frac{dr_1}{l} \cdots \frac{dr_{N-1}}{l}} \tag{7.160}$$

with the discrete action

$$S_N := \sum_{n=0}^{N-1} \left\{ \frac{m}{2} \left(\frac{r_{n+1} - r_n}{\Delta t} \right)^2 \right\} \Delta t$$

and the boundary values $r_0 = r_N := 0$.

(III) The generalized Gaussian integral. Let $a > 0$ or $a < 0$ and let $\beta \in \mathbb{R}$. According to (7.183) on page 561, we have the crucial Gaussian integral formula

$$\boxed{\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}\mathrm{i}ap^2}\, \mathrm{e}^{\mathrm{i}\beta p}\, \frac{dp}{\sqrt{2\pi}} := \frac{\mathrm{e}^{-\beta^2/2\mathrm{i}a}}{\sqrt{\mathrm{i}a}}.} \tag{7.161}$$

Here, the square root is to be understood as principal value. As we will discuss in Sect. 7.8, this definition has to be understood in the sense of analytic continuation.

(IV) Computation of the integrals from (7.160). Let us first integrate over the variable r_1. This yields the integral

$$\frac{1}{l^2} \int_{\mathbb{R}} \exp\left(\frac{\mathrm{i}m}{2\Delta t} \left((r_2 - r_1)^2 + (r_1 - r_0)^2 \right) \right) dr_1$$

which is equal to

$$\frac{1}{l^2} \exp\left(\frac{\mathrm{i}mr_2^2}{4\Delta t} \right) \int_{\mathbb{R}} \exp\left\{ \frac{\mathrm{i}m}{\Delta t} \left(r_1 - \frac{r_2}{2} \right)^2 \right\} dr_1 = \sqrt{\frac{m}{2\pi\mathrm{i}(2\Delta t)}} \exp\left(\frac{\mathrm{i}mr_2^2}{4\Delta t} \right).$$

Similarly, by induction, integrating over $r_1 \cdots r_n$ we get

$$\sqrt{\frac{m}{2\pi\mathrm{i}(n+1)\Delta t}}\, \exp\left(\frac{\mathrm{i}mr_{n+1}^2}{2(n+1)\Delta t} \right).$$

Choosing $n = N - 1$ and observing that $r_0 = r_N = 0$, we obtain

$$\sqrt{\frac{m}{2\pi\mathrm{i}N\Delta t}} = \sqrt{\frac{m}{2\pi\mathrm{i}t}}.$$

This expression does not depend on N. Thus, the limit $N \to \infty$ yields the same value. □

7.7.4 Application to the Harmonic Oscillator

> The path integral for the harmonic oscillator is closely related to the difference method for the classical harmonic oscillator in numerical analysis.
> Folklore

Consider the Schrödinger equation (7.154) above for the harmonic oscillator with mass m and angular frequency $\omega > 0$. This corresponds to the potential $U(x) = \frac{m\omega^2}{2}x^2$. The classical action is given by

$$S[q] = \int_s^t \tfrac{1}{2}m\dot{q}(\tau)^2 - \tfrac{1}{2}m\omega^2 q(\tau)^2 \, d\tau. \tag{7.162}$$

In Sect. 7.6.7 on page 537, we have computed the corresponding Feynman propagator kernel

$$\mathcal{K}(x,t;y,s) = \frac{1}{x_0\sqrt{2\pi \mathrm{i}\sin\omega(t-s)}} \exp\left(\mathrm{i}\frac{(x^2+y^2)\cos\omega(t-s) - 2xy}{2x_0^2\sin\omega(t-s)}\right)$$

for the harmonic oscillator. Here, we restrict ourselves to the first critical time interval $s < t < s + t_{1,\mathrm{crit}}$, where $t_{1,\mathrm{crit}} = \frac{\pi}{\omega}$. Furthermore, $x_0 = \sqrt{\frac{\hbar}{m\omega}}$.

Proposition 7.45 *For the quantized harmonic oscillator, the Feynman path integral coincides with the Feynman propagator kernel on the first critical time interval, that is, for all positions $x, y \in \mathbb{R}$, and all times $t \in]s, s + t_{1,\mathrm{crit}}[$, we have*

$$\mathcal{K}(x,t;y,s) = \int_{C\{s,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar} \, \mathcal{D}q. \tag{7.163}$$

This proposition is to be understood in a generalized sense which will be made precise in the following proof. First the Gaussian integrals have to be understood in a generalized sense by using analytic continuation. Secondly the limit $N \to \infty$ from (7.156) does not exist in the classical sense. Therefore, we will use a summation method.

Proof. (I) The classical trajectory. The boundary-value problem

$$\ddot{q}(\tau) + \omega^2 q(\tau) = 0, \quad s < \tau < t, \quad q(s) = y,\ q(t) = x \tag{7.164}$$

has the unique solution $q_{\mathrm{class}}(\tau) = y\cos\omega(\tau - s) + \big(x - y\cos\omega(\tau - s)\big)\,\frac{\sin\omega(\tau-s)}{\sin\omega(t-s)}$. This is a trajectory of the classical harmonic oscillator with the action

$$S[q_{\mathrm{class}}] = \hbar \cdot \frac{(x^2+y^2)\cos\omega(t-s) - 2xy}{2x_0^2\sin\omega(t-s))} \ .$$

(II) Decomposition of trajectories. Now we consider perturbations of the classical trajectory, by setting

$$q(\tau) := q_{\mathrm{class}}(\tau) + r(\tau)$$

where $r \in C_0^2[s,t]$. By definition, this notation means that the function $r : [s,t] \to \mathbb{R}$ is twice continuously differentiable and satisfies the following boundary condition $r(s) = r(t) = 0$. We have

$$S[q] = S[q_{\mathrm{class}}] + S[r] \qquad \text{for all} \quad r \in C_0^2[s,t]. \tag{7.165}$$

In fact, integration by parts yields

$$\int_0^t \dot{q}_{\text{class}}\dot{r} - \omega^2 q_{\text{class}} r \, d\tau = -\int_0^t (\ddot{q}_{\text{class}} + \omega^2 q_{\text{class}}) r \, d\tau = 0,$$

since q_{class} satisfies the classical equation of motion (7.164). Motivated by (7.165), let us slightly modify the definition (7.156) of the path integral by setting

$$\boxed{\int_{C\{s,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}q := \mathrm{e}^{\mathrm{i}S[q_{\text{class}}]/\hbar} \lim_{N\to\infty} \frac{1}{l} \int_{\mathbb{R}^{N-1}} \mathrm{e}^{\mathrm{i}S_N/\hbar}\, \frac{dr_1}{l} \cdots \frac{dr_{N-1}}{l}}$$

with the discrete action

$$S_N := \sum_{n=0}^{N-1} \left\{ \frac{m}{2} \left(\frac{r_{n+1} - r_n}{\Delta t} \right)^2 - \frac{m\omega^2}{2} r_n^2 \right\} \Delta t \tag{7.166}$$

and the boundary values $r_0 = r_N := 0$.

(III) The discrete action. To simplify notation, we set $s := 0$. The function S_N is a quadratic form. Explicitly,

$$\frac{\mathrm{i}S_N}{\hbar} = \frac{\mathrm{i}m\Delta t}{2\hbar} \cdot \langle r | A_N r \rangle. \tag{7.167}$$

Here, we set $\langle r|A_N r\rangle := r^d A_N r$ with the symmetric matrix

$$A_N := \frac{1}{(\Delta t)^2} \begin{pmatrix} a & -1 & 0 & \dots & 0 & 0 \\ -1 & a & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & a & -1 \\ 0 & 0 & 0 & \dots & -1 & a \end{pmatrix}$$

and $r^d := (r_1, \dots, r_{N-1})$. Furthermore, $a := 2 - (\omega\Delta t)^2$.

(IV) The discrete eigenvalue problem. The matrix equation $A_N r = \lambda r$ reads as

$$-\frac{r_{j+1} - 2r_j + r_{j-1}}{(\Delta t)^2} - \omega^2 r_j = \lambda_j r_j, \qquad r_0 = r_N = 0, \tag{7.168}$$

where $j = 1, \dots, N-1$. This equation has the eigensolutions

$$\lambda_n = \frac{n^2\pi^2}{t^2} \left(\frac{\sin\alpha(n)}{\alpha(n)} \right)^2 - \omega^2, \qquad n = 1, \dots, N-1,$$
$$r_n^d = \left(\sin\frac{n\pi\Delta t}{t}, \sin\frac{2n\pi\Delta t}{t}, \dots, \sin\frac{(N-1)n\pi\Delta t}{t} \right),$$

where $\alpha(n) := \frac{n\pi\Delta t}{2t}$. Using the limit $N \to +\infty$ (i.e., $\Delta t \to 0$), these eigensolutions go to the eigensolutions $\lambda_n = \frac{n^2\pi^2}{t^2} - \omega^2$ and $q(\tau) = \sin\frac{n\pi\tau}{t}$ of the boundary-eigenvalue

$$-\ddot{r}(\tau) - \omega^2 r(\tau) = \lambda r(\tau), \quad 0 < \tau < t, \quad r(0) = r(t) = 0$$

for the classical harmonic oscillator.

(IV) The Gaussian integral. By (7.167), we get

$$\int_{\mathbb{R}^{N-1}} \mathrm{e}^{\mathrm{i}S_N/\hbar}\,\frac{dr_1}{l}\cdots\frac{dr_{N-1}}{l} = (\sqrt{\gamma})^{N-1}\int_{\mathbb{R}^{N-1}} \mathrm{e}^{-\frac{1}{2}\gamma\langle r|A_N r\rangle}\,\frac{dr_1}{\sqrt{2\pi}}\cdots\frac{dr_{N-1}}{\sqrt{2\pi}},$$

where $\gamma := \frac{m\Delta t}{\hbar \mathrm{i}}$ and $l = \sqrt{\frac{2\pi \mathrm{i}\hbar\Delta t}{m}}$. By the key formula (7.190) for Gaussian integrals on page 564 (based on analytic continuation), we obtain

$$\frac{1}{l}\int_{\mathbb{R}^{N-1}} \mathrm{e}^{\mathrm{i}S_N/\hbar}\,\frac{dr_1}{l}\cdots\frac{dr_{N-1}}{l} = \frac{1}{l\sqrt{\det A_N}}.$$

(V) The problem of convergence. It remains to compute the limit

$$\lim_{N\to\infty} \frac{1}{l\sqrt{\det A_N}}.$$

Unfortunately, this limit does not exist in the classical sense. This follows immediately from

$$\det A_N = \lambda_1\lambda_2\cdots\lambda_{N-1} = \prod_{n=1}^{N-1}\left\{\frac{n^2\pi^2}{t^2}\left(\frac{\sin\alpha(n)}{\alpha(n)}\right)^2 - \omega^2\right\}$$

and $\frac{1}{l} = \sqrt{\frac{m}{2\pi\hbar \mathrm{i}\Delta t}} = \sqrt{N}\cdot\sqrt{\frac{m}{2\pi\hbar \mathrm{i}t}}$.

(VI) Summation method (generalized limit). We set $a(n\Delta t) := \Delta t\cdot\det A_n$ for the indices $n = 1, 2, \ldots, N$. Recall that $N\Delta t = t$. In addition let $a(0) := \Delta t$. By the definition of the matrix A_N, the Laplace expansion for determinants tells us that

$$\frac{a((n+1)\Delta t) - 2a(n\Delta t) + a((n-1)\Delta t)}{(\Delta t)^2} + \omega^2 a(n\Delta t) = 0$$

for all $n = 1, 2, \ldots, N-1$. Letting $\Delta t \to 0$, we obtain the ordinary differential equation

$$\ddot{a}(\tau) + \omega^2 a(\tau) = 0, \qquad 0 < \tau < t$$

with the initial condition $a(0) = 0$ and $\dot{a}(0) = 1$.[90] This initial-value problem has the unique solution

$$a(\tau) = \frac{\sin\omega\tau}{\omega}.$$

This motivates the definition

$$\lim_{N\to\infty} \Delta t\cdot\det A_N := \frac{\sin\omega t}{\omega},$$

as generalized limit. Therefore, we get

$$\lim_{N\to\infty}\frac{1}{l\sqrt{\det A_N}} = \sqrt{\frac{m}{2\pi\hbar \mathrm{i}}} \quad \lim_{N\to\infty}\frac{1}{\sqrt{\Delta t\cdot\det A_N}} = \sqrt{\frac{m\omega}{2\pi\hbar \mathrm{i}\sin\omega t}}.$$

□

The proof of Prop. 7.45 shows that the computation of the Feynman path integral for the harmonic oscillator is closely related to the eigenvalues of the corresponding classical boundary-value problem. Indeed, this is a crucial method for computing Feynman path integrals. We will study this in Sect. 7.9 on page 570 by means of the zeta-function regularization for infinite-dimensional Gaussian integrals.

[90] In fact, $a(0) = \lim_{\Delta t\to 0}\Delta t = 0$ and

$$\dot{a}(0) = \lim_{\Delta t\to 0}\frac{a(\Delta t) - a(0)}{\Delta t} = \lim_{\Delta t\to 0} A_1 - 1 = 2 - 1 = 1.$$

7.7.5 The Propagator Hypothesis

Motivated by Props. 7.44 and 7.45, let us formulate the following so-called propagator hypothesis:

(H) *For the Schrödinger equation* (7.154) *on the real line, the Feynman propagator kernel is given by the Feynman path integral, that is,*

$$\mathcal{K}(x,t;y,s) = \int_{C\{s,t\}} e^{\mathrm{i}mS[q]/\hbar}\, \mathcal{D}q.$$

Here, $S[q]$ is the classical action given by (7.155) *on page* 549.

7.7.6 Motivation of Feynman's Path Integral

It is our goal to motivate the propagator hypothesis (H) by using the formal Dirac calculus.

The classical Liouville measure in phase space. Consider a gas on the real line at high temperature $T > 0$. Let x and p denote the position and the momentum of a gas particle, respectively. In semi-classical statistical physics, the mean value $\bar{A}$ of a physical quantity $A = A(x,p)$ is given by

$$\bar{A} = \int_{\mathbb{R}} A(x,p)\varrho(x,p;T)d\mu(x,p).$$

Here, $\varrho = \varrho(x,p;T)$ denotes the density function from (7.152) on page 545. Furthermore, $d\mu := dxdp/h$ denotes the Liouville measure. This means that the Liouville measure $\mu(\Omega)$ of a compact subset Ω of the (x,p)-phase space is given by

$$\mu(\Omega) = \int_{\Omega} d\mu = \int_{\Omega} \frac{dxdp}{h}.$$

If the function $A = A(x)$ does not depend on the momentum, then we separately integrate with respect to the variable p. This yields

$$\bar{A} = \int_{\mathbb{R}} A(x)\varrho(x;T)dx$$

with $\varrho(x,T) := \int_{\mathbb{R}} \varrho(x,p;T)\frac{dp}{h}$. Note that the appearance of the Planck constant h as denominator guarantees that the Liouville measure $dxdp/h$ is dimensionless. We now want to show that the formal Feynman path integral is nothing else than an integral with respect to a formal infinite-dimensional Liouville measure. Alternatively, the Feynman path integral can also be viewed as a modified infinite-dimensional Gaussian integral.

The path integral in the phase space. Consider the motion $x = q(t)$ of a classical particle with mass m on the real line. For the momentum, $p(t) = m\dot{q}(t)$. The Hamiltonian function reads as

$$\mathcal{H}(q,p) := \frac{p^2}{2m} + U(q), \qquad q,p \in \mathbb{R}$$

with the potential U. The action along a trajectory during the time interval $[s,t]$ is given by

$$\boxed{S[q,p] := \int_s^t \frac{p(\tau)^2}{2m} - U(q(\tau))\, d\tau.} \tag{7.169}$$

Here, $p(\tau)^2/2m$ is the kinetic energy and $U(q(\tau))$ is the potential energy of the classical particle at time τ. Equivalently,

$$S[q,p] := \int_s^t p(\tau)\dot{q}(\tau) - \mathcal{H}(q(\tau),p(\tau))\, d\tau.$$

We now pass to the Schrödinger equation for the corresponding quantum particle on the real line,

$$\mathrm{i}\hbar\psi_t(x,t) = H\psi(x,t), \tag{7.170}$$

with the Hamiltonian operator $H = \frac{P^2}{2m} + U(Q)$. Recall from (7.88) on page 491 that the initial-value problem for the Schrödinger equation (7.170) has the solution

$$\boxed{\psi(x,t) = \int_{\mathbb{R}} \mathcal{K}(x,t;y,s)\psi_0(y)dy, \qquad t > s, \quad x \in \mathbb{R}}$$

with $\psi(x,s) = \psi_0(x)$ for all positions $x \in \mathbb{R}$ at the initial time s. The kernel $\mathcal{K}$ has the physical dimension of $[\text{length}]^{-1}$. In the elegant formal language of the Dirac calculus,

$$\mathcal{K}(x,t;y,s) = \langle x|\mathrm{e}^{-\mathrm{i}H(t-s)/\hbar}|y\rangle.$$

Our goal is to motivate *Feynman's magic formula*

$$\boxed{\mathcal{K}(x,t;y,s) = \int_{\mathcal{C}\{s,t\}} \mathrm{e}^{\mathrm{i}S[q,p]/\hbar}\ \mathcal{D}q\mathcal{D}p,} \tag{7.171}$$

which tells us that the propagator kernel $\mathcal{K}$ can be represented by a path integral. Here, we integrate over the space $\mathcal{C}\{s,t\}$ of all continuous paths $q,p : [s,t] \to \mathbb{R}$ with

$$q(s) = y, \qquad q(t) = x.$$

That is, we fix the initial time s, the initial point y, the final time t, and the final point x. Note that both the initial value $p(s)$ and the final value $p(t)$ of the momentum variable are unconstrained. Moreover, we use the classical action

$$S[q,p] := \int_s^t p(\tau)\dot{q}(\tau) - \mathcal{H}(q(\tau),p(\tau))\, d\tau$$

along the path $q = q(\tau), p = p(\tau),\ s \le \tau \le t$. The symbol $\mathcal{D}[q,p]$ represents a formal infinite-dimensional Liouville measure on the space $\mathcal{C}[s,t]$ of curves in the phase space. Formally,

$$\mathcal{D}[q,p] := \frac{dp(s)}{h} \prod_{s<\tau\le t} \frac{dq(\tau)dp(\tau)}{h}.$$

Since $dq(\tau)dp(\tau)/h$ is dimensionless, both the measure $\mathcal{D}[q,p]$ and the kernel from (7.171) have the same physical dimension of $[\text{length}]^{-1}$, as expected. The motivation given below will show that the magic formula (7.171) stands for the following formal limit

$$\boxed{\mathcal{K}(x,t;y,s) = \lim_{N\to\infty} \int_{\mathbb{R}^{2N-1}} \mathrm{e}^{\mathrm{i}S_N/\hbar} \, \frac{dp_0}{h} \prod_{n=1}^{N-1} \frac{dq_n dp_n}{h}} \tag{7.172}$$

along with the discrete action

$$S_N := \sum_{n=0}^{N-1} \Big(p_n \, \frac{q_{n+1} - q_n}{\Delta t} - \mathcal{H}(q_n, p_n) \Big) \, \Delta t.$$

Here, $q_0 := y$ and $q_N := x$.

Applying the Dirac calculus. We want to motivate formula (7.172). To simplify notation, set $\hbar = 1$. Then $h = 2\pi$. Let us decompose the time interval $[s,t]$ into N pieces of equal length by setting

$$s = t_0 < t_1 < \ldots < t_{N-1} < t_N = t$$

along with $t_n := s + n\Delta t$ and $\Delta t := (t-s)/N$. Recall the propagator kernel formula: $\mathcal{K}(x,t;y,s) := \langle x|\mathrm{e}^{-\mathrm{i}NH\Delta t}|y\rangle$.

(i) Causality and the product property of the propagator kernel: For all intermediate times τ with $s \le \tau \le t$, we get the following product formula

$$\boxed{\mathcal{K}(x,t;y,s) = \int_{\mathbb{R}} \mathcal{K}(x,t;q,\tau)\mathcal{K}(q,\tau;y,s)dq.} \tag{7.173}$$

In fact, by the addition theorem for the exponential function,

$$\mathrm{e}^{-\mathrm{i}H(t-s)} = \mathrm{e}^{-\mathrm{i}H(t-\tau)}\mathrm{e}^{-\mathrm{i}H(\tau-s)}.$$

Using the completeness relation $\int_{\mathbb{R}} |q\rangle\langle q| \, dq = I$, we obtain

$$\langle x|\mathrm{e}^{-\mathrm{i}H(t-s)}|y\rangle = \int_{\mathbb{R}} \langle x|\mathrm{e}^{-\mathrm{i}H(t-\tau)}|q\rangle\langle q|\mathrm{e}^{-\mathrm{i}H(\tau-s)}|y\rangle dq$$

which proves (7.173). From the physical point of view, the product formula (7.173) reflects nothing other than causality (see page 482).

(ii) The propagator kernel for small time intervals: We want to show that the propagator kernel $\mathcal{K}$ for the small time interval $[s, s+\Delta t]$ can be approximately represented by the following simple key formula

$$\boxed{\mathcal{K}(x, s+\Delta t; y, s) = \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}\mathcal{H}(y,p)\Delta t} \, \mathrm{e}^{\mathrm{i}p(x-y)} \, \frac{dp}{2\pi},} \tag{7.174}$$

up to terms of order $(\Delta t)^2$. To get this, note that the following is true, up to terms of order $(\Delta t)^2$. By Taylor expansion,

$$\mathrm{e}^{-\mathrm{i}H\Delta t} = I - \mathrm{i}H\Delta t.$$

Recall that $H = \frac{P^2}{2m} + \frac{m\omega^2 Q^2}{2}$. For the position operator, $Q|y\rangle = y|y\rangle$. Hence, by the completeness relation $\int_{\mathbb{R}} |p\rangle\langle p| \, dp = I$, we get

$$\langle x|Q^2|y\rangle = y^2\langle x|y\rangle = y^2 \int_{\mathbb{R}} \langle x|p\rangle\langle p|y\rangle dp.$$

Since $\langle x|p\rangle = \frac{\mathrm{e}^{\mathrm{i}xp}}{\sqrt{2\pi}}$, we obtain

$$\langle x|Q^2|y\rangle = \int_{\mathbb{R}} y^2 \mathrm{e}^{\mathrm{i}xp}\mathrm{e}^{-\mathrm{i}yp}\frac{dp}{2\pi}.$$

For the momentum operator, $P|p\rangle = p|p\rangle$. Therefore,

$$\langle x|P^2|y\rangle = \int \langle x|P^2 p\rangle\langle p|y\rangle dp = \int_{\mathbb{R}} p^2\langle x|p\rangle\langle p|y\rangle\frac{dp}{2\pi}.$$

Hence $\langle x|P^2|y\rangle = \int_{\mathbb{R}} p^2 \mathrm{e}^{\mathrm{i}p(x-y)}\ \frac{dp}{2\pi}$. Summarizing,

$$\langle x|H|y\rangle = \int_{\mathbb{R}}\left(\frac{p^2}{2m}+\frac{m\omega^2 y^2}{2}\right)\mathrm{e}^{\mathrm{i}p(x-y)}\ \frac{dp}{2\pi} = \int_{\mathbb{R}}\mathcal{H}(y,p)\mathrm{e}^{\mathrm{i}p(x-y)}\ \frac{dp}{2\pi}.$$

Thus

$$\langle x|I-\mathrm{i}H\Delta t|y\rangle = \int_{\mathbb{R}}(1-\mathrm{i}\mathcal{H}(y,p)\Delta t)\ \mathrm{e}^{\mathrm{i}p(x-y)}\ \frac{dp}{2\pi}.$$

Finally,

$$\langle x|\mathrm{e}^{-\mathrm{i}H\Delta t}|y\rangle = \int_{\mathbb{R}}\mathrm{e}^{-\mathrm{i}\mathcal{H}(y,p)\Delta t}\ \mathrm{e}^{\mathrm{i}p(x-y)}\ \frac{dp}{2\pi},$$

up to terms of order $(\Delta t)^2$. This finishes the formal proof of the claim.

(iii) The path integral: Consider first the case where $N = 2$. Set $q_0 := y$ and $q_2 := x$. By the product formula (7.173),

$$\mathcal{K}(q_2,t_2;q_0,t_0) = \int_{\mathbb{R}}\mathcal{K}(q_2,t_2;q_1,t_1)\mathcal{K}(q_1,t_1;q_0,t_0)dq_1.$$

Using (7.174), we get the approximative formula

$$\mathcal{K}(q_2,t_2;q_0,t_0) = \int_{\mathbb{R}^3}\mathrm{e}^{\mathrm{i}S_2}\ dq_1\cdot\frac{dp_1}{2\pi}\cdot\frac{dp_0}{2\pi} \tag{7.175}$$

with the discrete action

$$S_2 := \left(p_1\frac{q_2-q_1}{\Delta t}+p_0\frac{q_1-q_0}{\Delta t}-\mathcal{H}(q_1,p_1)-\mathcal{H}(q_0,p_0)\right)\Delta t.$$

Now let $N = 2,3,\ldots$. Similarly, the product formula (7.173) yields

$$\mathcal{K}(q_N,t_N;q_0,t_0) = \int_{\mathbb{R}^{N-1}} dq_{N-1}\cdots dq_1\prod_{n=1}^{N}\mathcal{K}(q_n,t_n;q_{n-1},t_{n-1}).$$

By (7.174), we obtain the approximative formula

$$\mathcal{K}(q_N,t_N;q_0,t_0) = \int_{\mathbb{R}^{2N-1}}\mathrm{e}^{\mathrm{i}S_N}\ \frac{dp_0}{2\pi}\prod_{n=1}^{N-1}\frac{dq_n dp_n}{2\pi} \tag{7.176}$$

with the discrete action

$$S_N := \sum_{n=0}^{N-1}\left(p_n\frac{q_{n+1}-q_n}{\Delta t}-\mathcal{H}(q_n,p_n)\right)\Delta t.$$

Finally, carry out the formal limit $N\to\infty$ which implies $\Delta t\to 0$. This way, we get Feynman's magic limit formula (7.171) which we mnemonically write as (7.172).

The path integral in the position space. Feynman's magic formula (7.171) can be simplified by integrating over the momentum variables $p_0, p_1, ...$ This yields the following modified magic formula

$$\boxed{\mathcal{K}(x,t;y,s) = \int_{C\{s,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}q.} \tag{7.177}$$

Here, we integrate over all continuous paths $q = q(\tau), s \le \tau \le t$, in the position space with

$$q(s) = y, \qquad q(t) = x$$

for given initial time s, initial point y, final time t, and final point x. Along this path, we use the classical action

$$S[q] := \int_s^t \left\{ \tfrac{1}{2} m\dot{q}(\tau)^2 - U(q(\tau)) \right\} d\tau.$$

Mnemonically, the magic formula (7.177) stands for the following formal limit

$$\boxed{\mathcal{K}(x,t;y,s) = \lim_{N\to\infty} \frac{1}{l} \int_{\mathbb{R}^{N-1}} \mathrm{e}^{\mathrm{i}S_N/\hbar}\, \frac{dq_1}{l} \cdots \frac{dq_{N-1}}{l}} \tag{7.178}$$

along with the discrete action

$$S_N := \sum_{n=0}^{N-1} \left\{ \frac{m}{2} \left(\frac{q_{n+1} - q_n}{\Delta t} \right)^2 - U(q_n) \right\} \Delta t,$$

and the characteristic length $l := \sqrt{\frac{2\pi\hbar \mathrm{i} \Delta t}{m}}$. Now let us motivate formula (7.178). Set $\hbar := 1$. Proceeding as above, formula (7.178) follows from the product formula

$$\mathcal{K}(q_N, t_N; q_0, t_0) = \int_{\mathbb{R}^{N-1}} dq_{N-1} \cdots dq_1 \prod_{n=1}^{N} \mathcal{K}(q_n, t_n; q_{n-1}, t_{n-1})$$

along with the approximation

$$\boxed{\mathcal{K}(x, s+\Delta t; y, s) = l^{-1} \mathrm{e}^{\mathrm{i}m(x-y)^2/2\Delta t}\, \mathrm{e}^{-\mathrm{i}U(y)\Delta t},} \tag{7.179}$$

up to terms of order $(\Delta t)^2$. It remains to justify formula (7.179). To this end, use

$$\mathcal{K}(x, s+\Delta t; y, s) = \mathrm{e}^{-\mathrm{i}U(y)\Delta t} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}p^2 \Delta t/2m}\, \mathrm{e}^{\mathrm{i}p(x-y)}\, \frac{dp}{2\pi}, \tag{7.180}$$

by (7.174). Applying the Gaussian integral formula (7.161) on page 551 to (7.180), we get the desired formula (7.179).

7.8 Finite-Dimensional Gaussian Integrals

The rigorous theory of finite-dimensional Gaussian integrals in classical probability theory represents the prototype of the formal theory of infinite-dimensional Gaussian integrals, which play a crucial role for describing correlations in quantum field theory.

> In order to understand the formal properties of infinite-dimensional Gaussian integrals, write the well-defined N-dimensional Gaussian integrals in such a way that the formal limit $N \to \infty$ can be easily performed.
>
> Folklore

Path integrals (or more general functional integrals) can be computed by regarding them as infinite-dimensional Gaussian integrals. To discuss this, in the present section we are going to study finite-dimensional Gaussian integrals. In the next section, we will generalize this to the infinite-dimensional case. Note that

- free systems (i.e., systems without any interaction) correspond to standard Gaussian integrals, whereas
- interacting systems are described by the perturbation of standard Gaussian integrals.

In the framework of perturbation theory, the computation of perturbed Gaussian integrals can be reduced to the computation of moments for standard Gaussian integrals. Analytically, this can be based on the Wick theorem. Graphically, this corresponds to Feynman diagrams.

7.8.1 Basic Formulas

One-dimensional Gaussian integrals. The starting point is the integral

$$\boxed{\int_{\mathbb{R}} \mathrm{e}^{-x^2} dx = \sqrt{\pi}.} \tag{7.181}$$

This formula elegantly follows from $\int_{\mathbb{R}} \mathrm{e}^{-x^2} dx \int_{\mathbb{R}} \mathrm{e}^{-y^2} dy = \int_{\mathbb{R}^2} \mathrm{e}^{-x^2-y^2} dxdy$. Passing to polar coordinates, the latter integral is equal to

$$2\pi \int_0^\infty \mathrm{e}^{-r^2}\, r dr = -\pi \mathrm{e}^{-r^2}|_0^\infty = \pi.$$

(i) Rescaling: Using translation and rescaling, for all $a > 0$ and $x_0 \in \mathbb{R}$, we get

$$\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}a(x-x_0)^2} \frac{dx}{\sqrt{2\pi}} = \frac{1}{\sqrt{a}}.$$

(ii) Quadratic supplement: Using (i) and setting $b := ax_0$, we obtain

$$\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}ax^2} \mathrm{e}^{bx}\, \frac{dx}{\sqrt{2\pi}} = \frac{\mathrm{e}^{b^2/2a}}{\sqrt{a}} \tag{7.182}$$

for all $a > 0$ and $b \in \mathbb{R}$. The reduction process from (7.182) to (i) is called the method of the quadratic supplement.

(iii) Analytic continuation: Introduce the set

$$\Omega := \{(a,b) \in \mathbb{C}^2 : a = r\mathrm{e}^{\mathrm{i}\varphi},\ \ r > 0,\ \ -\pi < \varphi < \pi\}.$$

For all $(a,b) \in \Omega$, the function

$$F(a,b) := \frac{\mathrm{e}^{b^2/2a}}{\sqrt{a}}$$

is well defined. Here, the square root is to be understood in the sense of the principal value, that is, $\sqrt{a} := \sqrt{r}\, \mathrm{e}^{\mathrm{i}\varphi/2}$. In fact, the function $F : \Omega \to \mathbb{C}$ is holomorphic. We define

$$\boxed{\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}ax^2} \mathrm{e}^{bx}\, \frac{dx}{\sqrt{2\pi}} := F(a,b) \qquad \text{for all} \quad (a,b) \in \Omega.} \tag{7.183}$$

This definition is based on the idea of analytic continuation. For example,

$$\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}x^2}\, \frac{dx}{\sqrt{2\pi}} := \frac{1}{\sqrt{-2\mathrm{i}}} = \frac{\mathrm{e}^{\mathrm{i}\pi/4}}{\sqrt{2}} = \frac{1+\mathrm{i}}{2}. \tag{7.184}$$

This Fresnel integral exists in the classical sense.[91] In the special case where $\Re(a) > 0$ and $b \in \mathbb{C}$, relation (7.183) is always valid in the classical sense (i.e., the integral exists).[92]

(iv) Fourier transform: Let $a > 0$. Then it follows from (iii) that

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-ax^2/2}\, \mathrm{e}^{-\mathrm{i}px}\, dx = \frac{\mathrm{e}^{-p^2/2a}}{\sqrt{a}} \qquad \text{for all} \quad p \in \mathbb{R}.$$

In the special case where $a = 1$, this relation shows that the Gaussian function $x \mapsto \mathrm{e}^{-x^2/2}$ is a fixed point of the Fourier transform.

(v) The method of stationary phase: Let us introduce the so-called phase function $\Phi(x) := -\frac{1}{2}ax^2 + bx$. The equation

$$\Phi'(x) = -ax + b = 0$$

has the unique solution $x_{\mathrm{crit}} := b/a$. For this point, the phase function Φ becomes stationary. Relation (7.183) can be written as

$$\boxed{\int_{\mathbb{R}} \mathrm{e}^{\Phi(x)}\, \frac{dx}{\sqrt{2\pi}} = \frac{\mathrm{e}^{\Phi(x_{\mathrm{crit}})}}{\sqrt{a}} \qquad \text{for all} \quad (a,b) \in \Omega.}$$

This so-called method of stationary phase tells us that the integral is determined by the integrand at the stationary point x_{crit}, up to a normalization constant.

(vi) Adiabatic regularization: Let $f : \mathbb{R} \to \mathbb{C}$ be a bounded function, that is, $\sup_{x\in\mathbb{R}} |f(x)| < \infty$, which is continuous (or continuous up to a set of Lebesgue measure zero). Then the integral

$$\int_{\mathbb{R}} f(x) \mathrm{e}^{-\frac{1}{2}\varepsilon x^2}\, dx, \qquad \varepsilon > 0$$

exists, which is called the adiabatic regularization of the integral $\int_{\mathbb{R}} f(x)dx$. For example, let $\alpha \in \mathbb{R}, b \in \mathbb{C}$, and let $\varepsilon > 0$. Then the integral

[91] For the computation of this integral by using Cauchy's residue method, we refer to page 734 of Vol. I.

[92] Here, $|\mathrm{e}^{-\frac{1}{2}ax^2} \mathrm{e}^{bx}| = \mathrm{e}^{-\frac{1}{2}\Re(a)x^2} \mathrm{e}^{\Re(b)x}$ for all $x \in \mathbb{R}$. Both the existence of the integral from (7.183) and its analytic dependence on the parameters a and b follow then from the majorant criterion for integrals (see Vol. I, p. 493).

$$\int_{\mathbb{R}} \left(\mathrm{e}^{-\frac{1}{2}\alpha \mathrm{i} x^2} \mathrm{e}^{bx} \right) \mathrm{e}^{-\frac{1}{2}\varepsilon x^2} \frac{dx}{\sqrt{2\pi}} = \frac{\mathrm{e}^{b^2/2(\varepsilon+\alpha \mathrm{i})}}{\sqrt{\varepsilon + \alpha \mathrm{i}}}$$

exists. If $\alpha \neq 0$, then we have the limit relation

$$\lim_{\varepsilon \to +0} \int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}\alpha \mathrm{i} x^2} \mathrm{e}^{bx} \mathrm{e}^{-\frac{1}{2}\varepsilon x^2} \frac{dx}{\sqrt{2\pi}} = \int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}\alpha \mathrm{i} x^2} \mathrm{e}^{bx} \frac{dx}{\sqrt{2\pi}} = \frac{\mathrm{e}^{b^2/2\alpha \mathrm{i}}}{\sqrt{\alpha \mathrm{i}}}.$$

(vii) Moments and the Wick trick: Let $a > 0$. We want to compute the moments

$$\langle x^k \rangle := \frac{\int_{\mathbb{R}} x^k \mathrm{e}^{-\frac{1}{2}ax^2} \frac{dx}{\sqrt{2\pi}}}{\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}ax^2} \frac{dx}{\sqrt{2\pi}}}, \qquad k = 0, 1, 2, \dots$$

To this end, for $J \in \mathbb{C}$, we introduce the so-called generating function

$$Z(J) := \frac{\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}ax^2} \mathrm{e}^{Jx} \frac{dx}{\sqrt{2\pi}}}{\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}ax^2} \frac{dx}{\sqrt{2\pi}}} = \mathrm{e}^{J^2/2a}.$$

Differentiation yields $Z'(0) = \langle x \rangle$. More generally,

$$\boxed{\langle x^k \rangle = \frac{d^k Z(0)}{dJ^k}, \qquad k = 0, 1, 2 \dots} \tag{7.185}$$

For example,

$$\langle x \rangle = Z'(0) = 0, \qquad \langle x^2 \rangle = Z''(0) = a^{-1}.$$

Note that if $\Re(a) > 0$, then the integrals $M_0, M_1, M_2, \dots$ exist, and the Wick trick formula (7.185) holds true, by the majorant criterion (see Vol. I, p. 493). The entire function $Z : \mathbb{C} \to \mathbb{C}$ has the power series expansion

$$Z(J) = M_0 + M_1 J + \frac{M_2 J^2}{2!} + \frac{M_3 J^3}{3!} + \dots$$

N-dimensional Gaussian integrals. In what follows, we choose the dimensions $N = 1, 2, \dots$ All the square roots are to be understood as principal values. Let $(\lambda_k, b_k) \in \Omega$ be given for $k = 1, \dots, N$. The prototype is the definition

$$\int_{\mathbb{R}^N} \prod_{k=1}^{N} \mathrm{e}^{-\frac{1}{2}\lambda_k x_k^2 + b_k x_k} \frac{dx_k}{\sqrt{2\pi}} := \prod_{k=1}^{N} \int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}\lambda_k x^2 + b_k x} \frac{dx}{\sqrt{2\pi}} = \prod_{k=1}^{N} \frac{\mathrm{e}^{b_k^2/2\lambda_k}}{\sqrt{\lambda_k}}. \tag{7.186}$$

The integrals are to be understood in the generalized sense. However, if $\Re(\lambda_k) > 0$ for $k = 1, \dots, N$, then the integrals exist, and relation (7.186) is to be understood in the classical sense. We make the following assumption.

(H) *All the eigenvalues of the real symmetric $(N \times N)$-matrix $A = (a_{kl})$ are positive, that is, $\lambda_1 > 0, \dots, \lambda_N > 0$.*

Then $\det A = \lambda_1 \lambda_2 \cdots \lambda_N$. By definition, the zeta function of the matrix A reads as

$$\zeta_A(s) := \sum_{k=1}^{N} \frac{1}{\lambda_k^s} \qquad \text{for all} \quad s \in \mathbb{C}.$$

For all $x, y \in \mathbb{R}^N$ and all $b \in \mathbb{C}^N$, we set

$$\langle y|Ax\rangle := \sum_{k,l=1}^{N} y_k a_{kl} x_l, \qquad \langle b|x\rangle := \sum_{k=1}^{N} b_k x_k.$$

Since $\lambda_k^{-s} = \mathrm{e}^{-s \ln \lambda_k}$, we obtain the derivative

$$\zeta_A'(s) = -\sum_{k=1}^{N} \frac{\ln \lambda_k}{\lambda_k^s}, \qquad s \in \mathbb{C}.$$

This implies the key formula

$$\boxed{\det A = \prod_{k=1}^{N} \lambda_k = \mathrm{e}^{-\zeta_A'(0)}.} \tag{7.187}$$

The following properties of finite-dimensional Gaussian integrals are crucial for the theory of infinite-dimensional Gaussian integrals.

(i) The standard Gaussian integral: For all $y \in \mathbb{R}^N$, we have

$$\int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\langle (x-y)|A(x-y)\rangle} \, \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}} = \frac{1}{\sqrt{\det A}}.$$

Proof. After a translation, we can choose $y = 0$. By the principal axis theorem on the real Hilbert space $\mathbb{R}^N$, there exists an orthogonal transformation which sends the integral to the normal form (7.186) with $b_k = 0$ for all k. □

(ii) Quadratic supplement: For all $b \in \mathbb{R}^N$, we have

$$\int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle} \, \mathrm{e}^{\langle b|x\rangle} \, \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}} = \frac{\mathrm{e}^{\frac{1}{2}\langle b|A^{-1}b\rangle}}{\sqrt{\det A}}. \tag{7.188}$$

This can be written as

$$\boxed{\int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle} \, \mathrm{e}^{\langle b|x\rangle} \, \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}} = \mathrm{e}^{\frac{1}{2}\langle b|A^{-1}b\rangle} \, \mathrm{e}^{\frac{1}{2}\zeta_A'(0)}.} \tag{7.189}$$

Proof. This is an easy consequence of (i) above. Since A is symmetric, we get

$$\begin{aligned} \langle (x-y)|A(x-y)\rangle &= \langle x|Ax\rangle - \langle y|Ax\rangle - \langle x|Ay\rangle + \langle y|Ay\rangle \\ &= \langle x|Ax\rangle - 2\langle Ay|x\rangle + \langle Ay|y\rangle. \end{aligned}$$

Finally, set $b := Ay$. By (i), the integral on the left-hand side of (7.188) is equal to

$$\frac{\mathrm{e}^{\frac{1}{2}\langle Ay|y\rangle}}{\sqrt{\det A}}.$$

Finally, observe that $y = A^{-1}b$. Hence $\langle Ay|y\rangle = \langle b|A^{-1}b\rangle$. □

(iii) Analytic continuation: Let γ be a nonzero complex number with the argument $-\pi < \arg(\gamma) < \pi$ (e.g., $\gamma = \pm \mathrm{i}$). Then, for all $b \in \mathbb{C}^N$, we define

$$\boxed{\int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\gamma\langle x|Ax\rangle}\, \mathrm{e}^{\langle b|x\rangle}\, \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}} := \frac{\mathrm{e}^{\frac{1}{2}\langle b|(\gamma A)^{-1}b\rangle}}{\sqrt{\det(\gamma A)}}.} \tag{7.190}$$

Here, the square root is to be understood as

$$\sqrt{\det(\gamma A)} := (\sqrt{\gamma})^N \sqrt{\det A}$$

where $\sqrt{\gamma}$ is the principal value of the square root. Note that equation is valid for all $\gamma > 0$, and the integral exists in the classical sense. Then we use analytic continuation.

(iv) Adiabatic regularization: Let A and b be given as in (iii) above. Furthermore, let $\alpha \in \mathbb{R}$ and $\varepsilon > 0$. Then the integral

$$\int_{\mathbb{R}^N} \left(\mathrm{e}^{-\frac{1}{2}\alpha\mathrm{i}\langle x|Ax\rangle}\, \mathrm{e}^{\langle b|x\rangle}\right) \mathrm{e}^{-\frac{1}{2}\varepsilon\langle x|x\rangle}\, \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}} := \frac{\mathrm{e}^{\frac{1}{2}\langle b|(\alpha\mathrm{i}A+\varepsilon I)^{-1}b\rangle}}{\sqrt{\det(\alpha\mathrm{i}A + \varepsilon I)}}$$

exists. If $\alpha \neq 0$, then we have the limit relation

$$\lim_{\varepsilon\to+0} \int_{\mathbb{R}^N} \left(\mathrm{e}^{-\frac{1}{2}\alpha\mathrm{i}\langle x|Ax\rangle}\, \mathrm{e}^{\langle b|x\rangle}\right) \mathrm{e}^{-\frac{1}{2}\varepsilon\langle x|x\rangle}\, \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}} = \frac{\mathrm{e}^{\frac{1}{2}\langle b|(\alpha\mathrm{i}A)^{-1}b\rangle}}{\sqrt{\det(\alpha\mathrm{i}A)}}.$$

(v) The method of stationary phase: Let A and b be given as in (iii) above. Introduce the phase function $\Phi(x) := -\frac{1}{2}\gamma\langle x|Ax\rangle + \langle b|x\rangle$. The equation

$$\Phi'(x) = -\gamma Ax + b = 0$$

has the unique solution $x_{\mathrm{crit}} := (\gamma A)^{-1}b$. Then relation (7.190) can be written as

$$\boxed{\int_{\mathbb{R}^n} \mathrm{e}^{\Phi(x)}\, \frac{dx^1}{\sqrt{2\pi}} \cdots \frac{dx^N}{\sqrt{2\pi}} = \frac{\mathrm{e}^{\Phi(x_{\mathrm{crit}})}}{\sqrt{\det(\gamma A)}}.}$$

7.8.2 Free Moments, the Wick Theorem, and Feynman Diagrams

In what follows, we will use a terminology which fits best the needs of quantum field theory. Our approach can be viewed as a discrete variant of quantum field theory. The basic notions are:

- free probability distribution (also called the Gaussian distribution in mathematics),
- free moments (free n-correlation functions or, briefly, called free n-point functions),
- generating function of the free moments,
- Feynman diagrams (i.e., graphic representation of free moments).

In the next section, we will generalize this to full probability distributions and full moments.

In terms of discrete quantum field theory, full moments (resp. free moments) describe particles under interaction (resp. free particles without any interaction).

Moments are fundamental quantities. The theory of moments in probability theory tells us that, roughly speaking, a probability distribution is uniquely determined by the knowledge of its moments (see Vol. I, page 751). Our main task is to reduce the computation of full moments to the computation of free moments. This is the basic trick of perturbation theory in quantum field theory.

The free probability distribution. Assume that the matrix A has the property (H) formulated on page 562. Introduce the key quantity

$$\varrho(x) := \frac{\mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle}}{\int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle} dx_1 \cdots dx_N}, \qquad x \in \mathbb{R}^N.$$

This is called the free probability density. The function $F : \mathbb{R}^N \to \mathbb{R}$ given by

$$F(x) := \int_{-\infty}^{x} \varrho(y) dy$$

is called the free probability distribution (or Gaussian distribution).

Free Moments. Choose the indices $k_1, k_2, \ldots, k_n = 1, 2, \ldots$, and fix the positive integer $n = 1, 2, \ldots$ Define

$$\boxed{\langle x_{k_1} x_{k_2} \cdots x_{k_n} \rangle := \int_{\mathbb{R}^N} x_{k_1} x_{k_2} \cdots x_{k_n} \cdot \varrho(x) dx_1 \cdots dx_N.}$$

These expectation values are called the moments of the probability density ϱ (or, briefly, the free moments). We also use the notation[93]

$$C_{n,\mathrm{free}}(x_{k_1}, x_{k_2}, \ldots, x_{k_n}) := \langle x_{k_1} x_{k_2} \cdots x_{k_n} \rangle,$$

and we call $C_{n,\mathrm{free}}$ a free discrete n-correlation function (or a free n-point function). Explicitly, we get[94]

$$\langle x_{k_1} x_{k_2} \cdots x_{k_n} \rangle := \frac{\int_{\mathbb{R}^N} x_{k_1} x_{k_2} \cdots x_{k_n} \mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle} \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}}}{\int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle} \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}}}.$$

The trick of the generating function. For all $J \in \mathbb{R}^N$, introduce the so-called generating function

$$Z_{\mathrm{free}}(J) := \int_{\mathbb{R}^N} \varrho(x) \mathrm{e}^{\langle J|x\rangle} \, dx_1 \cdots dx_N.$$

Explicitly, we get

[93] The value $C_{n,\mathrm{free}}(x_{k_1}, x_{k_2}, \ldots, x_{k_n})$ only depends on the indices $k_1, k_2, \ldots, k_n$. However, mnemonically, our notation is convenient for the passage to quantum field theory. Then we can use the same notation for the continuously varying variables $x_{k_1}, x_{k_2}, \ldots, x_{k_n}$.

[94] We introduce the rescaled differential $\frac{dx_k}{\sqrt{2\pi}}$ in order to prepare the limit $N \to \infty$ to path integrals (infinite-dimensional Gaussian integrals) later on.

$$Z_{\text{free}}(J) := \frac{\int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle}\, \mathrm{e}^{\langle J|x\rangle}\, \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}}}{\int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle}\, \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}}} = \mathrm{e}^{\frac{1}{2}\langle J|A^{-1}J\rangle}.$$

Differentiation with respect to J yields

$$\langle x_{k_1} x_{k_2} \cdots x_{k_n} \rangle = \frac{\partial^n}{\partial J_{k_1} \partial J_{k_2} \cdots \partial J_{k_n}}\, \mathrm{e}^{\frac{1}{2}\langle J|A^{-1}J\rangle}, \tag{7.191}$$

by setting $J = 0$ after differentiation. In particular, for the free 2-point function we get

$$\boxed{\langle x_k x_l \rangle = (A^{-1})_{kl}, \qquad k, l = 1, \ldots, N}$$

where $(A^{-1})_{kl}$ is the entry of the inverse matrix A^{-1} located in the kth row and in the lth column.

Theorem 7.46 *Let $k_1, \ldots, k_n = 1, 2, \ldots N$. If n is even, then*

$$\langle x_{k_1} x_{k_2} \cdots x_{k_n} \rangle = \sum \langle x_{i_1} x_{i_2} \rangle \langle x_{i_3} x_{i_4} \rangle \cdots \langle x_{i_{n-1}} x_{i_n} \rangle.$$

Here, we sum over all possible pairings of the indices $k_1, k_2, \ldots, k_n$. If n is odd, then $\langle x_{k_1} x_{k_2} \cdots x_{k_n} \rangle = 0$.

This so-called Wick theorem tells us that the Gaussian distribution has the following important property: the 2-point function determines all the other n-point functions.

Proof. Observe that the function $J \mapsto \langle J|A^{-1}J\rangle$ is quadratic. If n is even, then use (7.191) together with the chain rule. If n is odd, then note that the function $(x_1, x_2, x_3) \mapsto x_1 x_2 x_3$ is odd, and so on. □

Feynman diagrams. For example, the Wick theorem tells us that

$$\langle x_1 x_2 x_3 x_4 \rangle = \langle \underbrace{x_1}\, \underbrace{x_2}\, \underbrace{x_3}\, x_4 \rangle + \langle \overbrace{\underbrace{x_1 x_2}\, \underbrace{x_3}}^{}\, x_4 \rangle + \langle \underbrace{x_1}\, \overbrace{\underbrace{x_2}\, \underbrace{x_3}}^{}\, x_4 \rangle. \tag{7.192}$$

That is, we sum over all possible fully contracted symbols. Explicitly, this means

$$\langle x_1 x_2 x_3 x_4 \rangle = \langle x_1 x_2 \rangle \langle x_3 x_4 \rangle + \langle x_1 x_3 \rangle \langle x_2 x_4 \rangle + \langle x_1 x_4 \rangle \langle x_2 x_3 \rangle.$$

This is graphically represented in Table 7.1(c) by using so-called Feynman diagrams. Here, the contractions correspond to connections of the vertices. Similarly, we get

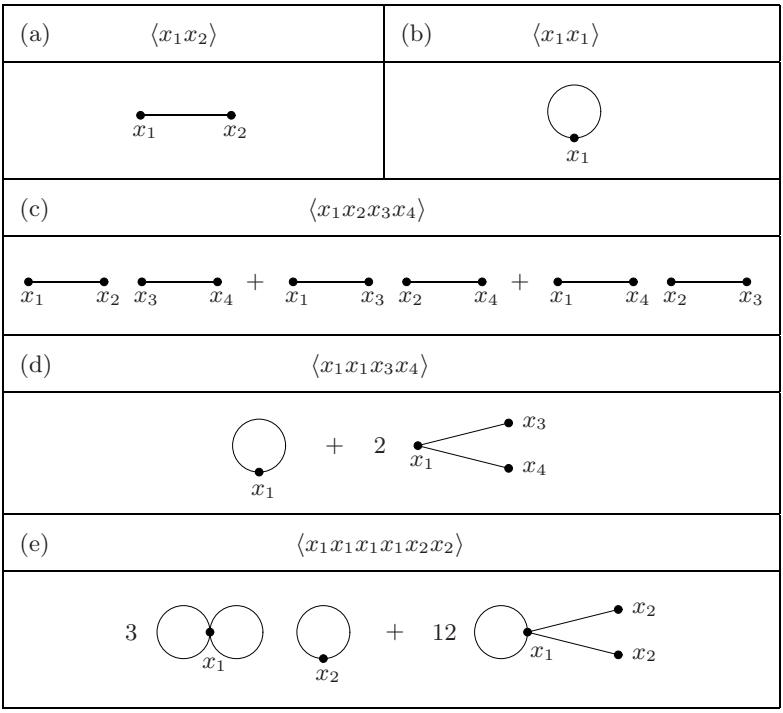
$$\langle x_1 x_1 x_3 x_4 \rangle = \langle \underbrace{x_1}\, \underbrace{x_1}\, \underbrace{x_3}\, x_4 \rangle + \langle \overbrace{\underbrace{x_1 x_1}\, \underbrace{x_3}}^{}\, x_4 \rangle + \langle \underbrace{x_1}\, \overbrace{\underbrace{x_1}\, \underbrace{x_3}}^{}\, x_4 \rangle. \tag{7.193}$$

This is graphically represented in Table 7.1(d). Naturally enough, the diagram corresponding to $\langle x_1 x_1 \rangle$ is called a loop. Analogously,

$$\langle x_1^4 x_2^2 \rangle = 3\langle x_1^2 \rangle^2 \langle x_2^2 \rangle + 12 \langle x_1^2 \rangle \langle x_1 x_2 \rangle^2. \tag{7.194}$$

This is computed in Problem 7.33 (see Table 7.1(e)).

Table 7.1. Feynman diagrams



7.8.3 Full Moments and Perturbation Theory

Distinguish strictly between free moments and full moments.
Folklore

Now we pass to probability distributions which are perturbations of Gaussian distributions. The strength of perturbation is measured by the coupling constant κ. This way, free moments (resp. free n-point functions) are replaced by full moments (resp. full n-point functions).

The full probability distribution under interaction. Assume that the matrix A has the property (H) formulated on page 562. Let

$$U : \mathbb{R}^N \to \mathbb{R}$$

be a polynomial with respect to the real variables $x_1, \dots, x_N$ (e.g., we choose $U(x) := -\langle x|x\rangle^2$). We are given the real nonnegative number κ called coupling constant. Introduce

$$\boxed{\varrho_\kappa(x) := \frac{\mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle}\mathrm{e}^{\kappa U(x)}}{\int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle}\mathrm{e}^{\kappa U(x)} dx_1 \cdots dx_N}, \qquad x \in \mathbb{R}^N.} \tag{7.195}$$

This is called the full probability density, which depends on the coupling constant κ. The function $F_\kappa : \mathbb{R}^N \to \mathbb{R}$ given by

$$F_\kappa(x) := \int_{-\infty}^{x} \varrho_\kappa(y)dy, \qquad x \in \mathbb{R}^N$$

is called the full probability distribution (or perturbed Gaussian distribution). The function κU measures the strength of the perturbation. As a rule, we will consider the case where the coupling constant κ is small. We assume that the function ϱ_κ is well defined, that is, the denominator of ϱ_κ in (7.195) is a finite integral. Note that the free probability density corresponds to the case where the coupling constant vanishes, $\kappa = 0$. Define

$$\boxed{\langle x_{k_1} x_{k_2} \cdots x_{k_n} \rangle_{\text{full}} := \int_{\mathbb{R}^N} x_{k_1} x_{k_2} \cdots x_{k_n} \cdot \varrho_\kappa(x) dx_1 \cdots dx_N.}$$

These expectation values are called the full moments. We also use the notation

$$C_{n,\text{full}}(x_{k_1}, x_{k_2}, \ldots, x_{k_n}) := \langle x_{k_1} x_{k_2} \cdots x_{k_n} \rangle_{\text{full}},$$

and we call $C_{n,\text{full}}$ a full discrete n-correlation function (or a full n-point function). Explicitly, we get

$$\langle x_{k_1} x_{k_2} \cdots x_{k_n} \rangle_{\text{full}} := \frac{\int_{\mathbb{R}^N} x_{k_1} x_{k_2} \cdots x_{k_n} \mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle} \, \mathrm{e}^{\kappa U(x)} \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}}}{\int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle} \mathrm{e}^{\kappa U(x)} \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}}}.$$

For all $J \in \mathbb{R}^N$, introduce the full generating function

$$Z_{\text{full}}(J) := \int_{\mathbb{R}^N} \varrho_\kappa(x) \, \mathrm{e}^{\langle J|x\rangle} \, dx_1 \cdots dx_N.$$

Explicitly,

$$Z_{\text{full}}(J) = \frac{\int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle} \, \mathrm{e}^{\kappa U(x)} \, \mathrm{e}^{\langle J|x\rangle} \, \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}}}{\int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle} \, \mathrm{e}^{\kappa U(x)} \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}}}.$$

Differentiation with respect to J yields

$$\langle x_{k_1} x_{k_2} \cdots x_{k_n} \rangle_{\text{full}} = \frac{\partial^n Z_{\text{full}}(0)}{\partial J_{k_1} \partial J_{k_2} \cdots \partial J_{k_n}}. \tag{7.196}$$

By Taylor expansion,

$$Z_{\text{full}}(J) = 1 + \sum_{n=1}^{\infty} \sum_{r_1 + r_2 + \ldots + r_N = n} \frac{\langle x_1^{r_1} x_2^{r_2} \cdots x_N^{r_N} \rangle_{\text{full}}}{r_1! r_2! \cdots r_N!} \, J_1^{r_1} J_2^{r_2} \cdots J_N^{r_N}.$$

Perturbation theory. We want to compute the following full moment:

$$\langle x_{k_1} x_{k_2} \cdots x_{k_n} \rangle_{\text{full}} = \frac{\int_{\mathbb{R}^N} x_{k_1} x_{k_2} \cdots x_{k_n} \mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle} \, \mathrm{e}^{\kappa U(x)} \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}}}{\int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle} \mathrm{e}^{\kappa U(x)} \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}}}$$

where U is a polynomial.To this end, we start with the Taylor expansion

$$\mathrm{e}^{\kappa U} = 1 + \kappa U + \tfrac{1}{2}\kappa^2 U^2 + \ldots$$

Setting $a := \int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle} \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}}$, we get

$$\int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle} \mathrm{e}^{\kappa U(x)} \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}} = a\left(1 + \kappa\langle U(x)\rangle + \tfrac{1}{2}\kappa^2\langle U(x)^2\rangle + \ldots\right).$$

Similarly, the integral

$$\int_{\mathbb{R}^N} x_{k_1} \cdots x_{k_n} \mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle} \mathrm{e}^{\kappa U(x)} \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_N}{\sqrt{2\pi}}$$

is equal to

$$a(\langle x_{k_1} \cdots x_{k_n}\rangle + \kappa\langle x_{k_1} \cdots x_{k_n} U(x)\rangle + \tfrac{1}{2}\kappa^2\langle x_{k_1} \cdots x_{k_n} U(x)^2\rangle + \ldots).$$

Hence

$$\boxed{\langle x_{k_1} \cdots x_{k_n}\rangle_{\mathrm{full}} = \langle x_{k_1} \cdots x_{k_n}\rangle + \kappa\big(\langle x_{k_1} \cdots x_{k_n} U(x)\rangle - \langle U(x)\rangle\big) \ldots}$$

Here, the dots stand for terms of order $O(\kappa^2)$ as $\kappa \to 0$. Since $U(x)$ is a polynomial with respect to $x_1, \ldots, x_n$, the right-hand side only contains free moments. Therefore the Wick theorem tells us that

> *The computation of full moments can be reduced to the computation of the special free moments* $\langle x_i x_j\rangle$.

This is the secret behind the success of perturbation theory in quantum field theory. For example, let $N \geq 2$. Choose $U(x) := x_1^4$. Then $\langle U(x)\rangle = 3\langle x_1^2\rangle^2$. By (7.194), we get

$$\langle x_2^2\rangle_{\mathrm{full}} = \langle x_2\rangle^2 + \kappa\big(3\langle x_1^2\rangle^2\langle x_2^2\rangle + 12\langle x_1^2\rangle\langle x_1 x_2\rangle^2 - 3\langle x_1^2\rangle^2\big) + O(\kappa^2)$$

as $\kappa \to 0$.

The reduced full moments (cumulants). In order to avoid redundant expressions, let us introduce the reduced full generating function

$$Z_{\mathrm{full,red}}(J) := \ln Z_{\mathrm{full}}(J).$$

Then

$$\boxed{Z_{\mathrm{full}}(J) = \mathrm{e}^{Z_{\mathrm{full,red}}(J)}.} \tag{7.197}$$

By definition, $Z_{\mathrm{full,red}}$ is the generating function for the so-called reduced full moments:[95]

$$\langle x_{k_1} x_{k_2} \ldots x_{k_n}\rangle_{\mathrm{full,red}} := \frac{\partial^n Z_{\mathrm{full,red}}(0)}{\partial J_{k_1} \partial J_{k_2} \cdots \partial J_{k_n}}.$$

Hence

$$Z_{\mathrm{full,red}}(J) = 1 + \sum_{n=1}^{\infty} \sum_{r_1+r_2+\ldots+r_N=n} \frac{\langle x_1^{r_1} x_2^{r_2} \cdots x_N^{r_N}\rangle_{\mathrm{full,red}}}{r_1! r_2! \cdots r_N!} J_1^{r_1} J_2^{r_2} \cdots J_N^{r_N}.$$

Using Taylor expansion with respect to κ, it follows from (7.197) that

> *The full moments can be uniquely computed by means of the reduced full moments.*

[95] In mathematics, reduced moments are also called cumulants.

In the special free case where $\kappa = 0$, we obtain

$$Z_{\text{free,red}}(J) = \ln Z_{\text{free}}(J) = \ln \mathrm{e}^{\frac{1}{2}\langle J|A^{-1}J\rangle} = \tfrac{1}{2}\langle J|A^{-1}J\rangle.$$

Hence

$$\langle x_k x_l\rangle_{\text{free red}} = \langle x_k x_l\rangle, \qquad k,l = 1,\ldots,N.$$

The remaining reduced free moments are equal to zero. This implies the following result.

The reduced full generating function satisfies the relation

$$Z_{\text{full,red}}(J) = Z_{\text{free,red}}(J) + O(\kappa) = 1 + \sum_{i,k=1}^{N} \tfrac{1}{2}\langle x_i x_k\rangle J_i J_k + O(\kappa)$$

as $\kappa \to 0$. Therefore, the function $Z_{\text{full,red}}$ describes the perturbation of the second free moments, under the influence of the coupling constant κ. In contrast to this, the formula

$$Z_{\text{full}}(J) = Z_{\text{free}}(J) + O(\kappa), \qquad \kappa \to 0$$

is full of redundance, since the function Z_{free} is redundant compared with $Z_{\text{free,red}}$. This is why physicists use reduced full correlation (or n-point) functions in quantum field theory.

7.9 Rigorous Infinite-Dimensional Gaussian Integrals

> The definition of infinite-dimensional Gaussian integrals depends on the spectrum of the linear symmetric dispersion operator.
>
> Folklore

In order to explain the basic idea, let us start with the finite-dimensional key formula

$$\int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}(\lambda_1 x_1^2 + \ldots + \lambda_N^2 x_N^2)} \mathrm{e}^{b_1 x_1 + \ldots + b_N x_N} \frac{dx_1}{\sqrt{2\pi}} \ldots \frac{dx_N}{\sqrt{2\pi}} = B_N$$

where

$$B_N := \frac{\mathrm{e}^{\frac{1}{2}\sum_{k=1}^N b_k^2 \lambda_k^{-1}}}{\left(\prod_{k=1}^N \lambda_k\right)^{1/2}}.$$

Here, $N = 1, 2, \ldots$. Furthermore, we assume that $\lambda_1, \lambda_2, \ldots$ are positive numbers, and $b_1, b_2 \ldots$ are real numbers. Now we want to study the limit $N \to \infty$. Obviously, we have the following result.

Proposition 7.47 *Suppose that* $\sum_{k=1}^\infty b_k^2 \lambda_k^{-1} < \infty$ *and* $0 < \prod_{k=1}^\infty \lambda_k < \infty$. *Then the following limit*

$$\lim_{N\to\infty} \int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}(\lambda_1 x_1^2 + \ldots + \lambda_N^2 x_N^2)}\, \mathrm{e}^{b_1 x_1 + \ldots + b_N x_N} \frac{dx_1}{\sqrt{2\pi}} \ldots \frac{dx_N}{\sqrt{2\pi}} = \frac{\mathrm{e}^{\frac{1}{2}\sum_{k=1}^\infty b_k^2 \lambda_k^{-1}}}{\left(\prod_{k=1}^\infty \lambda_k\right)^{1/2}}$$

exists in the classical sense. We briefly write

$$\int_{\mathbb{R}^\infty} \mathrm{e}^{-\frac{1}{2}\sum_{k=1}^\infty \lambda_k x_k^2}\, e^{\sum_{k=1}^\infty b_k x_k} \prod_{k=1}^\infty \frac{dx_k}{\sqrt{2\pi}} := \frac{\mathrm{e}^{\frac{1}{2}\sum_{k=1}^\infty b_k^2 \lambda_k^{-1}}}{\left(\prod_{k=1}^\infty \lambda_k\right)^{1/2}}.$$

We call this a normalized infinite-dimensional Gaussian integral.

7.9.1 The Infinite-Dimensional Dispersion Operator

We want to generalize the preceding formulas. To this end, we are given the linear symmetric operator

$$A : D(A) \to X$$

defined on the linear dense subspace $D(A)$ of the real infinite-dimensional separable Hilbert space X. Assume that we have the eigenvector equation

$$\boxed{A\varphi_k = \lambda_k \varphi_k, \qquad k = 1, 2, \ldots}$$

where $\lambda_k > 0$ for all k, and the eigenvectors $\varphi_1, \varphi_2 \ldots$ form a complete orthonormal system of the Hilbert space X (together with $\varphi_k \in D(A)$ for all k). Then we obtain $b = \sum_{k=1}^{\infty} \langle b|\varphi_k\rangle \varphi_k$ for all $b \in X$, and

$$A\varphi = \sum_{k=1}^{\infty} \lambda_k \langle \varphi|\varphi_k\rangle \varphi_k \qquad \text{for all} \quad \varphi \in D(A).$$

This implies $\langle \varphi|A\varphi\rangle = \sum_{k=1}^{\infty} \lambda_k \langle \varphi_k|\varphi\rangle^2$. If $A\varphi = 0$, then $\varphi = 0$. Thus the operator A is injective, and the inverse operator $A^{-1} : D(A^{-1}) \to X$ exists with

$$A^{-1}\varphi_k = \lambda_k^{-1} \varphi_k, \qquad k = 1, 2, \ldots$$

In particular, we get $D(A^{-1}) \subseteq D(A)$, and

$$\langle b|A^{-1}b\rangle = \sum_{k=1}^{\infty} \lambda_k^{-1} \langle b|\varphi_k\rangle^2 \qquad \text{for all} \quad b \in D(A).$$

Furthermore, for the dispersion operator A, we define

- the trace $\operatorname{tr} A := \sum_{k=1}^{\infty} \lambda_k$,
- the determinant $\det A := \prod_{k=1}^{\infty} \lambda_k$, and
- the zeta function $\zeta_A(s) = \sum_{k=1}^{\infty} \lambda_k^{-s}$ for suitable complex numbers s.

If the trace is finite, that is $\operatorname{tr}(A) < \infty$, then

$$\det A = \mathrm{e}^{\operatorname{tr} A}.$$

In what follows, we are given $b \in D(A)$. We have to distinguish the following two cases.

(C1) Regular case: $0 < \det A < \infty$ (the determinant exists).
(C2) Singular case: $\det A = \infty$ (the determinant does not exist in the usual sense).

Regular case. Here, we define the normalized infinite-dimensional Gaussian integral by setting

$$\boxed{\int_{D(A)} \mathrm{e}^{-\frac{1}{2}\langle \varphi|A\varphi\rangle} \mathrm{e}^{\langle b|\varphi\rangle} \mathcal{D}_G\varphi := \frac{\mathrm{e}^{\frac{1}{2}\langle b|A^{-1}b\rangle}}{\sqrt{\det A}}.} \tag{7.198}$$

Observe that in concrete situations, the domain of definition $D(A)$ of the operator A describes boundary conditions. Changing the boundary conditions means changing the operator A and its eigenvalues. Since the determinant $\det A$ depends on the eigenvalues, the integral depends on the domain of definition $D(A)$. Now let us study the singular case which is typically encountered in quantum physics.

7.9.2 Zeta Function Regularization and Infinite-Dimensional Determinants

> The definition $\ln \det A := -\zeta_A'(0)$ was first used by the mathematicians Ray and Singer (1971), when they tried to give a definition of the Reidemeister–Franz torsion in analytic terms... Later zeta function regularization was used by physicists in the context of dimensional regularization when applied to quantum field theory in curved space-time.[96]
>
> Klaus Kirsten, 2002

It is our goal to use (7.198) and to redefine the determinant $\det A$ by means of the zeta function ζ_A together with analytic continuation.

Singular case. Motivated by (7.187), the key formula reads as

$$\boxed{\det A = \mathrm{e}^{-\zeta_A'(0)}.} \tag{7.199}$$

Let us assume the following:

(H) *The zeta function* $\zeta_A(s) = \sum_{n=1}^{\infty} \lambda_n^{-s}$ *converges for all sufficiently large positive real numbers* s*, and it can be analytically continued to some neighborhood of the point* $s = 0$ *in the complex plane.*

Here, we define the determinant $\det A$ of the operator A by (7.199). This generates the definition of the normalized infinite-dimensional Gaussian integral in the singular case:

$$\int_{D(A)} \mathrm{e}^{-\frac{1}{2}\langle \varphi | A\varphi\rangle} \mathrm{e}^{\langle b|\varphi\rangle} \mathcal{D}_G\varphi := \mathrm{e}^{\frac{1}{2}\langle b|A^{-1}b\rangle}\ \mathrm{e}^{\frac{1}{2}\zeta_A'(0)}. \tag{7.200}$$

The rescaling trick. Let γ be a nonzero complex number with the property $-\pi < \arg(\gamma) < \pi$, and assume (H). We define the normalized infinite-dimensional Gaussian integral by setting

$$\boxed{\int_{D(A)} \mathrm{e}^{-\frac{1}{2}\gamma\langle \varphi | A\varphi\rangle} \mathrm{e}^{\langle b|\varphi\rangle} \mathcal{D}_G\varphi := \mathrm{e}^{-\frac{1}{2}\zeta_A(0)\ln\gamma} \cdot \frac{\mathrm{e}^{\frac{1}{2}\gamma^{-1}\langle b|A^{-1}b\rangle}}{\sqrt{\det A}}} \tag{7.201}$$

with $\sqrt{\det A} := \mathrm{e}^{-\frac{1}{2}\zeta_A'(0)}$, and $\ln\gamma$ is the principal value of the logarithm. Definition (7.201) is crucial for quantum physics, as we will show in the next section. In order to motivate (7.201), observe first that the following hold.

Proposition 7.48 *Let* $\gamma > 0$*. Assume that the hypothesis* (H) *above is valid. Then*

$$\det(\gamma A) = \gamma^{\zeta_A(0)} \det A.$$

---

[96] K. Kirsten, Spectral Functions in Mathematics and Physics, Chapman, Boca Raton, Florida, 2002 (see also the hints for further reading on page 671).
D. Ray and I. Singer, Reidemeister torsion and the Laplacian on Riemann manifolds, Advances in Math. **7**, (1971) 145–210.

It was independently proven by Werner Müller and Jeff Cheeger that the original combinatorial definition of the Reidemeister–Franz torsion is equivalent to the analytic definition:
W. Müller, Analytic torsion and Reidemeister torsion of Riemannian manifolds, Advances in Math. **28** (1978), 233–305.
J. Cheeger (1979), Analytic torsion and the heat equation, Ann. Math. **109**, 259–322.

This generalizes the classical relation $\det(\gamma A) = \gamma^N \det A$ which is valid in the N-dimensional Euclidean space with $N = 1, 2, \ldots$ The proof will be given in Problem 7.34 by using Euler's gamma function. Replacing A by γA it follows from (7.200) that

$$\int_{D(A)} \mathrm{e}^{-\frac{1}{2}\langle\varphi|(\gamma A)\varphi\rangle} \mathrm{e}^{\langle b|\varphi\rangle} \mathcal{D}_G\varphi := \frac{\mathrm{e}^{\frac{1}{2}\langle b|(\gamma A)^{-1}b\rangle}}{\sqrt{\det(\gamma A)}}, \qquad \gamma > 0.$$

This yields (7.201) if $\gamma > 0$. For general complex numbers γ (outside the negative real axis), the right-hand side of (7.201) makes sense after analytic continuation.

The quotient trick. Fortunately enough, in quantum field theory one frequently encounters quotients of Gaussian integrals which dramatically simplifies the approach. To illustrate this, note that, in the regular case, it follows from (7.198) that

$$\frac{\int_{D(A)} \mathrm{e}^{-\frac{1}{2}\langle\varphi|A\varphi\rangle} \mathrm{e}^{\langle b|\varphi\rangle} \mathcal{D}_G\varphi}{\int_{D(A)} \mathrm{e}^{-\frac{1}{2}\langle\varphi|A\varphi\rangle} \mathcal{D}_G\varphi} := \mathrm{e}^{\frac{1}{2}\langle b|A^{-1}b\rangle}. \tag{7.202}$$

This expression is independent of the determinant $\det A$. Therefore, we use this as a definition for all dispersion operators A and all $b \in D(A)$. This way, the use of the critical determinant $\det A$ is completely avoided.

Example. Let $m > 0$. The following example will be used below in order to study the free quantum particle on the real line. Consider the quadratic form

$$S[r] := \tfrac{1}{2} m \langle r|Ar\rangle, \qquad r \in D(A)$$

with the linear differential operator $A : D(A) \to X$ given by

$$Ar := -\frac{d^2 r}{d\tau^2}, \qquad r \in D(A).$$

Here, X is the real Hilbert space $L_2(\mathbb{R})$, and the domain of definition $D(A)$ consists of all twice continuously differentiable functions $r : [s,t] \to \mathbb{R}$ with $r(s) = r(t) = 0$. We write $C_0^2[s,t]$ instead of $D(A)$. Integration by parts yields

$$S[r] = -\int_s^t \tfrac{1}{2} m r(\tau) \ddot{r}(\tau) d\tau = \int_s^t \tfrac{1}{2} m \dot{r}(\tau)^2 \, d\tau$$

for all $r \in C_0^2[s,t]$. This is the action of a free quantum particle on the real line with the boundary condition $r(s) = r(t) = 0$.

Proposition 7.49 *There holds* $\int_{C_0^2[s,t]} e^{\mathrm{i}S[r]/\hbar} \, \mathcal{D}_G r = \frac{1}{\sqrt{2(t-s)}} \left(\frac{m}{\hbar}\right)^{1/4} e^{-\mathrm{i}\pi/8}$.

Proof. To simplify notation, set $s := 0$. The crucial eigenvalue problem

$$A\varphi = \lambda\varphi, \qquad \varphi \in D(A)$$

corresponds to the equation $-\ddot{\varphi}(\tau) = \lambda\varphi(\tau), 0 < \tau < t$ with the boundary condition $\varphi(0) = \varphi(t) = 0$. The solutions are

$$\varphi_n(\tau) := \text{const} \cdot \sin\sqrt{\lambda_n}\tau, \qquad \lambda_n := \left(\frac{n\pi}{t}\right)^2, \qquad n = 1, 2, \ldots$$

For all complex numbers z with $\Re(z) > \frac{1}{2}$, the zeta function ζ_A of the operator A is given by the convergent series

$$\zeta_A(z) = \sum_{n=1}^{\infty} \frac{1}{\lambda_n^z} = \left(\frac{t}{\pi}\right)^{2z} \sum_{n=1}^{\infty} \frac{1}{n^{2z}} = \left(\frac{t}{\pi}\right)^{2z} \zeta(2z).$$

Here, ζ denotes the Riemann zeta function. Note that ζ can be analytically continued to a holomorphic function on the pointed plane $\mathbb{C} \setminus \{1\}$. Here, $\zeta(0) = -\frac{1}{2}$ and $\zeta'(0) = -\frac{1}{2} \ln 2\pi$. Hence $\zeta_A(0) = -\frac{1}{2}$ and

$$\zeta_A'(0) = 2\zeta(0)(\ln t - \ln \pi) + 2\zeta'(0) = -\ln 2t.$$

This implies $\det A = \mathrm{e}^{-\zeta_A'(0)} = 2t$. Set $\gamma := \frac{m}{\hbar \mathrm{i}}$. By (7.201), the integral $\int_{C_0^2[0,t]} \mathrm{e}^{\mathrm{i}S[r]/\hbar}\, \mathcal{D}_G r$ is equal to

$$\int_{C_0^2[0,t]} \mathrm{e}^{-\frac{1}{2}\gamma\langle r|Ar\rangle} \mathcal{D}_G r = \frac{\mathrm{e}^{-\frac{1}{2}\zeta_A(0)\ln\gamma}}{\sqrt{\det A}} = \frac{1}{\sqrt{2t}} \left(\frac{m}{\hbar \mathrm{i}}\right)^{1/4}.$$

This is the desired result. □

7.9.3 Application to the Free Quantum Particle

Consider the motion of a free quantum particle on the real line. In Theorem 7.16 on page 488, we have computed the corresponding Feynman propagator kernel

$$\mathcal{K}(x,t;y,s) = \sqrt{\frac{m}{2\pi\hbar \mathrm{i}(t-s)}}\ \mathrm{e}^{\mathrm{i}m(x-y)^2/2\hbar(t-s)} \tag{7.203}$$

for all positions $x, y \in \mathbb{R}$ and all times $t > s$.[97] In addition, we have shown that the dynamics of the free quantum particle is governed by the formula

$$\boxed{\psi(x,t) := \int_{\mathbb{R}} \mathcal{K}(x,t;y,s)\psi(y,s)\, dy, \quad x \in \mathbb{R}, t > s.} \tag{7.204}$$

If we know the Schrödinger wave function ψ of the free particle at time s, then the kernel formula (7.204) tells us how to obtain the wave function at the later time t. This explains the importance of the Feynman propagator kernel. In Prop. 7.44 on page 550, we have proved that

$$\mathcal{K}(x,t;y,s) = \int_{C\{s,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}q.$$

That is, the Feynman propagator kernel can be represented by a Feynman path integral. In this section, it is our goal to prove that

[97] Recall that the square root is to be understood as principal value. Explicitly,

$$\sqrt{\frac{m}{2\pi\hbar \mathrm{i}(t-s)}} = \mathrm{e}^{-\mathrm{i}\pi/4} \sqrt{\frac{m}{2\pi\hbar(t-s)}}.$$

$$\mathcal{K}(x,t;y,s) = \mathcal{N} \int_{C\{s,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}_G q.$$

This implies the key formula

$$\boxed{\mathcal{K}(x,t;y,s) = \int_{C\{s,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}q = \mathcal{N} \int_{C\{s,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}_G q} \tag{7.205}$$

for all positions $x, y \in \mathbb{R}$ and all times $t > s$. This formula tells us the crucial fact that the Feynman path integral coincides with the corresponding normalized infinite-dimensional Gaussian integral, up to some normalization factor $\mathcal{N}$. Explicitly, $\mathcal{N} = \left(\frac{m}{\pi^2\hbar}\right)^{1/4} \mathrm{e}^{-\mathrm{i}\pi/8}$.

The classical trajectory. The action of a classical free particle of mass m on the real line is given by

$$S[q] := \int_s^t \tfrac{1}{2} m \dot{q}(\tau)^2 \, d\tau.$$

The boundary-value problem

$$m\ddot{q}(\tau) = 0, \quad s < \tau < t, \quad q(s) = y,\ q(t) = x$$

corresponds to the motion of the particle with given endpoints. The unique solution is $q_{\mathrm{class}}(\tau) = y + \frac{\tau - s}{t - s}(x - y)$ with the classical action

$$S[q_{\mathrm{class}}] = \int_s^t \tfrac{1}{2} m \dot{q}_{\mathrm{class}}(\tau)^2 d\tau = \frac{m(x-y)^2}{2(t-s)}.$$

Quantum fluctuations and the WKB relation. In order to study perturbations of the classical trajectory, we consider the trajectories

$$q(\tau) = q_{\mathrm{class}}(\tau) + r(\tau), \qquad \tau \in [s,t]$$

where $r \in C_0^2[s,t]$, that is, the function $r : [s,t] \to \mathbb{R}$ is twice continuously differentiable and satisfies the boundary condition $r(s) = r(t) = 0$. By (7.159) on page 550,

$$S[q] = S[q_{\mathrm{class}}] + S[r]. \tag{7.206}$$

For the Feynman propagator kernel, it follows from (7.203) that

$$\boxed{\mathcal{K}(x,t;y,s) = \mathrm{e}^{\mathrm{i}S[q_{\mathrm{class}}]/\hbar}\, \mathcal{K}_{\mathrm{fluct}}(t;s)} \tag{7.207}$$

for all positions $x, y \in \mathbb{R}$ and all times $t > s$, with the fluctuation term

$$\mathcal{K}_{\mathrm{fluct}}(t;s) := \sqrt{\frac{m}{2\pi\hbar \mathrm{i}(t-s)}}.$$

Equation (7.207) is called the WKB relation for the free quantum particle. It shows that the Feynman propagator is the product of the purely classical factor $\mathrm{e}^{\mathrm{i}S[q_{\mathrm{class}}]/\hbar}$ with a factor caused by quantum fluctuations.

The key relation. Motivated by the decomposition formula (7.206), we define

$$\int_{C\{s,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}_G q := \mathrm{e}^{\mathrm{i}S[q_{\mathrm{class}}]/\hbar} \int_{C_0^2[s,t]} \mathrm{e}^{\mathrm{i}S[r]/\hbar}\, \mathcal{D}_G r.$$

Prop. 7.49 on page 573 tells us that

$$\int_{C_0^2[s,t]} \mathrm{e}^{\mathrm{i}S[r]/\hbar}\, \mathcal{D}_G r = \frac{\mathcal{K}_{\mathrm{fluct}}(t;s)}{\mathcal{N}}$$

with the normalization constant $\mathcal{N} = \left(\frac{m}{\pi^2\hbar}\right)^{1/4} \mathrm{e}^{-\mathrm{i}\pi/8}$. This implies the key formula (7.205).

7.9.4 Application to the Quantized Harmonic Oscillator

Parallel to the free quantum particle in the preceding section, let us now study the harmonic oscillator of mass $m > 0$ and angular frequency $\omega > 0$ on the real line. Introduce the characteristic length $x_0 := \sqrt{\frac{\hbar}{m\omega}}$. Furthermore, choose the time parameter in such a way that

$$t \in \,]s + t_{n,\mathrm{crit}}, s + t_{n+1,\mathrm{crit}}[, \qquad n = 0, 1, 2, ... \tag{7.208}$$

Here, the critical points of time are defined by $t_{n,\mathrm{crit}} := \frac{n\pi}{\omega}$. In addition, we introduce the Maslov index by $\mu(s,t) := n$. By formula (7.144) on page 537, we have computed the Feynman propagator kernel for the quantized harmonic oscillator:

$$\mathcal{K}(x,t;y,s) = \frac{\mathrm{e}^{-\mathrm{i}\pi/4}\, \mathrm{e}^{-\mathrm{i}\pi\mu(s,t)/2}}{x_0\sqrt{2\pi|\sin\omega(t-s)|}} \exp\left(\mathrm{i}\frac{(x^2+y^2)\cos\omega(t-s) - 2xy}{2x_0^2 \sin\omega(t-s)}\right).$$

This formula is valid for all all positions $x, y \in \mathbb{R}$ and all non-critical times $t > s$ from (7.208).

The classical trajectory. The action of the classical harmonic oscillator is given by

$$S[q] := \int_s^t \tfrac{1}{2}m\dot{q}(\tau)^2 - \tfrac{1}{2}m\omega^2 q(\tau)^2\, d\tau.$$

The boundary-value problem

$$\ddot{q}(\tau) + \omega^2 q(\tau) = 0, \quad s < \tau < t, \quad q(s) = y,\ q(t) = x \tag{7.209}$$

has the solution $q_{\mathrm{class}}(\tau) = y\cos\omega(\tau - s) + \big(x - y\cos\omega(\tau - s)\big)\, \frac{\sin\omega(\tau-s)}{\sin\omega(t-s)}$. This is a classical trajectory with the action

$$S[q_{\mathrm{class}}] = \hbar \cdot \frac{(x^2+y^2)\cos\omega(t-s) - 2xy}{2x_0^2\sin\omega(t-s)}\ .$$

Note that the trajectory q_{class} is the unique solution of (7.209) if t is a non-critical point of time. The uniqueness is violated for critical points of time. In what follows, we only consider non-critical points of time (7.208).

Quantum fluctuations and the WKB relation. Now use the perturbed trajectory

$$q(t) = q_{\mathrm{class}}(\tau) + r(\tau), \qquad \tau \in [s,t],$$

where $r \in C_0^2[s,t]$, that is, the function $r : [s,t] \to \mathbb{R}$ is twice continuously differentiable and satisfies the boundary condition $r(s) = r(t) = 0$. By (7.165) on page 552, we get

$$S[q] = S[q_{\text{class}}] + S[r]. \tag{7.210}$$

The Feynman propagator kernel for the quantized harmonic oscillator can be written as

$$\boxed{\mathcal{K}(x,t;y,s) = \mathrm{e}^{\mathrm{i}S[q_{\text{class}}]/\hbar}\, \mathcal{K}_{\text{fluct}}(t;s)} \tag{7.211}$$

with the quantum fluctuation term

$$\mathcal{K}_{\text{fluct}}(t;s) := \frac{\mathrm{e}^{-\mathrm{i}\pi/4}\, \mathrm{e}^{-\mathrm{i}\pi\mu(s,t)/2}}{x_0\sqrt{2\pi|\sin\omega(t-s)|}}.$$

This is a special case of the WKB method (see (7.216) on page 581). Observe that the fluctuation term is independent of the position coordinates x and y.

Now we restrict ourselves to the first critical time interval, that is, we assume that $t \in]s, s + t_{1,\text{crit}}[$.

Our goal is the key relation (7.214) below. Let us first compute the following normalized infinite-dimensional Gaussian integral.

Proposition 7.50 *For all times* $t \in]s, s + t_{1,\text{crit}}[$, *we have*

$$\int_{C_0^2[s,t]} \mathrm{e}^{\mathrm{i}S[r]/\hbar}\, \mathcal{D}_G r = \frac{\mathcal{K}_{\text{fluct}}(t;s)}{\mathcal{N}(\omega)}. \tag{7.212}$$

The complex non-zero constant $\mathcal{N}(\omega)$ *will be determined below.*

Proof. We will proceed as in the proof of Prop. 7.49 on page 573. To simplify notation, set $s := 0$. For $r \in C_0^2[0,t]$, integration by parts yields

$$\frac{\mathrm{i}S[r]}{\hbar} = \frac{\mathrm{i}m}{2\hbar}\int_0^t r(\tau)\left(-\ddot{r}(\tau) - \omega^2 r(\tau)\right)\, d\tau = -\tfrac{1}{2}\gamma\langle r|Br\rangle$$

with $\gamma := \frac{m}{\hbar \mathrm{i}}$. Here, we introduce the differential operator $B : D(B) \to L_2(\mathbb{R})$ with

$$Br := -\frac{d^2 r}{d\tau^2} - \omega^2 r^2$$

and the domain of definition $D(B) := C_0^2[0,t]$.

(I) The infinite-dimensional Gaussian integral. By (7.201) on page 572, we get

$$\int_{C_0^2[0,t]} \mathrm{e}^{-\frac{1}{2}\gamma\langle r|Br\rangle} \mathcal{D}_G r := \frac{\mathrm{e}^{-\frac{1}{2}\zeta_B(0)\ln\gamma}}{\sqrt{\det B}}. \tag{7.213}$$

We have to compute the determinant $\det B = \mathrm{e}^{-\zeta_B'(0)}$.

(II) The eigenvalues. The crucial eigenvalue problem

$$B\varphi = \lambda\varphi, \qquad \varphi \in D(B)$$

corresponds to the equation $-\ddot{\varphi}(\tau) - \omega^2 q(\tau) = \lambda\varphi(\tau), 0 < \tau < t$ with the boundary condition $\varphi(0) = \varphi(t) = 0$. The solutions are

$$\varphi_n(\tau) := \text{const} \cdot \sin\sqrt{\lambda_n}\tau, \qquad \lambda_n := \left(\frac{n\pi}{t}\right)^2 - \omega^2, \qquad n = 1, 2, \ldots$$

Let us also introduce $\mu_n := \left(\frac{n\pi}{t}\right)^2$ which is obtained from λ_n by setting $\omega = 0$.

(III) The zeta function: For all complex numbers z with $\Re(z) > \frac{1}{2}$, the zeta function ζ_B of the operator B is given by the convergent series

$$\zeta_B(z) = \sum_{n=1}^{\infty} \frac{1}{\lambda_n^z} = \left(\frac{t}{\pi}\right)^{2z} \sum_{n=1}^{\infty} \frac{1}{\left(n^2 - \frac{t^2\omega^2}{\pi^2}\right)^z} \,.$$

Because of the boundary condition $r(0) = r(t) = 0$, the differential operator B can be regarded as an elliptic differential operator on a circle, which is the simplest example of a compact Riemannian manifold. There exists a general theory of elliptic operators on compact Riemannian manifolds which tells us that the corresponding zeta function can be analytically extended to a meromorphic function on the complex plane, and this extension is holomorphic at the origin $z = 0$ (see Gilkey (1995) and Kirsten (2002)). Therefore, $\zeta_B(0)$ and $\zeta_B'(0)$ are well-defined, and we can use the method of zeta-function regularization. In order to get quickly an explicit result, we will introduce a modified method which is used by physicists.

(IV) The determinant $\det B$. Formally, we get

$$\det B = \prod_{n=1}^{\infty} \lambda_n = \prod_{n=1}^{\infty} \mu_n \prod_{n=1}^{\infty} \left(1 - \frac{\omega^2}{\mu_n}\right).$$

By the classical Euler formula, we have the following convergent product

$$\sin z = z \prod_{n=1}^{\infty} \left(1 - \frac{z^2}{n^2\pi^2}\right), \qquad z \in \mathbb{C}.$$

Hence $\det B = \frac{\sin \omega t}{\omega t} \prod_{n=1}^{\infty} \mu_n$. The product $\prod_{n=1}^{\infty} \mu_n$ is divergent. In order to regularize $\det B$ it is sufficient to regularize $\prod_{n=1}^{\infty} \mu_n$. However, this product is the determinant of the operator B with $\omega = 0$ which coincides with the operator A from the proof of Prop. 7.49 on page 573. By this proof, $\det A = 2t$. Therefore, we define

$$\det B := \left(\prod_{n=1}^{\infty} \mu_n\right)_{\text{reg}} \prod_{n=1}^{\infty} \left(1 - \frac{\omega^2}{\mu_n}\right) = 2t \cdot \frac{\sin \omega t}{\omega t} = \frac{2 \sin \omega t}{\omega}.$$

(V) The constant $\mathcal{N}(\omega)$. By (7.213), the integral $\int_{C_0^2[0,t]} \mathrm{e}^{\mathrm{i}S[r]/\hbar} \mathcal{D}_G r$ is equal to

$$\frac{\mathrm{e}^{-\frac{1}{2}\zeta_B(0)\ln\gamma}\sqrt{\omega}}{\sqrt{2\sin\omega t}} = \frac{\mathcal{K}_{\text{fluct}}(t;0)}{\mathcal{N}(\omega)} = \frac{\mathrm{e}^{-\mathrm{i}\pi/4}}{\mathcal{N}(\omega) x_0 \sqrt{2\pi \sin \omega t}},$$

where $\gamma = \frac{m}{\hbar \mathrm{i}}$ and $x_0 = \sqrt{\frac{\hbar}{m\omega}}$. This yields $\mathcal{N}(\omega) = \mathrm{e}^{-\mathrm{i}\pi/4} \mathrm{e}^{\frac{1}{2}\zeta_B(0)\ln\gamma} \sqrt{\frac{m}{\pi\hbar}}$. □

The key relation. Motivated by the decomposition formula (7.210), we define

$$\int_{C\{s,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar} \, \mathcal{D}_G q := \mathrm{e}^{\mathrm{i}S[q_{\text{class}}]/\hbar} \int_{C_0^2[s,t]} \mathrm{e}^{\mathrm{i}S[r]/\hbar} \, \mathcal{D}_G r.$$

It follows from Prop. 7.50 together with (7.211) that

$$\mathcal{K}(x,t;y,s) = \mathcal{N}(\omega) \int_{C\{s,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar} \, \mathcal{D}_G q$$

for all $x, y \in \mathbb{R}$ and all $t \in]s, s + t_{1,\mathrm{crit}}[$. By Prop. 7.45 on page 552,

$$\mathcal{K}(x,t;y,s) = \int_{C\{s,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}q.$$

This implies the desired key relation

$$\boxed{\mathcal{K}(x,t;y,s) = \int_{C\{s,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}q = \mathcal{N}(\omega) \int_{C\{s,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}_G q} \tag{7.214}$$

for all positions $x, y \in \mathbb{R}$ and all times $t \in]s, s + t_{1,\mathrm{crit}}[$. Observe that for $\omega = 0$, the normalization factor $\mathcal{N}(0) = \left(\frac{m}{\pi^2\hbar}\right)^{1/4} \mathrm{e}^{-\mathrm{i}\pi/8}$ is the same as for the free quantum particle.

The free quantum particle as a limit. For all times $t \in]s, s + t_{1,\mathrm{crit}}[$ and all positions $x, y \in \mathbb{R}$, we have

$$\mathcal{K}(x,t;y,s) = \frac{1}{x_0\sqrt{2\pi\mathrm{i}\sin\omega(t-s)}} \quad \exp\left(\mathrm{i}\frac{(x^2+y^2)\cos\omega(t-s) - 2xy}{2x_0^2\sin\omega(t-s)}\right).$$

Noting that $\lim_{\omega\to+0} x_0^2 \sin\omega(t-s) = \frac{\hbar(t-s)}{m} \lim_{\omega\to+0} \frac{\sin\omega(t-s)}{\omega(t-s)} = \frac{\hbar(t-s)}{m}$, we obtain the limit relation

$$\lim_{\omega\to+0} \mathcal{K}(x,t;y,s) = \mathcal{K}_{\mathrm{free}}(x,t;y,s) = \sqrt{\frac{m}{2\pi\hbar\mathrm{i}(t-s)}}\ \mathrm{e}^{\mathrm{i}m(x-y)^2/2\hbar(t-s)}.$$

This tells us the quite natural fact that the Feynman propagator kernel of the quantized harmonic oscillator passes over to the Feynman propagator kernel of the free quantum particle if the angular frequency ω goes to zero.

7.9.5 The Spectral Hypothesis

Motivated by the rigorous results above for the free quantum particle and the quantized harmonic oscillator, we formulate the following general spectral hypothesis:

$$\int_{C\{s,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}q = \mathcal{N} \int_{C\{s,t\}} \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}_G q. \tag{7.215}$$

This hypothesis tells us that the Feynman path integral coincides with the corresponding normalized infinite-dimensional Gaussian integral, up to a normalization factor $\mathcal{N}$ which depends on the action functional S. Physicists take this spectral hypothesis for granted in both quantum mechanics and quantum field theory. The experience of physicists shows that this hypothesis works well as a universal tool. In terms of mathematics, it turns out that this heuristic tool also works well for conjecturing new topological invariants in the setting of topological quantum field theory and string theory. For example, this concerns knot theory, smooth manifolds in differential geometry, and algebraic varieties (generalized manifolds including singularities) in algebraic geometry.

7.10 The Semi-Classical WKB Method

> The WKB method in physics is the prototype of singular perturbation theory in mathematics.
>
> Folklore

> To the best of our knowledge, the first paper on path integrals, apart from Feynman's, written by a physicist was submitted by Cécile Morette in 1950.[98]
> During Pauli's stay at the Institute for Advanced Study in 1949, Morette and Van Hove presented to Pauli at the occasion of an appointment with him a semiclassical formula (S) for quantum mechanics based on Morette's approach to path integrals... Pauli wrote a number of research notes ... In these notes Pauli corrected a sign factor, and he obtained the important (exact) result that for small time intervals, the semiclassical propagator kernel from (S) satisfies the Schrödinger equation up to order $\hbar^2$ ...
> Pauli was, to the best of our knowledge, the first of the older generation, having laid the foundations of quantum mechanics, who fully appreciated the new approach developed by Feynman.[99]
>
> Christian Grosche and Frank Steiner, 1998

Approximation methods play an important role in physics in order to simplify computation and to get insight. Let us study an important approximation method in quantum mechanics called the WKB method.[100] The dynamics of a particle in quantum mechanics is governed by the equation

$$\psi(t) = \mathrm{e}^{-\mathrm{i}H(t-s)/\hbar}\psi(s), \qquad t \geq s.$$

The quantum particle behaves approximately like a classical particle if Planck's quantum of action is small, $\hbar \to 0$. More precisely, we have to assume that the dimensionless quotient $S/\hbar$ is large where S is the action (energy times $t-s$). The WKB method investigates the semi-classical approximation of quantum processes with respect to the limit

$$\hbar \to 0.$$

The two key formulas for the motion of quantum particles in the 3-dimensional Euclidean space read as follows:

(K) Time evolution of Schrödinger's wave function:

$$\psi(\mathbf{x},t) = \int_{\mathbb{R}^3} \mathcal{K}(\mathbf{x},t;\mathbf{y},s)\psi(\mathbf{y},s)d^3y, \qquad \mathbf{x} \in \mathbb{R}^3, \quad t > s.$$

We assume that the function $\mathbf{y} \mapsto \psi(\mathbf{y},s)$ is smooth with compact support (at the initial time s).

---

[98] C. Morette, On the definition and approximation of Feynman's path integral, Phys. Rev. **81** (1951), 848–852.

[99] This slightly modified quotation is taken from C. Grosche and F. Steiner, Handbook of Feynman Path Integrals, Springer, Berlin, 1998 (reprinted with permission).

[100] The three letters 'WKB' refer to the physicists 'Wentzel, Kramers, and Brioullin'. The basic papers are quoted on page 484.

Fig. 7.1. Classical trajectories

(A) Approximation of the propagator kernel as $\hbar \to 0$:

$$\boxed{\mathcal{K}(\mathbf{x},t;\mathbf{y},s) = \mathrm{e}^{\mathrm{i}S[\mathbf{q}]/\hbar}\ \frac{\mathrm{e}^{-3\mathrm{i}\pi/4}\ \mathrm{e}^{-\mathrm{i}\pi\mu(s,t)/2}}{h^{3/2}|\det J(t)|^{1/2}}\ (1+O(\hbar)).} \tag{7.216}$$

Here, $S[\mathbf{q}]$ is the action of the classical trajectory $\mathbf{q} = \mathbf{q}(\tau)$ which connects the point $\mathbf{y}$ at the initial time s with the point $\mathbf{x}$ at the final time t (Fig. 7.1(a)).[101] Furthermore, $\mu(s,t)$ denotes the Morse index (or Maslov index) of the trajectory $\mathbf{q} = \mathbf{q}(\tau)$ on the time interval $[s,t]$. Roughly speaking, the Morse index measures the number and the structure of the focal points on the trajectory. The use of the Morse index allows us to obtain a global formula for large times. As a rule, the Morse index jumps at focal points of the trajectory. Now let us discuss this more precisely.[102]

Classical particle. We start with the Newtonian equation of motion

$$m\ddot{\mathbf{q}}(\tau) = -U'(\mathbf{q}), \qquad s \le \tau \le t \tag{7.217}$$

for the trajectory

$$C : \mathbf{q} = \mathbf{q}(\tau), \qquad s \le \tau \le t$$

of a classical particle of mass m in the 3-dimensional Euclidean space. The potential $U = U(\mathbf{q})$ is assumed to be a smooth real-valued function. The action along the trajectory C is given by

$$\boxed{S[\mathbf{q}] := \int_s^t (\tfrac{1}{2}m\dot{\mathbf{q}}(\tau)^2 - U(\mathbf{q}(\tau))d\tau.}$$

For the trajectory C, we also study the corresponding Jacobi equation,

$$m\ddot{J}(\tau) + U''(\mathbf{q}(\tau))J(\tau) = 0, \qquad s \le \tau \le t,$$

along with the initial conditions $J(s) = 0$ and $\dot{J}(s) = m^{-1}I$.[103]

[101] The case where several trajectories connect the point $\mathbf{y}$ with the point $\mathbf{x}$ will be considered in (7.218) below. This corresponds to Fig. 7.1(b).

[102] The WKB method is always used in physics if a typical physical parameter goes to zero. For example, this concerns the following limits: $T \to 0$ (low temperature), $\lambda \to 0$ (short wavelength), $1/c \to 0$ (low velocity), $\nu \to 0$ (low viscosity). In terms of mathematics, the WKB method is part of singular perturbation theory.

[103] Explicitly, for the real symmetric (3×3)-matrix $J = (J_{kl})$, we get

$$m\ddot{J}_{kl}(\tau) + \sum_{r=1}^{3} \frac{\partial^2 U}{\partial x_k \partial x_r}(\mathbf{q}(\tau))J_{rl}(\tau) = 0, \qquad k,l = 1,2,3.$$

Morse index. By definition, the Morse index of the trajectory C is equal to the number of negative eigenvalues λ of the Jacobi eigenvalue problem

$$-m\ddot{\mathbf{h}}(\tau) - U''(\mathbf{q}(\tau))\mathbf{h}(\tau) = \lambda \mathbf{h}(\tau), \qquad s \le \tau \le t$$

along with the boundary condition $\mathbf{h}(s) = \mathbf{h}(t) = 0$.

Quantum particle. The Schrödinger equation for the corresponding quantum particle reads as

$$\mathrm{i}\hbar\psi_t(\mathbf{x}, t) = -\frac{\hbar^2}{2m}\,\Delta\psi(\mathbf{x}, t) + U(\mathbf{x})\psi(\mathbf{x}, t).$$

Semi-classical approximation. The approximation formula (7.216) is valid under the following assumptions.[104]

(H1) Uniqueness: There exists a unique solution $\mathbf{q} = \mathbf{q}(\tau)$, $s \le \tau \le t$, of the classical equation of motion (7.217) which satisfies the boundary condition

$$\mathbf{q}(s) = \mathbf{y}, \qquad \mathbf{q}(t) = \mathbf{x}$$

for given $\mathbf{y}, t, \mathbf{x}, s$ (Fig. 7.1(a) on page 581).

(H2) Regularity of the trajectory: At the final time t, the matrix $J(t)$ is invertible. Here, $\tau \mapsto J(\tau)$ is the solution of the Jacobi equation with respect to the trajectory from (H1).

Modifications. Replace (H1) by the assumption that the boundary value problem has not a unique solution, but at most a finite number of trajectories $\mathbf{q} = \mathbf{q}_n(\tau)$, $n = 1, \dots, N$ (Fig. 7.1(b) on page 581). In addition, assume that all of these trajectories are regular in the sense of (H2). Then, the formula (7.216) has to be replaced by the following sum formula as $\hbar \to 0$:

$$\boxed{\mathcal{K}(\mathbf{x}, t; \mathbf{y}, s) = \sum_{n=1}^{N} \mathrm{e}^{\mathrm{i}S[\mathbf{q}_n]/\hbar}\, \frac{\mathrm{e}^{-d\mathrm{i}\pi/4}\, \mathrm{e}^{-\mathrm{i}\pi\mu_n(s,t)/2}}{\hbar^{d/2}|\det J_n(t)|^{1/2}}\, (1 + O(\hbar))} \tag{7.218}$$

with $d = 3$. For motions of the particles on the real line and in the Euclidean plane, we have to choose $d = 1$ and $d = 2$, respectively. The formula (7.218) is precise (i.e., $O(\hbar) = 0$) if the potential U is a quadratic function.

Small time intervals. If the time interval $[s, t]$ is sufficiently small, then it follows from

$$J_n(t) = \frac{t - s}{m}\, I + O((t - s)^2)$$

that $\det J_n(t) \neq 1$. Moreover, $\mu_n(s, t) = 0$. This simplifies the key formula (7.218).

The quantized harmonic oscillator. To get insight, let us consider the equation of motion

$$\ddot{q}(\tau) = -\omega^2 q(\tau), \qquad 0 \le \tau \le t$$

for a classical harmonic oscillator on the real line. Here, $t > 0$. Add the boundary condition[105] $q(0) = y, q(t) = x$. This problem has the unique solution

[104] A sketch of the proof based on the path integral can be found in C. Grosche and F. Steiner, Handbook of Feynman Path Integrals, Sect. 5.2, Springer, Berlin, 1998. For the full proof embedded into a general setting, see the monograph by V. Guillemin and S. Sternberg, Geometric Asymptotics, Sect. II.7, Amer. Math. Soc., Providence, Rhode Island, 1989.

[105] To simplify notation, we set $s = 0$.

$$q(\tau) = y\cos\omega\tau + (x - y\cos\omega t)\frac{\sin\omega\tau}{\sin\omega t}$$

if the given time t is different from the critical time points $t_{n,\mathrm{crit}} := n\pi/\omega$ with $n = 1, 2, ...$ This yields the action

$$S[q] = \int_0^t (\tfrac{1}{2}m\dot{q}(\tau)^2 - \tfrac{1}{2}m\omega^2 q(\tau)^2)d\tau = \frac{(x^2+y^2)\cos\omega t - 2xy}{2x_0^2\sin\omega t}.$$

The Jacobi equation reads as

$$\ddot{J}(\tau) + \omega^2 J(\tau) = 0, \qquad 0 \le \tau \le t, \qquad J(0) = 0,\ \dot{J}(0) = \frac{1}{m}.$$

Hence

$$J(t) = \frac{\sin\omega t}{m}.$$

If $t \neq t_{n,\mathrm{crit}}$, then $J(t) \neq 0$. To compute the Morse index, consider the Jacobi eigenvalue problem

$$-\ddot{h}(\tau) - \omega^2 h(\tau) = \lambda h(\tau), \qquad 0 \le \tau \le t, \quad h(0) = h(t) = 0.$$

If $0 < t\omega < \pi$, then there is no negative eigenvalue. Hence $\mu(0,t) = 0$. However, if $n\pi < t\omega < (n+1)\pi$ with $n = 1, 2, ...$ then there are precisely n negative eigenvalues,

$$\lambda_k = \frac{k^2\pi^2}{t^2} - \omega^2, \qquad k = 1, \ldots, n$$

along with the eigenfunctions $q = \sin\tau\sqrt{\lambda_k + \omega^2}$, $k = 1, ..., n$. This way, for the harmonic oscillator, formula (7.218) reads as

$$\boxed{\mathcal{K}(x,t;y,0) = \frac{\mathrm{e}^{-\mathrm{i}\pi/4}\,\mathrm{e}^{-\mathrm{i}\pi n/2}}{x_0\sqrt{2\pi|\sin\omega t|}}\ \exp\left(\mathrm{i}\frac{(x^2+y^2)\cos\omega t - 2xy}{2x_0^2\sin\omega t}\right)} \tag{7.219}$$

for all times t with $n\pi < t\omega < (n+1)\pi$, $n = 0, 1, 2, ...$ Here, we introduce the characteristic length $x_0 := \sqrt{\hbar/m\omega}$. This is a precise formula for $\mathcal{K}$; it coincides with formula (7.144) on page 537.

The freely moving quantum particle on the real line. Let $t > 0$. We start with the classical equation of motion

$$\ddot{q}(\tau) = 0, \quad 0 \le \tau \le t.$$

Adding the boundary condition $q(0) = y, q(t) = x$, we get the unique solution $q(\tau) = y + \tau(x-y)/t$. This yields the classical action

$$S[q] = \int_0^t \tfrac{1}{2}m\dot{q}(\tau)^2 d\tau = \frac{m(x-y)^2}{2t}.$$

The Jacobi equation

$$\ddot{J}(\tau) = 0, \qquad 0 \le \tau \le t, \qquad J(0) = 0,\ \dot{J}(0) = \frac{1}{m}$$

yields $J(t) = t/m$. The Jacobi eigenvalue problem

$$-\ddot{h}(\tau) = \lambda h(\tau), \qquad 0 \le \tau \le t, \quad h(0) = h(t) = 0$$

has no negative eigenvalues. Hence $\mu(0,t) = 0$. By (7.218) with $d = 1$, we obtain

$$\mathcal{K}(x,t;y,0) = \mathrm{e}^{-\mathrm{i}\pi/4} \cdot \sqrt{\frac{m}{2\pi\hbar t}}\ \mathrm{e}^{\mathrm{i}m(x-y)^2/2\hbar t}.$$

This coincides with the Feynman propagator kernel (7.157) on page 550.

7.11 Brownian Motion

In order to understand the beauty of Feynman's approach to quantum mechanics, one has to understand the Brownian motion of immersed particles and its relation to diffusion processes.

Folklore

7.11.1 The Macroscopic Diffusion Law

We want to consider the diffusion of particles of mass $m > 0$ on the real line. Let $\varrho(x,t) > 0$ denote the mass density of the particles at the position x at time t. Then the basic diffusion equation reads as

$$\boxed{\varrho_t(x,t) = \kappa \varrho_{xx}(x,t), \qquad x \in \mathbb{R},\ t \in \mathbb{R}.} \tag{7.220}$$

Here, the positive number κ is called the diffusion coefficient. Let us motivate this.

Conservation of mass. Let $\mathbf{v}(x,t) = v(x,t)\mathbf{i}$ denote the velocity vector of the particles at the point x at time t. Here, the unit vector $\mathbf{i}$ points in direction of the positive x-axis. Furthermore, we introduce the mass current density vector

$$\mathbf{J}(x,t) := \varrho(x,t)\mathbf{v}(x,t).$$

We have $\mathbf{J}(x,t) = J(x,t)\mathbf{i}$ where

$$J(x,t) = \lim_{\Delta t \to 0} \frac{M(x;t,t+\Delta t)}{\Delta t}.$$

Here, $M(x;t,t+\Delta t)$ is the mass which flows through the point x from left to right during the time interval $[t,t+\Delta t]$. Conservation of mass tells us that the change of mass on the compact interval $[a,b]$ during the time interval $[t,t+\Delta t]$ is equal to the mass which flows through the boundary points during the time interval $[t,t+\Delta t]$. Explicitly, for small Δt, we obtain

$$\int_a^b (\varrho(x,t+\Delta t) - \varrho(x,t))dt = J(a,t)\Delta t - J(b,t)\Delta t,$$

up to terms of order $o(\Delta t)$ as $\Delta t \to 0$. Letting $\Delta t \to 0$, we get

$$\int_a^b \varrho_t(x,t)dx = J(a,t) - J(b,t) = -\int_a^b J_x(x,t)dx.$$

Contracting the interval $[a,b]$ to the point x, we obtain

$$\varrho_t(x,t) = -J_x(x,t). \tag{7.221}$$

Fick's empirical diffusion law. Motivated by physical experiments, we assume that

$$J(x,t) = -\kappa \varrho_x(x,t).$$

That is, the mass current density is proportional to the (negative) spatial derivative of the mass density. By (7.221), we get the diffusion equation (7.220).

In the three-dimensional case, the one-dimensional diffusion equation (7.220) passes over to the three-dimensional diffusion equation

$$\varrho_t(\mathbf{x},t) = -\kappa \Delta \varrho(\mathbf{x},t) \tag{7.222}$$

with the position vector $\mathbf{x}$ and time t. Furthermore, $\Delta\varrho = -\varrho_{xx} - \varrho_{yy} - \varrho_{zz}$.

7.11.2 Einstein's Key Formulas for the Brownian Motion

We are going to consider the three-dimensional motion of particles of mass $m > 0$ suspended in a resting fluid. We assume that the suspended particles have a much greater mass than the molecules of the ambient fluid. The irregular motion of the suspended particles is caused by a large number of collisions with the molecules of the ambient fluid. In 1828 the botanist Robert Brown (1773–1858) observed first such an irregular motion under the microscope, which is called Brownian motion nowadays. In his famous 1905 paper, the young Einstein (1879–1955) derived the following two key formulas for the random Brownian motion.[106]

(i) Fluctuation of the position vector $\mathbf{x}$ of a single suspended particle:

$$(\Delta \mathbf{x})^2 = 6\kappa t. \tag{7.223}$$

(ii) The Stokes–Einstein relation between the diffusion coefficient D of the suspended particles and the viscosity η of the ambient fluid:

$$\kappa = \frac{kT}{6\pi\eta r}. \tag{7.224}$$

Here, T is the absolute temperature, k is the Boltzmann constant, and r is the radius of the suspended particles.

The physical motivation of the Einstein formulas can be found in Chap. 4 of the monograph by R. Mazo, Brownian Motion: Fluctuations, Dynamics, and Applications, Oxford University Press, 2002.

7.11.3 The Random Walk of Particles

The random model. We want to investigate the random walk of a particle on the real line. To this end, we set

$$x_j := j\Delta x, \quad j = 0, \pm 1, \pm 2, \dots \quad \text{and} \quad t_k := k\Delta t, \quad k = 0, 1, 2, \dots$$

We define

$$P(x_j, t_k) := \text{probability of finding the particle at the point } x_j \text{ at time } t_k.$$

We assume the following.

(i) The initial condition: The particle is at the origin $x_0 = 0$ at the initial time $t_0 = 0$. That is, $P(0, 0) = 1$. Moreover, $P(x_j, t_k) = 0$ if $x_j \neq 0$ or $t_k > 0$.

(ii) The transition condition: Suppose that the particle is at the point x_j at time t_k. Then it will be at the point x_{j+1} (resp. x_{j-1}) at time t_{k+1} with probability $\frac{1}{2}$. Applying this to the motion from x_{j-1} to x_j and from x_{j+1} to x_j, we obtain that, for all j, k,

$$P(x_j, t_{k+1}) = \tfrac{1}{2}P(x_{j-1}, t_k) + \tfrac{1}{2}P(x_{j+1}, t_k). \tag{7.225}$$

[106] A. Einstein, Die von der molekular-kinetischen Theorie der Wärme geforderte Behandlung von in ruhenden Flüssigkeiten suspendierten Teilchen (On the motion of suspended particles in a resting fluid by using the methods of molecular kinetics), Ann. Phys. **17** (1905), 549–560 (in German). English translation: J. Stachel (Ed.), Einstein's Miraculous Year 1905: Five Papers that Changed the Universe, Princeton University Press, 1998.

The probability for the particle position. Set $p(x,t) := \frac{\mathrm{e}^{-x^2/4\kappa t}}{\sqrt{4\pi\kappa t}}$. We claim that the number

$$\int_a^b p(x,t)dx \tag{7.226}$$

equals the probability of finding the particle in the interval $[a,b]$ at time t.

Motivation. In order to motivate (7.226), let us introduce the (discrete) probability density

$$p(x_j,t_k) := \frac{P(x_j,t_k)}{\Delta x}.$$

Then the number

$$\sum_{j=0}^{j_b} p(x_j,t_k)\Delta x$$

equals the probability of finding the particle in the interval $[0,b]$ at time t_k. Here, we choose $j_b := b/\Delta x$. By (7.225),

$$P(x_j,t_{k+1}) - P(x_j,t_k) = \tfrac{1}{2}(P(x_{j+1},t_k) - 2P(x_j,t_k) + P(x_{j-1},t_k)).$$

This implies

$$p(x,t+\Delta t) - p(x,t) = \tfrac{1}{2}(p(x+\Delta x,t) - 2p(x,t) + p(x-\Delta x,t)).$$

Hence

$$\frac{p(x,t+\Delta t) - p(x,t)}{\Delta t} = \frac{p(x+\Delta x,t) - 2p(x,t) + p(x-\Delta x,t)}{(\Delta x)^2} \cdot \frac{(\Delta x)^2}{2\Delta t}.$$

Letting $\Delta x \to 0$ and $\Delta t \to 0$ such that the quotient $(\Delta x)^2/2\Delta t$ goes to the positive number κ, then

$$p_t(x,t) = \kappa p_{xx}(x,t), \quad x \in \mathbb{R},\ t > 0. \tag{7.227}$$

In addition, we obtain the formal initial condition $p(x,0) = \delta(x)$ for all points $x \in \mathbb{R}$.[107] By the study of the diffusion equation on page 487, the solution of (7.227) reads as $p(x,t) = \frac{\mathrm{e}^{-x^2/4\kappa t}}{\sqrt{4\pi\kappa t}}$.

7.11.4 The Rigorous Wiener Path Integral

Probabilities of a continuous random walk. Let us consider the random walk of a particle on the real line with diffusion coefficient $\kappa > 0$. Choose the function

$$p(x,t) := \frac{\mathrm{e}^{-x^2/4\kappa t}}{\sqrt{4\pi\kappa t}},$$

and choose the points of time $0 < t_1 < \ldots < t_N := T$. Suppose that the particle is at the point $x_0 := 0$ at time $t_0 := 0$.

- The real number $\int_{J_1} p(x_1 - x_0, t_1 - t_0)dx_1$ is the probability of finding the particle on the interval J_1 at time t_1.

[107] This follows from the discrete initial condition $p(x_j,0) = \frac{P(x_j,0)}{\Delta x} = \frac{\delta_{j0}}{\Delta x}$ by letting $\Delta x \to 0$.

- The real number $\int_{J_1}\int_{J_2} p(x_1-x_0,t_1-t_0)p(x_2-x_1,t_2-t_1)dx_1dx_2$ is the probability of finding the particle on the interval J_1 and J_2 at time t_1 and t_2, respectively.
- The real number

$$\int_{J_1}\dots\int_{J_N}\prod_{j=1}^{N} p(x_j-x_{j-1},t_j-t_{j-1})\,dx_1\cdots dx_N \tag{7.228}$$

  is the probability of finding the particle on the interval $J_1,\dots,J_N$ at time $t_1,\dots,t_N$, respectively.

The Wiener measure. We want to translate the preceding probabilities into the language of measure theory. Fix the time $T>0$. By definition, the function space $\mathcal{C}_0[0,T]$ consists of all continuous functions

$$q:[0,T]\to\mathbb{R}$$

with $q(0)=0$. Intuitively, $x=q(t)$, $0\le t\le T$, describes the trajectory of a Brownian particle on the real line. We want to construct a measure W on the space $\mathcal{C}_0[0,T]$ of trajectories such that, for each measurable subset Ω of $\mathcal{C}_0[0,T]$, the real number

$$W(\Omega)$$

equals the probability of finding the trajectory $q\in\mathcal{C}_0[0,T]$ in the set Ω. We will proceed in two steps.

Step 1: Pre-measure on cylindrical subsets. Let

$$\Omega_{\text{cyl}}:=\{q\in\mathcal{C}_0[0,T]: q(t_k)\in J_k,\ k=1,\dots,N\}$$

where $0<t_1<\dots<t_N:=T$, $N=1,2,\dots$, and $J_1,\dots J_N$ are intervals on the real line. We define the number $W(\Omega_{\text{cyl}})$ by (7.228). This number is called the Wiener pre-measure of the cylindrical set Ω_{cyl}.

Step 2: Extension of the pre-measure to the Wiener measure. The Wiener pre-measure on cylindrical sets can be extended to a measure on the function space $\mathcal{C}_0[0,T]$. This measure (called the Wiener measure) is uniquely determined on the smallest σ-algebra of $\mathcal{C}_0[0,T]$ which contains all the cylindrical sets. For general measure theory and measure integrals, see Sec. 10.2.1 of Vol. I. Furthermore, we refer to:

> H. Amann and J. Escher, Analysis, Vol. 3, Birkhäuser, Basel, 2001 (in German). (English edition in preparation.)
>
> E. Stein and R. Shakarchi, Princeton Lectures in Analysis, Vol. III: Measure Theory, Princeton University Press, 2003.

A detailed summary can be found in the Appendix to Zeidler (1986), Vol. IIB (see the references on page 1049).

Example. If $\mathcal{C}_0^1[0,T]$ denotes the set of all continuously differentiable functions $q:[0,T]\to\mathbb{R}$ with $q(0)=0$, then

$$W(\mathcal{C}_0^1[0,T])=0.$$

This tells us that the trajectory of a Brownian particle is continuously differentiable with probability zero. In fact, under the microscope one observes zigzag trajectories of Brownian motion.

The Wiener path integral. General measure theory tells us that the Wiener measure W on the function space $\mathcal{C}_0[0,T]$ induces the measure integral

$$\int_{\mathcal{C}_0[0,T]} F(q)\, dW(q)$$

for appropriate functions $F : \mathcal{C}_0[0,T] \to \mathbb{R}$. This integral is called the Wiener path integral. Here, we integrate over a set of trajectories. In particular, we have

$$\int_{\mathcal{C}_0[0,T]} F(q)\, dW(q) = \sum_{n=1}^{N} F_n W(\Omega_n)$$

if $\Omega_1, \ldots, \Omega_N$ is a collection of pairwise disjoint cylindrical sets of the function space $\mathcal{C}_0[0,T]$, and the real-valued function F has the constant values $F_1, \ldots, F_N$ on $\Omega_1, \ldots, \Omega_N$, respectively, and it vanishes outside these sets. If Ω is a measurable subset of the function space $\mathcal{C}_0[0,T]$ (e.g., a cylindrical set), then the Wiener measure of Ω is given by

$$W(\Omega) = \int_{\Omega} dW = \int_{\mathcal{C}_0[0,T]} \chi(q)\, dW(q)$$

where $\chi(q) := 1$ for all $q \in \Omega$ and $\chi(q) := 0$ for all $q \notin \Omega$.

7.11.5 The Feynman–Kac Formula

In 1947, Marc Kac (1914–1984) attended a lecture given by the young Richard Feynman (1918–1988) at Cornell University. He was amazed about the fact that Feynman's formula related the quantum mechanical propagator to classical mechanics in a very elegant way. He also noticed that Feynman's idea of the path integral was close to his own ideas about stochastic processes based on the Wiener integral due to Norbert Wiener (1894–1964). A few days later Kac rigorously proved a formula which is known nowadays as the Feynman–Kac formula. In his autobiography *Enigmas of Chance*, Harper & Row, New York, 1985, Marc Kac writes:

> It is only fair to say that I had Wiener's shoulders to stand on. Feynman as in everything else he has done, stood on its own, a trick of intellectual contortion that he alone is capable of.

In order to discuss the Feynman–Kac formula, let us consider the one-dimensional diffusion equation

$$\varrho_t(x,t) = \kappa \varrho_{xx}(x,t) - U(x)\varrho(x,t), \quad x \in \mathbb{R},\ t > 0 \tag{7.229}$$

with the initial condition $\varrho(x,0) = \varrho_0(x)$ for all $x \in \mathbb{R}$. We are given the positive diffusion constant κ, the real-valued potential $U \in C_0^\infty(\mathbb{R})$, and the real-valued initial mass density $\varrho_0 \in C_0^\infty(\mathbb{R})$. Define

$$\mathcal{H}\varrho := -\kappa \varrho_{xx} + U\varrho \qquad \text{for all} \quad \varrho \in C_0^\infty(\mathbb{R}).$$

The operator $\mathcal{H} : C_0^\infty(\mathbb{R}) \to L_2(\mathbb{R})$ can be uniquely extended to a self-adjoint operator $H : D(H) \to L_2(\mathbb{R})$ on the real Hilbert space $L_2(\mathbb{R})$. In terms of functional analysis, the solution of (7.229), that is, $\varrho_t = -H\varrho$, reads as

$$\varrho(t) = \mathrm{e}^{-tH}\varrho_0, \qquad t > 0. \tag{7.230}$$

The famous Feynman–Kac formula tells us the following.

Theorem 7.51 *For all times $T > 0$ and all positions $x \in \mathbb{R}$, the solution* (7.230) *of the diffusion equation* (7.229) *is given by*

$$\varrho(x,T) = \int_{\mathcal{C}_0[0,T]} \varrho_0(x + q(t))\, \mathrm{e}^{-\int_0^T U(x+q(t))dt}\, dW(q).$$

Intuitively, this is a statistics over all possible continuous trajectories of a particle which starts at the point x at time $t = 0$. The statistical weight is related to both the Wiener measure and an exponential function which depends on the potential U. The proof can be found in:

G. Johnson and M. Lapidus, The Feynman Integral and Feynman's Operational Calculus, Chap. 12, Clarendon Press, Oxford, 2000.

We also refer to:

M. Reed and B. Simon, Methods of Modern Mathematical Physics II: Fourier Analysis, Self-Adjointness, Academic Press, New York, 1975.

B. Simon, Functional Integration and Quantum Physics, Academic Press, New York, 1979.

In terms of the limit of classical N-dimensional integrals, the solution $\varrho(t) = \mathrm{e}^{-tH}\varrho_0$ of the diffusion equation (7.229) can be represented as

$$\varrho(x,T) = \lim_{N\to\infty} \left(\sqrt{\frac{1}{4\pi\kappa\Delta t}}\right)^N PV \int_{-\infty}^{\infty} \dots PV \int_{-\infty}^{\infty} \varrho_0(q_N) \mathrm{e}^{S_N^-} dq_1 \dots dq_N \tag{7.231}$$

with $S_N^- := \sum_{j=0}^{N-1} -\frac{m}{2}\left(\frac{q_{j+1}-q_j}{\Delta t}\right)^2 - U(q_j)$, as well as $\Delta t := T/N$, $\kappa = 1/2m$, and $q_0 := x$.

Corollary 7.52 *For all times $T > 0$ and all positions $x \in \mathbb{R}$, we have* (7.231).

Note that the principal value $PV \int_{-\infty}^{\infty} \dots$ means $\lim_{r\to\infty} \int_{-r}^{r} \dots$, and the limit $N \to \infty$ refers to the convergence on the real Hilbert space $L_2(\mathbb{R})$. The proof based on the Trotter product formula (see Sect. 8.3 of Vol. I) can be found in Reed and Simon (1975), Vol. II, Sect. X.11, quoted above.

The passage to the Schrödinger equation. We replace the diffusion equation (7.229) by the Schrödinger equation

$$\mathrm{i}\hbar\varrho_t(x,t) = \mathcal{H}\varrho(x,t), \qquad x \in \mathbb{R},\, t > 0 \tag{7.232}$$

with the initial condition $\varrho(x,0) = \varrho_0(x)$ for all $x \in \mathbb{R}$. Here, we use the differential operator $\mathcal{H}\varrho := -\kappa\varrho_{xx} + U\varrho$ with $\kappa := \frac{\hbar^2}{2m}$. In terms of the limit of classical N-dimensional integrals, the solution $\varrho(t) = \mathrm{e}^{-\mathrm{i}tH/\hbar}\varrho_0$ of the Schrödinger equation (7.232) can be represented as follows: the function $\varrho(x,T)$ at the point x at time T is equal to the limit

$$\lim_{N\to\infty} \left(\sqrt{\frac{\hbar}{4\pi\mathrm{i}\kappa\Delta t}}\right)^N PV \int_{-\infty}^{\infty} \dots PV \int_{-\infty}^{\infty} \varrho_0(q_N) \mathrm{e}^{\mathrm{i}S_N/\hbar} dq_1 \dots dq_N \tag{7.233}$$

with the discrete action

$$S_N := \sum_{j=0}^{N-1} \frac{m}{2} \left(\frac{q_{j+1} - q_j}{\Delta t} \right)^2 - U(q_j),$$

as well as $\Delta t := T/N$ and $q_0 := x$. The square root is to be understood as principal value.

Corollary 7.53 *For all times $T > 0$ and all positions $x \in \mathbb{R}$, we have* (7.233).

Here, the limits are to be understood as in Corollary 7.52. Naturally enough, formula (7.233) is obtained from (7.231) by rescaling. The proof of Corollary 7.53 can be found in Reed and Simon (1975), Vol. II, Sect. X.11, quoted on page 589.

Unfortunately, the Feynman–Kac formula from Theorem 7.51 cannot be rigorously extended to the Schrödinger equation, since the corresponding complex-valued measure does not exist. This is the statement of the famous Cameron non-existence theorem which can be found in Johnson and Lapidus (2000), Sect. 4.6, quoted on page 589.[108]

7.12 Weyl Quantization

> The use of the Moyal product for smooth functions avoids the use of Hilbert-space operators in quantum mechanics.
>
> Folklore

> We can say that quantum mechanics is a deformation of classical mechanics. The Planck constant h is the corresponding deformation parameter. This is for me the most concise formulation of the correspondence principle and explains what is meant by quantization.
> Beautiful results, which I learned from A. Lichnerowicz, M. Flato, and D. Sternheimer, allow one to say that classical mechanics is unstable and that quantum mechanics is essentially a unique deformation of it into a nonequivalent stable structure.[109]
>
> Ludwig Faddeev, 1999

[108] R. Cameron, A family of integrals serving to connect the Wiener and Feynman integrals. J. of Math. and Phys. Sci. of MIT **39** (1960), 126–140.

[109] L. Faddeev, Elementary introduction to quantum field theory, Vol. 1, pp. 513–552. In: P. Deligne, P. Etingof, D. Freed, L. Jeffrey, D. Kazhdan, J. Morgan, D. Morrison, and E. Witten (Eds.), Lectures on Quantum Field Theory: A course for mathematicians given at the Institute for Advanced Study in Princeton in 1996/97, Vols. 1, 2, Amer. Math. Soc., Providence, Rhode Island, 1999 (reprinted with permission).
We also refer to the beautiful book by L. Faddeev and A. Slavnov, Gauge Fields, Benjamin, Reading, Massachusetts, 1980. This book is based on the use of Feynman functional integrals; it represents the Faddeev–Popov approach to gauge theory which was a breakthrough in the quantization of the Standard Model in particle physics. See L. Faddeev and V. Popov, Feynman diagrams for the Yang–Mills field, Phys. Lett. **25B** (1967), 29–30.
Ludwig Faddeev made seminal contributions to mathematical physics. This is described in the book by L. Faddeev, 40 Years in Mathematical Physics, World Scientific, Singapore, 1995.
F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer, Deformation theory and quantization I, II, Annals of Physics **111** (1978), 61–110; 111–151.

The elegant method of deformation quantization is based on the use of classical smooth functions equipped with the Moyal star product. This star product represents a deformation of the classical product of functions. The deformation depends on the Planck constant h. The first quantum correction of the classical product is related to the Poisson bracket in classical mechanics. The relation between deformation quantization and the operator-theoretic approach to quantum mechanics in Hilbert spaces is given by the Weyl calculus.

In the following sections, we will only sketch the basic ideas. We will start with the formal language used by physicists. From the mnemonic point of view, the language of physicists is very convenient. Unfortunately, rigorous mathematical arguments are more involved. The rigorous Weyl calculus will be considered in Sect. 7.12.6; this represents a special case of the modern theory of pseudo-differential operators, which combines differential operators with integral operators in the setting of generalized functions. We would like to encourage the reader to learn both the language of physicists and the language of mathematicians.

7.12.1 The Formal Moyal Star Product

Let $C^\infty(\mathbb{R}^2)$ be the space of smooth functions $f : \mathbb{R}^2 \to \mathbb{C}$. For $f, g \in C^\infty(\mathbb{R}^2)$, the formal Moyal star product is defined by

$$\boxed{f * g := f \mathrm{e}^{\frac{\mathrm{i}\hbar}{2}(\partial'_q \partial_p - \partial'_p \partial_q)} g.}$$

Here, the functions f and g depend on the real variables q and p, and we set $\partial_q := \partial/\partial q$ and $\partial_p := \partial/\partial p$. In addition, the prime of ∂'_q indicates that the partial derivative acts on the left factor f. Explicitly,

$$f * g = \sum_{m,n=0}^{\infty} \left(\frac{\mathrm{i}\hbar}{2}\right)^{m+n} \frac{(-1)^m}{m!n!} (\partial_p^m \partial_q^n f)(\partial_p^n \partial_q^m g). \tag{7.234}$$

This is to be understood as a formal power series with respect to the variable $\hbar$. The Moyal star product has the following properties.

(i) The correspondence principle: For all $f, g \in C^\infty(\mathbb{R}^2)$,

$$\boxed{f * g = fg + \frac{\mathrm{i}\hbar}{2}\{f, g\} + O(\hbar^2), \qquad \hbar \to 0.}$$

Here, we use the Poisson bracket $\{f, g\} := f_q g_p - g_q f_p$. Hence

$$f * g - g * f = \mathrm{i}\hbar\{f, g\} + O(\hbar^2), \qquad \hbar \to 0.$$

Therefore, the star product $f * g$ represents a deformation of the classical product fg. This deformation depends on the Planck constant $\hbar$. In terms of physics, the difference $f * g - fg$ describes quantum fluctuations which depend on $\hbar$. For example, if we choose $f(q, p) := q$ and $g(q, p) := p$, then $q*p = qp + \frac{1}{2}\mathrm{i}\hbar$ and $p * q = pq - \frac{1}{2}\mathrm{i}\hbar$. Hence

$$\boxed{q * p - p * q = \mathrm{i}\hbar.}$$

This commutation rule (for the Moyal star product of classical smooth functions) corresponds to the Born–Heisenberg–Jordan commutation relation $QP - PQ = \mathrm{i}\hbar I$ (in the operator-theoretic formulation of quantum mechanics on Hilbert spaces). As we will show below, the use of the Moyal star product avoids the use of operators.

(ii) Associativity: For all $f, g, k \in C^\infty(\mathbb{R}^2)$, we have

$$(f * g) * k = f * (g * k).$$

7.12.2 Deformation Quantization of the Harmonic Oscillator

The basic equations of deformation quantization. We want to apply the method of deformation quantization to the motion of a particle on the real line. The classical trajectory $q = q(t)$ is described by the canonical equations

$$\dot{p}(t) = -H_q(q(t), p(t)), \qquad \dot{q}(t) = H_p(q(t), p(t)), \qquad t \in \mathbb{R}.$$

We are given the Hamiltonian $H \in C^\infty(\mathbb{R}^2)$.

The corresponding quantum motion is obtained by solving the following problem. We are looking for

- a nonempty index set $\mathcal{M}$,
- a measure μ on the set $\mathcal{M}$,
- functions $\varrho_m = \varrho_m(q, p)$ on the phase space $\mathbb{R}^2$ for each index $m \in \mathcal{M}$, and
- real values E_m for each index $m \in \mathcal{M}$

such that the following equations hold.

(E) Quantized energy levels E_m:

$$\boxed{H * \varrho_m = E_m \varrho_m \qquad \text{for all} \quad m \in \mathcal{M}.}$$

(D) Distribution function ϱ_m: For all indices $m, n \in \mathcal{M}$, we have the orthogonality relation

$$\varrho_m * \varrho_n = 0, \qquad m \neq n,$$

along with the idempotent law

$$\varrho_m * \varrho_m = \varrho_m,$$

and the normalization relation on the phase space,

$$\int_{\mathbb{R}^2} \varrho_m(q, p) \frac{dqdp}{h} = 1.$$

(Q) Quantized energy decomposition of the classical Hamiltonian function:[110]

$$H(q, p) = \int_{\mathcal{M}} E_m \varrho_m(q, p) d\mu(m) \qquad \text{for all} \quad q, p \in \mathbb{R}.$$

(M) Mean value of energy: For all $m \in \mathcal{M}$,

$$\boxed{E_m = \int_{\mathbb{R}^2} H(q, p) \varrho_m(q, p) \, \frac{dqdp}{h}.}$$

In terms of physics, this means that each of the functions $\varrho_m = \varrho_m(q, p)$ is a probability distribution on the phase space which has the quantized energy level E_m as energy mean value.

[110] In the special case where $\mathcal{M} := \{0, 1, , 2 \ldots\}$, the integral $\int_{\mathcal{M}} E_m \varrho_m(q, p) d\mu(m)$ is equal to the infinite series $\sum_{m=0}^{\infty} E_m \varrho_m(q, p) \mu_m$. Here, the nonnegative number μ_m is the measure of the point $\{m\}$ for all $m = 0, 1, \ldots$

Suppose that we know a solution of the equations (E) through (M) above. Then, to a given complex-valued function $F : \mathbb{R} \to \mathbb{C}$ we can assign the star function F_* defined by

$$F_*(q,p) := \int_{\mathcal{M}} F(E_m)\varrho_m(q,p)d\mu(m) \qquad \text{for all} \quad q,p \in \mathbb{R}.$$

For example, we may formally define the exponential star function

$$\mathrm{Exp}_*(\alpha tH)(q,p) := \int_{m\in\mathcal{M}} \mathrm{e}^{\alpha tH(q,p)}\varrho_m(q,p)d\mu(m)$$

for all $q,p \in \mathbb{R}$, all times $t \in \mathbb{R}$, and fixed complex number α. Formally, it follows from (E) above that

$$\frac{d}{dt}\mathrm{Exp}_*\left(\alpha tH\right) = \alpha H * \mathrm{Exp}_*\left(\alpha tH\right).$$

This equation is called the Schrödinger equation in quantum deformation. In concrete models, one has to check that all of the equations formulated above possess a rigorous meaning, in the sense of well-defined formal expansions with respect to $\hbar$. Let us show how quantum deformation works for the harmonic oscillator. In this case, we choose $\mathcal{M} = \{0,1,2,,...\}$ and $\mu_m := 1$ for all m.

Application to the harmonic oscillator. The classical function

$$H(q,p) := \frac{p^2}{2m} + \frac{m\omega^2 q^2}{2}$$

is the Hamiltonian for a harmonic oscillator of mass m and angular frequency ω on the real line. To simplify the computation, it is useful to introduce the new dimensionless variable

$$a := \sqrt{\frac{m\omega}{2\hbar}}\left(q + \frac{\mathrm{i}p}{m\omega}\right) \tag{7.235}$$

and the conjugate complex variable

$$a^\dagger = \sqrt{\frac{m\omega}{2\hbar}}\left(q - \frac{\mathrm{i}p}{m\omega}\right). \tag{7.236}$$

Hence

$$\boxed{H = \hbar\omega a a^\dagger.}$$

By the chain rule, the Moyal star product reads as

$$\boxed{f * g = f\mathrm{e}^{\frac{1}{2}(\partial'_a\partial_{a^\dagger} - \partial'_{a^\dagger}\partial_a)}g}$$

with respect to the new variables a and $a^\dagger$. Here, we set $\partial_a := \partial/\partial a$, as well as $\partial_{a^\dagger} := \partial/\partial a^\dagger$, and we regard f and g as functions of the variables a and $a^\dagger$. Explicitly, we obtain[111]

[111] If one wants to see the dependence on the parameter $\hbar$, then one has to replace a by $\sqrt{\hbar}\cdot b$. This yields $b * b^\dagger - b^\dagger * b = \hbar$, and

$$f * g = \sum_{m,n=0}^{\infty}\left(\frac{\hbar}{2}\right)^{m+n}\frac{(-1)^m}{m!n!}\,(\partial_{b^\dagger}^m\partial_b^n f)(\partial_{b^\dagger}^n\partial_b^m g).$$

$$f * g = \sum_{m,n=0}^{\infty} \frac{(-1)^m}{2^{m+n} m! n!} (\partial_{a^\dagger}^m \partial_a^n f)(\partial_{a^\dagger}^n \partial_a^m g). \tag{7.237}$$

For example, $a * a^\dagger = aa^\dagger + \frac{1}{2}$ and $a^\dagger * a = aa^\dagger - \frac{1}{2}$. This implies

$$\boxed{a * a^\dagger - a^\dagger * a = 1.} \tag{7.238}$$

For $m = 1, 2, , ..,$ define

- $E_0 := \frac{1}{2}\omega\hbar, \quad \varrho_0 := 2\mathrm{e}^{-2aa^\dagger}$;
- $E_m := \omega\hbar(m + \frac{1}{2})$;
- $\varrho_m := \frac{1}{m!}(a^\dagger)^m * \varrho_0 * a^m$.

Theorem 7.54 *For all $m, n = 0, 1, , 2, ...,$ the following hold.*
(E) *Quantized energy levels: $H * \varrho_m = E_m \varrho_m$.*
(D) *Distribution functions: $\varrho_m * \varrho_n = \delta_{nm}\varrho_m$.*
(Q) *Quantized energy decomposition of the classical Hamiltonian function:*

$$H(q,p) = \sum_{m=0}^{\infty} E_m \varrho_m(q,p) \qquad \text{for all} \quad q, p \in \mathbb{R}.$$

For the proof, see Problem 7.29.

The relation to the Laguerre polynomials. For all $w, z \in \mathbb{R}$ with $|w| < 1$, the Laguerre polynomials $L_0, L_1, ...$ are generated by the function

$$\frac{1}{1+w} \exp\left(\frac{wz}{1+w}\right) = \sum_{n=0}^{\infty} (-1)^n w^n L_n(z).$$

Explicitly, for $n = 0, 1, 2, ...,$

$$L_n(z) = \frac{\mathrm{e}^z}{n!} \frac{d^n(z^n \mathrm{e}^{-z})}{dz^n} = \sum_{m=0}^{n} \frac{(-1)^m\, n!}{(n-m)!\, m!\, n!}\, z^m.$$

The functions

$$\mathcal{L}_n(x) := \mathrm{e}^{-x/2} L_n(x) \qquad x \in \mathbb{R}, \qquad n = 0, 1, 2, ...$$

form a complete orthonormal system of the Hilbert space $L_2(0, \infty)$.

Theorem 7.55 *For all $m, n = 0, 1, , 2, ...,$ the following hold.*
(L) *Laguerre polynomials:*

$$\varrho_m = 2(-1)^m e^{-2H/\hbar\omega} L_m\left(\frac{4H}{\hbar\omega}\right)$$

with the normalization condition $\int_{\mathbb{R}^2} \varrho_m(q,p)\, \frac{dqdp}{h} = 1$.
(M) *Mean value: $E_m = \int_{\mathbb{R}^2} H(q,p)\varrho_m(q,p)\frac{dqdp}{h}$.*
(S) *The Schrödinger equation*

$$i\hbar F_t(q,p,t) = H(q,p) * F(q,p,t), \qquad q, p, t \in \mathbb{R}$$

has the solution

$$F(q,p,t) = \frac{1}{\cos\frac{\omega t}{2}} \exp\left(\frac{2H}{i\hbar\omega}\ \tan\frac{\omega t}{2}\right)$$

for all $t \in \mathbb{R}$ with $\omega t \neq 2n\pi,\ n = 0, \pm 1, \pm 2, ...$

For the proof, see Problem 7.30.

Motivation for the deformation quantization of the harmonic oscillator. We want to show how the method of deformation quantization considered above is related to Schrödinger's operator-theoretic treatment of the harmonic oscillator studied on page 534. Consider the operators

$$Q_{\mathrm{pre}}, P_{\mathrm{pre}}, H_{\mathrm{pre}} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$$

with $Q_{\mathrm{pre}}\varphi(q) := q\varphi(q)$ and $P_{\mathrm{pre}}\varphi(q) = -\mathrm{i}\hbar\varphi'(q)$ for all $q \in \mathbb{R}$, as well as

$$H_{\mathrm{pre}} := \frac{P_{\mathrm{pre}}^2}{2m} + \frac{m\omega^2 Q_{\mathrm{pre}}}{2}.$$

Using the Dirac calculus, let $|\varphi_0\rangle, |\varphi_1\rangle, \ldots$ denote the complete orthonormal system of eigenvectors of the Hamiltonian H_{pre}. That is,

$$H_{\mathrm{pre}}|\varphi_m\rangle = E_m|\varphi_m\rangle, \qquad m = 0, 1, 2, \ldots$$

with $E_m := \hbar\omega(m + \frac{1}{2})$. In addition, let us introduce the operator

$$\boldsymbol{\varrho}_m := |\varphi_m\rangle\langle\varphi_m|, \qquad m = 0, 1, \ldots$$

This is the von Neumann density operator corresponding to the eigenstate $|\varphi_m\rangle$. Then, for all indices $m, n = 0, 1, \ldots$ and all times $t \in \mathbb{R}$, the following hold:[112]

(a) $H_{\mathrm{pre}}\boldsymbol{\varrho}_m = E_m\boldsymbol{\varrho}_m$;
(b) $H_{\mathrm{pre}} = \sum_{m=0}^{\infty} E_m\boldsymbol{\varrho}_m$;
(c) $\boldsymbol{\varrho}_m\boldsymbol{\varrho}_n = \delta_{mn}\boldsymbol{\varrho}_m$;
(d) $\mathrm{i}\hbar\frac{d}{dt}\mathrm{e}^{-\mathrm{i}tH_{\mathrm{pre}}/\hbar} = H_{\mathrm{pre}}\mathrm{e}^{-\mathrm{i}tH_{\mathrm{pre}}/\hbar}$.

Relation (a) follows from

$$(H_{\mathrm{pre}}\boldsymbol{\varrho}_m)|\varphi_m\rangle = H_{\mathrm{pre}}|\varphi_m\rangle\langle\varphi_m|\varphi\rangle = E_m|\varphi_m\rangle\langle\varphi_m|\varphi\rangle = E_m\boldsymbol{\varrho}_m|\varphi\rangle.$$

Relations (b) and (d) are a consequence of

$$f(H_{\mathrm{pre}})\varphi = \sum_{m=0}^{\infty} f(E_m)|\varphi_m\rangle\langle\varphi_m|\varphi\rangle$$

for all $\varphi \in \mathcal{S}(\mathbb{R})$, where $f(x) := x$ or $f(x) := \mathrm{e}^{-\mathrm{i}xt/\hbar}$ for all $x \in \mathbb{R}$. Finally, relation (c) follows from

$$|\varphi_m\rangle\langle\varphi_m|\varphi_n\rangle\langle\varphi_m|\varphi\rangle = E_m\delta_{mn}|\varphi_m\rangle\langle\varphi_m|\varphi\rangle \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).$$

This finishes the proof of (a)–(d). In the following sections, we will introduce the Weyl calculus. Here,

[112] Explicitly, condition (b) means that $H_{\mathrm{pre}}\varphi = \sum_{m=0}^{\infty} E_m\boldsymbol{\varrho}_m\varphi$ for all $\varphi \in \mathcal{S}(\mathbb{R})$, and condition (d) is a short-hand writing for the equation

$$\mathrm{i}\hbar\frac{d}{dt}\sum_{m=0}^{\infty}\mathrm{e}^{-\mathrm{i}E_m t/\hbar}\boldsymbol{\varrho}_m\varphi = \sum_{m=0}^{\infty}E_m\mathrm{e}^{-\mathrm{i}E_m t/\hbar}\boldsymbol{\varrho}_m\varphi,$$

which is valid for all $\varphi \in \mathcal{S}(\mathbb{R})$. The limits are to be understood in the sense of the convergence on the Hilbert space $L_2(\mathbb{R})$.

- operators have to be replaced by their symbols, and
- operator products have to be replaced by the Moyal star product of the corresponding symbols.

Replacing the operators $H_{\mathrm{pre}}, \boldsymbol{\varrho}_m, \mathrm{e}^{-\mathrm{i}H_{\mathrm{pre}}/\hbar}$ by their symbols H, ϱ_m, F, the formulas (a)–(d) pass over to the following formulas:

(a*) $H * \varrho_m = E_m \varrho_m$;
(b*) $H = \sum_{m=0}^{\infty} E_m \varrho_m$;
(c*) $\varrho_m \varrho_n = \delta_{mn} \varrho_m$;
(d*) $\mathrm{i}\hbar F_t = H * F$.

This corresponds to Theorems 7.54 and 7.55 above. For the annihilation operator $\mathbf{a}$ and the creation operator $\mathbf{a}^\dagger$ given by

$$\mathbf{a} := \sqrt{\frac{m\omega}{2\hbar}} \left(Q_{\mathrm{pre}} + \frac{\mathrm{i}P_{\mathrm{pre}}}{m\omega} \right) \qquad \text{and} \qquad \mathbf{a}^\dagger := \sqrt{\frac{m\omega}{2\hbar}} \left(Q_{\mathrm{pre}} - \frac{\mathrm{i}P_{\mathrm{pre}}}{m\omega} \right),$$

the symbols a and $a^\dagger$ are given by (7.235) and (7.236), respectively. The operator commutation relation $\mathbf{a}\mathbf{a}^\dagger - \mathbf{a}^\dagger\mathbf{a} = I$ corresponds to the Moyal-star-product relation $a * a^\dagger - a^\dagger * a = 1$ for the symbols in the Weyl calculus. This coincides with (7.238).

7.12.3 Weyl Ordering

> The Moyal star product of classical symbols passes over to the operator product of the corresponding Weyl operators.
>
> Folklore

As a preparation for the general Weyl calculus, let us start with the rigorous theory of Weyl polynomials. In the quantum mechanics of particles on the real line, we encounter both[113]

- the position operator $Q : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ given by $(Q\psi)(q) := q\psi(q)$ and
- the momentum operator $P : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ given by $(P\psi)(q) := -\mathrm{i}\hbar\psi'(q)$

for all $\psi \in \mathcal{S}(\mathbb{R})$ and all $q \in \mathbb{R}$. These two basic operators are formally self-adjoint on the Hilbert space $L_2(\mathbb{R})$, that is,

$$\langle Q\psi|\varphi\rangle = \langle \psi|Q\varphi\rangle \quad \text{and} \quad \langle P\psi|\varphi\rangle = \langle \psi|P\varphi\rangle \qquad \text{for all} \quad \psi, \varphi \in \mathcal{S}(\mathbb{R}).$$

Here, we use the inner product $\langle \psi|\chi\rangle := \int_{\mathbb{R}} \psi^\dagger(q)\chi(q)dq$ on $L_2(\mathbb{R})$. In other words, $Q^\dagger = Q$ and $P^\dagger = P$.[114]

Weyl polynomials with respect to the operators Q and P on the linear function space $\mathcal{S}(\mathbb{R})$. Consider an arbitrary polynomial

$$\boxed{a(q,p) := \sum_{k,m=0}^{N} c_{km} q^k p^m \qquad \text{for all} \quad q, p \in \mathbb{R}} \tag{7.239}$$

with respect to the real variables q and p. Here, the coefficients c_{km} are complex numbers. It is our goal to assign to each polynomial a a linear operator

[113] To simplify notation, we write the operator symbol Q (resp. P) instead of Q_{pre} (resp. P_{pre}).

[114] In addition, the operators $Q, P : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ are essentially self-adjoint on the Hilbert space $L_2(\mathbb{R})$.

$$A(a) : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R}),$$

which is a polynomial with respect to Q and P, such that the following properties hold.

(W1) Linearity: For all polynomials a, b and all complex numbers α, β, we get

$$A(\alpha a + \beta b) = \alpha A(a) + \beta A(b).$$

In particular, if $a(q,p) := q$ and $b(q,p) := p$, then $A(a) := Q$ and $A(b) := P$. Furthermore, $A(1) = I$ (identity operator).

(W2) Weyl ordering: If $a(q,p) := qp$, then[115]

$$A(a) = \tfrac{1}{2}(QP + PQ).$$

(W3) Formal self-adjointness: If the coefficients of the polynomial a are real, then the Weyl operator $A(a)$ is formally self-adjoint. Explicitly,

$$\langle A(a)\psi|\varphi\rangle = \langle \psi|A(a)\varphi\rangle \qquad \text{for all} \quad \psi, \varphi \in \mathcal{S}(\mathbb{R}).$$

In other words, the Weyl polynomials $A(a)$ to real polynomials a are formal observables in quantum mechanics.

(W4) Composition rule: If a and b are polynomials, then[116]

$$A(a * b) = A(a)B(b).$$

This means that the Moyal star product of polynomials is translated into the operator product of Weyl polynomials on the space $\mathcal{S}(\mathbb{R})$. This is the characteristic property of the Moyal star product.

In about 1930, it was the idea of Weyl to introduce the symmetric Weyl polynomials $(q^k p^m)_W$ by setting

- $(q^k)_W := Q^k$ and $(p^m)_W := P^m$, where $m, k = 0, 1, \ldots$;
- $(qp)_W := \frac{1}{2}(QP + PQ)$;
- $(q^2 p)_W := \frac{1}{3}(Q^2 P + PQ^2 + QPQ)$;
- $(q^2 p^2)_W := \frac{1}{6}(Q^2P^2 + P^2Q^2 + QP^2Q + PQ^2P + QPQP + PQPQ)$.

In the general case, we proceed as follows. In order to obtain $(q^k p^m)_W$, we start with the symmetrized expression

$$(A_1 A_2 \cdots A_{k+m})_{\text{sym}} := \frac{1}{(k+m)!} \sum_{\pi} A_{\pi(1)} A_{\pi(2)} \cdots A_{\pi(k+m)}$$

where we sum over all possible permutations π of $1, 2, \ldots, k+m$. Finally, we set $A_1 = \ldots = A_k := Q$ and $A_{k+1} = \ldots = A_{k+m} := P$. For each polynomial a from (7.239), we now define the Weyl polynomial

[115] This expression is symmetric with respect to Q and P. Furthermore, the operator $A(a) : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ is formally self-adjoint, that is,

$$A(a)^\dagger := \tfrac{1}{2}(P^\dagger Q^\dagger + Q^\dagger P^\dagger) = \tfrac{1}{2}(PQ + QP) = A(a).$$

These properties would fail if we would assign to qp the operators QP or PQ .

[116] Note that the Moyal star product $a * b$ from (7.234) on page 591 is a finite sum if $a = a(q,p)$ and $b = b(q,p)$ are polynomials.

$$\boxed{A(a) := \sum_{k,m=0}^{N} c_{km}(q^k p^m)_W.} \tag{7.240}$$

The polynomial a is called the symbol of the Weyl polynomial $A(a)$.

Proposition 7.56 *The Weyl correspondence* (7.240) *possesses the properties* (W1) *through* (W4) *formulated above.*

In particular, it follows from (W4) above that the symbol of the operator product $A(a)A(b)$ is the Moyal star product $a*b$ of the symbols a and b of the operators $A(a)$ and $B(a)$, respectively.

The proof of Prop. 7.56 is elementary. For the Moyal star product one has to use an induction argument. For example, it follows from relation (7.237) on page 594 that $q*p = qp + \frac{1}{2}\mathrm{i}\hbar$. Hence

$$A(q*p) = A(qp) + \tfrac{1}{2}\mathrm{i}\hbar A(1) = \tfrac{1}{2}(QP+PQ) + \tfrac{1}{2}\mathrm{i}\hbar I.$$

Using the commutation relation $QP - PQ = \mathrm{i}\hbar I$, we obtain

$$A(q*p) = QP = A(q)A(p).$$

Proposition 7.57 *Let $k = 0, 1, 2, \ldots$ and $r, s \in \mathbb{C}$. The operator $(rQ+sP)^k$ is the Weyl operator to the polynomial $a(q,p) := (rq+sp)^k$.*

The proof is elementary. For example, we have

$$(rq+sp)^2 = r^2q^2 + 2rsqp + s^2p^2$$

and $(rQ+sP)^2 = (rQ+sP)(rQ+sP) = r^2Q^2 + rs(QP+PQ) + s^2P^2$. Hence

$$(rQ+sP)^2 = r^2(q^2)_W + 2rs(qp)_W + s^2(p^2)_W.$$

Standard example. Let $\hat{a} \in \mathcal{S}(\mathbb{R}^2)$, and $N = 0, 1, \ldots$[117] Then the polynomial

$$a(q,p) = \frac{1}{2\pi\hbar}\int_{\mathbb{R}^2}\sum_{k=0}^{N}\frac{\mathrm{i}^k(rq+sp)^k}{\hbar^k k!}\hat{a}(r,s)drds,$$

with respect to the real variables q and p, is well-defined. By Prop. 7.57, the Weyl operator to the symbol a reads as

$$A(a) = \frac{1}{2\pi\hbar}\int_{\mathbb{R}^2}\sum_{k=0}^{N}\frac{\mathrm{i}^k(rQ+sP)^k}{\hbar^k k!}\hat{a}(r,s)drds.$$

Formal generalization. Now consider the well-defined integral

$$a(q,p) = \frac{1}{2\pi\hbar}\int_{\mathbb{R}^2}\mathrm{e}^{\mathrm{i}(rq+sp)/\hbar}\hat{a}(r,s)drds.$$

Here, $a \in \mathcal{S}(\mathbb{R}^2)$. Explicitly, $\hat{a}$ is the Fourier transform of a. Using the formal limit $N \to \infty$, we get

[117] The definition of both the Schwartz function space $\mathcal{S}(\mathbb{R}^n)$ and the space of tempered distributions $\mathcal{S}'(\mathbb{R}^n)$ can be found on pages 537 and 615 of Volume I.

$$A(a) = \frac{1}{2\pi\hbar} \int_{\mathbb{R}^2} \mathrm{e}^{\mathrm{i}(rQ+sP)/\hbar} \hat{a}(r,s) dr ds. \tag{7.241}$$

This formal expression is frequently used by physicists.

Inductive construction of the Weyl operators. One can show that, for all polynomials a of the form (7.239), the following rigorous formulas hold:

$$QA(a) = A(qa + \tfrac{1}{2}\mathrm{i}\hbar a_p), \quad A(a)Q = A(qa - \tfrac{1}{2}\mathrm{i}\hbar a_q),$$
$$PA(a) = A(pa - \tfrac{1}{2}\mathrm{i}\hbar a_p), \quad A(a)P = A(pa + \tfrac{1}{2}\mathrm{i}\hbar a_p).$$

For example, if $a(q,p) := p$, then we get $QP = QA(p) = A(qp) + \frac{1}{2}\mathrm{i}\hbar I$. In addition, we have $PQ = A(p)Q = A(qp) - \frac{1}{2}\hbar\mathrm{i}I$. Hence $QP + PQ = 2A(qp)$.

7.12.4 Operator Kernels

> Operator kernels generalize matrix elements; they relate differential operators to integral operators, in a generalized sense. The formal approach was introduced by Paul Dirac in the late 1920s (Dirac calculus). The rigorous theory is based on the kernel theorem which was proved by Laurent Schwartz in the late 1940s (theory of tempered distributions).[118]
>
> Folklore

Classical kernels. For given function $\mathcal{A} \in \mathcal{S}(\mathbb{R}^2)$, we define

$$\boxed{(A\psi)(x) := \int_{\mathbb{R}^2} \mathcal{A}(x,y)\psi(y)dy, \qquad x \in \mathbb{R}}$$

for all functions $\psi \in \mathcal{S}(\mathbb{R})$. The function $\mathcal{A}$ is called the kernel of the linear, sequentially continuous operator

$$A : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R}). \tag{7.242}$$

Each function $\varphi \in \mathcal{S}(\mathbb{R})$ corresponds to a tempered distribution $T_\varphi \in \mathcal{S}(\mathbb{R})$ given by

$$T_\varphi(\chi) := \int_{\mathbb{R}} \varphi(x)\chi(x)dx \qquad \text{for all} \quad \chi \in \mathcal{S}(\mathbb{R}).$$

The map $\varphi \mapsto T_\varphi$ is an injective, linear, sequentially continuous map from $\mathcal{S}(\mathbb{R})$ into $\mathcal{S}'(\mathbb{R})$. Identifying φ with T_φ, we get $\mathcal{S}(\mathbb{R}) \subseteq \mathcal{S}'(\mathbb{R})$. In this sense, the map $\psi \mapsto A\psi \mapsto T_{A\psi}$ yields the linear, sequentially continuous operator

$$A : \mathcal{S}(\mathbb{R}) \to \mathcal{S}'(\mathbb{R}).$$

Explicitly, we obtain

$$(A\psi)(\chi) = \int_{\mathbb{R}^2} \mathcal{A}(x,y)\chi(x)\psi(y)dxdy \qquad \text{for all} \quad \psi, \chi \in \mathcal{S}(\mathbb{R}). \tag{7.243}$$

[118] L. Schwartz, Théorie des noyaux (Theory of kernels) (in French), Proceedings of the 1950 International Congress of Mathematicians in Cambridge, Massachusetts, Vol. I, pp. 220–230, Amer. Math. Soc., Providence, Rhode Island, 1952. At this congress, Laurent Schwartz (1915–2002) was awarded the Fields medal for creating the theory of distributions in about 1945.

Here, we briefly write $(A\psi)(\chi)$ instead of $T_{A\psi}(\chi)$. Introducing the tempered distribution $\mathsf{A} \in \mathcal{S}'(\mathbb{R}^2)$ by setting

$$\mathsf{A}(\varrho) := \int_{\mathbb{R}^2} \mathcal{A}(x,y)\varrho(x,y)dxdy \qquad \text{for all} \quad \varrho \in \mathcal{S}(\mathbb{R}^2),$$

equation (7.243) tells us that

$$\boxed{(A\psi)(\chi) = \mathsf{A}(\chi \otimes \psi) \qquad \text{for all} \quad \psi, \chi \in \mathcal{S}(\mathbb{R}).} \tag{7.244}$$

The product property of kernels. If the kernels $\mathcal{A}, \mathcal{B} \in \mathcal{S}(\mathbb{R}^2)$ correspond to the operators $A, B : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$, respectively, then the product operator AB has the kernel $\mathcal{C}$ given by the product formula

$$\boxed{\mathcal{C}(x,y) := \int_{\mathbb{R}} \mathcal{A}(x,z)\mathcal{B}(z,y)dz \qquad \text{for all} \quad x, y \in \mathbb{R}.} \tag{7.245}$$

This relation generalizes the matrix product. To prove (7.245), set $\chi := A\psi$ and $\psi := B\varphi$. Then $\chi = (AB)\varphi$. Hence

$$\chi(x) = \int_{\mathbb{R}} \mathcal{A}(x,z)(B\varphi)(z)dz = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathcal{A}(x,z)\mathcal{B}(z,y)dz \right) \varphi(y)dy.$$

The kernel of the position operator Q. For all $\chi, \psi \in \mathcal{S}(\mathbb{R})$,

$$(Q\psi)(\chi) = \int_{\mathbb{R}^2} \chi(x)x\psi(x)dx. \tag{7.246}$$

Using the Dirac delta function, the equation $(Q\psi)(x) = x\psi(x)$ can formally be written as

$$(Q\psi)(x) = \int_{\mathbb{R}} x\delta(x-y)\psi(y)dy \qquad \text{for all} \quad x \in \mathbb{R}.$$

Thus, the function $\mathcal{Q}(x,y) := x\delta(x-y)$ is the formal kernel of the position operator Q. Using the Dirac calculus[119], the formal kernel of the position operator Q can also be obtained by

$$\mathcal{Q}(x,y) = \langle x|Q|y\rangle = y\langle x|y\rangle = y\delta(x-y) = x\delta(x-y).$$

The kernel of the momentum operator P. For all $\chi, \psi \in \mathcal{S}(\mathbb{R})$,

$$(P\psi)(\chi) = \int_{\mathbb{R}^2} (-\mathrm{i}\hbar\psi'(x))\chi(x)dx. \tag{7.247}$$

In order to get the formal kernel $\mathcal{P}$ of the operator P used by physicists, we start with the (rescaled) Fourier transformation

$$(\mathcal{F}\varphi)(p) := \frac{1}{\sqrt{2\pi\hbar}} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}xp/\hbar}\varphi(x)dx, \qquad \varphi(x) = \frac{1}{\sqrt{2\pi\hbar}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}xp/\hbar}(\mathcal{F}\varphi)(p)dp.$$

Here, the operator $\mathcal{F} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ is bijective, linear, and sequentially continuous, and the inverse operator $\mathcal{F}^{-1}$ has the same properties. It follows from

[119] See page 596 of Volume I.

$$\mathcal{F}(P\psi)(p) = p(\mathcal{F}\psi)(p)$$

that we have the formal relation

$$\mathcal{F}(P\psi)(p) = \int_{\mathbb{R}} p\delta(p-r)(\mathcal{F}\psi)(r)dr \qquad \text{for all} \quad p \in \mathbb{R}.$$

This implies $(P\psi)(x) = \int_{\mathbb{R}} \mathcal{P}(x,y)\psi(y)dy$ with the formal kernel

$$\begin{aligned}\mathcal{P}(x,y) :&= \frac{1}{2\pi\hbar}\int_{\mathbb{R}^2} \mathrm{e}^{\mathrm{i}(xp-yr)/\hbar} p\delta(p-r)dpdr \\ &= \frac{1}{2\pi\hbar}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}p(x-y)/\hbar} p\, dp \qquad \text{for all} \quad x,y \in \mathbb{R}. \end{aligned} \tag{7.248}$$

Using the Dirac calculus (i.e., the completeness relation $\int_{\mathbb{R}} dp\, |p\rangle\langle p| = I$), the formal kernel can also be obtained by

$$\mathcal{P}(x,y) = \langle x|P|y\rangle = \int_{\mathbb{R}} dp \int_{\mathbb{R}} dr \;\; \langle x|p\rangle\langle p|P|r\rangle\langle r|y\rangle.$$

Noting that $\langle p|P|r\rangle = r\langle p|r\rangle = r\delta(p-r) = p\delta(p-r)$ and $\langle x|p\rangle = \mathrm{e}^{\mathrm{i}xp/\hbar}/\sqrt{2\pi\hbar}$, again we get (7.248).

The Schwartz kernel theorem. *Let $A : \mathcal{S}(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ be a linear, sequentially continuous operator (e.g., the Weyl operator $A(a)$ to the polynomial symbol a). Then there exists precisely one tempered distribution $\mathsf{A} \in \mathcal{S}'(\mathbb{R}^2)$ such that*

$$(A\psi)(\chi) = \mathsf{A}(\chi \otimes \psi) \qquad \textit{for all} \quad \psi, \chi \in \mathcal{S}(\mathbb{R}). \tag{7.249}$$

The tempered distribution A is called the kernel of the operator A.

This theorem generalizes (7.244). The kernels of the operators Q and P are given by (7.246) and (7.247), respectively.

Nuclear spaces. The Schwartz kernel theorem is the special case of a functional-analytic theorem about bilinear forms on nuclear spaces. A Hilbert space is nuclear iff its dimension is finite. Furthermore, the infinite-dimensional spaces $\mathcal{D}(\mathbb{R}^n)$ and $\mathcal{S}(\mathbb{R}^n)$ are nuclear for $n = 1, 2, \ldots$ For the theory of nuclear spaces and their important applications in harmonic analysis, we refer to the following monographs:

A. Pietsch, Nuclear locally convex spaces, Springer, Berlin, 1972.

A. Pietsch, Operator Ideals, Deutscher Verlag der Wissenschaften, Berlin, 1978.

A. Pietsch, History of Banach Spaces and Linear Operators, Birkhäuser, Boston, 2007.

I. Gelfand, G. Shilov, and N. Vilenkin, Generalized Functions, Vols. 1–5, Academic Press, New York, 1964.

K. Maurin, Generalized Eigenfunction Expansions and Unitary Representations of Topological Groups, Polish Scientific Publishers, Warsaw, 1968.

K. Maurin, Methods of Hilbert Spaces, Polish Scientific Publishers, Warsaw, 1972.

The theory of nuclear spaces was created by Grothendieck in the 1950s. In the 1955s, Grothendieck left analysis, and he moved to algebra and geometry. For his seminal contributions to algebraic geometry, homological algebra, and functional analysis, Alexandre Grothendieck (born 1928 in Berlin) was awarded the Fields medal in 1966. His childhood and youth was overshadowed by German fascism. His father died in the German concentration camp Auschwitz in 1942. We refer to:

A. Grothendieck, Récoltes et Semailles: réflexions et témoignage sur un passé de mathématicien, 1986 (ca. 1000 pages) (in French).(Reaping and Sowing: the life of a mathematician – reflections and bearing witness). Internet: `http://www.fermentmagazine.org/rands/recoltes1.html` Translations into English, Russian, and Spanish are ongoing.

P. Cartier, A mad day's work: from Grothendieck to Connes and Kontsevich. The evolution of concepts of space and symmetry. Bull. Amer. Math. Soc. **38**(4) (2001), 389–408.

W. Scharlau, Who is Alexander Grothendieck? Part I, 2007 (in German). Internet: `http://www.Scharlau-online.de/DOKS/ag`

7.12.5 The Formal Weyl Calculus

Our goal is to extend the relation between polynomial symbols $a = a(q,p)$ and Weyl operators $A(a)$ to more general symbols a. In order to motivate the rigorous approach to be considered in Sect. 7.12.6, let us start with purely formal arguments used by physicists. The key formulas read as follows.

(i) Superposition: For the symbol

$$a(q,p) = \frac{1}{2\pi\hbar} \int_{\mathbb{R}^2} \mathrm{e}^{\mathrm{i}(xq+yp)/\hbar}\, \hat{a}(x,y)dxdy, \qquad q,p \in \mathbb{R},$$

the Weyl operator is given by

$$\boxed{A(a) := \frac{1}{2\pi\hbar} \int_{\mathbb{R}^2} \mathrm{e}^{\mathrm{i}(xQ+yP)/\hbar}\, \hat{a}(x,y)dxdy.} \tag{7.250}$$

Here, $\hat{a} = \hat{a}(x,y)$ is the (rescaled) Fourier transform of $a = (q,p)$.

(ii) The kernel formula: We have

$$(A\psi)(x) = \int_{\mathbb{R}} \mathcal{A}(x,y)\psi(y)dy, \qquad x \in \mathbb{R}$$

with the formal kernel

$$\boxed{\mathcal{A}(x,y) = \frac{1}{2\pi\hbar} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}p(x-y)/\hbar}\, a\left(\frac{x+y}{2}, p\right) dp, \qquad x,y \in \mathbb{R}.} \tag{7.251}$$

The inverse Fourier transformation yields

$$a(q,p) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}rp/\hbar} \mathcal{A}(q - \tfrac{1}{2}r, q + \tfrac{1}{2}r)dr, \qquad q,p \in \mathbb{R}. \tag{7.252}$$

(iii) Formal self-adjointness: For the formally adjoint operator of the Weyl operator $A(a)$ on the Hilbert space $L_2(\mathbb{R})$, we get

$$A(a)^\dagger = A(a^\dagger).$$

In particular, if the function a is real-valued, then the corresponding Weyl operator $A(a)$ is formally self-adjoint on the Hilbert space $L_2(\mathbb{R})$.

(iv) The composition formula: If the symbols $a = a(q,p)$ and $b = b(q,p)$ correspond to the Weyl operators $A(a)$ and $A(b)$, then the operator product is given by

$$\boxed{A(a)A(b) = A(a * b)}$$

with the star product

$$(a * b)(q,p) := \frac{1}{\pi^2\hbar^2} \int_{\mathbb{R}^4} \mathrm{e}^{2\varrho\,\mathrm{i}/\hbar}\; a(q_1,p_1)b(q_2,p_2)\; dq_1dp_1dq_2dp_2$$

for all $q, p \in \mathbb{R}$. Here, the function $\varrho = \varrho(q,p,q_1,p_1,q_2,p_2)$ is defined by

$$\varrho := \begin{vmatrix} q & p & 1 \\ q_1 & p_1 & 1 \\ q_2 & p_2 & 1 \end{vmatrix} = q(p_1 - p_2) + p(q_2 - q_1) + (q_1p_2 - p_1q_2).$$

If $\hat{b}$ denotes the (rescaled) Fourier transform of b, that is,

$$\hat{b}(\xi,\eta) = \frac{1}{2\pi\hbar} \int_{\mathbb{R}^2} \mathrm{e}^{-\mathrm{i}(q\xi+p\eta)/\hbar}\; b(q,p)\; dqdp, \tag{7.253}$$

then

$$\boxed{(a * b)(q,p) = \frac{1}{2\pi\hbar} \int_{\mathbb{R}^2} \mathrm{e}^{\mathrm{i}(q\xi+p\eta)/\hbar}\; a\left(q - \frac{\eta}{2}, p + \frac{\xi}{2}\right) \hat{b}(\xi,\eta)\; d\xi d\eta.}$$

As we will show below by using the Fourier transform together with the Taylor expansion, this implies

$$(a * b)(q,p) = a\left(q + \frac{\mathrm{i}\hbar}{2}\frac{\partial}{\partial p_2}, p - \frac{\mathrm{i}\hbar}{2}\frac{\partial}{\partial q_2}\right) b(q_2,p_2)_{|q_2=q,p_2=p}.$$

Here, we have to assume that a is a polynomial (or a formal power series expansions with respect to q and p). Finally, note that the star product $a * b$ coincides with the formal Moyal star product, that is,

$$\boxed{a * b = \sum_{m,n=0}^{\infty} \left(\frac{\mathrm{i}\hbar}{2}\right)^{m+n} \frac{(-1)^m}{m!n!}\; \frac{\partial^{m+n}a}{\partial p^m \partial q^n} \frac{\partial^{m+n}b}{\partial p^n \partial q^m}.} \tag{7.254}$$

Let us motivate this in a formal manner. To simplify notation, we set $\hbar := 1$.

Ad (i) See formula (7.241) on page 599.

Ad (ii). (I) Commutation relation. It follows from $QP - PQ = \mathrm{i}I$ that

$$Q^nP - PQ^n = \mathrm{i}nQ^{n-1}, \qquad n = 1,2,\ldots$$

by induction. If $F(Q) = a_0I + a_1Q + a_2Q^2 + \ldots$, then we formally get

$$F(Q)P - PF(Q) = \mathrm{i}F'(Q).$$

In particular, $\mathrm{e}^{-\mathrm{i}trQ}P - P\mathrm{e}^{-\mathrm{i}trQ} = tr \cdot \mathrm{e}^{-\mathrm{i}trQ}$ for all $t, r \in \mathbb{R}$.

(II) Let us prove the key relation

$$\mathrm{e}^{\mathrm{i}t(rQ+sP)} = \mathrm{e}^{\mathrm{i}t^2rs/2} \cdot \mathrm{e}^{\mathrm{i}trQ}\mathrm{e}^{\mathrm{i}tsP}, \qquad r,s \in \mathbb{R}. \tag{7.255}$$

To this end, we set $U(t) := \mathrm{e}^{-\mathrm{i}tsP}\mathrm{e}^{-\mathrm{i}trQ}\mathrm{e}^{\mathrm{i}t(rQ+sP)}$ for all $t \in \mathbb{R}$. Differentiating with respect to time t and using (I), we obtain

$$U'(t) = -\mathrm{i}s\mathrm{e}^{-\mathrm{i}tsP}\left(P\mathrm{e}^{-\mathrm{i}trQ} - \mathrm{e}^{-\mathrm{i}trQ}P\right)\mathrm{e}^{\mathrm{i}t(rQ+sP)} = \mathrm{i}trsU(t).$$

Since $U(0) = I$, we get $U(t) = \mathrm{e}^{\mathrm{i}t^2rs/2}I$. This implies (7.255).

(III) Setting $t = 1$, we obtain

$$\mathrm{e}^{\mathrm{i}(rQ+sP)} = \mathrm{e}^{\mathrm{i}rs/2}\cdot\mathrm{e}^{\mathrm{i}rQ}\mathrm{e}^{\mathrm{i}sP}, \qquad r, s \in \mathbb{R}.$$

Recall that $\mathrm{i}P\psi = \psi'$. By Taylor expansion,

$$(\mathrm{e}^{\mathrm{i}sP}\psi)(x) = \psi(x) + s\psi'(x) + \tfrac{s^2}{2!}\psi''(x) + \ldots = \psi(x+s).$$

Similarly,

$$(\mathrm{e}^{\mathrm{i}rQ}\psi)(x) = \psi(x) + \mathrm{i}rx\psi(x) + \frac{(\mathrm{i}rx)^2}{2!}\psi(x) + \ldots = \mathrm{e}^{\mathrm{i}rx}\psi(x).$$

Hence $(\mathrm{e}^{\mathrm{i}(rQ+sP)}\psi)(x) = \mathrm{e}^{\mathrm{i}rs/2}\mathrm{e}^{\mathrm{i}rx}\psi(x+s)$ for all $x \in \mathbb{R}$.

(IV) We briefly write A instead of $A(a)$. By (7.250),

$$(A\psi)(x) = \frac{1}{2\pi}\int_{\mathbb{R}^2}\mathrm{e}^{\mathrm{i}rx}\mathrm{e}^{\mathrm{i}rs/2}\psi(x+s)\,\hat{a}(r,s)drds.$$

Inserting $\hat{a}(r,s) = \frac{1}{2\pi}\int_{\mathbb{R}^2}\mathrm{e}^{-\mathrm{i}(rq+sp)}\,a(q,p)dqdp$, we get

$$(A\psi)(x) = \frac{1}{(2\pi)^2}\int_{\mathbb{R}^2}\mathrm{e}^{\mathrm{i}r(x-q+\frac{1}{2}s)}\mathrm{e}^{-\mathrm{i}sp}\,a(q,p)\psi(x+s)drdsdqdp.$$

Since $\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}r(x-q+\frac{1}{2}s)}dr = 2\pi\delta(x-q+\frac{1}{2}s)$, we obtain

$$(A\psi)(x) = \frac{1}{2\pi}\int_{\mathbb{R}^2}\mathrm{e}^{-\mathrm{i}sp}\,a(x+\tfrac{1}{2}s,p)\psi(x+s)dpds.$$

Finally, the substitution $y = x + s$ yields the desired result

$$(A\psi)(x) = \frac{1}{2\pi}\int_{\mathbb{R}^2}\mathrm{e}^{\mathrm{i}(x-y)p}\,a\left(\frac{x+y}{2},p\right)\psi(y)dpdy.$$

Ad (iii). By (ii), the operator $A(a^\dagger)$ has the kernel

$$\mathcal{B}(x,y) = \frac{1}{2\pi}\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}p(x-y)/\hbar}\,a\left(\frac{x+y}{2},p\right)^\dagger dp, \qquad x, y \in \mathbb{R}.$$

Again by (ii), this is equal to $\mathcal{A}(y,x)^\dagger$. Hence $A(a^\dagger) = A(a)^\dagger$.

Ad (iv). (I) The kernel $\mathcal{C}$ of the operator product $C := A(a)A(b)$ is given by

$$\mathcal{C}(x,y) = \int_{\mathbb{R}}\mathcal{A}(x,z)\mathcal{B}(z,y)dz.$$

By (ii), we have the following relations between the symbols a, b and the kernels $\mathcal{A}, \mathcal{B}$, respectively:

$$\mathcal{A}(x,z) = \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}p_1(x-z)}\; a\left(\frac{x+z}{2}, p_1\right) dp_1,$$
$$\mathcal{B}(z,y) = \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}p_2(z-y)}\; a\left(\frac{z+y}{2}, p_2\right) dp_2.$$

Hence

$$\mathcal{C}(x,y) = \frac{1}{4\pi^2}\int_{\mathbb{R}^3} \mathrm{e}^{\mathrm{i}p_1(x-z)}\mathrm{e}^{\mathrm{i}p_2(z-y)} a\left(\frac{x+z}{2}, p_1\right) b\left(\frac{z+y}{2}, p_2\right) dp_1 dp_2 dz.$$

Let c be the symbol of the operator C. Again by (ii), after the rescaling $\eta = \frac{1}{2}r$, we get

$$c(q,p) = 2\int_{\mathbb{R}} \mathrm{e}^{2\mathrm{i}p\eta}\; \mathcal{C}(q-\eta, q+\eta) d\eta.$$

Therefore,

$$c(q,p) = \frac{1}{2\pi^2}\int_{\mathbb{R}^4} \mathrm{e}^{\mathrm{i}\sigma}\; a\left(\frac{q+z-\eta}{2}, p_1\right) b\left(\frac{q+z+\eta}{2}, p_2\right) dp_1 dp_2 dz d\eta$$

with $\sigma := (q-z-\eta)p_1 + (z-q-\eta)p_2 + 2p\eta$. Using the substitution

$$q_1 = \tfrac{1}{2}(q+z-\eta), \qquad q_2 = \tfrac{1}{2}(q+z+\eta)$$

and setting $\varrho := (q-q_2)p_1 + (q_1-q)p_2 + (q_2-q_1)p$, we obtain

$$c(q,p) = \frac{1}{\pi^2}\int_{\mathbb{R}^4} \mathrm{e}^{2\mathrm{i}\varrho}\; a(q_1,p_1) b(q_2,p_2) dp_1 dp_2 dq_1 dq_2. \tag{7.256}$$

(II) Moyal product. Using the substitution $q_1 = q - \frac{1}{2}\eta, p_1 = p + \frac{1}{2}\xi$, we get

$$c(q,p) = \frac{1}{4\pi^2}\int_{\mathbb{R}^4} \mathrm{e}^{\mathrm{i}(q-q_2)\xi}\mathrm{e}^{\mathrm{i}(p-p_2)\eta}\; a\left(q-\frac{\eta}{2}, p+\frac{\xi}{2}\right) b\,(q_2,p_2)\;\, dq_2 dp_2 d\xi d\eta.$$

If $\hat{b}$ denotes the (rescaled) Fourier transform (7.253) of the function b, then

$$c(q,p) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \mathrm{e}^{\mathrm{i}(q\xi+p\eta)}\; a\left(q-\frac{\eta}{2}, p+\frac{\xi}{2}\right) \hat{b}(\xi,\eta)\; d\xi d\eta. \tag{7.257}$$

Suppose now that the symbol a is a polynomial (or a formal power series expansion). By Fourier transform, we get the formal expression

$$c(q,p) = a\left(q + \frac{\mathrm{i}}{2}\frac{\partial}{\partial p_2}, p - \frac{\mathrm{i}}{2}\frac{\partial}{\partial q_2}\right) b(q_2,p_2)_{|q_2=q, p_2=p}. \tag{7.258}$$

Finally, using Taylor expansion, we obtain

$$c(q,p) = \sum_{m,n=0}^{\infty} \left(\frac{\mathrm{i}}{2}\right)^{m+n} \frac{(-1)^m}{m!n!}\; \frac{\partial^{m+n}a}{\partial p^m \partial q^n}\, \frac{\partial^{m+n}b}{\partial p^n \partial q^m}. \tag{7.259}$$

(III) Motivation of (7.258). First let $a(q,p) := q$. It follows from

$$b(q_2,p_2) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \mathrm{e}^{\mathrm{i}(q_2\xi+p_2\eta)}\; \hat{b}(\xi,\eta)\; d\xi d\eta$$

that

$$\left(q + \frac{\mathrm{i}}{2}\frac{\partial}{\partial p_2}\right) b(q_2, p_2) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \mathrm{e}^{\mathrm{i}(q_2\xi + p_2\eta)} \left(q - \frac{\eta}{2}\right) \hat{b}(\xi, \eta)\, d\xi d\eta.$$

Setting $q_2 = q$ and $p_2 = p$ and using (7.257), we obtain (7.258).

Similarly, if $a(q, p) := p$, then

$$\left(p - \frac{\mathrm{i}}{2}\frac{\partial}{\partial q_2}\right) b(q_2, p_2) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \mathrm{e}^{\mathrm{i}(q_2\xi + p_2\eta)} \left(p + \frac{\xi}{2}\right) \hat{b}(\xi, \eta)\, d\xi d\eta.$$

Again this yields (7.258).

(IV) Motivation of (7.259). This follows from

$$a(q + \alpha, p + \beta) = \sum_{m,n=0}^{\infty} \frac{\partial^{m+n} a(q, p)}{\partial p^m \partial q^n} \cdot \frac{\beta^m \alpha^n}{m! n!},$$

by setting $\beta := -\frac{\mathrm{i}}{2}\frac{\partial}{\partial q}$ and $\alpha := \frac{\mathrm{i}}{2}\frac{\partial}{\partial p}$. □

7.12.6 The Rigorous Weyl Calculus

It is possible to translate the formal Weyl calculus into a rigorous mathematical approach by using the language of generalized functions. It is our goal to assign to a general class of symbols Weyl operators in such a way that

- the theory of Weyl polynomials from Sect. 7.12.3 is generalized and
- the formal Weyl calculus from Sect. 7.12.5 gets a rigorous mathematical basis.

The proofs of the following statements can be found in the monographs by L. Hörmander, The Analysis of Linear Partial Differential Operators, Vol. 3, Springer, New York, 1983, and by M. de Gosson, Symplectic Geometry and Quantum Mechanics, Birkhäuser, Basel, 2006.

Smooth, rapidly decreasing symbols. Let $a, b \in \mathcal{S}(\mathbb{R})$. The functions a and b are called symbols. Then the following hold.

(i) Weyl operator: For given symbol a, define the Weyl operator

$$(A(a)\psi)(x) := \int_{\mathbb{R}} \mathcal{A}(x, y)\psi(y)dy, \qquad x \in \mathbb{R}$$

for all $\psi \in \mathcal{S}(\mathbb{R})$ with the kernel

$$\mathcal{A}(x, y) := \frac{1}{2\pi\hbar}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}p(x-y)/\hbar}\, a\left(\frac{x+y}{2}, p\right) dp, \qquad x, y \in \mathbb{R}.$$

Then $\mathcal{A} \in \mathcal{S}(\mathbb{R}^2)$, and the operator $A(a) : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ is linear and sequentially continuous.

(ii) Bilinear form: Let $\chi, \psi \in \mathcal{S}(\mathbb{R}^2)$. Then

$$(A(a)\psi)(\chi) = \int_{\mathbb{R}^2} \mathcal{A}(x, y)\varphi(x)\psi(y)dy, \qquad x, y \in \mathbb{R}.$$

Hence

$$(A(a)\psi)(\chi) = \frac{1}{2\pi\hbar}\int_{\mathbb{R}^2} \mathrm{e}^{\mathrm{i}p(x-y)/\hbar}\, a\left(\frac{x+y}{2}, p\right) \chi(x)\psi(y)dxdydp.$$

Using the substitution $y = 2q - x, x = x$, we get

$$(A(a)\psi)(\chi) = \frac{1}{\pi\hbar}\int_{\mathbb{R}^2} \mathrm{e}^{2\mathrm{i}p(x-q)/\hbar}\; a\,(q,p)\,\chi(x)\psi(2q-x)dxdqdp.$$

This implies

$$(A(a)\psi)(\chi) = \int_{\mathbb{R}^2} a(q,p)\varrho_{\chi,\psi}(q,p)dqdp \tag{7.260}$$

with $\varrho_{\chi,\psi}(q,p) := \frac{1}{\pi\hbar}\int_{\mathbb{R}} \mathrm{e}^{2\mathrm{i}p(x-q)/\hbar}\chi(x)\psi(2q-x)dx$.

(ii) Formal self-adjointnes: We get

$$A(a)^\dagger = A(a^\dagger).$$

This means that $\langle A(a^\dagger)\varphi|\psi\rangle = \langle\varphi|A(a)\psi\rangle$ for all $\psi,\varphi \in \mathcal{S}(\mathbb{R})$, where $\langle .|.\rangle$ is the inner product on the Hilbert space $L_2(\mathbb{R})$.

(iii) The composition formula and the rigorous Moyal star product: For the operator product, we have

$$A(a)A(b) = A(a*b)$$

together with the rigorous Moyal star product[120]

$$\boxed{(a*b)(q,p) := \frac{1}{\pi^2\hbar^2}\int_{\mathbb{R}^4} \mathrm{e}^{2\varrho\mathrm{i}/\hbar}\; a(q_1,p_1)\cdot b(q_2,p_2)\; dq_1dp_1dq_2dp_2}$$

for all $q,p \in \mathbb{R}$. Here, we use the determinant

$$\varrho := \begin{vmatrix} q & p & 1 \\ q_1 & p_1 & 1 \\ q_2 & p_2 & 1 \end{vmatrix} = q(p_1-p_2)+p(q_2-q_1)+(q_1p_2-p_1q_2). \tag{7.261}$$

This coincides with (7.256).

(iv) Associativity of the Moyal star product: For all $a,b,c \in \mathcal{S}(\mathbb{R})$, we have

$$(a*b)*c = a*(b*c).$$

Tempered distributions as symbols. Let $a \in \mathcal{S}'(\mathbb{R}^2)$. Motivated by (7.260), define

$$\boxed{(A(a)\psi)(\chi) := a(\varrho_{\chi,\psi}) \qquad \chi,\psi \in \mathcal{S}(\mathbb{R}).}$$

Then $A(a)\psi \in \mathcal{S}'(\mathbb{R})$, and the linear operator $A(a) : \mathcal{S}(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ is sequentially continuous. In particular, if $a = a(q,p)$ is a polynomial with respect to the variables q and p, then the corresponding tempered distribution is given by $a(\varrho) = \int_{\mathbb{R}^2} a(q,p)\varrho(q,p)dqdp$ for all $\varrho \in \mathcal{S}(\mathbb{R}^2)$.

[120] In the general case, the rigorous Moyal star product (7.261) differs from the formal Moyal star product (7.254). This is discussed in G. Piacitelli, Nonlocal theories: new rules for old diagrams, 2004. Internet: arXiv: hep-th/0403055

7.13 Two Magic Formulas

> According to one view, the Feynman path integral is simple a suitable hierolglyphic shorthand for an algorithm of perturbation theory. On the other hand, the traditional (Wiener) view of the path integral as an integral with respect to a measure in the function space runs into practically insurmountable difficulties here and is thus also imperfect. Our own view is that the Feynman path integral should be understood as the limit of finite-dimensional approximations. But which approximations? The path integral proves to be very sensitive to the choice of its approximations, the resulting ambiguity being of the same nature as the non-uniqueness of quantization.[121]
>
> Feliks Berezin and Mikhail Shubin, 1991

It is our goal to use the Weyl calculus in order to get the two magic formulas (7.274) on page 614 and (7.277) on page 615 for the kernel of the Feynman propagator operator and the kernel of the Heisenberg scattering operator, respectively. It turns out that the Weyl calculus relates the Feynman propagator kernel to the Feynman path integral in a quite natural manner.

Basic ideas. Consider the motion $q = q(t)$ of a classical particle on the real line with the equation of motion

$$\dot{p}(t) = -a_q(q(t), p(t)), \qquad \dot{q}(t) = a_p(q(t), p(t)), \qquad t \in \mathbb{R}.$$

Here, the given classical Hamiltonian $a : \mathbb{R}^2 \to \mathbb{R}$ is assumed to be smooth. Now we pass to the corresponding quantum particle. Then we have to study the Schrödinger equation

$$\boxed{\mathrm{i}\hbar\psi_t = H\psi, \qquad \psi(t_0) = \psi_0} \tag{7.262}$$

for the wave function $\psi = \psi(x, t)$ of the quantum particle on the Hilbert space $L_2(\mathbb{R})$.

In terms of Weyl quantization, the operator $H = A(a)$ is the Weyl operator related to the symbol $a = a(q, p)$. This operator is called the Hamiltonian (or energy operator) of the quantum particle. It is our goal to study both

- the full dynamics of the quantum particle (i.e., the Feynman propagator operator $P(t, t_0) := \mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}$), and
- scattering processes for the quantum particle (i.e., the Heisenberg scattering operator $\mathsf{S}(t, t_0) := \mathrm{e}^{\mathrm{i}tH_{\mathrm{free}}/\hbar}\mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}\mathrm{e}^{-\mathrm{i}t_0H_{\mathrm{free}}/\hbar}$). Here, we assume that the Hamiltonian H is a perturbation of the free Hamiltonian H_{free}. Explicitly,

$$H = H_{\mathrm{free}} + \kappa U. \tag{7.263}$$

A scattering process is characterized by the property that the motion of the quantum particle is free in the remote past ($t_0 \to -\infty$) and in the far future ($t \to +\infty$). The free Hamiltonian $H_{\mathrm{free}} = P^2/2m$ is the Weyl operator to the symbol $a_{\mathrm{free}}(p) := p^2/2m$, and the operator U is the Weyl operator to the symbol $q \mapsto U(q)$. The real number κ is called coupling constant. Summarizing, the Hamiltonian operator H has the symbol

$$a(q, p) = \frac{p^2}{2m} + \kappa U(q).$$

[121] F. Berezin and M. Shubin, The Schrödinger Equation, Kluwer, Dordrecht, 1991 (reprinted with permission).

We will proceed in the following manner.

(a) Evolution operators: We start with time-dependent operators in the Hilbert space $L_2(\mathbb{R})$ (i.e., the Feynman propagator and the Heisenberg scattering operator).
(b) Kernels: The evolution operators can be described by kernels depending on space and time coordinates.
(c) Causality: The kernel on a finite time interval is the superposition of kernels on small time intervals.
(d) Reduction to operator symbols: The kernel of a small time interval can be computed by using the kernel formula of the Weyl calculus, which depends on the symbol of the evolution operator.
(e) Limit: If the small time interval goes to zero, then the kernel of the evolution operator can be expressed by a Feynman path integral, which depends on the symbol a of the Hamiltonian operator.

This way, we obtain an elegant relation between classical mechanics described by the classical Hamiltonian a and

- the kernel $\mathcal{K}$ of the Feynman propagator operator (called the Feynman propagator kernel), and
- the kernel $\mathcal{S}$ of the Heisenberg scattering operator (called the scattering kernel).

In what follows, we will only use formal arguments. Let us first discuss the physical meaning of both the Feynman propagator operator and the Heisenberg scattering operator.

The Feynman propagator operator. The operator

$$P(t,t_0) := \mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}, \qquad t \geq t_0$$

is called the Feynman propagator. For given initial state $\psi_0 \in L_2(\mathbb{R})$, the state

$$\psi(t) = P(t,t_0)\psi_0$$

is a solution of the Schrödinger equation (7.262). From the physical point of view, the propagator $P(t,t_0)$ sends the particle state ψ_0 at the initial time t_0 to the particle state $\psi(t)$ at time t. Therefore, the propagator describes the dynamics of the quantum particle. Let

$$-\infty < t_0 < t_1 < \cdots < t_{N-1} < t_N < \infty.$$

Then the addition theorem for the exponential function tells us that we have the following operator product

$$\boxed{P(t_N,t_0) = P(t_N,t_{N-1}) \cdots P(t_2,t_1)P(t_1,t_0).} \tag{7.264}$$

This product property reflects causality. To understand this, note that it follows from $\psi(t_1) = P(t_1,t_0)\psi_0$ and $\psi(t_2) = P(t_2,t_1)\psi(t_1)$ that

$$\psi(t_2) = P(t_2,t_1)P(t_1,t_0)\psi_0 = P(t_2,t_0)\psi_0.$$

The propagator $t \mapsto P(t,t_0)$ satisfies the following equation

$$\boxed{\mathrm{i}\hbar P_t(t,t_0) = HP(t,t_0), \qquad t \geq t_0, \quad P(t_0,t_0) = I,} \tag{7.265}$$

which is called the propagator differential equation.

The Heisenberg scattering operator. Suppose that the Hamiltonian operator H is the perturbation of the free Hamiltonian H_{free} according to (7.263). Let us investigate scattering processes. The operator

$$\mathsf{S}(t,t_0) := \mathrm{e}^{\mathrm{i}tH_{\text{free}}/\hbar} P(t,t_0) \mathrm{e}^{-\mathrm{i}t_0 H_{\text{free}}/\hbar}, \qquad t \geq t_0$$

with $P(t,t_0) := \mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}$ is called the Heisenberg scattering operator (or the S-matrix operator). In order to understand the physical meaning of the scattering operator, consider the free motion

$$\psi_{\text{free,in}}(t) := \mathrm{e}^{-\mathrm{i}tH_{\text{free}}/\hbar}\varphi_{\text{in}}, \qquad t \in \mathbb{R}$$

with the initial state φ_{in} at time $t = 0$, and

$$\psi_{\text{free,out}}(t) := \mathrm{e}^{-\mathrm{i}tH_{\text{free}}/\hbar}\varphi_{\text{out}}, \qquad t \in \mathbb{R}$$

with the initial state φ_{out} at time $t = 0$. The transition amplitude

$$\tau := \langle \psi_{\text{free,out}}(t) | P(t,t_0) \psi_{\text{free,in}}(t_0) \rangle, \qquad t > t_0$$

is equal to

$$\tau = \langle \varphi_{\text{out}} | \left(\mathrm{e}^{-\mathrm{i}tH_{\text{free}}/\hbar} \right)^{\dagger} P(t,t_0) \mathrm{e}^{-\mathrm{i}t_0 H_{\text{free}}/\hbar} \varphi_{\text{in}} \rangle = \langle \varphi_{\text{out}} | \mathsf{S}(t,t_0) \varphi_{\text{in}} \rangle.$$

The real number

$$\boxed{|\tau|^2 = |\langle \varphi_{\text{out}} | \mathsf{S}(t,t_0) \varphi_{\text{in}} \rangle|^2, \qquad t > t_0} \tag{7.266}$$

is the transition probability from the incoming free state $\psi_{\text{free,in}}(t_0)$ at time t_0 to the outgoing free state $\psi_{\text{free,out}}(t)$ at time t.

The transition probability (7.266) *indexscattering matrix (S-matrix)!transition probability is the key for computing cross sections of scattering processes in particle accelerators.*

We also define

$$\langle \varphi_{\text{out}} | \mathsf{S} \varphi_{\text{in}} \rangle := \lim_{t \to +\infty} \lim_{t_0 \to -\infty} = \langle \varphi_{\text{out}} | \mathsf{S}(t,t_0) \varphi_{\text{in}} \rangle$$

if this limit exists. Here, the complex number $\langle \varphi_{\text{out}} | \mathsf{S} \varphi_{\text{in}} \rangle$ is called an S-matrix element. Parallel to (7.264), we get the causal product relation

$$\boxed{\mathsf{S}(t_N,t_0) = \mathsf{S}(t_N,t_{N-1}) \cdots \mathsf{S}(t_2,t_1)\mathsf{S}(t_1,t_0).} \tag{7.267}$$

Furthermore, we have the differential equation

$$\boxed{\mathrm{i}\hbar \mathsf{S}_t(t,t_0) = \kappa \mathsf{U}(t) \mathsf{S}(t,t_0), \qquad t \geq t_0, \quad \mathsf{S}(t_0,t_0) = I} \tag{7.268}$$

for the scattering operator. Here, we introduce the transformed perturbation

$$\mathsf{U}(t) := \mathrm{e}^{\mathrm{i}tH_{\text{free}}/\hbar} U \mathrm{e}^{-\mathrm{i}t_0 H_{\text{free}}/\hbar}.$$

Let us motivate (7.268). To simplify notation, choose $\hbar := 1$. Then

$$\mathsf{i}\mathsf{S}_t(t,t_0) = -\mathrm{e}^{\mathrm{i}tH_{\mathrm{free}}}H_{\mathrm{free}}P(t,t_0)\mathrm{e}^{-\mathrm{i}t_0H_{\mathrm{free}}} + \mathrm{i}\mathrm{e}^{\mathrm{i}tH_{\mathrm{free}}}P_t(t,t_0)\mathrm{e}^{-\mathrm{i}t_0H_{\mathrm{free}}},$$

which is equal to

$$\mathrm{e}^{\mathrm{i}tH_{\mathrm{free}}}(H - H_{\mathrm{free}})P(t,t_0)\mathrm{e}^{-\mathrm{i}t_0H_{\mathrm{free}}} = \kappa\mathrm{e}^{\mathrm{i}tH_{\mathrm{free}}}U\mathrm{e}^{-\mathrm{i}t_0H_{\mathrm{free}}}\mathsf{S}(t,t_0).$$

Dyson's magic S-matrix formula. Let us pass from differential equations to integral equations. From the differential equation (7.268) for the scattering operator, we get the equivalent Volterra integral equation

$$\boxed{\mathsf{S}(t,t_0) = I - \frac{\mathsf{i}\kappa}{\hbar}\int_{t_0}^{t}\mathsf{U}(\tau)\mathsf{S}(\tau,t_0)d\tau, \qquad t \geq t_0.} \tag{7.269}$$

We have shown in Sect. 7.17.4 of Vol. I that the integral equation (7.269) has the unique solution

$$\boxed{\mathsf{S}(t,t_0) = \mathcal{T}\mathrm{e}^{-\frac{\mathsf{i}\kappa}{\hbar}\int_{t_0}^{t}\mathsf{U}(\tau)d\tau}, \qquad t \geq t_0} \tag{7.270}$$

where $\mathcal{T}$ is the chronological operator (see page 382 of Vol. I). This is Dyson's magic S-matrix formula which plays the decisive role in the operator-theoretic approach to quantum field theory. Comparing the propagator equation (7.265) with the equation (7.268) for the scattering operator, we get the following:

> *The scattering operator* $\mathsf{S}(t,t_0)$ *coincides with the Feynman propagator* $\mathsf{P}(t,t_0)$ *in the Dirac interaction picture (with respect to the transformed perturbation* $\kappa\mathsf{U}(t)$ *of the Hamiltonian operator).*[122]

This fact is of fundamental importance for understanding the S-matrix theory in quantum field theory.

The integral equation for states. For given $\varphi_{\mathrm{in}} \in L_2(\mathbb{R})$, introduce the function $\varphi(t) := \mathsf{S}(t,t_0)\varphi_{\mathrm{in}}$. By (7.269), we obtain the integral equation

$$\varphi(t) = \varphi_{\mathrm{in}} - \frac{\mathsf{i}\kappa}{\hbar}\int_{t_0}^{t}\mathsf{U}(\tau)\varphi(\tau)d\tau, \qquad t \geq t_0.$$

Let $\varphi = \varphi(t)$ be a solution of this integral equation. Set $\psi(t) := \mathrm{e}^{-\mathrm{i}tH_{\mathrm{free}}/\hbar}\varphi(t)$ for all $t \geq t_0$. Then

$$\psi(t) = P(t,t_0)\mathrm{e}^{-\mathrm{i}t_0H_{\mathrm{free}}/\hbar}\varphi_{\mathrm{in}}, \qquad t \geq t_0.$$

By the propagator equation (7.265), this is a solution of the Schrödinger equation (7.262) with the initial condition $\psi(t_0) = \psi_{\mathrm{free,in}} = \mathrm{e}^{-\mathrm{i}t_0H_{\mathrm{free}}/\hbar}\varphi_{\mathrm{in}}$.

7.13.1 The Formal Feynman Path Integral for the Propagator Kernel

> The dynamics of a quantum system is described by a time-dependent operator called the Feynman propagator. The kernel of the propagator can be formally represented by a Feynman path integral which depends on the classical Hamiltonian (i.e., the symbol of the Hamiltonian operator). This is the first magic formula in quantum physics.
>
> Folklore

[122] The Schrödinger picture, the Heisenberg picture, and the Dirac (or interaction) picture are thoroughly discussed on page 393 of Vol. I.

Euler's polygon method. Set $t_k := t_0 + k\Delta t$, $k = 1, \ldots, N$ and $t_N := t$. This way, we get the decomposition

$$t_0 < t_1 < \ldots < t_{N-1} < t_N$$

of the time interval $[t_0, t]$. Let $b : \mathbb{R} \to \mathbb{R}$ be a given smooth function. We want to solve the ordinary differential equation

$$\psi'(t) = b(t)\psi(t), \qquad t \geq t_0, \qquad \psi(t_0) = \psi_0.$$

We are looking for a smooth solution $\psi : \mathbb{R} \to \mathbb{R}$. This uniquely determined solution is denoted by $\psi(t) = P(t, t_0)\psi_0$. Then

$$P(t, t_0)\psi_0 = P(t_N, t_{N-1}) \cdots P(t_2, t_1)P(t_1, t_0)\psi_0,$$

and $P_t(t, t_0)\psi_0 = b(t)P(t, t_0)\psi_0$ for all $\psi_0 \in \mathbb{R}$. Hence $P_t(t, t) = b(t)$. By Taylor expansion, linearization of the propagator yields

$$P(t_{k+1}, t_k) = P(t_k, t_k) + \Delta t \cdot P_t(t_k, t_k) + O((\Delta t)^2), \qquad \Delta t \to 0$$

with $P(t_k, t_k) = 1$ and $P_t(t_k, t_k) = b(t_k)$. Replacing the propagator by its linearization, we obtain the approximate solution

$$\psi_{\Delta t}(t) = (1 + b(t_{N-1}\Delta t)) \cdots (1 + b(t_1)\Delta t)(1 + b(t_0)\Delta t)\psi_0.$$

A standard result in numerical analysis tells us that this approximation method is convergent, that is,

$$\lim_{\Delta t \to 0} \psi_{\Delta t}(t) = \psi(t), \qquad t \geq t_0.$$

For example, fix the real number B, and set $b(t) := B$ for all t. Then we get the well-known classical formula for Euler's exponential function:

$$\lim_{\Delta t \to 0} (1 + B\Delta t)^N \psi_0 = \mathrm{e}^{B(t-t_0)}\psi_0, \tag{7.271}$$

which is valid for all times $t \in \mathbb{R}$ and all $\psi_0 \in \mathbb{R}$.

A general approximation principle for the propagators of time-depending processes. The argument above can be generalized to fairly general time-depending processes. For example, the limit (7.271) exists on a Banach space X for all $\psi_0 \in X$ if $B : X \to X$ is a linear bounded operator. More general functional-analytic results can be found in P. Lax, Functional Analysis, Sect. 34.3, Wiley, New York, 2002.[123] The situation is more subtle if B is an unbounded operator, as in quantum mechanics. In what follows, we will only use formal arguments.

From the propagator to the kernel. Let $\mathcal{K}$ be the kernel of the Feynman propagator operator $P(t, t_0) = \mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}$. Then the unique solution

$$\psi(t) = P(t, t_0)\psi_0$$

of the Schrödinger equation (7.262) on page 608 can be represented by the integral formula

$$\boxed{\psi(x, t) = \int_{\mathbb{R}} \mathcal{K}(x, t; y, t_0)\psi_0(y)dy, \qquad x \in \mathbb{R},\ t \geq t_0.}$$

It remains to compute the propagator kernel $\mathcal{K}$. Our goal is the key formula (7.274) below. The propagator possesses the linearization

[123] The proof uses the uniform boundedness theorem in functional analysis.

$$P(t_{k+1},t_k) = I - \frac{\mathrm{i}\Delta t}{\hbar}H + O((\Delta t))^2, \qquad \Delta t \to 0.$$

We set $P_{\Delta t}(t_{k+1},t_k) := I - \frac{\mathrm{i}\Delta t}{\hbar}H$. It follows from the causal product formula (7.264) on page 609 together with the approximation principle above that

$$P(t,t_0) = \lim_{\Delta t\to 0} P_{\Delta t}(t_N,t_{N-1})\cdots P_{\Delta t}(t_1,t_0).$$

Thus, we obtain

$$P(t,t_0) = \lim_{\Delta t\to 0}\left(I - \frac{\mathrm{i}\Delta t}{\hbar}H\right)^N.$$

The kernel product formula (7.245) on page 600 tells us that

$$\begin{aligned}\mathcal{K}(x,t;x_0,t_0) = \int_{\mathbb{R}^{N-1}} &\mathcal{K}(x,t;q_{N-1},t_{N-1})\times\cdots\\ &\times\mathcal{K}(q_2,t_2;q_1,t_1)\mathcal{K}(q_1,t_1;x_0,t_0)dq_{N-1}\cdots dq_2dq_1.\end{aligned} \tag{7.272}$$

From the kernel to the symbol. The Hamiltonian operator H has the symbol $a(q,p)$. Thus, the operator $P_{\Delta t}(t_{k+1},t_k)$ has the symbol $1-\frac{\mathrm{i}\Delta t}{\hbar}a(q,p)$. By the kernel formula (7.251) of the Weyl calculus on page 602, we obtain

$$\mathcal{K}_{\Delta t}(x,t_0+\Delta t;y,t_0) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}p(x-y)/\hbar}\ \left[1 - \frac{\mathrm{i}\Delta t}{\hbar}\ a\left(\frac{x+y}{2},p\right)\right]\frac{dp}{h}.$$

Up to terms of order $O(\Delta t)^2)$ as $\Delta t\to 0$, this yields

$$\mathcal{K}_{\Delta t}(x,t_0+\Delta t;y,t_0) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}p(x-y)/\hbar}\ \exp\left[-\frac{\mathrm{i}\Delta t}{\hbar}\ a\left(\frac{x+y}{2},p\right)\right]\frac{dp}{h}.$$

Since $t_{k+1} = t_k + \Delta t$, we also get the approximation $\mathcal{K}_{\Delta t}(q_{k+1},t_{k+1};q_k,t_k)$ being equal to

$$\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}p_{k+1}(q_{k+1}-q_k)/\hbar}\ \exp\left[-\frac{\mathrm{i}\Delta t}{\hbar}\ a\left(\frac{q_{k+1}+q_k}{2},p_{k+1}\right)\right]\frac{dp_{k+1}}{h}$$

where $k = 0,1,\ldots,N-1$.

The Feynman path integral. Using (7.272) and replacing $\mathcal{K}$ by $\mathcal{K}_\Delta$, we obtain the approximation

$$\mathcal{K}_{\Delta t}(x,t;y,t_0) = \int_{\mathbb{R}^{2N-1}} \mathrm{e}^{\mathrm{i}S_N/\hbar}\ \frac{dp_N}{h}\prod_{k=1}^{N-1}\frac{dq_kdp_k}{h}$$

with

$$\begin{aligned}S_N := \Big[&p_N\frac{q_N-q_{N-1}}{\Delta t} + \ldots + p_1\frac{q_1-q_0}{\Delta t}\\ &-a(\tfrac{1}{2}(q_N+q_{N-1}),p_N) + \ldots + a(\tfrac{1}{2}(q_1+q_0),p_1)\Big]\cdot\Delta t.\end{aligned}$$

Since the mid-point $\frac{1}{2}(q_k+q_{k-1})$ of the interval $[q_k,q_{k-1}]$ appears, we call this the mid-point approximation.

Now we pass over to the limit $\Delta t\to 0$ (i.e., $N\to\infty$) in a formal way. Let $S[q,p]$ denote the formal limit $\lim_{N\to\infty} S_N$. Then

$$S[q,p] = \int_{t_0}^{t} [p(\tau)\dot{q}(\tau) - a(q(\tau),p(\tau))]d\tau. \tag{7.273}$$

This is the action along the classical path $q = q(\tau), p = p(\tau)$ in the phase space on the time interval $t_0 \le \tau \le t$. Furthermore, we write the limit $\lim_{\Delta t \to 0} \mathcal{K}_{\Delta t}(x,t;y,t_0)$ in the following symbolic form:[124]

$$\boxed{\mathcal{K}(x,t;y,t_0) = \int_{\mathcal{C}\{t_0,t\}} \mathrm{e}^{\mathrm{i}S[q,p]/\hbar} \cdot \frac{dp(t_0)}{h} \prod_{t_0<\tau\le t} \frac{dq(\tau)dp(\tau)}{h}} \tag{7.274}$$

for all points $x, y \in \mathbb{R}$ and all time intervals $[t_0, t]$. Here, we formally sum over all continuous paths $q = q(\tau), p = p(\tau)$, $t_0 \le \tau \le t$, which satisfy the boundary condition

$$q(t_0) = y, \qquad q(t) = x.$$

The magic formula (7.274) relates classical mechanics to quantum mechanics by means of the classical action.

The crux with differentiable paths. The reader should note that the action $S[q,p]$ from (7.273) only makes sense if the path $q = q(t), p = p(t)$ is sufficiently smooth. However, our formal argument above also takes highly irregular paths into account, which are not differentiable at all. Such irregular paths are typical for the Brownian motion of tiny particles immersed in a liquid. In fact, in Wiener's theory of Brownian motion, the probability is equal to one for the realization of continuous, but not differentiable paths (see Sect. 7.11.4). Then the action $S[q,p]$ does not make any sense, in terms of classical analysis. This indicates that our formal approach is not well defined. Fortunately enough, it turns out that the main contribution to the Feynman path integral (7.274) comes from the paths which satisfy the classical equation of motion in mechanics. This is the main idea behind the WKB approximation method (see Sect. 7.10).

The symbol of the Feynman propagator. Let $\mathrm{sym}_P(q,p\,;t,t_0)$ denote the symbol of the propagator operator $P(t,t_0)$. By the kernel formula (7.252) of the Weyl calculus on page 602, we get

$$\mathrm{sym}_P(q,p\,;t,t_0) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}rp/\hbar}\, \mathcal{K}(q - \tfrac{1}{2}r, q + \tfrac{1}{2}r)dr \tag{7.275}$$

for all $q, p \in \mathbb{R}$ and all $t \ge t_0$. Recall that the propagator kernel $\mathcal{K}(x,t;y,t_0)$ can be represented by the Feynman path integral (7.274) above.

7.13.2 The Relation between the Scattering Kernel and the Propagator Kernel

> In perturbation theory, the scattering of free quantum particles under the action of a force is described by the Heisenberg scattering operator. The kernel of the scattering operator can be represented by the propagator kernel. This is the second magic formula in quantum physics.
>
> Folklore

Let $\mathcal{S}$ be the kernel of the propagator operator $\mathsf{S}(t,t_0)$. This means that the function $\varphi(t) := \mathsf{S}(t,t_0)\varphi_{\mathrm{in}}$ can be represented by the integral formula

[124] We also briefly write $\mathcal{K}(x,t;y,t_0) = \int_{\mathcal{C}\{t_0,t\}} \mathrm{e}^{\mathrm{i}S[q,p]/\hbar}\, \mathcal{D}q\mathcal{D}p$.

$$\varphi(x,t) = \int_{\mathbb{R}} \mathcal{S}(x,t;y,t_0)\varphi_{\rm in}(y)dy, \qquad x \in \mathbb{R},\ t \geq t_0.$$

This yields the transition amplitude

$$\boxed{\langle \varphi_{\rm out} | \mathsf{S}(t,t_0)\varphi_{\rm in}\rangle = \int_{\mathbb{R}^2} \varphi_{\rm out}(x)\mathcal{S}(x,t;y,t_0)\varphi_{\rm in}(y)dxdy,}$$

which generates the crucial transition probability $|\langle \varphi_{\rm out}|\mathsf{S}(t,t_0)\varphi_{\rm in}\rangle|^2$ from (7.266). It remains to compute the kernel $\mathcal{S}(x,t;y,t_0)$. Our goal is the key formula (7.276) below which relates the scattering kernel to the propagator kernel computed in the preceding section. The point is that there exists a simple relation between the symbol of the scattering operator and the symbol of the Feynman propagator operator. By the Weyl calculus, this implies the desired relation between the scattering kernel $\mathcal{S}(x,t;y,t_0)$ and the Feynman propagator kernel $\mathcal{K}(x,t;y,t_0)$.

The symbol of the scattering operator. Let $\mathrm{sym}_{\mathsf{S}}(q,p\,;t,t_0)$ denote the symbol of the scattering operator

$$\mathsf{S}(t,t_0) := \mathrm{e}^{\mathrm{i}tH_{\rm free}/\hbar}P(t,t_0)\mathrm{e}^{-\mathrm{i}t_0H_{\rm free}/\hbar}, \qquad t \geq t_0.$$

By the Weyl calculus, we have to replace this operator product by the Moyal star product for the corresponding symbols. Note that $\mathrm{e}^{-\mathrm{i}tp^2/2m\hbar}$ is the symbol of the free propagator $\mathrm{e}^{-\mathrm{i}tH_{\rm free}/\hbar}$. Hence

$$\mathrm{sym}_{\mathsf{S}}(q,p\,;t,t_0) = \mathrm{e}^{\mathrm{i}tp^2/2m\hbar} * \mathrm{sym}_P(q,p\,;t,t_0) * \mathrm{e}^{-\mathrm{i}t_0p^2/2m\hbar}.$$

Using formula (7.261) for the Moyal star product on page 607 together with the associativity of the Moyal star product, we obtain the key relation for the symbols:

$$\mathrm{sym}_{\mathsf{S}}(q,p\,;t,t_0) = \int_{\mathbb{R}^2} \mathcal{A}(q,p\,;t,t_0;q_1,p_1)\ \mathrm{sym}_P(q_1,p\,;t,t_0)\ dq_1dp_1. \tag{7.276}$$

Here, the kernel $\mathcal{A}(q,p\,;t,t_0;q_1,p_1)$ is given by the following formula:

$$\frac{1}{\pi\hbar}\exp\left[\frac{\mathrm{i}t(p_1-2p)^2}{2m\hbar} - \frac{\mathrm{i}t_0p_1^2}{2m\hbar} + \frac{2\mathrm{i}(q-q_1)(p-p_1)}{\hbar}\right].$$

The explicit computation of (7.276) will be performed in Problem 7.32. According to (7.275), the symbol sym_P is given by a Feynman path integral which depends on the classical action.

The kernel of the scattering operator. Finally, it follows from the kernel formula (7.251) of the Weyl calculus on page 602 that

$$\boxed{\mathcal{S}(x,t;y,t_0) = \frac{1}{2\pi\hbar}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}p(x-y)/\hbar}\ \mathrm{sym}_{\mathsf{S}}\left(\frac{x+y}{2},p\right)dp} \tag{7.277}$$

for all $x,y \in \mathbb{R}$ and all $t \geq t_0$. This is the magic formula for the kernel of the scattering operator.

7.14 The Poincaré–Wirtinger Calculus

> The Poincaré–Wirtinger calculus reformulates real analysis in terms of the language of complex analysis. This is very useful for modern quantum theory. Folklore

Let $f : \mathbb{R}^2 \to \mathbb{C}$ be a smooth complex-valued function on the real plane $\mathbb{R}^2$. We set

$$z := x + \mathrm{i}y, \quad \bar{z} := x - \mathrm{i}y,$$

and we write $f(x, y) := u(x, y) + \mathrm{i}v(x, y)$ where $u : \mathbb{R}^2 \to \mathbb{R}$ is the real part and $v : \mathbb{R}^2 \to \mathbb{R}$ is the imaginary part of f. The main idea of the Poincaré–Wirtinger calculus is to introduce the following two differential operators:

$$\boxed{\frac{\partial}{\partial z} := \frac{1}{2}\left(\frac{\partial}{\partial x} - \mathrm{i}\frac{\partial}{\partial y}\right), \quad \frac{\partial}{\partial \bar{z}} := \frac{1}{2}\left(\frac{\partial}{\partial x} + \mathrm{i}\frac{\partial}{\partial y}\right).} \tag{7.278}$$

This yields

$$\begin{aligned}\frac{\partial f(x,y)}{\partial z} &= \frac{1}{2}(u_x(x,y) + v_y(x,y)) + \frac{\mathrm{i}}{2}(v_x(x,y) - u_y(x,y)),\\ \frac{\partial f(x,y)}{\partial \bar{z}} &= \frac{1}{2}(u_x(x,y) - v_y(x,y)) + \frac{\mathrm{i}}{2}(v_x(x,y) + u_y(x,y)).\end{aligned}$$

Therefore, the following two conditions are equivalent:

(i) $\frac{\partial f}{\partial \bar{z}} = 0$ on $\mathbb{R}^2$.

(ii) $u_x = v_y$ and $u_y = -v_x$ on $\mathbb{R}^2$ (Cauchy–Riemann differential equations).

In this case, we say that the function f is holomorphic on $\mathbb{R}^2$. In terms of complex function theory, this means that the function $z \mapsto f(x, y)$ is holomorphic on the complex plane $\mathbb{C}$, in the classical sense. Similarly, the following two conditions are equivalent:

(i) $\frac{\partial f}{\partial z} = 0$ on $\mathbb{R}^2$.

(ii) $u_x = -v_y$ and $u_y = v_x$ on $\mathbb{R}^2$ (anti-Cauchy–Riemann differential equations).

In this case, we say that the function f is anti-holomorphic on $\mathbb{R}^2$. This is equivalent to the fact that the function $z \mapsto f(x, y)^\dagger$ is holomorphic on $\mathbb{C}$.

Example. (a) For $f(x, y) := x^2 + y^2$, we get

$$\frac{\partial f(x,y)}{\partial z} = x - \mathrm{i}y, \quad \frac{\partial f(x,y)}{\partial \bar{z}} = x + \mathrm{i}y.$$

The function f is neither holomorphic nor anti-holomorphic on $\mathbb{R}^2$.

(b) For $f(x, y) := (x + \mathrm{i}y)^2$, we get

$$\frac{\partial f(x,y)}{\partial z} = 2(x + \mathrm{i}y), \quad \frac{\partial f(x,y)}{\partial \bar{z}} = 0.$$

The function f is holomorphic on $\mathbb{R}^2$.

(c) For $f(x, y) := (x - \mathrm{i}y)^2$, we get

$$\frac{\partial f(x,y)}{\partial \bar{z}} = 2(x - \mathrm{i}y), \quad \frac{\partial f(x,y)}{\partial z} = 0.$$

The function f is anti-holomorphic on $\mathbb{R}^2$.

Mnemonic elegance. The results (a)–(c) above can be reformulated as follows.
(a) For $f(x,y) = x^2 + y^2 = z\bar{z}$, we get

$$\frac{\partial f(x,y)}{\partial z} = \bar{z} = x - \mathrm{i}y, \quad \frac{\partial f(x,y)}{\partial \bar{z}} = z = x + \mathrm{i}y.$$

(b) For $f(x,y) := (x+\mathrm{i}y)^2 = z^2$, we get

$$\frac{\partial f(x,y)}{\partial z} = 2z = 2(x+\mathrm{i}y), \quad \frac{\partial f(x,y)}{\partial \bar{z}} = 0.$$

(c) For $f(x,y) := (x-\mathrm{i}y)^2 = \bar{z}^2$, we get

$$\frac{\partial f(x,y)}{\partial \bar{z}} = 2\bar{z} = 2(x-\mathrm{i}y), \quad \frac{\partial f(x,y)}{\partial z} = 0.$$

These results are formally obtained by considering f as a function of the two independent variables z and $\bar{z}$ and by using formal partial differentiation with respect to z and $\bar{z}$.

For a general smooth function $f : \mathbb{R}^2 \to \mathbb{C}$, we proceed as follows. Using the representations $x = (z+\bar{z})/2$ and $y = (z-\bar{z})/2\mathrm{i}$, we define

$$F(z,\bar{z}) := f\left(\frac{z+\bar{z}}{2}, \frac{z-\bar{z}}{2\mathrm{i}}\right). \tag{7.279}$$

Considering formally the function F as a function of the independent variables z and $\bar{z}$, the chain rule tells us that

$$\begin{aligned}\frac{\partial F(z,\bar{z})}{\partial z} &= \frac{1}{2} f_x\left(\frac{z+\bar{z}}{2}, \frac{z-\bar{z}}{2\mathrm{i}}\right) + \frac{1}{2\mathrm{i}} f_y\left(\frac{z+\bar{z}}{2}, \frac{z-\bar{z}}{2\mathrm{i}}\right) \\ &= \frac{1}{2} f_x(x,y) - \frac{\mathrm{i}}{2} f_y(x,y),\end{aligned}$$

and

$$\frac{\partial F(z,\bar{z})}{\partial \bar{z}} = \frac{1}{2} f_x(x,y) + \frac{\mathrm{i}}{2} f_y(x,y).$$

This coincides with definition (7.278). The following observation is useful.

- The function f is holomorphic on $\mathbb{R}^2$ iff F is independent of $\bar{z}$ and $z \mapsto F(z)$ is holomorphic on $\mathbb{C}^2$. Then $\frac{\partial f(x,y)}{\partial z} = F'(z)$ for all $z = x + \mathrm{i}y$ on $\mathbb{C}$.
- The function f is anti-holomorphic on $\mathbb{R}^2$ iff F is independent of z and $\zeta \mapsto F(\zeta)$ is holomorphic on $\mathbb{C}^2$. Then $\frac{\partial f(x,y)}{\partial \bar{z}} = F'(\bar{z})$ for all $\bar{z} = x - \mathrm{i}y$ on the complex plane $\mathbb{C}$.

In later volumes, the Poincaré–Wirtinger calculus will play a crucial role in studying the following subjects: Kähler geometry, conformal field theory, and string theory.

7.15 Bargmann's Holomorphic Quantization

Our goal is to realize the commutation relation

$$\boxed{a^- a^+ - a^+ a^- = I} \tag{7.280}$$

together with $(a^-)^\dagger = a^+$ by elementary operators on a Hilbert space $B(\mathbb{C})$ of holomorphic functions. The precise formulation will be given in Theorem 7.58 below.

In terms of physics, the operator a^+ (resp. a^-) is a creation (resp. annihilation) operator.[125]

The Bargmann–Fock space $B(\mathbb{C})$**.** We start with the inner product

$$\langle F|G\rangle := \frac{1}{\pi}\int_{\mathbb{R}^2} F(z)^\dagger G(z)\mathrm{e}^{-zz^\dagger}\,dxdy. \tag{7.281}$$

By definition, the space $B(\mathbb{C})$ consists of all holomorphic functions $F:\mathbb{C}\to\mathbb{C}$ with $\langle F|F\rangle < \infty$. This is a complex Hilbert space with respect to the inner product (7.281). The set of polynomials $z\mapsto F(z)$ is a dense subset of $B(\mathbb{C})$. We define the operators $a^\pm : D(a^\pm)\to B(\mathbb{C})$ by setting

$$\boxed{(a^+F)(z) := zF(z) \qquad \text{for all} \quad z\in\mathbb{C},}$$

and

$$\boxed{(a^-F)(z) := \frac{d}{dz}F(z) \qquad \text{for all} \quad z\in\mathbb{C}.}$$

More precisely, the domain of definition $D(a^\pm)$ of the operator $a^\pm$ consists of all functions $F\in B(\mathbb{C})$ with $a^\pm F\in B(\mathbb{C})$. For example, this is satisfied for all polynomials F. Setting $F_0(z):=1$ for all $z\in\mathbb{C}$, we get $\langle F_0|F_0\rangle = 1$ and

$$\boxed{a^-F_0 = 0.}$$

In terms of physics, the function F_0 is called the ground state (or the vacuum state). This state does not contain any particles.

Theorem 7.58 (i) *For all polynomials $F\in B(\mathbb{C})$, we get*

$$(a^-a^+ - a^+a^-)F = F.$$

This is the precise formulation of the commutation relation (7.280) .

(ii) *For all polynomials $F,G\in B(\mathbb{C})$, we get*

$$\langle a^-F|G\rangle = \langle F|a^+G\rangle.$$

This means that $(a^-)^\dagger = a^+$, in the sense of a formally adjoint operator.

Proof. Ad (i). Note that $(a^+a^-F)(z) = zF'(z)$ and

$$(a^-a^+F)(z) = (zF(z))' = F(z) + zF'(z).$$

Ad (ii). We will use the Poincaré–Wirtinger calculus introduced on page 616. Recall that $\bar z := z^\dagger$. Since G is holomorphic, $\frac{\partial G(z)}{\partial \bar z} = 0$. By the product rule,

$$\frac{\partial}{\partial\bar z}\left(G(z)\mathrm{e}^{-z\bar z}\right) = \frac{\partial G(z)}{\partial\bar z}\,\mathrm{e}^{-z\bar z} - G(z)z\mathrm{e}^{-z\bar z} = -G(z)z\mathrm{e}^{-z\bar z}.$$

[125] The proofs can be found in the classical paper by V. Bargmann, On a Hilbert space of analytic functions and an associated integral transform, Commun. Pure and Appl. Math. **14** (1961), 187–214. See also the last chapter of the monograph by F. Berezin and M. Shubin, The Schrödinger Equation, Kluwer, Dordrecht, 1991.
The basic idea goes back to V. Fock, Generalizing and solving Dirac's statistical equation, Z. Phys. **49** (1928), 339–357 (in German).

Hence

$$\frac{\partial}{\partial\bar z}\left(F(z)^\dagger G(z)\mathrm{e}^{-z\bar z}\right) = \left(\frac{\partial}{\partial\bar z}F(z)^\dagger\right)G(z)\mathrm{e}^{-z\bar z} - F(z)^\dagger G(z)z\mathrm{e}^{-z\bar z}$$
$$= \left(\frac{\partial}{\partial z}F(z)\right)^\dagger G(z)\mathrm{e}^{-z\bar z} - F(z)^\dagger(zG(z))\mathrm{e}^{-z\bar z}.$$

Because of the polynomial growth of F and G at infinity, we get[126]

$$\lim_{R\to\infty}\int_{x^2+y^2\le R^2}\frac{\partial}{\partial\bar z}\left(F(z)^\dagger G(z)\mathrm{e}^{-z\bar z}\right)\,dxdy = 0.$$

This yields the claim. □

Orthonormal basis. The functions

$$F_n := \frac{(a^+)^n F_0}{\sqrt{n!}}, \qquad n = 0, 1, \ldots$$

form an orthonormal basis of the Hilbert space $B(\mathbb{C})$. Explicitly, $F_n(z) = z^n/\sqrt{n!}$ for all $z \in \mathbb{C}$. In terms of physics, F_n represents a (normalized) state of n particles. Intuitively, this state is generated from the vacuum state F_0 by n-fold application of the creation operator a^+ to F_0.

Wick operators. Let α_{kn} be complex numbers for $k, n = 0, \ldots, m$, where $m = 0, 1, \ldots$. For all polynomials $z \mapsto F(z)$, define

$$AF := \sum_{k,n=0}^{m}\alpha_{kn}(a^+)^k(a^-)^n F.$$

Note that the powers of the annihilation operator a^- stand on the right. In particular, $\alpha_{kn}(a^+)^k(a^-)^n F_0 = 0$ if $n = 1, 2, \ldots$. The operators A are called Wick operators on the Hilbert space $B(\mathbb{C})$ (or normally ordered operators). The polynomial

$$\mathrm{sym}_A(z,\zeta) := \sum_{k,n=0}^{m}\alpha_{kn}z^k\zeta^n$$

with respect to the complex variables z and ζ is called the symbol of the Wick operator A. As we will show later on, Wick operators play a crucial role for describing physical quantities in quantum field theory (e.g., collision processes) and in quantum statistics (e.g., superfluidity and superconductivity).

The Bargmann–Fock space $B(\mathbb{C}^s)$. We want to apply the preceding construction to s species of particles. To this end, we define $\langle F|G\rangle$ by

$$\frac{1}{\pi^s}\int_{\mathbb{R}^{2s}}F(z_1,\ldots,z_s)^\dagger G(z_1,\ldots,z_s)\mathrm{e}^{-\sum_{k=1}^s z_k z_k^\dagger}dx_1\cdots dx_s dy_1\cdots dy_s,$$

where $z_k := x_k + \mathrm{i}y_k$, and $x_k, y_k \in \mathbb{R}$ for all $k = 1, \ldots, s$. By definition, the Fock–Bargmann space $B(\mathbb{C}^s)$ consists of all holomorphic functions[127]

[126] Note the following. Since $\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - \mathrm{i}\frac{\partial}{\partial y}\right)$, the Gaussian integral theorem transforms this integral into a boundary integral (over the sphere of radius R), which goes to zero as $R \to \infty$.

[127] This means that $F : \mathbb{C}^s \to \mathbb{C}$ is a power series expansion (with complex coefficients) which is absolutely convergent for all complex numbers $z_1, \ldots, z_s$.

$$F : \mathbb{C}^s \to \mathbb{C}$$

with $\langle F|F\rangle < \infty$. The space $B(\mathbb{C}^s)$ is a complex Hilbert space equipped with the inner product $\langle F|G\rangle$. The set of polynomials $(z_1, \dots, z_s) \mapsto F(z_1, \dots, z_s)$ (with complex coefficients) is a dense subset of $B(\mathbb{C}^s)$. Let $k = 1, \dots, s$. For all polynomials F, define

$$(a_k^+ F)(z_1, \dots, z_s) := z_k F(z_1, \dots, z_s), \quad (a_k^- F)(z_1, \dots, z_s) := \frac{\partial F(z_1, \dots, z_s)}{\partial z_k},$$

where $z_1, \dots, z_s \in \mathbb{C}$. Then, for all polynomials F and all $j, k = 1, \dots, s$, we have the commutation relations

$$(a_j^- a_k^+ - a_k^+ a_j^-)F = \delta_{jk} F$$

together with $(a_j^- a_k^- - a_k^- a_j^-)F = (a_j^+ a_k^+ - a_k^+ a_j^+)F = 0$. We briefly write

$$\boxed{[a_j^-, a_k^+]_- = \delta_{jk} I, \qquad [a_j^-, a_k^-]_- = [a_j^+, a_k^+]_- = 0, \quad j, k = 1, \dots, s.}$$

Moreover, for all polynomials F, G, we have

$$\langle a_k^- F|G\rangle = \langle F|a_k^+ G\rangle, \qquad k = 1, \dots, s.$$

Hence $(a_k^-)^\dagger = a_k^+$ for $k = 1, \dots, s$ in the sense of formal adjoint operators.

In the monograph by L. Faddeev and A. Slavnov, Gauge Fields, Benjamin, Reading, Massachusetts, 1980, it is emphasized that the Feynman path integral based on Bargmann quantization is very convenient for studying the quantization of the Standard Model in particle physics (Faddev–Popov quantization of gauge theories). We will investigate this in Vol. V.

Application to the quantized harmonic oscillator. We want to show that the use of Bargmann's holomorphic quantization allows us immediately to obtain the energy spectrum of the quantized harmonic oscillator. Motivated by Sect. 7.3.1 on page 443, we use the Hamiltonian

$$H = \frac{P^2}{2m} + \frac{m\omega^2 Q^2}{2}$$

of the harmonic oscillator, and we set

$$Q := \frac{x_0}{\sqrt{2}}(a^+ + a^-), \qquad P := \frac{\mathrm{i}\hbar}{x_0\sqrt{2}}(a^+ - a^-)$$

with $x_0 := \sqrt{\frac{\hbar}{m\omega}}$. It follows from $a^- a^+ - a^+ a^- = I$ that $QP - PQ = \mathrm{i}\hbar I$ and

$$H = \hbar\omega(a^+ a^- + \tfrac{1}{2}).$$

Setting $F_n(z) := z^n$, we get $a^+ a^- F_n = z\frac{d}{dz}F_n = nF_n$. Therefore, introducing $E_n := \hbar\omega(n + \frac{1}{2})$, we obtain

$$HF_n = E_n F_n, \qquad n = 0, 1, \dots$$

7.16 The Stone–Von Neumann Uniqueness Theorem

> The name "Heisenberg commutation relation" is a bit of a misnomer; the relations were in fact first formulated in their modern form not by Heisenberg (1925), but by Born and Jordan (1925) and by Dirac (1925) in the one-dimensional case and in the "Dreimännerarbeit" (three-man work) by Born, Heisenberg and Jordan (1926) and by Dirac (1926) in the multi-dimensional case. However, it is true that they grew out of the original ground-breaking work of Heisenberg (1925), though one would have to examine Heisenberg's paper very carefully to find anything remotely suggesting the commutation relations.[128]
>
> Jonathan Rosenberg, 2004

In this chapter, we have based quantum mechanics on the Born–Heisenberg–Jordan commutation relation

$$\boxed{QP - PQ = \mathrm{i}\hbar I.} \tag{7.282}$$

We want to show that, in an appropriate sense, the construction of the theory is unique. That is, each realization of quantum mechanics is equivalent to Schrödinger's approach. This follows from the Stone–von Neumann uniqueness theorem below. Moreover, we want to show that this problem is closely related to the following mathematical topics: functional analysis (operator theory on Hilbert spaces), symplectic geometry, C^*-algebras, functors between categories (the Weyl quantization functor), Lie algebras (the Heisenberg algebra) and Lie groups (the Heisenberg group). The main trick is to replace (7.282) by the Weyl relation (7.283) below. This way, we circumvent the technical subtlety related to the fact that the operators Q and P are not defined on the total Hilbert space. The Weyl relation refers to the unitary operators $U(a) = \mathrm{e}^{\mathrm{i}aP/\hbar}$ and $V(b) = \mathrm{e}^{\mathrm{i}bQ}$ defined on the total Hilbert space. Here, a and b are real parameters. This exponentiation is an infinite-dimensional variant of the passage from Lie algebras to Lie groups.

7.16.1 The Prototype of the Weyl Relation

Prototype. Consider the motion of a quantum particle on the real line. Choose the real numbers a and b. For each wave function $\psi \in L_2(\mathbb{R})$, we define the translation operator

$$(U(a)\psi)(x) := \psi(x + a), \qquad x \in \mathbb{R},$$

and the phase operator

$$(V(b)\psi)(x) := \mathrm{e}^{\mathrm{i}bx}\psi(x), \qquad x \in \mathbb{R}.$$

Then, we have the so-called Weyl relation[129]

[128] See the references given on page 673ff. The fascinating (and surprising) discovery of the commutation relation by Born after reading Heisenberg's paper is described on page 64 of Vol. I.
J. Rosenberg, A selective history of the Stone–von Neumann Theorem. In: Operator algebras, quantization, and noncommutative geometry, Contemporary Mathematics **365**, pp. 123–158, Amer. Math. Soc., Providence, Rhode Island, 2004 (reprinted with permission).

[129] H. Weyl, Quantum mechanics and group theory, Z. Physik **46** (1928), 1–47 (in German).

$$\boxed{U(a)V(b) = \mathrm{e}^{\mathrm{i}ab}V(b)U(a) \qquad \text{for all} \quad a, b \in \mathbb{R}.} \tag{7.283}$$

In fact, $(U(a)V(b)\psi)(x) = U(a)(\mathrm{e}^{\mathrm{i}bx}\psi(x))$. This is equal to

$$\mathrm{e}^{\mathrm{i}b(x+a)}\psi(x+a) = \mathrm{e}^{\mathrm{i}ab}V(b)U(a)\psi(x).$$

Using the notion of strongly continuous one-parameter unitary group introduced on page 506, the following holds:

$\{U(a)\}_{a\in\mathbb{R}}$ and $\{V(b)\}_{b\in\mathbb{R}}$ are strongly continuous one-parameter unitary groups on the Hilbert space $L_2(\mathbb{R})$.

The Born–Heisenberg–Jordan commutation relation. Let the operators $Q : D(Q) \to L_2(\mathbb{R})$ and $P : D(P) \to L_2(\mathbb{R})$ be the self-adjoint position and momentum operator, respectively, introduced in Sect. 7.6.4 on page 518. Then, for any test function $\psi \in \mathcal{S}(\mathbb{R})$, we get

$$P\psi(x) = -\mathrm{i}\hbar\frac{d}{dx}\psi(x) = -\mathrm{i}\hbar\frac{d}{da}U(a)\psi(x)_{|a=0}$$

and

$$Q\psi(x) = x\psi(x) = -\mathrm{i}\frac{d}{db}V(b)\psi(x)_{|b=0}.$$

Differentiating successively the Weyl relation

$$(U(a)V(b) - \mathrm{e}^{\mathrm{i}ab}V(b)U(a))\psi(x) = 0$$

with respect to the parameter a at the point $a = 0$ and with respect to b at $b = 0$, we get

$$(QP - PQ)\psi = \mathrm{i}\hbar\psi. \tag{7.284}$$

Therefore, the Born–Heisenberg–Jordan commutation relation (7.284) can be regarded as the infinitesimal variant of the Weyl relation (7.283). In terms of the Stone theorem on page 506,

$$U(a) = \mathrm{e}^{\mathrm{i}aP/\hbar}, \qquad V(b) = \mathrm{e}^{\mathrm{i}bQ}, \qquad a, b \in \mathbb{R}.$$

The Heisenberg algebra $\mathcal{A}_{\mathrm{Heis}}$. Consider the linear operators

$$Q, P, \hbar\mathrm{i}I : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R}).$$

Explicitly, $(Q\psi)(x) := x\psi(x)$, $(P\psi)(x) := -\mathrm{i}\hbar\frac{d}{dx}\psi(x)$, and $\hbar\mathrm{i}I\psi(x) := \hbar\mathrm{i}\psi(x)$ for all $x \in \mathbb{R}$ and all $\psi \in \mathcal{S}(\mathbb{R})$. Set

$$\mathcal{A}_{\mathrm{Heis}} := \{aQ + bP + c\hbar\mathrm{i}I : a, b, c \in \mathbb{R}\}.$$

This is a 3-dimensional real Lie algebra with respect to the following Lie products[130]

$$[Q, P]_- = \hbar\mathrm{i}I, \quad [Q, \hbar\mathrm{i}I]_- = [P, \hbar\mathrm{i}I]_- = 0.$$

Trivially, we have $[Q, Q]_- = [P, P]_- = [\hbar\mathrm{i}I, \hbar\mathrm{i}I]_- = 0$. This Lie algebra is called the Heisenberg Lie algebra (or briefly the Heisenberg algebra).

The realization of the Heisenberg algebra as a matrix Lie algebra. Let us introduce the matrices

[130] Recall that $[A, B]_- := AB - BA$.

$$A := \begin{pmatrix} 0\ 1\ 0 \\ 0\ 0\ 0 \\ 0\ 0\ 0 \end{pmatrix}, \quad B := \begin{pmatrix} 0\ 0\ 0 \\ 0\ 0\ 1 \\ 0\ 0\ 0 \end{pmatrix}, \quad C := \begin{pmatrix} 0\ 0\ 1 \\ 0\ 0\ 0 \\ 0\ 0\ 0 \end{pmatrix}.$$

Then, we have the Lie products

$$[A, B]_- = C, \quad [A, C]_- = [B, C]_- = 0.$$

Consequently, the set of all matrices

$$aA + bB + cC = \begin{pmatrix} 0\ a\ c \\ 0\ 0\ b \\ 0\ 0\ 0 \end{pmatrix}, \qquad a, b, c \in \mathbb{R}$$

forms a real 3-dimensional Lie algebra denoted by $sut(3, \mathbb{R})$.

The Heisenberg Lie algebra $\mathcal{A}_{\text{Heis}}$ is isomorphic to the Lie algebra $sut(3)$.

This isomorphism is given by the map $aQ + bP + c\hbar iI \mapsto aA + bB + cC$. All the matrices

$$\begin{pmatrix} 1\ a\ c \\ 0\ 1\ b \\ 0\ 0\ 1 \end{pmatrix}, \qquad a, b, c \in \mathbb{R}$$

form a group (with respect to matrix multiplication). This Lie group is denoted by $SUT(3, \mathbb{R})$ (group of special upper triangular real (3×3)-matrices).

The Lie algebra of the Lie group $SUT(3, \mathbb{R})$ is equal to $sut(3, \mathbb{R})$.

For more details, we refer to both Sec. 7.6ff of Vol. I and to Baker (2002).

The universal enveloping algebra of the Heisenberg algebra. Again consider the operators $Q, P, \hbar iI : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$. Let $\mathcal{E}(\mathcal{A}_{\text{Heis}})$ denote the set of all polynomials in Q, P and $\hbar iI$ with complex coefficients. For example, the operator

$$a\hbar iI + bP^3 + cP^2Q + dQP$$

with complex coefficients a, b, c, d is an element of $\mathcal{E}(\mathcal{A}_{\text{Heis}})$.

- The set $\mathcal{E}(\mathcal{A}_{\text{Heis}})$ is a complex algebra (with respect to the sum and the product of operators).
- If $A, B \in \mathcal{A}_{\text{Heis}}$, then $A, B \in \mathcal{E}(\mathcal{A}_{\text{Heis}})$ and $[A, B]_- = AB - BA$.

That is, the Lie product $[.,.]_-$ on $\mathcal{A}_{\text{Heis}}$ can be represented by using the product on $\mathcal{E}(\mathcal{A}_{\text{Heis}})$. In terms of the general theory of Lie algebras, the algebra $\mathcal{E}(\mathcal{A}_{\text{Heis}})$ is called the universal enveloping algebra of the Heisenberg Lie algebra $\mathcal{A}_{\text{Heis}}$.

The Weyl system with respect to the symplectic form ω on the plane $\mathbb{R}^2$. For all $(a, b) \in \mathbb{R}^2$, define

$$\boxed{W(a, b) := \mathrm{e}^{-\frac{\mathrm{i}}{2}ab} U(a)V(b).} \tag{7.285}$$

Then, for all $(a, b), (c, d) \in \mathbb{R}^2$, the following hold:

(i) $W(a, b)W(b, c) = \mathrm{e}^{\frac{\mathrm{i}}{2}\omega(a,b;c,d)} W(a + c, b + d)$. Here, we set

$$\omega(a, b; c, d) := (a, b) \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Note that ω is the symplectic form on the plane $\mathbb{R}^2$.

(ii) $W(a,b)^\dagger = W(-a,-b)$ and $W(0,0) = I$.

The operator family $\{W(a,b)\}_{(a,b)\in\mathbb{R}^2}$ is called the Weyl system of $\mathbb{R}^2$. One checks easily that (7.283), (7.285) imply (i) and (ii). For example, we have

$$\begin{aligned} W(a,b)^\dagger &= \mathrm{e}^{\frac{\mathrm{i}}{2}ab} V(b)^\dagger U(a)^\dagger = \mathrm{e}^{\frac{\mathrm{i}}{2}ab} V(-b)U(-a) \\ &= \mathrm{e}^{-\frac{\mathrm{i}}{2}ab} U(-a)V(-b) = W(-a,-b). \end{aligned}$$

Conversely, if W is given, then we set $U(a) := W(a,0)$ and $V(b) := W(0,b)$. Then (i), (ii) imply (7.283), (7.285).

The Weyl algebra of the Hilbert space $L_2(\mathbb{R})$. The linear continuous operators $A : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ form a C^*-algebra $\mathcal{A}$. The C^*-subalgebra of $\mathcal{A}$ generated by $\{W(a,b) : (a,b) \in \mathbb{R}^2\}$ is called the Weyl algebra of the Hilbert space $L_2(\mathbb{R})$ (see page 628).

The Heisenberg group $\mathcal{G}_{\mathrm{Heis}}$**.** For all $a,b,\lambda \in \mathbb{R}$, modify the Weyl system by setting

$$H(a,b,\lambda) := \mathrm{e}^{\mathrm{i}\lambda} W(a,b), \qquad a,b,\lambda \in \mathbb{R}.$$

Then, for all $a,b,c,d,\lambda,\mu \in \mathbb{R}$, we have the product formula

$$H(a,b,\lambda)H(c,d,\mu) = H\big(a+c, b+d, \lambda+\mu+\tfrac{1}{2}\omega(a,b;c,d)\big).$$

This means that the set $\{H(a,b,\lambda)\}_{a,b,\lambda\in\mathbb{R}}$ forms a group $\mathcal{H}$. The space $\mathbb{R}^3$ is a group with respect to the product

$$(a,b,\lambda)(c,d,\mu) := (a+c, b+d, \lambda+\mu)$$

for all $(a,b,\lambda),(c,d,\mu) \in \mathbb{R}^3$. If we modify this product by setting

$$(a,b,\lambda)(c,d,\mu) := \big(a+c,\ b+d,\ \lambda+\mu+\tfrac{1}{2}\omega(a,b;c,d)\big),$$

then $\mathbb{R}^3$ becomes a group which is called the Heisenberg group $\mathcal{G}_{\mathrm{Heis}}$.[131] The additional term $\frac{1}{2}\omega(a,b;c,d)$ is called a twist. There exists a group epimorphism[132]

$$\chi : \mathcal{G}_{\mathrm{Heis}} \to \mathcal{H}$$

given by the map $(a,b,\lambda) \mapsto H(a,b,\lambda)$.

The Heisenberg group $\mathcal{G}_{\mathrm{Heis}}$ is a 3-dimensional Lie group whose Lie algebra is the Heisenberg algebra $\mathcal{A}_{\mathrm{Heis}}$.

Since the Heisenberg group $\mathcal{G}_{\mathrm{Heis}}$ is arcwise connected and simply connected, it represents the universal covering Lie group of the Heisenberg Lie algebra $\mathcal{A}_{\mathrm{Heis}}$. By the general theory of Lie groups, the universal covering Lie group $\mathcal{G}_{\mathrm{universal}}$ of a given Lie algebra $\mathcal{L}$ knows everything about all of the Lie groups $\mathcal{G}$ whose Lie algebra is equal to $\mathcal{L}$.

The Heisenberg group is isomorphic to the group $SUT(3,\mathbb{R})$.

[131] The definition of the Heisenberg group is not unique in the literature. The Heisenberg group is also called the Weyl group in the physical literature. In fact, the Heisenberg group never appears in the papers written by Heisenberg; this group was introduced by Weyl.

[132] A surjective (resp. injective) group morphism is called group epimorphism (resp. group monomorphism). The same is true for rings. The general definition of epimorphisms and monomorphisms in terms of category theory will be considered in Vol. IV.

This Lie group isomorphism is given by the map

$$\begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \mapsto (a, b, c - \tfrac{1}{2}ab).$$

Central extensions of groups. Let $\mathcal{G}$ be a group. The set of all elements A of $\mathcal{G}$ with the property that

$$AB = BA \qquad \text{for all} \quad B \in \mathcal{G}$$

is called the center $C(\mathcal{G})$ of the group $\mathcal{G}$. The center $C(\mathcal{G})$ is a normal subgroup of $\mathcal{G}$. The group $\mathcal{G}$ is said to be a central extension of the quotient group $\mathcal{G}/C(\mathcal{G})$.[133] For the Heisenberg group,

$$C(\mathcal{G}_{\text{Heis}}) := \{(0, 0, \lambda) : \lambda \in \mathbb{R}\}.$$

This center is isomorphic to the additive group $\mathbb{R}$. The map

$$(a, b, \lambda) \mapsto (a, b)$$

is a group morphism from the Heisenberg group $\mathcal{G}_{\text{Heis}}$ onto the additive group $\mathbb{R}^2$. The kernel of the unit element $(0, 0)$ of $\mathbb{R}^2$ is the center $C(\mathcal{G}_{\text{Heis}})$. Therefore, we have the group isomorphism

$$\mathcal{G}_{\text{Heis}}/C(\mathcal{G}_{\text{Heis}}) \simeq \mathbb{R}^2.$$

Consequently, the Heisenberg group $\mathcal{G}_{\text{Heis}}$ is a central extension of the additive group $\mathbb{R}^2$.

Central extensions of Lie algebras. Let X be a linear space. Introducing the Lie product $[A, B] := 0$ for all $A, B \in X$, we obtain the trivial Lie algebra X. In this sense, the linear spaces $\mathbb{R}^n$ $(n = 1, 2, \ldots)$ become trivial Lie algebras.

Let $\mathcal{L}$ be a Lie algebra. The set of all elements A of $\mathcal{L}$ with the property that

$$[A, B] = 0 \qquad \text{for all} \quad B \in \mathcal{L}$$

is called the center $C(\mathcal{L})$ of $\mathcal{L}$. The center $C(\mathcal{L})$ is an ideal of $\mathcal{L}$. The Lie algebra $\mathcal{L}$ is said to be a central extension of the quotient Lie algebra $\mathcal{L}/C(\mathcal{L})$.[134] For example, the center of the Heisenberg algebra is given by

$$C(\mathcal{A}_{\text{Heis}}) = \{c\hbar i I : c \in \mathbb{R}\}.$$

This center is isomorphic to $\mathbb{R}$. The map

$$aQ + bP + c\hbar i I \mapsto (a, b)$$

is a Lie algebra morphism from the Heisenberg algebra $\mathcal{A}_{\text{Heis}}$ onto the trivial Lie algebra $\mathbb{R}^2$. The kernel of the zero element $(0, 0)$ of $\mathbb{R}^2$ is the center $C(\mathcal{A}_{\text{Heis}})$. Therefore, we have the Lie algebra isomorphism

$$\mathcal{A}_{\text{Heis}}/C(\mathcal{A}_{\text{Heis}}) \simeq \mathbb{R}^2.$$

[133] More general, if H is a subgroup of $C(G)$, then the quotient group G/H is called a central extension of the group G by the group H.

[134] More general, if $\mathcal{J}$ is a subalgebra of $C(\mathcal{L})$, then the Lie algebra $\mathcal{L}$ is called a central extension of the Lie algebra $\mathcal{L}/\mathcal{J}$ by the Lie algebra $\mathcal{J}$.

Consequently, the Heisenberg algebra $\mathcal{A}_{\text{Heis}}$ is a central extension of the trivial Lie algebra $\mathbb{R}^2$.

Central extensions of Lie groups and Lie algebras play an important role in quantum physics (e.g., the Bargmann theorem on the lifting of projective quantum symmetries to unitary symmetries, and the Virasoro algebra in both conformal quantum field theory and string theory). As an introduction, we recommend M. Schottenloher, A Mathematical Introduction to Conformal Field Theory, Springer, Berlin, 1997.

7.16.2 The Main Theorem

Theorem 7.59 *Let $\{\mathcal{U}(a)\}_{a\in\mathbb{R}}$ and $\{\mathcal{V}(b)\}_{b\in\mathbb{R}}$ be strongly continuous one-parameter unitary groups on the complex separable non-trivial Hilbert space X.*[135] *Suppose that the Weyl relation*

$$\mathcal{U}(a)\mathcal{V}(b) = e^{\mathrm{i}ab}\mathcal{V}(b)\mathcal{U}(a) \qquad \text{for all} \quad a, b \in \mathbb{R}$$

is satisfied. Then the operators $\mathcal{U}(a)$ (resp. $\mathcal{V}(b)$) are unitarily equivalent to direct sums of translation (resp. phase) operators on $L_2(\mathbb{R})$.

More precisely, the following hold.

(i) Invariant subspaces: There exists a finite or countable family $X_1, X_2, \ldots$ of pairwise orthogonal, closed, linear subspaces of the Hilbert space X with the direct sum decomposition

$$X = \bigoplus_k X_k.$$

All of the spaces $X_1, X_2, \ldots$ are invariant under the operators $\mathcal{U}(a)$ and $\mathcal{V}(b)$.

(ii) Unitary equivalence: For all $a, b \in \mathbb{R}$, the operator $\mathcal{U}(a)$ (resp. $\mathcal{V}(b)$) is unitarily equivalent to the translation operator $U(a)$ (resp. the phase operator $V(b)$) on the Hilbert space $L_2(\mathbb{R})$ introduced on page 622. This means that there exist unitary operators $U_k : X_k \to L_2(\mathbb{R})$ such that, for all k, the following diagram is commutative:

$$\begin{array}{ccc} X_k & \xrightarrow{\mathcal{U}(a)} & X_k \\ {\scriptstyle U_k}\downarrow & & \downarrow{\scriptstyle U_k} \\ L_2(\mathbb{R}) & \xrightarrow{U(a)} & L_2(\mathbb{R}). \end{array}$$

The same is true if we replace $\mathcal{U}(a)$ and $U(a)$ by $\mathcal{V}(b)$ and $V(b)$, respectively.

The proof can be found in Putnam (1967), p. 65. Theorem 7.59 is called the Stone–von Neumann uniqueness theorem. This theorem was announced by Stone in 1930. The first proof was given by

J. von Neumann, On the uniqueness of the Schrödinger operators, Math. Ann. **104** (1931), 570–578 (in German).

[135] A Hilbert space X is called trivial iff $X = \{0\}$.

7.16.3 C^*-Algebras

> A crucial strategy in modern mathematical physics consists in using C^*-algebras.
>
> Folklore

The simplest case of a C^*-algebra is the set $\mathbb{C}$ of complex numbers equipped with the operations $z+w$, zw, $z^\dagger$, and $|z|$. Since $zw = wz$, this C^*-algebra is called commutative.

Let $n = 2, 3, \dots$ The prototype of a (noncommutative) C^*-algebra is the set of complex $(n \times n)$-matrices equipped with the operations $A+B$, AB, $A^\dagger$ (adjoint matrix), αA ($\alpha \in \mathbb{C}$), and $||A|| := \sqrt{\mathrm{tr}(AA^\dagger)}$ (norm). The unit matrix I is the unit element. Traditionally, instead of $A^\dagger$ and $z^\dagger$ we write A^* and z^*, respectively.

Definition of C^*-algebra. Let $\mathcal{A}$ be a complex associative algebra which is also a complex Banach space. In addition, suppose that there exists a map $A \mapsto A^*$ (called the $*$-map) such that the following hold for all $A, B \in \mathcal{A}$ and all complex numbers α, β:

(i) $A^* \in \mathcal{A}$ (adjoint element);
(ii) $(\alpha A + \beta B)^* = \alpha^\dagger A^* + \beta^\dagger B^*$ (the $*$-map is antilinear);
(iii) $(A^*)^* = A$ (the $*$-map is an involution);
(iv) $(AB)^* = B^* A^*$;
(iv) $||AB|| \leq ||A|| \cdot ||B||$;
(v) $||A^*|| = ||A||$ and $||A^* A|| = ||A||^2$.

Then $\mathcal{A}$ is called a C^*-algebra.

The C^*-algebra is called commutative iff $AB = BA$ for all $A, B \in \mathcal{A}$. Furthermore, the C^*-algebra is called unital iff there exists a unit element I of $\mathcal{A}$ with $||I|| = 1$.

C^*-subalgebra. A subset $\mathcal{S}$ of a C^*-algebra $\mathcal{A}$ is called a C^*-subalgebra of $\mathcal{A}$ iff it is a C^*-algebra with respect to the operations on $\mathcal{A}$.

If $\mathcal{S}$ is a subset of a C^*-algebra $\mathcal{A}$, then there exists a (uniquely determined) smallest C^*-subalgebra $\mathcal{B}$ of $\mathcal{A}$ which contains the set $\mathcal{S}$. Explicitly, $\mathcal{B}$ is the intersection of all C^*-subalgebras of $\mathcal{A}$ which contain the set $\mathcal{S}$. We say that $\mathcal{B}$ is generated by $\mathcal{S}$.

By definition, a C^*-ideal of the C^*-algebra $\mathcal{A}$ is a C^*-subalgebra $\mathcal{I}$ of $\mathcal{A}$ which has the additional property that $AB \in \mathcal{I}$ and $BA \in \mathcal{I}$ for all $A \in \mathcal{A}, B \in \mathcal{I}$.

Examples. (a) The function space $C(M)$. Let M be a nonempty compact separated topological space (e.g. $M = [0,1]$ or, more generally, M is a compact subset of $\mathbb{R}^n$, $n = 1, 2, \dots$). The space $C(M)$ of all continuous functions

$$f : M \to \mathbb{C}$$

is a unital C^*-algebra with respect to the norm

$$||f|| := \max_{x \in M} |f(x)|.$$

Moreover, we set $f^*(x) := f(x)^\dagger$ for all $x \in M$ (complex-conjugate function). The function $f(x) := 1$ for all $x \in M$ is the unit element of $C(M)$.

(b) The operator space $L(X, X)$. Let X be a complex Hilbert space. The space $L(X, X)$ of all linear continuous operators

$$A : X \to X$$

is a C^*-algebra with respect to the operator norm

$$||A|| := \sup_{||\psi||\leq 1} ||A\psi||$$

and $A^* := A^\dagger$ (adjoint operator). If $X \neq \{0\}$, then the C^*-algebra $L(X,X)$ is unital, where the unit operator I is the unit element.

(c) The Weyl algebra of the Hilbert space $L_2(\mathbb{R})$. Let $X := L_2(\mathbb{R})$. The smallest C^*-subalgebra of $L(X,X)$, which contains the Weyl operators $W(a,b)$, $a,b \in \mathbb{R}$, is called the Weyl algebra of $L_2(\mathbb{R})$. Explicitly, this is the closure of the set

$$\text{span}\{W(A,B) : a,b \in \mathbb{R}\}.$$

The closure is to be understood in the sense of the Banach space $L(X,X)$.

The Gelfand–Naimark theorem below shows that examples (a) and (b) above are typical for C^*-algebras and commutative C^*-algebras, respectively.

C^*-morphism. Let $\mathcal{A}$ and $\mathcal{B}$ be C^*-algebras. The map

$$\chi : \mathcal{A} \to \mathcal{B}$$

is called a C^*- morphism iff it respects the algebra structure, the $*$-operation, and the norm structure, that is, for all $A, B \in \mathcal{A}$ and all complex numbers α, β, we have

- $\chi(\alpha A + \beta B) = \alpha\chi(A) + \beta\chi(B)$,
- $\chi(AB) = \chi(A)\chi(B)$,
- $\chi(A)^* = \chi(A)$, and
- $||\chi(A)|| = ||A||$.

Bijective C^*-morphisms are called C^*-isomorphisms. Moreover, C^*-isomorphisms $\chi : \mathcal{A} \to \mathcal{A}$ from a C^*-algebra $\mathcal{A}$ onto itself, are called C^*-automorphisms.

The category of C^*-algebras. In order to describe the common features of mathematical structures, one uses categories in mathematics. A category consists of objects and morphisms.

- The objects of the category of C^*-algebras are the C^*-algebras,
- and the morphisms of the category of C^*-algebras are the C^*-morphisms.

The general setting of category theory will be investigated in Vol. IV on quantum mathematics.

The Gelfand–Naimark structure theorem. In 1943, Gelfand (born 1913) and Naimark (1909–1978) proved the following crucial result.

Theorem 7.60 (i) *Each C^*-algebra is C^*-isomorphic to some C^*-subalgebra of $L(X,X)$, where X is some Hilbert space.*

(ii) *Each commutative unital C^*-algebra $\mathcal{A}$ is C^*-isomorphic to a C^*-algebra $C(M)$ of continuous functions on some nonempty compact separated topological space M. Here, M is the space of maximal ideals of the algebra $\mathcal{A}$ equipped with an appropriate topology.*

The proof is based on the so-called Gelfand–Naimark–Segal (GNS) construction, which is basic for algebraic quantum field theory and quantum statistics.[136] This will be considered in Vol. IV. There, we will also show that (ii) above is crucial for the spectral theory of unitary and self-adjoint operators. For the proofs, we refer

[136] I. Gelfand, Normed rings of operators, Mat. Sbornik **9** (1941), 3–24 (in German).
I. Gelfand and M. Naimark, On the embedding of normed rings into the ring of operators in Hilbert space, Mat. Sbornik **12** (1943), 197–213.
I. Segal, Postulates for general quantum mechanics, Ann. Math. **48** (1947), 930–948.

to P. Kadison and J. Ringrose, Fundamentals of the Theory of Operator Algebras, Vols. 1, Academic Press, New York, 1983.

$*$-Algebras. Sometimes it is convenient to replace C^*-algebras by the weaker notion of $*$-algebra (star algebra). Here, all the properties of a C^*-algebra drop out which refer to the norm. For example, the complex associative algebra $\mathcal{A}$ is called a $*$-algebra iff conditions (i)–(iv) on page 627 are satisfied. Analogously, one obtains the following terms: $*$-subalgebra, $*$-morphism, $*$-isomorphism, $*$-automorphism, category of $*$-algebras.

7.16.4 Operator Ideals

In order to give the proof for special cases of Fermat's last theorem in number theory, Kummer (1810–1891) introduced so-called ideal numbers.[137] Generalizing this, Dedekind (1831–1916) created the theory of ideals in ring theory. The theory of operator ideals generalizes this to operator algebras.

Compact operators. Let $C : X \to X$ be a linear compact self-adjoint operator on the complex separable non-trivial Hilbert space X. Then the eigenvectors of C form a complete orthonormal system in X. Let $\lambda_1, \lambda_2, \ldots$ denote the eigenvalues of C. Then:

- The spectrum of C is a pure point spectrum.
- The operator C is called of trace class iff $\sum_n |\lambda_n| < \infty$. In this case, the trace $\mathrm{tr}(C) := \sum_n \lambda_n$ is finite. Operators of trace class are also called nuclear operators.
- The operator C is called a Hilbert–Schmidt operator iff $\sum_n \lambda_n^2 < \infty$.

Let us generalize this. The linear compact operator $A : X \to X$ is called a trace class (resp. Hilbert–Schmidt) operator iff $\mathrm{tr}(\sqrt{A^*A}) < \infty$ (resp. $\mathrm{tr}(A^*A) < \infty$).[138] Every trace class operator is a Hilbert–Schmidt operator. If the linear operator $A : X \to X$ is compact on the complex separable Hilbert space X, then there exist orthogonal systems $\varphi_1, \varphi_2, \ldots$ and $\psi_1, \psi_2, \ldots$ together with positive numbers $\mu_1, \mu_2, \ldots$ (called the singular values of the operator A) such that

$$A\varphi = \sum_n \mu_n \langle \varphi_n | \varphi \rangle \psi_n \qquad \text{for all} \quad \varphi_n \in X.$$

Explicitly, we choose a complete orthonormal system $\varphi_1, \varphi_2, \ldots$ of eigenvectors of the self-adjoint compact operator A^*A, that is, $A^*A\varphi_n = \lambda_n \varphi_n$. Here, $\lambda_n \geq 0$ for all n. Moreover, throw away the eigenvectors φ_m with $\lambda_m = 0$, and set $\mu_n := \sqrt{\lambda_n}$, as well as $\psi_n := \lambda_n^{-1} A\varphi_n$. Equivalently, we write

$$A = \sum_n \mu_n \psi_n \otimes \varphi_n .$$

C^*-operator ideals. Consider the C^*-algebra $L(X, X)$ of the linear continuous operators $A : X \to X$ on the complex separable Hilbert space X.

(i) The set of linear compact operators $A : X \to X$ forms a C^*-ideal of $L(X, X)$. This ideal is denoted by $\mathcal{I}_{\mathrm{compact}}$.

[137] The famous complete proof of Fermat's last theorem was given by Wiles in 1994 (see page 17 of the Prologue to Vol. I; we also refer to F. Diamond and J. Shurman, A First Course in Modular Forms, Springer, Berlin, 2005).

[138] Note that the operator A^*A is self-adjoint and compact, and its eigenvalues are nonnegative.

(ii) The trace-class operators $A : X \to X$ form a C^*-ideal of $L(X, X)$. This ideal is denoted by $\mathcal{I}_{\text{trace class}}$.
(iii) The Hilbert–Schmidt operators of $A : X \to X$ form a C^*-ideal of $L(X, X)$. This ideal is denoted by $\mathcal{I}_{\text{Hilbert}-\text{Schmidt}}$. We have the inclusions $\mathcal{I}_{\text{trace class}} \subseteq \mathcal{I}_{\text{Hilbert}-\text{Schmidt}} \subseteq \mathcal{I}_{\text{compact}} \subseteq L(X, X)$.
(iv) $A \in \mathcal{I}_{\text{trace class}}$ iff $A = BC$ with $B, C \in \mathcal{I}_{\text{Hilbert}-\text{Schmidt}}$.

For more details, we refer to M. Reed and B. Simon, Methods of Modern Mathematical Physics, Vol. I, Sects. VI.5ff, Academic Press, 1972, as well as to the monographs by R. Schatten, Norm Ideals of Completely Continuous Operators, Springer, Berlin, 1960, and by A. Pietsch (1978), (2007) quoted on page 601.

7.16.5 Symplectic Geometry and the Weyl Quantization Functor

Functors play a crucial role in modern mathematics and physics.
Folklore

Symplectic linear spaces. A symplectic linear space X is a real linear space equipped with a symplectic form

$$\omega : X \times X \to \mathbb{R}.$$

That is, for all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in X$ and all real numbers α, β, the following hold:

- $\omega(\mathbf{a}, \mathbf{b}) = -\omega(\mathbf{b}, \mathbf{a})$ (antisymmetry);
- $\omega(\alpha\mathbf{a} + \beta\mathbf{b}, \mathbf{c}) = \alpha\omega(\mathbf{a}, \mathbf{c}) + \beta\omega(\mathbf{b}, \mathbf{c})$ (bilinearity);
- $\omega(\mathbf{a}, \mathbf{v}) = 0$ for all $\mathbf{v} \in X$ implies $\mathbf{a} = 0$ (non-degeneracy).

For example, the space $\mathbb{R}^2$ is a symplectic linear space with respect to the symplectic form

$$\omega((a, b), (c, d)) := \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc, \quad (a, b), (c, d) \in \mathbb{R}^2. \tag{7.286}$$

Symplectic morphism. Let X and Y be symplectic linear spaces with the symplectic forms ω and μ, respectively. The map

$$\chi : X \to Y$$

is called a symplectic morphism iff it is linear and respects the symplectic forms, that is,

$$\mu(\chi(\mathbf{a}), \chi(\mathbf{b})) = \omega(\mathbf{a}, \mathbf{b}) \qquad \text{for all} \quad \mathbf{a}, \mathbf{b} \in X.$$

Bijective symplectic morphisms are called symplectic isomorphisms. Then the inverse map is also a symplectic morphism. For example, a symplectic isomorphism of the plane $\mathbb{R}^2$ onto itself (with respect to the symplectic form (7.286)) is a linear area-preserving map from $\mathbb{R}^2$ onto itself. The category of symplectic linear spaces is defined in the following way:

- The objects are symplectic linear spaces,
- and the morphisms are symplectic morphisms.

Weyl algebras. We want to generalize the Weyl algebra of the Hilbert space $L_2(\mathbb{R})$. Let X be a linear symplectic space, and let $\mathcal{A}$ be a C^*-algebra with unit element. The map

$$W : X \to \mathcal{A}$$

is called a Weyl map iff the following hold for all $\mathbf{a}, \mathbf{b} \in X$:

(i) $W(\mathbf{a})W(\mathbf{b}) = \mathrm{e}^{\frac{\mathrm{i}}{2}\omega(\mathbf{a},\mathbf{b})}W(\mathbf{a}+\mathbf{b})$;
(ii) $W(0) = I$;
(iii) $W(\mathbf{a})^* = W(-\mathbf{a})$.

The subset $\{W(\mathbf{a}) : \mathbf{a} \in X\}$ of the C^*-algebra $\mathcal{A}$ is called a Weyl system. The smallest C^*-algebra of $\mathcal{A}$ which contains a Weyl system is called a Weyl algebra $\mathcal{W}(X)$. This algebra is also called the CCR-algebra of the linear symplectic space X with the symplectic form ω. Here, 'CCR' stands for 'canonical commutation relation'.

The existence and uniqueness theorem for Weyl algebras. The following theorem generalizes the Stone–von Neumann uniqueness theorem.

Theorem 7.61 *For each symplectic linear space X, there exists a Weyl algebra $\mathcal{W}(X)$ which is unique, up to C^*-isomorphisms.*

Proof. (I) Existence. Let $l_2(X)$ denote the space of all functions $f : X \to \mathbb{C}$ with at most countable support[139] and the property that

$$\sum_{\mathbf{a}\in X} |f(\mathbf{a})|^2 < \infty.$$

The complex linear space $l_2(X)$ becomes a complex Hilbert space equipped with the inner product $\langle f|g\rangle := \sum_{\mathbf{a}\in X} f(\mathbf{a})^\dagger g(\mathbf{a})$.

Now let us choose the C^*-algebra $\mathcal{A}$ consisting of all the linear continuous operators $A : l_2(X) \to l_2(X)$. For all $f \in l_2(X)$, we define

$$(W(\mathbf{a})f)(\mathbf{b}) := \mathrm{e}^{-\frac{\mathrm{i}}{2}\omega(\mathbf{a},\mathbf{b})} f(\mathbf{a}+\mathbf{b}), \qquad \mathbf{a},\mathbf{b} \in X.$$

One checks directly that $W(\mathbf{a}) \in \mathcal{A}$ and that $W : X \to \mathcal{A}$ is a Weyl map. To finish the argument, let $\mathcal{W}(X)$ be the C^*-subalgebra of $\mathcal{A}$ generated by the set $\{W(\mathbf{a}) : \mathbf{a} \in X\}$.

(II) Uniqueness. See Bär et al (2007), p. 121 (see the reference on page 632). □

Theorem 7.62 *If $\sigma : X \to Y$ is a symplectic morphism between the symplectic linear spaces X and Y, then there exists a uniquely determined injective C^*-morphism $\mathcal{W}(\sigma) : \mathcal{W}(X) \to \mathcal{W}(Y)$ such that the following diagram is commutative:*

$$\begin{array}{ccc} X & \xrightarrow{\ \sigma\ } & Y \\ \downarrow & & \downarrow \\ \mathcal{W}(X) & \xrightarrow{\ \mathcal{W}(\sigma)\ } & \mathcal{W}(Y). \end{array}$$

For the proof, we refer to Bär et al. (2007), p. 122.

The Weyl quantization functor. We have the following two properties:

(F1) $\mathcal{W}(\tau \circ \sigma) = \mathcal{W}(\tau) \circ \mathcal{W}(\sigma)$;
(F2) $\mathcal{W}(\mathrm{id}) = \mathrm{id}$.

[139] This means that the function f vanishes outside an at most countable subset of the linear space X.

More precisely, this means the following. If

$$X \xrightarrow{\sigma} Y \xrightarrow{\tau} Z$$

is the composition of two symplectic morphisms σ and τ, then this is transformed into the composition

$$\mathcal{W}(X) \xrightarrow{\mathcal{W}(\sigma)} \mathcal{W}(Y) \xrightarrow{\mathcal{W}(\tau)} \mathcal{W}(Z)$$

of the corresponding C^*-morphisms $\mathcal{W}(\sigma)$ and $\mathcal{W}(\tau)$. Furthermore, the identical symplectomorphism

$$X \xrightarrow{\mathrm{id}} X$$

is transformed into the identical map

$$\mathcal{W}(X) \xrightarrow{\mathrm{id}} \mathcal{W}(X)$$

of the C^*-algebra $\mathcal{W}(X)$. In terms of mathematics, the situation (F1), (F2) above describes a functor $\mathcal{W}$ between the category of symplectic linear spaces and the category of C^*-algebras. This functor is called the Weyl quantization functor.

Perspectives. In general, functors between categories map objects to objects and morphisms to morphisms such that the two properties (F1), (F2) above are satisfied. Functors play a fundamental role in the modern theory of mathematical structures. Typical examples are:

- the homology functor which sends continuous maps between topological spaces to group morphisms between homology groups,
- and the de Rham cohomology cofunctor[140] which sends smooth maps between manifolds to group morphisms between cohomology groups (i.e. linear maps between real linear spaces).

This will be studied in later volumes.

It was discovered recently, that functors are the right tool in order to generalize Einstein's principle of general relativity (also called the covariance principle) to quantum field theories on curved space-times. This principle postulates that physics does not depend on the choice of observers. Roughly speaking, the basic idea is to assign C^*-algebras to the open subsets of globally hyperbolic space-time manifolds (realization of the Haag–Kastler axioms). The point is that the change of the space-time manifolds induces a natural change of the assigned C^*-algebras. Furthermore, two different quantization functors are related to each other by a natural transformation. We refer to:

R. Brunetti, K. Fredenhagen, and R. Verch, The generally covariant locality principle – a new paradigm for local quantum field theory, Commun. Math. Phys. **237** (2003), 31–68.

C. Bär, N. Ginoux, and F. Pfäffle, Wave Equations on Lorentzian Manifolds and Quantization, European Mathematical Society 2007.

J. Baez and J. Dolan, Categorification, Contemporary Mathematics **230** (1998), 1–36.

The monograph by Bär, Ginoux, and Pfäffle contains a detailed study of the initial-value problem for normally hyperbolic differential equations on globally hyperbolic manifolds, together with applications to quantum field theory. This sophisticated global theory due to Jacques Hadamard (1865–1963), Marcel Riesz (1886–1969) and Jean Leray (1906–1998) is based on modern differential geometry (the language of bundles) and the theory of distributions on manifolds. Distributions are needed in order to handle the strong singularities of the Green's functions.

[140] In contrast to the composition rule (F2) above, a cofunctor $\mathcal{F}$ is characterized by the reverse composition rule $\mathcal{F}(\tau \circ \sigma) = \mathcal{F}(\sigma) \circ \mathcal{F}(\tau)$.

7.17 A Glance at the Algebraic Approach to Quantum Physics

In this section, we want to discuss a few basic ideas about the algebraic approach to non-relativistic quantum physics. Further material can be found in the volumes to follow. The Haag–Kastler theory, that is, the relativistic approach based on local operator algebras will be studied in Vol. IV on quantum mathematics.

7.17.1 States and Observables

> The states and the observables are basic concepts in the description of a physical system and their description has undergone a drastic fundamental change in the transition from the classical theory to the quantum theory.[141]
>
> Huzihiro Araki, 1999

The prototypes of pure states and mixed states. Consider a complex non-trivial Hilbert space X.

(a) Pure state: Fix $\psi \in X$ with $||\psi|| = 1$. Define

$$\chi(A) := \langle \psi | A\psi \rangle \qquad \text{for all} \quad A \in L(X, X).$$

Then $\chi(I) = 1$, and $\chi(A^*A) = \langle \psi | A^*A\psi \rangle = \langle A\psi | A\psi \rangle \geq 0$. Moreover, we have

$$\chi(A)^\dagger = \langle A\psi | \psi \rangle = \langle \psi | A^*\psi \rangle = \chi(A^*).$$

We call the linear continuous functional $\chi : L(X, X) \to \mathbb{C}$ a vector state (or a pure state).

(b) Mixed state: Let $\psi_0, \psi_1, \ldots$ be a complete orthonormal system of the Hilbert space X, and let $p_0, p_1, \ldots$ be real numbers contained in the unit interval $[0, 1[$ such that $\sum_k p_k = 1$. Define[142]

$$\chi(A) := \sum_k p_k \langle \psi_k | A\psi_k \rangle \qquad \text{for all} \quad A \in L(X, X).$$

Again, $\chi(I) = 1$ and $\chi(A^*A) \geq 0$ together with $\chi(A)^\dagger = \chi(A^*)$ for all operators $A \in L(X, X)$. The linear continuous functional $\chi : L(X, X) \to \mathbb{C}$ is called a mixed state. In terms of physics, the pure state ψ_k is realized with the probability p_k.

(c) Dynamics. If $H : D(H) \to X$ is a linear self-adjoint Hamiltonian operator, then the dynamics of the initial state ψ_0 is given by $\psi(t) = U(t)\psi_0$ for all times $t \in \mathbb{R}$, where we set

$$U(t) := \mathrm{e}^{-\mathrm{i}tH/\hbar}, \qquad t \in \mathbb{R}.$$

Motivated by $\langle U(t)\psi_0 | AU(t)\psi_0 \rangle = \langle \psi_0 | U(t)^{-1} AU(t)\psi_0 \rangle$, we define the operator $\mathcal{U}_t : L(X, X) \to L(X, X)$ for all $t \in \mathbb{R}$ by setting

$$\mathcal{U}_t A := U(t)^{-1} AU(t).$$

[141] H. Araki, Mathematical Theory of Quantum Fields, Oxford University Press, 1999.

[142] Since $|\langle \psi_k | A\psi_k \rangle| \leq ||\psi_k|| \cdot ||A\psi_k|| \leq ||\psi_k|| \cdot ||A|| \cdot ||\psi_k|| \leq ||A||$, the series for $\chi(A)$ is convergent.

The operator $\mathcal{U}_t$ is a C^*-isomorphism from the C^*-algebra $L(X,X)$ onto itself. As usual, such C^*-isomorphisms are also called C^*-automorphisms. Define

$$\chi_t(A) := \chi(\mathcal{U}_t A) \qquad \text{for all} \quad A \in L(X,X), \;\; t \in \mathbb{R}.$$

Then the map $t \to \chi_t$ corresponds to the map $t \mapsto U(t)\psi_0$, which describes the time evolution of the state ψ_0.

The general definition in terms of C^*-algebras. Suppose that we are given a C^*-algebra $\mathcal{A}$ with unit element I.

(i) Observables: The self-adjoint elements A of $\mathcal{A}$ (i.e., $A^* = A$) are called observables. The C^*-algebra $\mathcal{A}$ is called the extended algebra of observables.[143]

(ii) States: The linear functionals $\chi : \mathcal{A} \to \mathbb{C}$ with the normalization condition $\chi(I) = 1$ and the positivity condition

$$\chi(A^*A) \geq 0 \qquad \text{for all} \quad A \in \mathcal{A}$$

are called states.[144] A state is called mixed iff there exist two different states χ_1 and χ_2 such that

$$\chi = \lambda\chi_1 + (1-\lambda)\chi_2 \quad \text{for some number } \lambda \in]0,1[.$$

Otherwise the state is called pure.

(iii) Measurements of an observable. Let A be an observable, and let χ be a state. The real number

$$\boxed{\bar{A} := \chi(A)}$$

is called the measured mean value of the observable A in the state χ. Similarly, the nonnegative number ΔA given by

$$(\Delta A)^2 := \chi((A-\bar{A})^2)$$

is the measured fluctuation of the observable A in the state χ.[145] If A and B are two observables, then the complex number

$$\gamma := \frac{\chi\big((A-\bar{A})(B-\bar{B})\big)}{\Delta A \, \Delta B}$$

is called the correlation coefficient in the state χ.[146]

(iv) Dynamics: By definition, a dynamics on $\mathcal{A}$ is a one-parameter group $\{\mathcal{U}_t\}_{t\in\mathbb{R}}$ of C^*-automorphisms of the algebra $\mathcal{A}$. Explicitly, this means that, for all times $t, s \in \mathbb{R}$, the map $\mathcal{U}_t : \mathcal{A} \to \mathcal{A}$ is a C^*-automorphism and

$$\mathcal{U}_{t+s} = \mathcal{U}_t\mathcal{U}_s, \qquad U_0 = \mathrm{id}.$$

This yields the time evolution $t \mapsto \chi_t$ of a state χ, namely, we define

$$\chi_t(A) := \chi(\mathcal{U}_t A)) \qquad \text{for all} \quad A \in \mathcal{A}, \;\; t \in \mathbb{R}.$$

[143] Note that the algebra $\mathcal{A}$ also contains elements which are not observables, in the sense of the definition given above.

[144] It can be shown that states are always continuous. We also have $\chi(A)^\dagger = \chi(A^*)$ for all $A \in \mathcal{A}$. In particular, if A is an observable, then $\chi(A)$ is real.

[145] Since $(A-\bar{A})^* = A - \bar{A}$, we get $\chi\big((A-\bar{A})^2\big) \geq 0$, by (ii).

[146] The Schwarz inequality for C^*-algebras tells us that $|\gamma| \leq 1$.

(v) Thermodynamical equilibrium states (KMS-states): Let $\beta := 1/kT$, where k is the Boltzmann constant, and T is the absolute temperature. By definition, the state χ is called a KMS-state of temperature T with respect to the dynamics $\{\mathcal{U}_t\}_{t\in\mathbb{R}}$ iff it satisfies the β-KMS condition

$$\chi(A\mathcal{U}_t(B)) = \chi(\mathcal{U}_{t-\mathrm{i}\beta\hbar}(B)A) \qquad \text{for all} \quad A, B \in \mathcal{A}, \quad t \in \mathbb{R}.$$

One of the main problems in thermodynamics is the characterization of thermodynamic equilibrium states. The C^*-algebra approach to thermodynamics is able to do this. The three letters KMS stand for the names of the physicists Kubo, Martin, and Schwinger. For the historical background, see the discussion on page 659.

The following is crucial for distinguishing between classical physics and quantum physics.

The passage from classical physics to quantum physics corresponds to the passage from commutative algebras to noncommutative algebras.

For example, this also corresponds to the passage from classical information to quantum information, which represents the theoretical framework for the intended construction of quantum computers in the future. This will be studied in Vol. IV. We refer to N. Nielsen and M. Chuang, Quantum Computation and Quantum Information, Cambridge University Press, 2001.

Orthogonal projections as fundamental observables related to questions posed by physical experiments. The simplest observables in $L(X, X)$ are orthogonal projection operators. Let us summarize elementary properties.[147] We set

$$\mathcal{A} := L(X, X).$$

We assume that X is a complex separable non-trivial Hilbert space. We will use such a language that later on we can replace $L(X, X)$ by a von Neumann algebra $\mathcal{A}$ which is a factor (see page 657). In the language of von Neumann algebras, the following properties of orthogonal projections will tell us that $L(X, X)$ is a von Neumann algebra (more precisely, a factor) of type I.

(i) Orthogonal projections: By definition, an orthogonal projection is an element of $\mathcal{A}$ with $P^* = P$ and $P^2 = P$. Let $\mathcal{P}(\mathcal{A})$ denote the set of all orthogonal projections in $\mathcal{A}$.
Geometrically, this means the following. For any $\psi \in X$, we have the decomposition

$$\psi = P\psi + (I - P)\psi$$

where $P\psi$ is contained in the closed linear subspace $P(X)$ of X, and $(I - P)\psi$ is contained in the orthogonal complement $P(X)^\perp$.[148] By the Pythagorean theorem,

$$||\psi||^2 = ||P\psi||^2 + ||(I - P)\psi||^2.$$

Hence if $P \neq 0$, then $||P|| = 1$. Conversely, let Y be a linear closed subspace Y of X. For given $\psi \in X$, the variational problem

$$||\psi - \varphi|| = \min!, \qquad \varphi \in Y$$

has a unique solution denoted by $P\psi$. Then $P : X \to Y$ is an orthogonal projection onto the subspace Y.

[147] For the missing proofs, we refer to Zeidler (1995a) (see the references on page 1049), and to F. Riesz and B. Nagy, Functional Analysis, Frederyck Ungar, New York, 1978.

[148] Recall that, for a subset L of the Hilbert space X, the orthogonal complement is given by $L^\perp := \{\psi \in X : \ \langle\psi|\varphi\rangle = 0 \ \text{ for all } \varphi \in L\}$.

(ii) Partial ordering on $\mathcal{P}(\mathcal{A})$: Let $Q, P \in \mathcal{P}(\mathcal{A})$. We write

$$Q \leq P \qquad \text{iff} \qquad PQ = Q.$$

This is a partial ordering on $\mathcal{P}(\mathcal{A})$. Geometrically, this is equivalent to the inclusion $Q(X) \subseteq P(X)$.

(iii) Expectation values: Let $P, Q \in \mathcal{P}(\mathcal{A})$. Then,
- $0 \leq \langle\psi|P\psi\rangle \leq ||\psi||^2$ for all $\psi \in X$,[149] and
- $Q \leq P$ iff $\langle\psi|Q\psi\rangle \leq \langle\psi|P\psi\rangle$ for all $\psi \in X$.

In terms of expectation values in physics, this means

$$\boxed{0 \leq \bar{P} \leq 1 \qquad \text{and} \qquad Q \leq P \Rightarrow \bar{Q} \leq \bar{P}.}$$

Here, we exclude the trivial case $X = \{0\}$.

(iv) Eigenvalues: Let $P \in \mathcal{P}(\mathcal{A})$ with $P \neq 0$. If

$$P\psi = \lambda\psi$$

with $||\psi|| = 1$, then either $\lambda = 1$ or $\lambda = 0$. The eigenspace to the eigenvalue $\lambda = 1$ (resp. $\lambda = 0$) is $P(X)$ (resp. the orthogonal complement $P(X)^\perp$). In terms of mathematical logic, we regard the observable P as a question and the eigenvalues $\lambda = 1$ (resp. $\lambda = 0$) correspond to the answers "yes" (resp. "no").

(v) Orthogonality: Let $P, Q \in \mathcal{P}(\mathcal{A})$. We say that P is orthogonal to Q iff $PQ = 0$. Geometrically, this means that $P(X)$ is orthogonal to $Q(X)$.

(vi) Partial isometry: The linear continuous operator $U : X \to X$ is called a partial isometry iff we have

$$||U\psi|| = ||\psi|| \qquad \text{for all} \quad \psi \in \ker(U)^\perp.$$

That is, if we use the orthogonal composition, $X = Y \oplus Y^\perp$ with respect to the subspace $Y := \ker(U)$, then $U = 0$ on Y, and $U : Y^\perp \to \text{im}(U)$ is an isometry. More precisely, it follows from the Fredholm alternative that $\text{im}(U) = \ker(U^*)^\perp$, and hence the operator

$$U : \ker(U)^\perp \to \ker(U^*)^\perp$$

is a unitary operator.

(vii) The Murray–von Neumann equivalence relation: Let $Q, P \in \mathcal{P}(\mathcal{A})$. We write

$$Q \sim P$$

iff there exists an operator $U \in \mathcal{A}$ with $Q = UU^*$ and $P = U^*U$. This is a equivalence relation.

Geometrically, this means that U is a partial isometry whose restriction

$$U : P(X) \to Q(X)$$

to the space $P(X)$ is a unitary operator onto the space $Q(X)$. Similarly, U^* is a partial isometry whose restriction $U^* : Q(X) \to P(X)$ is a unitary operator.[150]

If $Q, P \in \mathcal{P}(A)$, then

$$||Q - P|| < 1 \qquad \text{implies} \quad Q \sim P.$$

[149] Note that $\langle\psi|P\psi\rangle = \langle\psi|P^2\psi\rangle = \langle P\psi|P\psi\rangle = ||P\psi||^2 \leq ||\psi||^2$.

[150] To prove this, note that $||P\psi||^2 = \langle\psi|P\psi\rangle = \langle U\psi|U\psi\rangle = ||U\psi||^2$. Thus, we obtain $\ker(U) = P(X)^\perp$. Similarly, $\ker(U^*) = Q(X)^\perp$.

(viii) The trivial center: The algebra $\mathcal{A}$ has a trivial center. This means that if $B \in \mathcal{A}$ and $AB = BA$ for all $A \in \mathcal{A}$, then B is a multiple of the unit operator (i.e., $B = \lambda I$ for some complex number λ). This is called the Schur lemma. [151]

(ix) Invariant linear subspaces of X. The linear subspace Y of X is called invariant under $\mathcal{A}$ iff $A(Y) \subseteq Y$ for all $A \in \mathcal{A}$.

(x) The dimension function: Let $P, Q \in \mathcal{P}(\mathcal{A})$. Define $d(P) := \dim P(X)$. Then:
- $d(P) = d(Q)$ iff $P \sim Q$.
- If $P(X)$ is orthogonal to $Q(X)$, then $d(P + Q) = d(P) + d(Q)$.

In what follows, we will show that orthogonal projections play a crucial role concerning

- the Gleason theorem and
- the Murray–von Neumann classification of factors of von Neumann algebras (see page 657).

7.17.2 Gleason's Extension Theorem – the Main Theorem of Quantum Logic

> Among other features, the Gleason theorem means that:
> - Probabilities provide a tool for constructing the language of physics long before they can be considered as empirically meaningful quantities.
> - The expression for the probabilities first proposed by Max Born (in 1926) is an unavoidable part of an interpretation. If any probability should ever play a part in the theory, it can be only this one.
> - The density operator (introduced by von Neumann) is the basic notion one must associate with a quantum state and not simply a pure state represented by a wave function.[152]
>
> Roland Omnès, 1994

Let X be a complex separable Hilbert space. Let $\mathcal{P}$ denote the set of all orthogonal projections $P : X \to X$. By definition, a pre-state of the C^*-algebra $L(X, X)$ is a function $s : \mathcal{P} \to [0, 1]$ with the property that we have

$$s(P_1 + \ldots + P_n) = s(P_1) + \ldots + s(P_n)$$

for each finite family of orthogonal projections $P_1, \ldots, P_n \in \mathcal{P}$ with the additional property that $P_i(X)$ is orthogonal to $P_j(X)$ if $i \neq j$. Such a pre-state is also called a finitely additive measure on $\mathcal{P}$.

Theorem 7.63 *If the dimension of the Hilbert space X is 3 or larger, then each pre-state can be uniquely extended to a state of the C^*-algebra $L(X, X)$.*

Conversely, the restriction of any state on $L(X, X)$ to the space of orthogonal projections $\mathcal{P}$ is a pre-state.

From the philosophical point of view, roughly speaking, Gleason's theorem tells us the following: [153]

[151] Schur (1875–1941).

[152] R. Omnès, The Interpretation of Quantum Mechanics, Princeton University Press, Princeton, New Jersey, 1994. Reprinted by permission of Princeton University Press.

[153] J. von Neumann and G. Birkhoff, The logic of quantum mechanics, Ann. Math. **37** (1936), 823–843.
A. Gleason, Measures on the closed subspaces of a Hilbert space, J. Math. Mech. **6** (1957), 885–893.

A quantum state is completely determined by only knowing the answers to all of the possible yes/no questions.

The Gleason theorem was generalized to von Neumann algebras by Christensen and Yeadon.[154]

7.17.3 The Finite Standard Model in Statistical Physics as a Paradigm

The partition function knows all about the thermodynamic system. The Feynman path integral can be viewed as a generalized partition function.
Folklore

The mean value. Let us consider a physical system $\mathcal{S}$ which can be in the finite number of states

$$\mathsf{S}_1, ..., \mathsf{S}_M$$

with the probabilities $p_1, \ldots, p_M$, respectively. Suppose that the physical quantity A (e.g., energy) attains the value A_m in the state S_m. By definition, if we measure the physical quantity A of the system $\mathcal{S}$, then we get the mean value

$$\boxed{\chi(A) := \sum_{m=1}^{M} p_m A_m.} \tag{7.287}$$

We also write $\bar{A}$ instead of $\chi(A)$.

Fluctuations and correlations can be described by mean values.

In fact, the fluctuation $\Delta A \geq 0$ of the physical quantity A is defined by

$$(\Delta A)^2 = \overline{(A - \bar{A})^2} = \sum_{m=1}^{M} p_m (A - \bar{A})^2.$$

Obviously, $(\Delta A)^2 = \overline{A^2} - (\bar{A})^2$. For two physical quantities A and B, the correlation coefficient is defined by

$$\mathrm{cor}(A, B) := \frac{\overline{(A - \bar{A})(B - \bar{B})}}{\Delta A \cdot \Delta B} = \frac{\sum_{k=1}^{M} p_m (A - \bar{A})(B - \bar{B})}{\Delta A \cdot \Delta B}.$$

The fundamental quantity

$$\boxed{S := -k \sum_{k=1}^{M} p_m \ln p_m}$$

is called the entropy of the physical system $\mathcal{S}$. If $0 < p_1, \ldots, p_M < 1$, then $\bar{A}$ is called the mean value of the physical quantity A with respect to the mixed state $(\mathsf{S}_1, p_1; \ldots, \mathsf{S}_M, p_M)$. If $p_{m_0} = 1$ and $p_m = 0$ for all indices $m \neq m_0$, then $\bar{A}$ is called the mean value of A with respect to the pure state S_{m_0}.

The language of C^*-algebras. The set $\{\mathsf{S}_1, \ldots, \mathsf{S}_M\}$ is called the state space S. The set of functions

[154] S. Maeda, Probability measures on projections in von Neumann algebras, Rev. Math. Phys. **1** (1989), 235–290 (survey article).

$$A : \mathsf{S} \to \mathbb{C}$$

is denoted by $\mathcal{A}$. We write $A_m := A(\mathsf{S}_m)$, and $A^*(\mathsf{S}_m) := A(\mathsf{S}_m)^\dagger$. If $A, B \in \mathcal{A}$ and $\alpha, \beta \in \mathbb{C}$, then

$$\alpha A + \beta B, \quad AB, \quad A*$$

are also contained in $\mathcal{A}$. In addition, we introduce the norm

$$||A|| := \sup_{m=1,\ldots,M} |A(\mathsf{S}_m)|.$$

The set $\mathcal{A}$ is a commutative unital C^*-algebra which is called the extended algebra of observables.

- Observables: The real-valued functions in $\mathcal{A}$ (i.e., $A^* = A$) are called observables.
- States: Let $0 \le p_1, \ldots, p_M \le 1$ and $p_1 + \ldots + p_M = 1$. Define the function

  $$\chi : \mathcal{A} \to \mathbb{C}$$

  by the key relation (7.287) above. Precisely all such functions are called states of the C^*-algebra $\mathcal{A}$. These functions have the following positivity property $\chi(AA^*) = \sum_{m=1}^M p_m A_m A_m^\dagger \ge 0$.
- Mean value: The value $\chi(A)$ is called the mean value of the observable A in the state χ.

The grand canonical ensemble. Now let us consider special physical systems whose states are characterized by energy and particle number. This is typical for statistical physics. More precisely, assume that the physical system $\mathcal{S}$ can be in the finite number of states $\mathsf{S}_1, \ldots, \mathsf{S}_M$; each state S_m is characterized by the energy E_m and the particle number N_m. We will motivate below that it is reasonable to assume that the number

$$\boxed{p_m = \frac{\mathrm{e}^{(\mu N_m - E_m)/kT}}{\sum_{m=1}^M \mathrm{e}^{(\mu N_m - E_m)/kT}}, \qquad m = 1, \ldots, M} \tag{7.288}$$

is the probability for finding the physical system in the state S_m. Here, the parameter $T > 0$ is called the (absolute) temperature, the real parameter μ is called the chemical potential, and k is the Boltzmann constant. The mean energy $\bar{E}$ and the energy fluctuation $\Delta E \ge 0$ are given by

$$\bar{E} = \sum_{m=1}^M p_m E_m, \qquad (\Delta E)^2 = \sum_{m=1}^M p_m (E - \bar{E})^2.$$

Similarly, the mean particle number $\bar{N}$ and the particle number fluctuation $\Delta N \ge 0$ are given by

$$\bar{N} = \sum_{m=1}^M p_m N_m, \qquad (\Delta N)^2 = \sum_{m=1}^M p_m (N - \bar{N})^2.$$

Physical interpretation. The grand canonical ensemble describes a (large) many-particle system which is able to exchange energy and particles with its environment. However, we assume that this exchange is so weak that one can attribute a mean energy and a mean particle number to the system $\mathcal{S}$. Moreover, this exchange is governed by two macroscopic parameters, namely, the absolute temperature T and the chemical potential μ. This tells us that the many-particle system does not behave wildly, but regularly. Physicists say that the system is in *thermodynamic*

equilibrium. For example, the sun radiates photons into the universe at the fixed surface temperature of about 6000 K. The change of the particle number can be caused by chemical reactions. This motivates the designation 'chemical potential' for μ.

The special case where $\mu = 0$ corresponds to a fixed particle number (i.e., there are no chemical reactions or no particle exchange with the environment). The grand canonical ensemble with $\mu = 0$ is called canonical ensemble.

The importance of the partition function. The main trick of statistical physics is to introduce the function

$$\boxed{Z(T,\mu) := \sum_{m=1}^{M} \mathrm{e}^{(\mu N_m - E_m)/kT}} \tag{7.289}$$

which is called the partition function of the grand canonical ensemble. The following proposition tells us that

> *The knowledge of the partition function allows us to compute all of the crucial thermodynamic quantities in statistical physics.*

To this end, we introduce the so-called statistical potential

$$\Omega(T,\mu) := -kT \ln Z(T,\mu). \tag{7.290}$$

This function is also called the Gibbs potential. An elementary computation shows that the following relations hold for the partial derivatives of the statistical potential.

(i) Entropy: $S = -\Omega_T$.
(ii) Mean particle number: $\bar{N} = -\Omega_\mu$.
(iii) Particle number fluctuation: $(\Delta N)^2 = kT\bar{N}_\mu$.
(iv) Free energy: By definition, $F := \Omega + \mu\bar{N}$.
(v) Mean energy: $\bar{E} = F + TS$.[155]
(vi) Energy fluctuation: If the particle number is fixed (i.e., $\mu = 0$), then we obtain $(\Delta E)^2 = kT^2\bar{E}_T$.
(vii) Pressure: Suppose that the energies $E_1, \ldots, E_M$ and the particle numbers $N_1, \ldots, N_M$ depend on the volume V of the physical system. Then the statistical potential $\Omega(T,\mu,V)$ also depends on the volume V, and the pressure of the physical system is defined by $P := -\Omega_V$.

The reader should observe that the Feynman functional integral

$$Z = \int \mathrm{e}^{\mathrm{i}S[\psi]/\hbar} \mathcal{D}\psi$$

can be regarded as a (formal) continuous variant of the partition function.

7.17.4 Information, Entropy, and the Measure of Disorder

> Many-particle systems in nature are able to store information. This is equivalent to both the measure of disorder and the notion of entropy in physics. Folklore

[155] The mean energy is also called the inner energy.

Information and words. In order to get some information in daily life, it is useful to ask L questions which have to be answered by 'yes' or 'no'. Then the typical answer looks like

$$YN \dots NN. \tag{7.291}$$

This is a word of length L with the two letters Y (yes) and N (no). Intuitively, the minimal number L of questions measures information. For example, suppose we have n balls of different weight. We want to know the heaviest ball. Using a balance, if $n = 2$, then we need one experiment (question). If $n = 3$, then we need two experiments. Generally, it follows by induction that we need $n-1$ experiments for n balls in order to find out the heaviest ball. After knowing this, we gain the information $I = n - 1$.

Observe that in computers, we use words of the type (7.291) in order to transport information. It is our goal to generalize this simple approach to more general situations. Interestingly enough, it turns out that one has to use the methods of probability theory.

General definition of information. Let $M = 1, 2, \dots$. Consider a random experiment which has the possible M outcomes

$$O_1, O_2, \dots, O_M \tag{7.292}$$

where O_m appears with the probability p_m. Here, $0 \le p_1, p_2, \dots, p_M \le 1$ and $p_1 + p_2 + \dots + p_M = 1$. The nonnegative number

$$\boxed{I := -\sum_{m=1}^{M} p_m \log_2 p_m} \tag{7.293}$$

is called the information of the random experiment (7.292).[156] The unit of I is called bit. Moreover, 1 byte = 8 bits. Intuitively, we gain the information I after performing the random experiment and after knowing the outcome. For example, let us throw a coin L times. The outcome corresponds to a word of the form (7.291), where Y and N stand for head and tail, respectively. The number of words of type (7.291) is equal to 2^L. Thus, the probability for a single outcome of the random coin experiment is equal to

$$p_m = \frac{1}{2^L}, \qquad m = 1, \dots, 2^L.$$

After performing the coin experiment, we gain the information

$$I = -\sum_{m=1}^{2^L} p_m \log_2 p_m = \log_2 2^L = L.$$

This coincides with the intuitive information introduced above in terms of answering yes/no questions. The number 2^L is called the statistical weight of the event (7.291).

Suppose that we have $p_1 = 1$ and $p_2 = \dots = p_M = 0$. Then we know the outcome O_1 of our random experiment in advance. This means that we do not

[156] By convention, if $p_m := 0$ for some index m, then we set $p_m \log_2 p_m = 0$. Information theory was created by Claude Shannon (1916–2001) in his paper: A mathematical theory of communication, Bell System Techn. J. **27** (1948), 379–423; 623–656.

gain any information after knowing the outcome. In fact, by (7.293) we get $I = -\log_2 1 = 0$.

The genetic code. The DNA (desoxyribonucleic acid) encodes the genetic information. This is a double-stranded molecule held together by weak bounds between base pairs of nucleotides. The four nucleotides in DNA contain the bases: adenine (A), cytosine (C), guanine (G), and thymine (T). A single strand can be formally described by a word

$$AGCT \dots G \tag{7.294}$$

of length L with the four letters A, C, G, T. There are 4^L such words. Introducing the weight $p_m := 1/4^L$, the word (7.294) contains the information

$$I = -\sum_{m=1}^{4^L} p_m \log_2 p_m = \log_2 4^L = 2L.$$

In nature, base pairs are only formed between A and T and between C and G. Thus, the base sequence (7.294) of each single strand can be deduced from that of its partner. The crucial protein synthesis in a biological cell is encoded into the messenger RNA (ribonucleic acid). This can be formally described by a word

$$C_{m_1} C_{m_2} \dots C_{m_{\mathcal{L}}} \tag{7.295}$$

of length $\mathcal{L}$ with the twenty letters $C_1, C_2, \dots, C_{20}$. These letters are called codons. Each codon is a word of length 3 with the letters A, C, G, T. Consequently, there are $4^3 = 64$ codons. However, by redundance, only 20 codons are essential. This corresponds to the multiplicity of spectral lines in the spectroscopy of molecules. This analogy combined with supersymmetry can be used in order to model mathematically the redundance of codons.[157] The information encoded into the word (7.295) is equal to $I = \log_2 20^{\mathcal{L}} = \mathcal{L} \log_2 20$.

The properties of the information function. Let

$$\sigma_M := \{(p_1, \dots, p_M) : \ 0 \le p_1 + \dots + p_M \le 1, \ \ p_1 + \dots + p_M = 1\}$$

be an $(M-1)$-dimensional simplex in $\mathbb{R}^M$. This is the closed convex hull of the M extremal points (vertices) $(1, 0, \dots, 0), \dots, (0, \dots, 0, 1)$. The proof of the following statement will be given in Problem 7.37.

Proposition 7.64 *The function $I : \sigma_M \to \mathbb{R}$ given by (7.293) is continuous and concave.*[158] *The minimal value $I = 0$ is attained at the extremal points of σ_M. Furthermore, the maximal value $I = \log_2 M$ is attained at the point $p_k = \frac{1}{M}$ for all $k = 1, \dots, M$.*

Measure of disorder. Consider the following experiment. We are given N particles, and we want to distribute them into M boxes $\mathcal{B}_1, \dots, \mathcal{B}_M$. Each possible distribution can be described by the symbol

[157] See M. Forger and S. Sachse, Lie super-algebras and the multiplet structure of the genetic code, I. Codon representations, II. Branching rules, J. Math. Phys. **41** (2000), 5407–5422; 5423–5444.
F. Antonelli, L. Braggion, M. Forger, et al., Extending the search for symmetries in the genetic code, Intern. J. Modern Physics B **17** (2003), 3135–3204.

[158] Explicitly, $I(\lambda q + (1-\lambda)p) \ge \lambda I(q) + (1-\lambda) I(p)$ for all $q, p \in \sigma_M$ and $\lambda \in [0, 1]$.

$$N_1 N_2 \dots N_M \tag{7.296}$$

where N_m is the number of particles in the box $\mathcal{B}_m$. Then $N_1 + \dots + N_M = N$. Set $p_m := \frac{N_m}{N}$. By definition, the number

$$I := -\sum_{m=1}^{M} p_m \log_2 p_m$$

is called the measure of disorder of the distribution (7.296). In order to show that this definition is reasonable, consider the following special cases.

- By Prop. 7.64, $0 \le I \le \log_2 M$.
- If all of the particles are in the same box, say, $\mathcal{B}_1$, then we have $p_1 = 1$ and $p_2 = \dots = p_M = 0$ Hence $I = -p_1 \log_2 p_1 = 0$. This corresponds to minimal disorder.
- If each box contains the same number of particles, then $N_m = \frac{N}{M}$. Hence $p_m = \frac{1}{M}$ for $m = 1, \dots, M$. Therefore, $I = \log_2 M$. This corresponds to maximal disorder.

Entropy. For historical reasons, physicists replace the information I from (7.293) by the quantity

$$S = -k \sum_{m=1}^{M} p_m \ln p_m.$$

Here, we use the Boltzmann constant $k = 1.380 \cdot 10^{-23}$ J/K. This implies that the entropy S has the physical dimension (heat) energy per temperature (see Sect. 7.17.11 on page 654). Since $\ln p_m = \ln 2 \cdot \log_2 p_m$, the relation between entropy and information is given by

$$S = I \cdot k \ln 2.$$

Intuitively, the entropy S measures the disorder of a many-particle system in physics. We have $0 \le S \le k \ln M$. Recent astronomical observations show that our universe is expanding in an accelerated manner. This means that stars and black holes decay after a long time.[159] Hence the disorder of the universe increases, that is, the entropy increases. This was postulated by Clausius (1822–1888) in 1865. He called this the heat death of the universe.

Temperature and chemical potential as Lagrange multipliers. In order to motivate the grand canonical ensemble, let us study the following maximum problem:

$$S = -k \sum_{m=1}^{M} p_m \ln p_m = \max!, \qquad p \in C \tag{7.297}$$

with the unit cube $C := \{(p_1, \dots, p_M\} : 0 \le p_1, \dots, p_M \le 1\}$ and the constraints

$$\bar{E} = \sum_{m=1}^{M} p_m E_m, \quad \bar{N} = \sum_{m=1}^{M} p_m N_m, \quad p_1 + \dots + p_M = 1. \tag{7.298}$$

Let $M \ge 2$. We are given the positive numbers $E_1, \dots E_M, N_1, \dots, N_M$ and $\bar{E}, \bar{N}$. We are looking for a solution $(p_1, \dots, p_M)$.

[159] F. Adams and G. Laughlin, A dying universe: the long-term fate and evolution of astrophysical objects, Rev. Mod. Phys. **69** (1997), 337–372.
F. Adams and G. Laughlin, The Five Ages of the Universe: Inside the Physics of Eternity, Simon and Schuster, New York, 1999.

Theorem 7.65 *Consider* $(p_1, \ldots, p_M)$ *given by* (7.288) *on page 639. Suppose that the real parameter* μ *and the positive parameter* T *are fixed in such a way that the constraints* (7.298) *are satisfied. In addition, assume that* $0 < p_1, \ldots, p_M < 1$ *and that the matrix*

$$\begin{pmatrix} E_1 & \ldots & E_M \\ N_1 & \ldots & N_M \\ 1 & \ldots & 1 \end{pmatrix}$$

has rank three. Then $(p_1, \ldots, p_M)$ *is the unique solution of the maximum problem* (7.297), (7.298).

Proof. (I) Local existence. We will use the sufficient solvability condition for the Lagrangian multiplier rule (see Prop. 43.23 of Zeidler (1986), Vol. III (see the references on page 1049). To this end, set

$$L := S + \alpha \left(\bar{E} - \sum_m p_m E_m \right) + \beta \left(\bar{N} - \sum_m p_m N_m \right) + \gamma \left(1 - \sum_m p_m \right).$$

That is, we add the constraints (7.298) to the function S which has to be maximized. The real numbers α, β, γ (called Lagrange multipliers) will be chosen below. For the partial derivatives, we get

$$L_{p_m} = -k \ln p_m - k - \alpha E_m - \beta N_m - \gamma,$$

and

$$L_{p_j p_m} = -\frac{k \delta_{jm}}{p_m}.$$

By (7.288), we choose μ, T and $(p_1, \ldots, p_M)$ in such a way that the constraints (7.298) are satisfied. Moreover, we set

$$\alpha := \frac{1}{T}, \quad \beta := -\frac{\mu}{T}, \quad \gamma := -k + k \ln \sum_m \mathrm{e}^{(\mu N_m - E_m)/kT}.$$

Then $L_{p_m} = 0$ for all m, and the matrix $(-L_{p_j p_m})$ is positive definite. This guarantees that our choice $(p_1, \ldots, p_M)$ represents a local maximum of the entropy function S under the constraints (7.298).

(II) Global existence. Since the entropy function S is concave, each local maximum of S on a convex set is always a global maximum. (We refer to Prop. 42.3 of Zeidler (1986), Vol. III (see the references on page 1049), and note that $-S$ is convex.)

(III) Uniqueness. On the boundary of the cube C, the entropy function S vanishes. Therefore, any solution of (7.297), (7.298) lies in the interior of C. Since the matrix $(-S_{p_j p_m})$ is positive definite, the function S is strictly concave on the interior of C. This implies the uniqueness of the solution (see Theorem 38.C. of Zeidler (1986), Vol. III). □

In the special case where the particle numbers are fixed, we use the choice $N_1 = \ldots = N_M = \bar{N}$, and $\mu = 0$. Then we have merely to assume that the matrix

$$\begin{pmatrix} E_1 & \ldots & E_M \\ 1 & \ldots & 1 \end{pmatrix}$$

has rank two, that is, there exist at least two different energies.

7.17.5 Semiclassical Statistical Physics

> In semiclassical statistical physics, the extended algebra of observables is a commutative $*$-algebra of functions, and the states are generated by some probability measure.
>
> Folklore

The key relation reads as

$$\boxed{\bar{A} := \int_M A(q,p)\varrho(q,p)\frac{dqdp}{h}.}$$

Here, we use the product set $M := \mathcal{B} \times \mathbb{R}$, where $\mathcal{B}$ is a closed interval on the real line. We are given the bounded continuous function $A : M \to \mathbb{C}$ and the bounded continuous function $\varrho : M \to [0, \infty[$ with the normalization condition

$$\int_M \varrho(q,p)\frac{dqdp}{h} = 1.$$

Then the function ϱ represents a probability density on the phase space M, and $\bar{A}$ is the mean value of the function $A = A(q,p)$. Traditionally, this function is called a (physical) observable iff it is real-valued.[160] The square of the mean fluctuation is given by

$$(\Delta A)^2 = \int_M (A(q,p) - \bar{A})^2 \varrho(q,p)\frac{dqdp}{h}.$$

In terms of physics, we consider an ideal gas[161] on the interval $\mathcal{B}$, that is, the position coordinate q of a single gas particle lives on the interval $\mathcal{B}$, and the momentum coordinate p lives on the real line $\mathbb{R}$. If $H = H(q,p)$ is the Hamiltonian function of a single gas particle, then we choose the function

$$\varrho(q,p) := \frac{\mathrm{e}^{-H(q,p)/kT}}{\int_M \mathrm{e}^{-H(q,p)/kT}\frac{dqdp}{h}}.$$

This function generates the semiclassical Gibbs statistics.[162] Here, T is the absolute temperature, k is the Boltzmann constant, and h is Planck's quantum of action. For example, if the gas particles behave like harmonic oscillators, then we choose $H(q,p) = \frac{p^2}{2m} + \frac{m\omega^2 q^2}{2}$. We need the physical constants k and h in order to guarantee that both the quantities $\frac{H(q,p)}{kT}$ and $\frac{dqdp}{h}$ are dimensionless. This implies that the function $\mathrm{e}^{-H(q,p)/kT}$ makes sense, the probability density ϱ is dimensionless, and the mean value $\bar{A}$ has the same dimension as the physical observable $A(q,p)$. For example, if we choose $A(q,p) := H(q,p)$, then $N\bar{H}$ is the mean energy of the ideal gas at the temperature T, where N is the number of gas particles. The function

$$S(q,p) = -k\varrho(q,p)\ln\varrho(q,p)$$

[160] Note that the algebra of observables to be introduced below is not only based on real-valued functions, but on complex-valued functions in order to get a complex $*$-algebra.

[161] An ideal gas is characterized by the property that there are no interactions between the gas particles, that is, the single gas particles behave like independent random objects.

[162] Gibbs (1839–1903).

corresponds to the entropy, and $N\bar{S}$ is the entropy of the ideal gas at the temperature T. If C is a compact subset of the phase space M, then the integral

$$\int_C \varrho(q,p)\frac{dqdp}{h}$$

is the probability for finding the position-momentum coordinate (q,p) of a single gas particle in the set C. Let us translate this into the language of $*$-algebras.

The extended $*$-algebra $\mathcal{A}$ of observables. Let $\mathcal{A}$ denote the set of all bounded continuous functions $A : M \to \mathbb{C}$. With respect to the star operation $A^*(q,p) := A(q,p)^\dagger$ for all $(q,p) \in M$, the set $\mathcal{A}$ is a commutative $*$-algebra with unit element $\mathbf{1}$.[163] The $*$-algebra $\mathcal{A}$ is called the extended $*$-algebra of observables (of the gas). Precisely the real-valued functions A in $\mathcal{A}$ are called observables. In addition, equipped with the norm

$$||A|| := \sup_{(q,p)\in M} |A(q,p)|,$$

the $*$-algebra $\mathcal{A}$ becomes a normed space with

- $||A^*|| = ||A||$ and $||A^*A|| = ||A||^2$ for all $A \in \mathcal{A}$;
- $||\mathbf{1}|| = 1$.

Since the phase space M is an unbounded closed subset of $\mathbb{R}^2$ (i.e., M is not compact), the normed space $\mathcal{A}$ is not a Banach space. We call $\mathcal{A}$ an incomplete C^*-algebra (or a pre-C^*-algebra).

States. Generally, states are functionals χ which assign a real number $\chi(A)$ to each observable A. We define

$$\chi(A) := \int_M A(q,p)\varrho(q,p)\frac{dqdp}{h} \qquad \text{for all} \quad A \in \mathcal{A}.$$

Then, for all $A \in \mathcal{A}$, we have:

- $\chi(A^*A) = \int_M A(q,p)^\dagger A(q,p)\varrho(q,p)\frac{dqdp}{h} \geq 0$;
- $\chi(I) = \int_M 1 \cdot \varrho(q,p)\frac{dqdp}{h} = 1$.
- The map $\chi : \mathcal{A} \to \mathbb{C}$ is linear.
- $|\chi(A)| \leq \sup_{(q,p)\in\mathbb{R}^2} |A(q,p)| = ||A||$.

We call χ a state on the $*$-algebra $\mathcal{A}$. This state corresponds to the probability measure ν generated by the probability density ϱ (i.e., $d\nu = \varrho\frac{dqdp}{h}$).

Dynamics. To avoid technicalities, choose $J := \mathbb{R}$, that is, $M = \mathbb{R}^2$. Motivated by the classical equation of motion

$$\dot{q}(t) = H_p(q(t),p(t)), \qquad \dot{p}(t) = -H_q(q(t),p(t)), \qquad t \in \mathbb{R} \tag{7.299}$$

with the initial condition $q(0) = q_0, p(0) = p_0$, we define

$$(\mathcal{U}_t A)(q_0,p_0) := A(q(t),p(t))$$

for all times $t \in \mathbb{R}$ and all initial points $(q_0,p_0) \in \mathbb{R}^2$. We assume that, as for the harmonic oscillator, the trajectories $q = q(t), p = p(t)$ exist for all times. Then, for each time $t \in \mathbb{R}$, the map

$$\mathcal{U}_t : \mathcal{A} \to \mathcal{A}$$

[163] Here, $\mathbf{1}$ is given by the function $A(q,p) \equiv 1$.

is a $*$-automorphism. Thus, $\{\mathcal{U}_t\}_{t\in\mathbb{R}}$ is a one-parameter group of $*$-automorphisms of the $*$-algebra $\mathcal{A}$.

Our next goal is to prove that the dynamics of the gas corresponds to a family $\{U_t\}_{t\in\mathbb{R}}$ of unitary operators U_t on the Hilbert space $L_2(\mathbb{R}^2)$. To this end, let $\mathcal{A}_{\rm pre}$ denote the set of all smooth functions $A:\mathbb{R}^2\to\mathbb{C}$ with compact support. Obviously, $\mathcal{A}_{\rm pre}$ is a $*$-subalgebra of the $*$-algebra $\mathcal{A}$ of observables. In addition, $\mathcal{A}_{\rm pre}$ is a dense subset of the Hilbert space $L_2(\mathbb{R}^2)$ equipped with the inner product $\langle A|B\rangle := \int_{\mathbb{R}^2} A(q,p)^\dagger B(q,p)dqdp$.

Proposition 7.66 *Let $A, B\in\mathcal{A}_{\rm pre}$. Then $\langle \mathcal{U}_t A|\mathcal{U}_t B\rangle = \langle A|B\rangle$ for all $t\in\mathbb{R}$.*

This tells us that the dynamics of the gas respects the inner product on the Hilbert space $L_2(\mathbb{R}^2)$. Using this result and the extension theorem from Problem 7.21, we get the following.

Corollary 7.67 *For any time $t\in\mathbb{R}$, the operator $\mathcal{U}_t:\mathcal{A}_{\rm pre}\to\mathcal{A}$ can be uniquely extended to a unitary operator $U_t: L_2(\mathbb{R})\to L_2(\mathbb{R})$.*

It remains to prove Prop. 7.66. Using the equation (7.299) of motion, we get

$$\frac{d}{dt}(\mathcal{U}_t A)(q_0,p_0) = A_q(q(t),p(t))H_p(q(t),p(t)) - A_p(q(t),p(t))H_q(q(t),p(t)).$$

Noting that $H_{qp}=H_{pq}$, integration by parts yields

$$\int_{\mathbb{R}^2}(A_q^\dagger H_p - A_p^\dagger H_q)Bdqdp = -\int_{\mathbb{R}^2} A^\dagger(B_q H_p - B_p H_q)dqdp.$$

This implies $\frac{d}{dt}\langle \mathcal{U}_t A|\mathcal{U}_t B\rangle = \langle \frac{d}{dt}\mathcal{U}_t A|\mathcal{U}_t B\rangle + \langle \mathcal{U}_t A|\frac{d}{dt}\mathcal{U}_t B\rangle = 0$. □

Generalization. The simple special case considered above can be generalized to $2s$-dimensional phase space manifolds M by starting from the key formula

$$\bar{A} := \int_M A(q,p)d\nu(q,p)$$

with $\int_M d\nu = 1$. Here, $(q,p)=(q_1,\dots,q_s;p_1,\dots,p_s)$. As a rule, the Hamiltonian $H=H(q,p)$ describes interactions between the particles; this corresponds to so-called real gases.

For example, consider a gas consisting of N molecules in a box $\mathcal{B}$ of finite volume V in the 3-dimensional space. Then $s=3N$, and $M=\mathcal{B}^N\times\mathbb{R}^{3N}$. Moreover,

$$d\nu := \varrho(q,p)\frac{dq^{3N}dp^{3N}}{h^{3N}N!} \qquad \text{with} \qquad \varrho(q,p) := \frac{\mathrm{e}^{-H(q,p)/kT}}{\int_M \mathrm{e}^{-H(q,p)/kT}\frac{dq^{3N}dp^{3N}}{h^{3N}N!}}.$$

We assume that the function ϱ is invariant under permutations of the particles. The factorial $N!$ takes the Pauli principle into account (principle of indistinguishable particles). If we introduce the partition function

$$Z(T,V) := \int_M \mathrm{e}^{-H(q,p)/kT}\frac{dq^{3N}dp^{3N}}{h^{3N}N!},$$

then we obtain the following thermodynamic quantities:

- Free energy: $F(T,V) := -kT\ln Z(T,V)$.
- Entropy: $S(T,V) = -F_T(T,V)$.
- Pressure: $P(T,V) = -F_V(T,V)$.
- Mean energy: $\bar{E}(T,V) = F(T,V) + TS(T,V)$.

7.17.6 The Classical Ideal Gas

Let us consider an ideal gas which consists of N freely moving molecules of mass m. The fixed particle number N is assumed to be large (of magnitude 10^{23}). We assume that the molecules move in a 3-dimensional box $\mathcal{B}$ of volume V. Then the following hold:

(i) Free energy: $F = -NkT(1 + \ln \frac{V(2\pi mkT)^{3/2}}{Nh^3})$.

(ii) Entropy: $S = Nk\left(\frac{5}{2} + \ln \frac{V(2\pi mkT)^{3/2}}{Nh^3}\right)$.

(iii) Energy: $E = \frac{3}{2}NkT$.

(iv) Energy fluctuation: $\frac{\Delta E}{E} = \sqrt{\frac{2}{3N}}$.

(v) Pressure: $P = NkT/V$.

(vi) Maxwell's velocity distribution: Fix the origin $\mathcal{O}$ and consider the velocity vector $\mathbf{v} = \overline{\mathcal{OP}}$. The probability of finding the endpoint $\mathcal{P}$ of the velocity vector $\mathbf{v}$ of a single molecule in the open subset C of $\mathbb{R}^3$ is given by the Gaussian integral

$$\left(\frac{m}{2\pi kT}\right)^{3/2} \int_C \mathrm{e}^{-m\mathbf{v}^2/2kT}\, d^3\mathbf{v}. \tag{7.300}$$

Here, $m\mathbf{v}^2/2$ is the kinetic energy of the freely moving molecule, and the normalization factor guarantees that the probability is equal to one if $C = \mathbb{R}^3$.

The experience of physicists shows that these formulas are valid if the temperature T is sufficiently high.[164] Let us compute (i) through (vi). We start with the energy function $H = \sum_{j=1}^N \frac{\mathbf{p}_j^2}{2m}$. The partition function reads as

$$\begin{aligned} Z = \int_{\mathcal{B}^N \times \mathbb{R}^{3N}} \mathrm{e}^{-H(\mathbf{P})/kT}\ \frac{d^{3N}q\, d^{3N}p}{h^{3N}N!} &= \frac{V^N}{h^{3N}N!}\left(\int_{\mathbb{R}} \mathrm{e}^{-p^2/2mkT} dp\right)^{3N} \\ &= \frac{V^N(2\pi mkT)^{3N/2}}{h^{3N}N!} \sim \left(\frac{\mathrm{e}V(2\pi mkT)^{3/2}}{Nh^3}\right)^N. \end{aligned}$$

Here, to simplify computations, we use the approximation[165] $\frac{1}{N!} \sim \left(\frac{\mathrm{e}}{N}\right)^N$. Parallel to (i)–(vii) on page 640 with $\mu = 0$, we get the following formulas

$$F = -kT \ln Z, \quad S = -F_T, \quad E = F + TS, \quad P = -F_V$$

and $(\Delta E)^2 = kT^2 E_T$. By straightforward computations, we obtain the desired formulas (i) through (v). To get (vi), we start with the Gibbs distribution

[164] More precisely, we asssume that $\frac{V(2\pi mkT)^{3/2}}{Nh^3}$ is small. This means that the de Broglie wave length $\lambda := h/(2\pi mkT)^{1/2}$ is small compared with the mean distance $(V/N)^{1/3}$ of the molecules.

[165] This can be motivated by the Stirling formula

$$\frac{1}{N!} = \left(\frac{\mathrm{e}}{N}\right)^N \cdot \frac{1}{\mathrm{e}^{\vartheta(N)/12N}\sqrt{2\pi N}}, \qquad N = 1, 2, \ldots$$

where $0 < \vartheta(N) < 1$. Hence $\frac{1}{N!} = \frac{\mathrm{e}^{N(1+o(N))}}{N^N}$ as $N \to \infty$.

$$\varrho(\mathbf{p}_1,\dots,\mathbf{p}_n)=\frac{\mathrm{e}^{-\sum_{j=1}^n \mathbf{p}_j^2/2mkT}}{\int_{\mathcal{B}^N\times\mathbb{R}^{3N}} \mathrm{e}^{-\sum_{j=1}^N \mathbf{p}_j^2/2mkT}\ \frac{d^{3N}q\,d^{3N}p}{h^{3N}N!}}=\prod_{j=1}^N \nu(\mathbf{p}_j)$$

where

$$\nu(\mathbf{p}):=\frac{\mathrm{e}^{-\mathbf{p}^2/2mkT}}{V\int_{\mathbb{R}^3}\mathrm{e}^{-\mathbf{p}^2/2mkT}\ \frac{d^3\mathbf{p}}{h^3(N!)^{1/N}}}=\frac{h^3(N!)^{1/N}\mathrm{e}^{-\mathbf{p}^2/2mkT}}{V(2\pi kT)^{3/2}}.$$

We assume that the single molecules move independently. Thus, it is reasonable to regard the function ν as the distribution function for a single molecule. For the mean momentum of a single molecule, we obtain

$$\bar{\mathbf{p}}=\int_{\mathcal{B}\times\mathbb{R}^3}\nu(\mathbf{p})\mathbf{p}\,\frac{d^3\mathbf{q}\,d^3\mathbf{p}}{h^3(N!)^{1/N}}.$$

Using $\mathbf{p}=m\mathbf{v}$, we get the mean velocity $\bar{\mathbf{v}}=\int_{\mathbb{R}^3}\mathbf{v}\mathrm{e}^{-m\mathbf{v}^2/2kT}\left(\frac{m}{2\pi kT}\right)^{3/2}d^3\mathbf{v}$ which motivates (vi).

7.17.7 Bose–Einstein Statistics

Let us consider the following situation which frequently arises in quantum statistics. Suppose that the system Γ (e.g., a gas of photons) consists of particles that may assume one of the energy values $\varepsilon_0,\dots,\varepsilon_M$. By definition, a state of Γ is characterized by

$$\varepsilon_0,\varepsilon_1,\dots,\varepsilon_J;\qquad n_0,n_1,\dots,n_J. \tag{7.301}$$

This means that precisely n_j particles of Γ have the energy ε_j, where the index j runs from 0 to J. For each such state, the particle number N and the energy E are given by

$$N=\sum_{j=0}^J n_j,\qquad E=\sum_{j=0}^J n_j\varepsilon_j.$$

Therefore, the partition function is given by

$$Z(T,\mu):=\sum_{\Gamma}\mathrm{e}^{(\mu N-E)/kT}=\prod_{j=0}^J\sum_{n_j}\mathrm{e}^{(\mu n_j-n_j\varepsilon_j)/kT}.$$

Furthermore, we introduce the statistical potential

$$\Omega(T,\mu):=-kT\ln Z(T,\mu)=-kT\sum_{j=0}^J\ln\sum_{n_j}\left(\mathrm{e}^{(\mu-\varepsilon_j)/kT}\right)^{n_j}. \tag{7.302}$$

We now make the crucial assumption that

Each occupation number n_j may assume the values $0,1,\dots,n$.

This corresponds to bosons (that is, particles with integer spin, e.g., photons). Using the geometric series, $\Omega(T,\mu)$ is equal to

$$-kT\sum_{j=0}^J\ln\sum_{n_j=0}^n\left(\mathrm{e}^{(\mu-\varepsilon_j)/kT}\right)^{n_j}=-kT\sum_{j=0}^J\ln\frac{1-\mathrm{e}^{(n+1)(\mu-\varepsilon_j)/kT}}{1-\mathrm{e}^{(\mu-\varepsilon_j)/kT}}.$$

Furthermore, assume that the maximal occupation number n is very large and $\mu - \varepsilon_j < 0$ for all j. Letting $n \to \infty$, we get the final statistical potential

$$\Omega(T,\mu) = kT \sum_{j=0}^{J} \ln\left(1 - \mathrm{e}^{(\mu-\varepsilon_j)/kT}\right).$$

By (7.290) on page 640, $N = -\Omega_\mu, S = -\Omega_T, F = \Omega + \mu N$, and $E = F + TS$. This yields the following.

(i) Mean particle number: $N = \sum_{j=0}^{J} N_j$ where $N_j := \frac{\mathrm{e}^{(\mu-\varepsilon_j)/kT}}{1-\mathrm{e}^{(\mu-\varepsilon_j)/kT}}$.
(ii) Mean energy: $E = \sum_{j=0}^{J} N_j \varepsilon_j$.
(iii) Free energy: $F = \Omega + \mu N$.
(iv) Entropy: $S = \frac{E-F}{T}$.

In particular, if each particle behaves like a quantum harmonic oscillator of angular frequency ω, then $\varepsilon_j = \hbar\omega(j + \frac{1}{2})$. We have shown in Sect. 2.3.2 of Vol. I that Planck's radiation law is a consequence of the mean energy formula (ii) for the quantum harmonic oscillator.

The Maxwell–Boltzmann statistics as a limit case for high temperature. In the special case where $\mathrm{e}^{(\mu-\varepsilon_j)/kT} \ll 1$ (e.g., $\mu - \varepsilon_j < 0$ and T is large), we approximately obtain

$$N_j = \mathrm{e}^{(\mu-\varepsilon_j)/kT}. \tag{7.303}$$

This is called the classical Maxwell–Boltzmann statistics. which generalizes the Maxwell velocity distribution (7.300) on page 648.

Bose–Einstein condensation as a limit case at low temperature. Suppose that $0 \le \varepsilon_0 < \varepsilon_1 < \varepsilon_2 < \dots$. We expect that at low temperatures most of the bosons are located in the ground state. In fact, by (i), for the particle numbers we get

$$\lim_{T\to+0} \lim_{\mu\to\varepsilon_0-0} N_j(T,\mu) = \begin{cases} +\infty & \text{if} \quad j = 0, \\ 0 & \text{if} \quad j = 1,2,\dots \end{cases}$$

This phenomenon is called Bose–Einstein condensation.

7.17.8 Fermi–Dirac Statistics

In contrast to the preceding section, we now assume that

Each occupation number n_j may only assume the values $0, 1$.

This corresponds to the Pauli exclusion principle for fermions (that is, particles with half-integer spin, e.g., electrons).[166] This yields

$$\Omega(T,\mu) = -kT \sum_{j=0}^{J} \ln(1 + \mathrm{e}^{(\mu-\varepsilon_j)/kT}).$$

As in Sect. 7.17.7, we now obtain the following:s

[166] More precisely, if s is the spin of the particles, then each energy value ε_j has to be counted with the multiplicity $2s + 1$.

(i) Mean particle number: $N = \sum_{j=0}^{J} N_j$ where $N_j := \frac{e^{(\mu-\varepsilon_j)/kT}}{1+e^{(\mu-\varepsilon_j)/kT}}$.
(ii) Mean energy: $E = \sum_{j=0}^{J} N_j \varepsilon_j$.
(iii) Free energy: $F = \Omega + \mu N$.
(iv) Entropy: $S = \frac{E-F}{T}$.

If $e^{(\mu-\varepsilon_j)/kT} \ll 1$ (e.g., $\mu - \varepsilon_j < 0$ and T is large), then we obtain the classical Maxwell–Boltzmann statistics (7.303).

The Fermi ball as a limit case at low temperature. For the particle numbers, we get

$$\lim_{T \to +0} N_j(T, \mu) = \begin{cases} 1 & \text{if} \quad \varepsilon_j < \mu, \\ 0 & \text{if} \quad \mu < \varepsilon_j. \end{cases}$$

This means that at low temperature each of the lowest energy levels is occupied by precisely one particle. In contrast to Bose–Einstein condensation, by the Pauli principle it is impossible that all of the particles are in the ground state. For example, consider a gas of N electrons in a box of volume V in the limit case of temperature $T = 0$. Since the electron has spin $s = \frac{1}{2}$, each cell of volume h^3 in the phase space contains two electrons with different spin orientations. Thus, if P denotes the maximal momentum of the electrons at $T = 0$, then the phase space volume $\frac{4}{3}\pi P^3 \cdot V$ contains N particles where

$$N = \frac{2}{h^3} \cdot \frac{4}{3}\pi P^3 V.$$

The ball of radius P is called the Fermi ball of the N-particle electron gas at zero temperature, and the surface of the Fermi ball is called the Fermi surface.

Applications of the Bose–Einstein statistics and the Fermi–Dirac statistics to interesting physical phenomena can be found in Zeidler (1986), Vol. IV, Chap. 68 (see the references on page 1049). For example, this concerns Planck's radiation law for photon gases, as well as the Fermi ball which is crucial for computing the critical Chandrasekhar mass of special stars called white dwarfs (see also N. Straumann, General Relativity with Applications to Astrophysics, Springer, New York, 2004). Using the methods of quantum field theory, the structure of Fermi surfaces for electrons in a crystal is studied in M. Salmhofer, Renormalization: An Introduction, Springer, Berlin, 1999.

7.17.9 Thermodynamic Equilibrium and KMS-States

The grand canonical example in finite quantum statistics. Let X be a finite-dimensional complex Hilbert space, $X \neq \{0\}$. Choose the density operator

$$\varrho_0 := \frac{e^{\beta(\mu N - H)}}{\operatorname{tr} e^{\beta(\mu N - H)}}.$$

Here, $H, N : X \to X$ are self-adjoint operators, and $\beta > 0$ and μ are real parameters, with the temperature T, the Boltzmann constant k, the chemical potential μ, and $\beta = 1/kT$. In the language of C^*-algebras, the following hold.

- The extended C^*-algebra $\mathcal{A} = L(X, X)$ of observables consists of all linear operators $A : X \to X$ equipped with the norm $||A|| := \sqrt{\operatorname{tr}(A^* A)}$.
- The states are defined by $\chi_0(A) := \operatorname{tr}(\varrho_0 A)$ for all $A \in \mathcal{A}$.

- The dynamics of the state χ_0 is given by

$$\chi_t(A) := \mathrm{tr}(\varrho_0 \mathcal{U}_t A)$$

for all $A \in \mathcal{A}$ and all times $t \in \mathbb{R}$. Here, $\mathcal{U}_t(A) := \mathrm{e}^{\mathrm{i}tH/\hbar} A \mathrm{e}^{-\mathrm{i}tH/\hbar}$.

Proposition 7.68 *The state χ_0 corresponding to the density operator ϱ_0 and the dynamics $\{\chi_t\}_{t\in\mathbb{R}}$ is a KMS-state of temperature T.*

Proof. Set $Z := \mathrm{tr}(\mathrm{e}^{\beta(\mu N - H)})$. To simplify notation, choose $\mu := 0$ and $\hbar := 1$. Then

$$\chi_0(A) = \mathrm{tr}(\varrho_0 A) = Z^{-1}\,\mathrm{tr}(\mathrm{e}^{-\beta H} A).$$

Noting the commutativity property of the trace, $\mathrm{tr}(CD) = \mathrm{tr}(DC)$, we get

$$\begin{aligned} Z\chi_0(\mathcal{U}_{t-\mathrm{i}\beta}(B)A) &= \mathrm{tr}(\mathrm{e}^{-\beta H}\mathrm{e}^{\mathrm{i}(t-\mathrm{i}\beta)H} B \mathrm{e}^{-\mathrm{i}(t-\mathrm{i}\beta)H} A) = \mathrm{tr}(\mathrm{e}^{\mathrm{i}tH} B \mathrm{e}^{-\mathrm{i}tH}\mathrm{e}^{-\beta H} A) \\ &= \mathrm{tr}(\mathrm{e}^{-\beta H} A \mathrm{e}^{\mathrm{i}tH} B \mathrm{e}^{-\mathrm{i}tH}) = Z\chi_0(A\mathcal{U}_t(B)). \end{aligned}$$

□

Example. As a typical example, choose the operators H and N in such a way that

$$H\psi_j = E_j\psi_j, \qquad N\psi_j = N_j\psi_j, \qquad j = 1, \ldots, n$$

where $\psi_1, \ldots, \psi_n$ is an orthonormal basis of X, and E_j, N_j are nonnegative numbers for all j. Then $\varrho_0\psi_j = p_j\psi_j$ with

$$p_j = \frac{\mathrm{e}^{\beta(\mu N_j - E_j)}}{\sum_{j=1}^n \mathrm{e}^{\beta(\mu N_j - E_j)}}.$$

The operator H (resp. N) is called the Hamiltonian with the energy levels $E_1, \ldots, E_n$ (resp. the particle operator with the particle numbers $N_1, \ldots, N_n$.)

7.17.10 Quasi-Stationary Thermodynamic Processes and Irreversibility

> In the huge factory of natural processes, the principle of entropy occupies the position of manager, for it dictates the manner and method of the whole business, whilst the principle of energy merely does the bookkeeping, balancing debits and credits...
> Life on the earth needs the radiation of the sun. Our conditions of existence require a determinate degree of temperature, and for the maintenance of this there is needed not addition of energy, but addition of entropy.[167]
>
> Robert Emden, 1938

Let us study the sufficiently regular time-evolution of the grand canonical ensemble. By a quasi-stationary process of the grand canonical ensemble, we understand smooth time-depending functions of temperature, chemical potential, and volume:

$$T = T(t), \quad \mu = \mu(T), \quad V = V(t), \qquad t_0 \le t \le t_1. \tag{7.304}$$

By (7.290) on page 640, this yields the following quantities:

$$E = E(t), \quad N = N(t), \quad S = S(t), \quad P = P(t), \qquad t_0 \le t \le t_1.$$

[167] R. Emden, Why do we have winter heating? Nature **14** (1938), 908–909.

Here, E is the mean energy, N is the mean particle number, S is the entropy, and P is the pressure.[168] From the physical point of view, this is an idealization. We assume that the physical system is in thermodynamic equilibrium at each time t. In reality, a certain relaxation time is needed in order to pass from a thermodynamic equilibrium state to a new one. Let $Q(t)$ be the heat added to the physical system during the time interval $[t_0, t]$. We postulate that, for all times t in the interval $[t_0, t_1]$, the process (7.304) has the following properties.

(i) The first law of thermodynamics: $\dot{E}(t) = \dot{Q}(t) - P(t)\dot{V}(t) + \mu(t)\dot{N}(t)$.

(ii) The second law of thermodynamics: $T(t)\dot{S}(t) \geq \dot{Q}(t)$.

(iii) The third law of thermodynamics. Suppose that the temperature $T(t)$ goes to zero as $t \to t_1 - 0$. Then so do the entropy $S(t)$ and its partial derivatives $S_T(t), S_\mu(t), S_V(t)$.

The first law describes conservation of energy. To discuss the second law, let us introduce the external entropy

$$S_e(t) := S(t_0) + \int_{t_0}^{t} \frac{\dot{Q}(\tau)}{T(\tau)}\, d\tau, \qquad t_0 \leq t \leq t_1,$$

which depends on the heat added to the system. In addition, we introduce the remaining internal entropy $S_i(t) := S(t) - S_e(t)$. Then

$$\dot{S}_e(t) = \frac{\dot{Q}(t)}{T(t)}, \qquad \dot{S}_i(t) \geq 0, \qquad t_0 \leq t \leq t_1.$$

Assume that $t_0 = -t_1$ where $t_1 > 0$. If the quasi-stationary process (7.304) has the property that also the time-reflected process

$$T = T(-t), \quad \mu = \mu(-T), \quad V = V(-t), \qquad t_0 \leq t \leq t_1$$

is quasi-stationary, then the process is called reversible. In this case, because of $\frac{d}{dt}S(-t) = -(\frac{d}{dt}S)(-t)$, the second law tells us that

$$-T(-t)\dot{S}(-t) \geq \dot{Q}(-t), \qquad -t_1 \leq t \leq t_1.$$

This implies

$$T(t)\dot{S}(t) \geq \dot{Q}(t), \qquad -T(t)\dot{S}(t) \geq \dot{Q}(t), \qquad -t_1 \leq t \leq t_1.$$

Hence $T(t)\dot{S}(t) = \dot{Q}(t)$ for all $t \in [-t_1, t_1]$. This means that the internal entropy S_i vanishes on the time interval $[-t_1, t_1]$. Processes are called irreversible iff they are not reversible. Typically, the time-evolution of living beings is irreversible. A more detailed discussion can be found in Zeidler (1986), Vol. IV, Chap. 67 (see the references on page 1049).

The thermodynamic limit and phase transitions. If the volume V of the physical system goes to infinity, $V \to \infty$, then this limit is called the thermodynamic limit by physicists. Then it may happen that important thermodynamic quantities become singular for appropriate parameters (e.g., temperature T). These singularities correspond to phase transitions (e.g., the transition from water to ice). Phase transitions play a fundamental role for understanding critical phenomena in nature (e.g., the inflation of the very early universe and the emergence of the three fundamental forces during the cooling process of the hot universe after the Big

[168] To simplify notation, we write E and N instead of $\bar{E}$ and $\bar{N}$, respectively.

Bang).[169] In terms of statistical physics, phase transitions correspond to a strong increase of fluctuations. We will encounter this in later volumes. As an introduction to the rigorous theory of phase transitions, we recommend the classical survey article by Griffith.[170]

7.17.11 The Photon Mill on Earth

Living objects store a lot of information related to the genetic code. There arises the following question in physics: where does this information come from? The solution of this interesting problem is given by the entropy relation

$$\Delta S_e = \frac{\Delta Q}{T_{\text{in}}} - \frac{\Delta Q}{T_{\text{out}}},$$

which is called the photon mill on earth. In fact, the sun sends photons to the earth at the temperature $T_{\text{in}} = 5800$ K, which is the high surface temperature of the sun. Most of these photons are reflected by the surface of earth, and they are sent to the universe at the lower temperature $T_{\text{out}} = 260$ K. Since $T_{\text{in}} > T_{\text{out}}$, the earth radiates the amount of entropy ΔS_e into the universe. More precisely, during one second, the surface of earth gets the heat energy $\Delta Q = 10^{17}$ J from the sun. Hence the entropy loss of earth during one second is equal to

$$\Delta S_e = -4 \cdot 10^{14} \text{ J/K}.$$

This means that one square meter of the surface of earth radiates the entropy of about 1 J/K during one second into the universe. The radiated entropy decreases the disorder on earth, that is, the earth gains order. This is mainly the information stored in living objects. Physicists describe this by saying that energy at a higher temperature has a higher quality than the same amount of energy at a lower temperature.

7.18 Von Neumann Algebras

> In order to deeply understand the mathematical structure of quantum mechanics, John von Neumann studied a special class of operator algebras. Nowadays these algebras are called von Neumann algebras.[171]
> Each von Neumann algebra is a C^*-algebra. But the converse is seldom true.
>
> Folklore

[169] See G. Börner, The Early Universe: Facts and Fiction, Springer, Berlin, 2003.
Ø. Grøn and S. Hervik, Einstein's Theory of General Relativity: with Modern Applications in Cosmology, Springer, New York, 2007.
S. Weinberg, Cosmology, Oxford University Press, 2008.

[170] R. Griffith, Rigorous results and theorems. In: C. Domb and M. Green (Eds.), Phase transitions and critical phenomena, Academic Press, New York, 1970, pp. 9–108.

[171] F. Murray and J. von Neumann, On rings of operators, Ann. Math. **37** (1936), 116–229.
J. von Neumann, On rings of operators: reduction theory, Ann. Math. **50** (1949), 401–485.

> The theory of von Neumann algebras has been growing in leaps and bounds in the last 20 years. It has always had strong connections with ergodic theory and mathematical physics. It is now beginning to make contact with other areas such as differential geometry and K-theory... The book commences with the Murray–von Neumann classification of factors, proceeds through the basic modular theory (the Tomita–Takesaki theory) to the Connes classification of von Neumann algebras of type III_λ, and concludes with a discussion of crossed-products, Krieger's ratio set, examples of factors, and Takesaki's duality theorem.[172]
>
> Viakalathur Sunder, 1987

In what follows, X is a complex separable non-trivial Hilbert space (i.e., $X \neq \{0\}$).

7.18.1 Von Neumann's Bicommutant Theorem

Commutant. Commutation relations play a crucial role in quantum mechanics. In particular, if two observables commute, then it is possible that they have common eigenvectors, that is, they can be sharply measured at the same time. This motivated John von Neumann to investigate commutants of algebras. Consider the C^*-algebra $L(X, X)$ of the linear continuous operators

$$A : X \to X.$$

Let $\mathcal{S}$ be a subset of $L(X, X)$. By definition, the operator $A \in L(X, X)$ belongs to the commutant $\mathcal{S}'$ of the set $\mathcal{S}$ iff

$$AS = SA \qquad \text{for all} \quad S \in \mathcal{S}.$$

Naturally enough, we set $\mathcal{S}'' := (\mathcal{S}')'$ and call this the bicommutant of the set $\mathcal{S}$. Obviously, $\mathcal{S} \subseteq \mathcal{S}''$. Von Neumann studied the special case where $\mathcal{S} = \mathcal{S}''$. A subset $\mathcal{A}$ of $L(X, X)$ is called a $*$-subalgebra iff $A, B \in L(X, X)$ and $\alpha, \beta \in \mathbb{C}$ imply that the operators $\alpha A + \beta B, AB, A^*$ are also contained in $L(X, X)$.

> *By definition, a von Neumann algebra is a $*$-subalgebra $\mathcal{A}$ of $L(X, X)$ with unit element and $\mathcal{A}'' = \mathcal{A}$.*

This definition is purely algebraic. Equivalently, one can characterize von Neumann algebras in topological terms by using weak convergence. Let us discuss this.

Weak operator convergence. Let (A_n) be a sequence of linear operators $A_n : X \to X$ in $L(X, X)$, $n = 1, 2, \ldots$ We write

$$w - \lim_{n\to\infty} A_n = A$$

iff $A \in L(X, X)$ and $\lim_{n\to\infty} \langle\psi|A_n\varphi\rangle = \langle\psi|A\varphi\rangle$ for all $\psi, \varphi \in X$. This is called the weak operator convergence. This corresponds to the convergence of matrix elements. In terms of physics, this guarantees the convergence of expectation values. Generalizing this, let $(A_\nu)_{\nu\in\mathcal{N}}$ be a generalized sequence in $L(X, X)$ with a directed index set $\mathcal{N}$ (see page 240). We write

$$w - \lim_{\nu\to\infty} A_\nu = A \tag{7.305}$$

iff $A \in L(X, X)$ and $\lim_{\nu\to\infty} \langle\psi|A_\nu\varphi\rangle = \langle\psi|A\varphi\rangle$ for all $\psi, \varphi \in X$, in the sense of generalized convergence. In addition, let us introduce the following two notions of convergence.

[172] V. Sunder, An Invitation to von Neumann Algebras, Springer, Berlin, 1987 (reprinted with permission).

- $s-\lim_{\nu\to\infty} A_\nu = A$ iff $\lim_{\nu\to\infty} ||(A-A_\nu)\varphi|| = 0$ for all $\varphi \in X$ (strong operator convergence);
- $u-\lim_{\nu\to\infty} A_\nu = A$ iff $\lim_{\nu\to\infty} ||A_\nu - A|| = 0$ (uniform operator convergence).

Recall that $||A_\nu - A|| = \sup_{||\varphi||\le 1} ||(A_\nu - A)\varphi||$. This justifies the notion of uniform operator convergence.

Semi-norms. A map $p : L \to \mathbb{R}$ on the complex (resp. real) linear space L is called a semi-norm iff for all $A, B \in L$ and all complex (resp. real) numbers α the following hold:

- $p(A) \ge 0$,
- $p(\alpha A) = |\alpha| p(A)$, and
- $p(A+B) \le p(A) + p(B)$.

If, in addition, $p(A) = 0$ implies $A = 0$, then p is a norm.

Topologies on $L(X,X)$. (i) Weak operator topology. For fixed $\psi, \varphi \in X$, define

$$p_{\psi,\varphi}(A) := |\langle \psi | A\varphi \rangle| \qquad \text{for all} \quad A \in L(X,X).$$

This is a semi-norm on $L(X,X)$. A subset $\mathcal{S}$ of $L(X,X)$ is called *weakly open* iff, for each operator $A_0 \in \mathcal{S}$, there exist a finite family $\psi_1, \varphi_1, \ldots, \psi_n, \varphi_n$ of elements in X and a number $\varepsilon > 0$ such that the set

$$\{A \in L(X,X) : \ \ p_{\psi_j,\varphi_j}(A - A_0) < \varepsilon, \ j = 1, \ldots, n\}$$

is contained in $\mathcal{S}$. This generates a topology on $L(X,X)$ called the weak operator topology. A subset of $L(X,X)$ is called weakly closed iff its complement in $L(X,X)$ is weakly open.

A subset $\mathcal{S}$ of $L(X,X)$ is weakly closed iff, for all generalized sequences (A_ν) in $\mathcal{S}$, it follows from $w-\lim_{\nu\to\infty} A_\nu = A$ that $A \in \mathcal{S}$.

(ii) Strong operator topology. Similarly, we obtain the strong operator topology by replacing $p_{\psi,\varphi}$ by the semi-norm $p_\varphi(A) := ||A\varphi||$. A subset $\mathcal{S}$ of $L(X,X)$ is called *strongly open* iff, for each operator $A_0 \in \mathcal{S}$, there exist a finite family $\varphi_1, \ldots, \varphi_n$ of elements in X and a number $\varepsilon > 0$ such that the set

$$\{A \in L(X,X) : \ \ p_{\varphi_j}(A - A_0)\} < \varepsilon, \ j = 1, \ldots, n\}$$

is contained in $\mathcal{S}$. This generates a topology on $L(X,X)$ called the strong operator topology. A subset of $L(X,X)$ is called strongly closed iff its complement in $L(X,X)$ is strongly open.

A subset $\mathcal{S}$ of $L(X,X)$ is strongly closed iff, for all generalized sequences (A_ν) in $\mathcal{S}$, it follows from $s-\lim_{\nu\to\infty} A_\nu = A$ that $A \in \mathcal{S}$.

(iii) Uniform operator topology. This topology is obtained by replacing p_φ by the norm $p(A) := ||A||$. A subset $\mathcal{S}$ of $L(X,X)$ is called *uniformly open* iff, for each operator $A_0 \in \mathcal{S}$, there exists a number $\varepsilon > 0$ such that the set

$$\{A \in L(X,X) : \ \ p(A - A_0) < \varepsilon\}$$

is contained in $\mathcal{S}$. This generates a topology on $L(X,X)$ called the uniform operator topology.[173] A subset of $L(X,X)$ is called uniformly closed iff its complement in $L(X,X)$ is uniformly open.

[173] This topology coincides with the topology induced by the Banach space structure on $L(X,X)$.

A subset $\mathcal{S}$ of $L(X,X)$ is uniformly closed iff, for all classical sequences $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{S}$, it follows from $u-\lim_{n\to\infty} A_n = A$ that $A \in \mathcal{S}$.

The bicommutant theorem. The topological characterization of von Neumann algebras reads as follows.

Theorem 7.69 *Let X be a complex separable non-trivial Hilbert space. For a given $*$-subalgebra $\mathcal{A}$ of $L(X,X)$ with unit element, the following three statements are equivalent:*
(i) *$\mathcal{A}$ is a von Neumann algebra (i.e., $\mathcal{A}'' = \mathcal{A}$).*
(ii) *$\mathcal{A}$ is weakly closed in $L(X,X)$.*
(iii) *$\mathcal{A}$ is strongly closed in $L(X,X)$.*

More general, the following hold: If $\mathcal{A}$ is a $*$-subalgebra of $L(X,X)$ with unit element, then the closure of $\mathcal{A}$ in the weak (resp. strong) topology on $L(X,X)$ coincides with the bicommutant $\mathcal{A}''$.

Corollary 7.70 *A $*$-subalgebra algebra of $L(X,X)$ is a C^*-algebra iff it is uniformly closed in $L(X,X)$.*

Consequently, each von Neumann algebra is a C^*-algebra. But the converse is seldom true. For the proofs, we refer to P. Kadison and J. Ringrose, Fundamentals of the Theory of Operator Algebras, Vol. 1, Academic Press, New York, 1983. Many beautiful applications of von Neumann algebras to harmonic analysis can be found in

K. Maurin, Generalized Eigenfunction Expansions and Unitary Representations of Topological Groups, PWN, Warsaw, 1968.

Examples. Suppose that the operator $A \in L(X,X)$ is self-adjoint (i.e., $A^* = A$). The bicommutant $\mathcal{A}''$ of the one-point set $\mathcal{A} := \{A\}$ is the smallest von Neumann algebra in $L(X,X)$ containing the self-adjoint operator A.

Let $\mathcal{S}$ be a subset of $L(X,X)$ with the property that $A \in \mathcal{S}$ implies $A^* \in \mathcal{S}$. Then:

(i) The commutant $\mathcal{S}'$ is a von Neumann algebra.
(ii) The bicommutant $\mathcal{S}''$ is the smallest von Neumann algebra in $L(X,X)$ containing the set $\mathcal{S}$.
(iii) $\mathcal{S}' = \mathcal{S}'''$.

By induction, this implies $\mathcal{S}' = \mathcal{S}^{2n+1}$ and $\mathcal{S}'' = \mathcal{S}^{2n+2}$ for all $n = 1, 2, \ldots$ That is, all of the higher commutants are determined by $\mathcal{S}'$ and $\mathcal{S}''$.

A von Neumann algebra is called a factor iff its center $\mathcal{A} \cap \mathcal{A}'$ is trivial (i.e., it consists of the multiples of the unit operator, $\mathcal{A} \cap \mathcal{A}' = \{\alpha I : \alpha \in \mathbb{C}\}$).

The classification problem for von Neumann algebras. By von Neumann's spectral theory, a self-adjoint operator $A \in L(X,X)$ on the Hilbert space X can be represented by orthogonal projection operators E_λ ($\lambda \in \mathbb{R}$) called the spectral family of A. Now we consider the following generalization:

- self-adjoint operator $\Rightarrow$ von Neumann algebra,
- spectral family $\Rightarrow$ factors.

The building blocks of factors are orthogonal projections.

In contrast to general C^*-algebras, von Neumann algebras possess a rich structure of orthogonal projections.

Since orthogonal projections are observables corresponding to "questions," von Neumann algebras represent a nice tool for describing physical processes in quantum theory. Murray and von Neumann showed that each von Neumann algebra can be represented as a direct sum (or, more general, as a direct integral) of factors. Therefore it remains to classify the factors.[174]

7.18.2 The Murray–von Neumann Classification of Factors

Let X be a complex separable non-trivial Hilbert space, and let the subset $\mathcal{A}$ of $L(X,X)$ be a von Neumann algebra which is a factor. The factor $\mathcal{A}$ is said to be of type I, II, III iff it satisfies the following conditions, respectively:

Type I: $\mathcal{A}$ contains a minimal projection.
Type II: $\mathcal{A}$ contains no minimal projection, but does contain a non-zero projection.
Type III: $\mathcal{A}$ contains no non-zero finite projection.

Here, we use the following terminology. Let $\mathcal{P}(\mathcal{A})$ be the set of all orthogonal projections $P \in \mathcal{A}$. For $P, Q \in \mathcal{A}$, we write $Q \sim P$ iff there exists an operator $U \in \mathcal{A}$ such that $Q = UU^*$ and $P = U^*U$. This is an equivalence relation on $\mathcal{P}(\mathcal{A})$.

- The orthogonal projection P is called finite iff it follows from $Q(X) \subseteq P(X)$ and $Q \sim P$ that $Q = P$.
- The orthogonal projection P is called minimal iff the following three conditions are satisfied:
  (α) $P \neq 0$.
  (β) $P(X)$ is invariant under $\mathcal{A}'$.
  (γ) If a linear subspace Y of $P(X)$ is invariant under $\mathcal{A}'$, then Y is trivial (i.e., $Y = \{0\}$ or $Y = P(X)$).
  Minimal projections are always finite (and non-zero).

For example, if $\mathcal{A} = L(X,X)$, then P is finite iff the projection space $P(X)$ is finite-dimensional. Moreover, precisely the orthogonal projections onto one-dimensional linear subspaces are minimal.

The generalized dimension function of factors. For each factor $\mathcal{A}$, there exists a function $d : \mathcal{P}(\mathcal{A}) \to [0, \infty]$ which has the following properties:

(i) $Q \sim P$ iff $d(Q) = d(P)$.
(ii) If $P(X)$ is orthogonal to $Q(X)$, then $d(P+Q) = d(P) + d(Q)$.
(iii) P is finite iff $d(P) < \infty$, and $d(P) = 0$ iff $P = 0$.

The function d is uniquely determined, up to a positive multiplicative constant. For a suitable choice of this constant, the function d has the following range:

Type I_n : $\{0, 1, \ldots, n\}$, where $n = 1, 2, \ldots$ or $n = \infty$.
Type II_1: $[0, 1]$.
Type II_∞: $[0, \infty]$.
Type III: $\{0, \infty\}$.

A factor $\mathcal{A}$ is of type I_n iff $\mathcal{A} = L(X,X)$ where $\dim X = n$. In this simple case, $d(Q) = \dim P(X)$.

[174] Direct integrals of Hilbert spaces generalize direct sums of Hilbert spaces by summing over general index sets with respect to a measure. This will be considered in Vol. IV on quantum mathematics (see also K. Maurin, Generalized Eigenfunction Expansions and Unitary Representations of Topological Groups, PWN, Warsaw, 1968).

7.18.3 The Tomita–Takesaki Theory and KMS-States

> The Tomita–Takesaki theorem is a beautiful example of "prestabilized harmony" between physics and mathematics.[175]
> On the one hand, it is intimately related to the Kubo–Martin–Schwinger (KMS) condition. On the other hand it initiated a significant advance in the classification theory of von Neumann algebras and led to powerful computational techniques.
>
> Rudolph Haag, 1996

KMS-states in thermodynamic equilibrium. The physicists Kubo, Martin and Schwinger discovered in the late 1950s that states of thermodynamic equilibrium can be characterized by special analyticity properties of the Green's function.[176] In 1967 it was shown by Haag, Hugenholtz, and Winnink that this can be formulated in terms of von Neumann algebras. In fact, it turned out that this was closely related to the so-called Tomita–Takesaki theory for von Neumann algebras, which was created by the Japanese mathematician Tomita in the 1960s, by purely mathematical motivation.[177] Roughly speaking, the Tomita–Takesaki theory formulates conditions which guarantee the existence of a dynamics on a von Neumann algebra that can be used in order to describe the dynamics of a physical state in thermodynamic equilibrium.

The basic mathematical idea of the Tomita–Takesaki theory. Let $\mathcal{A}$ be a von Neumann algebra of operators on the complex separable non-trivial Hilbert space X. Suppose that there is a vector ψ_0 in X which has the following two properties:

- ψ_0 is cyclic, that is, the set $\{A\psi_0 : A \in \mathcal{A}\}$ is dense in X.
- ψ_0 is separating, that is, if $A, B \in \mathcal{A}$ and $A \neq B$, then $A\psi_0 \neq B\psi_0$.

Define the operator $\mathsf{S} : \mathrm{dom}(\mathsf{S}) \to X$ by setting[178]

$$\mathsf{S}(A\psi_0) := A^*\psi_0 \qquad \text{for all} \quad A \in \mathcal{A}.$$

Then, the operator S has a closure $\bar{\mathsf{S}}$. By Problem 7.24, there exists the unique polar decomposition

$$\mathsf{S} = J\Delta^{1/2}$$

with the following properties:

- The so-called modular operator $\Delta := \bar{\mathsf{S}}^*\bar{\mathsf{S}}$ is self-adjoint and $\langle\psi|\Delta\psi\rangle \geq 0$ for all $\psi \in \mathrm{dom}(\Delta)$.

---

[175] The term "prestabilized harmony" was introduced by Leibniz (1646–1716) in his philosophy of monads (which are ultimate units of being).
R. Haag, Local Quantum Physics: Fields, Particles, Algebras, Springer, Berlin, 1996 (reprinted with permission).

[176] R. Kubo, Statistical mechanical theory of irreversible processes, J. Math. Soc. Japan **12** (1957), 570–586.
P. Martin and J. Schwinger, Theory of many-particle systems. Phys. Rev. **115** (1959), 1342–1373.

[177] R. Haag, N. Hugenholtz, and M. Winnink, On the equilibrium states in quantum statistical mechanics, Commun. Math. Phys. **5** (1967), 215–236.
M. Takesaki, Tomita's Theory of Modular Hilbert Algebras and Its Applications, Springer, Berlin, 1970.

[178] Since $A^*(\alpha\varphi) = \alpha^\dagger A^*\varphi$ for all complex numbers α, the operator S is antilinear, that is, $\mathsf{A}(\alpha\varphi + \beta\psi) = \alpha^\dagger A + \beta^\dagger B$ for all $\varphi, \psi \in X$ and all $\alpha, \beta \in \mathbb{C}$.

- The so-called modular conjugation operator $J : X \to X$ is antiunitary, and it has the property $J^2 = I$.[179]

For all times $t \in \mathbb{R}$, we have

$$\boxed{\Delta^{it} \mathcal{A} \Delta^{-it} = \mathcal{A},}$$

and $J\mathcal{A}J = \mathcal{A}$.[180] Setting $\mathcal{U}_t(A) := \Delta^{it} \mathcal{A} \Delta^{-it}$ for all $A \in \mathcal{A}$, the map

$$\mathcal{U}_t : \mathcal{A} \to \mathcal{A}$$

is a C^*-automorphism of the von Neumann algebra $\mathcal{A}$, and the family $\{\mathcal{U}_t\}_{t \in \mathbb{R}}$ forms a one-parameter group of C^*-automorphisms on $\mathcal{A}$ (see page 634). These C^*-automorphisms are called modular automorphisms.

For the general mathematical theory of von Neumann algebras together with numerous applications to quantum physics, see the hints for further reading on page 677. We will come back to this in Vol. IV on quantum mathematics.

7.19 Connes' Noncommutative Geometry

> The abstract theory of commutative Banach algebras was initiated by Mazur (1905–1981) in 1936, but it blossomed in the hands of Gelfand (born 1913), who in one brilliant study gave it the final perfect shape. This was the Gelfand theory of maximal ideals, or the Gelfand spectral theory looking at it the other way... The Gelfand spectral theory soon became a powerful tool and a bonanza of new ideas. Gelfand himself, Naimark (1909–1978), and others of his co-workers found a multitude of models and applications.[181]
>
> Krysztof Maurin, 1968

> Noncommutative geometry amounts to a program of unification of mathematics under the aegis of the quantum apparatus, that is, the theory of operators and of C^*-algebras. Largely the creation of a single person, Alain Connes, noncommutative geometry is just coming of age as the new century opens.[182] The bible of the subject is, and will remain, Connes' *Noncommutative Geometry* (1994), itself the "3.8 expansion" of the French *Géométrie non commutative* from 1990. These are extraordinary books, a "tapestry" of physics and mathematics, in the words of Vaughan Jones, and the work of a "poet of modern science," according to Daniel Kastler, replete with subtle knowledge and insights apt to inspire several generations.

[179] That is, for all $\varphi, \psi \in X$ and all complex numbers, we have $\langle J\psi | J\varphi \rangle = \langle \psi | \varphi \rangle^\dagger$, and the operator J is antilinear.

[180] By definition, $B\mathcal{A}C := \{BAC : A \in \mathcal{A}.\}$.

[181] K. Maurin, Generalized Eigenfunction Expansions and Unitary Representations of Topological Groups, Polish Scientific Publishers, Warsaw, 1968 (reprinted with permission). See also the footnote on page 628.

[182] Alain Connes (born 1947) works at the *Collège de France*, Paris, and at the *l'Institut des Hautes Études Scientifiques* (IHES) (Institute of Advanced Scientific Studies), Bures-sur-Yvette (near Paris). For his contributions to the theory of von Neumann algebras of type III, Connes was awarded the Fields medal in 1983. See A. Connes, Une classification des facteurs de type III, Ann. Scient. École Norm. Sup. **6** (1973), 133–252 (in French).

> Despite an explosion of research by some of the world's leading mathematicians, and a bouquet of applications – to the reinterpretation of the phenomenological Standard Model of particle physics as a new space-time geometry, the quantum Hall effect, strings, renormalization and more in quantum field theory – the six years that have elapsed since the publication of *Noncommutative Geometry* have seen no sizeable book returning to the subject. This volume aspires to fit snugly in that gap, but does not pretend to fill it. It is rather meant to be an introduction to some of the core topics of *Noncommutative Geometry.*[183]
>
> José Gracia-Bondia, Joseph Várilly, and Héctor Figueroa, 2001

If M is a nonempty compact separated topological space (e.g., a bounded closed subset of $\mathbb{R}^n$), then the C^*-algebra $C(M)$ knows a lot about the geometry of M. For example, for a given point P, the set $\mathcal{J}_P$ of all continuous functions

$$f : M \to \mathbb{C}$$

with $f(P) = 0$ forms a maximal C^*-ideal of $C(M)$.[184] Conversely, each maximal C^*-ideal of $C(M)$ can be obtained this way. Thus,

$$P \mapsto \mathcal{J}_P$$

is a bijective map between the space M and the set of maximal ideals of the function algebra $C(M)$. Furthermore, the Gelfand–Naimark structure theorem tells us that, for each commutative C^*-algebra $\mathcal{A}$, we have the C^*-isomorphism

$$\mathcal{A} \simeq C(M)$$

where M is the set of maximal C^*-ideals of $\mathcal{A}$.

> *The basic idea of Connes' noncommutative geometry is to replace commutative C^*-algebras by noncommutative C^*-algebras.*

In this setting, properties of noncommutative geometry are identified with properties of noncommutative C^*-algebras. This identification is motivated by the corresponding identification between classical geometric properties and properties of commutative C^*-algebras. From the physical point of view, the idea is that states and observables are primary, but space and time are secondary. For example, physicists assume that space and time did not exist shortly after the Big Bang of our universe, but only physical states existed. The familiar structure of our space-time was only created later on by a stochastic process. Furthermore, below the Planck length it is assumed that space and time loose their classical properties in the setting of quantum gravity. Therefore, noncommutative geometry is one of the candidates for creating a mathematical theory of quantum gravity. Noncommutative geometry is based on so-called spectral triplets for elliptic Dirac operators. As an introduction to noncommutative geometry, we recommend:

> J. Várilly, Lectures on Noncommutative Geometry, European Mathematical Society 2006.
>
> M. Paschke, An essay on the spectral action principle and its relation to quantum gravity, pp. 127–150. In: B. Fauser, J. Tolksdorf, and E. Zeidler (Eds.), Quantum Gravity: Mathematical Models and Experimental Bounds, Birkhäuser, Basel, 2006.

[183] J. Gracia-Bondia, J. Várilly, and H. Figueroa, Elements of Noncommutative Geometry, Birkhäuser, Boston, 2001 (reprinted with permission).

[184] This means that the C^*-ideal $\mathcal{J}_P$ cannot be extended to a larger C^*-ideal of $C(M)$ which is different from the trivial ideal $C(M)$.

We also refer to the comprehensive monograph:

M. Gracia-Bondia, J. Várilly, and H. Figueroa, Elements of Noncommutative Geometry, Birkhäuser, Boston, 2001.

A detailed study of the applications of noncommutative geometry to the Standard Model in particle physics can be found in the comprehensive monograph:

A. Connes and M. Marcolli, Noncommutative Geometry, Quantum Fields, and Motives, Amer. Math. Soc., Rhode Island, 2008.
Internet: `http://www.math.fsu.edu/~marcolli/bookjune4.pdf`

In the 1930s, John von Neumann discovered that operator algebras play a fundamental role in the mathematical formulation of quantum mechanics. Noncommutative geometry stands in this tradition and allows us to approach the Standard Model in elementary particle physics.

In 2006 the first volume of the *Journal of Noncommutative Geometry* appeared. The editor-in-chief is Alain Connes. The following list of topics covered by the journal shows the scope of modern noncommutative geometry:

- operator algebras,
- Hochschild and cyclic cohomology,
- K-theory and index theory,
- measure theory and topology of noncommutative spaces,
- spectral geometry of noncommutative spaces,
- noncommutative algebraic geometry,
- Hopf algebras and quantum groups,
- foliations, gruppoids, stacks, gerbes,
- deformations and quantizations,
- noncommutative spaces in number theory and arithmetic geometry,
- noncommutative geometry in physics: quantum field theory, renormalization, gauge theory, string theory, gravity, mirror symmetry, solid state physics, statistical mechanics.

7.20 Jordan Algebras

Let $\mathcal{O}(X)$ denote the set of all observables in $L(X, X)$ (i.e., the set of all linear continuous self-adjoint operators $A : X \to X$), where X is a complex Hilbert space. If $A, B \in \mathcal{O}(X)$, then the usual operator product AB is contained in $\mathcal{O}(X)$ iff $AB = BA$. This follows from

$$(AB)^* = B^* A^* = BA.$$

Thus, as a rule, $\mathcal{O}(X)$ is not an algebra with respect to the operator product. In order to cure this defect, Pascal Jordan (1902–1980) introduced the product

$$A \circ B := \tfrac{1}{2}(AB + BA).$$

Then the set $\mathcal{O}(X)$ becomes a real algebra with respect to the real linear combinations $\alpha A + \beta B$ and the Jordan product $A \circ B$. This commutative algebra of observables is called the real Jordan algebra of the Hilbert space X. As a rule, Jordan algebras are not associative.[185] The theory of Jordan algebras is a branch

[185] P. Jordan, On the multiplication of quantum-mechanical quantities I, II, Z. Phys. **80** (1933), 285–291; **87** (1934), 505–512 (in German).
P. Jordan, J. von Neumann, and N. Wigner, On an algebraic generalization of the quantum mechanical formalism, Ann. Math. **35** (1934), 29–64.

of modern mathematics.[186] In 1994, Zelmanov was awarded the Fields medal for his contributions to Jordan algebras.

7.21 The Supersymmetric Harmonic Oscillator

> The foundations of the theory of commuting and anticommuting variables were laid by Schwinger in 1953, who presented the analysis for commuting and anticommuting variables on the physical level of strictness[187]...
> The first mathematical formalism that made it possible to operate with commuting and anticommuting coordinates was Martin's algebraic formalism proposed in 1959[188]...
> In 1974, Salam and Strathdee proposed a very apt name for a set of superpoints.[189] After this work and the work by Wess and Zumino[190] were published, the superspace became a foundation for the most important physical theories.[191]
>
> Andrei Khrennikov, 1997

> In contrast to Heisenberg's harmonic quantum oscillator, the ground state energy of the supersymmetric harmonic oscillator is equal to zero.
>
> The golden rule of supersymmetry

> Supersymmetry is a relativistic symmetry between bosons and fermions. This is the only known way available at the present to unify the four-dimensional space-time and internal symmetries of the S-matrix in relativistic particle theory. [192]
>
> Prem Srivasta, 1985

Supersymmetry describes bosons and fermions in a unified way. Recall that the bosonic harmonic oscillator has the energy values

[186] H. Upmeier, Jordan Algebras in Analysis, Operator Theory, and Quantum Mechanics, Amer. Math. Soc., Rhode Island, 1987.
T. Springer and F. Veldkamp, Octonions, Jordan Algebras, and Exceptional Groups, Springer, Berlin, 2000.
K. McCrimmon, A Taste of Jordan Algebras, Springer, New York, 2004.

[187] J. Schwinger, Note on the quantum dynamical principle, Phil. Mag. **44** (1953), 1171–1193.

[188] J. Martin, Generalized classical analysis and "classical" analogue of a Fermi oscillator, Proc. Royal Soc. **A251** (1959), 536–542; The Feynman principle for a Fermi system, Proc. Royal Soc. **A251** (1959), 543–549.

[189] A. Salam and J. Strathdee, Supergauge transformations, Nucl. Phys. **B76** (1974), 477–483; Feynman rules for superfields, Nucl. Phys. **B86** (1975), 142–152.

[190] J. Wess and B. Zumino, Supergauge transformations in four dimensions, Nucl. Phys. **B70** (1974), 39–50.

[191] A. Khrennikov, Superanalysis, Kluwer, Dordrecht, 1997 (reprinted with permission).

[192] R. Haag, J. Lopuszanski, and M. Sohnius, All possible generators of supersymmetries of the S-matrix, Nucl. Phys. **B88** (1975), 257–274.
The supersymmetric Standard Model in particle physics is studied in S. Weinberg, Quantum Field Theory, Vol. 3, Cambridge University Press, 1995.
P. Srivasta, Supersymmetry, Superfields and Supergravity, Adam Hilger, Bristol, 1985.

$$E_n = \hbar\omega(n + \tfrac{1}{2}), \qquad n = 0, 1, 2, \ldots \tag{7.306}$$

As we will show below, the energy levels of the supersymmetric harmonic oscillator are given by

$$\boxed{E_{n_b,n_f} = \hbar\omega(n_b + n_f), \qquad n_b = 0, 1, 2, \ldots,\ n_f = 0, 1.}$$

In terms of physics, this is the energy of n_b bosons and n_f fermions. The point is that an infinite number of bosonic harmonic oscillators has the ground state energy

$$\sum_{k=0}^{\infty} \tfrac{1}{2}\hbar\omega = +\infty.$$

This causes the main trouble in quantum field theory. In contrast to this pathological situation, the ground state energy of an arbitrary number of supersymmetric harmonic oscillators is equal to zero, since $E_{0,0} = 0$.

A supersymmetric harmonic oscillator is the superposition of a bosonic harmonic oscillator and a fermionic harmonic oscillator. The nonzero ground state energies of the two harmonic oscillators compensate each other.

Because of the Pauli principle, it is not possible that two fermions are in the same energy state of a harmonic oscillator. This motivates why the number n_f of fermions in an energy eigenstate only attains the values $n_f = 0, 1$.

The supersymmetric Hamiltonian. Let us introduce the following Hamiltonians:

(B) Bosonic Hamiltonian: $H_{\text{bosonic}} := \hbar\omega(a^\dagger a + \frac{1}{2})$.
(F) Fermionic Hamiltonian: $H_{\text{fermionic}} := \hbar\omega(b^\dagger b - \frac{1}{2})$.
(S) Supersymmetric Hamiltonian:

$$\boxed{H_{\text{super}} := \hbar\omega(a^\dagger a \otimes I + I \otimes b^\dagger b).} \tag{7.307}$$

As a rule, physicists briefly write $H_{\text{super}} = \hbar\omega(a^\dagger a + b^\dagger b)$.

Hilbert spaces. The bosonic Hamiltonian acts on the so-called bosonic Hilbert space $X_{\text{bosonic}} := L_2(\mathbb{R})$ with the energy eigenstates

$$H_{\text{bosonic}}|n_b\rangle = E_{n_b}|n_b\rangle, \qquad n_b = 0, 1, 2, \ldots$$

where $E_{n_b} := \hbar\omega(n_b + \frac{1}{2})$. The bosonic eigenstates

$$|n_b\rangle, \qquad n_b = 0, 1, 2, \ldots$$

form a complete orthonormal system on the Hilbert space X_{bosonic}. In terms of physics, the state $|n_b\rangle$ describes n_b bosons.

The fermionic Hamiltonian acts on the Hilbert space $X_{\text{fermionic}} := \mathbb{C}^2$ with the energy eigenstates

$$H_{\text{fermionic}}|n_f\rangle = \mathcal{E}_{n_f}|n_f\rangle, \qquad n_f = 0, 1$$

where $\mathcal{E}_{n_f} := \hbar\omega(n_f - \frac{1}{2})$. The explicit form of the states $|0\rangle$ and $|1\rangle$ will be given below. The state $|n_f\rangle$ corresponds to n_f fermions.

Bosonic-fermionic states. Let us now introduce the Hilbert space

$$X_{\text{super}} := X_{\text{bosonic}} \otimes X_{\text{fermionic}}.$$

The states

$$|n_b\rangle \otimes |n_f\rangle, \qquad n_b = 0, 1, 2, \ldots, \quad n_f = 0, 1$$

form a complete orthonormal system of X_{super}. In terms of physics, the state $|n_b\rangle \otimes |n_f\rangle$ corresponds to n_b bosons and n_f fermions.[193] For the supersymmetric Hamiltonian, we get

$$H_{\text{super}} = H_{\text{bosonic}} \otimes H_{\text{fermionic}}.$$

This operator acts on the Hilbert space X_{super}. Explicitly,

$$H_{\text{super}}(|n_b\rangle \otimes |n_f\rangle) = H_{\text{bosonic}}|n_b\rangle \otimes |n_f\rangle + |n_b\rangle \otimes H_{\text{fermionic}}|n_f\rangle.$$

Hence

$$H_{\text{super}}(|n_b\rangle \otimes |n_f\rangle) = E_{n_b,n_f}(|n_b\rangle \otimes |n_f\rangle)$$

along with the energies

$$\boxed{E_{n_b,n_f} := \hbar\omega(n_b + n_f)}$$

where $n_b = 0, 1, 2, \ldots$ and $n_f = 0, 1$. This implies that the ground state $|0\rangle \otimes |0\rangle$ of the supersymmetric harmonic oscillator has zero energy, that is,

$$H_{\text{super}}(|0\rangle \otimes |0\rangle) = 0.$$

Bosonic creation and annihilation operators. Set $a^- := a$ and $a^+ := a^\dagger$. For the bosonic annihilation operator a^- and the bosonic creation operator a^+, we have[194]

$$[a^-, a^+]_- = I, \qquad [a^-, a^-]_- = [a^+, a^+]_- = 0.$$

Furthermore, by Sect. 7.2 on page 432, for $n = 0, 1, 2, \ldots$ we have

$$a^-|n+1\rangle = \sqrt{n+1}\ |n\rangle, \qquad a^+|n\rangle = \sqrt{n+1}\ |n+1\rangle.$$

Fermionic creation and annihilation operators. For the fermionic annihilation operator b^- and the fermionic creation operator b^+, we get

$$[b^-, b^+]_+ = I, \qquad [b^-, b^-]_+ = [b^+, b^+]_+ = 0.$$

In particular, $(b^+)^2 = (b^-)^2 = 0$. Furthermore,

$$b^-|0\rangle = 0, \quad b^-|1\rangle = |0\rangle, \quad b^+|0\rangle = |1\rangle, \quad b^+|1\rangle = 0.$$

Explicitly, we set

$$|0\rangle := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |1\rangle := \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The states $|0\rangle$ and $|1\rangle$ form a complete orthonormal system of the Hilbert space $X_{\text{fermionic}}$. That is, each element $\chi \in X_{\text{fermionic}}$ can be uniquely represented as

$$\chi = \alpha|0\rangle + \beta|1\rangle, \qquad \alpha, \beta \in \mathbb{C}.$$

[193] Physicists briefly write $|n_b\rangle|n_f\rangle$.
[194] Recall that $[A, B]_\pm := AB \pm BA$.

In the language of matrices, $\chi = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$. Moreover, we define

$$b^- := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad b^+ := \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad N := b^+b^- = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Since $N|0\rangle = 0$ and $N|1\rangle = |1\rangle$, the operator N is called the fermionic particle number operator. Obviously, b^+ is the adjoint matrix to b^-, i.e., $b^+ = (b^-)^\dagger$.

Supersymmetric creation and annihilation operators. We want to write the supersymmetric Hamiltonian in the form

$$\boxed{H_{\text{super}} = \hbar\omega(Q^+Q^- + Q^-Q^+).} \tag{7.308}$$

To this end, we introduce the operators

$$Q^+ := a^- \otimes b^+, \qquad Q^- := a^+ \otimes b^-$$

called the supersymmetric creation operator Q^+ and the supersymmetric annihilation operator Q^-. Explicitly,

$$Q^+(|n_b\rangle \otimes |n_f\rangle) = a^-|n_b\rangle \otimes b^+|n_f\rangle$$

and

$$Q^-(|n_b\rangle \otimes |n_f\rangle) = a^+|n_b\rangle \otimes b^-|n_f\rangle.$$

For example,

$$Q^+(|1\rangle \otimes |0\rangle) = |0\rangle \otimes |1\rangle, \qquad Q^-(|0\rangle \otimes |1\rangle) = |1\rangle \otimes |0\rangle.$$

Thus, the operator Q^+ sends one boson to one fermion (resp. the operator Q^- sends one fermion to one boson). We have

$$\boxed{[Q^+, Q^+]_+ = [Q^-, Q^-]_+ = 0.}$$

This is equivalent to $(Q^+)^2 = (Q^-)^2 = 0$. In fact,

$$Q^+Q^+ = (a^- \otimes b^+)(a^- \otimes b^+) = a^-a^- \otimes b^+b^+ = 0,$$

since $(b^+)^2 = 0$. Similarly, we get $(Q^-)^2 = 0$.

Let us now prove (7.308). It follows from

$$Q^+Q^- = (a^- \otimes b^+)(a^+ \otimes b^-) = a^-a^+ \otimes b^+b^-$$

together with $a^-a^+ = I + a^+a^-$ that

$$Q^+Q^- = (I + a^+a^-) \otimes b^+b^-.$$

Similarly,

$$Q^-Q^+ = (a^+ \otimes b^-)(a^- \otimes b^+) = a^+a^- \otimes b^-b^+ = a^+a^- \otimes (I - b^+b^-).$$

Therefore,

$$Q^+Q^- + Q^-Q^+ = I \otimes b^+b^- + a^+a^- \otimes I.$$

This yields (7.307), (7.308).

Supersymmetric invariance of the supersymmetric Hamiltonian. We claim that

$$\boxed{[H_{\text{super}}, Q^+]_- = [H_{\text{super}}, Q^-]_- = 0.}$$

To prove this, observe that $Q^+Q^+ = 0$. Hence

$$(Q^+Q^- + Q^-Q^+)Q^+ - Q^+(Q^+Q^- + Q^-Q^+) = Q^+Q^-Q^+ - Q^+Q^-Q^+ = 0.$$

This implies $[H_{\text{super}}, Q^+]_- = 0$. Similarly, $[H_{\text{super}}, Q^-]_- = 0$.

Perspective. Supersymmetry plays an important role in modern quantum field theory. We will come back to this in later volumes. There exists a huge amount of literature on supersymmetric models in quantum theory. Some hints for further reading can be found on page 679.

7.22 Hints for Further Reading

Textbooks on Quantum Mechanics

We refer to the following classic textbooks which use the language of physicists:

P. Dirac, The Principles of Quantum Mechanics, Clarendon Press, Oxford, 1930.

V. Fock, Fundamentals of Quantum Mechanics, Nauka, Moscow, 1931 (in Russian). (English edition: Mir, Moscow, 1978.)

R. Feynman, R. Leighton, and M. Sands, The Feynman Lectures in Physics, Addison-Wesley, Reading, Massachusetts, 1963.

L. Landau and E. Lifschitz, Course of Theoretical Physics, Vol. 3: Non-Relativistic Quantum Mechanics, Butterworth-Heinemann, Oxford, 1982.

J. Schwinger, Quantum Mechanics, Springer, New York, 2001.

F. Dyson, Advanced Quantum Mechanics, Dyson's 1951 Cornell Lecture Notes on Quantum Electrodynamics, Cornell University, Ithaca, New York. World Scientific, Singapore, 2007.

Much material can be found in the following handbooks:

G. Drake (Ed.), Springer Handbook of Atomic, Molecular, and Optical Physics, Springer, Berlin, 2005.

Encyclopedia of Mathematical Physics, Vols. 1–5. Edited by J. Françoise, G. Naber, and T. Tsun, Elsevier, Amsterdam, 2006.

Modern Encyclopedia of Mathematical Physics, Vols. 1, 2. Edited by I. Araf'eva and D. Sternheimer, Springer, Berlin, 2009 (to appear).

Furthermore, we recommend:

A. Messiah, Quantum Mechanics, Vols. 1, 2, North-Holland, Amsterdam, 1961.

J. Sakurai, Advanced Quantum Mechanics, Reading, Massachusetts, 1967.

L. Schiff, Quantum Mechanics, McGraw-Hill, New York, 1968.

M. Mizushima, Quantum Mechanics of Atomic Spectra and Atomic Structure, Benjamin, New York, 1970.

A. Galindo and P. Pascual, Quantum Mechanics, Vols. 1, 2, Springer, Berlin, 1990.

A. Bohm, Quantum Mechanics: Foundations and Applications, Springer, Berlin, 1994.

J. Sakurai and San Fu Tuan, Modern Quantum Mechanics, Benjamin and Cummings, New York, 1994.

E. Merzbacher, Quantum Mechanics, Wiley, New York, 1998.

J. Basdevant and J. Dalibard, Quantum Mechanics, Springer, Berlin, 2002.

F. Schwabl, Quantum Mechanics, Springer, Berlin, 2002.

F. Schwabl, Advanced Quantum Mechanics, Springer, Berlin, 2003.

N. Straumann, A Basic Course on Non-relativistic Quantum Mechanics, Springer, Berlin, 2002 (in German).

K. Gottfried and Tung-Mow Yan, Quantum Mechanics: Fundamentals, Springer, New York, 2003.

F. Scheck, Quantum Physics, Springer, Berlin, 2007.

Exercises can be found in:

S. Flügge, Practical Quantum Mechanics, Vols. 1, 2, Springer, Berlin, 1979.

J. Basdevant, The Quantum-Mechanics Solver: How to Apply Quantum Theory to Modern Physics, Springer, Berlin, 2000.

V. Radanovic, Problem Book in Quantum Field Theory, Springer, New York, 2006.

Visualizations of solutions in quantum mechanics are represented in:

S. Brandt and H. Dahmen, The Picture Book of Quantum Mechanics, Springer, New York, 1995.

B. Thaller, Visual Quantum Mechanics, Springer, New York, 2000.

B. Thaller, Advanced Visual Quantum Mechanics, Springer, New York, 2005.

Mathematical Methods in Quantum Mechanics

The classic monograph was written by

J. von Neumann, Mathematical Foundations of Quantum Mechanics (in German), Springer, Berlin, 1932. (English edition: Princeton University Press, 1955.)

Furthermore, we recommend:

M. Reed and B. Simon, Methods of Modern Mathematical Physics, Vols. 1–4, Academic Press, New York, 1972ff.

M. Schechter, Operator Methods in Quantum Mechanics, North-Holland, Amsterdam, 1982.

H. Triebel, Higher Analysis, Barth, Leipzig, 1989.

F. Berezin and M. Shubin, The Schrödinger Equation, Kluwer, Dordrecht, 1991.

E. Zeidler, Applied Functional Analysis: Applications to Mathematical Physics, Springer, New York, 1995.

W. Steeb, Hilbert Spaces, Wavelets, Generalized Functions and Modern Quantum Mechanics, Kluwer, Dordrecht, 1998.

W. Thirring, Quantum Mathematical Physics: Atoms, Molecules, and Large Systems, Springer, New York, 2002.

S. Gustafson and I. Sigal, Mathematical Concepts of Quantum Mechanics, Springer, Berlin, 2003.

A. Komech, Lectures on Quantum Mechanics (nonlinear PDE point of view), Lecture Notes No. 25 of the Max Planck Institute for Mathematics in the Sciences, Leipzig. Internet: `http://mis.mpg.de/preprints/ln/`

F. Strocchi, An Introduction to the Mathematical Structure of Quantum Mechanics: A Short Course for Mathematicians, Lecture Notes, Scuola Normale Superiore, Pisa (Italy), World Scientific, Singapore, 2005.

V. Varadarajan, Geometry of Quantum Theory, Springer, New York, 2007.

For the applications of Lie group theory to the spectra of atoms and molecules, we refer to the following classic monographs:

H. Weyl, The Theory of Groups and Quantum Mechanics, Springer, Berlin, 1929 (in German). (English edition: Dover, New York, 1931.)

B. van der Waerden, Group Theory and Quantum Mechanics, Springer, Berlin, 1932 (in German). (English edition: Springer, New York, 1974.)

See also the hints for further reading on axiomatic quantum field theory given on page 454. We also refer to a series of fundamental papers on the mathematical foundations of quantum mechanics and statistical physics written by

E. Lieb, The Stability of Matter: From Atoms to Stars, Selecta of Elliott Lieb. Edited by W. Thirring, Springer, New York, 2002.

E. Lieb, Inequalities: Selecta of Elliott Lieb. Edited by M. Loss, Springer, New York, 2002.

The Path Integral

It is crucial that there exist specific methods for the explicit computation of path integrals. This way, it is possible to obtain all of the explicitly known propagator kernels in quantum mechanics. We especially recommend

C. Grosche and F. Steiner, Handbook of Feynman Path Integrals, Springer, Berlin, 1998.

This comprehensive handbook contains a large list of known path integrals (200 pages), about 1000 references, and a detailed discussion of the historical background. Much material about the computation of path integrals can also be found in:

D. Khandekar, S. Lawande, and K. Bhagwat, Path-Integral Methods and their Applications, World Scientific, Singapore, 1993.

H. Kleinert, Path Integrals in Quantum Mechanics, Statistics, and Polymer Physics, World Scientific, River Edge, New York, 2004.

W. Dittrich and M. Reutter, Classical and Quantum Dynamics, Springer, Berlin, 1999.

M. Chaichian and A. Demichev, Path Integrals in Physics. Vol. 1: Stochastic Processes and Quantum Mechanics; Vol. 2: Quantum Field Theory, Statistical Physics, and other Modern Applications, Institute of Physics Publishing, Bristol, 2001.

For the language used in physics, we recommend:

R. Feynman and R. Hibbs, Quantum Mechanics and Path Integrals, McGraw-Hill, New York, 1965.

R. Feynman, Statistical Mechanics: A Set of Lectures, 14th edn., Addison Wesley, Reading, Massachusetts, 1998.

A. Zee, Quantum Field Theory in a Nutshell, Princeton University Press, 2003.

L. Faddeev and A. Slavnov, Gauge Fields, Benjamin, Reading, Massachusetts, 1980.

L. Faddeev, Elementary Introduction to Quantum Field Theory, Vol. 1, pp. 513–552. In: P. Deligne et al. (Eds.), Lectures on Quantum Field Theory, Vols. 1, 2, Amer. Math. Soc., Providence, Rhode Island, 1999.

M. Masujima, Path Integral Quantization and Stochastic Quantization, Springer, Berlin, 2000.

M. Marinov, Path integrals in quantum theory: an outlook of basic concepts, Phys. Rep. **60** (1) (1980), 1–57.

L. Schulman, Techniques and Applications of Path Integrals, Wiley, New York, 1981.

G. Roepstorff, Path Integral Approach to Quantum Physics, Springer-Verlag, New York, 1996.

J. Zinn-Justin, Quantum Field Theory and Critical Phenomena, 4th edn., Clarendon Press, Oxford, 2003 (extensive presentation of about 1000 pages based on the path-integral technique).

K. Fujikawa and H. Suzuki, Path Integrals and Quantum Anomalies, Oxford University Press, Oxford 2004.

For the language used in mathematics, we recommend:

B. Simon, Functional Integration and Quantum Physics, Academic Press, New York, 1979.

J. Glimm and A. Jaffe, Mathematical Methods of Quantum Physics: A Functional Integral Point of View, Springer, New York, 1981.

G. Johnson and M. Lapidus, M., The Feynman Integral and Feynman's Operational Calculus, Clarendon Press, Oxford, 2000.

S. Albeverio, R. Høegh-Krohn, and S. Mazzucchi, Mathematical Theory of the Feynman Path Integral: An Introduction, Springer, Berlin, 2006.

P. Cartier and C. DeWitt-Morette, Functional Integration: Action and Symmetries, Cambridge University Press, 2006

M. Freidlin, Functional Integration and Partial Differential Equations, Princeton University Press, 1985

together with

M. Kac, Wiener and integration in function spaces, Bull. Amer. Math. Soc. **72** (1966), 52–68.

I. Daubechies and J. Klauder, Constructing measures for path integrals, J. Math. Phys. **23** (1982), 1806–1822.

I. Daubechies and J. Klauder, Quantum-mechanical path integrals with Wiener measure for all polynomial Hamiltonians, Math. Phys. **26** (1985), 2239–2256.

J. Klauder, Beyond Conventional Quantization, Cambridge University Press, 2000.

For the application of spectral methods in physics, we refer to:

K. Kirsten, Spectral Functions in Mathematics and Physics, Chapman, Boca Raton, Florida, 2002

together with

E. Elizalde, Ten Physical Applications of Spectral Zeta Functions, Springer, Berlin, 1995.

A. Bytsenko, G. Cognola, E. Elizalde, V. Moretti, and S. Zerbini, Analytic Aspects of Quantum Fields, World Scientific, Singapore, 2003.

D. Vassilievich, Heat Kernel Expansion: Users' Manual, Physics Reports **388** (2003), 279-360.

In terms of mathematics, we recommend:

H. Edwards, Riemann's Zeta Function, Academic Press, New York, 1974.

P. Gilkey, Invariance Theory, the Heat Equation, and the Atiyah–Singer Index Theorem, CRC Press, Boca Raton, Florida, 1995.

P. Gilkey, P., Asymptotic Formulae in Spectral Geometry, Chapman, CRC Press, Boca Raton, Florida, 2003.

P. Gilkey, The spectral geometry of Dirac and Laplace type, pp. 289–326. In: Handbook of Global Analysis. Edited by D. Krupka and D. Saunders, Elsevier, Amsterdam, 2008.

Brownian Motion and the Wiener Integral

As an introduction, we recommend:

M. Mazo, Brownian Motion: Fluctuations, Dynamics, and Applications, Oxford University Press, 2002.

Y. Rozanov, Introductory Probability Theory, Prentice-Hall, Englewood Cliffs, New Jersey 1969.

L. Arnold, Stochastic Differential Equations, Krieger, Malabar, Florida, 1992.

L. Evans, An Introduction to Stochastic Differential Equations, Lectures held at the University of California at Berkeley, 2005.
Internet: `http://math.berkeley.edu/∼evans/SDE.course.pdf`

Furthermore, we recommend the following books:

W. Hakenbroch and A. Thalmaier, Stochastische Analysis, Teubner, Stuttgart, 1994 (in German).

B. Øksendal, Stochastic Differential Equations, Springer, Berlin, 2003.

E. Nelson, Dynamical Theories of Brownian Motion, Princeton University Press, Princeton, New Jersey, 1967.

K. Chung and Z. Zhao, From Brownian Motion to Schrödinger's Equation, Springer, New York, 1995.

B. Hughes, Random Walks and Random Environments, Vols. 1, 2, Clarendon Press, Oxford, 1995.

A. Borodin and P. Salminen, Handbook of Brownian Motion: Facts and Formulas, Birkhäuser, Basel, 2002.

P. Del Moral, Feynman–Kac Formulae, Springer, New York, 2004.

The history of the Feynman–Kac formula is described in:

M. Kac, Enigmas of Chance: An Autobiography, Harper & Row, New York, 1985.

We also refer to the following classic survey article:

S. Chandrasekhar, Stochastic problems in physics and astronomy, Rev. Mod. Phys. **15** (1943), 1–89.

The WKB Method

As an introduction to singular perturbation theory, we recommend:

W. Eckhaus, Asymptotic Analysis of Singular Perturbation, North-Holland, Amsterdam, 1979.

J. Kevorkian and J. Cole, Perturbation Methods in Applied Mathematics, Springer, New York, 1981.

A. Nayfeh, Perturbation Methods, Wiley, New York, 1973.

A. Nayfeh and B. Balachandran, Applied Nonlinear Dynamics: Analytical, Computational, and Experimental Methods, Wiley, New York, 1995.

Simple variants of the WKB method can be found in most textbooks on quantum mechanics (see page 667). As an introduction to the relation between classical mechanics and quantum mechanics, we refer to

W. Dittrich and M. Reutter, Classical and Quantum Dynamics, Springer, Berlin, 1999.

This concerns the explicit computation of numerous physical examples related to Schwinger's action principle, the Kolmogorov–Arnold–Moser (KAM) theory, the Maslov index, the Berry phase, and the Feynman path integral. As an introduction to the mathematics of the WKB method, we recommend the monographs by

V. Guillemin and S. Sternberg, Geometric Asymptotics, Amer. Math. Soc., Providence, Rhode Island, 1989.

V. Maslov and M. Fedoryuk, Semiclassical Approximation in Quantum Mechanics, Reidel, Dordrecht, 1981.

B. Helffer, Semiclassical Analysis, World Scientific, Singapore, 2003.

V. Nazaikinskii, B. Schulze, and B. Sternin, Quantization Methods in Differential Equations, Taylor & Francis, London, 2002.

The relation between the path integral and the WKB method is studied in the following monographs:

L. Schulmann, Techniques and Applications of Path Integrals, Wiley, New York, 1981.

C. Grosche and F. Steiner, Handbook of Feynman Path Integrals, Springer, Berlin, 1998.

H. Kleinert, Path Integrals in Quantum Mechanics, Statistics, and Polymer Physics, World Scientific, River Edge, New York, 2004.

The intellectual father of the global WKB method is Victor Maslov (born 1930). We refer to the following monographs:

V. Maslov, Théorie des perturbations et méthodes asymptotiques, Dunod, Paris, 1972 (in French).

J. Leray, Analyse Lagrangien et mécanique quantique: une structure mathématique apparantée aux développements asymtotiques et à l'indice de Maslov, Strasbourgh, France, 1978 (in French). (English edition: MIT Press, Cambridge, Massachusetts, 1981.)

Quantum chaos. Observe that the WKB method can also be applied to quantum chaos. This means that the corresponding classical dynamical system is chaotic. Here, Choquardt's expansion formula and Gutzwiller's trace formula are crucial.[195] This can be found in:

M. Gutzwiller, Chaos in Classical and Quantum Mechanics, Springer, New York, 1990.

C. Grosche and F. Steiner, Handbook of Feynman Path Integrals, Springer, New York, 1998.

Commutation Relations and the Stone–von Neumann Uniqueness Theorem

We recommend:

J. Rosenberg, A selective history of the Stone–von Neumann Theorem, Contemporary Mathematics **365** (2004), 123–158.

S. Summers, On the Stone–von Neumann uniqueness theorem and its ramifications, pp. 135–172. In: M. Rédei and M. Stöltzner (Eds.), John von Neumann and the Foundations of Quantum Physics, Kluwer, Dordrecht, 2000.

C. Putnam, Commutation Properties of Hilbert Space Operators and Related Topics, Springer, Berlin, 1967.

D. Petz, An Invitation to the Algebra of Canonical Commutation Relations, Leuven University Press, Leuven (Belgium), 1990.

V. Guillemin and S. Sternberg, Symplectic Techniques in Physics, Cambridge University Press, 1990.

N. Woodhouse, Geometric Quantization, Oxford University Press, 1997.

F. Berezin, The Method of Second Quantization, Academic Press, New York, 1966. (Second Russian edition: Nauka, Moscow, 1986.)

N. Bogoliubov et al., General Principles of Quantum Field Theory, Kluwer, Dordrecht, 1990.

Yu. Berezansky and V. Kondratiev, Spectral Methods in Infinite-Dimensional Analysis, Vols. 1, 2, Kluwer Dordrecht, 1995.

C. Bratelli and D. Robinson, Operator Algebras and Quantum Statistical Mechanics, Vols. 1, 2, Springer, New York, 2002.

The relations between classical mechanics and geometric quantization are studied in:

[195] P. Choquardt, Semi-classical approach to general forces in the setting of Feynman's path integral, Helv. Phys. Acta **28** (1955), 89–157 (in French).

R. Abraham and J. Marsden, Foundations of Mechanics, Addison-Wesley, Reading, Massachusetts, 1978.

Classical papers on commutation relations for a finite and an infinite number of operators are:

W. Heisenberg, Quantum-theoretical re-interpretation of kinematics and mechanical relations, Z. Physik **33** (1925), 879–893 (in German).[196]

M. Born and P. Jordan, On quantum mechanics, Z. Phys. **35** (1925), 858–888 (in German).

P. Dirac, The fundamental equations of quantum mechanics, Proc. Royal Soc. London Ser. A **109** (1926), no. 752, 642–653.

M. Born, P. Jordan, and W. Heisenberg, On quantum mechanics II, Z. Physik **35** (1926), 557–615 (in German).

W. Pauli, On the hydrogen spectrum from the standpoint of the new quantum mechanics, Z. Phys. **36** (1926), 336–365 (in German).

P. Jordan and E. Wigner, On the Pauli exclusion principle, Z. Phys. **47** (1928), 631–651 (in German).

H. Weyl, Quantum mechanics and group theory, Z. Phys. **46** (1928), 1–47 (in German).

M. Stone, Linear transformations in Hilbert space III, Proc. Nat. Acad. Sci. U.S.A. **16** (1930), 172–175.

J. von Neumann, The uniqueness of the Schrödinger operators, Math. Ann. **104** (1931), 570–578 (in German).

V. Fock, Configuration space and second quantization, Z. Phys. **75** (1932), 622–647 (in German).

H. Groenewold, On the principles of elementary quantum mechanics, Physica **12** (1946), 405–460.

A. Wintner, The unboundedness of quantum-mechanical matrices, Phys. Rev. **71** (2) (1947), 738–739.

H. Wielandt, On the unboundedness of the operators in quantum mechanics, Math. Ann. **121** (1949), 21–23 (in German).

G. Mackey, A theorem of Stone and von Neumann, Duke Math. J. **16** (1949), 313–326.

L. van Hove, Sur certaines representations unitaires d'un groupe infini de transformations. Mem. Acad. Royal Belgium (1951), 61–102.

J. Cook, The mathematics of second quantization, Trans. Amer. Math. Soc. **74** (1953)(2), 222–245.

K. Friedrichs, Mathematical Aspects of the Quantum Theory of Fields, Interscience Publishers, New York, 1953.

L. Gårding and A. Wightman, Representations of the anticommutation relations, Proc. Nat. Acad. Sci. U.S.A. **40** (1954), 617–621.

L. Gårding and A. Wightman, Representations of the commutation relations, Proc. Nat. Acad. Sci. U.S.A. **40** (1954), 622–625.

[196] The English translation of the classical papers by Born, Dirac, Jordan, Heisenberg, and Pauli can be found in B. van der Waerden (Ed.), Sources of Quantum Mechanics (1917–1926), Dover, New York, 1968.

I. Segal, Distributions in Hilbert space and canonical systems of operators, Trans. Amer. Math. Soc. **88** (1958), 12–42.

I. Segal, Quantization of nonlinear systems, J. Math. Phys. **1** (1960), 468–488.

I. Segal, Mathematical Problems of Relativistic Physics, Amer. Math. Soc. Providence, Rhode Island, 1963.

V. Bargmann, On a Hilbert space of analytic functions and an associated integral transform, Commun. Pure and Appl. Math. **14** (1961), 187–214.

G. Mackey, The Mathematical Foundations of Quantum Mechanics, Benjamin, New York, 1963.

A. Weil, Sur certains groupes d'opérateurs unitaires, Acta Math. **111** (1964), 143–211 (in French).

D. Kastler, The C^*-algebras of a free Boson field, Commun. Math. Phys. **1** (1965), 14–48.

G. Mackey, Induced Representations of Groups and Quantum Mechanics, Benjamin, New York, 1968.

M. Rieffel, On the uniqueness of the Heisenberg commutation relations, Duke Math. J. **39** (1972), 745–752.

G. Mackey, Unitary Group Representations in Physics, Probability, and Number Theory, Benjamin, Reading, Massachusetts, 1978.

R. Howe, On the role of the Heisenberg group in harmonic analysis, Bull. Amer. Math. Soc. (N.S.) **3**(2) (1980), 821–843.

H. Grosse and L. Pittner, A supersymmetric generalization of von Neumann's theorem, J. Math. Phys. **29**(1) (1988), 110–118.

G. Mackey, The Scope and History of Commutative and Noncommutative Harmonic Analysis, Amer. Math. Soc., Providence, Rhode Island, 1992.

A generalized version of the Stone–von Neumann uniqueness theorem plays a fundamental role in Ashtekhar's loop gravity (which represents an approach to quantum gravity). We refer to:

C. Fleischhack, Kinematical uniqueness of loop gravity, pp. 203–218. In: B. Fauser, J. Tolksdorf, and E. Zeidler (Eds.), Quantum Gravity, Birkhäuser, Basel, 2006.

Weyl Quantization

As a comprehensive introduction to deformation quantization in mathematics and physics, we recommend the following textbook which is based on the language of modern differential geometry (vector bundles, symplectic geometry, Poisson geometry, pseudo-differential operators):

S. Waldmann, Poisson Geometry and Deformation Quantization, Springer, Berlin, 2007 (in German).

For formal proofs based on the language of physicists, we refer to:

A. Hirshfeld and P. Henselder, Deformation quantization in the teaching of quantum mechanics, Am. J. Phys. **70** (2002), 537–547.

A. Hirshfeld and P. Henselder, Star products and perturbative quantum field theory, Ann. Phys. **298** (2002), 352–393.

F. Berezin and M. Shubin, The Schrödinger Equation, Kluwer, Dordrecht, 1991.

For rigorous proofs based on the language of mathematicians, we refer to:

L. Hörmander, The Weyl calculus of pseudo-differential operators, Commun. Pure Appl. Math. **32** (1979), 359–443.

M. de Gosson, Symplectic Geometry and Quantum Mechanics, Birkhäuser, Basel, 2006.

V. Nazaikinskii, B. Schulze, and B. Sternin, Quantization Methods in Differential Equations, Taylor & Francis, London, 2002.

Kontsevich proved the fundamental result that each Poisson manifold can be equipped with a formal Moyal star product. We refer to:

M. Kontsevich, Deformation quantization of Poisson manifolds, Lett. Math. Phys. **66**(3) (2003), 157–216.

A. Cattaneo and G. Felder, A path integral approach to the Kontsevich quantization formula, Commun. Math. Phys. **212** (2000), 591–611.

Concerning deformation quantization, we recommend the following books:

B. Fedosov, Deformation Quantization and Index Theory, Akademie-Verlag, Berlin, 1996.

A. Perelomov, Generalized Coherent States and Their Applications, Springer, Berlin, 1986.

M. Majid, Foundations of Quantum Group Theory, Cambridge University Press, 1995.

J. Madore, An Introduction to Noncommutative Differential Geometry and Its Applications, Cambridge University Press, 1995.

In addition, we recommend:

S. Podleś and S. Woronowicz, Quantum deformation of Lorentz group, Commun. Math. Phys. **130** (1990), 381–453.

M. Rieffel, Deformation quantization for actions of $\mathbb{R}^d$, Mem. Amer. Math. Soc. **106** (1993).

J. Wess, Gauge theories on noncommutative space-time treated by the Seiberg–Witten method, pp. 179–192. In: U. Carow-Watamura et al. (Eds.), Quantum Field Theory and Noncommutative Geometry, Springer, Berlin, 2005.

H. Grosse and R. Wulkenhaar, Renormalisation of φ^4-theory on noncommutative $\mathbb{R}^4$ in the matrix base, Commun. Math. Phys. **256** (2005), 305–374.

A survey on different quantization methods can be found in:

P. Bandyopadhyay, Geometry, Topology, and Quantization, Kluwer, Dordrecht, 1996.

S. Ali and M. Engliš, Quantization methods: a guide for physicists and analysts, 2004. Internet: `http://arxiv.org/math-ph/0405065`

N. Woodhouse, Geometric Quantization, Oxford University Press, 1997.

We also recommend:

J. Śniatycki, Geometric Quantization and Quantum Mechanics, Springer, New York, 1980.

N. Hurt, Geometric Quantization in Action, Reidel, Dordrecht, 1983.

S. Bates and A. Weinstein, Lectures on the Geometry of Quantization, Amer. Math. Soc., Providence, Rhode Island, 1997.

Furthermore, we refer to the following classic papers:

H. Weyl, The Theory of Groups and Quantum Mechanics, Dover, New York, 1931.

H. Groenewold, On the principles of elementary quantum mechanics, Physica **12** (1946), 405–460.

J. Moyal, Quantum mechanics as a statistical theory, Proc. Cambridge Phil. Soc. **45** (1949), 99–124.

M. Gerstenhaber, On the deformation of rings and algebras, Ann. Math. **79** (1964), 59–103.

F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer, Deformation theory and quantization I, II, Annals of Physics **111** (1978), 61–110; 111–151.

D. Sternheimer, Deformation quantization: twenty years after, pp. 107–145. In: Particles, Fields, and Gravitation, AIP Conf. Proc. vol. 453, American Institute of Physics, Woodbury, 1998.

F. Berezin, General concept of quantization, Commun. Math. Phys. **40** (1975), 153–174.

Statistical Physics

As an introduction to the vast literature on statistical physics, we recommend:

O. Bühler, A Brief Introduction to Classical, Statistical, and Quantum Mechanics, Courant Lecture Notes, Amer. Math. Soc., Providence, Rhode Island, 2006.

K. Huang, Statistical Physics, Wiley, New York, 1987.

R. Feynman, Statistical Mechanics: A Set of Lectures, 14th edn., Addison Wesley, Reading, Massachusetts, 1998.

A. Fetter and J. Walecka, Quantum Theory of Many-Particle Systems, McGraw-Hill, New York, 1971.

G. Mahan, Many-Particle Physics, Plenum Press, New York, 1990.

P. Martin and F. Rothen, Many-Body Problems and Quantum Field Theory, Springer, Berlin, 2002.

For the second law of thermodynamics, we refer to:

E. Lieb and J. Yngvason, A guide to entropy and the second law of thermodynamics, Notices Amer. Math. Soc. **45** (1998), 571–581.

E. Lieb and J. Yngvason, The physics and mathematics of the second law of thermodynamics, Physics Reports **310**(1) (1999), 1–96.

C^*-Algebras and von Neumann Algebras

Much material can be found in:

C. Bratelli and D. Robinson, Operator Algebras and Quantum Statistical Mechanics, Vols. 1, 2, Springer, New York, 2002.

B. Blackadar, Operator Algebras: C^*-Algebras and von Neumann Algebras. Springer, Berlin, 2005.

As an introduction, we recommend:

R. Haag, Local Quantum Physics: Fields, Particles, Algebras, Springer, New York, 1996.

H. Araki, Mathematical Theory of Quantum Fields, Oxford University Press, New York, 1999.

W. Thirring, Quantum Mathematical Physics: Atoms, Molecules, and Large Systems, Springer, New York, 2002.

Further applications to physics can be found in:

R. Streater and R. Wightman, PCT, Spin, Statistics, and All That, Benjamin, New York, 1968.

G. Emch, Algebraic Methods in Statistical Physics and Quantum Field Theory, Wiley, New York, 1972.

J. Glimm and A. Jaffe, Quantum Field Theory and Statistical Mechanics: Expositions, Birkhäuser, Boston, 1985.

B. Simon, The Statistical Theory of Lattice Gases, Princeton University Press, 1993.

F. Strocchi, An Introduction to the Mathematical Structure of Quantum Mechanics: A Short Course for Mathematicians, Lecture Notes, Scuola Normale Superiore, Pisa (Italy), World Scientific, Singapore, 2005.

K. Fredenhagen, K. Rehren, and E. Seiler, Quantum field theory: where we are. Lecture Notes in Physics **721** (2007), 61–87.
Internet: `http://arxiv.org/hep-th/0603155`

For the mathematical theory, we refer to:

K. Maurin, Generalized Eigenfunction Expansions and Unitary Representations of Topological Groups, PWN, Warsaw, 1968.

K. Maurin, Methods of Hilbert Spaces, PWN, Warsaw, 1972.

V. Sunder, An Invitation to von Neumann Algebras, Springer, New York, 1987.

J. Diximier, Von Neumann Algebras, North Holland, Amsterdam, 1981.

J. Diximier, C^*-Algebras, North-Holland, Amsterdam, 1982.

P. Kadison and J. Ringrose, Fundamentals of the Theory of Operator Algebras, Vols. 1–4, Academic Press, New York, 1983.

M. Takesaki, Tomita's Theory of Modular Hilbert Algebras and Its Applications, Springer, Berlin, 1970.

M. Takesaki, Theory of Operator Algebras, Vols. 1–3, Springer, New York, 1979.

M. Karoubi, K-Theory: An Introduction, Springer, Berlin, 1978.

For applications to noncommutative geometry, we recommend:

J. Várilly, Lectures on Noncommutative Geometry, European Mathematical Society, 2006.

M. Gracia-Bondia, J. Várilly, and H. Figueroa, Elements of Noncommutative Geometry, Birkhäuser, Boston, 2001.

Connes, A., Marcolli, M., Noncommutative Geometry, Quantum Fields, and Motives, Amer. Math. Soc., Providence, Rhode Island, 2008.
Internet: `http://www.math.fsu.edu/∼ marcolli/bookjune4.pdf`

There exist close relations between operator algebras and the realization of quantum groups (i.e., deformations of classical groups and symmetries). We refer to:

M. Majid, A Quantum Groups Primer, Cambridge University Press, 2002.

M. Majid, Foundations of Quantum Group Theory, Cambridge University Press, 1995.

Applications to quantum information can be found in:

M. Nielsen and I. Chuang, Quantum Computation and Quantum Information, Cambridge University Press, 2001.

Supersymmetry

As an introduction to supersymmetry including the Wess–Zumino model, we recommend:

P. Srivasta, Supersymmetry, Superfields, and Supergravity: An Introduction, Adam Hilger, Bristol, 1985.

H. Kalka and G. Soff, Supersymmetrie, Teubner-Verlag, Stuttgart, 1997 (in German).

L. Ryder, Quantum Field Theory, Cambridge University Press, 1999.

Shi-Hai Dong, Factorization Method in Quantum Mechanics, Springer, Dordrecht, 2007 (700 references).

Furthermore, we refer to:

J. Wess and J. Bagger, Supersymmetry and Supergravity, Princeton University Press, 1991.

G. Juncker, Supersymmetric Methods in Quantum and Statistical Physics, Springer, Berlin, 1996.

A. Khrennikov, Superanalysis, Kluwer, Dordrecht, 1997.

I. Buchbinder and S. Kuzenko, Ideas and Methods of Supersymmetry and Supergravity or a Walk Through Superspace, Institute of Physics, Bristol, 1998.

D. Freed, Five Lectures on Supersymmetry, Amer. Math. Soc., Providence, Rhode Island, 1999.

P. Deligne et al. (Eds.), Lectures on Quantum Field Theory: A Course for Mathematicians Given at the Institute for Advanced Study in Princeton, Vols. 1, 2, Amer. Math. Soc., Providence, Rhode Island, 1999.

V. Varadarajan, Supersymmetry for Mathematicians, Courant Lecture Notes, Amer. Math. Soc., Providence, Rhode Island, 2004.

J. Jost, Geometry and Physics, Springer, Berlin, 2008.

The supersymmetric Standard Model in particle physics can be found in:

S. Weinberg, Quantum Field Theory, Vol. 3, Cambridge University Press, 1995.

Applications of supersymmetry to cosmology:

P. Binétruy, Supersymmetry: Theory, Experiment, and Cosmology, Oxford University Press, 2006.

Applications of supersymmetry to solid state physics:

K. Efetov, Supersymmetry in Disorder and Chaos, Cambridge University Press, 1997.

Applications of supersymmetry to the genetic code in biology:

M. Forger and S. Sachse, Lie superalgebras and the multiplet structure of the genetic code, I. Codon representations, II. Branching rules, J. Math. Phys. **41** (2000), 5407–5422; 5423–5444.

History of Quantum Mechanics

B. van der Waerden (Ed.), Sources of Quantum Mechanics (1917–1926), Dover, New York, 1968.

P. Dirac, The Development of Quantum Mechanics, Gordon and Breach, New York, 1970.

J. Dieudonné, History of Functional Analysis, 1900–1975, North-Holland, Amsterdam, 1983.

J. Mehra and H. Rechenberg, The Historical Development of Quantum Mechanics, Vols. 1–6, Springer, New York, 2002.

S. Antoci and D. Liebscher, The third way to quantum mechanics (due to Feynman) is the forgotten first, Annales de Fondation Louis de Broglie **21** (1996), 349–368.

The Philosophy of Quantum Physics

R. Omnès, The Interpretation of Quantum Mechanics, Princeton University Press, 1994.

W. Heisenberg, Physics and Beyond: Encounters and Conversations, Harper and Row, New York, 1970.

P. Dirac, Directions in Physics, Wiley, New York, 1978.

Tian Yu Cao, Conceptual Developments of 20th Century Field Theories, Cambridge University Press, 1998.

Tian Yu Cao (Ed.), Conceptual Foundations of Quantum Field Theory, Cambridge University Press, 1999.

R. Penrose, The Road to Reality: A Complete Guide to the Laws of the Universe, Jonathan Cape, London, 2004.

The Cambridge Dictionary of Philosophy. Edited by R. Audi, Cambridge University Press, 2005.

Problems

In the first group of problems we want to show how to apply von Neumann's theory of self-adjoint (and essentially self-adjoint) operators to quantum mechanics. As prototypes, we will study the position operator Q, the momentum operator P, and the Hamiltonian H_{free} of a free quantum particle on the real line in Problems 7.5 and 7.15–7.17. Further typical examples can be found in Problem 7.19. Observe that in Problem 7.5, we will show that

The basic idea behind the notion of self-adjoint operator is the integration-by-parts formula and the extension of the classical derivative for functions to distributions (generalized functions).

The key observation is that the classical integration-by-parts formula for smooth functions with compact support,

$$\boxed{\int_{\mathbb{R}} \psi(x)\varphi'(x)dx = -\int_{\mathbb{R}} \psi'(x)\varphi(x)dx \qquad \text{for all} \quad \varphi, \psi \in \mathcal{D}(\mathbb{R}),} \tag{7.309}$$

remains valid if the derivatives φ', ψ' are to be understood in the sense of tempered distributions and the functions φ, ψ, as well as φ', ψ' are contained in the Hilbert space $L_2(\mathbb{R})$ of square-integrable functions (see Problem 7.3). This can be written as

$$\int_{\mathbb{R}} \psi(x)\varphi'(x)dx = -\int_{\mathbb{R}} \psi'(x)\varphi(x)dx \qquad \text{for all} \quad \varphi, \psi \in W_2^1(\mathbb{R}). \tag{7.310}$$

Let us introduce the two operators $A_{\text{pre}}\varphi := \varphi'$ for all $\varphi \in \mathcal{S}(\mathbb{R})$ and

$$A\varphi := \varphi' \qquad \text{for all} \quad \varphi \in W_2^1(\mathbb{R}).$$

Using the inner product $\langle f|g\rangle := \int_{\mathbb{R}} f(x)^\dagger g(x)dx$ on the Hilbert space $L_2(\mathbb{R})$, we get

$$\langle \psi|A_{\text{pre}}\varphi\rangle = -\langle A\psi|\varphi\rangle \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}),\ \psi \in W_2^1(\mathbb{R}). \tag{7.311}$$

Setting $P_{\text{pre}} := -\mathrm{i}\hbar A_{\text{pre}}$ and $P := -\mathrm{i}\hbar A$, formula (7.311) implies

(i) $\langle \psi|P_{\text{pre}}\varphi\rangle = \langle P_{\text{pre}}\psi|\varphi\rangle$ for all $\varphi, \psi \in \mathcal{S}(\mathbb{R})$,
(ii) $\langle \psi|P_{\text{pre}}\varphi\rangle = \langle P\psi|\varphi\rangle$ for all $\varphi \in \mathcal{S}(\mathbb{R}), \psi \in W_2^1(\mathbb{R})$, and
(iii) $\langle \psi|P\varphi\rangle = \langle P\psi|\varphi\rangle$ for all $\varphi, \psi \in W_2^1(\mathbb{R})$.

The three formulas (i)–(iii) display the basic ideas of von Neumann's functional-analytic theory for self-adjoint operators. We will show in Problem 7.5 that the formulas (i)–(iii) imply that $P_{\text{pre}}^* = P = P^*$. In the general case, let us consider the linear operator

$$\boxed{A : D(A) \to X} \tag{7.312}$$

whose domain of definition $D(A)$ is a linear dense subspace of the complex Hilbert space X. The linearity of A means that

$$A(\alpha\varphi + \beta\psi) = \alpha A\varphi + \beta A\psi \qquad \text{for all} \quad \varphi, \psi \in D(A),\ \ \alpha, \beta \in \mathbb{C}.$$

The density of the set $D(A)$ in the Hilbert space X means that, for any element $\varphi \in X$, there exists a sequence (φ_n) in $D(A)$ such that $\lim_{n\to\infty} \varphi_n = \varphi$ in X. Suppose that we are given two operators $B : D(B) \to X$ and $C : D(C) \to X$, where $D(B)$ and $D(C)$ are subsets of the space X.

- We write $B = C$ iff $D(B) = D(C)$ and $A\varphi = B\varphi$ for all $\varphi \in D(A)$.
- We write $B \subseteq C$ iff the operator $B : D(B) \to X$ is an extension of the operator C, that is, we have $D(A) \subseteq D(B) \subseteq X$ and $A\varphi = B\varphi$ for all $\varphi \in D(A)$.

7.1 *The smoothing technique (Friedrichs' mollification).* Let $\varphi \in L_2(\mathbb{R})$. For any positive real number $\varepsilon > 0$, we define

$$\varphi_\varepsilon(x) := \frac{1}{\varepsilon} \int_{\mathbb{R}} K\left(\frac{x-y}{\varepsilon}\right) \varphi(y) dy, \qquad x \in \mathbb{R}.$$

Here, we choose $K(x) := c \cdot \mathrm{e}^{-(1-x^2)^{-1}}$ if $|x| < 1$ and $K(x) := 0$ if $|x| \geq 1$. The positive constant c is chosen in such a way that $\int_{\mathbb{R}} K(x) dx = 1$. Prove that, for all $\varepsilon > 0$, the following hold:
(i) The smooth function φ_ε is contained in the Hilbert space $L_2(\mathbb{R})$.
(ii) $\lim_{\varepsilon \to +0} \int_{\mathbb{R}} |\varphi_\varepsilon(x) - \varphi(x)|^2 dx = 0$.
Hint: We refer to Zeidler (1995a), p. 186 (see the references on page 1049).

7.2 *The Sobolev space* $W_2^1(\mathbb{R})$. By definition, the function $\varphi : \mathbb{R} \to \mathbb{C}$ is contained in the space $W_2^1(\mathbb{R})$ iff $\varphi \in L_2(\mathbb{R})$, and the derivative φ' (in the sense of distributions) is also contained in $L_2(\mathbb{R})$. This means that

$$\int_{\mathbb{R}} \varphi'(x)\chi(x) dx = -\int \varphi(x)\chi'(x) dx$$

for all test functions $\chi \in \mathcal{D}(\mathbb{R})$. Prove the following:
(i) The Sobolev space $W_2^1(\mathbb{R})$ is a Hilbert space equipped with the inner product

$$\langle \chi | \varphi \rangle_{1,2} := \langle \chi | \varphi \rangle + \langle \chi' | \varphi' \rangle = \int_R \chi(x)^\dagger \varphi(x) dx + \int_{\mathbb{R}} \chi'(x)^\dagger \varphi'(x) dx$$

for all functions $\chi, \varphi \in W_2^1(\mathbb{R})$.[197]
(ii) The sets $\mathcal{D}(\mathbb{R})$ and $\mathcal{S}(\mathbb{R})$ are dense in $W_2^1(\mathbb{R})$.
(iii) The sets $\mathcal{D}(\mathbb{R})$ and $\mathcal{S}(\mathbb{R})$ are proper linear subspaces of the Sobolev space $W_2^1(\mathbb{R})$.
(iv) The function $\varphi : \mathbb{R} \to \mathbb{C}$ is contained in $W_2^1(\mathbb{R})$ iff $\varphi \in L_2(\mathbb{R})$ and the Fourier transform $\hat{\varphi}$ satisfies the condition[198]

$$\int_{\mathbb{R}} \left(|\hat{\varphi}(p)|^2 + |p\hat{\varphi}(p)|^2 \right) dp < \infty.$$

(v) If $\varphi, \chi \in W_2^1(\mathbb{R})$, then $\langle \chi | \varphi \rangle_{1,2} = \int_{\mathbb{R}} \hat{\chi}(p)^\dagger \hat{\varphi}(p) + p^2 \hat{\chi}(p)^\dagger \hat{\varphi}(p)\, dp$.
Hint: Use Problem 7.1. Concerning (iii), note that the function $\varphi(x) := |x| \mathrm{e}^{-x^2}$ has a derivative on the pointed set $\mathbb{R} \setminus \{0\}$ which is square integrable. Hence $\varphi \in W_2^1(\mathbb{R})$, but $\varphi \notin \mathcal{S}(\mathbb{R})$. The proofs can be found in Zeidler (1986), Vol. IIA, Chap. 21 (see the references on page 1049), together with much additional material.

7.3 *Integration by parts.* Prove that the generalized integration-by-parts formula (7.310) holds true.
Solution: Let $\varphi, \psi \in W_2^1(\mathbb{R})$. Since $\mathcal{D}(\mathbb{R})$ is dense in the Hilbert space $W_2^1(\mathbb{R})$, there exist sequences (φ_n) and (ψ_n) in $\mathcal{D}(\mathbb{R})$ such that $\varphi_n \to \varphi$ and $\psi_n \to \psi$ in $W_2^1(\mathbb{R})$ as $n \to \infty$. This means that

[197] Two functions φ and ψ are considered as the same element of the Hilbert space $W_2^1(\mathbb{R})$ iff $\varphi(x) = \psi(x)$ and $\varphi'(x) = \psi'(x)$ for all $x \in \mathbb{R}$, up to a set of Lebesgue measure zero.

[198] Recall that the Fourier transform of the derivative φ' is the product function $p \mapsto \mathrm{i}p\hat{\varphi}(p)$.

$$\varphi_n \to \varphi, \quad \varphi_n' \to \varphi', \quad \psi_n \to \psi, \quad \psi_n' \to \psi' \quad \text{in } L_2(\mathbb{R}) \text{ as } n \to \infty.$$

Letting $n \to \infty$, it follows from

$$\int_{\mathbb{R}} \psi_n(x)\varphi_n'(x)dx = -\int_{\mathbb{R}} \psi_n'(x)\varphi_n(x)dx$$

that $\int_{\mathbb{R}} \psi(x)\varphi'(x)dx = -\int_{\mathbb{R}} \psi'(x)\varphi(x)dx$.

7.4 *The adjoint operator.* The linear operator $A^\dagger : D(A) \to X$ is called the formally adjoint operator to the linear operator A from (7.312) iff

$$\boxed{\langle \psi | A\varphi \rangle = \langle A^\dagger \psi | \varphi \rangle \qquad \text{for all} \quad \varphi, \psi \in D(A).}$$

The operator $A : D(A) \to X$ is called formally self-adjoint (or symmetric) iff

$$\langle \psi | A\varphi \rangle = \langle A\psi | \varphi \rangle \qquad \text{for all} \quad \varphi, \psi \in D(A).$$

The more sophisticated definition of the adjoint operator $A^* : D(A^*) \to X$ is based on the formula

$$\boxed{\langle \psi | A\varphi \rangle = \langle A^* \psi | \varphi \rangle \qquad \text{for all} \quad \varphi \in D(A), \psi \in D(A^*).} \tag{7.313}$$

More precisely, we first define the set $D(A^*)$. The element ψ is contained in $D(A^*)$ iff there exists an element χ in X such that

$$\langle \psi | A\varphi \rangle = \langle \chi | \varphi \rangle \qquad \text{for all} \quad \varphi \in D(A).$$

We then define $A^*\psi := \chi$. This yields (7.313). The following two definitions are basic for quantum mechanics. Let $A : D(A) \to X$ be a formally self-adjoint operator of the form (7.312).

- The operator A is called self-adjoint iff $A = A^*$.
- The operator A is called essentially self-adjoint iff it has precisely one self-adjoint extension.

Show that the following hold:

(i) Both the formally adjoint operator $A^\dagger$ and the adjoint operator A^* are uniquely determined by the given operator A.

(ii) The adjoint operator A^* is linear.

(iii) If the formally adjoint operator $A^\dagger$ exists, then $A^\dagger \subseteq A^*$, that is, the operator A^* is an extension of $A^\dagger$.

(iv) The operator A is formally self-adjoint iff $A \subseteq A^*$.

Hint: We refer to Zeidler (1995a), Sect. 5.2 (see the references on page 1049).

7.5 *The prototype of a self-adjoint differential operator.* Define

$$P_{\text{pre}}\varphi := -\mathrm{i}\hbar\varphi' \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}),$$

and

$$P\varphi := -\mathrm{i}\hbar\varphi' \qquad \text{for all} \quad \varphi \in W_2^1(\mathbb{R}).$$

In the latter equation, the derivative is to be understood in the sense of tempered distributions. Note that $\varphi \in W_2^1(\mathbb{R})$ implies $P\varphi \in L_2(\mathbb{R})$. Prove the following:

(i) The operator $P_{\text{pre}} : \mathcal{S}(\mathbb{R}) \to L_2(\mathbb{R})$ is formally self-adjoint.

(ii) The adjoint operator P_{pre}^* coincides with P.

(iii) The operator $P : W_2^1(\mathbb{R}) \to L_2(\mathbb{R})$ is self-adjoint.

(iv) $P_{\text{pre}} \subseteq P_{\text{pre}}^{**} = P_{\text{pre}}^* = P$.

(v) The closure $\overline{P}_{\rm pre}$ of $P_{\rm pre}$ coincides with $P^{**}_{\rm pre}$ (see Problem 7.9).
Solution: Set $\hbar := 1$. By Problem 7.3,

$$\int_{\mathbb{R}} \psi^\dagger(-\mathrm{i}\varphi')dx = \int_{\mathbb{R}} (-\mathrm{i}\psi')^\dagger \varphi dx \qquad \text{for all} \quad \varphi, \psi \in W_2^1(\mathbb{R}). \quad (7.314)$$

Ad (i). By (7.314), $\langle\psi|P_{\rm pre}\varphi\rangle = \langle P_{\rm pre}\psi|\varphi\rangle$ for all $\varphi, \psi \in \mathcal{S}(\mathbb{R})$.
Ad (ii). By definition of the adjoint operator $P^*_{\rm pre}$, we have $\chi = P^*_{\rm pre}\psi$ iff $\psi, \chi \in L_2(\mathbb{R})$ and

$$\langle\psi|P_{\rm pre}\varphi\rangle = \langle\chi|\varphi\rangle \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).$$

Equivalently,

$$\int_{\mathbb{R}} \psi^\dagger(-\mathrm{i}\varphi')dx = \int_{\mathbb{R}} \chi^\dagger \varphi dx \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).$$

Passing over to conjugate complex values and setting $\varrho := -\mathrm{i}\varphi^\dagger$, we get

$$\int_{\mathbb{R}} \psi(-\varrho')dx = \int_{\mathbb{R}} (\mathrm{i}\chi)\varrho dx \qquad \text{for all} \quad \varrho \in \mathcal{S}(\mathbb{R}).$$

This means that $\frac{d}{dx}\psi = \mathrm{i}\chi$, in the sense of tempered distributions. Hence $\psi \in W_2^1(\mathbb{R})$, and $\chi = -\mathrm{i}\frac{d}{dx}\psi$. Therefore, $\chi = P\psi$.
Ad (iii). By (7.314), $\langle\psi|P\varphi\rangle = \langle P\psi|\varphi\rangle$ for all $\varphi, \psi \in W_2^1(\mathbb{R})$. Hence the operator P is formally self-adjoint. Suppose that, for fixed $\psi, \chi \in L_2(\mathbb{R})$, we have

$$\langle\psi|P\varphi\rangle = \langle\chi|\varphi\rangle \qquad \text{for all} \quad \varphi \in W_2^1(\mathbb{R}).$$

The same argument as in (ii) above shows that $P\psi = \chi$. Hence $P^*\psi = P\psi$ for all $\psi \in W_2^1(\mathbb{R})$.
Ad (iv). By definition, $P_{\rm pre} \subseteq P$. By (ii), (iii), we get $P^*_{\rm pre} = P$ and $P^* = P$.
Ad (v). Let (φ_n) be a sequence in $D(P_{\rm pre})$ with

$$\lim_{n\to\infty} \varphi_n = \varphi, \qquad \lim_{n\to\infty} P_{\rm pre}\varphi_n = \chi. \quad (7.315)$$

Then $\overline{P}_{\rm pre}\psi = \chi$. Letting $n \to \infty$, it follows from

$$\langle\varrho|P_{\rm pre}\varphi_n\rangle = \langle P\varrho|\varphi_n\rangle \qquad \text{for all} \quad \varrho \in \mathcal{S}(\mathbb{R})$$

that $\langle\varrho|\chi\rangle = \langle P\varrho|\varphi\rangle$ for all $\varrho \in \mathcal{S}(\mathbb{R})$. Hence $\chi = P\varphi$.
Conversely, if $\chi = P\varphi$, then there exists a sequence (φ_n) in $\mathcal{S}(\mathbb{R})$ with (7.315), by Problem 7.2(ii). Summarizing, $P\varphi = \overline{P}\varphi$ for all $\varphi \in W_2^1(\mathbb{R})$.

7.6 *Closed operators.* The subset

$$\mathrm{graph}(A) := \{(\varphi, A\varphi) : \varphi \in D(A)\}$$

of the product space $X \times X$ is called the graph of the operator A from (7.312). The operator A is defined to be closed iff the set $\mathrm{graph}(A)$ is closed in $X \times X$. This means that if there exists a sequence (φ_n) in $D(A)$ with the convergence property

$$\lim_{n\to\infty} \varphi_n = \varphi \quad \text{and} \quad \lim_{n\to\infty} A\varphi_n = \psi,$$

then $\varphi \in D(A)$ and $A\varphi = \psi$. This generalizes the notion of continuity.[199] Show that the adjoint operator $A^* : D(A) \to X$ from Problem 7.4 is closed.
Solution: Let $\varphi_n \in D(A^*)$ for all n, and let

$$\lim_{n\to\infty} \varphi_n = \varphi \quad \text{and} \quad \lim_{n\to\infty} A^*\varphi_n = \psi.$$

Then $\langle A^*\varphi_n|\chi\rangle = \langle\varphi_n|A\chi\rangle$. Letting $n \to \infty$, we get

$$\langle\psi|\chi\rangle = \langle\varphi|A\chi\rangle \qquad \text{for all} \quad \chi \in D(A).$$

Hence $\varphi \in D(A^*)$ and $\psi = A^*\varphi$.

7.7 *The crucial symmetry criterion for self-adjoint operators.* Prove that the linear, densely defined operator $A : D(A) \to X$ on the complex Hilbert space X is self-adjoint iff the following two conditions are satisfied:
(i) $\langle\psi|A\varphi\rangle = \langle A\psi|\varphi\rangle$ for all $\varphi, \psi \in D(A)$.
(ii) If $\langle\psi|A\varphi\rangle = \langle\chi|\varphi\rangle$ for fixed $\psi, \chi \in X$ and all $\varphi \in D(A)$, then $\psi \in D(A)$.
Solution: (I) $\Rightarrow$: Assume that A is self-adjoint. Then $A = A^*$. This implies (i). If $\langle\psi|A\varphi\rangle = \langle\chi|\varphi\rangle$ for all $\varphi \in D(A)$, then $\psi \in D(A^*)$. Hence $\psi \in D(A)$.
(II) $\Leftarrow$: Assume that (i) and (ii) hold. By (i), $A \subseteq A^*$. In order to show $A^* \subseteq A$, let $\psi \in D(A^*)$. Then

$$\langle\psi|A\varphi\rangle = \langle A^*\psi|\varphi\rangle \qquad \text{for all} \quad \varphi \in D(A).$$

By (ii), $\psi \in D(A)$. It follows from (i) that $\langle A^*\psi|\varphi\rangle = \langle A\psi|\varphi\rangle$. Hence

$$\langle A^*\psi - A\psi|\varphi\rangle = 0 \qquad \text{for all} \quad \varphi \in D(A).$$

Since $D(A)$ is dense in X, we get $A^*\psi = A\psi$. □
In the following problems we want to show that

> *The properties of self-adjointness and essential self-adjointness are closely related to natural extension properties of formally self-adjoint operators A based on the inclusions $A \subseteq \overline{A} \subseteq A^*$, where $\overline{A}$ denotes the closure of A. In addition, $\overline{A} = A^{**}$.*

7.8 *Maximal extension and the adjoint operator.* Let $A : D(A) \to X$ be a formally self-adjoint operator of the form (7.312). Show the following:
(i) There exists a maximal linear extension $B : D(B) \to X$ of A with

$$\langle\psi|A\varphi\rangle = \langle B\psi|\varphi\rangle \qquad \text{for all} \quad \varphi \in D(A), \psi \in D(B).$$

This maximal extension B is equal to the adjoint operator A^*.
(ii) The operator A is self-adjoint iff the maximal extension B coincides with A, that is, $B = A$.
Hint: Convince yourself that this is merely a reformulation of the basic definitions.

7.9 *Minimal extension and the closure.* Let $A : D(A) \to X$ be a formally self-adjoint operator of the form (7.312). Show that the operator A can be minimally extended to a linear, closed, formally self-adjoint operator. This operator is denoted by $\overline{A} : D(\overline{A}) \to X$, and it is called the closure of A.
Hint: Let D_{cl} be the set of all $\varphi \in X$ for which a sequence (φ_n) exists in $D(A)$ such that

[199] Banach's closed graph theorem tells us the crucial fact that a linear closed operator $A : X \to X$ defined on the total Hilbert space X is continuous. However, the self-adjoint Hamiltonian operators arising in quantum mechanics are not defined on the total Hilbert space; as a rule, they are not continuous, but they are closed.

- $\lim_{n\to\infty} \varphi_n = \varphi$ and
- $(A\varphi_n)$ is convergent, that is, $\lim_{n\to\infty} A\varphi_n = \psi$.

Letting $n \to \infty$, it follows from $\langle \chi | A\varphi_n \rangle = \langle A\chi | \varphi_n \rangle$ that

$$\langle \chi | \psi \rangle = \langle A\chi | \varphi \rangle \qquad \text{for all} \quad \chi \in D(A).$$

Since $D(A)$ is dense in X, the element ψ of X is uniquely determined by φ. Now we set $\overline{A}\varphi := \psi$ and $D(\overline{A}) := D_{\mathrm{cl}}$. Since

$$\langle \chi | \overline{A}\varphi \rangle = \langle A\chi | \varphi \rangle \qquad \text{for all} \quad \chi \in D(A),\ \varphi \in D(\overline{A}), \tag{7.316}$$

we get $A \subseteq \overline{A} \subseteq A^*$. Let $\varrho \in D(\overline{A})$. Then there exists a sequence (ϱ_n) in $D(A)$ such that $\lim_{n\to\infty} \chi_n = \varrho$. Considering (7.316) with $\chi := \varrho_n$ and letting $n \to \infty$, we obtain

$$\langle \varrho | \overline{A}\varphi \rangle = \langle \overline{A}\varrho | \varphi \rangle \qquad \text{for all} \quad \varrho, \varphi \in D(\overline{A}).$$

Thus, the operator $\overline{A}$ is formally self-adjoint. Finally, it remains to show that the operator $\overline{A}$ is closed (see H. Triebel, Higher Analysis, Sect. 17, Barth, Leipzig, 1989).

7.10 *Properties of the closure.* Let $A : D(A) \to X$ and $B : D(B) \to X$ be formally self-adjoint operators of the form (7.312) on page 681. Show the following:
(i) $A \subseteq B$ implies $\overline{A} \subseteq \overline{B}$ and $B^* \subseteq A^*$.
(ii) $A \subseteq \overline{A} \subseteq A^*$.
(iii) $\overline{A} = A^{**}$ and $(\overline{A})^* = A^*$.
(iii) The operator A is essentially self-adjoint iff the closure $\overline{A}$ is self-adjoint.
Hint: We refer to Zeidler (1995a), p. 415ff (see the references on page 1049).

7.11 *General properties of self-adjoint operators.* For the formally self-adjoint operator A of the form (7.312), the following statements are equivalent:
(i) The operator A is self-adjoint.
(ii) All the non-real numbers z belong to the resolvent set $\varrho(A)$.
(iv) $\mathrm{im}(\pm \mathrm{i}I - A) = X$.
(iv) The operator A is closed and $\ker(\pm \mathrm{i}I - A^*) = \{0\}$.
(v) The operator A is essentially self-adjoint and closed.
Hint: See Zeidler (1995a), p. 416.

7.12 *General properties of essentially self-adjoint operators.* For the formally self-adjoint operator operator A of the form (7.312), the following statements are equivalent:
(i) The operator A is essentially self-adjoint.
(ii) The closure $\overline{A}$ is self-adjoint.
(iv) The two sets $\mathrm{im}(\pm \mathrm{i}I - A)$ are dense in X.
(iii) $\ker(\pm \mathrm{i}I - A^*) = \{0\}$.
Hint: See Zeidler (1995a), p. 424.

7.13 *Further properties of essentially self-adjoint operators.* Let $A : D(A) \to X$ be a linear, formally self-adjoint, and densely defined operator on the complex Hilbert space X. Assume that the operator A is essentially self-adjoint, and let $B : D(B) \to X$ be the uniquely determined self-adjoint extension of A. Prove that

$$A^* = \overline{A} = A^{**} = B. \tag{7.317}$$

Solution: By Problem 7.10, $B = \overline{A}$. Moreover, $\overline{A} = A^{**}$ and $A^* = (\overline{A})^* = \overline{A}$.

7.14 *Unitary invariance.* The linear operator $A : D(A) \to X$ is said to be unitarily equivalent to the linear operator $B : D(B) \to X$ iff there exists a unitary operator $U : X \to Y$ from the complex Hilbert space X onto the complex Hilbert space Y such that the diagram

$$\begin{array}{ccc} D(A) & \xrightarrow{A} & X \\ {\scriptstyle U}\downarrow & & \downarrow{\scriptstyle U} \\ D(B) & \xrightarrow{B} & Y \end{array}$$

is commutative. This means that $D(B) = UD(A)$ and $B = UAU^{-1}$. Show that the following notions are invariant under this transformation: formally self-adjoint, self-adjoint, essentially self-adjoint, and closed.
Hint: Use the corresponding definitions.

7.15 *The position operator on the real line.* Set

$$D(Q) := \{\varphi \in L_2(\mathbb{R}) : \int_{\mathbb{R}} |x\varphi(x)|^2 dx < \infty\}.$$

Fix $x \in \mathbb{R}$. Define $(Q_{\rm pre}\varphi)(x) := x\varphi(x)$ for all $\varphi \in \mathcal{S}(\mathbb{R})$, and $(Q\varphi)(x) := x\varphi(x)$ for all $\varphi \in D(Q)$. Prove the following:
(i) The operator $Q_{\rm pre} : \mathcal{S}(\mathbb{R}) \to L_2(\mathbb{R})$ is formally self-adjoint.
(ii) The operator $Q : D(Q) \to L_2(\mathbb{R})$ is self-adjoint.
(iii) The operator $Q_{\rm pre} : \mathcal{S}(\mathbb{R}) \to L_2(\mathbb{R})$ is essentially self-adjoint, but not self-adjoint.
(iv) $Q^*_{\rm pre} = \overline{Q}_{\rm pre} = Q^{**}_{\rm pre} = Q$.
Solution: Ad (i). For all $\varphi, \psi \in \mathcal{S}(\mathbb{R})$,

$$\int_{\mathbb{R}} (x\psi(x))^\dagger \varphi(x) dx = \int \psi(x)^\dagger x\varphi(x) dx.$$

Ad(ii), (iii). For given function $f \in L_2(\mathbb{R})$, the equation

$$\pm \mathrm{i}\varphi - Q\varphi = f, \qquad \varphi \in D(Q) \tag{7.318}$$

has the unique solution $\varphi(x) := \frac{f(x)}{\pm \mathrm{i} - x}$ for all $x \in \mathbb{R}$. In fact, $|\varphi(x)| \le \mathrm{const}|f(x)|$ for all $x \in \mathbb{R}$. This implies $\varphi \in L_2(\mathbb{R})$. Hence $\varphi \in D(Q)$. Thus, we get the key relation $\mathrm{im}(\pm \mathrm{i}I - Q) = L_2(\mathbb{R})$, that is, Q is self-adjoint.
In particular, if $f \in \mathcal{S}(\mathbb{R})$, then the solution of equation (7.318) is contained in $\mathcal{S}(\mathbb{R})$. Since the set $\mathcal{S}(\mathbb{R})$ is dense in $L_2(\mathbb{R})$, the sets $\mathrm{im}(\pm I - Q_{\rm pre})$ are dense in $L_2(\mathbb{R})$. Therefore, the operator $Q_{\rm pre}$ is essentially self-adjoint.
Finally, note that the set $D(Q_{\rm pre}) = \mathcal{S}(\mathbb{R})$ differs from $D(Q)$. For example, choose $\psi(x) := |x|\mathrm{e}^{-x^2}$. Then $\psi \in D(Q)$, but $\psi \notin \mathcal{S}(\mathbb{R})$.
Ad (iv). Use Problem 7.13.

7.16 *The momentum operator on the real line.* As in Problem 7.5, define

$$P_{\rm pre}\varphi := -\mathrm{i}\hbar\varphi' \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}),$$

and $P\varphi := -\mathrm{i}\hbar\varphi'$ for all $\varphi \in W_2^1(\mathbb{R})$. Use the Weyl equation

$$\boxed{\pm \mathrm{i}\varphi - P\varphi = f, \qquad \varphi \in \mathcal{S}(\mathbb{R})} \tag{7.319}$$

in order to prove the following:

(i) The operator $P_{\rm pre} : \mathcal{S}(\mathbb{R}) \to L_2(\mathbb{R})$ is formally self-adjoint.
(ii) The operator $P : W_2^1(\mathbb{R}) \to L_2(\mathbb{R})$ is self-adjoint.
(iii) The operator $P_{\rm pre} : \mathcal{S}(\mathbb{R}) \to L_2(\mathbb{R})$ is essentially self-adjoint, but not self-adjoint.
(iv) $P_{\rm pre}^* = \overline{P}_{\rm pre} = P_{\rm pre}^{**} = P$.
Solution: Ad (i). See Problem 7.5.
Ad (ii), (iii). For given $f \in \mathcal{S}(\mathbb{R})$, the equation (7.319) has a unique solution $\varphi \in \mathcal{S}(\mathbb{R})$. In fact, Fourier transformation yields

$$\pm \mathrm{i}\hat{\varphi}(p) - \hbar p \hat{\varphi}(p) = \hat{f}(p), \qquad p \in \mathbb{R}.$$

This yields $\hat{\varphi}(p) = \frac{\hat{f}(p)}{\pm \mathrm{i} - \hbar p}$ which is contained in $\mathcal{S}(\mathbb{R})$. Then the inverse Fourier transform yields the desired solution φ of (7.319). Since the set $\mathcal{S}(\mathbb{R})$ is dense in $L_2(\mathbb{R})$, the image set $\mathrm{im}(\pm I - P_{\rm pre})$ is dense in $L_2(\mathbb{R})$. Consequently, the operator $P_{\rm pre} : \mathcal{S}(\mathbb{R}) \to L_2(\mathbb{R})$ is essentially self-adjoint. Thus, it has a unique self-adjoint extension.
Using the extended Fourier transform $\mathcal{F} : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ together with Problem 7.2(iv), the same argument as above shows that, for given function $f \in L_2(\mathbb{R})$, the equation

$$\pm \mathrm{i}\varphi - P\varphi = f, \qquad \varphi \in W_2^1(\mathbb{R})$$

has a (unique) solution φ. Hence $\mathrm{im}(\pm \mathrm{i} I - P) = L_2(\mathbb{R})$. Consequently, the operator $P : W_2^1(\mathbb{R}) \to L_2(\mathbb{R})$ is self-adjoint. Furthermore, the operator P is the unique self-adjoint extension of the operator $P_{\rm pre}$. Since $\mathcal{S}(\mathbb{R}) \neq W_2^1(\mathbb{R})$, the operator $P_{\rm pre}$ differs from P.
Ad (iv). Use Problem 7.13.
Historical remarks. The importance of equations of the type (7.319) for the study of the spectral properties of ordinary differential equations was discovered by Weyl in 1910 and developed by von Neumann in his 1929 theory of deficiency indices.

- H. Weyl, On ordinary differential equations with singularities, Math. Ann. **68** (1910), 220–269 (in German).
- J. von Neumann, General spectral theory of Hermitean operators, Math. Ann. **102** (1929), 49–131 (in German).
- K. Kodaira, The eigenvalue problem for ordinary differential equations of the second order and Heisenberg's theory of S-matrices. Amer. J. Math. **71** (1949), 921–945.
- K. Jörgens and F. Rellich, Eigenvalue problems for ordinary differential equations, Springer, Berlin, 1976 (in German).

The Weyl–Kodaira theory will be studied in Vol. III, together with interesting physical applications.

7.17 *The Hamiltonian of the free quantum particle on the real line.* Define

$$H_{\rm pre}\varphi := -\frac{\hbar^2}{2m}\varphi'' \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}),$$

and

$$H_{\rm free}\varphi := -\frac{\hbar^2}{2m}\varphi'' \qquad \text{for all} \quad \varphi \in W_2^2(\mathbb{R}).$$

In the latter equation, the derivatives are to be understood in the sense of tempered distributions. If $\varphi \in W_2^2(\mathbb{R})$, then $H_{\rm free}\varphi \in L_2(\mathbb{R})$. Prove the following:
(i) The operator $H_{\rm pre} : \mathcal{S}(\mathbb{R}) \to L_2(\mathbb{R})$ is formally self-adjoint.
(ii) The operator $H_{\rm free} : W_2^2(\mathbb{R}) \to L_2(\mathbb{R})$ is self-adjoint.

(iii) $\overline{H}_{\rm pre} = H^*_{\rm pre} = H_{\rm free}$.
(iv) The operator $H_{\rm pre}$ is essentially self-adjoint, but not self-adjoint.
Hint: Apply integration by parts twice, and use analogous arguments as in Problem 7.7.

7.18 *Deficiency indices and von Neumann's extension theorem for self-adjoint operators.* Let $A : D(A) \to X$ be a linear, formally self-adjoint, densely defined, and closed operator on the complex Hilbert space X. The numbers

$$d_\pm := \dim\,(\pm \mathrm{i}I - A)^\perp$$

are called the deficiency indices of the operator A.[200] Prove the following:
(i) The operator A has a self-adjoint extension iff $d_+ = d_-$.
(ii) The operator A is self-adjoint iff $d_+ = d_- = 0$.
Hint: Use the Cayley transform in order to reduce this to the extension problem for isometric operators (see Problems 7.22 and 7.23).

7.19 *Formally self-adjoint operators which have no self-adjoint extension or infinitely many self-adjoint extensions.* Consider the operator

$$A\varphi := -\mathrm{i}\hbar \frac{d\varphi}{dx} \qquad \text{for all} \quad \varphi \in D(A)$$

where $D(A)$ is a linear dense subspace of the complex Hilbert space X. We will choose $X := L_2(0,\infty)$ or $X := L_2(0,1)$. We want to show that the properties of the operator A critically depend on the choice of the domain of definition $D(A)$. In turn, this depends on the choice of boundary conditions. Show that the following hold:
(i) Choose $D(A) := \mathcal{D}(0,\infty)$ and $X := L_2(0,\infty)$. [201] Then the operator A is formally self-adjoint, but it cannot be extended to a self-adjoint operator.
(ii) Fix the complex number α with $|\alpha| = 1$ and $\alpha \neq 1$. Choose[202]

$$D(A) := \{\varphi \in C^1[0,1] : \varphi(0) = \alpha\varphi(1)\}$$

and $X = L_2(0,1)$. Then the operator A is essentially self-adjoint.
(iii) Choose $D(A) := \{\varphi \in C^1[0,1] : \varphi(0) = \varphi(1) = 0\}$ and $X := L_2(0,1)$. Then the operator A is formally self-adjoint, but its closure $\overline{A}$ is not self-adjoint. However, the operator has an infinite number of self-adjoint extensions given by the operators from (ii).
Hint: Ad (i). Set $\hbar := 1$. Integration by parts yields

$$\langle\chi|A\varphi\rangle = \int_0^\infty \chi^\dagger(-\mathrm{i}\varphi')dx = \int_0^\infty (-\mathrm{i}\chi')^\dagger\varphi dx = \langle A\chi|\varphi\rangle$$

for all $\chi, \varphi \in \mathcal{D}(0,\infty)$. Thus, the operator A is formally self-adjoint. Now fix the non-real complex number z and study the equation $A - zI = f$, that is

$$-\mathrm{i}\varphi' - z\varphi = f, \qquad \varphi \in \mathcal{D}(0,\infty). \tag{7.320}$$

[200] If L is a linear subspace of X, then the orthogonal complement $L^\perp$ consists of all the elements φ of X which are orthogonal to L.
[201] Recall that $\varphi \in \mathcal{D}(0,\infty)$ iff the function $\varphi :]0,\infty[\to \mathbb{C}$ is smooth with compact support (i.e., it vanishes outside some interval $[a,b]$ with $0 < a < b < \infty$). Then the function φ satisfies the boundary condition $\varphi(0) = \varphi(+\infty) = 0$.
[202] The space $C^k[0,1]$, $k = 1, 2, \ldots$ consists of all continuous functions $\varphi : [0,1] \to \mathbb{C}$ which have continuous derivatives on the open interval $]0,1[$ up to order k, and all of these derivatives can be continuously extended to the closed interval $[0,1]$.

We are given $f \in L_2(0,\infty)$. If φ is a solution of (7.320), then

$$e^{-izx} f(x) = -i\frac{d}{dx}\left(\varphi(x)e^{-izx}\right).$$

Integration by parts tells us that

$$\int_0^\infty e^{-izx} f(x)dx = 0. \tag{7.321}$$

Choosing $z := -i$, we get $e^{-izx} = e^{-x}$. Then condition (7.321) is satisfied for all $f \in L_2(0,\infty)$. In contrast to this, if $z := i$, then $e^{-izx} = e^{x}$, and condition (7.321) is not valid for all $f \in L_2(0,\infty)$. Use this observation in order to show that the deficiency indices of A are given by $d_- = 0$ and $d_+ \neq 0$. By von Neumann's deficiency-index criterion (see Problem 7.18), the operator A has no self-adjoint extension.

For the complete proof of (i)–(iii), see P. Lax, Functional Analysis, Chap. 33, Wiley, New York, 2002.

7.20 *Continuity and boundedness.* Show that, for the linear operator $A : X \to X$ on the (real or complex) Hilbert space X, the following statements are equivalent:

(i) The operator A is continuous, that is, for any fixed element $\varphi_0 \in X$ and any number $\varepsilon > 0$, there exists a number $\delta(\varepsilon, \varphi_0) > 0$ such that

$$||\varphi - \varphi_0|| < \delta(\varepsilon, \varphi_0) \quad \text{implies} \quad ||A\varphi - A\varphi_0|| < \varepsilon.$$

(ii) The operator is sequentially continuous, that is, $\lim_{n\to\infty} \varphi_n = \varphi$ implies $\lim_{n\to\infty} A\varphi_n = A\varphi$.

(iii) The operator A is bounded, that is, $||A|| := \sup_{||\varphi||\leq 1} ||A\varphi|| < \infty$.

Hint: We refer to Zeidler (1995a), Sect. 1.9 (see the references on page 1049).

7.21 *Extension of a linear, densely defined, bounded operator.* Let $A : D(A) \to Y$ be a linear operator, where $D(A)$ is a linear dense subspace of the complex (resp. real) Hilbert space X, and Y is also a complex (resp. real) Hilbert space. Suppose that

$$||A\psi|| \leq \text{const } ||\psi|| \qquad \text{for all} \quad \psi \in D(A).$$

Show that the operator A can be uniquely extended to a linear bounded operator $A : X \to Y$. This statement remains true if X and Y are complex (resp. real) Banach spaces.

Hint: Let $\psi \in X$. Choose a sequence (ψ_n) in $D(A)$ with $\psi = \lim_{n\to\infty} \psi_n$. Using the Cauchy criterion, show that the sequence $(A\psi_n)$ is convergent. Set $A\psi := \lim_{n\to\infty} A\psi_n$. Finally, show that $A\psi$ is independent of the choice of the sequence (ψ_n). We refer to Zeidler (1995a), Sect. 3.6 (see the references on page 1049).

7.22 *Extension of isometric operators.* Let $A : D(A) \to X$ be a linear isometric operator on the linear subspace $D(A)$ of the complex Hilbert space X, that is, $\langle A\psi | A\varphi\rangle = \langle\psi|\varphi\rangle$ for all $\varphi, \psi \in D(A)$. Show that the operator A can be extended to a unitary operator $U : X \to X$ iff $\dim D(A)^\perp = \dim \text{im}(A)^\perp$.

Hint: (I) Assume first that $D(A)$ is a closed linear subspace of the separable Hilbert space X. Let $\dim D(A)^\perp = \dim \text{im}(A)^\perp$. Set

$$U\varphi_j := \psi_j \qquad \text{for all} \quad j,$$

where $\varphi_1, \varphi_2, \ldots$ (resp. $\psi_1, \psi_2, \ldots$) is an orthonormal basis in $D(A)^\perp$ (resp. $\text{im}(A)^\perp$).

(II) If $D(A)$ is not closed, then consider the closure $D_{\rm cl}$ of $D(A)$. This is a closed linear subspace of X. By Problem 7.21, the operator A can be uniquely extended to a linear isometric operator $B : D_{\rm cl} \to X$. Now apply argument (I) to the extension B.
(III) If the Hilbert space X is not separable, then replace $\varphi_1, \varphi_2, \dots$ (resp. $\psi_1, \psi_2, \dots$) by a generalized orthonormal basis, by using the Zorn lemma. As in Problem 7.19, see Lax (2002), Sect. 6.4.

7.23 *The Cayley transform.* The classical Möbius transformation

$$f(z) := \frac{z - \mathrm{i}}{z + \mathrm{i}}, \qquad z \in \mathbb{R}$$

generates a conformal map from the real line onto the unit circle. Generalizing this, we obtain the Cayley transformation

$$C_A := (A - \mathrm{i}I)(A + \mathrm{i}I)^{-1}$$

which was used for matrices A by Cayley.[203] In the late 1920s, von Neumann generalized this to operators in Hilbert spaces in order to solve the extension problem for self-adjoint operators (see Problem 7.18). Let $A : D(A) \to X$ be a linear, formally self-adjoint operator on the linear subspace $D(A)$ of the complex Hilbert space X. Show the following:
(i) $\mathrm{dom}(C_A) = \mathrm{im}(A + \mathrm{i}I)$ and $\mathrm{im}(C_A) = \mathrm{im}(A - \mathrm{i}I)$.
(ii) The operator C_A is isometric.
(iii) C_A is unitary on X iff A is self-adjoint.
(iv) C_A is closed iff A is closed.
(v) Let $B : D(B) \to X$ be linear and formally self-adjoint. Then, $A \subseteq B$ iff $C_A \subseteq C_B$.
(vi) If A is closed, then $\mathrm{dom}(C_A)$ and $\mathrm{im}(C_A)$ are closed linear subspaces of the Hilbert space X.
Hint: See F. Riesz and B. Nagy, Functional Analysis, Sect. 123, Frederyck Ungar, New York, 1978.

7.24 *Polar decomposition.* Each complex number z allows the polar decomposition $z = ur$ with $r := |z|$ and $u = \mathrm{e}^{\mathrm{i}\varphi}$. Here, $|u| = 1$. We want to generalize this to operators. Let $A : D(A) \to X$ be a linear (resp. antilinear), densely defined, closed operator on the complex Hilbert space X (e.g., a linear continuous operator $A : X \to X$.) Show the following:
(i) There exists a factorization

$$A = UR$$

where $R : D(R) \to X$ is a linear self-adjoint operator with $D(R) = D(A)$, and $\langle \psi | R\psi \rangle \geq 0$ for all $\psi \in D(R)$. In addition, $\ker(R) = \ker(A)$. Moreover, the operator $U : X \to X$ is linear (resp. antilinear) and the restriction

$$U : \ker(A)^{\perp} \to \mathrm{cl}(\mathrm{im}(A))$$

is unitary (resp. antiunitary), whereas $\ker(U) = \ker(A)$. Explicitly, $R = \sqrt{A^* A}$. The operator R is also called the absolute value of A (denoted by $|A|$). In particular, if the operator $A : X \to X$ is linear (resp. antilinear), continuous, and bijective, then the operator $U : X \to X$ is unitary (resp. antiunitary).
(ii) The operators R and U are uniquely determined by the properties formulated in (i).

[203] Möbius (1790–1868), Cayley (1821–1895).

(iii) If the linear operator $A : X \to X$ is continuous and normal, that is, $AA^* = A^*A$, then the operator $R : X \to X$ is linear, continuous, and self-adjoint, and the operator $U : X \to X$ is unitary. In addition, $UR = RU$.
Hint: See Reed and Simon, Methods of Modern Mathematical Physics, Vol. 1, Sect. VIII.9, Academic Press, as well as F. Riesz and B. Nagy, Functional Analysis, Sect. 110, Frederyck Ungar, New York, 1978.

7.25 *The theorem of Rolle on the zeros of functions.* Show the following for smooth functions $f : \mathbb{R} \to \mathbb{R}$:[204]

(i) If $f(a) = f(c) = 0$ with $a < c$, then there exists a number b with $a < b < c$ such that $f'(b) = 0$.
(ii) If $f(c) = 0$ and $\lim_{x\to+\infty} f(x) = 0$, then there exists a number $d > c$ such that $f'(d) = 0$.
(iii) Let $n \geq 1$. If the function f has at least n zeros on the compact interval J, then the derivative f' has at least $n-1$ zeros on J. If, in addition, the function f goes to zero as $x \to +\infty$ and $x \to -\infty$, then f' has at least $n+1$ zeros on $\mathbb{R}$.

Solution: Ad (i). By the classical mean theorem in calculus,

$$f(c) - f(a) = f'(b)(c-a) \qquad \text{for some } \ b \in]a, c[.$$

Ad (ii). Since $f(x) = \int_c^x f'(y)dy$, we get

$$\int_c^\infty f'(y)dy = \lim_{x\to+\infty} f(x) = 0.$$

Suppose that the function f' has no zeros on the interval $]c, \infty[$. Then, f' has constant sign on this interval, by the Bolzano theorem. Hence the integral of f' over $[c, \infty[$ does not vanish, a contradiction.
Ad (iii). For $n = 1$ the statement is trivial. Let $n \geq 2$. Suppose that $f(x_j) = 0$ for $j = 1, 2, ..., n$ with

$$x_1 < x_2 < ... < x_n.$$

By (ii), there exist numbers $y_1, y_2, ...$ with

$$x_1 < y_1 < x_2 < ... < x_{n-1} < y_{n-1} < x_n$$

such that $f'(y_k) = 0$ for $k = 1, ..., n-1$. In addition, if $f(x) \to 0$ as $x \to +\infty$, then there exists a number $y_n > x_n$ such that $f'(y_n) = 0$, by (ii). Similarly, it follows from $f(x) \to 0$ as $x \to -\infty$ that there exists a number $y_{-1} < x_1$ such that $f'(y_{-1}) = 0$.

7.26 *The zeros of the Hermite polynomials.* Show that, for $n = 0, 1, 2, ...$, the Hermite polynomial H_n of order n has precisely n zeros.[205]

Solution: Set $\mathcal{H}_n(x) := (-1)^n \mathrm{e}^{-x^2} H_n(x)$. By (7.7) on page 436,

$$\mathcal{H}_n(x) = \frac{d^n \mathrm{e}^{-x^2}}{dx^n}, \qquad n = 0, 1, 2, ...$$

Note that H_n is a polynomial of degree n. Thus, the maximal number of real zeros of H_n is equal to n. Moreover, $\mathcal{H}_n(x) \to 0$ as $x \to \pm\infty$ for $n = 0, 1, 2, , ...$
Using the recursive formula

[204] The French mathematician Michel Rolle (1652–1719) investigated the zeros of polynomials in his 1690 treatise *Traité d'algèbre*.
[205] This implies that the n zeros of H_n are simple.

$$\mathcal{H}_{n+1}(x) = \mathcal{H}_n'(x), \qquad n = 0, 1, 2, ...$$

and Problem 7.25, we proceed by induction. The function $H_0(x) = 1$ has no zeros. The polynomial H_1 of first order has precisely one zero. Now suppose that the polynomial H_n has n real zeros. Then, the function $\mathcal{H}_n$ has also n zeros. By Problem 7.25(iii), $\mathcal{H}_{n+1}$ has $n+1$ zeros. In turn, H_{n+1} has $n+1$ real zeros.

7.27 *The normal product* $: Q^n :$. Fix $x_0 := 1$ as on page 436. Let $m, n = 0, 1, 2, \ldots$ Define

$$P_n(x) := \sum_{k=0}^{[n/2]} (-1)^k c_{n,k} x^{n-2k}$$

where $c_{n,k} := n!/k!(n-2k)!2^k$. Here, $[n/2]$ denotes the largest integer j with $j \le n/2$. Using the normal product $: Q^n :$ introduced on page 438, prove the following:

(i) $H_n(x) = 2^{n/2} P_n(\sqrt{2}\, x)$.

(ii) $\int_{\mathbb{R}} H_n(x) H_m(x) \mathrm{e}^{-x^2} dx = 2^n n! \sqrt{\pi}\, \delta_{nm}$.

(iii) $x^n = \sum_{k=0}^{[n/2]} c_{n,k} P_{n-2k}(x)$.

(iv) $: Q^n := 2^{-n} H_n(x)$.

Hint: See J. Glimm and A. Jaffe, Mathematical Methods of Quantum Physics, Sect. 1.5, Springer, New York, 1981.

7.28 *The modified Moyal star product.* For all functions $f, g \in C^\infty(\mathbb{R}^2)$, define the modified Moyal product

$$f \star g := f \mathrm{e}^{\partial_a' \partial_{a^\dagger}} g = \sum_{m,n=0}^{\infty} \frac{1}{m!n!} \; (\partial_a^m f)(\partial_{a^\dagger}^n g).$$

Moreover, set $\pi_0 := \mathrm{e}^{-aa^\dagger}$ along with

$$\pi_n := \frac{1}{n!} \; (a^\dagger)^n \star \pi_0 \star a^n, \qquad n = 0, 1, 2, ...$$

Recall that $H := \hbar\omega a a^\dagger$ by page 593. Show that the following hold:

(i) $a^\dagger \star a = a a^\dagger, \;\; a \star a^\dagger = a a^\dagger + 1$.

(ii) $\pi_n = \pi_0 (a^\dagger)^n a^n / n!, \;\; n = 1, 2, \ldots$

(iii) $a \star \pi_0 = 0$.

(iv) $H \star \pi_n = n\hbar\omega \pi_n, \;\; n = 0, 1, 2, \ldots$

(v) The generalized Schrödinger equation

$$\mathrm{i}\hbar F_t(a, a^\dagger, t) = H \star F(a, a^\dagger, t), \qquad t \in \mathbb{R}, \;\; a \in \mathbb{C}$$

is equivalent to the equation $\mathrm{i}\hbar F_t(a, a^\dagger, t) = (H + \hbar\omega a^\dagger \partial_{a^\dagger}) F(a, a^\dagger, t)$. The solution is given by

$$F(a, a^\dagger, t) = \sum_{n=0}^{\infty} \pi_n(a, a^\dagger) \mathrm{e}^{-\mathrm{i}n\omega t}.$$

Hint: See A. Hirshfeld and P. Henselder, Deformation quantization in the teaching of quantum mechanics, Am. J. Phys. **70** (2002), 537–547.

7.29 *Proof of Theorem 7.54 on page 594.* Hint: Proceed similarly to Problem 7.28. See A. Hirshfeld and P. Henselder (2002), as above.

7.30 *Proof of Theorem 7.55 on page 594.* Hint: See A. Hirshfeld and P. Henselder (2002), as above.

7.31 *Weyl polynomials.* Prove Proposition 7.56 on page 598. Hint: Generalize the special argument given on page 598.

7.32 *The symbol of the scattering operator.* Motivate relation (7.276) on page 615, by using the Dirac delta function.
Solution: To simplify notation, we set $\hbar = m := 1$. Furthermore, choose

$$a(q,p) := \mathrm{e}^{\mathrm{i}tp^2/2}, \quad b(q,p) := \mathrm{sym}_P(q,p\,;t,t_0), \quad c(q,p) := \mathrm{e}^{-\mathrm{i}t_0p^2/2}.$$

Because of the associativity of the Moyal star product, we have to compute $(a*b)*c$.
(I) Computation of $a*b$. Set $f(q,p) := (a*b)(q,p)$. Choose the new notation $u := q_1, v := p_1, w := q_2$, and $z := p_2$. By definition of the Moyal star product (7.261) on page 607, we get

$$f(q,p) = \frac{1}{\pi^2}\int_{\mathbb{R}^4} \mathrm{e}^{2\mathrm{i}p(w-u)}\mathrm{e}^{2\mathrm{i}v(q-w)}\mathrm{e}^{2\mathrm{i}z(u-q)}\cdot\mathrm{e}^{\mathrm{i}tv^2/2}\,b(w,z)\,dudvdwdz.$$

Note that the substitution $x = 2u$ yields

$$\frac{1}{\pi}\int_{\mathbb{R}} \mathrm{e}^{2\mathrm{i}u(z-p)}du = \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}x(z-p)}dx = \delta(z-p).$$

Therefore, integration over the variable u yields

$$f(q,p) = \frac{1}{\pi}\int_{\mathbb{R}^3} \delta(z-p)\mathrm{e}^{2\mathrm{i}pw}\mathrm{e}^{2\mathrm{i}v(q-w)}\mathrm{e}^{-2\mathrm{i}zq}\cdot\mathrm{e}^{\mathrm{i}tv^2/2}\,b(w,z)\,dvdwdz.$$

Using $\int_{\mathbb{R}} F(z)\delta(z-p)dz = F(p)$, we get

$$f(q,p) = \frac{1}{\pi}\int_{\mathbb{R}^2} \mathrm{e}^{2\mathrm{i}pw}\mathrm{e}^{2\mathrm{i}v(q-w)}\mathrm{e}^{-2\mathrm{i}pq}\cdot\mathrm{e}^{\mathrm{i}tv^2/2}\,b(w,p)\,dvdw.$$

Changing the integration variables, $w \mapsto \xi, v \mapsto \eta$, we obtain

$$f(q,p) = \frac{1}{\pi}\int_{\mathbb{R}^2} \mathrm{e}^{2\mathrm{i}(p-\eta)(\xi-q)}\cdot\mathrm{e}^{\mathrm{i}t\eta^2/2}\,b(\xi,p)\,d\xi d\eta. \tag{7.322}$$

(II) Computation of $f*c$. Set $g := f*c$. Again by (7.261) on page 607,

$$\begin{aligned} g(q,p) &= \frac{1}{\pi^2}\int_{\mathbb{R}^4} \mathrm{e}^{2\mathrm{i}p(w-u)}\mathrm{e}^{2\mathrm{i}v(q-w)}\mathrm{e}^{2\mathrm{i}z(u-q)}\cdot f(u,v)\mathrm{e}^{-\mathrm{i}t_0z^2/2}\,dudvdwdz \\ &= \frac{1}{\pi}\int_{\mathbb{R}^3} \delta(p-v)\mathrm{e}^{-2\mathrm{i}pu}\mathrm{e}^{2\mathrm{i}vq}\mathrm{e}^{2\mathrm{i}z(u-q)}\cdot f(u,v)\mathrm{e}^{-\mathrm{i}t_0z^2/2}\,dudvdz, \end{aligned}$$

after integrating over w. Integration over v implies

$$g(q,p) = \frac{1}{\pi}\int_{\mathbb{R}^2} \mathrm{e}^{-2\mathrm{i}pu}\mathrm{e}^{2\mathrm{i}pq}\mathrm{e}^{2\mathrm{i}z(u-q)}\,f(u,p)\mathrm{e}^{-\mathrm{i}t_0z^2/2}\,dudz.$$

(III) Inserting $f(u,p)$ from (7.322), we obtain that $g(q,p)$ is equal to the integral

$$\frac{1}{\pi^2}\int_{\mathbb{R}^4} \mathrm{e}^{2\mathrm{i}(p-z)(q-u)}\mathrm{e}^{2\mathrm{i}(p-\eta)(\xi-u)}\mathrm{e}^{\mathrm{i}t\eta^2/2}\mathrm{e}^{-\mathrm{i}t_0 z^2/2}\, b(\xi,p)\, d\xi d\eta du dz.$$

After integrating over u, we get

$$\frac{1}{\pi}\int_{\mathbb{R}^3} \delta(z+\eta-2p)\, \mathrm{e}^{2\mathrm{i}(p-z)q}\, \mathrm{e}^{2\mathrm{i}(p-\eta)\xi}\, \mathrm{e}^{\mathrm{i}t\eta^2/2}\, \mathrm{e}^{-\mathrm{i}t_0 z^2/2}\, b(\xi,p)\, d\xi d\eta dz.$$

Consequently, integrating over η, we obtain

$$g(q,p) = \frac{1}{\pi}\int_{\mathbb{R}^2} \mathrm{e}^{2\mathrm{i}(p-z)(q-\xi)}\, \mathrm{e}^{\mathrm{i}t(z-2p)^2/2}\, \mathrm{e}^{-\mathrm{i}t_0 z^2/2}\, b(\xi,p)\, d\xi dz.$$

This is the claim (7.276) on page 615.

7.33 *The Wick theorem.* Compute the moment $\langle x_1^4 x_2^2\rangle$ by using the Wick theorem.
Solution: To simplify notation, we write (ij) instead of $\langle y_i y_j\rangle$. We first compute $\langle y_1 y_2 y_3 y_4 y_5 y_6\rangle$. This is equal to

$$\begin{aligned}
&(12)(34)(56) + (12)(35)(46) + (12)(36)(45)\\
&+ (13)(24)(56) + (13)(25)(46) + (13)(26)(45)\\
&+ (14)(23)(56) + (14)(25)(36) + (14)(26)(35)\\
&+ (15)(23)(46) + (15)(24)(36) + (15)(26)(34)\\
&+ (16)(23)(45) + (16)(24)(35) + (16)(25)(34).
\end{aligned}$$

Setting $y_1 = y_2 = y_3 = y_4 := x_1$ and $y_5 = y_6 := x_2$, we get

$$\langle x_1^4 x_2^2\rangle = 3\langle x_1^2\rangle^2\langle x_2^2\rangle + 12\langle x_1^2\rangle\langle x_1 x_2\rangle^2.$$

By induction, we obtain that $\langle x_1 x_2 \cdots x_{2n}\rangle$ contains $s(2n)$ summands where $s(0) := 1$ and

$$s(2n) = (2n-1)s(2n-2), \qquad n = 1, 2, 3, \ldots$$

For example, $s(2) = 1, s(4) = 3, s(6) = 15, s(8) = 7 \cdot 15 = 105$.

7.34 *The rescaling trick.* Prove Prop. 7.48 on page 572.
Solution: Let $s \geq s_0$. By assumption, there exists a number $s_0 > 1$ such that the series $\zeta_A(s) = \sum_{n=1}^{\infty} \lambda_n^{-s}$ converges. Using Euler's gamma function

$$\Gamma(s) = \int_0^\infty t^{s-1}\mathrm{e}^{-t} dt,$$

we get

$$\Gamma(s)\zeta_A(s) = \int_0^\infty t^{s-1}\mathrm{e}^t \sum_{n=1}^{\infty} \lambda_n^{-s} dt.$$

Here, it is allowed to interchange summation with integration, by the majorant criterion for integrals (see page 493 of Vol. I). The substitution $t = \lambda_n u$ yields

$$\zeta_A(s) = \frac{1}{\Gamma(s)}\int_0^\infty u^{s-1} \sum_{n=1}^{\infty} \mathrm{e}^{-\lambda_n u} du.$$

Let $\gamma > 0$. Replacing $A \mapsto \gamma A$ and $\lambda_n \mapsto \gamma\lambda_n$, we obtain

$$\zeta_{\gamma A}(s) = \frac{1}{\Gamma(s)}\int_0^\infty u^{s-1} \sum_{n=1}^{\infty} \mathrm{e}^{-\gamma\lambda_n u} du.$$

The substitution $v = \gamma u$ yields

$$\zeta_{\gamma A}(s) = \frac{\gamma^{-s}}{\Gamma(s)} \int_0^\infty v^{s-1} \sum_{n=1}^\infty \mathrm{e}^{-\lambda_n v} dv = \gamma^{-s} \zeta_A(s).$$

Differentiating this with respect to s, we obtain

$$\zeta_{\gamma A}(s) = -\ln \gamma \cdot \gamma^{-s} \zeta_A(s) + \gamma^{-s} \zeta_A'(s).$$

After analytic continuation of the zeta function ζ_A, we get

$$\zeta_{\gamma A}'(0) = -\zeta_A(0) \ln \gamma + \zeta_A'(0).$$

This implies the desired result

$$\det(\gamma A) = \mathrm{e}^{-\zeta_{\gamma A}'(0)} = \gamma^{\zeta_A(0)} \mathrm{e}^{-\zeta_A'(0)} = \gamma^{\zeta_A(0)} \det A.$$

7.35 *Special Fourier–Laplace integrals.* Let $E, H \in \mathbb{R}$, and $\varepsilon > 0$. Prove the following:

(i) $\int_{-\infty}^\infty \mathrm{e}^{\mathrm{i}(E+\mathrm{i}\varepsilon)t/\hbar} \mathrm{e}^{-\mathrm{i}Ht/\hbar} \theta(t) dt = \frac{\mathrm{i}\hbar}{E+\mathrm{i}\varepsilon - H}$.

(ii) $\theta(t) \mathrm{e}^{-\mathrm{i}Ht/\hbar} = \frac{\mathrm{i}}{2\pi} PV \int_{-\infty}^\infty \frac{\mathrm{e}^{-\mathrm{i}(E+\mathrm{i}\varepsilon)t/\hbar}}{E+\mathrm{i}\varepsilon - H}\, dE$ for all $t \in \mathbb{R} \setminus \{0\}$.

Solution: To simplify notation, set $\hbar := 1$. Since $\lim_{t\to+\infty} \mathrm{e}^{-\varepsilon t} = 0$,

$$\int_0^\infty \mathrm{e}^{\mathrm{i}Et} \mathrm{e}^{-\varepsilon t} \mathrm{e}^{-\mathrm{i}Ht} dt = \lim_{N\to\infty} \frac{\mathrm{e}^{\mathrm{i}Et}\, \mathrm{e}^{-\varepsilon t} \mathrm{e}^{-\mathrm{i}Ht}}{\mathrm{i}(E+\mathrm{i}\varepsilon) - \mathrm{i}H} \Big|_0^N = \frac{\mathrm{i}}{E+\mathrm{i}\varepsilon - H}.$$

In order to get the inverse transformation, we formally apply the Fourier transform to (i). This yields

$$\theta(t) \mathrm{e}^{-\varepsilon t}\, \mathrm{e}^{-\mathrm{i}Ht} = \frac{1}{2\pi} \int_{-\infty}^\infty \mathrm{e}^{-\mathrm{i}Et} \cdot \frac{\mathrm{i}}{E+\mathrm{i}\varepsilon - H}\, dE, \qquad t \in \mathbb{R}. \tag{7.323}$$

However, the crux is that this integral does not exist because of too slow decay at infinity. Therefore, we have to argue more carefully. Observe first that the function

$$f(t) := \theta(t) \mathrm{e}^{-\varepsilon t} \mathrm{e}^{-\mathrm{i}Ht}, \qquad t \in \mathbb{R}$$

is not smooth. This is the reason for the failing of the Fourier transform, in the classical sense. However, since $|f|$ is bounded, the function f is a tempered distribution, and its Fourier transform is well defined. Thus, we may regard equation (7.323) as a short-hand notation for the Fourier transform in the sense of tempered distributions. To refine this argument, note that $\int_{\mathbb{R}} |f(t)|^2 dt < \infty$, that is, $f \in L_2(\mathbb{R})$. The Plancherel theorem tells us that the Fourier transform

$$f(t) = \lim_{R\to+\infty} \int_{-\mathbb{R}}^R \mathrm{e}^{-\mathrm{i}Et} \cdot \frac{\mathrm{i}}{E+\mathrm{i}\varepsilon - H}\, dE, \qquad t \in \mathbb{R}$$

is valid in the sense of the convergence in the Hilbert space $L_2(\mathbb{R})$ (see page 514). More precisely, applying the residue theorem, Cauchy's integration method implies that (ii) is valid for all $t \neq 0$. Argue as in Problem 12.1 of Vol. I.

7.36 *The Fourier–Laplace transform.* Prove Prop. 7.17 on page 498.
Hint: Use Problem 7.35. For interchanging limits, construct absolutely convergent majorant series. To this end, observe that the inequality $2ab \le a^2 + b^2$ (for real numbers a, b) yields

$$2|\langle\chi|\varphi_k\rangle\langle\varphi_k|\varphi\rangle| \le |\langle\chi|\varphi_k\rangle|^2 + |\langle\chi|\varphi_k\rangle|^2.$$

Finally, use the Parseval equation.

7.37 *Proof of Proposition 7.64 on page 642.* Solution: It is convenient to use the function

$$J := -\sum_{k=1}^{M} p_m \ln p_m$$

which differs from I by a positive factor. (Note that $\log_2 a = \ln a \cdot \log_2 \mathrm{e}$.) Since $\lim_{x\to+0} x \ln x = 0$, the function J is continuous on the closed simplex σ_M. For the partial derivatives of J on the interior of σ_M, we get $J_{p_m} = -\ln p_m - 1$ and

$$J_{p_m p_n} = -\frac{\delta_{mn}}{p_m}, \qquad m, n = 1, \ldots, M.$$

Thus, the symmetric matrix $(-J_{p_m p_n})$ is positive definite on the interior of σ_M, and hence the function $-J$ is convex, that is, J is concave on the interior of σ_M. By continuity, this remains true on σ_M. One checks easily that the maximal value of J is attained at an inner point of σ_M. From $J_{p_m} = 0$ for $m = 1, \ldots, M$, we get $p_1 = \ldots = p_M$.

8. Quantum Particles on the Real Line – Ariadne's Thread in Scattering Theory

> The S-Matrix knows all about scattering processes and bound states of quantum particles.
>
> Folklore

The S-Matrix is the most important quantity in elementary particle physics. The fundamental role played by the S-matrix was emphasized by Heisenberg in 1943.[1] He was motivated by the following philosophy:

> *Use only such quantities in quantum physics which are closely related to physical experiments.*

In fact, the S-matrix encodes transition probabilities in scattering experiments. This way, one obtains the cross sections for scattering processes which can be measured in particle accelerators. Interestingly enough, scattering experiments also give information about bound states by using the poles of the S-matrix. It is also possible to detect unstable particles (also called resonances in particle physics) by carrying out scattering experiments. In the present chapter, we will explain the basic ideas of scattering theory by studying very simple models.

8.1 Classical Dynamics Versus Quantum Dynamics

In our solar system, there exist two different types of motions:

(a) Planets (and some comets) move on bounded orbits (Fig. 8.1(a)).
(b) In addition, there exist comets which move on unbounded orbits (Fig. 8.1(b)).

Concerning (b), the motion is asymptotically free in the remote past (i.e., $t \to -\infty$) and in the distant future (i.e., $t \to +\infty$). We speak of bound states (a) and scattering states (b). It turns out that the distinction between bound states and asymptotically free states (also called scattering states) is of fundamental importance in elementary particle physics. For example, consider the electron of a hydrogen atom which moves around the proton.

- The planets correspond to bound states of the electron. In contrast to celestial mechanics, the energies $E_1, E_2, \ldots$ of bound states are discrete. More precisely, $-\infty < E_1 < E_2 < \ldots < 0$.
- The asymptotically free moving comets correspond to the scattering of electrons by the proton. The possible energies of the scattered electrons fill the infinite interval $[0, \infty[$.

[1] W. Heisenberg, The observable quantities in particle physics, Z. Phys. **120** (1943), 513–538, 673–702; **123** (1944), 93–112 (in German).

(a) bound state (b) scattering state

Fig. 8.1. Classification of states

The distinction between bound states and scattering states is closely related to the spectral theory of self-adjoint operators in Hilbert spaces. Roughly speaking, we have the following situation:

- The possible energy values of a quantum particle form the spectrum of the Hamiltonian $H : D(H) \to X$ in the Hilbert space X. Here, H is a self-adjoint operator in X.
- The energy values of the bound states form the discrete spectrum of H.
- The energy values of the scattering states form the absolutely continuous spectrum of H.

The Hilbert space X allows the following orthogonal decomposition

$$X = X_1 \oplus X_2.$$

The closed subspace X_1 of X has an orthonormal basis $\varphi_1, \varphi_2, \ldots$ with

$$\boxed{H\varphi_j = E_j\varphi_j \qquad \text{for all} \quad j.}$$

The elements of X_1 (resp. X_2) are called bound states (resp. scattering states). In particular, each eigenvector φ_j represents a bound state with energy E_j. For given initial state $\psi(0) \in X$, the dynamics of the particle is given by

$$\boxed{\psi(t) = \mathrm{e}^{-\mathrm{i}tH/\hbar}\psi(0) \qquad \text{for all} \quad t \in \mathbb{R}.}$$

For the dynamics, we have the following situation:

- If $\psi(0) \in X_1$, then $\psi = \psi(t)$ represents a bound motion of the quantum particle.
- If $\psi(0) \in X_2$, then the motion $\psi = \psi(t)$ represents a scattering process for the quantum particle. Intuitively, this is the motion of a wave packet which is asymptotically free as time goes to $+\infty$ and $-\infty$.

In particle accelerators, the incoming particles form a homogeneous stream of fixed particle energy E. Such a particle stream is described by a current density vector. Note the following:

- Current densities do not live in the Hilbert space X, but they are costates of X (i.e., they are functionals on X).
- In terms of the Schrödinger equation, bound states φ_j of energy E_j are normalized functions, that is,

$$\int_{\mathbb{R}^3} |\varphi_j(x)|^2 d^3x = 1.$$

- The current densities φ are described by eigenfunctions of the Schrödinger equation which cannot be normalized, that is, $H\varphi = E\varphi$, and

$$\int_{\mathbb{R}^3} |\varphi(x)|^2 d^3x = \infty.$$

Such eigenfunctions are called generalized eigenfunctions.

Motion of a classical particle. Let us study a simple example. Consider the motion $x = x(t)$ of a classical particle of mass m on the real line governed by the following equation of motion

$$\boxed{m\ddot{x}(t) = -U'(x(t)).} \tag{8.1}$$

Assume first that the potential $U : \mathbb{R} \to \mathbb{R}$ is smooth. Then, for each solution of (8.1), there exists a real number E such that

$$\frac{m\dot{x}(t)^2}{2} + U(x(t)) = E \quad \text{for all times } \ t \in \mathbb{R}. \tag{8.2}$$

The number E represents the energy of the motion. Hence

$$U(x(t)) \leq E \qquad \text{for all} \quad t \in \mathbb{R}.$$

This restricts the possible motions of the particle. In fact, for given energy E, the motion of the particle is only possible in the subset

$$\{x \in \mathbb{R} : \ U(x) \leq E\}$$

of the real line. Now we consider the special case where the potential is given by the following function

$$U(x) := \begin{cases} U_0 & \text{if} \quad -r \leq x \leq r, \\ 0 & \text{otherwise.} \end{cases} \tag{8.3}$$

Here, $r > 0$. In this case, the potential U is piecewise continuous. By a solution $x = x(t)$ of the equation of motion (8.1), we understand a solution in the sense of distributions.[2] In addition, we assume that $x = x(t)$ is continuous and piecewise continuously differentiable on the time interval $]-\infty, \infty[$, and it satisfies equation (8.2) for fixed real energy E and all points in time where the velocity function $t \to \dot{x}(t)$ is continuous. It turns out that the velocity v of the particle is piecewise constant. If $v > 0$ (resp. $v < 0$), then the particle moves from left to right (resp. from right to left). We have to distinguish the following cases.

(i) Free motion: $U_0 = 0$. For given energy $E > 0$, the particle moves on the real line with the velocity v given by $\frac{mv^2}{2} = E$. For $E = 0$, the particle rests.

(ii) Potential well: $U_0 < 0$ (Fig. 8.2).
- For given energy $E > 0$, the particle moves with piecewise constant velocity on the real line.

[2] Explicitly, this means that $\int_{\mathbb{R}} mx(t)\ddot{\varphi}(t)dt = -\int_{\mathbb{R}} U'(x(t))\varphi(t)dt$ for all smooth test functions $\varphi : \mathbb{R} \to \mathbb{R}$ with compact support.

(a) Asymptotically free motion (b) Bound motion

Fig. 8.2. Motion of a classical particle of energy E on the real line

- For given energy $E \in]U_0, 0[$, the particle moves inside the interval $[-r, r]$ with the constant velocity v given by

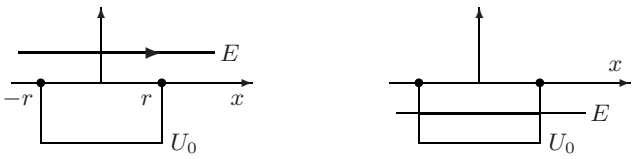
$$\frac{mv^2}{2} + U_0 = E. \tag{8.4}$$

If the particle hits one of the boundary points $x = \pm r$, then it will be reflected and it changes the direction of motion.
- For $E = U_0$, the particle rests on the interval $[-r, r]$.

(iii) Potential barrier: $U_0 > 0$ (Fig. 8.8 on page 728).
- For given energy $E > U_0$, the particle moves with piecewise constant velocity on the real line.
- For given energy $E \in]0, U_0[$, the particle moves either on the interval $]-\infty, -r]$ or on $[r, \infty[$. If the particle hits one of the boundary points $x = -r$ or $x = r$, then it is reflected.

Motion of a quantum particle. The quantum motion with respect to the potential U from (8.3) will be explicitly computed in the next section. Roughly speaking, there appear the following quantum modifications which are crucial from the physical point of view.

- In the case of the potential well, there exists only a finite number of bound-state energies. The corresponding wave functions are not concentrated on the interval $[-r, r]$. However, they are decaying exponentially as $x \to -\infty$ and $x \to +\infty$ (Fig. 8.3).
- In the case of the potential barrier, the quantum particle is able to pass through the interval $[-r, r]$. This means that quantum particles are able to reach regions which are forbidden by classical mechanics. This quantum effect is called tunnelling. For example, the radioactive decay of molecules is based on tunnelling (Fig. 8.9 on page 728).



(a) Scattering state (b) Bound state

Fig. 8.3. Motion of a quantum particle of energy E on the real line

8.2 The Stationary Schrödinger Equation

The motion of a quantum particle of mass m on the real line is governed by the Schrödinger equation

$$\mathrm{i}\hbar\psi_t(x,t) = -\frac{\hbar^2}{2m}\psi_{xx}(x,t) + U(x)\psi(x,t), \qquad x,t \in \mathbb{R}. \tag{8.5}$$

Choosing the real number E and setting

$$\psi(x,t) = \mathrm{e}^{-\mathrm{i}Et/\hbar}\varphi(x), \tag{8.6}$$

we get the stationary Schrödinger equation

$$\boxed{-\frac{\hbar^2}{2m}\varphi_{xx}(x,t) + U(x)\varphi(x) = E\varphi(x), \qquad x \in \mathbb{R}.} \tag{8.7}$$

If φ is a smooth solution of (8.7), then the function ψ from (8.6) satisfies the equation (8.5) and describes a quantum state of energy E. Furthermore, we have the conservation law

$$\varrho_t(x,t) + J_x(x,t) = 0 \qquad \text{for all} \quad x,t \in \mathbb{R} \tag{8.8}$$

with the particle number density $\varrho := |\psi|^2$ and the particle current density[3]

$$J := \Re\left(\psi^\dagger \frac{P}{m}\,\psi\right) \tag{8.9}$$

along with the momentum operator $P\psi := -\mathrm{i}\hbar\frac{\partial}{\partial x}\psi$. Hence

$$\varrho(x) = |\varphi(x)|^2, \qquad J(x) = \Re\left(\varphi(x)^\dagger \frac{P}{m}\,\varphi(x)\right).$$

By (8.8), $J'(x) = 0$ for all $x \in \mathbb{R}$. This implies the conservation of the particle current density:

$$\boxed{J(x) = \text{const} \qquad \text{for all} \quad x \in \mathbb{R}.} \tag{8.10}$$

Let us formulate a second important conservation law. If φ_1 and φ_2 are smooth solutions of (8.7), then a simple computation shows that the derivative of the Wronskian

$$\mathcal{W}(x) := \begin{vmatrix} \varphi_1(x) & \varphi_2(x) \\ \varphi_1'(x) & \varphi_2'(x) \end{vmatrix}$$

vanishes. Hence

$$\boxed{\mathcal{W}(x) = \text{const} \qquad \text{for all} \quad x \in \mathbb{R}.} \tag{8.11}$$

Classification of solutions of the stationary Schrödinger equation. Let φ be a solution of (8.7).

[3] The current density J has the dimension "(particle number/length) × velocity". In the SI system, this corresponds to s^{-1}.

- If $\int_{\mathbb{R}} |\varphi(x)|^2 dx = 1$, then the function φ describes a single quantum particle on the real line, and the integral

$$\int_a^b |\varphi(x)|^2 dx$$

  is equal to the probability of finding the particle in the interval $[a, b]$. This probability does not depend on time.
- If $\int_{\mathbb{R}} |\varphi(x)|^2 dx = \infty$, then the function φ describes a homogeneous particle stream on the real line with the constant particle current density

$$J = \Re \left(\varphi^\dagger \frac{P}{m} \, \varphi \right).$$

  If $J > 0$ (resp. $J < 0$), then the particles move from left to right (resp. from right to left). The number of particles that passes through a fixed, but otherwise arbitrary point x during the time interval $[t_0, t_1]$ is equal to $|J| \cdot (t_1 - t_0)$.

8.3 One-Dimensional Quantum Motion in a Square-Well Potential

We are going to show that the motion of a quantum particle on the real line, under the action of a square-well potential, can be explicitly computed (Fig. 8.4). Our investigations serve as a prototype for typical phenomena arising in elementary particle physics. This concerns the following:

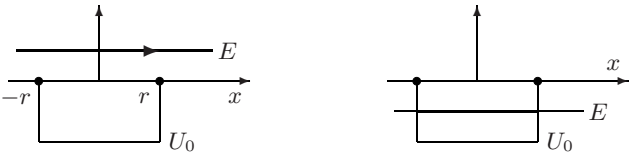
- the appearance of both bound states and unbound states (scattering states),
- scattering states are described by eigenfunctions of the energy operator which are not normalizable (i.e., generalized eigenfunctions in the Hilbert space setting),
- the physical and mathematical importance of the S-matrix,
- conservation of probability and the unitarity of the S-matrix,
- the relation between bound-state energies and singularities of the S-matrix,
- the Fourier–Stieltjes transform with respect to (normalized and non-normalizable) eigenfunctions of the energy operator,
- wave packets and wave operators,
- the dynamical meaning of the S-matrix and the relation to wave operators,
- the duality between partial differential equations (the Schrödinger equation) and integral equations (the Lippmann–Schwinger equation),
- the method of the Green's function in scattering theory (the Lippmann–Schwinger equation),
- the energetic Riemann surface and the existence of unstable particles (resonances and the Breit–Wigner formula).

To begin with, let us switch off the potential. That is, we want to study the free motion.

8.3.1 Free Motion

For vanishing potential, $U \equiv 0$, the stationary Schrödinger equation reads as

$$\boxed{-\frac{\hbar^2}{2m} \, \varphi''(x) = E\varphi(x), \qquad x \in \mathbb{R}.} \tag{8.12}$$



(a) Scattering state (b) Bound state

Fig. 8.4. Square-well potential in quantum mechanics

For fixed parameter $k > 0$, introduce the so-called free Jost functions

$$\varphi_{k,\mathrm{free}}(x) := \mathrm{e}^{\mathrm{i}kx}, \qquad \varphi_{-k,\mathrm{free}}(x) := \mathrm{e}^{-\mathrm{i}kx}, \qquad x \in \mathbb{R}.$$

Then

$$P\varphi_{\pm k,\mathrm{free}} = \pm\hbar k \varphi_{\pm k,\mathrm{free}}$$

and

$$J_\pm = \Re\left(\varphi^\dagger_{\pm k,\mathrm{free}} \frac{P}{m}\, \varphi_{\pm k,\mathrm{free}}\right) = \pm\frac{\hbar k}{m}, \qquad E = \frac{\hbar^2 k^2}{2m}.$$

Consequently, for the given parameter $k > 0$, the free Jost function $\varphi_{k,\mathrm{free}}$ describes a stream of particles where each particle has the momentum $\hbar k$ and the energy $E = \frac{\hbar^2 k^2}{2m}$. The particle current density is given by $J_+ = \frac{\hbar k}{m}$. Since $J_+ > 0$, the particles move from left to right.

Similarly, the free Jost function $\varphi_{-k,\mathrm{free}}$ describes a stream of particles with particle momentum $-\hbar k$, particle energy $E = \frac{\hbar^2 k^2}{2m}$, and particle current density $J_- = -\frac{\hbar k}{m}$. Since $J_- < 0$, the particles move from right to left.

General solution. For given $E > 0$, chose $k := \sqrt{\frac{2mE}{\hbar^2}}$. Then the general solution of the stationary Schrödinger equation (8.12) reads as

$$\varphi(x) = \alpha\varphi_{k,\mathrm{free}}(x) + \beta\varphi_{-k,\mathrm{free}}(x)$$

where α and β are arbitrary complex numbers. This shows the importance of the free Jost functions.

Convention. In order to simplify the notation, we will set $\hbar := 1$ and $m := 1$ in the remaining part of this chapter.

8.3.2 Scattering States and the *S*-Matrix

Square-well potential. As a typical case, let us now choose the square-well potential $U : \mathbb{R} \to \mathbb{R}$ from (8.3) with $U_0 < 0$ (Fig. 8.4 on page 705). We want to solve explicitly the stationary Schrödinger equation (8.7) with respect to U.

The scattering function φ_k. Fix the parameter $k > 0$. By definition, the scattering function φ_k has the form

$$\varphi_k(x) := \begin{cases} \mathrm{e}^{\mathrm{i}kx} + \varrho_{\leftarrow}(k)\mathrm{e}^{-\mathrm{i}kx} & \text{if} \quad x < -r, \\ A\mathrm{e}^{\mathrm{i}Kx} + B\mathrm{e}^{-\mathrm{i}Kx} & \text{if} \quad -r \le x \le r, \\ \tau_{\rightarrow}(k)\mathrm{e}^{\mathrm{i}kx} & \text{if} \quad x > r. \end{cases} \tag{8.13}$$

We have to match the complex coefficients $\varrho_{\leftarrow}, A, B, \tau_{\rightarrow}$ such that

$$\varphi_k \in C^1(\mathbb{R}),$$

that is, we get a continuous and continuously differentiable function φ_k on the real line. We will show below that the coefficients can be uniquely determined in such a way that we obtain a continuously differentiable solution (in the sense of distributions) of the stationary Schrödinger equation (8.7) with energy E. From the physical point of view, the scattering function φ_k describes the following situation:

- The function $x \mapsto \mathrm{e}^{\mathrm{i}kx}$ on $]-\infty, -r[$ represents an incoming stream of particles of particle momentum $p = k$, particle energy $E = k^2/2$, and particle current density
$$J_{\text{in}} = k.$$
The particles move from left to right.
- Reflection. The function $x \mapsto \varrho_{\leftarrow}(k)\mathrm{e}^{-\mathrm{i}kx}$ on $]-\infty, -r[$ represents a particle stream which is reflected at the point $x = -r$. The reflected particles have the momentum $p = -k$ and the energy $E = k^2/2$. The current density is given by
$$\boxed{J_{\text{refl}} = -|\varrho_{\leftarrow}(k)|^2 J_{\text{in}}.}$$
The reflected particles move from $x = -r$ to $x = -\infty$.
- Transmission. The function $x \mapsto \tau_{\rightarrow}(k)\mathrm{e}^{\mathrm{i}kx}$ on $]r, \infty[$ represents the transmitted particle stream. These particles have the momentum $p = k$ and the corresponding energy $E = \frac{k^2}{2}$. The particle current density is given by
$$\boxed{J_{\text{trans}} = |\tau_{\rightarrow}(k)|^2 J_{\text{in}}.}$$
The transmitted particles move from $x = -r$ to $x = +\infty$.

Naturally enough, the complex number $\varrho_{\leftarrow}(k)$ is called the reflection amplitude, and $\tau_{\rightarrow}(k)$ is called the transmission amplitude.

Using $\Re(z - z^\dagger) = 0$, a simple computation shows that the scattering function φ_k corresponds to the particle current density $J_{\text{in}} + J_{\text{refl}}$ on $]-\infty, -r[$ and the particle current density J_{trans} on $]r, \infty[$. Conservation of particle current density tells us that we have $J_{\text{in}} + J_{\text{refl}} = J_{\text{trans}}$. Hence

$$\boxed{|\varrho_{\leftarrow}(k)|^2 + |\tau_{\rightarrow}(k)|^2 = 1 \qquad \text{for all} \quad k > 0.} \tag{8.14}$$

For an incoming particle of momentum k, note that

- $|\varrho_{\leftarrow}(k)|^2$ is the probability for reflection at the point $x = -r$ where the potential well starts, and
- $|\tau_{\rightarrow}(k)|^2$ is the probability for transmission at the point $x = -r$.

Naturally enough, the sum of the two probabilities is equal to 1 by (8.14). We will show below that the crucial relation (8.14) between reflection probability and transmission probability is closely connected with the unitarity of the S-matrix.

The scattering function φ_{-k}. Let us now reverse the direction of scattering. This scattering process from right to left is described by the scattering function φ_{-k} of the form

$$\varphi_{-k}(x) := \begin{cases} \mathrm{e}^{-\mathrm{i}kx} + \varrho_{\rightarrow}(k)\mathrm{e}^{\mathrm{i}kx} & \text{if} \quad x > r, \\ A\mathrm{e}^{\mathrm{i}Kx} + B\mathrm{e}^{-\mathrm{i}Kx} & \text{if} \quad -r \le x \le r, \\ \tau_{\leftarrow}(k)\mathrm{e}^{-\mathrm{i}kx} & \text{if} \quad x < -r. \end{cases} \tag{8.15}$$

Because of the symmetry of the potential well U, we have $\varphi_{-k}(x) = \varphi_k(-x)$ for all points $x \in \mathbb{R}$ and all wave numbers $k > 0$.

Proposition 8.1 *Let $k > 0$. For the reflection and transmission amplitude of the scattering function φ_k with respect to the potential well U from Fig. 8.4 on page 705, we get*

$$\varrho_{\leftarrow}(k) = \frac{\frac{\mathrm{i}}{2}\left(\frac{K}{k} - \frac{k}{K}\right) e^{-2\mathrm{i}kr} \sin^2 2Kr}{\cos 2Kr - \frac{\mathrm{i}}{2}\left(\frac{K}{k} + \frac{k}{K}\right) \sin 2Kr},$$
$$\tau_{\rightarrow}(k) = \frac{e^{-2\mathrm{i}kr}}{\cos 2Kr - \frac{\mathrm{i}}{2}\left(\frac{K}{k} + \frac{k}{K}\right) \sin 2Kr}.$$

Here, $E = \frac{k^2}{2}$ and $K = \sqrt{2(E + |U_0|)}$. For the transmission probability, we have

$$|\tau_{\rightarrow}(k)|^2 = \left(1 + \frac{\sin^2 2Kr}{\frac{4E}{|U_0|}\left(1 + \frac{E}{|U_0|}\right)}\right)^{-1}.$$

This yields the reflection probability $|\varrho_{\leftarrow}(k)|^2 = 1 - |\tau_{\rightarrow}(k)|^2$.

For the scattering function φ_{-k}, we obtain

$$\varrho_{\rightarrow}(k) = \varrho_{\leftarrow}(k), \qquad \tau_{\leftarrow}(k) = \tau_{\rightarrow}(k). \tag{8.16}$$

This follows from the symmetry of the potential well U.

Proof. Consider the scattering function φ_k. We have to determine the coefficients $\varrho_{\leftarrow}, A, B, \tau_{\rightarrow}$ along with K. To this end, let us insert the ansatz for φ_k into the stationary Schrödinger equation (8.7).

Considering the intervals $]-\infty, -r[$ and $]r, \infty[$, we get $E = \frac{k^2}{2}$. On the interval $[-r, r]$, we obtain

$$E = \frac{K^2}{2} + U_0.$$

Hence $K = \sqrt{2(E - U_0)}$. The function φ_k is continuous at the points $x = -r$ and $x = r$ iff

$$\begin{aligned} \mathrm{e}^{-\mathrm{i}kr} + \varrho_{\leftarrow}\mathrm{e}^{\mathrm{i}kr} &= A\mathrm{e}^{-\mathrm{i}Kr} + B\mathrm{e}^{\mathrm{i}Kr}, \\ A\mathrm{e}^{\mathrm{i}Kr} + B\mathrm{e}^{-\mathrm{i}Kr} &= \tau_{\rightarrow}\mathrm{e}^{\mathrm{i}kr}. \end{aligned}$$

The derivative of φ_k is continuous at the points $x = -r$ and $x = r$ iff

$$\begin{aligned} k(\mathrm{e}^{-\mathrm{i}kr} - \varrho_{\leftarrow}\mathrm{e}^{\mathrm{i}kr}) &= K(A\mathrm{e}^{-\mathrm{i}Kr} - B\mathrm{e}^{\mathrm{i}Kr}), \\ K(A\mathrm{e}^{\mathrm{i}Kr} - B\mathrm{e}^{-\mathrm{i}Kr}) &= \tau_{\rightarrow}k\mathrm{e}^{\mathrm{i}kr}. \end{aligned}$$

To simplify notation, we set $\mathcal{R} := \varrho_{\leftarrow}\mathrm{e}^{2\mathrm{i}kr}$, $\mathcal{D} := \tau_{\rightarrow}$, $\mathcal{A} := A\mathrm{e}^{\mathrm{i}(k-K)r}$, as well as $\mathcal{B} := B\mathrm{e}^{\mathrm{i}(k+K)r}$. Then

$$\begin{aligned} 1 + \mathcal{R} &= \mathcal{A} + \mathcal{B}, & k(1 - \mathcal{R}) &= K(\mathcal{A} - \mathcal{B}), \\ \mathcal{D}\mathrm{e}^{2\mathrm{i}kr} &= \mathcal{A}\mathrm{e}^{2\mathrm{i}Kr} + \mathcal{B}\mathrm{e}^{-2\mathrm{i}Kr}, & k\mathcal{D}\mathrm{e}^{2\mathrm{i}kr} &= K(\mathcal{A}\mathrm{e}^{2\mathrm{i}Kr} - \mathcal{B}\mathrm{e}^{-2\mathrm{i}Kr}). \end{aligned}$$

Hence

$$1+\mathcal{R}+\frac{k}{K}(1-\mathcal{R})=2\mathcal{A}, \qquad \mathcal{D}\left(1+\frac{k}{K}\right)\mathrm{e}^{2\mathrm{i}(k-K)r}=2\mathcal{A},$$
$$1+\mathcal{R}-\frac{k}{K}(1-\mathcal{R})=2\mathcal{B}, \qquad \mathcal{D}\left(1-\frac{k}{K}\right)\mathrm{e}^{2\mathrm{i}(k+K)r}=2\mathcal{B}.$$

Eliminating $\mathcal{A}$ and $\mathcal{B}$, we get

$$1+\mathcal{R}+\frac{k}{K}(1-\mathcal{R})=\mathcal{D}\left(1+\frac{k}{K}\right)\mathrm{e}^{2\mathrm{i}(k-K)r},$$
$$1+\mathcal{R}-\frac{k}{K}(1-\mathcal{R})=\mathcal{D}\left(1-\frac{k}{K}\right)\mathrm{e}^{2\mathrm{i}(k-K)r}.$$

Computing the solution $\mathcal{D},\mathcal{R}$ of this system, we obtain the claim. □

The scattering matrix (S-matrix). Fix the parameter $k>0$ called wave number. The scattering process is characterized by the four transition amplitudes

- $\tau_{\leftarrow}(k), \varrho_{\rightarrow}(k)$ (scattering from right to left),
- $\varrho_{\leftarrow}(k), \tau_{\rightarrow}(k)$ (scattering from left to right).

Define the matrix

$$\hat{S}(k) := \begin{pmatrix} \tau_{\leftarrow}(k) & \varrho_{\rightarrow}(k) \\ \varrho_{\leftarrow}(k) & \tau_{\rightarrow}(k) \end{pmatrix}. \tag{8.17}$$

This is called the S-matrix with respect to the wave number k. We will show below that the matrix $\hat{S}(k)$ is unitary because of the conservation of the particle number. The crucial relation between the S-matrix and the quantum dynamics will be discussed in Sec. 8.3.9 on page 722.

Unitarity of the S-matrix. Let us prove that

For each wave number $k>0$, the S-matrix $\hat{S}(k)$ is unitary.

We have to show that $\hat{S}(k)\hat{S}(k)^\dagger = I$ (see Problem 9.5). Explicitly,

$$\begin{pmatrix} \tau_{\leftarrow}(k) & \varrho_{\rightarrow}(k) \\ \varrho_{\leftarrow}(k) & \tau_{\rightarrow}(k) \end{pmatrix} \begin{pmatrix} \tau_{\leftarrow}(k)^\dagger & \varrho_{\leftarrow}(k)^\dagger \\ \varrho_{\rightarrow}(k)^\dagger & \tau_{\rightarrow}(k)^\dagger \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

In fact, by (8.14), $|\varrho_{\leftarrow}(k)|^2+|\tau_{\rightarrow}(k)|^2=1$ (conservation of the particle number current). Because of (8.16), it remains to show that

$$\varrho_{\leftarrow}(k)\tau_{\leftarrow}(k)^\dagger + \tau_{\rightarrow}(k)\varrho_{\rightarrow}(k)^\dagger = 0. \tag{8.18}$$

To this end, note that the scattering functions φ_k and φ_{-k} are solutions of the stationary Schrödinger equation (8.7). Since E and $U(x)$ are real, the conjugate complex scattering functions $\varphi^\dagger_{\pm k}$ are also solutions of (8.7). Therefore, the Wronskian

$$\begin{vmatrix} \varphi_k(x) & \varphi_{-k}(x)^\dagger \\ \frac{d}{dx}\varphi_k(x) & \frac{d}{dx}\varphi_{-k}(x)^\dagger \end{vmatrix}$$

is constant on the real line. Considering the Wronskian at the two points $x=-r$ and $x=r$, we obtain (8.18), after an elementary computation. □

The transfer matrix. Let us consider a solution φ of the stationary Schrödinger equation (8.7) on page 703 which has the following form:

$$\varphi(x) = \begin{cases} A_- \mathrm{e}^{\mathrm{i}kx} + B_- \mathrm{e}^{-\mathrm{i}kx} & \text{if} \quad x < -r, \\ A\mathrm{e}^{\mathrm{i}Kx} + B\mathrm{e}^{-\mathrm{i}Kx} & \text{if} \quad -r \le x \le r, \\ A_+ \mathrm{e}^{\mathrm{i}kx} + B_+ \mathrm{e}^{-\mathrm{i}kx} & \text{if} \quad x > r. \end{cases} \tag{8.19}$$

We are looking for a function $\varphi \in C^1(\mathbb{R})$. The function φ is continuous at the points $x = -r$ and $x = r$ iff

$$\begin{aligned} A_- \mathrm{e}^{-\mathrm{i}kr} + B_- \mathrm{e}^{\mathrm{i}kr} &= A\mathrm{e}^{-\mathrm{i}Kr} + B\mathrm{e}^{\mathrm{i}Kr}, \\ A\mathrm{e}^{\mathrm{i}Kr} + B\mathrm{e}^{-\mathrm{i}Kr} &= A_+ \mathrm{e}^{\mathrm{i}kr} + B_+ \mathrm{e}^{-\mathrm{i}kr}. \end{aligned}$$

The derivative φ' is continuous at the points $x = -r$ and $x = r$ iff

$$\begin{aligned} k(A_- \mathrm{e}^{-\mathrm{i}kr} - B_- \mathrm{e}^{\mathrm{i}kr}) &= K(A\mathrm{e}^{-\mathrm{i}Kr} - B\mathrm{e}^{\mathrm{i}Kr}), \\ K(A\mathrm{e}^{\mathrm{i}Kr} - B\mathrm{e}^{-\mathrm{i}Kr}) &= k(A_+ \mathrm{e}^{\mathrm{i}kr} - B_+ \mathrm{e}^{-\mathrm{i}kr}). \end{aligned}$$

This is a linear system for the complex coefficients $A_\pm, B_\pm$.

Proposition 8.2 *Let $k > 0$. For given complex numbers A_+ and B_+, the coefficients A_- and B_- are uniquely determined. Explicitly,*

$$\begin{pmatrix} A_- \\ B_- \end{pmatrix} = \begin{pmatrix} \alpha(k) & \beta(k) \\ \beta(k)^\dagger & \alpha(k)^\dagger \end{pmatrix} \begin{pmatrix} A_+ \\ B_+ \end{pmatrix} \tag{8.20}$$

where

$$\alpha(k) := \frac{1}{\tau_\rightarrow(k)}, \qquad \beta(k) := -\frac{\varrho_\rightarrow(k)}{\tau_\rightarrow(k)}. \tag{8.21}$$

For the square-well potential U from Fig. 8.4 on page 705, the transition amplitude $\tau_\rightarrow(k)$ and the reflection amplitude $\varrho_\rightarrow(k)$ are given by Prop. 8.1. The complex (2×2)-matrix from (8.20) is called the transfer matrix of the potential well U.

Proof. We start with the ansatz

$$\begin{pmatrix} A_- \\ B_- \end{pmatrix} = \begin{pmatrix} \alpha(k) & \beta(k) \\ \gamma(k) & \delta(k) \end{pmatrix} \begin{pmatrix} A_+ \\ B_+ \end{pmatrix}. \tag{8.22}$$

In particular, for the scattering function φ_k introduced in (8.13) on page 705, we obtain

$$A_- = 1, \ B_- = \varrho_\leftarrow(k) \quad \text{and} \quad A_+ = \tau_\rightarrow(k), \ B_+ = 0.$$

Hence

$$\begin{pmatrix} 1 \\ \varrho_\leftarrow(k) \end{pmatrix} = \begin{pmatrix} \alpha(k) & \beta(k) \\ \gamma(k) & \delta(k) \end{pmatrix} \begin{pmatrix} \tau_\rightarrow(k) \\ 0 \end{pmatrix}.$$

Similarly, the scattering function φ_{-k} from (8.15) on page 706 tells us that

$$\begin{pmatrix} 0 \\ \tau_\leftarrow(k) \end{pmatrix} = \begin{pmatrix} \alpha(k) & \beta(k) \\ \gamma(k) & \delta(k) \end{pmatrix} \begin{pmatrix} \varrho_\rightarrow(k) \\ 1 \end{pmatrix}.$$

This implies $1 = \alpha\tau_\rightarrow$ and $0 = \alpha\varrho_\rightarrow + \beta$. Passing to the conjugate complex equation of (8.19), we have to replace A_-, B_- by $B_-^\dagger, A_-^\dagger$, respectively, and so on. This implies

$$\begin{pmatrix} B_-^\dagger \\ A_-^\dagger \end{pmatrix} = \begin{pmatrix} \alpha(k) & \beta(k) \\ \gamma(k) & \delta(k) \end{pmatrix} = \begin{pmatrix} B_+^\dagger \\ A_+^\dagger \end{pmatrix}. \tag{8.23}$$

Comparing this with the first line of (8.22), we get $\gamma(k) = \beta(k)^\dagger$, and $\delta(k) = \alpha(k)^\dagger$. □

The transfer matrix plays a crucial role for the approximative computation of quantum scattering processes on the computer. This can be found in R. Gilmore, Elementary Quantum Mechanics in One Dimension, John Hopkins University Press, Baltimore, Maryland, 2004. The idea is to approximate general potentials by step functions and to compute the transfer matrix for each step.

8.3.3 Bound States

Consider again the square-well potential U from Fig. 8.4 on page 705. A solution φ of the stationary Schrödinger equation (8.7) on page 703 represents a bound state iff

$$\int_{\mathbb{R}} |\varphi(x)|^2 dx = 1. \tag{8.24}$$

For the energy E of the quantum particle, we have to distinguish the following three cases.

- Case 1: $E > 0$. Since $U = 0$ on the interval $]-\infty, -r[$, the general solution of equation (8.7) on the interval $]-\infty, -r[$ has the form

$$A_- \mathrm{e}^{\mathrm{i}kx} + B_- \mathrm{e}^{-\mathrm{i}kx} \tag{8.25}$$

  with complex coefficients A_-, B_- and $E = \frac{k^2}{2}$.
- Case 2: $E = 0$. The general solution of equation (8.7) on $]-\infty, -r[$ is given by

$$A_- + B_- x.$$

  Thus, bound states are impossible in cases 1 and 2.
- Case 3: $U_0 \le E < 0$. Let $\kappa > 0$. The general solution of (8.7) on $]-\infty, -r[$ reads as

$$A_- \mathrm{e}^{-\kappa x} + B_- \mathrm{e}^{\kappa x} \tag{8.26}$$

  where $E = -\frac{\kappa^2}{2}$. If this represents a bound state on the interval $]-\infty, -r[$, then $A_- = 0$. That is, the term $A_- \mathrm{e}^{-\kappa x}$ (with the exponential growth as $x \to -\infty$) vanishes.

Note that the solution (8.26) is obtained from (8.25) by choosing the imaginary wave number $k := \mathrm{i}\kappa$.

The bound-state energies. In order to compute all of the bound states φ, let us start with the ansatz

$$\varphi(x) = \begin{cases} B_- \mathrm{e}^{\kappa x} & \text{if} \quad x < -r, \\ A\mathrm{e}^{\mathrm{i}Kx} + B\mathrm{e}^{-\mathrm{i}Kx} & \text{if} \quad -r \le x \le r, \\ A_+ \mathrm{e}^{-\kappa x} & \text{if} \quad x > r. \end{cases} \tag{8.27}$$

Here, we choose $\kappa := \sqrt{-2E}$ and $K := \sqrt{2(E - U_0)}$.

Proposition 8.3 (i) *There exists precisely a finite number of bound-state energies* $E_1, \dots E_n$ *with* $n = 1, 2, \dots$ *Here,*

$$U_0 < E_1 < E_2 < \dots < E_n < 0.$$

All of the energy eigenvalues $E_1, \dots, E_n$ *are simple. The bound states* φ *are* $C^1(\mathbb{R})$*-functions. They are either even or odd functions with respect to the variable* x*.*

(ii) *The even bound states* φ *are given by*

$$\varphi(x) = \begin{cases} e^{\kappa x} & \text{if} \quad x < -r, \\ A \cos Kx & \text{if} \quad -r \le x \le r, \\ e^{-\kappa x} & \text{if} \quad x > r, \end{cases} \tag{8.28}$$

up to a multiplicative constant. The corresponding energy E *satisfies the transcendental equation*

$$\tan Kr = \frac{\kappa}{K}.$$

There exists at least one even bound state. The bound state of least energy (i.e., the ground state) is always even.

(iii) *The odd bound states* φ *are given by*

$$\varphi(x) = \begin{cases} e^{\kappa x} & \text{if} \quad x < -r, \\ B \sin Kx & \text{if} \quad -r \le x \le r, \\ -e^{-\kappa x} & \text{if} \quad x > r, \end{cases} \tag{8.29}$$

up to a multiplicative constant. The energy E *satisfies the transcendental equation*

$$\cot Kr = -\frac{\kappa}{K}.$$

Proof. (I) Simplicity of the bound-state eigenfunctions. Let $\varphi_1, \varphi_2 \in C^1(\mathbb{R})$. Suppose that φ_1 and φ_2 are bound-state solutions of the stationary Schrödinger equation (8.7) for fixed energy $E < 0$. Consider the interval $]-\infty, -r[$. By (8.27), there exists a complex number μ such that

$$\varphi_1(x) = \mu \varphi_2(x) \qquad \text{for all} \quad x \in]-\infty, -r[.$$

Hence $\varphi_1'(x) = \mu \varphi_2'(x)$ for all $x \in]-\infty, -r[$. By continuity,

$$\varphi_1(-r) = \mu \varphi_2(-r), \quad \varphi_1'(-r) = \mu \varphi_2'(-r).$$

Solving the Schrödinger equation (8.7) on the interval $[-r, r]$ for given initial values $\varphi_j(-r), \varphi_j'(-r)$, we get $\varphi_1(x) = \mu \varphi_2(x)$ for all $x \in [-r, r]$. The same argument shows that $\varphi_1 = \mu \varphi_2$ on $\mathbb{R}$.

(II) Symmetry of the bound-state eigenfunctions. Choose the function φ_1 as in (I). Define the reflection operator

$$R\varphi_1(x) := \varphi_1(-x) \qquad \text{for all} \quad x \in \mathbb{R}.$$

Since $U(-x) = U(x)$ for all $x \in \mathbb{R}$, the function $R\varphi_1$ is also a solution of (8.7). Hence $R\varphi_1 = \mu \varphi_1$ for some complex number μ. Therefore,

$$\varphi_1 = R^2 \varphi_1 = R(R\varphi_1) = \mu^2 \varphi_1.$$

If $\varphi_1 \neq 0$, then $\mu^2 = 1$. This implies either $\mu = 1$ or $\mu = -1$.

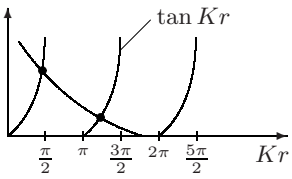


Fig. 8.5. Computation of bound-state energies

(III) Computation of the even bound-state eigenfunctions. Use the ansatz for the function φ from (8.27). Suppose that the function φ is even. Hence $A = B$. This yields (8.28). Then φ and φ' are continuous at the point $x = r$ iff

$$A \cos Kr = \mathrm{e}^{-\kappa r}, \qquad AK \sin Kr = \kappa \mathrm{e}^{-\kappa r}.$$

Hence $\tan Kr = \kappa/K$. This is the claim (ii). Analogously, we get (iii). □

The equation $\tan Kr = \frac{\kappa}{K}$ with $\kappa = \sqrt{-2E}$ and $K = \sqrt{2(E - U_0)}$ can be solved graphically. To this end, setting $x := Kr$ we get the equation

$$\tan x = \frac{\sqrt{\xi^2 - x^2}}{x}, \qquad 0 < x < \xi.$$

Here, $\xi := r\sqrt{2|U_0|}$. The solutions x of this equation are the intersection points between the two curves plotted in Fig. 8.5. There exists at least one intersection point, and the number of intersection points is finite. For the intersection point x, the corresponding bound-state energy is given by

$$E = U_0 + \frac{K^2}{2} = U_0 \left(1 - \frac{x^2}{\xi^2}\right).$$

The corresponding state function φ is even. It can be shown that the number of intersection points in Fig. 8.5 increases if the parameter ξ increases. In particular, if the width $2r$ or the depth $|U_0|$ of the potential well goes to infinity (i.e., $\xi \to \infty$), then the number of bound-state energies goes to infinity as well.

Similarly, we can treat the equation $\cot Kr = -\frac{\kappa}{K}$ which yields the bound-state energies for odd state functions φ. It turns out that the ground state of energy E_1 is always even with $Kr \in [0, \frac{\pi}{2}]$. The first excited state of energy E_2 is always odd with $Kr \in [\frac{\pi}{2}, \pi]$.

8.3.4 Bound-State Energies and the Singularities of the *S*-Matrix

We now want to compute the bound-state energies by using analyticity properties of the S-matrix. Set

$$k := \mathrm{i}\kappa$$

with $\kappa > 0$. Consider the function φ from (8.27) with $B_- = 1$. Since Prop. 8.2 on page 709 remains valid by using analytic continuation, we obtain

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha(\mathrm{i}\kappa) & \beta(\mathrm{i}\kappa) \\ \gamma(\mathrm{i}\kappa) & \delta(\mathrm{i}\kappa) \end{pmatrix} \begin{pmatrix} A_+ \\ 0 \end{pmatrix}. \tag{8.30}$$

For a bound state φ, we have $A_+ \neq 0$. Therefore, it follows from (8.30) that φ is a bound state iff $\alpha(\mathrm{i}\kappa) = 0$. Since $\tau_{\rightarrow} = 1/\alpha$, we get the following.

Proposition 8.4 *There exists a bound state for the energy $E_b < 0$ iff the function $\kappa \mapsto \tau_{\rightarrow}(\mathrm{i}\kappa)$ has a pole at the point $\kappa = \sqrt{-2E_b}$.*

Recall from (8.16) and (8.17) that the S-matrix is given by

$$\hat{S}(\mathrm{i}\kappa) := \begin{pmatrix} \tau_{\rightarrow}(\mathrm{i}\kappa) & \varrho_{\rightarrow}(\mathrm{i}\kappa) \\ \varrho_{\rightarrow}(\mathrm{i}\kappa) & \tau_{\rightarrow}(\mathrm{i}\kappa) \end{pmatrix}.$$

Thus, Proposition 8.4 tells us the crucial fact that the energies of the bound states are related to the singularities of the S-matrix. Explicitly, by Prop. 8.1 on page 707,

$$\tau_{\rightarrow}(\mathrm{i}\kappa) = \frac{\mathrm{e}^{2\kappa r}}{\cos 2Kr + \frac{1}{2}\left(\frac{\kappa}{K} - \frac{K}{\kappa}\right)\sin 2Kr} \tag{8.31}$$

along with $\kappa = \sqrt{-2E}$ and $K = \sqrt{2(E + |U_0|)}$. Consequently, we obtain the equation

$$\boxed{\cos 2Kr + \frac{1}{2}\left(\frac{\kappa}{K} - \frac{K}{\kappa}\right)\sin 2Kr = 0} \tag{8.32}$$

for determing the bound-state energy E. Using the well-known addition theorem $\cot 2x = \frac{1}{2}(\cot x - \tan x)$, equation (8.32) is equivalent to

$$\tan Kr - \cot Kr = \frac{\kappa}{K} - \frac{K}{\kappa}.$$

This is satisfied iff [4]

$$\boxed{\tan Kr = \frac{\kappa}{K} \qquad \text{or} \qquad \cot Kr = -\frac{\kappa}{K}.}$$

By Prop. 8.3 on page 711, these two equations determine precisely the bound-state energies.

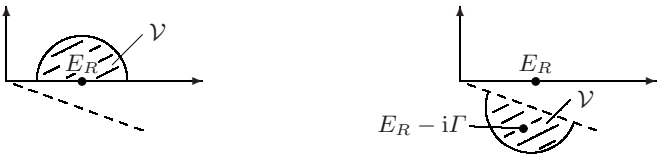
8.3.5 The Energetic Riemann Surface, Resonances, and the Breit–Wigner Formula

Again let us consider the motion of a homogeneous stream of quantum particles on the real line from left to right under the action of the square-well potential U pictured in Fig. 8.4 on page 705. We are given the particle energy $E > 0$. According to Prop. 8.1 on page 707, the transmission probability is given by

$$|\tau_{\rightarrow}(k)|^2 = \left(1 + \frac{\sin^2 2Kr}{\frac{4E}{|U_0|}\left(1 + \frac{E}{|U_0|}\right)}\right)^{-1}$$

where $E = \frac{k^2}{2}$ and $K = \sqrt{2(E + |U_0|)}$. This yields the reflection probability $|\varrho_{\leftarrow}(k)|^2 = 1 - |\tau_{\rightarrow}(k)|^2$.

[4] Observe that an inspection of the graph of the function $f(x) := x - \frac{1}{x}$ tells us that for given $\xi > 0$, the equation $f(x) = f(\xi)$ has precisely two solutions. Explicitly, the solutions are $x = \xi$ and $x = -\frac{1}{\xi}$.



(a) first sheet (b) second sheet

Fig. 8.6. The energetic Riemann surface

Resonances. Physicists say that the particle stream is in resonance at the energy E_R iff the transmission probability is equal to one, that is,

$$|\tau_{\rightarrow}(k)|^2 = 1$$

with $E_R = \frac{k^2}{2}$. Consequently, the reflection probability is equal to zero,

$$|\varrho_{\rightarrow}(k)|^2 = 0.$$

Explicitly, we have $|\tau_{\rightarrow}(k)|^2 = 1$ iff $\sin^2 2Kr = 0$. Hence $2Kr = n\pi$ where n is a positive integer. Therefore, the resonance energies are given by

$$E_R = \frac{K^2}{2} + U_0 = \frac{n^2\pi^2}{8r^2} - |U_0|, \qquad n = n_0, n_0+1, n_0+2, \ldots$$

Here, n_0 is the smallest positive integer with $\frac{n_0^2\pi^2}{8r^2} \geq |U_0|$.

The energetic Riemann surface. Transmission amplitudes depend on the wave number $k = \sqrt{2E}$. In order to allow complex parameters k via analytic continuation, we have to study the function $E \mapsto \sqrt{E}$ on its Riemann surface (Fig. 8.6). Explicitly, we consider the first sheet

$$\Sigma_1 := \{E \in \mathbb{C} : E = |E| \cdot \mathrm{e}^{\mathrm{i}\alpha},\ 0 \leq \alpha < 2\pi\}$$

and the second sheet

$$\Sigma_2 := \{E \in \mathbb{C} : E = |E| \cdot \mathrm{e}^{\mathrm{i}\beta},\ 2\pi \leq \beta < 4\pi\}.$$

Observe that the arguments α and β of E are different on the first and second sheet. Intuitively, we cut two exemplars of the complex plane along the positive real axis, and we glue together the two sheets in such a way that the set $\mathcal{V}$ pictured in Fig. 8.6 represents a neighborhood of the point E_R on the Riemann surface. To illustrate this, define

$$E(t) := \varrho \mathrm{e}^{\mathrm{i}t}, \qquad 0 \leq t \leq 4\pi,\ \varrho > 0.$$

Consider this as a motion $E = E(t)$ on the two sheets Σ_1 and Σ_2. Here, we move counterclockwise along the circle of radius ϱ around the origin. More precisely,

- we start at the point $E(0) = \varrho$ on the first sheet at time $t = 0$;
- we pass through the point $E(\pi) = -\varrho$ on the first sheet at time $t = \pi$;
- we then arrive at the point $E(2\pi) = \varrho$ on the second sheet at time $t = 2\pi$,
- and we finish the trip at the point $E(4\pi) = \varrho$ on the first sheet at time $t = 4\pi$.

We now define

$$\sqrt{E} := \begin{cases} \sqrt{|E|} \cdot \mathrm{e}^{\mathrm{i}\alpha/2} & \text{if} \quad E \in \Sigma_1, \\ \sqrt{|E|} \cdot \mathrm{e}^{\mathrm{i}\beta/2} & \text{if} \quad E \in \Sigma_2. \end{cases}$$

Similarly, $\sqrt{2E} := \sqrt{2}\sqrt{E}$. In particular, this definition is chosen in such a way that the function $E \mapsto \sqrt{E}$ is continuous on the Riemann surface. Note that the values of $\sqrt{E}$ on the two sheets differ by the sign. For example, if we consider the point $E = 1$ on the first (resp. second) sheet, then $\sqrt{E} = 1$ (resp. $\sqrt{E} = -1$).

Analytic continuation of the transmission amplitude. From Prop. 8.1, we get

$$\tau_{\rightarrow}(\sqrt{2E}) = \frac{\mathrm{e}^{-2\mathrm{i}k(E)r}}{\cos 2K(E)r - \frac{\mathrm{i}}{2}\left(\frac{K(E)}{k(E)} + \frac{k(E)}{K(E)}\right)\sin 2K(E)r} \tag{8.33}$$

for all $E > 0$. Here, we set $k(E) := \sqrt{2E}$ and $K(E) := \sqrt{2(E + |U_0|)}$. However, by analytic continuation, this formula makes sense for all complex numbers E on the Riemann surface of the function $E \mapsto \sqrt{E}$. The singularities of the function $E \mapsto \tau_{\rightarrow}(\sqrt{2E})$ encode important properties of the quantum particle on the real line.

(i) The transmission amplitude $E \mapsto \tau_{\rightarrow}(\sqrt{2E})$ has poles at the points

$$E_1 < E_2 < \ldots < E_n < 0$$

on the negative real axis of the first sheet of the energetic Riemann surface. These poles are the energies of bound states of the particle.

(ii) The transmission probability function $E \mapsto |\tau_{\rightarrow}(\sqrt{2E})|^2$ is equal to one at the points

$$E_{R_n} = \frac{n^2\pi^2}{8r^2} - |U_0|, \qquad n = n_0, n_0 + 1, n_0 + 2, \ldots$$

where n_0 is the smallest positive integer with $\frac{n_0^2\pi^2}{8r^2} \geq |U_0|$. The points $E_{R_1}, E_{R_2}, \ldots$ correspond to the energies of so-called resonances.

(iii) The transmission amplitude $\tau_{\rightarrow}(\sqrt{2E})$ allows the following approximation

$$\tau_{\rightarrow}(\sqrt{2E}) = \frac{\mathrm{i}(-1)^n \Gamma_n \mathrm{e}^{-2\mathrm{i}k(E_{R_n})r}}{E - (E_{R_n} - \mathrm{i}\Gamma_n)}, \qquad n = n_0, n_0 + 1, \ldots \tag{8.34}$$

for all complex energies E in some open neighborhood $\mathcal{V}$ of the real resonance energy $E_R = E_{R_n}$ on the energetic Riemann surface (see Fig. 8.6 on page 714). Here,

$$\Gamma_n := \frac{\sqrt{2E_{R_n}}}{r}\left(1 + \frac{E_{R_n}}{E_{R_n} + |U_0|}\right)^{-1}.$$

This will be proved below. Thus, the transmission amplitude has poles at the energy points

$$E_{R_n} - \mathrm{i}\Gamma_n, \qquad n = n_0, n_0 + 1, n_0 + 2, \ldots$$

which lie on the second sheet of the energetic Riemann surface, by the construction of the open neighborhood $\mathcal{V}$ in Fig. 8.6. In terms of elementary particle physics, this serves as a model for an unstable particle of energy E_{R_n} and mean lifetime Γ_n. For the transmission probability, we get

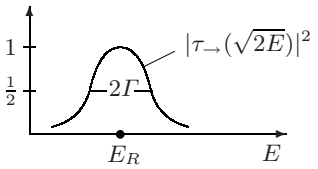


Fig. 8.7. The Breit–Wigner formula

$$\boxed{|\tau_{\rightarrow}(\sqrt{2E})|^2 = \frac{\Gamma_n^2}{(E-E_{R_n})^2+\Gamma_n^2}.} \tag{8.35}$$

This is the famous Breit–Wigner formula. The function $E \mapsto |\tau_{\rightarrow}(\sqrt{2E})|^2$ is pictured in Fig. 8.7 with $E_R = E_{R_n}$ and $\Gamma = \Gamma_n$.
Observe that resonances can be measured in particle accelerator experiments. As a rule, physicists measure curves of the type pictured in Fig. 8.7. Such curves are used in order to determine the energy, E_R, and the mean lifetime Γ of unstable particles.

The Breit–Wigner formula. Let us motivate relation (8.34). By equation (8.33),

$$\tau_{\rightarrow}(\sqrt{2E}) = \frac{1}{\cos 2K(E)r} \cdot \frac{\mathrm{e}^{-2\mathrm{i}k(E)r}}{1-\frac{\mathrm{i}}{2}\left(\frac{K(E)}{k(E)}+\frac{k(E)}{K(E)}\right)\tan 2K(E)r} \tag{8.36}$$

where $k(E) := \sqrt{2E}$ and $K(E) := \sqrt{2(E+|U_0|)}$. Consider the point E_R on the positive real axis of the first sheet Σ_1 of the energetic Riemann surface, and choose an open neighborhood $\mathcal{V}$ of the point E_R on the energetic Riemann surface as pictured in Fig. 8.6 on page 714.

Now to the point. The function $E \mapsto \sqrt{E}$ is holomorphic on the open set $\mathcal{V}$. Consequently, the function $E \mapsto \tau_{\rightarrow}(\sqrt{2E})$ is holomorphic on $\mathcal{V}$. For the points $E \in \mathcal{V}$, Taylor expansion yields

$$\tau_{\rightarrow}(\sqrt{2E}) = \tau_{\rightarrow}(\sqrt{2E_R}) + \frac{d\tau_{\rightarrow}(\sqrt{2E})}{dE}_{|E=E_R}(E-E_R) + o(E-E_R),$$

as $E \to E_R$. Let us compute this. Recalling that $2K(E_R)r = n\pi$, we get

$$\cos 2K(E_R)r = (-1)^n, \qquad \tan 2K(E_R)r = 0$$

and

$$\frac{d(2K(E)r)}{dE}_{|E=E_R} = \frac{2r}{K(E_R)}.$$

Therefore, up to terms of order $o(E-E_R)$ as $E \to E_R$, we obtain

$$\tan 2K(E)r = \frac{2r}{K(E_R)} \cdot (E-E_R),$$

and hence

$$\frac{1}{2}\left(\frac{K(E)}{k(E)}+\frac{k(E)}{K(E)}\right)\tan 2K(E)r = \frac{E-E_R}{\Gamma(E_R)}$$

with

$$\frac{1}{\Gamma(E_R)} := \frac{1}{2}\left(\frac{K(E_R)}{k(E_R)} + \frac{k(E_R)}{K(E_R)}\right)\frac{2r}{K(E_R)} = \frac{r}{\sqrt{2E_R}}\left(1 + \frac{E_R}{E_R + |U_0|}\right).$$

By (8.36), this implies

$$\tau_{\rightarrow}(\sqrt{2E}) = \frac{(-1)^n \mathrm{e}^{-2\mathrm{i}k(E_R)r}}{1 - \frac{\mathrm{i}(E-E_R)}{\Gamma(E_R)}} = \frac{\mathrm{i}(-1)^n \Gamma(E_R)\mathrm{e}^{-2\mathrm{i}k(E_R)r}}{E - E_R + \mathrm{i}\Gamma(E_R)},$$

up to terms of order $o(E - E_R)$ as $E \to E_R$. This yields the claim (8.34).

The relation between energy operators, non-real energy eigenvalues, and resonances (unstable particles) is studied in

P. Hislop and I. Sigal, Introduction to Spectral Theory: with Applications to Schrödinger Operators, Springer, New York, 1996.

This monograph uses the functional-analytic approach to operators which are *not* self-adjoint. We also refer to

S. Gustafson and I. Sigal, Mathematical Concepts of Quantum Mechanics, Springer, Berlin, 2003.

Our discussion above shows that scattering processes are closely related to analyticity properties of the S-matrix elements with respect to the wave number. There exist far-reaching generalizations in the context of analytic S-matrix theory. We recommend the monograph by

A. Barut, The Theory of the Scattering Matrix, MacMillan, New York, 1967.

Summary. Let us summarize the results on the motion of a quantum particle on the real line under the action of the square-well potential U introduced in (8.3) on page 701 with $U_0 < 0$. The stationary Schrödinger equation

$$\boxed{-\tfrac{1}{2}\varphi_{xx}(x) + U(x)\varphi(x) = E\varphi(x), \qquad x \in \mathbb{R}} \tag{8.37}$$

has the following solutions.

(i) Bound states. There exist the eigenfunctions χ_j, $j = 1, \ldots, n$, with the corresponding energies E_j, $j = 1, \ldots n$, where

$$U_0 < E_1 < \ldots < E_n < 0.$$

The eigenvalues are simple. The eigenfunctions $\chi_1, \ldots, \chi_n$ form an orthonormal system in the Hilbert space $L_2(\mathbb{R})$. Explicitly,

$$\langle \chi_j | \chi_k \rangle_2 = \delta_{jk}, \qquad j, k = 1, \ldots, n.$$

Here, we use the inner product

$$\langle \varphi | \chi \rangle_2 := \int_{\mathbb{R}} \varphi(x)^\dagger \chi(x) dx \tag{8.38}$$

for all functions $\varphi, \chi \in L_2(\mathbb{R})$. In what follows, we will also use the symbol $\langle \varphi | \chi \rangle_2$ in the case where at least one of the functions φ, χ does not lie in the Hilbert space $L_2(\mathbb{R})$, but only the integral $\int_{\mathbb{R}} \varphi(x)^\dagger \chi(x) dx$ exists. For example, if $\varphi(x) := 1$ and $\chi(x) := \mathrm{e}^{-x^2/2}$ for all $x \in \mathbb{R}$, then

$$\langle \varphi | \chi \rangle_2 = \int_{\mathbb{R}} \varphi(x)^\dagger \chi(x) dx = \int_{\mathbb{R}} \chi(x) dx = \sqrt{2\pi}.$$

(ii) Scattering eigenfunctions. For each wave number $k > 0$, there exist the two scattering eigenfunctions φ_k and φ_{-k} with the energy $E = \frac{k^2}{2}$. The eigenfunctions $\varphi_{\pm k}$ do not lie in the Hilbert space $L_2(\mathbb{R})$. In fact, we have

$$\langle \varphi_{\pm k} | \varphi_{\pm k} \rangle_2 = \int_{\mathbb{R}} |\varphi_{\pm k}(x)|^2 dx = \infty \qquad \text{for all} \quad k > 0.$$

The explicit form of the eigenfunctions $\varphi_{\pm k}$ can be found in Prop. 8.1 on page 707. The scattering eigenfunctions $\varphi_{\pm k}$ have the following asymptotic behavior

$$\varphi_{\pm k}(x) \simeq \tau_{\rightarrow}(k)\, \mathrm{e}^{\pm \mathrm{i} k x} \qquad \text{as} \quad x \to \pm\infty.$$

8.3.6 The Jost Functions

Fix the wave number $k > 0$ and choose the square-well potential U from (8.3) on page 701 with $U_0 \leq 0$. Consider the stationary Schrödinger equation (8.37) on page 717 with the energy

$$E := \frac{k^2}{2}.$$

Observe that we have the following general solution principle.

(P) *If two $C^1(\mathbb{R})$-solutions φ and χ of equation (8.37) (in the sense of distributions) coincide on some open interval, then they coincide on the real line.*

This follows from the unique solvability of the initial-value problem for (8.37). Since the square-well potential U vanishes outside the interval $[-r, r]$, the equation (8.37) has the special solutions

$$J_k(x) := \mathrm{e}^{\mathrm{i} k x} \qquad \text{for all} \quad x > r$$

and

$$J_{-k}(x) := \mathrm{e}^{-\mathrm{i} k x} \qquad \text{for all} \quad x < -r.$$

These functions can be uniquely extended to solutions of the stationary Schrödinger equation (8.37) on the real line. Explicitly,

$$J_{\pm k}(x) = \frac{\varphi_{\pm k}(x)}{\tau_{\rightarrow}(k)} \qquad \text{for all} \quad x \in \mathbb{R}.$$

The functions J_k and J_{-k} are called the Jost functions.[5]

Proposition 8.5 *Let $k > 0$. For all $x \in \mathbb{R}$, we have*

$$\begin{pmatrix} J_{-k}(x) \\ J_{-k}(x)^\dagger \end{pmatrix} = \begin{pmatrix} a(k) & b(k) \\ b(k)^\dagger & a(k)^\dagger \end{pmatrix} \begin{pmatrix} J_k(x)^\dagger \\ J_k(x) \end{pmatrix}$$

where

$$a(k) := \frac{1}{\tau_{\rightarrow}(k)}, \qquad b(k) := \frac{\varrho_{\rightarrow}(k)}{\tau_{\rightarrow}(k)}.$$

[5] If the potential vanishes identically, $U \equiv 0$, then we get the free Jost functions $J_{\pm k}(x) = \mathrm{e}^{\pm \mathrm{i} k x}$ for all $x \in \mathbb{R}$.

Therefore, if we know the Jost functions, then we also know the coefficients $a(k), b(k)$ which imply the reflection and transmission amplitudes. This tells us that

The Jost functions know all about the S-matrix.

Proof. Let $x > r$. Then $J_k(x) = \mathrm{e}^{\mathrm{i}kx}$. By (8.15) on page 706, we get

$$J_{-k}(x) = \frac{\varphi_{-k}(x)}{\tau_{\rightarrow}(k)} = \frac{\mathrm{e}^{-\mathrm{i}kx} + \varrho_{\rightarrow}(k)\mathrm{e}^{\mathrm{i}kx}}{\tau_{\rightarrow}(k)}.$$

Hence $a(k) = 1/\tau_{\rightarrow}(k)$ and $b(k) = \varrho_{\rightarrow}(k)/\tau_{\rightarrow}(k)$. □

8.3.7 The Fourier–Stieltjes Transformation

> The Fourier transformation and its generalizations lie at the heart of mathematics.
>
> Folklore

Let us again consider the square-well potential U from (8.3) on page 701 with $U_0 < 0$. Let us use the notation summarized on page 717. The key formula reads as

$$\boxed{\varphi(x) = \sum_{j=1}^{n} c_j \chi_j(x) + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} c(k)\varphi_k(x)dk} \tag{8.39}$$

for all $x \in \mathbb{R}$. For the so-called Fourier–Stieltjes coefficients $c_j, c(k)$, we will get

$$c_j := \int_{-\infty}^{\infty} \chi_j(x)^{\dagger}\varphi(x)dx, \qquad c(k) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \varphi_k(x)^{\dagger}\varphi(x)dx$$

where $j = 1, \ldots, n$ and $k \in \mathbb{R} \setminus \{0\}$. Moreover, in the special case where $k = 0$, we set $\varphi_0(x) := 0$ for all $x \in \mathbb{R}$, and $c(0) := 0$. Let us add the Parseval equation[6]

$$\int_{-\infty}^{\infty} \varphi(x)^{\dagger}\chi(x)dx = \sum_{j=1}^{n} c_j^{\dagger} d_j + \int_{-\infty}^{\infty} c(k)^{\dagger} d(k)dk. \tag{8.40}$$

Here, $c_j, c(k)$ and $d_j, d(k)$ are the Fourier–Stieltjes coefficients of the functions φ and χ, respectively. We now use the space $\mathcal{S}(\mathbb{R})$ of smooth, rapidly decreasing functions $\varphi : \mathbb{R} \to \mathbb{C}$.[7]

Theorem 8.6 *For all functions $\varphi, \chi \in \mathcal{S}(\mathbb{R})$, we have the Fourier–Stieltjes expansion formula* (8.39) *and the Parseval equation* (8.40).

For the proof, we refer to Berezin and Shubin (1991), p. 126. This theorem is the prototype of general expansion theorems in mathematics which date back to Hermann Weyl's 1908 dissertation in Göttingen on singular integral operators and his 1910 habilitation thesis in Göttingen on singular differential operators.[8] Weyl's

[6] Parseval des Chénes (1755–1836), Fourier (1768–1830), Stieltjes (1856–1894).

[7] The precise definition of $\mathcal{S}(\mathbb{R})$ can be found in Vol. I, Sec. 10.3.3.

[8] H. Weyl, On ordinary differential equations with singularities, Math. Ann. **68** (1910), 220–269 (in German). Weyl's theory was completed by the following fundamental paper:
K. Kodaira, The eigenvalue problem for ordinary differential equations of the second order and Heisenberg's theory of S-matrices. Amer. J. Math. **71** (1949), 921–945.

dissertation was supervised by Hilbert. If the potential vanishes identically, $U \equiv 0$, then the eigenfunctions χ_j drop out and we have $\varphi_k(x) = \mathrm{e}^{\mathrm{i}kx}$ for all $x \in \mathbb{R}$. In this special case, equation (8.39) passes over to the classical Fourier transform. Let us now discuss the relation to both the theory of unitary operators and the theory of generalized eigenfunctions in Hilbert spaces.

Unitary extension. Motivated by the Fourier–Stieljes expansion formula (8.39), we set

$$\boxed{\mathcal{U}\varphi := (c_1, \ldots, c_n; c(k))_{k\in\mathbb{R}}.}$$

This defines a linear operator $\mathcal{U} : \mathcal{S}(\mathbb{R}) \to Y$ from the space $\mathcal{S}(\mathbb{R})$ of rapidly decreasing test functions into the product Hilbert space

$$Y := \mathbb{C}^n \times L_2(\mathbb{R})$$

equipped with the inner product

$$\sum_{j=1}^{n} c_j^\dagger d_j + \int_{-\infty}^{\infty} c(k)^\dagger d(k) dk.$$

The Parseval equation (8.40) tells us that

$$\langle \mathcal{U}\varphi | \mathcal{U}\chi \rangle_Y = \langle \varphi | \chi \rangle_2 \qquad \text{for all} \quad \varphi, \chi \in \mathcal{S}(\mathbb{R}).$$

Since the set $\mathcal{S}(\mathbb{R})$ is dense in the Hilbert space $L_2(\mathbb{R})$, the operator $\mathcal{U}$ can be uniquely extended to a unitary operator[9]

$$\boxed{\mathcal{U} : L_2(\mathbb{R}) \to Y.} \tag{8.41}$$

This operator is called the Fourier–Stieltjes transform generated by the stationary Schrödinger equation (8.37).

8.3.8 Generalized Eigenfunctions of the Hamiltonian

Motivated by the stationary Schrödinger equation (8.37), we introduce the Hamiltonian

$$H\varphi := -\frac{1}{2}\frac{d^2\varphi}{dx^2} + U\varphi$$

for all test functions $\varphi \in \mathcal{S}(\mathbb{R})$. Moreover, for any test function $\varphi \in \mathcal{S}(\mathbb{R})$, we define the functionals

- $F_j(\varphi) := \langle \chi_j | \varphi \rangle_2$ for $j = 1, \ldots, n$, and
- $G_k(\varphi) := \frac{1}{\sqrt{2\pi}} \langle \varphi_k | \varphi \rangle_2$ for all $k \in \mathbb{R}$. [10]

In terms of the Fourier–Stieltjes transformation, this means that we assign the Fourier–Stieltjes coefficients to the test function φ. Briefly,

$$F_j(\varphi) = c_j, \quad G_k(\varphi) = c(k), \qquad j = 1, \ldots, n, \ k \in \mathbb{R}.$$

This way, we obtain linear, sequentially continuous functionals

[9] See Zeidler (1995a), Sec. 3.6, quoted on page 1049.

[10] Note that the integral $\langle \varphi_k | \varphi \rangle_2 = \int_{\mathbb{R}} \varphi_k(x)^\dagger \varphi(x) dx$ exists for all $\varphi \in \mathcal{S}(\mathbb{R})$ because of the boundedness of the functions φ_k.

$$F_j, G_k : \mathcal{S}(\mathbb{R}) \to \mathbb{C}.$$

In other words, F_j and G_k are tempered distributions[11] which lie in the space $\mathcal{S}'(\mathbb{R})$, and we have

$$F_j(H\varphi) = E_j F_j(\varphi), \qquad G_k(H\varphi) = E(k) G_k(\varphi) \tag{8.42}$$

for all $\varphi \in \mathcal{S}(\mathbb{R})$ with the energies E_j and $E(k) := \frac{k^2}{2}$. Here, $j = 1, \ldots, n$ and $k \in \mathbb{R}$. In fact, since E_j is real, we get

$$\langle \chi_j | H\varphi \rangle_2 = \langle H\chi_j | \varphi \rangle_2 = E_j \langle \chi_j | \varphi \rangle \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).$$

Furthermore, note that $H\varphi_k = E(k)\varphi_k$ for all $k \in \mathbb{R}$. Thus, for all $\varphi \in \mathcal{S}(\mathbb{R})$, integration by parts yields

$$\langle \varphi_k | H\varphi \rangle_2 = \langle H\varphi_k | \varphi \rangle_2 = E(k) \langle \varphi_k | \varphi \rangle_2.$$

Note that the boundary terms of the integration-by-parts formula vanish, since the function φ decreases rapidly at $\pm\infty$.

Theorem 8.7 *The family $\{F_1, \ldots, F_n, G_k\}_{k\in\mathbb{R}}$ represents a complete orthonormal system of generalized eigenfunctions of the Hamiltonian H.*

Explicitly, this means that the following three conditions are satisfied.

(i) Generalized eigenfunctions. In terms of the theory of tempered distributions, we have

$$H\chi_j = E_j \chi_j, \qquad HG_k = E(k) G_k, \qquad j = 1, \ldots, n, \ k \in \mathbb{R}.$$

(ii) Completeness. For any given $\varphi \in \mathcal{S}(\mathbb{R})$, it follows from

$$F_j(\varphi) = G_k(\varphi) = 0$$

for all indices $j = 1, \ldots, n$, $k \in \mathbb{R}$ that $\varphi = 0$.

(iii) Orthonormality. The operator $\mathcal{U} : \mathcal{S}(\mathbb{R}) \to Y$ given by

$$\mathcal{U}(\varphi) := (F_1(\varphi), \ldots, F_n(\varphi); G_k(\varphi))_{k\in\mathbb{R}}$$

can be uniquely extended to a unitary operator $\mathcal{U} : L_2(\mathbb{R}) \to Y$.

Proof. Ad (i). This is a reformulation of (8.42).

Ad (ii). This follows from the Parseval equation (8.40),

$$\int_{-\infty}^{\infty} |\varphi(x)|^2 dx = \sum_{j=1}^{n} |c_j|^2 + \int_{-\infty}^{\infty} |c(k)|^2 dk,$$

with $c_j = 0$ and $c(k) = 0$ for all indices $j = 1, \ldots, n$, $k \in \mathbb{R}$.

Ad (iii). See (8.41). □

[11] See Sec. 10.3.3 of Vol. I.

8.3.9 Quantum Dynamics and the Scattering Operator

> Wave operators describe the motion of wave packets; the scattering operator is closely related to the S-matrix.
>
> Folklore

Again let U denote the square-well potential from (8.3) with $U_0 < 0$ (see Fig. 8.4 on page 705). The preceding results admit the following physical interpretation. Consider again the Hamiltonian H given by

$$(H\varphi)(x) := \left(-\frac{1}{2}\frac{d^2}{dx^2} + U(x)\right)\varphi(x), \qquad x \in \mathbb{R}$$

for all $\varphi \in \mathcal{S}(\mathbb{R})$. The operator $H : \mathcal{S}(\mathbb{R}) \to L_2(\mathbb{R})$ is an essentially self-adjoint operator on the Hilbert space $L_2(\mathbb{R})$, which can be extended to the self-adjoint operator

$$H : D(H) \to L_2(\mathbb{R})$$

on the Hilbert space $L_2(\mathbb{R})$. By definition, the free Hamiltonian H_{free} is obtained from H by switching off the potential, $U \equiv 0$.

(i) Dynamics: For each given initial state $\psi_0 \in L_2(\mathbb{R})$, the motion of the quantum particle is given by

$$\boxed{\psi(t) = \mathrm{e}^{-\mathrm{i}Ht}\psi_0 \qquad \text{for all} \quad t \in \mathbb{R}.} \tag{8.43}$$

(ii) Bound states. We have the eigensolutions

$$\boxed{H\chi_j = E_j\chi_j, \qquad j = 1, \ldots, n}$$

with $U_0 < E_1 < \ldots < E_n < 0$. The energy eigenstates $\chi_1, \ldots, \chi_n$ form an orthonormal system in $L_2(\mathbb{R})$. For given initial state χ_j with $j = 1, \ldots, n$, the corresponding dynamics is given by $\psi_j(t) = \mathrm{e}^{-\mathrm{i}E_j t}\chi_j$ for all $t \in \mathbb{R}$.

(iii) Scattering states. Let B be the linear hull of the energy eigenstates $\chi_1, \ldots, \chi_n$. The elements of B are called bound states. We have the orthogonal decomposition

$$L_2(\mathbb{R}) = B \oplus B^{\perp}.$$

By definition, $\varphi \in B^{\perp}$ iff $\langle\varphi|\chi_j\rangle_2 = 0$ for all $j = 1, \ldots, n$. The elements of the orthogonal complement $B^{\perp}$ to the space B of bound states are called scattering states of the stationary Schrödinger equation (8.37). It turns out that the following holds.

> *If the initial state ψ_0 is a scattering state, then the dynamics (8.43) is asymptotically free in both the distant future and the remote past.*

Explicitly, there exist states φ_{out} and φ_{in} in $L_2(\mathbb{R})$ such that

$$\lim_{t \to +\infty} ||\mathrm{e}^{-\mathrm{i}tH}\psi_0 - \mathrm{e}^{-\mathrm{i}tH_{\text{free}}}\varphi_{\text{out}}|| = 0$$

and

$$\lim_{t \to -\infty} ||\mathrm{e}^{-\mathrm{i}tH}\psi_0 - \mathrm{e}^{-\mathrm{i}tH_{\text{free}}}\varphi_{\text{in}}|| = 0.$$

This justifies the designation 'scattering states' for the states in $B^{\perp}$.

(iv) Wave operators. Setting

$$W_{\text{out}}\varphi_{\text{out}} := \psi_0, \qquad W_{\text{in}}\varphi_{\text{in}} := \psi_0,$$

we get the wave operators $W_{\text{out}}, W_{\text{in}} : L_2(\mathbb{R}) \to B^\perp$ which are unitary from the original Hilbert space $L_2(\mathbb{R})$ onto the space $B^\perp$ of scattering states. In terms of limits,[12]

$$\boxed{W_{\text{out}}\varphi_{\text{out}} = \lim_{t\to+\infty} \mathrm{e}^{\mathrm{i}tH}\mathrm{e}^{-\mathrm{i}tH_{\text{free}}}\varphi_{\text{out}}.}$$

Similarly, we get $W_{\text{in}}\varphi_{\text{in}} = \lim_{t\to-\infty} \mathrm{e}^{\mathrm{i}tH}\mathrm{e}^{-\mathrm{i}tH_{\text{free}}}\varphi_{\text{in}}$.

(v) The scattering operator S. Set

$$\boxed{\mathsf{S}\varphi_{\text{in}} := \varphi_{\text{out}}.}$$

In other words, the scattering operator connects the initial state of the free motion in the remote past with the initial state of the free motion in the distant future. This way, the time-dependent scattering process is reduced to the investigation of initial states at time $t = 0$ for free motions. In terms of the wave operators, the scattering operator is given by

$$\boxed{\mathsf{S} = W_{\text{out}}^{-1}W_{\text{in}}.} \tag{8.44}$$

This means that $W_{\text{out}}\mathsf{S} = W_{\text{in}}$. In other words, the following diagram

$$\begin{array}{ccc} L_2(\mathbb{R}) & \xrightarrow{\ \mathsf{S}\ } & L_2(\mathbb{R}) \\ & {\scriptstyle W_{\text{in}}}\searrow & \downarrow{\scriptstyle W_{\text{out}}} \\ & & L_2(\mathbb{R}) \end{array} \tag{8.45}$$

is commutative.[13] It follows from the unitarity of the wave operators that

The scattering operator $\mathsf{S} : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ *is unitary.*

Summarizing, we get the following.

- Suppose that we are given the scattering state $\psi_0 \in B^\perp$ at the initial time $t = 0$.
- The dynamics $\psi(t) = \mathrm{e}^{-\mathrm{i}tH}\psi_0$ of the quantum particle under the action of the potential U behaves like the free dynamics

$$\psi_-(t) = \mathrm{e}^{-\mathrm{i}tH_{\text{free}}}\varphi_{\text{in}} \quad \text{as} \quad t \to -\infty.$$

Here, $\varphi_{\text{in}} = W_{\text{in}}^{-1}\psi(0)$.

[12] Note that $||\mathrm{e}^{-\mathrm{i}tH}\psi_0 - \mathrm{e}^{-\mathrm{i}tH_{\text{free}}}\varphi_{\text{out}}||$ is equal to

$$||\mathrm{e}^{-\mathrm{i}tH}(\psi_0 - \mathrm{e}^{\mathrm{i}tH}\mathrm{e}^{-\mathrm{i}tH_{\text{free}}}\varphi_{\text{out}})|| = ||\psi_0 - \mathrm{e}^{\mathrm{i}tH}\mathrm{e}^{-\mathrm{i}tH_{\text{free}}}\varphi_{\text{out}}||,$$

by the unitarity of the operator $\mathrm{e}^{\mathrm{i}tH}$ for all $t \in \mathbb{R}$. This remains true if we replace φ_{out} by φ_{in}.

[13] In order to prove (8.44), note that $\varphi_{\text{out}} = \mathsf{S}\varphi_{\text{in}}$ is equivalent to

$$W_{\text{out}}^{-1}\psi_0 = \mathsf{S}W_{\text{in}}^{-1}\psi_0 \qquad \text{for all} \quad \psi_0 \in B^\perp.$$

Hence $W_{\text{out}}^{-1} = \mathsf{S}W_{\text{in}}^{-1}$. This implies $W_{\text{out}}^{-1}W_{\text{in}} = \mathsf{S}$.

- The dynamics $\psi(t) = \mathrm{e}^{-\mathrm{i}tH}\psi_0$ of the quantum particle under the action of the potential U behaves like the free dynamics

$$\psi_+(t) = \mathrm{e}^{-\mathrm{i}tH_{\mathrm{free}}}\varphi_{\mathrm{out}} \quad \text{as } t \to +\infty.$$

Here, $\varphi_{\mathrm{out}} = \mathsf{S}\varphi_{\mathrm{in}}$.

(vi) The relation between wave operators and the motion of wave packets. The harmonic wave function $x \mapsto \mathrm{e}^{\mathrm{i}kx}$ does not lie in the Hilbert space $L_2(\mathbb{R})$.

In order to improve the situation, the trick is to construct wave packets which lie in $L_2(\mathbb{R})$.

More precisely, by a free wave packet we understand a function of the form

$$\varphi_{\mathrm{free}}(x) := \int_{-\infty}^{\infty} A(k)\mathrm{e}^{\mathrm{i}kx}dk \qquad \text{for all} \quad x \in \mathbb{R}$$

where the smooth amplitude function $A : \mathbb{R} \to \mathbb{C}$ has compact support. In terms of physics, this is a superposition of harmonic waves. Since $A \in \mathcal{S}(\mathbb{R})$, the theory of the Fourier transform tells us that $\varphi_{\mathrm{free}} \in \mathcal{S}(\mathbb{R})$. Hence $\varphi_{\mathrm{free}} \in L_2(\mathbb{R})$. Moreover, for each wave number $k \in \mathbb{R}$, we have

$$H_{\mathrm{free}}\mathrm{e}^{\mathrm{i}kx} = E(k)\mathrm{e}^{\mathrm{i}kx}, \qquad x \in \mathbb{R}$$

with the energy $E(k) = \frac{k^2}{2}$.
Now switch on the potential U. For each wave number $k \in \mathbb{R} \setminus \{0\}$, let us pass to the function φ_k with

$$H\varphi_k(x) = E(k)\varphi_k(x), \qquad x \in \mathbb{R}.$$

By a wave packet induced by the potential U, we understand a function of the form

$$\varphi_U(x) := \int_{-\infty}^{\infty} A(k)\varphi_k(x)dk \qquad \text{for all} \quad x \in \mathbb{R}.$$

That is, we replace the harmonic wave by the solution φ_k of the stationary Schrödinger equation (8.37) on page 717. The proofs of the following statements can be found in Berezin and Shubin (1991).
It turns out that the wave operator W_{out} sends the free wave packet φ_{free} to the wave packet φ_U. Explicitly, for all amplitude functions $A \in \mathcal{D}(\mathbb{R})$, we get

$$\boxed{W_{\mathrm{out}}\left(\int_{-\infty}^{\infty} A(k)\mathrm{e}^{\mathrm{i}kx}dk\right) = \int_{-\infty}^{\infty} A(k)\varphi_k(x)dk.}$$

Similarly, for the wave operator W_{in}, we get

$$\boxed{W_{\mathrm{in}}\left(\int_{-\infty}^{\infty} A(k)\mathrm{e}^{-\mathrm{i}kx}dk\right) = \int_{-\infty}^{\infty} A(k)\varphi_k(x)^{\dagger}dk.}$$

This yields the scattering operator $\mathsf{S} = W_{\mathrm{out}}^{-1}W_{\mathrm{in}}$.

(vii) The relation between the scattering operator S and the S-matrix $\hat{S}$. Fix the wave number $k > 0$. Recall that the functions φ_k and φ_{-k} have the following asymptotical behavior

$$\varphi_k(x) := \begin{cases} \mathrm{e}^{\mathrm{i}kx} + \varrho_{\leftarrow}(k)\mathrm{e}^{-\mathrm{i}kx} & \text{if} \quad x < -r, \\ \tau_{\rightarrow}(k)\mathrm{e}^{\mathrm{i}kx} & \text{if} \quad x > r \end{cases}$$

and

$$\varphi_{-k}(x) := \begin{cases} \mathrm{e}^{-\mathrm{i}kx} + \varrho_{\rightarrow}(k)\mathrm{e}^{\mathrm{i}kx} & \text{if} \quad x > r, \\ \tau_{\leftarrow}(k)\mathrm{e}^{-\mathrm{i}kx} & \text{if} \quad x < -r. \end{cases}$$

For the unitary scattering operator $\mathsf{S} : L_2(\mathbb{R}) \to L_2(\mathbb{R})$, set

$$\boxed{\chi := \mathsf{S}\varphi \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}).}$$

If $\hat{\varphi}$ (resp. $\hat{\chi}$) denotes the Fourier transform of φ (resp. χ), then

$$\boxed{\begin{pmatrix} \hat{\chi}(k) \\ \hat{\chi}(-k) \end{pmatrix} = \hat{S}(k) \begin{pmatrix} \hat{\varphi}(k) \\ \hat{\varphi}(-k) \end{pmatrix} \qquad \text{for all} \quad k > 0}$$

with the S-matrix

$$\hat{S}(k) := \begin{pmatrix} \tau_{\leftarrow}(k) & \varrho_{\rightarrow}(k) \\ \varrho_{\leftarrow}(k) & \tau_{\rightarrow}(k) \end{pmatrix}.$$

The explicit form of the S-matrix elements can be found in Prop. 8.1 on page 707. It turns out that, for all $A \in \mathcal{D}(\mathbb{R})$, the scattering operator S sends the free wave packet

$$\int_0^\infty A(k)\mathrm{e}^{\mathrm{i}kx} + A(-k)\mathrm{e}^{-\mathrm{i}kx} \; dk$$

to the free wave packet

$$\int_0^\infty B(k)\mathrm{e}^{\mathrm{i}kx} + B(-k)\mathrm{e}^{-\mathrm{i}kx} \; dk$$

with the transformed amplitudes

$$\boxed{\begin{pmatrix} B(k) \\ B(-k) \end{pmatrix} := \hat{S}(k) \begin{pmatrix} A(k) \\ A(-k) \end{pmatrix}.}$$

In Problem 9.7, we will prove that for all wave numbers $k > 0$, we have

$$\boxed{\begin{pmatrix} \varphi_{-k}(x) \\ \varphi_k(x) \end{pmatrix} = \hat{S}(k) \begin{pmatrix} \varphi_k(x)^\dagger \\ \varphi_{-k}(x)^\dagger \end{pmatrix}, \qquad x \in \mathbb{R}.} \tag{8.46}$$

Finally, let us introduce the global S-matrix $\hat{S}$ by setting

$$\boxed{\hat{S} := \{\hat{S}(k)\}_{k>0}.} \tag{8.47}$$

Here, $\hat{S}(k)$ is called the k-component of the global S-matrix with respect to the wave number k.

The formulas above show that one has to distinguish between the scattering operator S and the scattering matrix $\hat{S}$. However, for historical reasons and by abuse of language, the two notions 'scattering operator' and 'S-matrix' are frequently used in a synonymous manner.

8.3.10 The Feynman Propagator

It is our goal to describe the quantum dynamics in terms of the Fourier–Stieltjes transformation. Recall that $Y := \mathbb{C}^n \times L_2(\mathbb{R})$. The key formulas read as follows:

- $\mathcal{U}\varphi := (c_j, c(k))$ and $\mathcal{U}\chi := (d_j, d(k))$ for all $\varphi, \chi \in L_2(\mathbb{R})$.
- $\hat{H}(c_j, c(k)) := (E_j c_j, E(k)c(k))$ on $D(\hat{H})$.
- For all times $t \in \mathbb{R}$ and all $(c_j, c(k)) \in Y$, we have

$$\mathrm{e}^{-\mathrm{i}t\hat{H}}(c_j, c(k)) := (\mathrm{e}^{-\mathrm{i}tE_j}c_j, \mathrm{e}^{-\mathrm{i}tE(k)}c(k)).$$

More precisely, $(c_j, c(k))$ stands for $(c_j, c(k))_{j=1,\ldots,n,k\in\mathbb{R}}$, and so on. By definition, the set $D(\hat{H})$ consists of all $(c_j, c(k)) \in Y$ with the property

$$\sum_{j=1}^{n} E_j^2 |c_j|^2 + \int_{-\infty}^{\infty} E(k)^2 |c(k)|^2 dk < \infty$$

where $E(k) := \frac{k^2}{2}$. We also set $D(H) := \mathcal{U}^{-1}D(\hat{H})$. Then the following hold true:

(i) The operator $\hat{H} : D(\hat{H}) \to Y$ is self-adjoint on the Hilbert space Y, and it is the Fourier–Stieltjes transform of the self-adjoint Hamiltonian $H : D(H) \to L_2(\mathbb{R})$ on the Hilbert space $L_2(\mathbb{R})$. This means that the following diagram

$$\begin{array}{ccc} L_2(\mathbb{R}) & \xrightarrow{H} & L_2(\mathbb{R}) \\ {\scriptstyle\mathcal{U}}\downarrow & & \downarrow{\scriptstyle\mathcal{U}} \\ Y & \xrightarrow{\hat{H}} & Y \end{array} \tag{8.48}$$

is commutative. In other words, $\mathcal{U}H = \hat{H}\mathcal{U}$ on $D(H)$. Equivalenty, we have $H = \mathcal{U}^{-1}\hat{H}\mathcal{U}$ on $D(H)$.

(ii) For each time $t \in \mathbb{R}$, the unitary operator $\mathrm{e}^{-\mathrm{i}t\hat{H}} : Y \to Y$ is the Fourier–Stieltjes transform of the unitary operator $\mathrm{e}^{-\mathrm{i}tH} : L_2(\mathbb{R}) \to L_2(\mathbb{R})$. Explicitly, $\mathrm{e}^{-\mathrm{i}tH} = \mathcal{U}^{-1}\mathrm{e}^{-\mathrm{i}t\hat{H}}\mathcal{U}$.

The propagator. Set $P(t) := \mathrm{e}^{-\mathrm{i}tH}$ for all $t \in \mathbb{R}$. The operator $P(t)$ is called the propagator of the Hamiltonian H at time t. Since the operator H is self-adjoint, the Stone theorem tells us that the operator

$$P(t) : L_2(\mathbb{R}) \to L_2(\mathbb{R})$$

is a well-defined unitary operator. For all given functions $\varphi, \chi \in \mathcal{S}(\mathbb{R})$ and each time $t \in \mathbb{R}$, we have

$$\boxed{\langle\varphi|P(t)\chi\rangle_2 = \sum_{j=1}^{n} \mathrm{e}^{-\mathrm{i}tE_j} c_j^\dagger d_j + \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}tE(k)} c(k)^\dagger d(k) dk.} \tag{8.49}$$

Using the Heaviside function, define

$$P^+(t) := \theta(t)P(t) \qquad \text{for all} \quad t \in \mathbb{R}$$

and $P^-(t) := -\theta(-t)P(t)$.[14] This implies the splitting

[14] Recall that $\theta(t) := 1$ if $t \geq 0$ and $\theta(t) := 0$ if $t < 0$. .

$$P(t) = P^+(t) - P^-(t) \qquad \text{for all} \quad t \in \mathbb{R}$$

of the propagator into the retarded propagator P^+ and the advanced propagator P^-. In the literature, the retarded propagator $P^+(t)$ is also called the Feynman propagator at time t.

Formal approach. Physicists formally write

$$\boxed{\langle \varphi | P(t)\chi \rangle_2 = \int_{\mathbb{R}^2} \varphi(x)^\dagger \mathcal{P}(x,y;t)\chi(y)dxdy} \tag{8.50}$$

with the propagator kernel

$$\mathcal{P}(x,y;t) := \sum_{j=1}^{n} \mathrm{e}^{-\mathrm{i}tE_j} \chi_j(x)\chi_j(y)^\dagger + \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}tE(k)} \varphi_k(x)\varphi_k(y)^\dagger dk.$$

Here, the functions χ_j and φ_k are taken from (8.39) on page 719. In physics textbooks, formula (8.50) is formally motivated in the following way. The integral

$$\int_{\mathbb{R}^2} \varphi(x)^\dagger \mathcal{P}(x,y;t)\chi(y)dxdy$$

is equal to

$$\sum_{j=1}^{n} \mathrm{e}^{-\mathrm{i}tE_j} \int_{\mathbb{R}} \varphi(x)^\dagger \chi_j(x)dx \int_{\mathbb{R}} \chi_j(y)^\dagger \chi(y)dy$$
$$+\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}tE(k)} \left(\int_{\mathbb{R}} \varphi^\dagger(x)\varphi_k(x)dx \int_{\mathbb{R}} \varphi_k(y)^\dagger \chi(y)dy \right) dk.$$

This coincides with

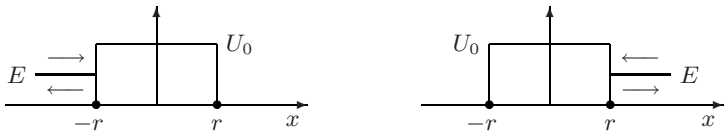
$$\sum_{j=1}^{n} \mathrm{e}^{-\mathrm{i}tE_j} c_j^\dagger d_j + \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}tE(k)} c(k)^\dagger d(k)dk.$$

In turn, this is equal to $\langle \varphi | P(t)\chi \rangle_2$.

The kernel strategy of physicists. The mathematical approach to quantum physics is based on the language of self-adjoint operators. In order to get more information about the operators, it is useful to pass to the representation of operators by kernels which depend on space and time. The kernels are closely related to Green's functions. In mathematics, the operator approach was introduced by von Neumann in the late 1920s. In physics, the formal kernel approach was invented by Dirac in the late 1920s and further developed by Feynman in the early 1940s (representation of the kernels by path integrals). In the late 1940s, Laurent Schwartz proved his kernel theorem which gives the kernel approach of physicists a rigorous basis (see Vol. I, Sec. 12.2.7). It was shown by Grothendieck in the 1950s that the kernel approach is closely related to the theory of nuclear spaces.

8.4 Tunnelling of Quantum Particles and Radioactive Decay

Radioactive decay. It was discovered in about 1900 that there exist atoms which decay. This is the phenomenon of radioactive decay under the action of the weak force. We distinguish between



(a) (b)

Fig. 8.8. Classical motion on the real line

(i) α-decay (emission of α-particles consisting of two protons and two neutrons) and
(ii) β-decay (emission of electrons).

In terms of classical mechanics, the α-particles cannot leave the atom because of a high potential barrier. However, in contrast to the classical situation, the α-particles can leave the atom by means of the tunnelling effect in quantum mechanics. Let us discuss this by considering a one-dimensional model.

Classical motion. We consider the potential barrier

$$U(x) := \begin{cases} U_0 & \text{if} \quad -r \le x \le r, \\ 0 & \text{otherwise} \end{cases} \tag{8.51}$$

where $U_0 > 0$. For a classical particle of energy E, we have $E \ge U(x)$ for all positions x of the particle. Let $0 \le E < U_0$. Then the particle is able to move on the interval $]r, \infty[$ or on $]-\infty, -r[$ with the constant velocity v where $E = \frac{v^2}{2}$. If the particle reaches the potential barrier at the point $x = r$ or $x = -r$, then it is reflected (Fig. 8.8).

Motion of a quantum particle. In contrast to the classical motion, a quantum particle is able to pass the potential barrier (Fig. 8.9). This quantum effect is called tunnelling. Let us study this. For the given energy E of the quantum particle, let us distinguish the following two cases.

Case 1: $E > U_0$. We set $k := \sqrt{2E}$ and $K := \sqrt{2(E - U_0)}$. Moreover, for the state function, we make the ansatz

$$\varphi_k(x) = \begin{cases} \mathrm{e}^{\mathrm{i}kx} + \varrho_{\leftarrow}(k)\mathrm{e}^{-\mathrm{i}kx} & \text{if} \quad x < -r, \\ A\mathrm{e}^{\mathrm{i}Kx} + B\mathrm{e}^{-\mathrm{i}Kx} & \text{if} \quad -r \le x \le r, \\ \tau_{\rightarrow}(k)\mathrm{e}^{\mathrm{i}kx} & \text{if} \quad x > r. \end{cases} \tag{8.52}$$

From Prop. 8.1 on page 707 we obtain the following formulas for the coefficients

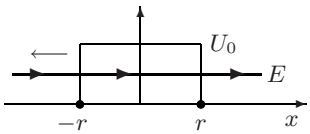


Fig. 8.9. Tunnelling of a quantum particle on the real line

$$\varrho_{\leftarrow}(k) = \frac{\frac{\mathrm{i}}{2}\left(\frac{K}{k} - \frac{k}{K}\right)\mathrm{e}^{-2\mathrm{i}kr}\sin^2 2Kr}{\cos 2Kr - \frac{\mathrm{i}}{2}\left(\frac{K}{k} + \frac{k}{K}\right)\sin 2Kr},$$
$$\tau_{\rightarrow}(k) = \frac{\mathrm{e}^{-2\mathrm{i}kr}}{\cos 2Kr - \frac{\mathrm{i}}{2}\left(\frac{K}{k} + \frac{k}{K}\right)\sin 2Kr}. \tag{8.53}$$

Case 2: $0 < E < U_0$. We now set

$$\varphi_k(x) = \begin{cases} \mathrm{e}^{\mathrm{i}kx} + \varrho_{\leftarrow}(k)\mathrm{e}^{-\mathrm{i}kx} & \text{if} \quad x < -r, \\ A\mathrm{e}^{-\mathcal{K}x} + B\mathrm{e}^{\mathcal{K}x} & \text{if} \quad -r \le x \le r, \\ \tau_{\rightarrow}(k)\mathrm{e}^{\mathrm{i}kx} & \text{if} \quad x > r. \end{cases}$$

Recall that $\sin \mathrm{i}z = \mathrm{i}\sinh z$ and $\cos \mathrm{i}z = \cosh z$ for all $z \in \mathbb{C}$. Letting $k = \sqrt{2E}$ and $K = \mathrm{i}\mathcal{K}$ with $\mathcal{K} := \sqrt{2(U_0 - E)}$, it follows from (8.53) that

$$\varrho_{\leftarrow}(k) = \frac{\left(\frac{\mathcal{K}}{k} + \frac{k}{\mathcal{K}}\right)\mathrm{e}^{-2\mathrm{i}kr}\sinh^2 2\mathcal{K}}{\cosh 2\mathcal{K}r + \frac{\mathrm{i}}{2}\left(\frac{\mathcal{K}}{k} - \frac{k}{\mathcal{K}}\right)\sinh 2\mathcal{K}r},$$
$$\tau_{\rightarrow}(k) = \frac{\mathrm{e}^{-2\mathrm{i}kr}}{\cosh 2\mathcal{K}r + \frac{\mathrm{i}}{2}\left(\frac{\mathcal{K}}{k} - \frac{k}{\mathcal{K}}\right)\sinh 2\mathcal{K}r}. \tag{8.54}$$

This is an analytic continuation of the solution (8.53) of the stationary Schrödinger equation (8.37), and hence it is also a solution of (8.37).

8.5 The Method of the Green's Function in a Nutshell

> The universal method of the Green's function is based on integration by parts, which corresponds to the fundamental theorem of calculus due to Newton and Leibniz in one dimension and to the Gauss integral theorem in higher dimensions.[15] This method allows us the passage from differential equations to integral equations.
> In particular, the Schrödinger differential equation from the year 1926 passes over to the Lippmann–Schwinger integral equation from the year 1950. In contrast to the Schrödinger differential equation, the Lippmann–Schwinger integral equation automatically yields the right asymptotic behavior of the wave function for scattering processes and allows us to compute approximations, like the Born approximation in lowest order. This is related to Picard's method of successive iterations and Banach's fixed point theorem.
> The two approaches of Feynman and Schwinger to quantum field theory from the 1940s are based on Green's method combined with the Dirac delta function. The modern version of Green's method uses the theory of distributions created by Laurent Schwartz in the 1940s.[16]
>
> Folklore

Let us now replace the square-well potential and the potential barrier by a more general potential U. To simplify the considerations, we assume that the following hold true.

[15] Newton (1743–1727), Leibniz (1746–1716), Gauss (1777–1855), George Green (1793–1843), Picard (1856–1941), Banach (1892–1945), Dirac (1902–1984), Feynman (1918–1988), Schwinger (1918–1994), Laurent Schwartz (1915–2004).

[16] For the theory of distributions, we refer to Chap. 11 of Vol. I.

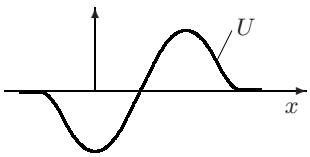


Fig. 8.10. Potential of a quantum particle on the real line

(A) The function $U : \mathbb{R} \to \mathbb{R}$ is smooth and has compact support, that is, there exists a number $r > 0$ such that $U(x) = 0$ if $|x| > r$ (Fig. 8.10).

The corresponding Schrödinger equation reads as

$$\mathrm{i}\hbar\psi_t(x,t) = -\frac{\hbar^2}{2m}\psi_{xx}(x,t) + \kappa U(x)\psi(x,t), \qquad x,t \in \mathbb{R} \tag{8.55}$$

where $\kappa \geq 0$ is called the coupling constant. The time-dependent wave function $\psi = \psi(x,t)$ describes the motion of a quantum particle on the real line under the action of the potential U.

8.5.1 The Inhomogeneous Helmholtz Equation

Fix the wave number $k > 0$, and set $E(k) := \frac{\hbar k^2}{2m}$. According to the classical Fourier method, we make the ansatz

$$\psi(x,t) := \mathrm{e}^{-\mathrm{i}E(k)t/\hbar}\varphi(x), \qquad x,t \in \mathbb{R}.$$

To simplify notation, we set $\hbar := 1$ and $m := \frac{1}{2}$ (mass of the particle). This way, we obtain the stationary Schrödinger equation

$$\boxed{-\frac{d^2\varphi(x)}{dx^2} - k^2\varphi(x) = -\kappa U(x)\varphi(x), \qquad x \in \mathbb{R}.} \tag{8.56}$$

Our goal is to reduce the stationary Schrödinger equation to the following inhomogeneous Helmholtz equation[17]

$$\boxed{-\frac{d^2\varphi(x)}{dx^2} - k^2\varphi(x) = f(x), \qquad x \in \mathbb{R}.} \tag{8.57}$$

Again the parameter $k > 0$ is fixed. We are given the smooth function

$$f : \mathbb{R} \to \mathbb{C}$$

with compact support, that is, $f \in \mathcal{D}(\mathbb{R})$. We are looking for the most general smooth solution $\varphi : \mathbb{R} \to \mathbb{C}$ of (8.57).

In what follows, we will explain the relation between the formal language of physicists and the rigorous language of mathematicians.

[17] Helmholtz (1821–1894).

Note that the formal language used by physicists is distinguished by mnemonic elegance. For a mathematician it is very useful to know both the formal and the rigorous language.

Our plan. In the next section, we will introduce the retarded fundamental solution F_+ and the retarded Green's function $\mathcal{G}^+(x,y) = F_+(x-y)$, and we will use this function in order to solve the Helmholtz equation (8.57).

In Section 8.6, we will use the retarded Green's function in order to replace the stationary Schrödinger equation (8.56) by the following Lippmann–Schwinger integral equation

$$\boxed{\varphi(x) = \mathrm{e}^{\mathrm{i}kx} - \kappa \int_{\mathbb{R}} \mathcal{G}^+(x,y)U(y)\varphi(y)dy, \qquad x \in \mathbb{R}} \tag{8.58}$$

for the unknown function φ. Setting $\varphi = 0$ on the right-hand side of (8.58), we get the zeroth approximation $\varphi_0(x) := \mathrm{e}^{\mathrm{i}kx}$. Replacing φ by φ_0 on the right-hand side of (8.58), we obtain the first approximation

$$\varphi_1(x) := \varphi_0(x) - \kappa \int_{\mathbb{R}} \mathcal{G}^+(x,y)U(y)\varphi_0(y)dy$$

which is called the Born approximation by physicists.[18] This is a good approximation if the coupling constant κ is sufficiently small.

The solutions of the stationary Schrödinger equation (8.56) may possess different asymptotics at infinity, $x \to \pm\infty$.

> *The advantage of the Lippmann–Schwinger equation is that the solutions of this integral equation (8.58) are solutions of the Schrödinger equation which possess the right asymptotic behavior with respect to a scattering process from left to right.*

This is guaranteed by the appropriate choice of the retarded Green's function $\mathcal{G}^+$.

8.5.2 The Retarded Green's Function, and the Existence and Uniqueness Theorem

In order to simplify the notation, let us introduce the Helmholtz operator

$$L := -\frac{d^2}{dx^2} - k^2$$

for fixed $k > 0$. Thus, the Helmholtz equation (8.57) reads as $L\varphi = f$.

Retarded fundamental solution. To begin with, let us study the equation $L\varphi = f$ in the special case where $f(x) := \delta(x)$. This means that we want to solve the special inhomogeneous Helmholtz equation

$$\boxed{L\varphi(x) = \delta(x), \qquad x \in \mathbb{R},} \tag{8.59}$$

in the sense of tempered distributions on the real line. To this end, set

$$\boxed{F_+(x) := \frac{\mathrm{i}\mathrm{e}^{\mathrm{i}|x|k}}{2k} \qquad \text{for all} \quad x \in \mathbb{R}.}$$

[18] M. Born, On collision processes in quantum theory, Z. Physik **37** (1926), 863-867 (in German).

The corresponding tempered distribution reads as $\mathcal{F}_+(\chi) := \int_{\mathbb{R}} F_+(x)\chi(x)dx$ for all test functions $\chi \in \mathcal{S}(\mathbb{R})$. The function F_+ satisfies the following asymptotic condition

$$\lim_{x\to\pm\infty}(F_+'(x) \mp \mathrm{i}kF_+(x)) = 0.$$

This is a special case of the Sommerfeld radiation condition[19] to be discussed below on page 734.

Proposition 8.8 *For fixed $k > 0$, the function F_+ is a solution of equation (8.59).*

The function F_+ is called the retarded fundamental solution of the Helmholtz operator L.

Proof. We have to show that

$$\int_{\mathbb{R}} F_+(x)L\chi(x)dx = \chi(0) \qquad \text{for all} \quad \chi \in \mathcal{S}(\mathbb{R}).$$

To this end, we will use integration by parts, that is,

$$\int_a^b u'vdx = -\int_a^b uv'dx + uv|_a^b$$

for smooth functions $u, v : [a, b] \to \mathbb{C}$ where $-\infty < a < b < \infty$.

(I) Green's key formula. Let $\chi \in \mathcal{D}(\mathbb{R})$. Fix $\eta > 0$, and set $U_\eta(0) :=]-\eta, \eta[$. Integration by parts yields Green's key formula

$$\int_{\mathbb{R}\setminus U_\eta(0)} (F_+L\chi - \chi LF_+)dx = (F_+\chi' - \chi F_+')|_{-\eta}^{\eta}. \tag{8.60}$$

Note that the smooth function χ vanishes outside some bounded interval. Thus, the additional boundary terms vanish.

(II) The limit $\eta \to +0$. For the classical derivative of F_+ outside the origin, we get

$$F_+'(x) = \begin{cases} -\frac{1}{2}\mathrm{e}^{\mathrm{i}xk} & \text{if} \quad x > 0, \\ \frac{1}{2}\mathrm{e}^{-\mathrm{i}xk} & \text{if} \quad x < 0. \end{cases}$$

Consequently, the function F_+ has the following three crucial properties:

(i) Solution of the differential equation outside the origin: $LF_+(x) = 0$ for all $x \in \mathbb{R} \setminus \{0\}$.
(ii) Continuity: The function $F_+ : \mathbb{R} \to \mathbb{C}$ is continuous.
(iii) Jump of the first derivative at the origin: $F_+'(-0) - F_+'(+0) = 1$.[20]

Applying this to the key formula (8.60), we get

$$\int_{\mathbb{R}\setminus U_\eta(0)} F_+L\chi dx = (F_+\chi' - \chi F_+'|_{-\eta}^{\eta}.$$

Letting $\eta \to +0$, we obtain

$$\int_{\mathbb{R}} F_+L\chi dx = \chi(0)(F_+'(-0) - F_+'(+0)) = \chi(0).$$

[19] Sommerfeld (1868–1951).
[20] As usually, we write $F(x_0 \pm 0) := \lim_{\varepsilon\to+0} F(x_0 \pm \varepsilon)$.

This is the desired relation.

The same argument applies to each test function $\chi \in \mathcal{S}(\mathbb{R})$. In fact, the function χ vanishes rapidly as $x \to \pm\infty$. Thus, applying integration by parts to the large interval $[-K, K]$, we obtain boundary terms at the points $\pm K$ which vanish as $K \to +\infty$. □

Formal approach to the Helmholtz equation via the Dirac delta function. We now want to use the fundamental solution F_+ in order to construct a special solution of the inhomogeneous Helmholtz equation (8.57). We will use a general formal method which is always applied by physicists to linear, inhomogeneous differential equations. The idea is to elegantly use the convolution

$$\varphi_{\text{special}} = F_+ * f.$$

Explicitly, we obtain

$$\boxed{\varphi_{\text{special}}(x) = \int_{\mathbb{R}} F_+(x-y) f(y)\,dy \qquad \text{for all} \quad x \in \mathbb{R}.} \tag{8.61}$$

In fact, formally applying the relation $LF_+(x) = \delta(x)$, we get

$$L\varphi_{\text{special}}(x) = \int_{\mathbb{R}} L_x F_+(x-y) f(y)\,dy = \int_{\mathbb{R}} \delta(x-y) f(y)\,dy = f(x).$$

That is, the function φ_{special} is a special solution of the Helmholtz equation $L\varphi = f$. The general solution of the homogeneous equation $L\psi = 0$ reads as $\psi(x) := a_+ \mathrm{e}^{\mathrm{i}kx} + a_- \mathrm{e}^{-\mathrm{i}kx}$ where $a_\pm$ are arbitrary complex numbers. Consequently, the general solution of the Helmholtz equation $L\varphi = f$ is given by

$$\varphi(x) = a_+ \mathrm{e}^{\mathrm{i}kx} + a_- \mathrm{e}^{-\mathrm{i}kx} + \varphi_{\text{special}}(x).$$

Rigorous approach. We now want to show how the preceding elegant, but formal argument can be turned into a rigorous proof. To this end, we have to carefully compute the operator $L\varphi$ by using the limit from the difference quotient to the differential quotient. The point is that the first derivative of F_+ jumps at the origin. Naively, one expects that

$$L_x \int_{\mathbb{R}} F_+(x-y) f(y)\,dy = \int_{\mathbb{R}} L_x F_+(x-y) f(y)\,dy.$$

However, observe that this formula is only valid if the derivatives of the function $x \mapsto F_+(x-y)$ behave regularly (see Problem 9.2). The discontinuities of LF_+ are responsible for the modified formula

$$L_x \int_{\mathbb{R}} F_+(x-y) f(y)\,dy = \int_{\mathbb{R}} L_x F_+(x-y) f(y)\,dy + C(x)$$

with an additional non-vanishing term $C(x)$. Let us define the retarded Green's function $\mathcal{G}^+(x,y) := F_+(x-y)$. Explicitly,

$$\mathcal{G}^+(x,y) = \frac{\mathrm{i}\mathrm{e}^{\mathrm{i}k|x-y|}}{2k} \qquad \text{for all} \quad x, y \in \mathbb{R}.$$

Theorem 8.9 *Fix $k > 0$. We are given the smooth function $f : \mathbb{R} \to \mathbb{C}$ with compact support. Then the following hold.*

(i) *General solution: The smooth solutions of the inhomogeneous Helmholtz equation (8.57) are given by*

$$\varphi(x) = a_+ e^{\mathrm{i}kx} + a_- e^{-\mathrm{i}kx} + \int_{\mathbb{R}} \mathcal{G}^+(x,y) f(y) dy, \quad x \in \mathbb{R} \tag{8.62}$$

with the retarded Green's function $\mathcal{G}^+$ and arbitrary complex numbers a_+ and a_-.

(ii) *Unique solution: If we add the Sommerfeld radiation condition for outgoing waves*

$$\frac{d\varphi(x)}{dx} = \pm \mathrm{i}k\varphi(x) + o(1), \qquad x \to \pm\infty,$$

then the inhomogeneous Helmholtz equation (8.57) has the unique smooth solution (8.62) with $a_+ = a_- = 0$.

Proof. Ad (i). It remains to show that the function

$$\varphi_{\text{special}}(x) := \int_{-\infty}^{\infty} \mathcal{G}^+(x,y) f(y) dy, \qquad x \in \mathbb{R}$$

is a special solution of $L\varphi = f$.

(I) By the proof of Prop. 8.8, we know that

$$\int_{\mathbb{R}} F_+(y) L\chi(y) dy = \chi(0) \qquad \text{for all} \quad \chi \in \mathcal{D}(\mathbb{R}).$$

Replacing the origin $x = 0$ by the point $x \in \mathbb{R}$, the same argument tells us that

$$\int_{\mathbb{R}} F_+(y-x) L\chi(y) dy = \chi(x) \qquad \text{for all} \quad \chi \in \mathcal{D}(\mathbb{R}).$$

Because of $F_+(x-y) = F_+(y-x)$ and $f \in \mathcal{D}(\mathbb{R})$, we get

$$\int_{\mathbb{R}} F_+(x-y) Lf(y) dy = f(x).$$

This is true for all $x \in \mathbb{R}$.

(II) Thus, it remains to show that

$$L\varphi_{\text{special}}(x) = \int_{\mathbb{R}} F_+(x-y) Lf(y) dy.$$

To prove this, fix $x \in \mathbb{R}$.

(III) First let us prove that

$$\varphi'_{\text{special}}(x) = \int_{\mathbb{R}} F_+(x-y) f'(y) dy. \tag{8.63}$$

In other words, we have to show that

$$\lim_{h \to 0} \frac{\varphi_{\text{special}}(x+h) - \varphi_{\text{special}}(x)}{h} = \int_{\mathbb{R}} F_+(x-y) f'(y) dy.$$

To this end, we set

$$\frac{\varphi_{\text{special}}(x+h)-\varphi_{\text{special}}(x)}{h} = A + B$$

along with

$$A := \int_{]x-\eta,x+\eta[} \frac{F_+(x+h-y)-F_+(x-y)}{h}\, f(y)dy,$$
$$B := \int_{\mathbb{R}\setminus]x-\eta,x+\eta[} \frac{F_+(x+h-y)-F_+(x-y)}{h}\, f(y)dy$$

where $\eta > 0$.

(III-1) Inspection of A. We are given $\varepsilon > 0$. We want to show that there exist numbers $h_0 > 0$ and $\eta_0 > 0$ such that

$$|A| < \varepsilon \qquad \text{for all} \quad h \in]-h_0, h_0[\ \text{ and } \ \eta \in]-\eta_0, \eta_0[. \tag{8.64}$$

This follows from the triangle inequality $|\,|x+h-y|-|x-y|\,| \le |h|$ and the Taylor expansion

$$\mathrm{e}^{\mathrm{i}k|x+h-y|} - \mathrm{e}^{\mathrm{i}k|x-y|} = \mathrm{i}k(|x+h-y|-|x-y|) + \text{remainder}.$$

Hence

$$\left|\frac{\mathrm{e}^{\mathrm{i}k|x+h-y|} - \mathrm{e}^{\mathrm{i}k|x-y|}}{h}\right| \le k + \dots$$

The dots denote terms which go to zero if $y \to x$ and $h \to 0$. Thus, we obtain $|A| \le 2\eta k + \dots$ This implies (8.64).

(III-2) Inspection of B. Since the point x lies outside the domain of integration $\mathbb{R}\setminus]x-\eta, x+\eta[$, we can use the classical rule for differentiating parameter integrals (see Problem 9.2). Therefore,

$$\lim_{h\to 0} B = \int_{\mathbb{R}\setminus]x-\eta,x+\eta[} \frac{\partial}{\partial x} F_+(x-y)\, f(y)dy.$$

Since $\frac{\partial}{\partial x}F_+(x-y) = -\frac{\partial}{\partial y}F_+(x-y)$, integration by parts yields

$$\lim_{h\to 0} B = \int_{\mathbb{R}\setminus]x-\eta,x+\eta[} F_+(x-y)\, f'(y)dy + C(\eta),$$

where $R(\eta) := F_+(-\eta)f(x-\eta) - F_+(\eta)f(x+\eta)$. Letting $\eta \to 0$ and noting that $\lim_{\eta\to+0} R(\eta) = 0$, we get the claim (8.63), up to some term whose modulus is less than ε. Since the number ε can be chosen arbitrarily small, we obtain the claim (8.63).

(IV) Repeating the argument from (III), we get

$$\varphi''_{\text{special}}(x) = \int_{\mathbb{R}} F_+(x-y) f''(y)dy.$$

This yields immediately the desired result (II).

Ad (ii). Observe that the function $\varphi(x) := \mathrm{e}^{\mathrm{i}kx}$ (resp. $\mathrm{e}^{-\mathrm{i}kx}$) violates the Sommerfeld radiation condition at $x = -\infty$ (resp. $x = +\infty$). However, the argument from the proof of Prop. 8.10 below shows that the function $x \mapsto \int_{\mathbb{R}} \mathcal{G}^+(x,y)f(y)dy$ satisfies the Sommerfeld radiation condition. □

Asymptotic behavior. Fix the parameter $k > 0$. For given function $f \in \mathcal{D}(\mathbb{R})$, set

$$\varphi(x) := \mathrm{e}^{\mathrm{i}kx} + \int_{\mathbb{R}} \mathcal{G}^+(x,y) f(y) dy \qquad \text{for all} \quad x \in \mathbb{R}.$$

The function φ satisfies the Helmholtz equation $L\varphi = f$ on $\mathbb{R}$, and it has the following asymptotic behavior.

Proposition 8.10 (i) $\varphi(x) \simeq e^{\mathrm{i}kx} + \varrho_+(k) e^{-\mathrm{i}kx}$ *as* $x \to -\infty$ *with the reflection coefficient*

$$\varrho_+(k) := \frac{\mathrm{i}}{2k} \int_{\mathbb{R}} f(y) e^{\mathrm{i}ky} dy.$$

(ii) $\varphi(x) \simeq \tau_+(k) e^{\mathrm{i}kx}$ *as* $x \to +\infty$ *with the transmission coefficient*

$$\tau_+(k) := 1 + \frac{\mathrm{i}}{2k} \int_{\mathbb{R}} f(y) e^{-iky} dy.$$

This means that the solution asymptotically behaves like

- an incoming wave of wave number $k > 0$ at $x = -\infty$, and
- an outgoing wave of wave number $k > 0$ at $x = +\infty$.

This wave is reflected under the influence of the external source f.

Proof. For each $x \in \mathbb{R}$, we have the following decomposition

$$\varphi(x) = \mathrm{e}^{\mathrm{i}kx} + \frac{\mathrm{i}\mathrm{e}^{\mathrm{i}kx}}{2k} \int_{-\infty}^{x} f(y) \mathrm{e}^{-\mathrm{i}ky} dy + \frac{\mathrm{i}\mathrm{e}^{-\mathrm{i}kx}}{2k} \int_{x}^{\infty} f(y) \mathrm{e}^{\mathrm{i}ky} dy.$$

Now consider the limits $x \to +\infty$ and $x \to -\infty$. □

8.5.3 The Advanced Green's Function

Let $k > 0$. If the function φ is a solution of the Helmholtz equation

$$\left(-\frac{d^2}{dx^2} - k^2 \right) \varphi(x) = f(x), \qquad x \in \mathbb{R},$$

then the substitution $k \mapsto -k$ generates a new solution of the Helmholtz equation. For example, the retarded fundamental solution $F_+(x) := \frac{\mathrm{i}\mathrm{e}^{\mathrm{i}k|x|}}{2k}$ passes over to F_- where

$$F_-(x) := -\frac{\mathrm{i}\mathrm{e}^{-\mathrm{i}k|x|}}{2k}, \qquad x \in \mathbb{R}$$

is called the advanced fundamental solution. We have

$$LF_\pm(x) = \delta(x), \qquad x \in \mathbb{R}.$$

In terms of physics,

- the retarded fundamental function F_+ describes an outgoing wave at the infinite points $x = -\infty$ and $x = +\infty$;
- the advanced fundamental solution F_- describes an incoming wave at the infinite points $x = -\infty$ and $x = +\infty$.

Suppose that we are given the function $f \in \mathcal{D}(\mathbb{R})$. Replacing the retarded Green's function $\mathcal{G}^+(x,y) := F_+(x-y)$ by the advanced Green's function

$$\mathcal{G}^-(x,y) := F_-(x-y) \qquad x,y \in \mathbb{R},$$

the most general smooth solution of the Helmholtz equation $L\varphi = f$ on $\mathbb{R}$ reads as

$$\varphi(x) = a_+ \mathrm{e}^{\mathrm{i}kx} + a_- \mathrm{e}^{-\mathrm{i}kx} + \int_{-\infty}^{\infty} \mathcal{G}^-(x,y) f(y) dy, \quad x \in \mathbb{R}$$

where $a_\pm$ are arbitrary complex numbers. If we add the Sommerfeld radiation condition for incoming waves

$$\frac{d\varphi(x)}{dx} = \mp \mathrm{i}k\varphi(x) + o(1), \qquad x \to \pm\infty,$$

then we obtain the unique smooth solution of the Helmholtz equation $L\varphi = f$ by setting $a_\pm = 0$. The retarded Green's function $\mathcal{G}^+$ and the advanced Green's function satisfy $\mathcal{G}^-$ satisfy the following equation

$$\left(-\frac{\partial^2}{\partial x^2} - k^2\right) \mathcal{G}^\pm(x,y) = \delta(x-y), \qquad x,y \in \mathbb{R}.$$

8.5.4 Perturbation of the Retarded and Advanced Green's Function

Fix $k > 0$. We know that the equation

$$-\left(\frac{d^2}{dx^2} + k^2\right) F(x) = \delta(x), \qquad x \in \mathbb{R} \tag{8.65}$$

has the two solutions

$$F_\pm(x) := \pm \frac{\mathrm{i}\mathrm{e}^{\pm \mathrm{i}|x|k}}{2k}, \qquad x \in \mathbb{R},$$

in the sense of tempered distributions. We want to obtain F_+ and F_- by using two different perturbation methods with respect to the small parameter $\varepsilon > 0$, and by carrying out the limit $\varepsilon \to +0$.

The perturbed retarded fundamental solution. Let us replace the original problem (8.65) by the perturbed problem

$$-\left(\frac{d^2}{dx^2} + k^2 + \varepsilon\mathrm{i}\right) F_{+,\varepsilon}(x) = \delta(x), \qquad x \in \mathbb{R} \tag{8.66}$$

for fixed parameter $\varepsilon > 0$.

Proposition 8.11 *In the sense of tempered distributions, equation* (8.66) *has the solution*

$$F_{+,\varepsilon}(x) := \frac{\mathrm{i}e^{\mathrm{i}|x|\sqrt{k^2+\varepsilon\mathrm{i}}}}{2\sqrt{k^2+\varepsilon\mathrm{i}}}, \qquad x \in \mathbb{R},$$

and we have the limit $\lim_{\varepsilon \to +0} F_{+,\varepsilon} = F_+$.

Here, we use the principal value of the square root. The perturbed retarded Green's function is defined by

$$\mathcal{G}_\varepsilon^+(x,y) := F_{+,\varepsilon}(x-y), \qquad x,y \in \mathbb{R}.$$

Proof. Replacing k by $\sqrt{k^2+\varepsilon \mathrm{i}}$ and using the same argument as in the proof of Prop. 8.8, we obtain $F_{+,\varepsilon}$. It remains to show that

$$\lim_{\varepsilon \to +0} \int_{\mathbb{R}} F_{+,\varepsilon}(x)\chi(x)dx = \int_{\mathbb{R}} F_+(x)\chi(x)dx \tag{8.67}$$

for all test functions $\chi \in \mathcal{S}(\mathbb{R})$. Because of

$$\sqrt{k^2+\varepsilon \mathrm{i}} = \sqrt{k^2\left(1+\frac{\varepsilon \mathrm{i}}{k^2}\right)} = k\left(1+\frac{\varepsilon \mathrm{i}}{2k^2}+O(\varepsilon^2)\right), \quad \varepsilon \to +0, \tag{8.68}$$

we obtain

$$F_{+,\varepsilon}(x) = \frac{\mathrm{i}\mathrm{e}^{\mathrm{i}k|x|}\mathrm{e}^{-\frac{1}{2}\varepsilon|x|/2k}}{2k} + O(\varepsilon), \qquad \varepsilon \to +0.$$

Note that the damping factor $\mathrm{e}^{-\frac{1}{2}\varepsilon|x|/2k}$ appears if $\varepsilon > 0$. This implies

$$\lim_{\varepsilon \to +0} F_{+,\varepsilon}(x) = F_+(x) \qquad \text{for all} \quad x \in \mathbb{R}.$$

Finally, we get (8.67). □

The perturbed advanced fundamental solution. Again fix $k > 0$ and choose the parameter $\varepsilon > 0$. Consider now the perturbed equation

$$-\left(\frac{d^2}{dx^2}+k^2-\varepsilon \mathrm{i}\right)F_{-,\varepsilon}(x) = \delta(x), \qquad x \in \mathbb{R} \tag{8.69}$$

which differs from (8.66) by changing the sign of the perturbation term.

Proposition 8.12 *In the sense of tempered distributions, equation* (8.69) *has the solution*

$$F_{-,\varepsilon}(x) := -\frac{\mathrm{i}e^{-\mathrm{i}|x|\sqrt{k^2-\varepsilon \mathrm{i}}}}{2\sqrt{k^2-\varepsilon \mathrm{i}}}, \qquad x \in \mathbb{R},$$

and we have the limit $\lim_{\varepsilon \to +0} F_{-,\varepsilon} = F_-$.

The proof proceeds as above. The perturbed advanced Green's function is defined by

$$\mathcal{G}_\varepsilon^-(x,y) := F_{-,\varepsilon}(x-y), \qquad x,y \in \mathbb{R}.$$

Damped time-depending waves and the principle of limiting absorption. Let $\varepsilon > 0$. We want to discuss the physics behind the perturbation method above.

(i) The distant future. Suppose that the function φ_ε satisfies the perturbed Helmholtz equation

$$-\left(\frac{d^2}{dx^2}+k^2+\varepsilon \mathrm{i}\right)\varphi_\varepsilon(x) = 0.$$

Then the time-dependent function

$$\psi(x,t) = \mathrm{e}^{\mathrm{i}t\sqrt{k^2+\mathrm{i}\varepsilon}}\varphi_\varepsilon(x), \qquad x,t \in \mathbb{R}$$

satisfies the wave equation $\psi_{tt} - \psi_{xx} = 0$. Because of (8.68), the function ψ is approximated by the wave

$$\psi_+(x,t) := \mathrm{e}^{\mathrm{i}kt}\mathrm{e}^{-\frac{1}{2}\varepsilon t/2k}\varphi_\varepsilon(x), \qquad x \in \mathbb{R}$$

which is damped as time t goes to $+\infty$. Furthermore, we have the following asymptotic behavior: $\lim_{t\to+\infty}\psi(x,t) = \varphi_\varepsilon(x)$ for all $x \in \mathbb{R}$.

(ii) The remote past. Similarly, suppose that the function φ_ε satisfies the perturbed Helmholtz equation

$$-\left(\frac{d^2}{dx^2} + k^2 - \varepsilon\mathrm{i}\right)\varphi_\varepsilon(x) = 0.$$

Then, the time-dependent function

$$\psi(x,t) = \mathrm{e}^{\mathrm{i}t\sqrt{k^2-\mathrm{i}\varepsilon}}\varphi_\varepsilon(x), \qquad x,t \in \mathbb{R}$$

satisfies the wave equation $\psi_{tt} - \psi_{xx} = 0$. Because of

$$\sqrt{k^2-\varepsilon\mathrm{i}} = k\left(1 - \frac{\varepsilon\mathrm{i}}{2k^2} + O(\varepsilon^2)\right), \qquad \varepsilon \to +0,$$

the function ψ is approximated by the wave

$$\psi_-(x,t) := \mathrm{e}^{\mathrm{i}kt}\mathrm{e}^{\frac{1}{2}\varepsilon t/2k}\varphi_\varepsilon(x), \qquad x,t \in \mathbb{R}$$

which is damped as time t goes to $-\infty$. Furthermore, we have the following asymptotic behavior: $\lim_{t\to-\infty}\psi(x,t) = \varphi_\varepsilon(x)$ for all $x \in \mathbb{R}$.

This shows that the perturbation terms $\pm\varepsilon\mathrm{i}$ arising in (8.66) and (8.69) correspond to damping (or absorption) effects.

This motivates the choice of the sign of the perturbation terms.

The limit $\varepsilon \to +0$ describes the vanishing of absorption. This explains why Propositions 8.11 and 8.12 are called the limiting absorption principle in mathematics. In physics, the limit $\varepsilon \to +0$ is also called the adiabatic limit.

8.5.5 Feynman's Regularized Fourier Method

Formal approach. The following method is due to Feynman. This method and its straightforward generalization to higher dimensions is frequently used in the physics textbooks on quantum field theory. The basic idea is to use the Fourier method in order to compute fundamental solutions of the Helmholtz equation. To begin with, recall that the Fourier transform $\hat{f}$ of the function f is given by

$$\hat{f}(p) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(x)\mathrm{e}^{-\mathrm{i}px}dx, \qquad p \in \mathbb{R}, \tag{8.70}$$

and we have the inversion formula

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \hat{f}(p) \mathrm{e}^{\mathrm{i}px} dx, \qquad x \in \mathbb{R}. \tag{8.71}$$

It follows from

$$\frac{d}{dx} f(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{i}p \hat{f}(p) \mathrm{e}^{\mathrm{i}px} \, dx, \qquad x \in \mathbb{R}$$

that

- the operation of differentiation $f(x) \mapsto \frac{d}{dx} f(x)$ is transformed into
- the operation of multiplication $\hat{f}(p) \mapsto \mathrm{i}p\hat{f}(p)$.

The function $p \mapsto \mathrm{i}p$ is called the symbol of the differential operator $\frac{d}{dx}$. Finally, using the Dirac delta function, we formally get

$$\hat{\delta}(p) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \delta(x) \mathrm{e}^{-\mathrm{i}px} dx = \frac{1}{\sqrt{2\pi}}, \qquad p \in \mathbb{R}.$$

Formal inversion yields s

$$\delta(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}px} dp, \qquad x \in \mathbb{R}.$$

Instead of $\hat{f}$, we also write the symbol $\mathcal{F}f$.[21] After these preparations, let us start with the equation

$$-\left(\frac{d^2}{dx^2} + k^2\right) F(x) = \delta(x), \qquad x \in \mathbb{R}. \tag{8.72}$$

Using the Fourier transform, we get

$$(p^2 - k^2)\hat{F}(p) = \hat{\delta}(p) = \frac{1}{\sqrt{2\pi}}.$$

Hence

$$\hat{F}(p) = \frac{1}{\sqrt{2\pi}\,(p^2 - k^2)}.$$

This implies

$$F(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\mathrm{e}^{\mathrm{i}px}}{p^2 - k^2} \, dp. \tag{8.73}$$

Unfortunately, this approach completely fails. In fact, the integral (8.73) does not exist because of the singularity at the points $p = \pm k$.

To cure this defect by regularization, let us pass to the perturbed equations

$$-\left(\frac{d^2}{dx^2} + k^2 \pm \varepsilon \mathrm{i}\right) F_{\pm,\varepsilon}(x) = \delta(x), \qquad x \in \mathbb{R} \tag{8.74}$$

where $\varepsilon > 0$ is sufficiently small. Replacing k^2 by $k^2 \pm \varepsilon \mathrm{i}$, the Fourier method yields

$$\boxed{F_{\pm,\varepsilon}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\mathrm{e}^{\mathrm{i}px}}{p^2 - k^2 \mp \varepsilon \mathrm{i}} \, dp.} \tag{8.75}$$

Rigorous approach. We now want to investigate the properties of the integrals (8.75) in rigorous terms.

[21] The rigorous Fourier transformation for both functions and tempered distributions is thoroughly studied in Chap. 11 of Vol. I.

Theorem 8.13 *Let $\varepsilon > 0$ be sufficiently small. Then the integrals (8.75) are equal to*

$$F_{\pm,\varepsilon}(x) = \pm \frac{\mathrm{i} e^{\pm \mathrm{i}|x|\sqrt{k^2 \pm \varepsilon \mathrm{i}}}}{2\sqrt{k^2 \pm \varepsilon \mathrm{i}}}, \qquad x \in \mathbb{R}.$$

Surprisingly enough, this coincides with the functions obtained in Sect. 8.5.4 by using a completely different method. Before giving the proof, let us discuss this result. By Sect. 8.5.4, the following are true in the sense of tempered distributions.

(i) Adiabatic limit: $\lim_{\varepsilon \to +0} F_{\pm,\varepsilon} = F_{\pm}$.
(ii) Fundamental solution: F_+ and F_- are solutions of the equation (8.72).
(iii) Fourier transform: $\lim_{\varepsilon \to +0} \mathcal{F}(F_{\pm,\varepsilon}) = \mathcal{F}(F_{\pm})$.[22]

Statement (ii) can be written as

$$\lim_{\varepsilon \to +0} \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\mathrm{e}^{\mathrm{i}px}}{p^2 - k^2 \mp \varepsilon \mathrm{i}}\, dp = F_{\pm}(x), \qquad x \in \mathbb{R}. \tag{8.76}$$

This is a rigorous statement in the language of tempered distributions. If we formally interchange the limit with the integration, then we get

$$F_{\pm}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \lim_{\varepsilon \to +0} \frac{\mathrm{e}^{\mathrm{i}px}}{p^2 - k^2 \mp \varepsilon \mathrm{i}}\, dp = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\mathrm{e}^{\mathrm{i}px}}{p^2 - k^2}\, dp.$$

In contrast to (8.76), this is only a mnemonic formula.

Proof of Theorem 8.13. We will use Cauchy's residue method from Section 4.4 of Volume I. Let $k > 0$ and $\varepsilon > 0$.

(I) Retarded case: Fix the real number x and consider the integral

$$F_{+,\varepsilon}(x) := \int_{-\infty}^{\infty} f(p) dp$$

with the integrand

$$f(p) := \frac{1}{2\pi} \cdot \frac{\mathrm{e}^{\mathrm{i}xp}}{p^2 - (k^2 + \varepsilon \mathrm{i})}.$$

We have to study the poles of the function f on the complex plane. Setting $p_+ := \sqrt{k^2 + \varepsilon \mathrm{i}}$ and $p_- := -p_+$, we get

$$f(p) = \frac{\mathrm{e}^{\mathrm{i}xp}}{2\pi} \cdot \frac{1}{(p - p_+)(p - p_-)} = \frac{\mathrm{e}^{\mathrm{i}xp}}{4\pi p_+} \left(\frac{1}{p - p_+} - \frac{1}{p - p_-} \right).$$

(I-1) Let $x \geq 0$. Choosing the curve C_+ as pictured in Fig. 8.11(a), Cauchy's residue theorem tells us that

$$\int_{C_+} f(p) dp = 2\pi \mathrm{i} \cdot \mathrm{res}_{p_+}(f) = \frac{\mathrm{i} \mathrm{e}^{\mathrm{i}xp_+}}{2p_+}.$$

Since the oriented curve C_+ is the union of the interval $[-R, R]$ with the semicircle S_R^+ in the upper half-plane, we have

$$\int_{-R}^{R} f(p) dp + \int_{S_R^+} f(p) dp = \frac{\mathrm{i} \mathrm{e}^{\mathrm{i}xp_+}}{2p_+}.$$

[22] Since the Fourier transform $\mathcal{F} : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ is sequentially continuous, statement (iii) is an immediate consequence of (ii).

(a) (b)

Fig. 8.11. Cauchy's residue method

Letting $R \to +\infty$, we obtain

$$\int_{-\infty}^{\infty} f(p)dp + \lim_{R\to+\infty} \int_{S_{R+}} f(p)dp = \frac{\mathrm{i}e^{\mathrm{i}xp_+}}{2p_+}.$$

It remains to show that $\lim_{R\to+\infty} \int_{S_R^+} f(p)dp = 0$. In fact, the semicircle S_R^+ can be parameterized by $p = Re^{\mathrm{i}\gamma}$, $0 \le \gamma \le \pi$. Hence

$$\int_{S_R^+} f(p)dp = \frac{1}{2\pi} \int_0^{\pi} \frac{e^{\mathrm{i}xR\cos\gamma} e^{-xR\sin\gamma}}{R^2 e^{2\mathrm{i}\gamma} - (k^2 + \varepsilon\mathrm{i})} \cdot \mathrm{i}Re^{\mathrm{i}\gamma} d\gamma.$$

Since $\sin\gamma \ge 0$ if $0 \le \gamma \le \pi$, we get $|e^{\mathrm{i}xR\cos\gamma} e^{-xR\sin\gamma} e^{\mathrm{i}\gamma}| = e^{-xR\sin\gamma} \le 1$. This implies

$$\int_{S_R^+} f(p)dp = O\left(\frac{1}{R}\right), \qquad R \to +\infty.$$

(I-2) Let $x < 0$. Use now the same argument with respect to the curve C_- as pictured in Fig. 8.11(b). Observing the negative orientation of the curve C_-, we get

$$\int_{C_-} f(p)dp = -2\pi\mathrm{i}\,\mathrm{res}_{p_-}(f) = \frac{\mathrm{i}e^{\mathrm{i}xp_-}}{2p_+}.$$

Recall that $p_+ = -p_-$. Since $x < 0$, we have $e^{-xR\sin\gamma} \ge 1$ if $-\pi \le \gamma \le 0$. Hence $\lim_{R\to+\infty} \int_{S_R^-} f(p)dp = 0$. Letting $R \to +\infty$, we obtain

$$\int_{-\infty}^{\infty} f(p)dp = \frac{\mathrm{i}e^{\mathrm{i}xp_-}}{2p_+} = \frac{\mathrm{i}e^{-\mathrm{i}xp_+}}{2p_+}.$$

(II) Advanced case: We now consider the integral $F_{-,\varepsilon}(x) = \int_{-\infty}^{\infty} f(p)dp$ with the integrand

$$f(p) := \frac{1}{2\pi} \cdot \frac{e^{\mathrm{i}xp}}{p^2 - (k^2 - \varepsilon\mathrm{i})}.$$

Setting $p_- := \sqrt{k^2 - \varepsilon\mathrm{i}}$ and $p_+ := -p_-$, we get

$$f(p) = \frac{e^{\mathrm{i}xp}}{2\pi} \cdot \frac{1}{(p-p_+)(p-p_-)} = \frac{e^{\mathrm{i}xp}}{4\pi p_+} \left(\frac{1}{p-p_+} - \frac{1}{p-p_-}\right).$$

Let $x \ge 0$ (resp. $x < 0$). Choosing the curve C_+ (resp. C_-) as pictured in Fig. 8.11, we obtain

$$\int_{-\infty}^{\infty} f(p)dp = \frac{\mathrm{i}e^{\mathrm{i}xp_+}}{2p_+} = -\frac{\mathrm{i}e^{-\mathrm{i}xp_-}}{2p_-}$$

(resp. $\int_{-\infty}^{\infty} f(p)dp = \frac{\mathrm{i}e^{\mathrm{i}xp_-}}{2p_+} = -\frac{\mathrm{i}e^{\mathrm{i}xp_-}}{2p_-}$). □

8.6 The Lippmann–Schwinger Integral Equation

We are given the smooth potential $U : \mathbb{R} \to \mathbb{R}$ with compact support. For fixed $k > 0$ and all real values x, set $\varphi_{\text{free}}(x) := \mathrm{e}^{\mathrm{i}kx}$. The Lippmann–Schwinger integral equation

$$\boxed{\varphi(x) = \varphi_{\text{free}}(x) - \kappa \int_{\mathbb{R}} \mathcal{G}^+(x,y)U(y)\varphi(y)dy, \qquad x \in \mathbb{R}} \tag{8.77}$$

plays a crucial role for studying scattering processes of a quantum particle on the real line under the influence of the potential U. Recall that $\mathcal{G}^+(x,y)$ is equal to $\frac{\mathrm{i}}{2k}\mathrm{e}^{\mathrm{i}k|x-y|}$.

Proposition 8.14 *If $\varphi : \mathbb{R} \to \mathbb{C}$ is a smooth solution of the integral equation (8.77), then it is a solution of the stationary Schrödinger equation*[23]

$$-\varphi''(x) - k^2\varphi(x) + \kappa U(x)\varphi(x) = 0, \qquad x \in \mathbb{R}.$$

Furthermore, the solution φ of (8.77) has the following asymptotics:

(i) $\varphi(x) \simeq e^{\mathrm{i}kx} + \varrho_+(k)e^{-\mathrm{i}kx}$ *as $x \to -\infty$ with the reflection coefficient*

$$\varrho_+(k) := \frac{1}{2k\mathrm{i}} \int_{\mathbb{R}} U(y)\varphi(y)e^{\mathrm{i}ky}dy.$$

(ii) $\varphi(x) \simeq \tau_+(k)e^{\mathrm{i}kx}$ *as $x \to +\infty$ with the transmission coefficient*

$$\tau_+(k) := 1 + \frac{1}{2k\mathrm{i}} \int_{\mathbb{R}} U(y)\varphi(y)e^{-iky}dy.$$

This follows immediately from Propositions 8.9 and 8.10.

8.6.1 The Born Approximation

Recall that $\varphi_{\text{free}}(x) := \mathrm{e}^{\mathrm{i}kx}$ for fixed $k > 0$ and all $x \in \mathbb{R}$. Replacing φ by φ_{free}, we obtain the following approximative solution

$$\varphi(x) = \varphi_{\text{free}}(x) - \kappa \int_{\mathbb{R}} \mathcal{G}^+(x,y)U(y)\varphi_{\text{free}}(y)dy, \qquad x \in \mathbb{R}$$

of the Lippmann–Schwinger integral equation (8.77). Furthermore, the reflection and transition coefficients of φ read as

$$\varrho_+(k) = \frac{1}{2k\mathrm{i}} \int_{\mathbb{R}} U(y)\mathrm{e}^{2\mathrm{i}ky}dy, \qquad \tau_+(k) = 1 + \frac{1}{2k\mathrm{i}} \int_{\mathbb{R}} U(y)dy.$$

This is called the Born approximation of the scattering process under the action of the potential U. As we will show in the next section, the Born approximation works well if the coupling constant $\kappa > 0$ is sufficiently small.

[23] Recall that we set $\hbar := 1$ and $m := \frac{1}{2}$.

8.6.2 The Existence and Uniqueness Theorem via Banach's Fixed Point Theorem

Define the norm $||\varphi||_\infty := \sup_{x\in\mathbb{R}} |\varphi(x)|$. By definition, the space $C_b(\mathbb{R})$ consists of all continuous functions $\varphi : \mathbb{R} \to \mathbb{C}$ with $||\varphi||_\infty < \infty$. This space is a complex Banach space equipped with the norm $||.||_\infty$.[24]

Theorem 8.15 *There exists a number $\kappa_0 > 0$ such that, for each given coupling constant $\kappa \in]0, \kappa_0[$, the Lippmann–Schwinger integral equation (8.77) has a unique solution φ in the Banach space $C_b(\mathbb{R})$. This solution can be obtained by the iterative method*

$$\varphi_{n+1}(x) = \varphi_0(x) - \kappa \int_\infty^\infty \mathcal{G}^+(x,y)U(y)\varphi_n(y)dy, \qquad x \in \mathbb{R}, \;\; n = 0, 1, \ldots$$

with $\varphi_0 := \varphi_{\text{free}}$. The sequence (φ_n) converges to φ in the Banach space $C_b(\mathbb{R})$, that is,

$$\lim_{n\to\infty} ||\varphi_n - \varphi||_\infty = 0.$$

The first approximation φ_1 coincides with the Born approximation.

Proof. Define

$$(B\varphi)(x) := -\int_{-\infty}^\infty \mathcal{G}^+(x,y)U(y)\varphi(y)dy \qquad \text{for all} \quad x \in \mathbb{R}.$$

Set $X := C_b(\mathbb{R})$. We have to solve the operator equation

$$\varphi = \varphi_{\text{free}} + \kappa B\varphi, \qquad \varphi \in X. \tag{8.78}$$

The operator $B : X \to X$ is linear and bounded. In fact, since $|\mathcal{G}^+(x,y)|$ is uniformly bounded and $\int_{-\infty}^\infty |U(y)|dy < \infty$, we get

$$|(B\varphi)(x)| \leq C \cdot \sup_{y\in\mathbb{R}} |\varphi(y)| \qquad \text{for all} \quad x \in \mathbb{R}$$

where C denotes some constant. Hence $||B\varphi||_\infty \leq C||\varphi||_\infty$ for all $\varphi \in X$. This implies $||B|| \leq C$. If $||\kappa B|| = |\kappa| \cdot ||B|| < 1$, then the inverse operator

$$(I - \kappa B)^{-1} : X \to X$$

exists and is bounded. Thus, the equation (8.78) has the unique solution

$$\varphi = (I - \kappa B)^{-1}\varphi_{\text{free}}.$$

Here, $(I - \kappa B)^{-1} = I + \kappa B + \kappa^2 B^2 + \ldots$ This geometric series (also called Neumann series)[25] is convergent with respect to the norm on the Banach space X. Consequently, the infinite series

$$\varphi = \lim_{n\to\infty} \left(\varphi_{\text{free}} + \sum_{m=1}^n \kappa^m B^m \varphi_{\text{free}} \right)$$

[24] This follows from the fact that the uniform limit of a sequence of continuous functions is again a continuous function.

[25] Carl Neumann (1832–1925), Stefan Banach (1892–1945).

is convergent in the Banach space X. This corresponds to the convergence

$$\varphi = \lim_{n\to\infty} \varphi_{n+1}$$

of the iterative method $\varphi_{n+1} = \varphi_0 + \kappa B\varphi_n, n = 1, 2, \ldots$ with $\varphi_0 := \varphi_{\text{free}}$.

This argument is a special case of the Banach fixed-point theorem (also called the contraction principle) (see Zeidler (1995a), Sect. 1.6, quoted on page 1049). □

8.6.3 Hypoellipticity

Let $N = 1, 2, \ldots$ We want to study the smoothness properties of the distributive solutions φ of the linear differential equation

$$\boxed{L\varphi = f \qquad \text{on } \mathbb{R}^N.} \tag{8.79}$$

Here, we assume that the symbol L denotes a general linear differential operator of order $m = 1, 2, \ldots$ with constant complex coefficients[26]

$$L := \sum_{|\alpha|\le m} a_\alpha \partial^\alpha.$$

We are given the distribution $f \in \mathcal{D}'(\mathbb{R}^N)$, and we are looking for a distribution $\varphi \in \mathcal{D}'(\mathbb{R}^N)$ such that equation (8.79) is valid. Recall from Volume I that the solutions $F \in \mathcal{D}'(\mathbb{R}^N)$ of the equation

$$LF = \delta$$

are called fundamental solutions of the differential operator L. Here, the symbol δ denotes the Dirac delta distribution given by $\delta(\chi) := \chi(0)$ for all test functions $\chi \in \mathcal{D}(\mathbb{R}^N)$. The symbol $\hat{L}$ of the differential operator L is defined by

$$\hat{L}(p) := \sum_{|\alpha|\le m} a_\alpha \mathrm{i}^{|\alpha|} p^\alpha \qquad \text{for all} \quad p \in \mathbb{R}^N.$$

This is obtained from the operator L by replacing the differential operator $\frac{\partial}{\partial x_j}$ by the complex variable $\mathrm{i}p_j$. In terms of the Fourier transform, we get

$$\mathcal{F}(L\varphi) = \hat{L}\mathcal{F}(\varphi) \qquad \text{for all} \quad \varphi \in \mathcal{S}(\mathbb{R}^N).$$

Here, $\mathcal{F}(\varphi)$ denotes the Fourier transform of the function φ. In other words, the symbol $\hat{L}$ tells us how the differential operator L acts in the Fourier space (momentum space) by multiplication.

Theorem 8.16 *The following three statements are equivalent.*

(i) *Hypoellipticity: The differential operator L is hypoelliptic. By definition, this means that*

$$\text{sing supp } \varphi = \text{sing supp } L\varphi \qquad \textit{for all} \quad \varphi \in \mathcal{D}'(\Omega)$$

and all open subsets Ω of $\mathbb{R}^N$.

[26] We exclude the trivial case where $L = 0$.

(ii) *Fundamental solution: The differential operator L has a fundamental solution F with trivial singular support, that is,* $\operatorname{sing\,supp} F = \{0\}$.

(iii) *Symbol: The partial derivatives $\partial^\beta \hat{L}$ of order $1, 2, \ldots$ of the symbol have the asymptotic behavior*

$$\lim_{|p|\to\infty} \left| \frac{\partial^\beta \hat{L}(p)}{\hat{L}(p)} \right| = 0, \qquad p \in \mathbb{R}^N.$$

The proof can be found in Hörmander (1983), Vol. 2, Sect. 11.1. Hypoelliptic differential operators were characterized by Hörmander in 1955.[27] For his contributions to the theory of differential operators, Hörmander was awarded the Fields medal in 1962.

Recall from Sect. 11.5.1 of Vol. I that the singular support of the distribution $\varphi \in \mathcal{D}'(\mathbb{R}^N)$ is the largest open subset Ω of $\mathbb{R}^N$ such that φ is smooth on Ω. In particular, if condition (ii) is met, then the smoothness of the function $f : \mathbb{R}^N \to \mathbb{C}$ implies the smoothness of all the solutions $\varphi : \mathbb{R}^N \to \mathbb{C}$ of equation (8.79). In terms of the wavefront set,[28] the linear differential operator L is hypoelliptic iff

$$WF(\varphi) = WF(L\varphi) \qquad \text{for all} \quad \varphi \in \mathcal{D}'(\Omega)$$

and all open subsets Ω of $\mathbb{R}^N$.

Example. Consider the Helmholtz equation

$$\left(-\frac{d^2}{dx^2} - k^2 \right) \varphi = f \qquad \text{on } \mathbb{R} \tag{8.80}$$

for fixed $k > 0$. The symbol of the Helmholtz operator $L = -\frac{d^2}{dx^2} - k^2$ is obtained by the substitution $\frac{d}{dx} \Rightarrow \mathrm{i}p$. Hence

$$\hat{L}(p) = p^2 - k^2.$$

Obviously,

$$\lim_{p\to\pm\infty} \frac{\hat{L}^{(n)}(p)}{\hat{L}(p)} = 0, \qquad n = 1, 2, \ldots$$

The fundamental solution $F_+(x) = \frac{\mathrm{i}\mathrm{e}^{\mathrm{i}k|x|}}{2k}$ of the Helmholtz operator is smooth on $\mathbb{R} \setminus \{0\}$. Hence $\operatorname{sing\,supp} F_+ = \{0\}$. Thus, the Helmholtz operator is hypoelliptic. This means that all of the solutions $\varphi \in \mathcal{D}'(\mathbb{R})$ of equation (8.80) are smooth functions on $\mathbb{R}$ if the given right-hand side $f : \mathbb{R} \to \mathbb{C}$ is smooth.

Further examples of hypoelliptic differential operators are given by the Laplace equation and the heat equation (resp. diffusion equation) in $\mathbb{R}^N$ with $N = 1, 2, \ldots$ In contrast to this, the wave equation is not hypoelliptic. This reflects the typical fact that the wave equation describes the propagation of waves which may possess discontinuities (e.g. jumps of the electromagnetic field).

---

[27] L. Hörmander, On the theory of general linear differential operators, Acta Math. **94** (1955), 161–248.

[28] See Section 11.5.2 of Volume I.

9. A Glance at General Scattering Theory

> The time-dependent picture of the scattering process has been described in 1945 by Christian Møller (1902–1980). A stationary picture has been formulated by John Wheeler (born 1911) in 1937. It was elaborated by Werner Heisenberg (1901–1976) in 1943.[1]
> In 1930 Heisenberg lectured at the University of Chicago on the new quantum mathematics. His assistant there was the young American physicist Frank Hoyt, who helped prepare the English lecture notes. During the Second World War, Hoyt joined the Manhattan project for building nuclear weapons; one of his assignments was to scrutinize every wartime publication of Heisenberg and see if it could be a by-product of bomb research. Hoyt studied very thoroughly the two papers Heisenberg published in 1943 on scattering theory, and as he told me later, concluded that they had no bearing on nuclear weapon. This may have saved Heisenberg's life, for the OSS, the wartime precursor of the CIA, had been training an agent to assassinate him.[2]
>
> Peter Lax, 2003

The main tool for studying experimentally the properties of elementary particles are scattering processes performed in particle accelerators. Physicists measure cross sections and the mathematical theory has to develop methods for computing cross sections.

> *The computation of cross sections for scattering processes is based on the computation of transition probabilities.*

In turn, the transition probabilities are encoded into a specific unitary operator, which is called the scattering operator (or the Heisenberg S-matrix). The mathematics of scattering processes in quantum mechanics and quantum field theory is sophisticated. Until now, there exists a complete mathematical scattering theory for the scattering of N particles in quantum mechanics (i.e., the scattering of atoms and molecules), but such a complete theory is missing in quantum field theory.

> *The reason for the mathematical difficulties is the physical phenomenon of clustering in scattering processes.*

---

[1] J. Wheeler, On the mathematical description of light nuclei by the method of resonating group structure, Phys. Rev. **52** (1937), 107–122.
W. Heisenberg, The observable quantities in particle physics I–III, Z. Phys. **120** (1943), 513–538, 673–702; **123** (1944), 93–112 (in German).
C. Møller, General properties of the characteristic matrix in the theory of elementary particles I,II, Det. Kgl. Danske Videnskab. Selskab. Mat.-Phys. Medd. **22** (1945), 1–48; **23** (1946), 1–46.

[2] P. Lax, Functional Analysis, Wiley, New York, 2003 (reprinted with permission).

In a particle accelerator, the incoming free particles might form bound states for a finite time by clustering. At the end of the process, the particles are again free particles. The intermediate cluster states of finite lifetime are also called resonances in physics. In the quantum-mechanical scattering of N particles, the complexity of the cluster states and the number of different scattering channels becomes more and more complex if N increases. Here the critical case $N \geq 4$ is much more complicated than the case $N = 3$.

The rigorous scattering theory for quantum N-particle systems is a highlight in mathematical physics.

From the mathematical point of view, scattering theory studies the perturbation of the absolutely continuous spectrum of self-adjoint operators on Hilbert spaces. In what follows, we make the following hypotheses.

(H1) The perturbed Hamiltonian: The operator $H : D(H) \to X$ is self-adjoint on the dense subset $D(H)$ of the complex infinite-dimensional separable Hilbert space X. The domain of definition $D(H)$ is dense in X. We have the splitting

$$\boxed{H = H_{\text{free}} + U.}$$

(H2) The free Hamiltonian: The operator $H_{\text{free}} : D(H_{\text{free}}) \to X$ is self-adjoint on the dense subset $D(H_{\text{free}})$ of X. The spectrum of H_{free} coincides with its absolutely continuous spectrum (i.e., both the pure point spectrum and the singular spectrum of H_{free} are empty).

(H3) The perturbation: The operator $U : D(U) \to X$ is symmetric on the dense subset $D(U)$ of X. The singular spectrum of the perturbed Hamiltonian H is empty. We distinguish between short-range potentials U (e.g., the Yukawa potential) and long-range potentials (e.g., the Coulomb potential).[3]

From the physical point of view, the free Hamiltonian H_{free} describes the unperturbed free motion in the absence of bound states. This corresponds to the missing pure point spectrum of H_{free}. Our goal is to study the perturbed motion governed by the perturbation potential U (e.g., the perturbed motion on the real line as investigated below). We have the decomposition

$$\boxed{X = X_{\text{bound}} \oplus X_{\text{scatt}}}$$

into two orthogonal closed subspaces of the Hilbert space X, where X_{bound} and X_{scatt} correspond to the pure point spectrum and the absolutely continuous spectrum of the perturbed Hamiltonian H, respectively (see (7.106) on page 503). Recall that if $\varphi \in X_{\text{scatt}}$, then the corresponding spectral measure has a density ϱ, that is, the expectation value of the energy measured in the normalized state φ can be represented as

$$\bar{E} = \langle \varphi | H\varphi \rangle = \int_{\mathbb{R}} \varrho(E) dE.$$

In terms of physics, the elements $\psi(0)$ of the subspace X_{bound} (resp. X_{scatt}) are the initial states of bound motion (resp. scattering motion) (see Fig. 8.1 on page 700).

The perturbed dynamics. Let us introduce

- the free Feynman propagator $P_{\text{free}}(t, t_0) := \mathrm{e}^{-\mathrm{i}(t-t_0)H_{\text{free}}/\hbar}$, $t \geq t_0$, and
- the perturbed Feynman propagator $P(t, t_0) := \mathrm{e}^{-\mathrm{i}t(t-t_0)H/\hbar}$, $t \geq t_0$.

[3] The mathematical scattering theory for long-range potentials is much more complicated than for short-range potentials.

In order to understand scattering processes, one has to study the following motions of the quantum system under consideration.

- Free motion in the future: $\psi_{\text{out}}(t) = P_{\text{free}}(t,0)\varphi_{\text{out}}$ for all times $t \geq 0$.
- Free motion in the past: $\psi_{\text{in}}(s) = P_{\text{free}}(0,s)\varphi_{\text{in}}$ for all times $s \leq 0$.
- Perturbed motion: $\psi(t) = P(t,t_0)\psi(t_0)$ for all times $t \geq t_0$.

The transition probability. Let $t > 0$. The real number

$$a(t,-t) := \langle \psi_{\text{out}}(t) | P(t,-t)\psi_{\text{in}}(-t) \rangle \tag{9.1}$$

is called the transition amplitude with respect to the time interval $[-t,t]$. Furthermore, the real number

$$p(t,-t) = |a(t,-t)|^2$$

is called the transition probability with respect to the time interval $[-t,t]$. This allows the following physical interpretation:

- Fix the time $t > 0$. Suppose that the physical system is in the free motion state $\psi_{\text{in}}(-t) = P_{\text{free}}(0,-t)\varphi_{\text{in}}$ at the negative time $-t$.
- Then the system is in the state $P(t,-t)\psi_{\text{in}}(-t)$ at the positive time t.
- Suppose that we measure the system at the positive time t, and we observe the free motion state $\psi_{\text{out}}(t) = P(t,0)\varphi_{\text{out}}$ at time t.

Then the number $p(t,-t)$ is the probability for this measurement. To be precise, we assume that $\psi_{\text{in}}(-t)$ and $\psi_{\text{out}}(t)$ are normalized quantum states, that is,

$$||\psi_{\text{in}}(-t)|| = ||\psi_{\text{out}}(t)|| = 1 \qquad \text{for all} \quad t \geq 0. \tag{9.2}$$

Since the propagators are unitary operators, it is sufficient to assume that the condition (9.2) is satisfied, say, at time $t = 0$ (i.e., $||\varphi_{\text{in}}|| = ||\varphi_{\text{out}}|| = 1$). This implies $||P(t,-t)\psi_{\text{in}}(-t)|| = 1$ for all $t \geq 0$. Our goal is to compute the limit

$$a(+\infty,-\infty) := \lim_{t\to+\infty} a(t,-t)$$

which yields the desired transition probability

$$\boxed{p(+\infty,-\infty) := |a(+\infty,-\infty)|^2}$$

over the infinite time interval $[-\infty,+\infty]$. In particle accelerators, this corresponds to the transition probability over a fairly 'long' time interval.

9.1 The Formal Basic Idea

The scattering operator. We have $(\mathrm{e}^{-\mathrm{i}tH_{\text{free}}/\hbar})^\dagger = \mathrm{e}^{\mathrm{i}tH_{\text{free}}/\hbar}$ for the adjoint operator. This implies

$$a(t,-t) = \langle \varphi_{\text{out}} | \mathrm{e}^{\mathrm{i}tH_{\text{free}}t/\hbar} P(t,-t) \mathrm{e}^{\mathrm{i}tH_{\text{free}}/\hbar} \varphi_{\text{in}} \rangle.$$

Motivated by this expression, we define

$$\mathsf{S}(t,-t) := \mathrm{e}^{\mathrm{i}tH_{\text{free}}t/\hbar} P(t,-t) \mathrm{e}^{\mathrm{i}tH\text{free}/\hbar}.$$

This is called the scattering operator with respect to the time interval $[-t,t]$. Suppose that the limit

$$\mathsf{S}_{\text{formal}} := \lim_{t\to+\infty} \mathsf{S}(t,-t) \tag{9.3}$$

exists. Then we get

$$a(+\infty,-\infty) := \lim_{t\to+\infty} a(t,-t) = \langle \varphi_{\text{out}} | \mathsf{S}_{\text{formal}} \varphi_{\text{in}} \rangle,$$

which is called the S-matrix element with respect to the states φ_{in} and φ_{out}. This yields the transition probability

$$p(+\infty,-\infty) = |\langle \varphi_{\text{in}} | \mathsf{S}_{\text{formal}} \varphi_{\text{out}} \rangle|^2.$$

The main problem is that

The limit (9.3) does not exist, as a rule.

To overcome this difficulty, we will introduce the notion of wave operators.

The trick of Møller's wave operators. Observe that $P(t,-t)$ is equal to the product $P(t,0)P(0,-t)$. Therefore, the transition amplitude can be written as

$$a(t,-t) = \langle \mathrm{e}^{-\mathrm{i}tH_{\text{free}}/\hbar} \varphi_{\text{out}} | P(t,0)P(0,-t) \mathrm{e}^{\mathrm{i}tH_{\text{free}}/\hbar} \varphi_{\text{in}} \rangle.$$

Hence

$$a(t,-t) = \langle \mathrm{e}^{\mathrm{i}tH/\hbar} \mathrm{e}^{-\mathrm{i}tH_{\text{free}}/\hbar} \varphi_{\text{out}} | \mathrm{e}^{\mathrm{i}sH/\hbar} \mathrm{e}^{-\mathrm{i}sH_{\text{free}}/\hbar} \varphi_{\text{in}} \rangle,$$

where we set $s := -t$. Suppose that the following limits exist:

- $W_{\text{out}}\varphi_{\text{out}} := \lim_{t\to+\infty} \mathrm{e}^{\mathrm{i}tH/\hbar} \mathrm{e}^{-\mathrm{i}tH_{\text{free}}/\hbar} \varphi_{\text{out}}$.
- $W_{\text{in}}\varphi_{\text{out}} := \lim_{s\to-\infty} \mathrm{e}^{\mathrm{i}sH/\hbar} \mathrm{e}^{-\mathrm{i}sH_{\text{free}}/\hbar} \varphi_{\text{in}}$.

Letting $t \to +\infty$, we obtain

$$a(+\infty,-\infty) = \langle W_{\text{out}}\varphi_{\text{out}} | W_{\text{in}}\varphi_{\text{in}} \rangle.$$

Finally, suppose that the element $W_{\text{in}}\varphi_{\text{in}}$ is contained in the domain of the adjoint operator $W_{\text{out}}^\dagger$. Then

$$a(+\infty,\infty) = \langle \varphi_{\text{out}} | W_{\text{out}}^\dagger W_{\text{in}} \varphi_{\text{in}} \rangle.$$

Introducing the scattering operator

$$\mathsf{S}\varphi_{\text{in}} := W_{\text{out}}^\dagger W_{\text{in}} \varphi_{\text{in}}, \tag{9.4}$$

we obtain $a(+\infty,\infty) = \langle \varphi_{\text{out}} | \mathsf{S}\varphi_{\text{in}} \rangle$. Finally, we get the transition probability

$$\boxed{p(+\infty,-\infty) := |\langle \varphi_{\text{out}} | \mathsf{S}\varphi_{\text{in}} \rangle|^2.} \tag{9.5}$$

In particular, this procedure replaces the formal limit (9.3) by the rigorous limit

$$a(+\infty,-\infty) = \lim_{t\to\infty} \langle \varphi_{\text{out}} | \mathsf{S}(t,-t)\varphi_{\text{in}} \rangle = \langle \varphi_{\text{out}} | \mathsf{S}\varphi_{\text{in}} \rangle. \tag{9.6}$$

This is the limit of the corresponding matrix elements. We will show below that for complete wave operators, the scattering operator

$$\mathsf{S} : X \to X$$

is unitary, and the limit (9.6) exists for all $\varphi_{\text{in}}, \varphi_{\text{out}} \in X$. In terms of weak convergence, this means that

$$\boxed{w - \lim_{t\to+\infty} \mathsf{S}(t,-t)\varphi_{\text{in}} = \mathsf{S}\varphi_{\text{in}} \qquad \text{for all} \quad \varphi_{\text{in}} \in X.} \tag{9.7}$$

Rigorous scattering theory studies the properties of the Møller wave operators and uses the two key relations (9.4) and (9.5). Let us summarize some crucial results. The proofs can be found in:

S. Gustafson and I. Sigal, Mathematical Concepts of Quantum Mechanics, Springer, Berlin, 2003.

I. Sigal, Scattering Theory for Many-Body Quantum Mechanical Systems: Rigorous Results, Springer, New York, 1983.

M. Schechter, Operator Methods in Quantum Mechanics, North-Holland, Amsterdam, 1982,

and in the standard textbook on mathematical scattering theory by M. Reed and B. Simon, Methods of Modern Mathematical Physics, Vol. III, Academic Press, 1979. The language used by physicists in scattering theory can be found in:

R. Newton, Scattering Theory of Waves and Particles, Springer, New York, 1982.

A. Bohm, Quantum Mechanics: Foundations and Applications, Springer, Berlin, 1994.

9.2 The Rigorous Time-Dependent Approach

A mathematical basis for the quantum-mechanical scattering theory was laid out in terms of wave operators about 30 years ago. For instance the scattering operator was expressed as a product of wave operators. As a result, the main mathematical problem of the mathematical quantum-mechanical scattering theory was defined as the proof of establishing their properties, the latter being
(i) isometry,
(ii) mutual orthogonality, and
(iii) completeness.
The existence proof was found very fast by Cook.[4] At the same time it was shown by Jauch that the existence implies readily the isometry and mutual orthogonality... The completeness however proved, from the very beginning, to be a hard nut and required a certain mathematical sophistication in tackling it... These notes are devoted to the mathematical foundations of the quantum N-body problem. A small part of them describes essentially known results in the field, while the rest is concentrated on the crux of the problem; the N-body asymptotic completeness.[5]

Israel Michael Sigal, 1983

Our monograph would be much simpler, shorter and less interesting if we restricted ourselves to short range potentials. In fact, long-range scattering (including the Coulomb potential) is the central subject of our monograph. We study it under very general conditions, which are motivated by the mathematical structure of the problem.[6]

Jan Dereziński and Christian Gérard, 1997

[4] J. Cook, Convergence of the Møller wave matrix, J. Math. Phys. **36** (1957), 82–87. J. Jauch, Theory of the scattering operator, Helv. Phys. Acta **31** (1958), I, II (multichannel scattering), **31** (1958), 127–158, 661–684.

[5] I. Sigal, Scattering Theory for Many-Body Quantum Mechanical Systems: Rigorous Results, Springer, New York, 1983 (reprinted with permission).

[6] J. Dereziński, C. Gérard, Scattering Theory of Classical and Quantum N-Particle Systems, Springer, New York, 1997 (reprinted with permission).

Asymptotically free motion in the distant future. The theory of Møller's wave operators is based on the following simple observation.

Proposition 9.1 *For the perturbed dynamics* $\psi(t) = \mathrm{e}^{-\mathrm{i}tH/\hbar}\psi(0)$*, the following two conditions are equivalent:*

(i) *There exists a state* $\varphi_{\mathrm{out}} \in X$ *such that*

$$\lim_{t\to+\infty}(\mathrm{e}^{-\mathrm{i}tH/\hbar}\psi(0) - \mathrm{e}^{-\mathrm{i}tH_{\mathrm{free}}/\hbar}\varphi_{\mathrm{out}}) = 0.$$

(ii) *The limit* $\psi(0) = \lim_{t\to+\infty} e^{\mathrm{i}tH/\hbar}e^{-\mathrm{i}tH_{\mathrm{free}}/\hbar}\varphi_{\mathrm{out}}$ *exists.*

Proof. We have the identity

$$\mathrm{e}^{-\mathrm{i}tH/\hbar}\psi(0) - \mathrm{e}^{-\mathrm{i}tH_{\mathrm{free}}/\hbar}\varphi_{\mathrm{out}} = \mathrm{e}^{-\mathrm{i}tH/\hbar}(\psi(0) - \mathrm{e}^{\mathrm{i}tH/\hbar}\mathrm{e}^{-\mathrm{i}tH_{\mathrm{free}}/\hbar}\varphi_{\mathrm{out}}).$$

Since the operator $\mathrm{e}^{-\mathrm{i}tH/\hbar} : X \to X$ is unitary, we get

$$||\mathrm{e}^{-\mathrm{i}tH/\hbar}\psi(0) - \mathrm{e}^{-\mathrm{i}tH_{\mathrm{free}}/\hbar}\varphi_{\mathrm{out}}|| = ||\psi(0) - \mathrm{e}^{\mathrm{i}tH/\hbar}\mathrm{e}^{-\mathrm{i}tH_{\mathrm{free}}/\hbar}\varphi_{\mathrm{out}}||.$$

Letting $t \to +\infty$, we get the claim. □

By definition, the motion $\psi(t) = \mathrm{e}^{-\mathrm{i}tH/\hbar}\psi(0), t \in \mathbb{R}$ is asymptotically free at time $t = +\infty$ iff condition (i) is satisfied. We say that the wave operator W_{out} exists iff the limit

$$\boxed{W_{\mathrm{out}}\varphi_{\mathrm{out}} := \lim_{t\to+\infty}\mathrm{e}^{\mathrm{i}tH/\hbar}\mathrm{e}^{-\mathrm{i}tH_{\mathrm{free}}/\hbar}\varphi_{\mathrm{out}}}$$

exists for all $\varphi_{\mathrm{out}} \in X$.

Asymptotically free motion in the remote past. We now replace $t = +\infty$ by $t = -\infty$. For the perturbed dynamics $\psi(t) = \mathrm{e}^{-\mathrm{i}tH/\hbar}\psi(0)$, the following two conditions are equivalent:

(i) There exists a state $\varphi_{\mathrm{in}} \in X$ such that

$$\lim_{t\to-\infty}(\mathrm{e}^{-\mathrm{i}tH/\hbar}\psi(0) - \mathrm{e}^{-\mathrm{i}tH_{\mathrm{free}}/\hbar}\varphi_{\mathrm{in}}) = 0.$$

(ii) The limit $\psi(0) = \lim_{t\to+\infty}\mathrm{e}^{\mathrm{i}tH/\hbar}\mathrm{e}^{-\mathrm{i}tH_{\mathrm{free}}/\hbar}\varphi_{\mathrm{int}}$ exists.

By definition, the motion $\psi(t) = \mathrm{e}^{-\mathrm{i}tH/\hbar}\psi(0), t \in \mathbb{R}$ is asymptotically free at time $t = -\infty$ iff condition (i) is satisfied. We say that the wave operator W_{in} exists iff the limit

$$\boxed{W_{\mathrm{in}}\varphi_{\mathrm{in}} := \lim_{t\to-\infty}\mathrm{e}^{\mathrm{i}tH/\hbar}\mathrm{e}^{-\mathrm{i}tH_{\mathrm{free}}/\hbar}\varphi_{\mathrm{in}}}$$

exists for all $\varphi_{\mathrm{in}} \in X$.

The motion $\psi(t) = \mathrm{e}^{-\mathrm{i}tH/\hbar}\psi(0), t \in \mathbb{R}$ is called a scattering process iff it is asymptotically free at times $t = -\infty$ and $t = +\infty$, that is, we have $\psi(0) \in \mathrm{im}(W_{\mathrm{in}}) \cap \mathrm{im}(W_{\mathrm{out}})$.

Completeness of the wave operators. The wave operators W_{in} and W_{out} are called complete iff these wave operators exist and

$$\mathrm{im}(W_{\mathrm{out}}) = \mathrm{im}(W_{\mathrm{in}}) = X_{\mathrm{scatt}}.$$

In this case, the following hold: The Hilbert space $X = X_{\mathrm{bound}} \oplus X_{\mathrm{scatt}}$ is decomposed into the orthogonal closed subspaces X_{bound} and X_{scatt}. Consider the motion $\psi(t) := \mathrm{e}^{-\mathrm{i}tH/\hbar}\psi(0)$ for all times $t \in \mathbb{R}$.

- If $\psi(0) \in X_{\text{bound}}$, then $\psi(t) \in X_{\text{bound}}$ for all $t \in \mathbb{R}$. This is called a bound motion.
- The motion $t \mapsto \psi(t)$ is a scattering process iff $\psi(0) \in X_{\text{scatt}}$.

The main theorem on the scattering operator. We set $\mathsf{S} := W_{\text{out}}^{\dagger} W_{\text{in}}$ provided this operator exists. Furthermore, set $\mathsf{S}(t,s) := \mathrm{e}^{-\mathrm{i}(t-s)H/\hbar}$ for all $t, s \in \mathbb{R}$.

Theorem 9.2 *If the wave operators W_{in} and W_{out} are complete, then the following hold:*

(i) *The two wave operators $W_{\text{in}}, W_{\text{out}} : X \to X_{\text{scatt}}$ are unitary. That is, the Hilbert spaces X and X_{scatt} are unitarily equivalent.*

(ii). *The scattering operator $\mathsf{S} : X \to X$ is unitary. Explicitly, we have the relation $\mathsf{S} = W_{\text{out}}^{-1} W_{\text{in}}$.*

(iii) *We have the weak limit*

$$\mathsf{S}\varphi_{\text{in}} = w - \lim_{t \to +\infty} \mathsf{S}(t,-t)\varphi_{\text{in}} \qquad \text{for all} \quad \varphi_{\text{in}} \in X.$$

This implies the limit of the corresponding matrix elements, that is,

$$\lim_{t \to +\infty} \langle \varphi_{\text{out}} | \mathsf{S}(t,-t)\varphi_{\text{in}} \rangle = \langle \varphi_{\text{out}} | \mathsf{S}\varphi_{\text{in}} \rangle \qquad \text{for all} \quad \varphi_{\text{in}}, \varphi_{\text{out}} \in X.$$

(iv) *The one-parameter unitary groups generated by the Hamiltonians H and H_{free} on X_{scatt} and X, respectively, are unitarily equivalent. Explicitly,*

$$e^{-\mathrm{i}tH/\hbar}\varphi = W_{\text{in}} e^{-\mathrm{i}tH_{\text{free}}/\hbar} W_{\text{in}}^{-1}\varphi \qquad \text{for all} \quad \varphi \in X_{\text{scatt}},\ t \in \mathbb{R}.$$

The same is true if we replace W_{in} by W_{out}.

The proof of this theorem together with the proof of the following Theorem 9.3 can be found in M. Reed and B. Simon (1979), Methods of Modern Mathematical Physics, Vol. III, Sect. XI.3, Academic Press, New York, 1979.

The completeness theorem for wave operators. The following theorem relates the completeness of wave operators to compact perturbations.

Theorem 9.3 *The wave operators W_{in} and W_{out} are complete if one of the following two conditions is satisfied:*

(i) *The difference $H - H_{\text{free}}$ is of trace class (Kato–Rosenblum condition).*

(ii) *The difference $(H + \mathrm{i}I)^{-1} - (H_{\text{free}} + \mathrm{i}I)^{-1}$ is of trace class (Kuroda–Birman condition).*

9.3 The Rigorous Time-Independent Approach

The definition of the wave operators above has been based on the time limits $t \to +\infty$ and $t \to -\infty$. This is the time-dependent approach to scattering theory (also called instationary scattering theory). In mathematics and physics, one also uses the time-independent (or stationary) approach. The following theorem shows that the wave operators can be computed by the time-independent formulas

$$W_{\text{out}}\varphi = \lim_{\varepsilon \to +0} \int_{\mathbb{R}} R(E - \mathrm{i}\varepsilon) R_{\text{free}}(E + \mathrm{i}\varepsilon)\varphi \, dE,$$

$$W_{\text{in}}\varphi = \lim_{\varepsilon \to -0} \int_{\mathbb{R}} R(E - \mathrm{i}\varepsilon) R_{\text{free}}(E + \mathrm{i}\varepsilon)\varphi \, dE. \tag{9.8}$$

Here, $R(\lambda) := (\lambda I - H)^{-1}$ and $R_{\text{free}}(\lambda) := (\lambda I - H_{\text{free}})^{-1}$ are the resolvents of the Hamiltonian H and the free Hamiltonian H_{free}.

Theorem 9.4 *If the wave operators exist, then we have the formulas* (9.8) *for all elements* φ *of the complex Hilbert space* X.

The proof can be found in Sigal (1983), p. 16, quoted on page 751. The proof essentially uses the operator-valued Dirac delta function and the Abel limit

$$A - \lim_{t\to+\infty} f(t) := \lim_{\varepsilon\to+0} \int_0^\infty \varepsilon \mathrm{e}^{-\varepsilon t} f(t)dt$$

dating back to Abel (1802–1829). The Abel limit generalizes the classical limit $\lim_{t\to+\infty} f(t)$. In fact, if this classical limit exists for the given continuous function $f : [0,\infty[\to X$, then

$$A - \lim_{t\to+\infty} f(t) = \lim_{t\to+\infty} f(t). \tag{9.9}$$

For the proof, see Problem 9.10.[7] The discrete variant of the Abel limit reads as

$$A - \lim_{n\to+\infty} s_n := \lim_{\varepsilon\to+0} \sum_{n=0}^\infty \varepsilon \mathrm{e}^{-\varepsilon n} s_n.$$

If the sequence (s_n) is bounded in the Hilbert space X and the limit $\lim_{n\to\infty} s_n$ exists, then

$$A - \lim_{n\to\infty} s_n = \lim_{n\to\infty} s_n.$$

In particular, for an infinite series $\sum_{k=0}^\infty a_k$, we set $s_n := \sum_{k=0}^n a_k$, and we define the Abel summation

$$A - \sum_{k=0}^\infty a_k := A - \lim_{n\to\infty} s_n.$$

If the series $\sum_{k=0}^\infty a_k$ is convergent, then $A - \sum_{k=0}^\infty a_k = \sum_{k=0}^\infty a_k$.

Theorem 9.4 indicates that scattering theory is closely related to analytic functions with values in a Banach space and their boundary values.

9.4 Applications to Quantum Mechanics

Let us consider the Schrödinger equation

$$\boxed{\mathrm{i}\hbar\psi_t(x,t) = -\frac{\hbar^2}{2m}\psi_{xx}(x,t) + U(x)\psi(x,t), \qquad x,t\in\mathbb{R}} \tag{9.10}$$

for the motion of a quantum particle of mass $m > 0$ on the real line. We assume that the potential function $U : \mathbb{R} \to \mathbb{R}$ is smooth. We set

$$U_0 := \inf_{x\in\mathbb{R}} U(x),$$

and we postulate that $U_0 > -\infty$.[8]

[7] Far reaching generalizations of the Abel limit and their applications to scattering theory can be found in H. Baumgärtel and M. Wollenberg, Mathematical Scattering Theory, Birkhäuser, Basel, 1983.

[8] For the energy $E_{\text{class}} := \frac{1}{2}m\dot{x}^2 + U(x)$ of a classical particle on the real line, we have $E_{\text{class}} \geq U_0$.

We are interested in getting information on the possible energy values of the quantum particle.

As we will see below, the structure of the energy spectrum of (9.10) strongly depends on the shape of the potential U. For any function $\varphi \in \mathcal{D}(\mathbb{R})$, define

$$\boxed{(H_{\mathrm{pre}}\varphi)(x) := -\frac{\hbar^2}{2m}\varphi''(x) + U(x)\varphi(x) \qquad \text{for all} \quad x \in \mathbb{R}.}$$

The linear operator $H_{\mathrm{pre}} : \mathcal{D}(\mathbb{R}) \to L_2(\mathbb{R})$ on the complex Hilbert space $L_2(\mathbb{R})$ is called the pre-Hamiltonian for the motion of the quantum particle. We have[9]

$$\langle\varphi|H_{\mathrm{pre}}\varphi\rangle \geq U_0\langle\varphi|\varphi\rangle \qquad \text{for all} \quad \varphi \in \mathcal{D}(\mathbb{R}).$$

That is, the symmetric operator H_{pre} is semi-bounded from below, in the sense of Friedrichs. Let us summarize some important results. The proofs can be found in the textbooks by M. Schechter, Operator Methods in Quantum Mechanics, North-Holland, Amsterdam, 1982, and F. Berezin and M. Shubin, The Schrödinger Equation, Kluwer, Dordrecht, 1991. The operator H_{pre} is *essentially self-adjoint* on the Hilbert space $L_2(\mathbb{R})$. This means that the operator H_{pre} has a unique self-adjoint extension

$$H : D(H) \to L_2(\mathbb{R})$$

which is called the Hamiltonian of the quantum dynamics. Observe that we have the following inclusions: $D(H_{\mathrm{pre}}) = \mathcal{D}(\mathbb{R}) \subseteq D(H) \subseteq L_2(\mathbb{R})$. Our goal is

- to characterize the domain of definition $D(H)$ and
- the *spectrum* $\sigma(H)$ of H. Recall that precisely the points E in the spectrum $\sigma(H)$ represent the possible energy values of the quantum particle. In particular, we will get

$$E \geq U_0 \qquad \text{for all} \quad E \in \sigma(H).$$

That is, the lower bound U_0 for the energies of the classical particle is also a lower bound for the energies of the quantum particle.

In what follows we will use the classification of the spectrum $\sigma(H)$ introduced on page 504. In all the situations considered below, the singular spectrum of the Hamiltonian H is empty, and we have the orthogonal decomposition

$$\boxed{X = X_{\mathrm{bound}} \oplus X_{\mathrm{scatt}}}$$

of the Hilbert space $X := L_2(\mathbb{R})$.

- If the Hamiltonian H has no eigenvectors in the Hilbert space X, then we obtain $X_{\mathrm{bound}} = \{0\}$ and $X_{\mathrm{scatt}} = X$.

[9] In fact, let $\varphi \in \mathcal{D}(\mathbb{R})$. Then integration by parts tells us that the inner product

$$\langle\varphi|H_{\mathrm{pre}}\varphi\rangle = \int_{\mathbb{R}} \varphi(x)^\dagger \Big(-\frac{\hbar^2}{2m}\varphi''(x) + U(x)\varphi(x)\Big)dx$$

is equal to the integral $J := \int_{\mathbb{R}} \big(\frac{\hbar^2}{2m}\varphi'(x)^\dagger\varphi'(x) + \varphi(x)^\dagger U(x)\varphi(x)\big)dx$, and

$$J \geq \int_{\mathbb{R}} \varphi(x)^\dagger U(x)\varphi(x)dx \geq \int_{\mathbb{R}} U_0\varphi(x)^\dagger\varphi(x)dx = \langle\varphi|\varphi\rangle.$$

- If the Hamiltonian H has at least one eigenvector in X, then the space X_{bound} is the closed linear hull of the set of eigenvectors of H, and the linear subspace space $X_{\text{scatt}} = X_{\text{bound}}^{\perp}$ of X is the orthogonal complement to X_{bound}.

For any $\varphi_0 \in X_{\text{scatt}}$, the dynamics $\psi(t) = \mathrm{e}^{-\mathrm{i}tH/\hbar}\varphi_0$ for all $t \in \mathbb{R}$ corresponds to a scattering motion (i.e., asymptotically free motion as $t \to \pm\infty$). The pure point spectrum $\sigma_{\text{pp}}(H)$ of the Hamiltonian H is the closure of the set of eigenvalues of H. The absolutely continuous spectrum $\sigma_{\text{ac}}(H)$ of H is the spectrum of the restriction $H : X_{\text{scatt}} \to X_{\text{scatt}}$.

Free motion. Let $U \equiv 0$. The unique self-adjoint extension of the pre-Hamiltonian H_{pre} is given by $H_{\text{free}} : W_2^2(\mathbb{R}) \to L_2(\mathbb{R})$. Here, for any element φ of $W_2^2(\mathbb{R})$, we have

$$(H_{\text{free}}\varphi)(x) = -\frac{\hbar^2}{2m}\varphi''(x), \qquad x \in \mathbb{R}$$

in the sense of distributions.[10] The operator H_{free} is called the free Hamiltonian. For the spectrum, we have $X_{\text{bound}} = \{0\}$, as well as $X = X_{\text{scatt}}$ and

$$\sigma(H_{\text{free}}) = \sigma_{\text{ac}}(H_{\text{free}}) = \sigma_{\text{ess}}(H_{\text{free}}) = [0, +\infty[, \quad \sigma_{\text{pp}}(H) = \emptyset,$$

where $\sigma_{\text{ess}}(H)$ denotes the essential spectrum of H.

Bound states. Suppose that $U(x) \to +\infty$ as $|x| \to +\infty$. For the spectrum of the Hamiltonian H, we have $X = X_{\text{bound}}$ and $X_{\text{scatt}} = \{0\}$, as well as

$$\sigma(H) = \sigma_{pp}(H) = \{E_0, E_1, \ldots\}, \quad \sigma_{\text{ac}}(H) = \sigma_{\text{ess}}(H) = \emptyset.$$

The operator H has the eigenvalues $E_0, E_1, \ldots$ with $\lim_{n\to+\infty} E_j = +\infty$ and

$$H\varphi_j = E_j\varphi_j, \qquad j = 0, 1, 2, \ldots$$

All the eigenvalues $E_0, E_1, \ldots$ are simple. The smooth functions $\varphi_0, \varphi_1, \ldots$ form a complete orthonormal system in $X = L_2(\mathbb{R})$. The function φ_0 has no zeros, and φ_n has precisely n zeros if $n = 1, 2, \ldots$ We have $\varphi \in D(H)$ iff $\varphi \in L_2(\mathbb{R})$ and $\sum_{j=0}^{\infty} E_j^2 |\langle\varphi_j|\varphi\rangle|^2 < \infty$.

The special case where $U(x) := \frac{1}{2}\omega^2 x^2$ for all $x \in \mathbb{R}$ describes the harmonic quantum oscillator.

Short-range potential. Let $U \in \mathcal{D}(\mathbb{R})$. Then the operator H_{pre} has the unique self-adjoint extension $H : W_2^2(\mathbb{R}) \to L_2(\mathbb{R})$. For any $\varphi \in W_2^2(\mathbb{R})$, we have

$$\boxed{(H_{\text{pre}}\varphi)(x) := -\frac{\hbar^2}{2m}\varphi''(x) + U(x)\varphi(x) \qquad \text{for all} \quad x \in \mathbb{R},}$$

in the sense of distributions.

(i) Spectrum: $\sigma(H) \subseteq [U_0, +\infty[$, and $\sigma_{\text{ac}}(H) = \sigma_{\text{ess}}(H) = [0, \infty[$. The pure point spectrum $\sigma_{pp}(H)$ is either empty or finite with

$$\sigma_{pp}(H) = \{E_0, \ldots, E_N\}$$

and $U_0 \le E_0 \le E_1 \le \ldots \le E_N < 0$. The finite-dimensional space X_{bound} is the linear hull of the eigenvectors of H, and X_{scatt} is the orthogonal complement to X_{bound}.

[10] The definition of the Sobolev space $W_2^m(\mathbb{R}), m = 1, 2, \ldots$ can be found in (7.120) on page 514.

- If $U(x) \geq 0$ for all $x \in \mathbb{R}$, then $X_{\text{bound}} = \{0\}$. That is, $\sigma_{pp}(H)$ is empty, and $\sigma(H) = \sigma_{\text{ac}}(H) = \sigma_{\text{ess}}(H) = [0, \infty[$.
- If $\int_{\mathbb{R}} U(x)dx < 0$, then $\sigma_{\text{pp}}(H)$ is not empty.

(ii) Scattering processes: The wave operators $W_{\text{in}}, W_{\text{out}}$ with respect to H, H_{free} are complete. In particular, the operators

$$W_{\text{in}} : X \to X_{\text{scatt}}, \qquad W_{\text{out}} : X \to X_{\text{scatt}}$$

are unitary. If $\varphi_0 \in X_{\text{scatt}}$, then

$$\mathrm{e}^{-\mathrm{i}tH/\hbar}\varphi_0 = W_{\text{in}}\mathrm{e}^{-\mathrm{i}tH_{\text{free}}/\hbar}W_{\text{in}}^{-1}\varphi_0 \qquad \text{for all} \quad t \in \mathbb{R}.$$

This tells us that the scattering dynamics under the influence of the potential U is unitarily equivalent to the free motion with vanishing potential.

Long-range potential. Suppose that $\int_{\mathbb{R}} U(x)dx = -\infty$ and $U \leq 0$ outside a compact interval.[11] Then the Hamiltonian H has an infinite number of negative eigenvalues with $U_0 \leq E_0 \leq E_1 \leq \ldots < 0$.

9.5 A Glance at Quantum Field Theory

In quantum field theory, one encounters the crucial situation where

- the free Hamiltonian $H_{\text{free}} : D(H_{\text{free}}) \to X_{\text{free}}$ and
- the Hamiltonian $H : D(H) \to X$ of interaction

are defined on different Hilbert spaces X_{free} and X, respectively. The corresponding Haag–Ruelle theory is studied in the monograph by Reed and Simon (1979), Vol. III, Sect. XI.16, quoted on page 759.[12] We also refer to the monograph by Iagolnitzer (1993), quoted on page 759. Let us briefly discuss the basic idea of the modified wave operators. We consider the following motions of the quantum system:

- $\psi(t) := \mathrm{e}^{-\mathrm{i}tH/\hbar}\psi(0)$ (scattering process on the Hilbert space X of interaction).
- $\psi_{\text{out}}(t) := \mathrm{e}^{-\mathrm{i}tH_{\text{free}}/\hbar}\varphi_{\text{out}}, t \geq 0$ (free motion on the Hilbert space X_{free} in the future).
- $\psi_{\text{in}}(t) := \mathrm{e}^{-\mathrm{i}tH_{\text{free}}/\hbar}\varphi_{\text{in}}, t \leq 0$ (free motion on the Hilbert space X_{free} in the past).

We also need a linear continuous operator $J : X_{\text{free}} \to X$ which intertwines the two Hilbert spaces. In this case, we replace the transition amplitude (9.1) by the expression

$$a(t, -t) := \langle J\psi_{\text{out}}(t) | P(t, -t)J\psi_{\text{in}}(-t)\rangle \tag{9.11}$$

where $P(t, t_0) := \mathrm{e}^{-\mathrm{i}(t-t_0)H/\hbar}$. Hence

$$a(t, -t) = \langle \mathrm{e}^{\mathrm{i}tH/\hbar}J\mathrm{e}^{-\mathrm{i}tH_{\text{free}}/\hbar}\varphi_{\text{out}} | \mathrm{e}^{\mathrm{i}tH/\hbar}J\mathrm{e}^{\mathrm{i}tH_{\text{free}}\hbar}\varphi_{\text{in}}\rangle.$$

Introducing the wave operators

[11] For example, $U(x) = -\frac{1}{|x|+1}$ for all $x \in \mathbb{R}$.

[12] R. Haag, Quantum field theories with particles and asymptotic conditions, Phys. Rev. **112** (1958), 669–673.
D. Ruelle, On the asymptotic condition in quantum field theory, Helv. Phys. Acta **35** (1962), 147–163.
J. Dereziński and C. Gérard, Asymptotic completeness in quantum field theory. Massive Pauli–Fierz Hamiltonians, Rev. Math. Phys. **11** (4) (1999), 383–450.

- $W_{\rm out}\varphi_{\rm out} := \lim_{t\to+\infty} {\rm e}^{{\rm i}tH/\hbar} J {\rm e}^{-{\rm i}tH_{\rm free}/\hbar}\varphi_{\rm out}$,
- $W_{\rm in}\varphi_{\rm in} := \lim_{s\to-\infty} {\rm e}^{{\rm i}sH/\hbar} J {\rm e}^{-{\rm i}sH_{\rm free}/\hbar}\varphi_{\rm in}$,

we get $a(+\infty,-\infty) = \lim_{t\to+\infty}\langle W_{\rm out}\varphi_{\rm out}|W_{\rm in}\varphi_{\rm in}\rangle$. Finally, introducing the scattering operator $\mathsf{S} := W_{\rm out}^\dagger W_{\rm in}$, we obtain

$$a(+\infty,-\infty) = \langle\varphi_{\rm out}|\mathsf{S}\varphi_{\rm in}\rangle.$$

This yields the transition probability $p(+\infty,-\infty) = |\langle\varphi_{\rm out}|\mathsf{S}\varphi_{\rm in}\rangle|^2$.

9.6 Hints for Further Reading

As an introduction to the relations between quantum mechanics and mathematics we recommend:

G. Gustafson and I. Sigal, Mathematical Concepts of Quantum Mechanics, Springer, New York, 2003;

P. Hislop and I. Sigal, Introduction to Spectral Theory With Applications to Schrödinger Operators, Springer, New York, 1996 (including the spectral theory of non-self-adjoint operators for describing resonances)

together with:

M. Schechter, Operator Methods in Quantum Mechanics, North-Holland, Amsterdam, 1982.

L. Faddeev and O. Yakubovskii, Lectures on Quantum Mechanics for Students of Mathematics, Springer, New York, 1986.

A. Komech, Lectures on Quantum Mechanics (nonlinear PDE – partial differential equation – point of view), Lecture Notes no. 25 of the Max Planck Institute for Mathematics in the Sciences, Leipzig, 2005.
Internet: `http://mis.mpg.de/preprints/ln/`

Numerous explicit computations for concrete problems in quantum mechanics can be found in:

S. Flügge, Practical Quantum Mechanics, Vols. 1, 2, Springer, Berlin, 1979.

Mathematical Scattering theory

,
We refer to the standard textbook:

M. Reed and B. Simon, Methods of Mathematical Physics III: Scattering Theory, Academic Press, New York, 1979

together with:

T. Kato, Perturbation Theory for Linear Operators, Springer, Berlin, 1966.

P. Lax and R. Phillips, Scattering Theory, Academic Press, New York, 1967.

B. Simon, Quantum Mechanics for Hamiltonians Defined on Quadratic Forms, Princeton University Press, 1971.

P. Lax and R. Phillips, Scattering Theory for Automorphic Functions, Princeton University Press, 1976.

W. Amrein, J. Jauch, and K. Sinha, Scattering Theory in Quantum Mechanics, Benjamin, Reading, Massachusetts, 1977.

H. Baumgärtel and M. Wollenberg, Mathematical Scattering Theory, Birkhäuser, Basel, 1983.

A. Galindo and P. Pascual, Quantum Mechanics, Vols. 1, 2, Springer, Berlin, 1990.

P. Lax, Functional Analysis, Wiley, New York, 2002.

The mathematics of scattering processes in quantum field theory is studied in:

M. Reed and B. Simon, Methods of Modern Mathematical Physics III: Scattering Theory, Academic Press, New York, 1978 (Haag–Ruelle theory).

N. Bogoliubov, A. Logunov, and I. Todorov, I. (1975), Introduction to Axiomatic Quantum Field Theory, Benjamin, Reading, Massachusetts, 1975.

N. Bogoliubov et al., General Principles of Quantum Field Theory, Kluwer, Dordrecht, 1990 (1200 references).

D. Iagolnitzer, Scattering in Quantum Field Theory, Princeton University Press, 1993.

The Physics of Scattering Processes

M. Goldberger and K. Watson, Collision Theory, Wiley, New York. 1964.

J. Taylor, Scattering Theory, Wiley, New York, 1972.

R. Newton, Scattering Theory of Waves and Particles, Springer, New York, 1982.

A. Bohm, Quantum Mechanics: Foundations and Applications, Springer, Berlin, 1994.

F. Schwabl, Quantum Mechanics, Springer, Berlin, 2000.

K. Gottfried and Tung-Mow Yan, Quantum Mechanics: Fundamentals, Springer, New York, 2003.

Analytic S-matrix theory: In the late 1950s, Chew emphasized the importance of the analyticity properties of the S-matrix for understanding the strong interaction in particle accelerators. For this analytic S-matrix theory, we refer to:

G. Chew, S-Matrix Theory of Strong Interaction, Benjamin, New York, 1962.

G. Chew, The Analytic S-Matrix: A Basis for Nuclear Democracy, Benjamin, New York, 1966.

A. Barut, The Theory of the Scattering Matrix, MacMillan, New York, 1967 (symmetry and analytic continuation properties of the S-matrix, dispersion relations and Mandelstam theory, complex angular momentum and Regge poles).

Analytic S-matrix theory is a mixture of rigorous mathematical results and formal argumentations which are strongly motivated by physical intuition. We also refer to:

N. Bogolyubov, B. Medvedev, and M. Polivanov, Problems of the Theory of Dispersion Relations, Fizmatgiz, Moscow, 1958 (in Russian).[13]

R. Eden et al., The Analytic S-Matrix Theory, Cambridge University Press, 1966.

R. Eden, High Energy Collisions of Elementary Particles, Cambridge University Press, 1967.

I. Todorov, Analytic Properties of Feynman Diagrams in Quantum Field Theory, Pergamon Press, London, 1971.

Explicit Solutions

W. de Melo and S. van Strien, One-Dimensional Dynamics, Springer, Berlin, 1993 (classical dynamical systems and chaos).

M. Takahashi, Thermodynamics of One-dimensional Solvable Models, Cambridge University Press, 1999.

S. Brandt and H. Dahmen, The Picture Book of Quantum Mechanics, Springer, New York, 1995.

E. Lieb and D. Mattis, Mathematical Physics in One Dimension: Exactly Soluble Models of Interacting Particles (a collection of reprints), Academic Press, New York, 1986.

D. Mattis, The Many-Body Problem: An Encyclopedia of Exactly Solved One-Dimensional Models, World Scientific, Singapore, 1993.

R. Gilmore, Elementary Quantum Mechanics in One Dimension, John Hopkins University Press, Baltimore, Maryland, 2004 (numerical computations).

Concerning explicitly solvable models, we also refer to:

Baxter, Exactly Solved Models in Statistical Mechanics, Academic Press, New York, 1982.

L. Faddeev, Integrable models in $1+1$-dimensional quantum field theory. Les Houches, session XXXIX, pp. 561–608, Elsevier, Amsterdam, 1984.

M. Berry, Riemann's Zeta Function: a Model of Quantum Chaos, Lecture Notes in Physics **263**, Springer, Berlin, 1986.

S. Albeverio, Solvable Models in Quantum Mechanics, Springer, New York, 1988.

H. Grosse, Models in Statistical Physics and Quantum Field Theory, Springer, New York, 1988.

B. Hughes, Random Walks and Random Environments, Vols. 1, 2, Clarendon Press, Oxford, 1995.

B. Felsager, Geometry, Particles, and Fields, Springer, New York, 1997.

B. McCoy and Tai-Tsu Wu, The Two-Dimensional Ising Model, Harvard University Press, Cambridge, Massachusetts, 1997.

[13] See also H. Bremmermann, R. Oehme, and J. Taylor, Proof of dispersion relations in quantized field theories, Phys. Rev. **109** (1958), 2178–2190.
N. Bogolyubov and D. Shirkov, Introduction to Quantum Field Theory, Interscience, New York, 1980 (Chap. IX studies dispersion relations).

D. Thouless, (Ed.), Topological Quantum Numbers in Non-Relativistic Physics, World Scientific, Singapore, 1998 (collection of 40 important articles on superfluidity, quantum Hall effect, phase transitions, etc.)

W. Dittrich and M. Reutter, Classical and Quantum Dynamics, Springer, Berlin, 1999.

S. Ketov, Quantum Non-Linear Sigma Models: From Quantum Field Theory to Supersymmetry, Conformal Field Theory, Black Holes and Strings, Springer, Berlin, 2000.

E. Abdalla, M. Abdalla, and K. Rothe, Non-perturbative Methods in Two-Dimensional Quantum Field Theory, World Scientific, Singapore, 2001.

A. Tsvelik, Quantum Field Theory in Condensed Matter Physics, Cambridge University Press, 2003 (many applications, e.g., Fermi liquids, Tomonaga–Luttinger liquids, electrodynamics in metals, quantum Hall effect, the Schwinger model in (1+1)-quantum electrodynamics, spin models, nonlinear sigma models, and the Wess–Zumino–Novikov–Witten model).

G. Volovik, The Universum in a Helium Droplet, Clarendon Press, Oxford, 2003.

T. Asselmeyer and C. Brans, Exotic Structures and Physics: Differential Topology and Space-Time Models, World Scientific, Singapore, 2004.

B. Sutherland, Beautiful Models: 70 Years of Exactly Solved Quantum Many-Body Problems, World Scientific, Singapore, 2004.

S. Bellucci, S. Ferrara, and A. Marrani, Supersymmetric Mechanics, Vol. 1: Supersymmetry, Noncommutativity and Matrix Models, Vol. 2: The Attractor Mechanism and Space Time Singularities, Springer, Berlin, 2006.

T. Dauxois and M. Peyrard, Physics of Solitons, Cambridge University Press, 2006 (very stable nonlinear waves).

G. Mussardo, Il Modello di Ising: Introduzioni alla Teoria dei Campi e telle Tranzioni di Fase (The Ising model: an introduction to (quantum) field theory and phase transitions), Torino, Bollati–Bozinglieri, Italia, 2007 (in Italian).

In 1983 Subramanyan Chandrasekhar was awarded the Nobel prize in physics for his theoretical studies of the physical processes of importance to the structure and evolution of stars. Chandrasekhar (1910–1995) was a master of computing explicit solutions in physics. We recommend the reader to have a look at the following classic survey article

S. Chandrasekhar, Stochastic problems in physics and astronomy, Rev. Mod. Phys. **15** (1943), 1–89,

and at the following classic books:

S. Chandrasekhar, An Introduction to the Study of Stellar Structure, Dover, New York, 1939. Reprinted in 1957.

S. Chandrasekhar, Principles of Stellar Dynamics, University of Chicago Press, 1943. Reprinted by Dover, New York, 1960.

S. Chandrasekhar, Hydrodynamic and Hydromagnetic Stability, Oxford University Press, 1961.

S. Chandrasekhar, Ellipsoidal Figures of Equilibrium, Yale University Press, New Haven, Connecticut, 1969.

S. Chandrasekhar, The Mathematical Theory of Black Holes, Clarendon Press, Oxford, 1983.

Special Functions

Singular ordinary differential operators and the Fourier–Stieltjes transformation:

N. Dunford and J. Schwartz, Linear Operators, Vol. II, Wiley, New York, 1958.

K. Yosida, Lectures on Differential- and Integral Equations, Interscience, New York, 1960.

M. Naimark, Linear Differential Operators, Frederyck Ungar, New York, 1967.

K. Jörgens and F. Rellich, Eigenvalue Problems for Ordinary Differential Equations, Springer, Berlin, 1976 (in German).

G. Rozenblum, M. Solomiak, and M. Shubin, Spectral Theory of Differential Operators, Springer, Berlin, 1991 (Encyclopedia of Mathematical Sciences).

Special functions in quantum mechanics:

E. Whittaker and G. Watson, A Course of Modern Analysis, Cambridge University Press, 1944. (The first edition of the American Mathematical Society was published in 1979.)

A. Sommerfeld, Lectures on Theoretical Physics VI: Partial Differential Equations in Mathematical Physics, Academic Press, New York, 1949.

P. Morse and H. Feshbach, Methods of Theoretical Physics, Vols. 1, 2, McGraw-Hill, New York, 1953.

V. Smirnov, A Course of Higher Mathematics, Vol. III/2, Pergamon Press, New York, 1964.

H. Jeffreys and B. Jeffreys, Methods of Mathematical Physics, Cambridge University Press, 1972.

A. Wawrzyńczyk, Group Representations and Special Functions, Reidel, Dordrecht, 1984.

H. Triebel, Higher Analysis, Barth, Leipzig, 1989.

N. Vilenkin and A. Klimyk, Special Functions and Representations of Lie Groups, Vols. I–IV, Kluwer, Dordrecht, 1991.

Handbooks of special functions:

M. Abramowitz and I. Stegun (Eds.), Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Wiley, New York, and National Bureau of Standards, Washington, DC, 1984 (see also NIST Digital Library of Mathematical Functions (2007) below).

I. Gradshtein and I. Ryshik, Tables of Integrals, Series, and Products, Academic Press, New York, 1980.

A. Prudnikov, Yu. Brychkov, and O. Manichev, Integrals and Series, Vols. I–V, Gordon and Breach, New York, 1986.

A. Erdély et al., Higher Transcendental Functions, Vols. I–III, McGraw-Hill, New York, 2006.

NIST Digital Library of Mathematical Functions (2007).
Internet: `http://dlmf.nist.gov`

Sophisticated N-Particle Scattering in Quantum Mechanics

L. Faddeev and S. Merkuryev, Quantum Scattering Theory for Several Particle Systems, Kluwer, Dordrecht, 1996. (In 1963 Faddeev solved the quantum three-body problem.)[14]

I. Sigal, Scattering Theory for Many-Body Quantum Mechanical Systems: Rigorous Results, Springer, New York, 1983. (Sigal solved the quantum N-particle problem for short-range potentials.)

J. Dereziński and C. Gérard, Scattering Theory of Classical and Quantum N-Particle Systems, Springer, New York, 1997. (In 1993 Dereziński solved the quantum N-particle problem for the most difficult case of long-range potentials.)

Problems

9.1 *Continuity of parameter integrals.* We consider the function

$$F(p) := \int_M f(x,p)dx \tag{9.12}$$

for all parameters $p \in P$. We are given the function

$$f : M \times P \to \mathbb{C}$$

where M is a nonempty Lebesgue measurable subset of $\mathbb{R}^N$ (e.g., an open set) with $N = 1, 2, \ldots$, and the parameter space P is a nonempty open subset of the real line $\mathbb{R}$. Show that the function $F : P \to \mathbb{C}$ is continuous if the following conditions are satisfied.

(i) For each parameter $p \in P$, the function $x \mapsto f(x,p)$ is measurable on M (e.g., $x \mapsto f(x,p)$ is continuous for almost all $x \in M$).[15] For the elements of measure theory and integration theory, we refer to Zeidler (1995a), Vol. 1, Appendix (see the references on page 1049.

(ii) There exists an integrable function $g : M \to \mathbb{R}$ such that

$$|f(x,p)| \le g(x) \qquad \text{for all} \quad p \in P$$

and almost all $x \in M$.

(iii) The function $p \mapsto f(x,p)$ is continuous on P for almost all $x \in M$.

[14] L. Faddev, Mathematical Aspects of the Three-Body Problem in Quantum Mechanics, Steklov-Institute, Leningrad, 1963 (in Russian). English edition: Israel Program for Scientific Translation, 1965.
J. Dereziński, Asymptotic completeness for N-particle long-range quantum systems, Ann. Math. **138** (1993), 427–476.

[15] This means that there exists a subset S of M which has the N-dimensional Lebesgue measure zero such that $x \mapsto f(x,p)$ is continuous for all $x \in M \setminus S$. The definition of sets of measure zero can be found in Vol. I, Sec. 10.2.3. Note that, roughly speaking, each set of dimension $< N$ has the N-dimensional Lebesgue measure zero. The prototype of a set of N-dimensional Lebesgue measure in $\mathbb{R}^N$ is a finite or countable set of points.

Solution: Let $p_n \to p$ as $n \to \infty$. Then, we get

$$|f(x, p_n)| \leq g(x) \qquad \text{for all} \quad n = 1, 2, \ldots$$

and almost all $x \in M$. By (ii), the principle of dominated convergence[16] tells us that

$$\lim_{n\to\infty} F(p_n) = \int_M \lim_{n\to\infty} f(x, p_n) dx.$$

Hence $\lim_{n\to\infty} F(p_n) = \int_M f(x, p) dx = F(p)$.

9.2 *Differentiability of parameter integrals.* Show that the function $F : P \to \mathbb{R}$ is differentiable and

$$F'(p) = \int_M f_p(x, p) dx \qquad \text{for all} \quad p \in P$$

if the following additional conditions are satisfied.

(iv) For each parameter $p \in P$, the partial derivative $f_p(x, p)$ exists for almost all $x \in M$.

(v) There exists an integrable function $h : M \to \mathbb{R}$ such that

$$|f_p(x, p)| \leq h(x) \qquad \text{for all} \quad p \in P$$

and almost all $x \in M$.

Solution: Observe that

$$\lim_{h\to 0} \frac{F(p+h) - F(p)}{h} = \int_M \lim_{h\to 0} \frac{f(x, p+h) - f(x, p)}{h}\, dx = \int_M f_p(x, p) dx.$$

This is justified by the principle of dominated convergence. In fact, by the mean value theorem, there exists a number $\vartheta \in]0, 1[$ such that

$$\left| \frac{f(x, p+h) - f(x, p)}{h} \right| = |f_p(x, p + \vartheta h)| \leq h(x).$$

9.3 *Dispersion relation.* Recall that the subset

$$\mathbb{C}_> := \{z \in \mathbb{C} : \Im(z) > 0\}$$

of the complex plane $\mathbb{C}$ is called the open upper-half plane. Moreover, the set

$$\mathbb{C}_\geq := \{z \in \mathbb{C} : \Im(z) \geq 0\}$$

is called the closed upper-half plane. Let us make the following assumptions.

- The function $f : \mathbb{C}_\geq \to \mathbb{C}$ is continuous on the closed upper-half-plane, up to a finite number of points on the real axis.
- The function f is holomorphic on the open upper-half plane $\mathbb{C}_>$.
- The function f has precisely the zeros $z_1, \ldots, z_n$ on $\mathbb{C}_>$ and no zeros on the real axis. All of the zeros z_j are simple.
- $\sup_{y\geq 0} \int_{\mathbb{R}} \left|\ln |f(x + \mathrm{i}y)|\right|^2 dy < \infty$.

[16] See Zeidler (1995a), Vol. 1, p. 437, quoted on page 1049

Show that for all points $x \in \mathbb{R}$ on the real line, we have

$$\arg f(x) = \frac{1}{\mathrm{i}} \sum_{j=1}^{n} \ln \frac{x - z_j}{x - z_j^\dagger} + \frac{1}{\pi}\, PV \int_{\mathbb{R}} \frac{\ln |f(\xi)|}{x - \xi}\, d\xi. \tag{9.13}$$

The symbol $\frac{1}{\pi}\, PV \int_{\mathbb{R}} \ldots$ is to be understood in the sense of $H(\ln |f|)$ with the Hilbert transform $H : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ (see Sec. 11.9.3 of Volume I). Equation (9.13) shows that there exists a relation between the modulus $|f(x)|$ and the argument $\arg f(x)$ of the boundary values of the function f on the real line. Relations of this type play a fundamental role in elementary particle physics. They are called dispersion relations.
Solution: The trick is to introduce the auxiliary function

$$g(z) := f(z) \prod_{j=1}^{n} \frac{z - z_j^\dagger}{z - z_j} \qquad \text{for all} \quad z \in \mathbb{C}_{\geq}.$$

We will reduce the proof to the Hilbert transform of g. The function g has no zeros on $\mathbb{C}_{\geq}$. Therefore,

$$f(z) := g(z) \prod_{j=1}^{n} \frac{z - z_j}{z - z_j^\dagger} \qquad \text{for all} \quad z \in \mathbb{C}_{\geq}. \tag{9.14}$$

Recall that for all $z \in \mathbb{C} \setminus \{0\}$, we have

$$\ln z := \ln |z| + \mathrm{i} \arg z \qquad \text{where} \quad -\pi < \arg z \leq \pi.$$

It follows from $\left| \frac{x - z_j}{x - z_j^\dagger} \right| = 1$ for all $x \in \mathbb{R}$ that $|f(x)| = |g(x)|$ on the real line. By (9.14),

$$\ln f(x) := \ln g(x) + \ln \prod_{j=1}^{n} \frac{x - z_j}{x - z_j^\dagger} \qquad \text{for all} \quad x \in \mathbb{R}.$$

Noting that $\ln \prod_{j=1}^{n} \left| \frac{x - z_j}{x - z_j^\dagger} \right| = \ln 1 = 0$, this implies

$$\mathrm{i} \arg f(x) = \mathrm{i} \arg g(x) + \ln \prod_{j=1}^{n} \frac{x - z_j}{x - z_j^\dagger} = \mathrm{i} \arg g(x) + \sum_{j=1}^{n} \ln \frac{x - z_j}{x - z_j^\dagger}.$$

By the Hilbert transform, $\Im(\ln g) = H\Re(\ln g)$ on the real line. That is,

$$\arg g(x) = \frac{1}{\pi}\, PV \int_{\mathbb{R}} \frac{\ln |g(\xi)|}{x - \xi}\, d\xi \qquad \text{for all} \quad x \in \mathbb{R}.$$

9.4 *Ariadne's thread in matrix theory.* We want to study the complex (2×2)-matrix

$$S = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix}.$$

Recall that by definition, the dual (or transposed) matrix S^d, the adjoint matrix $S^\dagger$, and the conjugate-complex matrix S^c read as[17]

$$S^d = \begin{pmatrix} s_{11} & s_{21} \\ s_{12} & s_{22} \end{pmatrix}, \quad S^\dagger = \begin{pmatrix} s_{11}^\dagger & s_{21}^\dagger \\ s_{12}^\dagger & s_{22}^\dagger \end{pmatrix}, \quad S^c = \begin{pmatrix} s_{11}^\dagger & s_{12}^\dagger \\ s_{21}^\dagger & s_{22}^\dagger \end{pmatrix}. \tag{9.15}$$

Here, $s_{11}^\dagger$ denotes the conjugate-complex number to s_{11}, and so on. In particular, $S^c = (S^\dagger)^d$ and $S^\dagger = (S^c)^d$. Furthermore, the unit matrix I and the diagonal matrix $\mathrm{diag}(\mu, \nu)$ are given by

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathrm{diag}(\mu, \nu) = \begin{pmatrix} \mu & 0 \\ 0 & \nu \end{pmatrix}.$$

Finally, the determinant $\det S$ and the trace $\mathrm{tr}\, S$ of the matrix S are defined by

$$\det S := \begin{vmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{vmatrix} = s_{11}s_{22} - s_{12}s_{21}, \quad \mathrm{tr}\, S := s_{11} + s_{22}. \tag{9.16}$$

The matrix S is called self-adjoint (resp. skew-adjoint) iff we have $S^\dagger = S$ (resp. $S^\dagger = -S$). The matrix S is invertible iff $\det S \neq 0$. Explicitly,

$$S^{-1} = \frac{1}{\det S} \begin{pmatrix} s_{22} & -s_{12} \\ -s_{21} & s_{11} \end{pmatrix}. \tag{9.17}$$

Then $SS^{-1} = S^{-1}S = I$. The complex number λ is an eigenvalue of the matrix S iff the equation

$$Sx = \lambda x, \qquad x \in \mathbb{C}^2, \quad x \neq 0$$

has a solution. This is equivalent to the fact that the matrix $\lambda I - S$ is not invertible. Thus, the eigenvalues λ_1, λ_2 of the matrix S are the solutions of the equation $\det(\lambda I - S) = 0,\ \lambda \in \mathbb{C}$. Explicitly,

$$\lambda^2 - \lambda\, \mathrm{tr}\, S + \det S = 0, \qquad \lambda \in \mathbb{C}.$$

This so-called characteristic equation of S is equal to $(\lambda - \lambda_1)(\lambda - \lambda_2) = 0$.[18] Hence

$$\mathrm{tr}\, S = \lambda_1 + \lambda_2, \qquad \det S = \lambda_1\lambda_2. \tag{9.18}$$

The set of eigenvalues λ_1, λ_2 is called the spectrum $\sigma(S)$ of the matrix S. The complement $\varrho(S) := \mathbb{C} \setminus \sigma(S)$ is called the resolvent set of the matrix S. Let λ be a complex number. The inverse matrix

$$R(\lambda) = (\lambda I - S)^{-1}$$

[17] Instead of S^d one also writes S^t or ${}^tS$. In terms of functional analysis, dual matrices correspond to dual operators A^d on dual spaces X^d. Therefore, we use the symbol 'd' which refers to 'duality'.

[18] In celestial mechanics, the characteristic equation is also called the secular equation.

exists iff $\lambda \in \varrho(\lambda)$. The matrix $R(\lambda)$ is called the resolvent of the matrix S at the point λ. By definition, the space $\mathbb{C}^2$ consists of all matrices of the form

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

where x_1 and x_2 are complex numbers. The space $\mathbb{C}^2$ becomes a 2-dimensional complex Hilbert space equipped with the inner product

$$\langle x|y\rangle := x^\dagger y, \qquad x, y \in \mathbb{C}^2.$$

Explicitly, $\langle x|y\rangle = (x_1^\dagger, x_2^\dagger)\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = x_1^\dagger y_1 + x_2^\dagger y_2$. If S is self-adjoint, then

$$\langle x|Sy\rangle = \langle Sx|y\rangle \qquad \text{for all} \quad x, y \in \mathbb{C}^2. \tag{9.19}$$

In fact, $x^\dagger Sy = x^\dagger S^\dagger y = (Sx)^\dagger y$. Conversely, the symmetry condition (9.19) implies that the matrix S is self-adjoint.

9.5 *Unitary matrices.* By definition, the matrix S is called unitary iff it is invertible and it preserves the inner product, that is,

$$\langle Sx|Sy\rangle = \langle x|y\rangle \qquad \text{for all} \quad x, y \in \mathbb{C}^2. \tag{9.20}$$

Show that the following hold true.

(i) The matrix S is unitary iff $S^\dagger S = I$. This is equivalent to

$$\begin{aligned} s_{11}s_{11}^\dagger + s_{21}s_{21}^\dagger = 1, \quad & s_{12}s_{12}^\dagger + s_{22}s_{22}^\dagger = 1, \\ s_{11}s_{12}^\dagger + s_{21}s_{22}^\dagger = 0, \quad & s_{12}s_{11}^\dagger + s_{22}s_{21}^\dagger = 0. \end{aligned} \tag{9.21}$$

This means that the two columns of the matrix S form an orthonormal basis of $\mathbb{C}^2$.

(ii) The matrix S is unitary iff $SS^\dagger = I$. This is equivalent to

$$\begin{aligned} s_{11}s_{11}^\dagger + s_{12}s_{12}^\dagger = 1, \quad & s_{21}s_{21}^\dagger + s_{22}s_{22}^\dagger = 1, \\ s_{11}s_{21}^\dagger + s_{12}s_{22}^\dagger = 0, \quad & s_{21}s_{11}^\dagger + s_{22}s_{12}^\dagger = 0. \end{aligned} \tag{9.22}$$

This means that the two rows of the matrix S form an orthonormal basis of $\mathbb{C}^2$.

(iii) The matrix S is unitary iff it is invertible and $S^{-1} = S^\dagger$.

(iv) The matrix S is unitary iff the adjoint matrix $S^\dagger$ is unitary.

(v) If the matrix S is unitary, then $|\det S| = 1$.

(vi) If the matrix S is unitary, then the eigenvalues of S lie on the unit circle. This tells us that there exist real numbers $\delta_1, \delta_2 \in]-\pi, \pi]$ with

$$\lambda_j = \mathrm{e}^{\mathrm{i}\delta_j}, \qquad j = 1, 2.$$

In quantum physics, the numbers δ_1, δ_2 are called the phases of the unitary matrix S.

(vii) If the matrix S is self-adjoint, then its eigenvalues lie on the real axis.

(viii) Cayley transform. If the complex (2×2)-matrix H is self-adjoint, then the matrix

$$S := (H - \mathrm{i}I)(H + \mathrm{i}I)^{-1} \tag{9.23}$$

is well-defined and unitary. Conversely,

$$H = \mathrm{i}(I - S)^{-1}(I + S).$$

The matrix S is called the Cayley transform of H.[19] The Cayley transform $H \mapsto S$ is a bijective map from the set of self-adjoint matrices onto the set of unitary matrices which do not have the eigenvalue $\lambda = 1$.

Solution: Ad (i)–(v). The condition (9.20) is equivalent to

$$(Sx)^\dagger(Sy) = x^\dagger S^\dagger S y = x^\dagger y \qquad \text{for all} \quad x, y \in \mathbb{C}^2.$$

Thus, the matrix S is unitary iff $S^\dagger S = I$ and S is invertible. However, the relation $S^\dagger S = I$ implies the invertibility of S. In fact,

$$1 = \det I = \det(S^\dagger S) = \det S^\dagger \cdot \det S = (\det S)^\dagger \det S = |\det S|^2.$$

Finally, note that (9.21) is equivalent to the equation

$$S^\dagger S = \begin{pmatrix} s_{11}^\dagger & s_{21}^\dagger \\ s_{12}^\dagger & s_{22}^\dagger \end{pmatrix} \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Ad (vi). If $Sx = \lambda x$ with $x \neq 0$, then the relation (9.20) tells us that $|\lambda|^2 = 1$. In fact,

$$\langle x|x\rangle = \langle Sx|Sx\rangle = \lambda^\dagger \lambda \langle x|x\rangle = |\lambda|^2 \langle x|x\rangle.$$

Ad (vii). If $Sx = \lambda x$ with $x \neq 0$, then

$$\lambda\langle x|x\rangle = \langle x|Sx\rangle = \langle Sx|x\rangle^\dagger = \lambda^\dagger \langle x|x\rangle.$$

Ad (viii). Since the eigenvalues of the self-adjoint matrix H are real, we have $\pm\mathrm{i} \in \varrho(H)$. Thus, the matrices $H \pm \mathrm{i}I$ are invertible. The commutativity relation

$$(H - \mathrm{i}I)(H + \mathrm{i}I) = (H + \mathrm{i}I)(H - \mathrm{i}I)$$

implies $(H+\mathrm{i}I)^{-1}(H-\mathrm{i}I) = (H-\mathrm{i}I)(H+\mathrm{i}I)^{-1}$. It follows from $(AB)^\dagger = B^\dagger A^\dagger$ and $(A^{-1})^\dagger = (A^\dagger)^{-1}$ along with $(H \pm \mathrm{i}I)^\dagger = H^\dagger \mp \mathrm{i}I^\dagger = H \mp \mathrm{i}I$ that

$$S^\dagger = (H - \mathrm{i}I)^{-1}(H + \mathrm{i}I).$$

Hence $S^\dagger S = (H - \mathrm{i}I)^{-1}(H + \mathrm{i}I)(H + \mathrm{i}I)^{-1}(H - \mathrm{i}I) = I$. Let us now study the inverse Cayley transform. It follows from (9.23) that $S(H + \mathrm{i}I) = H - \mathrm{i}I$. Hence

$$(I - S)H = \mathrm{i}(I + S).$$

It remains to show that $\lambda = 1$ is not an eigenvalue of the Cayley transform S. In fact, suppose that $Sx = x$. By (9.23),

$$(H + \mathrm{i}I)x = (H - \mathrm{i}I)x.$$

Hence $x = 0$. This implies $1 \notin \sigma(H)$. Consequently, the matrix $I - S$ is invertible. This yields $H = \mathrm{i}(I - S)^{-1}(I + S)$.

[19] Cayley (1821–1895)

9.6 *Factorization of unitary matrices.* If the complex (2×2)-matrix $S = (s_{ij})$ is unitary and $s_{11} \neq 0$, then

$$S = \begin{pmatrix} 1 & s_{12} \\ 0 & s_{22} \end{pmatrix} \begin{pmatrix} s_{11}^\dagger & 0 \\ s_{12}^\dagger & 1 \end{pmatrix}^{-1}.$$

It turns out that this factorization of the S-matrix lurks behind the factorization $\mathsf{S} = W_{\text{out}}^{-1} W_{\text{in}}$ of the scattering operator.
Solution: It follows from the unitarity condition

$$SS^\dagger = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} \begin{pmatrix} s_{11}^\dagger & s_{21}^\dagger \\ s_{12}^\dagger & s_{22}^\dagger \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

that

$$\begin{pmatrix} 1 & s_{12} \\ 0 & s_{22} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -s_{12}^\dagger & s_{11}^\dagger \end{pmatrix} \frac{1}{s_{11}^\dagger} = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix}.$$

9.7 *The S-matrix.* Prove that

$$\begin{pmatrix} \varphi_{-k}(x) \\ \varphi_k(x) \end{pmatrix} = \hat{S}(k) \begin{pmatrix} \varphi_k(x)^\dagger \\ \varphi_{-k}(x)^\dagger \end{pmatrix} \qquad \text{for all} \quad x \in \mathbb{R}.$$

Solution: For the S-matrix $\hat{S}(k) = (s_{ij})$, we have

$$s_{11} = \tau_{\leftarrow}(k), \quad s_{12} = \varrho_{\rightarrow}(k), \quad s_{21} = \varrho_{\leftarrow}(k), \quad s_{22} = \tau_{\rightarrow}(k),$$

and $s_{11} = s_{22}$. For the square-well potential, we also get $s_{12} = s_{21}$. However, this symmetry property is not needed for our argument. For all real numbers x with $x > r$,

$$\begin{pmatrix} \varphi_{-k}(x) \\ \varphi_k(x) \end{pmatrix} = \begin{pmatrix} 1 & \varrho_{\rightarrow}(k) \\ 0 & \tau_{\rightarrow}(k) \end{pmatrix} \begin{pmatrix} \mathrm{e}^{-\mathrm{i}kx} \\ \mathrm{e}^{\mathrm{i}kx} \end{pmatrix}, \qquad \begin{pmatrix} \varphi_k(x)^\dagger \\ \varphi_{-k}(x)^\dagger \end{pmatrix} = \begin{pmatrix} \tau_{\rightarrow}(k)^\dagger & 0 \\ \varrho_{\rightarrow}(k)^\dagger & 1 \end{pmatrix} \begin{pmatrix} \mathrm{e}^{-\mathrm{i}kx} \\ \mathrm{e}^{\mathrm{i}kx} \end{pmatrix}.$$

Finally, use Problem 9.6 and the solution principle (P) on page 718.

9.8 *Matrix functions.* The space X of all complex (2×2)-matrices S is a complex linear space of dimension four with the basis

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

For all matrices $S, T \in X$, define

$$\langle S|T\rangle := \operatorname{tr}(S^\dagger T), \tag{9.24}$$

and $||S|| := \sqrt{\langle S|S\rangle}$. Furthermore, define

$$\mathrm{e}^S := \sum_{n=0}^{\infty} \frac{S^n}{n!} = I + S + \frac{S^2}{2} + \ldots, \tag{9.25}$$

and

$$\ln(I+S) := \sum_{n=0}^{\infty} (-1)^n \frac{S^n}{n} = I - S + \frac{S^2}{2} + \ldots \tag{9.26}$$

Show the following:

(a) The space X is a Hilbert space equipped with the inner product (9.24).
(b) $||ST|| \leq ||S|| \cdot ||T||$ for all $S, T \in X$.
(c) The series (9.25) converges in the Hilbert space X, and hence the corresponding series for each matrix element converges. That is,

$$\mathrm{e}^S = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = I + S + \frac{S^2}{2!} + \ldots = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} + \ldots$$

Hence $a_{11} = 1 + s_{11} + \ldots$, and so on.
(d) If $||S|| < 1$, then the series (9.26) converges in the Hilbert space X, and hence the corresponding series for each matrix element converges.

Hint: For proving (c), use (b) and the majorant criterion for classical power series expansions. See Zeidler (1995a), Sec. 1.23, quoted on page 1049.

Matrix calculus shows that the elementary results for (2×2)-matrices summarized above can be generalized to complex $(n \times n)$-matrices with $n = 2, 3, \ldots$

9.9 *Operator-valued approximate delta function.* Let $H : D(H) \to X$ be a self-adjoint operator defined on the dense subset $D(H)$ of the complex separable infinite-dimensional Hilbert space X. Prove that

$$w - \lim_{R \to +\infty} \int_{-R}^{R} \delta_\varepsilon(EI - H) dE = I \qquad \text{for all} \quad \varepsilon > 0$$

where $\delta_\varepsilon(EI - H) := \frac{\varepsilon}{\pi}(H - EI + \mathrm{i}\varepsilon I)^{-1}(H - EI - \mathrm{i}\varepsilon I)^{-1}$.
Hint: See I. Sigal, Scattering Theory for Many-Body Quantum Mechanical Systems, p. 8, Springer, New York, 1983.

9.10 *The Abel limit.* Prove the Abel theorem (9.9).
Solution: The key to the proof is the simple formula

$$\int_0^\infty \varepsilon \mathrm{e}^{-\varepsilon t} f(t) dt = 1.$$

Let $a := \lim_{t \to +\infty} f(t)$. Then $\sup_{t \geq 0} ||f(t)|| < \infty$. For all $\varepsilon > 0$, set

$$J_\varepsilon := \int_0^\infty \varepsilon \mathrm{e}^{-\varepsilon t} f(t) dt - a = \int_0^T \varepsilon \mathrm{e}^{-\varepsilon t} \, (f(t) - a) dt + \int_T^\infty \varepsilon \mathrm{e}^{-\varepsilon t} \, (f(t) - a) dt.$$

We have to show that $\lim_{\varepsilon \to +0} J_\varepsilon = 0$. In fact,

$$||J_\varepsilon|| \leq \varepsilon T \cdot (||a|| + \sup_{t \geq 0} ||f(t)||) + \sup_{t \geq T} ||f(t) - a||.$$

Finally, choose $T := 1/\sqrt{\varepsilon}$.

9.11 *The time-independent approach to wave operators.* Use Problems 9.9 and 9.10 in order to prove Theorem 9.4.
Hint: As in Problem 9.9, see Sigal (1983), p. 16.

10. Creation and Annihilation Operators

> Use tensor products and Fock spaces for describing mathematically the states of many-particle systems.
>
> Folklore

We want to study a mathematical formalism which describes creation and annihilation operators for many-particle systems. We have to distinguish between

- bosons (particles with integer spin like photons, gluons, vector bosons, and gravitons) and
- fermions (particles with half-integer spin like electrons, neutrinos, and quarks).

The point is that the possible number of identical bosons being in the same physical state is unlimited. In contrast to this, the behavior of fermions is governed by the Pauli exclusion principle. This principle tells us that:

Two identical fermions cannot be in the same physical state.

Furthermore, we have the following general principle of indistinguishability for both bosons and fermions:

It is impossible to distinguish between n identical particles.

Roughly speaking, in contrast to planets, elementary particles do not possess any individuality. Fock spaces were introduced by Vladimir Fock (1898–1974) in 1932.[1]

10.1 The Bosonic Fock Space

The elements of the bosonic Fock space X are infinite tuples of physical fields. To display the main idea, we restrict ourselves to the prototype of complex-valued fields[2]

$$\psi : \mathbb{R}^4 \to \mathbb{C}.$$

[1] V. Fock, Configuration space and second quantization, Z. Phys. **75** (1932), 622–647 (in German). See also P. Jordan and E. Wigner, On the Pauli equivalence principle, Z. Phys. **47** (1928), 631–658 (in German).

[2] Important generalizations will be studied later on in connection with quantum electrodynamics. This refers to fields

$$\psi : \mathbb{R}^4 \to \mathbb{C}^m$$

with m components. For describing photons, it will be necessary to pass from Hilbert spaces to indefinite inner product spaces.

The point $x = (x^0, x^1, x^2, x^3)$ describes space and time in an inertial system, that is, the position coordinates x^1, x^2, x^3 are right-handed Cartesian coordinates. Moreover, we introduce the time-like coordinate

$$x^0 := ct$$

where t is time, and c is the velocity of light in a vacuum. The position vector $\mathbf{x} = x^1\mathbf{i} + x^2\mathbf{j} + x^3\mathbf{k}$ refers to the right-handed orthonormal system $\mathbf{i}, \mathbf{j}, \mathbf{k}$.

The Hilbert space $L_2(\mathbb{R}^{4n})$. Choose $n = 1, 2, \ldots$ Let us introduce the inner product

$$\boxed{\langle\psi|\varphi\rangle_n := \int_{\mathbb{R}^{4n}} \psi(x_1, \ldots, x_n)^\dagger \varphi(x_1, \ldots, x_n)\, d^4x_1 \cdots d^4x_n}$$

along with the corresponding norm $||\psi||_n := \sqrt{\langle\psi|\psi\rangle_n}$. Hence

$$||\psi||_n^2 = \int_{\mathbb{R}^{4n}} |\psi(x_1, \ldots, x_n)|^2\, d^4x_1 \cdots d^4x_n.$$

Here, the arguments $x_1, \ldots, x_n$ live in $\mathbb{R}^4$. By definition, the space $L_2(\mathbb{R}^{4n})$ consists of all the functions[3]

$$\psi : \mathbb{R}^{4n} \to \mathbb{C}$$

with $||\psi||_n < \infty$. The space $L_2(\mathbb{R}^{4n})$ becomes a complex Hilbert space equipped with the inner product $\langle\psi|\varphi\rangle_n$. We have the direct sum decomposition

$$\boxed{L_2(\mathbb{R}^{4n}) = L_{2,\mathrm{sym}}(\mathbb{R}^{4n}) \oplus L_{2,\mathrm{antisym}}(\mathbb{R}^{4n})}$$

where the space $L_{2,\mathrm{sym}}(\mathbb{R}^{4n})$ (resp. $L_{2,\ \mathrm{antisym}}(\mathbb{R}^{4n})$) contains all the functions

$$\psi = \psi(x_1, \ldots, x_n), \qquad x_1, \ldots, x_n \in \mathbb{R}^4$$

from $L_2(\mathbb{R}^{4n})$ which are symmetric (resp. antisymmetric) with respect to the arguments $x_1, \ldots, x_n$.

Tensor products of fields. Suppose that the two functions $\psi, \varphi : \mathbb{R}^4 \to \mathbb{C}$ live in the Hilbert space $L_2(\mathbb{R}^4)$. Set

$$(\psi \otimes \varphi)(x_1, x_2) := \psi(x_1)\varphi(x_2) \qquad \text{for all} \quad x_1, x_2 \in \mathbb{R}^4.$$

Then, the tensor product $\psi \otimes \varphi$ lives in the Hilbert space $L_2(\mathbb{R}^8)$. Introducing the symmetrization

$$\mathrm{sym}(\psi \otimes \varphi) := \tfrac{1}{2}(\psi \otimes \varphi + \varphi \otimes \psi)$$

and the antisymmetrization

$$\mathrm{antisym}(\psi \otimes \varphi) := \tfrac{1}{2}(\psi \otimes \varphi - \varphi \otimes \psi) = \tfrac{1}{2}(\psi \wedge \varphi)$$

of the tensor product $\psi \otimes \varphi$, we get the decomposition

[3] We tacitly assume that the functions ψ are measurable with respect to the Lebesgue measure on $\mathbb{R}^{4n}$. This only excludes highly pathological functions having extremely wild discontinuities. In addition, observe that two functions $\psi, \varphi : \mathbb{R}^{4n} \to \mathbb{C}$ are identified with each other if they only differ on a subset of $\mathbb{R}^{4n}$ which has the $4n$-dimensional Lebesgue measure zero.

$$\psi \otimes \varphi = \mathrm{sym}(\psi \otimes \varphi) + \mathrm{antisym}(\psi \otimes \varphi)$$

with $\mathrm{sym}(\psi \otimes \varphi) \in L_{2,\mathrm{sym}}(\mathbb{R}^8)$ and $\mathrm{antisym}(\psi \otimes \varphi) \in L_{2,\mathrm{antisym}}(\mathbb{R}^8)$. Moreover, the set of all the finite linear combinations

$$\alpha_1 \psi_1 \otimes \varphi_1 + \ldots + \alpha_n \psi_n \otimes \varphi_n, \qquad n = 1, 2, \ldots$$

with $\psi_1, \varphi_1, \ldots \in L_2(\mathbb{R}^4)$ and complex numbers $\alpha_1, \alpha_2, \ldots$ is dense in the Hilbert space $L_2(\mathbb{R}^8)$. We write

$$L_2(\mathbb{R}^8) = L_2(\mathbb{R}^4) \otimes L_2(\mathbb{R}^4) = L_2(\mathbb{R}^4)^{\otimes 2}.$$

This is the prototype of the tensor product of two Hilbert spaces. We also write

$$L_{2,\mathrm{antisym}}(\mathbb{R}^8) = L_2(\mathbb{R}^4) \wedge L_2(\mathbb{R}^4) = L_2(\mathbb{R}^4)^{\wedge 2}.$$

For the inner product, we have[4]

$$\boxed{\langle \varphi_1 \otimes \varphi_2 | \psi_1 \otimes \psi_2 \rangle = \langle \varphi_1 | \psi_1 \rangle \langle \varphi_2 | \psi_2 \rangle.}$$

Finally, for two given linear operators $A, B : L_2(\mathbb{R}^4) \to L_2(\mathbb{R}^4)$, we define the tensor product $A \otimes B$ by setting

$$\boxed{(A \otimes B)(\psi \otimes \varphi) := A\psi \otimes \varphi + \psi \otimes B\varphi}$$

for all $\varphi, \psi \in L_2(\mathbb{R}^4)$.

Definition of the bosonic Fock space. The bosonic Fock space X is defined to be the direct sum

$$\boxed{X = \bigoplus_{n=0}^{\infty} X_n}$$

of the complex Hilbert spaces $X_0 := \mathbb{C}$, $X_1 := L_2(\mathbb{R}^4)$, and

$$X_n := L_{2,\mathrm{sym}}(\mathbb{R}^{4n}), \qquad n = 2, 3, \ldots$$

Explicitly, this means the following. The bosonic Fock space X consists of all the infinite sequences

$$(\psi_0, \psi_1, \psi_2, \ldots)$$

with $\sum_{n=1}^{\infty} ||\psi_n||_n^2 < \infty$. The function ψ_n is called an n-particle function. Here,

- ψ_0 is an arbitrary complex number,
- the one-particle functions $\psi_1 : \mathbb{R}^4 \to \mathbb{C}$ are of the form $\psi_1 = \psi_1(x)$ with $x \in \mathbb{R}^4$, and they live in the complex Hilbert space $L_2(\mathbb{R}^4)$,
- the n-particle functions $\psi_n : \mathbb{R}^{4n} \to \mathbb{C}$,

$$\psi_n = \psi_n(x_1, x_2, \ldots, x_n), \qquad n = 2, 3, \ldots,$$

are symmetric with respect to the n arguments $x_1, x_2, ..., x_n \in \mathbb{R}^4$, and they live in the complex Hilbert space $L_{2,\mathrm{sym}}(\mathbb{R}^{4n})$.

[4] Explicitly, the integral $\int (\varphi_1(x_1)\varphi_2(x_2))^\dagger \psi_1(x_1)\psi_2(x_2) d^4x_1 d^4x_2$ is equal to

$$\int \varphi_1(x_1)^\dagger \psi_1(x_1) d^4x_1 \int \varphi_2(x_2)^\dagger \psi_2(x_2) d^4x_2.$$

The symmetry of the n-particle functions ψ_n reflects the principle of indistinguishability for n bosons. The bosonic Fock space X is an infinite-dimensional complex Hilbert space equipped with the inner product

$$\boxed{\langle\psi|\varphi\rangle := \psi_0^\dagger\varphi_0 + \sum_{n=1}^{\infty}\langle\psi_n|\varphi_n\rangle_n.}$$

10.1.1 The Particle Number Operator

For $\psi := (\psi_0, \psi_1, \psi_2, ...)$, set

$$\boxed{N\psi := (0, \psi_1, 2\psi_2, \ldots, n\psi_n, \ldots).}$$

More precisely, the linear operator $N : D(N) \to X$ is defined for all states $\psi \in X$ of the bosonic Fock space X with

$$\sum_{n=1}^{\infty} n^2 ||\psi_n||_n^2 < \infty.$$

For example, choose $\psi_n \in X_n$ for fixed index n with $||\psi_n||_n = 1$. Define

$$\Psi_n := (0, \ldots, 0, \psi_n, 0, 0, \ldots)$$

where ψ_n stands at the nth place. Then, Ψ_n is a normalized state in the bosonic Fock space X with

$$N\Psi_n = n\Psi_n.$$

In terms of physics, the state Ψ_n describes n bosons.

10.1.2 The Ground State

The state $|0\rangle := (1, 0, 0, \ldots)$ is a normalized state in the bosonic Fock space X with

$$\boxed{N|0\rangle = 0.}$$

The state $|0\rangle$ is called the normalized vacuum state (or briefly the vacuum), since the number of bosons is equal to zero in this state.

Dense linear subspace X_{fin} of the bosonic Fock space X. Let X_{fin} denote the set of all the states

$$\psi = (\psi_0, \psi_1, \ldots)$$

in the bosonic Fock space X for which at most a finite number of the functions $\psi_1, \psi_2, \ldots$ does not vanish identically. For example, the state

$$\psi = (\psi_0, \psi_1, \ldots, \psi_4, 0, 0, \ldots)$$

with $\psi_0 \in \mathbb{C}$ and $\psi_j \in L_2(\mathbb{R}^4)$ for $j = 1, ..., 4$ lies in the subspace X_{fin}.

Composition of particle functions. We are given the one-particle function $f \in L_2(\mathbb{R}^4)$. For each n-particle function

$$\psi_n \in L_{2,\text{sym}}(\mathbb{R}^{4n}),$$

the symmetrization of the tensor product $f \otimes \psi_n$ yields the $(n+1)$-particle function[5]

$$\boxed{\varrho_{n+1} := \sqrt{n+1} \cdot \mathrm{sym}(f \otimes \psi_n).} \tag{10.1}$$

Intuitively, this is the composition of the one-particle state f with the n-particle state ψ_n. Explicitly,

$$\varrho_{n+1}(x_1, \ldots, x_{n+1}) = \frac{\sqrt{n+1}}{(n+1)!} \sum_{\pi} f(x_1) \psi_n(x_2, \ldots, x_{n+1})$$

where we sum over all permutations π of the arguments $x_1, \ldots, x_{n+1}$.

Creation operator $a^+(f)$. Fix again the one-particle function $f \in L_2(\mathbb{R}^4)$. We want to construct a linear operator

$$a^+(f) : X_{\text{fin}} \to X$$

which describes the creation of particles. Explicitly, for each sequence

$$\psi := (\psi_0, \psi_1, \psi_2, \ldots)$$

in the linear subspace X_{fin} of the bosonic Fock space X, we define

$$\boxed{a^+(f)\psi := (0, \rho_1, \rho_2, \ldots)} \tag{10.2}$$

where the functions $\varrho_1, \varrho_2, \ldots$ are given by (10.1). In particular,

$$\varrho_1(x_1) := f(x_1)\psi_0 \qquad \varrho_2(x_1, x_2) = \frac{f(x_1)\psi_1(x_2) + f(x_2)\psi_1(x_1)}{\sqrt{2}}.$$

Annihilation operator $a^-(f)$. We want to construct a linear operator

$$a^-(f) : X_{\text{fin}} \to X$$

which is formally adjoint to the creation operator $a^+(f)$, that is,

$$\boxed{\langle a^-(f)\varphi | \varphi \rangle = \langle \varphi | a^+(f)\psi \rangle \qquad \text{for all} \quad \varphi, \psi \in X_{\text{fin}}.}$$

In other words, we want to get $a^-(f) = (a^+(f))^\dagger$ on X_{fin}. To this end, for each sequence $\varphi := (\varphi_0, \varphi_1, \varphi_2, \ldots)$ in X_{fin} we define

$$\boxed{a^-(f)\varphi := (\chi_0, \chi_1, \chi_2, \ldots)} \tag{10.3}$$

along with

$$\chi_n(x_1, \ldots, x_n) := \sqrt{n+1} \int_{\mathbb{R}^4} f(x)^\dagger \varphi_{n+1}(x, x_1, \ldots, x_n) d^4x$$

for all indices $n = 0, 1, 2, \ldots$ In particular, we have

$$\chi_0 := \int_{\mathbb{R}^4} f(x)^\dagger \varphi_1(x) d^4x, \qquad \chi_1(x_1) := \sqrt{2} \int_{\mathbb{R}^4} f(x)^\dagger \varphi_2(x, x_1) d^4x.$$

[5] It is convenient to add the normalization factor $\sqrt{n+1}$ (see Theorem 10.1(iii) below).

For the vacuum state, we get

$$\boxed{a^-(f)|0\rangle = 0 \qquad \text{for all} \quad f \in L_2(\mathbb{R}^4).}$$

Fundamental commutation relations. Fix $f, g \in L_2(\mathbb{R}^4)$. Recall that $[A,B]_- := AB - BA$. In particular, $[A,B]_- = 0$ is equivalent to $AB = BA$.

Theorem 10.1 *For all states ψ, φ in the linear subspace $X_{\rm fin}$ of the bosonic Fock space X, the following relations hold:*
(i) *Creation operators:* $[a^+(f), a^+(g)]_- \psi = 0$.
(ii) *Annihilation operators:* $[a^-(f), a^-(g)]_- \psi = 0$.
(iii) *Creation and annihilation operators:*

$$[a^-(f), a^+(g)]_- \psi = \langle f|g\rangle_1 \psi. \tag{10.4}$$

(iv) *Duality:* $\langle a^-(f)\varphi|\psi\rangle = \langle \varphi|a^+(f)\psi\rangle$.

Proof. To display the main ideas of the proof, we restrict ourselves to some special cases. Then the proof of the general case proceeds similarly by induction.

Ad (i)–(iii). Choose the functions $f, g \in L_2(\mathbb{R}^4)$ with $||f||_1 = ||g||_1 = 1$. Since $a^+(g)|0\rangle = (0, g, 0, \ldots)$, we have

$$a^+(f)a^+(g)|0\rangle = \frac{1}{\sqrt{2}}\,(0, 0, f(x_1)g(x_2) + f(x_2)g(x_1), 0, \ldots).$$

Using symmetry, $a^+(f)a^+(g)|0\rangle - a^+(g)a^+(f)|0\rangle = 0$.

By $a^-(f)|0\rangle = a^-(g)|0\rangle = 0$, we get $a^-(f)a^-(g)|0\rangle - a^-(g)a^-(f)|0\rangle = 0$. Finally, it follows from

$$a^-(f)a^+(g)|0\rangle = \left(\int_{\mathbb{R}^4} f(x)^\dagger g(x) d^4x, 0, 0, \ldots\right)$$

and $a^+(g)a^-(f)|0\rangle = 0$ that

$$a^-(f)a^+(g)|0\rangle - a^+(g)a^-(f)|0\rangle = \langle f|g\rangle_1 |0\rangle.$$

Ad (iv). Choosing the two special states

$$\psi := (0, \psi_1, 0, \ldots), \qquad \varphi := (0, 0, \varphi_2, 0, \ldots),$$

we obtain

$$a^+(f)\psi = \frac{1}{\sqrt{2}}\,(0, 0, f(x_1)\psi_1(x_2) + f(x_2)\psi_1(x_1), 0, \ldots)$$

and

$$a^-(f)\varphi = (0, \sqrt{2}\int_{\mathbb{R}^4} f^\dagger(x_1)\varphi_2(x_1, x_2)\; d^4x_1, 0, \ldots).$$

Therefore, the inner product $\langle \varphi|a^+(f)\psi\rangle$ is equal to

$$\frac{1}{\sqrt{2}}\int_{\mathbb{R}^8} \varphi_2(x_1, x_2)^\dagger \{f(x_1)\psi_1(x_2) + f(x_2)\psi_1(x_1)\}\; d^4x_1 d^4x_2.$$

Since the function φ_2 is symmetric,

$$\langle \varphi|a^+(f)\psi\rangle = \sqrt{2}\int_{\mathbb{R}^8} \varphi_2(x_1, x_2)^\dagger f(x_1)\psi_1(x_2)\; d^4x_1 d^4x_2.$$

Furthermore,

$$\langle a^-(f)\varphi|\psi\rangle = \sqrt{2}\left(\int_{\mathbb{R}^8} f^\dagger(x_1)\varphi_2(x_1,x_2)\,d^4x_1\right)^\dagger \psi_1(x_2)\,d^4x_2.$$

Consequently, $\langle a^-(f)\varphi|\psi\rangle = \langle\varphi|a^+(f)\psi\rangle$. □

Physical interpretation. Choose one-particle functions $f_1,\dots,f_s$ in the space $L_2(\mathbb{R}^4)$ such that $||f_j||_1 = 1$ for $j = 1,\dots,s$. Set

$$\boxed{\psi := a^+(f_1)a^+(f_2)\cdots a^+(f_s)|0\rangle.} \tag{10.5}$$

This is a state in the bosonic Fock space X. Observe that

$$a^+(f_j)|0\rangle = (0, f_j, 0, \dots)$$

and $Na^+(f_j)|0\rangle = a^+(f_j)|0\rangle$. We say that

- the function f_j represents a normalized one-particle state of a boson, and
- the operator $a^+(f_j)$ generates the normalized one-particle state $a^+(f_j)|0\rangle$ from the vacuum $|0\rangle$.

Note that the state ψ from (10.5) has the form $(\psi_0, \psi_1, \dots)$ where $\psi_j = 0$ if $j \neq s$. Hence

$$\boxed{N\psi = s\psi.}$$

This tells us that if $\psi \neq 0$, then the state ψ from (10.5) represents s bosons being in one-particle states corresponding to $f_1,\dots,f_s$. Because of Theorem 10.1, the state ψ from (10.5) is invariant under permutations of $f_1,\dots,f_s$. This reflects the principle of indistinguishability for s bosons.

Important special case. Consider a system of functions

$$f_1, f_2, f_3, \dots$$

from $\mathbb{R}^4$ to $\mathbb{C}$ which forms an orthonormal system in the Hilbert space $L_2(\mathbb{R}^4)$, that is, $\langle f_k|f_l\rangle_1 = \delta_{kl}$ for all $k,l = 1,2,\dots$ Define

$$a_j^+ := a^+(f_j), \qquad a_j^- := a^-(f_j), \qquad j = 1,2,...$$

For all $\psi \in X_{\text{fin}}$ with $j,k = 1,2,\dots$, we then have the following commutation relations:

$$\boxed{\begin{aligned}\left[a_j^+, a_k^+\right]_-\ \psi = \left[a_j^-, a_k^-\right]_-\ \psi = 0,\\ \left[a_j^-, a_k^+\right]_-\ \psi = \delta_{jk}\psi.\end{aligned}} \tag{10.6}$$

This follows immediately from Theorem 10.1. Moreover, for $j,k = 1,2,\dots$, the following states are normalized in the bosonic Fock space X:

(i) $a_j^+|0\rangle$;
(ii) $a_j^+a_k^+|0\rangle$ if $j \neq k$;
(iii) $\frac{1}{\sqrt{2}}(a_j^+)^2|0\rangle$.

Proof. Ad (i). Note that $a_j^+|0\rangle = (0, f_j, 0, \ldots)$.

Ad (ii). We have $a_j^+ a_k^+|0\rangle = (0, 0, \varphi_2, 0, \ldots)$ with

$$\varphi_2(x_1, x_2) := \frac{f_j(x_1)f_k(x_2) + f_j(x_2)f_k(x_1)}{\sqrt{2}}.$$

Since $j \neq k$, $\int_{\mathbb{R}^4} f_j(x)^\dagger f_k(x) d^4x = 0$. Hence $\langle\varphi_2|\varphi_2\rangle_2$ is equal to

$$\int_{\mathbb{R}^8} \varphi_2(x,y)^\dagger \varphi_2(x,y) d^4x d^4y = \int_{\mathbb{R}^4} |f_j(x)|^2 d^4x \int_{\mathbb{R}^4} |f_k(y)|^2 d^4y = 1.$$

Ad (iii). Since $j = k$, we obtain

$$\langle\varphi_2|\varphi_2\rangle_2 = 2\int_{\mathbb{R}^4} |f_j(x)|^2 d^4x \int_{\mathbb{R}^4} |f_j(y)|^2 d^4y = 2.$$

This argument finishes the proof. □

More generally, if $1 \leq j_1 < \ldots < j_k$ and $m_1, \ldots, m_k = 1, 2, \ldots$, then

$$\boxed{\frac{(a_{j_1}^+)^{m_1}}{\sqrt{m_1!}} \frac{(a_{j_2}^+)^{m_2}}{\sqrt{m_2!}} \cdots \frac{(a_{j_k}^+)^{m_k}}{\sqrt{m_k!}} |0\rangle}$$

is a normalized state in the bosonic Fock space X. States of this form are basic

- in the scattering theory for elementary particles,
- in the theory of many-particle systems in solid state physics, and in
- quantum optics (laser beams).

The rigorous language of operator-valued distributions in quantum field theory. The space of linear operators

$$A : X_{\text{fin}} \to X$$

is denoted by $L(X_{\text{fin}}, X)$. Set[6]

$$\boxed{A^+(f) := a^+(f) \qquad \text{for all} \quad f \in \mathcal{D}(\mathbb{R}^4).}$$

Then, $A^+ : \mathcal{D}(\mathbb{R}^4) \to L(X_{\text{fin}}, X)$ is a linear map from the space $\mathcal{D}(\mathbb{R}^4)$ of test functions to the operator space $L(X_{\text{fin}}, X)$. That is,

$$A^+(\alpha f + \beta g) = \alpha A^+(f) + \beta A^+(g)$$

for all $f, g \in \mathcal{D}(\mathbb{R}^4)$ and all complex numbers α and β. We call A^+ a distribution with values in the operator space $L(X_{\text{fin}}, X)$. Similarly, we define

$$A^-(f) := a^-(f) \qquad \text{for all} \quad f \in \mathcal{D}(\mathbb{R}^4).$$

The map $A^- : \mathcal{D}(\mathbb{R}^4) \to L(X_{\text{fin}}, X)$ is antilinear, that is,

$$A^-(\alpha f + \beta g) = \alpha^\dagger A^-(f) + \beta^\dagger A^-(g)$$

[6] Recall that the space $\mathcal{D}(\mathbb{R}^4)$ consists of all the smooth functions $f : \mathbb{R}^4 \to \mathbb{C}$ which vanish outside some ball, which depends on f. Such functions are called test functions. The space $\mathcal{D}(\mathbb{R}^4)$ is also denoted by $C_0^\infty(\mathbb{R}^4)$.

for all $f, g \in \mathcal{D}(\mathbb{R}^4)$ and all complex numbers α and β. We call A^- an antidistribution with values in the operator space $L(X_{\text{fin}}, X)$.

The formal language of physicists. Physicists introduce the formal creation operators $a^-(x)$ and the formal annihilation operators $a^-(x)$ along with the formal commutation relations

$$\begin{aligned} \big[a^+(x), a^+(y)\big]_- &= \big[a^-(x), a^-(y)\big]_- = 0, \\ \big[a^-(x), a^+(y)\big]_- &= \delta(x-y)I \end{aligned} \tag{10.7}$$

and the duality relations

$$(a^+(x))^\dagger = a^-(x), \qquad (a^-(x))^\dagger = a^+(x).$$

These relations are assumed to be valid for all $x, y \in \mathbb{R}^4$. Intuitively, the operator $a^+(x)$ describes the creation of a boson at the given space-time point $x = (ct, \mathbf{x})$. This corresponds to the creation of a boson at the position $\mathbf{x}$ at time t. Similarly, the operator $a^-(x)$ describes the annihilation of a boson at the position $\mathbf{x}$ at time t. Furthermore, we formally write

$$\boxed{a^+(f) := \int_{\mathbb{R}^4} f(x) a^+(x) d^4x,}$$

and $a^-(f) := \int_{\mathbb{R}^4} f(x)^\dagger a^-(x) dx$ along with

$$a^-(f)a^+(g) := \int_{\mathbb{R}^8} f(x)^\dagger g(y) a^-(x) a^+(y) d^4x d^4y,$$

and so on. Mnemonically, this yields the rigorous approach introduced above. For example,

$$\begin{aligned} a^-(f)a^+(g) - a^+(g)a^-(f) &= \int_{\mathbb{R}^8} f(x)^\dagger g(y) \left[a^-(x), a^+(y)\right]_- d^4x d^4y \\ &= \int_{\mathbb{R}^8} f(x) g(y) \delta(x-y) I \cdot d^4x d^4y = \left(\int_{\mathbb{R}^4} f(x)^\dagger g(x) d^4x \right) I. \end{aligned}$$

Furthermore,

$$a^+(f)a^+(g) - a^+(g)a^+(f) = \int_{\mathbb{R}^8} f(x)^\dagger g(y) [a^+(x), a^+(y)]_- \, d^4x d^4y = 0.$$

Similarly, $a^-(f)a^-(g) - a^-(g)a^-(f) = 0$. Finally, we formally get

$$(a^+(f))^\dagger = \left(\int_{\mathbb{R}^4} f(x) a^+(x) d^4x \right)^\dagger = \int_{\mathbb{R}^4} f(x)^\dagger a^-(x) d^4x = a^-(f).$$

10.2 The Fermionic Fock Space and the Pauli Principle

In contrast to the bosonic Fock space, the components $\psi_2, \psi_3, \ldots$ of a state in the fermionic Fock space are not symmetric, but antisymmetric functions. As we will see below, this forces the Pauli exclusion principle principle. Let us consider the

prototype of a fermionic Fock space based on the one-particle function $\psi : \mathbb{R}^4 \to \mathbb{C}$. The fermionic Fock space Y is defined to be the direct sum

$$\boxed{Y = \bigoplus_{n=0}^{\infty} Y_n}$$

of the complex Hilbert spaces $Y_0 := \mathbb{C}$, $Y_1 := L_2(\mathbb{R}^4)$, and

$$Y_n := L_{2,\,\text{antisym}}(\mathbb{R}^{4n}), \qquad n = 2, 3, \ldots$$

Explicitly, the fermionic Fock space Y consists of all the infinite sequences

$$(\psi_0, \psi_1, \psi_2, \ldots)$$

with $\sum_{n=1}^{\infty} ||\psi_n||_n^2 < \infty$. More precisely,

- ψ_0 is an arbitrary complex number,
- the one-particle functions $\psi_1 : \mathbb{R}^4 \to \mathbb{C}$ are of the form $\psi_1 = \psi_1(x)$ with $x \in \mathbb{R}^4$, and they live in the complex Hilbert space $L_2(\mathbb{R}^4)$,
- the n-particle functions $\psi_n : \mathbb{R}^{4n} \to \mathbb{C}$,

$$\psi_n = \psi_n(x_1, x_2, \ldots, x_n), \qquad n = 2, 3, \ldots,$$

  are antisymmetric with respect to the n arguments $x_1, x_2, \ldots, x_n \in \mathbb{R}^4$, and they live in the complex Hilbert space $L_{2,\text{antisym}}(\mathbb{R}^{4n})$.

The antisymmetry of the functions ψ_n reflects the principle of indistinguishability for fermions. The fermionic Fock space Y is an infinite-dimensional complex Hilbert space equipped with the inner product

$$\boxed{\langle \psi | \varphi \rangle := \psi_0^\dagger \varphi_0 + \sum_{n=1}^{\infty} \langle \psi_n | \varphi_n \rangle_n.}$$

Recall that $\langle \psi | \varphi \rangle_n := \int_{\mathbb{R}^{4n}} \psi(x_1, \ldots, x_n)^\dagger \varphi(x_1, \ldots, x_n)\, dx_1 \cdots dx_n$.

Particle number operator N. For $\psi := (\psi_0, \psi_1, \psi_2, \ldots)$, set

$$\boxed{N\psi := (0, \psi_1, 2\psi_2, \ldots, n\psi_n, \ldots).}$$

More precisely, the linear operator $N : D(N) \to Y$ is defined for all states $\psi \in Y$ of the fermionic Fock space Y with

$$\sum_{n=1}^{\infty} n^2 ||\psi_n||_n^2 < \infty.$$

For example, let $\psi_n \in Y_n$ for fixed index n with $||\psi_n||_n = 1$. Choose

$$\Psi_n := (0, \ldots, 0, \psi_n, 0, 0, \ldots)$$

where ψ_n stands at the nth place. Then, Ψ_n is a normalized state in the fermionic Fock space Y with

$$N\Psi_n = n\Psi_n.$$

In terms of physics, the state Ψ_n describes n fermions.

Vacuum. The state $|0\rangle := (1, 0, 0, \ldots)$ is a normalized state in the fermionic Fock space Y with

$$\boxed{N|0\rangle = 0.}$$

The state $|0\rangle$ is called the normalized vacuum state (or briefly vacuum), since the number of fermions is equal to zero in this state.

Dense linear subspace Y_{fin} of the fermionic Fock space Y. Let Y_{fin} denote the set of all states $\psi = (\psi_0, \psi_1, \ldots)$ in the fermionic Fock space Y for which at most a finite number of the functions $\psi_1, \psi_2, \ldots$ does not vanish identically.

Composition of particle functions. We are given the one-particle function $f \in L_2(\mathbb{R}^4)$. For each n-particle function

$$\psi_n \in L_{2,\text{antisym}}(\mathbb{R}^{4n}),$$

the antisymmetrization of the tensor product $f \otimes \psi_n$ yields the $(n+1)$-particle function

$$\boxed{\varrho_{n+1} := \sqrt{n+1} \cdot \text{antisym}(f \otimes \psi_n).} \tag{10.8}$$

Intuitively, this is the composition of the one-particle state f with the n-particle state ψ_n. Explicitly,

$$\varrho_{n+1}(x_1, \ldots, x_{n+1}) = \frac{\sqrt{n+1}}{(n+1)!} \sum_{\pi} \text{sgn}\,\pi \cdot f(x_1)\psi_n(x_2, \ldots, x_{n+1})$$

where we sum over all permutations π of the arguments $x_1, ..., x_{n+1}$, and $\text{sgn}\,\pi$ denotes the sign of the permutation π.

Creation operator $b^+(f)$. Fix again the one-particle function $f \in L_2(\mathbb{R}^4)$. We want to construct a linear operator

$$b^+(f) : Y_{\text{fin}} \to Y$$

which describes the creation of particles. Explicitly, for each sequence

$$\psi := (\psi_0, \psi_1, \psi_2, \ldots)$$

in the linear subspace Y_{fin} of the fermionic Fock space Y, we define

$$b^+(f)\psi := (0, \rho_1, \rho_2, \ldots) \tag{10.9}$$

where the functions $\varrho_1, \varrho_2, \ldots$ are given by (10.8). In particular, we have

$$\varrho_1(x_1) := f(x_1)\psi_0, \qquad \varrho_2(x_1, x_2) = \frac{f(x_1)\psi_1(x_2) - f(x_2)\psi_1(x_1)}{\sqrt{2}}.$$

Annihilation operator $b^-(f)$. We want to construct a linear operator

$$b^-(f) : Y_{\text{fin}} \to Y$$

which is formally adjoint to the creation operator $b^+(f)$, that is,

$$\boxed{\langle b^-(f)\varphi|\psi\rangle = \langle\varphi|b^+(f)\psi\rangle \qquad \text{for all} \quad \varphi, \psi \in Y_{\text{fin}}.}$$

In other words, we want to get $b^-(f) = (b^+(f))^\dagger$ on Y_{fin}. To this end, for each sequence $\varphi := (\varphi_0, \varphi_1, \varphi_2, \ldots)$ in Y_{fin}, we define

$$b^-(f)\varphi := (\chi_0, \chi_1, \chi_2, \ldots) \tag{10.10}$$

along with

$$\boxed{\chi_n(x_1, \ldots, x_n) := \sqrt{n+1} \int_{\mathbb{R}^4} f(x)^\dagger \varphi_{n+1}(x, x_1, \ldots, x_n) d^4x}$$

for all indices $n = 0, 1, 2, \ldots$ In particular, we have

$$\chi_0 := \int_{\mathbb{R}^4} f(x)^\dagger \varphi_1(x) d^4x, \qquad \chi_1(x_1) := \sqrt{2} \int_{\mathbb{R}^4} f(x)^\dagger \varphi_2(x, x_1) d^4x.$$

For the vacuum state, we get

$$\boxed{b^-(f)|0\rangle = 0 \qquad \text{for all} \quad f \in L_2(\mathbb{R}^4).}$$

Fundamental anticommutation relations. Fix $f, g \in L_2(\mathbb{R}^4)$. Recall that $[A, B]_+ := AB + BA$. In particular, $[A, B]_+ = 0$ is equivalent to the anticommutativity relation $AB = -BA$.

Theorem 10.2 *For all states ψ, φ in the linear subspace Y_{fin} of the fermionic Fock space Y, the following relations hold:*
(i) *Creation operators:* $[b^+(f), b^+(g)]_+ \, \psi = 0$.
(ii) *Annihilation operators:* $[b^-(f), b^-(g)]_+ \, \psi = 0$.
(iii) *Creation and annihilation operators:*

$$[b^-(f), b^+(g)]_+ \, \psi = \langle f|g\rangle_1 \psi. \tag{10.11}$$

(iv) *Duality:* $\langle b^-(f)\varphi|\psi\rangle = \langle \varphi|b^+(f)\psi\rangle$.

Proof. Let us start with a special case. Choose functions $f, g \in L_2(\mathbb{R}^4)$ with $||f||_1 = ||g||_1 = 1$. Since $b^+(g)|0\rangle = (0, g, 0, ...)$, we get

$$b^+(f)b^+(g)|0\rangle = \frac{1}{\sqrt{2}} \, (0, 0, f(x_1)g(x_2) - f(x_2)g(x_1), 0, \ldots).$$

By antisymmetry, $b^+(f)b^+(g)|0\rangle + b^+(g)b^+(f)|0\rangle = 0$.

From $b^-(f)|0\rangle = b^-(g)|0\rangle = 0$ we get $b^-(f)b^-(g)|0\rangle + b^-(g)b^-(f)|0\rangle = 0$. Finally, it follows from

$$b^-(f)b^+(g)|0\rangle = \left(\int_{\mathbb{R}^4} f(x)^\dagger g(x) d^4x, 0, 0, \ldots \right)$$

and $b^+(g)b^-(f)|0\rangle = 0$ that

$$b^-(f)b^+(g)|0\rangle + b^+(g)b^-(f)|0\rangle = \langle f|g\rangle_1 |0\rangle.$$

The proof of the general case proceeds similarly by induction. □

Physical interpretation. Choose functions $f_1, \ldots, f_s \in L_2(\mathbb{R}^4)$ with the normalization condition $||f_j||_1 = 1$ for $j = 1, \ldots, s$. Set

$$\boxed{\psi := b^+(f_1)b^+(f_2) \cdots b^+(f_s)|0\rangle.} \tag{10.12}$$

This is a state in the fermionic Fock space Y. Observe that

$$b^+(f_j)|0\rangle = (0, f_j, 0, \ldots)$$

and $Nb^+(f_j)|0\rangle = b^+(f_j)|0\rangle$. We say that

- the function f_j represents a normalized one-particle state of one fermion, and
- the operator $b^+(f_j)$ generates the normalized one-particle state $b^+(f_j)|0\rangle$ from the vacuum state $|0\rangle$.

In the general case, we get

$$\boxed{N\psi = s\psi.}$$

Therefore, if $\psi \neq 0$, then the state ψ from (10.12) represents s fermions which are in one-particle states corresponding to $f_1, \ldots, f_s$.

The Pauli exclusion principle. Because of Theorem 10.2 above, the state ψ from (10.12) changes sign under odd permutations of $f_1, \ldots, f_s$. Thus, we get

$$b^+(f_1)b^+(f_2)\cdots b^+(f_s)|0\rangle = 0$$

if two one-particle states f_j and f_k coincide. For example,

$$b^+(f)b^+(f)|0\rangle = 0.$$

Important special case. Consider a system of functions $f_1, f_2, f_3, \ldots$ which form an orthonormal system in the Hilbert space $L_2(\mathbb{R}^4)$, that is,

$$\langle f_k | f_l \rangle_1 = \delta_{kl}, \qquad k, l = 1, 2, \ldots$$

Define

$$b_j^+ := b^+(f_j), \qquad b_j^- := b^-(f_j), \qquad j = 1, 2, \ldots$$

For all $\psi \in Y_{\text{fin}}$ and all $j, k = 1, 2, \ldots$, we then have the following anticommutation relations:

$$\boxed{\begin{aligned} \left[b_j^+, b_k^+\right]_+ \psi = \left[b_j^-, b_k^-\right]_+ \psi = 0, \\ \left[b_j^-, b_k^+\right]_+ \psi = \delta_{jk}\psi. \end{aligned}} \tag{10.13}$$

If $1 \leq j_1 < \ldots < j_k$, then the symbol

$$b_{j_1}^+ b_{j_2}^+ \cdots b_{j_k}^+ |0\rangle$$

represents a normalized state in the fermionic Fock space Y.

The rigorous language of operator-valued distributions. The space of linear operators $B : Y_{\text{fin}} \to Y$ is denoted by $L(Y_{\text{fin}}, Y)$. Set

$$\boxed{B^+(f) := b^+(f) \qquad \text{for all} \quad f \in \mathcal{D}(\mathbb{R}^4).}$$

Then, $B^+ : \mathcal{D}(\mathbb{R}^4) \to L(X_{\text{fin}}, X)$ is a linear map from the space $\mathcal{D}(\mathbb{R}^4)$ of test functions to the operator space $L(Y_{\text{fin}}, Y)$. We call B^+ a distribution with values in the operator space $L(Y_{\text{fin}}, Y)$. Similarly, we define

$$B^-(f) := b^-(f) \qquad \text{for all} \quad f \in \mathcal{D}(\mathbb{R}^4).$$

The map $B^- : \mathcal{D}(\mathbb{R}^4) \to L(Y_{\text{fin}}, Y)$ is antilinear. We call B^- an antidistribution with values in the operator space $L(Y_{\text{fin}}, Y)$.

The formal language of physicists. Physicists introduce the formal fermionic creation operators $b^+(x)$ and the formal fermionic annihilation operators $b^-(x)$ along with the formal commutation relations

$$\begin{aligned}\left[b^+(x), b^+(y)\right]_+ &= \left[b^-(x), b^-(y)\right]_+ = 0,\\ \left[b^-(x), b^+(y)\right]_+ &= \delta(x-y)I\end{aligned} \tag{10.14}$$

and the duality relations

$$(b^+(x))^\dagger = b^-(x), \qquad (b^-(x))^\dagger = b^+(x).$$

These relations are assumed to be valid for all $x, y \in \mathbb{R}^4$. Intuitively, the operator $b^+(x)$ describes the creation of a fermion at the space-time point x (resp. the operator $b^-(x)$ describes the annihilation of a fermion at x). Furthermore, we formally write

$$\boxed{b^-(f) := \int_{\mathbb{R}^4} f(x)^\dagger b^-(x) d^4x, \qquad b^+(f) := \int_{\mathbb{R}^4} f(x) b^+(x) d^4x}$$

along with

$$b^-(f)b^+(g) := \int_{\mathbb{R}^8} f(x)^\dagger g(y) b^-(x) b^+(y) d^4x d^4y,$$

and so on. Mnemonically, this yields the rigorous approach introduced above. For example,

$$\begin{aligned}b^-(f)b^+(g) + b^+(g)b^-(f) &= \int_{\mathbb{R}^8} f(x)^\dagger g(y) \left[b^-(x), b^+(y)\right]_+ d^4x d^4y\\ &= \int_{\mathbb{R}^8} f(x)^\dagger g(y)\delta(x-y)I \cdot d^4x d^4y = \left(\int_{\mathbb{R}^4} f(x)^\dagger g(x) d^4x\right) I.\end{aligned}$$

Furthermore,

$$b^+(f)b^+(g) + b^+(g)b^+(f) = \int_{\mathbb{R}^8} f(x)g(y)[b^+(x), b^+(y)]_+ \, d^4x d^4y = 0.$$

Similarly, $b^-(f)b^-(g) + b^-(g)b^-(f) = 0$. Finally, we formally get

$$(b^-(f))^\dagger = \left(\int_{\mathbb{R}^4} f(x)^\dagger b^-(x) d^4x\right)^\dagger = \int_{\mathbb{R}^4} f(x) b^+(x) d^4x = b^+(f).$$

10.3 General Construction

In a straightforward manner, we now want to generalize the construction of bosonic and fermionic Fock spaces to one–particle functions $\psi : \mathbb{R}^4 \to \mathbb{C}^d$ which possess d degrees of freedom:

$$\psi(x) = \begin{pmatrix}\psi^1(x)\\ \vdots\\ \psi^d(x)\end{pmatrix}, \qquad x \in \mathbb{R}^4.$$

We briefly write $\psi(x) = (\psi^j(x))$. The desired generalization can be easily obtained by using systematically the language of tensor products.

The one-particle Hilbert space $L_2(\mathbb{R}^4, \mathbb{C}^d)$. To begin with, let us introduce the following inner product:

$$\boxed{\langle\psi|\varphi\rangle_1 := \int_{\mathbb{R}^4} \sum_{j=1}^{d} \psi^j(x)^\dagger \varphi^j(x)\, d^4x.}$$

By definition, the space $L_2(\mathbb{R}^4, \mathbb{C}^d)$ consists of all the functions $\psi : \mathbb{R}^4 \to \mathbb{C}^d$ with $\langle\psi|\psi\rangle_1 < \infty$.[7] The space $L_2(\mathbb{R}^4, \mathbb{C}^d)$ becomes a complex Hilbert space equipped with the inner product $\langle\psi|\varphi\rangle_1$.

Bosonic two-particle functions. Let $\psi, \varphi \in L_2(\mathbb{R}^4, \mathbb{C}^d)$ be one-particle functions. The prototype of a two-particle function is the tensor product $\psi \otimes \varphi$. Explicitly, this is the tuple

$$(\psi \otimes \varphi)(x_1, x_2) := (\psi^i(x_1)\varphi^j(x_2))_{i,j=1,\dots,d}, \qquad x_1, x_2 \in \mathbb{R}^4.$$

Naturally enough, the inner product is defined by

$$\langle\psi \otimes \varphi|\psi_* \otimes \varphi_*\rangle_2 := \int_{\mathbb{R}^8} \sum_{i,j=1}^{d} \Psi^{ij}(x_1, x_2)^\dagger \Psi_*^{ij}(x_1, x_2)\, d^4x_1 d^4x_2$$

where $\Psi^{ij}(x_1, x_2) := \psi^i(x_1)\varphi^j(x_2)$, and $\Psi_*^{ij}(x_1, x_2) := \psi_*^i(x_1)\varphi_*^j(x_2)$, and

In order to get a bosonic two-particle function, we have to symmetrize. This means that we have to pass from $\psi \otimes \varphi$ to

$$\mathrm{sym}(\psi \otimes \varphi) := \tfrac{1}{2}(\psi \otimes \varphi + \varphi \otimes \psi).$$

In general, by a bosonic two-particle function we understand a tuple

$$\Psi(x_1, x_2) = (\Psi^{ij}(x_1, x_2))_{i,j=1,\dots,d}, \quad x_1, x_2 \in \mathbb{R}^4,$$

which is symmetric with respect to both the indices i, j and the arguments x_1, x_2. Explicitly, we obtain

$$\Psi^{ij}(x_1, x_2) = \Psi^{ji}(x_2, x_1), \qquad i, j = 1, \dots, d, \quad x_1, x_2 \in \mathbb{R}^4.$$

In addition, we assume that all the components Ψ^{ij} live in the space $L_2(\mathbb{R}^8)$. We briefly write

$$\Psi \in L_{2,\mathrm{sym}}(\mathbb{R}^8, \mathbb{C}^{d^2}).$$

In particular, for the bosonic two-particle functions $\Psi, \Phi \in L_{2,\mathrm{sym}}(\mathbb{R}^8, \mathbb{C}^{d^2})$, the inner product is given by

$$\langle\Psi|\Phi\rangle_2 := \int_{\mathbb{R}^8} \sum_{i,j=1}^{d} \Psi^{ij}(x_1, x_2)^\dagger \Phi^{ij}(x_1, x_2)\, d^4x_1 d^4x_2.$$

Fermionic two-particle functions. We now replace symmetry by antisymmetry. For given one-particle functions $\psi, \varphi \in L_2(\mathbb{R}^4, \mathbb{C}^d)$, antisymmetrization yields the special fermionic two-particle function

$$\mathrm{antisym}(\psi \otimes \varphi) = \tfrac{1}{2}(\psi \otimes \varphi - \varphi \otimes \psi) = \tfrac{1}{2}(\psi \wedge \varphi).$$

[7] We tacitly assume that the components of the functions ψ are measurable with respect to the Lebesgue measure on $\mathbb{R}^4$. In addition, observe that two functions $\psi, \varphi : \mathbb{R}^4 \to \mathbb{C}^d$ are identified with each other if they only differ on a subset of $\mathbb{R}^4$ which has the 4-dimensional Lebesgue measure zero.

Generally, by a fermionic two-particle function we understand a tuple

$$\Psi(x_1, x_2) = (\Psi^{ij}(x_1, x_2)), \qquad i, j = 1, ..., d, \quad x_1, x_2 \in \mathbb{R}^4$$

which is antisymmetric with respect to both the indices i, j and the arguments x_1, x_2. Explicitly,

$$\Psi^{ij}(x_1, x_2) = -\Psi^{ji}(x_2, x_1), \qquad i, j = 1, ..., d, \quad x_1, x_2 \in \mathbb{R}^4.$$

In addition, we assume that all the components Ψ^{ij} live in the space $L_2(\mathbb{R}^8)$. We briefly write $\Psi \in L_{2,\text{antisym}}(\mathbb{R}^8, \mathbb{C}^{d^2})$. Next we want to introduce

- the bosonic Fock space, and
- the fermionic Fock space.

The bosonic Fock space. The direct sum

$$\boxed{X = \bigoplus_{n=0}^{\infty} X_n}$$

of the Hilbert spaces $X_0 := \mathbb{C}$, $X_1 := L_2(\mathbb{R}^4, \mathbb{C})$, and

$$X_n := L_{2,\text{sym}}(\mathbb{R}^{4n}, \mathbb{C}^{d^n}), \qquad n = 2, 3, \ldots$$

is called the bosonic Fock space to the one-particle Hilbert space $L_2(\mathbb{R}^4, \mathbb{C}^d)$. Let $i_1, \ldots, i_n = 1, \ldots, d$ and $x_1, \ldots, x_n \in \mathbb{R}^4$. By definition, the elements of the space X_n are tuples

$$\Psi(x_1, \ldots, x_n) = (\Psi^{i_1 \ldots i_n}(x_1, \ldots, x_n))$$

which are symmetric with respect to both the indices $i_1, \ldots, i_n = 1, \ldots, n$ and the n space-time variables $x_1, \ldots, x_n$. Moreover, all of the components $\Psi^{i_1 \ldots i_n}$ live in the space $L_2(\mathbb{R}^{4n})$. The elements of the bosonic Fock space X are infinite tuples

$$\Psi = (\Psi_0, \Psi_1, \ldots)$$

where Ψ_0 is a complex number, and $\Psi_n \in X_n$ for $n = 1, 2, \ldots$ In addition, we postulate that $\sum_{n=1}^{\infty} \langle \Psi_n | \Psi_n \rangle_n < \infty$ where we define

$$\langle \Psi_n | \Phi_n \rangle_n := \int_{\mathbb{R}^{4n}} \sum_{i_1, \ldots, i_n = 1}^{d} (\Psi^{i_1 \ldots i_n})^\dagger \Phi^{i_1 \ldots i_n} \, d^4x_1 \cdots d^4x_n.$$

The bosonic Fock space X becomes a complex Hilbert space equipped with the inner product

$$\langle \Psi | \Phi \rangle := \Psi_0^\dagger \Phi_0 + \sum_{n=1}^{\infty} \langle \Psi_n | \Phi_n \rangle_n.$$

The linear subspace X_{fin} and the vacuum state $|0\rangle := (1, 0, 0, \ldots)$ are defined as in Sect. 10.1. For each given one-particle function $f \in X_1$, the creation operator

$$a^+(f) : X_{\text{fin}} \to X$$

is defined by $a^+(f)\Psi := (0, \rho_1, \rho_2, \ldots)$ where

$$\boxed{\rho_{n+1} := \sqrt{n+1} \cdot \text{sym}(f \otimes \Psi_n), \qquad n = 0, 1, 2, \ldots}$$

Explicitly, $\varrho_{n+1}^{i_1 \dots i_{n+1}}(x_1, \dots x_{n+1})$ is equal to

$$\frac{\sqrt{n+1}}{(n+1)!} \sum_{\pi} \pi\big(f^{i_1}(x_1) \Psi_n^{i_2 \dots i_{n+1}}(x_2, \dots, x_{n+1})\big)$$

where we sum over all permutations π of $1, \dots, n+1$. The operation $\pi(\dots)$ refers to permutations of both the indices $i_1, \dots, i_{n+1}$ and the arguments $x_1, \dots, x_{n+1}$. The annihilation operator

$$a^-(f) : X_{\text{fin}} \to X$$

is the formally adjoint operator to the creation operator $a^+(f)$, that is,

$$\boxed{\langle a^-(f)\Phi | \Psi \rangle = \langle \Phi | a^+(f) \Psi \rangle \qquad \text{for all} \quad \Phi, \Psi \in X_{\text{fin}}.}$$

In other words, $a^-(f) = (a^+(f))^\dagger$ on X_{fin}. Explicitly, for each given sequence $\Phi := (\Phi_0, \Phi_1, \Phi_2, \dots)$ in X_{fin}, we define

$$a^-(f)\psi := (\chi_0, \chi_1, \chi_2, \dots)$$

where $\chi_n^{i_1 \dots i_n}(x_1, \dots, x_n)$ is given by

$$\sqrt{n+1} \int_{\mathbb{R}^4} \sum_{i=1}^{d} f^i(x)^\dagger \Phi_{n+1}^{i i_1 \dots i_n}(x, x_1, \dots, x_n) d^4x.$$

In particular, $\chi_0 = \int_{\mathbb{R}^4} \sum_{i=1}^d f^i(x)^\dagger \Psi^i(x)\, d^4x$.

The fermionic Fock space. The direct sum

$$\boxed{Y = \bigoplus_{n=0}^{\infty} Y_n}$$

of the Hilbert spaces $Y_0 := \mathbb{C}$, $Y_1 := L_2(\mathbb{R}^4, \mathbb{C})$, and

$$Y_n := L_{2,\text{antisym}}(\mathbb{R}^{4n}, \mathbb{C}^{d^n}), \qquad n = 2, 3, ...$$

is called the fermionic Fock space to the one-particle Hilbert space $L_2(\mathbb{R}^4, \mathbb{C}^d)$. Let $i_1, \dots, i_n = 1, \dots, d$ and $x_1, \dots, x_n \in \mathbb{R}^4$. By definition, the elements of the space Y_n are tuples

$$\Psi(x_1, \dots, x_n) = (\Psi^{i_1 \dots i_n}(x_1, \dots, x_n))$$

which are antisymmetric with respect to both the indices $i_1, \dots, i_n$ and the n space-time variables $x_1, \dots, x_n$. Moreover, the components $\Psi^{i_1 \dots i_n}$ live in the space $L_2(\mathbb{R}^{4n})$. Explicitly, the elements of the fermionic Fock space Y are infinite tuples

$$\Psi = (\Psi_0, \Psi_1, \dots)$$

where Ψ_0 is a complex number, and $\Psi_n \in Y_n$ for $n = 1, 2, \dots$ In addition, we postulate that $\sum_{n=1}^{\infty} \langle \Psi_n | \Psi_n \rangle_n < \infty$. The space Y becomes a complex Hilbert space equipped with the inner product

$$\langle \Psi | \Phi \rangle := \Psi_0^\dagger \Phi_0 + \sum_{n=1}^{\infty} \langle \Psi_n | \Phi_n \rangle_n.$$

For each given one-particle function $f \in Y_1$, the creation operator

$$b^+(f) : Y_{\text{fin}} \to Y$$

is defined by $b^+(f)\Psi := (0, \rho_1, \rho_2, \ldots)$ where

$$\boxed{\rho_{n+1} := \sqrt{n+1} \cdot \text{antisym}(f \otimes \Psi_n), \qquad n = 0, 1, 2, \ldots}$$

Explicitly, $\varrho_{n+1}^{i_1 \ldots, i_{n+1}}(x_1, \ldots, x_{n+1})$ is equal to

$$\frac{\sqrt{n+1}}{(n+1)!} \sum_\pi \text{sgn}\, \pi \cdot \pi\big(f^{i_1}(x_1) \Psi_n^{i_2 \ldots i_{n+1}}(x_2, \ldots, x_{n+1})\big)$$

where we sum over all permutations π of $1, \ldots, n+1$. The operation $\pi(\ldots)$ refers to permutations of both the indices $i_1, \ldots, i_{n+1}$ and the arguments $x_1, \ldots, x_{n+1}$. The annihilation operator

$$b^-(f) : Y_{\text{fin}} \to Y$$

is the formally adjoint operator to the creation operator $b^+(f)$, that is,

$$\boxed{\langle b^-(f)\Phi | \Psi \rangle = \langle \Phi | b^+(f)\Psi \rangle \qquad \text{for all} \quad \Phi, \Psi \in Y_{\text{fin}}.}$$

In other words, $b^-(f) = (b^+(f))^\dagger$ on Y_{fin}. Explicitly, for each sequence

$$\Phi := (\Phi_0, \Phi_1, \Phi_2, \ldots)$$

in the space Y_{fin}, we define

$$b^-(f)\Phi := (\chi_0, \chi_1, \chi_2, \ldots)$$

where $\chi_n^{i_1 \ldots i_n}(x_1, \ldots, x_n)$ is equal to

$$\sqrt{n+1} \int_{\mathbb{R}^4} \sum_{i=1}^d f^i(x)^\dagger \Phi_{n+1}^{i i_1 \ldots i_n}(x, x_1, \ldots, x_n)\, d^4x.$$

10.4 The Main Strategy of Quantum Electrodynamics

The most important experiments in elementary particle physics are scattering experiments carried out in huge high-energy particle accelerators. Physicists characterize the outcome of such experiments by cross sections. If $J = \varrho v$ is the current density of the incoming particle stream with velocity v and particle density ϱ, then

$$N = \sigma T \cdot J$$

is the number of scattered particles observed during the time interval $[-\frac{T}{2}, \frac{T}{2}]$. In the SI system of physical units, J has the physical dimension of particle density times velocity, $1/\text{m}^2\text{s}$. Therefore, the quantity σ has the physical dimension of area, m^2, and σ is called the *cross section* of the scattering process. Observe the following:

- Cross sections follow from transition probabilities.
- Transition probabilities result from transition amplitudes.

- Transition amplitudes can be computed by using Feynman diagrams and the corresponding Feynman rules.

Our goal is to motivate the Feynman rules in quantum electrodynamics and to apply them to the computation of cross sections.

The Feynman rules represent the hard core of quantum field theory.

For quantum electrodynamics, the Feynman rules will be summarized in Sect. 14.3. Applications to scattering processes can be found in Chap. 15.

Quantum electrodynamics studies the interaction between the following particles: electrons, positrons, and photons. Here, photons represent quantized electromagnetic waves. Note that:

- The electron is called the basic particle of quantum electrodynamics.
- The positron is the antiparticle to the electron.
- The massless photon is responsible for the interaction between electrons and positrons. Therefore, photons are called the interacting particles of quantum electrodynamics.

In order to understand quantum field theory, one has to start with quantum electrodynamics. Let us discuss the main ideas of quantum electrodynamics. We will proceed in the following four steps:

(C) Classical field theory: We first consider the classical principle of critical action for the Maxwell–Dirac field which is obtained by coupling the classical electromagnetic field to the Dirac field for the relativistic electron.

(F) The free quantum field: For the free electromagnetic field and the free Dirac field of the electron, we find solutions in the form of finite Fourier series. Replacing Fourier coefficients by creation and annihilation operators, we get the corresponding free quantum fields for electrons, positrons, and photons (the method of Fourier quantization).

These free quantum fields depend on the choice of both a finite box in position space and a finite lattice in momentum space.

(I) The interacting quantum field: We use the interaction term between the electrodynamic field and the Dirac field for the electron in order to formulate the Dyson series for the S-matrix of quantum electrodynamics. The S-matrix is a formal power series expansion with respect to the dimensionless coupling constant in the SI system of physical units:

$$\boxed{\kappa := \sqrt{4\pi\alpha}.} \tag{10.15}$$

Here, α denotes the so-called fine structure constant in quantum electrodynamics:

$$\boxed{\alpha = \frac{1}{137.04} = 0.007297} \tag{10.16}$$

which is dimensionless. In addition, $-e$ is the negative electric charge of the electron.[8] In the SI system, we have

$$\alpha = \frac{e^2}{4\pi\varepsilon_0\hbar c}.$$

[8] If we want to emphasize that κ and α refer to quantum electrodynamics, but not to strong and weak interaction in the Standard Model, then we write κ_{QED} and α_{QED}, respectively.

The point is that the Dyson series for the S-matrix depends nonlinearly on the free quantum fields for electrons, positrons, and photons. Using the approximation of the S-matrix in lowest nontrivial order, we are able to compute approximately scattering processes for electrons, positrons, and photons.

(R) Renormalization: Using higher-order approximations of the S-matrix together with the high-energy limit (resp. the low-energy limit), we get divergent expressions for scattering processes. In order to extract physical information from those divergent expressions, we have to use the crucial method of renormalization. The final results are cross sections for scattering processes of the form

$$\boxed{\sigma = \sigma_1 \kappa + \sigma_2 \kappa^2 + \dots}$$

This is a power series expansion with respect to the small dimensionless coupling constant κ given by (10.15). The coefficients $\sigma_1, \sigma_2, \dots$ are real numbers (equipped with the physical dimension of area) coming from divergent integrals by using a regularization procedure. The coincidence between theory and physical experiment is extremely precise in quantum electrodynamics.

The smallness of the dimensionless (electromagnetic) fine structure constant α is responsible for the incredible success of perturbation theory in quantum electrodynamics.

The situation changes completely in strong interaction where the coupling constant is approximately equal to one, $\kappa = 1$. Then the results of perturbation theory are only crude approximations of reality.

In string theory, there exists a duality transformation between certain models which allows us to transform some models having large coupling constant into dual models having small coupling constant. In the future, physicists hope to establish such a beautiful duality method for strong interaction in nature.

Convention for the choice of the system of physical units. To simplify notation, in the following chapters we will use the energetic system of units, that is, we set

$$\boxed{\hbar = c = \varepsilon_0 = \mu_0 = k := 1.} \tag{10.17}$$

Then, the dimension of an arbitrary physical quantity is some power of energy (see the Appendix A.2 of Vol. I). In particular, the electric charge $-e$ of the electron is dimensionless, and we have

$$e = \sqrt{4\pi\alpha}.$$

The gauge condition. It is a typical feature of quantum electrodynamics that we do not start with the electromagnetic field $\mathbf{E}, \mathbf{B}$, but with the four-potential $U, \mathbf{A}$. The electromagnetic field is then given by

$$\mathbf{E} = -\operatorname{\mathbf{grad}} U - \dot{\mathbf{A}}, \qquad \mathbf{B} = \operatorname{\mathbf{curl}} \mathbf{A}.$$

The point is that the four-potential is only determined up to a gauge transformation of the form

$$U \mapsto U - \frac{\partial f}{\partial t}, \qquad \mathbf{A} \mapsto \mathbf{A} + \operatorname{\mathbf{grad}} f$$

where f is a smooth function. This causes some trouble. We will overcome the difficulties in the following sections by using the following trick:

(i) We first destroy the gauge invariance by passing to a modified Lagrangian.

(ii) The corresponding free quantum fields include virtual photons which do not possess an obvious physical meaning.

(iii) In classical theory, virtual photons are eliminated by adding the Lorenz gauge condition.
(iv) In quantum field theory, virtual photons are eliminated by adding a weak Lorenz gauge condition (Gupta–Bleuler quantization).

Nevertheless, we will see that virtual photons essentially influence physical processes proceeding in our real world. The point is that there arise terms in perturbation theory which depend on the photon propagator, and this photon propagator contains contributions coming from virtual photons. In general, quantum electrodynamics adds new physical effects to classical electrodynamics which can be summarized under the sketch word *quantum fluctuations of the ground state* (also called the vacuum). In particular, this concerns the so-called vacuum polarization.

11. The Basic Equations in Quantum Electrodynamics

Quantum electrodynamics couples the Maxwell equation for the photon to the Dirac equation for the electron.

Folklore

11.1 The Classical Lagrangian

The Einstein convention. Let us choose a fixed inertial system with

$$x = (x^0, x^1, x^2, x^3).$$

Here, $\mathbf{x} = x^1\mathbf{i} + x^2\mathbf{j} + x^3\mathbf{k}$ is the position vector of a Cartesian coordinate system with the right-handed orthonormal basis $\mathbf{i}, \mathbf{j}, \mathbf{k}$. We also set $x^0 := t$ where t denotes time. For the indices $\mu, \nu = 0, 1, 2, 3$, we set

$$\eta_{00} := 1, \quad \eta_{11} = \eta_{22} = \eta_{33} := -1, \quad \eta_{\mu\nu} := 0 \quad \text{if } \mu \neq \nu$$

along with $\eta^{\mu\nu} := \eta_{\mu\nu}$. As usual, we use the η-symbol for lifting and lowering of indices.

By the Einstein convention, we sum over equal lower and upper Greek indices from 0 to 3.

For example, according to the Einstein convention we write

$$A^\mu := \eta^{\mu\nu} A_\nu = \sum_{\nu=0}^{3} \eta^{\mu\nu} A_\nu, \qquad A_\mu = \eta_{\mu\nu} A^\nu = \sum_{\nu=0}^{3} \eta_{\mu\nu} A^\nu.$$

Introducing the partial derivative $\partial_\mu := \partial/\partial x^\mu$, we get

$$(\partial_\mu A_\nu)(\partial^\mu A^\nu) = \sum_{\mu,\nu=0}^{3} (\partial_\mu A_\nu)(\partial^\mu A^\nu) = \sum_{\mu,\nu,\alpha,\beta=0}^{3} \partial_\mu A_\nu \eta^{\mu\alpha} \eta^{\nu\beta} \partial_\alpha A_\beta.$$

The Dirac–Pauli matrices $\gamma^0, \gamma^1, \gamma^2, \gamma^3$ satisfy the following fundamental Clifford anticommutation rules:

$$\boxed{\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2\eta^{\mu\nu} I, \qquad \mu, \nu = 0, 1, 2, 3.} \tag{11.1}$$

Explicitly, the Dirac–Pauli matrices are given by

$$\gamma^0 := \begin{pmatrix} \sigma^0 & 0 \\ 0 & -\sigma^0 \end{pmatrix}, \quad \gamma^j := \begin{pmatrix} 0 & \sigma^j \\ -\sigma^j & 0 \end{pmatrix}, \qquad j = 1, 2, 3, \tag{11.2}$$

along with the Pauli matrices

$$\sigma^0 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma^1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^2 := \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma^3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Furthermore, we introduce the chiral matrix

$$\gamma^5 := \mathrm{i}\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} 0 & \sigma^0 \\ \sigma^0 & 0 \end{pmatrix}.$$

The choice of the Dirac–Pauli matrices is called the standard representation of the Clifford anticommutation relations (11.1). The chiral representation via Weyl matrices can be found in Problem 15.6 on page 936.

The principle of critical action in quantum electrodynamics. Let Ω be a nonempty bounded open subset of $\mathbb{R}^4$ with the closure $\mathrm{cl}(\Omega)$. Introduce the Dirac bispinor function $\psi = \psi(x)$ with the components

$$\psi(x) := \begin{pmatrix} \psi^0(x) \\ \psi^1(x) \\ \psi^2(x) \\ \psi^3(x) \end{pmatrix},$$

and the electromagnetic four-potential $A(x) := (A_0(x), A_1(x), A_2(x), A_3(x))$. Furthermore, we introduce the Dirac adjoint

$$\overline{\psi}(x) := \psi^\dagger(x)\gamma^0 = (\psi^0(x)^\dagger, \psi^1(x)^\dagger, -\psi^2(x)^\dagger, -\psi^3(x)^\dagger).$$

For example,

$$\overline{\psi}\psi = (\psi^0)^\dagger\psi^0 + (\psi^1)^\dagger\psi^1 - (\psi^2)^\dagger\psi^2 - (\psi^3)^\dagger\psi^3.$$

The principle of critical action in quantum electrodynamics reads as follows. We are looking for smooth functions $A : \mathrm{cl}(\Omega) \to \mathbb{R}^4$ and $\psi : \mathrm{cl}(\Omega) \to \mathbb{C}^4$ such that

$$\boxed{\int_\Omega \mathcal{L}(A(x), \psi(x), \psi^\dagger(x))\, d^4x = \text{critical!}} \tag{11.3}$$

along with the boundary condition "A = fixed and ψ = fixed on $\partial\Omega$". Here, the Lagrangian density

$$\mathcal{L} = \mathcal{L}_{\text{free}} + \mathcal{L}_{\text{int}}$$

decomposes into the free Lagrangian density[1]

$$\mathcal{L}_{\text{free}} := -\tfrac{1}{2}(\partial_\mu A_\nu)(\partial^\mu A^\nu) + \tfrac{1}{2}\left(1 - \frac{1}{\xi}\right)(\partial_\mu A^\mu)^2 + \mathrm{i}\overline{\psi}\gamma^\mu\partial_\mu\psi - m_e\overline{\psi}\psi$$

and the Lagrangian density of interaction

$$\boxed{\mathcal{L}_{\text{int}} := -J^\mu_{\text{QED}} A_\mu}$$

[1] The fixed real nonzero parameter ξ is called the gauge parameter of quantum electrodynamics. This terminology will be explained in Sect. 11.2. As a rule, we will choose the value $\xi = 1$ which is called the Feynman gauge.

along with the electric 4-current density vector

$$J^{\mu}_{\mathrm{QED}} := -e\overline{\psi}\gamma^{\mu}\psi.$$

Recall that $-e$ is the negative electric charge of the electron, and m_e is the positive rest mass of the electron. We will show in Chap. 13 that:

The interaction Lagrangian, $\mathcal{L}_{\mathrm{int}}$ is crucial for the construction of Dyson's S-matrix, and hence for the computation of scattering processes between electrons, positrons, and photons.

The Euler–Lagrange equations in quantum electrodynamics. Let us introduce the wave operator

$$\Box := \partial_{\mu}\partial^{\mu} = \frac{\partial^2}{\partial t^2} - \sum_{j=1}^{3} \frac{\partial^2}{\partial x^j \partial x^j},$$

and the covariant derivative $\nabla_{\mu} := \partial_{\mu} - \mathrm{i}eA_{\mu}$.

Theorem 11.1 *Each smooth solution A, ψ of the critical action problem (11.3) satisfies the following system of partial differential equations:*
(i) *Wave equations: For $\mu = 0, 1, 2, 3$,*

$$\Box A^{\mu} + \left(\frac{1}{\xi} - 1\right) \partial^{\mu}(\partial_{\nu} A^{\nu}) = J^{\mu}_{\mathrm{QED}}. \tag{11.4}$$

(ii) *Dirac equation:* $\mathrm{i}\gamma^{\nu}\nabla_{\nu}\psi = m_e\psi$.

The proof can be found in Problem 15.5 on page 935. The relation to the electromagnetic field $\mathbf{E}, \mathbf{B}$ will be studied in Sect. 11.2. We will also show in Problem 15.1 on page 932 that each solution of the Dirac equation (ii) satisfies the adjoint Dirac equation

$$-\mathrm{i}\nabla^{-}_{\mu}\ \overline{\psi}\gamma^{\mu} = m_e\overline{\psi}$$

where we introduce $\nabla^{-}_{\mu} := \partial_{\mu} + \mathrm{i}eA_{\mu}$. In particular, if we choose the Feynman gauge, $\xi = 1$, then the system of wave equations (11.4) passes over to the following system of classical wave equations:

$$\boxed{\Box A^{\mu} = J^{\mu}_{\mathrm{QED}}, \qquad \mu = 0, 1, 2, 3.} \tag{11.5}$$

The total energy. Choose the Feynman gauge, $\xi = 1$. Recall that, in classical mechanics, the energy H is given by

$$H := p\dot{q} - L$$

with the conjugate momentum $p := L_{\dot{q}}$. We want to generalize this to quantum electrodynamics. To this end, we replace q by $A = (A_0, A_1, A_2, A_3)$ and ψ. Moreover, we replace the time derivative $\dot{q}$ by the partial time derivatives $\dot{A}_0, \dot{A}_1, \dot{A}_2, \dot{A}_3$ and $\dot{\psi}$. Finally, instead of p we introduce the quantities

$$\Pi^{\mu} := \frac{\partial \mathcal{L}}{\partial \dot{A}_{\mu}}, \qquad \pi := \frac{\partial \mathcal{L}}{\partial \dot{\psi}}, \qquad \mu = 0, 1, 2, 3.$$

This motivates the following definition of the energy density:

$$\mathcal{H} := \Pi^{\mu}\dot{A}_{\mu} + \pi\dot{\psi} - \mathcal{L}.$$

From this we obtain the field energy

$$\boxed{H := \int_G \mathcal{H}\, d^3\mathbf{x}}$$

contained in the bounded open subset G of the position space $\mathbb{R}^3$. Let us introduce the operator

$$\mathbf{H} := m_e\gamma^0 + \sum_{j=1}^{3} \gamma^0\gamma^j(-\mathrm{i}\partial_j).$$

Theorem 11.2 *For the energy density $\mathcal{H}$ of the classical Dirac–Maxwell field, we get the decomposition*

$$\mathcal{H} = \mathcal{H}_{\mathrm{phot}} + \mathcal{H}_{\mathrm{el}} + \mathcal{H}_{\mathrm{int}}$$

with the photon energy density $\mathcal{H}_{\mathrm{phot}} := -\frac{1}{2}\sum_{\mu,\nu=0}^{3}(\partial_\nu A_\mu)^2$, the electron energy density

$$\mathcal{H}_{\mathrm{el}} := \psi^\dagger \mathbf{H}\psi,$$

and the interaction energy density $\mathcal{H}_{\mathrm{int}} := -e(\overline{\psi}\gamma^\mu\psi)A_\mu = J^\mu_{\mathrm{QED}}A_\mu$.

Proof. Recall that the Lagrangian density is given by

$$\mathcal{L} := -\tfrac{1}{2}(\partial_\mu A_\nu)(\partial^\mu A^\nu) + \overline{\psi}(\mathrm{i}\gamma^\mu\partial_\mu - m_e)\psi + e(\overline{\psi}\gamma^\mu\psi)A_\mu.$$

Noting that $\mathcal{L} = -\frac{1}{2}\dot{A}_0^2 + \frac{1}{2}\sum_{j=1}^{3}\dot{A}_j^2 + \ldots$ and $A_0 = A^0$, $A_j = -A^j$ for $j = 1, 2, 3$, we obtain

$$\Pi^\mu = \frac{\partial\mathcal{L}}{\partial\dot{A}_\mu} = -\dot{A}^\mu, \qquad \mu = 0, 1, 2, 3.$$

Moreover, it follows from $\overline{\psi} = \psi^\dagger\gamma^0$ and $(\gamma^0)^2 = I$ that

$$\mathcal{L} = \mathrm{i}\overline{\psi}\gamma^0\partial_0\psi + \ldots = \mathrm{i}\psi^\dagger\dot{\psi} + \ldots$$

Hence $\pi = \frac{\partial\mathcal{L}}{\partial\dot{\psi}} = \mathrm{i}\psi^\dagger$. □

Next we want to study the relation between the solutions of the Euler–Lagrange equations from Theorem 11.1 and the electromagnetic field.

11.2 The Gauge Condition

The Danish physicist Ludvig Valentin Lorenz (1829–1891) is perhaps best known for his pairing with the more famous Dutch physicist Hendrik Antoon Lorentz (1853–1928) in the Lorenz–Lorentz relation between index of refraction and density. In fact, Lorenz was a pioneer in the theory of light and in electrodynamics, contemporaneous with Maxwell (1831–1879). In 1862 he developed a mathematical theory of light, using the basic known facts (transversality of vibrations, Fresnel's law), but avoiding the (unnecessary to him) physical modelling of a mechanistic ether in favor of a purely phenomenological model.[2] ...

[2] L. Lorenz, On the theory of light, Philos. Mag. Ser. 4, **26** (1863), pp. 81–93, 205–219.

> Lorenz proposed the condition (11.6) below in the mid-1860s, but this constraint is generally *misattributed* to the better known Lorentz.[3]
>
> John Jackson and Lev Okun, 2001

Let ξ be a fixed nonzero real gauge parameter. Assume that the four-potential A_0, A_1, A_2, A_3 is a solution of the wave equations from Theorem 11.1,

$$\boxed{\Box A^\mu + \left(\frac{1}{\xi} - 1)\right) \partial^\mu(\partial_\nu A^\nu) = J^\mu_{\mathrm{QED}}, \qquad \mu = 0, 1, 2, 3,}$$

and that it satisfies the Lorenz gauge condition

$$\boxed{\partial_\nu A^\nu = 0.} \tag{11.6}$$

This implies the following.

Theorem 11.3 *The electromagnetic field*

$$F_{\mu\nu} := \partial_\mu A_\nu - \partial_\nu A_\mu \tag{11.7}$$

satisfies the Maxwell equations

$$\partial_\mu F^{\mu\nu} = J^\nu_{\mathrm{QED}}, \qquad \partial_\alpha F_{\beta\gamma} + \partial_\beta F_{\gamma\alpha} + \partial_\gamma F_{\alpha\beta} = 0 \tag{11.8}$$

for all $\nu, \alpha, \beta, \gamma = 0, 1, 2, 3$.

Proof. Because of the Lorenz gauge condition, we get

$$\Box A^\mu = J^\mu_{\mathrm{QED}}, \qquad \mu = 0, 1, 2, 3.$$

Again by the Lorenz gauge condition, we have

$$\partial_\mu F^{\mu\nu} = \partial_\mu(\partial^\mu A^\nu - \partial^\nu A^\mu) = \Box A^\nu - \partial^\nu(\partial_\mu A^\mu) = \Box A^\nu = J^\nu_{\mathrm{QED}}.$$

The equation $\partial_\alpha F_{\beta\gamma} + \partial_\beta F_{\gamma\alpha} + \partial_\gamma F_{\alpha\beta} = 0$ follows easily from (11.7). □

Explicitly, set $U := A^0$ and $\mathbf{A} := A^1\mathbf{i} + A^2\mathbf{j} + A^3\mathbf{k}$, as well as

$$\varrho := J^0_{\mathrm{QED}}, \qquad \mathbf{J}_{\mathrm{QED}} := J^1_{\mathrm{QED}}\mathbf{i} + J^2_{\mathrm{QED}}\mathbf{j} + J^3_{\mathrm{QED}}\mathbf{k}.$$

For the Feynman gauge $\xi = 1$, the following hold true. If the wave equations

$$\Box\mathbf{A} = \mathbf{J}_{\mathrm{QED}}, \qquad \Box U = \varrho$$

are fulfilled and the Lorenz gauge condition

[3] In 1902 Hendrik Antoon Lorentz and Pieter Zeeman were awarded the Nobel prize in physics, in recognition of the extraordinary service they rendered by their research into the influence of magnetism upon radiation phenomena.
H. Lorentz, Theory of Electrons, Teubner, Leipzig, and Stechert, New York, 1909. Second edition 1916. Reprinted by Dover, New York, in 1952.
J. Jackson and L. Okun, Historical roots of gauge invariance, Rev. Mod. Phys. **73** (2001), 663–680 (reprinted with permission; copyright by the Amer. Phys. Soc.). I would like to thank Christoph Dehne for drawing my attention to the necessary distinction between Lorenz (e.g., the Lorenz gauge condition) and Lorentz (e.g., the Lorentz transformation in the theory of special relativity).

$$\boxed{\dot{U} + \operatorname{div} \mathbf{A} = 0}$$

is satisfied, then the electromagnetic field

$$\mathbf{E} = -\operatorname{\mathbf{grad}} U - \dot{\mathbf{A}}, \qquad \mathbf{B} = \operatorname{\mathbf{curl}} \mathbf{A}$$

is a solution of the Maxwell equations

$$\operatorname{div} \mathbf{E} = \varrho, \qquad \operatorname{div} \mathbf{B} = 0,$$
$$\operatorname{\mathbf{curl}} \mathbf{B} = \mathbf{J}_{\mathrm{QED}} + \dot{\mathbf{E}}, \qquad \operatorname{\mathbf{curl}} \mathbf{E} = -\dot{\mathbf{B}}.$$

Summarizing, the solutions of the Euler–Lagrange equations from Theorem 11.1 on page 795 do not always possess a physical meaning, since the Maxwell equations are violated. This situation changes if we add the Lorenz gauge condition. Then we obtain an electromagnetic field which satisfies the Maxwell equations.

Relativistic invariance. In this chapter, the basic equations of quantum electrodynamics were formulated in a fixed inertial system. It remains to show that these equations have the same form in any inertial system. From the mathematical point of view, we have to indicate the transformation laws for the physical quantities under a change of the inertial system via a Poincaré transformation, and we have to show the invariance of the basic equations of quantum electrodynamics under the Poincaré group. This will be thoroughly studied in Vol. III by using both classical tensor analysis (including van der Waerden's spinor calculus) and Cartan's exterior calculus of differential forms (including Kähler's interior calculus of differential forms).

12. The Free Quantum Field of Electrons, Positrons, and Photons

> Quantum fields possess an infinite number of degrees of freedom. However, in order to overcome serious difficulties, it is wise to start with a finite number of degrees of freedom and to study the lattice limit (continuum limit) for such quantities which can be measured in physical experiments.
>
> Folklore

12.1 Classical Free Fields

For vanishing coupling constant, $\kappa_{\mathrm{QED}} = 0$ (free fields), we want to consider solutions of the classical field equations in the form of a finite Fourier series.

Plane electromagnetic waves. Let $\omega_{\mathbf{p}} := |\mathbf{p}|$. For the field

$$\boxed{U(\mathbf{x},t) := U_0 \mathrm{e}^{\mathrm{i}\mathbf{p}\mathbf{x}}\, \mathrm{e}^{-\mathrm{i}\omega_{\mathbf{p}} t}, \qquad \mathbf{A}(\mathbf{x},t) := \mathbf{A}_0 \mathrm{e}^{\mathrm{i}\mathbf{p}\mathbf{x}}\, \mathrm{e}^{-\mathrm{i}\omega_{\mathbf{p}} t},} \tag{12.1}$$

the following are true.

(W) Wave equations: $\Box U = 0$ and $\Box \mathbf{A} = 0$ on $\mathbb{R}^4$.

(G) Lorenz gauge condition: If $\omega_{\mathbf{p}} U_0 = \mathbf{p}\mathbf{A}_0$, then

$$\dot{U} + \operatorname{div} \mathbf{A} = 0 \qquad \text{on} \quad \mathbb{R}^4.$$

In fact, this implies $\Delta U = -U_{xx} - U_{yy} - U_{zz} = \mathbf{p}^2 U$. Hence

$$\Box U = \ddot{U} + \Delta U = (-\omega_{\mathbf{p}}^2 + \mathbf{p}^2) U = 0.$$

The statements (W) and (G) remain true if we replace $U, \mathbf{A}$ by the conjugate-complex fields $U^\dagger, \mathbf{A}^\dagger$. Explicitly,

$$U(\mathbf{x},t)^\dagger := U_0^\dagger \mathrm{e}^{-\mathrm{i}\mathbf{p}\mathbf{x}}\, \mathrm{e}^{\mathrm{i}\omega_{\mathbf{p}} t}, \qquad \mathbf{A}(\mathbf{x},t)^\dagger := \mathbf{A}_0^\dagger \mathrm{e}^{-\mathrm{i}\mathbf{p}\mathbf{x}}\, \mathrm{e}^{\mathrm{i}\omega_{\mathbf{p}} t}.$$

Our next goal is to consider finite sums of such plane wave solutions. To this end, we need some preparations.

12.1.1 The Lattice Strategy in Quantum Electrodynamics

Our basic strategy reads as follows:

(i) Periodicity of free fields: Classical free fields are represented by *finite Fourier series*. To this end, we put the free fields of photons and electrons in a box of length L and volume $\mathcal{V} := L^3$, and we assume that the fields have the period L with respect to the position variables. Moreover, we introduce a lattice in momentum space. This yields finite Fourier series.

(ii) The crucial method of Fourier quantization:
We replace the Fourier coefficients of the finite Fourier series by creation and annihilation operators.
The advantage of this construction is the fact that a fixed finite number of creation and annihilation operators can be uniquely realized, up to unitary equivalence. For bosons (i.e., photons), this follows from the Stone–von Neumann theorem on the realization of a finite number of commutation relations (see Sect. 7.16).[1]

(iii) The Dyson series for the S-matrix: Concerning a finite number of creation and annihilation operators, the S-matrix is a well-defined finite series in each order of perturbation theory. This leads to well-defined discrete Feynman propagators in the form of discrete integrals (i.e., finite sums) (see Chap. 13).

(iv) Feynman rules: The Wick theorem provides us with well-defined discrete integrals. This way, for example, we get well-defined cross sections in each order of perturbation theory.

(v) Continuum limit for the cross sections: Finally, we have to carry out the delicate limit
$$L \to +\infty,$$
that is, the period L goes to infinity. As a typical example, we will study this limit for the cross section of Compton scattering in order to get the Klein–Nishina formula (see Sects. 14.4 and 15.1).

(vi) Renormalization: In general, that is, in higher order of perturbation theory, an immediate computation of the continuum limit is not possible, since the corresponding limit integrals are divergent. Therefore, we have to apply the procedure of renormalization theory (see Chap. 16).

(vii) Relativistic invariance: The choice of the lattice in momentum space destroys the relativistic invariance of the approach. Therefore, we have to check that the final cross sections are relativistically invariant, that is, the final formulas for the cross sections do not depend on the choice of the inertial system. Roughly speaking, this can be ensured by using discrete formulas which are obtained by discretizing relativistically invariant formulas. We will proceed this way.

(viii) Gauge invariance: Our computations will be based on a fixed choice of the gauge condition for the four-potential of the electromagnetic field. A perfect theory has to show that the physically relevant quantities measured in physical experiments (e.g., the cross section) are gauge invariant, that is, they do not depend on the choice of the gauge condition. We will come back to this problem in later volumes in connection with the study of the Standard Model in particle physics.

Our mathematical language will be close to the language used by physicists. To this end, we will use the discrete Dirac calculus introduced in Sect. 12.1 of Vol. I. Let us recall some basic definitions.

The normalization volume $\mathcal{V}$. Choose a fixed inertial system. The point $x = (x^0, x^1, x^2, x^3)$ describes space and time in the inertial system. That is, the

[1] Note that, for an infinite number of commutation relations, the uniqueness statement fails. There exist unitarily inequivalent representations, as was shown by Gårding and Wightman.
L. Gårding and A. Wightman, Representations of the anticommutation relations, Proc. Nat. Acad. Sci. U.S.A. **40** (1954), 617–621.
L. Gårding and A. Wightman, Representations of the commutation relations, Proc. Nat. Acad. Sci. U.S.A. **40** (1954), 622–625.

position coordinates x^1, x^2, x^3 are right-handed Cartesian coordinates.[2] Moreover, we introduce the time coordinate $x^0 := t$. The position vector

$$\mathbf{x} = x^1\mathbf{i} + x^2\mathbf{j} + x^3\mathbf{k}$$

refers to the right-handed orthonormal system $\mathbf{i}, \mathbf{j}, \mathbf{k}$. The same is true for the momentum vector

$$\mathbf{p} = p^1\mathbf{i} + p^2\mathbf{j} + p^3\mathbf{k}.$$

We fix the length parameter $L > 0$, and we introduce the cube

$$\mathcal{C}(L) := \left\{(x^1, x^2, x^3) \in \mathbb{R}^3 : \ -\tfrac{L}{2} \le x^1, x^2, x^3 \le \tfrac{L}{2}\right\} \tag{12.2}$$

of side length L. The volume of the cube, $\mathcal{V} := L^3$, is called the normalization volume.

The truncated lattice $\mathcal{G}(N)$ in momentum space. Fix the natural number $N = 1, 2, \dots$ Let m^1, m^2, m^3 be integers. By definition, the symbol $\mathcal{G}(N)$ denotes the set of all lattice momentum vectors

$$\mathbf{p} := \frac{2\pi}{L} \cdot (m^1\mathbf{i} + m^2\mathbf{j} + m^3\mathbf{k})$$

with $|m^j| \le N$, $j = 1, 2, 3$. Moreover, we set

$$\Delta^3 p := \Delta p^1 \Delta p^2 \Delta p^3, \qquad \Delta p^j := \frac{2\pi}{L}, \qquad j = 1, 2, 3.$$

The crucial limit

$$L \to +\infty, \qquad N \to +\infty$$

sends the truncated grid $\mathcal{G}(N)$ to $\mathbb{R}^3$. Therefore, we call this the continuum limit.

Discrete Fourier series. For given complex numbers $a_{\mathbf{p}}$, the function

$$\boxed{f(\mathbf{x}) := \frac{1}{\sqrt{(2\pi)^3}} \sum_{\mathbf{p} \in \mathcal{G}(N)} a_{\mathbf{p}}\, \mathrm{e}^{\mathrm{i}\mathbf{p}\mathbf{x}}} \tag{12.3}$$

has the period L with respect to the variables x^1, x^2, x^3. The Fourier coefficients are related to the function f by

$$a_{\mathbf{p}} = \frac{1}{\sqrt{(2\pi)^3}} \int_{\mathcal{C}(L)} f(\mathbf{x})\, \mathrm{e}^{-\mathrm{i}\mathbf{p}\mathbf{x}}\, d^3\mathbf{x}, \qquad \mathbf{p} \in \mathcal{G}(N).$$

The map $f \mapsto \{a_{\mathbf{p}}\}_{\mathbf{p} \in \mathcal{G}(N)}$ is called the discrete Fourier transform of the given function f.

The discrete Dirac delta function in position space. Define

$$\delta_{\mathcal{C}(L)}(\mathbf{y} - \mathbf{x}) := \frac{1}{(2\pi)^3} \sum_{\mathbf{p} \in \mathcal{G}(N)} \mathrm{e}^{\mathrm{i}(\mathbf{y}-\mathbf{x})\mathbf{p}}\, \Delta^3 p$$

for all position vectors $\mathbf{x}$ and $\mathbf{y}$. This is called the discrete Dirac delta function in position space. To simplify notation, we do not indicate that $\delta_{\mathcal{C}(L)}$ also depends on N (the size of the grid $\mathcal{G}(N)$ in momentum space). For each function f of the form (12.3), we obtain the following integral formula:

[2] By our general convention (10.17) for the choice of the physical units, we set $c = 1$ and $\hbar = 1$.

$$f(\mathbf{y}) = \int_{\mathcal{C}(L)} \delta_{\mathcal{C}(L)}(\mathbf{y}-\mathbf{x}) f(\mathbf{x})\, d^3\mathbf{x}.$$

Here, $\mathbf{y}$ denotes an arbitrary position vector.

The discrete Dirac delta function in 3-dimensional momentum space. Define

$$\delta_{\mathcal{G}(N)}(\mathbf{p}-\mathbf{q}) := \frac{\delta_{\mathbf{pq}}}{\Delta^3 p}$$

for all lattice vectors $\mathbf{p}, \mathbf{q} \in \mathcal{G}(N)$. This is a rescaled Kronecker symbol. For all functions $g : \mathcal{G}(N) \to \mathbb{C}$, we get

$$g(\mathbf{q}) = \sum_{\mathbf{p}\in\mathcal{G}(N)} \delta_{\mathcal{G}(N)}(\mathbf{q}-\mathbf{p}) g(\mathbf{p})\, \Delta^3 p.$$

The definition and the properties of the 4-dimensional discrete Dirac delta function δ_{dis} can be found in Sect. 12.1.2 of Vol. I.

12.1.2 The High-Energy Limit and the Low-Energy Limit

The classical limit. Fix the position vector $\mathbf{y}$.

Proposition 12.1 *For all test functions $\varphi \in \mathcal{D}(\mathbb{R}^3)$, we have the classical limit*

$$\lim_{N\to\infty} \lim_{L\to+\infty} \int_{\mathcal{C}(L)} \delta_{\mathcal{C}(L)}(\mathbf{y}-\mathbf{x})\, \varphi(\mathbf{x})\, d^3\mathbf{x} := \varphi(\mathbf{y}). \tag{12.4}$$

Proof. Since the function φ vanishes outside a sufficiently large ball, we obtain

$$\int_{\mathcal{C}(L)} \delta_{\mathcal{C}(L)}(\mathbf{y}-\mathbf{x})\, \varphi(\mathbf{x})\, d^3\mathbf{x} := \int_{\mathbb{R}^3} \delta_{\mathcal{C}(L)}(\mathbf{y}-\mathbf{x})\, \varphi(\mathbf{x})\, d^3\mathbf{x}$$

if L is sufficiently large. Since the function φ lives in the space $\mathcal{S}(\mathbb{R}^3)$ of rapidly decreasing smooth functions at infinity, the same is true for the classical Fourier transform $\hat{\varphi}$ of φ. By Fourier transformation, we get

$$\begin{aligned}\int_{\mathbb{R}^3} \delta_{\mathcal{C}(L)}(\mathbf{y}-\mathbf{x})\varphi(\mathbf{x}) d^3\mathbf{x} &= \frac{1}{(2\pi)^3} \sum_{\mathbf{p}\in\mathcal{G}(N)} \Delta^3 p \int_{\mathbb{R}^3} \mathrm{e}^{\mathrm{i}(\mathbf{y}-\mathbf{x})\mathbf{p}} \varphi(\mathbf{x})\, d^3\mathbf{x} \\ &= \frac{1}{\sqrt{(2\pi)^3}} \sum_{\mathbf{p}\in\mathcal{G}(N)} \mathrm{e}^{\mathrm{i}\mathbf{y}\mathbf{p}}\, \hat{\varphi}(\mathbf{p})\, \Delta^3 p.\end{aligned}$$

The limit $L \to +\infty$ implies $\Delta^3 p \to 0$. Hence

$$\lim_{L\to+\infty} \int_{\mathbb{R}^3} \delta_{\mathcal{C}(L)}(\mathbf{y}-\mathbf{x})\varphi(\mathbf{x}) d^3\mathbf{x} = \frac{1}{\sqrt{(2\pi)^3}} \int_{|\mathbf{p}|\le N} \mathrm{e}^{\mathrm{i}\mathbf{y}\mathbf{p}} \hat{\varphi}(\mathbf{p})\, d^3\mathbf{p}.$$

Consequently,

$$\lim_{N\to\infty} \lim_{L\to+\infty} \int_{\mathbb{R}^3} \delta_{\mathcal{C}(L)}(\mathbf{y}-\mathbf{x})\varphi(\mathbf{x}) d^3\mathbf{x} = \frac{1}{\sqrt{(2\pi)^3}} \int_{\mathbb{R}^3} \mathrm{e}^{\mathrm{i}\mathbf{y}\mathbf{p}} \hat{\varphi}(\mathbf{p})\, d^3\mathbf{p}.$$

Using the inverse Fourier transformation, we obtain

$$\lim_{N\to\infty}\lim_{L\to+\infty}\int_{\mathbb{R}^3}\delta_{\mathcal{C}(L)}(\mathbf{y}-\mathbf{x})\varphi(\mathbf{x})d^3\mathbf{x}=\varphi(\mathbf{y}). \tag{12.5}$$

for all test functions $\varphi\in\mathcal{D}(\mathbb{R}^3)$. This finishes the proof. □

The key relation (12.5) tells us that, for fixed position vector $\mathbf{y}$, we have

$$\lim_{N\to+\infty}\lim_{L\to+\infty}\delta_{\mathcal{C}(L)}(\mathbf{y}-\mathbf{x})=\delta(\mathbf{y}-\mathbf{x}),$$

in the sense of the distribution space $\mathcal{D}'(\mathbb{R}^3)$.

The generalized limit in scattering theory. Again fix the position vector $\mathbf{y}$. For all continuous functions $\varphi:\mathbb{R}^3\to\mathbb{C}$, we define the following generalized limit:

$$\boxed{\lim_{N\to\infty}\lim_{L\to+\infty}\int_{\mathcal{C}(L)}\delta_{\mathcal{C}(L)}(\mathbf{y}-\mathbf{x})\ \varphi(\mathbf{x})\ d^3\mathbf{x}:=\varphi(\mathbf{y}).} \tag{12.6}$$

This definition is motivated by Prop. 12.1. Mnemonically, in the spirit of physicists, we use the following formal argument: $L\to+\infty$ implies $\Delta^3 p\to 0$. Hence

$$\lim_{N\to\infty}\lim_{L\to+\infty}\delta_{\mathcal{C}(L)}(\mathbf{y}-\mathbf{x})=\frac{1}{(2\pi)^3}\int_{\mathbb{R}^3}\mathrm{e}^{\mathrm{i}(\mathbf{y}-\mathbf{x})}d^3\mathbf{p}=\delta(\mathbf{y}-\mathbf{x}).$$

Note that the Dirac delta function is even, that is, $\delta(\mathbf{y}-\mathbf{x})=\delta(\mathbf{x}-\mathbf{x})$. Therefore,

$$\lim_{N\to\infty}\lim_{L\to+\infty}\int_{\mathcal{C}(L)}\delta_{\mathcal{C}(L)}(\mathbf{y}-\mathbf{x})\varphi(\mathbf{x})\ d^3\mathbf{x}=\int_{\mathbb{R}^3}\delta(\mathbf{y}-\mathbf{x})\varphi(\mathbf{x})d^3\mathbf{x}=\varphi(\mathbf{y}).$$

This is the key formula (12.6).

We will use similar arguments for computing the following physical effects in Chap. 15 via the continuum limit $L\to+\infty$ and $N\to+\infty$:

- the Klein–Nishina cross section formula for Compton scattering,
- the cross section for Cherenkov scattering,
- the scattering of electrons in the Yukawa potential,
- the scattering of electrons in the Coulomb potential,
- the emission rate formula for the spontaneous emission of photons by atoms, and
- the Heisenberg radiation formula for the intensity of spectral lines of atoms.

In physics, one uses the following terminology:

- The limit $N\to+\infty$ is called the large-momenta limit (or the high-energy limit).
- The limit $L\to+\infty$ (i.e., the normalization volume $\mathcal{V}$ goes to infinity) is called the small-momenta limit, $\Delta p\to 0$ (or the low-energy limit).

This will be studied in Chap 16.

12.1.3 The Free Electromagnetic Field

Let us start with the special free electromagnetic field (12.1) on page 799. Our goal is to use the superposition principle in order to construct more general free fields in the form of finite Fourier series. To this end, for given momentum vector $\mathbf{p}$, choose three vectors

$$\mathbf{e}_3(\mathbf{p}):=\frac{\mathbf{p}}{|\mathbf{p}|},\quad \mathbf{e}_1(\mathbf{p}),\quad \mathbf{e}_2(\mathbf{p})$$

which form a right-handed orthonormal system (Fig. 12.1). Fix $N=1,2,\ldots$, and

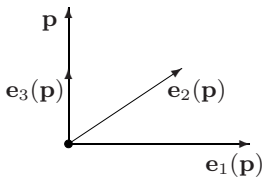


Fig. 12.1. Polarization of photons

choose the grid $\mathcal{G}(N)$ in momentum space. Define the real field $U, \mathbf{A}$ given by

$$U(\mathbf{x},t) := \sum_{\mathbf{p}\in\mathcal{G}(N)} (a_{\mathbf{p},0}\, \mathrm{e}^{\mathrm{i}(\mathbf{px}-\omega_{\mathbf{p}}t)} + a^{\dagger}_{\mathbf{p},0}\, \mathrm{e}^{-\mathrm{i}(\mathbf{px}-\omega_{\mathbf{p}}t)})\, \mathcal{N}_{\mathbf{p}}, \tag{12.7}$$

and

$$\mathbf{A}(\mathbf{x},t) := \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=1}^{3} (a_{\mathbf{p},s}\, \mathrm{e}^{\mathrm{i}(\mathbf{px}-\omega_{\mathbf{p}}t)} + a^{\dagger}_{\mathbf{p},s}\, \mathrm{e}^{-\mathrm{i}(\mathbf{px}-\omega_{\mathbf{p}}t)})\, \mathcal{N}_{\mathbf{p}}\, \mathbf{e}_s(\mathbf{p})$$

where $a_{\mathbf{p},s}$ are given complex numbers. The index s refers to the polarization of the plane waves. We use the normalization factor

$$\mathcal{N}_{\mathbf{p}} := \sqrt{\frac{1}{2L^3\omega_{\mathbf{p}}}} \qquad \text{for all} \quad \mathbf{p}\neq 0$$

in order to simplify the formulas of quantized free fields to be considered in Sect. 12.2.1. For $\mathbf{p}=0$, we set $\mathcal{N}_0 := 0$.

Theorem 12.2 *The field $U, \mathbf{A}$ satisfies the wave equations*

$$\Box U = 0, \qquad \Box \mathbf{A} = 0 \qquad \text{on} \quad \mathbb{R}^4.$$

If $a_{\mathbf{p},0} = a_{\mathbf{p},3}$ for all momentum vectors $\mathbf{p}\in\mathcal{G}(N)$, then the Lorenz gauge condition $\dot{U} + \operatorname{div}\mathbf{A} = 0$ is satisfied on $\mathbb{R}^4$.

Proof. Concerning the Lorenz gauge condition, note that

$$\begin{aligned} \frac{\partial}{\partial t}\, \mathrm{e}^{\mathrm{i}(\mathbf{px}-\omega_{\mathbf{p}}t)} &= -\mathrm{i}\omega_{\mathbf{p}}\, \mathrm{e}^{\mathrm{i}(\mathbf{px}-\omega_{\mathbf{p}}t)}, \\ \operatorname{div}_{\mathbf{x}} \mathbf{e}_s(\mathbf{p})\mathrm{e}^{\mathrm{i}(\mathbf{px}-\omega_{\mathbf{p}}t)} &= \mathrm{i}\mathbf{e}_s(\mathbf{p})\mathbf{p}\, \mathrm{e}^{\mathrm{i}(\mathbf{px}-\omega_{\mathbf{p}}t)} \end{aligned}$$

along with $\mathbf{e}_3(\mathbf{p})\mathbf{p} = |\mathbf{p}| = \omega_{\mathbf{p}}$, and $\mathbf{e}_1(\mathbf{p})\mathbf{p} = \mathbf{e}_2(\mathbf{p})\mathbf{p} = 0$. □

Introducing the components

$$\mathbf{e}_s(\mathbf{p}) := e^1_s(\mathbf{p})\mathbf{i} + e^2_s(\mathbf{p})\mathbf{j} + e^3_s(\mathbf{p})\mathbf{k}, \qquad s = 1,2,3,$$

and $\mathbf{A} = A^1\mathbf{i} + A^2\mathbf{j} + A^3\mathbf{k}$, we get

$$\mathbf{A}^j(\mathbf{x},t) := \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=1}^{3} (a_{\mathbf{p},s}\, \mathrm{e}^{\mathrm{i}(\mathbf{px}-\omega_{\mathbf{p}}t)} + a^{\dagger}_{\mathbf{p},s}\, \mathrm{e}^{-\mathrm{i}(\mathbf{px}-\omega_{\mathbf{p}}t)})\, \mathcal{N}_{\mathbf{p}}\, \mathbf{e}^j_s(\mathbf{p})$$

for $j = 1, 2, 3$. In order to obtain a 4-dimensional formulation, we define

$$e_0^0(\mathbf{p}) := 1, \quad e_0^j(\mathbf{p}) = e_j^0(\mathbf{p}) := 0, \qquad j = 1, 2, 3.$$

Letting $A^0 := U$, we get

$$\boxed{A^\mu(\mathbf{x}, t) := \sum_{\mathbf{p} \in \mathcal{G}(N)} \sum_{s=0}^{3} (a_{\mathbf{p},s}\, \mathrm{e}^{\mathrm{i}(\mathbf{px} - \omega_\mathbf{p} t)} + a_{\mathbf{p},s}^\dagger\, \mathrm{e}^{-\mathrm{i}(\mathbf{px} - \omega_\mathbf{p} t)})\, \mathcal{N}_\mathbf{p}\, \mathbf{e}_s^\mu(\mathbf{p})}$$

for $\mu = 0, 1, 2, 3$. From the orthogonality relation, $\mathbf{e}_r(\mathbf{p})\mathbf{e}_s(\mathbf{p}) = \delta_{rs}$, we obtain the so-called completeness relation:

$$\boxed{\sum_{s=0}^{3} e_s^\mu e_s^\nu = \delta^{\mu\nu}, \qquad \mu, \nu = 0, 1, 2, 3.} \tag{12.8}$$

The importance of virtual photons. The index s indicates the polarization of the field A^0, A^1, A^2, A^3. We have to distinguish the following three cases.

- For $s = 1, 2$, the field $\mathbf{A}$ is transversal to the direction $\mathbf{p}$ of the propagation of the plane wave (Fig. 12.1). Therefore, we speak of transversal polarization. For $s = 3$, the field $\mathbf{A}$ is parallel to the direction $\mathbf{p}$ of the propagation of the plane wave. We speak of longitudinal polarization.
- For $s = 0$, we speak of scalar polarization.

After quantization to be carried out below, plane electromagnetic waves pass over to photons (i.e, light particles).

> *Transversal (resp. longitudinal, scalar) polarization will then correspond to transversal (resp. longitudinal, scalar) photons.*

Longitudinal and scalar photons are also called virtual photons. Observe that virtual photons are nonphysical objects. They have to be eliminated by the Gupta–Bleuler procedure which will be considered in Sect. 12.4.4 on page 831.

> *The introduction of virtual physical states, also called ghosts, is typical for the quantization of gauge field theories like quantum electrodynamics and the more general Standard Model in particle physics.*

Note that this is more than a mathematical trick. It will be shown below that

- virtual particles do never appear as incoming and outgoing particles in scattering processes;
- nevertheless, virtual particles are responsible for real physical effects in quantum field theory via quantum fluctuations.

Mathematically, this depends on the fact that virtual particles contribute to the Feynman propagators which govern the computations of physical effects in perturbation theory (see Sect. 13.4.1). Intuitively, virtual particles correspond to internal lines of Feynman diagrams (see Chap. 14).

12.1.4 The Free Electron Field

Plane wave solutions of the Dirac equation for the free electron. The following explicit solutions $\psi^{\pm}_{\mathbf{p},s}$ of the Dirac equation play a fundamental role for computing scattering processes in quantum electrodynamics and, more generally, in the Standard model in particle physics. For fixed momentum vector $\mathbf{p}$, define

$$\boxed{\psi^{+}_{\mathbf{p},s}(\mathbf{x},t) := u_{\mathbf{p},s}\, \mathrm{e}^{\mathrm{i}\mathbf{p}\mathbf{x}} \mathrm{e}^{-\mathrm{i}E_{\mathbf{p}}t}}$$

and

$$\boxed{\psi^{-}_{\mathbf{p},s}(\mathbf{x},t) := v_{\mathbf{p},s}\, \mathrm{e}^{-\mathrm{i}\mathbf{p}\mathbf{x}} \mathrm{e}^{\mathrm{i}E_{\mathbf{p}}t}.}$$

Here, the energy is given by $E_{\mathbf{p}} := \sqrt{\mathbf{p}^2 + m_e^2}$ where m_e denotes the mass of the electron. Furthermore,

$$\chi_{\frac{1}{2}} := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \chi_{-\frac{1}{2}} := \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

along with

$$u_{\mathbf{p},s} := \mathcal{N} \begin{pmatrix} \chi_s \\ \frac{\mathbf{p}\boldsymbol{\sigma}}{E_{\mathbf{p}}+m_e}\, \chi_s \end{pmatrix}, \qquad v_{\mathbf{p},s} := \mathcal{N} \begin{pmatrix} \frac{\mathbf{p}\boldsymbol{\sigma}}{E_{\mathbf{p}}+m_e}\, \chi_{-s} \\ \chi_{-s} \end{pmatrix}, \tag{12.9}$$

and the normalization factor $\mathcal{N} := \sqrt{E_{\mathbf{p}} + m_e}$. We also set $\mathbf{p} = p^1\mathbf{i} + p^2\mathbf{j} + p^3\mathbf{k}$, and

$$\mathbf{p}\boldsymbol{\sigma} = \sum_{j=1}^{3} p^j \sigma^j = \begin{pmatrix} p^3 & p^1 - \mathrm{i}p^2 \\ p^1 + \mathrm{i}p^2 & -p^3 \end{pmatrix}.$$

Explicitly, we obtain the following four matrices:[3]

$$u_{\mathbf{p},\frac{1}{2}} := \mathcal{N} \begin{pmatrix} 1 \\ 0 \\ \frac{p^3}{E_{\mathbf{p}}+m_e} \\ \frac{p^1+\mathrm{i}p^2}{E_{\mathbf{p}}+m_e} \end{pmatrix}, \qquad u_{\mathbf{p},-\frac{1}{2}} := \mathcal{N} \begin{pmatrix} 0 \\ 1 \\ \frac{p^1-\mathrm{i}p^2}{E_{\mathbf{p}}+m_e} \\ -\frac{p^3}{E_{\mathbf{p}}+m_e} \end{pmatrix},$$

$$v_{\mathbf{p},\frac{1}{2}} := \mathcal{N} \begin{pmatrix} \frac{p^1-\mathrm{i}p^2}{E_{\mathbf{p}}+m_e} \\ -\frac{p^3}{E_{\mathbf{p}}+m_e} \\ 0 \\ 1 \end{pmatrix}, \qquad v_{\mathbf{p},-\frac{1}{2}} := \mathcal{N} \begin{pmatrix} \frac{p^3}{E_{\mathbf{p}}+m_e} \\ \frac{p^1+\mathrm{i}p^2}{E_{\mathbf{p}}+m_e} \\ 1 \\ 0 \end{pmatrix}.$$

[3] Alternatively, the modified normalization factor $\mathcal{N} := \sqrt{E_{\mathbf{p}} + m_e}/\sqrt{2m_e}$ is used in the literature. Since this normalization becomes meaningless for neutrinos with vanishing mass, $m_e = 0$, we do not use this modified normalization.

The Dirac equation $\mathrm{i}\gamma^\alpha\partial_\alpha\psi = m_e\psi$ can equivalently be written in the form of the following Schrödinger equation

$$\boxed{\mathrm{i}\dot{\psi} = \mathbf{H}\psi} \tag{12.10}$$

with the Hamiltonian

$$\mathbf{H} := \gamma^0 m_e + \sum_{j=1}^{3} \gamma^0\gamma^j(-\mathrm{i}\partial_j) = m_e \begin{pmatrix} \sigma^0 & 0 \\ 0 & -\sigma^0 \end{pmatrix} - \mathrm{i} \begin{pmatrix} 0 & \boldsymbol{\sigma\partial} \\ \boldsymbol{\sigma\partial} & 0 \end{pmatrix}$$

where $\boldsymbol{\sigma\partial} := \sum_{j=1}^{3} \sigma^j\partial_j$. To this end, multiply the Dirac equation by γ^0 from left and use $(\gamma^0)^2 = I$.[4] Analogously, introduce the operator

$$\mathbf{H_p} := \gamma^0 m_e + \sum_{j=1}^{3} \gamma^0\gamma^j p^j = m_e \begin{pmatrix} \sigma^0 & 0 \\ 0 & -\sigma^0 \end{pmatrix} + \begin{pmatrix} 0 & \mathbf{p\sigma} \\ \mathbf{p\sigma} & 0 \end{pmatrix}.$$

For $\varphi_{\mathbf{p}}(\mathbf{x}) := \mathrm{e}^{\mathrm{i}\mathbf{px}}$ and the constant column matrix $u \in \mathbb{C}^4$, note that

$$\mathbf{H}(u\varphi_{\mathbf{p}}) = (\mathbf{H_p}u)\varphi_{\mathbf{p}}.$$

In Problem 15.9 on page 938, we will prove the following result which is valid for all momentum vectors $\mathbf{p}, \mathbf{q}$, and all spin numbers $r, s = \pm\frac{1}{2}$.

Theorem 12.3 (i) *The functions $\psi^+_{\mathbf{p},s}$ and $\psi^-_{\mathbf{p},s}$ are solutions of the Dirac equation (12.10) above. Furthermore, we have*

$$\mathbf{H}\psi^\pm_{\mathbf{p},s} = \pm E_{\mathbf{p}}\psi^\pm_{\mathbf{p},s}$$

along with

$$\mathbf{H_p}u_{\mathbf{p},s} = E_{\mathbf{p}}u_{\mathbf{p},s}, \qquad \mathbf{H_p}v_{-\mathbf{p},s} = -E_{\mathbf{p}}v_{-\mathbf{p},s}.$$

(ii) *The four column matrices*

$$u_{\mathbf{p},\frac{1}{2}},\ u_{\mathbf{p},-\frac{1}{2}},\ v_{-\mathbf{p},\frac{1}{2}},\ v_{-\mathbf{p},-\frac{1}{2}}$$

form an orthogonal basis of eigenvectors of the self-adjoint matrix $H_{\mathbf{p}}$ with respect to the inner product $u^\dagger u$ on the complex Hilbert space $\mathbb{C}^4$. The normalization conditions read as

$$u^\dagger_{\mathbf{p},s}u_{\mathbf{p},s} = v^\dagger_{-\mathbf{p},s}v_{-\mathbf{p},s} = 2E_{\mathbf{p}}, \qquad s = \pm\tfrac{1}{2}.$$

There holds the following completeness relation:

$$\sum_{s=\pm\frac{1}{2}} u_{\mathbf{p},s}u^\dagger_{\mathbf{p},s} + v_{-\mathbf{p},s}v^\dagger_{-\mathbf{p},s} = 2E_{\mathbf{p}}I.$$

(iii) *Additionally, we have the generalized orthogonality relations*

$$\overline{u}_{\mathbf{p},r}v_{\mathbf{p},s} = \overline{v}_{\mathbf{p},r}u_{\mathbf{p},s} = 0, \qquad \overline{u}_{\mathbf{p},r}u_{\mathbf{p},s} = -\overline{v}_{\mathbf{p},r}v_{\mathbf{p},s} = 2m_e\delta_{rs},$$

along with the spin sum formulas [5]

$$\sum_{s=\pm\frac{1}{2}} u_{\mathbf{p},s}\overline{u}_{\mathbf{p},s} = \not{p} + m_e, \qquad \sum_{s=\pm\frac{1}{2}} v_{\mathbf{p},s}\overline{v}_{\mathbf{p},s} = \not{p} - m_e.$$

[4] Briefly, $\mathbf{H} = m_e\gamma^0 - \mathrm{i}\gamma^0\boldsymbol{\gamma\partial}$.

[5] We use Feynman's slash symbol, $\not{p} := \gamma^\mu p_\mu$.

Physical properties of plane electron waves. Consider first the wave function

$$\psi^+_{\mathbf{p},s}(\mathbf{x},t) = u_{\mathbf{p},s}\mathrm{e}^{\mathrm{i}\mathbf{p}\mathbf{x}}\mathrm{e}^{-\mathrm{i}E_{\mathbf{p}}t}, \qquad s = \pm\tfrac{1}{2}.$$

The physics of plane electron waves can be described in terms of the following four operators:

(M) Momentum operator: $\mathbf{P} := -\mathrm{i}\boldsymbol{\partial}$.
(E) Energy operator: $\mathbf{H} = m_e\gamma^0 + \gamma^0\boldsymbol{\gamma}\mathbf{P}$.
(S) Spin operator: $\mathbf{S} = S^1\mathbf{i} + S^2\mathbf{j} + S^3\mathbf{k}$. Here,

$$S^j := \frac{1}{2}\begin{pmatrix}\sigma^j & 0\\ 0 & \sigma^j\end{pmatrix}, \qquad j = 1,2,3.$$

(H) Helicity operator (spin projection): For nonzero momentum vector $\mathbf{p}$, we define the helicity operator in direction of $\mathbf{p}$ by

$$\mathbf{S}_{\mathbf{p}} := \frac{\mathbf{p}\mathbf{S}}{|\mathbf{p}|} = \frac{1}{2|\mathbf{p}|}\begin{pmatrix}\mathbf{p}\boldsymbol{\sigma} & 0\\ 0 & \mathbf{p}\boldsymbol{\sigma}\end{pmatrix}.$$

Using the Dirac–Pauli matrices, we set

$$\sigma^{\alpha\beta} := \frac{\mathrm{i}}{2}[\gamma^\alpha,\gamma^\beta]_-, \qquad \alpha,\beta = 0,1,2,3.$$

Then, $S^1 = \frac{1}{2}\sigma^{23}$, $S^2 = \frac{1}{2}\sigma^{31}$, and $S^3 = \frac{1}{2}\sigma^{12}$. Note that

$$\boxed{\mathbf{P}\psi^+_{\mathbf{p},s} = \mathbf{p}\psi^+_{\mathbf{p},s}, \qquad \mathbf{H}\psi^+_{\mathbf{p},s} = E_{\mathbf{p}}\psi^+_{\mathbf{p},s}, \qquad s = \pm\tfrac{1}{2}.}$$

In terms of physics, the field $\psi^+_{\mathbf{p},s}$ possesses the momentum vector $\mathbf{p}$ and the energy $E_{\mathbf{p}}$. For example, suppose that the momentum vector $\mathbf{p}$ points in direction of the z-axis, that is, $\mathbf{p} = p^3\mathbf{k}$ with $p^3 > 0$. Then

$$\boxed{\mathbf{S}\psi^+_{\mathbf{p},s} = (s\mathbf{k})\psi^+_{\mathbf{p},s}, \qquad \mathbf{S}_{\mathbf{p}}\psi^+_{\mathbf{p},s} = s\psi^+_{\mathbf{p},s}, \qquad s = \pm\tfrac{1}{2}.}$$

We say that the electron state $\psi^+_{\mathbf{p},s}$ with $\mathbf{p} = p^3\mathbf{k}$, $p^3 > 0$ has the spin vector $s\mathbf{k}$ and the helicity s. For $s = \frac{1}{2}$ (resp. $s = -\frac{1}{2}$) we speak of a spin-up (resp. spin-down) state.

The trouble with negative energies. Consider now the wave function

$$\psi^-_{\mathbf{p},s}(\mathbf{x},t) = v_{\mathbf{p},s}\mathrm{e}^{-\mathrm{i}\mathbf{p}\mathbf{x}}\mathrm{e}^{\mathrm{i}E_{\mathbf{p}}t}.$$

From the mathematical point of view, this is a solution of the Dirac equation $\mathrm{i}\dot{\psi} = \mathbf{H}\psi$ with

$$\boxed{\mathbf{P}\psi^-_{\mathbf{p},s} = -\mathbf{p}\psi^-_{\mathbf{p},s}, \qquad \mathbf{H}\psi^-_{\mathbf{p},s} = -E_{\mathbf{p}}\psi^-_{\mathbf{p},s}, \qquad s = \tfrac{1}{2}.}$$

Suppose that the momentum vector $\mathbf{p}$ points in direction of the z-axis, i.e., $\mathbf{p} = p^3\mathbf{k}$ with $p^3 > 0$. Then

$$\boxed{\mathbf{S}\psi^-_{\mathbf{p},s} = -(s\mathbf{k})\psi^-_{\mathbf{p},s}, \qquad \mathbf{S}_{\mathbf{p}}\psi^-_{\mathbf{p},s} = -s\psi^-_{\mathbf{p},s}, \qquad s = \pm\tfrac{1}{2}.}$$

From the physical point of view, we encounter the trouble that because of

$$\mathbf{H}\psi^-_{\mathbf{p},s} = -E_{\mathbf{p}}\psi^-_{\mathbf{p},s},$$

the function $\psi^-_{\mathbf{p},s}$ corresponds to the negative energy $-E_{\mathbf{p}}$. In 1928, Dirac did overcome this difficulty by postulating that there exists an antiparticle to the electron which is called the positron nowadays.

The function $\psi^-_{\mathbf{p},s}$ corresponds to a positron of positive energy $E_{\mathbf{p}}$ and momentum vector $\mathbf{p}$.

In the case where $\mathbf{p} = p^3\mathbf{k}$ with $p^3 > 0$, the function $\psi^-_{\mathbf{p},s}$ describes a positron of spin vector $s\mathbf{k}$ and helicity s with $s = \pm\frac{1}{2}$. In the 1940s, Stueckelberg (1905–1984) emphasized the following point of view:

The positron is an electron running backwards in time.

This refers to the following mathematical fact. For given positron function $\psi^-_{\mathbf{p},s}$, define

$$\psi(\mathbf{x}, t) := \psi^-_{\mathbf{p},s}(\mathbf{x}, -t), \qquad t \in \mathbb{R}.$$

Then, the function ψ is a solution of the Dirac equation $\mathrm{i}\dot{\psi} = \mathbf{H}\psi$ along with

$$\mathbf{H}\psi = E_{\mathbf{p}}\psi, \qquad \mathbf{P}\psi = -\mathbf{p}\psi.$$

In the case where $\mathbf{p} = p^3\mathbf{k}$ with $p^3 > 0$, we get

$$\mathbf{S}\psi = (-s\mathbf{k})\psi, \qquad \mathbf{S}_{\mathbf{p}}\psi = -s\psi.$$

The passage $\mathbf{p} \mapsto -\mathbf{p}$ and $s \mapsto -s$ is motivated by the fact that under the time reflection $t \mapsto -t$, the velocity vector changes sign, and hence both the momentum vector $\mathbf{p}$ and the angular momentum vector (spin) $\mathbf{a}$ of a classical particle change sign, i.e., $\mathbf{p} \mapsto -\mathbf{p}$ and $\mathbf{a} \mapsto -\mathbf{a}$. On the classical level, the approach to positrons sounds artificially. The natural setting for positrons is the passage to quantum field theory via creation and annihilation operators. This will be studied in Sect. 12.2.1 on page 812.

Orthogonality relations. Recall the definition of the cube

$$\mathcal{C}(L) := \left\{(x^1, x^2, x^3) \in \mathbb{R}^3 : \ -\tfrac{L}{2} \le x^1, x^2, x^3 \le \tfrac{L}{2}\right\}$$

of side length L. Introduce the inner product

$$\langle\psi|\varphi\rangle_{\mathcal{C}(L)} := \int_{\mathcal{C}(L)} \psi(\mathbf{x})^\dagger \varphi(\mathbf{x})\, d^3\mathbf{x}.$$

Let $L_{2,\,\mathrm{per}}(\mathcal{C}(L), \mathbb{C}^4)$ denote the space of all of the functions $\psi : \mathcal{C}(L) \to \mathbb{C}^4$ which have the following properties:

- $\langle\psi|\psi\rangle_{\mathcal{C}(L)} < \infty$;
- the function ψ has the period L with respect to the variables x^1, x^2, x^3;
- the components of ψ are measurable with respect to the Lebesgue measure on $\mathbb{R}^3$;
- two functions ψ and φ in $L_{2,\,\mathrm{per}}(\mathcal{C}(L), \mathbb{C}^4)$ are identified with each other iff they differ on a subset of Lebesgue measure zero on $\mathbb{R}^3$.

The space $L_{2,\,\mathrm{per}}(\mathcal{C}(L), \mathbb{C}^4)$ is a complex Hilbert space with respect to the inner product $\langle\psi|\varphi\rangle_{\mathcal{C}(L)}$.

Theorem 12.4 *For all lattice momentum vectors* $\mathbf{p} \in \mathcal{G}(N)$ *and all spin numbers* $s = \pm\frac{1}{2}$*, the fields* $\psi^{\pm}_{\mathbf{p},s}$ *form an orthogonal system in the Hilbert space* $L_{2,\,\mathrm{per}}(\mathcal{C}(L), \mathbb{C}^4)$ *with the normalization condition*

$$\langle \psi^{\pm}_{\mathbf{p},s} | \psi^{\pm}_{\mathbf{p},s} \rangle_{\mathcal{C}(L)} = 2E_{\mathbf{p}} L^3.$$

Explicitly, the orthogonality relations read as

$$\langle \psi^{\pm}_{\mathbf{p},r} | \psi^{\pm}_{\mathbf{q},s} \rangle_{\mathcal{C}(L)} = 2\delta_{\mathbf{pq}}\, \delta_{rs} E_{\mathbf{p}} L^3, \qquad \langle \psi^{\pm}_{\mathbf{p},r} | \psi^{\mp}_{\mathbf{q},s} \rangle_{\mathcal{C}(L)} = 0$$

for all $\mathbf{p}, \mathbf{q} \in \mathcal{G}(N)$ and all $r, s = \pm\frac{1}{2}$.

Proof. This follows by using an explicit computation based on the orthogonality relation

$$\frac{1}{L^3} \int_{\mathcal{C}(L)} \mathrm{e}^{\mathrm{i}(\mathbf{p}-\mathbf{q})\mathbf{x}} d^3\mathbf{x} = \delta_{\mathbf{pq}}, \qquad \mathbf{p}, \mathbf{q} \in \mathcal{G}(N),$$

along with the matrix orthogonality relations

$$u^{\dagger}_{\mathbf{p},\frac{1}{2}} u_{\mathbf{p},-\frac{1}{2}} = v^{\dagger}_{\mathbf{p},\frac{1}{2}} v_{\mathbf{p},-\frac{1}{2}} = 0, \qquad u^{\dagger}_{\mathbf{p},s} v_{-\mathbf{p},s} = 0$$

for all $\mathbf{p} \in \mathcal{G}(N)$ and all $s, r = \pm\frac{1}{2}$.

Alternatively, we can use the following standard argument from functional analysis. Let $\psi, \varphi \in L_{2,\,\mathrm{per}}(\mathcal{C}(L), \mathbb{C}^4)$ be smooth functions. Observing the periodicity of ψ and φ, integration by parts yields

$$\int_{\mathcal{C}(L)} \psi(\mathbf{x})^{\dagger} \partial_j \varphi(\mathbf{x}) d^3\mathbf{x} = -\int_{\mathcal{C}(L)} \partial_j \psi(\mathbf{x})^{\dagger} \varphi(\mathbf{x}) d^3\mathbf{x}$$

for $j = 1, 2, 3$. Hence

$$\boxed{\langle \psi | \mathbf{H}\varphi \rangle_{\mathcal{C}(L)} = \langle \mathbf{H}\psi | \varphi \rangle_{\mathcal{C}(L)}.}$$

This shows that the differential operator $\mathbf{H}$ is formally self-adjoint on the set of smooth functions in the Hilbert space $L_{2,\,\mathrm{per}}(\mathcal{C}(L), \mathbb{C}^4)$. Consequently, the eigenfunctions of the operator $\mathbf{H}$ with respect to different eigenvalues are orthogonal to each other. Finally, note that $\mathbf{H}\psi^{\pm}_{\mathbf{p},s} = \pm E_{\mathbf{p}} \psi^{\pm}_{\mathbf{p},s}$. □

The free Dirac field. We want to construct more general solutions of the Dirac equation by superposition. Fix $N = 1, 2, 3, \ldots$ If $b_{\mathbf{p},s}, c_{\mathbf{p},s}$ are arbitrary complex numbers, then the field

$$\psi := \sum_{\mathbf{p} \in \mathcal{G}(N)} \sum_{s=\pm\frac{1}{2}} (b_{\mathbf{p},s} \psi^{+}_{\mathbf{p},s} + c^{\dagger}_{\mathbf{p},s} \psi^{-}_{\mathbf{p},s}) N_{\mathbf{p}}$$

with the normalization factor

$$N_{\mathbf{p}} := \sqrt{\frac{1}{2L^3 E_{\mathbf{p}}}}$$

is a solution of the Dirac equation $\mathrm{i}\dot{\psi} = \mathbf{H}\psi$. Explicitly,

$$\boxed{\begin{aligned} \psi(\mathbf{x}, t) := \sum_{\mathbf{p} \in \mathcal{G}(N)} \sum_{s=\pm\frac{1}{2}} (b_{\mathbf{p},s} u_{\mathbf{p},s}\, \mathrm{e}^{\mathrm{i}(\mathbf{px} - E_{\mathbf{p}} t)} + \\ + c^{\dagger}_{\mathbf{p},s} v_{\mathbf{p},s}\, \mathrm{e}^{-\mathrm{i}(\mathbf{px} - E_{\mathbf{p}} t)}) \, N_{\mathbf{p}}. \end{aligned}} \tag{12.11}$$

The adjoint Dirac field. Recall that $\overline{\psi} := \psi^\dagger \gamma^0$. Therefore, it follows from (12.11) that

$$\boxed{\overline{\psi}(\mathbf{x},t) := \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=\pm\frac{1}{2}} (b^\dagger_{\mathbf{p},s}\overline{u}_{\mathbf{p},s}\, \mathrm{e}^{-\mathrm{i}(\mathbf{px}-E_\mathbf{p}t)} + c_{\mathbf{p},s}\overline{v}_{\mathbf{p},s}\, \mathrm{e}^{\mathrm{i}(\mathbf{px}-E_\mathbf{p}t)})\, N_\mathbf{p}.} \tag{12.12}$$

After quantization, the fields ψ and $\overline{\psi}$ will describe free electrons and positrons.

12.2 Quantization

> Quantization is not a handicraft, but an art.
> Folklore

We now use the method of Fourier quantization in order to describe free quantum fields of photons, electrons, and positrons which are the basic particles in quantum electrodynamics. The idea is to consider

- the Maxwell equations in classical electrodynamics with vanishing external electric charge density and vanishing electric current, and
- the Dirac equation for the classical relativistic electron with vanishing external electromagnetic field.

It turns out that there exist solutions in the form of Fourier series. The corresponding free quantum fields are then obtained by replacing

- the Fourier coefficients $a_\mathbf{p}$ by annihilation operators $a^-_\mathbf{p}$, and
- the conjugate-complex Fourier coefficients $a^\dagger_\mathbf{p}$ by creation operators $a^+_\mathbf{p}$.

The normalization factors are chosen in such a way that we obtain the appropriate expression for the energy operator. The classical free fields concern plane electromagnetic waves and plane waves for the classical relativistic electron.

> *The procedure of Fourier quantization leads quite naturally to the antiparticle of the electron called positron.*

The existence of the positron was conjectured by Dirac in 1928. Experimentally, the positron was discovered in cosmic rays by Anderson in 1932.

> *As we will show in Chap. 13 on the Dyson series for the S-matrix, the knowledge of the free quantum fields for photons, electrons, and positrons is basic for computing the interaction between photons, electrons, and positrons in terms of perturbation theory.*

Observe the following peculiarity. In order to simplify the approach, we first include photons which violate the Lorenz gauge condition. In Sect. 12.4.4, we will add the Gupta–Bleuler quantization condition,

$$\boxed{a^-_{\mathbf{p},0}|\Psi\rangle = 0, \qquad a^-_{\mathbf{p},3}|\Psi\rangle = 0,}$$

for characterizing physical states Ψ. This condition eliminates nonphysical photons which do not possess transversal polarization. The Gupta–Bleuler method is the prototype for the quantization of gauge theories by eliminating nonphysical quantum states via the BRST quantization methods based on cohomology.

12.2.1 The Free Photon Quantum Field

Fix $N = 1, 2, 3, \ldots$ Let $\mu = 0, 1, 2, 3$. Motivated by the free electromagnetic field in Sect. 12.1.3 on page 805, we define the free quantum fields of photons in the following way:

$$\boxed{A^\mu(\mathbf{x}, t) := \sum_{\mathbf{p} \in \mathcal{G}(N)} \sum_{s=0}^{3} (a^-_{\mathbf{p},s}\, \mathrm{e}^{\mathrm{i}(\mathbf{p}\mathbf{x} - \omega_{\mathbf{p}} t)} + a^+_{\mathbf{p},s}\, \mathrm{e}^{-\mathrm{i}(\mathbf{p}\mathbf{x} - \omega_{\mathbf{p}} t)})\, \mathcal{N}_{\mathbf{p}} e^\mu_s(\mathbf{p})}$$

along with the following normalization factor

$$\mathcal{N}_{\mathbf{p}} := \sqrt{\frac{1}{2L^3 \omega_{\mathbf{p}}}} \quad \text{for all} \quad \mathbf{p} \neq 0.$$

For $\mathbf{p} = 0$, we set $\mathcal{N}_0 := 0$. We postulate that for all given lattice momentum vectors $\mathbf{p}, \mathbf{q} \in \mathcal{G}(N)$ and all polarization indices $s, r = 0, 1, 2, 3$,

- the photon creation operators $a^+_{\mathbf{p},s}$, and
- the photon annihilation operators $a^-_{\mathbf{p},s}$

satisfy the following commutation relations which will be used frequently:[6]

(P1) $[a^-_{\mathbf{p},s}, a^+_{\mathbf{p},s}]_- = I$ if $s = 1, 2, 3$.
(P2) $[a^-_{\mathbf{p},0}, a^+_{\mathbf{p},0}]_- = -I$.
(P3) $[a^-_{\mathbf{p},s}, a^+_{\mathbf{q},r}]_- = 0$ if $\mathbf{p} \neq \mathbf{q}$ or $s \neq r$.
(P4) $[a^-_{\mathbf{p},s}, a^-_{\mathbf{q},r}]_- = [a^+_{\mathbf{p},s}, a^+_{\mathbf{q},r}]_- = 0$.
(P5) $a^-_{\mathbf{p},s}|0\rangle = 0$.
(P6) $(a^-_{\mathbf{p},s})^\dagger = a^+_{\mathbf{p},s}$.

To motivate these commutation relations, we will show next that (P1)–(P6) imply

- reasonable commutation relations for the free quantum field A^μ of photons and its conjugate field $\Pi^\mu = -\dot{A}^\mu$, and
- a reasonable expression for the energy operator.

The crucial realization of the commutation relations (P1)–(P6) will be studied in Sect. 12.4.3 on page 826. Note the following:

- The polarization index $s = 1, 2$ corresponds to transversal photons.
- The polarization index $s = 0$ (resp. $s = 3$) corresponds to scalar (resp. longitudinal photons.

Transversal photons are called real photons (or physical photons). Both scalar and longitudinal physical photons are called virtual photons (or unphysical photons).

The canonical commutation relations for the free quantum field of photons. Fix time t. For $\mu, \nu = 0, 1, 2, 3$ and all position vectors $\mathbf{x}$ and $\mathbf{y}$, we obtain

$$\boxed{[A^\mu(\mathbf{x}, t), \Pi^\nu(\mathbf{y}, t)]_- = \mathrm{i}\delta_{\mathcal{C}(L)}(\mathbf{x} - \mathbf{y}) \cdot \eta^{\mu\nu} I} \tag{12.13}$$

and

$$\boxed{[A^\mu(\mathbf{x}, t), A^\nu(\mathbf{y}, t)]_- = 0, \qquad [\Pi^\mu(\mathbf{x}, t), \Pi^\nu(\mathbf{y}, t)]_- = 0.} \tag{12.14}$$

[6] Recall that $[A, B]_- := AB - BA$.

Naturally enough, the commutation relations (12.13) for the free quantum field A^μ of photons and its conjugate field $\Pi^\mu = -\dot{A}^\mu$ generalize the Heisenberg commutation relation

$$[q,p]_- = \mathrm{i}I$$

for position q and and the conjugate momentum p in quantum mechanics. Observe the crucial fact that the appearance of the symbol $\eta^{\mu\nu}$ guarantees the appropriate behavior under Lorentz transformations (relativistic invariance). In order to obtain the symbol $\eta^{\mu\nu}$, but not $\delta^{\mu\nu}$, we postulate different signs for the commutation relations in (P1) and (P2) above.

Proof. Let us prove (12.13). The key is the completeness relation

$$\sum_{s=0}^{3} \eta_{ss} e_s^\mu e_s^\nu = \eta^{\mu\nu}, \qquad \mu,\nu = 0,1,2,3$$

which follows from (12.8). By the commutation relation $[a^-_{\mathbf{p},s}, a^+_{\mathbf{p},s}]_- = -\eta_{ss} I$ for $s = 0,1,2,3$, we obtain the following equation:

$$\begin{aligned}[A^\mu(\mathbf{x},t), \Pi^\nu(\mathbf{y},t)]_- &= \sum_{\mathbf{p}\in\mathcal{G}(N)} -\mathrm{i}\omega_{\mathbf{p}} \sum_{s=0}^{3} 2[a^-_{\mathbf{p},s}, a^+_{\mathbf{p},s}]_- \ \mathrm{e}^{\mathrm{i}\mathbf{p}(\mathbf{x}-\mathbf{y})} \ e_s^\mu e_s^\nu \, \mathcal{N}_{\mathbf{p}}^2 \\ &= \frac{\mathrm{i}\eta^{\mu\nu}}{(2\pi)^3} \left(\sum_{\mathbf{p}\in\mathcal{G}(N)} \mathrm{e}^{\mathrm{i}\mathbf{p}(\mathbf{x}-\mathbf{y})} \Delta^3\mathbf{p} \right) I.\end{aligned}$$

This yields the desired relation (12.13). Analogously, we get (12.14). □

The energy operator for the free quantum field of photons. The energy operator $\mathsf{H}_{\mathrm{phot}}$ of the free quantum field for photons is given by

$$\boxed{\mathsf{H}_{\mathrm{phot}} = \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=1}^{3} \omega_{\mathbf{p}} (a^+_{\mathbf{p},s} a^-_{\mathbf{p},s} + \tfrac{1}{2} I) + \omega_{\mathbf{p}} (\tfrac{1}{2} I - a^+_{\mathbf{p},0} a^-_{\mathbf{p},0}).}$$

Let us motivate this. By Theorem 11.2 on page 796, we define

$$\mathsf{H}_{\mathrm{phot}} := \int_{\mathcal{C}(L)} \mathcal{H}_{\mathrm{phot}} \, d^3\mathbf{x} = -\tfrac{1}{2} \int_{\mathcal{C}(L)} \sum_{\mu,\nu=0}^{3} (\partial_\nu A_\mu)^2 d^3\mathbf{x}.$$

This means that[7]

$$\mathsf{H}_{\mathrm{phot}} = -\tfrac{1}{2} \int_{\mathcal{C}(L)} \left(\dot{U}^2 + (\boldsymbol{\partial} U)^2 + \sum_{j=1}^{3} \dot{A}_j^2 + (\boldsymbol{\partial} A_j)^2 \right) d^3\mathbf{x}.$$

The time derivative of $U = A^0$ reads as

$$\dot{U} = \sum_{\mathbf{p}\in\mathcal{G}(N)} (-a^-_{\mathbf{p},0} \, \mathrm{e}^{\mathrm{i}(\mathbf{p}\mathbf{x} - \omega_{\mathbf{p}} t)} + a^+_{\mathbf{p},0} \mathrm{e}^{-\mathrm{i}(\mathbf{p}\mathbf{x} - \omega_{\mathbf{p}} t)}) \, \mathrm{i}\omega_{\mathbf{p}} \mathcal{N}_{\mathbf{p}}.$$

By the orthogonality relations on page 810, we get

[7] Explicitly, $(\boldsymbol{\partial} U)^2 = U_x^2 + U_y^2 + U_z^2$.

$$\int_{\mathcal{C}(L)} \dot{U}^2 \, d^3\mathbf{x} = \sum_{\mathbf{p}\in\mathcal{G}(N)} (a^-_{\mathbf{p},0}a^+_{\mathbf{p},0} + a^+_{\mathbf{p},0}a^-_{\mathbf{p},0}) \, \omega_{\mathbf{p}}^2 \mathcal{N}_{\mathbf{p}}^2 L^3. \tag{12.15}$$

Since $a^-_{\mathbf{p},0}a^+_{\mathbf{p},0} = a^+_{\mathbf{p},0}a^-_{\mathbf{p},0} - I$, we have

$$\int_{\mathcal{C}(L)} \dot{U}^2 \, d^3\mathbf{x} = \sum_{\mathbf{p}\in\mathcal{G}(N)} \omega_{\mathbf{p}}(a^+_{\mathbf{p},0}a^-_{\mathbf{p},0} - \tfrac{1}{2}I).$$

Furthermore, set $\mathbf{p} = p^1\mathbf{i} + p^2\mathbf{j} + p^2\mathbf{k}$. The partial derivative of U with respect to x reads as

$$U_x = \sum_{\mathbf{p}\in\mathcal{G}(N)} (a^-_{\mathbf{p},0}\mathrm{e}^{\mathrm{i}(\mathbf{p}\mathbf{x}-\omega_{\mathbf{p}}t)} - a^+_{\mathbf{p},0}\mathrm{e}^{-\mathrm{i}(\mathbf{p}\mathbf{x}-\omega_{\mathbf{p}}t)}) \, \mathrm{i}p^1\mathcal{N}_{\mathbf{p}}.$$

Hence

$$\int_{\mathcal{C}(L)} U_x^2 \, d^3\mathbf{x} = \sum_{\mathbf{p}\in\mathcal{G}(N)} (a^+_{\mathbf{p},0}a^-_{\mathbf{p},0} + a^-_{\mathbf{p},0}a^+_{\mathbf{p},0})(p^1)^2\mathcal{N}_{\mathbf{p}}^2 L^3.$$

Similarly, noting that $\sum_{j=1}^3 (p^j)^2 = \mathbf{p}^2 = \omega_{\mathbf{p}}^2$, we obtain

$$-\tfrac{1}{2}\int_{\mathcal{C}(L)} (\dot{U}^2 + (\boldsymbol{\partial}U)^2) \, d^3\mathbf{x} = \sum_{\mathbf{p}\in\mathcal{G}(N)} \omega_{\mathbf{p}}(-a^+_{\mathbf{p},0}a^-_{\mathbf{p},0} + \tfrac{1}{2}I).$$

Analogously, we get

$$-\tfrac{1}{2}\int_{\mathcal{C}(L)} (\dot{\mathbf{A}}^2 + \sum_{j=1}^3 (\boldsymbol{\partial}A_j)^2) \, d^3\mathbf{x} = \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=1}^3 \omega_{\mathbf{p}}(a^+_{\mathbf{p},s}a^-_{\mathbf{p},s} + \tfrac{1}{2}I).$$

This finishes the motivation of $\mathsf{H}_{\mathrm{phot}}$. For physical reasons, the energy operator $\mathsf{H}_{\mathrm{phot}}$ will be redefined below by setting

$$H_{\mathrm{phot}} := \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=1}^3 \omega_{\mathbf{p}}a^+_{\mathbf{p},s}a^-_{\mathbf{p},s} - \omega_{\mathbf{p}}a^+_{\mathbf{p},0}a^-_{\mathbf{p},0}.$$

That is, we separate the vacuum energy. In addition, by the Gupta–Bleuler quantization condition below, we will eliminate the contributions of nonphysical longitudinal and scalar photons. More precisely, for all physical states $|\Psi\rangle$, we will obtain the following crucial energy relation:

$$\boxed{H_{\mathrm{phot}}|\Psi\rangle = \left(\sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=1}^2 \omega_{\mathbf{p}}a^+_{\mathbf{p},s}a^-_{\mathbf{p},s}\right)|\Psi\rangle.}$$

12.2.2 The Free Electron Quantum Field and Antiparticles

Recall that the energy of an electron of mass m_e and momentum vector $\mathbf{p}$ is given by

$$E_{\mathbf{p}} := \sqrt{m_e^2 + \mathbf{p}^2}.$$

The positron is the antiparticle to the electron. The positron has the same mass as the electron. But, in contrast to the negative electric charge $-e$ of the electron, the positron has the positive electric charge e.[8]

The quantum field of free electrons and positrons. Consider now the classical solution ψ of the free Dirac equation from (12.11) on page 810.

- Replace the Fourier coefficient $b_{\mathbf{p},s}$ by the annihilation operator $b^-_{\mathbf{p},s}$ of an electron with momentum vector $\mathbf{p}$ and spin $s = \pm\frac{1}{2}$ in direction of $\mathbf{p}$.[9]
- Replace the Fourier coefficient $c_{\mathbf{p},s}$ by the annihilation operator $c^-_{\mathbf{p},s}$ of a positron with momentum vector $\mathbf{p}$ and spin $s = \pm\frac{1}{2}$ in direction of $\mathbf{p}$.
- Replace the conjugate-complex Fourier coefficient $b^\dagger_{\mathbf{p},s}$ by the creation operator $b^+_{\mathbf{p},s}$ of an electron with momentum vector $\mathbf{p}$ and spin $s = \pm\frac{1}{2}$ in direction of $\mathbf{p}$.
- Replace the conjugate-complex Fourier coefficient $c^\dagger_{\mathbf{p},s}$ by the creation operator $c^+_{\mathbf{p},s}$ of a positron with momentum vector $\mathbf{p}$ and spin $s = \pm\frac{1}{2}$ in direction of $\mathbf{p}$.

This way, from (12.11) we get the following free electron-positron quantum field

$$\boxed{\psi(\mathbf{x},t) := \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=\pm\frac{1}{2}} (b^-_{\mathbf{p},s} u_{\mathbf{p},s}\, \mathrm{e}^{\mathrm{i}(\mathbf{px}-E_\mathbf{p}t)} + c^+_{\mathbf{p},s} v_{\mathbf{p},s}\, \mathrm{e}^{-\mathrm{i}(\mathbf{px}-E_\mathbf{p}t)})\, N_\mathbf{p}} \tag{12.16}$$

with the normalization factor

$$N_\mathbf{p} := \sqrt{\frac{1}{2L^3 E_\mathbf{p}}}\,.$$

For the adjoint field from (12.12), we obtain

$$\boxed{\overline{\psi}(\mathbf{x},t) := \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=\pm\frac{1}{2}} (b^+_{\mathbf{p},s} \overline{u}_{\mathbf{p},s}\, \mathrm{e}^{-\mathrm{i}(\mathbf{px}-E_\mathbf{p}t)} + c^-_{\mathbf{p},s} \overline{v}_{\mathbf{p},s}\, \mathrm{e}^{\mathrm{i}(\mathbf{px}-E_\mathbf{p}t)})\, N_\mathbf{p}.} \tag{12.17}$$

The anticommutation relations for the electron. For the creation and annihilation operators of a free electron we postulate the following anticommutation relations for all $\mathbf{p},\mathbf{q} \in \mathcal{G}(N)$ and all $s, r = \pm\frac{1}{2}$:

(E1) $[b^-_{\mathbf{p},s}, b^+_{\mathbf{p},s}]_+ = I$.[10]
(E2) $[b^-_{\mathbf{p},s}, b^+_{\mathbf{q},r}]_+ = 0$ if $\mathbf{p} \neq \mathbf{q}$ or $s \neq r$.
(E3) $[b^-_{\mathbf{p},s}, b^-_{\mathbf{q},r}]_+ = [b^+_{\mathbf{p},s}, b^+_{\mathbf{q},r}]_+ = 0$.
(E4) $b^-_{\mathbf{p},s}|0\rangle = 0$.

[8] Observe that we work in the energetic system. In the SI system, the electron energy (resp. the positron energy) is given by

$$E_\mathbf{p} := \sqrt{m_e^2 c^4 + \mathbf{p}^2 c^2},$$

and the photon energy is given by $E_{\mathrm{phot}} := \hbar\omega_\mathbf{p}$.

[9] In the SI system, the electron has the spin $\hbar s$ in direction of the momentum vector $\mathbf{p}$, where $s = \pm\frac{1}{2}$.

[10] Recall that $[A,B]_+ := AB + BA$.

(E5) $(b^-_{\mathbf{p},s})^\dagger = b^+_{\mathbf{p},s}$.

The anticommutation relations for the positron. Similarly, for the creation and annihilation operators of a free positron, we postulate the following anticommutation relations:

(E^+1) $[c^-_{\mathbf{p},s}, c^+_{\mathbf{p},s}]_+ = I$.
(E^+2) $[c^-_{\mathbf{p},s}, c^+_{\mathbf{q},r}]_+ = 0$ if $\mathbf{p} \neq \mathbf{q}$ or $s \neq r$.
(E^+3) $[c^-_{\mathbf{p},s}, c^-_{\mathbf{q},r}]_+ = [c^+_{\mathbf{p},s}, c^+_{\mathbf{q},r}]_+ = 0$.
(E^+4) $c^-_{\mathbf{p},s}|0\rangle = 0$.
(E^+5) $(c^-_{\mathbf{p},s})^\dagger = c^+_{\mathbf{p},s}$.

Since photons, electrons, and positrons are independent particles, we assume that the corresponding creation and annihilation operators always commute. Explicitly, we postulate that for all momentum vectors $\mathbf{p}, \mathbf{q} \in \mathcal{G}(N)$ and all polarizations $s = 0, 1, 2, 3$ and $r, r' = \pm\frac{1}{2}$, the following are true:

(F1) $[a^\pm_{\mathbf{p},s}, b^\pm_{\mathbf{q},r}]_- = [a^\pm_{\mathbf{p},s}, b^\mp_{\mathbf{q},r}]_- = 0$.
(F2) $[a^\pm_{\mathbf{p},s}, c^\pm_{\mathbf{q},r}]_- = [a^\pm_{\mathbf{p},s}, c^\mp_{\mathbf{q},r}]_- = 0$.
(F3) $[b^\pm_{\mathbf{p},r}, c^\pm_{\mathbf{q},r'}]_- = [b^\pm_{\mathbf{p},r}, c^\mp_{\mathbf{q},r'}]_- = 0$.

The realization of the anticommutation relations (E1)–(E5) and (E^+1)–(E^+5) for electrons and positrons along with (F1)–(F3) will be discussed in Sect. 12.4.3 on page 826.

The canonical anticommutation relations for the free quantum field of electrons and positrons. Fix time t. For $j, k = 1, ..., 4$ and all position vectors $\mathbf{x}$ and $\mathbf{y}$, we obtain

$$\boxed{[\psi^j(\mathbf{x}, t), \pi^k(\mathbf{y}, t)]_+ = \mathrm{i}\delta_{\mathcal{C}(L)}(\mathbf{x} - \mathbf{y}) \cdot \delta^{jk} I} \tag{12.18}$$

and

$$\boxed{[\psi^j(\mathbf{x}, t), \psi^k(\mathbf{y}, t)]_+ = 0, \qquad [\pi^j(\mathbf{x}, t), \pi^k(\mathbf{y}, t)]_+ = 0.} \tag{12.19}$$

Here, $\pi = \mathcal{L}_{\dot\psi} = \mathrm{i}\psi^\dagger$ denotes the conjugate field to ψ. Using the anticommutation relations for the creation and annihilation operators $b^\pm, c^\pm$, this follows analogously as in the proof of (12.13) on page 812.

The energy of the free quantum field for electrons and positrons. The energy operator of the free quantum field for electrons is given by

$$\boxed{\mathsf{H}_{\mathrm{el}} := \sum_{\mathbf{p} \in \mathcal{G}(N)} \sum_{s = \pm\frac{1}{2}} E_{\mathbf{p}} (b^+_{\mathbf{p},s} b^-_{\mathbf{p},s} - \tfrac{1}{2} I).}$$

Similarly, the energy operator of the free quantum field for positrons is given by

$$\boxed{\mathsf{H}_{\mathrm{pos}} := \sum_{\mathbf{p} \in \mathcal{G}(N)} \sum_{s = \pm\frac{1}{2}} E_{\mathbf{p}} (c^+_{\mathbf{p},s} c^-_{\mathbf{p},s} - \tfrac{1}{2} I).}$$

Let us motivate this. By Theorem 11.2, we start with the energy operator

$$\mathsf{H}_{\mathrm{el/pos}} := \int_\Omega \psi^\dagger(\mathbf{x}, t) \mathbf{H} \psi(\mathbf{x}, t)\, d^3\mathbf{x}.$$

According to (12.16), we choose the free electron-positron quantum field

$$\psi(\mathbf{x},t) := \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=\pm\frac{1}{2}} (b^-_{\mathbf{p},s}\psi^+_{\mathbf{p},s} + c^+_{\mathbf{p},s}\psi^-_{\mathbf{p},s})\, N_{\mathbf{p}}.$$

Noting that $(b^-_{\mathbf{p},s})^\dagger = b^+_{\mathbf{p},s}$ and $(c^+_{\mathbf{p},s})^\dagger = c^-_{\mathbf{p},s}$, the adjoint fields reads as

$$\psi^\dagger(\mathbf{x},t) := \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=\pm\frac{1}{2}} (b^+_{\mathbf{p},s}\psi^{+\dagger}_{\mathbf{p},s} + c^-_{\mathbf{p},s}\psi^{-\dagger}_{\mathbf{p},s})\, N_{\mathbf{p}}.$$

Since $\mathbf{H}\psi^\pm_{\mathbf{p},s} = \pm E_{\mathbf{p}}\psi^\pm_{\mathbf{p},s}$, we get

$$\mathbf{H}\psi = \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=\pm\frac{1}{2}} E_{\mathbf{p}}(b^-_{\mathbf{p},s}\psi^+_{\mathbf{p},s} - c^+_{\mathbf{p},s}\psi^-_{\mathbf{p},s})\, N_{\mathbf{p}}.$$

The orthogonality relation from Theorem 12.4 tells us that

$$\begin{aligned}\mathsf{H}_{\mathrm{el/pos}} = \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=\pm\frac{1}{2}} E_{\mathbf{p}}(&b^+_{\mathbf{p},s}b^-_{\mathbf{p},s}\,\langle\psi^+_{\mathbf{p},s}|\psi^+_{\mathbf{p},s}\rangle_{\mathcal{C}(L)} \\ &- c^-_{\mathbf{p},s}c^+_{\mathbf{p},s}\,\langle\psi^-_{\mathbf{p},s}|\psi^-_{\mathbf{p},s}\rangle_{\mathcal{C}(L)})\, N^2_{\mathbf{p}}.\end{aligned}$$

By $c^-_{\mathbf{p},s}c^+_{\mathbf{p},s} = -c^+_{\mathbf{p},s}c^-_{\mathbf{p},s} + I$, we get

$$\mathsf{H}_{\mathrm{el/pos}} = \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=\pm\frac{1}{2}} 2E^2_{\mathbf{p}}(b^+_{\mathbf{p},s}b^-_{\mathbf{p},s} + c^+_{\mathbf{p},s}c^-_{\mathbf{p},s} - I)N^2_{\mathbf{p}}L^3. \tag{12.20}$$

Finally, it follows from $2E^2_{\mathbf{p}}N^2_{\mathbf{p}}L^3 = 2E_{\mathbf{p}}$ that $\mathsf{H}_{\mathrm{el/pos}} = \mathsf{H}_{\mathrm{el}} + \mathsf{H}_{\mathrm{pos}}$.

The total energy operator of free quantum fields in quantum electrodynamics. Using the superposition principle, we define the total energy operator of the free quantum field in quantum electrodynamics by

$$\mathsf{H}_{\mathrm{free}} := \mathsf{H}_{\mathrm{phot}} + \mathsf{H}_{\mathrm{el}} + \mathsf{H}_{\mathrm{pos}}. \tag{12.21}$$

For further investigations, it is useful to separate the vacuum energy. To this end, we redefine $\mathsf{H}_{\mathrm{phot}}, \mathsf{H}_{\mathrm{el}}, \mathsf{H}_{\mathrm{pos}}$ by setting

$$\boxed{\mathsf{H}_{\mathrm{free}} := H_{\mathrm{phot}} + H_{\mathrm{el}} + H_{\mathrm{pos}} + H_{\mathrm{vac}}.}$$

Here, we define the photon energy operator

$$H_{\mathrm{phot}} := H_{\text{real photons}} + H_{\text{virtual photons}}$$

where

$$H_{\text{real photons}} := \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=1,2} \omega_{\mathbf{p}} a^+_{\mathbf{p},s}a^-_{\mathbf{p},s}$$

and

$$H_{\text{virtual photons}} = \sum_{\mathbf{p}\in\mathcal{G}(N)} \omega_{\mathbf{p}} a^+_{\mathbf{p},3}a^-_{\mathbf{p},3} - \omega_{\mathbf{p}} a^+_{\mathbf{p},0}a^-_{\mathbf{p},0}.$$

We will show in Sect. 12.4.4 that only the energy operator $H_{\text{real photons}}$ of transversal photons possesses a physical meaning, whereas the energy operator $H_{\text{virtual photons}}$

of virtual photons drops out in Gupta–Bleuler quantization. Furthermore, let us introduce the following three operators: the electron energy operator

$$H_{\mathrm{el}} := \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=\pm\frac{1}{2}} E_{\mathbf{p}} b^+_{\mathbf{p},s} b^-_{\mathbf{p},s},$$

the positron energy operator

$$H_{\mathrm{pos}} := \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=\pm\frac{1}{2}} E_{\mathbf{p}} c^+_{\mathbf{p},s} c^-_{\mathbf{p},s},$$

and the vacuum energy operator

$$H_{\mathrm{vac}} := \sum_{\mathbf{p}\in\mathcal{G}(N)} (2\omega_{\mathbf{p}} - 2E_{\mathbf{p}})I. \tag{12.22}$$

Recall that $E_{\mathbf{p}} = \sqrt{m_e^2 + \mathbf{p}^2}$ and $\omega_{\mathbf{p}} = |\mathbf{p}|$.

Pauli's spin-statistics principle. In order to explain the basic idea behind this fundamental principle in elementary particle physics, consider the state

$$\psi := c^+_{\mathbf{q},r}\,|0\rangle$$

which corresponds to one positron of momentum vector $\mathbf{q}$ and spin $r = \pm\frac{1}{2}$ in direction of $\mathbf{q}$. Then

$$\boxed{H_{\mathrm{pos}}\psi = E_{\mathbf{q}}\psi,} \tag{12.23}$$

that is, the positron state ψ has the energy $E_{\mathbf{q}}$. To prove this note the following. Letting $\varepsilon := 1$ it follows from the anticommutation relation

$$c^-_{\mathbf{q},r} c^+_{\mathbf{q},r} = I - \varepsilon c^+_{\mathbf{q},r} c^-_{\mathbf{q},r} \tag{12.24}$$

and $c^-_{\mathbf{p},r}|0\rangle = 0$ that

$$c^+_{\mathbf{q},r} c^-_{\mathbf{q},r} c^+_{\mathbf{q},r}|0\rangle = c^+_{\mathbf{q},r}|0\rangle - \varepsilon (c^+_{\mathbf{q},r})^2 c^-_{\mathbf{q},r}|0\rangle = c^+_{\mathbf{q},r}|0\rangle.$$

Moreover, if $\mathbf{p} \neq \mathbf{q}$ or $s \neq r$, then

$$c^+_{\mathbf{p},s} c^-_{\mathbf{p},s} c^+_{\mathbf{q},r}|0\rangle = -\varepsilon c^+_{\mathbf{p},s} c^+_{\mathbf{q},r} c^-_{\mathbf{p},s}|0\rangle = 0.$$

In order to get insight, let us replace the anticommutation relations $[.\ ,\ .]_+$ for the creation and annihilation operators of electrons and positrons by analogous commutation relations $[.\ ,\ .]_-$. We want to show that this leads to a contradiction in physics (i.e., electron states and positron states of negative energy). In fact, if we assume that

$$[c^-_{\mathbf{p},r}, c^+_{\mathbf{p},r}]_- = I,$$

then equation (12.24) holds with $\varepsilon = -1$. According to (12.20) above, we get

$$H_{\mathrm{pos}} := \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=\pm\frac{1}{2}} \varepsilon E_{\mathbf{p}} c^+_{\mathbf{p},s} c^-_{\mathbf{p},s}$$

with $\varepsilon := -1$. The same argument as above yields

$$H_{\text{pos}}\psi = -E_{\mathbf{q}}\psi.$$

This negative energy $-E_{\mathbf{q}}$ of the positron is a nonphysical result. Summarizing, positivity of energy forces the anticommutation relations for electrons and positrons. In Sect. 10.2 on page 779, we have shown that anticommutation relations in the Fock space are responsible for the Pauli exclusion principle. Pauli formulated the following general spin-statistics principle for quantum fields in 1955:[11]

- Particles with integer spin (bosons) have to be quantized by commutation relations. The number of identical bosons in a fixed physical state is unlimited.
- Particles with half-integer spin (fermions) have to be quantized by anticommutation relations. Two identical fermions cannot be in the same physical state.

In terms of statistical physics, one has to use

- the Bose–Einstein statistics for bosons (e.g., photons), and
- the Fermi–Dirac statistics for fermions (e.g., electrons and positrons).

We refer to :

R. Streater and A. Wightman, PCT, Spin, Statistics, and All That, Benjamin, New York, 1968.

I. Duck and E. Sudarshan, Pauli and the Spin-Statistics Theorem, World Scientific, Singapore, 1997.

12.2.3 The Spin of Photons

The spin of elementary particles is related to infinitesimal rotations.
Folklore

Classical electromagnetic waves possess a polarization. Motivated by the electron spin, we want to reformulate this by introducing the photon spin.

Infinitesimal rotation. Consider a rotation

$$\mathbf{x}' = U\mathbf{x}, \qquad \mathbf{x} \in V_3,$$

about the origin in the 3-dimensional Euclidean vector space V_3.[12] In particular, consider a right-handed (x, y, z)-Cartesian coordinate with the right-handed orthonormal basis $\mathbf{i}, \mathbf{j}, \mathbf{k}$. The position vector reads as

$$\mathbf{x} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}.$$

Let $\mathbf{n} = n^1\mathbf{i} + n^2\mathbf{j} + n^3\mathbf{k}$ be a unit vector. The symbol $U_{\mathbf{n}}(\varphi)$ denotes a counterclockwise rotation about the axis $\mathbf{n}$ of the angle φ. In particular, $U_{\mathbf{k}}(\varphi)$ represents a rotation about the z-axis. For small angle φ, we have the approximation formula

$$U_{\mathbf{n}}(\varphi)\mathbf{x} = \mathbf{x} + \varphi(\mathbf{n} \times \mathbf{x}) + O(\varphi^2), \qquad \varphi \to 0.$$

Set $I_{\mathbf{n}}\mathbf{x} := \mathbf{n} \times \mathbf{x}$. The operator $I_{\mathbf{n}}$ is called an infinitesimal rotation about the axis $\mathbf{n}$. Each rotation U induces the following transformation T_U :

(i) Schrödinger's wave function: $T_U\psi(\mathbf{x}) = \psi(U^{-1}\mathbf{x})$.
(ii) Photon wave function: $T_U\mathbf{A}(\mathbf{x}) := U\mathbf{A}(U^{-1}\mathbf{x})$.

[11] W. Pauli, On the connection between spin and statistics, Progr. Theor. Phys. **5** (1955), 526–543.

[12] All of these rotations form a group called the special unitary group, $SU(V_3)$, of the space V_3.

(iii) Electron wave function: $T_U\psi(\mathbf{x}) = \mathcal{T}_U\psi(U^{-1}\mathbf{x})$.

Naturally enough, we postulate that the operator $\mathcal{T}_U$ possesses the following product property $\mathcal{T}_U\mathcal{T}_V = \mathcal{T}_{UV}$ for all $U, V \in SU(V_3)$. Concerning (i)–(iii), this implies[13]

$$\boxed{T_{UV} = T_U T_V \qquad \text{for all} \quad U, V \in SU(V_3).}$$

We say that the family of operators $\{T_U\}$ forms a representation of the group $SU(V_3)$.

Angular momentum of the Schrödinger wave function. Consider first the case (i) above. For smooth functions ψ, define the operator $I_{\mathbf{n}}$ by the following approximation formula

$$T_{U_{\mathbf{n}}(\varphi)}\psi(\mathbf{x}) = \psi(\mathbf{x}) + \varphi I_{\mathbf{n}}\psi(\mathbf{x}) + O(\varphi^2), \qquad \varphi \to 0.$$

The Taylor formula tells us that

$$\boxed{I_{\mathbf{n}} = \mathbf{n}(\mathbf{x} \times \boldsymbol{\partial}).}$$

Explicitly, $I_{\mathbf{n}}\psi(\mathbf{x}) = \mathbf{n}(\mathbf{x} \times \boldsymbol{\partial}\psi(\mathbf{x}))$. To prove this, consider first the special case where $\mathbf{n} := \mathbf{k}$. Then, up to terms of order $O(\varphi^2)$ as $\varphi \to 0$, we get

$$U_{\mathbf{k}}(\varphi)^{-1}\mathbf{x} = U_{\mathbf{k}}(-\varphi)\mathbf{x} = \mathbf{x} - \varphi \cdot (\mathbf{k} \times \mathbf{x}) + \ldots = \mathbf{x} + \varphi \cdot (y\mathbf{i} - x\mathbf{j}) + \ldots$$

This implies

$$\psi(U_{\mathbf{k}}(\varphi)^{-1}\mathbf{x}) = \psi(\mathbf{x}) + \varphi \cdot (y\psi_x(\mathbf{x}) - x\psi_y(\mathbf{x})) + \ldots$$

Hence $I_{\mathbf{k}}\psi(\mathbf{x}) = \mathbf{k}(\mathbf{x} \times \boldsymbol{\partial}\psi(\mathbf{x}))$. The case for general unit vectors $\mathbf{n}$ proceeds similarly. Using the infinitesimal operator $I_{\mathbf{n}}$, we define the angular momentum operator $\mathbf{L}_{\mathbf{n}}$ in direction of the unit vector $\mathbf{n}$ by setting

$$\boxed{\mathbf{L}_{\mathbf{n}} := -\mathrm{i}I_{\mathbf{n}}.}$$

Specializing $\mathbf{n} = \mathbf{i}, \mathbf{j}, \mathbf{k}$, we define

$$L^1 := -\mathrm{i}I_{\mathbf{i}}, \quad L^2 := -\mathrm{i}I_{\mathbf{j}}, \quad L^3 := -\mathrm{i}I_{\mathbf{k}} = \mathrm{i}(y\partial_x - x\partial_y).$$

Thus, for the operator $\mathbf{L} := L^1\mathbf{i} + L^2\mathbf{j} + L^3\mathbf{k}$, we get[14]

$$\boxed{\mathbf{L} = -\mathrm{i}(\mathbf{x} \times \boldsymbol{\partial}).}$$

Explicitly, $\mathbf{L}\psi(\mathbf{x}) = -\mathrm{i}(\mathbf{x} \times \boldsymbol{\partial}\psi(\mathbf{x}))$. This is the angular momentum operator in quantum mechanics.

Spin of the photon wave function. Consider now the case (ii) above. Up to terms of order $O(\varphi^2)$ as $\varphi \to 0$, we get

[13] For example, consider (iii). Then

$$(T_{UV}\psi)(\mathbf{x}) = \mathcal{T}_{UV}\psi((UV)^{-1}\mathbf{x}) = \mathcal{T}_U\mathcal{T}_V\psi(U^{-1}V^{-1}\mathbf{x}) = T_U(T_V\psi)(\mathbf{x}).$$

[14] Recall that we work in the energetic system. In the SI system, $L_{\mathbf{n}} := -\mathrm{i}\hbar I_{\mathbf{n}}$. Hence $\mathbf{L} := -\mathrm{i}\hbar(\mathbf{x} \times \boldsymbol{\partial})$.

$$T_{U_{\mathbf{n}}(\varphi)}\mathbf{A}(\mathbf{x}) = \mathbf{A}(\mathbf{x}') + \varphi \cdot \mathbf{n} \times \mathbf{A}(\mathbf{x}') + \dots$$

with $\mathbf{x}' = \mathbf{x} - \varphi \cdot (\mathbf{n} \times \mathbf{x}) + \dots$ Hence

$$T_{U_{\mathbf{n}}(\varphi)}\mathbf{A}(\mathbf{x}) = \mathbf{A}(\mathbf{x}) + \varphi I_{\mathbf{n}}\mathbf{A}(\mathbf{x}) + O(\varphi^2), \qquad \varphi \to 0$$

with the infinitesimal operator

$$\boxed{I_{\mathbf{n}}\mathbf{A}(\mathbf{x}) := \mathbf{n} \times \mathbf{A}(\mathbf{x}) + L_{\mathbf{n}}\mathbf{A}(\mathbf{x}).}$$

We now introduce the total angular momentum operator[15]

$$J_{\mathbf{n}} := -\mathrm{i}I_{\mathbf{n}}$$

in direction of the unit vector $\mathbf{n}$. Note that we obtain the decomposition

$$\boxed{J_{\mathbf{n}} = S_{\mathbf{n}} + L_{\mathbf{n}}.} \tag{12.25}$$

Here, the operator $L_{\mathbf{n}} = \mathbf{n}(\mathbf{x} \times \boldsymbol{\partial})$ is called the orbital angular momentum operator, and the operator $S_{\mathbf{n}}$ given by

$$\boxed{S_{\mathbf{n}}\mathbf{A} := -\mathrm{i}(\mathbf{n} \times \mathbf{A})}$$

is called the spin operator of the vector field $\mathbf{A}$ in direction of $\mathbf{n}$. For given nonzero momentum vector $\mathbf{p}$, introduce the unit vector $\mathbf{n} := \mathbf{p}/|\mathbf{p}|$. Choose the three vectors

$$\mathbf{e}_1(\mathbf{p}), \quad \mathbf{e}_2(\mathbf{p}), \quad \mathbf{e}_3(\mathbf{p}) := \mathbf{n}$$

in such a way that they form a right-handed orthonormal system. Then, the vectors

$$\mathbf{e}_\pm(\mathbf{p}) := \frac{\mathbf{e}_1(\mathbf{p}) \mp \mathrm{i}\mathbf{e}_2(\mathbf{p})}{\sqrt{2}}$$

are eigenvectors of the spin operator $S_{\mathbf{n}}$. Explicitly,[16]

$$S_{\mathbf{n}}\mathbf{e}_\pm(\mathbf{p}) = \pm\mathbf{e}_\pm(\mathbf{p}).$$

Therefore, we assign the spin $s = \pm 1$ to the vector $\mathbf{e}_\pm(\mathbf{p})$, respectively. Let $a_{\mathbf{p},\pm}$ be complex numbers. For the following photon wave functions

$$\mathbf{A}_\pm(\mathbf{x}, t) := a_{\mathbf{p},\pm}(\mathbf{p})\ \mathbf{e}_\pm(\mathbf{p})\ \mathrm{e}^{\pm\mathrm{i}(\mathbf{p}\mathbf{x} - E_{\mathbf{p}}t)},$$

we get the wave equation $\Box\mathbf{A} = 0$ and

$$S_{\mathbf{n}}\mathbf{A}_\pm = \pm\mathbf{A}_\pm.$$

We say that the photon wave function $\mathbf{A}_\pm$ has the spin $s = \pm 1$ in direction of the momentum vector $\mathbf{p}$, respectively.

Spin of the electron wave function. Consider the case (iii) above. Let $U := U_{\mathbf{n}}(\varphi)$. According to the representation theory for the Lorentz group, we get[17]

[15] In the SI system of units, $J_{\mathbf{n}} := -\mathrm{i}\hbar I_{\mathbf{n}}$ and $S_{\mathbf{n}}\mathbf{A} := -\mathrm{i}\hbar(\mathbf{n} \times \mathbf{A})$.

[16] Note that $\mathbf{n} \times (\mathbf{e}_1 + \mathrm{i}\mathbf{e}_2) = \mathbf{e}_2 - \mathrm{i}\mathbf{e}_1$.

[17] The representation theory of the Lorentz group will be thoroughly studied in Vol. III.

$$T_U\psi(\mathbf{x}) = \mathrm{e}^{\mathrm{i}\varphi\mathbf{nS}}\psi(U^{-1}\mathbf{x}).$$

Here, $\mathbf{nS} = \sum_{j=1}^{3} n^j S^j$ and $\mathbf{S} = S^1\mathbf{i} + S^2\mathbf{j} + S^3\mathbf{k}$ along with

$$S^3 := \tfrac{1}{2}\mathrm{i}\gamma^1\gamma^2 = \tfrac{1}{2}\begin{pmatrix} \sigma^3 & 0 \\ 0 & \sigma^3 \end{pmatrix}.$$

The remaining operators S^1, S^2 are obtained from this by using the cyclic permutation $1 \mapsto 2 \mapsto 3 \mapsto 1$. Therefore,

$$T_U\psi(\mathbf{x}) = \psi(\mathbf{x}) + I_{\mathbf{n}}\psi(\mathbf{x}) + O(\varphi^2), \qquad \varphi \to 0$$

with $I_{\mathbf{n}} = \mathrm{i}(\mathbf{nS} + L_{\mathbf{n}})$. Similarly as in (12.25) above, we introduce the total angular momentum operator $J_{\mathbf{n}} = -\mathrm{i}I_{\mathbf{n}}$. Hence

$$J_{\mathbf{n}} = \mathbf{nS} + L_{\mathbf{n}}.$$

The operator $S_{\mathbf{n}}$ (resp. $L_{\mathbf{n}}$) is called the spin operator (resp. the orbital angular momentum operator) of the electron.

Perspectives in representation theory. We will show in Vol. III that the irreducible representations of the rotation group $SU(V_3)$ can be classified by a number $s = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \ldots$ For the rotation group $SU(V_3)$ itself, we get $s = 1$. Each elementary particle corresponds to such an irreducible representation of $SU(V_3)$. In particular, the following are true:

- $s = 0$: gluons, mesons π^+, π^0, π^-;
- $s = \frac{1}{2}$: electron, positron, neutrinos, muon, tau, quarks, proton, neutron;
- $s = 1$: photon, vector bosons W^+, W^-, Z^0;
- $s = 2$: graviton.

The number s is called the spin number of the elementary particle. For given spin number s and given unit vector $\mathbf{n}$, there always exist states of the elementary particle where the spin attains either the value s or $-s$ in direction of the vector $\mathbf{n}$.[18]

The spin quantum number is the prototype for the mathematical construction of quantum numbers in elementary particle physics. This will be thoroughly studied in Vols. IIIff. For example, the group $SU(3)$ is responsible for the fact that the proton consists of three quarks. Moreover, mesons are quark-antiquark pairs. The representation theory of the compact Lie group $SU(3)$ due to Hermann Weyl helps to classify the possible reactions between elementary particles.

12.3 The Ground State Energy and the Normal Product

> In formal terms, the ground state energy (vacuum energy) of the electromagnetic quantum field is infinite. This causes mathematical trouble in quantum electrodynamics.
>
> Folklore

It is typical for quantum field theory that the ground state of a quantum field has nonzero energy. This so-called vacuum energy is responsible for physical effects which can be measured in physical experiments. The prototype for such a vacuum effect is the Casimir effect considered in Sect. 2.23.1 of Vol. I.

[18] In the SI system, the values are $s\hbar$ and $-s\hbar$.

From the mathematical point of view, the vacuum energy is a nontrivial phenomenon related to infinities.

This can be seen by increasing the number of grid points N. In fact, it follows from (12.22) that

$$\lim_{N\to\infty} H_{\mathrm{vac}} = -\infty.$$

Therefore, the vacuum energy becomes infinite in the continuum limit. In quantum electrodynamics, the simplest method is to ignore the vacuum term H_{vac} by using normal products of creation and annihilation operators. Let us discuss this.

The normal product of creation and annihilation operators. For photon creation and annihilation operators, we define the so-called normal product by setting

$$\boxed{:a^-_{\mathbf{p},s}a^+_{\mathbf{q},r}: = a^+_{\mathbf{q},r}a^-_{\mathbf{p},s}}$$

along with the following additional conventions:

- $:a^+_{\mathbf{p},s}a^-_{\mathbf{q},r}: = a^+_{\mathbf{p},s}a^-_{\mathbf{q},r}$,
- $:a^+_{\mathbf{p},s}a^+_{\mathbf{q},r}: = a^+_{\mathbf{p},s}a^+_{\mathbf{q},r}$,
- $:a^-_{\mathbf{p},s}a^-_{\mathbf{q},r}: = a^-_{\mathbf{p},s}a^-_{\mathbf{q},r}$.

Mnemonically, annihilation operators act first. The normal product can be extended to linear combinations in a natural way. For example,

$$:\alpha a^-_{\mathbf{p},s}a^+_{\mathbf{q},r} + \beta a^+_{\mathbf{p},s}a^+_{\mathbf{q},r}: = \alpha a^+_{\mathbf{q},r}a^-_{\mathbf{p},s} + \beta a^+_{\mathbf{p},s}a^+_{\mathbf{q},r}.$$

For electron creation and annihilation operators, there appears a sign change. Explicitly, we define

$$\boxed{:b^-_{\mathbf{p},s}b^+_{\mathbf{q},r}: = -b^+_{\mathbf{q},r}b^-_{\mathbf{p},s}}$$

along with the following conventions:

- $:b^+_{\mathbf{p},s}b^-_{\mathbf{q},r}: = b^+_{\mathbf{p},s}b^-_{\mathbf{q},r}$,
- $:b^+_{\mathbf{p},s}b^+_{\mathbf{q},r}: = b^+_{\mathbf{p},s}b^+_{\mathbf{q},r}$,
- $:b^-_{\mathbf{p},s}b^-_{\mathbf{q},r}: = b^-_{\mathbf{p},s}b^-_{\mathbf{q},r}$.

The same relations are true if we replace $b^\pm_{\mathbf{p},s}$ by $c^\pm_{\mathbf{p},s}$, respectively.

The modified energy operator. We replace the operator $\mathsf{H}_{\mathrm{free}}$ by the normal product $:\mathsf{H}_{\mathrm{free}}:$. The linearity of the normal product yields

$$:\mathsf{H}_{\mathrm{free}}: \;=\; :\mathsf{H}_{\mathrm{phot}}: + :\mathsf{H}_{\mathrm{el}}: + :\mathsf{H}_{\mathrm{pos}}:.$$

We claim that

$$\boxed{:\mathsf{H}_{\mathrm{free}}: \;= H_{\mathrm{phot}} + H_{\mathrm{el}} + H_{\mathrm{pos}}.} \tag{12.26}$$

This means that the modified energy operator $:\mathsf{H}_{\mathrm{free}}:$ does not contain the vacuum energy anymore. In the future, we will use the energy operator $:\mathsf{H}_{\mathrm{free}}:$ instead of the operator $\mathsf{H}_{\mathrm{free}}$. This means that the normal product is part of the quantization procedure.

Proof. By Sect. 12.2.1 on page 812, the normal product $:\mathsf{H}_{\mathrm{phot}}:$ is equal to

$$-\tfrac{1}{2}\int_{\mathcal{C}(L)} :\left(\dot U^2 + (\boldsymbol{\partial} U)^2 + \sum_{j=1}^{3}\dot A_j^2 + (\boldsymbol{\partial} A_j)^2\right): \, d^3\mathbf{x}.$$

We now proceed as in Sect. 12.2.1. The normal product replaces the original product $a^-_{\mathbf{p},s}a^+_{\mathbf{p},s}$ by $a^+_{\mathbf{p},s}a^-_{\mathbf{p},s}$. It follows from (12.15) that

$$\int_{\mathcal{C}(L)} :\dot{U}^2: \, d^3\mathbf{x} = \sum_{\mathbf{p}\in\mathcal{G}(N)} 2a^+_{\mathbf{p},0}a^-_{\mathbf{p},0}\,\omega^2_{\mathbf{p}}\mathcal{N}^2_{\mathbf{p}}L^3.$$

This is equal to $\sum_{\mathbf{p}} \omega_{\mathbf{p}} a^+_{\mathbf{p},0}a^-_{\mathbf{p},0}$. Similarly,

$$:\mathsf{H}_{\text{phot}}:= \sum_{\mathbf{p}\in\mathcal{G}(N)} \left(-\omega_{\mathbf{p}}a^+_{\mathbf{p},0}a^-_{\mathbf{p},0} + \sum_{s=1}^{3} \omega_{\mathbf{p}}a^+_{\mathbf{p},s}a^-_{\mathbf{p},s}\right) = H_{\text{phot}}.$$

Furthermore,

$$:\mathsf{H}_{\text{el/pos}}:= \int_{\mathcal{C}(L)} :\psi^\dagger \mathbf{H}\psi: \, d^3\mathbf{x}.$$

Applying the same argument as in Sect. 12.2.2 on page 814, we get

$$:\mathsf{H}_{\text{el/pos}}:= \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=\pm\frac{1}{2}} E_{\mathbf{p}}(: b^+_{\mathbf{p},s}b^-_{\mathbf{p},s}: \, \langle\psi^+_{\mathbf{p},s}|\psi^+_{\mathbf{p},s}\rangle_{\mathcal{C}(L)} - \\ - :c^-_{\mathbf{p},s}c^+_{\mathbf{p},s}: \, \langle\psi^-_{\mathbf{p},s}|\psi^-_{\mathbf{p},s}\rangle_{\mathcal{C}(L)})\, N^2_{\mathbf{p}}.$$

This implies

$$:\mathsf{H}_{\text{el/pos}}:= \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=\pm\frac{1}{2}} 2E^2_{\mathbf{p}}(b^+_{\mathbf{p},s}b^-_{\mathbf{p},s} + c^+_{\mathbf{p},s}c^-_{\mathbf{p},s})N^2_{\mathbf{p}}L^3.$$

Hence

$$:\mathsf{H}_{\text{el/pos}}:= \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=\pm\frac{1}{2}} E_{\mathbf{p}}b^+_{\mathbf{p},s}b^-_{\mathbf{p},s} + E_{\mathbf{p}}c^+_{\mathbf{p},s}c^-_{\mathbf{p},s} = H_{\text{el}} + H_{\text{pos}}.$$

This finishes the proof of (12.26). □

12.4 The Importance of Mathematical Models

Quantize carefully.
Folklore

In the history of quantum field theory, physicists have encountered mathematical contradictions several times. The trouble comes from the quantization procedure. In order to quantize a classical field theory, there exist several possibilities. In any case, one has to pass from classical physical quantities to operators by adding commutation and anticommutation relations as well as appropriate side conditions. It may happen that all of these relations cannot be realized without producing contradictions. Let us consider a simple example. Suppose that we want to quantize the motion $q = q(t)$ of a classical particle on the real line. To this end, we add the commutation relation

$$qp - pq = \mathrm{i}I, \tag{12.27}$$

and we regard q and p as $(n \times n)$-matrices, $n = 1, 2, ...$ However, this postulate can never be realized. In fact, using the trace of matrices, it follows from (12.27) that

$$\operatorname{tr}(qp) - \operatorname{tr}(pq) = \mathrm{i}\operatorname{tr}(I) = \mathrm{i}n.$$

Since $\operatorname{tr}(qp) = \operatorname{tr}(pq)$, we get $n = 0$. This is a contradiction.

In order to avoid contradictions, one has to construct mathematical models.

Those mathematical models are also called representations of the abstractly given quantities. In the case of the commutation relation (12.27), the standard model reads as follows. We construct the complex linear space $\mathcal{S}(\mathbb{R})$, and we choose the operators $q, p : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ defined by

$$(q\psi)(x) := x\psi(x), \qquad (p\psi)(x) = -\mathrm{i}\frac{d\psi(x)}{dx}, \qquad x \in \mathbb{R}$$

for all functions $\psi \in \mathcal{S}(\mathbb{R})$. Summarizing, the commutation relation (12.27) can never be realized in a finite-dimensional linear space, but only in infinite-dimensional linear spaces. The same argument shows that the commutation relation

$$aa^\dagger - a^\dagger a = I$$

cannot be realized by a finite-dimensional $(n \times n)$-matrix. An infinite-dimensional model can be obtained by setting

$$a := \frac{q + \mathrm{i}p}{\sqrt{2}}, \qquad a^\dagger := \frac{q - \mathrm{i}p}{\sqrt{2}},$$

and by using the linear operators $a, a^\dagger : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$.

12.4.1 The Trouble with Virtual Photons

Scalar photons can never exist in a Hilbert space.
Folklore

We want to show that the commutation relations (P1)–(P6) concerning creation operators and annihilation operators $a^\pm_{\mathbf{p},s}$ for photons from Sect. 12.2.1 cannot be realized within a Hilbert space structure. To this end, consider the state

$$\Psi := a^+_{\mathbf{p},3}\,|0\rangle.$$

Let us assume the following:

(i) The commutation relations

$$\boxed{a^-_{\mathbf{p},0}a^+_{\mathbf{p},0} - a^+_{\mathbf{p},0}a^-_{\mathbf{p},0} = -I}$$

can be realized by operators living in some Hilbert space X.

(ii) There exists a unit vector $|0\rangle$ such that $a^-_{\mathbf{p},0}|0\rangle = 0$.

(iii) $(a^-_{\mathbf{p},0})^\dagger = a^+_{\mathbf{p},0}$.

We want to prove that this implies

$$\boxed{\langle\Psi|\Psi\rangle < 0.}$$

This is the desired contradiction. For simplifying notation, write $a^\pm$ instead of $a^\pm_{\mathbf{p},0}$. Then,

$$a^-a^+|0\rangle = a^+a^-|0\rangle - |0\rangle = -|0\rangle.$$

Hence $\langle\Psi|\Psi\rangle = \langle 0|(a^+)^\dagger a^+|0\rangle = \langle 0|a^-a^+|0\rangle = -\langle 0|0\rangle = -1$.

This argument tells us that the commutation relations for scalar photons force the use of indefinite inner product spaces.

12.4.2 Indefinite Inner Product Spaces

Relativistic invariance forces the use of indefinite inner products.
Folklore

Definition. A complex linear space X is called a generalized inner product space iff there exists a symbol $\langle\psi|\varphi\rangle$ such that for all $\psi, \varphi, \chi \in X$ and all complex numbers α, β, the following are met:

(i) $\langle\psi|\varphi\rangle \in \mathbb{C}$.
(ii) $\langle\psi|\alpha\varphi + \beta\chi\rangle = \alpha\langle\psi|\varphi\rangle + \beta\langle\psi|\chi\rangle$.
(iii) $\langle\psi|\varphi\rangle^\dagger = \langle\varphi|\psi\rangle$.
(iv) $\langle\psi|\varphi\rangle = 0$ for all $\varphi \in X$ implies $\psi = 0$.

We speak of an inner product space (or pre-Hilbert space) iff the inner product is definite, i.e., $\langle\psi|\psi\rangle > 0$ for all nonzero elements ψ in X. Otherwise, we speak of an indefinite inner product space.[19]

Standard example. Let $X = \mathbb{C}^n$. For all $\psi, \varphi \in \mathbb{C}^n$, we set

$$\langle\psi|\varphi\rangle := \lambda_1(\psi^1)^\dagger\varphi^1 + \ldots + \lambda_n(\psi^n)^\dagger\varphi^n$$

where $\lambda_1, ..., \lambda_n$ are nonzero real numbers. This inner product is definite iff all the numbers $\lambda_1, ..., \lambda_n$ are positive. Otherwise the inner product is indefinite.

Minkowski space. The linear space $\mathbb{R}^4$ equipped with the indefinite inner product

$$\langle x|y\rangle := x^0y^0 - x^1y^1 - x^2y^2 - x^3y^3, \qquad x, y \in \mathbb{R}^4,$$

is called Minkowski space. Introducing the definite inner product

$$\langle x|y\rangle = x^0y^0 + x^1y^1 + x^2y^2 + x^3y^3, \qquad x, y \in \mathbb{R}^4,$$

on the linear space $\mathbb{R}^4$, we obtain the 4-dimensional Euclidean space.

12.4.3 Representation of the Creation and Annihilation Operators in QED

Photons, electrons, and positrons are described by the tensor product of the corresponding indefinite and definite Fock spaces.
Folklore

In Sects. 12.2.1 and 12.2.2, we have introduced creation and annihilation operators for photons, electrons, and positrons by postulating commutation and anticommutation relations along with additional properties concerning both the vacuum and adjoint operators. Motivated by the discussion in Sect. 12.4, we want to show that all of these relations can be realized by an *infinite-dimensional* mathematical model. To this end, we will use a modification of the Fock space introduced in Chap. 10.

The idea is to replace definite inner products by indefinite inner products.

In what follows, let us fix

- the cube $\mathcal{C}(L)$ of side length $L > 0$ in position space,
- the time interval $[-\frac{T}{2}, \frac{T}{2}]$ of length $T > 0$, and
- the finite lattice $\mathcal{G}(N)$ in momentum space.

[19] If X is a real linear space, then we assume that $\langle\psi|\varphi\rangle$ is always real, and α, β are real numbers in (i).

Choose the space-time point $x = (\mathbf{x}, t)$ along with $\mathbf{x} = x^1\mathbf{i} + x^2\mathbf{j} + x^3\mathbf{k}$, $t = x^0$, and

$$-\frac{L}{2} \le x^1, x^2, x^3 \le \frac{L}{2}, \qquad -\frac{T}{2} \le x^0 \le \frac{T}{2}.$$

Set $\Omega := \mathcal{C}(L) \times [-\frac{T}{2}, \frac{T}{2}]$. By Ω^n, we understand the product set $\Omega \times \cdots \times \Omega$ with n factors.

The photon Fock space. A one-photon function $A : \Omega \to \mathbb{C}^4$ has the form

$$A(x) := \begin{pmatrix} A^0(x) \\ A^1(x) \\ A^2(x) \\ A^3(x) \end{pmatrix}$$

where each of the components $A^\mu : \Omega \to \mathbb{C}$ lives in the complex Hilbert space $L_2(\Omega)$, that is,

$$\int_\Omega |A^\mu(x)|^2 d^4x < \infty, \qquad \mu = 0, 1, 2, 3.$$

We also introduce the indefinite inner product[20]

$$\langle A|B\rangle_1 := -\int_\Omega \eta_{\mu\nu} A^\mu(x)^\dagger B^\nu(x) d^4x.$$

More precisely, X_{phot} is a bosonic Fock space. This means that, by definition, the elements of X_{phot} are infinite sequences

$$\mathcal{A} = (\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2, \ldots)$$

where $\mathcal{A}_0 \in \mathbb{C}$. Moreover, $\mathcal{A}_1 \in L_2(\Omega, \mathbb{C}^4)$ along with

$$\mathcal{A}_n \in L_{2,\text{sym}}(\Omega^n, \mathbb{C}^{4^n}), \qquad n = 2, 3, \ldots$$

This means the following. The photon functions

$$\mathcal{A}_n(x_1, \ldots, x_n) = (\mathcal{A}^{\mu_1 \cdots \mu_n}(x_1, \ldots, x_n))$$

are symmetric with respect to both the indices $\mu_1, \ldots, \mu_n = 0, 1, 2, 3$ and the n arguments $x_1, \ldots, x_n \in \Omega$, and the components $\mathcal{A}_n^{\mu_1 \cdots \mu_n}$ live in $L_2(\Omega)$. We equip the space X_{phot} with the following indefinite inner product:

$$\boxed{\langle \mathcal{A}|\mathcal{B}\rangle := \mathcal{A}_0^\dagger \mathcal{B}_0 + \sum_{n=1}^{\infty} \langle \mathcal{A}_n|\mathcal{B}_n\rangle_n}$$

where

$$\langle \mathcal{A}_n|\mathcal{B}_n\rangle_n := (-1)^n \int_{\Omega^n} \eta_{\mu_1\nu_1} \cdots \eta_{\mu_n\nu_n} (\mathcal{A}_n^{\mu_1 \cdots \mu_n})^\dagger \mathcal{B}_n^{\nu_1 \cdots \nu_n} d^4x_1 \cdots d^4x_n.$$

More precisely, we only consider such tuples $(\mathcal{A}_n)$ and $(\mathcal{B}_n)$ of X_{phot} for which the indefinite inner product $\langle \mathcal{A}|\mathcal{B}\rangle$ is finite. In particular, the symbol $X_{\text{phot,fin}}$ contains all the elements $(\mathcal{A}_0, \mathcal{A}_1, \ldots)$ of X_{phot} where at most a finite number of elements $\mathcal{A}_0, \mathcal{A}_1, \ldots$ is different from zero. The state

[20] Recall that $-\eta_{\mu\nu} A^{\mu\dagger} B^\nu = -(A^0)^\dagger B^0 + \sum_{j=1}^3 (A^j)^\dagger B^j$.

$$\boxed{|0\rangle_{\text{phot}} := (1, 0, 0, \ldots)}$$

is called the photon vacuum. For the given momentum vector $\mathbf{p} \in \mathcal{G}(N)$ and the given polarization index $s = 0, 1, 2, 3$, we define special photon functions

$$A^{\mu}_{\mathbf{p},s}(\mathbf{x}, t) := \frac{\mathrm{e}^{\mathrm{i}(\mathbf{px}-\omega_{\mathbf{p}}t)}}{\sqrt{TL^3}}\, e^{\mu}_s, \qquad \mu = 0, 1, 2, 3.$$

From the orthogonality relation and the orthogonality relations for the polarization vectors e_s introduced in Sect. 12.2.1 on page 812, we get the key relation

$$\boxed{\langle A_{\mathbf{p},s} | A_{\mathbf{q},r} \rangle_1 = -\eta_{sr}\delta_{\mathbf{pq}}} \tag{12.28}$$

for all momentum vectors $\mathbf{p}, \mathbf{q} \in \mathcal{G}(N)$ and all polarization indices $s, r = 0, 1, 2, 3$.

Creation operators and annihilation operators for photons. We now proceed similarly as in Sect. 10.3 on page 784. For each given one-photon function $A_{\mathbf{p},s} \in X_1$, the creation operator

$$a^{+}_{\mathbf{p},s} : X_{\text{phot,fin}} \to X_{\text{phot}}$$

is defined by $a^{+}_{\mathbf{p},s}\mathcal{A} := (0, \mathcal{B}_1, \mathcal{B}_2, \ldots)$ where

$$\boxed{\mathcal{B}_{n+1} := \sqrt{n+1} \cdot \operatorname{sym}(A_{\mathbf{p},s} \otimes \mathcal{A}_n), \qquad n = 0, 1, 2, \ldots}$$

Explicitly, $\mathcal{B}^{\mu_1 \ldots \mu_{n+1}}_{n+1}(x_1, \ldots x_{n+1})$ is equal to

$$\frac{\sqrt{n+1}}{(n+1)!} \sum_{\pi} \pi\big(A^{\mu_1}_{\mathbf{p},s}(x_1)\mathcal{A}^{\mu_2 \ldots \mu_{n+1}}_n(x_2, \ldots, x_{n+1})\big)$$

where we sum over all permutations π of $1, \ldots, n+1$. The operation $\pi(\ldots)$ refers to permutations of both the indices $\mu_1, \ldots, \mu_{n+1}$ and the arguments $x_1, \ldots, x_{n+1}$. The annihilation operator

$$a^{-}_{\mathbf{p},s} : X_{\text{phot,fin}} \to X_{\text{phot}}$$

is the formally adjoint operator to the creation operator $a^{+}_{\mathbf{p},s}$, that is,

$$\boxed{\langle a^{-}_{\mathbf{p},s}\mathcal{A} | \mathcal{B} \rangle = \langle \mathcal{A} | a^{+}_{\mathbf{p},s}\mathcal{B} \rangle \qquad \text{for all} \quad \mathcal{A}, \mathcal{B} \in X_{\text{phot,fin}}.}$$

In other words, $a^{-}_{\mathbf{p},s} = (a^{+}_{\mathbf{p},s})^{\dagger}$ on $X_{\text{phot,fin}}$. Explicitly, for each given sequence $\mathcal{A} := (\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2, \ldots)$ in $X_{\text{phot,fin}}$ we define

$$a^{-}_{\mathbf{p},s}\mathcal{A} := (\mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2, \ldots)$$

where $\mathcal{C}^{\mu_1 \ldots \mu_n}_n(x_1, \ldots, x_n)$ is given by

$$-\sqrt{n+1} \int_{\Omega} \eta_{\mu\nu} A^{\mu}_{\mathbf{p},s}(x)^{\dagger} \mathcal{A}^{\nu\mu_1 \ldots \mu_n}_{n+1}(x, x_1, \ldots, x_n) d^4x.$$

In particular, $\mathcal{C}_0 = -\int_{\Omega} \eta_{\mu\nu} \mathcal{A}^{\mu}_{\mathbf{p},s}(x)^{\dagger} \mathcal{A}^{\nu}_1(x)\, dx$. For all $\mathcal{A} \in X_{\text{phot}}$ and all momentum vectors $\mathbf{p}, \mathbf{q} \in \mathcal{G}(N)$ as well as $s, r = 0, 1, 2, 3$, it follows analogously to Theorem 10.1 on page 776 that

$$\boxed{[a^-_{\mathbf{p},s}, a^+_{\mathbf{q},r}]_- \mathcal{A} = \langle A_{\mathbf{p},s} | A_{\mathbf{q},r} \rangle_1 \mathcal{A} = -\eta_{sr}\delta_{\mathbf{pq}}\mathcal{A}.}$$

Similarly,

$$[a^+_{\mathbf{p},s}, a^+_{\mathbf{q},r}]_- \mathcal{A} = [a^-_{\mathbf{p},s}, a^-_{\mathbf{q},r}]_- \mathcal{A} = 0.$$

Finally, $a^+_{\mathbf{p},s}|0\rangle_{\text{phot}} = 0$.

The electron-positron Fock space. By an electron-positron function, we understand a function $\psi : \Omega \to \mathbb{C}^4$ of the form

$$\psi(x) := \begin{pmatrix} \psi^1(x) \\ \psi^2(x) \\ \psi^3(x) \\ \psi^4(x) \end{pmatrix}$$

where each component $\psi^j : \Omega \to \mathbb{C}$ lives in the complex Hilbert space $L_2(\Omega)$, i.e.,

$$\int_\Omega |\psi^j(x)|^2 d^4x < \infty, \qquad j = 1, 2, 3, 4.$$

We also introduce the inner product

$$\langle \psi | \varphi \rangle_1 := \int_\Omega \psi(x)^\dagger \varphi(x) d^4x.$$

More precisely, Y_{el} is a fermionic Fock space. This means that, by definition, the elements of Y_{el} are infinite sequences

$$\Psi = (\Psi_0, \Psi_1, \Psi_2, \ldots)$$

where $\Psi_0 \in \mathbb{C}$. Moreover, $\Psi_1 \in L_2(\Omega, \mathbb{C}^4)$ along with

$$\Psi_n \in L_{2,\text{antisym}}(\Omega^n, \mathbb{C}^{4^n}), \qquad n = 2, 3, \ldots$$

This means the following. The electron-positron functions

$$\Psi_n(x_1, \ldots, x_n) = (\Psi_n^{j_1 \cdots j_n}(x_1, \ldots, x_n))$$

are antisymmetric with respect to both the indices $j_1, \ldots, j_n = 1, 2, 3, 4$ and the n arguments $x_1, \ldots, x_n \in \Omega$, and the components $\Psi_n^{j_1 \cdots j_n}$ live in $L_2(\Omega)$. We equip the space Y_{el} with the following inner product

$$\boxed{\langle \Psi | \Phi \rangle := \Psi_0^\dagger \Phi_0 + \sum_{n=1}^\infty \langle \Psi_n | \Phi_n \rangle_n}$$

where

$$\langle \Psi_n | \Phi_n \rangle_n := \int_{\Omega^n} \sum_{j_1, \ldots, j_n = 1}^4 (\Psi_n^{j_1 \cdots j_n})^\dagger \Phi_n^{j_1 \cdots j_n} d^4x_1 \cdots d^4x_n.$$

More precisely, we only consider such tuples (Ψ_n) and (Φ_n) of Y_{el} for which the inner product $\langle \Psi | \Phi \rangle$ is finite. In particular, the symbol $Y_{\text{el,fin}}$ contains all the elements $(\Psi_0, \Psi_1, \ldots)$ of Y_{el} where at most a finite number of elements $\Psi_0, \Psi_1, \ldots$ is different from zero. The state

$$\boxed{|0\rangle_{\text{el}} := (1, 0, 0, \ldots)}$$

is called the electron-positron vacuum. For the momentum vector $\mathbf{p} \in \mathcal{G}(N)$ and the spin index $s = \pm\frac{1}{2}$, we define the following special electron-positron functions:[21]

$$\psi^+_{\mathbf{p},s}(\mathbf{x},t) := \frac{1}{\sqrt{2E_{\mathbf{p}}L^3T}}\; u_{\mathbf{p},s}\mathrm{e}^{\mathrm{i}\mathbf{p}\mathbf{x}}\mathrm{e}^{-\mathrm{i}E_{\mathbf{p}}t},$$
$$\psi^-_{\mathbf{p},s}(\mathbf{x},t) := \frac{1}{\sqrt{2E_{\mathbf{p}}L^3T}}\; v_{\mathbf{p},s}\mathrm{e}^{-\mathrm{i}\mathbf{p}\mathbf{x}}\mathrm{e}^{\mathrm{i}E_{\mathbf{p}}t}.$$

From Theorem 12.4 on page 810 we get the key orthogonality relations

$$\langle\psi^\pm_{\mathbf{p},s}|\psi^\pm_{\mathbf{q},r}\rangle_1 = \delta_{sr}\delta_{\mathbf{pq}}, \qquad \langle\psi^\pm_{\mathbf{p},s}|\psi^\mp_{\mathbf{q},r}\rangle_1 = 0 \tag{12.29}$$

for all momentum vectors $\mathbf{p},\mathbf{q} \in \mathcal{G}(N)$ and all spin numbers $s, r = \pm\frac{1}{2}$.

Creation and annihilation operators for electrons. We proceed as in Sect. 10.3. For each given one-electron function $\psi^+_{\mathbf{p},s} \in X_1$, the creation operator

$$b^+_{\mathbf{p},s} : X_{\mathrm{el,fin}} \to X_{\mathrm{el}}$$

is defined by $b^+_{\mathbf{p},s}\Psi := (0, \Phi_1, \Phi_2, \ldots)$ where

$$\boxed{\Phi_{n+1} := \sqrt{n+1}\cdot \mathrm{antisym}(\psi^+_{\mathbf{p},s}\otimes\Psi_n), \qquad n = 0,1,2,\ldots}$$

Explicitly, $\Phi^{j_1\ldots j_{n+1}}_{n+1}(x_1,\ldots x_{n+1})$ is equal to

$$\frac{\sqrt{n+1}}{(n+1)!}\;\sum_\pi \mathrm{sgn}\,\pi\cdot\pi\big(\psi^{j_1}_{\mathbf{p},s}(x_1)\Psi^{j_2\ldots j_{n+1}}_n(x_2,\ldots,x_{n+1})\big)$$

where we sum over all permutations π of $1,\ldots,n+1$. The operation $\pi(\ldots)$ refers to permutations of both the indices $j_1,\ldots,j_{n+1}$ and the arguments $x_1,\ldots,x_{n+1}$. The annihilation operator

$$b^-_{\mathbf{p},s} : X_{\mathrm{el,fin}} \to X_{\mathrm{el}}$$

is the formally adjoint operator to the creation operator $b^+_{\mathbf{p},s}$, that is,

$$\boxed{\langle b^-_{\mathbf{p},s}\Psi|\Phi\rangle = \langle\Psi|b^+_{\mathbf{p},s}\Phi\rangle \qquad \text{for all} \quad \Psi,\Phi \in X_{\mathrm{el,fin}}.}$$

Explicitly, for each given sequence $\Psi := (\Psi_0,\Psi_1,\Psi_2,\ldots)$ in $X_{\mathrm{el,fin}}$ we define

$$b^-_{\mathbf{p},s}\Psi := (\Lambda_0,\Lambda_1,\Lambda_2,\ldots)$$

where $\Lambda^{j_1\ldots j_n}_n(x_1,\ldots,x_n)$ is given by

$$\sqrt{n+1}\int_\Omega \sum_{j=1}^n \psi^{+j}_{\mathbf{p},s}(x)^\dagger\Psi^{jj_1\ldots j_n}_{n+1}(x,x_1,\ldots,x_n)\,d^4x.$$

In particular, $\Lambda_0 = \int_\Omega \psi^+_{\mathbf{p},s}(x)^\dagger\Psi(x)\,d^4x$. For all $\Psi \in X_{\mathrm{el}}$ and all momentum vectors $\mathbf{p},\mathbf{q} \in \mathcal{G}(N)$ as well as $r,s = \pm\frac{1}{2}$, it follows analogously to Theorem 10.2 that

$$\boxed{[b^-_{\mathbf{p},s}, b^+_{\mathbf{q},r}]_-\Psi = \langle\psi^+_{\mathbf{p},s}|\psi^+_{\mathbf{q},r}\rangle_1\Psi = \delta_{sr}\delta_{\mathbf{pq}}\Psi.}$$

[21] Recall that $E_{\mathbf{p}} = \sqrt{m_e^2 + \mathbf{p}^2}$.

Similarly,

$$[b^+_{\mathbf{p},s}, b^+_{\mathbf{q},r}]_- \Psi = [b^-_{\mathbf{p},s}, b^-_{\mathbf{q},r}]_- \Psi = 0.$$

Finally, $b^-_{\mathbf{p},s}|0\rangle_{\mathrm{el}} = 0$.

Creation and annihilation operators for positrons. Replacing $\psi^+_{\mathbf{p},s}$ by $\psi^-_{\mathbf{p},s}$, we obtain the operators $c^\pm_{\mathbf{p},s}$ instead of $b^\pm_{\mathbf{p},s}$.

The combined Fock space. In order to describe combined states of photons, electrons, and positrons, we introduce the combined Fock space

$$\boxed{X := X_{\mathrm{phot}} \otimes X_{\mathrm{el}}.}$$

The elements of X are tuples of the form

$$(\mathcal{A}_m \otimes \Psi_n)_{m,n=0,1,\ldots}$$

Here, the tensor product $(\mathcal{A}_m \otimes \Psi_n)(x_1, \ldots, x_m, y_1, \ldots, y_n)$ is equal to the tuple

$$(\mathcal{A}^{\mu_1 \cdots \mu_m}_m(x_1, \ldots, x_m)\, \Psi^{j_1 \cdots j_n}_n(y_1, \ldots, y_n))$$

where $\mathcal{A}_m \in L_{2,\mathrm{sym}}(\Omega^m, \mathbb{C}^{4^m})$ and $\Psi_n \in L_{2,\mathrm{antisym}}(\Omega^n, \mathbb{C}^{4^n})$. The state

$$|0\rangle := |0\rangle_{\mathrm{phot}} \otimes |0\rangle_{\mathrm{el}}$$

is called the total vacuum state of X. In a natural way, the creation and annihilation operators for photons can be extended from the photon Fock space X_{phot} to the combined Fock space X by setting

$$\boxed{a^\pm_{\mathbf{p},s}(\mathcal{A}_m \otimes \Psi_n) := (a^\pm_{\mathbf{p},s}\mathcal{A}_m) \otimes \Psi_n.}$$

For the creation and annihilation operators of electrons and positrons, we obtain

$$b^\pm_{\mathbf{p},s}(\mathcal{A}_m \otimes \Psi_n) := \mathcal{A}_m \otimes (b^\pm_{\mathbf{p},s}\Psi_n)$$

and $c^\pm_{\mathbf{p},s}(\mathcal{A}_m \otimes \Psi_n) := \mathcal{A}_m \otimes (c^\pm_{\mathbf{p},s}\Psi_n)$, respectively.

12.4.4 Gupta–Bleuler Quantization

Eliminate virtual photons as external particles.
Folklore

The Gupta-Bleuler approach to quantum electrodynamics was invented in 1950.[22] The basic idea reads as follows. For states with negative norm $\langle\Psi|\Psi\rangle < 0$, there fails the usual probabilistic interpretation of states in quantum theory. Such states are called ghost states.

(i) Introduction of ghosts: To simplify the formulation, besides real transversal photons we introduce virtual photons (i.e., scalar and longitudinal photons) which are related to ghost states.

[22] N. Gupta, Theory of longitudinal photons in quantum electrodynamics, Proc. Phys. Soc. (London) **A63** (1950), 681–691.
K. Bleuler, A new method for treating longitudinal and scalar photons, Helv. Phys. Acta **23** (1950), 567–586 (in German).

(ii) Elimination of explicit ghosts: We distinguish between physical and nonphysical states. The restriction to physical states eliminates processes with virtual photons as incoming or outgoing particles.
(iii) The implicit influence of ghosts: Observe the crucial fact that the influence of the ghosts cannot be completely neglected. Interestingly enough, the cross sections measured in scattering processes do not only depend on the ghost-free incoming and out-going states, but also on the photon Feynman propagator which depends on both real and virtual photons.

The Gupta-Bleuler method is the prototype for quantizing gauge theories in the BRST setting. This will be considered in Vol. IV.

The trouble with the gauge condition. Consider first the classical plane wave

$$A^\mu(\mathbf{x},t) := \sum_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=0}^{3} (a^-_{\mathbf{p},s}\, \mathrm{e}^{\mathrm{i}(\mathbf{px}-\omega_{\mathbf{p}}t)} + a^+_{\mathbf{p},s}\, \mathrm{e}^{-\mathrm{i}(\mathbf{px}-\omega_{\mathbf{p}}t)})\, \mathcal{N}_{\mathbf{p}} \mathbf{e}^\mu_s(\mathbf{p})$$

for $\mu = 0,1,2,3$. Here, $a^-_{\mathbf{p},s}$ are complex numbers, and $a^+_{\mathbf{p},s}$ is the conjugate complex number to $a^-_{\mathbf{p},s}$. By Sect. 12.2.1 on page 812, the gauge condition

$$a^-_{\mathbf{p},0} = a^-_{\mathbf{p},3} \tag{12.30}$$

for all vectors $\mathbf{p} \in \mathcal{G}(N)$ implies that the electromagnetic field

$$\mathbf{E} := -\,\mathbf{grad}\, U - \dot{\mathbf{A}}, \qquad \mathbf{B} := \mathbf{curl}\, \mathbf{A} \tag{12.31}$$

satisfies the Maxwell equations. We now pass to the corresponding free quantum field $A^\mu(x)$ by using creation operators $a^+_{\mathbf{p},s}$ and annihilation operators $a^-_{\mathbf{p},s}$. Suppose that the gauge condition (12.30) remains valid on the operator level. Then, the same computation as for the classical field shows that the field operators (12.31) satisfy the Maxwell equations.

Unfortunately, the gauge condition (12.30) *contradicts the commutation relations for the creation and annihilation operators.*

In fact, according to Sect. 12.2.1, we have

$$a^-_{\mathbf{p},3}a^+_{\mathbf{p},3} - a^+_{\mathbf{p},3}a^-_{\mathbf{p},3} = I, \qquad a^-_{\mathbf{p},0}a^+_{\mathbf{p},0} - a^+_{\mathbf{p},0}a^-_{\mathbf{p},0} = -I$$

along with $a^+_{\mathbf{p},s} = (a^-_{\mathbf{p},s})^\dagger$ for $s = 0,1,2,3$. Thus, setting $a^-_{\mathbf{p},3} = a^-_{\mathbf{p},0}$, we get the contradiction $I = -I$. Therefore, we have to replace the gauge condition (12.30) by a weaker condition. This problem was solved by Gupta and Bleuler in the following way.

The physical space. By definition, the physical space X_{phys} is a linear subspace of the state space X which consists of all the states Ψ in X with the property

$$\boxed{a^-_{\mathbf{p},0}\Psi = 0, \qquad a^-_{\mathbf{p},3}\Psi = 0}$$

along with $\langle\Psi|\Psi\rangle \geq 0$. By a physical state Ψ, we understand an element of X_{phys} with $\langle\Psi|\Psi\rangle > 0$. Two elements Ψ and Φ of X_{phys} are called equivalent iff

$$\langle\Psi - \Phi|\Psi - \Phi\rangle = 0.$$

We write $\Psi \sim \Phi$. In particular, $\Psi \sim 0$ iff $\langle\Psi|\Psi\rangle = 0$.

Examples of physical states. Recall that a photon is called transversal iff its polarization index s attains the values $s = 1$ or $s = 2$. We are given the vectors $\mathbf{p}, \mathbf{q}, \mathbf{r} \in \mathcal{G}(N)$, the polarization indices

$$s, s_1, s_2 = 1, 2,$$

and the spin indices $\sigma, \tau = \pm\frac{1}{2}$. Then the following states are physical states, and they form an orthonormal system in the linear subspace X_{phys} of the combined Fock space X:

(i) Vacuum: $|0\rangle$.
(ii) One-particle states of transversal photons, electrons, and positrons:

$$a^+_{\mathbf{p},s}|0\rangle, \qquad b^+_{\mathbf{q},\sigma}\,|0\rangle, \qquad c^+_{\mathbf{r},\tau}\,|0\rangle.$$

For example, $a^+_{\mathbf{p},s}|0\rangle$ describes a photon of momentum vector $\mathbf{p}$ and transversal polarization $s = 1, 2$.
(iii) Two identical transversal photons:

$$\frac{(a^+_{\mathbf{p},s})^2\,|0\rangle}{\sqrt{2}}.$$

(iv) Two different transversal photons: $a^+_{\mathbf{p},s_1}a^+_{\mathbf{q},s_2}\,|0\rangle$.[23]
(v) Two different electrons: $b^+_{\mathbf{p},\sigma}b^+_{\mathbf{q},\tau}\,|0\rangle$.
(vi) Two different positrons: $c^+_{\mathbf{p},\sigma}c^+_{\mathbf{q},\tau}\,|0\rangle$.
(vii) Mixed two-particle states: $a^+_{\mathbf{p},s}b^+_{\mathbf{q},\sigma}\,|0\rangle$, $a^+_{\mathbf{p},s}c^+_{\mathbf{q},\tau}\,|0\rangle$, and $b^+_{\mathbf{p},\sigma}c^+_{\mathbf{q},\tau}\,|0\rangle$.
(viii) Three identical transversal photons:

$$\frac{(a^+_{\mathbf{p},s})^3\,|0\rangle}{\sqrt{3}}.$$

Proof. Ad (i). Set $\Psi_0 := |0\rangle$. Observe that $\langle\Psi_0|\Psi_0\rangle = 1$, and $a^-_{\mathbf{p},0}\Psi_0 = a^-_{\mathbf{p},3}\Psi_0 = 0$.

Ad (ii). For $s = 1, 2$, the commutation rule yields

$$(a^-_{\mathbf{p},s}a^+_{\mathbf{p},s} - a^+_{\mathbf{p},s}a^-_{\mathbf{p},s})\Psi_0 = I\Psi_0 = \Psi_0.$$

Since $a^-_{\mathbf{p},s}\Psi_0 = 0$, we obtain

$$\langle a^+_{\mathbf{p},s}\Psi_0|a^+_{\mathbf{p},s}\Psi_0\rangle = \langle\Psi_0|a^-_{\mathbf{p},s}a^+_{\mathbf{p},s}\Psi_0\rangle = \langle\Psi_0|\Psi_0\rangle = 1.$$

Moreover, we get $a^-_{\mathbf{p},0}(a^+_{\mathbf{p},s}\Psi_0) = a^+_{\mathbf{p},s}a^-_{\mathbf{p},0}\Psi_0 = 0$. Similarly, $a^-_{\mathbf{p},3}(a^+_{\mathbf{p},s}\Psi_0) = 0$. An analogous argument yields

$$(b^-_{\mathbf{p},\sigma}b^+_{\mathbf{p},\sigma} + b^+_{\mathbf{p},\sigma}b^-_{\mathbf{p},\sigma})\Psi_0 = \Psi_0.$$

Since $b^-_{\mathbf{p},\sigma}\Psi_0 = 0$, we get

$$\langle b^+_{\mathbf{p},\sigma}\Psi_0|b^+_{\mathbf{p},\sigma}\Psi_0\rangle = \langle\Psi_0|b^-_{\mathbf{p},\sigma}b^+_{\mathbf{p},\sigma}\Psi_0\rangle = \langle\Psi_0|\Psi_0\rangle = 1.$$

Moreover, $a^-_{\mathbf{p},r}(b^+_{\mathbf{p},\sigma}\Psi_0) = b^+_{\mathbf{p},\sigma}a^-_{\mathbf{p},r}\Psi_0 = 0$ for $r = 0, 3$.

Ad (iii). It follows from the commutation rule

[23] Naturally enough, we exclude the case where $\mathbf{p} = \mathbf{q}$ and $s_1 = s_2$. In (v) and (vi), we exclude the case where $\mathbf{p} = \mathbf{q}$ and $\sigma = \tau$.

$$(a^-_{\mathbf{p},s}a^+_{\mathbf{p},s} - a^+_{\mathbf{p},s}a^-_{\mathbf{p},s})\Psi = \Psi$$

and $a^-_{\mathbf{p},s}\Psi_0 = 0$ that $a^-_{\mathbf{p},s}a^+_{\mathbf{p},s}\Psi_0 = \Psi_0$. Hence

$$a^-_{\mathbf{p},s}(a^+_{\mathbf{p},s})^2\Psi_0 = a^+_{\mathbf{p},s}\Psi_0 + a^+_{\mathbf{p},s}(a^-_{\mathbf{p},s}a^+_{\mathbf{p},s})\Psi_0 = 2a^+_{\mathbf{p},s}\Psi_0.$$

This implies

$$\langle(a^+_{\mathbf{p},s})^2\Psi_0|(a^+_{\mathbf{p},s})^2\Psi_0\rangle = \langle a^+_{\mathbf{p},s}\Psi_0|a^-_{\mathbf{p},s}(a^+_{\mathbf{p},s})^2\Psi_0\rangle = 2\langle a^+_{\mathbf{p},s}\Psi_0|a^+_{\mathbf{p},s}\Psi_0\rangle = 2.$$

Finally, $a^-_{\mathbf{p},r}(a^+_{\mathbf{p},r})^2\Psi_0 = (a^+_{\mathbf{p},s})^2a^-_{\mathbf{p},r}\Psi_0 = 0$ for $r = 0, 3$. The remaining proofs proceed analogously. □

Examples of nonphysical states (ghost states). Let

$$s, s_1, s_2 = 0, 3$$

and $\sigma, \tau = \pm\frac{1}{2}$. The following states are not physical states:

(a) One-particle state of virtual photons: $a^+_{\mathbf{p},3}\,|0\rangle$, $a^+_{\mathbf{p},0}\,|0\rangle$.
(b) Two identical electrons: $(b^+_{\mathbf{p},\sigma})^2\,|0\rangle$.
(c) Two identical positrons: $(c^+_{\mathbf{p},\sigma})^2\,|0\rangle$.
(d) Two-particle states including one virtual photon: $a^+_{\mathbf{p},s}b^+_{\mathbf{q},\sigma}|0\rangle$.
(e) Two virtual photons: $a^+_{\mathbf{p},s_1}a^+_{\mathbf{q},s_2}|0\rangle$.

Proof. Ad (a). Again let $\Psi_0 := |0\rangle$. Note that

$$a^-_{\mathbf{p},0}(a^+_{\mathbf{p},0}\Psi_0) = -\Psi_0 - a^+_{\mathbf{p},0}a^-_{\mathbf{p},0}\Psi_0 = -\Psi_0.$$

Since this is different from zero, $a^+_{\mathbf{p},0}\Psi_0$ is not a physical state. Similarly, we obtain

$$a^-_{\mathbf{p},3}(a^+_{\mathbf{p},3}\Psi_0) = \Psi_0.$$

Ad (b), (c). We have $(b^+_{\mathbf{p},\sigma})^2\Psi_0 = 0$. This follows from the anticommutation rule

$$2(b^+_{\mathbf{p},\sigma})^2 = [b^+_{\mathbf{p},\sigma}, b^+_{\mathbf{p},\sigma}]_+ = 0.$$

Similarly, $(c^+_{\mathbf{p},\sigma})^2\Psi_0 = 0$.

Ad (d), (e). Noting that

$$a^-_{\mathbf{p},0}a^+_{\mathbf{p},0} = -I + a^+_{\mathbf{p},0}a^-_{\mathbf{p},0}$$

along with $a^-_{\mathbf{p},0}b^+_{\mathbf{q},\sigma} = b^+_{\mathbf{q},\sigma}a^-_{\mathbf{p},0}$ and $a^-_{\mathbf{p},0}\Psi_0 = 0$, we obtain

$$a^-_{\mathbf{p},0}(a^+_{\mathbf{p},0}b^+_{\mathbf{q},\sigma}\Psi_0) = -b^+_{\mathbf{q},\sigma}\Psi_0.$$

This is different from zero. The remaining cases are treated similarly. □

13. The Interacting Quantum Field, and the Magic Dyson Series for the *S*-Matrix

> Whoever understands the *S*-matrix (scattering matrix) can understand everything in the theory of scattering processes for elementary particles.[1]
>
> Folklore

> Both Kaiser's admirable *Drawing Theories Apart* and Schweber's *QED and the Men Who Made It* refer frequently to the famous lectures on quantum electrodynamics given by Freeman Dyson at Cornell University (Ithaca, New York) in 1951.[2] Two generations ago, graduate students and their professors wishing to learn the new techniques of QED passed around copies of Dyson's Cornell lectures, then the best and fullest treatment available.[3]
>
> David Derbes, 2007

Let us first summarize the two key formulas (13.1) and (13.6). The motivation will be given below.

13.1 Dyson's Key Formula

Fix the time interval $[-\frac{T}{2}, \frac{T}{2}]$ of length $T > 0$. By definition, Dyson's *S*-matrix operator reads as

$$\boxed{\mathsf{S}(T) := \mathcal{T} \exp\left(-\mathrm{i}\int_{-T/2}^{T/2} H_{\mathrm{int}}(t)dt\right)} \tag{13.1}$$

with $H_{\mathrm{int}}(t) := \int_{\mathcal{C}(L)} : \mathcal{H}_{\mathrm{int,free}}(\mathbf{x}, t) : d^3\mathbf{x}$, and

$$\boxed{\mathcal{H}_{\mathrm{int,free}} := -e\overline{\psi}_{\mathrm{free}} \not{A}_{\mathrm{free}} \psi_{\mathrm{free}}.}$$

[1] For the fascinating history of Dyson's discovery, we refer to the beginning of Section 1.2 in Volume I where we quote from Dyson's book *Disturbing the Universe*, Harper and Row, New York, 1979. See also F. Dyson, The *S*-matrix in quantum electrodynamics, Phys. Rev. **75** (1949), 1736–1755.

[2] D. Kaiser, Drawing Theories Apart: The Dispersion of Feynman Diagrams in Postwar Physics, University of Chicago Press, 2005.
S. Schweber, QED (Quantum Electrodynamics) and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga, Princeton University Press, 1994.

[3] From the preface to: F. Dyson, Advanced Quantum Mechanics. Reprinted by permission of World Scientific Publishing Co. Pte. Ltd. Singapore, 2007.

Explicitly, we have[4]

$$\mathcal{H}_{\text{int,free}}(\mathbf{x},t) = -e\overline{\psi}_{\text{free}}(\mathbf{x},t)\gamma_\mu A^\mu_{\text{free}}(\mathbf{x},t)\psi_{\text{free}}(\mathbf{x},t).$$

The meaning of the time ordering operator $\mathcal{T}$ will be explained in (T) below.[5]

The magic Dyson formula (13.1) *represents the key formula of quantum electrodynamics.*

Using the classical power series expansion $\exp x = 1 + x + \frac{1}{2}x^2 + \ldots$, formula (13.1) reads explicitly as

$$\begin{aligned}\mathsf{S}(T) = I - \mathrm{i}\int_{-T/2}^{T/2} H_{\text{int}}(t)dt \\ -\frac{1}{2}\int_{-T/2}^{T/2} dt \int_{-T/2}^{T/2} ds\ \ \mathcal{T}(H_{\text{int}}(t)H_{\text{int}}(s)) + \ldots\end{aligned} \tag{13.2}$$

Since $H_{\text{int}} = e\ldots$, this is an expansion with respect to the small parameter e. In general, $\mathsf{S}(T)$ is equal to

$$\sum_{n=0}^{\infty} \frac{(-\mathrm{i})^n}{n!} \int_{-T/2}^{T/2} \cdots \int_{-T/2}^{T/2} \mathcal{T}(H_{\text{int}}(t_1)\cdots H_{\text{int}}(t_n))\ dt_1 \cdots dt_n.$$

This is the magic Dyson series in quantum electrodynamics. Unfortunately, the convergence of the Dyson series is an open problem, and there exist heuristic arguments which indicate the divergence of this series. The finite series

$$\sum_{n=0}^{m} \frac{(-\mathrm{i})^n}{n!} \int_{-T/2}^{T/2} \cdots \int_{-T/2}^{T/2} \mathcal{T}(H_{\text{int}}(t_1)\cdots H_{\text{int}}(t_n))\ dt_1 \cdots dt_n$$

is called the finite Dyson series of order m. Setting $H_{\text{int}} := -eV_{\text{int}}$, we obtain

$$\sum_{n=0}^{m} \frac{\kappa^n \mathrm{i}^n}{n!} \int_{-T/2}^{T/2} \cdots \int_{-T/2}^{T/2} \mathcal{T}(V_{\text{int}}(t_1)\cdots V_{\text{int}}(t_n))\ dt_1 \cdots dt_n$$

with the small coupling constant $\kappa := e = \sqrt{4\pi\alpha_{\text{QED}}}$.

The graphical representation of the S-matrix operator $\mathsf{S}(T)$ *in terms of Feynman diagrams will be considered in Chap. 14.*

It is crucial that the formula (13.1) above for Dyson's S-matrix operator is explicitly known, since $\mathcal{H}_{\text{int,free}}$ refers to the interaction term in quantum electrodynamics expressed by free fields of electrons, positrons, and photons. In fact, A_{free} denotes the free photon quantum field from Sect. 12.2.1:

[4] Recall that $\gamma_\mu A^\mu = \gamma^\mu A_\mu = \sum_{\mu=0}^{3} \gamma^\mu A_\mu$, where $A_0 := A^0$, and $A_j := -A^j$ for $j = 1, 2, 3$.

[5] According to our convention (10.17), we work in the energetic system, that is, $\hbar = c = \varepsilon_0 = \mu_0 = k := 1$. Hence $e = \sqrt{4\pi\alpha_{\text{QED}}}$ with the electromagnetic fine structure constant $\alpha_{\text{QED}} = 1/137.04$. This means that e and α_{QED} are small dimensionless quantities. In order to obtain the formula for $\mathsf{S}(T)$ in the SI system, we have to replace $-\mathrm{i}$ by the quotient $-\mathrm{i}/\hbar$, and e by the product ec.

$$A^{\mu}_{\text{free}}(\mathbf{x},t) := \sum_{\mathbf{p}\in\mathcal{G}(N)}\sum_{s=0}^{3} (a^{-}_{\mathbf{p},s}\, \mathrm{e}^{\mathrm{i}(\mathbf{px}-\omega_{\mathbf{p}}t)} + \\ + a^{+}_{\mathbf{p},s}\, \mathrm{e}^{-\mathrm{i}(\mathbf{px}-\omega_{\mathbf{p}}t)})\, \mathcal{N}_{\mathbf{p}} \mathbf{e}^{\mu}_{s}(\mathbf{p}) \tag{13.3}$$

along with the normalization factor $\mathcal{N}_{\mathbf{p}} := 1/\sqrt{2L^3\omega_{\mathbf{p}}}$, and the photon energy $\omega_{\mathbf{p}} = |\mathbf{p}|$. We set $\mathcal{N}_{\mathbf{p}} := 0$ if $\mathbf{p} = 0$. Furthermore, ψ_{free} denotes the free electron-positron quantum field from (12.11), that is,

$$\psi_{\text{free}}(\mathbf{x},t) := \sum_{\mathbf{p}\in\mathcal{G}(N)}\sum_{s=\pm\frac{1}{2}} (b^{-}_{\mathbf{p},s} u_{\mathbf{p},s}\, \mathrm{e}^{\mathrm{i}(\mathbf{px}-E_{\mathbf{p}}t)} + \\ + c^{+}_{\mathbf{p},s} v_{\mathbf{p},s}\, \mathrm{e}^{-\mathrm{i}(\mathbf{px}-E_{\mathbf{p}}t)})\, N_{\mathbf{p}} \tag{13.4}$$

with the normalization factor $N_{\mathbf{p}} := 1/\sqrt{2L^3E_{\mathbf{p}}}$, and the particle energy

$$E_{\mathbf{p}} = \sqrt{m_e^2 + \mathbf{p}^2}.$$

For the adjoint free field, we obtain

$$\overline{\psi}_{\text{free}}(\mathbf{x},t) := \sum_{\mathbf{p}\in\mathcal{G}(N)}\sum_{s=\pm\frac{1}{2}} (b^{+}_{\mathbf{p},s} \overline{u}_{\mathbf{p},s}\, \mathrm{e}^{-\mathrm{i}(\mathbf{px}-E_{\mathbf{p}}t)} + \\ + c^{-}_{\mathbf{p},s} \overline{v}_{\mathbf{p},s}\, \mathrm{e}^{\mathrm{i}(\mathbf{px}-E_{\mathbf{p}}t)})\, N_{\mathbf{p}}. \tag{13.5}$$

Furthermore, observe the following:

(N) Normal operator product: The symbol $:\mathcal{H}_{\text{int,free}}:$ denotes the normal operator product. By definition, the factors are reordered in such a way that all of the creation operators stand to the left of the annihilation operators. For fermionic operators, we add a sign factor which is the sign of the permutation needed for obtaining the desired reordering. For example, in the case of photons we get

$$: a^{+}_{\mathbf{p},s} a^{-}_{\mathbf{q},r} : \; = \; : a^{-}_{\mathbf{q},r} a^{+}_{\mathbf{p},s} : \; = a^{+}_{\mathbf{p},s} a^{-}_{\mathbf{q},r},$$

whereas we obtain

$$: b^{+}_{\mathbf{p},\sigma} b^{-}_{\mathbf{q},\tau} : \; = - : b^{-}_{\mathbf{q},\tau} b^{+}_{\mathbf{p},\sigma} : \; = b^{+}_{\mathbf{p},\sigma} b^{-}_{\mathbf{q},\tau}$$

in the case of electrons (or positrons).[6]

(T) Time ordering: The symbol $\mathcal{T}$ refers to time-ordering. By definition, the time-depending operators are reordered in such way that the time arguments increase from right to left. For fermionic operators, we add a sign factor which is the sign of the permutation needed for obtaining the desired reordering. For example, if $t > s$, then

$$\mathcal{T}(B(\mathbf{x},t)B(\mathbf{y},s)) = \mathcal{T}(B(\mathbf{y},s)B(\mathbf{x},t)) = B(\mathbf{x},t)B(\mathbf{y},s)$$

for bosonic operators (e.g., the photon field), and

$$\mathcal{T}(F(\mathbf{x},t)F(\mathbf{y},s)) = -\mathcal{T}(F(\mathbf{y},s)F(\mathbf{x},t)) = F(\mathbf{x},t)F(\mathbf{y},s)$$

for fermionic operators (e.g., electron fields or positron fields). For equal times, $t = s$, the $\mathcal{T}$-product coincides with the usual product. The symbol $\mathcal{T}$ is also called the chronological operator.

[6] The general definition of the normal product can be found on page 823.

The key formula for the transition probability in scattering processes between electrons, positrons, and photons. Consider a scattering process where Φ_{in} and Φ_{out} are unit vectors which correspond to the incoming and outgoing particles, respectively. By definition, the number

$$\boxed{\mathcal{W}(T) := |\langle \Phi_{\text{in}} | \mathsf{S}(T) \Phi_{\text{out}} \rangle|^2} \tag{13.6}$$

represents the transition probability from the incoming particles to the outgoing particles during the time-interval $[-\frac{T}{2}, \frac{T}{2}]$. For example, the symbol

$$\Phi_{\text{in}} := a^+_{\mathbf{p},s} b^+_{\mathbf{q},\sigma} |0\rangle$$

describes two incoming particles, namely,

- one incoming photon of momentum vector $\mathbf{p}$ and polarization $s = 1, 2$, and
- one incoming electron of momentum vector $\mathbf{q}$ and spin number $\sigma = \pm\frac{1}{2}$.

Similarly, the symbol

$$\Phi_{\text{out}} := a^+_{\mathbf{p}',s'} b^+_{\mathbf{q}',\sigma'} |0\rangle$$

corresponds to one outgoing photon and one outgoing electron. The scattering process described by Φ_{in} and Φ_{out} is called Compton scattering in physics. The transition probability and the related cross section for Compton scattering will be computed in Sect. 15.1. In general, for scattering processes in quantum electrodynamics, we will use the normalized states

$$\Phi_{\text{in}} = \frac{(a^+_{\mathbf{p}_1,s_1})^{n_1}}{\sqrt{n_1!}} \cdots \frac{(a^+_{\mathbf{p}_A,s_A})^{n_A}}{\sqrt{n_A!}} b^+_{\mathbf{q}_1,\sigma_1} \cdots b^+_{\mathbf{q}_B,\sigma_B} c^+_{\mathbf{r}_1,\tau_1} \cdots c^+_{\mathbf{r}_C,\tau_C} |0\rangle$$

which correspond to photons, electrons, and positrons. The particles live in the cubic box $\mathcal{C}(L)$ of volume $\mathcal{V} := L^3$. The particle density ϱ of the photons corresponding to $a^+_{\mathbf{p}_1,s_1}$ is equal to

$$\varrho = \frac{n_1}{\mathcal{V}},$$

and so on. Similar expressions are used for Φ_{out}.

Observe that Φ_{in} and Φ_{out} do not depend on time t. This is motivated by formula (13.13) below.

Motivation for Dyson's S-matrix operator. Let us motivate the key formula (13.1) above. The idea is the following:

- First consider rigorously the finite-dimensional situation.
- Then pass to the infinite-dimensional case by using a formal argument in a straightforward manner. This motivates the definition (13.1).

To begin with, let us start with the rigorous Schrödinger equation

$$\mathrm{i}\dot{\Psi}(t) = (\mathsf{H}_{\text{free}} + \mathsf{H}_{\text{int}})\Psi(t), \qquad t \in \mathbb{R}, \tag{13.7}$$

in the n-dimensional complex Hilbert space $\mathbb{C}^n$. Here,

$$\Psi(t) = \begin{pmatrix} \Psi^1(t) \\ \vdots \\ \Psi^n(t) \end{pmatrix}.$$

The given complex $(n \times n)$-matrices H_{free} and H_{int} are self-adjoint; they do not depend on time t. Now we use the following rigorous arguments:

(i) The propagator of the Schrödinger equation: For given initial state $\Psi(t_0)$, the unique solution of the Schrödinger equation (13.7) reads as

$$\Psi(t) = P(t,t_0)\Psi(t_0), \qquad t \in \mathbb{R}, \tag{13.8}$$

with the propagator $P(t,t_0) := \exp\left(-\mathrm{i}\mathsf{H}_{\mathrm{free}} - \mathrm{i}\mathsf{H}_{\mathrm{int}})(t-t_0)\right)$. It is our goal to separate the free dynamics. We want to show that

$$P(t,t_0) = \mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\mathrm{free}}}\mathsf{S}(t,t_0)\mathrm{e}^{\mathrm{i}t_0\mathsf{H}_{\mathrm{free}}} \tag{13.9}$$

along with

$$\boxed{\mathsf{S}(t,t_0) := \mathcal{T}\exp\left(-\mathrm{i}\int_{t_0}^{t} H_{\mathrm{int}}(\tau)d\tau\right)} \tag{13.10}$$

where $H_{\mathrm{int}}(t) := \mathrm{e}^{\mathrm{i}t\mathsf{H}_{\mathrm{free}}}\mathsf{H}_{\mathrm{int}}\mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\mathrm{free}}}$.

(ii) The propagator of the transformed Schrödinger equation (the S-matrix): The key idea is to use the transformation

$$\Psi(t) = \mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\mathrm{free}}}\,\Phi(t). \tag{13.11}$$

It follows from the Schrödinger equation (13.7) that

$$\mathsf{H}_{\mathrm{free}}\mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\mathrm{free}}}\Phi(t) + \mathrm{i}\mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\mathrm{free}}}\dot{\Phi}(t) = (\mathsf{H}_{\mathrm{free}} + \mathsf{H}_{\mathrm{int}})\mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\mathrm{free}}}\Phi(\mathsf{t}).$$

Hence

$$\mathrm{i}\mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\mathrm{free}}}\dot{\Phi}(t) = \mathsf{H}_{\mathrm{int}}\mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\mathrm{free}}}\Phi(t).$$

This way, we get the transformed Schrödinger equation

$$\mathrm{i}\dot{\Phi}(t) = H_{\mathrm{int}}(t)\Phi(t), \qquad t \in \mathbb{R}.$$

For given initial state $\Phi(t_0)$, the unique solution reads as

$$\Phi(t) = \mathsf{S}(t,t_0)\Phi(t_0)$$

along with (13.10). This follows from Sect. 7.17.4 of Vol. I. By (13.8), we have $\Psi(t) = P(t,t_0)\Psi(t_0)$. Using (13.11), we obtain

$$\Phi(t) = \mathrm{e}^{\mathrm{i}t\mathsf{H}_{\mathrm{free}}}P(t,t_0)\mathrm{e}^{-\mathrm{i}t_0\mathsf{H}_{\mathrm{free}}}\Phi(t_0).$$

This tells us the validity of (13.9).

(iii) The transition probability: According to the rules of quantum mechanics, the real number

$$\boxed{\mathcal{W}(t,t_0) := |\langle\Psi_{\mathrm{out}}(t)|P(t,t_0)\Psi_{\mathrm{in}}(t_0)\rangle|^2} \tag{13.12}$$

is the transition probability from the state Ψ_{in} at time t_0 to the state Ψ_{out} at time t provided $||\Psi_{\mathrm{in}}(t_0)|| = 1$ and $||\Psi_{\mathrm{out}}(t)|| = 1$. Let us consider the special case where

$$\Psi_{\mathrm{in}}(t) := \mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\mathrm{free}}}\Psi_{\mathrm{in}}(0), \qquad \Psi_{\mathrm{out}}(t) := \mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\mathrm{free}}}\Psi_{\mathrm{out}}(0)$$

with $||\Psi_{\mathrm{in}}(0)|| = ||\Psi_{\mathrm{out}}(0)|| = 1$. The functions $t \mapsto \Psi_{\mathrm{in}}(t)$ and $t \mapsto \Psi_{\mathrm{out}}(t)$ are solutions of the Schrödinger equation (13.7) with vanishing interaction, $\mathsf{H}_{\mathrm{int}} = 0$. Furthermore,

$$||\Psi_{\text{in}}(t)|| = ||\Psi_{\text{out}}(t)|| = 1 \qquad \text{for all} \quad t \in \mathbb{R},$$

since the operator $\mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\text{free}}}$ is unitary for any time t. Intuitively, the functions $t \mapsto \Psi_{\text{in}}(t)$ and $t \mapsto \Psi_{\text{out}}(t)$ represent an incoming and outgoing free particle, respectively. Using (13.9), we get

$$\mathcal{W}(t, t_0) = |\langle \Psi_{\text{out}}(0) | \mathsf{S}(t, t_0) \Psi_{\text{in}}(0) \rangle|^2.$$

We are given the time interval $[-\frac{T}{2}, \frac{T}{2}]$. Choosing the initial time $t_0 := -\frac{T}{2}$ and the final time $t := \frac{T}{2}$, and setting $\mathsf{S}(T) := \mathsf{S}(t, t_0)$, we obtain the key formula

$$\boxed{\mathcal{W}(T) := |\langle \Psi_{\text{out}}(0) | \mathsf{S}(T) \Psi_{\text{in}}(0) \rangle|^2.} \tag{13.13}$$

Now let us use the rigorous finite-dimensional approach (i)–(iii) above in order to motivate the definition of Dyson's S-matrix operator $\mathsf{S}(T)$ from (13.1) on page 835.

Step 1: The Schrödinger equation for states in quantum electrodynamics: Let $\Psi(t)$ denote the state of a system of electrons, positrons, and photons at time t. Then the Schrödinger equation reads as

$$\mathrm{i}\dot{\Psi}(t) = (\mathsf{H}_{\text{free}} + \mathsf{H}_{\text{int}})\Psi(t), \qquad t \in \mathbb{R}. \tag{13.14}$$

Here, the Hamiltonian H_{free} of the interaction-free situation is given by (12.21) on page 817. Furthermore, motivated by Theorem 11.2 on page 796, the interaction term of the Hamiltonian is given by

$$\mathsf{H}_{\text{int}} := \int_{\mathcal{C}(L)} \mathcal{H}_{\text{int}}(\mathbf{x}) d^3\mathbf{x}$$

with

$$\boxed{\mathcal{H}_{\text{int}}(\mathbf{x}) := -e\overline{\psi}(\mathbf{x})\, \not{A}(\mathbf{x})\psi(\mathbf{x}).}$$

Recall Feynman's slash symbol $\not{A}(x) := \gamma_\mu A^\mu(x)$. In this connection, observe the following crucial fact. In the Schrödinger approach to quantum physics, the state vectors $\Psi(t)$ depend on time t whereas the observables (i.e., the physical quantities) B do not depend on time t. This motivates our assumption that H_{int} depends on the time-independent quantum field ψ, A.

Step 2: The free Schrödinger equation: Consider the special case where the interaction vanishes, $\mathsf{H}_{\text{int}} = 0$. Then, the dynamics of the state vectors is given by

$$\Psi(t) = \mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\text{free}}}\Psi(0), \qquad t \in \mathbb{R}.$$

If we define

$$B(t) := \mathrm{e}^{\mathrm{i}t\mathsf{H}_{\text{free}}} B \mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\text{free}}}, \qquad t \in \mathbb{R}, \tag{13.15}$$

then we obtain

$$\langle \Psi(t) | B\Psi(t) \rangle = \langle \Psi(0) | B(t)\Psi(0) \rangle, \qquad t \in \mathbb{R}.$$

This shows the invariance of the inner product, and hence the invariance of expectation values. Physicists call the time-dependent operator function $B = B(t)$ the Heisenberg picture of the observable B with respect to the free Hamiltonian.

Step 3: The transformed Schrödinger equation: As in (ii) above, the transformation $\Psi(t) = \mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\mathrm{free}}}\Phi(t)$ sends the Schrödinger equation (13.14) of quantum electrodynamics to the equation

$$\mathrm{i}\dot{\Phi}(t) = H_{\mathrm{int}}(t)\Phi(t), \qquad t \in \mathbb{R}$$

with $H_{\mathrm{int}}(t) := \mathrm{e}^{\mathrm{i}t\mathsf{H}_{\mathrm{free}}}\mathsf{H}_{\mathrm{int}}\mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\mathrm{free}}}$. Explicitly,

$$H_{\mathrm{int}}(t) := \int_{\mathcal{C}(L)} \mathrm{e}^{\mathrm{i}t\mathsf{H}_{\mathrm{free}}}\mathcal{H}_{\mathrm{int}}(\mathbf{x})\mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\mathrm{free}}} d^3\mathbf{x}.$$

Now to the point. Observe that $\mathrm{e}^{\mathrm{i}t\mathsf{H}_{\mathrm{free}}}\mathcal{H}_{\mathrm{int}}(\mathbf{x})\mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\mathrm{free}}}$ is equal to the product

$$-e(\mathrm{e}^{\mathrm{i}t\mathsf{H}_{\mathrm{free}}}\overline{\psi}(\mathbf{x})\mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\mathrm{free}}}) \cdot (\mathrm{e}^{\mathrm{i}t\mathsf{H}_{\mathrm{free}}}\, \not{A}(\mathbf{x})\mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\mathrm{free}}}) \cdot (\mathrm{e}^{\mathrm{i}t\mathsf{H}_{\mathrm{free}}}\psi(\mathbf{x})\mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\mathrm{free}}}).$$

Motivated by the Heisenberg picture (13.15), we define

$$\boxed{\mathrm{e}^{\mathrm{i}t\mathsf{H}_{\mathrm{free}}}\mathcal{H}_{\mathrm{int}}(\mathbf{x})\mathrm{e}^{-\mathrm{i}t\mathsf{H}_{\mathrm{free}}} := -e\overline{\psi}_{\mathrm{free}}(\mathbf{x},t)\, \not{A}_{\mathrm{free}}(\mathbf{x},t)\psi_{\mathrm{free}}(\mathbf{x},t)}$$

where $\psi_{\mathrm{free}}(\mathbf{x},t)$ and $A_{\mathrm{free}}(\mathbf{x},t)$ describe the dynamics of free quantum fields in quantum electrodynamics.

Step 4: Dyson's S-matrix operator: Using (13.10), we obtain the S-matrix operator $\mathsf{S}(T)$ from (13.1) on page 835. More precisely, additionally, we have to replace the operator product by the normal operator product.

Step 5: The normal operator product: Motivated by Sect. 12.2.2 on page 814, we replace the Hamiltonian operator of quantum electrodynamics by its normal product. This way we obtain the modified Schrödinger equation

$$\mathrm{i}\dot{\Psi}(t) = (:\mathsf{H}_{\mathrm{free}}: + :\mathsf{H}_{\mathrm{int}}:)\Psi(t), \qquad t \in \mathbb{R} \tag{13.16}$$

in quantum electrodynamics. With respect to Dyson's S-matrix operator $\mathsf{S}(T)$, this means that we have to replace the operator $\mathsf{H}_{\mathrm{int}}$ by the normal product $:\mathsf{H}_{\mathrm{int}}:$ in (13.1).

This finishes the motivation of the key definition (13.1) of Dyson's S-matrix.

13.2 The Basic Strategy of Reduction Formulas

> The computation of scattering processes in quantum field theory can be elegantly reduced to the computation of propagators by using the Wick theorem.
>
> Folklore

For scattering processes during the time-interval $[-\frac{T}{2}, \frac{T}{2}]$, our goal is to compute the transition probability

$$\boxed{\mathcal{W}(T) := |\langle\Psi_{\mathrm{out}}(0)|\mathsf{S}(T)\Psi_{\mathrm{in}}(0)\rangle|^2}$$

based on Dyson's S-operator S. According to Feynman, the inner product

$$\langle\Psi_{\mathrm{out}}(0)|\mathsf{S}(T))\Psi_{\mathrm{in}}(0)\rangle$$

is called the transition amplitude. We will show that the transition amplitudes can be reduced to vacuum expectation values

$$\langle 0|\mathcal{T}(A_1 A_2 \cdots A_n)|0\rangle$$

where $A_1, ..., A_n$ are linear combinations of creation and annihilation operators, and $\mathcal{T}$ denotes the operator of time-ordering. In turn, by reduction formulas, these vacuum expectation values can be reduced to the simpler vacuum expectation values

$$\langle 0|\mathcal{T}(AB)|0\rangle$$

of two factors. These expressions are called propagators.

Therefore, it remains to compute the propagators.

The reduction formulas can be represented graphically by using Feynman diagrams. This will be studied in Chap. 14. In this chapter, we emphasize the algebraic aspects of the approach. This should help the reader to understand that the elegant language of Feynman diagrams has a sound mathematical basis in terms of combinatorics. In particular, the algebraic approach clearly justifies the appearance and the specific values of the symmetry factors attributed to Feynman diagrams.

For the computation of propagators, there exist the following two different universal methods:

(i) The Fourier method based on Fourier series for classical free fields.
(ii) The method of Feynman functional integrals (or path integrals) based on Gaussian integrals.

In what follows, we will use method (i) in quantum electrodynamics, whereas method (ii) will be thoroughly studied in Vol. IV on quantum mathematics.

Prototype of the Wick theorem. To explain the simple basic idea, let us consider the typical transition amplitude

$$\boxed{\tau := \langle b^+ \Omega | ABa^+ \Omega\rangle}$$

where $\Omega := |0\rangle$ denotes the vacuum state, and a^+, b^+ are two (bosonic) creation operators with the corresponding annihilation operators a^-, b^-, respectively. Furthermore, A, B are complex linear combinations of creation and annihilation operators. First let us represent the transition amplitude τ as a vacuum expectation value. Since we have the key relation $(b^+)^\dagger = b^-$ between creation and annihilation operators, we get

$$\tau = \langle 0|b^- ABa^+|0\rangle.$$

The main trick is to define the so-called contraction

$$\boxed{\mathcal{C}(AB) := \langle 0|AB|0\rangle} \tag{13.17}$$

between the operators A and B. We claim the validity of the following reduction formula for vacuum expectation values:

$$\boxed{\langle 0|b^- ABa^+|0\rangle = \mathcal{C}(b^- A)\mathcal{C}(Ba^+) + \mathcal{C}(b^- B)\mathcal{C}(Aa^+).} \tag{13.18}$$

As another example, we mention the following reduction formula:

$$\langle 0|a^- ABa^+|0\rangle = \mathcal{C}(a^- A)\mathcal{C}(Ba^+) + \mathcal{C}(a^- B)\mathcal{C}(Aa^+) + \mathcal{C}(AB). \tag{13.19}$$

Before proving this, we need some preparations. Decompose

$$A = A^+ + A^-, \qquad B = B^+ + B^-.$$

Here, A^+, B^+ (resp. A^-, B^-) are linear combinations of creation (resp. annihilation) operators.

Special contractions for creation and annihilation operators. For the creation operators a^+, b^+ and the annihilation operators a^-, b^-, we have the following commutation relations:

$$\boxed{[a^-, a^+]_- = I, \qquad [b^-, b^+]_- = I,} \tag{13.20}$$

where I denotes the identical operator. In addition, we get

$$[a^-, b^\pm]_- = 0, \qquad [a^+, b^\pm]_- = 0.$$

Furthermore, $(a^\pm)^\dagger = (a^\mp)^\dagger$ and $(b^\pm)^\dagger = (b^\mp)^\dagger$. Finally, we have

$$\boxed{a^-|0\rangle = 0, \qquad b^-|0\rangle = 0.} \tag{13.21}$$

Proposition 13.1 *There hold* $\mathcal{C}(a^+a^-) = \mathcal{C}(a^+a^+) = \mathcal{C}(a^-a^-) = 0$, *and*

$$\mathcal{C}(a^-a^+) = 1,$$

as well as $\mathcal{C}(a^\pm b^\pm) = \mathcal{C}(a^\pm b^\mp) = 0$.

Proof. Set $\Omega := |0\rangle$. It follows from $a^-\Omega = 0$ that

$$\mathcal{C}(a^\pm a^-) = \langle \Omega | a^\pm a^- \Omega \rangle = 0.$$

Moreover, noting that $(a^+)^\dagger = a^-$, we get

$$\mathcal{C}(a^+a^\pm) = \langle \Omega | a^+ a^\pm \Omega \rangle = \langle a^-\Omega | a^\pm \Omega \rangle = 0.$$

In addition, it follows from $a^-a^+ = a^+a^- + I$ and $\mathcal{C}(I) = \langle \Omega | I\Omega \rangle = \langle \Omega | \Omega \rangle = 1$ that

$$\mathcal{C}(a^-a^+) = \mathcal{C}(a^+a^-) + \mathcal{C}(I) = 1.$$

Since $a^-b^\pm = b^\pm a^-$, we obtain

$$\mathcal{C}(a^-b^\pm) = \mathcal{C}(b^\pm a^-) = \langle \Omega | b^\pm a^- \Omega \rangle = 0.$$

Finally, $\mathcal{C}(a^+b^\pm) = \langle \Omega | a^+ b^\pm \Omega \rangle = \langle a^-\Omega | b^\pm \Omega \rangle = 0$. □

Further relations. We claim that Prop. 13.1 implies the following key relations:

$$\boxed{[A^-, B^+]_- = \mathcal{C}(AB)I} \tag{13.22}$$

and

$$[A^-, a^+]_- = \mathcal{C}(Aa^+)I, \qquad [b^-, A^+] = \mathcal{C}(b^-A)I.$$

Furthermore, $[A^+, B^+]_- = [A^-, B^-]_- = 0$. Finally, we have

$$[c^+, C^+]_- = [c^-, C^-]_- = 0$$

for $C = A, B$ and $c = a, b$.

Proof of (13.22). By the commutation relations for creation and annihilation operators, there exists a complex number $\mathcal{A}$ such that

$$[A^-, B^+]_- = \mathcal{A}I.$$

Moreover, $A^-|0\rangle = 0$. This implies

$$\langle 0|A^+|0\rangle = 0.$$

In fact, if A^+ is a creation operator, then $(A^+)^\dagger = A^-$. Set $\Omega := |0\rangle$. Then

$$\langle \Omega|A^+\Omega\rangle = \langle A^-\Omega|\Omega\rangle = 0.$$

The same argument tells us that $\langle 0|A^+A_1A_2\cdots A_n|0\rangle = 0$. Hence

$$\begin{aligned}\mathcal{A} &= \langle 0|\,[A^-, B^+]_-|0\rangle = \langle 0|A^-B^+ - B^+A^-|0\rangle = \langle 0|A^-B^+|0\rangle \\ &= \langle 0|(A^- + A^+)(B^+ + B^-)|0\rangle = \langle 0|AB|0\rangle = \mathcal{C}(AB).\end{aligned}$$

□

Proof of (13.18). We will use a brute force method. The more elegant approach via the Wick theorem will be considered in the next section. The idea is to use the operator identity[7]

$$\boxed{CD = [C, D]_- + DC}$$

for moving the annihilation operators a^-, A^-, B^- to the right. Finally, we will apply the relations $A^-|0\rangle = B^-|0\rangle = a^-|0\rangle = b^-|0\rangle = 0$. This implies

$$\boxed{\langle 0|\cdots C^-|0\rangle = 0} \tag{13.23}$$

for $C = A, B, a, b$. Furthermore, since $(C^+)^\dagger = C^-$, we also get

$$\boxed{\langle 0|C^+\cdots|0\rangle = 0.} \tag{13.24}$$

The point is that this procedure cancels a lot of terms.

In what follows, the two key relations (13.23) and (13.24) will be used again and again. In particular, we get

$$\mathcal{C}(A^+B) = \mathcal{C}(AB^-) = 0.$$

This implies $\mathcal{C}(AB) = \mathcal{C}((A^+ + A^-)(B^+ + B^-)) = \mathcal{C}(A^-B^+)$. To begin with, let us use the following decomposition:

$$\begin{aligned}b^-ABa^+ &= b^-(A^+ + A^-)(B^+ + B^-)a^+ \\ &= b^-A^+B^+a^+ + b^-A^+B^-a^+ + b^-A^-B^+a^+ + b^-A^-B^-a^+.\end{aligned}$$

For example, consider the product $b^-A^+B^-a^+$. From the identity

$$B^-a^+ = [B^-, a^+]_- + a^+B^- = \mathcal{C}(Ba^+)I + a^+B^-$$

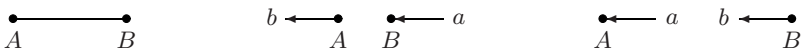
we get

$$b^-A^+B^-a^+ = \mathcal{C}(Ba^+)b^-A^+ + b^-A^+a^+B^-.$$

Since $\langle 0|b^-A^+a^+B^-|0\rangle = 0$, it remains to consider the term $\mathcal{C}(Ba^+)b^-A^+$. From

$$\mathcal{C}(Ba^+)b^-A^+ = \mathcal{C}(Ba^+)\mathcal{C}(b^-A)I + \mathcal{C}(Ba^+)A^+b^-$$

[7] Recall that $[C, D]_- := CD - DC$ and $[C, D]_+ := CD + DC$.



(a) $\mathcal{C}(AB)$ (b) $\mathcal{C}(b^- A)\mathcal{C}(Ba^+)$ (c) $\mathcal{C}(Aa^+)\mathcal{C}(b^- B)$

Fig. 13.1. Feynman diagrams

and $\langle 0|\cdots b^-|0\rangle = 0$ we get

$$\langle 0|b^- A^+ B^- a^+|0\rangle = \mathcal{C}(Ba^+)\mathcal{C}(b^- A)\langle 0|I|0\rangle = \mathcal{C}(Ba^+)\mathcal{C}(b^- A).$$

The other products are handled similarly. □

In an analogous manner, one obtains the reduction formula (13.19). Using the Wick theorem, a more elegant proof will be given on page 849.

Feynman diagrams. The terms from the reduction formulas (13.18) and (13.19) can be represented graphically as pictured in Fig. 13.1. These graphs are prototypes of Feynman diagrams containing both

- external lines corresponding to incoming and outgoing particles (also called legs),
- and internal lines (also called propagators).

In contrast to external lines, internal lines connect two nodes. We distinguish between

- connected Feynman graphs (Fig. 13.1(a)), and
- disconnected Feynman graphs (Fig. 13.1(b),(c)).

The Feynman graphs of quantum electrodynamics will be considered in Chap. 14. A Feynman graph without external lines is called an amputated graph.

Propagators, Wightman's functions, Green's functions, and reduction formulas. The reduction formulas (13.18) and (13.19) on page 842 can be equivalently written as

$$\langle 0|b^- ABa^+|0\rangle = \langle 0|b^- A|0\rangle \cdot \langle 0|Ba^+|0\rangle + \langle 0|b^- B|0\rangle \cdot \langle 0|Aa^+|0\rangle \qquad (13.25)$$

and

$$\begin{aligned}\langle 0|a^- ABa^+|0\rangle \; &= \langle 0|a^- A|0\rangle \cdot \langle 0|Ba^+|0\rangle + \langle 0|a^- B|0\rangle \cdot \langle 0|Aa^+|0\rangle \\ &\quad + \langle 0|AB|0\rangle. \end{aligned} \qquad (13.26)$$

Physicists use the following terminology:

- $\mathcal{C}(AB) = \langle 0|AB|0\rangle$ (Wightman propagator or 2-point Wightman function).
- $\mathcal{C}(b^- A) = \langle 0|b^- A|0\rangle$ and $\mathcal{C}(Ba^+) = \langle 0|Ba^+|0\rangle$ (external legs).
- If the operators A and B depend on time, and if the symbol $\mathcal{T}$ denotes the time-ordering operator, then
$$\langle 0|\mathcal{T}(AB)|0\rangle$$
is called the 2-point Green's function (or the Feynman propagator). This crucial quantity describes correlations in quantum field theory.

Formulas (13.25) and (13.26) are prototypes of reduction formulas in quantum field theory.[8] These formulas tells us that:

[8] Important reduction formulas were introduced by H. Lehmann, K. Symanzik, and W. Zimmermann, On the formulation of quantized field theories, Nuovo Cimento **1** (1955), 205–225; **6** (1957), 319–333. They are called the LSZ reduction formulas.

The transition amplitudes of the S-matrix can be reduced to propagators and external legs.

The external legs correspond to the free classical fields of incoming and outgoing particles. Propagators describe the interactions of the particles during the scattering process.The symmetry of general reduction formulas will be revealed in the next section.

13.3 The Wick Theorem

Wick's main theorem is a simple consequence of the algebraic relation

$$CD = [C, D]_{\mp} \pm DC.$$

This allows the effective computation of transition probabilities in perturbative scattering theory for all kinds of quantum field theories including quantum electrodynamics and the Standard Model in particle physics. Feynman diagrams graphically represent Wick's main theorem.

Folklore

Wick's main theorem on vacuum expectation values. Consider the vacuum expectation value

$$\tau := \langle 0|\mathcal{T}(A_1 A_2 \cdots A_n)|0\rangle, \qquad n = 1, 2, \ldots$$

of a time-ordered operator product. We assume that each time-dependent operator function $A_j = A_j(t)$ is either bosonic or fermionic. By definition, a bosonic (resp. fermionic) operator is a finite complex linear combination of bosonic (resp. fermionic) creation and annihilation operators. Wick's main theorem on general reduction formulas reads as follows.

Theorem 13.2 *If n is odd, then $\tau = 0$. If n is even, then τ is equal to the sum of all the products*

$$\eta \cdot \langle 0|\mathcal{T}(A_{i_1} A_{i_2})|0\rangle \cdot \langle 0|\mathcal{T}(A_{i_3} A_{i_4})|0\rangle \cdots \langle 0|\mathcal{T}(A_{i_{n-1}} A_{i_n})|0\rangle$$

where $i_1, \ldots, i_n$ is a permutation of $1, \ldots, n$, and η is the sign of this permutation by taking only the fermionic factors A_j into account. Moreover, the sequence $i_1, i_3, \ldots, i_{n-1}$ is increasing, and $i_1 < i_2$, $i_3 < i_4$, $\ldots$, $i_{n-1} < i_n$.

Let us first consider a few examples. To simplify notation, set $A_1 := A, \ldots$ For $n = 3$, Theorem 13.2 tells us that

$$\langle 0|\mathcal{T}(ABC)|0\rangle = 0.$$

Now let us consider the case where $n = 4$. Set $\mathcal{T}(AB)_0 := \langle 0|\mathcal{T}(AB)|0\rangle$. Theorem 13.2 yields the following:

(i) Bosonic case: Let A, B, C, D be bosonic operators. Then, we get the following reduction formula:

$$\mathcal{T}(ABCD)_0 \;=\; \mathcal{T}(AB)_0\mathcal{T}(CD)_0 + \mathcal{T}(AC)_0\mathcal{T}(BD)_0 + \mathcal{T}(AD)_0\mathcal{T}(BC)_0.$$

(ii) Fermionic case: Let A, B, C, D be fermionic operators. Then

$$\mathcal{T}(ABCD)_0 \;=\; \mathcal{T}(AB)_0\mathcal{T}(CD)_0 - \mathcal{T}(AC)_0\mathcal{T}(BD)_0 + \mathcal{T}(AD)_0\mathcal{T}(BC)_0.$$

Note that $ACBD$ is an odd permutation of $ABCD$.

(iii) Mixed case: Let A, B, C be fermionic, and let D be bosonic. Then

$$\mathcal{T}(ABCD)_0 \;=\; \mathcal{T}(AB)_0\mathcal{T}(CD)_0 - \mathcal{T}(AC)_0\mathcal{T}(BD)_0 + \mathcal{T}(AD)_0\mathcal{T}(BC)_0.$$

Observe that $ACBD$ is an odd permutation of the fermionic operators A, B, C.

Before giving the proof of Theorem 13.2 on page 851 (bosonic case) and page 855 (fermionic case and bosonic-fermionic case), we need some preparations concerning paired normal products and time-ordered products. This leads to the first and second Wick theorem below. We have to distinguish between bosonic and fermionic operators.

Bosonic Operators

Suppose that $A_1, \ldots, A_n, A, B, \ldots$ are bosonic operators. Decompose

$$A_j = A_j^+ + A_j^-, \qquad j = 1, \ldots, n$$

where A_j^+ (resp. A_j^-) contains creation (resp. annihilation) operators. By Sect. 13.2,

$$\cdots A^-|0\rangle = 0 \quad \text{and} \quad \langle 0|A^+ \cdots |0\rangle = 0.$$

Consequently, for the computation of vacuum expectation values, normal operator products play a fundamental role. Recall that the normal operator product

$$: A_1 A_2 \cdots A_n :$$

represents a reordering of the operators such that the annihilation operators stand to the right of the creation operators. For example, consider the product

$$AB = (A^+ + A^-)(B^+ + B^-) = A^+B^+ + A^+B^- + A^-B^- + A^-B^+.$$

Only the product A^-B^+ has to be reordered to B^+A^-. Hence

$$\boxed{: AB := A^+B^+ + A^+B^- + A^-B^- + B^+A^-.} \tag{13.27}$$

By (13.22) on page 843, $A^-B^+ = B^+A^- + \mathcal{C}(AB) \cdot I$. Therefore,

$$\boxed{AB = \; : AB : + \mathcal{C}(AB) \cdot I.} \tag{13.28}$$

This is the prototype of the first Wick theorem below. The normal product principle tells us that:

Vacuum expectation values of normal operator products vanish.

For example, $\langle 0| : AB : |0\rangle = 0$, since $\langle \cdots A^-|0\rangle = \langle 0|A^+ \cdots \rangle = 0$, and so on. The idea of the first Wick theorem is to reduce operator products to both normal operator products and multiples of the unit operator. By the normal product principle, only the multiples of the unit operator contribute to vacuum expectation values. To formulate the first Wick theorem, it is convenient to introduce the notion of paired normal product:

- One pairing: We define $: \underbrace{AB}: := \mathcal{C}(AB) \cdot I$, and

$$: \underbrace{ABC}: := \mathcal{C}(AB) \cdot C, \qquad : A\underbrace{BC}: := \mathcal{C}(AC) \cdot B, \qquad : \underbrace{ABC}: := \mathcal{C}(BC) \cdot A.$$

  Furthermore, we define

$$: \underbrace{ABCD}: := \mathcal{C}(AB) : CD :, \qquad : A\underbrace{BC}D: := \mathcal{C}(AC) : BD :,$$

  and so on.
- Two pairings: We define $: \underbrace{AB}\,\underbrace{CD}: := \mathcal{C}(AB)\mathcal{C}(CD) \cdot I$, and

$$: \underbrace{\underbrace{ABCD}}: := \mathcal{C}(AC)\mathcal{C}(BD) \cdot I, \qquad : \underbrace{A\underbrace{BC}D}: := \mathcal{C}(AD)\mathcal{C}(BC) \cdot I.$$

Analogously, we define three pairings, four pairings, and so on. Let us now formulate the *first Wick theorem*.

Theorem 13.3 *The operator product $A_1 A_2 \cdots A_n$ is equal to the sum over all possible paired normal operator products:*

$$: A_1 A_2 A_3 \cdots A_n : + : \underbrace{A_1 A_2} A_3 \cdots A_n : + \ldots$$
$$+ : \underbrace{A_1 A_2 A_3 \cdots A_n} : + : \underbrace{A_1 A_2} \underbrace{A_3 A_4} \cdots A_n : + \ldots$$

Examples. Before proving Theorem 13.3, let us consider some examples:

(i) One factor: $: A := A$, and $\langle 0|A|0\rangle = \langle 0|A^+|0\rangle + \langle 0|A^-|0\rangle = 0$.

(ii) Two factors:

$$\boxed{AB = \, : AB : + : \underbrace{AB} : .} \tag{13.29}$$

This stands for $AB = \, : AB : + \mathcal{C}(AB) \cdot I$ which was proved in (13.28).

(iii) Three factors:

$$\boxed{ABC = \, : ABC : + : \underbrace{AB}C : + : \underbrace{ABC} : + : A\underbrace{BC} : .}$$

This stands for

$$ABC = \, : ABC : + \mathcal{C}(AB)C + \mathcal{C}(AC)B + \mathcal{C}(BC)A.$$

By the normal product principle, we get

$$\boxed{\langle 0|ABC|0\rangle = 0.}$$

(iv) Four factors:

$$\begin{aligned} ABCD = \, & : ABCD : + : \underbrace{AB}CD : + : \underbrace{ABC}D : + : \underbrace{ABCD} : \\ & + : \underbrace{AB}\underbrace{CD} : + : \underbrace{\underbrace{ABC}D} : + : A\underbrace{\underbrace{BCD}} : \\ & + : \underbrace{AB}\,\underbrace{CD} : + : \underbrace{A\underbrace{BC}D} : + : \underbrace{AB\underbrace{CD}} : . \end{aligned}$$

By the normal product principle, only the totally paired normal products, namely,

$$:\underbrace{AB}\,\underbrace{CD}: := \mathcal{C}(AB)\mathcal{C}(CD)\cdot I, \qquad :\underbrace{AB\,CD}: := \mathcal{C}(AC)\mathcal{C}(BD)\cdot I,$$

and $:\underbrace{AB\,CD}: := \mathcal{C}(AD)\mathcal{C}(BC)$ contribute to the vacuum expectation value. Hence

$$\boxed{\langle 0|ABCD|0\rangle = \mathcal{C}(AB)\mathcal{C}(CD) + \mathcal{C}(AC)\mathcal{C}(BD) + \mathcal{C}(AD)\mathcal{C}(BC).}$$

In particular, noting that $\mathcal{C}(b^-a^+) = 0$, we get

$$\langle 0|b^-BCa^+|0\rangle = \mathcal{C}(b^-B)\mathcal{C}(Ca^+) + \mathcal{C}(b^-C)\mathcal{C}(Ba^+).$$

This is the reduction formula (13.18) on page 842. Noting that $\mathcal{C}(a^-a^+) = 1$, we also obtain the following reduction formula:

$$\langle 0|a^-BCa^+|0\rangle = \mathcal{C}(a^-B)\mathcal{C}(Ca^+) + \mathcal{C}(a^-C)\mathcal{C}(Ba^+) + \mathcal{C}(BC) \tag{13.30}$$

which coincides with (13.19).

Proof of the first Wick Theorem 13.3 in the bosonic case.

Step 1: For $n = 2$, we get the statement from (ii) above.

Step 2: Consider now the case where $n = 3$. We will reduce this to the statement for $n = 2$.

(I) To this end, we start with the decomposition

$$ABC = A^+(BC) + A^-(BC).$$

Moving A^- to the right, we obtain[9]

$$(A^-B)C = [A^-, B]_+C - (BA^-)C = \mathcal{C}(AB)C - (BA^-)C.$$

Analogously, $B(A^-C) = \mathcal{C}(AC)B + B(CA^-)$. Hence

$$A^-(BC) = (BC)A^- + \mathcal{C}(AB)C + \mathcal{C}(AC)B.$$

This implies

$$ABC = A^+(BC) + (BC)A^- + \mathcal{C}(AB)C + \mathcal{C}(AC)B.$$

(II) Using the known case (13.29) for two factors, $n = 2$, we get

$$A^+(BC) = A^+ : BC : + A^+ : \underbrace{BC}: ,$$

and $(BC)A^- = :BC:A^- + :\underbrace{BC}:A^-$. Hence

$$ABC = (A^+ : BC : + : BC : A^-) + A^+ : \underbrace{BC}: + : \underbrace{BC}: A^- + : \underbrace{ABC}: + : \underbrace{ABC}: .$$

[9] Note that $[A^-, B^-]_- = 0$. Thus, by (13.22), we get

$$[A^-, B^+ + B^-]_- = [A^-, B^+]_- = \mathcal{C}(AB)\cdot I.$$

(III) By definition of the normal operator product, we have

$$A^+ : BC := : A^+BC : \quad \text{and} \quad : BC : A^- = : BCA^- : .$$

Note that the bosononic normal product is invariant under permutations of the factors. Hence $: BC : A^- = : BCA^- : = : A^-BC : .$ Consequently,

$$A^+ : BC : + : BC : A^- = : (A^+ + A^-)BC : = : ABC : .$$

Finally, this yields

$$ABC = : ABC : + : \underbrace{ABC} : + : \underbrace{ABC} : + : \underbrace{ABC} : .$$

This is precisely the claim for $n = 3$. The general case $n = 4, 5, \ldots$ follows analogously by induction. □

Time-ordered operator products. Now we want to replace operator products by time-ordered operator products. To this end, we assume that the bosonic operators $A_1, A_2, \ldots, A, B, \ldots$ depend on time t. Recall that time-ordering of an operator product refers to a reordering of the factors such that time increases from right to left. This is motivated by causality. In what follows, time-ordered products always refer to points in time $t_1, t_2, \ldots, t_n$ which are pairwise different. By definition, we have

$$\mathcal{T}(A_1(t_1)A_2(t_2)\cdots A_n(t_n)) := A_{i_1}(t_{i_1})A_{i_2}(t_{i_2})\cdots A_{i_n}(t_{i_n})$$

where $i_1, i_2, \ldots, i_n$ is a permutation of $1, 2, \ldots, n$ such that $t_{i_1} > t_{i_2} > \ldots > t_{i_n}$. For example, if $t > s$, then

$$\mathcal{T}(A(t)B(s)) = \mathcal{T}(B(s)A(t)) := A(t)B(s).$$

Let us introduce the time-ordered contraction

$$\boxed{\mathcal{C}_\mathcal{T}(A(t)B(s)) := \langle 0|\mathcal{T}(A(t)B(s))|0\rangle.}$$

Time-ordered paired normal products are defined as above by replacing $\mathcal{C}(AB)$ by $\mathcal{C}_\mathcal{T}(AB)$. For example, we define:

- $: \underbrace{AB} : \; = \mathcal{C}_\mathcal{T}(AB) \cdot I,$
- $: \underbrace{AB}\underbrace{CD} : \; = \mathcal{C}_\mathcal{T}(AB)\mathcal{C}_\mathcal{T}(CD) \cdot I.$

Let us first consider the prototype of two factors. We claim that there holds

$$\boxed{\mathcal{T}(A(t)B(s)) = : A(t)B(s) : + : \underbrace{A(t)B}(s) : .} \tag{13.31}$$

This stands for

$$\mathcal{T}(A(t)B(s)) = : A(t)B(s) : + \mathcal{C}_\mathcal{T}(A(t)B(s)) \cdot I.$$

To prove (13.31), let us start with the case where $t > s$. Then

$$A(t)B(s) = : A(t)B(s) : + : \underbrace{A(t)B}(s) : .$$

This is identical with (13.31). Observe now that both the left-hand side and the right-hand side of (13.31) are invariant under a permutation of t and s. Therefore, equation (13.31) remains valid if $s > t$. The same argument yields the following *second Wick theorem*.

Theorem 13.4 *The time-ordered operator product $\mathcal{T}(A_1A_2\cdots A_n)$ is equal to the sum over the normal product and all the possible time-ordered pairings of the normal product,*

$$: A_1A_2A_3\cdots A_n : + : \underbrace{A_1}A_2A_3\cdots A_n : + \ldots$$
$$+ : \underbrace{A_1A_2A_3\cdots A_n} : + : \underbrace{A_1}\underbrace{A_2}\underbrace{A_3}A_4\cdots A_n : + \ldots$$

Proof of Wick's Main Theorem 13.2 in the bosonic case. Using the normal product principle, we obtain that the vacuum expectation value

$$\langle 0|\mathcal{T}(A_1 ... A_n)|0\rangle$$

is equal to the sum over all vacuum expectation values of totally paired normal products. For example, we get

$$\begin{aligned}\mathcal{T}(ABCD) = {} & : ABCD : + : \underbrace{A}BCD : + : A\underbrace{B}CD : + : AB\underbrace{C}D : + : ABC\underbrace{D} : \\ & + : \underbrace{A}\underbrace{B}CD : + : \underbrace{A}B\underbrace{C}D : + : \underbrace{A}BC\underbrace{D} : + : A\underbrace{B}\underbrace{C}D : + : AB\underbrace{C}\underbrace{D} : \\ & + : \underbrace{A}\underbrace{B}\underbrace{C}D : + : \underbrace{A}\underbrace{B}C\underbrace{D} : + : \underbrace{A}B\underbrace{C}\underbrace{D} : + : A\underbrace{B}\underbrace{C}\underbrace{D} : + : \underbrace{A}\underbrace{B}\underbrace{C}\underbrace{D} : .\end{aligned}$$

The totally paired normal products stand in the last line. This is equal to

$$\mathcal{C}_{\mathcal{T}}(AB)\mathcal{C}_{\mathcal{T}}(CD)\cdot I + \mathcal{C}_{\mathcal{T}}(AC)\mathcal{C}_{\mathcal{T}}(BD)\cdot I + \mathcal{C}_{\mathcal{T}}(AD)\mathcal{C}_{\mathcal{T}}(BC)\cdot I.$$

By the normal product principle, the vacuum expectation value

$$\langle 0|\mathcal{T}(ABCD)|0\rangle$$

is equal to

$$\mathcal{C}_{\mathcal{T}}(AB)\mathcal{C}_{\mathcal{T}}(CD) + \mathcal{C}_{\mathcal{T}}(AC)\mathcal{C}_{\mathcal{T}}(BD) + \mathcal{C}_{\mathcal{T}}(AD)\mathcal{C}_{\mathcal{T}}(BC).$$

Since $\mathcal{C}_{\mathcal{T}}(AB) = \langle 0|\mathcal{T}(AB)|0\rangle$, we get the claim of Theorem 13.2. The general case proceeds analogously. □

Fermionic Operators

Now assume that A, B are fermionic operators. As in the bosonic case, we define the contraction $\mathcal{C}(AB)$ by setting

$$\boxed{\mathcal{C}(AB) := \langle 0|AB|0\rangle.} \tag{13.32}$$

Special contractions for fermionic creation and annihilation operators. For two fermionic creation operators a^+, b^+ and the corresponding fermionic annihilation operators a^-, b^-, we have the following anticommutation relations:

$$\boxed{[a^-, a^+]_+ = I, \qquad [b^-, b^+]_+ = I,} \tag{13.33}$$

where I denotes the identical operator. In addition, we get

$$[a^-, b^\pm]_+ = 0, \qquad [a^+, b^\pm]_+ = 0.$$

Furthermore, $(a^\pm)^\dagger = (a^\mp)^\dagger$ and $(b^\pm)^\dagger = (b^\mp)^\dagger$. Finally, we have

$$\boxed{a^-|0\rangle = 0, \qquad b^-|0\rangle = 0.} \tag{13.34}$$

Proposition 13.5 *There hold* $\mathcal{C}(a^+a^-) = \mathcal{C}(a^+a^+) = \mathcal{C}(a^-a^-) = 0$, *and*

$$\mathcal{C}(a^-a^+) = 1,$$

as well as $\mathcal{C}(a^\pm b^\pm) = \mathcal{C}(a^\pm b^\mp) = 0$.

Proof. Set $\Omega := |0\rangle$. It follows from $a^-\Omega = 0$ that

$$\mathcal{C}(a^\pm a^-) = \langle \Omega | a^\pm a^- \Omega \rangle = 0.$$

Moreover, noting that $(a^+)^\dagger = a^-$, we get

$$\mathcal{C}(a^+a^\pm) = \langle \Omega | a^+ a^\pm \Omega \rangle = \langle a^-\Omega | a^\pm \Omega \rangle = 0.$$

In addition, it follows from $a^-a^+ = -a^+a^- + I$ and $\mathcal{C}(I) = \langle \Omega | I\Omega \rangle = \langle \Omega | \Omega \rangle = 1$ that

$$\mathcal{C}(a^-a^+) = -\mathcal{C}(a^+a^-) + \mathcal{C}(I) = 1.$$

Since $a^-b^\pm = -b^\pm a^-$, we obtain

$$\mathcal{C}(a^-b^\pm) = -\mathcal{C}(b^\pm a^-) = -\langle \Omega | b^\pm a^- \Omega \rangle = 0.$$

Finally, $\mathcal{C}(a^+b^\pm) = \langle \Omega | a^+ b^\pm \Omega \rangle = \langle a^-\Omega | b^\pm \Omega \rangle = 0$. □

General construction. Suppose that the operators $A_1, \ldots, A_n, A, B, \ldots$ are fermionic.

The first Wick Theorem 13.3 *and the second Wick Theorem* 13.4 *remain valid for fermionic operators.*

Let us discuss this. The point is that we have to modify the definition of normal products, paired normal products, and time-ordered products. The reason for that is the relation

$$CD = [C, D]_+ - DC$$

which forces us to introduce additional signs related to the sign of permutations of fermionic operators. The modified notions read as follows:

(a) Normal product: The normal operator product

$$: A_1 A_2 \cdots A_n :$$

represents a reordering of the fermionic operators such that the annihilation operators stand to the right of the creation operators. In contrast to the bosonic case, there appears an additional sign which is the sign of the necessary permutation of factors. For example, $: A := A$, and

$$: A^+B^- := A^+B^-, \qquad : B^-A^+ := -A^+B^-.$$

From the product formula

$$AB = (A^+ + A^-)(B^+ + B^-) = A^+B^+ + A^+B^- + A^-B^- + A^-B^+$$

we get

$$\boxed{: AB := A^+B^+ + A^+B^- + A^-B^- - B^+A^-.} \tag{13.35}$$

This implies

$$AB = \ : AB : + [A^-, B^+]_+. \tag{13.36}$$

Furthermore, for three factors, we obtain

$$:A^+B^+C^-:=A^+B^+C^-,\qquad :A^+C^-B^+:=-A^+B^+C^-.$$

In general, the normal product remains unchanged (resp. changes sign) under even (resp. odd) permutations of the fermionic factors. The normal product principle remains valid, that is, the vacuum expectation values of normal products vanish.

(b) Contraction: Using (13.32), we get[10]

$$[A^-,B^+]_+=\mathcal{C}(AB)\cdot I. \tag{13.37}$$

Moreover, $[A^-,B^-]_+=0$, by Prop. 13.5. Hence

$$[A^-,B]_+=[A^-,B^++B^-]_+=\mathcal{C}(AB)\cdot I. \tag{13.38}$$

(c) Paired normal product: Defining

$$:\underbrace{AB}:\ :=\mathcal{C}(AB)\cdot I,$$

we obtain the following key relation:

$$\boxed{AB=\ :AB:+:\underbrace{AB}:\,.} \tag{13.39}$$

This formula follows immediately from (13.36) and (13.37). More generally, we define

$$:\underbrace{ABC}:\ :=\mathcal{C}(AB)\cdot C,\qquad :\underbrace{ABC}:\ :=-\mathcal{C}(AC)\cdot B,\qquad :A\underbrace{BC}:\ :=\mathcal{C}(BC)\cdot A.$$

Furthermore, $:\underbrace{AB}\,\underbrace{CD}:\ :=\mathcal{C}(AB)\mathcal{C}(CD)\cdot I$, and

$$:\underbrace{A\underbrace{BC}D}:\ :=-\mathcal{C}(AC)\mathcal{C}(BD)\cdot I,\qquad :\underbrace{A\underbrace{BC}D}:\ :=\mathcal{C}(AD)\mathcal{C}(BC)\cdot I.$$

The general definition goes like this:

- For one pairing, set

$$\boxed{:A_1\cdots\underbrace{A_i\cdots A_j}\cdots A_n:\ :=(-1)^{j-i+1}\mathcal{C}(A_iA_j):A_1\cdots\dot{A}_i\cdots\dot{A}_j\cdots A_n:}$$

  where the dotted factors have to be cancelled.
- The same definition will be used for several pairings where we assume that A_i stands to the left of all the other pairings.

[10] To prove this, note that the anticommutation relations tell us that $[A^-,B^+]_+$ is equal to $\mathcal{A}I$ with a real number $\mathcal{A}$. Hence $\langle 0|\,[A^-,B^+]_+|0\rangle=\langle 0|\mathcal{A}I|0\rangle=\mathcal{A}$. Therefore, $\mathcal{A}$ is equal to

$$\langle 0|A^-B^++B^+A^-|0\rangle=\langle 0|A^-B^+|0\rangle=\langle 0|(A^-+A^+)(B^++B^-)|0\rangle=\mathcal{C}(AB).$$

This allows us to define paired normal products in an inductive way. For example,

$$:\underbrace{ABCD}: \;=-\mathcal{C}(AC):\underbrace{BD}:=-\mathcal{C}(AC)\mathcal{C}(BD)\cdot I,$$

and

$$:\underbrace{ABCD}: \;=\mathcal{C}(AD)\;:\underbrace{BC}:=\mathcal{C}(AD)\mathcal{C}(BC)\cdot I.$$

Observe that the following hold:

The paired normal product remains unchanged (resp. changes sign) under even (resp. odd) permutations of the fermionic factors.

Note that, in the bosonic case, the paired normal product is invariant under all kind of permutations of the bosonic factors.

(d) Time-ordered product: For $t > s$, we define

$$\mathcal{T}(A(t)B(s)) = -\mathcal{T}(B(s)A(t)) := A(t)B(s).$$

In general, let $t_1, ..., t_n$ be n pairwise different time points. By definition, the time-ordered product

$$\mathcal{T}(A_1(t_1)\cdots A_n(t_n))$$

is equal to

$$\eta\cdot A_{i_1}(t_{i_1})\cdots A_{i_n}(t_{i_n})$$

such that $t_{i_1} > t_{i_2} > \ldots > t_{i_n}$, and η is the sign of the permutation from $1\ldots n$ to $i_1\ldots i_n$. As in the bosonic case, we define the time-ordered contraction by

$$\mathcal{C}_{\mathcal{T}}(A(t)B(s)) := \langle 0|\mathcal{T}(A(t)B(s))|0\rangle.$$

(e) Time-ordered paired normal product: This product is defined parallel to the paired normal product by replacing each contraction $\mathcal{C}(A_jA_k)$ by $\mathcal{C}_{\mathcal{T}}(A_jA_k)$. For example, we define

$$:\underbrace{A\,B\,C\,D}: \;=\mathcal{C}_{\mathcal{T}}(AD)\mathcal{C}_{\mathcal{T}}(BC)\cdot I.$$

As another example, consider

$$:\underbrace{A\,B\,C\,D\,E}: \;=-\mathcal{C}_{\mathcal{T}}(AC):\underbrace{B\,D\,E}: \;=\mathcal{C}_{\mathcal{T}}(AC)\mathcal{C}_{\mathcal{T}}(BE)\cdot D.$$

Proof of the first Wick Theorem 13.3 in the fermionic case. We will proceed as in the bosonic case. The only difference is that we replace commutation relations $[A, B]_-$ by anticommutation relations $[A, B]_+$, and that we observe the sign convention for the permutations of the fermionic factors of normal products.
Step 1: For $n = 2$, the statement follows from (13.39).
Step 2: Consider now the case where $n = 3$. We will reduce this to the statement for $n = 2$.

(I) To this end, we start with the decomposition

$$ABC = A^+(BC) + A^-(BC).$$

Moving A^- to the right, it follows from (13.38) that

$$(A^-B)C = [A^-, B]_+C - (BA^-)C = \mathcal{C}(AB)C - (BA^-)C.$$

Analogously, $B(A^-C) = \mathcal{C}(AC)B - B(CA^-)$. Hence

$$A^-(BC) = (BC)A^- + \mathcal{C}(AB)C - \mathcal{C}(AC)B.$$

This implies

$$ABC = A^+(BC) + (BC)A^- + \mathcal{C}(AB)C - \mathcal{C}(AC)B.$$

(II) Using the known case (13.39) for two factors, $n = 2$, we get

$$A^+(BC) = A^+ : BC : + A^+ : \underbrace{BC} :,$$

and $(BC)A^- = : BC : A^- + : \underbrace{BC} : A^-$. Hence

$$\begin{aligned} ABC = (A^+ : BC : + : BC : A^-) + A^+ : \underbrace{BC} : + : \underbrace{BC} : A^- \\ + : \underbrace{ABC} : + : \underbrace{ABC} : . \end{aligned}$$

(III) By definition of the normal operator product, we have

$$A^+ : BC := : A^+BC : \quad \text{and} \quad : BC : A^- = : BCA^- : .$$

Note that the fermionic normal product is invariant under even permutations of the factors. Hence $: BC : A^- = : BCA^- : = : A^-BC : .$ Consequently,

$$A^+ : BC : + : BC : A^- = : (A^+ + A^-)BC : = : ABC : .$$

Finally, this yields

$$ABC = : ABC : + : \underbrace{ABC} : + : \underbrace{ABC} : + : \underbrace{ABC} : .$$

This is precisely the claim for $n = 3$. The general case $n = 4, 5, \ldots$ follows analogously by induction. □

Proof of Wick's Main Theorem 13.2 in the fermionic case. The proofs of both the second Wick Theorem 13.4 and the Main Wick Theorem 13.2 in the fermionic case proceed analogously to the bosonic case above.

The Bosonic-Fermionic Case

We now consider mixed operator products

$$A_1 A_2 \cdots A_n$$

where each factor is either bosonic or fermionic. Such products occur in quantum electrodynamics.

> *The first Wick theorem 13.3 and the second Wick Theorem 13.4 remain valid for operator products of mixed type.*

This follows from the fact that bosonic operators commute with fermionic operators. This implies that:

> *Pairings between bosonic and fermionic operators drop out.*

For example, if A is bosonic and B is fermionic, then $\mathcal{C}(AB) = 0$. In fact,

$$\begin{aligned}\mathcal{C}(AB) &= \langle 0|AB|0\rangle = \langle 0|(A^+ + A^-)(B^+ + B^-)|0\rangle \\ &= \langle 0|(A^+B^+ + A^+B^- + B^+A^- + A^-B^-)|0\rangle = 0.\end{aligned}$$

This follows from $A^-B^+ = B^+A^-$.

Internal normal products. Suppose that AB is a normal operator product. By the normal product principle, we get

$$\mathcal{C}(AB) = \langle 0|AB|0\rangle = 0.$$

This implies that pairings drop out if the factors belong to a normal product. For example, the product $A : BCDE : FG$ is equal to

$$A : BCDE : FG =: A : BCDE : FG : + : \underbrace{A : BCDE : FG : + \ldots}$$

This corresponds to the formula for $A : BCDE : FG$ where the pairings between the operators B, C, D, E drop out. This fact will simplify the Feynman diagrams in quantum electrodynamics.

Invariance properties of products. Note that the following hold true:

- Normal products or paired normal products are invariant under both permutations of bosonic operators and even permutations of fermionic operators.
- Normal products or paired normal products change sign under odd permutations of fermionic operators.
- Time-ordered products or time-ordered paired normal products are invariant under both permutations of bosonic operators and even permutations of fermionic operators.
- Time-ordered products or time-ordered paired normal products change sign under odd permutations of fermionic operators.

13.4 Feynman Propagators

> In terms of physics, Feynman propagators describe the propagation of physical effects in quantum field theory by taking both causality and antiparticles into account.
> In terms of mathematics, Feynman propagators are distinguished fundamental solutions of the wave equation, the Klein–Gordon equation, and the Dirac equation. In the Fourier space, Feynman propagators are inverse differential operators, after regularization.
>
> Folklore

It was discovered by Feynman in the 1940s that the propagation of quantum effects is governed by specific functions called propagators. From the modern point of view, propagators arise in a natural way from the Wick technique in perturbation theory.

13.4.1 Discrete Feynman Propagators for Photons and Electrons

As a typical example, let us compute the Feynman propagators of quantum electrodynamics, namely,

- the Feynman photon propagator $\mathcal{D}_F^{\mu\nu}(x - y)$, and

Table 13.1. Feynman propagators on the lattice

Regularized Feynman photon propagator (m_{ph} virtual photon mass, $\varepsilon > 0$)
$\mathcal{D}^{\mu\nu}_{F,\varepsilon,m_{ph}}(x) := \int_{\mathbb{R}} dp^0 \sum_{\mathbf{p}\in\mathcal{G}(N)} \Delta^3\mathbf{p}\; D^{\mu\nu}(p)\mathrm{e}^{-\mathrm{i}px}$
$D^{\mu\nu}(p) := -\dfrac{\mathrm{i}\eta^{\mu\nu}}{(2\pi)^4} \cdot \dfrac{1}{(p^0)^2 - (\mathcal{E}_{\mathbf{p}} - \varepsilon\mathrm{i})^2}$
Regularized Feynman electron propagator (m_e bare electron mass, $\varepsilon > 0$)
$\mathcal{S}_{F,\varepsilon}(x) := \int_{\mathbb{R}} dp^0 \sum_{\mathbf{p}\in\mathcal{G}(N)} \Delta^3\mathbf{p}\; S(p)\mathrm{e}^{-\mathrm{i}px}$
$S(p) := \dfrac{\mathrm{i}}{(2\pi)^4} \cdot \dfrac{\not{p} + m_e I}{(p^0)^2 - (E_{\mathbf{p}} - \varepsilon\mathrm{i})^2}$
$p := (p^0, \mathbf{p}), \qquad x := (t, \mathbf{x}), \qquad px := p^0 t - \mathbf{px}, \qquad \not{p} := \gamma_\mu p^\mu,$ $\mathcal{E}_{\mathbf{p}} := \sqrt{\mathbf{p}^2 + m_{ph}^2}\,, \qquad E_{\mathbf{p}} := \sqrt{p^2 + m_e^2}$

- the Feynman electron propagator $\mathcal{S}_F^{\alpha\beta}(x-y)$.

Here, $\mu, \nu, \alpha, \beta = 0, 1, 2, 3$. As we are going to show, these propagators are related to the Feynman functions $\mathcal{G}_F$ and $\mathcal{G}_{F,m_e}$ by the crucial formulas

$$\boxed{\mathcal{D}_F^{\mu\nu}(x) = -\eta^{\mu\nu}\mathcal{G}_F(x), \qquad \mathcal{S}_F(x) = (\mathrm{i}\gamma^\mu\partial_\mu + m_e)\mathcal{G}_{F,m_e}(x)}$$

for all space-time points $x \in \mathbb{R}^4$. Explicitly, for positive mass m_e, we define

$$\boxed{\mathcal{G}_{F,m_e}(x) := \sum\nolimits_{\mathbf{p}\in\mathcal{G}(N)} \frac{\Delta^3\mathbf{p}}{2(2\pi)^3 E_{\mathbf{p}}}\,(\theta(t)\mathrm{e}^{-\mathrm{i}px} + \theta(-t)\mathrm{e}^{\mathrm{i}px}).}$$

In this connection, we use the standard notation[11]

$$px := E_{\mathbf{p}}t - \mathbf{px}, \qquad E_{\mathbf{p}} := \sqrt{\mathbf{p}^2 + m_e^2}, \qquad \Delta^3\mathbf{p} := \frac{(2\pi)^3}{L^3}.$$

This refers to the truncated lattice $\mathcal{G}(N)$ in momentum space introduced in Sect. 12.1.1 on page 799. Recall that $E_{\mathbf{p}}$ is the energy of a particle having momentum vector $\mathbf{p}$ and mass m_e, and recall that the cubic cells of the lattice $\mathcal{G}(N)$ have the volume L^3. For vanishing mass, $m = 0$, we define

$$\mathcal{G}_F(x) := \sum\nolimits'_{\mathbf{p}\in\mathcal{G}(N)} \frac{\Delta^3\mathbf{p}}{2(2\pi)^3\omega_{\mathbf{p}}}\,(\theta(t)\mathrm{e}^{-\mathrm{i}px} + \theta(-t)\mathrm{e}^{\mathrm{i}px})$$

[11] Recall that $\theta(t) = 1$ (resp. $=0$) if $t \geq 0$ (resp. $t < 0$).

with the photon energy $\omega_{\mathbf{p}} := |\mathbf{p}|$. The symbol Σ' indicates that we do not sum over $\mathbf{p} = 0$. The reader, who wants to become familiar with the Feynman rules in quantum electrodynamics, may immediately pass to Chap. 14.

The discrete Feynman propagators are finite Fourier series.

The regularized discrete Feynman propagators and the continuum limit of Feynman propagators will be studied in Sects. 13.4.2 and 13.4.3, respectively. The continuum limit of Feynman propagators leads to well-defined mathematical objects, namely, tempered distributions. Applications to interesting physical processes can be found in Chap. 15.

As a preparation for the definition of the photon propagator below, recall the free photon quantum field

$$A^{\mu}_{\text{free}}(\mathbf{x},t) := \sum\nolimits'_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=0}^{3} (a^{-}_{\mathbf{p},s}\, \mathrm{e}^{\mathrm{i}(\mathbf{px}-\omega_{\mathbf{p}}t)} + \\ + a^{+}_{\mathbf{p},s}\, \mathrm{e}^{-\mathrm{i}(\mathbf{px}-\omega_{\mathbf{p}}t)})\, \mathcal{N}_{\mathbf{p}} e^{\mu}_{s}(\mathbf{p}) \tag{13.40}$$

along with the normalization factor $\mathcal{N}_{\mathbf{p}} := 1/\sqrt{2L^3\omega_{\mathbf{p}}}$, and the photon energy $\omega_{\mathbf{p}} = |\mathbf{p}|$. Furthermore, recall the free electron-positron quantum field

$$\psi_{\text{free}}(\mathbf{x},t) := \sum\nolimits_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=\pm\frac{1}{2}} (b^{-}_{\mathbf{p},s} u_{\mathbf{p},s}\, \mathrm{e}^{\mathrm{i}(\mathbf{px}-E_{\mathbf{p}}t)} + \\ + c^{+}_{\mathbf{p},s} v_{\mathbf{p},s}\, \mathrm{e}^{-\mathrm{i}(\mathbf{px}-E_{\mathbf{p}}t)})\, N_{\mathbf{p}} \tag{13.41}$$

with the normalization factor $N_{\mathbf{p}} := 1/\sqrt{2L^3 E_{\mathbf{p}}}$, and the particle energy

$$E_{\mathbf{p}} = \sqrt{m_e^2 + \mathbf{p}^2}.$$

For the adjoint free field, we obtain

$$\overline{\psi}_{\text{free}}(\mathbf{x},t) := \sum\nolimits_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=\pm\frac{1}{2}} (b^{+}_{\mathbf{p},s} \overline{u}_{\mathbf{p},s}\, \mathrm{e}^{-\mathrm{i}(\mathbf{px}-E_{\mathbf{p}}t)} + \\ + c^{-}_{\mathbf{p},s} \overline{v}_{\mathbf{p},s}\, \mathrm{e}^{\mathrm{i}(\mathbf{px}-E_{\mathbf{p}}t)})\, N_{\mathbf{p}}. \tag{13.42}$$

Definition of the discrete Feynman photon propagator. Choose the space-time points $x = (\mathbf{x},t)$ and $y = (\mathbf{y},s)$. Let $\mu,\nu = 0,1,2,3$. The most important quantity for describing photons in quantum electrodynamics is defined by

$$\boxed{\mathcal{D}^{\mu\nu}_{F}(x,y) := \langle 0|\mathcal{T}(A^{\mu}_{\text{free}}(x) A^{\nu}_{\text{free}}(y))|0\rangle}$$

for all times t and s with $t \neq s$. This vacuum expectation value of the time-ordered operator product

$$A^{\mu}_{\text{free}}(x) A^{\nu}_{\text{free}}(y)$$

is called the discrete Feynman photon propagator $\mathcal{D}^{\mu\nu}_{F}(x,y)$ of the free photon field. Synonymously, we also speak of the discrete free 2-point Green function $\mathcal{D}^{\mu\nu}_{F}$ of the photon field.[12] Intuitively,

[12] The relation to classical Green's functions (fundamental solutions of partial differential equations) will be studied below.

- the Feynman propagator $\mathcal{D}_F^{\mu\nu}(x,y)$ measures the correlation between the components
$$A^\mu_{\text{free}}(\mathbf{x},t) \quad \text{and} \quad A^\nu_{\text{free}}(\mathbf{y},s)$$
of the free photon field if A^ν_{free} acts first, that is, $t > s$.
- If $t < s$, then the Feynman propagator $\mathcal{D}_F^{\mu\nu}(x,y)$ measures the correlation between the components
$$A^\nu_{\text{free}}(\mathbf{y},s) \quad \text{and} \quad A^\mu_{\text{free}}(\mathbf{x},t)$$
of the free photon field. Here, A^μ_{free} acts first.
- If $t = s$ and $\mathbf{x} \neq \mathbf{y}$, we define
$$\mathcal{D}_F^{\mu\nu}(\mathbf{x},t;\mathbf{y},t) := 0.$$
This is motivated by the fact that, in the theory of special relativity, physical effects propagate maximally with the speed of light. This implies that there are no correlations between the free photon field at different space points, $\mathbf{x} \neq \mathbf{y}$, at equal time, $t = s$.
- In the singular case where $\mathbf{x} = \mathbf{y}$ and $t = s$, there does not exist a generic definition for the correlations between $A^\mu_{\text{free}}(x)$ and $A^\nu_{\text{free}}(x)$.[13]

Theorem 13.6 *For all space-time points $x, y \in \mathbb{R}^4$ corresponding to different times, $t \neq s$, the discrete Feynman propagator of the free photon field reads as*

$$\mathcal{D}_F^{\mu\nu}(x,y) = -\eta^{\mu\nu}\mathcal{G}_F(x-y), \qquad \mu,\nu = 0,1,2,3.$$

This tells us that the propagator $\mathcal{D}_F^{\mu\nu}(x,y)$ only depends on the difference $x-y$. In what follows, we will write $\mathcal{D}_F^{\mu\nu}(x-y)$ instead of $\mathcal{D}_F^{\mu\nu}(x,y)$. Before proving this theorem, let us discuss the physical interpretation. In terms of physics, the function

$$\theta(t)\mathrm{e}^{-\mathrm{i}px} = \begin{cases} 0 & \text{if} \quad t < 0, \\ \mathrm{e}^{\mathrm{i}\mathbf{p}\mathbf{x}}\mathrm{e}^{-\mathrm{i}\omega_{\mathbf{p}}t} & \text{if} \quad t > 0 \end{cases}$$

describes the creation of one photon at the initial time $t = 0$. This photon has the momentum vector $\mathbf{p}$ and the energy $\omega_{\mathbf{p}} = |\mathbf{p}|$. Furthermore, the function

$$\theta(-t)\mathrm{e}^{\mathrm{i}px} = \begin{cases} \mathrm{e}^{-\mathrm{i}\mathbf{p}\mathbf{x}}\mathrm{e}^{\mathrm{i}\omega_{\mathbf{p}}t} & \text{if} \quad t < 0, \\ 0 & \text{if} \quad t > 0 \end{cases}$$

describes the annihilation of one incoming photon at time $t = 0$. This photon has the momentum vector $-\mathbf{p}$ and the energy $\omega_{\mathbf{p}}$. This tells us that:

[13] For example, we could use one of the following definitions

$$\langle 0|A^\mu(x)A^\nu(x)|0\rangle, \qquad \langle 0|A^\nu(x)A^\mu(x)|0\rangle$$

or the mean value $\frac{1}{2}\langle 0|\ (A^\mu(x)A^\nu(x) + A^\nu(x)A^\mu(x))\ |0\rangle$. This element of arbitrariness is always related to substantial trouble caused in quantum field theory by the operator $\mathcal{T}$ of time-ordering. From the mathematical point of view, the Epstein–Glaser approach to quantum field theory overcomes this difficulty in rigorous terms by using the modern theory of tempered distributions (generalized functions). This will be thoroughly studied in Vol. IV. We also refer to the monograph by G. Scharf, Finite Quantum Electrodynamics: the Causal Approach, Springer, Berlin, 1995.

The discrete Feynman photon propagator $\mathcal{D}_F^{\mu\nu}(x-y)$ describes the superposition of creation and annihilation processes for photons.

Proof. Let us introduce the discrete Wightman function

$$\mathcal{W}^{\mu\nu}(x,y) := \langle 0|A^\mu_{\text{free}}(x)A^\nu_{\text{free}}(y)|0\rangle.$$

Obviously, if $t \neq s$, then

$$\mathcal{D}_F^{\mu\nu}(x,y) = \theta(t-s)\mathcal{W}^{\mu\nu}(x,y) + \theta(s-t)\mathcal{W}^{\nu\mu}(y,x).$$

It remains to show that

$$\boxed{\mathcal{W}^{\mu\nu}(x,y) = -\eta^{\mu\nu}\sum\nolimits'_{\mathbf{p}\in\mathcal{G}(N)} \frac{\Delta^3\mathbf{p}}{2(2\pi)^3\omega_{\mathbf{p}}}\, \mathrm{e}^{-\mathrm{i}p(x-y)}.}$$

In fact, it follows from $a^-_{\mathbf{p},s}|0\rangle = 0$ and the commutation relation

$$a^-_{\mathbf{p},s}a^+_{\mathbf{p},s} = -\eta_{ss}I + a^+_{\mathbf{p},s}a^-_{\mathbf{p},s}$$

that $\langle 0|a^-_{\mathbf{p},s}a^+_{\mathbf{p},s}|0\rangle = -\eta_{ss}$. Furthermore, let us use the completeness relation

$$\sum_{s=0}^{3} \eta_{ss}e^\mu_s(\mathbf{p})e^\nu_s(\mathbf{p}) = \eta^{\mu\nu}$$

for the polarization 4-vectors. By (13.40), $\mathcal{W}^{\mu\nu}$ is equal to

$$-\sum\nolimits'_{\mathbf{p}\in\mathcal{G}(N)} \sum_{s=0}^{3} \eta_{ss}\mathcal{N}^2_{\mathbf{p}}e^\mu_s(\mathbf{p})e^\nu_{\mathbf{p},s}(\mathbf{p})\, \mathrm{e}^{-\mathrm{i}\omega_{\mathbf{p}}(t-s)}\, \mathrm{e}^{\mathrm{i}\mathbf{p}(\mathbf{x}-\mathbf{y})}.$$

Finally, recall that $\mathcal{N}^2_{\mathbf{p}} = 1/2L^3\omega_{\mathbf{p}} = \Delta^3\mathbf{p}/2(2\pi)^3\omega_{\mathbf{p}}$. □

Definition of the discrete Feynman electron propagator. Choose the space-time points $x = (\mathbf{x}, t)$ and $y = (\mathbf{y}, s)$. Let $\alpha, \beta = 0, 1, 2, 3$. The most important quantity for describing electrons and positrons in quantum electrodynamics is defined by

$$\boxed{\mathcal{S}_F^{\alpha\beta}(x,y) := \langle 0|\mathcal{T}(\psi^\alpha_{\text{free}}(x)\overline{\psi}^{\,\beta}_{\text{free}}(y))|0\rangle}$$

for all times t and s with $t \neq s$. This vacuum expectation value of the time-ordered operator product $\psi^\alpha_{\text{free}}(x)\overline{\psi}^{\,\beta}_{\text{free}}(y)$ is called the discrete Feynman electron propagator $\mathcal{S}_F^{\alpha\beta}(x,y)$ of the free electron-positron field. Synonymously, we also speak of the discrete free 2-point Green function $\mathcal{S}_F^{\alpha\beta}$ of the electron-positron field. Interestingly enough, it turns out that $\mathcal{S}_F^{\alpha\beta}$ is closely related to the Feynman function $\mathcal{G}_F$ via the Dirac equation. To discuss this, define the matrix function

$$\boxed{\mathcal{S}_F(x) := (\mathrm{i}\gamma^\mu\partial_\mu + m_e)\mathcal{G}_{F,m_e}(x)}$$

where m_e denotes the electron mass. To simplify the notation, we use the same symbol for the two different functions $(x,y) \mapsto \mathcal{S}_F(x,y)$ and $x \mapsto \mathcal{S}_F(x)$. This is justified by the following result.

Theorem 13.7 *For all space-time points $x, y \in \mathbb{R}^4$ corresponding to different times, $t \neq s$, the discrete Feynman electron propagator reads as*

$$\mathcal{S}_F^{\alpha\beta}(x,y) = \mathcal{S}_F^{\alpha\beta}(x-y), \qquad \alpha,\beta = 0,1,2,3.$$

Proof. The key to the proof are the two matrix relations

$$\sum_{s=\pm\frac{1}{2}} u_{\mathbf{p},s}\overline{u}_{\mathbf{p},s} = \not{p} + m_e, \qquad \sum_{s=\pm\frac{1}{2}} v_{\mathbf{p},s}\overline{v}_{\mathbf{p},s} = \not{p} - m_e$$

for the solutions of the Dirac equation considered in Theorem 12.3 on page 807. Furthermore, we will use the anticommutation relations for creation and annihilation operators of electrons and positrons. For example,

$$b^-_{\mathbf{p},s} b^+_{\mathbf{p},s} = I - b^+_{\mathbf{p},s} b^-_{\mathbf{p},s}$$

along with $b^-_{\mathbf{p},s}|0\rangle = 0$.

(I) Consider first the case where $t > s$. By definition of the time-ordering operator $\mathcal{T}$, we obtain

$$\mathcal{S}_F^{\alpha\beta}(x,y) = \langle 0|\psi^\alpha_{\text{free}}(x)|\overline{\psi}^{\ \beta}_{\text{free}}(y)|0\rangle.$$

It follows from the expression (13.41) for free fields that

$$\mathcal{S}_F^{\alpha\beta}(x,y) = \sum_{\mathbf{p}\in\mathcal{G}(N)} N_{\mathbf{p}}^2 \sum_{s=\pm\frac{1}{2}} u^\alpha_{\mathbf{p},s}\overline{u}^\beta_{\mathbf{p},s} \mathrm{e}^{-\mathrm{i}p(x-y)}.$$

Hence

$$\mathcal{S}_F(x,y) = \sum_{\mathbf{p}\in\mathcal{G}(N)} N_{\mathbf{p}}^2 (\not{p} + m_e)\mathrm{e}^{-\mathrm{i}p(x-y)}.$$

This implies the following representation formula

$$\mathcal{S}_F(x,y) = \sum_{\mathbf{p}\in\mathcal{G}(N)} N_{\mathbf{p}}^2 (\mathrm{i}\gamma^\mu\partial_\mu + m_e)\mathrm{e}^{-\mathrm{i}p(x-y)}$$

where the derivative ∂_μ refers to the variable x. Since $N_{\mathbf{p}}^2 = \Delta^3\mathbf{p}/2(2\pi)^3 E_{\mathbf{p}}$, we finally obtain

$$\mathcal{S}_F(x,y) = (\mathrm{i}\gamma^\mu\partial_\mu + m_e)\mathcal{G}_{F,m_e}(x-y).$$

(II) Now assume that $t < s$. Then

$$\mathcal{S}_F^{\alpha\beta}(x,y) = -\langle 0|\overline{\psi}^\beta_{\text{free}}(y)|\psi^\alpha_{\text{free}}(x)|0\rangle = -\sum_{\mathbf{p}\in\mathcal{G}(N)} N_{\mathbf{p}}^2 \sum_{s=\pm\frac{1}{2}} v^\alpha_{\mathbf{p},s}\overline{v}^\beta_{\mathbf{p},s}\mathrm{e}^{\mathrm{i}p(x-y)}.$$

Therefore, we get

$$\mathcal{S}_F(x,y) = -\sum_{\mathbf{p}\in\mathcal{G}(N)} N_{\mathbf{p}}^2 (\not{p} - m_e)\mathrm{e}^{\mathrm{i}p(x-y)},$$

telling us that $\mathcal{S}_F(x,y) = (\mathrm{i}\gamma^\mu\partial_\mu + m_e)\mathcal{G}_{F,m_e}(x-y)$. □

Definition of the discrete Feynman positron propagator. Let us interchange the role of ψ_{free} and $\overline{\psi}_{\text{free}}$ by defining

$$\mathcal{P}_F^{\alpha\beta}(x,y) := \langle 0|\mathcal{T}(\overline{\psi}^{\ \alpha}_{\text{free}}(x)\psi^\beta_{\text{free}}(y))|0\rangle, \qquad \alpha,\beta = 0,1,2,3.$$

This is called the discrete Feynman positron propagator. Using the definition of the time-ordering operator $\mathcal{T}$ for fermionic fields, we immediately get

$$\boxed{\mathcal{P}_F^{\alpha\beta}(x,y) = -\mathcal{S}_F^{\beta\alpha}(y-x)}$$

for all space-time points $x, y \in \mathbb{R}^4$ with different times $t \neq s$, and all indices $\alpha,\beta = 0,1,2,3$.

13.4.2 Regularized Discrete Propagators

Definition of the regularized discrete photon propagator. For computing scattering processes in quantum electrodynamics, it is crucial to replace the discrete photon propagator $\mathcal{D}_F^{\mu\nu}$ by the regularized version

$$\mathcal{D}^{\mu\nu}_{F,\varepsilon,m_{ph}}(x-y) := -\eta^{\mu\nu}\mathcal{G}_{F,\varepsilon,m_{ph}}(x-y)$$

with the regularized Feynman function

$$\boxed{\mathcal{G}_{F,\varepsilon,m_{ph}}(x) := \sum_{\mathbf{p}\in\mathcal{G}(N)} \frac{\Delta^3\mathbf{p}}{2(2\pi)^3\mathcal{E}_{\mathbf{p}}}\,(\theta(t)\mathrm{e}^{-\mathrm{i}px}\,\mathrm{e}^{-\varepsilon t} + \theta(-t)\mathrm{e}^{\mathrm{i}px}\,\mathrm{e}^{\varepsilon t}).}$$

Here, we use the lattice $\mathcal{G}(N)$ introduced in Sect. 12.1.1 on page 799. Furthermore, we set $px := \mathcal{E}_{\mathbf{p}}t - \mathbf{px}$, and $\mathcal{E}_{\mathbf{p}} := \sqrt{\mathbf{p}^2 + m_{ph}^2}$ with $m_{ph} > 0$. This means that

- we replace the photon energy $\omega_{\mathbf{p}} = \sqrt{\mathbf{p}^2}$ by the regularized photon energy $\mathcal{E}_{\mathbf{p}}$ where m_{ph} represents a small (artificial) photon mass, and
- we introduce the small damping coefficient $\varepsilon > 0$.

Note that the photon energy $\omega_{\mathbf{p}}$ vanishes if the momentum vector of the photon vanishes, $\mathbf{p} = 0$. In contrast to this, the regularized photon energy $\mathcal{E}_{\mathbf{p}}$ never vanishes. Thus, the denominator $\mathcal{E}_{\mathbf{p}}$ appearing in the regularized Feynman function is well defined for all momentum vectors $\mathbf{p}$. In addition, the oscillations $\mathrm{e}^{\mp\mathrm{i}\omega_{\mathbf{p}}t}$ are replaced by damped oscillations which vanish if time goes to infinity. Explicitly,

$$\lim_{t\to\pm\infty} \mathrm{e}^{\mp\mathrm{i}\mathcal{E}_{\mathbf{p}}t}\mathrm{e}^{\mp\varepsilon t} = 0.$$

The damping coefficient ε arises by shifting the photon energy $\mathcal{E}(\mathbf{p})$ to the complex value

$$\mathcal{E}_{\mathbf{p}} - \mathrm{i}\varepsilon, \qquad \varepsilon > 0$$

which lives in the upper complex plane.[14]

Feynman's integral trick for propagators. Let us simplify the expressions for the Feynman propagators by passing over to the 4-dimensional Fourier space.

Theorem 13.8 *The regularized discrete Feynman propagator for photons allows the following integral representation*

$$\mathcal{D}^{\mu\nu}_{F,\varepsilon,m_{ph}}(x) = -\frac{\mathrm{i}\eta^{\mu\nu}}{(2\pi)^4}\int_{\mathbb{R}} dp^0 \sum_{\mathbf{p}\in\mathcal{G}(N)} \Delta^3\mathbf{p}\;\frac{e^{-ipx}}{p^2 - m_{ph}^2 + \varepsilon_{\mathbf{p}}\mathrm{i}}$$

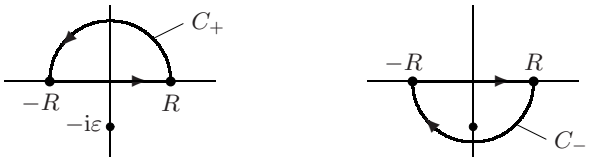
for all space-time points $x = (\mathbf{x}, t)$ with nonzero time, $t \neq 0$, and all parameters $\varepsilon > 0, m_{ph} > 0$.

Here, we set $p^2 := (p^0)^2 - \mathbf{p}^2$, and $px := p^0t - \mathbf{px}$, along with

$$\varepsilon_{\mathbf{p}} := 2\mathcal{E}_{\mathbf{p}}\varepsilon - \mathrm{i}\varepsilon^2.$$

The proof given below will be based on the following key relation:

[14] To simplify computations, physicists frequently use complex energy values.



(a) $t < 0$ (b) $t > 0$

Fig. 13.2. The Feynman trick

$$\boxed{\int_{\mathbb{R}} \frac{\mathrm{e}^{-\mathrm{i}tz}}{z + \varepsilon\mathrm{i}}\, dz = -2\pi\mathrm{i}\theta(t)\mathrm{e}^{-\varepsilon t}} \tag{13.43}$$

which is valid for all parameters $\varepsilon > 0$ and all times $t \neq 0$. In fact, suppose first that $t > 0$. The function

$$\frac{\mathrm{e}^{-\mathrm{i}tz}}{z + \varepsilon\mathrm{i}}, \qquad z \in \mathbb{C}, \quad z \neq -\varepsilon\mathrm{i}$$

is holomorphic on the complex plane except for the simple pole at the point $z = -\mathrm{i}\varepsilon$. Using the closed curve C_- pictured in Fig. 13.2(b) below, Cauchy's residue theorem tells us that

$$\int_{C_-} \frac{\mathrm{e}^{-\mathrm{i}tz}}{z + \varepsilon\mathrm{i}}\, dz = -2\pi\mathrm{i}\mathrm{e}^{-\varepsilon t}.$$

This integral is equal to the sum $A + B$ with

$$A := \int_{-R}^{R} \frac{\mathrm{e}^{-\mathrm{i}tz}}{z + \varepsilon\mathrm{i}}\, dz \quad \text{and} \quad B := \int_{\mathcal{C}_-} \frac{\mathrm{e}^{-\mathrm{i}tz}}{z + \varepsilon\mathrm{i}}\, dz$$

where $\mathcal{C}_-$ denotes the semicircle $\{z = R\mathrm{e}^{\mathrm{i}\varphi} : \ -\pi < \varphi < 0\}$. In order to get the claim (13.43) for $t > 0$, it remains to show that $B \to 0$ as $R \to \infty$. Indeed, noting the Euler formula $\mathrm{e}^{i\varphi} = \cos\varphi + \mathrm{i}\sin\varphi$, we get

$$B = \int_0^{-\pi} \mathrm{e}^{Rt\sin\varphi} \cdot \frac{\mathrm{e}^{-\mathrm{i}Rt\cos\varphi}}{R\mathrm{e}^{\mathrm{i}\varphi} + \varepsilon\mathrm{i}}\, \mathrm{i}\mathrm{e}^{\mathrm{i}\varphi} d\varphi.$$

Using the decomposition $\int_0^{-\pi} = \int_0^{-\delta} + \int_{-\delta}^{-\pi+\delta} + \int_{-\pi+\delta}^{-\pi}$ for sufficiently small $\delta > 0$ and

$$\left| \int_{\delta}^{-\pi+\delta} \mathrm{e}^{Rt\sin\varphi} d\varphi \right| \ \leq \ \pi\mathrm{e}^{-Rt\sin\delta},$$

we obtain $B \to 0$ as $R \to +\infty$. For proving (13.43) in the case where $t < 0$, we use a similar argument based on

$$\int_{C_+} \frac{\mathrm{e}^{-\mathrm{i}tz}}{z + \varepsilon\mathrm{i}}\, dz = 0.$$

Proof. Let $t \neq 0$. In order to prove Theorem 13.8, we start with the decomposition

$$\frac{1}{(p^0)^2 - (\mathcal{E}_{\mathbf{p}} - \varepsilon\mathrm{i})^2} = \frac{1}{2\mathcal{E}_{\mathbf{p}}} \left(\frac{1}{p^0 - \mathcal{E}_{\mathbf{p}} + \varepsilon\mathrm{i}} - \frac{1}{p^0 + \mathcal{E}_{\mathbf{p}} - \varepsilon\mathrm{i}} \right).$$

By the key formula (13.43) along with the substitution $z = p^0 - \mathcal{E}_{\mathbf{p}}$, we get

$$\int_{\mathbb{R}} \frac{\mathrm{e}^{-\mathrm{i}\mathcal{E}_{\mathbf{p}}t}}{p^0 - \mathcal{E}_{\mathbf{p}} + \varepsilon\mathrm{i}}\ dp^0 = -2\pi\mathrm{i}\theta(t)\mathrm{e}^{-\mathrm{i}\mathcal{E}_{\mathbf{p}}t}\ \mathrm{e}^{-\varepsilon t}.$$

Similarly, we obtain

$$-\int_{\mathbb{R}} \frac{\mathrm{e}^{\mathrm{i}\mathcal{E}_{\mathbf{p}}t}}{p^0 + \mathcal{E}_{\mathbf{p}} - \varepsilon\mathrm{i}} = 2\pi\mathrm{i}\theta(-t)\mathrm{e}^{\mathrm{i}\mathcal{E}_{\mathbf{p}}t}\mathrm{e}^{\varepsilon t}.$$

Hence the regularized Feynman function reads as

$$\boxed{\mathcal{G}_{F,\varepsilon,m_{ph}}(x) = \frac{\mathrm{i}}{(2\pi)^4}\int_{\mathbb{R}} dp^0 \sum_{\mathbf{p}\in\mathcal{G}(N)} \Delta^3\mathbf{p}\ \ \frac{\mathrm{e}^{-\mathrm{i}px}}{(p^0)^2 - (\mathcal{E}_{\mathbf{p}} - \varepsilon\mathrm{i})^2}.}$$

Finally, note that $(p^0)^2 - (\mathcal{E}_{\mathbf{p}} - \varepsilon\mathrm{i})^2 = p^2 - m_{ph}^2 + \varepsilon_{\mathbf{p}}\mathrm{i}$. □

Definition of the regularized discrete electron propagator. Let $\varepsilon > 0$. Motivated by Sect. 13.4.1, we define the regularized discrete electron propagator by setting $\mathcal{S}_{F,\varepsilon}(x,y) := \mathcal{S}_{F,\varepsilon}(x-y)$ where

$$\boxed{\mathcal{S}_{F,\varepsilon}(x) := (\mathrm{i}\gamma^\mu\partial_\mu + m_e)\mathcal{G}_{F,\varepsilon,m_e}(x)}$$

for all space-time points x, y with different times $t \neq s$.

Definition of the regularized discrete positron propagator. Similarly, we set

$$\boxed{\mathcal{P}^{\alpha\beta}_{F,\varepsilon}(x,y) := -\mathcal{S}^{\beta\alpha}_{F,\varepsilon}(y-x), \qquad \alpha,\beta = 0,1,2,3.}$$

13.4.3 The Continuum Limit of Feynman Propagators

> Physics does not depend on the choice of the discretization method.
> The golden heuristic rule

The regularized discrete photon propagator $\mathcal{D}^{\mu\nu}_{F,\varepsilon,m_{ph}}$ and the regularized discrete electron propagator $\mathcal{S}^{\mu\nu}_{F,\varepsilon}$ will enter the Feynman rules in Chap. 14.

This way, we will obtain well-defined expressions in each order of perturbation theory.

However, observe that these propagators depend on the choice of the grid $\mathcal{G}(N)$ in the momentum space introduced in Sect. 12.1.1. From the physical point of view, it is quite natural to study the limit

$$\mathcal{G}(N) \to \mathbb{R}^3 \tag{13.44}$$

where the grid $\mathcal{G}(N)$ passes over to the 3-dimensional continuous momentum space $\mathbb{R}^3$. It is our goal to obtain the formulas

$$\mathsf{D}^{\mu\nu}_F = -\eta^{\mu\nu}\mathsf{G}_F,$$

and

$$\mathsf{S}_F = (\mathrm{i}\gamma^\mu\partial_\mu + m_e)\mathsf{G}_{F,m_e}$$

for the Feynman photon propagator distributions $\mathsf{D}_F^{\mu\nu}$ and the Feynman electron propagator distribution S_F. Here, G_F and G_{F,m_e} denote the Feynman propagator distribution for the wave equation and the Klein–Gordon equation, respectively, defined in (13.51) on page 868. All of these distributions are tempered distributions on the space-time manifold $\mathbb{R}^4$, that is, they live in the space $\mathcal{S}'(\mathbb{R}^4)$. Let us discuss the basic ideas.

Step 1 : Low-energy limit $\Delta^3\mathbf{p} \to 0$. Fix the maximal momentum $P_{\max}$ of the grid $\mathcal{G}(N)$, and let the volume of the grid cells go to zero, $L \to 0$. This corresponds to the limit $\Delta^3\mathbf{p} \to 0$. Motivated by Theorem 13.8, the definition of the truncated Feynman propagator $\mathcal{D}^{\mu\nu}_{F,\varepsilon,m_{ph},P_{\max}}(x)$ reads as

$$\boxed{\mathcal{D}^{\mu\nu}_{F,\varepsilon,m_{ph},P_{\max}}(x) := -\frac{\mathrm{i}\eta^{\mu\nu}}{(2\pi)^4}\int_{\mathbb{R}} dp^0 \int_{|\mathbf{p}|\le P_{\max}} d^3\mathbf{p}\ \ \frac{\mathrm{e}^{-\mathrm{i}px}}{p^2 - m_{ph}^2 + \varepsilon_{\mathbf{p}}\mathrm{i}}.}$$

This integral exists for all space-time points $x \in \mathbb{R}^4$.

Step 2 : High-energy limit $P_{\max} \to +\infty$. Formally, we get

$$\mathcal{D}^{\mu\nu}_{F,\varepsilon,m_{ph},\infty}(x) := -\frac{\mathrm{i}\eta^{\mu\nu}}{(2\pi)^4}\int_{\mathbb{R}} dp^0 \int_{\mathbb{R}^3} d^3\mathbf{p}\ \ \frac{\mathrm{e}^{-\mathrm{i}px}}{p^2 - m_{ph}^2 + \varepsilon_{\mathbf{p}}\mathrm{i}}. \tag{13.45}$$

This tells us that the Fourier–Minkowski transform of $\mathcal{D}_{F,\varepsilon,m_{ph},\infty}$ is given by

$$(\mathcal{F}_M \mathcal{D}_{F,\varepsilon,m_{ph},\infty})(p) = -\frac{\mathrm{i}\eta^{\mu\nu}}{(2\pi)^2(p^2 - m_{ph}^2 + \varepsilon_{\mathbf{p}}\mathrm{i})} \tag{13.46}$$

for all $p \in \mathbb{R}^4$. However, this is only a formal argument, since the Fourier integral (13.45) does not exist, in the classical sense. The trouble is that the integrand decays too slowly for large energies, $|\mathbf{p}| \to \infty$.

To overcome this difficulty in a quite natural way, we will use the theory of tempered distributions together with the Fourier–Minkowski transform (see Sects. 11.3 and 14.1 of Vol. I). Let us discuss this. Recall first that, for each function $g \in \mathcal{S}(\mathbb{R}^4)$, the Fourier–Minkowski transform $\mathcal{F}_M g$ is given by

$$\boxed{(\mathcal{F}_M g)(p) := \frac{1}{(2\pi)^2}\int_{\mathbb{R}^4} \mathrm{e}^{\mathrm{i}px} g(x)\, d^4x, \qquad p \in \mathbb{R}^4} \tag{13.47}$$

with $px := p^0 t - \mathbf{px}$. The inverse Fourier–Minkowski transformation reads as

$$g(x) := \frac{1}{(2\pi)^2}\int_{\mathbb{R}^4} \mathrm{e}^{-\mathrm{i}px}(\mathcal{F}_M g)(p)\, d^4p, \qquad x \in \mathbb{R}^4.$$

The map $\mathcal{F}_M : \mathcal{S}(\mathbb{R}^4) \to \mathcal{S}(\mathbb{R}^4)$ is linear and bijective. For each tempered distribution $\mathsf{T} \in \mathcal{S}'(\mathbb{R}^4)$, the Fourier–Minkowski transform is defined by

$$\boxed{(\mathcal{F}_M \mathsf{T})(\varphi) = \mathsf{T}(\mathcal{F}_M \varphi)}$$

for all test functions $\varphi \in \mathcal{S}(\mathbb{R}^4)$. The induced map $\mathcal{F}_M : \mathcal{S}'(\mathbb{R}^4) \to \mathcal{S}'(\mathbb{R}^4)$ is linear and bijective. The function from (13.46) is locally integrable. Therefore, it represents a tempered distribution T given by

$$\mathsf{T}(\varphi) := -\frac{\mathrm{i}\eta^{\mu\nu}}{(2\pi)^2}\int_{\mathbb{R}^4} \frac{\varphi(p)}{p^2 - m_{ph}^2 + \varepsilon_{\mathbf{p}}\mathrm{i}}\ d^4p$$

for all test functions $\varphi \in \mathcal{S}(\mathbb{R}^4)$. Using the inverse Fourier–Minkowski transformation for tempered distributions, we define

$$\boxed{\mathsf{D}^{\mu\nu}_{F,\varepsilon,m_{ph},\infty} := \mathcal{F}_M^{-1}(\mathsf{T})} \tag{13.48}$$

for all test functions $\varphi \in \mathcal{S}(\mathbb{R}^4)$. The tempered distribution $\mathsf{D}^{\mu\nu}_{F,\varepsilon,m_{ph},\infty}$ is called the regularized Feynman propagator distribution for photons. It follows from (13.48) that

$$\mathsf{D}^{\mu\nu}_{F,\varepsilon,m_{ph},\infty}(\varphi) = -\frac{\mathrm{i}\eta^{\mu\nu}}{(2\pi)^4}\int_{\mathbb{R}^4}\left(\int_{\mathbb{R}^4}\frac{\mathrm{e}^{-\mathrm{i}px}}{p^2 - m_{ph}^2 + \varepsilon_{\mathbf{p}}\mathrm{i}}\; d^4p\right)\varphi(x)d^4x$$

for all test functions $\varphi \in \mathcal{S}(\mathbb{R}^4)$, and all indices $\mu, \nu = 0, 1, 2, 3$.

Proposition 13.9 *In the sense of tempered distributions on $\mathbb{R}^4$, we have the limit relation*

$$\lim_{(\varepsilon,m_{ph})\to+(0,0)} \Box\, \mathsf{D}^{\mu\nu}_{F,\varepsilon,m_{ph},\infty} = \mathrm{i}\eta^{\mu\nu}\delta,$$

where δ denotes the Dirac delta distribution.

The limit refers to sequences of pairs (ε, m_{ph}) with $\varepsilon > 0$ and $m_{ph} > 0$. We will show below that the much stronger result holds: it is possible to define the Feynman propagator distribution for photons, $\mathsf{D}^{\mu\nu}_F$, in such a way that

$$\boxed{\Box\, \mathsf{D}^{\mu\nu}_F = \mathrm{i}\eta^{\mu\nu}\delta,}$$

and the definition of $\mathsf{D}^{\mu\nu}_F$ is independent of the choice of the regularization method.
Proof of Proposition 13.9. By definition, for a sequence (T_n), of tempered distributions T_n, the convergence

$$\lim_{n\to\infty} \mathsf{T}_n = \mathsf{T}$$

is equivalent to $\lim_{n\to\infty} \mathsf{T}_n(\varphi) = \mathsf{T}(\varphi)$ for all test functions $\varphi \in \mathcal{S}(\mathbb{R}^4)$. Since the Fourier–Minkowski transform respects the convergence of tempered distributions, we have to show that

$$\lim_{(\varepsilon,m_{ph})\to+(0,0)} \mathcal{F}_M(\Box\, \mathsf{D}_{F,\varepsilon,m_{ph},\infty}) = \mathrm{i}\eta^{\mu\nu}\mathcal{F}_M(\delta).$$

Note that, in the Fourier space, the wave operator $\Box$ corresponds to the multiplication with the factor $-p^2 = -(p^0)^2 + \mathbf{p}^2$, and δ corresponds to the function $1/(2\pi)^2$. Therefore, it remains to show that

$$\lim_{(\varepsilon,m_{ph})\to+(0,0)} \int_{\mathbb{R}^4} \frac{p^2\varphi(p)d^4p}{p^2 - m_{ph}^2 + \varepsilon_{\mathbf{p}}\mathrm{i}} = \int_{\mathbb{R}^4} \varphi(p)d^4p.$$

Obviously, this is true. In this connection recall that $\mathcal{E}_{\mathbf{p}}^2 := \mathbf{p}^2 + m_{ph}^2$, and observe that the denominator

$$(p^0)^2 - (\mathcal{E}_{\mathbf{p}} - \varepsilon\mathrm{i})^2 = p^2 - m_{ph}^2 + \varepsilon_{\mathbf{p}}\mathrm{i}$$

is different from zero if $\varepsilon > 0$ and $m_{ph} > 0$. □

The importance of the Feynman propagator as an inverse differential operator. For quantum field theory, it is crucial that the Feynman propagator represents an inverse differential operator, in some sense. This is responsible for the surprising success of the Feynman functional integral method in quantum field theory, as we will show in Vol. IV. The classical background is the Gaussian integral formula

$$\boxed{\int_{\mathbb{R}^n} \mathrm{e}^{-\frac{1}{2}\langle x|Ax\rangle} \mathrm{e}^{\langle J|x\rangle} \, \frac{d^n x}{(2\pi)^{n/2}} = \mathrm{e}^{\frac{1}{2}\langle J|A^{-1}J\rangle} \, (\det A^{-1})^{1/2}} \tag{13.49}$$

where A denotes a real nonsingular symmetric $(n \times n)$-matrix, and $J \in \mathbb{R}^n$. Explicitly,

$$\langle x|Ax\rangle = \sum_{j,k=1}^{n} a_{jk} x^j x^k, \qquad \langle J|x\rangle = \sum_{k=1}^{n} J^k x^k.$$

Note that the value of the integral (13.49) depends essentially on the inverse operator A^{-1}. The Feynman functional integral can be viewed as an infinite-dimensional version of (13.49) where

- $\langle x|Ax\rangle$ corresponds to the action integral of the classical field, and
- the differential operator A corresponds to the Euler–Lagrange equation of the classical field.

Consider first a simple finite-dimensional model. For the diagonal matrix

$$A = \begin{pmatrix} p_1 & 0 \\ 0 & p_2 \end{pmatrix}$$

with nonzero numbers p_1, p_2, the inverse matrix is given by

$$A^{-1} = \begin{pmatrix} \frac{1}{p_1} & 0 \\ 0 & \frac{1}{p_2} \end{pmatrix}.$$

In the singular case where $p_1 = 0$ and , say, $p_2 \neq 0$, we can regularize the situation by replacing $p_1 = 0$ with the small positive number ε.

In the infinite-dimensional case of differential operators, this method can be generalized by using the Fourier–Minkowski transform. Let us discuss this.

Step 1 : Formal Fourier–Minkowski transformation. Consider the wave equation

$$\Box \, \mathcal{G}_F = -\mathrm{i}\delta$$

on the 4-dimensional space-time manifold. By Fourier–Minkowski transformation, we obtain the relation $\mathcal{F}_M(\Box \, \mathcal{G}_F) = -\mathrm{i}\mathcal{F}_M(\delta)$. Hence

$$((p^0)^2 - \mathbf{p}^2)\mathcal{F}_M(\mathcal{G}_F) = \frac{\mathrm{i}}{(2\pi)^2}.$$

This yields

$$\mathcal{F}_M(\mathcal{G}_F)(p) = \frac{\mathrm{i}}{(2\pi)^2 p^2}.$$

Formally, in the Fourier space the differential operator $\Box$ corresponds to the multiplication by $-p^2$, and the Feynman function $\mathcal{G}_F$ corresponds to the inverse operation. Hence

$$\mathcal{G}_F(x) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^4} \frac{\mathrm{i}}{(2\pi)^2 p^2}\, \mathrm{e}^{-\mathrm{i}px}\, d^4p. \tag{13.50}$$

Symbolically, let us write

$$\boxed{\mathcal{G}_F = (\Box)^{-1}(-\mathrm{i}\delta).}$$

However, there arises the following serious problem. The equation $p^2 = 0$, that is,

$$(p^0)^2 - \mathbf{p}^2 = 0$$

has the real solutions $p^0 = \pm|\mathbf{p}|$ which correspond to the photon energy. Therefore, the Fourier–Minkowski transform $\mathcal{F}_M(\mathcal{G}_F)$ has singularities. In addition, the Fourier integral (13.50) does not exist in the classical sense, since the integrand decays too slowly at infinity. This is related to the fact that, for given smooth function f, the classical wave equation

$$\Box\, \mathcal{G} = f$$

has an infinite number of solutions which depend on both the initial values of $\mathcal{G}$ and the partial time derivative $\mathcal{G}_t$ at the initial time $t = 0$.

Step 2 : Regularization. As discussed above, to overcome the trouble we replace the Fourier–Minkowski transform $\mathcal{F}_M(\mathcal{G}_F)$ by the regularized Fourier–Minkowski transform

$$\frac{\mathrm{i}}{(2\pi)^2} \cdot \frac{1}{(p^0)^2 - (\mathcal{E}_{\mathbf{p}} - \varepsilon\mathrm{i})^2}$$

with $\mathcal{E}_{\mathbf{p}} := \sqrt{\mathbf{p}^2 + m_{ph}^2}$ and $\varepsilon > 0, m_{ph} > 0$. In addition, the inverse Fourier–Minkowski transform is to be understood in the sense of tempered distributions.

Observe that there exist many possibilities for choosing the regularization of the Fourier–Minkowski transform. There arises the question whether it is possible to define the Feynman propagator in such a way that it is independent of the special choice of the regularization procedure. This is possible, as we will show in Vol. IV. At this point, let us only sketch the basic ideas.

The Feynman propagator distributions. We define the following tempered distributions:

(i) The Feynman propagator distribution for the wave equation:

$$\mathsf{G}_F(x) := \frac{\mathrm{i}}{(2\pi)^2} \int_{\mathbb{R}^4} \frac{1}{(2\pi)^2(p^2 + 0_+\mathrm{i})}\, \mathrm{e}^{-\mathrm{i}px}\, d^4p. \tag{13.51}$$

(ii) The Feynman propagator distribution for the Klein-Gordon equation with mass $m > 0$:

$$\mathsf{G}_{F,m}(x) := \frac{\mathrm{i}}{(2\pi)^2} \int_{\mathbb{R}^4} \frac{1}{(2\pi)^2(p^2 - m^2 + 0_+\mathrm{i})}\, \mathrm{e}^{-\mathrm{i}px}\, d^4p. \tag{13.52}$$

(ii) The Feynman propagator distribution for photons: For $\mu, \nu = 0, 1, 2, 3$, we set

$$\boxed{\mathsf{D}_F^{\mu\nu} := -\eta^{\mu\nu}\mathsf{G}_F.} \tag{13.53}$$

(iii) The Feynman propagator distribution for electrons: We set

$$\boxed{\mathsf{S}_F := (\mathrm{i}\gamma^\mu \partial_\mu + m_e)\mathsf{G}_{F,m_e}.} \tag{13.54}$$

Explicitly, we have

$$\mathsf{S}_F = \frac{\mathrm{i}}{(2\pi)^2} \int_{\mathbb{R}^4} \frac{\gamma^\mu p_\mu + m_e I}{(2\pi)^2(p^2 - m_e^2 + 0_+\mathrm{i})}\ \mathrm{e}^{-\mathrm{i}px}\, d^4p.$$

These mnemonical formulas are to be understood in the following sense. The tempered distribution G_F is the inverse Fourier–Minkowski transform of the tempered distribution

$$\boxed{\frac{1}{(2\pi)^2(p^2 + 0_+\mathrm{i})}.} \tag{13.55}$$

This distribution is defined by combining regularization, analytic continuation, and limits of tempered distributions. More generally, our goal is to construct tempered distributions denoted by

$$\boxed{\mathsf{T}_\lambda = \frac{1}{(p^2 - m^2 + 0_+\mathrm{i})^\lambda}} \tag{13.56}$$

for each mass parameter $m \geq 0$ and each complex exponent λ different from $-2, -3, \ldots$ To this end, fix the regularization parameter $\varepsilon > 0$, and the positive exponent $\lambda > 0$. For each complex point z in the open upper half-plane, $\Im z > 0$, we choose the argument $0 < \arg z < \pi$, and we set

$$z^\lambda := \mathrm{e}^{\lambda(\ln|z| + \mathrm{i}\arg z)}.$$

In this sense, for each test function $\varphi \in \mathcal{S}(\mathbb{R}^4)$, the integral

$$\mathsf{T}_{\lambda,\varepsilon}(\varphi) := \int_{\mathbb{R}^4} (p^2 + \varepsilon\mathrm{i}\sum_{j=1}^4 (p^j)^2)^\lambda \varphi(p) d^4p$$

is well-defined. The function

$$\lambda \mapsto \mathsf{T}_{\lambda,\varepsilon}(\varphi)$$

can be analytically continued to a function which is holomorphic on the punctured complex plane $\mathbb{C} \setminus \{-2, -3, -4, ...\}$. For these λ-values, the limit

$$\mathsf{T}_\lambda(\varphi) := \lim_{\varepsilon \to +0} \mathsf{T}_{\varepsilon,\lambda}(\varphi)$$

exists for all $\varphi \in \mathcal{S}(\mathbb{R}^4)$. This way, we obtain the desired tempered distributions from (13.56). Let us summarize the Fourier–Minkowski transform of propagators:

(i) Feynman propagator of the wave equation:

$$(\mathcal{F}_M \mathsf{G}_F)(p) = \frac{\mathrm{i}}{(2\pi)^2(p^2 + 0_+\mathrm{i})}.$$

(ii) Feynman propagator of the Klein–Gordon equation with mass $m > 0$:

$$(\mathcal{F}_M \mathsf{G}_{F,m})(p) = \frac{\mathrm{i}}{(2\pi)^2(p^2 - m^2 + 0_+\mathrm{i})}.$$

(iii) Feynman propagator of photons:

$$(\mathcal{F}_M \mathsf{D}_F^{\mu\nu})(p) = -\frac{\mathrm{i}\eta^{\mu\nu}}{(2\pi)^2(p^2 + 0_+\mathrm{i})}, \qquad \mu, \nu = 0, 1, 2, 3.$$

(iv) Feynman propagator of electrons:

$$(\mathcal{F}_M \mathsf{S}_F)(p) = \frac{\mathrm{i}(\gamma^\mu p_\mu + m_e I)}{(2\pi)^2(p^2 - m_e^2 + 0_+\mathrm{i})}.$$

Using the inverse Fourier–Minkowski transform, it can be shown that the Feynman propagator distribution for photons has the following form:

$$\boxed{\mathsf{D}_F^{\mu\nu} = -\frac{\eta^{\mu\nu}}{4\pi^2} \cdot \frac{1}{x^2 + 0_+\mathrm{i}}, \qquad \mu, \nu = 0, 1, 2, 3.}$$

This is a tempered distribution on the space-time $\mathbb{R}^4$ which is defined as (13.56). Here, we have to replace $p^2 = (p^0)^2 - \mathbf{p}^2$ by $x^2 = t^2 - \mathbf{x}^2$.

Fundamental solutions. For a linear differential operator L with constant coefficients, consider the equation

$$LU = \delta$$

in the sense of tempered distributions. Each solution U is called a tempered fundamental solution of L. Interestingly enough, the Feynman propagators for photons and electrons are tempered fundamental solutions of the wave equation and the Dirac equation, up to normalization factors. Explicitly, we get the following:

(i) The Feynman propagator distribution for the wave equation: $\Box\, \mathsf{G}_F = -\mathrm{i}\delta$.
(ii) The Feynman propagator distribution for the Klein-Gordon equation with mass $m > 0$: $(\Box + m^2)\, \mathsf{G}_{F,m} = -\mathrm{i}\delta$.
(iii) The Feynman propagator distribution for photons: $\Box\, \mathsf{D}_F^{\mu\nu} = \mathrm{i}\eta^{\mu\nu}\delta$. Here, the indices μ, ν run from 0 to 3.
(iv) The Feynman propagator distribution for electrons:

$$(\mathrm{i}\gamma^\mu \partial_\mu - m_e)\mathsf{S}_F = \mathrm{i}I\delta.$$

13.4.4 Classical Wave Propagation versus Feynman Propagator

> Propagation of quantum effects differs substantially from classical physics.
> Folklore

The Feynman propagator distribution G_F for the wave equation allows the following representation:

$$\boxed{\mathsf{G}_F = \mathsf{G}_{\mathrm{ret}} - \mathsf{G}^- = \mathsf{G}_{\mathrm{adv}} + \mathsf{G}^+.}$$

This shows that the Feynman propagator distribution differs from both the classical retarded propagator distribution $\mathsf{G}_{\mathrm{ret}}$ and the classical advanced propagator distribution $\mathsf{G}_{\mathrm{adv}}$ of the wave equation. This tells us that

Quantum wave propagation governed by the Feynman propagator distribution differs from classical wave propagation.

Let us discuss this. The tempered distributions $\mathsf{G}_{\rm ret}, \mathsf{G}_{\rm adv}, \mathsf{G}^{\pm}$ can be characterized by their Fourier–Minkowski transforms:

$$\mathcal{F}_M \mathsf{G}_{\rm ret} = \frac{1}{4\pi^2} \lim_{\varepsilon \to +0} \frac{\mathrm{i}}{(p^0 + \mathrm{i}\varepsilon)^2 - \mathbf{p}^2},$$
$$\mathcal{F}_M \mathsf{G}_{\rm adv} = \frac{1}{4\pi^2} \lim_{\varepsilon \to +0} \frac{\mathrm{i}}{(p^0 - \mathrm{i}\varepsilon)^2 - \mathbf{p}^2}$$

and

$$\boxed{\mathcal{F}_M \mathsf{G}^{\pm} = \pm \frac{\theta(\pm p^0)}{2\pi} \ \delta(p^2).} \tag{13.57}$$

These symbols are to be understood in the following way. The limit $\varepsilon \to +0$ refers to the convergence of tempered distributions. Explicitly, for all test functions ψ in the function space $\mathcal{S}(\mathbb{R}^4)$, we get

$$\mathcal{F}_M \mathsf{G}_{\rm ret}(\psi) = \frac{1}{4\pi^2} \lim_{\varepsilon \to +0} \int_{\mathbb{R}^4} \frac{\mathrm{i}\psi(p)}{(p^0 + \mathrm{i}\varepsilon)^2 - \mathbf{p}^2} \ d^4p, \tag{13.58}$$

and

$$\mathcal{F}_M \mathsf{G}^{+}(\psi) := \int_0^{\infty} M_{p^0}(\psi) p^0 dp^0,$$
$$\mathcal{F}_M \mathsf{G}^{-}(\psi) := \int_{-\infty}^{0} M_{-p^0}(\psi)(-p^0) dp^0$$

along with the spherical mean value

$$M_r(\psi) := \int_{|\mathbf{p}|=r} \psi(\mathbf{p}, p^0) dS_r.$$

Here, dS_r refers to the surface measure on the sphere of radius r in the $\mathbf{p}$-space. Let us discuss in which sense the symbol (13.57) (taken from the language of physicists) implies mnemonically the formula (13.58). To begin with, let $a > 0$. Recall the formula

$$\delta(a^2 - x^2) = \frac{\delta(x-a)}{2a} + \frac{\delta(x+a)}{2a}, \qquad x \in \mathbb{R}$$

(see Sect. 11.2.4 of Vol. I). Hence

$$\delta(p^2) = \delta((p^0)^2 - |\mathbf{p}|^2) = \frac{p^0 - |\mathbf{p}|}{2p^0} + \frac{p^0 + |\mathbf{p}|}{2p^0}.$$

To compute the integral

$$\mathsf{G}_{\rm ret}(\psi) = \int_{\mathbb{R}^4} \frac{\theta(p^0)}{2\pi} \delta(p^2) \psi(p) \ d^4p,$$

we use spherical coordinates, that is, $d^4p = dp^0(r^2 dr d\Omega)$ with $r := |\mathbf{p}|$ and the solid angle Ω. Then, $\mathsf{G}_{\rm ret}(\psi)$ is equal to

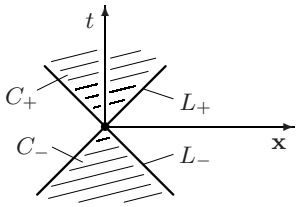


Fig. 13.3. Light cone

$$\int_0^\infty \frac{dp^0}{4\pi p^0} \int_{|\mathbf{p}|<\infty} \delta(p^0 - |\mathbf{p}|)\psi(\mathbf{p}, p^0)|\mathbf{p}|^2 d|\mathbf{p}| d\Omega$$
$$= \int_0^\infty \frac{dp^0}{4\pi p^0} \int_{|\mathbf{p}|=p^0} \psi(\mathbf{p}, p^0)(p^0)^2 d\Omega = \int_0^\infty M_{p^0}(\psi) p^0 dp^0.$$

For the following, let us change the notation. Replace the momentum 4-vector $p = (\mathbf{p}, p^0)$ by the space-time point $x = (\mathbf{x}, t)$. In terms of mathematics, the symbol $\frac{1}{2\pi}\delta(x^2)$ is a tempered distribution on $\mathbb{R}^4$ which corresponds to a measure μ on the light cone

$$L = L_+ \cup L_-$$

with the forward light cone $L_+ := \{(\mathbf{x}, t) \in \mathbb{R}^4 : \ |\mathbf{x}| = t, \ \ t \geq 0\}$ (Fig. 13.3) and the backward light cone

$$L_- := \{(\mathbf{x}, t) \in \mathbb{R}^4 : \ |\mathbf{x}| = -t, \ \ t \leq 0\}.$$

Explicitly,

$$\int_L \psi d\mu := \int_0^\infty M_t(\psi) t dt + \int_{-\infty}^0 M_{-t}(\psi)(-t) dt.$$

Instead of $\int_L \psi d\mu$ we also write $\int_{\mathbb{R}^4} \frac{1}{2\pi}\delta(x^2) d^4x$. Mnemonically, physicists use the following formulas for the Fourier–Minkowski transform considered above:

$$\mathsf{G}_{\mathrm{ret}}(x) = \frac{\mathrm{i}}{(2\pi)^4} \int_{\mathbb{R}^4} \frac{\mathrm{e}^{-\mathrm{i}px}}{(p^0 + 0_+\mathrm{i})^2 - \mathbf{p}^2} \ d^4p,$$
$$\mathsf{G}_{\mathrm{adv}}(x) = \frac{\mathrm{i}}{(2\pi)^4} \int_{\mathbb{R}^4} \frac{\mathrm{e}^{-\mathrm{i}px}}{(p^0 - 0_+\mathrm{i})^2 - \mathbf{p}^2} \ d^4p,$$

and

$$\mathsf{G}^{\pm}(x) = \pm \frac{1}{(2\pi)^3} \int_{\mathbb{R}^4} \theta(\pm p^0)\delta(p^2)\mathrm{e}^{-\mathrm{i}px} \ d^4p.$$

Let us now explain the meaning of the tempered distributions $\mathsf{G}_{\mathrm{ret}}, \mathsf{G}_{\mathrm{adv}}, \mathsf{G}^{\pm}$ in terms of fundamental solutions of the wave equation.

(i) The retarded propagator distribution: There exists precisely one tempered distribution $\mathsf{G}_{\mathrm{ret}} \in \mathcal{S}'(\mathbb{R}^4)$ such that the wave equation

$$\Box \, \mathsf{G}_{\mathrm{ret}} = -\mathrm{i}\delta \qquad \text{on } \mathbb{R}^4$$

is satisfied, and the support of $\mathsf{G}_{\mathrm{ret}}$ is contained in the forward light cone L_+. Explicitly,

$$\mathsf{G}_{\mathrm{ret}} = \frac{\theta(t)}{2\pi\mathrm{i}} \ \delta(x^2).$$

This means that

$$\mathsf{G}_{\mathrm{ret}}(\psi) = \frac{1}{\mathrm{i}} \int_0^\infty M_t(\psi) t dt, \qquad \psi \in \mathcal{S}(\mathbb{R}).$$

(ii) The advanced propagator distribution: By the time reflection $t \mapsto -t$, the retarded propagator distribution, $\mathsf{G}_{\mathrm{ret}}$, is transformed into the advanced propagator distribution, $\mathsf{G}_{\mathrm{adv}} \in \mathcal{S}'(\mathbb{R}^4)$. This tempered distribution is uniquely determined by the fact that the wave equation

$$\Box\, \mathsf{G}_{\mathrm{adv}} = -\mathrm{i}\delta \qquad \text{on } \mathbb{R}^4$$

is satisfied, and the support of $\mathsf{G}_{\mathrm{adv}}$ is contained in the backward light cone L_-. Explicitly,

$$\mathsf{G}_{\mathrm{adv}} = \frac{\theta(-t)}{2\pi\mathrm{i}}\, \delta(x^2).$$

This means that

$$\mathsf{G}_{\mathrm{adv}}(\psi) = \frac{1}{\mathrm{i}} \int_{-\infty}^0 M_{-t}(\psi)(-t) dt, \qquad \psi \in \mathcal{S}(\mathbb{R}).$$

(iii) The Jordan–Pauli distribution: The tempered distribution

$$\boxed{\mathsf{G} := \mathsf{G}_{\mathrm{ret}} - \mathsf{G}_{\mathrm{adv}}} \tag{13.59}$$

is called the Jordan–Pauli distribution. Explicitly, we obtain

$$\mathsf{G} = \frac{\operatorname{sgn} t}{2\pi\mathrm{i}}\, \delta(x^2).$$

Using the tempered distributions G^+ and G^-, it turns out that

$$\boxed{\mathsf{G} = \mathsf{G}^+ + \mathsf{G}^-.} \tag{13.60}$$

Furthermore, $\Box \mathsf{G}^\pm = \Box \mathsf{G} = 0$ on $\mathbb{R}^4$. The two decompositions (13.59) and (13.60) of the Jordan–Pauli function G are dual to each other in terms of the Fourier–Minkowski transform. This is to be understood in the sense of the formulas

$$\mathsf{G}_{\mathrm{ret}} = \frac{\theta(t)}{2\pi\mathrm{i}}\, \delta(x^2), \qquad \mathsf{G}_{\mathrm{adv}} = \frac{\theta(-t)}{2\pi\mathrm{i}}\, \delta(x^2)$$

and the following formulas for the Fourier–Minkowski transform

$$\mathcal{F}_M \mathsf{G}^\pm = \pm \frac{\theta(\pm p^0)}{2\pi}\, \delta(p^2).$$

We are speaking of propagator duality. The decomposition

$$\mathsf{G}_F = \mathsf{G}_{\mathrm{ret}} - \mathsf{G}^- = \mathsf{G}_{\mathrm{adv}} + \mathsf{G}^+$$

shows that the Feynman propagator G_F is obtained by mixing of propagator duality.

Perspectives. The Feynman propagators play a fundamental role in the approaches to quantum field theory by using

- the Dyson S-matrix operator,

- the method of moments (Green's functions and the LSZ reduction formalism), and
- the Feynman functional integral.

These three apparently different methods are indeed equivalent. This will be thoroughly studied in Vol. IV on quantum mathematics.

Historical remarks The role of highly singular Green functions for linear hyperbolic partial differential equations was systematically studied by Hadamard (regularization of divergent integrals) and Marcel Riesz (method of analytic continuation). For this, we recommend:

J. Hadamard, The Initial-Value Problem for Linear Hyperbolic Partial Differential Equations, Hermann, Paris, 1932 (in French).

M. Riesz, The Riemann–Liouville integral and the initial-value problem, Acta Math. **81** (1948), 1–223 (in French).

Hints for further reading. The modern formulation of Hadamard's theory in terms of tempered distributions and its applications can be found in the following monographs:

I. Gelfand and G. Shilov, Generalized Functions, Vol. I, Chap. 3, Academic Press, New York, 1964.

Yu. Egorov, A. Komech, and M. Shubin, Elements of the Modern Theory of Partial Differential Equations, Springer, New York, 1999.

P. Günther, Huygens' Principle and Hyperbolic Differential Equations, Academic Press, San Diego, 1988.

C. Bär, N. Ginoux, and F. Pfäffle, Wave Equations on Lorentzian Manifolds and Quantization, European Mathematical Society, 2007.

The latter monograph emphasizes the relations between the classical theory and modern applications to quantum field theory. We also recommend:

P. Günther, Huygens' principle and Hadamard's conjecture, Math. Intelligencer **13** (1991), 56–63.

R. Radzikowski, Micro-local approach to the Hadamard condition in quantum field theory on curved space-time. Commun. Math. Phys. **179** (1996), 529–553.

The explicit relations between Green's functions (Feynman propagators) and quantum field theory in terms of tempered distributions are studied in:

N. Bogoliubov and D. Shirkov, Introduction to Quantum Field Theory, Interscience, New York, 1980.

N. Bogoliubov and D. Shirkov, Quantum Fields. Lectures given at the Moscow Lomonosov University, Benjamin, Reading, Massachusetts, 1983.

G. Scharf, Finite Quantum Electrodynamics: the Causal Approach, Springer, Berlin, 1995.

See also Gelfand and Shilov (1964), Vol. 1, above, and A. Komech, Linear Partial Differential Equations with Constant Coefficients, pp. 121–256. In: Egorov, Komech, and Shubin (1999) above. Much material can be found in:

N. Ortner and P. Wagner, A survey on explicit representation formulas for fundamental solutions of partial differential operators, Acta Applicandae Mathematicae **47** (1997), 101–124.

N. Ortner and P. Wagner, Distribution-Valued Analytic Functions: Theory and Applications, Lecture Notes 37/2008, Max Planck Institute for Mathematics in the Sciences, Leipzig, 2008.
Internet: `http://www.mis.de/preprints`

14. The Beauty of Feynman Diagrams in QED

> Whoever understands Feynman diagrams can understand everything in quantum field theory.
>
> Folklore

In elementary particle physics, physicists use the highly intuitive language of Feynman diagrams as a universal tool. It is crucial to know that the geometric Feynman diagrams come from well-defined analytic expressions generated by applying the Wick theorem to the Dyson series.

> *The Feynman rules translate the Feynman diagrams into the corresponding analytic expressions and vice versa.*

The following section is crucial for understanding the language of physicists used in elementary particle physics. The simple basic idea behind Feynman diagrams can be found in Sect. 13.2. We are going now to consider the general case.

> *For the convenience of the reader, we will only use expressions which are well-defined in terms of mathematics.*

However, the notation is chosen in such a way that our language is very close to the language used by physicists. For mnemonical reasons, physicists do not hesitate to use ill-defined formulas like divergent integrals and the square, $\delta(x)^2$, of Dirac's delta function. In Sect. 12.6.1 of Vol. I, we have shown that it is impossible to define a reasonable product for all distributions (the counterexample of Laurent Schwartz). However, based on the principle of causality and the notion of the wave front set, there exists a subclass of distributions for which the product is well defined. This is the starting point of the Epstein–Glaser approach to quantum electrodynamics. The computation of scattering processes on the time interval $[-\frac{T}{2}, \frac{T}{2}]$ is governed by the transition probabilities

$$\boxed{\mathcal{W}(T) := |\langle \Phi_{\mathrm{out}} | \mathsf{S}(T) \Phi_{\mathrm{in}} \rangle|^2.} \tag{14.1}$$

To this end, we have to compute the transition amplitude

$$\langle \Phi_{\mathrm{out}} | \mathsf{S}(T) \Phi_{\mathrm{in}} \rangle.$$

Recall from Sect. 13.1 that, in Nth order approximation, Dyson's S-matrix

$$\boxed{\mathsf{S}(T) = I + \sum_{n=1}^{N} \mathsf{S}_n(T)}$$

is given by $\mathsf{S}_1(T) := \int_{-T/2}^{T/2} \mathsf{V}(t) dt$ and

$$\mathsf{S}_n(T) := \frac{1}{n!}\int_{-T/2}^{T/2}\cdots\int_{-T/2}^{T/2}\mathcal{T}(\mathsf{V}(t_1)\cdots\mathsf{V}(t_n))\,dt_1\cdots dt_n$$

for $n = 2, 3, \ldots$ together with

$$\mathsf{V}(t) := \int_{\mathcal{C}(L)} :\mathcal{V}(\mathbf{x},t): d^3\mathbf{x}$$

and

$$\boxed{\mathcal{V}(x) := \mathrm{i}e\overline{\psi}(x)\gamma_\mu A^\mu(x)\psi(x).}$$

Here, ψ, A^μ denote free fields, and $x = (\mathbf{x}, t)$. Moreover, the electron has the electric charge $-e$.[1] In general, incoming particles are described by the following normalized states

$$\Phi_{\mathrm{in}} = \frac{(a^+_{\mathbf{p}_1,s_1})^{n_1}}{\sqrt{n_1!}}\cdots\frac{(a^+_{\mathbf{p}_A,s_A})^{n_A}}{\sqrt{n_A!}}b^+_{\mathbf{q}_1,\sigma_1}\cdots b^+_{\mathbf{q}_B,\sigma_B}c^+_{\mathbf{r}_1,\tau_1}\cdots c^+_{\mathbf{r}_C,\tau_C}|0\rangle$$

which correspond to photons, electrons, and positrons. Similar expressions are used for Φ_{out}. For the transition amplitude, we get

$$\langle 0|\ldots\mathsf{S}(T)\ldots|0\rangle \tag{14.2}$$

where ... represent products of creation operators for incoming and outgoing particles.

Perturbation theory uses the approximation $I+\mathsf{S}_1+\ldots+\mathsf{S}_N$ *of the S-matrix* S *in* N*th order.*

For computing physical processes via Feynman diagrams, the crucial Feynman rules are summarized in Table 14.5 on page 897. In this chapter, we explain how these very convenient Feynman rules arise from the Dyson series by applying the Wick theorem. For example, this should help the reader to understand the appearance and the size of the appropriate symmetry factors.

14.1 Compton Effect and Feynman Rules in Position Space

The transition amplitudes (14.2) are computed by means of the Main Wick Theorem 13.2 on page 846. This leads to the Feynman rules from Table 14.5 on page 897. To illustrate this, let us consider a typical example, namely, the scattering process between one photon and one electron (Compton scattering) in second order. In this case, the normalized incoming state

$$\boxed{\Phi_{\mathrm{in}} := a^+_{\mathbf{p},s}b^+_{\mathbf{q},\sigma}|0\rangle}$$

consists of one photon having the momentum vector $\mathbf{p}$ and the polarization number $s = 1, 2$, along with one electron having the momentum vector $\mathbf{q}$ and the spin number $\sigma = \pm\frac{1}{2}$. Similarly, the normalized outgoing state

[1] Recall that, by our convention (10.17), we work in the energetic system of physical units, that is, $c = \hbar = \varepsilon_0 = \mu_0 = 1$. Using the notation introduced in Sect. 13.1, we have $\mathsf{V}(t) = -\mathrm{i}H_{\mathrm{int}}$.

$$\boxed{\Phi_{\text{out}} := a^+_{\mathbf{p}',s'} b^+_{\mathbf{q}',\sigma'} |0\rangle}$$

consists of one photon having the momentum vector $\mathbf{p}'$ and the polarization number $s' = 1, 2$, along with one electron having the momentum vector $\mathbf{q}'$ and the spin number $\sigma' = \pm\frac{1}{2}$. In order to simplify notation in the following, we suppress polarization and spin indices, that is, we write

$$a^+_{\mathbf{p}}, a^-_{\mathbf{p}'}, b^+_{\mathbf{q}}, b^-_{\mathbf{q}'}$$

instead of $a^+_{\mathbf{p},s}, a^-_{\mathbf{p}',s'}, b^+_{\mathbf{q},\sigma}, b^-_{\mathbf{q}',\sigma'}$, respectively. We also write S instead of $\mathsf{S}(T)$. It is our goal to compute the following transition amplitude

$$\tau := \langle \Phi_{\text{out}} | \mathsf{S} \Phi_{\text{in}} \rangle = \langle 0 | a^-_{\mathbf{p}'} b^-_{\mathbf{q}'} \mathsf{S} a^+_{\mathbf{p}} b^+_{\mathbf{q}} | 0 \rangle.$$

In second-order approximation, we get

$$\boxed{\tau = \langle 0 | a^-_{\mathbf{p}'} b^-_{\mathbf{q}'} (I + \mathsf{S}_1 + \mathsf{S}_2) a^+_{\mathbf{p}} b^+_{\mathbf{q}} | 0 \rangle.}$$

Here, we set $\mathsf{S}_0 := I$ along with

$$\mathsf{S}_1 := \int_\Omega d^4x : \mathrm{i} e \overline{\psi}(x) \gamma_\mu A^\mu(x) \psi(x) :$$

and

$$\boxed{\mathsf{S}_2 := \tfrac{1}{2} \mathcal{T} \int_{\Omega \times \Omega} d^4x d^4y : \mathrm{i} e \overline{\psi}(x) \gamma_\mu A^\mu(x) \psi(x) : \, : \mathrm{i} e \overline{\psi}(y) \gamma_\nu A^\nu(y) \psi(y) : .}$$

Here, $\Omega := [-\frac{T}{2}, \frac{T}{2}] \times \mathcal{C}(L)$, and $x = (t, \mathbf{x})$. Furthermore, $d^4x = dt \, d^3\mathbf{x}$. The operator $\mathcal{T}$ organizes the time ordering of the integrand. Let us compute the single terms.

(i) Zero-order approximation: Naturally enough, we assume that the incoming particles are different from the outgoing particles. Hence

$$\langle \Phi_{\text{out}} | \Phi_{\text{in}} \rangle = 0.$$

This implies $\tau_0 := \langle 0 | a^-_{\mathbf{p}'} b^-_{\mathbf{q}'} \mathsf{S}_0 a^+_{\mathbf{p}} b^+_{\mathbf{q}} | 0 \rangle = 0$.

(ii) First-order approximation: Since S_1 represents an operator product of three factors $\overline{\psi}, A^\mu, \psi$, the transition amplitude

$$\tau_1 := \langle 0 | a^-_{\mathbf{p}'} b^-_{\mathbf{q}'} \mathsf{S}_1 a^+_{\mathbf{p}} b^+_{\mathbf{q}} | 0 \rangle$$

contains an operator product of $2 + 3 + 2$ factors. This is an odd number. Therefore, $\tau_1 = 0$, by the Main Wick Theorem 13.2 on page 846.

(iii) Second-order approximation: Again by the Main Wick Theorem 13.2, the transition amplitude

$$\tau_2 := \langle 0 | a^-_{\mathbf{p}'} b^-_{\mathbf{q}'} \mathsf{S}_2 a^+_{\mathbf{p}} b^+_{\mathbf{q}} | 0 \rangle$$

can be represented as the sum

$$\boxed{\tau_2 = \tfrac{1}{2} (\tau_{21} + \tau_{22} + \tau_{23} + \tau_{24})}$$

on pairings. This is represented graphically in Table 14.1(a)–(d) on page 879 by using the conventions introduced in Table 14.5 on page 897.

Let us discuss this. Using the formula

$$\boxed{\tau_{2j} := -e^2 \int_{\Omega\times\Omega} \rho_{2j}(x,y)d^4xd^4y, \qquad j=1,2,3,4,} \tag{14.3}$$

we get the following expressions:

(a) Transition amplitude τ_{21} (Table 14.1(a)):

$$\varrho_{21} := \langle 0|a^-_{\mathbf{p}'}b^-_{\mathbf{q}'} : \overline{\psi}(x)\gamma_\mu A^\mu(x)\psi(x) : \, : \overline{\psi}(y)\gamma_\nu A^\nu(y)\psi(y) : a^+_{\mathbf{p}}b^+_{\mathbf{q}}|0\rangle.$$

(b) Transition amplitude τ_{22} (Table 14.1(b)):

$$\varrho_{22} := \langle 0|a^-_{\mathbf{p}'}b^-_{\mathbf{q}'} : \overline{\psi}(x)\gamma_\mu A^\mu(x)\psi(x) : \, : \overline{\psi}(y)\gamma_\nu A^\nu(y)\psi(y) : a^+_{\mathbf{p}}b^+_{\mathbf{q}}|0\rangle.$$

(c) Transition amplitude τ_{23} (Table 14.1(c)):

$$\varrho_{23} := \langle 0|a^-_{\mathbf{p}'}b^-_{\mathbf{q}'} : \overline{\psi}(x)\gamma_\mu A^\mu(x)\psi(x) : \, : \overline{\psi}(y)\gamma_\nu A^\nu(y)\psi(y) : a^+_{\mathbf{p}}b^+_{\mathbf{q}}|0\rangle.$$

(d) Transition amplitude τ_{24} (Table 14.1(d)):

$$\varrho_{24} := \langle 0|a^-_{\mathbf{p}'}b^-_{\mathbf{q}'} : \overline{\psi}(x)\gamma_\mu A^\mu(x)\psi(x) : \, : \overline{\psi}(y)\gamma_\nu A^\nu(y)\psi(y) : a^+_{\mathbf{p}}b^+_{\mathbf{q}}|0\rangle.$$

The symbols from (a)–(d) above are obtained systematically in the following way:

- Use only the pairings $\overline{\psi}\psi$, $\psi\overline{\psi}$, a^-A, $A\,a^+$, and $b^-\overline{\psi}$, ψb^+ (see page 880).
- Consider all completely paired expressions.

The Feynman graphs from Table 14.1(a)–(d) on page 879 are in one-to-one correspondence to the analytical expressions (a)–(d) above. This method is a universal one. It applies to all kinds of Feynman diagrams in quantum electrodynamics.

Justification. The key result is the Main Wick Theorem 13.2 on page 846. The pairings of ϱ_{21} are to be understood in the following sense:

$$\varrho_{21}(x,y) = b^-_{\mathbf{q}'}\overline{\psi}(x)\gamma_\mu a^-_{\mathbf{p}'}A^\mu(x)\psi(x)\overline{\psi}(y)\gamma_\nu A^\nu(y)a^+_{\mathbf{p}}\psi(y)b^+_{\mathbf{q}}.$$

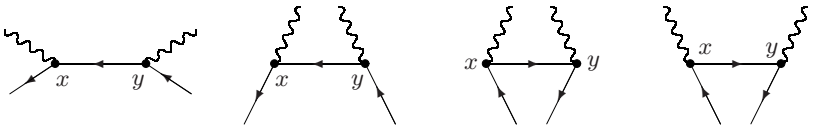
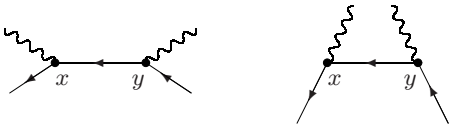
Here, we use the time-ordered contractions defined by

$$\boxed{B(x)D(y) = \overbrace{B(x)D(y)} := \mathcal{C}_{\mathcal{T}}(B(x)D(y)) = \langle 0|\mathcal{T}(B(x)D(y))|0\rangle.}$$

Furthermore, we obtain the following product of time-ordered contractions:

Table 14.1. Electron-photon scattering (Compton scattering)

Vertex diagrams

Composition of vertex diagrams: $\tau_2 = \frac{1}{2}(\tau_{21} + \tau_{22} + \tau_{23} + \tau_{24})$
 (a) τ_{21} (b) τ_{22} (c) $\tau_{23} = \tau_{21}$ (d) $\tau_{24} = \tau_{22}$
Normal forms: $\tau = \tau_{21} + \tau_{22}$
 (a) τ_{21} (b) τ_{22}

$$\varrho_{22}(x,y) = b^-_{\mathbf{q}'}\overline{\psi}(x)\gamma_\mu \overbrace{A^\mu(x)a^+_{\mathbf{p}}}\psi(x)\overline{\psi}(y)\psi(y)b^+_{\mathbf{q}}\gamma_\nu \overbrace{a^-_{\mathbf{p}'}A^\nu}(y).$$

Similarly,

$$\varrho_{23}(x,y) = -b^-_{\mathbf{q}'}\overline{\psi}(y)\gamma_\mu \overbrace{A^\mu(x)a^+_{\mathbf{p}}}\overline{\psi}(x)\psi(y)\gamma_\nu \overbrace{a^-_{\mathbf{p}'}A^\nu}(y)\psi(x)b^+_{\mathbf{q}}$$

where the sign is generated by an odd permutation of fermionic operators $\overline{\psi}, \psi$. Finally,

$$\varrho_{24}(x,y) = -b^-_{\mathbf{q}'}\overline{\psi}(y)\gamma_\mu \overbrace{a^-_{\mathbf{p}'}A^\mu}(x)\overline{\psi}(x)\psi(y)\psi(x)b^+_{\mathbf{q}}\gamma_\nu \overbrace{A^\nu(y)a^+_{\mathbf{p}}}.$$

Since $\gamma_\mu, \psi, \overline{\psi}$ are matrices, one has to read these equations with care. In fact, in order to get the explicit expressions one has to use components. This will be done below.

Furthermore, note that symmetry substantially simplifies the computation of Feynman diagrams. As we will show below, we have

$$\tau_{23} = \tau_{12}, \qquad \tau_{24} = \tau_{22}. \tag{14.4}$$

Hence

$$\boxed{\tau = \tau_{21} + \tau_{22}.}$$

Thus, one has only to consider the Feynman diagrams (a) and (b) in Table 14.1 on page 879. This is the basic idea behind normal forms of Feynman diagrams. The proof of (14.4) is an immediate consequence of the symmetry relations

$$\varrho_{23}(x,y) = \varrho_{21}(y,x), \qquad \varrho_{23}(x,y) = \varrho_{22}(y,x). \tag{14.5}$$

Mnemonically, this follows from interchanging x with y and using

$$\boxed{\psi(x)\overline{\psi}(y) = -\overline{\psi}(y)\psi(x).}$$

Intuitively, this reflects symmetry properties of the graphs from Table 14.1 on page 879, as will be discussed below.

Time-ordered pairings. It remains to discuss Table 14.5 on page 897. As a typical example, let us show that

$$\boxed{\underbrace{\psi^\beta(x) b^+_{\mathbf{q},s}} = \langle 0|\psi^\beta(x) b^+_{\mathbf{q},s}|0\rangle = N_{\mathbf{q}} u_{\mathbf{q},s} \mathrm{e}^{\mathrm{i}(\mathbf{qx} - E_{\mathbf{q}} t)}.}$$

Proof. For the free electron field, we have the expression

$$\psi(x) := \sum\nolimits'_{\mathbf{q}\in\mathcal{G}(N), s=\pm\frac{1}{2}} (b^-_{\mathbf{q},s} u_{\mathbf{q},s}\ \mathrm{e}^{\mathrm{i}(\mathbf{qx}-E_{\mathbf{q}}t)} + c^+_{\mathbf{q},s} v_{\mathbf{q},s}\ \mathrm{e}^{-\mathrm{i}(\mathbf{qx}-E_{\mathbf{q}}t)})\ N_{\mathbf{q}}$$

with $x = (\mathbf{x}, t)$. It follows from the anticommutation relation

$$b^-_{\mathbf{q},s} b^+_{\mathbf{q},s} = I - b^+_{\mathbf{q},s} b^-_{\mathbf{q},s}$$

along with $b^-_{\mathbf{q},s}|0\rangle = 0$ that

$$\langle 0|b^-_{\mathbf{q},s} b^+_{\mathbf{q},s}|0\rangle = \langle 0|I|0\rangle\langle 0|0\rangle = 1.$$

If $(\mathbf{q}', s') \neq (\mathbf{q}, s)$, then

$$b^-_{\mathbf{q}',s'} b^+_{\mathbf{q},s} = -b^+_{\mathbf{q},s} b^-_{\mathbf{q}',s'}.$$

Hence $\langle 0|b^-_{\mathbf{q}',s'} b^+_{\mathbf{q},s}|0\rangle = 0$. Finally, $c^-_{\mathbf{q}',s'}|0\rangle = 0$ and $(c^+)^\dagger_{\mathbf{q}',s'} = c^-_{\mathbf{q}',s}$ for all $\mathbf{q}', s'$ implies that

$$\langle 0|c^+_{\mathbf{q}',s'} b^+_{\mathbf{q},s}|0\rangle = \langle (c^-_{\mathbf{q}',s'}|0\rangle)|b^+_{\mathbf{q},s}|0\rangle = 0.$$

□

Similarly, we obtain the other expressions from Table 14.5 on 897. In addition, observe the following peculiarities. By the Main Wick Theorem 13.2 on page 846, we have to take all the completely paired expressions into account. However, some of them vanish. In particular, the following pairings drop out:

- pairings between bosons and fermions, $\underbrace{\psi A}$, $\underbrace{A\psi}$, $\underbrace{\overline{\psi} A}$, $\underbrace{A\overline{\psi}}$;
- pairings inside a normal product, $\underbrace{\overline{\psi}(x)\psi(x)}$;
- the pairings $\underbrace{\psi(x)\psi(y)}$ and $\underbrace{\overline{\psi}(x)\overline{\psi}(y)}$;
- pairings between creation and annihilation operators corresponding to different particles.

Therefore, only the time-ordered pairings listed in Table 14.5 on page 897 have to be taken into account.

14.2 Symmetry Properties

It happens quite naturally that different graphs yield the same transition amplitudes, by symmetry. This simplifies the computation of transition amplitudes. To explain this, let us show that

$$\boxed{\tau_{23} = \tau_{21}.} \tag{14.6}$$

This is a consequence of the symmetry property (14.5) and the fact that the integral (14.3) is invariant under a permutation of x and y. To prove (14.5), observe that $\varrho_{21}(x, y)$ is equal to

$$-\underbrace{b^-_{\mathbf{q}'}\overline{\psi}^{\alpha}}(x)(\gamma_\mu)_{\alpha\beta}\overbrace{a^-_{\mathbf{p}'}A^{\mu}}(x)\underbrace{\psi^{\beta}(x)\overline{\psi}^{\kappa}}(y)(\gamma_\nu)_{\kappa\lambda}\overbrace{A^{\nu}(y)a^+_{\mathbf{p}}}\underbrace{\psi^{\lambda}(y)b^+_{\mathbf{q}}}.$$

By Table 14.5 on page 897, we get the following key relation

$$\boxed{\underbrace{\psi^{\beta}(x)\overline{\psi}^{\kappa}}(y) = -\underbrace{\overline{\psi}^{\kappa}(y)\psi^{\beta}}(x).}$$

On the other hand, $\varrho_{23}(x, y)$ is equal to

$$-\underbrace{b^-_{\mathbf{q}'}\overline{\psi}^{\lambda}}(y)(\gamma_\mu)_{\alpha\beta}\overbrace{A^{\mu}(x)a^+_{\mathbf{p}}}\underbrace{\overline{\psi}^{\alpha}(x)\psi^{\lambda}}(y)(\gamma_\nu)_{\kappa\lambda}\overbrace{a^-_{\mathbf{p}'}A^{\nu}}(y)\underbrace{\psi^{\beta}(x)b^+_{\mathbf{q}}}.$$

Changing variables,

$$\boxed{x \Leftrightarrow y, \quad \mu \Leftrightarrow \nu, \quad \alpha \Leftrightarrow \kappa, \quad \beta \Leftrightarrow \lambda,}$$

we get the desired relation $\varrho_{21}(x, y) = \varrho_{23}(y, x)$. Similarly, we obtain

$$\boxed{\tau_{24} = \tau_{21}.}$$

Physical intuition. Table 14.1(a) on page 879 represents

- an incoming photon and an incoming electron at the vertex y,
- an outgoing photon and an outgoing electron at the vertex x, and
- a virtual electron moving from y to x.

Furthermore, Table 14.1(c) on page 879 describes

- an incoming photon and an incoming electron at the vertex x,
- an outgoing photon and an outgoing electron at the vertex y, and
- a virtual electron moving from x to y.

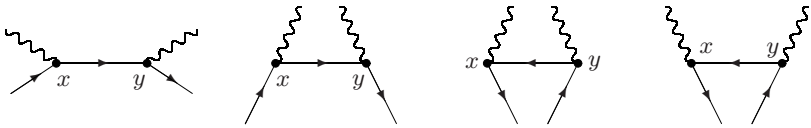
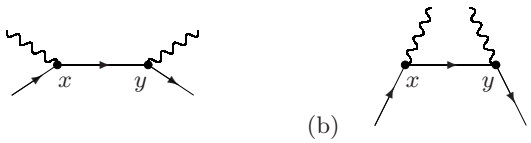
This is precisely Table 14.1(a) on page 879 by interchanging the role of the vertices x and y. Since the two Feynman diagrams from Table 14.1(a) and Table 14.1(c) describe the same physical situation, we expect that the corresponding transition amplitudes coincide. Our rigorous proof for $\tau_{23} = \tau_{12}$ given above shows that this is indeed true.

Similarly, the two Feynman diagrams from Table 14.1(b) and Table 14.1(d) represent the same physical situation. Therefore, we may restrict ourselves to the normal forms from Table 14.1(a) and Table 14.1(b) by counting them twice.

The positron trick for Feynman diagrams. If we replace creation and annihilation operators for electrons by creation and annihilation operators for positrons, that is, if we use the substitution

Table 14.2. Positron-photon scattering

Vertex diagrams

Composition of vertex diagrams: $\tau_2 = \frac{1}{2}(\tau_{21} + \tau_{22} + \tau_{23} + \tau_{24})$
 (a) τ_{21} (b) τ_{22} (c) $\tau_{23} = \tau_{21}$ (d) $\tau_{24} = \tau_{22}$
Normal forms: $\tau = \tau_{21} + \tau_{22}$
 (a) (b)

$$b^{\pm}_{\mathbf{q},\sigma} \Rightarrow c^{\pm}_{\mathbf{q},\sigma},$$

then electron-photon scattering passes over to positron-photon scattering. From Table 14.1 on page 879 we obtain Table 14.2. Here, the external electrons are replaced by external positrons, by simply reversing the arrows. Mnemonically, physicists say that:

Positrons are electrons running backward in time.

This very fruitful idea was introduced by Stueckelberg in the 1940s and extensively used by Feynman in connection with Feynman diagrams.

In order to avoid any ambiguities, we always read Feynman diagrams from right to left corresponding to increasing time.

14.3 Summary of the Feynman Rules in Momentum Space

It is worth noting that the notation facilitates discovery. This, in a most wonderful way, reduces the mind's labors.

Gottfried Wilhelm Leibniz (1646-1717)

Let us consider the Feynman diagrams of quantum electrodynamics in nth order of perturbation theory. We want to study a fixed scattering process for given incoming and outgoing particles. As a prototype think of electron-photon scattering (Compton effect) from Table 14.1 on page 879. To fix the terminology, we put the physical system into a finite cubic box $\mathcal{C}(L)$ of side length L and volume[2]

$$\mathcal{V} = L^3.$$

The discrete volume element of the 3-dimensional momentum space is equal to

$$\Delta^3\mathbf{p} = \frac{(2\pi)^3}{\mathcal{V}}.$$

Let us observe the given scattering process during the time interval $[-\frac{T}{2}, \frac{T}{2}]$ of length T.

(i) Composition rule for vertex diagrams: Consider n numbered vertices

(14.7)

Construct all possible combinations in such a way that all the processes are generated which fit the situation given by the prescribed incoming and outgoing particles. This construction corresponds to all possible complete pairings of the form

$$\langle 0|a^-b^-\ldots : \overline{\psi}(x_1)\, \not{A}(x_1)\psi(x_1) : \cdots : \overline{\psi}(x_n)\, \not{A}(x_n)\psi(x_n) : a^+b^+\ldots\, |0\rangle,$$

similarly as in (14.3). By convention, the diagrams have to be read from right to left for increasing time.

(ii) Normal forms: Two Feynman diagrams from (i) are called equivalent iff they represent the same physical situation (up to a permutation of the vertices). Select a complete family of inequivalent Feynman diagrams in a fixed, but otherwise arbitrary manner. The members of this family are called the normal forms of the Feynman diagrams from (i). To each normal form, assign the factor

$$\frac{m}{n!}\,,$$

where m is the multiplicity of the normal form (i.e., the number of corresponding equivalent Feynman diagrams). As a rule, $m = n!$. If there exist internal symmetries of the normal form, than the multiplicity m can be less than $n!$.

[2] Explicitly, the cube $\mathcal{C}(L)$ in 3-dimensional position space consists of all the points $(x^1, x^2, x^3) \in \mathbb{R}^3$ with $|x^j| \leq L$ for $j = 1, 2, 3$. The grid $\mathcal{G}(N)$ in 3-dimensional momentum space consists of all the momentum vectors

$$\mathbf{p} = p^1\mathbf{i} + p^2\mathbf{j} + p^3\mathbf{k}, \qquad p^j = m^j\Delta p, \qquad j = 1, 2, 3$$

where m^1, m^2, m^3 are integers with $|m^j| \leq N$ for $j = 1, 2, 3$, and $\Delta p := 2\pi/L$.

(iii) Vertex x: Assign the factor[3]

$$\boxed{\mathrm{i}e(\gamma_\mu)_{\alpha\beta}\cdot(2\pi)^4\,\delta_{\mathrm{dis}}(q'-q\pm p)}$$

where $+p$ (resp. $-p$) refers to an outgoing (resp. incoming) photon at the vertex x. Moreover, we use the following notation:

- 4-momentum vector $q=(E_{\mathbf{q}},\mathbf{q})$ of the incoming fermion with the given 3-momentum vector $\mathbf{q}$ and the energy $E_{\mathbf{q}}:=\sqrt{\mathbf{q}^2+m_e^2}$;
- 4-momentum vector q' of the outgoing fermion;
- 4-momentum vector $p=(\omega_{\mathbf{p}},\mathbf{p})$ of the incoming (resp. outgoing) photon with the 3-momentum vector $\mathbf{p}$ and the energy $\omega_{\mathbf{p}}:=|\mathbf{p}|$.

The properties of discrete Dirac functions are studied in Sect. 15.1.1 of Vol. I. As we will discuss below, the appearance of the discrete Dirac delta function ensures conservation of energy and momentum at each vertex.

(vi) External lines at the vertex x: Assign the following factors:[4]

$\mathcal{N}_{\mathbf{p}}e_r^\mu$	(incoming or outgoing photon)
$N_{\mathbf{q}}u^\beta_{\mathbf{q},s}$	(incoming electron)
$N_{\mathbf{q}'}\overline{u}^\alpha_{\mathbf{q}',s'}$	(outgoing electron)
$N_{\mathbf{q}'}\overline{v}^\alpha_{\mathbf{q}',s'}$	(incoming positron)
$N_{\mathbf{q}}v^\beta_{\mathbf{q},s}$	(outgoing positron).

The normalization factors are given by

$$\mathcal{N}_{\mathbf{p}}:=\sqrt{\frac{1}{2\omega_{\mathbf{p}}\mathcal{V}}},\qquad N_{\mathbf{q}}:=\sqrt{\frac{1}{2E_{\mathbf{q}}\mathcal{V}}}.$$

(vi) Internal fermion line pointing from the vertex x_k to the vertex x_j: Assign the semi-discrete integral

$$\int_{\mathbb{R}}dq^0\sum_{\mathbf{q}\in\mathcal{G}(N)}\Delta^3\mathbf{q}\,S^{\beta_j\alpha_k}(q)$$

where

$$\boxed{S(q):=\frac{\mathrm{i}}{(2\pi)^4}\cdot\frac{\not{q}+m_eI}{(q^0)^2-(E_{\mathbf{q}}-\varepsilon\mathrm{i})^2}}\tag{14.8}$$

along with $\not{q}:=\gamma_\mu q^\mu$, the regularization parameter $\varepsilon>0$, and the fermion energy $E_{\mathbf{q}}:=\sqrt{\mathbf{q}^2+m_e^2}$.

[3] To simplify notation, we write x instead of x_j. Similarly, we write $p,q,\mu,\alpha,\beta,...$ instead of $p_j,q_j,\mu_j,\alpha_j,\beta_j,...$, respectively.

[4] For the definition of the Dirac spinors $u_{\mathbf{q},s},v_{\mathbf{q},s}$ with the spin number $s=\pm\frac{1}{2}$, we refer to (12.10) on page 807. Furthermore, recall that $\overline{u}_{\mathbf{q},s}:=u^\dagger_{\mathbf{q},s}\gamma^0$. Analogously, the symbol $\overline{v}_{\mathbf{q},s}$ is related to $v_{\mathbf{q},s}$. For the polarization vectors, we have s

$$\mathbf{e}_r=e_r^1\mathbf{i}+e_r^2\mathbf{j}+e_r^3\mathbf{k},\quad e_r^0=0,\qquad r=1,2.$$

Here, the three vectors $\mathbf{e}_1,\mathbf{e}_2$, and $\mathbf{p}/|\mathbf{p}|$ form a right-handed orthonormal system.

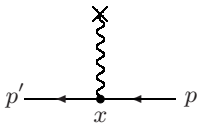


Fig. 14.1. Electron in an external electromagnetic field

(vii) Internal photon line connecting the two vertices x_j and x_k: Assign the semi-discrete integral

$$\int_{\mathbb{R}} dp^0 \sum_{\mathbf{p}\in\mathcal{G}(N)} \Delta^3\mathbf{p}\, D^{\mu_j\mu_k}(p)$$

where

$$\boxed{D^{\mu_j\mu_k}(p) := -\frac{\mathrm{i}\eta^{\mu_j\mu_k}}{(2\pi)^4}\cdot\frac{1}{(p^0)^2-(\mathcal{E}_{\mathbf{p}}-\varepsilon\mathrm{i})^2}} \tag{14.9}$$

along with the virtual photon mass $m_{ph} > 0$, the regularization parameter $\varepsilon > 0$, and the regularized photon energy $\mathcal{E}_{\mathbf{p}} := \sqrt{\mathbf{p}^2 + m_{ph}^2}$.

(viii) External electromagnetic field at the vertex x: Assign the factor A^μ. Graphically, an external electromagnetic field is represented by an external photon line equipped with a cross. Fig. 14.1 shows the motion of an electron under the influence of an external electromagnetic field with the four-potential (A^1, A^2, A^3, A^0). The cross section for this process will be computed in Sect. 15.2 on page 914.

(ix) Furry's rule: Consider a fermion loop consisting of l fermion lines. If l is even, assign the factor $(-1)^l$. If l is odd, the Feynman diagram can be cancelled, since it does not contribute to the transition amplitude.

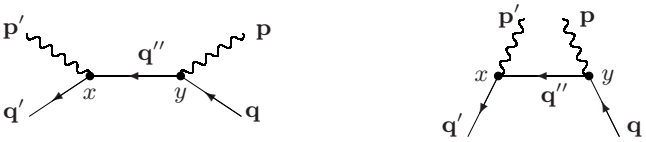
Justification of the Feynman rules. These rules are a simple consequence of the rules from Table 14.5 on page 897 by carrying out the integrations in position space. These integrations produce the discrete Dirac delta functions at the vertices. The typical argument can be found in the proof of Example 14.3 on page 888.

Furry's rule allows us to reduce positron lines to electron lines (plus an additional sign). This rule depends on both the main Wick theorem and symmetry properties for the trace of products of Dirac matrices. The proof of Fury's rule will be given in Problem 15.13.

14.4 Typical Examples

> In terms of physics, Feynman propagators describe the propagation of physical effects in quantum field theory by taking both causality and antiparticles into account.
> In terms of mathematics, Feynman propagators are distinguished fundamental solutions of the wave equation, the Klein–Gordon equation, and the Dirac equation. In the Fourier space, Feynman propagators are inverse differential operators, after regularization.
>
> Folklore

Let us illustrate the Feynman rules by considering some examples.



(a) τ_{21} (b) τ_{22}

Fig. 14.2. Crossing symmetry of Compton scattering

Example 14.1 (electron-photon scattering – Compton scattering). We are given

- an incoming photon with momentum vector $\mathbf{p}$ and polarization number $r = 1, 2$,
- an incoming electron with momentum vector $\mathbf{q}$ and spin number $s = \pm\frac{1}{2}$,
- an outgoing photon with momentum vector $\mathbf{p}$ and polarization number $r' = 1, 2$,
- and an outgoing electron with momentum vector $\mathbf{q}'$ and spin number $s' = \pm\frac{1}{2}$.

We assume that the incoming particles are different from the outgoing particles. By momentum conservation and energy conservation, we get

$$\mathbf{p}' + \mathbf{q}' = \mathbf{p} + \mathbf{q}, \qquad \omega_{\mathbf{p}'} + E_{\mathbf{q}'} = \omega_{\mathbf{p}} + E_{\mathbf{q}}.$$

The normal forms of the corresponding Feynman diagrams are pictured in Fig. 14.2. Following the convention introduced by Dirac, the symbol

$$\langle \mathbf{p}', r'; \mathbf{q}', s' | \mathsf{S}(T) | \mathbf{p}, r; \mathbf{q}, s \rangle$$

stands for the following transition amplitude:

$$\langle 0 | a^-_{\mathbf{p}',r'} b^-_{\mathbf{q}',s'} \mathsf{S}(T)\ a^+_{\mathbf{p},r} b^+_{\mathbf{q},s} | 0 \rangle.$$

In second-order approximation, the transition amplitude

$$\boxed{\tau := \langle \mathbf{p}', r'; \mathbf{q}', s' | I + \mathsf{S}_1(T) + \mathsf{S}_2(T)\ | \mathbf{p}, r; \mathbf{q}, s \rangle}$$

of Compton scattering is given by $\tau = \tau_{21} + \tau_{22}$ where

$$\tau_{2j} := \int_{\mathbb{R}} dq''_0 \sum_{\mathbf{q}'' \in \mathcal{G}(N)} \Delta^3 \mathbf{q}''\ f_{2j}(q''), \qquad j = 1, 2.$$

The integrand $f_{21}(q'')$ is equal to

$$-\frac{(2\pi)^8 e^2}{2\mathcal{V}^2 \sqrt{\omega_{\mathbf{p}'} \omega_{\mathbf{p}}}} \quad \delta_{\mathrm{dis}}(p' + q' - q'')\ \delta_{\mathrm{dis}}(q'' - p - q)$$
$$\times \frac{\overline{u}_{\mathbf{q}',s'}}{\sqrt{2E_{\mathbf{p}'}}}\ \not{\epsilon}_{r'}\ S(q'')\ \not{\epsilon}_r \frac{u_{\mathbf{q},s}}{\sqrt{2E_{\mathbf{p}}}}.$$

Here, we set $\not{\epsilon}_r := e^\mu_r \gamma_\mu$. The matrix $S(q'')$ can be found in (14.8). Recall that $\mathsf{S}_0 = I$ and S_1 do not contribute to the scattering amplitude. Hence

$$\tau := \langle \mathbf{p}', r'; \mathbf{q}', s' | \mathsf{S}_2(T)\ | \mathbf{p}, r; \mathbf{q}, s \rangle.$$

The integrand f_{22} is obtained from f_{21} by using the transformation

$$(p, r) \Rightarrow (-p', r'), \qquad (p', r') \Rightarrow (-p, r).$$

Intuitively, the incoming (resp. outgoing) photon is interchanged with an outgoing (resp. incoming) photon. This symmetry property of the Feynman diagram depicted in Fig. 14.2 above is called *crossing symmetry*.

Proof. Consider first τ_{21}. Using the Feynman diagram from Fig. 14.2(a) above, it follows from the Feynman rules in momentum space that

$$\tau_{21} = \int_{\mathbb{R}} dq_0'' \sum_{\mathbf{q}'' \in \mathcal{G}(N)} \Delta^3 \mathbf{q}''\ f_{21}(q'').$$

Here, the integrand $f_{21}(q'')$ is equal to

$$-e^2 \mathcal{N}_{\mathbf{p}'} N_{\mathbf{q}'} e^{\mu_1}_{r'} \overline{u}^{\alpha_1}_{\mathbf{q}',s'} (\gamma_{\mu_1})_{\alpha_1\beta_1} S^{\beta_1\alpha_2}(q'') (\gamma_{\mu_2})_{\alpha_2\beta_2} u^{\beta_2}_{\mathbf{q},s} e^{\mu_2}_{r} \mathcal{N}_{\mathbf{p}} N_{\mathbf{q}} g(q'')$$

along with

$$g(q'') := (2\pi)^8\ \delta_{\text{discrete}}(q'' - p' - p') \cdot \delta_{\text{discrete}}(p + q - q'').$$

Argue similarly for τ_{22} related to Fig. 14.2(b) above. □

The following observation is crucial for the computation of the cross section in Sect. 14.6. For large time intervals $[-\frac{T}{2}, \frac{T}{2}]$ and vanishing regularization parameter, $\varepsilon = 0$, the transition amplitude has the following asymptotic form, as $T \to +\infty$:

$$\boxed{\begin{aligned}\tau = -\frac{(2\pi)^4 \mathrm{i} e^2}{4\mathcal{V}^2(\omega_{\mathbf{p}'}\omega_{\mathbf{p}} E_{\mathbf{q}'} E_{\mathbf{q}})^{1/2}} \cdot \delta_{\text{dis}}(p' + q' - p - q) \times \\ \times \overline{u}_{\mathbf{p}',s'} M u_{\mathbf{0},s} \cdot (1 + o(1))\end{aligned}}$$

with the matrix

$$M := e_{s'} \frac{\not{q} + \not{p} + m_e}{(q+p)^2 - m_e^2}\ e_s + e_s \frac{\not{q} - \not{p}' + m_e}{(q-p')^2 - m_e^2}\ e_{s'}.$$

Furthermore, as $T \to +\infty$,

$$\boxed{\begin{aligned}\frac{|\tau|^2}{T} = \frac{(2\pi)^4 e^4}{16\mathcal{V}^3 \omega_{\mathbf{p}'}\omega_{\mathbf{p}} E_{\mathbf{q}'} E_{\mathbf{q}}} \cdot \delta_{\text{dis}}(p' + q' - p - q) \times \\ \times |\overline{u}_{\mathbf{q}',s'} M u_{\mathbf{0},s}|^2 \cdot (1 + o(1)).\end{aligned}} \tag{14.10}$$

Proof. We will critically use the properties of the discrete Dirac delta function (see Sect. 12.1.2 of Vol. I). Obviously, the sum

$$\sum_{\mathbf{q}' \in \mathcal{G}(N)} \Delta^3 \mathbf{q}''\ \delta_{\mathcal{G}(N)}(\mathbf{p}' + \mathbf{q}' - \mathbf{q}'') \cdot \delta_{\mathcal{G}(N)}(\mathbf{q}'' - \mathbf{p} - \mathbf{q})$$

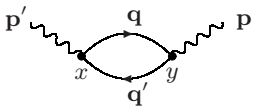
is equal to $\delta_{\mathcal{G}(N)}(\mathbf{p}' + \mathbf{q}' - \mathbf{p} - \mathbf{q})$. By Prop. 12.1 on page 672 of Vol. I, as $T \to +\infty$, we get

$$\int_{\mathbb{R}} dq_0''\ \delta_T(p_0' + q_0' - q_0'') \cdot \delta_T(q_0'' - p_0 - q_0) = \delta_T(p_0' + q_0' - p_0 - q_0)(1 + o(1)),$$

and $\delta_{\text{dis}}(p)^2 = \frac{T\mathcal{V}}{(2\pi)^4} \cdot \delta_{\text{dis}}(p)(1 + o(1))$. □

Table 14.3. Self-interaction of a photon

Vertex diagrams

Composition of vertex diagrams: $\tau_2 = \frac{1}{2}(\tau_{21} + \tau_{22})$
 (a) τ_{21} (b) $\tau_{22} = \tau_{21}$
Normal form: $\tau_2 = \tau_{21}$


Example 14.2 (positron-photon scattering). Replace the incoming and outgoing electron in Example 14.1 by positrons. Then, we get the transition amplitudes τ_{21} and τ_{22} as above by using the replacement

$$u_{\mathbf{q},s} \Rightarrow \overline{v}_{\mathbf{q},s}, \qquad \overline{u}_{\mathbf{q}',s'} \Rightarrow v_{\mathbf{q}',s'}.$$

Example 14.3 (Self-interaction of a photon). Consider the interaction between

- one photon of momentum vector $\mathbf{p}$ and polarization number r and
- another photon of momentum vector $\mathbf{p}'$ and polarization number r' (see Table 14.3).

For the transition amplitude in second-order approximation, we are going to show that

$$\boxed{\langle \mathbf{p}', r' | I + \mathsf{S}_1 + \mathsf{S}_2 | \mathbf{p}, r \rangle = \delta_{\mathbf{p}'\mathbf{p}} + \tau_2}$$

along with

$$\tau_2 = \int_{\mathbb{R}^2} dq_0 dq_0' \sum_{\mathbf{q},\mathbf{q}' \in \mathcal{G}(N)} \Delta^3\mathbf{q} \Delta^3\mathbf{q}' \ f(q, q').$$

The integrand $f(q, q')$ is given by

$$\frac{e^2}{2\mathcal{V}} \sqrt{\frac{1}{\omega_{\mathbf{p}'}\omega_{\mathbf{p}}}} \cdot \delta_{\mathrm{dis}}(p' + q' - q)\ \delta_{\mathrm{dis}}(q - q' - p) \times$$
$$\times \operatorname{tr}(\not{\epsilon}_{s'} S(q')\ \not{\epsilon}_s S(q)).$$

The matrix $S(q)$ can be found in (14.8) on page 884.

Proof. We will similarly argue as for the Compton scattering in Example 14.1. In terms of creation and annihilation operators,

$$\tau = \langle 0|a^-_{\mathbf{p}'}(\mathsf{S}_0 + \mathsf{S}_1 + \mathsf{S}_2)a^+_{\mathbf{p}}|0\rangle$$

with $\mathsf{S}_0 := I$. Set $\tau_j := \langle 0|a^-_{\mathbf{p}'}\mathsf{S}_j a^+_{\mathbf{p}}|0\rangle$. Since the particle states $a^+_{\mathbf{p},s}|0\rangle$ and $a^+_{\mathbf{p}',s'}|0\rangle$ are normalized, $\tau_0 = \delta_{\mathbf{p}'\mathbf{p}}$. By the Wick Theorem 13.2, we get $\tau_1 = 0$, since there appear five (i.e., an odd number) of factors. Again by Theorem 13.2 on page 846, we obtain $\tau_2 = \frac{1}{2}(\tau_{21} + \tau_{22})$ with

$$\tau_{2j} = -e^2 \int_{\Omega\times\Omega} \varrho_{2j}(x,y)d^4x d^4y.$$

Here,

$$\varrho_{21}(x,y) := \langle 0|a^-_{\mathbf{p}'} : \overline{\psi}(x)\gamma_\mu A^\mu(x)\psi(x) : : \overline{\psi}(y)\gamma_\nu A^\nu(y)\psi(y) : a^+_{\mathbf{p}}|0\rangle,$$

and

$$\varrho_{22}(x,y) := \langle 0|a^-_{\mathbf{p}'} : \overline{\psi}(x)\gamma_\mu A^\mu(x)\psi(x) : : \overline{\psi}(y)\gamma_\nu A^\nu(y)\psi(y) : a^+_{\mathbf{p}}|0\rangle.$$

$$\varrho_{22}(x,y) := \langle 0|a^-_{\mathbf{p}'} : \overline{\psi}(x)\gamma_\mu A^\mu(x)\psi(x) : : \overline{\psi}(y)\gamma_\nu A^\nu(y)\psi(y) : a^+_{\mathbf{p}}|0\rangle.$$

Hence

$$\varrho_{21}(x,y) = a^-_{\mathbf{p}'}A^\mu(x)A^\nu(y)a^+_{\mathbf{p}}(\gamma_\mu)_{\alpha\beta}(\gamma_\nu)_{\kappa\lambda}\overline{\psi}^\alpha(x)\psi^\lambda(y)\psi^\beta(x)\overline{\psi}^\lambda(y)$$

and

$$\varrho_{22}(x,y) = a^-_{\mathbf{p}'}A^\nu(y)A^\mu(x)a^+_{\mathbf{p}}(\gamma_\mu)_{\alpha\beta}(\gamma_\nu)_{\kappa\lambda}\overline{\psi}^\alpha(x)\psi^\lambda(y)\psi^\beta(x)\overline{\psi}^\lambda(y).$$

By Table 14.5 on 897, $\overline{\psi}^\alpha(x)\psi^\lambda(y) = -\,\psi^\lambda(y)\overline{\psi}^\alpha(x)$. Hence

$$\overline{\psi}^\alpha(x)\psi^\lambda(y)\psi^\beta(x)\overline{\psi}^\kappa(y) = \psi^\lambda(y)\overline{\psi}^\alpha(x)\overline{\psi}^\kappa(y)\psi^\beta(x).$$

Using the replacement $x \Leftrightarrow y, \mu \Leftrightarrow \nu, \alpha, \beta \Leftrightarrow \kappa, \lambda$, we get $\varrho_{21}(x,y) = \varrho_{22}(y,x)$. This implies $\tau_{12} = \tau_{22}$. Consequently,

$$\boxed{\tau_2 = \tau_{21}.}$$

From Table 14.5 on page 897, we obtain[5]

[5] The minus sign of $-\mathcal{S}^{\lambda\alpha}(y-x)$ is responsible for Furry's rule in the present situation.

$$\varrho_{21}(x,y) = -e^2 \mathcal{N}_{\mathbf{p}'} \mathrm{e}^{\mathrm{i}p'x} e^{\mu}_{r'} (\gamma_\mu)_{\alpha\beta} \times \\ \times \mathcal{S}^{\beta\kappa}(x-y)(\gamma_\nu)_{\kappa\lambda}(-\mathcal{S}^{\lambda\alpha}(y-x)) \mathcal{N}_{\mathbf{p}} e^{\nu}_{r} \mathrm{e}^{-\mathrm{i}py}$$

along with the matrix

$$\mathcal{S}(x) := \int_{\mathbb{R}} dq^0 \sum_{\mathbf{q}\in\mathcal{G}(N)} \Delta^3\mathbf{q}\, S(q) \mathrm{e}^{-\mathrm{i}qx}.$$

Using the trace of matrices, this can be elegantly written as

$$\varrho_{21}(x,y) = e^2 \mathcal{N}_{\mathbf{p}'} \mathcal{N}_{\mathbf{p}} \mathrm{e}^{\mathrm{i}p'x} \mathrm{e}^{-\mathrm{i}py} \cdot \mathrm{tr}(\not{e}_{r'} \mathcal{S}(x-y)\ \not{e}_r \mathcal{S}(y-x)).$$

The point is that the integration over x and y can be carried out explicitly in terms of the discrete Dirac delta function. In fact, it follows from

$$\int_{-T/2}^{T/2} \mathrm{e}^{\mathrm{i}p^0 t} dt \int_{\mathcal{C}(L)} \mathrm{e}^{-\mathrm{i}\mathbf{p}\mathbf{x}} d^3\mathbf{x} = (2\pi)^4 \delta_T(p^0) \cdot \delta_{\mathcal{G}(N)}(\mathbf{p})$$

that

$$\boxed{\frac{1}{(2\pi)^4} \int_{\Omega} \mathrm{e}^{\mathrm{i}px} d^4x = \delta_{\mathrm{dis}}(p).}$$

Hence the integral $\tau_{21} = \int_{\Omega\times\Omega} \varrho_{21}(x,y) d^4x d^4y$ is equal to

$$\int_{\Omega\times\Omega} \mathrm{e}^{\mathrm{i}(p'+q'-q)x}\ \mathrm{e}^{\mathrm{i}(q-q'-p)y} f(q,q',p,p') d^4x d^4y \\ = (2\pi)^8 \delta_{\mathrm{dis}}(p'+q-q') \cdot \delta_{\mathrm{dis}}(q'-q+p) f(q,q',p,p'). \qquad (14.11)$$

□

14.5 The Formal Language of Physicists

The Feynman rules summarized above pass over to the language used in physics in the following way.

(D) Dirac delta function: Replace the discrete Dirac functions $\delta_T, \delta_{\mathcal{G}(N)}$, and

$$\delta_{\mathrm{dis}}(p) = \delta_T(E) \cdot \delta_{\mathcal{G}(N)}$$

by the corresponding formal Dirac functions $E \mapsto \delta(E)$, $\mathbf{p} \mapsto \delta(\mathbf{p})$, and

$$\delta(p) = \delta(E) \cdot \delta(\mathbf{p}).$$

(I) Integrals: Replace the discrete integral $\sum_{\mathbf{p}\in\mathcal{G}(N)} \Delta^3\mathbf{p}...$ by the integral

$$\int_{\mathbb{R}^3} d^3\mathbf{p}...,$$

and replace the integral $\int_{-T}^{T} dt...$ by the limit $\int_{-\infty}^{\infty} dt...$ for $T \to +\infty$.

(S) Squares of the Dirac delta function:

$$\delta(p)^2 = \frac{T\mathcal{V}}{(2\pi)^4} \cdot \delta(p) \qquad \text{for all} \quad p \in \mathbb{R}^4. \qquad (14.12)$$

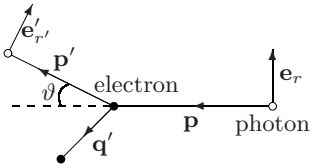


Fig. 14.3. Compton scattering

The point is that this formal translation is nice from the mnemonical point of view, but it leads frequently to meaningless expressions. Physicists repair this defect by using the method of renormalization theory.

Concerning (14.12), the following comment is in order. Laurent Schwartz proved in 1954 that the square of the Dirac delta function does not exist as a distribution. As a standard tool in scattering theory, formula (14.12) is used by physicists in a mnemonic manner. We will show below that this formula is a consequence of our rigorous lattice approach.

14.6 Transition Probabilities and Cross Sections of Scattering Processes

> Physicists measure real numbers in experiments. The theory has to predict these numbers.
>
> Folklore

In particle accelerators, physicists measure cross sections of scattering processes. We want to show how cross sections can be computed by using Feynman diagrams. The idea is to compute first the transition amplitudes by means of the Feynman rules. This way, we get the transition probabilities which imply the cross sections in a quite natural manner. Let us illustrate this by investigating electron-photon scattering (Compton effect).

We consider a fixed inertial system with the right-handed orthonormal basis $\mathbf{i}, \mathbf{j}, \mathbf{k}$.

Compton effect. Let us consider electron-photon scattering in the following special situation (Fig. 14.3):

- We work in the rest system of the incoming electron, i.e., we set $\mathbf{q} = 0$. We assume that the resting electron has the spin vector $s\hbar\mathbf{k}$ with the spin number $s = \pm\frac{1}{2}$. The rest energy of the electron is equal to $m_e c^2$. Since we work in the energetic system, $\hbar = c = 1$.
- The incoming photon moves with the velocity c from right to left along the x-axis. The photon has the momentum vector $\mathbf{p} = -p\mathbf{i}$ with $p > 0$, the energy $\omega_{\mathbf{p}} = p$, and the transversal polarization vector $\mathbf{e}_r$ with $r = 1, 2$. Here, the vectors $\mathbf{e}_1, \mathbf{e}_2$, and $\mathbf{p}/|\mathbf{p}|$ form a right-handed orthonormal system.
- The incoming photon scatters the resting electron. The scattered (outgoing) electron has the momentum vector $\mathbf{q}'$, the energy

$$E_{\mathbf{q}'} = \sqrt{\mathbf{q}'^2 + m_e^2},$$

and the spin vector $s'\hbar\mathbf{k}$ with the spin number $s' = \pm\frac{1}{2}$.

- The outgoing photon has the momentum vector $\mathbf{p}'$, the energy $\omega_{\mathbf{p}'}$, and the polarization vector $\mathbf{e}'_{r'}$ with the polarization number $r' = 1, 2$. The vectors $\mathbf{e}'_{1'}, \mathbf{e}'_{2'}$, and $\mathbf{p}'/|\mathbf{p}'|$ form a right-handed orthonormal system. By definition, the scattering angle ϑ is the angle between the momentum vector $\mathbf{p}$ of the incoming photon and the momentum vector $\mathbf{p}'$ of the outgoing photon where $-\frac{\pi}{2} \le \vartheta \le \frac{\pi}{2}$ (Fig. 14.3 above). Our final goal is the Klein–Nishina formula for the cross section of the Compton scattering to be considered in (15.1) on page 899.

By conservation of momentum and energy, we get

$$\boxed{\mathbf{p}' + \mathbf{q}' = \mathbf{p}, \qquad E_{\mathbf{q}'} + \omega_{\mathbf{p}'} = m_e + \omega_{\mathbf{p}}.} \tag{14.13}$$

For the energy $\omega_{\mathbf{p}'}$ of the scattered photon, we obtain the key relation

$$\boxed{\omega_{\mathbf{p}'} = \frac{\omega_{\mathbf{p}}}{1 + \eta(1 - \cos\vartheta)}} \tag{14.14}$$

with $\eta := (\omega_{\mathbf{p}}/m_e)$. Here, ϑ denotes the scattering angle of the photon (Fig. 14.3 above). Indeed, for the energy of the scattered electron, we obtain

$$E^2_{\mathbf{q}'} = (m_e + \omega_{\mathbf{p}} - \omega_{\mathbf{p}'})^2 = m_e^2 + 2m_e(\omega_{\mathbf{p}} - \omega_{\mathbf{p}'}) + (\omega_{\mathbf{p}} - \omega_{\mathbf{p}'})^2$$

and

$$E^2_{\mathbf{q}'} = m_e^2 + \mathbf{q}'^2 = m_e^2(\mathbf{p} - \mathbf{p}')^2 = m_e^2 + \omega_{\mathbf{p}}^2 + \omega_{\mathbf{p}'}^2 - 2\omega_{\mathbf{p}}\omega_{\mathbf{p}'}\cos\vartheta.$$

Hence $\omega_{\mathbf{p}}\omega_{\mathbf{p}'}(1 - \cos\vartheta) = m_e(\omega_{\mathbf{p}} - \omega_{\mathbf{p}'})$.

Note that we investigate the scattering process in a cubic box $\mathcal{C}(L)$ of length L and volume $\mathcal{V} = L^3$ during the time interval $[-\frac{T}{2}, \frac{T}{2}]$.[6] Moreover, the lattice in momentum space is chosen in such a way that there exists a maximal momentum $P_{\max}$ with

$$|\mathbf{p}| \le P_{\max} \qquad \text{for all} \quad \mathbf{p} \in \mathcal{G}(N).$$

Recall that in Sect. 12.4.3 the construction of the creation operators

$$a^+_{\mathbf{p},r},\ b^+_{\mathbf{q},s},\ a^+_{\mathbf{p}',r'},\ b^+_{\mathbf{q}',s'}$$

for the photons and electrons depends on the choice of the box $\mathcal{C}(L)$. By definition, the state

$$a^+_{\mathbf{p},r}|0\rangle$$

represents precisely the situation where one photon of momentum vector $\mathbf{p}$ and polarization number r is in the box $\mathcal{C}(L)$.

Cross section. In a physical experiment, we consider a fixed resting electron and a homogeneous stream of identical incoming photons described by the current density

$$j_{\text{in}} := c\varrho$$

with the velocity of light, $c = 1$, and the particle density

$$\varrho := \frac{N_{\text{in}}}{\mathcal{V}}.$$

That is, there are N_{in} identical photons with momentum vector $\mathbf{p}$ and polarization number r in the box $\mathcal{C}(L)$. Let N_{out} be the number of scattered photons during the time interval $[-\frac{T}{2}, \frac{T}{2}]$. Define

[6] The volume $\mathcal{V}$ of the box is called the normalization volume.

$$\sigma(\mathcal{V}, T, P_{\max}) := \frac{N_{\text{out}}}{T\mathcal{V}j_{\text{in}}}.$$

This number has the physical dimension of area (in the SI system); it is called the cross section of the scattering process, by physicists. This yields the mnemonic formula

$$\boxed{N_{\text{out}} = \sigma(\mathcal{V}, T, P_{\max}) \cdot T\mathcal{V}j_{\text{in}}.}$$

Transition probability. By definition, the real number

$$\mathcal{W}(T) := |\langle \mathbf{p}', r'; \mathbf{q}', s' | I + \mathsf{S}_1(T) + \mathsf{S}_2(T) | \mathbf{p}, s; \mathbf{0}, s \rangle|^2$$

is called the transition probability of the scattering process during the time interval $[-\frac{T}{2}, \frac{T}{2}]$. Since the contributions of I and S_1 drop out, we get

$$\mathcal{W}(T) := |\langle \mathbf{p}', r'; \mathbf{q}', s' | \mathsf{S}_2(T) | \mathbf{p}, r; \mathbf{0}, s \rangle|^2.$$

For the number of outgoing photons, we obtain

$$N_{\text{out}} = \sum_{\mathbf{p}', \mathbf{q}'} \frac{1}{2} \sum_{s=\pm\frac{1}{2}} \mathcal{W}(T) N_{\text{in}}.$$

Here, we sum over all final particle states and we average over the two possible spin positions $s = \pm\frac{1}{2}$ of the resting electron. More precisely, we use the summation

$$\sum_{\mathbf{p}', \mathbf{q}'} \ldots = \sum_{\mathbf{p}', \mathbf{q}' \in \mathcal{G}(N)} \frac{\mathcal{V}\Delta^3\mathbf{p}'}{h^3} \cdot \frac{\mathcal{V}\Delta^3\mathbf{q}'}{h^3} \ldots \tag{14.15}$$

This is motivated by the following physical argument from quantum statistics. The quantity

$$\sum_{\mathbf{p}' \in \mathcal{G}(N)} \frac{\mathcal{V}\Delta^3\mathbf{p}'}{h^3}$$

is a dimensionless real number; this normalized volume of the phase space is equal to the number of particle states in the corresponding region of the phase space. In addition, note that, in the case of N_{out}, we have to take both the outgoing photons and electrons into account. Recall that $h = 2\pi\hbar = 2\pi$ in the energetic system. Summarizing,

$$N_{\text{out}} = \sum_{\mathbf{p}', \mathbf{q}' \in \mathcal{G}(N)} \frac{\mathcal{V}\Delta^3\mathbf{p}'}{h^3} \cdot \frac{\mathcal{V}\Delta^3\mathbf{q}'}{h^3} \cdot \frac{1}{2} \sum_{s=\frac{1}{2}} \mathcal{W}(T) N_{\text{in}}.$$

This implies the following key formula for the cross section:

$$\boxed{\sigma(\mathcal{V}, T, P_{\max}) = \frac{\mathcal{V}^3}{(2\pi)^6} \sum_{\mathbf{p}', \mathbf{q}' \in \mathcal{G}(N)} a \cdot \Delta^3\mathbf{p}' \Delta^3\mathbf{q}'} \tag{14.16}$$

with

$$a := \frac{1}{2} \sum_{s=\pm\frac{1}{2}} \frac{|\langle \mathbf{p}', r'; \mathbf{q}', s' | \mathsf{S}_2(T) | \mathbf{p}, r; \mathbf{0}, s \rangle|^2}{T}.$$

14.7 The Crucial Limits

Naturally enough, we consider the following three limits:

$$\boxed{L \to +\infty, \qquad P_{\max} \to +\infty, \qquad T \to +\infty.} \tag{14.17}$$

That is, the box $\mathcal{C}(L)$ becomes the 3-dimensional space $\mathbb{R}^3$ (i.e., the normalization volume $\mathcal{V} = L^3$ becomes infinite), the maximal momentum $P_{\max}$ becomes infinite, and we observe the scattering process over the infinite time interval $]-\infty, \infty[$. From the physical point of view, we will regard the following three limits:

(H) High-energy limit: $P_{\max} \to +\infty$.
(L) Low-energy-limit: $\Delta^3\mathbf{p} = \frac{(2\pi)^3}{\mathcal{V}} \to 0$.
(T) Long-time limit: $T \to +\infty$.

In addition, we will carry out the following two limits:

(RE) Regularization of the electron propagator: $\varepsilon \to +0$.
(RP) Regularization of the photon propagator: $m_{ph} \to +0$ (the virtual photon mass goes to zero).

The asymptotic cross section. The cross section $\sigma(\mathcal{V}, T, P_{\max})$ depends on the volume $\mathcal{V}$ of the box, the time interval $[-\frac{T}{2}, \frac{T}{2}]$, and the maximal momentum $P_{\max}$. To free ourselves from this arbitrary choice, we consider the limit (14.17), and we define the asymptotic cross section by

$$\boxed{\sigma := \lim_{\mathcal{V}, T, P_{\max} \to +\infty} \frac{\sigma(\mathcal{V}, T, P_{\max})}{\mathcal{V}T}.}$$

For physical reasons, we expect that this limit exists, at least in some generalized sense.

Example 14.4 For the cross section of the Compton effect, we have

$$\boxed{\sigma = \frac{\alpha^2}{8m_e^2} \int_{|\mathbf{p}'|=1} \left(\frac{\omega_{\mathbf{p}'}}{\omega_{\mathbf{p}}}\right)^2 \sum_{s,s'=\pm\frac{1}{2}} |\overline{u}_{\mathbf{q}',s'} M u_{\mathbf{0},s}|^2 \, d\Omega_{\mathbf{p}'}.} \tag{14.18}$$

Here, we introduce the surface measure differential $d\Omega_{\mathbf{p}'} = 2\pi \cos\vartheta \, d\vartheta$ of the unit sphere, the fine structure constant $\alpha = e^2/4\pi = 1/137.04$, and the matrix

$$M := e_{s'} \frac{\not{q} + \not{p} + m_e}{(q+p)^2 - m_e^2} \, e_s + e_s \frac{\not{q} - \not{p}\,' + m_e}{(q-p')^2 - m_e^2} \, e_{s'}.$$

Naturally enough, we average over the two possible spin positions, $s = \pm\frac{1}{2}$, of the resting electron, and we sum over the two possible spin positions, $s' = \pm\frac{1}{2}$, of the scattered electron. Observe that the asymptotic cross section, σ, s is independent of both the time interval $[-\frac{T}{2}, \frac{T}{2}]$ and the normalization volume $\mathcal{V}$, as expected by physicists.

Motivation. We will use conservation of energy and momentum from (14.13) along with formal arguments based on the calculus for Dirac's delta function. By (14.10) and (14.16), as $T \to +\infty$, we get

$$\sigma(\mathcal{V},T,P_{\max}) = \frac{\alpha^2}{2} \sum_{\mathbf{p}',\mathbf{q}'\in\mathcal{G}(N)} \frac{\Delta^3\mathbf{p}'\Delta^3\mathbf{q}'}{\omega_{\mathbf{p}'}\omega_{\mathbf{p}}E_{\mathbf{q}'}E_{\mathbf{q}}} \times \\ \times\, \delta_{\mathrm{dis}}(p'+q'-p-q)\cdot|\overline{u}_{\mathbf{q}',s'}Mu_{\mathbf{q},s}|(1+o(1)).$$

The three limits $\mathcal{V}, P_{\max}, T \to +\infty$ send sums to integrals (resp. discrete Dirac functions to Dirac functions). Formally, we obtain

$$\sigma = \frac{\alpha^2}{2} \int_{\mathbf{p}',\mathbf{q}'\in\mathbb{R}^3} \frac{d^3\mathbf{p}'d^3\mathbf{q}'}{\omega_{\mathbf{p}'}\omega_{\mathbf{p}}E_{\mathbf{q}'}E_{\mathbf{q}}} \times \\ \times\, \delta(p'+q'-p-q)\cdot|\overline{u}_{\mathbf{q}',s'}Mu_{\mathbf{q},s}|^2.$$

Starting with the decomposition

$$\delta(p'+q'-p-q) = \delta(\mathbf{p}'+\mathbf{q}'-\mathbf{p}-\mathbf{q})\cdot\delta(\omega_{\mathbf{p}'}+E_{\mathbf{q}'}-\omega_{\mathbf{p}}-E_{\mathbf{q}}),$$

where $\mathbf{q}=0$ and $E_{\mathbf{q}}=m_e$, and carrying out the integration over the variable $\mathbf{q}'$, we get

$$\sigma = \frac{\alpha^2}{2}\int_{\mathbb{R}^3} \frac{d^3\mathbf{p}'}{\omega_{\mathbf{p}'}\omega_{\mathbf{p}}E_{\mathbf{q}'}m_e}\cdot\delta(\omega_{\mathbf{p}'}+E_{\mathbf{q}'}-\omega_{\mathbf{p}}-m_e)\cdot|\overline{u}_{\mathbf{q}',s'}Mu_{\mathbf{q},s}|^2.$$

Here, $\mathbf{q}'=\mathbf{p}-\mathbf{p}'$. Using spherical coordinates and noting that $\omega_{\mathbf{p}'}=|\mathbf{p}'|$, we obtain $d^3\mathbf{p}'=\omega_{\mathbf{p}'}d\Omega_{\mathbf{p}'}d\omega_{\mathbf{p}'}$. Hence

$$\sigma = \frac{\alpha^2}{2}\int_0^\infty d\omega_{\mathbf{p}'}\int_{|\mathbf{p}'|=1}\frac{d\Omega_{\mathbf{p}'}}{\omega_{\mathbf{p}'}\omega_{\mathbf{p}}E_{\mathbf{q}'}m_e}\cdot\delta(f(\omega_{\mathbf{p}}))\cdot|\overline{u}_{\mathbf{p}-\mathbf{p}',s'}Mu_{\mathbf{q},s}|^2$$

along with the function

$$f(\omega_{\mathbf{p}'}) := \omega_{\mathbf{p}'}+E_{\mathbf{q}'}-\omega_{\mathbf{p}}-m_e.$$

Furthermore, $E_{\mathbf{q}'}=\sqrt{m_e^2+(\mathbf{p}-\mathbf{p}')^2}=\sqrt{m_e^2+\omega_{\mathbf{p}'}^2+\omega_{\mathbf{p}}^2-\omega_{\mathbf{p}'}\omega_{\mathbf{p}}\cos\vartheta}$. We now use the following key relation for the Dirac delta function:

$$\boxed{\int_0^\infty \delta(f(x))g(x)dx = \frac{g(x_0)}{|f'(x_0)|}.} \tag{14.19}$$

Here, the functions $f,g:[0,\infty]\to\mathbb{R}$ are smooth, and we assume that the positive number x_0 is the only zero of the function f (see Sect. 11.2.4 of Vol. I). By (14.14), the equation $f(\omega_{\mathbf{p}'})=0$ has the unique solution

$$\omega_{\mathbf{p}'} = \frac{1}{1+(\omega_{\mathbf{p}}/m_e)(1-\cos\vartheta)}.$$

For the derivative,

$$f'(\omega_{\mathbf{p}'}) = \frac{E_{\mathbf{q}'}+\omega_{\mathbf{p}'}-\omega_{\mathbf{p}}\cos\vartheta}{E_{\mathbf{q}'}} = \frac{\omega_{\mathbf{p}}+m_e-\omega_{\mathbf{p}}\cos\vartheta}{E_{\mathbf{q}'}} = \frac{\omega_{\mathbf{p}}m_e}{E_{\mathbf{q}'}\omega_{\mathbf{p}'}}.$$

Now the claim (14.18) follows from (14.19).

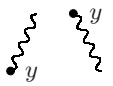
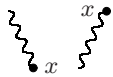
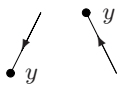
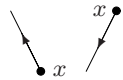
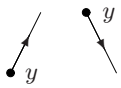
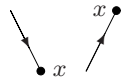
14.8 Appendix: Table of Feynman Rules

In the following two tables, we summarize both the Feynman rules and the formulas for the discrete propagators. The details are explained in Sect. 14.3 on page 882. The reader should note that we will use these tables again and again. In later volumes we will use similar rules for computing processes in the Standard Model in particle physics.

Table 14.4. Feynman propagators (lattice setting)

Regularized Feynman photon propagator (m_{ph} virtual photon mass, $\varepsilon > 0$)
$\mathcal{D}^{\mu\nu}_{F,\varepsilon,m_{ph}}(x) := \int_{\mathbb{R}} dp^0 \sum_{\mathbf{p}\in\mathcal{G}(N)} \Delta^3\mathbf{p}\, D^{\mu\nu}(p)\mathrm{e}^{-\mathrm{i}px}$
$D^{\mu\nu}(p) := -\dfrac{\mathrm{i}\eta^{\mu\nu}}{(2\pi)^4} \cdot \dfrac{1}{(p^0)^2 - (\mathcal{E}_{\mathbf{p}} - \varepsilon\mathrm{i})^2}$
Regularized Feynman electron propagator (m_e bare electron mass, $\varepsilon > 0$)
$\mathcal{S}_{F,\varepsilon}(x) := \int_{\mathbb{R}} dp^0 \sum_{\mathbf{p}\in\mathcal{G}(N)} \Delta^3\mathbf{p}\, S(p)\mathrm{e}^{-\mathrm{i}px}$
$S(p) := \dfrac{\mathrm{i}}{(2\pi)^4} \cdot \dfrac{\not{p} + m_e I}{(p^0)^2 - (E_{\mathbf{p}} - \varepsilon\mathrm{i})^2}$
$p := (p^0, \mathbf{p}), \quad x := (t, \mathbf{x}), \quad px := p^0 t - \mathbf{px}, \quad \not{p} := \gamma_\mu p^\mu,$ $\mathcal{E}_{\mathbf{p}} := \sqrt{\mathbf{p}^2 + m_{ph}^2}\,, \quad E_{\mathbf{p}} := \sqrt{p^2 + m_e^2}$

Table 14.5. Feynman rules for Feynman diagrams (lattice setting)

particle	pairing $\underbrace{AB}$	factor $\mathcal{C}_{\mathcal{T}}(AB)$	symbols
vertex		$\mathrm{i}e(\gamma_\mu)_{\alpha\beta}\int_\Omega d^4x...$	$\bullet\, x$
		$\mathrm{i}e(\gamma_\nu)_{\kappa\lambda}\int_\Omega d^4x...$	$\bullet\, y$
		$x := (t, \mathbf{x}), \quad \Omega := [-T, T] \times \mathcal{C}(L)$	
incoming photon	$\overbrace{A^\nu(y)a^+_{\mathbf{p},r}}$	$\mathcal{N}_{\mathbf{p}}\mathrm{e}^{-\mathrm{i}py}$	
outgoing photon	$\overbrace{a^-_{\mathbf{p},r}A^\mu(x)}$	$\mathcal{N}_{\mathbf{p}}\mathrm{e}^{\mathrm{i}px}e^\mu_r(\mathbf{p})$	
$px := \omega_{\mathbf{p}}t - \mathbf{px}, \quad \omega_{\mathbf{p}} := \lvert\mathbf{p}\rvert, \quad \mathcal{N}_{\mathbf{p}} := \sqrt{\frac{1}{2L^3\omega_{\mathbf{p}}}}, \qquad s = 1, 2$			
incoming electron	$\underbrace{\psi^\lambda(y)b^+_{\mathbf{q},s}}$	$N_{\mathbf{q}}u^\lambda_{\mathbf{q},s}\mathrm{e}^{-\mathrm{i}qy}$	
outgoing electron	$\underbrace{b^-_{\mathbf{q},s}\overline{\psi}^\alpha(x)}$	$N_{\mathbf{q}}\overline{u}^\alpha_{\mathbf{q},s}\mathrm{e}^{\mathrm{i}qx}$	
incoming positron	$\underbrace{\overline{\psi}^\kappa(y)c^+_{\mathbf{q},s}}$	$N_{\mathbf{q}}\overline{v}^\kappa_{\mathbf{q},s}\mathrm{e}^{-\mathrm{i}qy}$	
outgoing positron	$\underbrace{c^-_{\mathbf{q},s}\psi^\beta(x)}$	$N_{\mathbf{q}}v^\beta_{\mathbf{q},s}\mathrm{e}^{\mathrm{i}qx}$	
$qx := E_{\mathbf{q}}t - \mathbf{qx}, \quad E_{\mathbf{q}} := \sqrt{\mathbf{q}^2 + m_e^2}, \quad N_{\mathbf{q}} = \sqrt{\frac{1}{2L^3E_{\mathbf{q}}}}, \qquad \sigma = \pm\frac{1}{2}$			
virtual photon	$\overbrace{A^\mu(x)A^\nu(y)}$	$\mathcal{D}^{\mu\nu}_{F,\varepsilon,m_{ph}}(x-y)$	x ∿∿∿ y
virtual electron	$\underbrace{\psi^\beta(x)\overline{\psi}^\kappa(y)}$	$\mathcal{S}^{\beta\kappa}_{F,\varepsilon}(x-y)$	x ⟵ y
virtual positron	$\underbrace{\overline{\psi}^\alpha(x)\psi^\lambda(y)}$	$-\mathcal{S}^{\lambda\alpha}_{F,\varepsilon}(y-x)$	x ⟶ y
global factor for n vertices	$\frac{1}{n!}$		

15. Applications to Physical Effects

We want to apply the method of Feynman diagrams to the following problems:

- the scattering of electrons and photons (Compton effect);
- scattering of particles in an external electromagnetic field;
- the spontaneous emission of photons by atoms, and
- the Cherenkov effect.

According to our convention formulated on page 790, we will work in the energetic system with $c = \hbar = \varepsilon_0 = \mu_0 := 1$.

15.1 Compton Effect

The Compton effect lies at the heart of modern physics.
Folklore

Let us consider the scattering process between a resting electron and a homogenous stream of incoming photons. We use the notation introduced in Sect. 14.6. The goal is to compute the relation between the (asymptotic) cross section σ and the scattering angle ϑ for the photon (see Fig. 14.3 on page 891). By convention, we choose $-\frac{\pi}{2} \le \vartheta \le \frac{\pi}{2}$. Then

$$d\Omega = 2\pi \cos\vartheta \, d\vartheta$$

represents the surface measure differential on the unit sphere.

The Klein–Nishina formula. For the (asymptotic) cross section of the Compton effect in second order of perturbation theory, we get

$$\boxed{\sigma = \int_{|\mathbf{p}'|=1} f(\vartheta) d\Omega_{\mathbf{p}'}.} \tag{15.1}$$

Here,

$$f(\vartheta) := \frac{r_e^2}{2} \left(\frac{\omega_{\mathbf{p}'}}{\omega_{\mathbf{p}}}\right)^2 \left(\frac{\omega_{\mathbf{p}'}}{\omega_{\mathbf{p}}} + \frac{\omega_{\mathbf{p}}}{\omega_{\mathbf{p}'}} - \sin^2\vartheta\right).$$

This 1929 Klein–Nishina formula is one of the most beautiful formulas in quantum electrodynamics which goes beyond the classical theory of electromagnetism based on the Maxwell equations. In particular, $\alpha = 1/137.04$ is the fine structure constant, m_e is the electron mass, and $r_e = \alpha/m_e$ is the so-called classical electron radius which will be discussed below. The relation between the energies $\omega_{\mathbf{p}}$ and $\omega_{\mathbf{p}'}$ of the incoming and outgoing photons, respectively, is given by

$$\omega_{\mathbf{p}'} = \frac{\omega_{\mathbf{p}}}{1 + \eta(1 - \cos\vartheta)}.$$

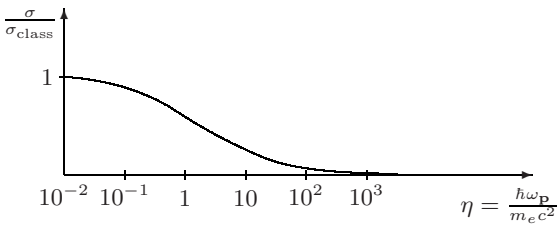


Fig. 15.1. Cross section for the Compton effect

Here, the dimensionless parameter $\eta := \omega_{\mathbf{p}}/m_e$ is the ratio between the energy of the incoming photon and the energy of the resting electron. By convention, the total cross section σ averages over both the spin of the electrons and the polarization of the photons. Naturally enough, we write

$$\boxed{\frac{d\sigma}{d\Omega} = f(\vartheta).}$$

The differential $d\sigma = f(\vartheta)d\Omega$ is called the differential cross section of the Compton effect. Explicitly,

$$\boxed{\sigma = \frac{8}{3}\cdot\pi r_e^2\left\{\left(\frac{3}{8\eta} - \frac{3}{4\eta^2} - \frac{3}{4\eta^3}\right)\ln(1+2\eta) + \frac{3(1+\eta)}{4(1+2\eta)^2} + \frac{6}{4\eta^2}\right\}.}$$

For small photon energies (i.e., small parameters $\eta = \omega_{\mathbf{p}}/m_e$), we get the asymptotic formula[1]

$$\sigma = \frac{8}{3}\cdot\pi r_e^2\left(1 - 2\eta + \frac{312}{60}\,\eta^2 + \dots\right), \qquad \eta \to 0.$$

The zeroth approximation

$$\boxed{\sigma_{\text{class}} = \frac{8}{3}\cdot\pi r_e^2}$$

is a classical formula which was obtained by John Thompson (1856–1940) at the end of the 19th century based on the Maxwell equations.[2] For the scattering of a photon stream, this formula tells us the fundamental result that the electron behaves approximately like a disc of radius r_e given by

$$r_e := \frac{\alpha}{m_e} = \alpha\lambda\!\!\!^{-}_C = \frac{\alpha\lambda_C}{2\pi}.$$

Therefore, about 1900, the quantity r_e was coined the classical electron radius. Moreover, $\lambda_C := 2\pi/m$ (resp. $\lambda\!\!\!^{-}_C = 1/m_e$) is called the Compton wavelength (resp. the reduced Compton wavelength) of the electron. In the SI system,

$$\alpha = \frac{e^2}{4\pi\varepsilon_0\hbar c}, \qquad \lambda_C = \frac{h}{m_e c}, \qquad \lambda\!\!\!^{-}_C = \frac{\hbar}{m_e c}, \qquad r_e = \alpha\lambda\!\!\!^{-}_C, \quad \eta = \frac{\hbar\omega_{\mathbf{p}}}{m_e c^2}.$$

[1] For visible light, $\eta \sim 10^{-5}$.

[2] See J. Jackson, Classical Electrodynamics, Wiley, 1975, Sect. 14.7.

Explicitly, $r_e = 2.818 \cdot 10^{-15}$ m, $\lambda_C = 2.426 \cdot 10^{-12}$ m, and $\alpha = 1/137.04$. For large photon energies, that is, large η, we obtain the asymptotic formula

$$\sigma = \frac{\pi r_e^2}{2} \cdot \frac{1 + 2\ln 2\eta}{\eta} + o\left(\frac{1}{\eta}\right), \qquad \eta \to +\infty.$$

The cross section is pictured in Fig. 15.1.

Proof. (I) Key identity. In Sect. 15.1.2, we will prove the following matrix identity:

$$\boxed{\sum_{s,s'=\pm\frac{1}{2}} |\overline{u}_{\mathbf{q}',s'} M u_{\mathbf{0},s}|^2 = 2\left(\frac{(\omega_{\mathbf{p}'} - \omega_{\mathbf{p}})^2}{\omega_{\mathbf{p}'}\omega_{\mathbf{p}}} + 4(\mathbf{e}_{\mathbf{p}',r'}\mathbf{e}_{\mathbf{p},r})^2\right).} \tag{15.2}$$

Here, $\mathbf{q}' = \mathbf{p} - \mathbf{p}'$ is the momentum vector of the scattered electron. Explicitly, the left-hand side of (15.2) contains thousands of terms. Therefore, we need a clever method for computing this expression. In this connection, physicists invented the elegant trace method for Dirac matrices.

(II) Cross section. Define

$$\sigma_{r',r} = \frac{\alpha^2}{m_e^2} \int_{|\mathbf{p}'|=1} d\Omega_{\mathbf{p}'} \left(\frac{\omega_{\mathbf{p}'}}{\omega_{\mathbf{p}}}\right)^2 \left(\frac{(\omega_{\mathbf{p}'} - \omega_{\mathbf{p}})^2}{4\omega_{\mathbf{p}'}\omega_{\mathbf{p}}} + (\mathbf{e}_{\mathbf{p}',r'}\mathbf{e}_{\mathbf{p},r})^2\right).$$

It follows from (14.18) that the total cross section σ is given by

$$\sigma = \frac{1}{2} \sum_{r,r'=\pm\frac{1}{2}} \sigma_{r,r'}. \tag{15.3}$$

(II) Polarization averaging. In (15.3) we average over the two polarization states of the incoming photon. In addition, we sum over the two polarization states of the outgoing photon. Explicitly, we choose

$$\mathbf{e}_{\mathbf{p}',1} = \mathbf{e}_{\mathbf{p},1} = \frac{\mathbf{p}' \times \mathbf{p}}{|\mathbf{p}' \times \mathbf{p}|}, \quad \mathbf{e}_{\mathbf{p},2} = \mathbf{e}_{\mathbf{p},1} \times \frac{\mathbf{p}}{|\mathbf{p}|}, \quad \mathbf{e}_{\mathbf{p}',2} := \mathbf{e}_{\mathbf{p}',1} \times \frac{\mathbf{p}'}{|\mathbf{p}'|}.$$

An elementary computation yields $\sum_{s,s'=\pm\frac{1}{2}} (\mathbf{e}_{\mathbf{p}',s'}\mathbf{e}_{\mathbf{p},s})^2 = 1 + \cos^2\vartheta$. Hence

$$\sigma = \frac{\alpha^2}{2m_e} \int_{|\mathbf{p}'|=1} d\Omega_{\mathbf{p}'} \left(\frac{\omega_{\mathbf{p}'}}{\omega_{\mathbf{p}}}\right)^2 \left(1 + \cos^2\vartheta + \frac{(\omega_{\mathbf{p}'} - \omega_{\mathbf{p}})^2}{\omega_{\mathbf{p}'}\omega_{\mathbf{p}}}\right).$$

(III) The basic integral. In order to compute the cross section σ, we set

$$\eta := \frac{\omega_{\mathbf{p}}}{m_e}, \quad \zeta := \cos\vartheta, \quad u := \eta(1 - \zeta).$$

By (14.14), we get

$$\frac{\omega_{\mathbf{p}'}}{\omega_{\mathbf{p}}} = \frac{1}{1+u}, \qquad d\Omega_{\mathbf{p}'} = 2\pi d\zeta.$$

Hence

$$\sigma = \frac{\pi\alpha^2}{m_e^2} \int_{-1}^{1} d\zeta \left(\frac{1}{1+u} + \frac{1}{(1+u)^3} - \frac{1-\zeta^2}{(1+u)^2}\right).$$

Passing over to the variable u, we obtain $\sigma = \int_0^{2\eta} f(u)du$ with

$$f(u) := \frac{\pi\alpha^2}{\eta m_e^2}\Big[\frac{1}{\eta^2} + \frac{1}{1+u}\left(1 - \frac{2}{\eta} - \frac{2}{\eta^2}\right) + \frac{1}{(1+u)^2}\left(\frac{2}{\eta} + \frac{1}{\eta^2}\right) + \frac{1}{(1+u)^3}\Big].$$

Elementary integration yields then the desired cross section formula for σ in terms of the dimensionless parameter η. □

15.1.1 Duality between Light Waves and Light Particles in the History of Physics

Light represents the most important physical phenomenon. The experimental and theoretical investigation of light has played a fundamental role in the historical development of physics and mathematics. The light of our sun is also fundamental for the existence of life on earth. Laser light helps in medicine and high technology. For studying the properties of our universe, we rely on analyzing the spectrum of the light coming from stars, galaxies, and distant quasars.

The principle of least action. The reflection law of light was deduced by Heron of Alexandria (100 B.C.). In 1636, there appeared the seminal book *Discourse sur la méthode de bien conduire sa raison* by Descartes (1596–1650).[3] Among other subjects, Descartes' *Discourse* treated analytic geometry and dioptrics. Motivated by Descartes' book, Fermat (1601–1665) invented his *principle of shortest time* for light rays; he used this principle in order to derive Descartes' principle for the refraction of light.

(i) Light as a wave: In his book on the foundations of optics, *Traité de la lumière*, Huygens (1629–1695) had the ingenious idea of replacing Fermat's long-range principle for the propagation of light by a contact principle. In particular, Huygens postulated that light consists of waves.[4]

(ii) Light as particles: Newton (1643–1727) showed that Huygens' wave concept contradicts physical experiments. Therefore, Newton postulated that light consists of particles.[5]

(iii) Diffraction of light: Experimentally, if light passes through a small slit, interference patterns are observed on a screen behind the slit (i.e., the light intensity on the screen varies from point to point). The diffraction of light was thoroughly studied from the experimental and theoretical point of view by Fresnel (1788–1827). He convinced physicists that light possesses a wave character.

(iv) The duality between waves and particles: From the modern point of view, light is a quantum phenomenon. Light quanta are quantum states described by vectors in an appropriate Hilbert space. Quanta possess features of both waves and particles. This is the final answer to the question about the nature of light. In quantum electrodynamics, we start with a classical wave theory based on the Maxwell equations. We add the particle picture by quantizing the classical wave theory with the help of creation and annihilation operators.

[3] In the same year, 1636, the oldest university of the United States of America, the Harvard University, was founded in Cambridge near Boston, Massachusetts.

[4] In about 1870, Lie (1842–1899) created the mathematical theory of contact transformations which represents a far-reaching generalization of Huygens' ideas in terms of mathematics.

[5] At Newton's time, the concept of transversal vector waves was not known to physicists. Newton's argument only excluded longitudinal waves.

In 1744, Euler (1707–1783) founded the calculus of variations as a far-reaching mathematical generalization of Fermat's principle of shortest time for light. Lagrange (1736–1813) extended this to multidimensional integrals. Based on the work of Fermat, Leibniz, Maupertius, Euler, and Lagrange, the final form of the *principle of least action* was formulated by Hamilton (1805–1865).

The idea of physical fields. Gauss (1777–1855) based his theory of magnetism on long-range forces, similarly to Newton's long-range forces in gravitation. Faraday (1791–1867) created the ingenious idea of an electromagnetic field; this way, he replaced Gauss' long-range electromagnetic forces by a contact principle.[6]

Faraday's notion of physical fields is crucial for modern physics.

Based on his fundamental equations for electromagnetism from 1864, Maxwell (1831–1879) postulated that light consists of transversal waves of the electromagnetic field. Electromagnetic waves were experimentally established by Heinrich Hertz (1857–1894), shortly after Maxwell's death. In the second half of the 19th century, John Thomson (1856–1940) experimentally proved the existence of the electron, and he computed the classical cross section for the scattering of electromagnetic waves (light) based on Maxwell's equations.

Light quanta. In 1900, Planck (1858–1947) formulated his famous law for the energy spectrum of black-body radiation. To this end, Planck postulated that the energy of light is quantized. In 1905, Einstein (1879–1955) explained the photoelectric effect by postulating the existence of light quanta;[7] this way, Einstein could derive Planck's radiation law by the methods of statistical physics. In 1917, Einstein formulated his stochastic theory for the emission of light quanta in molecules; this is the basis for modern laser physics.[8] In 1922, the Compton effect was experimentally discovered by Compton (1892–1962). The Klein–Nishina cross section formula for the Compton effect was obtained in 1929.[9] For their contributions to the physics of electrons and light, John Thomson, Planck, Einstein, and Compton were awarded the Nobel prize in physics in the years 1906, 1918, 1921, and 1927, respectively.

Quantum electrodynamics was founded in the late 1940s by Feynman (1918–1979), Schwinger (1918–1994), and Tomonaga (1906–1979); these three physicists were awarded the Nobel prize in physics in 1965.

15.1.2 The Trace Method for Computing Cross Sections

> Computation of cross sections means computation of traces for products of Dirac matrices.
>
> Folklore

Trace properties. Let $A = (a_{ij})$ be a complex-valued $(n \times n)$-matrix. The complex number

[6] In 1915, Einstein replaced Newton's long-range gravitational forces by a contact principle, in the framework of his theory of general relativity.

[7] The notion of photon was coined by the physical chemist Gilbert Lewis in 1926.

[8] LASER stands for *Light Amplification by Stimulated Emission of Radiation*. Laser beams were experimentally produced in about 1960. For their fundamental work in the field of quantum electronics (laser physics), Basov, Prokhorov, and Townes were awarded the Nobel prize in physics in 1964.

[9] O. Klein and Y. Nishina, Z. Phys. **52** (1929), pp. 853ff; pp. 869ff. I. Tamm, Z. Phys. **62** (1930), pp. 545ff.

$$\boxed{\operatorname{tr}(A) := \sum_{i=1}^{n} a_{ii}}$$

is called the trace of the matrix A. In particular, for the $(n \times n)$-unit matrix, $\operatorname{tr}(I) = n$. Let $n = 1, 2, \ldots$

Proposition 15.1 *For complex-valued $(n \times n)$ matrices $A, B, A_1, A_2, \ldots$, the following properties are valid:*

(i) *Homogeneity:* $\operatorname{tr}(\alpha A) = \alpha \operatorname{tr} A$ *for all complex numbers* α.

(ii) *Product property:* $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

(iii) *Duality:* $(\operatorname{tr}(A))^\dagger = \operatorname{tr}(A^\dagger)$. *The trace of a self-adjoint matrix is real.*

(iv) *Cyclic permutations: The trace of a matrix product is invariant under cyclic permutations of the factors:*

$$\operatorname{tr}(A_1 A_2 A_3 \cdots A_N) = \operatorname{tr}(A_2 A_3 \cdots A_N A_1).$$

Moreover, $(\operatorname{tr}(A_1 A_2 A_3 \cdots A_N))^\dagger = \operatorname{tr}(A_N^\dagger A_{N-1}^\dagger \cdots A_1^\dagger)$.

(v) *Invariance:* $\operatorname{tr}(SAS^{-1}) = \operatorname{tr}(A)$ *for each invertible $(n \times n)$-matrix S.*

Trace rules for Dirac matrices. Let $A, B, A_1, A_2 \ldots$ be arbitrary complex (4×4)-matrices, and let $\gamma^0, \gamma^1 \ldots \gamma^5$ be the Dirac matrices introduced in Sect. 11.1 on page 793. For $j = 1, 2, 3$, we have

$$\gamma_0 = \gamma^0, \quad \gamma_j = -\gamma^j, \quad \not{A} := \gamma_\mu A^\mu, \quad \overline{A} := \gamma_0 A^\dagger \gamma_0.$$

In contrast to this definition, we set $\overline{\psi} := \psi^\dagger \gamma^0$ for each complex column matrix ψ which has four rows. The matrices $\overline{A}$ and $\overline{\psi}$ are called the Dirac adjoint of A and ψ, respectively. Recall that $\gamma^5 := \mathrm{i}\gamma^0\gamma^1\gamma^2\gamma^3$.

Example 15.2 For all the indices $\mu, \nu, \kappa, \lambda, \mu_1, \mu_2, \ldots = 0, 1, 2, 3$, the following are met:

(a) $(\gamma^0)^\dagger = \gamma^0$ and $(\gamma^j)^\dagger = -\gamma^j$ for $j = 1, 2, 3$.

(b) $(\overline{\gamma})^\mu = \gamma^\mu$.

(c) We have the crucial Clifford relation $\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2\eta^{\mu\nu}$.

(d) $\gamma^5\gamma^\mu = -\gamma^\mu\gamma^5$ and $\gamma^5\gamma^5 = I$.

(e) For $N = 1, 2, 3, ..$, the following two Furry relations hold:[10]

$$\boxed{\operatorname{tr}(\gamma^{\mu_1}\gamma^{\mu_2}\cdots\gamma^{\mu_N}) = \operatorname{tr}(\gamma^{\mu_N}\cdots\gamma^{\mu_2}\gamma^{\mu_1}),}$$

and

$$\operatorname{tr}(\gamma^{\mu_1}\gamma^{\mu_2}\cdots\gamma^{\mu_N}) = \operatorname{tr}((-\gamma^{\mu_1})(-\gamma^{\mu_2})\cdots(-\gamma^{\mu_N})).$$

(f) For odd $N = 1, 3, 5, \ldots$, we have

$$\boxed{\operatorname{tr}(\gamma^{\mu_1}\cdots\gamma^{\mu_N}) = 0.}$$

This key relation critically simplifies the computation of cross sections.

(g) $\operatorname{tr}(\gamma^\mu\gamma^\nu) = 4\eta^{\mu\nu}$.

(h) For even $N = 4, 6, \ldots$, the computation of the trace can be based on the following recursive formula:

[10] We will show in Problem 15.13 that the Furry relations are responsible for the Furry rule.

$$\operatorname{tr}(\gamma^{\mu_1}\cdots\gamma^{\mu_N}) = \sum_{r=2}^{N}(-1)^r \operatorname{tr}(\gamma^{\mu_1}\gamma^{\mu_r})\cdot\operatorname{tr}(\gamma^{\mu_1}\cdots\gamma^{\mu_{r-1}}\gamma^{\mu_{r+1}}\cdots\gamma^{\mu_N}).$$

For example, this yields

$$\operatorname{tr}(\gamma^\mu\gamma^\nu\gamma^\kappa\gamma^\lambda) = 4(\eta^{\mu\nu}\eta^{\kappa\lambda} - \eta^{\mu\kappa}\eta^{\nu\lambda} + \eta^{\mu\lambda}\eta^{\nu\kappa}).$$

(i) $\operatorname{tr}(\gamma^\mu\gamma^5) = \operatorname{tr}(\gamma^\mu\gamma^\nu\gamma^5) = 0$.

(j) For $N = 3, 4, \ldots$, the computation of the trace can be based on the product formula

$$\gamma^\mu\gamma^\nu\gamma^\kappa = \eta^{\mu\nu}\gamma^\kappa - \eta^{\mu\kappa}\gamma^\nu + \gamma^{\nu\kappa}\gamma^\mu - \mathrm{i}\epsilon^{\mu\nu\kappa\sigma}\gamma_\sigma\gamma^5.$$

For example, $\operatorname{tr}(\gamma^\mu\gamma^\nu\gamma^\kappa\gamma^5) = 0$, and

$$\operatorname{tr}(\gamma^\mu\gamma^\nu\gamma^\kappa\gamma^\lambda\gamma^5) = -4\mathrm{i}\epsilon^{\mu\nu\kappa\lambda}.$$

In particular, $\operatorname{tr}(\gamma^0\gamma^1\gamma^2\gamma^3\gamma^5) = -\mathrm{i}\operatorname{tr}(\gamma^5\gamma^5) = -\mathrm{i}\operatorname{tr} I = -4\mathrm{i}$.

(k) $(\operatorname{tr} A)^\dagger = \operatorname{tr} A^\dagger = \operatorname{tr}\overline{A}$.

(l) For $N = 1, 2, ...$, we get

$$(\operatorname{tr}(A_1A_2\cdots A_N))^\dagger = \operatorname{tr}(A_N^\dagger A_{N-1}^\dagger\cdots A_1^\dagger) = \operatorname{tr}(\overline{A}_N\overline{A}_{N-1}\cdots\overline{A}_1).$$

(m) $\overline{AB} = \overline{B}\cdot\overline{A}$.

The proof can be found in Problem 15.12 on page 942.

Trace rules for Feynman's slash matrices. Let $a = (a^0, a^1, a^2, a^3)$ and $b = (b^0, b^1, b^2, b^3)$ where all the numbers a^μ, b^ν are real. We define[11]

$$\boxed{\not{a} := a_\mu\gamma^\mu, \qquad ab := a_\mu b^\mu.}$$

The following rules are critically used for computing cross sections in quantum field theory.

Proposition 15.3 (T1) *Dirac duality:* $\overline{\not{a}} = \not{a}$ *and* $\overline{\not{a}\not{b}} = \not{b}\not{a}$.

(T2) *Anticommutativity: We have the key relation*[12]

$$\boxed{\not{a}\not{b} = 2ab - \not{b}\not{a}.}$$

(T3) *Reversion and cyclic permutation: For* $N = 2, 3, \ldots,$ *we get*

$$\operatorname{tr}(\not{a}_1\cdots\not{a}_N) = \operatorname{tr}(\not{a}_N\cdots\not{a}_1).$$

In addition, the trace is invariant under cyclic permutations, that is,

$$\operatorname{tr}(\not{a}_1\cdots\not{a}_{N-1}\not{a}_N) = \operatorname{tr}(\not{a}_N\not{a}_1\cdots\not{a}_{N-1}).$$

[11] We sum over equal lower and upper indices from 0 to 3. For example,

$$ab = \sum_{\mu=0}^{3} a_\mu b^\mu = a^0b^0 - \sum_{j=1}^{3} a^j b^j.$$

[12] To simplify notation, the symbol ab stands for the multiple, $(ab)I$, of the unit matrix, I. Note that $\operatorname{tr}(I) = 4$.

(T4) *Odd number of factors: For odd* $N = 1, 3, 5, \ldots$,

$$\operatorname{tr}(\not{a}_1 \cdots \not{a}_N) = 0.$$

In particular, $\operatorname{tr} \not{a} = \operatorname{tr}(\not{a}\, \not{b}\, \not{c}) = 0$.

(T5) *Even number of factors: For two and four factors,*

$$\operatorname{tr}(\not{a}\, \not{b}) = 4ab,$$
$$\frac{1}{4} \operatorname{tr}(\not{a}\, \not{b}\, \not{c}\, \not{d}) = (ab)(cd) - (ac)(bd) + (ad)(bc).$$

Generally, for even $N = 4, 6, 8, \ldots$, *the trace* $\operatorname{tr}(\not{a}_1\ \not{a}_2 \cdots \not{a}_N)$ *is given by the key reduction formula*

$$\sum_{r=2}^{N} (-1)^r (a_1 a_r) \cdot \operatorname{tr}(\not{a}_2 \cdots \not{a}_{r-1}\, \not{a}_{r+1} \cdots \not{a}_N).$$

(T6) *Square trick: If* $\not{a}_r = \not{a}_{r+1}$ *for fixed index* r, *then*

$$\operatorname{tr}(\not{a}_1 \cdots \not{a}_r\ \not{a}_{r+1} \cdots \not{a}_N) = a_r^2 \operatorname{tr}(\not{a}_1 \cdots \not{a}_{r-1}\ \not{a}_{r+2} \cdots \not{a}_N).$$

In particular, this trace vanishes if $a_r^2 = 0$.

(T7) *Special reduction formula: For even* $N = 4, 6, 8, \ldots$, *transposition of the two neighbored factors* $\not{a}_r$ *and* $\not{a}_{r+1}$ *yields*

$$\operatorname{tr}(\not{a}_1 \cdots \not{a}_r\ \not{a}_{r+1} \cdots \not{a}_N) = -\operatorname{tr}(\not{a}_1 \cdots \not{a}_{r+1}\ \not{a}_r \cdots \not{a}_N) + \\ +2(a_r a_{r+1}) \cdot \operatorname{tr}(\not{a}_1 \cdots \not{a}_{r-1}\ \not{a}_{r+2} \cdots \not{a}_N).$$

(T8) *Chirality:* $\operatorname{tr}(\not{a} \cdot \gamma^5) = \operatorname{tr}(\not{a}\, \not{b} \cdot \gamma^5) = 0$, *and*

$$\frac{\mathrm{i}}{4} \operatorname{tr}(\not{a}\, \not{b}\, \not{c}\, \not{d} \cdot \gamma^5) = a \wedge b \wedge c \wedge d = \epsilon^{\mu\nu\kappa\lambda} a_\mu b_\nu c_\kappa d_\lambda.$$

Proof. Ad (T1). $\overline{\not{a}} = a_\mu \overline{\gamma}^\mu = a_\mu \gamma^\mu$.

Ad (T2). By the Clifford relation,

$$\not{a}\, \not{b} + \not{b}\, \not{a} = a_\mu b_\nu [\gamma^\mu, \gamma^\nu]_+ = 2a_\mu b_\nu \eta^{\mu\nu} = 2ab.$$

Hence $\not{a}\, \not{b} = 2ab - \not{b}\, \not{a}$.

Ad (T3), (T4). Apply Prop. 15.2.

Ad(T5). By the Clifford relation,

$$\operatorname{tr}(\not{a}\, \not{b}) = a_\mu b_\nu \operatorname{tr}(\gamma^\mu \gamma^\nu) = 4a_\mu b_\nu \eta^{\mu\nu} = 4a_\mu b^\mu = 4ab.$$

The proof of the general recursive formula can be found in Problem 1.12.

Ad (T6). Note that $\not{a}_r^2 = \frac{1}{2}[\not{a}_r, \not{a}_r]_+ = a_r^2 I$.

Ad (T7). Use (T2). □

The key reduction formula (T5) above allows us to reduce the computation of traces to the computation of inner products. This can be done by using computer algebra, since (T5) represents a universal algorithm. For doing computations by hand, the special reduction formula (T7) above is usefully combined with the square trick. For example,

$$\operatorname{tr}(\not{a}\, \not{b}\, \not{c}\, \not{d}) = 2(ab) \cdot \operatorname{tr}(\not{c}\, \not{d}) - \operatorname{tr}(\not{b}\, \not{a}\, \not{c}\, \not{d}).$$

If $a = c$ and $c^2 = 0$, then the last term vanishes, by the square trick. Thus,

$$\mathrm{tr}(\not a\ \not b\ \not c\ \not d) = 8(ab)(cd).$$

This argument will be frequently used in the proof of Prop. 15.4 below.

The key identity. Using the preceding trace relations, we are going to prove the following identity:

Proposition 15.4 *There holds the following matrix identity:*

$$\sum_{s,s'=\pm\frac{1}{2}} |\overline{u}_{\mathbf{q}',s'} M u_{\mathbf{0},s}|^2 = 2\left(\frac{\omega_{\mathbf{p}'}}{\omega_{\mathbf{p}}} + \frac{\omega_{\mathbf{p}}}{\omega_{\mathbf{p}'}} - 2 + 4(\mathbf{e}_{\mathbf{p}',r'}\mathbf{e}_{\mathbf{p},r})^2\right).$$

Here, $\mathbf{q}' = \mathbf{p} - \mathbf{p}'$ *is the momentum vector of the outgoing electron.*

Proof. (I) Reduction to trace. Observe that the complex number

$$(a^1\ a^2)\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22}\end{pmatrix}\begin{pmatrix} b^1 \\ b^2\end{pmatrix}$$

is equal to the trace of the matrix

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22}\end{pmatrix}\begin{pmatrix} b^1 \\ b^2\end{pmatrix}(a^1\ a^2).$$

Let us suppose that

- A is an $(n \times n)$-matrix,
- a is a $(1 \times n)$-row matrix, and
- b is an $(n \times 1)$-column matrix.

Then

$$\boxed{aAb = \mathrm{tr}(Aba).} \tag{15.4}$$

This is the decisive trick. In fact, $a^i a_{ij} b^j = a_{ij} b^j a^i$.[13] By (15.4), we obtain the key relation

$$\boxed{\left|\overline{u}_{\mathbf{q}',s'} M u_{\mathbf{q},s}\right|^2 = \mathrm{tr}\left(M u_{\mathbf{q},s}\overline{u}_{\mathbf{q},s}\overline{M}u_{\mathbf{q}',s'}\overline{u}_{\mathbf{q}',s'}\right).} \tag{15.5}$$

This follows from $(\overline{u}_{\mathbf{q}',s'})^\dagger = (u^\dagger_{\mathbf{q}',s'}\gamma_0)^\dagger = \gamma_0 u^\dagger_{\mathbf{q}',s'}$ along with

$$\left(\overline{u}_{\mathbf{q}',s'} M u_{\mathbf{q},s}\right)^\dagger = u^\dagger_{\mathbf{q},s} M^\dagger \gamma_0 u_{\mathbf{q}',s'} = u^\dagger_{\mathbf{q},s}\gamma_0(\gamma_0 M^\dagger \gamma_0) u_{\mathbf{q}',s'} = \overline{u}_{\mathbf{q},s}\overline{M}u_{\mathbf{q}',s'},$$

as well as

$$\left|\overline{u}_{\mathbf{q}',s'} M u_{\mathbf{q},s}\right|^2 = \left(\overline{u}_{\mathbf{q}',s'} M u_{\mathbf{q},s}\right)\left(\overline{u}_{\mathbf{q}',s'} M u_{\mathbf{q},s}\right)^\dagger.$$

By Theorem 12.3 on page 807,

$$\sum_{s=\pm\frac{1}{2}} u_{\mathbf{q},s}\overline{u}_{\mathbf{q},s} = \not q + m_e.$$

Consequently, we get

$$\sum_{s,s'=\pm\frac{1}{2}} \left|u_{\mathbf{q}',s'} M u_{\mathbf{q},s}\right|^2 = \mathsf{T}$$

[13] We sum over equal lower and upper indices from 1 to n.

with the trace

$$\mathsf{T} := \mathrm{tr}\left(M(\not{q} + m_e)\overline{M}(\not{q}' + m_e)\right)$$

and $q' = q + p - p'$. To simplify notation, set

$$m := m_e, \quad \omega := \omega_{\mathbf{p},r}, \quad \omega' := \omega_{\mathbf{p}',r'},$$

and $e := e_r$, $e' := e_{r'}$ for $r, r' = 1, 2$.[14] Moreover, choose

$$P := \not{p}, \quad Q := \not{q}, \quad E = \not{e}, \quad E' = \not{e}'.$$

In terms of this notation, it is our goal to compute the trace

$$\mathsf{T} = \mathrm{tr}\left(M(Q + m)\overline{M}(Q' + m)\right)$$

with $Q' = Q + P - P'$, and

$$M = E'\frac{Q + P + m}{(q+p)^2 - m^2}E + E\frac{Q - P' + m}{(q-p')^2 - m^2}E'.$$

The trace T can be computed systematically by using the general reduction formula from Prop. 15.3 on page 905. This is the brute force method.

We are going to simplify critically this approach by using symmetry properties.

(II) Inner products. We want to compute the inner products between the given 4-vectors p, q, p', q', e, and e':

(i) $p^2 = p'^2 = 0$.
(ii) $q^2 = q'^2 = m^2$.
(iii) $pq = m\omega$ and $p'q = m\omega'$.
(iv) $pp' = m(\omega - \omega')$.
(v) $(p+q)^2 - m^2 = 2m\omega$ and $(q-p')^2 - m^2 = -2m\omega'$.
(vi) $e^2 = e'^2 = -1$ and $ee' = -\mathbf{e}\mathbf{e}'$.
(vii) $eq = ep = e'q = 0$ and $e'p' = 0$.
(viii) $e'q' = e'p$ and $q'p = qp' = m\omega'$.

Let us prove this.

Ad (i). By photon energy, $p^2 = (\omega, \mathbf{p})(\omega, \mathbf{p}) = \omega^2 - \mathbf{p}^2 = 0$.
Ad (ii). By electron energy, $q^2 = (m, \mathbf{0})(m, \mathbf{0}) = m^2$.
Ad (iii). $pq = (\omega, \mathbf{p})(m, \mathbf{0}) = m\omega$.
Ad (iv). It follows from

$$0 = q'^2 - m^2 = (q + p - p')^2 - m^2 = q^2 + 2qp - 2qp' - 2pp' - m^2$$

that $pp' = qp - qp' = m(\omega - \omega')$.
Ad (v). $(p+q)^2 - m^2 = p^2 + 2pq + q^2 - m^2 = 2pq = 2m\omega$.
Ad (vi). $e^2 = (0, \mathbf{e})(0, \mathbf{e}) = -\mathbf{e}^2 = -1$, and

$$ee' = (0, \mathbf{e})(0, \mathbf{e}') = -\mathbf{e}\mathbf{e}'.$$

Ad (vii). We have $eq = (0, \mathbf{e})(m, 0) = 0$. Hence, by transversality,

$$ep = (0, \mathbf{e})(\omega, \mathbf{p}) = -\mathbf{e}\mathbf{p} = 0.$$

[14] Only in this proof, the symbol e represents a polarization 4-vector, but not the electron charge.

Ad (viii). We have $e'q' = e'(q + p - p') = e'q$. Using momentum conservation, we get $p + q = p' + q'$. Hence

$$(p - q')^2 = (q - p')^2.$$

Thus, $p^2 + q'^2 - 2pq' = q^2 - 2qp' + p'^2$. □

(III) Anticommutation relations. We claim that the following relations hold:

(a) $[E, Q]_+ = [E', Q]_+ = [E, P]_+ = [E', P']_+ = 0$.
(b) $(Q + m)E(Q + m) = (Q + m)E'(Q + m) = 0$.
(c) $(Q + m)EPE'(Q + m) = 2(pq)E(Q - m)E'$.

Let us prove this.

Ad (a). $[E, Q]_+ = 2eq = 0$.

Ad (b). By (a), $EQ = -QE$. Hence $(Q + m)E(Q + m)$ ist equal to

$$-E(Q - m)(Q + m) = Q^2 - m^2 = \tfrac{1}{2}[Q, Q]_+ - m^2 = q^2 - m^2 = 0.$$

Ad (c). The matrix $(Q + m)EPE'(Q + m)$ is equal to

$$A := -E(Q - m)PE'(Q + m).$$

Using $QP = 2qp - PQ$, we get

$$\begin{aligned} A &= -E\big(2qp - P(Q + m)\big)E'(Q + m) \\ &= -2(qp)EE'(Q + m) = 2(qp)E(Q - m)E'. \end{aligned}$$

□

(IV) Matrix $\overline{M}$. By (II),

$$M = E'\frac{Q + P + m}{2m\omega}E + E\frac{Q - P' + m}{2m\omega'}E'.$$

Setting $A, B = E, E', Q, Q', P, P'$ and using $\overline{AB} = \overline{B} \cdot \overline{A} = B \cdot A$, we obtain

$$\overline{M} = E\frac{Q + P + m}{2m\omega}E' + E'\frac{Q - P' + m}{2m\omega'}E.$$

(V) Matrix N. Define

$$N := \frac{E'PE}{2m\omega} + \frac{EP'E'}{2m\omega'}.$$

Then

$$\overline{N} = \frac{EPE'}{2m\omega} + \frac{E'P'E}{2m\omega'}.$$

By (III)(b), we get

$$\boxed{\mathsf{T} = \operatorname{tr}\big(N(Q + m)\overline{N}(Q' + m)\big)}$$

with $Q' = Q + P - P'$. This tells us that the matrix M can be replaced by the simpler matrix N.

(VI) Decomposition of the trace. Set

$$\boxed{\mathsf{T} = \frac{1}{4m^2}\left(\frac{\mathsf{T}_{11}}{\omega^2} + \frac{\mathsf{T}_{22}}{\omega'^2} + \frac{\mathsf{T}_{12} + \mathsf{T}_{21}}{\omega\omega'}\right)}$$

along with

$$\begin{aligned}
\mathsf{T}_{11} &:= \operatorname{tr} E'PE(Q+m)EPE'(Q'+m),\\
\mathsf{T}_{22} &:= \operatorname{tr} EP'E'(Q+m)E'P'E(Q'+m),\\
\mathsf{T}_{12} &:= \operatorname{tr} E'PE(Q+m)E'P'E(Q'+m),\\
\mathsf{T}_{21} &:= \operatorname{tr} EP'E'(Q+m)EPE'(Q'+m).
\end{aligned}$$

Here, we set $Q' := Q + P - P'$.

(VII) Computation of T_{11}. We use the following decomposition:

$$\mathsf{T}_{11} = \mathsf{T}_{110} + m\mathsf{T}_{111} + m^2\mathsf{T}_{112}.$$

Here, we set

$$\begin{aligned}
\mathsf{T}_{110} &:= \operatorname{tr}(E'PEQEPE'Q'),\\
\mathsf{T}_{112} &:= \operatorname{tr}(E'PEEPE').
\end{aligned}$$

Now we use the trace rules (T1ff) from Prop. 15.3 on page 905. Since T_{111} contains an odd number of factors, $\mathsf{T}_{111} = 0$.

By the square trick (T6) along with $p^2 = 0$,

$$\mathsf{T}_{112} = e^2 \operatorname{tr}(E'P^2E') = e^2p^2 \operatorname{tr}(E'^2) = 0.$$

By the special reduction formula (T7) on page 905 along with $qe = 0$, we get

$$\mathsf{T}_{110} = -\operatorname{tr}(E'PQE^2PE'Q').$$

By the square trick along with $e^2 = -1$,

$$\mathsf{T}_{110} = \operatorname{tr}(E'PQPE'Q').$$

By the special reduction formula,

$$\mathsf{T}_{110} = 2pq \operatorname{tr}(E'PE'Q') - \operatorname{tr}(E'PQP^2E'Q').$$

Since $p^2 = 0$, the last term vanishes. By Prop. 15.3(T5) on page 905,

$$\mathsf{T}_{110} = 8pq\big(2(e'p)(e'q') - e'^2(pq')\big).$$

By (II)(viii),

$$\boxed{\mathsf{T}_{11} = \mathsf{T}_{110} = 16m\omega(\mathbf{pe}')^2 + 8m^2\omega\omega'.}$$

(VIII) Crossing symmetry. The transformation

$$p \;\Leftrightarrow\; -p', \qquad e \;\Leftrightarrow\; e'$$

induces $P \;\Leftrightarrow\; -P'$, $E \;\Leftrightarrow\; E'$. Thus, T_{11} is transformed into T_{22}. Hence

$$\boxed{\mathsf{T}_{22} = -16m\omega'(\mathbf{p}'\mathbf{e})^2 + 8m^2\omega\omega'.}$$

(IX) Reversion symmetry. We want to show that

$$\boxed{\mathsf{T}_{12} = \mathsf{T}_{21}.}$$

Indeed, by Prop. 15.3(T3), the trace of a product of Feynman slash matrices does not change if we reverse the order of the factors. Hence

$$\mathsf{T}_{21} = \operatorname{tr}(Q' + m)E'PE(Q + m)E'P'E.$$

By cyclic permutation,

$$\mathsf{T}_{21} = \operatorname{tr}\big(E'PE(Q + m)E'P'E(Q' + m)\big) = \mathsf{T}_{12}.$$

(X) Computation of T_{12}. Using $Q' = Q + P - P'$, we obtain

$$\begin{aligned}\mathsf{T}_{12} = \operatorname{tr} EP'E'(Q + m)EPE'(Q + m)\\ + \operatorname{tr} EP'E'(Q + m)EPE'P - \operatorname{tr} EP'E'(Q + m)EPE'P'.\end{aligned}$$

In the last two summands, the mass term drops out because of an odd number of factors. Therefore, we get the decomposition

$$\mathsf{T}_{12} = \mathsf{T}_{120} + \mathsf{T}_{121} + \mathsf{T}_{122}.$$

Here, we set

$$\begin{aligned}\mathsf{T}_{120} &:= \operatorname{tr} EP'E'(Q + m)EPE'(Q + m),\\ \mathsf{T}_{121} &:= \operatorname{tr} EP'E'QEPE'P,\\ \mathsf{T}_{122} &:= -\operatorname{tr} EP'E'QEPE'P'.\end{aligned}$$

(XI) Computation of T_{120}. By (III)(c),

$$(Q + m)EPE'(Q + m) = 2m\omega E(Q - m)E'.$$

Hence

$$\mathsf{T}_{120} = 2m\omega \operatorname{tr} EP'E'E(Q - m)E'.$$

The mass terms drops out because of an odd number of factors. Using the relation $EP' = -P'E$, and so on, we get

$$\mathsf{T}_{120} = 2m\omega \operatorname{tr}(EE'P'QEE').$$

Hence

$$\mathsf{T}_{120} = 2m\omega \operatorname{tr}\big((2ee' - E'E)P'QEE'\big).$$

By the square trick (T6), we obtain

$$\operatorname{tr} E'EP'QEE' = \operatorname{tr} P'QE'E^2E' = e^2e'^2 \operatorname{tr} P'Q = 4p'q = 4m\omega'.$$

Since $qe' = qe = 0$, we get

$$\operatorname{tr}(P'QEE') = 4(p'q)(ee') = -4m\omega'(\mathbf{e}\mathbf{e}').$$

This implies

$$\boxed{\mathsf{T}_{120} = 8m^2\omega\omega'\big(2(\mathbf{e}\mathbf{e}')^2 - 1\big).}$$

(XII) Computation of T_{121}. Using the special reduction formula (T7) on page 905, we obtain

$$\mathsf{T}_{121} = 2(e'p) \operatorname{tr}(EP'E'QEP) - \operatorname{tr}(EP'E'QEP^2E').$$

Since $p^2 = 0$, the last term drops out. Using $EP = -PE$ along with cyclic permutation,

$$\begin{aligned}\mathsf{T}_{121} &= -2(e'p) \operatorname{tr}(EP'E'QPE)\\ &= -2(e'p) \operatorname{tr}(P'E'QPE^2) = 2(e'p) \operatorname{tr}(P'E'QP).\end{aligned}$$

Since $e'p' = e'q = 0$,

$$\mathsf{T}_{121} = -8(e'p)^2(p'q) = -8m\omega'(\mathbf{e}'\mathbf{p})^2.$$

Similarly, we obtain

$$\mathsf{T}_{122} = -8m\omega(\mathbf{p}'\mathbf{e})^2.$$

This implies the final relation

$$\boxed{\mathsf{T}_{12} = \mathsf{T}_{21} = 8m^2\omega\omega'\big(2(\mathbf{e}\mathbf{e}')^2 - 1\big) - 8m\omega'(\mathbf{p}\mathbf{e}')^2 + 8m\omega(\mathbf{p}'\mathbf{e})^2.}$$

Summarizing, we get the claim

$$\boxed{\mathsf{T} = 2\left(\frac{\omega'}{\omega} + \frac{\omega}{\omega'} - 2 + 4(\mathbf{e}\mathbf{e}')^2\right).} \tag{15.6}$$

□

This finishes our computation of the cross section for the Compton effect in second order. Despite the use of symmetry, the argument has been lengthy.

It is typical for elementary particle physics that the computations of physical effects are rather involved.

Nowadays sophisticated computer programs are used by physicists (see Sect. 18.4 on page 977).

15.1.3 Relativistic Invariance

Physical processes proceed the same way in each inertial system.
Einstein's 1905 principle of special relativity

Our computation for the cross section of the Compton effect has been carried out in the resting system of the electron. We now want to compute the cross section in an arbitrary inertial system. Analyzing the preceding proof, it turns out that it can be carried out in an relativistically invariant manner by only using inner products between 4-vectors. The trace formula (15.6) can be written as

$$\boxed{\mathsf{T} = 2\left(\frac{p'q}{pq} + \frac{pq}{p'q} - 2 + 4(e_r e_{r'})^2\right).}$$

This is a relativistically invariant formula which coincides with our result in the rest frame. This formula is valid in each inertial system. Here, e and m is the charge and the rest mass of the electron. Moreover, $r_e = \alpha/m$ is the classical electron radius.[15]

The Mandelstam variables. We set

$$\boxed{s := (p+q)^2, \qquad t = (p'-p)^2, \qquad u := (p'-q)^2.}$$

Conservation of 4-momentum tells us that $p + q = p' + q'$. Hence

$$\begin{aligned} s &= (p'+q')^2 = m^2 + 2pq = m^2 + 2p'q', \\ t &= (q'-q)^2 = 2m^2 - 2pp' = -2q'q, \\ u &= (q'-p)^2 = m^2 - 2qp' = m^2 - 2q'p. \end{aligned}$$

[15] To simplify notation, we denote the electron mass m_e by m.

The variables s, t, u are called Mandelstam variables. We have

$$u = 2m^2 - s - t.$$

Thus, the variable u depends on s and t. However, in order to get symmetric formulas, it is convenient to use the three variables u, s, t. The crossing symmetry transformation $p \Leftrightarrow -p'$ corresponds to

$$s \Leftrightarrow u, \qquad t \Leftrightarrow t. \tag{15.7}$$

The relativistically invariant cross section. In second order, the differential cross section for the Compton effect is given by

$$d\sigma = 8\pi r_e^2 \cdot \frac{m^2 dt}{(s-m^2)^2} \left(\frac{m^2}{s-m^2} + \frac{m^2}{u-m^2} \right)^2 \times \\ \times \left(\frac{m^2}{s-m^2} + \frac{m^2}{u-m^2} - \frac{s-m^2}{4(u-m^2)} - \frac{u-m^2}{4(s-m^2)} \right) d\Omega.$$

This formula is invariant under the crossing symmetry transformation (15.7). The corresponding total cross section reads as

$$\sigma = \int_{\mathbb{S}^2} \frac{d\sigma}{d\Omega} \cdot d\Omega.$$

This relativistically invariant formula is valid in each inertial system. In the rest frame of the electron, we get

$$s = m^2 + 2m\omega, \quad u = m^2 - 2m\omega', \quad t = 2\omega\omega'(\cos\vartheta - 1).$$

This yields the famous Klein–Nishina formula:[16]

$$\boxed{\frac{d\sigma}{d\Omega} = \frac{r_e^2}{2} \left(\frac{\omega'}{\omega} \right)^2 \left(\frac{\omega'}{\omega} + \frac{\omega}{\omega'} - \sin^2\vartheta \right).}$$

Equivalently,

$$\frac{d\sigma}{d\Omega} = \frac{r_e^2}{2} \left(\frac{\omega'}{\omega} \right)^2 \left(1 + \cos^2\vartheta + \frac{(w'-w)^2}{\omega\omega'} \right).$$

Here, the incoming (resp. outgoing) photon has the frequency ω (resp. ω'). Moreover, ϑ is the angle between the incoming and the outgoing photon. Explicitly, introducing the dimensionless parameter $\eta := \omega/m$, we get

$$\omega' = \frac{\omega}{1 + \eta\cos\vartheta}.$$

Recall that the Klein–Nishina formula above averages over both the spin of the electrons and the polarization of the photons. Physicists speak of the cross section for the scattering of an electron with unpolarized light.

The classical Thomson scattering as an approximation. For visible light, we have $\eta \sim 10^{-5}$. In this case, the frequency relation $\omega' = \omega$ is a good approximation, and we get the following differential cross section formula:

[16] This formula was obtained in 1929, long before the foundation of quantum electrodynamics in the late 1940s.

$$\frac{d\sigma}{d\Omega} = \frac{r_e^2}{2}\left(1+\cos^2\vartheta\right).$$

For the total cross section, we obtain

$$\sigma = \int_{\mathbb{S}^2} \frac{d\sigma}{d\Omega}\cdot d\Omega = \frac{8}{3}\cdot \pi r_e^2.$$

This is the classical Thomson scattering formula obtained at the end of the 19th century. Recall that r_e is the classical electron radius with $r_e = 2.8\cdot 10^{-15}$ m in the SI system.

15.2 Asymptotically Free Electrons in an External Electromagnetic Field

15.2.1 The Key Formula for the Cross Section

In this section, we work again in the energetic system. Consider a right-handed Cartesian (x,y,z)-coordinate system with right-handed orthonormal basis vectors $\mathbf{i},\mathbf{j},\mathbf{k}$. Suppose a particle P_0 of electric charge Q_0 rests at the origin. This particle generates the electrostatic potential $U = U(\mathbf{x})$ with the Fourier representation

$$U(\mathbf{x}) = \frac{1}{(2\pi)^3}\int_{\mathbb{R}^3} \mathcal{U}(\mathbf{p})\mathrm{e}^{\mathrm{i}\mathbf{p}\mathbf{x}}d^3\mathbf{p}.$$

We want to study the situation where a homogenous stream of incoming particles is scattered at the particle P_0 with the charge Q_0. (Fig. 15.2). We assume that the incoming particles have the rest mass m, the electric charge Q, the velocity vector $\mathbf{v} := v\mathbf{i}$ with $v>0$, the momentum vector $\mathbf{p} = p\mathbf{i}$ with $p = mv/\sqrt{1-v^2}$. Moreover, suppose that the incoming particle stream possesses the current density vector

$$\mathbf{J}_{\mathrm{in}} = \varrho\mathbf{v}$$

with the particle density

$$\varrho_{\mathrm{in}} := \frac{N_{\mathrm{in}}}{\mathcal{V}}.$$

That is, the number of incoming particles is equal to N_{in} in a box of volume $\mathcal{V}$. The momentum vector of the outgoing particles has the form

$$\mathbf{p}' = p'\cos\vartheta\cdot(\cos\varphi\,\mathbf{i}+\sin\varphi\,\mathbf{j}) + p'\sin\vartheta\,\mathbf{k},$$

with respect to spherical coordinates $\pi\in]-\pi,\pi]$ and $\vartheta\in[-\frac{\pi}{2},\frac{\pi}{2}]$. Let Ω_0 be a subset of the unit sphere. We will motivate below that the cross section with respect to the solid angle Ω_0 is given by the formula

$$\boxed{\sigma(\Omega_0) = \int_{\frac{\mathbf{p}'}{|\mathbf{p}'|}\in\Omega_0}\frac{d\sigma}{d\Omega}\cdot d\Omega}$$

along with the differential cross section

$$\boxed{\frac{d\sigma}{d\Omega} = \frac{Q^2}{8\pi^2}\cdot\left|\mathcal{U}\left(\mathbf{p}-\mathbf{p}'\right)\right|^2(2m^2+\mathbf{p}^2+\mathbf{p}\mathbf{p}').} \tag{15.8}$$

Let N_{out} denote the number of scattered particles during the time interval $[t_0,t_1]$ such that the unit vector $\mathbf{p}'/|\mathbf{p}'|$ lives in the set Ω_0. Then

$$\boxed{N_{\mathrm{out}} = \sigma(\Omega_0)J_{\mathrm{in}}(t_1-t_0).}$$

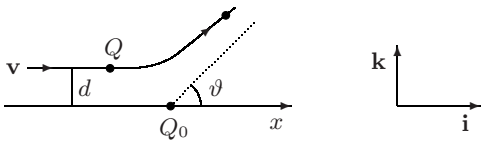


Fig. 15.2. Scattering of a charged homogeneous particle stream

15.2.2 Application to Yukawa Scattering

Let us first choose the Yukawa potential

$$U(\mathbf{x}) := \frac{Q_0}{4\pi} \cdot \frac{\mathrm{e}^{-\kappa r}}{r}.$$

Here, $r_0 = 1/\kappa$ is the typical range of the Yukawa potential. For the Fourier transform of the Yukawa potential, we get

$$\mathcal{U}(\mathbf{p}) = \frac{Q_0}{\mathbf{p}^2 + \kappa^2}.$$

This yields the following differential cross section:

$$\frac{d\sigma}{d\Omega} = \left(\frac{QQ_0m}{2\pi}\right)^2 \cdot \frac{1 + \frac{\mathbf{p}^2 + \mathbf{p}\mathbf{p}'}{4m^2}}{(\kappa^2 + (\mathbf{p} - \mathbf{p}')^2)^2}.$$

15.2.3 Application to Coulomb Scattering

The Coulomb potential corresponds to the parameter $\kappa = 0$. In the non-relativistic case, the momentum vector is given by $\mathbf{p} = m\mathbf{v}$ (resp. $\mathbf{p}' = m\mathbf{v}'$) and $p \ll m$ (resp. $p' \ll m$).[17] Moreover, conservation of kinetic energy tells us that $\frac{1}{2}m_0\mathbf{v}^2 = \frac{1}{2}m_0\mathbf{v}'^2$. Hence $\mathbf{p}^2 = \mathbf{p}'^2$, and

$$(\mathbf{p} - \mathbf{p}')^2 = \mathbf{p}^2 + \mathbf{p}'^2 - 2\mathbf{p}\mathbf{p}' = 2p^2(1 - \cos\vartheta) = 4p^2 \sin^2 \frac{\vartheta}{2}.$$

This implies the following classical Rutherford formula for Coulomb scattering:

$$\frac{d\sigma}{d\Omega} = \left(\frac{QQ_0m}{2\pi}\right)^2 \cdot \frac{1}{(\mathbf{p} - \mathbf{p}')^4} = \left(\frac{QQ_0}{8\pi m v^2}\right)^2 \cdot \frac{1}{\sin^4 \frac{\vartheta}{2}}.$$

In particular, for the scattering of a stream of electrons at a proton,

$$\boxed{\frac{d\sigma}{d\Omega} = \left(\frac{\alpha}{4E_{\mathrm{kin}}}\right)^2 \cdot \frac{1}{\sin^4 \frac{\vartheta}{2}}.}$$

Here $\alpha = e^2/4\pi = 1/137.04$ denotes the fine structure constant. Furthermore, $E_{\mathrm{kin}} = \frac{1}{2}mv^2$ represents the kinetic energy of an incoming electron. This differential cross section has a singularity at the angle $\vartheta = 0$ (no scattering). This is responsible for the divergence of the following integral

$$\sigma = \int_{\mathbb{S}^2} \frac{d\sigma}{d\Omega} \cdot d\Omega = +\infty$$

which represents the total cross section of Coulomb scattering.

[17] Note that, in the SI system, the momentum is small compared with mc.

15.2.4 Motivation of the Key Formula via S-Matrix

The transition amplitude. The idea is to use the transition amplitude from (14.1) by replacing the photon field by the 4-potential A^0, A^1, A^2, A^3 of an external electromagnetic field. The incoming particle is described by the normalized state

$$\Phi_{\text{in}} := b^+_{\mathbf{p},s}|0\rangle,$$

and the outgoing particle corresponds to the normalized state

$$\Phi_{\text{out}} := b^+_{\mathbf{p}',s'}|0\rangle.$$

The transition probability from the state Φ_{in} to the state Φ_{out} during the time interval $[-\frac{T}{2}, \frac{T}{2}]$ is given by

$$\boxed{\mathcal{W}(T) := |\langle p'|\mathsf{S}_N(T)|p\rangle|^2} \tag{15.9}$$

along with the transition amplitude

$$\langle p'|\mathsf{S}_N(T)|p\rangle := \langle \Phi_{\text{out}}|\mathsf{S}_N(T)\Phi_{\text{in}}\rangle = \langle 0|b^-_{\mathbf{p}',s'}|\mathsf{S}_N(T)b^+_{\mathbf{p},s}|0\rangle.$$

In Nth order approximation, Dyson's S-matrix reads as

$$\boxed{\mathsf{S}(T) = \mathsf{S}_0 + \sum_{n=1}^{N} \mathsf{S}_n(T)} \tag{15.10}$$

with $\mathsf{S}_0 := I$. Explicitly, we obtain

$$\mathsf{S}_1(T) := -\mathrm{i}\int_{-T/2}^{T/2} H_{\text{int}}(t)dt.$$

For $n = 2, 3, \ldots$, we get

$$\mathsf{S}_n(T) := \frac{(-\mathrm{i})^n}{n!}\int_{-T/2}^{T/2}\cdots\int_{-T/2}^{T/2} \mathcal{T}(H_{\text{int}}(t_1)\cdots H_{\text{int}}(t_n))dt_1\cdots dt_n,$$

along with

$$H_{\text{int}}(t) := \int_{\mathcal{C}(L)} :\mathcal{H}(\mathbf{x},t): d^3\mathbf{x}$$

and

$$\boxed{\mathcal{H}(x) := Q\overline{\psi}(x)\gamma_\mu A^\mu(x)\psi(x).} \tag{15.11}$$

Here, $A^\mu = A^\mu(x)$, $\mu = 0, 1, 2, 3$, denotes the 4-potential of the external electromagnetic field, and $\psi = \psi(x)$ denotes the free electron-positron field with $x = (\mathbf{x}, t)$.

This setting is obtained by completely replacing the photon quantum field by the external electromagnetic field.

Note that this is only an approximation of reality. This approach completely ignores the possible quantum effects caused by photons (see Sect. 15.2.5).

Concerning (15.11), the electron has the electric charge $Q = -e$. However, this expression is also valid for incoming particles with arbitrary electric charge Q. For example, the case $Q = e$ corresponds to incoming protons. Let us discuss some approximations.

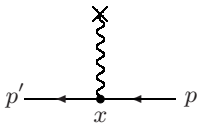


Fig. 15.3. First-order Feynman diagram for an external electromagnetic field

(i) Zero-order approximation: We assume that the incoming state Φ_{in} is different from the outgoing state Φ_{out}. Hence

$$\langle p'|\mathsf{S}_0|p\rangle = \langle \Phi_{\text{out}}|\Phi_{\text{in}}\rangle = 0.$$

(ii) First-order approximation: By the Main Wick Theorem 13.2 on page 846, we have

$$\langle p'|\mathsf{S}_1|p\rangle = \int_{-T/2}^{T/2} dt \int_{\mathcal{C}(L)} d^3\mathbf{x}\ \varrho_1(\mathbf{x},t)$$

with

$$\varrho_1(\mathbf{x},t) := \mathrm{i}Q\langle 0|\underbrace{b^-_{\mathbf{p}',s'} : \overline{\psi}}(x)\gamma_\mu A^\mu(x)\underbrace{\psi(x) : b^+_{\mathbf{p},s}}|0\rangle.$$

Table 14.5 on page 897 tells us that

$$\boxed{\varrho_1(x) = N_{\mathbf{p}'}\overline{u}_{\mathbf{p},s}\gamma_\mu A^\mu(\mathbf{x},t)N_{\mathbf{p}}u_{\mathbf{p},s}\mathrm{e}^{\mathrm{i}(\mathbf{p}-\mathbf{p}')\mathbf{x}}.} \tag{15.12}$$

The normalization factors are given by

$$N_{\mathbf{p}'} := \sqrt{\frac{1}{2E_{\mathbf{p}'}\mathcal{V}}}, \qquad N_{\mathbf{p}} := \sqrt{\frac{1}{2E_{\mathbf{p}}\mathcal{V}}}, \qquad E_{\mathbf{p}} := \sqrt{\mathbf{p}^2+m^2}.$$

(iii) Second order: Again by the Wick theorem, we get

$$\langle p'|\mathsf{S}_2|p\rangle = \frac{1}{2}\int_{-T/2}^{T/2}\int_{-T/2}^{T/2} dtds \int_{\mathcal{C}(L)\times\mathcal{C}(L)} d^3\mathbf{x}d^3\mathbf{y}\ \varrho_2(\mathbf{x},t,\mathbf{y},s)$$

with $\varrho_2 = \varrho_{21} + \varrho_{22}$. Here, $\varrho_{21}(x,y)$ is equal to

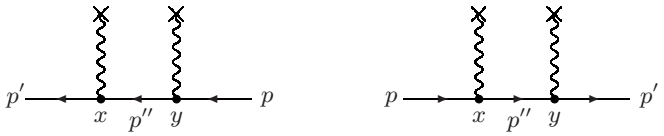
$$-Q^2\langle 0|\underbrace{b^-_{\mathbf{p}',s'} : \overline{\psi}}(x)\gamma_\mu A^\mu(x)\underbrace{\psi(x) :: \overline{\psi}}(y)\gamma_\mu A^\mu(x)\underbrace{\psi(x) : b^+_{\mathbf{p},s}}|0\rangle.$$

Moreover, $\varrho_{22}(x,y)$ is equal to

$$-Q^2\langle 0|\underbrace{b^-_{\mathbf{p}',s'} : \overline{\psi}}(x)\gamma_\mu A^\mu(x)\psi(x) : \overline{\psi}(y)\gamma_\nu A^\nu(y)\psi(y) : b^+_{\mathbf{p},s}}|0\rangle.$$

By Table 14.5 on page 897, the quantity $\varrho_{21}(x,y)$ reads explicitly as

$$-\frac{Q^2}{N_{\mathbf{p}}N_{\mathbf{p}'}}\cdot\overline{u}^\alpha_{\mathbf{p}',s'}(\gamma_\mu)_{\alpha\beta}A^\mu(x)S^{\beta\kappa}_{F,\varepsilon}(x-y)(\gamma_\nu)_{\kappa\lambda}A^\nu(y)u^\lambda_{\mathbf{p},s}\ \mathrm{e}^{\mathrm{i}(p'x-py)}.$$



(a) ϱ_{21} (b) ϱ_{22}

Fig. 15.4. Second-order Feynman diagrams for an external electromagnetic field

In the language of matrices, we obtain

$$\boxed{\varrho_{21}(x,y) = -\frac{Q^2}{N_{\mathbf{p}} N_{\mathbf{p}'}} \cdot \overline{u}_{\mathbf{p}',s'} \gamma_\mu A^\mu(x) \mathcal{S}_{F,\varepsilon}(x-y) \gamma_\nu A^\nu(y) u_{\mathbf{p},s} \mathrm{e}^{\mathrm{i}(p'x-py)}.}$$

Moreover, $\varrho_{22}(x,y)$ reads explicitly as

$$-\frac{Q^2}{N_{\mathbf{p}} N_{\mathbf{p}'}} \cdot \overline{u}^{\kappa}_{\mathbf{p}',s'} (\gamma_\mu)_{\alpha\beta} A^\mu(x) \mathcal{S}^{\lambda\alpha}_{F,\varepsilon}(y-x) u^{\beta}_{\mathbf{p}',s'} (\gamma_\nu)_{\kappa\lambda} A^\nu(y) u^{\beta}_{\mathbf{p},s} \, \mathrm{e}^{\mathrm{i}(p'y-px)}.$$

The transformation $x, \mu, \alpha, \beta \Leftrightarrow y, \nu, \kappa, \lambda$ shows us that

$$\boxed{\varrho_{21}(x,y) = \varrho_{22}(y,x).} \tag{15.13}$$

The corresponding Feynman diagrams can be found in Figs. 15.3 and 15.4.

The cross section. Let us compute the cross section in first-order approximation. This is related to ϱ_1 computed in (15.12).We will proceed similarly to the Compton effect above. The cross section σ is given by

$$N_{\mathrm{out}} = \sigma J_{\mathrm{in}} T$$

with $J_{\mathrm{in}} = \varrho v = v/\mathcal{V}$. Here, N_{out} is the number of outgoing particles during the time interval $[-\frac{T}{2}, \frac{T}{2}]$. Explicitly,

$$N_{\mathrm{out}} = \sum_{\mathbf{p}' \in \mathcal{G}(N)} \frac{\mathcal{V} \Delta^3 \mathbf{p}'}{(2\pi)^3} \cdot \frac{1}{2} \sum_{s,s'=\pm\frac{1}{2}} |\langle p' | \mathsf{S}_1 | p \rangle|^2.$$

Here, we count the number of outgoing particles by using the cell quantization of the phase space, as explained in (14.15) on page 893. Furthermore, we average over the spin states of the incoming particles, and we sum over the spin states of the outgoing particles. For the total cross section, we get

$$\boxed{\sigma = \sum_{\mathbf{p}' \in \mathcal{G}(N)} \frac{\mathcal{V}^2 \Delta^3 \mathbf{p}'}{2(2\pi)^3 \, vT} \sum_{s,s'=\pm\frac{1}{2}} |\langle p' | \mathsf{S}_1 | p \rangle|^2.} \tag{15.14}$$

It remains to compute the square $|\langle p' | \mathsf{S}_1 | p \rangle|^2$.

The transition amplitude. Assume that the external electromagnetic field has the period L with respect to the Cartesian variables x, y, z, and it does not

depend on time t. Then, the four-potential allows the representation in terms of the following Fourier series:

$$A^\mu(\mathbf{x}) = \frac{1}{(2\pi)^3} \sum_{\mathbf{q}\in\mathcal{G}(N)} \Delta^3\mathbf{q}\, \mathcal{A}^\mu(\mathbf{q})\, \mathrm{e}^{\mathrm{i}\mathbf{q}\mathbf{x}}, \qquad \mu = 0,1,2,3.$$

For the transition amplitude, this implies the following formula:

$$\langle p'|\mathsf{S}_1|p\rangle = -\frac{\mathrm{i}Q}{(2\pi)^3 N_{\mathbf{p}'} N_{\mathbf{p}}} \int_{-T/2}^{T/2} dt\, \mathrm{e}^{\mathrm{i}(E_{\mathbf{p}'}-E_{\mathbf{p}})t} \times \\ \times \int_{\mathcal{C}(L)} d^3\mathbf{x}\, \mathrm{e}^{\mathrm{i}(\mathbf{q}+\mathbf{p}-\mathbf{p}')\mathbf{x}} \sum_{\mathbf{q}\in\mathcal{G}(N)} \Delta^3\mathbf{q} \cdot \overline{u}_{\mathbf{p}',s'}\, \mathcal{A}(\mathbf{q}) u_{\mathbf{p},s}.$$

By definition of the discrete Dirac delta function $\delta_{\mathcal{G}(N)}$, we have[18]

$$\frac{1}{(2\pi)^3} \int_{\mathcal{C}(L)} d^3\mathbf{x}\, \mathrm{e}^{\mathrm{i}(\mathbf{q}+\mathbf{p}-\mathbf{p}')\mathbf{x}} = \delta_{\mathcal{G}(N)}(\mathbf{q}+\mathbf{p}-\mathbf{p}').$$

Therefore, the transition amplitude $\langle p'|\mathsf{S}_1|p\rangle$ is equal to

$$-\frac{2\pi\mathrm{i}Q}{N_{\mathbf{p}'}N_{\mathbf{p}}} \cdot \delta_T(E_{\mathbf{p}'}-E_{\mathbf{p}}) \sum_{\mathbf{q}\in\mathcal{G}(N)} \Delta^3\mathbf{q} \cdot \delta_{\mathcal{G}(N)}(\mathbf{q}+\mathbf{p}-\mathbf{p}') \cdot \overline{u}_{\mathbf{p}',s'}\, \mathcal{A}(\mathbf{q}) u_{\mathbf{p},s}.$$

Hence

$$\langle p'|\mathsf{S}_1|p\rangle = -\frac{2\pi\mathrm{i}Q}{N_{\mathbf{p}'}N_{\mathbf{p}}} \cdot \delta_T(E_{\mathbf{p}'}-E_{\mathbf{p}}) \cdot \overline{u}_{\mathbf{p}',s'}\, \mathcal{A}(\mathbf{p}'-\mathbf{p}) u_{\mathbf{p},s}.$$

For large time T it follows from Sect. 12.1.2 of Vol. I that

$$\delta_T(E_{\mathbf{p}'}-E_{\mathbf{p}})^2 = \frac{T}{2\pi} \cdot \delta_T(E_{\mathbf{p}'}-E_{\mathbf{p}}) + o(T), \qquad T\to+\infty.$$

Therefore,

$$|\langle p'|\mathsf{S}_1|p\rangle|^2 = \frac{Q^2 T}{2\mathcal{V}^2 E_{\mathbf{p}'}E_{\mathbf{p}}} \cdot |\overline{u}_{\mathbf{p}',s'}\, \mathcal{A}(\mathbf{p}'-\mathbf{p}) u_{\mathbf{p},s}|^2 + o(T), \quad T\to+\infty.$$

The trace trick. Introduce the sum

$$a(\mathbf{p}',\mathbf{p}) := \sum_{s,s'=\pm\frac{1}{2}} |\overline{u}_{\mathbf{p}',s'}\, \mathcal{A}(\mathbf{p}'-\mathbf{p}) u_{\mathbf{p},s}|^2.$$

By the trace trick from the proof to Prop. 15.4 on page 907, we obtain

$$a(\mathbf{p}',\mathbf{p}) = \sum_{s,s'=\pm\frac{1}{2}} \mathrm{tr}(\mathcal{A}(\mathbf{p}'-\mathbf{p}) u_{\mathbf{p},s}\overline{u}_{\mathbf{p},s}\, \mathcal{A}(\mathbf{p}'-\mathbf{p}) u_{\mathbf{p}',s'}\overline{u}_{\mathbf{p}',s'} \\ = \mathrm{tr}(\mathcal{A}(\mathbf{p}'-\mathbf{p})(\not{p}+m)\, \mathcal{A}(\mathbf{p}'-\mathbf{p})(\not{p}'+m)).$$

[18] In what follows, we will use the properties of the discrete Dirac delta functions $\delta_{\mathcal{G}(N)}$ and δ_T proved in Sect. 12.1.2 of Vol. I.

The Dirac delta function and energy conservation. Set $\wp := |\mathbf{p}|$ and $\wp' := |\mathbf{p}'|$. It follows from

$$\mathbf{p} = \frac{m\mathbf{v}}{\sqrt{1-v^2}}, \qquad E_{\mathbf{p}} = \sqrt{\wp^2 + m^2}$$

that $\wp^2(1-v^2) = m^2v^2$. Hence $\wp^2 = E_{\mathbf{p}}^2 v^2$. This implies

$$\frac{dE_{\mathbf{p}}}{d\wp} = \frac{\wp}{E_{\mathbf{p}}} = v.$$

Similarly,

$$\frac{dE_{\mathbf{p}'}}{d\wp'} = \frac{\wp'}{E_{\mathbf{p}'}} = v'.$$

By (15.14), we obtain

$$\sigma = \frac{Q^2}{8\pi^2} \sum_{\mathbf{p}' \in \mathcal{G}(N)} \frac{\Delta^3 \mathbf{p}'}{v E_{\mathbf{p}'} E_{\mathbf{p}}} \cdot \delta_T(E_{\mathbf{p}'} - E_{\mathbf{p}}) a(\mathbf{p}', \mathbf{p}).$$

Carrying out the limits $T \to +\infty$ and $L \to +\infty$, both the box $\mathcal{C}(L)$ and the lattice $\mathcal{G}(N)$ go to $\mathbb{R}^3$. Thus, we formally obtain

$$\sigma = \frac{Q^2}{8\pi^2} \int_{\mathbb{R}^3} \frac{d^3\mathbf{p}'}{v E_{\mathbf{p}'} E_{\mathbf{p}}} \cdot \delta(E_{\mathbf{p}'} - E_{\mathbf{p}})\ a(\mathbf{p}', \mathbf{p}).$$

Using spherical coordinates, we get

$$d^3\mathbf{p} = d\Omega_{\mathbf{p}'} \wp'^2 d\wp' = \wp' E_{\mathbf{p}'} d\Omega_{\mathbf{p}'} dE_{\mathbf{p}'}.$$

This implies

$$\frac{d^3\mathbf{p}}{v E_{\mathbf{p}'} E_{\mathbf{p}}} = \frac{\wp'}{\wp} \cdot d\Omega_{\mathbf{p}'} dE_{\mathbf{p}}.$$

Hence

$$\sigma = \frac{Q^2}{8\pi^2} \int_0^\infty dE_{\mathbf{p}'} \delta(E_{\mathbf{p}'} - E_{\mathbf{p}}) \int_{|\mathbf{p}'|=1} \frac{\wp' d\Omega_{\mathbf{p}'}}{\wp}\ \ a(\mathbf{p}', \mathbf{p}).$$

By definition of the Dirac delta function, we have to set

$$\boxed{E_{\mathbf{p}'} = E_{\mathbf{p}}}$$

which corresponds to energy conservation. This implies $\wp' = \wp$. Hence

$$\sigma = \frac{Q^2}{8\pi^2} \int_{|\mathbf{p}'|=1} d\Omega_{\mathbf{p}'}\ a(\mathbf{p}', \mathbf{p}).$$

The final cross section formula. Summarizing, for the total cross section, we obtain

$$\sigma = \frac{Q^2}{8\pi^2} \int_{\mathbb{S}^2} d\Omega_{\mathbf{p}'}\ \operatorname{tr}(\mathcal{A}(\mathbf{p}' - \mathbf{p})(\not{p} + m)\, \mathcal{A}(\mathbf{p}' - \mathbf{p})(\not{p}' + m)).$$

Similarly, the cross section with respect to the solid angle Ω_0 reads as

$$\sigma(\Omega_0) = \frac{Q^2}{8\pi^2} \int_{\Omega_0} d\Omega_{\mathbf{p}'} \ \mathrm{tr}(\not{\mathcal{A}}(\mathbf{p}' - \mathbf{p})(\not{p} + m)\, \not{\mathcal{A}}(\mathbf{p}' - \mathbf{p})(\not{p}' + m)).$$

The trace can be computed by using the rules (T3), (T4) from Prop. 15.3 on page 905. Explicitly,

$$\boxed{\sigma(\Omega_0) = \frac{Q^2}{8\pi^2} \int_{\Omega_0} d\Omega_{\mathbf{p}'} (m^2 \mathcal{A}\mathcal{A}^\dagger + (p\mathcal{A}^\dagger)p'\mathcal{A} + (p\mathcal{A})p'\mathcal{A}^\dagger - (p'p)\mathcal{A}\mathcal{A}^\dagger).}$$

Here, the four-vector $\mathcal{A}$ depends on $\mathbf{p}' - \mathbf{p}$, and we integrate over $\mathbf{p}'$. Furthermore,

$$p\mathcal{A} = E_{\mathbf{p}} \mathcal{U}(\mathbf{p}' - \mathbf{p}) - \mathbf{p}\boldsymbol{\mathcal{A}}(\mathbf{p}' - \mathbf{p}).$$

Special case. Let us consider the special case where $\mathcal{A} = (\mathcal{U}, \mathbf{0})$. This corresponds to an electric field with the potential U. By energy conservation, $E_{\mathbf{p}} = E_{\mathbf{p}'}$. Then

$$(pA)p'A^\dagger = E_{\mathbf{p}}^2 \cdot |\mathcal{U}(\mathbf{p}' - \mathbf{p})|^2 = (m^2 + \mathbf{p}^2) \cdot |\mathcal{U}(\mathbf{p}' - \mathbf{p})|^2.$$

This way, we get the desired cross section:

$$\boxed{\sigma(\Omega_0) = \frac{Q^2}{8\pi^2} \int_{\Omega_0} d\Omega_{\mathbf{p}'} \cdot |\mathcal{U}(\mathbf{p}' - \mathbf{p})|^2 (2m^2 + \mathbf{p}^2 + \mathbf{p}\mathbf{p}').}$$

15.2.5 Perspectives

External electromagnetic fields and Hilbert–Schmidt operators. An extensive rigorous functional analytic investigation of electrons and positrons in an external electromagnetic field can be found in

> G. Scharf, Finite Quantum Electrodynamics: the Causal Approach, Springer, Berlin, 1995, Sect. 2.4.

Here, the theory makes essentially use of compact Hilbert–Schmidt operators in Hilbert spaces. We also refer to:

> R. Seiler, Quantum theory of particles with spin zero and one half in external fields, Commun. Math. Phys. **25** (1972), 127–151.

The reader should observe that this nice functional-analytic approach fails for general photon fields in quantum electrodynamics.

Radiative corrections. The S-matrix introduced in (15.10) on page 916 only incompletely describes the electron in an external electromagnetic field. In the language of physicists, the full description in terms of quantum electrodynamics has to take into account the interaction of the electron with virtual photons, virtual electrons, and virtual positrons (vacuum fluctuations). This means that, besides the Feynman graphs depicted in Figs. 15.3 and 15.4 on page 917, there appear additional crucial Feyman graphs as depicted in Fig. 17.8 on page 960. This leads to so-called radiative corrections based on renormalization theory. Such processes are responsible for the anomalous magnetic moment of the electron (see Chap. 17). Here, theory and experiment coincide with extremely high accuracy. In the next section, we will use first-order perturbation theory in order to compute a physical effect which is well known from quantum mechanics. In this connection, the methods of renormalization theory are not necessary.

15.3 Bound Electrons in an External Electromagnetic Field

> We shall consider the transitions produced in an atomic system by an arbitrary perturbation. The method we shall adopt will be that previously given by the author, which leads in a simple way to equations which determine the probability of the system being in any stationary state of the unperturbed system at any time. This, of course, gives immediately the probable number of systems in that state at that time for an assembly of the systems that are independent of one another and are all perturbed in the same way.
>
> Paul Dirac, 1927
> *The quantum theory of the emission and absorption of radiation*[19]

Bound states of an electron in an atom. In an arbitrary, but fixed inertial system, set $x = (\mathbf{x}, t)$. Let us start with the Dirac equation

$$\gamma^\mu(\mathrm{i}\partial_\mu + eA_\mu)\psi = m_e\psi \tag{15.15}$$

for an electron of electric charge $-e$ and mass m_e which is bound in an atom under the influence of the time-independent electrostatic potential

$$U = U(\mathbf{x}).$$

The function U describes the electric forces caused by the atomic nucleus and the remaining electrons. This corresponds to the 4-potential $A_0 := U$ and $A_j := 0$ for $j = 1, 2, 3$. Suppose that there exist solutions

$$\boxed{\psi(\mathbf{x}, t) = \varphi_n(\mathbf{x})\mathrm{e}^{-\mathrm{i}E_n t}, \qquad n = 1, 2, \dots} \tag{15.16}$$

of the Dirac equation (15.15). These solutions describe bound states of the electron having the energy E_n. We postulate that the following normalization condition is satisfied:

$$\int_{\mathbb{R}^3} |\varphi_n(\mathbf{x})|^2 \, d^3\mathbf{x} = 1.$$

15.3.1 The Spontaneous Emission of Photons by the Atom

The physical experiment shows that atoms radiate photons. This is caused by random jumps of the electrons of the atom. Such jumps represent typical quantum processes. Explicitly, if the electron jumps from the energy level E_n to the lower energy level $E_{n'}$, then one photon is emitted which has the energy

$$\boxed{\omega' = E_n - E_{n'}.}$$

Suppose that there are N atoms in a cubic box $\mathcal{C}(L)$ of volume $\mathcal{V}$. Let N_{out} (resp. E_{out}) be the number of spontaneously emitted photons (resp. the emitted energy) during the time interval $[-\frac{T}{2}, \frac{T}{2}]$. Then

[19] Proceedings of the Royal Society of London, Series A, Vol. **114** (1927), 243–265. This is the first paper reprinted in a collection of 34 fundamental papers on quantum electrodynamics edited by J. Schwinger, Quantum Electrodynamics, Dover, New York, 1958.

$$N_{\text{out}} = \tau N T,$$

and

$$\boxed{E_{\text{out}} = \omega' N_{\text{out}}.}$$

Here, we introduce the coefficient τ which is called the emission rate of the spontaneous radiation. We will motivate below that

$$\boxed{\tau = \frac{e^2}{8\pi^2} \int_{|\mathbf{p}'|=1} d\Omega_{\mathbf{p}'} \cdot (E_n - E_{n'}) \sum_{r=1}^{2} |(\varphi_{n'}|\gamma_0\gamma_\mu e_r^\mu \mathrm{e}^{-\mathrm{i}\mathbf{p}'\mathbf{x}}|\varphi_n)|^2} \tag{15.17}$$

along with

$$(\varphi_{n'}|\gamma_0\gamma_\mu e_r^\mu \mathrm{e}^{-\mathrm{i}\mathbf{p}'\mathbf{x}}|\varphi_n) := \int_{\mathbb{R}^3} d^3\mathbf{x}\, \varphi_{n'}(\mathbf{x})^\dagger \cdot \gamma_0\gamma_\mu e_r^\mu \mathrm{e}^{-\mathrm{i}\mathbf{p}'\mathbf{x}} \cdot \varphi_n(\mathbf{x}).$$

In this connection, note that the 4-potential

$$A_{ph}^\mu(x) := \frac{e_r^\mu}{\mathcal{N}_{\mathbf{p}'}} \cdot \mathrm{e}^{\mathrm{i}p'x}, \qquad \mu = 0,1,2,3, \quad r = 1,2$$

describes one photon in the box $\mathcal{C}(L)$. The photon has the momentum vector $\mathbf{p}'$, the energy $\omega_{\mathbf{p}'} = |\mathbf{p}'|$, and the polarization vector $\mathbf{e}_r$ with

$$\mathbf{e}_r = e_r^1\mathbf{i} + e_r^2\mathbf{j} + e_r^3\mathbf{k}, \qquad r = 1,2.$$

Here, the three vectors $\mathbf{e}_1, \mathbf{e}_2$, and $\mathbf{p}'/|\mathbf{p}'|$ form a right-handed orthonormal basis. Moreover, $e_r^0 := 0$. We set $p'x := \omega_{\mathbf{p}'}t - \mathbf{p}'\mathbf{x}$. The normalization factor reads as

$$\mathcal{N}_{\mathbf{p}'} := \sqrt{\frac{1}{2\mathcal{V}\omega_{\mathbf{p}'}}}.$$

15.3.2 Motivation of the Key Formula

> The trick is to replace the free photon quantum field by the wave functions of bound electrons.
>
> Folklore

We define the transition amplitude from the bound electron state ψ_n with energy E_n to the bound electron state $\psi_{n'}$ with energy $E_{n'}$ and the photon state A_{ph} with 4-momentum p' by setting

$$\langle E_{n'}p'|\mathsf{S}_1(T)|E_n\rangle := -\mathrm{i}\int_{-T/2}^{T/2} dt\; H_{\text{int}}(t),$$

where

$$H_{\text{int}}(t) := -e\int_{\mathcal{C}(L)} d^3\mathbf{x}\; \overline{\psi}(\mathbf{x},t)\gamma_\mu A_{ph}(\mathbf{x},t)\psi(\mathbf{x},t).$$

Here, the function ψ is taken from (15.16). This ansatz for H_{int} is motivated by (15.11) on page 916. For the total transition probability during the time interval $[-\frac{T}{2}, \frac{T}{2}]$, we get

$$W(T) = \sum_{\mathbf{p}' \in \mathcal{G}(N)} \frac{\mathcal{V}\Delta^3\mathbf{p}'}{(2\pi)^3} \sum_{r=1,2} |\langle E_{n'} p' | \mathsf{S}_1(T) | E_n \rangle|^2.$$

As usual, the states of the outgoing photons are counted according to the cell quantization of the phase space, as explained in Sect. 14.15 on page 893. This implies the *emission rate*

$$\boxed{\tau = \lim_{T \to +\infty} \frac{W(T)}{T}.}$$

Our goal is to compute this limit. To simplify notation, let us introduce the following symbols:

$$\alpha_{n'n,r} := \int_{\mathcal{C}(L)} d^3\mathbf{x}\ \varphi_{n'}(\mathbf{x})^\dagger \gamma_0 \gamma_\mu e_r^\mu\ \mathrm{e}^{-\mathrm{i}\mathbf{p}'\mathbf{x}} \varphi_n(\mathbf{x}),$$

and

$$a_{n'n,r} := \int_{\mathbb{R}^3} d^3\mathbf{x}\ \varphi_{n'}(\mathbf{x})^\dagger \gamma_0 \gamma_\mu e_r^\mu\ \mathrm{e}^{-\mathrm{i}\mathbf{p}'\mathbf{x}} \varphi_n(\mathbf{x}).$$

Thus, the transition amplitude $\langle E_{n'} p' | \mathsf{S}_1(T) | E_n \rangle$ is equal to

$$\mathrm{i}e \int_{-T/2}^{T/2} dt\ \mathrm{e}^{\mathrm{i}(E_{n'} - E_n + \omega')t}\ \frac{\alpha_{n'n,r}}{\sqrt{2\mathcal{V}\omega'}} = 2\pi \mathrm{i} e\ \delta_T(E_{n'} - E_n + \omega')\ \ \frac{\alpha_{n'n,r}}{\sqrt{2\mathcal{V}\omega'}}.$$

Since $2\pi\delta_T(E_{n'} - E_n + \omega')^2 = T\ \delta_T(E_{n'} - E_n + \omega')$ by Sect. 12.1.2 of Vol. I, we obtain

$$\frac{W(T)}{T} = \frac{e^2}{8\pi^2} \sum_{\mathbf{p}' \in \mathcal{G}(N)} \frac{\Delta^3\mathbf{p}'}{\omega'}\ \delta_T(E_{n'} - E_n + \omega') \sum_{r=1}^{2} |\alpha_{n'n,r}|^2.$$

Carrying out the limits $T \to +\infty$ and $L \to +\infty$, both the box $\mathcal{C}(L)$ and the lattice $\mathcal{G}(N)$ go to the total space $\mathbb{R}^3$. Formally, we obtain

$$\tau = \frac{e^2}{8\pi^2} \int_{\mathbb{R}^3} \frac{d^3\mathbf{p}'}{\omega'}\ \delta(E_{n'} - E_n + \omega') \sum_{r=1}^{2} |a_{n'n,r}|^2.$$

Using spherical coordinates, we have $d^3\mathbf{p}' = d\Omega_{\mathbf{p}'} \omega'^2 d\omega'$. Hence

$$\tau = \frac{e^2}{8\pi^2} \int_0^\infty d\omega'\ \omega'\ \delta(E_{n'} - E_n + \omega') \int_{|\mathbf{p}'|=1} d\Omega_{\mathbf{p}'}\ \sum_{r=1}^{2} |a_{n'n,r}|^2.$$

The Dirac delta function forces us to set $\omega' = E_n - E_{n'}$ which corresponds to energy conservation. Explicitly,

$$\tau = \frac{e^2}{8\pi^2} \int_{|\mathbf{p}'|=1} d\Omega_{\mathbf{p}'} (E_n - E_{n'}) \sum_{r=1}^{2} |a_{n'n,r}|^2.$$

This is the claim.

15.3.3 Intensity of Spectral Lines

The emitted mean energy power. The intensity of a spectral line depends on the radiated energy E_{out}. Suppose that there are N atoms in a sufficiently large box $\mathcal{C}(L)$. Physicists introduce the quantity

$$\eta := \frac{E_{\text{out}}}{T} \cdot \frac{1}{N} = \tau\omega'.$$

This is the emitted photon power of angular frequency $\omega' = E_n - E_{n'}$ per atom. Recall that

$$\tau = \frac{e^2}{8\pi^2} \int_{|\mathbf{p}'|=1} d\Omega_{\mathbf{p}'}\ \omega' \sum_{r=1}^{2} |\langle \varphi_{n'} | \boldsymbol{\alpha} \mathbf{e}_r \cdot \varphi_n \rangle|^2.$$

By definition, $\boldsymbol{\alpha} := \gamma_0\gamma^1\mathbf{i} + \gamma_0\gamma^2\mathbf{i} + \gamma_0\gamma^3\mathbf{k}$.

Heisenberg's famous radiation formula. As an approximation, we get

$$\boxed{\eta = \frac{e^2}{3\pi} \cdot \omega'^4 \cdot |\mathbf{x}_{n'n}|^2} \tag{15.18}$$

along with the dipol moment $\mathbf{x}_{n'n} := \langle \varphi_{n'} | \mathbf{x} \varphi_n \rangle = \int_{\mathbb{R}^3} d^3\mathbf{x}\ \varphi_{n'}(\mathbf{x})^\dagger \mathbf{x} \varphi_n(\mathbf{x})$.

Motivation. We want to motivate (15.18). The wave functions φ_n and φ'_n of a bound electron in an atom are concentrated approximately in a ball about the origin of Bohr radius $r_0 = 10^{-8}$ cm. The basic idea of our approximation is the fact that the wave length $\lambda \sim 10^{-4}$ cm of visible light is much larger than the Bohr radius r_0. Since $|\mathbf{p}'| = 2\pi/\lambda$ is small compared with r_0, the integral

$$\langle \varphi_{n'} | \boldsymbol{\alpha} \mathbf{e}_r \cdot \mathrm{e}^{-\mathrm{i}\mathbf{p}'\mathbf{x}} \varphi_n \rangle = \int_{\mathbb{R}^3} d^3\mathbf{x}\ \varphi_{n'}(\mathbf{x})^\dagger \boldsymbol{\alpha} \mathbf{e}_r \cdot \mathrm{e}^{-\mathrm{i}\mathbf{p}'\mathbf{x}} \varphi_n(\mathbf{x})$$

can be approximately computed by setting $\mathrm{e}^{-\mathrm{i}\mathbf{p}'\mathbf{x}} \simeq 1$. Hence

$$\boxed{\langle \varphi_{n'} | \boldsymbol{\alpha} \mathbf{e}_r \cdot \mathrm{e}^{-\mathrm{i}\mathbf{p}'\mathbf{x}} \varphi_n \rangle \simeq \langle \varphi_{n'} | \boldsymbol{\alpha} \mathbf{e}_r \cdot \varphi_n \rangle.}$$

To compute this, we introduce the momentum operator $\mathbf{P} = -\mathrm{i}\boldsymbol{\partial}$ and the Hamiltonian

$$H := \gamma_0 m_e + \alpha \mathbf{P} - eU.$$

Then, the Dirac equation for the electron can be written as

$$\mathrm{i}\dot{\psi} = H\psi.$$

In particular, $H\varphi_n = E_n\varphi_n$ and $H\varphi_{n'} = E_{n'}\varphi_{n'}$. By Heisenberg's commutation relation, $[x_j, -\mathrm{i}\partial_k]_- = \mathrm{i}\delta_{jk}$. Hence

$$[\mathbf{x}, H]_- = [\mathbf{x}, \boldsymbol{\alpha}\mathbf{P}]_- = \mathrm{i}\boldsymbol{\alpha}.$$

Consequently, the inner product $\mathrm{i}\langle \varphi_{n'} | \mathbf{e}_r \boldsymbol{\alpha} \cdot \varphi_n \rangle$ is equal to

$$\begin{aligned} \langle \varphi_{n'} | \mathbf{e}_r [\mathbf{x}, H]_- \varphi_n \rangle &= \langle \varphi_{n'} | \mathbf{e}_r \mathbf{x} \cdot H\varphi_n - \mathbf{e}_r H(\mathbf{x}\varphi_n) \rangle \\ &= \langle \varphi_{n'} | \mathbf{e}_r \mathbf{x} \cdot H\varphi_n \rangle - \langle H\varphi_{n'} | \mathbf{e}_r \mathbf{x} \cdot \varphi_n \rangle = (E_n - E_{n'}) \langle \varphi_{n'} | \mathbf{e}_r \mathbf{x} \cdot \varphi_n \rangle. \end{aligned}$$

Therefore,

$$\boxed{|\langle \varphi_{n'} | \mathbf{e}_r \boldsymbol{\alpha} \cdot \varphi_n \rangle|^2 = \omega'^2\ |\langle \varphi_{n'} | \mathbf{e}_r \mathbf{x} \cdot \varphi_n \rangle|^2.}$$

This implies the decomposition $\tau = \tau_1 + \tau_2$ by setting

$$\tau_r := \frac{e^2\omega'^3}{8\pi^2} \int_{|\mathbf{p}'|=1} d\Omega_{\mathbf{p}'} \, |\langle \varphi_{n'} | \mathbf{e}_r \mathbf{x} \cdot \varphi_n \rangle|^2.$$

Now we want to average over the polarization vector $\mathbf{e}_1$. To this end, fix the momentum vector $\mathbf{p}'$, and choose the vector $\mathbf{e}_1$ in such a way that it is perpendicular to $\mathbf{p}'$. Using spherical coordinates, we have

$$\mathbf{e}_1 = \cos\vartheta(\cos\varphi\, \mathbf{i} + \sin\varphi\, \mathbf{j}) + \sin\vartheta\, \mathbf{k}.$$

Define the special transition amplitudes

$$a_x := \langle \varphi_{n'} | x \varphi_n \rangle, \quad a_y := \langle \varphi_{n'} | y \varphi_n \rangle, \quad a_z := \langle \varphi_{n'} | z \varphi_n \rangle,$$

and set $b_{xy} := a_x a_y^\dagger + a_x^\dagger a_y$. Then

$$|\langle \varphi_{n'} | \mathbf{x} \varphi_n \rangle|^2 = |a_x|^2 + |a_y|^2 + |a_z|^2.$$

Moreover,

$$\mathbf{e}_1 \mathbf{x} = x \cos\vartheta \cos\varphi + y \cos\vartheta \sin\varphi + z \sin\vartheta.$$

Thus, $|(\varphi_{n'} | \langle \mathbf{e}_1 \mathbf{x}) \varphi_n \rangle|^2$ is equal to

$$\begin{aligned} &|a_x|^2 \cos^2\vartheta \cos^2\varphi + |a_y|^2 \cos^2\vartheta \sin^2\varphi + |a_z|^2 \sin^2\vartheta \\ &\quad + b_{xy} \cos^2\vartheta \cos\varphi \sin\varphi + b_{xz} \cos\vartheta \sin\vartheta \cos\varphi + b_{yz} \cos\vartheta \sin\vartheta \sin\varphi. \end{aligned}$$

Noting that

$$\int_{|\mathbf{p}'|=1} d\Omega_{\mathbf{p}'} \ldots = \int_{-\pi/2}^{\pi/2} d\vartheta \cos\vartheta \int_{-\pi}^{\pi} d\varphi \ldots$$

and using the orthogonality relations for trigonometric functions like

$$\int_{-\pi}^{\pi} \cos\varphi \, d\varphi = 0, \qquad \int_{-\pi}^{\pi} \cos^2\varphi \, d\varphi = \pi,$$

we get the key relation

$$\tau_1 = \frac{e^2\omega'^3}{8\pi^2} \cdot \frac{4\pi}{3} \, (|a_x|^2 + |a_y|^2 + |a_z|^2) = \frac{e^2\omega'^3}{6\pi} \, |\langle \varphi_{n'} | \mathbf{x} \varphi_n \rangle|^2.$$

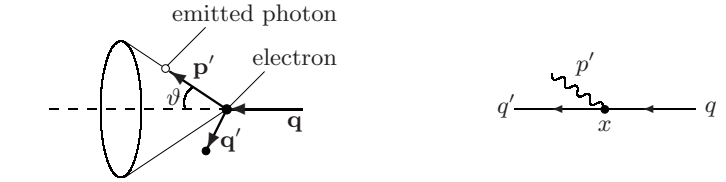
Interchanging the polarization vector $\mathbf{e}_1$ with $\mathbf{e}_2$, we obtain $\tau_2 = \tau_1$. Finally, we get $\eta = \tau\omega' = 2\tau_1\omega'$.

15.4 Cherenkov Radiation

In 1934, Pavel Cherenkov discovered a new kind of radiation called Cherenkov radiation.[20] This is a radiation in the form of bluish light, produced by charged particles moving in a transparent medium.

> *Einstein's principle of special relativity postulates that, in a vacuum, the speed of massive particles is less than the speed of light.*

[20] For the discovery and the interpretation of the Cherenkov effect, Cherenkov, Frank, and Tamm were awarded the Noble prize in physics in 1958.



(a) (b) Feynman diagram

Fig. 15.5. Cherenkov effect

The point is that the situation may change if we replace the vacuum by special media. In fact, there exist media where $c_m < c$, i.e., the speed of light, c_m, in the medium is less than the speed of light, c, in a vacuum. In this case, it is possible that massive particles (e.g., electrons) move with a velocity v which satisfies the key inequality

$$\boxed{c_m < v < c.} \tag{15.19}$$

The condition $v < c$ is forced by Einstein's principle of special relativity. The point is that the particle speed v is greater than the speed of light c_m in the medium. If the inequality (15.19) is satisfied, then the Cherenkov effect appears. The Cherenkov radiation is emitted in a cone around the direction in which the particle is travelling (Fig. 15.5(a)).

Let us work in the energetic system of units, i.e., we set $c = \hbar = \varepsilon_0 := 1$. We will use the following notation:

(i) Medium
- velocity of light in the medium: $0 < c_m < 1$;
- refraction index of the medium: $n := 1/c_m > 1$.

(ii) Incoming electron
- rest mass m_e;
- velocity vector: $\mathbf{v} = -v\mathbf{i}$ with $c_m < v < 1$;
- momentum vector: $\mathbf{q} = m_e\mathbf{v}/\sqrt{1-v^2}$; moreover, $v = |\mathbf{q}|/E_{\mathbf{q}}$;
- energy: $E_{\mathbf{q}} = \sqrt{m_e^2 + \mathbf{q}^2}$;
- 4-momentum vector: $q = (\mathbf{q}, E_{\mathbf{q}})$.

(iii) Emitted photon
- momentum vector $\mathbf{p}'$;
- energy (frequency): $\omega_{\mathbf{p}'} = c_m|\mathbf{p}'|$;
- 4-momentum: $p' = (\mathbf{p}', \omega_{\mathbf{p}'})$;
- the angle between the incoming electron and the emitted photon is denoted by ϑ.

(iv) Outgoing electron after the emission of the photon
- momentum vector: $\mathbf{q}' = \mathbf{q} - \mathbf{p}'$;
- energy: $E_{\mathbf{q}'} = \sqrt{m_e^2 + \mathbf{q}'^2} = E_{\mathbf{q}} - \omega_{\mathbf{p}'}$;
- 4-momentum vector: $q' = (\mathbf{q}', E_{\mathbf{q}'})$.

The emission angle ϑ of the photon is restricted by the following inequality:

$$\boxed{0 \le \vartheta < \vartheta_{\max}.}$$

Here, the maximal angle is determined by the equation

$$\cos\vartheta_{\max} = \frac{c_m}{v}, \qquad 0 < \vartheta_{\max} < \frac{\pi}{2}.$$

If there are N incoming electrons, then the total energy of the emitted photons during the sufficiently large time interval $[-\frac{T}{2}, \frac{T}{2}]$ is equal to

$$\boxed{E_{\text{emission}} = \tau N T.}$$

The emission rate is given by

$$\tau = \alpha \cdot \frac{2m_e v^2 c_m}{(1-c_m^2)\sqrt{1-v^2}} \int_0^{\vartheta_{\max}} d\vartheta \ \sin\vartheta \left(\sin^2\vartheta + \frac{2(\cos\vartheta - \cos\vartheta_{\max})^2}{1-c_m^2} \right)$$

where $\alpha = e^2/4\pi = 1/137.04$ denotes the fine structure constant. This formula for τ shows the angle distribution of the emitted photon energy, too. Computation of the integral yields

$$\boxed{\tau = \alpha \cdot \frac{2m_e c_m (v-c_m)^2}{(1-c_m^2)\sqrt{1-v^2}} \left(1 + \frac{v-c_m}{3v} \cdot \frac{1+c_m^2}{1-c_m^2} \right).} \tag{15.20}$$

Motivation for the emission angle. By energy-momentum conservation,

$$\boxed{q' + p' = q.}$$

Thus, $\mathbf{q}' + \mathbf{p}' = \mathbf{q}$ and $E_{\mathbf{q}'} + \omega_{\mathbf{p}'} = E_{\mathbf{q}}$. It follows from $E_{\mathbf{q}'}^2 = (E_{\mathbf{q}} - \omega_{\mathbf{p}'})^2$ that

$$m_e^2 + (\mathbf{p}' - \mathbf{q})^2 = E_{\mathbf{q}}^2 - 2\omega_{\mathbf{p}'} E_{\mathbf{q}} + \omega_{\mathbf{p}'}^2.$$

This implies $\mathbf{p}'^2 - 2\mathbf{p}'\mathbf{q} = -2c_m|\mathbf{p}'|E_{\mathbf{q}} + c_m^2\mathbf{p}'^2$. Consequently,

$$|\mathbf{p}'|(|\mathbf{p}'| - 2|\mathbf{q}|\cos\vartheta + 2c_m E_{\mathbf{q}} - c_m^2|\mathbf{p}'|) = 0.$$

Assuming that the momentum vector $\mathbf{p}'$ of the photon is nontrivial, we get

$$|\mathbf{p}'|(1-c_m^2) - 2|\mathbf{q}|\cos\vartheta + 2c_m E_{\mathbf{q}} = 0.$$

This yields the key relation

$$\boxed{|\mathbf{p}'| = \frac{2(|\mathbf{q}|\cos\vartheta - c_m E_{\mathbf{q}})}{1-c_m^2}.} \tag{15.21}$$

Since $|\mathbf{p}'| > 0$, the angle ϑ satisfies the following condition

$$\cos\vartheta > \frac{c_m E_{\mathbf{q}}}{|\mathbf{q}|} = \frac{c_m}{v}.$$

This shows that $0 \le \vartheta < \vartheta_{\max}$ along with $\cos\vartheta_{\max} = c_m/v$.

Motivation for the emission rate. We will use similar arguments as in the preceding examples. Set $x := (\mathbf{x}, t)$. The Cherenkov effect corresponds to a passage from the incoming electron state

$$\Phi_{\text{in}} := b_{\mathbf{q},s}^+ |0\rangle$$

to the outgoing electron-photon state

$$\Phi_{\text{out}} := b^+_{\mathbf{q}'s'} a^+_{\mathbf{p}',r'} |0\rangle.$$

In first-order approximation, the corresponding transition amplitude during the time interval $[-\frac{T}{2}, \frac{T}{2}]$ is given by

$$\boxed{\langle q'p'|I + \mathsf{S}_1(T)|q\rangle := \langle \Phi_{\text{out}}|I + \mathsf{S}_1(T)|\Phi_{\text{in}}\rangle}$$

along with

$$\mathsf{S}_1(T) := \mathrm{i}e \int_{-T/2}^{T/2} dt \int_{\mathcal{C}(L)} d^3\mathbf{x} \ : \overline{\psi}(x) \, \not{A}(x) \psi(x) : .$$

Here ψ (resp. $\not{A} = \gamma_\mu A^\mu$) corresponds to the free electron-positron field (resp. the free photon field) contained in the box $\mathcal{C}(L)$ with the volume $\mathcal{V}$. Our goal is to compute explicitly the transition amplitude. Note that $\langle \Phi_{\text{out}}|\Phi_{\text{in}}\rangle = 0$. By the Main Wick Theorem 13.2 on page 846,

$$\langle \Phi_{\text{out}}|\mathsf{S}_1(T)|\Phi_{\text{in}}\rangle = \langle 0|a^-_{\mathbf{p}',r'} b^-_{\mathbf{q}',s'} \mathsf{S}_1(T) b^+_{\mathbf{q},s}|0\rangle = \int_{-T/2}^{T/2} dt \int_{\mathcal{C}(L)} d^3\mathbf{x} \varrho(x)$$

with the function

$$\boxed{\varrho(x) := \mathrm{i}e\langle 0|\overbrace{a^-_{\mathbf{p}',r'} b^-_{\mathbf{q}',s'} : \overline{\psi}(x) \not{A}(x) \psi(x) : b^+_{\mathbf{q},s}}|0\rangle.} \tag{15.22}$$

Now we use Table 14.5 on 897 which summarizes the Feynman rules. Introducing the Feynman slash symbol $\not{e}_{r'} := \gamma_\mu e^\mu_{r'}$, we obtain

$$\varrho(x) = \mathrm{i}e \mathcal{N}_{\mathbf{p}'} N_{\mathbf{q}'} N_{\mathbf{q}} \cdot \overline{u}_{\mathbf{q}',s'} \ \not{e}_{r'} u_{\mathbf{q},s} \cdot \mathrm{e}^{\mathrm{i}(p'+q'-q)x}$$

with the normalization factors[21]

$$\mathcal{N}_{\mathbf{p}'} := c_m \sqrt{\frac{1}{2\mathcal{V}\omega_{\mathbf{p}'}}}, \quad N_{\mathbf{q}'} := \sqrt{\frac{1}{2\mathcal{V}E_{\mathbf{q}'}}}, \qquad N_{\mathbf{q}} := \sqrt{\frac{1}{2\mathcal{V}E_{\mathbf{q}}}}.$$

The Feynman diagram corresponding to (15.22) is pictured in Fig. 15.5(b) above.

The integration over space and time generates discrete Dirac delta functions. Explicitly,

$$\langle q'p'|\mathsf{S}_1(T)|q\rangle = (2\pi)^4 \delta_{\text{dis}}(p' + q' - q) \cdot \mathrm{i}e \mathcal{N}_{\mathbf{p}'} N_{\mathbf{q}'} N_{\mathbf{q}} \cdot \overline{u}_{\mathbf{q}',s'} \ \not{e}_{r'} u_{\mathbf{q},s}.$$

For the transition probability during the time interval $[-\frac{T}{2}, \frac{T}{2}]$, we get

$$\boxed{W(T) = \sum_{\mathbf{p}',\mathbf{q}' \in \mathcal{G}(N)} \frac{\mathcal{V}^2 \Delta^3\mathbf{p}' \Delta^3\mathbf{q}'}{(2\pi)^6} \cdot \frac{1}{2} \sum_{r',s',s} |\langle q'p'|\mathsf{S}_1(T)|q\rangle|^2.}$$

Here, we sum over $r' = 1, 2$ and $s', s = \pm\frac{1}{2}$. As usual, we are counting the outgoing states by using a cell decomposition of the phase space. In addition, we average over the two spin states of the incoming electron. Recall that

[21] In the SI system, $\mathcal{N}_{\mathbf{p}'} = c/\sqrt{2\mathcal{V}\omega_{\mathbf{p}'}}$. Now we have to replace the velocity of light in a vacuum, c, by the velocity of light in the medium, c_m.

$$\delta_{\rm dis}(p'+q'-q)^2 = \frac{\mathcal{V}T}{(2\pi)^4}\,\delta_{\rm dis}(p'+q'-q).$$

To simplify notation, we set

$$U(\mathbf{q}',\mathbf{q}) := \frac{1}{8}\sum_{r'=1,2}\sum_{s',s=\pm\frac{1}{2}} |\overline{u}_{\mathbf{q}',s'}\,\not{\varepsilon}_{r'}u_{\mathbf{q},s}|^2.$$

Then

$$\frac{W(T)}{T} = \frac{e^2}{8\pi^2}\sum_{\mathbf{p}',\mathbf{q}'\in\mathcal{G}(N)} \frac{c_m^2\Delta^3\mathbf{p}'\Delta^3\mathbf{q}'}{\omega_{\mathbf{p}'}E_{\mathbf{q}'}E_{\mathbf{q}}}\,\delta_{\rm dis}(p'+q'-q)\,U(\mathbf{q}',\mathbf{q}).$$

Consider now the limits

$$T\to+\infty, \qquad L\to+\infty.$$

This means that the time interval becomes infinite, and the box $\mathcal{C}(L)$ converges to the total space $\mathbb{R}^3$. The quantity

$$\boxed{\tau := \lim_{T\to+\infty}\frac{W(T)}{T}}$$

is called the *emission rate* for the photon. Our goal is to compute τ. The formal limit from the lattice to the continuum yields the key formula

$$\boxed{\tau = \frac{e^2}{8\pi^2}\int_{\mathbb{R}^3\times\mathbb{R}^3}\frac{d^3\mathbf{p}'d^3\mathbf{q}'}{\omega_{\mathbf{p}'}E_{\mathbf{q}'}E_{\mathbf{q}}}\,\delta(p'+q'-q)\,U(\mathbf{q}',\mathbf{q}).}$$

Step 1: Computation of the spin sum $U(\mathbf{q}',\mathbf{q})$: As in the proof to Prop. 15.4 on page 907 concerning Compton scattering, we use the trace trick in order to get the key formula

$$\begin{aligned}U(\mathbf{q}',\mathbf{q}) &= \frac{1}{8}\sum_{r'=1,2}\mathrm{tr}\Big(\sum_{s=\pm\frac{1}{2}}u_{\mathbf{q},s}\overline{u}_{\mathbf{q},s}\,\not{\varepsilon}_{r'}\sum_{s'=\pm\frac{1}{2}}u_{\mathbf{q}',s'}\overline{u}_{\mathbf{q}',s'}\,\not{\varepsilon}_{r'}\Big)\\ &= \frac{1}{8}\sum_{r'=1,2}\mathrm{tr}\big((\not{q}+m_e)\,\not{\varepsilon}_{r'}(\not{q}\,'+m_e)\,\not{\varepsilon}_{r'}\big).\end{aligned}$$

In addition, we have to set $q'=q-p'$. This implies

$$\boxed{U(\mathbf{q}',\mathbf{q}) = \frac{1}{8}\,\mathrm{tr}(m_e^2\,\not{\varepsilon}_{r'}^2 + \not{q}\,\not{\varepsilon}_{r'}\,\not{q}\,\not{\varepsilon}_{r'} - \not{q}\,\not{\varepsilon}_{r'}\,\not{p}\,'\,\not{\varepsilon}_{r'}).}$$

By the trace property (T4) from Prop. 15.3 on page 905, we get

$$U(\mathbf{q}',\mathbf{q}) = \frac{1}{2}\sum_{r'=1,2} m_e^2e_{r'}^2 + 2(qe_{r'})^2 - q^2e_{r'}^2 - 2(qe_{r'})(p'e_{r'}) + (qp')e_{r'}^2.$$

Noting that $q^2=m_e^2$ and $p'e_{r'} = -\mathbf{p}'\mathbf{e}_{r'} = 0$, we obtain

$$\boxed{U(\mathbf{q}',\mathbf{q}) = (\mathbf{q}\mathbf{e}_{1'})^2 + (\mathbf{q}\mathbf{e}_{2'})^2 + \mathbf{p}'\mathbf{q} - E_{\mathbf{q}}\omega_{\mathbf{p}'}.}$$

Step 2: Integration over the momentum of the outgoing electron: Recall that, for the 4-dimensional Dirac delta function, we have the factorization

$$\delta(p' + q' - q) = \delta(\omega_{\mathbf{p}'} + E_{\mathbf{q}'} - E_{\mathbf{q}}) \cdot \delta(\mathbf{p}' + \mathbf{q}' - \mathbf{q}).$$

Integrating over $\mathbf{q}'$, we get

$$\tau = \frac{e^2}{8\pi^2} \int_{\mathbb{R}^3} \frac{c_m^2 d^3\mathbf{p}'}{\omega_{\mathbf{p}'} E_{\mathbf{p}'-\mathbf{q}} E_{\mathbf{q}}} \, \delta(\omega_{\mathbf{p}'} + E_{\mathbf{q}'} - E_{\mathbf{q}}) \cdot U(\mathbf{p}' - \mathbf{q}, \mathbf{q}).$$

Step 3: Integration over the energy of the outgoing photon: To simplify notation, let us set

$$n := \frac{1}{c_m}, \qquad \omega := \frac{|\mathbf{p}'|}{n}, \qquad \zeta := \cos\vartheta.$$

Then, $\omega_{\mathbf{p}'} = n|\mathbf{p}'|$, and

$$\begin{aligned} E_{\mathbf{p}'-\mathbf{q}}(\omega, \vartheta) &= \sqrt{m_e^2 + (\mathbf{p}' - \mathbf{q})^2} \\ &= \sqrt{m_e^2 + n^2\omega^2 - n\omega|\mathbf{q}|\zeta + \mathbf{q}^2}. \end{aligned}$$

Choose the following polarization vectors

$$\mathbf{e}_{1'} := \frac{\mathbf{p}' \times \mathbf{q}}{|\mathbf{p}' \times \mathbf{q}|}, \qquad \mathbf{e}_{2'} := \frac{\mathbf{p}'}{|\mathbf{p}'|} \times \mathbf{e}_{1'}.$$

From $\mathbf{q}\mathbf{e}_{1'} = 0$ and $\mathbf{q}\mathbf{e}_{2'} = -n\omega\sin\vartheta$ we obtain

$$\mathcal{U}(\omega, \vartheta) := U(\mathbf{p}' - \mathbf{q}, \mathbf{q}) = n^2\omega^2(1 - \zeta^2) + 2n\omega\zeta - E_{\mathbf{q}}\omega.$$

Using spherical coordinates, we have

$$d^3\mathbf{p}' = d\Omega_{\mathbf{p}'}\mathbf{p}'^2 d|\mathbf{p}'| = n^3 d\Omega_{\mathbf{p}'}\omega^2 d\omega.$$

This yields

$$\boxed{\tau = \frac{e^2}{8\pi^2} \int_{|\mathbf{p}'|=1} d\Omega_{\mathbf{p}'} \int_0^\infty d\omega \;\; \delta(f(\omega, \vartheta)) \cdot F(\omega, \vartheta).}$$

Here, we set

$$F(\omega, \vartheta) := \frac{n\omega\mathcal{U}(\omega, \vartheta)}{E_{\mathbf{p}'-\mathbf{q}}(\omega, \vartheta) E_{\mathbf{q}}},$$

and $f(\omega, \vartheta) := E_{\mathbf{p}'-\mathbf{q}}(\omega, \vartheta) + \omega - E_{\mathbf{q}}$. The momentum vector $\mathbf{q}$ of the incoming electron is fixed. Let us also fix the direction of the momentum vector $\mathbf{p}'$, i.e., let us fix the angle ϑ. We want to integrate over the photon energy, ω. By (15.21), the equation

$$f(\omega, \vartheta) = 0$$

has the solution $\omega = \omega_{\mathbf{p}'}$ along with

$$\omega_{\mathbf{p}'} = \frac{2(n|\mathbf{q}|\zeta - E_{\mathbf{q}})}{n^2 - 1}. \tag{15.23}$$

This corresponds to $E_{\mathbf{p}'-\mathbf{q}}(\omega_{\mathbf{p}'}, \vartheta) = E_{\mathbf{q}} - \omega_{\mathbf{p}'}$ which describes conservation of energy for the outgoing photon. Consider the integral

$$J(\vartheta) := \int_0^\infty d\omega \;\; \delta(f(\omega,\vartheta)) \cdot F(\omega,\vartheta)$$

over all possible photon energies, ω. The integrand is concentrated at the energy, $\omega_{\mathbf{p}'}$, which corresponds to energy conservation for the Cherenkov process. A standard property of the Dirac delta function (see (14.19)) tells us that

$$J(\vartheta) = \frac{F(\omega_{\mathbf{p}'},\vartheta)}{|f_\omega(\omega_{\mathbf{p}'},\vartheta)|.}$$

For the partial derivative, we get

$$f_\omega(\omega,\vartheta) = 1 + \frac{n^2\omega - n|\mathbf{q}|\zeta}{E_{\mathbf{p}'-\mathbf{q}}(\omega,\vartheta)}.$$

By (15.23), we obtain

$$f_\omega(\omega_{\mathbf{p}'},\vartheta) = \frac{E_{\mathbf{q}} - \omega_{\mathbf{p}'} + n^2\omega_{\mathbf{p}'} - n|\mathbf{q}|\zeta}{E_{\mathbf{q}'}} = \frac{\omega_{\mathbf{p}'}(n^2-1)}{2E_{\mathbf{q}'}}.$$

Hence

$$J(\vartheta) = \frac{2E_{\mathbf{q}'}F(\omega_{\mathbf{p}'},\vartheta)}{\omega_{\mathbf{p}'}(n^2-1)}.$$

Since $d\Omega_{\mathbf{p}'} = \sin\vartheta \; d\vartheta d\varphi$, we finally obtain

$$\tau = \frac{e^2}{4\pi}\int_0^{\vartheta_{\max}} d\vartheta \;\; \sin\vartheta \;\; \frac{2n}{E_{\mathbf{q}}(n^2-1)}\left(\mathbf{q}^2\sin^2\vartheta + \frac{2(|\mathbf{q}|n\cos\vartheta - E_{\mathbf{q}})^2}{n^2-1}\right).$$

Naturally enough, we only sum over those angles ϑ which are admissible for outgoing photons. This finishes the motivation for the emission rate, τ, of the Cherenkov radiation.

Problems

15.1 *The principle of critical action for the Dirac equation of the relativistic electron.* Let Ω be a nonempty bounded open subset of $\mathbb{R}^4$. We are given the continuous functions $A_0, A_1, A_2, A_3 : \mathrm{cl}(\Omega) \to \mathbb{R}$. Introduce the Lagrangian density

$$\mathcal{L}_{\mathrm{Dirac}}(\psi, \partial\psi, \overline{\psi}) := \overline{\psi}(\mathrm{i}\gamma^\alpha\nabla_\alpha - m_e)\psi$$

with the covariant derivative $\nabla_\alpha := \partial_\alpha - \mathrm{i}eA_\alpha$, and $\nabla_\alpha^- := \partial_\alpha + \mathrm{i}eA_\alpha$. Recall the definition of the Dirac adjoint $\overline{\psi} := \psi^\dagger\gamma^0$. Show that each smooth solution $\psi : \mathrm{cl}(\Omega) \to \mathbb{C}^4$ of the variational problem

$$\int_\Omega \mathcal{L}_{\mathrm{Dirac}}(\psi, \partial\psi, \overline{\psi})\, d^4x = \text{critical!}$$

along with the boundary condition "ψ = fixed on $\partial\Omega$" satisfies the Dirac equation

$$\boxed{\mathrm{i}\gamma^\alpha\nabla_\alpha\psi = m_e\psi \qquad \text{on} \quad \Omega} \tag{15.24}$$

and the adjoint equation

$$-\mathrm{i}\nabla_\alpha^- \overline{\psi}\gamma^\alpha = m_e\overline{\psi} \qquad \text{on} \quad \Omega. \tag{15.25}$$

Prove that (15.25) is a consequence of (15.24).
Solution: Choose smooth test functions $h^\mu : \Omega \to \mathbb{C}$ which have compact support, i.e., $h^\mu \in C_0^\infty(\Omega)$ for $\mu = 0, 1, 2, 3$. Replacing ψ^μ by $\psi^\mu + \tau h^\mu$, we get

$$\mathcal{J}(\tau) := \int_\Omega \mathcal{L}(\psi + \tau h, \psi^\dagger + \tau h^\dagger)\, d^4x, \qquad \tau \in \mathbb{R}.$$

Suppose that ψ is a solution of the variational problem. Then $\dot{\mathcal{J}}(0) = 0$. Hence

$$\int_\Omega \{h^\dagger\gamma^0(\mathrm{i}\gamma^\alpha\nabla_\alpha - m_e)\psi + \overline{\psi}(\mathrm{i}\gamma^\alpha\nabla_\alpha - m_e)h\}\, d^4x = 0.$$

Integration by parts yields

$$\int_\Omega \{h^\dagger\gamma^0(\mathrm{i}\gamma^\alpha\nabla_\alpha - m_e)\psi + (-\mathrm{i}\partial_\alpha\overline{\psi}\gamma^\alpha + eA_\alpha\overline{\psi}\gamma^\alpha - \overline{\psi}m_e)h\}\, d^4x = 0.$$

By the complex variational lemma (see Sect. 10.4.1 of Vol. I), we obtain

$$\gamma^0(\mathrm{i}\gamma^\alpha\nabla_\alpha - m_e)\psi = 0, \qquad -\mathrm{i}\partial_\alpha\overline{\psi}\gamma^\alpha + eA_\alpha\overline{\psi}\gamma^\alpha - m_e\overline{\psi} = 0.$$

Since the inverse matrix to γ^0 exists, we get the Dirac equation (15.24) and its adjoint equation (15.25). Finally, let us show that (15.25) is a consequence of (15.24). In fact, equation (15.24) reads as

$$\gamma^\alpha(\mathrm{i}\partial_\alpha\psi + eA_\alpha)\psi = m_e\psi.$$

Applying the operator † to (15.24), we get

$$(-\mathrm{i}\partial_\alpha\psi^\dagger + eA_\alpha\psi^\dagger)\gamma^{\alpha\,\dagger} = m_e\psi^\dagger.$$

Multiplying this by γ^0 from the right and using

$$\gamma^{0\,\dagger} = \gamma^0, \qquad \gamma^{j\,\dagger} = -\gamma^j, \qquad \gamma^j\gamma^0 = -\gamma^j\gamma^0, \qquad j = 1, 2, 3,$$

we obtain

$$(-\mathrm{i}\partial_\alpha\psi^\dagger\gamma^0 + eA_\alpha\psi^\dagger\gamma^0)\gamma^\alpha = m_e\psi^\dagger\gamma^0.$$

This is the desired relation (15.25).

15.2 *The principle of critical action for the Maxwell equations in classical electrodynamics.* Let Ω be a nonempty bounded open subset of $\mathbb{R}^4$. We are given the continuous functions $J^0, J^1, J^2, J^3 : \mathrm{cl}(\Omega) \to \mathbb{R}$. Introduce the Lagrangian density

$$\mathcal{L}_{\mathrm{Maxwell}}(A, \partial A) := -\tfrac{1}{4}F_{\alpha\beta}F^{\alpha\beta} - J^\alpha A_\alpha$$

where the components of the electromagnetic field tensor $F_{\alpha\beta}$ are given by

$$F_{\alpha\beta} := \partial_\alpha A_\beta - \partial_\beta A_\alpha \tag{15.26}$$

in terms of the four-potential A_α, $\alpha = 0, 1, 2, 3$. Show that each smooth solution $A_0, A_1, A_2, A_3 : \mathrm{cl}(\Omega) \to \mathbb{R}$ of the variational problem

$$\int_\Omega \mathcal{L}_{\mathrm{Maxwell}}(A, \partial A)\, d^4x = \text{critical!}$$

with fixed values of A_0, A_1, A_2, A_3 on the boundary $\partial\Omega$ satisfies the Maxwell equations

$$\boxed{\partial_\alpha F^{\alpha\beta} = J^\beta \qquad \text{on} \quad \Omega, \qquad \beta = 0, 1, 2, 3.}$$

In addition, we have the Bianchi identity

$$\partial_\alpha F_{\beta\gamma} + \partial_\beta F_{\gamma\alpha} + \partial_\gamma F_{\alpha\beta} = 0, \qquad \alpha, \beta, \gamma = 0, 1, 2, 3$$

which follows immediately from (15.26) by antisymmetry.
Solution: Choose smooth test functions $h_\alpha : \Omega \to \mathbb{R}$, $\alpha = 0, 1, 2, 3$, which have compact support on the open set Ω. Replacing A_α by $A_\alpha + \tau h_\alpha$, we get

$$\mathcal{J}(\tau) := \int_\Omega -\tfrac{1}{4}(\partial_\alpha A_\beta + \tau\partial_\alpha h_\beta - \partial_\beta A_\alpha - \tau\partial_\beta h_\alpha) \times$$
$$\times(\partial^\alpha A^\beta + \tau\partial^\alpha h^\beta - \partial^\beta A^\alpha - \tau\partial^\beta h^\alpha) - J^\beta(A_\beta + \tau h_\beta))\, d^4x$$

for $\tau \in \mathbb{R}$. Suppose that A_0, A_1, A_2, A_3 is a solution of the variational problem. Then $\dot{\mathcal{J}}(0) = 0$. Hence

$$\int_\Omega \{-\tfrac{1}{4}(\partial_\alpha h_\beta - \partial_\beta h_\alpha)F^{\alpha\beta} - \tfrac{1}{4}F_{\alpha\beta}(\partial^\alpha h^\beta - \partial^\beta h^\alpha) - J^\beta h_\beta\}\, d^4x = 0.$$

It follows from $F_{\alpha\beta} = -F_{\beta\alpha}$ and $F_{\alpha\beta}h^{\alpha\beta} = F^{\alpha\beta}h_{\alpha\beta}$ that

$$\int_\Omega (-F^{\alpha\beta}\partial_\alpha h_\beta - J^\beta h_\beta)\, d^4x = 0.$$

Integration by parts yields

$$\int_\Omega (\partial_\alpha F^{\alpha\beta} - J^\beta)h_\beta\, d^4x = 0.$$

By the variational lemma, $\partial_\alpha F^{\alpha\beta} - J^\beta = 0$ (see Sect. 7.20.2 of Vol. I).

15.3 *First approach to quantum electrodynamics.* Let Ω be a bounded open subset of $\mathbb{R}^4$. Introduce the Lagrangian density

$$\mathcal{L} = -\tfrac{1}{4}F_{\alpha\beta}F^{\alpha\beta} + \overline{\psi}(\mathrm{i}\nabla_\alpha\gamma^\alpha - m_e)\psi + \mathcal{L}_{\text{int}}$$

with $\mathcal{L}_{\text{int}} := -J^\alpha_{\text{QED}}A_\alpha$ and $J^\alpha_{\text{QED}} := -e\overline{\psi}\gamma^\alpha\psi$. Show that each smooth solution tuple $A_0, A_1, A_2, A_3 : \mathrm{cl}(\Omega) \to \mathbb{R}$ and $\psi : \mathrm{cl}(\Omega) \to \mathbb{C}^4$ of the variational problem

$$\int_\Omega \mathcal{L}\, d^4x = \text{critical!}$$

with fixed boundary values of A_0, A_1, A_2, A_3 and ψ satisfies the Maxwell–Dirac system

$$\partial_\alpha F^{\alpha\beta} = J^\beta_{\text{QED}}, \qquad \mathrm{i}\nabla_\beta\psi = m_e\psi \qquad \text{on} \quad \Omega, \qquad \beta = 0, 1, 2, 3.$$

Solution: Use Problems 15.1 and 15.2.

15.4 *Wave system for the four-potential.* Fix the nonzero real gauge parameter ξ. Let Ω be a bounded open subset of $\mathbb{R}^4$. We are given the continuous functions $J^\alpha : \mathrm{cl}(\Omega) \to \mathbb{R}$, $\alpha = 0, 1, 2, 3$. Introduce the Lagrangian density

$$\mathcal{L} := -\tfrac{1}{2}(\partial_\alpha A_\beta)(\partial^\beta A^\alpha) + \tfrac{1}{2}(1 - \xi^{-1})(\partial_\alpha A^\alpha)^2 - J^\beta h_\beta.$$

Show that each smooth solution $A_0, A_1, A_2, A_3 : \mathrm{cl}(\Omega) \to \mathbb{R}$ of the variational problem

$$\int_\Omega \mathcal{L}\, d^4x = \text{critical!}$$

with the boundary conditions "A_α = fixed on $\partial\Omega$ for $\alpha = 0, 1, 2, 3$" satisfies the equations

$$\Box A^\beta + (\xi^{-1} - 1)\partial^\beta(\partial_\alpha A^\alpha) = J^\beta, \qquad \beta = 0, 1, 2, 3.$$

Solution: Choose smooth test functions $h_\alpha : \Omega \to \mathbb{R}$, $\alpha = 0, 1, 2, 3$, which have compact support on the open set. Replacing A_α by $A_\alpha + \tau h_\alpha$, we get

$$\mathcal{J}(\tau) : = \int_\Omega \{-\tfrac{1}{2}(\partial_\alpha A_\beta + \tau\partial_\alpha h_\beta)(\partial^\alpha A^\beta + \tau\partial^\alpha h^\beta) + \\ +\tfrac{1}{2}(1 - \xi^{-1})(\partial_\alpha A^\alpha + \tau\partial_\alpha h^\alpha)^2 - J^\beta(A_\beta + \tau h_\beta)\}\, d^4x$$

for $\tau \in \mathbb{R}$. Suppose that A_0, A_1, A_2, A_3 is a solution of the variational problem. Then $\dot{\mathcal{J}}(0) = 0$. Hence

$$\int_\Omega \{-\tfrac{1}{2}\ \partial_\alpha A_\beta \partial^\alpha h^\beta - \tfrac{1}{2}\partial_\alpha h_\beta \partial^\alpha A^\beta + \\ +(1 - \xi^{-1})(\partial_\alpha A^\alpha)(\partial_\beta h^\beta) - J^\beta h_\beta\}\, d^4x = 0.$$

Hence

$$\int_\Omega \{-\partial_\alpha h_\beta \partial^\alpha A^\beta + (1 - \xi^{-1})(\partial_\alpha A^\alpha)(\partial^\beta h_\beta) - J^\beta h_\beta\}\, d^4x = 0.$$

Integration by parts yields

$$\int_\Omega \{\partial_\alpha \partial^\alpha A^\beta + (\xi^{-1} - 1)\partial^\beta(\partial_\alpha A^\alpha) - J^\beta\} h_\beta\, d^4x = 0.$$

By the variational lemma, $\partial_\alpha \partial^\alpha A^\beta + (\xi^{-1} - 1)\partial^\beta(\partial_\alpha A^\alpha) - J^\beta = 0$.

15.5 *Second approach to quantum electrodynamics.* Fix the real nonzero gauge parameter ξ. Let Ω be a bounded open subset of $\mathbb{R}^4$. Introduce the Lagrangian density

$$\mathcal{L} = -\tfrac{1}{2}(\partial_\alpha A_\beta)(\partial^\alpha A^\beta) + \tfrac{1}{2}(1 - \xi^{-1})(\partial_\alpha A^\alpha)^2 \\ +\overline{\psi}(\mathrm{i}\gamma^\alpha \nabla_\alpha - m_e)\psi + \mathcal{L}_{\mathrm{int}}$$

with $\mathcal{L}_{\mathrm{int}} := -J^\alpha_{\mathrm{QED}} A_\alpha$ and $J^\alpha_{\mathrm{QED}} := -e\overline{\psi}\gamma^\alpha\psi$. Show that each smooth solution $A_0, A_1, A_2, A_3 : \mathrm{cl}(\Omega) \to \mathbb{R}$ and $\psi : \mathrm{cl}(\Omega) \to \mathbb{C}^4$ of the variational problem

$$\int_\Omega \mathcal{L}\, d^4x = \text{critical!}$$

with fixed boundary values of A_0, A_1, A_2, A_3 and ψ satisfies the system

$$\boxed{\Box A^\beta + (\xi^{-1} - 1)\partial^\beta(\partial_\alpha A^\alpha) = J^\beta_{\mathrm{QED}}, \qquad \mathrm{i}\gamma^\alpha \nabla_\alpha \psi = m_e \psi \quad \text{on } \Omega}$$

for all indices $\beta = 0, 1, 2, 3$.
Solution: Use Problems 15.1 and 15.4.
Remark: It turns out that the second approach to quantum electrodynamics is more convenient than the first approach discussed in Problem 15.3. The reason is that the wave equation appears for the four-potential in the second approach. This simplifies the method of Fourier quantization also called Gupta–Bleuler quantization in quantum electrodynamics (see Sects. 11.2 and 12.4.4).

15.6 *The Weyl matrices.* Define the (4×4)-Weyl matrices γ^α_W by

$$\gamma^\alpha_W := U\gamma^\alpha U^{-1}, \qquad \alpha = 0, 1, 2, 3$$

where

$$\gamma^0 := \begin{pmatrix} \sigma^0 & 0 \\ 0 & -\sigma^0 \end{pmatrix}, \quad \gamma^j := \begin{pmatrix} 0 & \sigma^j \\ -\sigma^j & 0 \end{pmatrix}, \qquad j = 1, 2, 3$$

are called the Dirac–Pauli matrices along with the (2×2)-Pauli matrices

$$\sigma^0 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma^1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^2 := \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma^3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

In addition, we introduce the matrix

$$U := \frac{1}{\sqrt{2}} \begin{pmatrix} \sigma^0 & \sigma^0 \\ -\sigma^0 & \sigma^0 \end{pmatrix}.$$

Show the following:
(i) The matrix U is unitary, i.e., $U^\dagger = U^{-1}$.
(ii) The Weyl matrices read as

$$\gamma^0_W = \begin{pmatrix} 0 & \sigma^0 \\ \sigma^0 & 0 \end{pmatrix}, \qquad \gamma^j_W = \gamma^j, \qquad j = 1, 2, 3.$$

(iii) Introducing the chiral Dirac–Pauli matrix

$$\gamma^5 := \mathrm{i}\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} 0 & \sigma^0 \\ \sigma^0 & 0 \end{pmatrix},$$

the chiral Weyl matrix reads as

$$\gamma^5_W := \mathrm{i}\gamma^0_W\gamma^1_W\gamma^2_W\gamma^3_W = \begin{pmatrix} -\sigma^0 & 0 \\ 0 & \sigma^0 \end{pmatrix}.$$

(iv) The Clifford anticommutation relations for the Dirac–Pauli matrices,

$$\gamma^\alpha\gamma^\beta + \gamma^\beta\gamma^\alpha = 2\eta^{\alpha\beta} I, \qquad \alpha, \beta = 0, 1, 2, 3,$$

pass over to $\gamma^\alpha_W\gamma^\beta_W + \gamma^\beta_W\gamma^\alpha_W = 2\eta^{\alpha\beta} I$.

(v) The Dirac equation

$$\mathrm{i}\gamma^\alpha \partial_\alpha \psi = m_e \psi$$

passes over to $\mathrm{i}\gamma_W^\alpha \partial_\alpha \psi_W = m_e \psi_W$ where $\psi_W := U\psi$.

15.7 *The Majorana matrices.* Define the (4×4)-Majorana matrices γ_M^α by setting

$$\gamma_M^\alpha := U\gamma^\alpha U^{-1}, \qquad \alpha = 0, 1, 2, 3,$$

where

$$U := \frac{1}{\sqrt{2}} \begin{pmatrix} \sigma^0 & \sigma^2 \\ \sigma^2 & -\sigma^0 \end{pmatrix}.$$

Show the following:

(i) The matrix U is unitary, i.e., $U^\dagger = U^{-1}$. In addition, $U = U^\dagger$.

(ii) The Majorana matrices read as[22]

$$\gamma_M^0 = \begin{pmatrix} 0 & \sigma^2 \\ \sigma^2 & 0 \end{pmatrix}, \quad \gamma_M^1 = \begin{pmatrix} \mathrm{i}\sigma^3 & 0 \\ 0 & \mathrm{i}\sigma^3 \end{pmatrix}, \quad \gamma_M^2 = \begin{pmatrix} 0 & -\sigma^2 \\ \sigma^2 & 0 \end{pmatrix}, \quad \gamma_M^3 = \begin{pmatrix} -\mathrm{i}\sigma^1 & 0 \\ 0 & -\mathrm{i}\sigma^1 \end{pmatrix}.$$

(iii) The chiral Majorana matrix reads as

$$\gamma_M^5 := \mathrm{i}\gamma_M^0 \gamma_M^1 \gamma_M^2 \gamma_M^3 = \begin{pmatrix} \sigma^2 & 0 \\ 0 & -\sigma^2 \end{pmatrix}.$$

(iv) The Clifford anticommutation relations for the Dirac–Pauli matrices pass over to

$$\gamma_M^\alpha \gamma_M^\beta + \gamma_M^\beta \gamma_M^\alpha = 2\eta^{\alpha\beta} I, \qquad \alpha, \beta = 0, 1, 2, 3.$$

(v) The Dirac equation $\mathrm{i}\gamma^\alpha \partial_\alpha \psi = m_e \psi$ is transformed into

$$\mathrm{i}\gamma_M^\alpha \partial_\alpha \psi_M = m_e \psi_M$$

where we set $\psi_M := U\psi$. The choice of the Majorana matrices has the advantage that the coefficients of the transformed Dirac equation become real numbers.

15.8 *Pauli matrices.* Choose the momentum vector $\mathbf{p} = p^1\mathbf{i} + p^2\mathbf{j} + p^3\mathbf{k}$. Determine the eigensolutions of the self-adjoint matrix

$$\mathbf{p}\boldsymbol{\sigma} = \sum_{j=1}^{3} p^j \sigma^j = \begin{pmatrix} p^3 & p^1 - \mathrm{i}p^2 \\ p^1 + \mathrm{i}p^2 & -p^3 \end{pmatrix},$$

and show that $(\mathbf{p}\boldsymbol{\sigma})^2 = \mathbf{p}^2 I$.

Solution: The eigenvalues of the matrix $\mathbf{p}\boldsymbol{\sigma}$ are $\lambda = \pm|\mathbf{p}|$. This follows from

[22] In terms of the Kronecker product for matrices, we get

$$\gamma_M^0 = \sigma^1 \otimes \sigma^2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \sigma^2 = \begin{pmatrix} 0 & \sigma^2 \\ \sigma^2 & 0 \end{pmatrix},$$

and $\gamma_M^1 = \mathrm{i}\sigma^0 \otimes \sigma^3$, $\gamma_M^2 = -\mathrm{i}\sigma^2 \otimes \sigma^2$, $\gamma_M^3 = -\mathrm{i}\sigma^0 \otimes \sigma^1$, $\gamma_M^5 = \sigma^3 \otimes \sigma^2$.

$$\det(\mathbf{p}\boldsymbol{\sigma} - \lambda I) = \begin{vmatrix} p^3 - \lambda & p^1 - \mathrm{i}p^2 \\ p^1 + \mathrm{i}p^2 & -p^3 - \lambda \end{vmatrix} = \lambda^2 - \mathbf{p}^2 = 0.$$

Let $\mathbf{p} \neq 0$. Then

$$(\mathbf{p}\boldsymbol{\sigma})\chi_{\pm\frac{1}{2}}(\mathbf{p}) = \pm|\mathbf{p}|\,\chi_{\pm\frac{1}{2}}(\mathbf{p}). \tag{15.27}$$

Here, for $\mathbf{p} \neq -\mathbf{k}$, we introduce the eigenvectors

$$\chi_{\frac{1}{2}}(\mathbf{p}) := \tfrac{1}{2}\begin{pmatrix} p^3 + |\mathbf{p}| \\ p^1 + \mathrm{i}p^2 \end{pmatrix}, \qquad \chi_{-\frac{1}{2}}(\mathbf{p}) := \tfrac{1}{2}\begin{pmatrix} \mathrm{i}p^2 - p^1 \\ p^3 + |\mathbf{p}| \end{pmatrix}.$$

For $\mathbf{p} = -\mathbf{k}$, we get

$$\chi_{-\frac{1}{2}}(-\mathbf{k}) = \chi_{\frac{1}{2}}(\mathbf{k}) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \chi_{\frac{1}{2}}(-\mathbf{k}) = \chi_{-\frac{1}{2}}(\mathbf{k}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

15.9 *Plane wave solutions of the Dirac equation.* Prove Theorem 12.3 on page 807.
Solution: Recall that the Dirac equation $\mathrm{i}\gamma^\alpha\partial_\alpha\psi = m_e\psi$ is equivalent to

$$\mathrm{i}\dot{\psi} = \mathbf{H}\psi$$

with the Hamiltonian

$$\mathbf{H} := \gamma^0 m_e + \sum_{j=1}^{3}\gamma^0\gamma^j(-\mathrm{i}\partial_j) = m_e\begin{pmatrix} \sigma^0 & 0 \\ 0 & -\sigma^0 \end{pmatrix} - \mathrm{i}\begin{pmatrix} 0 & \boldsymbol{\partial}\boldsymbol{\sigma} \\ \boldsymbol{\partial}\boldsymbol{\sigma} & 0 \end{pmatrix}.$$

Analogously, introduce the operator

$$\mathbf{H}_\mathbf{p} := \gamma^0 m_e + \sum_{j=1}^{3}\gamma^0\gamma^j p^j = m_e\begin{pmatrix} \sigma^0 & 0 \\ 0 & -\sigma^0 \end{pmatrix} + \begin{pmatrix} 0 & \mathbf{p}\boldsymbol{\sigma} \\ \mathbf{p}\boldsymbol{\sigma} & 0 \end{pmatrix}.$$

Setting $px := p_\alpha x^\alpha = E_\mathbf{p}t - \mathbf{p}\mathbf{x}$, we make the ansatz

$$\psi^\pm(x) := \mathrm{e}^{\mp\mathrm{i}px}\begin{pmatrix} \chi \\ \varphi \end{pmatrix}. \tag{15.28}$$

Inserting this into the Dirac equation $\mathrm{i}\dot{\psi}^\pm = \mathbf{H}\psi^\pm$, we get $\pm E_\mathbf{p}\psi^\pm = \mathbf{H}_\mathbf{p}\psi^\pm$. Therefore, we obtain the following system:

$$\begin{aligned} (E_\mathbf{p} \pm m_e)\varphi &= (\mathbf{p}\boldsymbol{\sigma})\chi, \\ (E_\mathbf{p} \mp m_e)\chi &= (\mathbf{p}\boldsymbol{\sigma})\varphi. \end{aligned} \tag{15.29}$$

Consider first this equation equipped with the upper sign. Choose χ arbitrarily. By the first equation from (15.29), we obtain

$$\varphi = \frac{\mathbf{p}\boldsymbol{\sigma}}{E_\mathbf{p} + m_e}\,\chi.$$

The second equation from (15.29) is then satisfied automatically. In fact,

$$(\mathbf{p}\boldsymbol{\sigma})\varphi = \frac{|\mathbf{p}|^2}{E_\mathbf{p} + m_e}\,\chi.$$

Choosing $\chi := \chi_s(\mathbf{k})$, we get the two solutions

$$\begin{pmatrix}\chi\\ \varphi\end{pmatrix} = \mathcal{N}\begin{pmatrix}\chi_s(\mathbf{k})\\ \frac{\mathbf{p}\boldsymbol{\sigma}}{E_{\mathbf{p}}+m_e}\ \chi_s(\mathbf{k})\end{pmatrix}, \qquad s=\pm\tfrac{1}{2}$$

with the normalization constant $\mathcal{N}$. For example, choose $\mathcal{N} := \sqrt{E_{\mathbf{p}}+m_e}$. Consider now the system (15.29) equipped with the lower sign. We then get the solution

$$\begin{pmatrix}\chi\\ \varphi\end{pmatrix} = \mathcal{N}\begin{pmatrix}\frac{\mathbf{p}\boldsymbol{\sigma}}{E_{\mathbf{p}}+m_e}\ \chi_s(\mathbf{k})\\ \chi_s(\mathbf{k})\end{pmatrix}, \qquad s=\pm\tfrac{1}{2}.$$

15.10 *Helicity of free electrons.* For given nonzero momentum vector $\mathbf{p}$, the helicity operator (or spin projection operator) is given by

$$\mathbf{S}_{\mathbf{p}} := \frac{\mathbf{pS}}{|\mathbf{p}|}$$

with the spin operator $\mathbf{S} := S^1\mathbf{i} + S^2\mathbf{j} + S^3\mathbf{k}$ where

$$S^1 := \tfrac{1}{2}\sigma^{23}, \qquad S^2 := \tfrac{1}{2}\sigma^{31}, \qquad S^3 := \tfrac{1}{2}\sigma^{12}$$

along with $\sigma^{\alpha\beta} := \frac{\mathrm{i}}{2}[\gamma^\alpha,\gamma^\beta]_-$. Explicitly,

$$S^j := \tfrac{1}{2}\begin{pmatrix}\sigma^j & 0\\ 0 & \sigma^j\end{pmatrix}, \qquad j=1,2,3, \qquad \mathbf{pS} = \tfrac{1}{2}\begin{pmatrix}\mathbf{p}\boldsymbol{\sigma} & 0\\ 0 & \mathbf{p}\boldsymbol{\sigma}\end{pmatrix}.$$

By $(\mathbf{p}\boldsymbol{\sigma})^2 = \mathbf{p}^2 I$, we get $\mathbf{S}_{\mathbf{p}}^2 = \frac{1}{4}I$, and we have the commutation relation

$$\boxed{[\mathbf{H}_{\mathbf{p}}, \mathbf{S}_{\mathbf{p}}]_- = 0.}$$

Show that there exist solutions of the Dirac equation $\mathrm{i}\gamma^\alpha\partial_\alpha\psi = m_e\psi$ of the form (15.29) which are also eigensolutions of the helicity operator $\mathbf{S}_{\mathbf{p}}$.
Solution: For $s=\pm\frac{1}{2}$, define

$$\varphi^+_{\mathbf{p},s}(\mathbf{x},t) := \mathcal{N}\begin{pmatrix}\chi_s(\mathbf{p})\\ \frac{2s|\mathbf{p}|}{E_{\mathbf{p}}+m_e}\ \chi_s(\mathbf{p})\end{pmatrix}\mathrm{e}^{\mathrm{i}\mathbf{px}}\mathrm{e}^{-\mathrm{i}E_{\mathbf{p}}t}$$

and

$$\varphi^-_{\mathbf{p},s}(\mathbf{x},t) := \mathcal{N}\begin{pmatrix}\frac{-2s|\mathbf{p}|}{E_{\mathbf{p}}+m_e}\ \chi_{-s}(\mathbf{p})\\ \chi_{-s}(\mathbf{p})\end{pmatrix}\mathrm{e}^{-\mathrm{i}\mathbf{px}}\mathrm{e}^{\mathrm{i}E_{\mathbf{p}}t}.$$

Using the key relation (15.27) for $\mathbf{p}\boldsymbol{\sigma}$, it follows from Problem 15.9 that the Dirac equation

$$\mathrm{i}\gamma^\alpha\partial_\alpha\varphi^\pm_{\mathbf{p},s} = m_e\varphi^\pm_{\mathbf{p},s}, \qquad s=\pm\tfrac{1}{2}$$

is satisfied. Equivalently,

$$\mathrm{i}\dot{\varphi}^\pm_{\mathbf{p},s} = \mathbf{H}\varphi^\pm_{\mathbf{p},s}, \qquad s=\pm\tfrac{1}{2}.$$

Finally, for each fixed nonzero momentum vector $\mathbf{p}$, we get

$$\boxed{\mathbf{H}\varphi^\pm_{\mathbf{p},s} = \pm E_{\mathbf{p}}\varphi^\pm_{\mathbf{p},s}, \qquad \mathbf{S}_{\mathbf{p}}\varphi^\pm_{\mathbf{p},s} = \pm s\varphi^\pm_{\mathbf{p},s}, \qquad s=\pm\tfrac{1}{2}.}$$

15.11 *Relativistically invariant projection operators for electron-positron waves.* We want to construct plane wave solutions of the Dirac equation

$$\mathrm{i}\not{\partial}\psi = m_e\psi$$

which are relativistically invariant. Here, we use Feynman's slash symbol[23]

$$\not{\partial} := \gamma^\alpha\partial_\alpha, \qquad \not{p} := p_\alpha\gamma^\alpha, \qquad \not{\Pi} := \Pi_\alpha\gamma^\alpha.$$

The idea is to introduce first projection operators $P_\alpha(p,\Pi)$, $\alpha = 0,1,2,3$. Then, the eigenvectors of these operators to the eigenvalue $\lambda = 1$ generate the plane wave solutions

$$\psi^+_{p,\Pi}(x) = u(p,\Pi)\mathrm{e}^{-\mathrm{i}px}, \qquad \psi^-_{p,\Pi}(x) = v(p,\Pi)\mathrm{e}^{\mathrm{i}px} \tag{15.30}$$

of the Dirac equation where $px := E_{\mathbf{p}}t - \mathbf{px}$. These solutions depend on the momentum 4-vector p, the polarization 4-vector Π, and the sign $\pm$ of energy. To fix notation, let $p = (p^0,p^1,p^2,p^3)$ and $\Pi = (\Pi^0,\Pi^1,\Pi^2,\Pi^3)$ be 4-vectors with

$$p^2 = m_e^2, \qquad \Pi^2 = -1, \qquad p\Pi = 0.$$

The general form of p,Π can be parameterized by the arbitrary momentum vector $\mathbf{p} = p^1\mathbf{i} + p^2\mathbf{j} + p^3\mathbf{k}$. Explicitly, for $\mathbf{p} \neq \mathbf{0}$ let

$$p := (E_{\mathbf{p}},\ \mathbf{p}), \qquad E_{\mathbf{p}} := \sqrt{m_e^2 + \mathbf{p}^2}, \qquad \Pi := \left(\frac{|\mathbf{p}|}{m_e},\ \frac{E_{\mathbf{p}}}{m_e}\cdot\frac{\mathbf{p}}{|\mathbf{p}|}\right). \tag{15.31}$$

Then, $p^2 = p^\alpha p^\beta\eta_{\alpha\beta} = E_{\mathbf{p}}^2 - \mathbf{p}^2 = m_e^2$. Moreover,

$$\Pi^2 = \Pi^\alpha\Pi^\beta\eta_{\alpha\beta} = (\Pi^0)^2 - \boldsymbol{\Pi}^2 = \frac{\mathbf{p}^2}{m_e^2} - \frac{E_{\mathbf{p}}^2}{m_e^2} = -1.$$

Finally, we obtain

$$p\Pi = p^\alpha\Pi^\beta\eta_{\alpha\beta} = p^0\Pi^0 - \mathbf{p}\boldsymbol{\Pi} = \frac{E_{\mathbf{p}}|\mathbf{p}|}{m_e} - \frac{E_{\mathbf{p}}\mathbf{p}^2}{m_e|\mathbf{p}|} = 0.$$

For $\mathbf{p} = 0$, let

$$p := (m_e,\ \mathbf{0}), \qquad \Pi := (0,\ \boldsymbol{\Pi})$$

where $\boldsymbol{\Pi}$ is an arbitrary unit vector. This corresponds to a resting particle having the spin vector $\boldsymbol{\Pi}$. Now to the point. Define the two key operators

$$\boxed{\Lambda_\pm(p) := \frac{m_e I \pm \not{p}}{2m_e}, \qquad \Sigma(\Pi) := \frac{I + \gamma^5\not{\Pi}}{2}.}$$

Using this, let us introduce the following four operators:

$$P_0(p,\Pi) := \Lambda_+(p)\Sigma(\Pi), \qquad P_1(p,\Pi) := \Lambda_+(p)\Sigma(-\Pi),$$
$$P_2(p,\Pi) := \Lambda_-(p)\Sigma(\Pi), \qquad P_3(p,\Pi) := \Lambda_-(p)\Sigma(-\Pi).$$

Show that the following hold:
(i) Commutation rules: $[\Lambda_\pm(p),\Sigma(\Pi)]_- = 0$.

[23] Recall that $\gamma^\alpha\partial_\alpha = \sum_{\alpha=0}^3\gamma^\alpha\partial_\alpha$ by the Einstein convention.

(ii) Projection operators: The operators $\Lambda_\pm(p), \Sigma(\Pi), P_\alpha(p,\Pi)$ are projection operators. That is, for $\alpha = 0,1,2,3$, we have

$$\Lambda_\pm(p)^2 = \Lambda_\pm(p), \qquad \Sigma(\Pi)^2 = \Sigma(\Pi), \qquad P_\alpha(p,\Pi)^2 = P_\alpha(p,\Pi).$$

Moreover, $\Lambda_-(p) = I - \Lambda_+(p)$, and

$$\sum_{\alpha=0}^{3} P_\alpha(p,\Pi) = I,$$

as well as $P_\alpha(p,\Pi)P_\beta(p,\Pi) = 0$ if $\alpha \neq \beta$, and $\alpha,\beta = 0,1,2,3$.

(iii) For each $\alpha = 0,1,2,3$, the image space of the operator

$$P_\alpha(p,\Pi) : \mathbb{C}^4 \to \mathbb{C}^4$$

is one-dimensional. Thus, there exists an eigenvector $w_\alpha(p,\Pi)$ of the form

$$P_\alpha w_\alpha(p,\Pi) = w_\alpha(p,\Pi), \qquad \alpha = 0,1,2,3.$$

This eigenvector is uniquely determined up to a nonzero complex factor.

(iv) Plane wave solutions: Set

$$\begin{aligned} u(p,\Pi) &:= w_0(p,\Pi), & u(p,-\Pi) &:= w_1(p,\Pi), \\ v(p,\Pi) &:= w_2(p,-\Pi), & v(p,-\Pi) &:= w_3(p,\Pi). \end{aligned}$$

Then, the four functions $\psi^+_{p,\Pi}$, $\psi^+_{p,-\Pi}$, $\psi^-_{p,\Pi}$, $\psi^-_{p,-\Pi}$ introduced in (15.30) are solutions of the Dirac equation $\mathrm{i}\partial\!\!\!/\psi = m_e\psi$. Furthermore,

$$\mathbf{H}\psi^\pm_{p,\Pi} = \pm E_{\mathbf{p}}\psi^\pm_{p,\Pi}, \qquad \Sigma(\Pi)\psi^\pm_{p,\Pi} = \psi^\pm_{p,\Pi}.$$

In terms of the amplitudes u and v, we get

$$\begin{aligned} (\mathrm{i}p\!\!\!/ - m_e)\psi^\pm_{p,\Pi} &= 0, & (\mathrm{i}p\!\!\!/ + m_e)\psi^\pm_{p,\Pi} &= 0, \\ \Lambda_\pm(p)u(p,\Pi) &= u(p,\Pi), & \Lambda_\pm(p)v(p,\Pi) &= v(p,\Pi), \\ \Sigma(\Pi)u(p,\Pi) &= u(p,\Pi), & \Sigma(\Pi)v(p,\Pi) &= v(p,\Pi). \end{aligned}$$

Naturally enough, all of these relations remain true if we replace the polarization 4-vector Π by $-\Pi$.

Explicitly, for given nonzero momentum vector $\mathbf{p}$, choose p and Π as in (15.31). Using the helicity solutions, $\varphi^\pm$, from Problem 15.10, we obtain

$$\psi^\pm_{p,\Pi} = \varphi^\pm_{\mathbf{p},\frac{1}{2}}, \qquad \psi^\pm_{p,-\Pi} = \varphi^\pm_{\mathbf{p},-\frac{1}{2}}.$$

This corresponds to particles which possess the spin vector $\mathbf{p}/|\mathbf{p}|$ in the rest system.

(v) The rest system: Introduce the following basis column matrices:

$$e_0 := \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad e_1 := \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad e_2 := \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad e_3 := \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

in the space $\mathbb{C}^4$. For a given particle, consider its rest system, that is, we have $p = (m_e, \mathbf{0})$ and $\Pi = (0, \mathbf{k}$. Here, the unit vector $\mathbf{k}$ in direction of the z-axis represents the spin vector of the resting particle. Then

$$u(p, \Pi) = \sqrt{m_e}\ e_0, \quad u(p, -\Sigma) = \sqrt{m_e}\ e_1$$
$$v(p, \Sigma) = \sqrt{m_e}\ e_2, \qquad v(p, -\Pi) = \sqrt{m_e}\ e_3.$$

(vi) Lorentz transformation: Consider a fixed inertial system Σ and suppose that the particle has the momentum vector $\mathbf{p}$ in Σ. In addition, let Σ_{rest} denote that rotated rest system in which the momentum vector of the particle points in direction of the z-axis. Then, the two inertial systems Σ_{rest} and Σ are related to each other by a Lorentz transformation

$$\begin{pmatrix} t \\ \mathbf{x} \end{pmatrix} = L \begin{pmatrix} t_{\text{rest}} \\ \mathbf{x}_{\text{rest}} \end{pmatrix}.$$

For the corresponding momentum vectors, we get

$$\begin{pmatrix} E_{\mathbf{p}} \\ \mathbf{p} \end{pmatrix} = L \begin{pmatrix} m_e \\ \mathbf{0} \end{pmatrix}.$$

Finally, set $\Pi_z = Le_3$. Then, the plane wave solution to the polarization 4-vector Π_z is given by

$$\psi^+_{p,\,\pm\Pi_z} = u_{\mathbf{p},\,\pm\frac{1}{2}}\ \mathrm{e}^{-\mathrm{i}px}, \qquad \psi^-_{p,\,\pm\Pi_z} = v_{\mathbf{p},\,\pm\frac{1}{2}}\ \mathrm{e}^{\mathrm{i}px}.$$

Here, the explicit form of $u_{\mathbf{p},s}$, $v_{\mathbf{p},s}$ can be found in Sect. 12.1 on page 799. In the language of matrices, we get

$$(u_{\mathbf{p},\,\frac{1}{2}}, u_{\mathbf{p},\,-\frac{1}{2}}, v_{\mathbf{p},\,\frac{1}{2}}, v_{\mathbf{p},\,-\frac{1}{2}}) = \sqrt{2m_e} \exp\left(-\frac{\omega}{2} \cdot \frac{\boldsymbol{\alpha}\mathbf{p}}{|\mathbf{p}|}\right)\ (e_0, e_1, e_2, e_3).$$

Here, $\boldsymbol{\alpha}\mathbf{p} = \sum_{j=1}^3 p^j \gamma^0 \gamma^j$, and $\tanh \frac{\omega}{2} = -|\mathbf{p}|/(E_{\mathbf{p}} + m_e)$. Explicitly,

$$\exp\left(-\frac{\omega}{2} \cdot \frac{\boldsymbol{\alpha}\mathbf{p}}{|\mathbf{p}|}\right) = \cosh\frac{\omega}{2} - \frac{\boldsymbol{\alpha}\mathbf{p}}{|\mathbf{p}|}\ \sinh\frac{\omega}{2}.$$

Hint: See Bjorken and Drell, Relativistic Quantum Mechanics, McGraw-Hill, New York, 1964. See also W. Greiner, Relativistic Quantum Mechanics, Springer, Berlin, 1997.

15.12 *Proof of Example 15.2.* Solution:

Ad (a)–(d). This follows from the definition of the Dirac matrices.

Ad (e), (f), (n). Obviously, $\operatorname{tr}\gamma^\mu = 0$. Now the trick is to use the matrix γ^5. By property (d),

$$\gamma^5\gamma^\mu\gamma^\nu\gamma^\kappa\gamma^5 = -\gamma^5\gamma^5(\gamma^\mu\gamma^\nu\gamma^\kappa) = -\gamma^\mu\gamma^\nu\gamma^\kappa.$$

Since $(\gamma^5)^{-1} = \gamma^5$,

$$\operatorname{tr}(\gamma^\mu\gamma^\nu\gamma^\kappa) = \operatorname{tr}(\gamma^5\gamma^\mu\gamma^\nu\gamma^\kappa\gamma^5) = -\operatorname{tr}(\gamma^\mu\gamma^\nu\gamma^\kappa).$$

Taking (g)–(l) for granted, the trace $\operatorname{tr}(\gamma^{\mu_1}\cdots\gamma^{\mu_N})$ is real, by (h). Thus, it follows from (b) and (l) that

$$\operatorname{tr}(\gamma^{\mu_1}\cdots\gamma^{\mu_N}) = \operatorname{tr}(\overline{\gamma}^{\mu_N}\cdots\overline{\gamma}^{\mu_1}) = \operatorname{tr}(\gamma^{\mu_N}\cdots\gamma^{\mu_1}).$$

Ad (g). By the Clifford relation (c),

$$2\operatorname{tr}(\gamma^\mu\gamma^\nu) = 2\eta^{\mu\nu}\operatorname{tr} I = 8\eta^{\mu\nu}.$$

Ad (h), (n), (o). To simplify notation, we set $A := a_\mu\gamma^\mu$, $B := b_\nu\gamma^\nu$,... By (g),

$$\operatorname{tr}(AB) = a_\mu b_\nu \operatorname{tr}(\gamma^\mu\gamma^\nu) = 4a_\mu b_\nu\eta^{\mu\nu} = 4a_\mu b^\mu.$$

By the Clifford relation (c),

$$\operatorname{tr}(ABCD) + \operatorname{tr}(BACD) = \operatorname{tr}(AB + BA)\operatorname{tr}(CD).$$

By permutation,

$$\begin{aligned}
\operatorname{tr}(ABCD) &= 2\operatorname{tr}(AB)\operatorname{tr}(CD) - \operatorname{tr}(BACD),\\
-\operatorname{tr}(BACD) &= -2\operatorname{tr}(AC)\operatorname{tr}(BD) + \operatorname{tr}(BCAD),\\
\operatorname{tr}(BCAD) &= 2\operatorname{tr}(AD)\operatorname{tr}(BC) - \operatorname{tr}(BCDA).
\end{aligned}$$

Adding this, we obtain[24]

$$\operatorname{tr}(ABCD) = \operatorname{tr}(AB)\operatorname{tr}(CD) - \operatorname{tr}(AC)\operatorname{tr}(BD) + \operatorname{tr}(AD)\operatorname{tr}(BC).$$

This is the claim for $N = 4$. The general case can be treated similarly.
Ad (i). By the Clifford relation (c),

$$\operatorname{tr}(\gamma^\mu\gamma^\nu\gamma^5) + \operatorname{tr}(\gamma^\nu\gamma^\mu\gamma^5) = 2\eta^{\mu\nu}\operatorname{tr}(\gamma^5) = 0.$$

Thus, $\operatorname{tr}(\gamma^\mu\gamma^\nu\gamma^5)$ is antisymmetric with respect to the indices μ, ν. Finally, for $(\mu,\nu) = (0,1),\ (0,2),\ldots$, an explicit computation yields $\operatorname{tr}(\gamma^\mu\gamma^\nu\gamma^5) = 0$.
Ad (j). An explicit computation yields the following key relation

$$\gamma^\mu\gamma^\nu\gamma^\kappa = \eta^{\mu\nu}\gamma^\kappa - \eta^{\mu\kappa}\gamma^\nu + \gamma^{\nu\kappa}\gamma^\mu - \mathrm{i}\epsilon^{\mu\nu\kappa\sigma}\gamma_\sigma\gamma^5.$$

Using $(\gamma^5)^2 = I$ and statement (i), we get

$$\operatorname{tr}(\gamma^\mu\gamma^\nu\gamma^\kappa\gamma^\lambda\gamma^5) = -\mathrm{i}\epsilon^{\mu\nu\kappa\sigma}\operatorname{tr}(\gamma_\sigma\gamma^\lambda) = -4\mathrm{i}\epsilon^{\mu\nu\kappa\sigma}\eta_\sigma^\lambda = -4\mathrm{i}\epsilon^{\mu\nu\kappa\lambda}.$$

Ad (k). Since $\gamma_0^{-1} = \gamma_0$, we have $\operatorname{tr}(\gamma_0 A^\dagger\gamma_0) = \operatorname{tr} A^\dagger = (\operatorname{tr} A)^\dagger$.
Ad (l). Note that

$$\operatorname{tr}(\overline{A}_N\cdots\overline{A}_1) = \operatorname{tr}(\gamma_0 A_N^\dagger A_{N-1}^\dagger\cdots A_1^\dagger\gamma_0) = \operatorname{tr}(A_N^\dagger\cdots A_1^\dagger) = \big(\operatorname{tr}(A_1\cdots A_N)\big)^\dagger.$$

Ad (m). The matrix $\overline{AB}$ is equal to

$$\gamma_0(AB)^\dagger\gamma_0\gamma_0 B^\dagger A^\dagger\gamma_0 = (\gamma_0 B^\dagger\gamma_0)(\gamma_0 A^\dagger\gamma_0) = \overline{B}\cdot\overline{A}.$$

15.13 *Proof of Furry's rule.* Solution: Let us start with a 2-vertex electron loop (see Table 14.3(a), (b) on page 888). By the Wick theorem, we have to consider the pairing[25]

$$\overline{\psi}(x)\psi(x)\overline{\psi}(y)\psi(y).$$

[24] Note that $\operatorname{tr}(BCDA) = \operatorname{tr}(ABCD)$ by cyclic permutation.
[25] To simplify notation, the pairings for the photon are not indicated.

(a) (b)

Fig. 15.6. Furry's rule

After an even permutation of the factors, we get

$$\underbrace{\psi(x)\overline{\psi}}(y)\underbrace{\overline{\psi}(x)}\psi(y).$$

Since $\underbrace{\overline{\psi}(x)\psi}(y) = -\underbrace{\psi(y)\overline{\psi}}(x)$, we obtain

$$-\underbrace{\psi(x)\overline{\psi}}(y)\underbrace{\psi(y)\overline{\psi}}(x).$$

Graphically, this represents two electron lines from y to x and from x to y, along with a minus sign.

This way, positron lines can be reduced to electron lines (plus an additional minus sign).

This is the simplest form of Furry's rule.

Now consider the case of a 3-vertex electron loop. The idea is that the contributions coming from the two graphs from Fig. 15.6 compensate each other after carrying out the substitution $\gamma_\mu \mapsto -\gamma_\mu$ for the Dirac matrices. In fact, the graphs (a) and (b) of Fig. 15.6 yield the transition amplitude $L + R$. Symbolically,

$$L = \operatorname{tr}\big(A(x)S(x,y)A(y)S(y,z)A(z)S(z,x)\big)$$

and

$$R = \operatorname{tr}\big(A(x)S(x,z)A(z)S(z,y)A(y)S(y,x)\big).$$

Now consider the term R and use the following identities:
(i) $\operatorname{tr}(\gamma_\mu\gamma_\nu\gamma_\lambda) = \operatorname{tr}(\gamma_\lambda\gamma_\nu\gamma_\mu)$;
(ii) $\operatorname{tr}(\gamma_\mu\gamma_\nu\gamma_\lambda) = \operatorname{tr}\big((-\gamma_\mu)(-\gamma_\nu)(-\gamma_\lambda)\big)$;
(iii) $A(x;-\gamma) = -A(x;\gamma)$ and $S(x,y;-\gamma) = S(x,y;\gamma)$.
Replacing the Dirac matrix γ_μ by $-\gamma_\mu$ and applying cyclic permutations, we get

$$\begin{aligned} R &= (-1)^3 \operatorname{tr}\big(A(x)S(z,x)A(z)S(y,z)A(y)S(x,y)\big) \\ &= -\operatorname{tr}\big(S(x,y)A(y)S(y,z)A(z)S(z,x)A(x)\big) \\ &= -\operatorname{tr}\big(A(x)S(x,y)A(y)S(y,z)A(z)S(z,x)\big). \end{aligned}$$

Hence $L + R = 0$. In the general case, we argue similarly.

16. The Continuum Limit

> Scattering processes for elementary particles are based on the long-time limit, the high-energy limit, and the low-energy limit. Try to compute these limits only for quantities which are measurable in physical experiments (e.g., cross sections).
>
> The golden rule of quantum field theory

In Sect. 14.7 we applied this golden rule to the cross section of the Compton scattering between photons and electrons. This way, we obtained the famous Klein–Nishina formula in lowest order of perturbation theory. This formula can be established by physical experiments. In this connection, we argued as follows: We started with a lattice. Then we studied the three fundamental limits: $P_{\max} \to +\infty$ (high-energy limit), $\mathcal{V} \to +\infty, \Delta p \to 0$ (low-energy limit), and $T \to +\infty$ (long-time limit), by using the language of distributions. In addition, we included the two limits $\varepsilon \to +0$ (electron propagator regularization) and $m_{ph} \to +0$ (photon propagator regularization – the virtual photon mass goes to zero). In this chapter, we want to discuss how this approach can be applied to general problems in quantum electrodynamics.

16.1 The Fundamental Limits

Choice of the lattice. Recall that we put the physical system in a cubic box of side length L and finite volume $\mathcal{V} = L^3$. Furthermore, we observe the physical system during the finite time interval $[-\frac{T}{2}, \frac{T}{2}]$. In addition, as in Sect. 12.1.1 on page 799, we consider a finite lattice $\mathcal{G}(N)$ in momentum space with minimal momentum $\Delta p := 2\pi/L$ and maximal momentum

$$P_{\max} := N \cdot \Delta p.$$

Feynman rules. Suppose that we have computed the transition probability $\mathcal{W}(T)$ for a specific process in fixed order of perturbation theory by using the Feynman rules for a lattice in momentum space, as formulated in Table 14.5 on page 897. The analytic expressions are well defined, but the transition probability depends on

- the choice of the finite lattice $\mathcal{G}(N)$ characterized by the minimal momentum Δp and the maximal momentum $P_{\max}$,
- the regularization parameter ε for the electron propagator,
- the artificial photon mass m_{ph}, and
- the length of the time interval $[-\frac{T}{2}, \frac{T}{2}]$.

The limits. Note that the limit $L \to +\infty$ yields the infinite box $\mathbb{R}^3$ in position space, that is, the normalization volume $\mathcal{V}$ goes to infinity, $\mathcal{V} \to \infty$. Moreover, $\Delta p \to 0$.

The limit $N \to +\infty$ yields $P_{\max} \to +\infty$. In a specific physical experiment, the parameters $\Delta p, P_{\max}$ and T are fixed. However, in order to get universal formulas, we are interested in performing the following three fundamental limits: $L \to +\infty$, $N \to \infty$, and $T \to +\infty$. In terms of physics, this means the following:

(H) High-energy limit: $P_{\max} \to +\infty$ and $\varepsilon \to +0$.
(L) Low-energy limit: $\Delta p \to 0$ and $m_{ph} \to +0$.
(T) Large-time limit: $T \to +\infty$.

Recall that $p = (E, \mathbf{p})$ denotes the 4-momentum, that is, E (resp. $\mathbf{p}$) denotes the energy (resp. the momentum vector) of the particles. Thus, the high-energy limit concerns the physical behavior at large 4-momenta, $p \to +\infty$. In contrast to this, the low-energy limit refers to small 4-momenta, $p \to 0$.

16.2 The Formal Limits Fail

It turns out that the limits (H), (L), (T) above do not always exist for higher-order terms in perturbation theory. In particular, the high-energy limit causes trouble. Here, the discrete algebraic Feynman integrals may pass to divergent integrals. To cure this defect, physicists invented the method of renormalization.

There exist several approaches to renormalization theory. Physicists prefer computations which are based on the regularization of divergent algebraic Feynman integrals. For a mathematician, it is important to understand this approach. Otherwise, he is lost in the physics literature, and he does not understand the language of physicists. In Chap. 19 we will discuss different approaches to renormalization theory. We want to emphasize that there exists a very elegant approach which attacks the high-energy limit by using the modern theory of tempered distributions. This is the Epstein–Glaser approach which completely avoids the use of lattices and divergent integrals (see Sect. 19.2).

In order to master complicated problems in both mathematics and physics, one tries to discover symmetries behind the problems and to use simplifications based on symmetry arguments. Concerning renormalization theory, there are two fundamental symmetries, namely,

- the gauge invariance and
- the Hopf algebra structure of the set of Feynman graphs.

The gauge invariance leads to crucial symmetry relations called the Slavnov–Taylor identities in gauge theory. In quantum electrodynamics, the Ward identity (and the more general Ward–Takahashi identities) are special cases of the Slavnov–Taylor identities. The Hopf algebra structure behind Feynman graphs helps to master the complicated combinatorics of renormalization theory in higher-order perturbation theory (also called multi-loop perturbation theory).

The procedure of renormalization depends on the choice of appropriate normalization conditions. A change of these normalization conditions is governed by an additional internal symmetry which is described by the renormalization group. This group is used by physicists in order to extrapolate the computations to high energies. This leads to effective coupling constants which depend on the energy of the elementary particles colliding in particle accelerators (so-called running coupling constants).

The beginner in renormalization theory is confronted with an ocean of messy formulas. We want to help the beginner to understand renormalization theory by first discussing the basic ideas. The unavoidable lengthy computations will be postponed to Vol. III.

16.3 Basic Ideas of Renormalization

> Nature is not embarrassed by difficulties of mathematical analysis.
> Augustin Fresnel (1788–1827)

Our goal is to discuss the following crucial notions:

- effective electron mass m_{eff},
- bare electron mass m_e,
- effective electron charge $-e_{\text{eff}}$,
- bare electron charge $-e$,
- counterterms of the Lagrangian (additional quantum fluctuations),
- the compensation principle,
- the invariance principles,
- dimensional regularization of discrete algebraic Feynman integrals, and
- multiplicative renormalization constants.

16.3.1 The Effective Mass and the Effective Charge of the Electron

For the electron, physicists measure the mass m_{eff} and the electric charge $-e_{\text{eff}}$. Explicitly, the measurements yield the following values in the SI system:

$$m_{\text{eff}} = 0.511 \text{ MeV}/c^2, \qquad e_{\text{eff}} = \sqrt{4\pi\varepsilon_0\hbar c\alpha} = 1.602 \cdot 10^{-19} \text{ As} \tag{16.1}$$

with the dimensionless (electromagnetic) fine structure constant $\alpha = 1/137.04$. Physicists call m_{eff} (resp. e_{eff}) the effective electron mass (resp. the effective electron charge). In the energetic system with $\hbar = c = \varepsilon_0 = 1$, we get

$$m_{\text{eff}} = 0.511 \text{ MeV}, \qquad e = \sqrt{4\pi\alpha}.$$

Physicists assume that these two quantities are the result of complicated interactions between parts of the quantum field in quantum electrodynamics. Nowadays nobody knows how to compute these values from first principles. Renormalization theory only provides us with finite expressions in each order of perturbation theory, and these expressions depend on two free parameters m_{eff} and e_{eff} which have to be measured by physical experiments.

16.3.2 The Counterterms of the Modified Lagrangian

The bare Lagrangian. In (11.3) on page 794 we introduced the following Lagrangian density of quantum electrodynamics:

$$\mathcal{L} = -\frac{1}{2}(\partial_\mu A_\nu)(\partial^\mu A^\nu) + \frac{1}{2}\left(1 - \frac{1}{\xi}\right)(\partial_\mu A^\mu)^2 + \mathrm{i}\overline{\psi}\,\partial\!\!\!/\psi - m_e\overline{\psi}\psi + e\overline{\psi}\,A\!\!\!/\psi.$$

Here, we use the Feynman slash symbols $\partial\!\!\!/ := \gamma^\mu\partial_\mu$ and $A\!\!\!/ = \gamma_\mu A^\mu$. The real parameter ξ is a fixed nonzero real number called the gauge parameter. The choice $\xi = 1$ is called the Feynman gauge. The parameter m_e is called the bare mass of the electron, and the parameter $-e$ is called the bare electric charge of the electron.

Additive renormalization. Physicists assume that the bare quantities m_e and e cannot be measured in physical experiments. One has to add contributions coming from complicated interaction processes. Therefore, we introduce the following two parameters:

$$\boxed{m_{\text{eff}} := m_e + \delta m, \qquad e_{\text{eff}} := e + \delta e.}$$

Hence

$$m_e = m_{\text{eff}} - \delta m, \quad e = e_{\text{eff}} - \delta e.$$

This way we get the new Lagrangian density:

$$\begin{aligned}\mathcal{L}_{\text{ren}} := &-\frac{1}{2}(\partial_\mu A_\nu)(\partial^\mu A^\nu) + \frac{1}{2}\left(1 - \frac{1}{\xi}\right)(\partial_\mu A^\mu)^2 \\ &+\mathrm{i}\overline{\psi}\ \partial\!\!\!/\psi - m_{\text{eff}}\overline{\psi}\psi + e_{\text{eff}}\overline{\psi}\ A\!\!\!/\psi + \mathcal{L}_{\text{counter}}\end{aligned}$$

along with

$$\mathcal{L}_{\text{counter}} := \delta m\ \overline{\psi}\psi - \delta e\ \overline{\psi}\ A\!\!\!/\psi.$$

Counterterms. The additional terms of $\mathcal{L}_{\text{counter}}$ are called counterterms. Now if one uses the modified Lagrangian density, then the S-matrix $\mathsf{S}(T)$ changes. Applying the Wick theorem to the new S-matrix, we get modified Feynman rules.

16.3.3 The Compensation Principle

The main idea is to choose the parameters δm and δe in such a way that the continuum limit exists in an appropriate sense. The prototype for this technique is the proof of the Mittag-Leffler theorem for the representation of meromorphic functions (see Sect. 2.1.3 on page 56). In this classical mathematical proof from the 19th century, Mittag-Leffler used additional subtraction terms in order to force the convergence from a finite sum approximation to an infinite series. In general terms, this corresponds to a passage from a finite number of degrees of freedom to an infinite number of degrees of freedom. In renormalization theory, physicists try to give the subtraction terms a physical meaning by using additional terms of the Lagrangian density called counterterms. The philosophy of physicists reads as follows:

- The Feynman rules are based on the classical Lagrangian density of quantum electrodynamics.
- This classical Lagrangian density does not see all of the complicated physical interactions between electrons, positrons, and photons. In particular, this concerns the so-called radiative corrections (e.g., the Lamb shift in the spectrum of the hydrogen atom and the anomalous magnetic moment of the electron).
- Therefore, we have to supplement systematically the classical Lagrangian density by adding so-called counterterms.
- From the physical point of view, the counterterms describe additional quantum fluctuations.
- From the mathematical point of view, the counterterms enforce the convergence of the expressions with respect to the limits (H), (L), and (T) above.

It turns out that this procedure is quite reasonable from the physical point of view. The crucial point is that:

- After finishing the procedure of renormalization, the quantities δm and δe completely disappear.
- We get well-defined expressions for physical quantities which depend on the two free real parameters m_{eff} and e_{eff}.

- These two parameters cannot be determined by the theory. They have to be measured by physical experiments.

The reader should have in mind that, in contrast to the proof of the Mittag-Leffler theorem in complex function theory, a direct attack of the continuum limit by means of subtraction terms is extremely clumsy. For simplifying the computations, we will modify the approach by using the method of dimensional regularization (see Sects. 16.3.5 and 17.4).

16.3.4 Fundamental Invariance Principles

In quantum field theory, one has to realize the following three fundamental postulates:

(R) relativistic invariance,
(U) unitarity of the S-matrix, and
(G) gauge invariance.

Condition (R) guarantees that the theory does not depend on the choice of the inertial system. The property (U) ensures that the elements of the S-matrix allow a reasonable probabilistic interpretation, and hence the computed cross sections possess a well-defined physical meaning for scattering processes. The gauge invariance postulates that the theory is invariant under gauge transformations (generalized phase transitions). In electrodynamics, gauge invariance means that electromagnetic phenomena only depend on the electromagnetic field $\mathbf{E}, \mathbf{B}$, but not on the different choices of the four-potential $U, \mathbf{A}$.

Unfortunately, the compatibility between (R), (U), *and* (G) *causes trouble.*

In order to avoid this trouble, a careful approach to quantum field theory is needed. In particular, it turns out that the postulate of gauge invariance leads to the so-called Ward identities which are crucial for the process of renormalization.

16.3.5 Dimensional Regularization of Discrete Algebraic Feynman Integrals

Let us sketch how the modern method of dimensional regularization is used in renormalization theory. The Feynman rules from Table 14.8 on page 896 generate discrete algebraic Feynman integrals denoted by

$$J(T, \mathcal{V}, \Delta p, P_{\max}; \varepsilon, m_{ph}).$$

In order to compute cross sections, as a rule, we are interested in time averages of the form

$$\frac{J(T, \mathcal{V}, \Delta p, P_{\max}; \varepsilon, m_{ph})}{T}.$$

We now proceed in the following steps:

Step 1: The continuum limit: The limits $L \to \infty, N \to \infty, T \to \infty$ correspond to the limits $\mathcal{V} \to \infty, \Delta p \to 0, P_{\max} \to \infty$, and $T \to \infty$. Performing these limits, we get the integral

$$\mathcal{J}_n(\varepsilon, m_{ph}) = \int_{\mathbb{R}^n} \dots$$

Note that this integral may diverge, that is, our limit is regarded as a formal limit. However, this procedure yields a uniquely determined convergent or divergent integral.

Step 2: Dimensional regularization: If the integral $\mathcal{J}$ is divergent, then we replace the dimension n by $n-\nu$, and we apply the method of dimensional regularization (see Sect. 2.2.9 on page 73). This way, we get

$$\mathcal{J}_{n-\nu}(\varepsilon, m_{ph}) = \sum_{k=-r}^{\infty} \frac{a_k}{\nu^k}.$$

Here, the natural number $r = 1, 2, \ldots$ tells us the order of the pole at the point $\nu = 0$.

Step 3: Truncation of the singular part (Rota–Baxter projection operator): Define

$$\boxed{\mathsf{P}\left(\sum_{k=-r}^{\infty} \frac{a_k}{\nu^k}\right) := \frac{a_{-1}}{\nu} + \frac{a_{-2}}{\nu^2} + \ldots + \frac{a_{-r}}{\nu^r}.} \tag{16.2}$$

This yields the regularization

$$(1-\mathsf{P}) \sum_{k=-r}^{\infty} \frac{a_k}{\nu^k} := a_0 + a_1\nu + a_2\nu^2 + \ldots.$$

Letting $\nu \to +0$, we get the regularized value

$$\boxed{\mathcal{J}_{\mathrm{reg}}(\varepsilon, m_{ph}) = \lim_{\nu \to +0} (1-\mathsf{P})\mathcal{J}_{n-\nu}(\varepsilon, m_{ph}) = a_0.}$$

Step 4: Choice of the counterterms via compensation principle: We try to choose δm and δe as functions of the small parameter ν in such a way that the terms containing $\delta m(\nu)$ and $\delta e(\nu)$ replace the action of the operator $1-\mathsf{P}$, that is, these terms cancel the singular terms with respect to the parameter ν. After that we carry out the limit $\nu \to +0$.

This way, we obtain a final result which only depends on the regularization parameter ε (shift of the electron rest energy to the upper half-plane) and the virtual photon mass m_{ph}. It remains to study the limits $\varepsilon \to +0$ and $m_{ph} \to +0$. Roughly speaking, this is a problem of the theory of generalized functions (including Cauchy's residue calculus, Wick rotation, and so on). Moreover, it turns out that the existence of the limit $m_{ph} \to 0$ is related to special physical effects. Roughly speaking, for a fixed order of perturbation theory, one has to take all possible Feynman diagrams into account. Then surprising cancellations appear (in particular, see the discussion of braking radiation on page 976).

16.3.6 Multiplicative Renormalization

As noticed by Dyson in his fundamental paper on renormalization theory,[1] it is convenient to replace the additive relation $m_e = m_{\mathrm{eff}} - \delta m$ by the multiplicative relation $m_e := Z_m m_{\mathrm{eff}}$. We will also change ψ, A, and ξ. To this end, we choose five real parameters Z_m, Z_e, Z_ξ, Z_ψ, and Z_A. We set

$$\boxed{m_e = Z_m m_{\mathrm{eff}}, \qquad e = Z_e e_{\mathrm{eff}},}$$

[1] F. Dyson, The S-matrix in quantum electrodynamics, Phys. Rev. **75** (1949), 1736–1755.

and we are going to use the replacement

$$A^\mu \mapsto Z_A^{1/2} A^\mu, \qquad \psi \mapsto Z_\psi^{1/2} \psi, \qquad \xi \mapsto Z_\xi\, \xi.$$

We then get the following new Lagrangian density:

$$\begin{aligned}\mathcal{L}_{\text{ren}} : = & -\frac{1}{2} Z_A (\partial_\mu A_\nu)(\partial^\mu A^\nu) + \frac{1}{2}\left(Z_A - \frac{Z_A}{Z_\xi\, \xi}\right)(\partial_\mu A^\mu)^2 \\ & + \mathrm{i} Z_\psi \overline{\psi}\, \not{\partial} \psi - Z_m Z_\psi m_{\text{eff}} \overline{\psi} \psi + Z_e Z_\psi Z_A^{1/2} e_{\text{eff}} \overline{\psi}\, \not{A} \psi.\end{aligned}$$

This can be written as

$$\begin{aligned}\mathcal{L}_{\text{ren}} : = & -\frac{1}{2} (\partial_\mu A_\nu)(\partial^\mu A^\nu) + \frac{1}{2}\left(1 - \frac{1}{\xi}\right)(\partial_\mu A^\mu)^2 \\ & + \mathrm{i} \overline{\psi}\, \not{\partial} \psi - m_{\text{eff}} \overline{\psi} \psi + e_{\text{eff}} \overline{\psi}\, \not{A} \psi + \mathcal{L}_{\text{counter}}\end{aligned}$$

along with the counterterms

$$\begin{aligned}\mathcal{L}_{\text{counter}} : = & -\tfrac{1}{2}(Z_A - 1)\partial_\mu A_\nu)(\partial^\mu A^\nu) + \tfrac{1}{2}\left(Z_A - 1 - \frac{Z_A}{Z_\xi \xi} + \frac{1}{\xi}\right)(\partial_\mu A^\mu)^2 \\ & + \mathrm{i}(Z_\psi - 1)\overline{\psi}\, \not{\partial} \psi - (Z_m Z_\psi - 1)\, m_{\text{eff}} \overline{\psi} \psi \\ & + (Z_e Z_\psi Z_A^{1/2} - 1)\, e_{\text{eff}} \overline{\psi}\, \not{A} \psi.\end{aligned}$$

We postulate that

$$\boxed{Z_e = Z_A^{-1/2}, \quad Z_\xi = Z_A.} \tag{16.3}$$

In Vol. III we will show that this postulate can be motivated by the postulate of gauge invariance via the crucial Ward identities. In fact, the postulate (16.3) ensures the renormalizability of quantum electrodynamics (see the quotation on page 997).

16.4 The Theory of Approximation Schemes in Mathematics

Consider the operator equation

$$Ax = b, \qquad x \in X. \tag{16.4}$$

We are given the linear or nonlinear operator $A : X \to Y$, where X and Y are, say, Hilbert spaces or Banach spaces. For given $b \in Y$, we are looking for a solution x of the equation (16.4). In order to get approximate solutions of (16.4), we add the equations

$$A_n x_n = b_n, \qquad x_n \in X_n, \qquad n = 1, , 2 \ldots \tag{16.5}$$

There arises the following question: Which general class of operators has the following property: It is possible to construct an approximation scheme (16.5) such that the approximate solutions converge against the unique solution of the original equation (16.4), that is, we have the convergence

$$\lim_{n\to\infty} x_n = x$$

in an appropriate sense. There exists a nontrivial answer in terms of the class of so-called A-proper operators introduced by Browder and Petryshyn in the late 1960s by summarizing a long development in numerical analysis. This theory can be found in the last three chapters of the author's monograph:

> E. Zeidler, Nonlinear Functional Analysis and its Applications, Vol. IIB: Monotone Operators, Springer, New York, 1990.

We also refer to the survey article by W. Petryshyn, On the approximation-solvability of equations involving A-proper and pseudo-A-proper mappings, Bull. Amer. Math. Soc. **81** (1975), 223–312.

The situation completely changes in renormalization theory. Here, we have a sophisticated approximation scheme in perturbed quantum field theory, but we do not know the correct global operator equation for the quantum field. It is a task for the future to clarify this situation.

17. Radiative Corrections of Lowest Order

> Unfortunately, the computations of radiative corrections in quantum electrodynamics are rather lengthy. For getting high accuracy by including multi-loop corrections, one needs sophisticated supercomputer programs.
>
> Folklore

In Chap. 15 we have studied some applications to physical processes by using the lowest order of perturbation theory. The next step is to compute corrections in higher order of perturbation theory. Such so-called radiative corrections suffer under the appearance of divergent algebraic Feynman integrals. Therefore, we have to use the method of renormalization in order to get finite expressions which depend on the free parameters m_{eff} and e_{eff}. These parameters must be determined by physical experiments. Typical radiative corrections in lowest possible order (so-called one-loop corrections) are:

- the radiative correction of the Coulomb potential (Uehling 1935);
- the Lamb shift of the spectrum of the hydrogen atom (Bethe 1947);
- the anomalous magnetic moment of the electron (Schwinger 1947);
- scattering of photons by photons (Euler 1936, Karpus and Neumann 1950).

For these physical effects, theoretical and experimental values are in extremely good agreement. The rather lengthy computations will be postponed to Vol. III. In this chapter, we will only discuss the main ideas.

17.1 Primitive Divergent Feynman Graphs

A Feynman graph is called divergent iff the corresponding algebraic Feynman integral is divergent. In quantum electrodynamics, the following divergent Feynman graphs play a special role:

- the primitive divergent self-energy graph of the photon (Fig. 17.1),
- the primitive divergent self-energy graph of the electron (Fig. 17.2),
- the primitive divergent vertex graph (Fig. 17.3), and

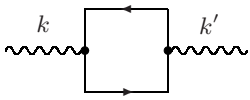


Fig. 17.1. Primitive divergent self-energy graph of the photon (vacuum polarization)

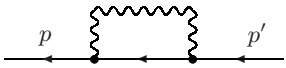


Fig. 17.2. Primitive self-energy graph of the electron

- the primitive divergent photon-photon scattering graph (see Fig. 17.9 on page 964).

These Feynman graphs are called primitive divergent graphs, since they are subgraphs of more complicated divergent Feynman graphs in quantum electrodynamics. Observe that these Feynman graphs have the following important property:

> *They are connected, and they remain connected after cutting any internal line.*

Such graphs are called one-particle irreducible graphs.

17.2 Vacuum Polarization

> Quantum theory so perfectly illustrates the fact that one might have understood a certain subject with complete clarity, yet at the same time knows that one can speak of it only allegorically and in pictures.
>
> Werner Heisenberg (1901–1976)

Physicists use Feynman diagrams in order to formulate the processes of quantum electrodynamics in a highly intuitive manner. Such pictures are very helpful for understanding typical quantum field effects in a heuristic way. The intuitive language used by physicists tells us the following:

- The internal lines of Feynman graphs correspond to virtual particles, since the Feynman rules tell us that conservation of energy and momentum are not valid for the integration over the functions corresponding to internal lines.
- The self-energy graph of the photon pictured in Fig. 17.1 describes the interaction of a photon with a virtual electron-positron pair. The existence of virtual electron-positron pairs is called *vacuum polarization* by physicists. Using an analogy with classical electrodynamics, physicists assume heuristically that the vacuum is filled with virtual electron-positron dipoles which cause a polarization of the vacuum.[1] We cannot directly observe the vacuum (i.e., the ground state of the quantum field); this is a virtual world. But, by quantum fluctuations, the virtual electron-positron dipoles of the vacuum may jump for a short time to our real world; this way, they add quantum corrections to real physical processes.
- The self-energy graph of the electron pictured in Fig. 17.2 describes the interaction of an electron with itself by emitting virtual photons "for a short time" and absorbing them again. Intuitively, this process is mainly responsible for the "electromagnetic mass" of the electron described by mass renormalization. This

[1] In classical electrodynamics, a pair of two charges, namely, a negative charge $-Q$ at the point $\mathbf{x}$ and a positive charge Q at the point $\mathbf{x} + \Delta\mathbf{x}$ is called a dipole if the distance $|\Delta\mathbf{x}|$ between the two charges is small. The polarization of macroscopic material used in technology corresponds to the existence of many molecular dipoles.

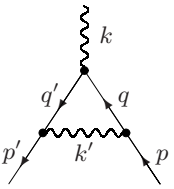


Fig. 17.3. Primitive divergent vertex graph

mass renormalization was first computed by Kramers in 1938 and by Weisskopf in 1939.[2]

17.3 Radiative Corrections of the Propagators

In 1949 Dyson discovered that both the regularized perturbed photon propagator and the regularized perturbed electron propagator possess a special multiplicative structure. This, together with the Ward identity for the vertex diagram, is responsible for the success of the renormalization of mass and electric charge of the electron in quantum electrodynamics. From the computational point of view, dimensional regularization is the most elegant regularization method. We have to distinguish between

- free Feynman propagators and
- full Feynman propagators.

Feynman propagators are also called 2-point correlation functions (or 2-point Green's functions). The free propagators describe the free quantum fields for photons, electrons, and positrons. These free propagators are basic ingredients of the Feynman rules summarized in Table 14.8 on page 896. The corresponding full propagators take interactions into account. This leads to corrections of the free propagators which are responsible for important physical effects called radiative corrections.

The basic formula for full propagators is the Gell-Mann–Low formula.

This formula, which we already encountered in Vol. I, expresses the full propagator by means of the S-matrix. Applying the Wick theorem to the Gell-Mann–Low formula, one obtains Feynman diagrams which graphically describe the full propagator. This will be studied in Vol. III. At this point, we will only indicate the Feynman diagrams for the corrections. More precisely, we will picture the perturbed propagators in lowest possible order, namely:

- the perturbed photon propagator (Fig. 17.4),
- the perturbed electron propagator (Fig. 17.5), and
- the perturbed vertex graph (Fig. 17.6).

[2] H. Kramers, The interaction between charged particles and the radiation field, Nuovo Cimento **15** (1938), 108–114 (in German).
V. Weisskopf, On the self-energy and the electromagnetic field of the electron, Phys. Rev. **56** (1939), 72–81.

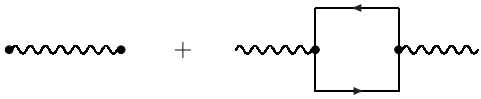


Fig. 17.4. Perturbed photon propagator

17.3.1 The Photon Propagator

According to the Feynman rules summarized in Table 14.8 on page 896, the free photon propagator is based on the family of functions

$$D^{\alpha\beta}(p) := -\frac{\mathrm{i}\eta^{\alpha\beta}}{(2\pi)^4} \cdot \frac{1}{(p^0)^2 - (\mathcal{E}_{\mathbf{p}} - \varepsilon\mathrm{i})^2}, \qquad \alpha, \beta = 0, 1, 2, 3,$$

with the photon energy $\mathcal{E}_{\mathbf{p}} := \sqrt{\mathbf{p}^2 + m_{ph}}$. Here, m_{ph} is the virtual photon mass, and $\varepsilon > 0$ is the regularization parameter. The perturbed photon propagator corresponds to the Feynman diagram depicted in Fig. 17.4. Applying the Feynman rules to the Feynman diagram and using dimensional regularization for the corresponding algebraic Feynman integrals, a lengthy computation shows that we have to replace $D^{\alpha\beta}$ by the functions

$$D^{\alpha\beta}_{\mathrm{pert}}(p) = Z_D(\nu) D^{\alpha\beta}_{\mathrm{ren}}(p; \alpha), \qquad \alpha, \beta = 0, 1, 2, 3.$$

Here, the so-called renormalized propagator function $D^{\alpha\beta}_{\mathrm{ren}}$ is a first-order perturbation of $D^{\alpha\beta}$ (with respect to the fine structure constant α). The explicit computation can be found in Vol. III.

From the physical point of view, the functions $D^{\alpha\beta}_{\mathrm{ren}}$ take the vacuum polarization into account.

Observe that the real number Z_D depends on the parameter ν of dimensional regularization. We have the limit

$$\lim_{\nu \to +0} Z_D(\nu) = \infty. \tag{17.1}$$

The point is that the divergent behavior of the Feynman diagram is concentrated in the factor $Z_D(\nu)$. We will discuss in Sect. 17.4 that this infinite limit can be compensated by counterterms of the Lagrangian density. In other words, the infinite limit (17.1) does not effect quantities which can be measured in physical experiments.

17.3.2 The Electron Propagator

Parallel to the preceding section, the free electron propagator is based on the function

$$S(p) := \frac{\mathrm{i}}{(2\pi)^4} \cdot \frac{\not{p} + m_e I}{(p^0)^2 - (E_{\mathbf{p}} - \varepsilon\mathrm{i})^2}$$

by Table 14.8 on page 896. Here, $E_{\mathbf{p}} := \sqrt{p^2 + m_e^2}$ is the electron energy with respect to the bare electron mass m_e, and ε is the positive regularization parameter. The perturbed electron propagator corresponds to the Feynman diagram depicted in Fig. 17.5. Applying the Feynman rules to the Feynman diagram and using dimensional regularization for the corresponding algebraic Feynman integrals, a lengthy computation shows that we have to replace $S(p)$ by the function

Fig. 17.5. Perturbed electron propagator

$$S_{\text{pert}}(p) = Z_S(\nu) S_{\text{ren}}(p; m_{\text{eff}}, \alpha).$$

Here, S_{ren} is a first-order perturbation of S (with respect to the fine structure constant α). Note that:

> *The renormalized propagator function S_{ren} does not depend on the bare electron mass m_e, but on the new mass parameter m_{eff}.*

This fact is called mass renormalization. The explicit computation can be found in Vol. III. There holds the limit

$$\lim_{\nu \to +0} Z_S(\nu) = \infty. \tag{17.2}$$

17.3.3 The Vertex Correction and the Ward Identity

The approach to quantum electrodynamics is substantially simplified by the fact that there exists a relation between the algebraic Feynman integrals corresponding to the perturbed electron propagator (Fig. 17.5) and the perturbed vertex graph (Fig. 17.6). This crucial relation is called the Ward identity. The explicit form will be given in Vol. III.

> *The Ward identity is closely related to the gauge invariance of quantum electrodynamics.*

In particular, the Ward identity implies the crucial relation (17.3) below concerning the multiplicative renormalization constants.

17.4 The Counterterms of the Lagrangian and the Compensation Principle

In Sect.16.3.6 we have introduced the multiplicative renormalization constants

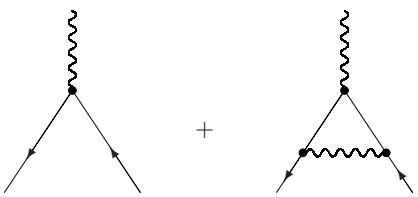


Fig. 17.6. Perturbed vertex graph

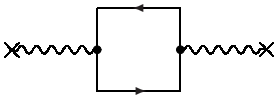


Fig. 17.7. Radiative correction of the external electromagnetic field

$$Z_e, Z_m, Z_A, Z_\psi$$

related to counterterms of the Lagrangian density in quantum electrodynamics.

The crucial point is that these renormalization constants can be used to compensate the factors $Z_D(\nu)$ and $Z_S(\nu)$ related to the perturbed photon propagator and the perturbed electron propagator, respectively.

In addition, the Ward identity yields the following crucial relation:

$$\boxed{Z_e = Z_A^{-1/2}.} \tag{17.3}$$

Observe the following two key points:

(i) The transition amplitudes and hence the transition probabilities for real physical processes do not depend on the multiplicative renormalization constants $Z_e, Z_m, Z_A, Z_\psi, Z_D, Z_S$; they cancel each other

(ii) The renormalization constants go to infinity as the parameter ν of dimensional regularization goes to zero. However, because of the cancellation of the multiplicative renormalization constants, the limit $\nu \to 0$ yields finite values of the transition amplitudes.

Computations in quantum electrodynamics are frequently simplified by surprising cancellations. As a rule, the Ward identity (and the more general Ward–Takahashi identities) are responsible for these cancellations.

17.5 Application to Physical Problems

The ways of people to the laws of nature are not less admirable than the laws themselves.

Johannes Kepler (1571–1630)

17.5.1 Radiative Correction of the Coulomb Potential

Consider the primitive divergent self-energy graph of the photon concerning vacuum polarization depicted in Fig. 17.1. Replacing the photon line by an external electromagnetic field described by the 4-potential $U, \mathbf{A}$, we get the Feynman graph depicted in Fig. 17.7. This diagram describes the radiative correction of an external electromagnetic field by vacuum polarization. Using the Coulomb potential of an electric charge Q,

$$U(\mathbf{x}) = \frac{Q}{4\pi|\mathbf{x}|},$$

the perturbed potential under the influence of vacuum polarization reads as

$$U_{\text{pert}}(\mathbf{x}) = \frac{Q}{4\pi|\mathbf{x}|}\left(1 + \frac{2\alpha}{3\pi}\int_1^\infty \frac{\sqrt{\xi^2-1}}{\xi^2}\Big(1+\frac{1}{2\xi^2}\Big)\,\mathrm{e}^{-2m_{\text{eff}}|\mathbf{x}|\cdot\xi}\,d\xi\right).$$

This is the so-called Uehling potential.[3] The corresponding electric field is given by $\mathbf{E}(\mathbf{x}) = -\,\mathbf{grad}\,U(\mathbf{x})$. Let us assume that the point $\mathbf{x}$ is far away from the origin, that is, $m_{\text{eff}}|\mathbf{x}| \gg 1$. Then we obtain the following approximation of the Uehling potential:

$$\boxed{U_{\text{pert}}(\mathbf{x}) = \frac{Q}{4\pi|\mathbf{x}|}\left(1 + \frac{\alpha}{4\sqrt{\pi}}\frac{\mathrm{e}^{-2m_{\text{eff}}|\mathbf{x}|}}{(m_{\text{eff}}|\mathbf{x}|)^{3/2}}\right).}$$

The correction term decays exponentially as $|\mathbf{x}| \to \infty$. The detailed computation of U_{pert} will be given in Vol. III.

17.5.2 The Anomalous Magnetic Moment of the Electron

> The most spectacular experiment for the anomalous magnetic moment of the electron is based on measurements performed on a single electron, caught in an electromagnetic trap.[4]
>
> Martinus Veltman, 2003

The relativistic Dirac equation implies that the electron possesses an internal angular momentum called spin (see Vol. III). This leads to a magnetic moment of the electron, which can be measured if the electron moves in an external magnetic field. Explicitly, the magnetic moment vector of the electron is given by the following formula in the SI system:

$$\mathbf{M} = -\frac{e_{\text{eff}}}{2m_{\text{eff}}}\cdot g_e\mathbf{S}$$

with the so-called gyromagnetic factor

$$g_e = 2(1+a).$$

Here, m_{eff} (resp. $-e_{\text{eff}}$) is the effective mass (resp. effective charge) of the electron. The spin vector $\mathbf{S}$ has the length $\hbar/2$.

As we will show in Vol. III, the relativistic Dirac equation yields the value $a = 0$. In 1947, using sophisticated arguments in the setting of quantum electrodynamics, Schwinger computed the value

$$a = \frac{\alpha}{2\pi}. \tag{17.4}$$

The deviation of the gyromagnetic factor g_e from Dirac's value $g_e = 2$ is called the *anomalous* magnetic moment of the electron.

---

[3] We use the energetic system of physical units.

[4] For trapping single atoms, Hans Dehmelt (born 1922) and Wolfgang Paul (1913–1993) were awarded the Nobel prize in physics in 1989.
M. Veltman, Facts and Mysteries in Elementary Particle Physics, World Scientific, Singapore, 2003.
In 1999, Gerardus 't Hooft (born 1946) and Martinus Veltman (born 1931) were awarded the Nobel prize in physics for elucidating the quantum structure of electroweak interaction in physics via renormalization theory.

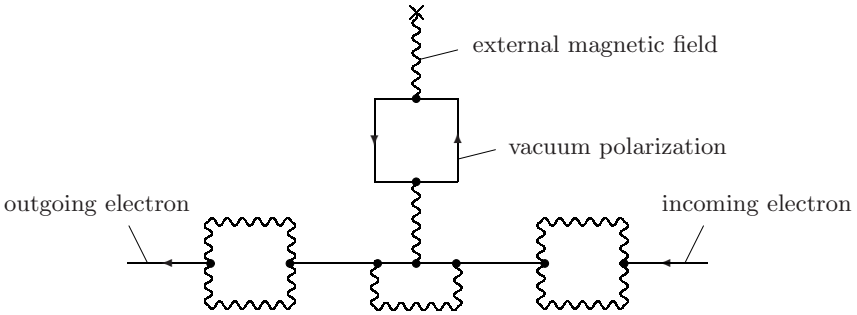


Fig. 17.8. Electron in an external magnetic field

The theoretical value via renormalization. As we will explicitly show in Vol. III, Schwinger's famous correction term (17.4) can be obtained by using the techniques of renormalization theory. The idea is to study the motion of an electron in an external magnetic field. One has to use Feynman graphs of the type schematically pictured in Fig. 17.8.

Using renormalization theory in higher orders of perturbation theory, one gets the following value:

$$a_{\text{theor}} = \frac{\alpha}{2\pi} - 0.328\,478\,965\left(\frac{\alpha}{\pi}\right)^2 + (1.175\,62 \pm 0.000\,56)\left(\frac{\alpha}{\pi}\right)^3 - (1.472 \pm 0.152)\left(\frac{\alpha}{\pi}\right)^4 .$$

For computing such a high-accuracy result based on about 900 complicated Feynman diagrams, one needs sophisticated computer programs and years of supercomputer time (see Sect. 18.4).

The experimental value. In 1947 Kusch, working at Columbia University, New York, measured the following value:[5]

$$a_{\text{exp}} = 0.001.$$

In 2006 the following unbelievable precise value was measured:

$$a_{\text{exp}} = 0.001\,159\,652\,180\,85(\pm 76).$$

See the references Odom et al. (2006) and Gabrielse et al. (2006) quoted on page 961. The American Institute of Physics called this "the outstanding physics achievement of 2006."

> *There is a fantastic coincidence between the experimental value* a_{exp} *and the theoretical value* a_{theo}.

This is a highlight in quantum electrodynamics.

[5] In 1955, Polykarp Kusch (1911–1993) was awarded the Nobel prize in physics for his precision determination of the magnetic moment of the electron.

17.5.3 The Anomalous Magnetic Moment of the Muon

> The sophistication in both theory and experiment for the anomalous magnetic moment of the muon is mind boggling. The computation of the coefficient of α^2 has taken some 20 years, involving some 72 Feynman diagrams, while the computation of the α^4 term (891 Feynman diagrams) has been done mainly by numerical approximation methods, using up years of supercomputer time.[6]
>
> Martinus Veltman, 2003

For the magnetic moment vector of the muon, renormalized perturbation theory yields the value

$$\mathbf{M}_\mu = -\frac{e_{\text{eff}}}{2m_{\mu,\text{eff}}} \cdot g_\mu \mathbf{S}$$

along with the gyromagnetic factor $g_\mu = 2(1 + a)$ with

$$a = 0.001\,165\,920\ \mathit{500} \pm 0.000\,000\,000\ \mathit{460}.$$

Here, e_{eff} is the effective charge of the electron, and $m_{\mu,\text{eff}}$ is the effective mass of the muon. The necessary huge computations take all of the interactions of the Standard Model in particle physics into account. The main contribution comes from quantum electrodynamics. Corrections are due to strong and weak interaction. The experimental value reads as

$$a_{\text{exp}} = 0.001\,165\,915\ \mathit{97} \pm 0.000\,000\,000\ \mathit{67}.$$

Therefore, the experimental and the theoretical result agree within 9 significant digits. This serves as a high-precision test for the Standard Model in particle physics and the method of renormalized Feynman diagrams. This is a highlight in the Standard Model in particle physics.

Hints for further reading. For high-precision tests in quantum electrodynamics and in the Standard Model of particle physics, we refer to:

F. Scheck, Radiative corrections confronted with experiment: a survey, 15 pages. In: F. Scheck (Ed.), Theory of Renormalization and Regularization, Lecture Notes, Hesselberg Workshop 2002 (Germany).
Internet: `http://www.thep.physik.uni.mainz.de~scheck/Hessbg`

T. Kinoshita, Quantum Electrodynamics, World Scientific, Singapore, 1990.

B. Odom, D. Hanneke, B. D'Urso, and G. Gabrielse, New measurement of the electron magnetic moment using a one-electron quantum cyclotron, Phys. Rev. Lett. **97** (2006), 030801.

G. Gabrielse, D. Hanneke, T. Kinoshita, M. Nio, and B. Odom, New determination of the fine structure constant from the electron g value and QED, Phys. Rev. Lett. **97** (2006), 030802, and Erratum (2007).

M. Knecht, The anomalous moment of the electron and the muon. In: Poincaré Seminar 2002: Vacuum Energy – Renormalization. Edited by B. Duplantier and V. Rivasseau, Birkhäuser, Basel, 2003, pp. 265–310.

W. Hollik and G. Duckeck, Electroweak Precision Tests at LEP, Springer, Berlin, 2000.

[6] M. Veltman, Facts and Mysteries in Elementary Particle Physics. Reprinted by permission of World Scientific Publishing Co. Pte. Ltd., Singapore, 2003.

S. Narison, QCD (Quantum Chromodynamics) as a Theory of Hadrons: From Partons to Confinement, Cambridge University Press, 2004. (Chapter 14 is devoted to high-precision tests in QED.)

M. Awramik and M. Czakon, Complete two-loop electroweak contributions to the muon lifetime in the standard model, Phys. Lett. **B568** (2003), p. 48.

17.5.4 The Lamb Shift

On April 26, 1947, Lamb and Retherford experimentally discovered some fine structure in the spectrum of the hydrogen atom, which is called the Lamb shift nowadays; they used microwave technique.[7] This spectacular experimental result was reported at the famous Shelter Island Conference in 1947 (Long Island, New York State). A few days after the conference, Hans Bethe (1906–2005), who was amazed by the discovery, did the nonrelativistic computation of the Lamb shift during a train ride from New York to Schenectady. Bethe used second-order perturbation theory for the Schrödinger equation. In order to master divergent integrals, Bethe applied Kramers' approach of mass renormalization to the electron of the hydrogen atom.[8] In his beautiful book on the history of quantum electrodynamics, Silvan Schweber reports the following on Bethe's computation of the Lamb shift:[9]

> Bethe was not quite confident of the accuracy of his formula, because he was not quite sure of the correctness of a factor of $\sqrt{2}$ in his expansion of the radiation operator in terms of creation and annihilation operators. This he checked on Monday morning in Heitler's book.[10] He also got Miss Steward and Dr. Stehn to evaluate numerically the missing mean excitation energy for the $2s$ state of the electron of the hydrogen atom. It was found to be 17.5 R (Rydberg constant), "an amazingly high value." Inserting this in his formula, Bethe found
>
> $$\Delta E_{\text{Lamb}} = 1040 \text{ MHz}$$
>
> for the Lamb shift "in excellent agreement with the observed 1000 MHz (Mega Hertz)"...
> On the occasion of Lamb's sixtieth birthday in 1973, Freeman Dyson wrote to congratulate him. He noted: "Your work on the hydrogen fine structure led directly to the wave of progress in quantum electrodynamics on which I took a ride to fame and fortune. You did the hard, tedious, exploratory work... Those years, when the Lamb shift was the central theme of physics, were golden years for the physicists of my generation. You were the first

[7] In 1955, Willis Eugene Lamb (born 1913) was awarded the Nobel prize in physics for his discoveries concerning the fine structure of the hydrogen spectrum.

[8] See the paper by Kramer (1938) quoted on page 955. Bethe's argument can be found in F. Gross, Relativistic Quantum Mechanics and Field Theory, Wiley, New York, 1993.

[9] Schweber, Silvan, QED and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga, Princeton University Press, Princeton, New Jersey, 1994. Reprinted by permission of Princeton University Press.

[10] W. Heitler, The Theory of Radiation, Clarendon Press, Oxford, 1936. For one generation of physicists, Heitler's book was the bible in radiation theory. The modern version of Heitler's classics is represented by E. Pike and S. Sarkar, The Quantum Theory of Radiation, Clarendon Press, Oxford, 1995.

to see that that tiny shift, so elusive and hard to measure, would clarify in a fundamental way our thinking about particles and fields."

Using the relativistic renormalization technique in quantum electrodynamics, Weisskopf and French obtained the value 1051.13 MHZ in 1948, whereas Feynman and Schwinger computed a slightly smaller (incorrect) value. Today, the theoretical value is equal to

$$\Delta E_{\text{Lamb,theor}} = (1057.855 \pm 0.014) \text{ MHz}.$$

Here, vacuum polarization contributes the value of -27.1 MHz.[11] The modern experimental value is equal to

$$\Delta E_{\text{Lamb,exp}} = (1057.845 \pm 0.009)) \text{ MHz}.$$

Welton's semiclassical model on vacuum effects and pseudo-Brownian motion. It is quite interesting that the Lamb shift can be intuitively understood as a kind of Brownian motion of the electron. Using some heuristic model, Welton argued as follows:[12] The electron moves stochastically in the sea of virtual photons of the vacuum under the influence of quantum fluctuations (like the classical Brownian particle moves under the influence of the collisions with the fluid particles). Using the energy of the photon quantum field, the computation shows that the averaged square of the position fluctuation of the pseudo-Brownian motion of the electron is given by the following formula:[13]

$$\overline{(\Delta \mathbf{x})^2} = \frac{2}{\pi} \left(\frac{e_{\text{eff}}}{m_{\text{eff}}} \right)^2 \int_0^\infty \frac{dE}{E}.$$

Note that the energy integral is divergent. In order to get a finite result, we use an energy cut-off, that is, we replace the integral by

$$\overline{(\Delta \mathbf{x})^2} = \frac{2}{\pi} \left(\frac{e_{\text{eff}}}{m_{\text{eff}}} \right)^2 \int_{E_{\min}}^{E_{\max}} \frac{dE}{E}.$$

We argue that the energy of the electron is actually limited by the specific physical situation. As upper bound, we choose the rest energy of the electron, $E_{\max} := m_{\text{eff}}$, and as lower bound we choose the lowest possible energy of the electron in the hydrogen atom on the nth orbit,

$$E_{\min} = \frac{m_{\text{eff}} e_{\text{eff}}^4}{2n^2}, \qquad n = 1, 2, \ldots$$

The position fluctuation changes the potential energy of the electron; this energy change is responsible for the Lamb shift. The explicit computation for $n = 2$ and the s-state of the electron (i.e., the angular momentum vanishes, $l = 0$) yields indeed Bethe's value of 1040 MHz (see Vol. III for details).

Perspectives. In Vol. III we will thoroughly study the spectrum of the hydrogen atom including the Lamb shift. This is a fascinating story in both physics and mathematics. We will study the following approaches:

- the classical Kepler motion based on the rotational symmetry of the gravitational field of the sun (the $SO(3)$-Poisson algebra);

[11] We refer to W. Greiner and J. Reinhardt, Quantum Electrodynamics, Springer, Berlin, 1996.

[12] T. Welton, Some observable effects of the quantum-mechanical fluctuations of the electromagnetic field, Phys. Rev. **74** (1948), 1157–1167.

[13] We use the energetic system of physical units.

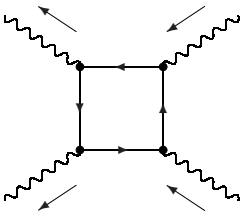


Fig. 17.9. The primitive divergent photon-photon scattering graph

- the classical Kepler motion and the Runge–Lenz vector as conserved quantity (the $SO(4)$-Poisson algebra);
- the classical Rutherford scattering formula from 1911,
- the semiclassical approach via the Bohr–Sommerfeld quantization rule in the phase space from 1916;
- Pauli's nonrelativistic quantum approach to the bound states of the hydrogen atom from 1926 based on the commutation relations for the $SO(4)$-Lie algebra;
- the Weyl–Kodaira spectral theory for singular differential operators and Schrödinger's nonrelativistic approach to both bound states and scattering states of the hydrogen atom from 1926 via the Schrödinger equation;
- Darwin's and Gordon's relativistic approach to the hydrogen atom via the Dirac equation from 1928;
- von Neumann's spectral theory for self-adjoint operators and the Hamiltonian of both the nonrelativistic and the relativistic hydrogen atom;
- Gelfand's theory of generalized eigenfunctions (distributions) for self-adjoint operators and the nonrelativistic hydrogen atom;
- Schwinger's Green's function for the nonrelativistic hydrogen atom from 1951 (including both bound states and scattering states), and the cross section for electron scattering processes;
- the Duru–Kleinert approach to the Feynman path integral for the nonrelativistic hydrogen atom from 1979 (including both bound states and scattering states);
- the Steiner approach to the Feynman path integral for the nonrelativistic hydrogen atom from 1984;
- the tricky supersymmetric approach to the bound states of the nonrelativistic hydrogen atom;
- Welton's semiclassical vacuum-energy model for the Lamb shift;
- renormalization theory in relativistic quantum electrodynamics and the Lamb shift.

Interestingly enough, Feynman, himself, did not succeed in computing the path integral for the nonrelativistic hydrogen atom; he gave this task to the young physicist Hagen Kleinert in 1972. To solve this problem, the main trick of Kleinert was to use a pseudo-time and a canonical transformation known in celestial mechanics in order to get a Hamiltonian of quadratic type which leads to a Gaussian integral.

17.5.5 Photon-Photon Scattering

The Maxwell equations represent a linear system of partial differential equations for the electromagnetic field. In this classical setting, the superposition principle is valid for electromagnetic waves; therefore, light-light scattering is impossible.

Light-light scattering (i.e., photon-photon scattering) is a typical quantum effect which can be described by quantum electrodynamics. A typical Feynman graph is pictured in Fig. 17.9. Since the corresponding algebraic Feynman integral is divergent, we have to apply the methods of renormalization theory. For a detailed study, we refer to L. Landau and M. Lifshitz, Course of Theoretical Physics, Vol 4: Quantum Electrodynamics, Sect. 124, Butterworth–Heinemann, Oxford, 1982. The cross section for photon-photon scattering is given by

$$\sigma = 10^{-72}\ \mathrm{cm}^2.$$

This is an extremely small value; therefore, an experimental verification of photon-photon scattering is still missing.

18. A Glance at Renormalization to all Orders of Perturbation Theory

The devil rides high on detail.
Folklore

In this chapter, we will sketch the basic ideas of general renormalization theory. A detailed study will be postponed to Vol. IV on quantum mathematics. Renormalization theory cannot be understood without knowing its long and strange history. At this point we merely restrict ourselves to a few quotations and comments. In the next chapter, we will sketch some basic ideas together with detailed hints for further reading.

The following quotation is taken from the famous article *Orientation* written by Arthur Wightman as an introduction to the volume *Renormalization Theory* edited by G. Velo and A. Wightman, Reidel, Dordrecht, pp. 1–24 (reprinted with permission). This volume describes the state of the art in renormalization theory in 1976.

> Renormalization theory is a notoriously complicated and technical subject... The prudent student would do well to distinguish sharply between what has been proved and what has been plausible, and in general he should watch out!
> My first cautionary tale has to do with the early days of renormalization theory. When Freeman Dyson analyzed the renormalization theory of the S-matrix for the quantum electrodynamics of spin one-half particles in his two great papers,[1] he laid the foundations for most later work on the subject, but this treatment of one phenomenon, *overlapping divergences*, was incomplete. Among the methods offered to clarify the situation, that of John Ward[2] seemed outstandingly simple, so much so that it was adopted in Jauch and Rohrlich's standard textbook.[3] Several years later, Mills and Yang noticed that unless further refinements are introduced the method does not work for the photon self-energy.[4] The lowest order for which the trouble manifests itself is the fourteenth (e.g., this concerns the graph depicted in Fig. 18.1). Mills and Yang repaired the method and sketched

[1] F. Dyson, The radiation theories of Tomonaga, Schwinger, and Feynman, Phys. Rev. **75** (1949), 406–502.
F. Dyson, The S-matrix in quantum electrodynamics, Phys. Rev. **75** (1949), 1736–1755.

[2] J. Ward, On the renormalization of quantum electrodynamics, Proc. Phys. Soc. London **A64** (1951), 54–56.

[3] J. Jauch and F. Rohrlich, The Theory of Photons and Electrons, Addison Wesley, New York, 1955.

[4] See T. Wu, Theory of pion-pion interaction. I. Renormalization, Phys. Rev. **125** (1962), 1436–1450.

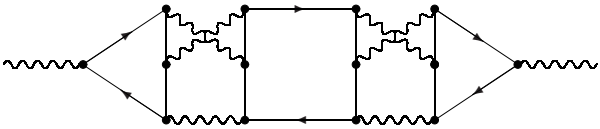


Fig. 18.1. Critical self-energy graph of the photon

some of the steps in a proof that it would actually yield a finite renormalized amplitude.[5] An innocent, reading the book by Jauch and Rohrlich, would never suspect that such refinements are necessary.

Another attempt to cope with the overlapping divergences was made by Salam.[6] I will not describe it, if for no other reason that I never have succeeded in understanding it... The belief is widespread that when Salam's work is combined with significant work by Steven Weinberg,[7] the result should be a mathematically coherent version of renormalization theory. At least that is what one reads in the textbook by Bjorken and Drell for quantum electrodynamics[8]...

Another foundation for renormalization theory with a rather different starting point was put forward by Stueckelberg and Green.[9] It was refounded and brought to a certain stage of completion in the standard textbook of Bogoliubov and Shirkov in 1957.[10] The mathematical nut that had to be cracked is in the paper of Bogoliubov and Parasiuk[11] (amazingly, not quoted in the 1959 English translation of the 1957 monograph by Bogoliubov and Shirkov). This paper introduces a systematical and analytic scheme for overcoming the overlapping divergence problem. This paper is very important for later developments. Unfortunately, it was found by Klaus Hepp[12] that Theorem 4 of the Bogoliubov/Parasiuk paper is false, and that consequently the proof of the main result is incomplete as it stands. However, Hepp showed that Theorem 4 is not essential to derive the main result and he could fill all the gaps. Thus, it is appropriate to introduce the initials BPH to stand for the renormalization method described in Bogoliubov/Parasiuk (1957) and Hepp (1966).[13]

[5] R. Mills and C. Yang, Progr. Theor. Phys. Supp. **37** (1966), 507–511.

[6] A. Salam, Phys. Rev. **82** (1951), 217–227; **84** (1951), 426–431.

[7] S. Weinberg, High energy behavior in quantum field theory, Phys. Rev. **118** (1960), 838–849.

[8] J. Bjorken and S. Drell, Relativistic Quantum Fields, McGraw-Hill, New York, 1965.

[9] E. Stueckelberg and T. Green, Helv. Phys. Acta **24** (1951), 153–174.

[10] N. Bogoliubov and D. Shirkov, Introduction to Quantum Field Theory, Nauka, Moscow, 1957. Fourth expanded Russian edition 1984. English translation: Interscience, New York, 1980 (translation of the third Russian edition published in 1973).

[11] N. Bogoliubov and O. Parasiuk, On the multiplication of propagators in quantum field theory, Acta Math. **97** (1957), 227–326 (in German).

[12] K. Hepp, Proof of the Bogoliubov–Parasiuk theorem on renormalization, Commun. Math. Phys. **2** (1966), 301–326.

[13] In 1969 Wolfhart Zimmermann discovered an explicit formula for Bogoliubov's iterative method called the *Zimmermann forest formula* (see page 984). There-

> An alternative method to BPHZ is *analytic renormalization* due to Eugene Speer.[14]... Analytic renormalization was later shown to be equivalent to BPHZ renormalization.[15] ...
> To make the Feynman history integral rigorous,[16] Gelfand and Yaglom proposed to regard it as the limit of integrals in which Planck's constant h has $-i\delta$. They argued that this gives rise to a *complex measure* on paths: ... "It is natural that such a complex measure for arbitrary $\delta > 0$ will be just as 'good' as Wiener measure, that is, it will have just a precise a meaning as measure in the space of continuous functions, and it will allow integration over it of a wide class of functionals including all continuous and bounded functionals"[17]... It turned out, alas, that this statement is wrong. It was shown by Cameron[18] that this proposal defines a completely additive complex measure only when $h = i\delta$ is purely imaginary ($\delta > 0$), that is, for the case considered by Wiener.[19] From a practical point of view this means that one does not have available all the powerful analytical devices of the theory of integration.
>
> Arthur Wightman, 1976

The following quotation is taken from J. Collins, *Renormalization*, Cambridge University Press, New York, 1984 (reprinted with permission). We recommend this book as an introduction to renormalization theory (see also the hints for further reading on page 1029). However, the reader should note that this book was written before Kreimer's Hopf algebra revolution in renormalization theory in 1998 (see Sect. 19.3).

> The structure of a quantum field theory often simplifies when one considers processes involving large momenta (i.e., large energies) or short distances. These simplifications are important in improving one's ability to calculate predictions from the theory, and in essence form the subject of this book... The subjects of renormalization, the renormalization group, and the operator product expansion are intimately linked, and we will treat them all in this book. The aim will be to explain the general methods that are not only applicable to the examples we will examine but in many other situations. We will not aim at complete rigor. However, there are many pitfalls and traps ready to ensnare an unwary physicist. Thus, a precise set of concepts and notations is necessary, for many of the dangers are es-

fore, nowadays the approach is characterized by the initials BPHZ (Bogoliubov, Parasiuk, Hepp, Zimmermann). See W. Zimmermann, Convergence of Bogoliubov's method of renormalization in momentum space, Commun. Math. Phys. **15** (1969), 208–234.

[14] E. Speer, Generalized Feynman Amplitudes, Annals of Mathematical Studies **62**, Princeton University Press (see Sect. 19.7).

[15] K. Hepp, Commun. Math. Phys. **14** (1969), 67–69.

[16] This integral is also called the Feynman path integral or the Feynman functional integral.

[17] I. Gelfand and A. Yaglom, Integration in functional spaces and its applications in quantum physics, J. Math. Phys. **1** (1960), 48–69 (translated from Uspekhi Mat. Nauk **11** (1956), 77–114 (in Russian).

[18] R. Cameron, A family of integrals serving to connect the Wiener and Feynman integrals, J. of Math. and Phys. Sci. of MIT **39** (1960), 126–140 (see also G. Johnson and M. Lapidus, The Feynman Integral and Feynman's Operational Calculus, Clarendon Press, Oxford, 2000).

[19] N. Wiener, Differential space, J. Math. and Phys. of MIT **2** (1923), 131–174.

sentially *combinatorial*. The appropriate basis is then that of Zimmermann (1969)/(1973).[20]
One other problem is that of choice of an ultra-violet (i.e., high-energy) cut-off. From a fundamental point of view, the lattice cut-off seems best as it appears in nonperturbative treatment using the functional integral.[21] In perturbation theory one can arrange to use no regulator whatsoever.[22] In practice, dimensional regularization has deservedly become very popular. This consists of replacing the physical space-time dimensionality 4 by an arbitrary complex number d. The main attraction of this method is that virtually no violence is done to the structure of a Feynman graph; a second attraction is that it also regulates infra-red (i.e., low-energy) divergences... The disadvantage is that this method has not been formulated outside of perturbation theory (at least not yet). Much of the treatment in this book, especially the examples, will be based on the use of dimensional regularization. However, it cannot be emphasized too strongly that none of the fundamental results depend on this choice.

John Collins, 1984

18.1 One-Particle Irreducible Feynman Graphs and Divergences

Divergences in quantum field theory are only caused by one-particle irreducible Feynman subgraphs.

The rule of thumb

Intuitively, a Feynman graph in quantum electrodynamics (also called Feynman diagram) consists of a finite number of vertices and connecting internal lines, as well as external lines. The electron and positron lines are oriented, whereas the photon lines are not oriented:

- external lines describe incoming and outgoing real particles;
- internal lines describe virtual particles.

In contrast to internal lines, external lines have only one vertex. The graph is called amputated iff it has no external lines. Such graphs describe processes for virtual particles in the vacuum (ground state). Intuitively, such processes correspond to quantum fluctuations of the vacuum which essentially influence the behavior of elementary particles in our real world. Analytically, correlation functions (also called

[20] W. Zimmermann, Convergence of Bogoliubov's method of renormalization in momentum space, Commun. Math. Phys. **15** (1969), 208–234.
W. Zimmermann, Local operator products and renormalization in quantum field theory. In: S. Deser et al. (Eds.), Lectures on Elementary Particles and Quantum Field Theory, Proceedings of the 1970 Brandeis Summer Institute in Theoretical Physics, MIT Press, Cambridge, Massachusetts, pp. 399–589.
W. Zimmermann, Normal products and the short-distance expansion in the perturbation theory of renormalizable interactions, Annals of Physics **77** (1973), 536–569; 570–601.

[21] See J. Glimm and A. Jaffe, Mathematical Methods of Quantum Physics: A Functional Integral Point of View, Springer, New York, 1981.

[22] See O. Piguet and A. Rouet, Symmetries in perturbative quantum field theory (pQFT), Phys. Rep. **76** (1981), 1–77.

(a) connected (b) disconnected

Fig. 18.2. Connectedness of graphs

Green's functions) correspond to amputated graphs. Furthermore, for two Feynman graphs Γ_1 and Γ_2, we write

$$\Gamma_1 \subseteq \Gamma_2$$

iff Γ_1 is a subgraph of Γ_2, that is, the vertices and lines of Γ_1 are also vertices and lines of Γ_2 (by taking the orientation of the lines into account). Moreover, we write

$$\Gamma_1 \cap \Gamma_2 = \emptyset$$

iff Γ_1 and Γ_2 have no common vertices and lines.

Connectedness of Feynman graphs. One has to distinguish between the following types of Feynman graphs:

(i) Connectedness: connected or disconnected.
(ii) Irreducibility: one-particle irreducible or reducible.
(iii) Overlapping subgraphs.

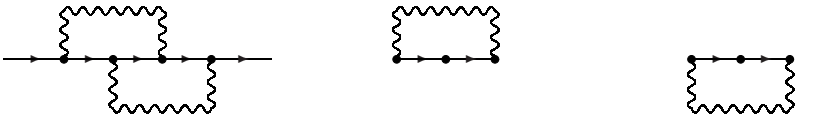
Let us discuss this. The graph is called disconnected iff it consists of at least two disjoint parts (Fig. 18.2(b)). Otherwise the graph is called connected (Fig. 18.2 (a)).

The graph is called one-particle irreducible iff it remains connected after cutting any internal line. Otherwise the graph is called reducible. Let us consider some examples:

(R) The graph depicted in Fig. 18.3(b) is reducible.
(I) The graph depicted in Fig. 18.3(a) is one-particle irreducible. The same property have the following Feynman graphs in quantum electrodynamics:
- the primitive divergent photon self-energy graph (vacuum polarization) (Fig. 17.1 on page 953),
- the primitive divergent electron self-energy graph (Fig. 17.2 on page 954),
- the primitive divergent vertex graph (Fig. 17.3 on page 955),
- the primitive divergent photon-photon scattering graph (Fig. 17.9 on page 964),
- the critical photon self-energy graph depicted in Fig. 18.1 on page 968 is one-particle irreducible.

(a) one-particle irreducible (b) one-particle reducible

Fig. 18.3. Irreducibility of graphs



(a) graph Γ (b) subgraph Γ_1 (c) subgraph Γ_2

Fig. 18.4. Overlapping subgraphs Γ_1 and Γ_2

The importance of one-particle irreducible graphs for quantum field theory will be explained in Sect. 19.6. Roughly speaking, the following hold:

- The one-particle irreducible graphs form the basis elements of the quantum field theory under consideration.
- Intuitively, these basis elements represent the basic correlations of the quantum fluctuations.
- If we know these basis elements, all the important information for the quantum processes can be obtained from them.
- This concerns the correlation functions (i.e., the Green's functions) and the transition amplitudes of the S-matrix for computing scattering processes of quantum particles.
- Only the algebraic Feynman integrals corresponding to these basic graphs have to be renormalized.

Two subgraphs Γ_1 and Γ_2 of the graph Γ are called overlapping iff the following three situations are excluded:

$$\Gamma_1 \subseteq \Gamma_2, \quad \Gamma_2 \subseteq \Gamma_1, \quad \Gamma_1 \cap \Gamma_2 = \emptyset.$$

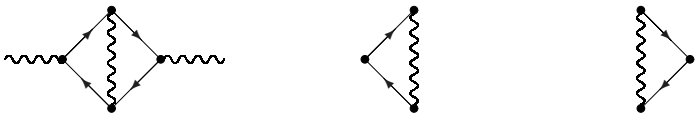
The prototype of a graph Γ with overlapping subgraphs Γ_1 and Γ_2 is depicted in Fig. 18.4. Furthermore, the subgraphs Γ_1 and Γ_2 of Γ depicted in Fig 18.5 are also overlapping subgraphs of Γ. Overlapping subgraphs caused a lot of trouble in the history of renormalization theory. In what follows we want to explain this.

18.2 Overlapping Divergences and Manoukian's Equivalence Principle

The following three quotations should help the reader to understand the difficulties that physicists encountered in the past in order to understand the concept of renormalization. The first quotation is taken from the beautiful history of quantum field theory written by S. Schweber, *QED and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga,* Princeton University Press, 1994 (reprinted by permission of both Princeton University Press and World Scientific Publishing Co. Pte. Ltd. Singapore).

> Overlapping divergences had reared their ugly head in Dyson's QED paper.[23] He had noted that a general self-energy graph could be regarded as an insertion of a modified vertex at either end of the lowest order self-energy graph. Insertion of modified vertices at both ends would correspond

[23] F. Dyson, The S-matrix in quantum electrodynamics, Phys. Rev. **75** (1949), 1736–1755.



(a) graph Γ (b) subgraph Γ_1 (c) subgraph Γ_2

Fig. 18.5. Overlapping subgraphs Γ_1 and Γ_2

to double-counting. "In his paper, while discussing these, Dyson had recommended precisely this – that one should subtract the vertex-part subinfinities twice before subtracting the final self-energy infinity."[24] But no proof of his assertion was given.

Salam decided to make the overlapping divergences his problem.[25] He thought that the best way to solve the problem "would be to ask Dyson's direct help."[26] So he rang Dyson up and said: "I am a beginning research student; I would like to talk with you. I am trying to renormalize meson theories, and there is the problem of overlapping divergences which you have solved. Could you give me some time?" Dyson indicated that he was leaving for the United States on the next day, so if Salam wanted to talk to him he had to come to Birmingham "tonight." This Salam did. The two of them got together the next morning. This was the first time that Salam met Dyson. He asked him "What is your solution of the overlapping infinity problem?" Dyson answered, "But, I have no solution. I only made a conjecture. Salam recalled that for a young student who had just started on research, "this was a terrible shock. Dyson was our hero. His papers were classics. For him to say that he had only made a conjecture made me feel that my support of certainty in the subject was slipping away." But Salam notes that "Dyson was being characteristically modest about his own work. He explained to me what the basis of his conjecture was. What he told me was enough to build on and show that he was absolutely right."

During the summer of 1950, Salam tackled the overlapping divergences problem and "using a generalization of Dyson's remarks," was able to show that QED and spin zero meson theories were indeed renormalizable to all orders.

Silvan Schweber, 1994

The following quotation is taken from the foreword and the introduction of the monograph by E. Manoukian, *Renormalization*, Academic Press, New York, 1983 (reprinted with permission). This monograph emphasizes the mathematical aspects of renormalization theory. It is based on the sophisticated proof of a general version of Weinberg's famous power-counting theorem (see Sect. 11.6.3 of Vol. I). This

[24] A. Salam, Overlapping divergences and the S-matrix, Phys. Rev. **82** (1951), 217–227.

[25] In 1979, Sheldon Glashow (born 1932), Abdus Salam (1926–1996) and Steven Weinberg (born 1933) were awarded the Nobel prize in physics for their contributions to the theory of electroweak interaction in the 1960s.

[26] A. Salam, Physics and the excellence of the life it brings, pp. 291–303. In: A. Salam, Ideals and Realities: Selected Essays of Abdus Salam. Edited by C. Lai, World Scientific, Singapore, 1987.

ensures the crucial convergence of the algebraic Feynman integrals which appear during the renormalization procedure.

> Renormalization theory is still with us and very much alive since its birth over three decades ago. It has reached such a high level of sophistication that any book on the subject has to be mathematically rigorous to do any justice on it... The first systematic study of renormalization historically was carried out by Salam in 1951 in a classic paper where the subtraction scheme of renormalization, in a general form, was formally sketched.[27] Surprisingly, this classic paper was not carefully reexamined until much later. In 1960 Weinberg established and proved one of the *most* important theorems in field theory. This theorem, popularly known as the "power-counting theorem" embodied a power-counting criterion to establish the absolute convergence of (algebraic) Feynman integrals.[28] Salam's work was first reexamined and brought to a mathematically consistent form in Manoukian (1976).[29] The absolute convergence of corresponding renormalized Feynman amplitudes was then proved by the author[30] by explicitly verifying in the process that the power-counting criterion of Weinberg was satisfied, thus completing the Dyson–Salam program:
>
> (DS) : Dyson → Salam → Weinberg → Manoukian (completion).
>
> Shortly, after the appearance of Salam's work, Bogoliubov, together with Parasiuk (1957),[31] in a classic paper developed a subtraction scheme and outlined a proof of its convergence. In 1966 Hepp[32] gave a convergence proof of the Bogoliubov–Parasiuk scheme by using in the intermediate stages ultraviolet cutoffs, and in 1969 Zimmermann[33] formulated the Bogoliubov scheme in momentum space with no ultraviolet cutoffs and gave a convergence proof of this subtraction scheme. Thus, these two latter authors completed the Bogoliubov–Parasiuk (BP) program:
>
> (BP) : Bogoliubov/Parasiuk → Hepp/Zimmermann (completion).
>
> Finally, the equivalence of the Bogoliubov scheme, in the Zimmermann form, and our scheme was then proved, after some systematic cancellations in the subtractions, by the author (see Manoukian (1976)) in a theorem that we have called the "unifying theorem of renormalization." Since the Zimmermann form grew out of Bogoliubov's work and our form grew out

[27] A. Salam, Overlapping divergences and the S-matrix, Phys. Rev. **82** (1951), 217–227.

[28] S. Weinberg, High energy behavior in quantum field theory, Phys. Rev. **118** (1960), 838–849.

[29] E. Manoukian, Generalization and improvement of the Dyson–Salam scheme and equivalence with other schemes, Phys. Rev. **D14** (1976), 966–971, 2202(E).

[30] E. Manoukian, Convergence of the generalized and improved Dyson–Salam renormalization scheme, Phys. Rev. **D15** (1977), 535–537; **D25** (1982), 1157(E). E. Manoukian, Class B_n-functions: Convergence of subtractions, Nuovo Cimento **A67** (1982), 101–120.

[31] N. Bogoliubov and O. Parasiuk, On the multiplication of propagators in quantum field theory, Acta Math. **97** (1957), 227–326 (in German).

[32] K. Hepp, Proof of the Bogoliubov–Parasiuk theorem on renormalization, Commun. Math. Phys. **2** (1966), 301–326.

[33] W. Zimmermann, Convergence of Bogoliubov's method of renormalization in momentum space, Commun. Math. Phys. **15** (1969), 208–234.

of Salam's work, this theorem establishes the long-standing problem of the equivalence between the paths taken in the ingenious approaches of Salam and Bogoliubov (in momentum space):

$$\text{(DS) program} \Longleftrightarrow \text{(BP) program (Manoukian).}$$

This book deals with a mathematically rigorous formulation of renormalization presented in a unified manner and a model independent way.

From the analytical point of view, the prototype of a divergent integral with overlapping divergences can be found in Sect. 2.2.5 on page 65. In terms of Feynman graphs, Fig. 18.4 on page 972 shows the prototype of an overlapping divergence. Bogoliubov's key recursion formula for the handling of overlapping divergences in the BPHZ approach can be found on page 982.

Unfortunately, it is hard to read papers and books on renormalization theory, since the *Salam criterion* is violated, as a rule: The following quotation describes this rule:

In 1951 Matthews and Salam formulated a requirement for renormalization procedures that has become popularly known as the *Salam criterion*:[34] "The difficulty, as in all this work, is to find a notation which is both concise and intelligible to at least two persons, of whom one may be an author." Possibly there are many proofs of the renormalizability of quantum electrodynamics which satisfy the Salam criterion. But we must confess that none of us has yet qualified as that other person who is the guarantor of the criterion. While there are today many standard texts which discuss the renormalizability of quantum electrodynamics, we are not aware of any which represents a complete proof and in particular justifies the claim that only gauge invariant counterterms are required. We here submit to you a direct and complete proof and we invite you to judge whether you can vouch for the Salam criterion.[35]

Joel Feldman, Thomas Hurd, Lon Rosen, and Jill Wright, 1988

18.3 The Renormalizability of Quantum Electrodynamics

Quantum electrodynamics is renormalizable by a passage to both the effective mass and the effective charge of the electron.[36]

Freemann Dyson, 1949

Renormalization is the readjustment of the higher order terms of the formal power series of quantum field theory so that the resulting quantities satisfy the axioms of quantum field theory in every order of perturbation theory

[34] P. Matthews and A. Salam, Renormalization, Rev. Mod. Phys. **23** (1951), 311–314.

[35] J. Feldman, T. Hurd, L. Rosen, and J. Wright, QED: A Proof of Renormalizability, Springer, Berlin 1988 (reprinted with permission).

[36] F. Dyson, The S-matrix in quantum electrodynamics, Phys. Rev. **75** (1949), 1736–1755.
F. Dyson, Advanced Quantum Mechanics. Dyson's Cornell Lecture Notes from 1951, Cornell University, Ithaca, New York. Transcribed by D. Derbes, World Scientific, Singapore, 2007.

> or, more ambitiously in constructive quantum field theory, in the exact, non-perturbative solution.[37]
>
> Klaus Hepp, 1971

The main goal of the renormalization of quantum electrodynamics reads as follows:

> *Show that there exists a general algorithm such that, in each order of perturbation theory, one gets finite S-matrix elements which only depend on two free parameters, namely,*
> - m_{eff} *(effective mass of the electron) and*
> - $-e_{\text{eff}}$ *(effective charge of the electron).*

These effective values cannot be determined by theoretical arguments. They have to be determined by physical experiments. Roughly speaking, the main goal formulated above can be realized. There exist algorithms used by physicists on supercomputers in order to compute high-precision results (e.g., the anomalous magnetic moment of the electron). However, the arguments for justifying the algorithms are highly technical and full of mathematical pitfalls. In particular, one has to show that

- the choice of the gauge condition doesn't matter (gauge invariance),
- the S-matrix elements generate transition probabilities (unitarity of the renormalized S-matrix), and
- the choice of the inertial system doesn't matter (relativistic covariance).

Note that there exist different approaches to renormalization theory. In the setting of the lattice approach, the main idea is to choose the free parameters of the counterterms of the Lagrangian in such a way that the continuum limit exists. For example, physicists have shown that the mathematical trick of introducing an artificial photon mass does not influence the computation of important physical effects. In fact, the artificial mass terms cancel each other if one takes all of the essential physical effects into account. For example, the additional production of low-energy (soft) photons compensates infrared divergences of the algebraic Feynman integrals. For example, when electrons scatter at protons or in the field of a nucleus, they can emit real photons. This physical phenomenon is called Bremsstrahlung. This German word means *breaking radiation*. In 1931 this effect was first studied by Sommerfeld in a non-relativistic setting. In 1934 Bethe and Heitler computed the relativistic cross section for breaking radiation related to the incoming stream of electrons in the Coulomb field of an atom and the outgoing photons. We refer to the study of breaking radiation in:

> L. Landau and E. Lifshitz, Course of Theoretical Physics, Vol 4: Quantum Electrodynamics, Butterworth–Heinemann, Oxford, 1982.
>
> C. Itzykson and J. Zuber, Quantum Field Theory, MacGraw-Hill, New York, 1908.

A detailed modern study can be found in Sect. 3.11 of the monograph by

> G. Scharf, Finite Quantum Electrodynamics: the Causal Approach, Springer, Berlin, 1995.

Here, the causal Epstein–Glaser approach for constructing the S-matrix is used in order to show that the adiabatic limit (i.e., the infrared limit) exists. This is based on surprising cancellations of singularities by taking Feynman graphs into account which correspond to soft photons of breaking radiation.

[37] K. Hepp, Renormalization theory, pp. 429–500. In: C. DeWitt and R. Stora (Eds.), Statistical Mechanics and Quantum Field Theory, Les Houches 1970, Gordon and Breach, New York, 1971.

Ambiguity of renormalization schemes. Fortunately enough, it seems that the mathematical ambiguity of the choice of renormalization schemes does not matter the physics.

> *The computational experience of physicists shows that different mathematical renormalization schemes yield the same results for the crucial values measured in physical experiments.*

This miracle is closely related to both

- Hepp's axiomatic approach to renormalization theory (including the equivalence of renormalization schemes) (see Hepp (1971) quoted on page 976)
- and the method of the renormalization group.

Here, we refer to:

> N. Bogoliubov and D. Shirkov, Introduction to Quantum Field Theory, Interscience, New York, 1980.
>
> V. Rivasseau, From Perturbative to Constructive Renormalization, Princeton University Press, Princeton, New Jersey, 1991.
>
> L. Brown, Quantum Field Theory, Cambridge University Press, New York, 1996.
>
> W. McComb, Renormalization Methods: A Guide for Beginners, Oxford University Press, Oxford, 2007.

Additional hints for further reading can be found on page 1029ff.

18.4 Automated Multi-Loop Computations in Perturbation Theory

In the near future, the most spectacular experiments in elementary particle physics will be performed at the LHC (Large Hadron Collider) of CERN (European Organization for Nuclear Research at Geneva, Switzerland). The maximal energy per particle of the LHC will be near 10 TeV. This corresponds to the rest energy of 10^4 protons. For mastering theoretically the experiments, one needs extremely complicated computations in the framework of multi-loop perturbation theory. This can only be done with the help of

- highly sophisticated computer programs
- based on deep theoretical insight into the procedure of renormalization.

We recommend the following references:[38]

(i) Software package (Max Planck Institute for Physics, Werner Heisenberg, in Munich): Internet `http://www.feynarts.de`

(ii) Classic review article from 1998: `http://arxiv.org/abs/hep-ph/9812357` together with

`http://www-ttp.particle.uni-karlsruhe.de/Links/algprog.html`

(iii) Recent review article from 2008 emphasizing supersymmetric computations which are crucial for the LHC: `http://arxiv.org/abs/0805.2088`

[38] The author would like to thank Thomas Hahn from the Max Planck Institute for Physics, Werner Heisenberg, in Munich for informing him about the references quoted above.

(iv) Computation of cross sections for a given Lagrangian:
- GRACE: `http://minami-home.kek.jp`
- Amegic++/Sherpa: `projects.hepforge.org/sherpa/dokuwiki/doku.php`
- MadGraph: `http://madgraph.hep.uiuc.edu`
- CompHEP: `http://comphep.sinp.msu.ru/`
- Pythia: `http://home.thep.lu.se/∼ torbjorn/Pythia.html`
- Feynarts: see (i) above.

(v) Computer algebra system FORM: `http://www.nikhef.nl/∼ form/`

(vi) The solution of the outstanding problem of finding a basis of N-loop integrals by the Laporta algorithm (55 pages): `arxiv.org/abs/hep-ph/0102033`

(vii) Highly effective computation of transition amplitudes by using the pole structure of Feynman diagrams and the Cauchy residue calculus (33 pages):

`http.://arxiv.org/abs/hep-th/0410179`

In algebraic geometry, one has to solve systems of polynomial equations. For doing this, one uses a so-called Gröbner basis. The most popular algorithm for solving polynomial equations and other problems in computer algebra is the Buchberger algorithm (see Cox et al., Using Algebraic Geometry, Springer, New York, 1998). The Laporta algorithm in renormalization theory works similarly as the Buchberger algorithm.

19. Perspectives

> Before doing the hard work of climbing a high mountain, go to the top by cable railway and enjoy looking at the beautiful landscape. This will motivate the later efforts.
>
> Folklore

In this chapter we will take the cable railway. We postpone the mountain climbing to Vol. IV on quantum mathematics. There exist the following two basic approaches to quantum field theory:

(S) the S-matrix approach (scattering matrix), and
(G) the Green's function approach (correlation functions).

The main ideas are discussed in Chaps. 14 and 15 of Vol. I. A detailed study will be carried out in Vol. IV. Roughly speaking, the two approaches (S) and (G) are equivalent to each other.

In the present volume, we use the S-matrix approach based on the Dyson series. The Epstein–Glaser approach (also called the causal approach) represents a refinement of the Dyson series in terms of tempered distributions (see Sect. 19.2).

The Green's function approach (G) uses the generating functional $Z(J)$ (with the source J) for the Green's functions of the interacting quantum field. Alternatively, one can describe $Z(J)$ by either

- some Feynman functional integral $Z(J) = \mathcal{N} \int \mathrm{e}^{-\mathrm{i}S[\psi,J]/\hbar}\, \mathcal{D}\psi$ (depending on the classical action S) or
- some functional-differential equation in terms of the functional derivative $\frac{\delta Z}{\delta J}$ (the quantum action principle).

Here, the quantum action principle (QA) represents a functional-differential equation for a fundamental basis of Green's functions of the interacting quantum field theory (the so-called vertex functions which correspond to one-particle irreducible Feynman graphs). The functional-differential equation (QA) is also called the Dyson–Schwinger equation, which can be regarded as the fundamental equation of motion for the Green's functions.

All the information on the interacting quantum field can be obtained from either the S-matrix or the generating functional $Z(J)$. In fact, $Z(J)$ can be replaced by the simpler generating functional $V(J)$ for the vertex functions.

- The passage from the S-matrix to the generating functional $Z(J)$ is obtained by the magic Gell-Mann–Low formula.
- Conversely, the passage from $Z(J)$ to the S-matrix is obtained by the magic LSZ (Lehmann, Szymanzik, Zimmermann) reduction formula.

In fact, it is possible to introduce a generating functional for the S-matrix (e.g., in terms of some Feynman functional integral which is related to $Z(J)$). Besides the S-matrix approach and the Green's function approach, there exist different approaches by using the following functions:

- the Wightman functions in axiomatic quantum field theory, and
- the retarded functions in the GLZ (Glaser, Lehmann, and Zimmermann) setting.[1]

For a detailed study, we refer to Vol. IV.

Symmetries of the classical action functional lead to both conservation laws and the fundamental Slavnov–Taylor identities for the Green's functions in gauge theories; these identities represent constraints for the generating functional $Z(J)$. Note that the Slavnov–Taylor identities generalize the Ward–Takahashi identities in quantum electrodynamics. These constraints can be handled by methods which are similar to Hilbert's theory of syzygies in commutative algebra (theory of invariants and algebraic geometry).[2] It is possible that classical symmetries do not survive the process of quantization. We call this an anomaly.

In renormalization theory, one has to distinguish between the following notions:

(i) regularization,
(ii) renormalization,
(iii) renormalizability, and
(iv) super-renormalizability.

By (i), we understand mathematical methods which replace divergent integrals by convergent integrals or which modify sums by subtracting terms such that, say, the lattice limit exists. Typical regularization methods for integrals are:

- truncation of the domain of integration (e.g., momentum cut-off),
- Pauli–Villars regularization,
- dimensional regularization, and
- analytic regularization.

This is discussed in Sect. 2.1. If the regularization method (i) leads to a reasonable S-matrix, which describes transition probabilities for scattering processes in terms of physics, then we speak of renormalization. This includes that, in each order of perturbation theory, all the S-matrix elements are finite, and they depend on a finite number of parameters which can be determined by physical experiments. In addition, the S-matrix has to be unitary. Otherwise, the interpretation of the S-matrix elements as transition probabilities fails.

In the important special case where the number of free parameters is bounded by a fixed integer in each order of perturbation theory, the quantum field theory is called renormalizable. For example, quantum electrodynamics is renormalizable; in each order of perturbation theory, the (renormalized) S-matrix depends on two free parameters, namely, the effective mass m_{eff} and the effective charge $-e_{\text{eff}}$ of the electron. These free parameters have to be measured by physical experiment.

Furthermore, the theory is called super-renormalizable iff there exists only a finite number of Feynman graphs which generate divergent integrals (e.g., in the lattice limit). For example, quantum electrodynamics is not super-renormalizable.

For gauge theories it is crucial that the renormalization procedure survives the symmetries of the classical action functional. For handling this in an elegant manner, the following three approaches exist:

[1] See O. Steinmann, Perturbative Quantum Electrodynamics and Axiomatic Field Theory, Springer, Berlin, 2000. See also the recent papers by Fredenhagen and his collaborators quoted on page 455.

[2] See. D. Eisenbud, Commutative Algebra with a View to Algebraic Geometry, Springer, New York, 1994.
D. Eisenbud, The Geometry of Syzygies: A Second Course in Commutative Algebra and Algebraic Geometry, Springer, New York, 2005.
M. Henneaux and C. Teitelboim, Quantization of Gauge Systems, Princeton University Press, 1993.

- the Faddeev–Popov method by using the orbit space of the gauge group in order to factorize the "measure" of the Feynman functional integral (see Sect. 16.6 of Vol. I),
- algebraic renormalization – the BRST (Becchi, Rouet, Stora, Tyutin) symmetry (see both Sect. 16.7 of Vol. I and Sect. 19.6 of the present volume), and
- the recent Master Ward identity due to Boas and Dütsch (see Sect. 19.9).

Detailed hints for further reading concerning the huge field of renormalization theory will be given in Sect. 19.11 on page 1029.

19.1 BPHZ Renormalization

19.1.1 Bogoliubov's Iterative *R*-Method

Dimensional regularization. Recall that a Feynman graph Γ is called divergent iff the corresponding n-dimensional algebraic Feynman integral $J_n(\Gamma)$ is divergent. Divergent graphs are one-particle irreducible. Replacing n by $n-\nu$, the method of dimensional regularization yields

$$J_{n-\nu}(\Gamma) = \sum_{k=-r}^{\infty} a_k \nu^k$$

where $r = 1, 2, \ldots$ (see Sect. 2.2.9). In the special case where the algebraic Feynman integral is convergent, we have $a_{-1} = \ldots = a_{-r} = 0$, and we get

$$J_n(\Gamma) = \lim_{\nu \to +0} J_{n-\nu}(\Gamma).$$

In the divergent case, we introduce the following Rota–Baxter truncation operator:

$$\boxed{\mathsf{P}_\Gamma J_{n-\nu}(\Gamma) := \sum_{k=-r}^{-1} \frac{a_k}{\nu^k}.} \tag{19.1}$$

This yields the regularization

$$(1-\mathsf{P}_\Gamma) J_{n-\nu}(\Gamma) := \sum_{k=0}^{\infty} \frac{a_k}{\nu^k}.$$

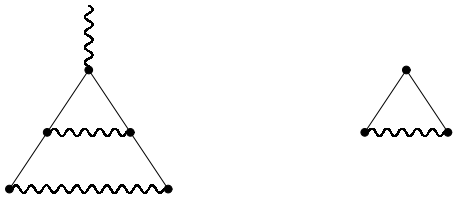
Hence

$$\lim_{\nu \to +0} (1-\mathsf{P}_\Gamma) J_{n-\nu}(\Gamma) = a_0.$$

This is the regularized value of the divergent integral $J_n(\Gamma)$. In what follows we will briefly write the symbol $(1-\mathsf{P}_\Gamma)J(\Gamma)$ instead of $(1-\mathsf{P}_\Gamma)J(\Gamma)_{n-\nu}$. All of the following expressions depend on the small real parameter ν in a regular way. The limit $\nu \to +0$ will be carried out at the end of Bogoliubov's iterative method. In what follows we will only consider one-particle irreducible graphs (briefly called 1PI graphs). This class comprehends the divergent graphs.

Example 1 (no proper 1PI subgraph). Consider a 1PI graph Γ which has no proper 1PI subgraph. We define the regularization $R(\Gamma)$ of the corresponding Feynman integral $J(\Gamma)$ by adding the subtraction term $S(\Gamma)$.[3] Explicitly,

[3] The subtraction term is also called counterterm in the literature. However, in order to avoid misunderstandings, we will use the designation 'counterterm' only for the corresponding terms of the Lagrangian density.



(a) graph Γ (b) subgraph γ

Fig. 19.1. Two-loop vertex diagram

$$R(\Gamma) := J(\Gamma) + S(\Gamma)$$

along with $S(\Gamma) := -\mathsf{P}_\Gamma J(\Gamma)$. Hence

$$\boxed{R(\Gamma) = (1 - \mathsf{P}_\Gamma)J(\Gamma).} \tag{19.2}$$

This is the usual regularization procedure of the overall integral $J(\Gamma)$. In order to handle graphs with an arbitrary number of divergent subgraphs, Bogoliubov invented the following general algorithm.

Bogoliubov's iterative method. We will proceed in two steps.

Step 1: The preparation of the algebraic Feynman integral. We replace the algebraic Feynman integral $J(\Gamma)$ by the so-called preparation integral $\overline{R}(\Gamma)$. Explicitly, we set

$$\overline{R}(\Gamma) := J(\Gamma) + \sum_{\gamma} J(\Gamma/\gamma) \cdot S(\gamma). \tag{19.3}$$

Here, we sum over all families

$$\gamma := \{\gamma_1, \ldots, \gamma_s\}, \qquad s = 1, 2, \ldots,$$

of pairwise disjoint, proper 1PI subgraphs $\gamma_1, \ldots, \gamma_s$ of the given graph Γ. In addition, we use the product rule

$$S(\gamma) := S(\gamma_1)S(\gamma_2)\cdots S(\gamma_s).$$

Finally, the symbol Γ/γ denotes a special subgraph of Γ. More precisely, Γ/γ is obtained from the graph Γ by shrinking the subgraph γ to a point. We also assume the validity of the following product rule:

$$J(\Gamma/\gamma) \cdot J(\gamma) = J(\Gamma), \tag{19.4}$$

which is a consequence of the Feynman rules.

Step 2: The subtraction term: The regularization of the Feynman integral $J(\Gamma)$ is given by

$$R(\Gamma) := \overline{R}(\Gamma) + S(\Gamma) \tag{19.5}$$

with the subtraction term $S(\Gamma) := -\mathsf{P}_\Gamma \overline{R}(\Gamma)$. Hence $R(\Gamma) = (1 - \mathsf{P}_\Gamma)\overline{R}(\Gamma)$.

This is a recursive procedure. Starting with the given graph Γ, Bogoliubov's method reduces the computation to proper 1PI subgraphs of Γ. After a finite number of steps, one arrives at some 1PI subgraph which has no proper 1PI subgraphs anymore. In this favorite case, we can apply Example 1.

Example 2 (one proper 1PI subgraph). Consider the two-loop vertex graph Γ depicted in Fig. 19.1 together with the proper 1PI subgraph γ. We claim that

$$\boxed{R(\Gamma) = (1 - \mathsf{P}_\Gamma)(1 - \mathsf{P}_\gamma)J(\Gamma).} \tag{19.6}$$

Proof. By (19.3), we get

$$\overline{R}(\Gamma) = J(\Gamma) + J(\Gamma/\gamma) \cdot S(\gamma).$$

By Example 1, $S(\gamma) = -\mathsf{P}_\gamma J(\gamma)$. Hence

$$\begin{aligned}\overline{R}(\Gamma) &= J(\Gamma) - J(\Gamma/\gamma)\mathsf{P}_\gamma J(\gamma) \\ &= J(\Gamma) - \mathsf{P}_\gamma\big(J(\Gamma/\gamma) \cdot J(\gamma)\big) = J(\Gamma) - \mathsf{P}_\gamma J(\Gamma).\end{aligned}$$

Therefore,

$$\overline{R}(\Gamma) = (1 - \mathsf{P}_\gamma)J(\Gamma).$$

This means that the preparation of the integral $J(\Gamma)$ consists in regularizing the subintegral which corresponds to the subgraph γ. By (19.5),

$$R(\Gamma) = \overline{R}(\Gamma) + S(\Gamma) = (I - \mathsf{P}_\Gamma)\overline{R}(\Gamma) = (I - \mathsf{P}_\Gamma))(I - \mathsf{P}_\Gamma).$$

□

Observe that formula (19.6) corresponds to a nested regularization. First we regularize the subintegral with respect to the subgraph γ, then we regularize the overall integral.

Example 3 (two proper 1PI subgraphs). Consider the two-loop vacuum polarization graph Γ depicted in Fig. 19.2 together with the two overlapping 1PI subgraphs γ_1 and γ_2. We claim that

$$R(\Gamma) = (1 - \mathsf{P}_\Gamma)(1 - \mathsf{P}_{\gamma_1} - \mathsf{P}_{\gamma_2})J(\Gamma). \tag{19.7}$$

Proof. By (19.3), we get

$$\overline{R}(\Gamma) = J(\Gamma) - J(\Gamma/\gamma_1)S(\gamma_1) - J(\Gamma/\gamma_2)S(\gamma_2)$$

with $S(\gamma_j) = -\mathsf{P}_{\gamma_j} J(\gamma_j)$, $j = 1, 2$. Hence

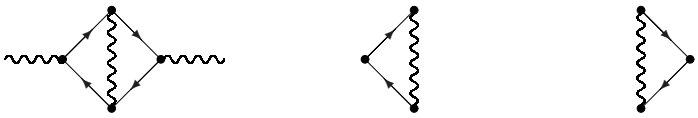
$$\overline{R}(\Gamma) = J(\Gamma) - \mathsf{P}_{\gamma_1}(J(\Gamma/\gamma_1)J(\gamma_1)) - \mathsf{P}_{\gamma_2}(J(\Gamma/\gamma_2)J(\gamma_2) = (1 - \mathsf{P}_{\gamma_1} - \mathsf{P}_{\gamma_1})J(\Gamma).$$

□

Note that formula (19.7) differs from the naive regularization formula

$$R(\Gamma) = (1 - \mathsf{S}_\Gamma)(1 - \mathsf{S}_{\gamma_1})(1 - \mathsf{S}_{\gamma_2})J(\Gamma).$$

This is related to the fact that overlapping 1PI subgraphs complicate the regularization procedure.



(a) graph Γ (b) subgraph γ_1 (c) subgraph γ_2

Fig. 19.2. Overlapping subgraphs γ_1 and γ_2

19.1.2 Zimmermann's Forest Formula

Our goal is to replace the iterative Bogoliubov formula (19.3) by a global formula. To this end, we need the concept of a forest.

Forest. Let Γ be a 1PI graph, and let $\Gamma_1, \dots, \Gamma_n$ be proper 1PI subgraphs of Γ.[4] By definition, a forest F of the graph Γ is a nonempty subset of the set

$$\{\Gamma, \Gamma_1, \dots, \Gamma_n\}$$

which does not contain overlapping graphs. The symbol $\mathcal{F}(\Gamma)$ denotes the family of all the forests F of the graph Γ. Let us consider three examples.

(i) The graph pictured in Fig. 19.3: We have the 1PI graph Γ and the proper 1PI subgraphs Γ_1 and Γ_2. Therefore, the corresponding family of forests consists of the following five sets:

$$\{\Gamma\},\ \ \{\Gamma_1\},\ \ \{\Gamma_2\},\ \ \{\Gamma, \Gamma_1\},\ \ \{\Gamma, \Gamma_2\}. \tag{19.8}$$

These five sets form the forest family $\mathcal{F}(\Gamma)$ of the graph Γ. Observe that the sets $\{\Gamma_1, \Gamma_2\}$ and $\{\Gamma, \Gamma_1, \Gamma_2\}$ are not forests, since Γ_1 and Γ_2 are overlapping subgraphs.

(ii) The two-loop vertex graph pictured in Fig. 19.1: The graphs Γ and γ are 1PI. Since $\gamma \subseteq \Gamma$, the graphs Γ and γ are not overlapping. The family $\mathcal{F}(\Gamma)$ of forests consists of the following three sets:

$$\{\Gamma\},\ \ \{\gamma\},\ \ \{\Gamma, \gamma\}. \tag{19.9}$$

(iii) The two-loop vacuum polarization graph pictured in Fig. 19.2: We have the 1PI graphs $\Gamma, \gamma_1, \gamma_2$. The corresponding family of forests consists of the following five sets:

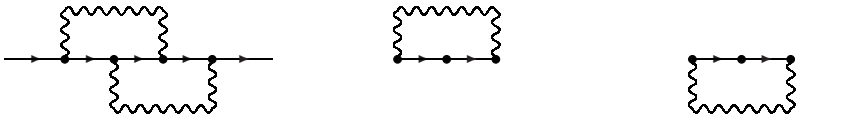
$$\{\Gamma\},\ \ \{\gamma_1\},\ \ \{\gamma_2\},\ \ \{\Gamma, \gamma_1\},\ \ \{\Gamma, \gamma_2\}. \tag{19.10}$$

The forest formula. Zimmermann discovered in 1969 that Bogoliubov's formula (19.3) can be obtained by the following simpler formula called Zimmermann's forest formula:

$$\boxed{R(\Gamma) = J(\Gamma) + \sum_{F \in \mathcal{F}(\Gamma)} \prod_{\gamma \in F} -\mathsf{P}_\gamma J(\Gamma).} \tag{19.11}$$

The sum is taken over all forests F of the graph Γ, whereas the product is taken over all subgraphs γ of the forest F. For nested subgraphs γ, the operator P should be

[4] We tacitly assume that Γ and $\Gamma_1, \dots, \Gamma_n$ are not empty.



(a) graph Γ (b) subgraph Γ_1 (c) subgraph Γ_2

Fig. 19.3. Overlapping subgraphs Γ_1 and Γ_2

applied inside to outside. This formula displays clearly that the subtraction terms depend on the forests F of the original graph Γ.

Example 1. Consider the two-loop vertex graph Γ depicted in Fig. 19.1 on page 982. The forests are the following sets: $\{\Gamma\}, \{\gamma\}, \{\Gamma, \gamma\}$. By the forest formula (19.11), we get

$$R(\Gamma) = J(\Gamma) - \mathsf{P}_\Gamma J(\Gamma) - \mathsf{P}_\gamma J(\Gamma) + \mathsf{P}_\Gamma \mathsf{P}_\gamma J(\Gamma).$$

Equivalently, $R(\Gamma) = (1 - \mathsf{P}_\Gamma)(1 - \mathsf{P}_\gamma)J(\Gamma)$. This is identical with Bogoliubov's formula (19.6).

Example 2. Consider the two-loop vacuum polarization graph Γ depicted in Fig. 19.2. Here, the forests are $\{\Gamma\}, \{\gamma_1\}, \{\gamma_2\}, \{\Gamma, \gamma_1\}, \{\Gamma, \gamma_2\}$. By the forest formula (19.11), we obtain

$$R(\Gamma) = J(\Gamma) - \big(\mathsf{P}_\Gamma + \mathsf{P}_{\gamma_1} + \mathsf{P}_{\gamma_2}\big)J(\Gamma) + \big(\mathsf{P}_\Gamma \mathsf{P}_{\gamma_1} + \mathsf{P}_\Gamma \mathsf{P}_{\gamma_2}\big)J(\Gamma).$$

Equivalently,

$$R(\Gamma) = (1 - \mathsf{P}_\Gamma)(1 - \mathsf{P}_{\gamma_1} - \mathsf{P}_{\gamma_2})J(\Gamma).$$

This is identical with Bogoliubov's formula (19.7).

19.1.3 The Classical BPHZ Method

> The main problem of the BPHZ method is the convergence of the algebraic Feynman integrals after applying the forest formula.
>
> Folklore

In his classical paper from 1969, Zimmermann did not use the method of dimensional regularization, but the method of Taylor subtraction in momentum space (see Sect. 2.2.4).[5] Let us sketch the basic ideas.

Superficial degree of divergence of an algebraic Feynman integral. Fix the real numbers a, b, and $\varepsilon > 0$. Let $m = 0, 1, 2, \ldots$, and $n = 1, 2, \ldots$ The integral

$$J(a, b, \varepsilon) := \int_{\mathbb{R}} \frac{q^m + a}{q^n + b + \mathrm{i}\varepsilon}\, dq$$

is convergent if $n - m > 1$. This is the simplest model for an algebraic Feynman integral. Here, the parameters a and b model the momenta of the external particles, and the parameter $\varepsilon > 0$ regularizes the integral by cancelling the possible zeros of the denominator on the real line. Introducing the degree of divergence,

$$d := m - n + 1, \tag{19.12}$$

[5] W. Zimmermann, Convergence of Bogoliubov's method of renormalization in momentum space, Commun. Math. Phys. **15** (1969), 208–234.

the integral J is convergent if $d < 0$. Fix $N = 1, 2, \ldots$, and let us consider the N-dimensional integral

$$\mathcal{J}(a, b; \varepsilon) := \int_{\mathbb{R}^N} \frac{f(q_1, \ldots, q_N; a, b)}{g(q_1, \ldots, q_N; a, b, \varepsilon)}\, dq_1 dq_2 \cdots dq_N. \tag{19.13}$$

Suppose that f (resp. g) is a real polynomial of degree m (resp. n) with respect to the real variables $q_1, \ldots, q_N$. In addition, suppose that the denominator g does not vanish on $\mathbb{R}^N$. This goal can be reached, if necessary, by replacing $g(q)$ by $g(q) + \mathrm{i}\varepsilon$ (or by similar expressions). The number

$$d := m - n + N$$

is called the superficial degree of divergence of the integral $\mathcal{J}$. It turns out that, for higher dimensions $N > 1$, the global degree d does not contain enough information about the convergence or divergence of the integral $\mathcal{J}$; it is possible that nested subdivergences appear. However, Weinberg's famous, highly nontrivial power-counting theorem tells us, roughly speaking, the following:

> *The algebraic Feynman integral $J(\Gamma)$ corresponding to a given Feynman graph Γ is absolutely convergent if the superficial degree of divergence of the integral $J(\Gamma)$ itself is negative, and the same is true for all the subintegrals corresponding to subgraphs of Γ.*

The main idea is to reach this favorite situation by adding counterterms to the original Lagrangian density in order to get additional Feynman graphs, which improve the superficial degrees of divergence.[6]

Zimmermann's fundamental convergence result. Let $m > 0$ be the particle mass. In order to avoid singularities caused by zeros of the denominator, Zimmermann used the following replacement:

$$\frac{1}{(p_0^2 - \mathbf{p}^2 - m^2)^r} \quad \Rightarrow \quad \frac{1}{(p_0^2 - \mathbf{p}^2 - m^2 + \mathrm{i}\varepsilon(\mathbf{p}^2 + m^2))^r}. \tag{19.14}$$

Here, $\varepsilon > 0$ is a fixed regularization parameter which cancels the zeros of the denominator on $\mathbb{R}^4$. Zimmermann's convergence result reads as follows:

(i) For fixed $\varepsilon > 0$, the Taylor subtraction method in the momentum space together with the forest formula (19.11) yields convergent algebraic Feynman integrals. These integrals are parameter integrals with respect to the regularization parameter ε and the external momenta.

(ii) As $\varepsilon \to +0$, the limit of the parameter integrals exists in the sense of tempered distributions.

In addition, the limit tempered distributions are relativistically invariant.

Lowenstein's fundamental convergence result. Lowenstein proved that Zimmermann's result can be generalized to the case of vanishing particle mass, $m = 0$. This extension of the BPHZ method is called the BPHZL (Bogoliubov, Parasiuk, Hepp, Zimmermann, Lowenstein) method.[7]

---

[6] S. Weinberg, High energy behavior in quantum field theory, Phys. Rev. **118** (1960), 838–849. This classical paper studies both the convergence of algebraic Feynman integrals and the asymptotic behavior of such integrals if the momenta of the external particles go to infinity.
For the Weinberg power-counting theorem, we refer to Sect. 11.6.3 of Vol. I. In a general setting, detailed proofs can be found in the monograph by E. Manoukian, Renormalization, Academic Press, New York, 1983.

[7] See J. Lowenstein, Convergence theorems for renormalized Feynman integrals with zero-mass propagators, Commun. Math. Phys. **47** (1976), 53–68.

19.2 The Causal Epstein–Glaser S-Matrix Approach

> A brilliant example of the creation and application of new mathematical methods was the development of an axiomatic approach to quantum field theory undertaken by Bogoliubov in the 1950s. He always strove to work on the latest and hottest topics of theoretical physics. At that time, the ultraviolet divergence was an important problem in quantum field theory when using the Hamiltonian formalism (canonical quantization). Bogoliubov proposed a new approach to this problem.[8] First of all, he abandoned the Hamiltonian formalism and took as a basis of the theory the S-matrix introduced by Heisenberg in 1943...
> Bogoliubov required that the S-matrix satisfies the following fundamental postulates: *it must be relativistically covariant, unitary, causal (local) and spectral.*[9]
>
> Vasilii Vladimirov, 2005

> It is shown how an inductive construction of the renormalized perturbation series of quantum field theory automatically yields, at each order, finite terms satisfying the requirement of locality. This method whose result is equivalent to the Bogoliubov–Parasiuk–Hepp prescriptions, also establishes the usual classification between renormalizable and non-renormalizable theories.[10]
>
> Henri Epstein and Vladimir Glaser, 1973

The latter is one of the most important papers in quantum field theory. However, for a long time, only a few specialists noticed this important approach to quantum field theory. Quantum electrodynamics in terms of the elegant Epstein–Glaser approach is thoroughly studied in the monograph by G. Scharf, Finite Quantum Electrodynamics: the Causal Approach, Springer, Berlin, 1995. The S-matrix is a formal power series whose terms are operator-valued tempered distributions. The starting point is the representation of the S-matrix by a formal series expansion

$$\mathsf{S}(g) = 1 + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathbb{R}^{4n}} S_n(x_1, x_2, \ldots, x_n) \prod_{k=1}^{n} g(x_k)\, dx_1 dx_2 \cdots dx_n, \qquad (19.15)$$

where g is a test function living in the space $\mathcal{S}(\mathbb{R}^4)$ of smooth functions which are rapidly decreasing at infinity. This symbolic notation means that the terms of

J. Lowenstein and E. Speer, Distributional limits of renormalized integrals with zero-mass denominator, Commun. Math. Phys. **47** (1976), 43–51.
J. Lowenstein, BPHZ Renormalization, pp. 95–160. In: G. Velo and A. Wightman (Eds.), Renormalization Theory, Reidel, Dordrecht, 1976.

[8] See N. Bogoliubov and D. Shirkov, Introduction to Quantum Field Theory, Nauka, Moscow, 1957 (in Russian). Fourth Russian edition, 1984. English edition: Interscience, New York, 1959/1980.
A similar approach was formulated by E. Stueckelberg and D. Rivier, Causalité et structure de la matrice S, Helv. Phys. Acta **23** (1950), 215–222 (in French).

[9] V. Vladimirov, Nikolai Nikolaevich Bogoliubov (1909–1992) – Mathematician by the Grace of God. In: A. Bolibruch, Yu. Osipov, and Ya. Sinai (Eds.), Mathematical Events in the 20th Century, Springer, Berlin, and PHASIS, Moscow, 2006 (reprinted with permission).

[10] H. Epstein and V. Glaser, The role of locality in perturbation theory, Ann. Inst. Poincaré **A19**(3) (1973), 211–295.

(19.15) are tempered distributions. The Epstein–Glaser approach to quantum field theory is based on the following axioms:[11] (L) locality, (R) relativistic invariance, (U) unitarity of the S-matrix, and (C) causality. Let us sketch the basic ideas. Since $\mathsf{S}(g)$ introduced in (19.15) is only a formal power series expansion, the following conditions only serve as a rough orientation. The precise formulation will be given in Vol. IV.

Locality. In terms of physics, locality means that physical interactions are localized in space and time. In the Epstein–Glaser setting, this locality is realized by the fact that the S-matrix is an operator-valued tempered distribution which depends on test functions. In particular, if the support of the test function g is concentrated on a small region of the four-dimensional space-time manifold, then the interaction is localized. More precisely, there exists a complex Hilbert space X such that, for any test function $g \in \mathcal{S}(\mathbb{R}^4)$, we have the linear operator

$$\mathsf{S}(g) : X \to X. \tag{19.16}$$

Relativistic invariance. According to Einstein's principle of special relativity, physics is independent of the choice of the inertial system. A change between two inertial systems Σ and Σ^+ is given by a Poincaré transformation

$$x^+ = \Lambda x + a,$$

where $x = (\mathbf{x}, t)$ and $x^+ = (\mathbf{x}^+, t^+)$ are the space-time coordinates in Σ and Σ^+, respectively. We postulate that

$$\mathsf{S}(g^+) = U(\Lambda, a) \cdot \mathsf{S}(g) \cdot U(\Lambda, a)^{-1}$$

for all test functions $g \in \mathcal{S}(\mathbb{R}^4)$, and all Poincaré transformations (Λ, a). Here, g^+ is defined by the condition $g^+(x^+) := g(x)$ for all space-time points x in Σ. Moreover, we assume that the map

$$(\Lambda, a) \mapsto U(\Lambda, a)$$

is a unitary representation of the Poincaré group in the Hilbert space X (see Vol. III). In 1939 Wigner showed that such (nontrivial) representations only exist in infinite-dimensional Hilbert spaces.[12]

Intuitively, this tells us that quantum field must possess an infinite number of degrees of freedom. From the mathematical point, this is related to the fact that the Lorentz group is not a compact Lie group, but only a locally compact Lie group.

Unitarity of the S-matrix. We assume that the operator (19.16) is unitary.

Causality. Two test functions f and g are called separated by causality iff there exist both an inertial system and a real number t_0 such that the following hold: If $f(\mathbf{x}, t) \neq 0$ (resp. $g(\mathbf{x}, t) \neq 0$), then $t < t_0$ (resp. $t > t_0$). The key causality axiom demands that this causal property of the test functions f and g implies the following crucial product property of the S-matrix:

$$\boxed{\mathsf{S}(f + g) = \mathsf{S}(f)\mathsf{S}(g).} \tag{19.17}$$

[11] This theory should be called the Stueckelberg–Bogoliubov–Epstein–Glaser theory. For brevity, we speak of the Epstein–Glaser theory.

[12] E. Wigner, On unitary representations of the inhomogeneous Lorentz group, Ann. Math. **40** (1939), 149–204. In 1963 Eugene Wigner (1902–1995) was awarded the Nobel prize in physics for his contributions to the theory of the atomic nucleus and the elementary particles, particularly through the discovery and application of fundamental symmetry principles.
The importance of symmetry in physics and the corresponding mathematical approach will be thoroughly studied in Vols. IIIff.

It can be shown that this functional equation for the S-matrix allows the construction of an iterative method which yields the tempered distributions corresponding to $S_n(x_1, \ldots, x_n)$ for $n = 1, 2, \ldots$ This method starts with the classical Lagrangian density of the field theory. The most delicate step is the separation of tempered distributions with causal support into retarded and advanced parts.

The Dyson series is based on the time ordered-product $\mathcal{T}(\psi(t)\psi(s))$ of a quantum field. It turns out that, in quantum field theory, typical difficulties arise from the fact that the time-ordered product is undetermined for equal time, $t = s$, since serious singularities appear for equal time points. The Epstein–Glaser approach overcomes this difficulty by using the sophisticated splitting technique for tempered distributions with causal support.

Renormalization and the Hahn–Banach theorem. Observe the following key point:

> *The Epstein–Glaser approach completely avoids the regularization of divergent integrals as in the BPHZ approach.*

The main idea reads as follows:

- We start with a special class of test functions which vanish on critical subsets of the space-time manifold.
- With respect to these special test functions g, the functional $g \mapsto \mathsf{S}(g)$ behaves nicely, that is, no singularities appear.
- Finally, we extend the restricted functional S to the full space of test functions. This way, we obtain additional real parameters which have to be identified with physical parameters (see Sect. 15.4.4 of Vol. I).

Surprisingly enough, in this setting, renormalization becomes a quite natural variant of the Hahn–Banach extension theorem for functionals, which is standard in functional analysis.[13]

Low-energy limit. The low-energy limit corresponds to limits of tempered distributions where the test functions go to the constant function, $g \equiv 1$. The latter function is not a test function anymore. This means that, in the low-energy limit, we have to leave the space of rapidly decreasing test functions.

Curved space-time manifold. It turns out that a variant of the Epstein–Glaser approach can also be applied to models in quantum field theory on special curved space-time manifolds (i.e., globally hyperbolic pseudo-Riemannian manifolds).[14]

The relation to the classical exponential function. What is the secret behind the great success of the Epstein–Glaser approach? In order to answer this question, consider Euler's classical exponential function

$$f(t) = \mathrm{e}^{At}, \qquad t \in \mathbb{R}. \tag{19.18}$$

Here, A is a fixed real number. There exist two different approaches to this function in classical analysis, namely:

[13] Many applications of the Hahn–Banach theorem to optimization and optimal control can be found in E. Zeidler, Nonlinear Functional Analysis and its Applications, Vol. III: Variational Methods and Optimization, Springer, New York, 1986.

[14] See R. Brunetti and K. Fredenhagen, Micro-local analysis and interacting quantum field theories: renormalization on physical backgrounds, Commun. Math. Phys. **208** (2000), 623–661.

(ODE) the use of the ordinary differential equation

$$\frac{df(t)}{dt} = Af(t), \qquad t \in \mathbb{R}, \qquad f(0) = 1; \tag{19.19}$$

(FE) the use of the functional equation

$$\boxed{f(t+s) = f(t)f(s) \qquad \text{for all} \quad t, s \in \mathbb{R}.} \tag{19.20}$$

Let us discuss this.

Ad (ODE). Equation (19.19) has a unique smooth solution given by the exponential function (19.18). Replacing the real number A by a linear operator, the exponential function (19.18) describes

- one-parameter groups (reversible causal processes in nature), and
- semi-groups (irreversible causal processes in nature)

in terms of functional analysis.[15] In particular, the fundamental Dyson series for the S-matrix is related to this setting; in fact, the Dyson series is a generalization of Lagrange's variation-of-parameter method in celestial mechanics (see Sect. 7.17.4 of Vol. I).

Ad (FE). If the continuous function $f : \mathbb{R} \to \mathbb{R}$ satisfies the functional equation (19.20), then it is given by (19.18). Consequently, causal processes in nature can be described by equations of the form (19.20). In particular, the Epstein–Glaser approach is of this type: the functional equation (19.20) corresponds to the causality axiom (19.17).

In Chap. 7 we have studied the following approaches to quantum mechanics:

- Heisenberg's method (commutation relations, creation and annihilation operators),
- Schrödinger's method (partial differential equations),
- Feynman's method (path integral and Brownian motion in imaginary time),
- Weyl's method (pseudo-differential operators and deformation quantization),
- von Neumann's method (spectral theory of self-adjoint operators, and von Neumann operator algebras),
- the Gelfand–Naimark–Segal method (C^*-operator algebras).

> *There exists an additional approach to quantum mechanics based on the functional equation* (19.20).

This is the causal Epstein–Glaser approach which can be generalized to quantum field theory. This will be thoroughly studied in Vol. IV on quantum mathematics.

19.3 Kreimer's Hopf Algebra Revolution

> The symmetry behind renormalization theory can be described by Hopf algebras.
>
> Folklore

[15] See E. Zeidler, Nonlinear Functional Analysis, Vol. IIA: Linear Monotone Operators, Springer, New York, 1997.
P. Lax, Functional Analysis, Wiley, New York, 2002.

19.3.1 The History of the Hopf Algebra Approach

The following quotation is taken from the survey article by K. Ebrahimi-Fard and D. Kreimer, The Hopf algebra approach to Feynman diagram calculations. Topical Review. J. Phys. A: Mathematical and General **38** (2005), R385–R407:[16]

> Quantum field theory (QFT) has a long and outstandingly successful history in all theories of physics. Merging the two major revolutionary achievements of the early twentieth century physics, quantum mechanics and special relativity, the founding fathers of QFT were setting out for a unified description of elementary particles phenomena. Its ideas and techniques found far reaching applications in different and very distinct areas of theoretical physics and pure and applied mathematics. Several approaches to QFT have been developed so far. We mention
> - Wightman's early axiomatic setting leading to constructive QFT,[17]
> - together with Haag's mathematically elegant and rigorous algebraic formulation of QFT in terms of von Neumann algebras,[18]
>
> These two approaches describe best the nowadays common belief of what should be the general physical principles underlying any QFT. Still, despite the enormous and mathematically rigorous progress which has been made using these formulations, both approaches have several problems in making fruitful contact with experimental results, whilst they give a crucial insight into the structure of free quantum fields.
>
> The *perturbative approach* to quantum field theory is the most successful. Theoretical predictions of physical quantities made by using their expansion in terms of renormalized Feynman graphs match experimental results with a vertiginous high precision. Nevertheless, in most, if not all, of the interesting and relevant four-dimensional quantum field theories, performing even simple perturbative calculations one cannot avoid facing ill-defined integrals. The removal of these divergences in a sound way is the process of renormalization, better known by the illustrative description as 'sweeping under the carpet'. The basic idea of perturbative renormalization goes back to Kramers in 1938,[19] and was successfully applied for the first time in a 1947 seminal paper by Bethe (Phys. Rev. **72** (1947), 339–341) dealing with the concrete problem of the self-energy contribution for the Lamb shift in perturbative quantum electrodynamics (QED). The latter can nowadays be regarded as one of the best tested physics theories. Its modern extension to the Standard Model of elementary particles represents one of the cornerstones of our present understanding of the physical world. Here again

[16] Reprinted by permission of IOP Publishing, Bristol, United Kingdom. Internet: `www.iop.org/journals/jphysa`

[17] R. Streater and A. Wightman, PCT, Spin and Statistics, and All That, Addison-Wesley, Redwood City, California, 1968.
J. Glimm and A. Jaffe, Mathematical Methods of Quantum Physics: A Functional Integral Point of View, Springer, New York, 1981.
J. Glimm and A. Jaffe, Quantum Field Theory and Statistical Mechanics: Expositions, Birkhäuser, Boston, 1985.

[18] R. Haag, Local Quantum Physics, Springer, Berlin, 1996.

[19] See H. Kramers, Collected Scientific Papers, North-Holland, Amsterdam, 1956. The fascinating history of classical renormalization theory, including Kramers contributions, is described in detail by the collection of survey articles edited by L. Brown, Renormalization, Springer, New York, 1993.

the perturbative treatment together with renormalization is the bread and butter of the practitioner in high energy physics.

Maintaining the physical principles of locality, unitarity and Lorentz invariance, renormalization theory may be summed up by the statement that to all orders in perturbation theory the (ultraviolet) divergencies can be absorbed in a redefinition of the parameters defining the QFT. Here two distinct concepts enter,

- that of renormalizability, and
- the process of renormalization.

The former distinguishes those theories with only a finite number of parameters, lending them considerably more predictive power. The process of renormalization instead, works indifferently of the number of parameters.[20]

Soon after Bethe's paper on perturbative QED, there were several approaches to establish that quantum field theories are renomalizable in general.

(i) Dyson was the first to do so, using integral equations and skeleton expansions for Green's functions (Phys. Rev. **75** (1949), 1736–1755). His work was then continued by Salam and Weinberg. Unfortunately, this attempt failed in the first instance, due to a problem related to a particular 14th order QED graph, but could be cured later (see Fig. 18.1 on page 968).

(ii) The second approach, based on earlier work of Stueckelberg and Green (Helv. Phys. Acta **24** (1951), 153–174) was taken by Bogoliubov and Parasiuk (Acta Math. **97** (1957), 227–326) using a recursive subtraction method, Bogoliubov's R-method. Also their proof contained a loop-hole, but eventually found its final and satisfying form with the work of Hepp (Commun. Math. Phys. **2** (1966), 301–326) and later Zimmermann (Commun. Math. Phys. **15** (1969), 208–234). This standard result is nowadays well known under the name Bogoliubov–Parasiuk–Hepp–Zimmermann (BPHZ) renormalization.

(iii) Later, Epstein and Glaser presented a rigorous proof of renormalizability in the realm of the axiomatic treatment of QFT (Ann. Inst. Poincaré **A19**(3) (1973), 211–295).

(iv) A fourth approach was taken by Blaer and Young (Nucl. Phys. **B63** (1974), 493–514) using the Callan–Szymanzik renormalization group equations, going back to a suggestion by Callan.

Unfortunately, despite its accomplishments, renormalization theory was stigmatized, especially for its lack of a firm mathematical underpinning. Indeed, examining the current introductory and advanced literature on renormalization, as it is used in every day applications in many branches of physics, one feels the need for a more conceptual picture unifying mathematical and computational aspects. A possible reason for this situation might have been the fact that the building blocks, the (one-particle irreducible) Feynman graphs in itself appeared to be unrelated to a sound mathematical structure that may underlie the renormalization prescription in perturbative QFT.

[20] In any order n of perturbation theory, the procedure of renormalization yields an expression which only depends on a finite number $m(n)$ of parameters. However, it may happen that $\lim_{n\to+\infty} m(n) = \infty$. In this case, we say that the quantum field theory is not renormalizable.

Almost five decades after Bethe's work, this changed to a great extent with the original paper by Kreimer introducing the notion of Hopf algebra.[21] The ensuing work by Kreimer and his collaborators,[22] especially Broadhurst and Connes, explored this new approach both in terms of its mathematical and physical content, as well as its computational aspects. The Hopf algebraic approach captures the combinatorial and algebraic aspects of the process for renormalization by organizing the Feynman graphs into a combinatorial Hopf algebra, $\mathcal{H}_F$, which is a connected graded commutative bialgebra, essentially characterized by its non-cocommutative coproduct structure map. The Hopf algebra formulation of renormalization was completed in the work by Connes and Kreimer. It gives rise to an elegant and useful disentanglement of analytic and algebraic aspects of perturbative renormalization in general QFT, affirming the remark that *Few physicists object nowadays to the idea that diagrams contain more truth than the underlying formalism* by Veltman and 't Hooft (Diagrammar, CERN, Report 1973/9).

19.3.2 Renormalization and the Iterative Birkhoff Factorization for Complex Lie Groups

The program of obtaining a characterization of a function in simple descriptive terms, which are independent of the equations of definition of the function, is a familiar one. To Riemann is due the formulation of this characterization for the algebraic functions and for the functions defined by ordinary differential equations without irregular points. In both of these instances the characterization involves a certain number of characteristic constants – *the monodromic group constants* in the last mentioned instances. Riemann also proposed the associated problem of assigning these constants at pleasure.

During the last few years I have discovered that the program admits of extension in a number of directions. The aim of the present paper is to solve the generalized problem of Riemann for ordinary linear differential

[21] D. Kreimer, On the Hopf algebra structure of perturbative quantum field theories, Adv. Theor. Math. Phys. **2** (1999), 303–334.

[22] D. Kreimer, Chen's iterated integral represents the operator product expansion, Adv. Theor. Math. Phys. **3** (1999), 627–670.
D. Kreimer, On overlapping divergences, Commun. Math. Phys. **204** (1999), 669–698.
D. Broadhurst and D. Kreimer, Combinatorial explosion of renormalization tamed by Hopf algebra. 30-loop Padé–Borel resummation, Phys. Lett **B475**, 63–70.
D. Broadhurst and D. Kreimer, Towards cohomology of renormalization: bigrading the combinatorial Hopf algebra of rooted trees, Commun. Math. Phys. **215** (2000), 217–236.
D. Broadhurst and D. Kreimer, Exact solutions of Dyson–Schwinger equations for iterated one-loop integrals and propagator-coupling duality, Nucl. Phys. **B600** (2001), 403–422.
A. Connes and H. Moscovici, Hopf algebras, cyclic cohomology, and the transverse index theorem, Commun. Math. Phys. **198** (1998), 199–246.
A. Connes and D. Kreimer, Hopf algebras, renormalization and noncommutative geometry, Commun. Math. Phys. **119** (1998), 203–242. See also the footnote on page 995.

equations with irregular singular points, and the analogous problem for linear difference equations and for linear q-difference equations... The problem of Riemann for linear differential equations in its classic form was first solved by Hilbert in 1905.[23] His treatment and Plemel's elegant completion thereof reposed alike upon a certain theorem whose proof was made by means of the Fredholm theory for linear integral equations. Owing to the deep-seated analogy between linear differential and difference and q-difference equations, I have been able to apply a convenient extension of the same theorem in all cases; my proof is based on a *method of successive approximations.*[24]

George Birkhoff, 1913

It turned out that the whole iterative and intricate structure of renormalization theory could be mapped to the theory of Hopf algebras, with Zimmermann's *forest formula* for the counterterm coming along as the *antipode* (coinverse). This will be the starting point of our tour through the realms of perturbative quantum field theory (pQFT). While the next chapter summarizes some basics about pQFT, in Chapter 3 we will progress towards the introduction of this Hopf algebra. This Hopf algebra succinctly summarizes the combinatorics imposed on pQFT by the desire to obtain local counterterms. Meanwhile it turned out that this Hopf algebra is the classifying space for Hopf algebras of this kind,[25] a result emphasizing the beauty and naturalness of local point particle quantum field theories.[26]

Dirk Kreimer, 2000

This paper gives a complete self-contained proof of our result announced in 1999[27] showing that renormalization in quantum field theory is a special instance of a mathematical procedure of extraction of finite values based on the *Riemann–Hilbert problem*. We shall first show that for any quantum field theory, the combinatorics of Feynman graphs gives rise to a Hopf algebra $\mathcal{H}$ which is commutative as an algebra. It is the *dual* Hopf algebra of the enveloping algebra of a Lie algebra $\mathcal{G}$ whose basis is labelled by the one-particle irreducible Feynman graphs. The Lie bracket of two such graphs is computed from insertions of one graph in the other and vice versa. The corresponding Lie group G is the group of characters of $\mathcal{H}$. We shall then show that, using dimensional regularization, the bare (unrenormalized) theory gives rise to a loop

$$\gamma(z) \in G, \qquad z \in C$$

[23] D. Hilbert, Foundations of the general theory of linear integral equations, Part III, Göttinger Nachr. Ges. Wiss. Göttingen, 1905, 307–338 (in German).
J. Plemelj, Riemann's families of functions for given monodromy group, Monatshefte für Math. und Physik **19** (1908), 205–246 (in German). Plemelj's solution of Hilbert's 21th problem was incomplete, as we will discuss on page 1006.

[24] G. Birkhoff, The generalized Riemann problem for linear differential equations and the allied problems for linear difference and q-difference equations, Proc. Amer. Math. Acad. Arts and Sci. **49** (1913), 521–568.

[25] A. Connes and D. Kreimer, Hopf algebras, renormalization and noncommutative geometry, Commun. Math. Phys. **119** (1998), 203–242.

[26] D. Kreimer, Knots and Feynman Diagrams, Cambridge University Press, 2000 (reprinted with permission).

[27] A. Connes and D. Kreimer, J. High Energy Phys. **9** (1999), 024.
Internet: `http://arxiv.org/hep-th/9909126`

> where C is a a small circle of complex dimensions around the integer dimension D of space-time. Our main result is that the renormalized theory is just the evaluation at the point $z = D$ of the holomorphic part γ_+ of the Birkhoff decomposition of γ. We begin to analyze the group G and show that it is a semi-direct product of an easily understood Abelian group by a highly non-trivial group closely tied up with groups of diffeomorphisms. The analysis of this latter group as well as the interpretation of the renormalization group and of anomalous dimensions are the content of our second paper with the same overall title.[28]
>
> Alain Connes and Dirk Kreimer, 2000

This is one of the most beautiful papers ever written in mathematical physics. For a comprehensive representation, we refer to the monograph by A. Connes and M. Marcolli, Noncommutative Geometry, Quantum Fields, and Motives, Amer. Math. Soc., Providence, Rhode Island, 2008. In what follows let us sketch the basic ideas of this new approach to renormalization theory.

Freely generated commutative algebras. Let x be a symbol. By definition, the commutative algebra $\mathbb{C}[x]$ freely generated by x (over the field of complex numbers $\mathbb{C}$) consists of all the symbols

$$a_0 + a_1 x + a_2 x^2 + \dots$$

where $a_0, a_1, \dots$ are complex numbers, and only a finite number of these coefficients is different form zero. Similarly, let x and y be symbols. Then the commutative algebra $\mathbb{C}[x, y]$ freely generated by x and y (over $\mathbb{C}$) consists of all the symbols

$$a_{00} + a_{10} x + a_{01} y + a_{20} x^2 + a_{12} xy + a_{02} y^2 + \dots,$$

where only a finite number of the complex coefficient $a_{00}, a_{10}, \dots$ is different from zero. Multiplying such symbols, we use the relation $xy = yx$. Analogously, we can define the commutative algebra $\mathbb{C}[x_1, x_2, \dots]$ freely generated by the countable set of symbols $x_1, x_2, \dots$

The Hopf algebra of Feynman graphs. We consider one-particle irreducible (1PI) Feynman graphs related to a fixed quantum field theory. Consider the set of all the symbols

$$(\Gamma, \chi(\Gamma))$$

where $\chi(\Gamma)$ is the number of external lines of the Feynman graph Γ. Let $\mathcal{H}$ be the commutative algebra freely generated by these symbols (over $\mathbb{C}$). The crucial point is the definition of the coproduct by setting

$$\boxed{\Delta(\Gamma) := 1 \otimes \Gamma + \Gamma \otimes 1 + \sum_{\gamma} \gamma \otimes \Gamma/\gamma.} \tag{19.21}$$

Here, we sum over all proper subgraphs γ of the given graph Γ, which are the union of pairwise disjoint 1PI graphs. In the setting of the algebra $\mathcal{H}$, formula (19.21) is to be understood in the following sense: we briefly write Γ instead of $(\Gamma, \chi(\Gamma))$, and γ is the algebra product of the pairwise disjoint 1PI components of the graph γ. The graph Γ/γ is obtained from Γ by shrinking the subgraph γ to a point.

[28] A. Connes and D. Kreimer, Renormalization in quantum field theory and the Riemann–Hilbert problem I: The Hopf algebra structure of graphs and the main theorem. II: The beta function, diffeomorphisms, and the renormalization group. Commun. Math. Phys. **210** (2000), 249–273; **216** (2001), 215–241 (reprinted with permission).

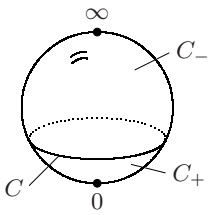


Fig. 19.4. Birkhoff factorization

It turns out that $\mathcal{H}$ is a Hopf algebra.

The definition (19.21) resembles the Bogoliubov iterative formula (19.3). In fact, the definition of $\Delta(\Gamma)$ is motivated by (19.3) on page 982.

The Birkhoff factorization for complex Lie groups and the general Riemann–Hilbert problem. Consider the decomposition

$$\mathbb{P}^1_{\mathbb{C}} = C_+ \cup C \cup C_-$$

of the Riemann sphere $\mathbb{P}^1_{\mathbb{C}}$ (Fig. 19.4). Here, by stereographic projection,

- the set C_+ corresponds to an open disc centered at the origin $z = 0$ in the complex plane $\mathbb{C}$;
- the curve C is the boundary of C_+, and
- the open set C_- is the complement to $C_+ \cup C$ on the Riemann sphere.

Let G be a connected complex Lie group. We are given a smooth map

$$l : C \to G.$$

We say that the loop l admits a Birkhoff factorization iff it can be written as a product

$$\boxed{l(z) = l_-(z)^{-1} l_+(z) \qquad \text{for all} \quad z \in C.} \tag{19.22}$$

Here, l_+ and l_- are the boundary values of holomorphic functions $l_\pm : C_\pm \to G$ with the normalization condition $l_-(\infty) = 1$. This factorization problem is not always solvable. The Riemann–Hilbert problem for ordinary differential equations and more general problems can be reduced to the Birkhoff factorization problem. In 1913 Birkhoff invented an iterative method for solving such factorization problems (see the quotation on page 993).

Renormalization and Birkhoff factorization. Connes and Kreimer proved that Bogoliubov's iterative R-method (19.3) on page 982 can be reformulated as the Birkhoff iterative method for a factorization problem with respect to the group $G := \mathrm{Hom}(\mathcal{H}, \mathbb{C})$. This group consists of algebra morphisms $\mu : \mathcal{H} \to \mathbb{C}$ with the normalization condition $\mu(\mathbf{1}) = 1$. The product on G is chosen in such a way that it is dual to the coproduct on the Hopf algebra $\mathcal{H}$ of Feynman graphs. In this setting, the *renormalization group* is a one-parameter subgroup of G.

19.3.3 The Renormalization of Quantum Electrodynamics

The relation between renormalization theory and the Birkhoff factorization was proven by Connes and Kreimer for a scalar field theory (see Sect. 19.3.2). This approach was applied to quantum electrodynamics in the paper by

D. Volovich and I. Prokhorenko, Renormalizations in quantum electrodynamics, and Hopf algebras, Trudy Mat. Inst. Steklova **245** (2004), 288–295. Internet: `http://arxiv.org/hep-th/0611178`

The authors proved the renormalizability of quantum electrodynamics by solving a Riemann–Hilbert problem. This research was extended by Suiljekom. Let us quote from his papers:

> We report on the Hopf algebraic description of renormalization theory of quantum electrodynamics. The Ward–Takahashi (WT) identities are implemented as linear relations on the (commutative) Hopf algebra of Feynman graphs of quantum electrodynamics. Compatibility of these relations with the Hopf algebra structure is the formulation of the physical fact that WT identities are compatible with renormalization. As a result, the counterterms and the renormalized Feynman amplitudes automatically satisfy the WT identities, which leads in particular to the well-known Ward identity $Z_e = Z_A^{-1/2}$.[29]
>
> Walter van Suijlekom, 2006

> We study the Connes–Kreimer Hopf algebra of renormalization in the case of gauge theories. We show that the Ward–Takahashi identities and the Slavnov–Taylor identities (in the Abelian and Non-Abelian case respectively) are compatible with the Hopf algebra structure, in that they generate a Hopf ideal. Consequently, the *quotient Hopf algebra* is well-defined and has those identities built in. This provides a purely combinatorial and rigorous proof of compatibility of the Slavnov–Taylor identities with renormalization.[30]
>
> Walter van Suijlekom, 2006

19.4 The Scope of the Riemann–Hilbert Problem

The Riemann–Hilbert problem is related to the following important problems in mathematical physics:

- renormalization (see Sect. 19.3.2 above),
- Fuchsian differential equations (generalizations of the Gauss hypergeometric differential equation) and the 21th Hilbert problem,
- solitons.

The point is that the solution of the Riemann–Hilbert problem allows us to construct analytic functions. Moreover, Riemann–Hilbert problems are closely related to singular integral equations and to holomorphic vector bundles. Let us briefly discuss some important applications. This fascinating line of historical development connects the Gaussian hypergeometric differential equation with modern complex function theory, modern algebraic geometry, and renormalization theory.

[29] W. van Suijlekom, The Hopf algebra of Feynman graphs in quantum electrodynamics, Letters in Mathematical Physics **77** (2006), 265–281 (reprinted with permission). Internet: `http://arxiv.org/hep-th/0602126`

[30] W. van Suijlekom, Renormalization of gauge fields: the Hopf algebra approach, 2006. Internet: `http://arxiv.org/hep-th/0610137`.

19.4.1 The Gaussian Hypergeometric Differential Equation

Let us introduce the so-called hypergeometric series

$$F(a,b,c;z) := 1 + \frac{ab}{c}\,z + \frac{a(a+1)b(b+1)}{2!\,c(c+1)}\,z^2 + \frac{a(a+1)(a+2)b(b+1)(b+2)}{3!\,c(c+1)(c+2)}\,z^3 + \dots \tag{19.23}$$

We assume that a, b, c are complex numbers with $c \neq 0, -1, -2, \dots$ In the special case where $a = b = c = 1$, this is the geometric series. In 1769 Euler studied the second-order differential equation

$$\boxed{w''(z) + \frac{c-(a+b+1)z}{z(z-1)}\,w'(z) - \frac{ab}{z-1}\,w(z) = 0} \tag{19.24}$$

for real values z, and he found that (19.23) is a solution of (19.24), in the sense of a formal power series expansion.[31] In 1813 Gauss published his paper *Disquisitiones generales circa seriem infinitam* (On infinite series). He proved the following:

- If $|z| < 1$, then the series $F(a,b,c;z)$ is absolutely convergent. This was the first convergence proof for a power series expansion in the history of mathematics.
- If a or b is equal to one of the integers $0, -1, -2, \dots$, then $F(a,b,c;z)$ is a polynomial. Otherwise the series $F(a,b,c;z)$ is divergent if $|z| > 1$.
- If $|z| = 1$ and the real part of $a+b-c$ is negative, then $F(a,b,c;z)$ is absolutely convergent.

Gauss noticed that many of the functions known at his time were special cases of the hypergeometric function F. For example,

- $(1+z)^n = F(-n,1,1;-z)$ for all $z \in \mathbb{C}$;
- $\ln(1+z) = zF(1,1,2;-z)$ for all complex numbers z with $|z| < 1$;
- $\mathrm{e}^z = \lim_{m\to+\infty} F(1,m,1,z/m)$ for all complex numbers z.

Moreover, Gauss discovered the limiting formula

$$\lim_{x\to 1-0} F(a,b,c;x) = F(a,b,c;1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} \tag{19.25}$$

for all real numbers a, b, c with $a+b-c < 0$ and $c \neq 0, -1, -2, \dots$ Furthermore, Gauss proved the integral formula

$$\int_0^x t^{\lambda-1}(1-t^\mu)^\nu dt = \frac{x^\lambda}{\lambda}\,F\left(-\nu,\ \frac{\lambda}{\nu},\ \frac{\lambda}{\mu}+1;\ x^\mu\right), \qquad 0 < x < 1$$

for all real numbers $\lambda > 0, \mu > 0, \nu \neq 0$. Letting $x \to 1-0$ and using (19.25), Gauss obtained the key integral formula

$$\int_0^1 t^{\lambda-1}(1-t^\mu)^\nu dt = \frac{\Gamma(\frac{\lambda}{\mu}+1)\Gamma(\nu+1)}{\lambda\Gamma\left(\frac{\lambda}{\mu}+nu+1\right)}. \tag{19.26}$$

[31] Euler (1707–1783), Gauss (1777–1855), Kummer (1810–1893), Riemann (1826–1866), Fuchs (1839–1902), Klein (1842–1925), Poincaré (1854–1912), Painlevé (1863–1933), Hilbert (1862–1943), Koebe (1882–1945), Birkhoff (1884–1944).

provided the limit exists as indicated above. In particular, this yields the (lemniscatic) elliptic integral

$$\int_0^1 \frac{dt}{(1-t^4)^{1/2}} = F\left(-\tfrac{1}{2}, 2, \tfrac{5}{4}; 1\right) = \frac{\Gamma\left(\frac{5}{4}\right)\Gamma\left(-\frac{1}{4}\right)}{\Gamma\left(\frac{3}{4}\right)}.$$

Using analytic continuation, formula (19.26) remains true as a generalized integral formula if the right-hand side of (19.26) makes sense (i.e., the gamma function has no poles). Explicitly, we assume that λ, μ, ν are complex numbers with the property $\nu \neq -1, -2, \ldots$ and $\mu \neq 0$, as well as $\frac{\lambda}{\mu} + \nu + 1 \neq 0, -1, -2, \ldots$ If $\lambda > 0$ and $\nu > -1$, formula (19.26) yields the Euler beta function $B(\lambda, \nu + 1)$.

Finally, let us mention that the two functions w_1, w_2 given by

$$w_1(z) := F(a, b, c; z), \quad w_2(z) := z^{1-c} F(1 + a - c, 1 + b - c, 2 - c; z), \quad |z| < 1$$

form a basis for the solutions of the Gauss hypergeometric differential equation (19.24) on the open unit disc. In other words, each solution of (19.24) can be represented by the formula

$$w(z) = C_1 w_1(z) + C_2 w_2(z), \qquad |z| < 1$$

where C_1 and C_2 are complex constants. In 1836, using symmetry transformations of the hypergeometric function F, Kummer found 24 different local solutions in different regions of the complex plane outside the singular points $z = 0, 1, \infty$.[32]

Analytic Continuation and Riemann's monodromy method. In 1857 Riemann wrote a famous paper on the global solution of the Gaussian hypergeometric differential equation. In modern terminology, Riemann studied second-order Fuchsian differential equations which have three singular points, say $z = 0, 1, \infty$. The Riemann surface of this global solution has two sheets and the three branch points $z = 0, 1, \infty$. Riemann introduced the symbol

$$P \begin{pmatrix} 0 & 1 & \infty \\ \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \end{pmatrix}, \tag{19.27}$$

which characterizes the local behavior of the global solution near the branch points (see (19.32) below). For example, the Gaussian equation (19.29) has the symbol

$$P \begin{pmatrix} 0 & 1 & \infty \\ 0 & 0 & a \\ 1-c & c-a-b & b \end{pmatrix}. \tag{19.28}$$

Riemann used analytic continuation in order to construct the Riemann surface of the global solution. This way, Riemann invented the monodromy method which studies the analytic continuation of local solutions along loops around the singular points (see Sect. 19.4.3 below.) In particular, Riemann showed that, roughly speaking, Kummer's 24 local solutions represent all possible local elements of the global

[32] R. Prosser, On the Kummer solutions of the hypergeometric equation, Amer. Math. Monthly **101** (1994), 535–543.

solution of the hypergeometric equation.[33] Riemann used a minimum of information on the singular points of the differential equation in order to get a maximum of information. Hilbert said:

> *Riemann has shown us that proofs are better achieved through ideas than through long calculations.*

Fuchsian Differential Equations

Singular points. In the late 1860s Fuchs generalized Riemann's approach to ordinary differential equations of nth order with a finite number of singular points. For example, consider the second-order differential equation

$$\boxed{w''(z) + p(z)w'(z) + q(z) = 0.} \tag{19.29}$$

We assume that the coefficient functions p and q are holomorphic on the complex plane up to a finite number of points. The point z_0 is called a regular point of (19.29) iff p and q are holomorphic at z_0. Otherwise the point z_0 is called a singular point of (19.29). To classify the point $z = \infty$, we use the coordinate transformation $z = 1/\zeta$. The point $z = \infty$ is called a regular (resp. singular) point of (19.29) iff $\zeta = 0$ is a regular (resp. singular) point of the transformed differential equation. By definition, the equation (19.29) is of Fuchsian type iff there exists a finite number of points $z_1, \ldots, z_n$ in the complex plane $\mathbb{C}$ such that

$$p(z) = \sum_{k=1}^{n} \frac{A_k}{z - z_k}, \qquad q(z) = \sum_{k=1}^{n} \frac{B_k}{(z - z_k)^2} + \frac{C_k}{z - z_k}. \tag{19.30}$$

Here, A_k, B_k, C_k $(k = 1, \ldots, n)$ are complex numbers with $\sum_{k=1}^{n} C_k = 0$. Let us consider two examples:

- $n = 1$: Using the Euler transformation $u = \ln(z - z_1)$, the Euler equation

$$w''(z) + \frac{A_1}{z - z_1}\, w(z) + \frac{B_1}{(z - z_1)^2}\, w(z) = 0$$

  passes over to a differential equation with constant coefficients. This is a trivial case from the point of view of singularities.
- $n = 2$: The Gaussian hypergeometric equation (19.23) is of Fuchsian type with the three singular points $z_1 = 0, z_2 = 1$ and $z = \infty$.

Local solutions at singular points. Let us motivate the notion of Fuchsian differential equation.[34] Suppose that the point $z_0 \in \mathbb{C}$ is a singular point of the differential equation (19.29). Then the general solution of (19.29) in an open neighborhood of z_0 can be written in the form

$$w(z) = C_1 w_1(z) + C_2 w_2(z), \qquad 0 < |z - z_0| < r,$$

[33] Much material can be found in the classic textbook by E. Whittaker and G. Watson, A Course of Modern Analysis: An Introduction to the General Theory of Infinite Processes and of Analytic Functions; with an Account of the Principal Transcendental Functions, Cambridge University Press, 1944.

[34] A detailed motivation can be found in V. Smirnov, A Course of Higher Mathematics, Vol. 3, Pergamon Press, New York, 1964.

with complex constants C_1 and C_2 and a positive number r.
Case 1: The functions w_1 and w_2 have the form

$$w_1(z) = (z - z_0)^\alpha \mathcal{L}_1(z - z_0), \qquad w_2(z) = (z - z_0)^\beta \mathcal{L}_2(z - z_0)$$

with real coefficients α and β.
Case 2: The function w_1 is given by Case 1, while the function w_2 has the following form:

$$w_2(z) = (z - z_0)^\alpha \mathcal{L}_3(z - z_0) + w_1(z) \ln(z - z_0).$$

Here, $\mathcal{L}_j(z - z_0) = \sum_{k=-\infty}^{\infty} a_{jk}(z - z_0)^k$ denotes a Laurent series at the point z_0.

The singular point z_0 is called weakly singular iff $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3$ are power series expansions at the point z_0.[35]

This is the case if and only if

- p has a pole of order at most one at z_0, and
- q has a pole of order at most two at z_0.

By definition, the differential equation (19.29) is of Fuchsian type iff it has a finite number of singular points and all of them are weakly singular. This definition includes the point $z = \infty$.

Construction of local solutions. The local solutions of the Fuchsian differential equation (19.29), (19.30) at the point $z_0 := z_j (j = 1, \ldots, n)$ can be constructed in the following way: The ansatz

$$w(z) = (z - z_0)^\varrho (a_0 + a_1(z - z_0) + a_2(z - z_0)^2 + \ldots) \tag{19.31}$$

leads in (19.29) to the so-called index equation

$$\varrho^2 + \varrho(A - 1) + B = 0$$

with the two solutions $\varrho = \alpha$ and $\varrho = \beta$.
Case A: The difference $\alpha - \beta$ is not an integer. Then one gets the solutions w_1 and w_2 by using the ansatz (19.29) with $\varrho = \alpha, \beta$ and comparing coefficients.
Case B: The difference $\alpha - \beta$ is an integer. Then we get w_1 as in the first case, while the second solution w_2 is obtained through integration, applying the formula

$$\frac{d}{dz}\left(\frac{w_2}{w_1}\right)(z) = \frac{1}{w_1(z)^2} \exp\left(-\int_{z_0}^{z} p(t)dt\right).$$

The local solutions at the point $z = \infty$ are obtained by using the transformation $z = 1/\zeta$ and applying the method above to the transformed differential equation at the point $\zeta = 0$. The symbol

$$P\begin{pmatrix} z_1 & z_2 & \ldots & z_n & \infty \\ \alpha_1 & \alpha_2 & \ldots & \alpha_n & \alpha_\infty \\ \beta_1 & \beta_2 & \ldots & \beta_n & \beta_\infty \end{pmatrix}, \tag{19.32}$$

introduced by Riemann in his fundamental 1857 paper, tells us that $z_1, \ldots, z_n, \infty$ are singular points with the indices α_j, β_j at the point z_j. If $z = \infty$ is a regular point, then the last row of (19.32) drops out.

[35] Otherwise it is called strongly singular.

Poincaré's Automorphic Functions and the Uniformization Theorem for Compact Riemann Surfaces

> The development starting with Gauss culminated in Poincaré's creation of the theory of Fuchsian groups and automorphic functions.
>
> Folklore

One of the high-lights in mathematics is the uniformization theorem. Let us briefly discuss this.

Real algebraic curves. To begin with, let $\mathbb{K} = \mathbb{R}$. In order to explain the basic idea, consider the curve

$$x^2 + y^2 = 1, \qquad (x, y) \in \mathbb{K}^2. \tag{19.33}$$

This is the unit circle which has the global parametrization

$$x = \cos t, \quad y = \sin t. \tag{19.34}$$

Here, the parameter t varies on the real line $\mathbb{R}$. Algebraic geometry studies curves which are given by algebraic equations of the form

$$P(x, y) = 0, \qquad (x, y) \in \mathbb{K}^2,$$

where P is a polynomial with respect to the variables x and y in $\mathbb{K}$ and coefficients in $\mathbb{K}$.

Complex algebraic curves. Riemann emphasized that one should study algebraic curves in terms of complex variables, that is, we set $\mathbb{K} = \mathbb{C}$. Then equation (19.34) describes a one-dimensional complex manifold. Such manifolds are also called Riemann surfaces. The global parametrization (19.34) remains valid if the parameter t varies on $\mathbb{C}$.

The uniformization theorem. In 1907 Poincaré and Koebe proved independently that every compact Riemann surface allows a global parametrization, which can be obtained by automorphic functions. This is the famous uniformization theorem.

Automorphic functions. Recall that transformations of the form

$$w = \frac{az + b}{cz + d}$$

are called Möbius transformations iff the complex coefficients a, b, c, d satisfy the condition $ad - bc \neq 0$. Every Möbius transformation generates a conformal diffeomorphism of the Riemann sphere $\mathbb{P}^1_{\mathbb{C}}$ onto itself. These transformations form a group called the automorphic group $\mathrm{Aut}(\mathbb{P}^1_{\mathbb{C}})$. Meromorphic functions which are invariant under a discrete subgroup of the automorphic group are called automorphic functions. This generalizes periodic and double-periodic (i.e., elliptic) functions. The theory of automorphic functions was created by Klein and Poincaré. The latter used solutions of Fuchsian differential equations and series of (Poincaré) theta functions in order to construct automorphic functions. We refer to:

L. Ford, Automorphic Functions, Chelsea, New York, 1972.

M. Farkas and I. Kra, Riemann Surfaces, Springer, New York, 1992.

M. Waldschmidt et al. (Eds.), From Number Theory to Physics, Springer, New York, 1995 (e.g., see the article by Bost on compact Riemann surfaces and algebraic curves, Jacobians, and Abelian varieties).

The Six Nonlinear Painlevé Equations

Consider the second-order ordinary differential equation

$$w''(z) = P(z, w(z), w'(z)),$$

where we assume that the function P is meromorphic in the complex variable z and rational in the complex variables w and w'. This equation is said to have the Painlevé property iff every solution $w = w(z)$ has a meromorphic continuation to the universal covering of a punctured Riemann sphere which is determined by the equation only. In about 1900 Painlevé and Gambier determined the complete list of all the equations which possess the Painlevé property, up to proper transformations of the independent and dependent variables. This complete list of 50 equations can be found in E. Ince, Ordinary Differential Equations, Dover, New York, 1956. There are six distinguished nonlinear equations among the list. The solutions of these six equations are called the Painlevé transcendents. The first (resp. second) Painlevé equation reads as

$$w''(z) = 6w^2(z) + z \qquad (\text{resp. } w''(z) = 2w^3(z) + zw(z) + a),$$

where a is a parameter.

Hints for further reading. The history of the theory of ordinary differential equations in the sense of Fuchs and Painlevé together with a discussion of the main results including the relations to algebraic geometry and Galois theory can be found in:

J. Gray, Linear Differential Equations and Group Theory: From Riemann to Poincaré, Birkhäuser, Boston, 2000.

K. Iwasaki, From Gauss to Painlevé: A Modern Theory of Special Functions. Vieweg, Wiesbaden, 1991.

S. Kichenassamy, Fuchsian Reduction: Applications to Geometry, Cosmology, and Mathematical Physics, Birkhäuser, Boston, 2007 (applications to nonlinear partial differential equations).

We also recommend:

F. Klein, Development of Mathematics in the 19th Century, Math. Sci. Press, New York, 1979.

The further development of these topics in the 20th century is described in:

J. Dieudonné, History of Functional Analysis, 1900–1975, North-Holland, Amsterdam, 1983.

J. Dieudonné, History of Algebraic Geometry, 400 B.C.-1985 A.C., Chapman, New York.

J. Dieudonnè, A History of Algebraic and Differential Topology, 1900–1960, Birkhäuser, Boston, 1989.

We also refer to:

E. Zeidler, Reflections on the future of mathematics, 25pp. In: H. Wußing, 6000 Years of Mathematics: a Cultural Journey through Time, Vol. II, Springer, Heidelberg, 2008 (in German).

The development of mathematics in the 20th century was strongly influenced by Hilbert's 23 problems which he formulated in 1900. We refer to:

B. Yandell, The Honors Class: Hilbert's Problems and Their Solvers, Peters Ltd, Natick, Massachusetts, 2001.

J. Gray, The Hilbert Challenge: A Perspective on 20th Century Mathematics, Oxford University Press, Oxford, 2002.

P. Odifreddi, The Mathematical Century: The 30 Greatest Problems of the Last 100 Years, Princeton University Press, 2004.

Hilbert's 21th problem will be discussed in Sect. 19.4.3.

19.4.2 The Confluent Hypergeometric Function and the Spectrum of the Hydrogen Atom

The confluent hypergeometric function. Parallel to the hypergeometric series (19.23), we define the so-called confluent hypergeometric series

$$\mathcal{F}(a,c;z) := 1 + \frac{a}{c}\,z + \frac{a(a+1)}{2!\;c(c+1)}\,z^2 + \frac{a(a+1)(a+2))}{3!\;c(c+1)(c+2)}\,z^3 + \dots, \quad (19.35)$$

which was studied by Kummer in the 1830s. We assume that a and c are complex numbers with $c \neq 0, -1, -2, \dots$ The series (19.35) is convergent for all z. The function $z \mapsto \mathcal{F}(a,c;z)$ satisfies the differential equation

$$w''(z) + \left(\frac{c}{z} - 1\right) w'(z) - aw(z) = 0, \qquad z \in \mathbb{C}, \quad (19.36)$$

which is called the confluent hypergeometric equation. Replacing z by z/b in the Gaussian hypergeometric equation (19.23) and letting $b \to +\infty$, we obtain (19.36). In contrast to the three singular points $0, 1, \infty$ for the hypergeometric equation (19.23), the confluent hypergeometric differential equation (19.36) has only the two singular points 0 and ∞. In addition, this equation is not of Fuchsian type. The singular point $z = 0$ is weak, but the singular point $z = \infty$ is strong.

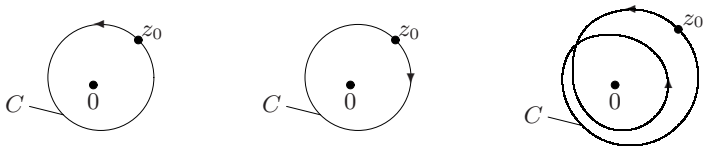
The spectrum of the hydrogen atom. In Vol. III we will thoroughly study the spectrum of the hydrogen atom (both bound and scattering states). The corresponding solutions of the Schrödinger equation are closely related to the confluent hypergeometric function. To compute these solutions in 1926, Erwin Schrödinger asked his colleague Hermann Weyl in Zurich for help; Weyl had created the spectral theory of singular differential equations in 1910. This theory was perfected by von Neumann in 1929 (general spectral theory of self-adjoint operators), and by Kodaira in 1949.[36]

19.4.3 Hilbert's 21th Problem

Linear Fuchsian systems of ordinary differential equations. Fix the dimension $n = 2, 3, \dots$ Let $z_1, \dots, z_k$ be a family of k different complex numbers, $k = 1, 2 \dots$ Consider the linear system

$$\boxed{\frac{dw(z)}{dz} = A(z)w(z), \qquad z \in \mathbb{C},\ z \neq z_1, \dots, z_k} \quad (19.37)$$

[36] K. Kodaira, The eigenvalue problem for ordinary differential equations of the second order and Heisenberg's theory of S-matrices. Amer. J. Math. **71** (1949), 921–945.



(a) $\gamma = 1$ (b) $\gamma = -1$ (c) $\gamma = 2$

Fig. 19.5. Monodromy

of ordinary differential equations. We assume that the complex $(n \times n)$-matrix $A(z)$ has the form

$$A(z) = \sum_{j=1}^{k} \frac{A_j}{z - z_j}$$

where $A_1, \ldots, A_k$ are constant complex nonzero $(n \times n)$-matrices. The points $z_1, \ldots, z_k$ are called singular, whereas the complementary points $z \in \mathbb{C} \setminus \{z_1, \ldots z_k\}$ are called regular. The point $z = \infty$ is called regular iff $\sum_{j=1}^{k} A_j = 0$. Otherwise, $z = \infty$ is called singular. The system (19.37) is called a Fuchsian system.[37]

For example, if we write the Gaussian hypergeometric differential equation (19.24) as a first-order system, then this is a Fuchsian system with the three singular points $z = 0, 1, \infty$. In what follows, we assume that $z = \infty$ is a regular point of (19.37).[38]

Riemann's monodromy group. Choose a regular point $z_0 \in \mathbb{C}$. Then there exist n linearly independent local solutions $w_1, \ldots, w_n$ of (19.37) which are holomorphic on some open neighborhood of the point z_0. Consider the punctured Riemann sphere

$$\mathcal{R} := \mathbb{P}^1_{\mathbb{C}} \setminus \{z_1, \ldots, z_k\},$$

and choose a loop C on $\mathcal{R}$. Fig. 19.5 shows some loops that wind around the origin (γ is the winding number). The point is that

- analytic continuation of the functions $w_1, \ldots, w_n$ along the loop C yields
- the new local solutions $w_1^+, \ldots, w_n^+$ in some open neighborhood of the point z_0.
- Since the functions $w_1, \ldots, w_n$ form a local solution basis, there exists a constant $(n \times n)$-matrix (a_{ij}) such that

$$w_i^+(z) = \sum_{j=1}^{n} a_{ij} w_j(z), \qquad i = 1, \ldots, n$$

for all complex numbers z on some open neighborhood of the point z_0.

One can show that the map $C \mapsto (a_{ij})$ has the following two crucial properties:

[37] Alternatively, such system can be characterized in the following way: The entries of the matrix function $z \mapsto A(z)$ are rational functions with weak singularities (i.e., the poles have order one), and the coordinate transformation $\zeta = 1/z$ yields a differential equation which has at most a weak singularity at the point $\zeta = 0$. The latter property describes the behavior of the differential equation at infinity. Finally, we exclude the case of constant coefficients.

[38] This situation can always be arranged by using a suitable Möbius transformation $\zeta = \frac{az+b}{cz+d}$ with $ad - bc \neq 0$.

- The matrix (a_{ij}) remains unchanged under deformations of the loop in the manifold $\mathcal{R}$.
- The composition of loops corresponds to the matrix product.

This way we obtain a group morphism

$$\chi : \pi_1(\mathcal{P}) \to GL(n, \mathbb{C}) \tag{19.38}$$

from the fundamental group $\pi_1(\mathcal{R})$ of the manifold $\mathcal{R}$ onto a subgroup of the group $GL(n, \mathbb{C})$.[39] If we change the regular point z_0 or the local basis $w_1, \ldots, w_n$, then the corresponding map χ changes in a natural way. Therefore, the monodromy map (19.37) describes the global solutions of the Fuchsian differential system (19.37).

The Riemann–Hilbert problem. We are given the k points $z_1, \ldots, z_k \in \mathbb{C}$ with $k \geq 3$. Let $n \geq 2$. The 21th Hilbert problem reads as follows:

> *Is it true that every group morphism* (19.37) *is the monodromy morphism of an* $(n \times n)$*-Fuchsian system with the singularities* $z_1, \ldots, z_k$*?*

In 1956 Krylow proved that the answer is 'yes' for (2×2)-Fuchsian systems and three singular points. In 1989 Bolibruch proved that the answer is 'no' in the general case. However, the answer is 'yes' if we pass to a larger class of first-order differential systems with a finite number of singularities called regular systems. The proofs can be found in:

> D. Anosov and A. Bolibruch, The Riemann–Hilbert Problem, Vieweg, Wiesbaden, 1994.

The history of this problem is thoroughly discussed in:

> A. Bolibruch, Inverse monodromy problems of the analytic theory of differential equations, pp. 49–74. In: A. Bolibruch, Yu. Osipov, and Ya. Sinai (Eds.), Mathematical Events of the Twentieth Century, Springer, Berlin/Phasis, Moscow, 2006.

In 1908 Plemelj proved that the answer is 'yes' for regular systems. He reduced the monodromy problem to a Riemann–Hilbert problem and solved this by using singular integral equations. Plemelj also claimed that his argument could be translated to Fuchsian systems. However, this claim is not true; Bolibruch constructed a counterexample. In 1957 Röhrl generalized Plemelj's results from the Riemann sphere to more general Riemann surfaces by using the modern theory of fiber bundles. For further results, we refer to:

> N. Muskhelishvili, Singular Integral Equations: Boundary Problems of Function Theory and Their Applications to Mathematical Physics, Nordhoff, Groningen, 1953.
>
> P. Deligne, Équations différentielles á point singuliers réguliers. Lecture Notes in Math. **163**, Springer, Berlin, 1970.
>
> A. Beauville, Monodromie des systèmes différentiels linéaire à pôles simple sur la sphère de Riemann, Séminaire de Bourbaki, n. 765, 1992/93, pp. 1–17 (in French).

[39] Recall that the group $GL(n, \mathbb{C})$ consists of all complex invertible $(n \times n)$-matrices.

19.4.4 The Transport of Information in Nature

The philosophy of modern physics is that the transport of physical information is carried out by transport along paths in the space-time manifold.

> *Monodromy and holonomy in mathematics model the transport of information along loops and the possible gain of information. This allows us to describe the strength of interactions.*

In fact, in gauge theory the parallel transport along small loops yields the curvature, and hence the strength of the fundamental forces in nature. This concerns

- the Standard Model in particle physics for the strong and electroweak force, and
- Einstein's theory of general relativity for gravitation.

Monodromy also plays the decisive role in Ashtekar's loop gravity (see Vol. VI on quantum gravity). For a survey, we refer to:

> A. Ashtekar and J. Lewandowski, Background independent quantum gravity: a status report, Class. Quant. Grav. **21** (2004), R53–R152.
>
> A. Ashtekar, M. Bojewald, and J. Lewandowski, Mathematical structure of loop quantum cosmology, Adv. Theor. Math. Phys. **7** (2003), 233–268.
>
> C. Fleischhack, Kinematical uniqueness of loop gravity, pp. 203–218. In: B. Fauser, J. Tolksdorf, and E. Zeidler (Eds.), Quantum Gravitation: Mathematical Models and Experimental Bounds, Birkhäuser, Basel, 2006.

19.4.5 Stable Transport of Energy and Solitons

> Waves are used in nature in order to transport energy and information.
>
> Folklore

Solitons are particle-like very stable waves. Such solitons appear in many fields of physics (e.g., water waves – catastrophic ocean waves called tsunami, laser optics, waves in optical fibers, plasma physics, conducting polymers, dislocations and plastic deformation of crystals, ferromagnetic and anti-ferromagnetic material, Bose–Einstein condensates, energy localization and energy transfer in proteins (DNA), blood pressure waves). Many beautiful applications in physics, chemistry, and biology can be found in:

> T. Dauxois and M. Peyrard, Physics of Solitons, Cambridge University Press, 2006.

The mathematical theory is based on inverse scattering theory, Lax pairs, the Riemann–Hilbert problem, Riemann surfaces, the Painlevé equation, the Lie–Bäcklund transformation, and Frobenius manifolds.

> *This combines sophisticated tools from algebra, analysis, and geometry with physics in a fascinating way.*

For example, this concerns spectral theory in functional analysis, algebraic geometry, and algebraic topology. We refer to the following two survey articles:

> B. Dubrovin, V. Matveev, and S. Novikov, Nonlinear equations of Korteweg–de Vries type, finite-zone linear operators, and Abelian varieties, Russian Mathematical Surveys **31**(1) (1976), 59–146.
>
> A. Its, The Riemann–Hilbert problem and integrable systems, Notices Amer. Math. Soc. **50**(11) (2003), 1389–1400.

A collection of reprints of 76 important articles on both classical solitons and quantized solitons (quantum field theory, elementary particles, superconductivity, magnetic monopoles in ccosmology) can be found in:

C. Rebbi and G. Soliani (Eds.), Solitons and Particles, World Scientific, Singapore, 1984.

We also refer to the following monographs:

S. Novikov, S. Manakov, L. Pitaevskii, and V. Zakharov, Solitons: The Inverse Scattering Method, Consultant Bureau, New York, 1984 (including the Riemann–Hilbert method).

Y. Yang, Solitons in Field Theory and Nonlinear Analysis, Springer, New York, 2001.

M. Toda, Theory of Nonlinear Lattices, Springer, New York, 1978.

M. Toda, Nonlinear Waves and Solitons, Kluwer, Dordrecht, 1989.

P. Deift, Orthogonal Polynomials and Random Matrices: A Riemann–Hilbert Approach, Courant Lecture Notes in Mathematics, vol. 3, CIMS, New York, 1999.

A. Scott, Nonlinear Science, Oxford University Press, 1999.

The crucial point is that:

The mathematical methods used in soliton theory can be regarded as a generalization of the classical Fourier transform to nonlinear problems describing interactions in physics (nonlinear harmonic analysis).

This allows us to compute explicit solutions for complicated nonlinear problems in physics. As an introduction to the computational aspects of soliton theory, we recommend:

R. Kaushal and D. Parashar, Advanced Methods of Mathematical Physics, Alpha Science, Pangbourne, India, 2000.

B. Felsager, Geometry, Particles, and Fields, Springer, New York, 1997.

G. Lamb, Elements of Soliton Theory, Wiley, New York, 1980.

G. Eilenberger, Solitons: Mathematical Methods for Physicists, Springer, New York, 1981.

As high-lights, we mention

- the construction of instantons (models for elementary particles) by using the Penrose twistor theory,
- magnetic monopoles generalizing the classic Dirac magnetic monopole, and
- the construction of solutions for the Einstein equations in general relativity (reduced to the Ernst equation), which describe thin rotating galaxies in the universe:

In this connection, we refer to:

M. Atiyah, V. Drinfeld, M. Hitchin, and Yu. Manin, Construction of instantons, Phys. Letters **65A** (1978), 185–187.

M. Atiyah, Geometry of Yang–Mills Fields, Lezioni Fermiani, Academia Nazionale dei Lincei Scuola Normale Superiore, Pisa, Italia, 1979 (instantons).

A. Jaffe and C. Taubes, Vortices and Monopoles: Structure of Static Gauge Theories, Birkhäuser, Boston, 1980.

C. Klein and O. Richter, Ernst Equation and Riemann Surfaces: Analytical and Numerical Methods, Springer, Berlin, 2006 (thin discs of rotating galaxies).

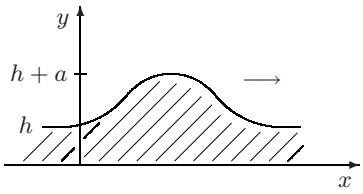


Fig. 19.6. Solitary water wave in a channel

In the latter monograph, the solutions are represented by Poincaré theta series. The parameters are obtained from the structure of hyperelliptic Riemann surfaces of genus $g > 1$. The difficulties are caused by collapsing pairs of branching points of the Riemann surfaces. Further material can be found on page 1015.

19.4.6 Ariadne's Thread in Soliton Theory

We want to sketch some basic ideas by following the line of historical development.

Solitary water waves. In nature, solitary waves were observed first by John Scott Russel at the Edinburgh–Glasgow channel in the United Kingdom in 1834. He published his observations in 1844 (Fig. 19.6). We will use the following terminology:

- h (depth of the channel),
- a (amplitude of the solitary wave), $g = 9.81\text{m/s}^2$ (acceleration constant on earth),
- $c_0 = \sqrt{gh}$ (basic velocity of the solitary wave),
- L (typical length of the solitary wave in the experimental situation),
- T (typical time of the observation),
- $\varepsilon = L/h$ (dimensionless parameter).

We assume that ε is sufficiently small, that is, $0 < \varepsilon \ll 1$. This corresponds to *shallow water*. In 1895, Korteweg and de Vries studied this problem. They started with the nonlinear Euler equation together with the nonlinear Bernoulli boundary condition for water waves.[40] Using the approximation for small ε in lowest order, Korteweg and de Vries obtained the nonlinear shallow water approximation

$$\frac{y_t}{c_0} + y_x + \frac{3(y-h)y_x}{2h} + \frac{h^2 y_{xxx}}{6} = 0 \tag{19.39}$$

for the surface $y = y(x,t)$ of the water wave, and they found the following solution

$$\boxed{y(x,t) = h + a \cdot \operatorname{sech}^2\left(\sqrt{\frac{3a}{4h^3}}\left(x - c\left[1 + \frac{a}{2h}\right]\,t\right)\right)} \tag{19.40}$$

of the nonlinear partial differential equation (19.39). This solution describes a travelling solitary wave as pictured in Fig. 19.6 with $y(0,0) = h + a$.[41] Observe that the propagation speed

[40] These equations can be found in E. Zeidler, Nonlinear Functional Analysis, Vol. IV: Applications to Mathematical Physics, Sect. 71.1, Springer, New York, 1995. The point is that the boundary is not known; it has to be determined together with the velocity field of the fluid and the pressure. This is a so-called free boundary-value problem.

[41] Here, we use the notation $\operatorname{sech} x := \frac{1}{\cosh x} = \frac{2}{e^x + e^{-x}}$.

Fig. 19.7. Interaction between two solitary waves

$$c_0\left[1+\frac{a}{2h}\right]$$

is larger than the basic speed parameter $c_0=\sqrt{gh}$. We would like to emphasize again that the solitary wave (19.40) only represents a solution of the approximative problem.[42] The full problem was solved by Lavrentev in 1946 and later by

K. Friedrichs and D. Hyers, The existence of solitary waves, Comm. Pure Appl. Math. **7** (1954), 517–550.

Friedrichs and Hyers used methods from nonlinear functional analysis (i.e., the implicit function theorem). The rigorous solution for the surface of the solitary wave has the form

$$y(x,t)=y_0(x,t)+\varepsilon y_1(x,t)+\varepsilon^2 y_2(x,t)+\dots,$$

with respect to the small parameter ε. Here, the first approximation y_0 is given by (19.40).

The normalized Korteweg–de Vries equation. If we introduce dimensionless quantities by using the rescaling

$$\xi:=\frac{x-c_0t}{L},\qquad \tau:=\frac{t}{T},\qquad U:=1-\frac{y}{h},$$

then we get the standard form

$$\boxed{U_\tau-6UU_\xi+U_{\xi\xi\xi}=0} \tag{19.41}$$

of the Korteweg–de Vries (KDV) equation.

The discovery of the stability of solitons in 1964. By computer experiments, Kruskal and Zabusky found out that two colliding solitary waves are extremely stable, that is, both the shape and the velocity of the waves remain unchanged after the collision (Fig. 19.7).[43]

The discovery of the inverse scattering method in 1967. Gardner, Greene, Kruskal, and Miura made the pioneering discovery that the computation of the solutions of the *nonlinear* Korteweg–de Vries equation (19.41) can be reduced to the solution of an inverse scattering problem for the *linear* Schrödinger equation.[44] Peter Lax noticed quickly that a general functional-analytic principle is behind the inverse spectral method by introducing appropriate 'Lax pairs' $\{S,K\}$ of differential operators K and S. Here,

[42] A detailed derivation of this subtle approximation can be found in Dauxois and Peyrard (2006), quoted on page 1007.

[43] N. Zabusky and M. Kruskal, Interaction of solitons in a collisionless plasma and the recurrence of initial states, Phys. Rev. Lett. **15** (1965), 240–243.

[44] C. Gardner, J. Greene, M. Kruskal, and R. Miura, Method for solving the Korteweg–de Vries equation, Phys. Rev. Lett. **19** (1967), 1095–1097.
C. Gardner, J. Greene, M. Kruskal, and R. Miura, Korteweg–de Vries equations and generalizations: methods for exact solutions, Comm. Pure Appl. Math. **27** (1974), 97–133.

- S is the linear Schrödinger operator, and K is a nonlinear operator chosen in such a way that the Korteweg–de Vries equsation (or another nonlinear evolution equation) can be written in the modified Heisenberg form[45]

$$\dot{S} = SK - KS.$$

Let us sketch the main ideas. We will proceed in the following steps:

- inverse scattering theory and the Gelfand–Levitan–Marchenko integral equation introduced in the 1950s,
- the inverse scattering method, and
- Lax pairs.

Inverse scattering theory. We consider the stationary Schrödinger equation

$$-\psi''(x) + U(x)\psi(x) = k^2\psi(x), \qquad -\infty < x < \infty. \tag{19.42}$$

We assume that the smooth potential $U : \mathbb{R} \to \mathbb{R}$ vanishes sufficiently fast at infinity, that is,

$$\int_{\mathbb{R}} |U(x)|(1 + |x|)\, dx < \infty.$$

We are looking for complex eigenfunctions ψ of (19.42). The spectrum of (19.42) has the following structure:

(i) *Continuous spectrum:* For each nonzero real number k (i.e., $k \in \mathbb{R}^\times$), the number k is a double eigenvalue of (19.42) with two linearly independent eigenfunctions ψ_1 and ψ_2, which are uniquely characterized by the following asymptotic behavior as $x \to -\infty$:

$$\psi_1(x) = \mathrm{e}^{-\mathrm{i}kx} + o(1), \quad \psi_2(x) = \mathrm{e}^{\mathrm{i}kx} + o(1).$$

Additionally, as $x \to +\infty$, we obtain the following asymptotic behavior:

$$\begin{aligned} \psi_1(x) &= a(k)\mathrm{e}^{-\mathrm{i}kx} + b(k)\mathrm{e}^{\mathrm{i}kx} + o(1), \\ \psi_2(x) &= b(k)^\dagger \mathrm{e}^{-\mathrm{i}kx} + a(k)^\dagger \mathrm{e}^{\mathrm{i}kx} + o(1). \end{aligned}$$

(ii) *Discrete spectrum:* Equation (19.42) has either no negative eigenvalues or a finite number of negative eigenvalues

$$-\infty < k_1^2 < k_2^2 < \cdots < k_N^2 < \infty.$$

All these eigenvalues are simple. Letting $k_j = \mathrm{i}q_j$ with $q_j > 0$, the corresponding eigenfunction $\psi^{[j]}$ is uniquely characterized by the following asymptotic behavior as $x \to -\infty$:

$$\psi^{[j]}(x) = \mathrm{e}^{q_j x} + o(\mathrm{e}^{q_j x}), \qquad j = 1, \dots, N.$$

Additionally, as $x \to +\infty$, we have

$$\psi^{[j]}(x) = c_j \mathrm{e}^{q_j x} + o(\mathrm{e}^{-q_j x}), \qquad j = 1, \dots, N,$$

where c_j is a real number.

[45] P. Lax, Integrals of nonlinear equations of evolution and solitary waves, Comm. Pure and Appl. Math. **21** (1967), 467–490.

In terms of quantum mechanics, (i) and (ii) correspond to scattered particles and to bound states of particles, respectively. We call

$$\boxed{a(k), b(k), q_j, c_j, \qquad k \in \mathbb{R}^\times, \;\; j = 1, \dots, N} \tag{19.43}$$

the spectral data of the potential U. The mapping

$$U \mapsto (a(k), b(k), q_j, c_j)$$

from the potential U to the spectral data is called the spectral transform (generated by the one-dimensional Schrödinger equation). We call this a nonlinear Fourier transform.

Let the spectral data (19.43) be given. The main task of *inverse scattering theory* consists in constructing the potential U. This can be done as follows. We introduce the kernel function

$$\mathsf{K}(\mathsf{x}) := \sum_{j=1}^{N} \frac{c_j \mathrm{e}^{-q_j x}}{\mathrm{i} a'(\mathrm{i} q_j)} + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{b(k)}{a(k)} \, \mathrm{e}^{\mathrm{i} k x} dk, \qquad x \in \mathbb{R},$$

and we consider the Gelfand–Levitan–Marchenko integral equation

$$W(x, y) + \mathsf{K}(x + y) + \int_x^\infty W(x, z) \mathsf{K}(z, y) \, dz = 0, \qquad x, y \in \mathbb{R}.$$

If we know a solution W of this linear integral equation, then we obtain the unknown potential U by the relation

$$U(x) = -2 \frac{d}{dx} W(x, x), \qquad x \in \mathbb{R}.$$

The proof can be found in F. Berezin and M. Shubin, The Schrödinger Equation, Kluwer, Dordrecht, 1991.

Inverse scattering method (nonlinear Fourier transform). We want to solve the initial-value problem for the Korteweg–de Vries equation:

$$\boxed{U_t - 6UU_x + U_{xxx} = 0, \;\; x \in \mathbb{R}, \;\; t > 0, \quad U(x, 0) = U_0(x).} \tag{19.44}$$

We are given the initial function U_0.[46] The following result due to Gardner, Greene, Kruskal, and Miura is crucial.

Theorem 19.1 *If the potential function $U = U(x, t), x \in \mathbb{R}, t \geq 0$, is a smooth solution of the Korteweg–de Vries equation (19.44) which vanishes sufficiently fast as $|x| \to \infty$, then the spectral transform of U at time $t > 0$ is uniquely determined by the spectral transform of U at the initial time $t = 0$. Explicitly, we have*

$$\begin{aligned} a(k, t) &= a(k, 0), \qquad & b(k, t) &= b(k, 0) e^{8\mathrm{i}k^3 t}, \\ q_j(t) &= q_j(0), \qquad & c_j(t) &= c_j(0) e^{8 q_j(0)^3 t} \end{aligned} \tag{19.45}$$

for all times $t > 0$ and all parameters $k \in \mathbb{R}^\times, j = 1, \dots, N$.

[46] The initial condition is to be understood in the sense of the limit $\lim_{t \to +0} U(x, t) = U_0(x)$ for all $x \in \mathbb{R}$.

For the proof, we refer to S. Novikov et al., Solitons, Consultant Bureau, New York, 1984. The proof is based on the use of the Lax pair $\{S, K\}$ of differential operators to be introduced below. Using this theorem, the solution of the initial-value problem (19.44) can be computed as follows:

- Step 1: We compute the spectral data of the initial function U_0.
- Step 2: We compute the spectral data at time $t > 0$ by (19.45).
- Step 3: By using the Gelfand–Levitan–Marchenko integral equation, we construct $U(.,t)$ at time $t > 0$. This is the desired solution of (19.44).

The Lax pair $\{S, K\}$. Introducing the differential operator $D := \frac{\partial}{\partial x}$, we define the following two differential operators:

- $S(t)\psi := -D^2\psi + U\psi$ (Schrödinger operator),
- $K(t)\psi := 4D^3\psi - 3U \cdot D\psi - D(U\psi)$ (Korteweg–de Vries operator)

Here, the wave function $\psi = \psi(x)$ depends on position x, whereas the potential function $U = U(x,t)$ depends on position x and time t.

We assume that the fixed potential function $U = U(x,t)$ is a solution of the Korteweg–de Vries equation (19.43).

Differentiating the operators S and K with respect to time t, we get the key relation

$$\boxed{\dot{S}(t) = S(t)K(t) - K(t)S(t), \qquad t \geq 0.} \tag{19.46}$$

The proof of Theorem 19.1 can be based on this equation, which is typical for Lax pairs. Let us only sketch one typical argument. Set

$$\mathcal{K}(t) := \mathrm{e}^{\int_0^t K(\tau)d\tau}, \qquad t \geq 0.$$

Defining the operators on suitable dense subsets of the Hilbert space $L_2(\mathbb{R})$, it turns out that the operator $\mathcal{K}(t) : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ is unitary for any $t \geq 0$ with $\mathcal{K}(0) = I$, and the unique solution of the linear differential equation (19.46) is given by

$$S(t) = \mathcal{K}(t)^{-1} S(0) \mathcal{K}(t), \quad t \geq 0.$$

Consequently, the operator $S(t)$ is unitarily equivalent to the operator $S(0)$. In particular, the two operators $S(t)$ and $S(0)$ have the same eigenvalues. In other words, the eigenvalues of the Schrödinger operator $S(t)$ with the time-dependent potential $U = U(x,t)$ are independent of time t.

The reduction of inverse scattering problems to Riemann–Hilbert problems. In the 1970s it was discovered that the inverse scattering method can be reduced to the solution of Riemann–Hilbert problems. Roughly speaking, the Gelfand–Levitan–Marchenko integral equation is replaced by a suitable Riemann–Hilbert problem. This universal modern approach to integrable systems can be found in the following monographs: Novikov et al., Solitons, New York, 1984, Beals et al. (1988), and Ablowitz and Clarkson (1991) (see the hints for further reading on page 1015).

We also refer to C. Klein and O. Richter, Ernst Equation and Riemann Surfaces, Springer, Berlin, 2005. Here, the Ernst equation is obtained as the integrability condition for an overdetermined linear system of ordinary differential equations for some matrix-valued function ψ. The point is that the matrix ψ depends on an additional spectral parameter (e.g., k^2 in the case of the Korteweg–de Vries equation). The main idea is to solve an appropriate Riemann–Hilbert problem for ψ with respect to the spectral parameter. Then the function ψ is used in order to construct the desired solutions of the Ernst equation. These solutions of the Ernst equation allow us to construct solutions of the Einstein equations in general relativity for describing thin discs of rotating galaxies in the universe.

19.4.7 Resonances

The mathematical analysis of standing water waves represents a difficult problem because of the appearance of resonances between an infinite number of eigenmodes of oscillations. This problem was solved in the following two papers by:

> P. Plotnikov and J. Toland, Nash–Moser theory for standing water waves, Arch. Rat. Mech. Anal. **159** (2001), 1–83.
>
> G. Iooss, P. Plotnikov, and J. Toland, Standing waves on an infinitely deep perfect fluid under gravity, Arch. Rat. Mech. Anal. **177**(3) (2005), 367–478.

The authors used the Moser–Nash technique (the hard implicit function theorem – see page 314). We expect that similar phenomena appear for quantum fields, which represent an infinite number of coupled harmonic oscillators, too.

19.4.8 The Role of Integrable Systems in Nature

> Integrable systems are only approximations of realistic systems in nature.
>
> Folklore

Classical integrable systems of ordinary differential equations. Consider the Hamiltonian canonical system

$$\begin{aligned} \dot{p}_k(t) &= -H_{q_k}(q_1(t),\dots,q_n(t),p_1(t),\dots,p_n(t)),\\ \dot{q}_k(t) &= H_{p_k}(q_1(t),\dots,q_n(t),p_1(t),\dots,p_n(t)), \qquad k=1,\dots,n \end{aligned} \tag{19.47}$$

with n degrees of freedom. This system is called *integrable* iff there exists a canonical transformation

$$q_k = A(Q_1,\dots,Q_n,P_1,\dots,P_n), \qquad p_k = B(Q_1,\dots,Q_n,P_1,\dots,P_n) \tag{19.48}$$

with $k=1,\dots,n$ such that the transformed system with respect to the new variables has the following simple normal form:

$$\boxed{\dot{P}_k(t)=0, \qquad \dot{Q}_k(t)=\text{const}=\omega_k, \qquad k=1,\dots,n.}$$

This system has the solution

$$Q_k(t)=\omega_k t+Q_{0k}, \qquad P_k(t)=P_{0k}, \qquad k=1,\dots,n, \quad t\in\mathbb{R}.$$

Here, Q_{0k}, P_{0k} with $k=1,\dots,n$ are real constants. Using the transformation (19.48), we obtain the solution

$$q_k(t)=A(\mathcal{P}(t)), \qquad p_k(t)=B(\mathcal{P}(t)), \qquad k=1,\dots,n, \quad t\in\mathbb{R} \tag{19.49}$$

of the original problem (19.47). Here, we set

$$\mathcal{P}(t) := (\omega_1 t+Q_{01},\dots,\omega_n t+Q_{0n},P_{01},\dots,P_{0n}).$$

We assume that $Q_1,\dots,Q_n$ are angle variables, that is, the functions A and B have the period 2π with respect to each of the variables $Q_1,\dots,Q_n$. The solution (19.49) depends on $2n$ real constants. In other words, this is the general solution of the original problem (19.47).

General integrable systems. For general physical systems in nature with a finite or infinite number of degrees of freedom, we will not give a technical definition of integrability, but only a heuristic definition. The system Σ is called integrable iff the following hold: There exists a transformation such that

- the transformed system Σ' is decomposed into subsystems $\Sigma'_1, \Sigma'_2, \ldots$ which do not interact with each other, and
- the dynamics of the subsystems $\Sigma'_1, \Sigma'_2, \ldots$ is fairly simple and explicitly known.

The experience of physicists shows that realistic systems in nature are never integrable, since there always exist interactions between the subsystems. As typical examples, let us consider the following two prototypes:

(i) Our solar system: Consider the motion of n celestial bodies (e.g., planets and asteroids) around the sun. If we switch off the gravitational forces between the celestial bodies, then we obtain the superposition of Kepler motions of the celestial bodies. This is an integrable system. If we switch on the small gravitational forces between the celestial bodies, then the system looses its integrability. It is possible that resonances appear between the celestial bodies which cause chaotic motions. In fact, chaotic motions of asteroids have been observed by astronomers.

(ii) Quantum electrodynamics: The system of the free fields of electrons, positrons, and photons is integrable. If we switch on the interaction between the free fields, then the system looses its integrability. Physicists use the method of renormalization in order to handle the complexities of the quantized electromagnetic interaction.

Let us formulate the following principle as a rule of thumb:

- Integrable systems are merely (reasonable) approximations of interacting systems in nature.
- The interaction may cause resonances which are responsible for highly complicated behavior of the realistic system.

One possibility of handling resonances is the Moser–Nash technique in mathematics. We refer to:

J. Feldman and E. Trubowitz, Renormalization in classical mechanics and many-body quantum field theory, Jerusalem J. d'Analyse Mathématique **52** (1992), 213–247.
Internet: `http://www.math.ubc.ca.~feldman/research.html`

J. Bricmont and A. Kupiainen, Renormalizing of partial differential equations, pp. 83–115. In: V. Rivasseau (Ed.), Constructive Physics, Springer, Berlin, 1995.

J. Bricmont, K. Gawędzki, and A. Kupiainen, Kolmogorov–Arnold–Moser (KAM) theorem and quantum field theory, Commun. Math. Phys. **201**(3) (1999), 699–727.

Hints for Further Reading

Ordinary differential equations of Fuchsian type:

E. Coddington and N. Levinson, The Theory of Ordinary Differential Equations, McGraw-Hill, New York, 1955.

P. Hartmann, Ordinary Differential Equations, Wiley, New York, 1964.

V. Smirnov, A Course of Higher Mathematics, Vols. I–V, Pergamon Press, New York, 1964 (applications to special functions can be found in Vol. III).

The six nonlinear Painlevé equations and the Painlevé transcendences:

E. Ince, Ordinary Differential Equations, Dover, New York, 1956.

K. Iwasaki et al., From Gauss to Painlevé: A Modern Theory of Special Functions, Vieweg, Wiesbaden, 1991.

Solitons:

G. Lamb, Elements of Soliton Theory, Wiley, New York, 1980.

R. Dodd et al., Solitons and Nonlinear Wave Equations, Academic Press, New York, 1982.

A. Its and V. Novorshenov, The Isomonodromic Deformation Method in the Theory of the Painlevé Equations, Springer, Berlin, 1986.

L. Faddeev and L. Takhtadzhian, Hamiltonian Method in the Theory of Solitons, Springer, New York, 1987.

R. Beals, P. Deift, and C. Tomei, Direct and Inverse Scattering on the Line, Amer. Math. Soc., Providence, Rhode Island, 1988.

J. Ablowitz and P. Clarkson, Solitons, Nonlinear Evolution Equations and Inverse Scattering, Cambridge University Press, 1991.

V. Zakharov, What is Integrability? Springer, Berlin, 1991.

D. Levi and P. Winternitz (Eds.), Painlevé Transcendents: Their Asymptotics and Physical Applications, Plenum Press, New York, 1992.

V. Korepin, N. Bogoliubov jr., and A. Izergin, Quantum Inverse Scattering Method and Correlation Functions, Cambridge University Press, 1993.

V. Matveev, Algebro-Geometrical Approach to Nonlinear Evolution Equations, Springer, New York, 1994.

R. Donagi et al. (Eds.), Integrable Systems and Quantum Groups, Springer, Berlin, 1993.

B. Dubrovin, Geometry of two-dimensional field theories. In: Donagi et al. (1993), pp. 120–348 (solitons and the foundations of the theory of Frobenius manifolds).

N. Manton and P. Sutcliffe, Topological Solitons, Cambridge University Press, 2004.

The reprints of important articles can be found in:

C. Rebbi and G. Soliani (Eds.), Solitons and Particles, World Scientific, Singapore, 1984.

D. Thouless (Ed.), Topological Quantum Numbers in Non-Relativistic Physics, World Scientific, Singapore, 1998 (collection of 40 articles on superfluidity, quantum Hall effect, phase transitions, and so on).

Nonlinear Riemann–Hilbert problems:

E. Wegert, Nonlinear Boundary Value Problems for Holomorphic Functions and Singular Integral Equations, Akademie-Verlag, Berlin, 1992.

19.5 The BFFO Hopf Superalgebra Approach

Let us summarize the main ideas of the BFFO (Brouder, Fauser, Frabetti, Oeckl) approach to quantum field theory.[47]

[47] C. Brouder, B. Fauser, A. Frabetti, and R. Oeckl, Quantum field theory and Hopf algebra cohomology, J. Phys. A: Mathematical and General **37** (2004),

We exhibit a Hopf superalgebra structure of the algebra of field operators of quantum field theory (QFT) with the normal product. Based on this we construct the operator product and the time-ordered product as a *twist deformation* in the sense of Drinfeld.[48]

The purpose of this paper is to present a new approach to the algebraic and combinatorial structures at the heart of quantum field theory (QFT). This approach has merits on the practical as well on the conceptual side. On the practical side, it allows for a major computational enhancement based on an efficient description of the combinatorics and on non-recursive closed formulae. On the conceptual side, it gives new insights into the algebraic structure of the QFT. We evidence this through applications to *non-perturbative* QFT and non-trivial vacua.

The starting point is the identification of a Hopf algebraic structure at the core of QFT. That is, the algebra of field operators with the

normal product

is a *Hopf superalgebra*. This means that besides the product there is a *coproduct* that describes, intuitively speaking, the different ways in which a product of field operators might be partitioned into two sets. Indeed it is this coproduct that plays the key role in a closed description of combinatorial structures and that allows for computationally efficient algorithms. Another key structure of the Hopf superalgebra is the *counit*. This turns out to describe the standard

vacuum expectation value.

Algebraically, the Hopf superalgebra is the graded symmetric Hopf algebra. The conceptional origin of this is rather simple. Identifying the normal ordered products with functionals on field configurations, the coproduct is induced by the linear addition of fields.

The second main step consists in identifying the standard canonical quantization with a *twist* in the sense of Drinfeld. More precisely,

the operator product

emerges as a twist deformation of the normal product. As is common we deal here at first with the free QFT. The twist is induced by a Laplace pairing which in turn is determined by a suitable propagator. Furthermore, the

time-ordered product

can be obtained similarly as a direct twist deformation of the normal product. In this case, the Laplace pairing is determined by the Feynman propagator. Since vacuum expectation values of time-ordered products are the main ingredients of physical scattering amplitudes, this allows the use of our methods in actual calculations of physical quantities...

It is one of the basic facts in quantum field theory that Wick's theorem relates normal and time-ordered correlation functions. It was only recently noted by Fauser that this transformation can be advantageously described in Hopf algebraic terms.[49]... While the twisted products described so far

5895–5927. Reprinted by permission of IOP Publishing, Bristol, UK. Internet: `www.iop.org/journals/jphysa`

[48] V. Drinfeld, Quasi-Hopf algebras, Leningrad Math. J. **1** (1990), 1419–1457.

[49] B. Fauser, On the Hopf algebraic origin of Wick normal-ordering, J. Phys. A: Math. Gen. **34** (2001), 105–116.

> are the products of the free quantum field theory, our framework is naturally compatible with the usual perturbation theory and thus applicable to it. This implies that the computational advantages directly apply to perturbative quantum field theory.
>
> The third step consists in exploiting the *Hopf algebra cohomology* due to Sweedler, which underlies the twisted product.[50] Besides affording conceptual insight this yields immediate practical benefits. Among these is the realization of the time-ordering prescription of the QFT as an algebra isomorphism. This in turn can be used on the computational side...
>
> *A quantum field theory is free iff the 2-cocycle is a Laplace pairing.*
>
> A further application of the cohomology that we develop is to non-trivial vacua. We show that changing the choice of the vacuum can also be encoded through a twist. Indeed, it turns out that there is a 'duality' or correspondence between the choice of vacuum and that of product. We exemplify this result by solving a problem posed by Kutzelnigg and Mukherjee regarding 'adapted normal products' in quantum chemistry.[51] While they were able to give only examples for low orders, our framework yields closed formulas for all orders. Our method is capable of describing condensates, too.[52]...
>
> A twist in the sense used here is automatically an (equivariant) deformation quantization. Indeed, this was one of the original motivations for Drinfeld to introduce this concept. This means that our approach is thus inherently connected to the deformation quantization approach to quantum field theory. This approach starts also with the normal ordered product and views the other products as deformations.[53]

The following primer in quantum field theory shows that the basic concepts in quantum field theory are intimately related to Hopf algebras. We will thoroughly study this in Vol. IV on quantum mathematics.

> This paper provides a primer in quantum field theory (QFT) based on Hopf algebra and describes new Hopf algebraic constructions inspired by QFT concepts. The following concepts are introduced: time-ordered products, S-matrix, Feynman diagrams, connected diagrams, Green functions, renormalization. The use of Hopf algebra for their definition allows for simple recursive derivations and leads to a correspondence between Feynman diagrams and semi-standard Young-tableaux.[54]
>
> Christian Brouder, 2006

B. Fauser, A treatise on quantum Clifford algebras, postdoctoral thesis, University of Konstanz (Germany), 2002.
Internet: `http://arxiv.org/math.QA/0202059`

[50] M. Sweedler, Cohomology of algebras over Hopf algebras, Trans. Amer. Math. Soc. **133** (1968), 204–239.

[51] W. Kutzelnigg and D. Mukherjee, Normal order and extended Wick theorem for a multiconfiguration reference wave function, J. Chem. Phys. **107** (1997), 432–449.

[52] B. Fauser, Clifford geometric parametrization of inequivalent vacua, Math. Methods Appl. Sci. **24** (2001), 885–912.

[53] A. Hirshfeld and P. Henselder, Star products and perturbative quantum field theory, Annals of Physics **298** (2002), 382–393.

[54] C. Brouder, Quantum field theory meets Hopf algebra: A Primer, 2006.
Internet: `http:// arxiv:hep-th/0611153`

19.6 The BRST Approach and Algebraic Renormalization

The Standard Model of electroweak interactions has been tested to high accuracy with the precision experiments at the Z-resonance at LEP (Large Electron-Positron Collider at CERN, Geneva, Switzerland). The degree of precision enforces to take into account also contributions beyond the tree approximation in the perturbative renormalization. For this reason, an extensive calculation of 1-loop processes and also 2-loop processes has been carried out in the past years and compared to the experimental results. A careful analysis shows that the theoretical predictions and the experiments are in excellent agreement with each other.[55]

A necessary prerequisite for carrying out precision tests of the Standard Model is the consistent mathematical and physical formulation of the Standard Model in the framework of its perturbative construction. Explicitly, one has to prove the following properties in order to bring it into the predictive power, which the Standard Model is expected to have:

- The (renormalized) Green's functions of the theory are *uniquely* determined as functions of a *finite* number of free parameters to all orders of perturbation theory. This property is called renormalizability.
- The physical scattering matrix constructed from the Green's functions is unitary and gauge parameter independent. In particular, these properties ensure a probability interpretation of S-matrix elements and guarantee at the same time that unphysical particles (i.e., ghosts) are cancelled in physical scattering processes. Only then the theory has indeed a physical interpretation.
- It has to be shown that the theory is in agreement with the experiments by calculating different processes as accurately as possible.

In the present lecture we only treat the first point, the unique construction of the Green's function to all orders of perturbation theory. We want to point out that the crucial unitarity and gauge parameter independence of the S-matrix are not rigorously derived in the Standard Model by now, but are commonly assumed to hold. However, its analysis includes the important problem of unstable particles, whose solution will have far-reaching consequences in phenomenological applications.

Renormalizability of gauge theories was first shown in the framework of dimensional regularization.[56] Dimensional regularization can be used as an invariant scheme for gauge and BRST (Becchi, Rouet, Stora, Tyutin) invariance, respectively, as long as parity is conserved. In this scheme it has

[55] Reports of the Working Group on precision calculations for the Z-resonance, CERN, Yellow Report, 1995, CERN 95-03. Edited by D. Bardin, W. Hollik, and G. Passarino.
W. Hollik and G. Duckeck, Electroweak Precision Tests at LEP, Springer, Berlin, 2000.

[56] L. Faddeev and V. Popov, Feynman diagrams for the Yang–Mills field, Phys. Lett. **25B** (1967), 29–30.
G. 't Hooft, Renormalizable Lagrangians for massive Yang–Mills fields, Nucl. Phys. **B35** (1971), 167–188.
G. 't Hooft and M. Veltman, Regularization and renormalization of gauge fields, Nucl. Phys. **B44** (1972), 189–213 (dimensional regularization).
G. 't Hooft and M. Veltman, Combinatorics of gauge fields, Nucl. Phys. **B50** (1972), 318–353 (the S-matrix is independent of the choice of the gauge).

> been proven that all the divergencies can be absorbed into gauge-invariant counterterms to the coupling, the field redefinitions and the masses of the classical action. This method implies the unique construction of the (renormalized) Green's functions. These proofs are not applicable to the Standard Model, since there *parity is broken*. It is also well-known that the group structure of the Standard Model allows the presence of anomalies.[57]
>
> *The algebraic method of renormalization provides a proof of renormalizability also in such cases where a gauge-invariant regularization scheme does not exist.*
>
> In a scheme-independent way, the algebraic method of renormalization gives the symmetry relations of finite Green's functions to all orders...
> Necessary prerequisite for the algebraic method to work was the discovery of the BRST symmetry. In its functional form BRST symmetry is called the Slavnov–Taylor identity. This identity is the defining symmetry of gauge theories and Lorentz invariant gauges and includes the gauge-fixing action and the action of the Faddeev–Popov ghosts.[58]
>
> Elisabeth Kraus, 1997

Let us sketch the main philosophy of algebraic renormalization theory in very rough terms. We will thoroughly study this in Vol. IV on quantum mathematics. Let us restrict ourselves to a scalar interacting quantum field φ.

(i) Correlation functions: We start with the correlation functions

$$C_n = C_n(x_1, \dots, x_n)$$

as fundamental quantities. Intuitively, they describe the correlations of the fluctuations of the quantum field with respect to different space-time points. The correlation functions are also called the Green's functions of the quantum field. Define the generating functional by setting

$$Z(J) = \sum_{n=0}^{\infty} \frac{(-1)^n}{\hbar^n n!} \int C(x_1, x_2, \dots, x_n) \prod_{k=1}^{n} J(x_k) \; dx_1 \cdots dx_n.$$

The correlation functions can be computed by means of the Gell-Mann–Low formula which relates C_n to the correlation functions of the free quantum field. Alternatively, one can describe $Z(J)$ by a Feynman functional integral which depends on the classic action. The computation formulas for correlation functions can be depicted by Feynman graphs.
Our next goal is to simplify the computation of correlation functions by showing that one has only to compute special correlation functions called vertex functions. First let us restrict to correlation functions $C_{n,c}$ which correspond to connected graphs. This yields

$$Z_c(J) = \sum_{n=0}^{\infty} \frac{(-1)^n}{\hbar^n n!} \int C_{n,c}(x_1, x_2, \dots, x_n) \prod_{k=1}^{n} J(x_k) \, dx_1 dx_2 \cdots dx_n.$$

It turns out that $Z(J) = \mathrm{e}^{-\mathrm{i}Z_c(J)/\hbar}$.

---

[57] Anomalies are symmetries of the classical field theory which do not survive the procedure of quantization.

[58] E. Kraus, Renormalization of the electroweak standard model, Lectures given at the Saalburg summer school (Germany) in 1997.
Internet: `http://ariv.org/hep-th/9809069`.
E. Kraus, Renormalization of the electroweak standard model to all orders, Annals of Physics **262** (1998), 155–259.

(ii) Vertex functions as a basis of correlation functions: Recall that a connected graph is called one-particle irreducible iff it remains connected after cutting any internal line. The correlation functions corresponding to one-particle irreducible graphs are called vertex functions and denoted by V_n. The functional

$$V(J) = \sum_{n=0}^{\infty} \frac{(-1)^n}{\hbar^n n!} \int_{\mathbb{R}^{4n}} V_n(x_1, x_2, \ldots, x_n) \prod_{k=1}^{n} J(x_k)\, dx_1 \cdots dx_n$$

is called the vertex functional. It turns out that V is related to Z_c by a Legendre transformation (see Sect. 7.24.12 of Vol. I). Furthermore, we have the power series expansion

$$V(\varphi) = S(\varphi) + \sum_{n=1}^{\infty} \hbar^n V_n(\varphi)$$

in terms of the Planck constant $\hbar$. Here, φ is a classical field, which satisfies the Euler–Lagrange equation, and $S(\varphi)$ is the classical action of the field φ. We call $V(\varphi)$ the effective quantum action of the classical φ. Intuitively, $V(\varphi)$ adds quantum fluctuations to the classical action. The reader should have the following general principle in mind:

The crucial vertex functional V knows all about the quantum field.

In particular, if we know the vertex functional V, then we know all the correlation functions, and all the S-matrix elements (see Vol. IV).[59]

(iv) The fundamental quantum action principle: In physics, one always formulates basic equations for the fundamental quantities. In the present approach to quantum field theory, the fundamental equation is the quantum action principle which represents an equation for the functional derivative $\frac{\delta V(J)}{\delta J}$ of the vertex functional V with respect to the source function J (or other external parameters).

Intuitively, the quantum action principle tells us the response of quantum fluctuations to external influences.

This principle generalizes the classical Dyson–Schwinger equation.[60]

(v) Symmetries: If the classical field theory possesses symmetries (e.g., gauge symmetries), then there exist symmetry relations between the correlation functions, which we call general Ward identities.[61] Roughly speaking, these identities follow from invariance properties of the effective quantum action under infinitesimal BRST (Becchi, Rouet, Stora, Tyutin) transformations.

(vi) Renormalization: In order to renormalize the theory, one critically uses the general Ward identities. The idea is to construct an iterative method in order to renormalize the vertex functions in such a way that the general Ward identities

[59] The effective quantum action was introduced by J. Goldstone, A. Salam, and S. Weinberg, Broken symmetry, Phys. Rev. **127** (1962), 965–970.
The fact that the vertex functional V is the Legendre transformation of Z_c was discovered by G. Jona-Lasinio, Relativistic field theories with symmetry-breaking solutions, Nuovo Cimento **34** (1964), 1790–1795.

[60] The quantum action principle was formulated by:
J. Lowenstein, Differential vertex operations in Lagrangian field theory, Commun. Math. Phys. **24** (1971), 1–21.
Y. Lam, Perturbation Lagrangian theory for scalar fields: Ward–Takahashi identity and current algebra, Phys. Rev. **D6** (1972), 2145–2161.

[61] This comprehends the classical Ward identities in quantum electrodynamics and the Slavnov–Taylor identities in general gauge theories.

remain valid. This is a purely algebraic procedure which is closely related to cohomology.

(vii) Renormalization group: The renormalization procedure depends on normalization conditions. The change of the normalization constants is governed by the renormalization group. The change of the constants leads to a change of the renormalized vertex functional. The quantum action principle tells us that this change can be described by an equation which is called the renormalization group equation for the vertex functional.

(viii) Anomalies: It is possible that some symmetry of the classical field theory does not survive the passage from the classical action to the effective quantum action. This is called an anomaly of the quantization procedure. Here, topological invariants play a crucial role.

(ix) Energies of bound states: Considering models in quantum mechanics, Heisenberg conjectured that the S-matrix knows all about both scattering processes and bound states. In particular, the energies of bound states correspond to singularities of the S-matrix in the energy space. This heuristic principle is very useful for computing bound states of elementary particles. However, the reader should note that this principle may fail. This was shown by R. Jost, On the false zeros of the eigenvalues of the S-matrix, Helv. Phys. Acta **20** (1947), 256–266 (in German).

We recommend the following references:

O. Piguet and S. Sorella, Algebraic Renormalization: Perturbative Renormalization, Symmetries, and Anomalies, Springer, Berlin, 1995.

V. Parameswaran Nair, Quantum Field Theory: A Modern Perspective, Springer, New York, 2005 (Green's functions, the iterative construction of the S-matrix by using vertex functions, anomalies).

O. Piguet, Rénormalisation en théorie quantique des champs (Renormalization in quantum field theory), Lecture Notes I, Department of Theoretical Physics, University of Geneva, Switzerland, 1982 (in French).

O. Piguet, Rénormalisation des théories de jauge (Renormalization in gauge theory), Lecture Notes II, Department of Theoretical Physics, University of Geneva, Switzerland, 1983 (in French) (algebraic renormalization of quantum electrodynamics in terms of Ward identities).

A. Barut, The Theory of the Scattering Matrix, MacMillan, New York, 1967.

19.7 Analytic Renormalization and Distribution-Valued Analytic Functions

The ideas of analytic renormalization go back to Marcel Riesz' generalization[62] of the Riemann–Liouville definition of fractional differentiation to treat the relativistic wave equation.[63] Riesz proposed to his physics colleagues that they try to use his technique to obtain finite answers in the quantum theory of fields. After some initial successes, it was found that

[62] Marcel Riesz (1886–1969).

[63] M. Riesz, L'intégrale de Riemann–Liouville et le problème de Cauchy, Acta Math. **81** (1948), 1–223.

> the method did not work.[64] It was pointed out by G. Källén that if instead of trying to use the method to calculate the values of quantities like self-energies one calculates *Green's functions,* the Riesz method should yield a convenient method of doing renormalization calculations... A version of the method for general graphs was worked out by E. Speer.[65]
>
> Arthur Wightman, 1976[66]

The basic ideas of analytic regularization are discussed in Sect. 2.2.11ff on page 77ff.

19.8 Computational Strategies

19.8.1 The Renormalization Group

> It was correctly realized by Wilson (born 1935) and his followers that in a quantum theory with many scales involved, the change of parameters from bare to renormalized values is a phenomenon too complex to be described in a single step.
> Just like the trajectory of a complicated dynamical system, it must be instead studied step by step through a local evolution rule. The change of scale in the renormalization group plays the role of time in dynamical systems. This analogy is deep. There is a natural arrow of time, related to the second principle of thermodynamics, and there is similarly a natural arrow for the renormalization group evolution: microscopic laws are expected to determine macroscopic laws, not the converse. The renormalization group erases unnecessary detailed short scale information...
> If we consider the universal character of the action principle both at the classical and quantum level, and observe that the relation between microscopic and macroscopic laws is perhaps the most central of all physical questions, it is probably not an exaggeration to conclude that the renormalization group is in some deep sense the "soul" of physics.[67]
>
> Vincent Rivasseau, 2002

> About twenty years ago Wilson and his collaborators published their ideas on the renormalization group and effective Lagrangians,[68] which have stimulated quantum field theory and statistical mechanics ever since. In 1984

[64] S. Nilson, Interaction of electrons and an electromagnetic field treated by analytic continuation, Arkiv Fysik **1** (1950), 369–423.

[65] E. Speer, Generalized Feynman Amplitudes, Princeton University Press, 1969.
E. Speer, Dimensional and analytic renormalization. In: G. Velo and A. Wightman (Eds.), Renormalization Theory, Reidel, Dordrecht, 1976, pp. 25–94.

[66] A. Wightman, Orientation, pp. 1–24 . In: G. Velo and A. Wightman (Eds.), Renormalization Theory, Reidel, Dordrecht, 1976 (reprinted with permission).

[67] V. Rivasseau, An introduction to renormalization. In: B. Duplantier and V. Rivasseau (Eds.), Poincaré Seminar 2002: Vacuum Energy – Renormalization. Birkhäuser, Basel, 2003, pp. 139–177 (reprinted with permission).

[68] K. Wilson and and J. Kogut, The renormalization group and the ε-expansion, Phys. Rep. **12C** (1974), 75–199.
J. Polchinski, Renormalization and effective Lagrangians, Nucl. Phys. **B231** (1984), 269–295.

> Polchinski showed that these ideas are suited for a treatment of the renormalization problem of perturbative field theory which does not make any use of Feynman diagrams and in particular sidesteps the complicated analysis of the divergence/convergence properties of the general bare or renormalized Feynman diagram. Instead he showed that the problem can be solved by bounding the solutions of a system of first order partial differential equations, the *flow equations*, which are a reduction of the Wilson flow equations to their perturbative content...
> We prove the perturbative renormalizability of the Euclidean quantum electrodynamics in four dimensions using flow equations. As compared to the φ^4-model, the additional difficulty to overcome is that the regularization violates gauge invariance. We prove that there exists a class of renormalization conditions such that the renormalized Green's functions satisfy the Ward identities and such that they are infrared finite at nonexceptional momenta. We give bounds on the singular behavior at exceptional momenta (due to the massless photon).[69]
>
> Georg Keller and Christoph Kopper, 1996

The original basic idea of renormalization group theory is the fact that the classical renormalization procedure leads to additional degrees of freedom which have to be fixed in a concrete calculation of a physical process (e.g., see Sect. 2.1.1 on page 48). Naturally enough, we expect that the physics is independent of this fixing. The renormalization group describes the change of fixings. Indeed, it turns out that the physics is invariant under the renormalization group. This allows us to extrapolate the behavior of physical processes from low to high energies (see Chap. 3 of Vol. I on scale changing). As a rule, one considers two equations:

(CS) the Callan–Szymanzik equation (change of scale), and
(RG) the renormalization group equation (e.g., change of the normalization energy in the renormalization procedure).

In particular, equation (RG) tells us the dependence, $\kappa = \kappa(E)$, of the coupling constant on the energy of the physical experiment under consideration (running coupling constant – see Sect. 3.2.2 of Vol. I). We will study this in Vol. IV on quantum mathematics.

19.8.2 Operator Product Expansions

> The Wilson operator product expansion (OPE) states that a product of n local quantum fields can be expanded at *short distances* as an asymptotic series, each term of which is given by a model dependent coefficient function $C^k_{i_1\dots i_n}$ of the n space-time arguments, $x_1, \dots, x_n$, times a local field at a nearby reference point y:

$$A_{i_1}(x_1)A_{i_2}(x_2)\cdots A_{i_n}(x_n) \sim \sum_k C^k_{i_1 i_2 \dots i_n}(x_1, x_2, \dots, x_n)A_k(y)$$

[69] G. Keller and C. Kopper, Renormalizability proof of quantum electrodynamics based on flow equations, Commun. Math. Phys. **176** (1996), 193–226 (reprinted with permission).

> as $x_1, \ldots, x_n \to y$.[70] This expansion has been established in perturbative quantum field theory on Minkowski space-time[71] and is by now a standard tool, for example, in the analysis of quantum gauge theories such as quantum chromodynamics...
> We present an algorithm for constructing the operator product expansion for perturbative interacting quantum field theory in general Lorentzian curved space-time, to arbitrary orders in perturbation theory. The remainder in this expansion is shown to go to zero at short distances in the sense of expectation values in arbitray Hadamard states.[72]
>
> Stefan Hollands, 2007

Singularities of operator products. The point is that the operator product

$$A_1(x_1)A_2(x_2)\cdots A_n(x_n)$$

of interacting quantum field becomes singular if two arguments coincide (e.g., $x_1 = x_2$). Intuitively, these singularities describe the short-distance behavior of the correlations corresponding to the fluctuations of the interacting quantum fields under consideration. In order to regularize (and to renormalize) local operator products, one has to understand the structure of the singular limits when two arguments approach each other (e.g., $x_1 \to x_2$). Hollands studies systematically such limits by applying a recent result from algebraic geometry. This concerns the following paper:

> W. Fulton and R. MacPherson, A compactification of configuration spaces, Ann. of Math. **139** (1994), 183–225.

Let us briefly discuss the basic idea behind the sophisticated mathematical approach.

One-point compactification of the complex plane. The complex plane $\mathbb{C}$ is not a compact topological space. However, adding the point ∞ to the complex plane, the extended complex plane

$$\overline{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$$

can be equipped with the topology of a compact topological space. To this end, we use the stereographic projection

$$\chi : \mathbb{P}^1_{\mathbb{C}} \setminus N \to \mathbb{C}$$

from the *Riemann sphere minus North Pole* onto the complex plane (Fig. 5.17 on page 308). Adding the North Pole N to the punctured sphere, we obtain the compact Riemann sphere $\mathbb{P}^1_{\mathbb{C}}$ which is in one-to-one correspondence to the extended complex plane $\overline{\mathbb{C}}$. Here, the North Pole N corresponds to the point ∞. Finally, a subset of $\overline{\mathbb{C}}$ is called open iff the corresponding set on the Riemann sphere is open. This way, the extended plane $\overline{\mathbb{C}}$ becomes a compact topological space.

---

[70] K. Wilson, On products of field operators at short distances, Cornell Report 1964.
K. Wilson, Non-Lagrangian models of current algebra, Phys. Rev. **179** (1969), 1499–1512.

[71] W. Zimmermann, Normal products and the short-distance expansion in the perturbation theory of renormalizable interactions, Annals of Physics **77** (1973), 536–601.

[72] S. Hollands, The operator product expansion for perturbative quantum field theory in curved space-time, Commun. Math. Phys. **273** (2007), 1–36 (reprinted with permission).

n-point compactification. Let X denote a nonsingular algebraic variety (e.g., the complex plane). Let us label n distinct points $z_1, \ldots, z_n$ of X. As usual, let X^n denote the n-fold Cartesian product $X \times \cdots \times X$. Finally, introduce the set

$$\Delta_{jk} := \{(x_1, x_2, \ldots, x_n) \in X^n : \ x_j = x_k\}$$

for fixed indices j and k with $j \neq k$ (e.g., $j = 1$ and $k = 2$). The paper by Fulton and MacPherson mentioned above studies all the possible compactifications of the set

$$X^n \setminus \bigcup_{j \neq k} \Delta_{jk}.$$

The coefficient functions of an operator product expansion as fundamental quantities. Passing from flat space-time manifolds to curved space-time manifolds $\mathcal{M}$, the following two serious difficulties arise:

- the lack of space-time translations on $\mathcal{M}$, and
- the lack of vacuum states (in contrast to the Minkowski space-time manifold).

To overcome these difficulties, in the paper quoted above, Hollands makes the proposal that quantum field theories on curved space-time manifolds should be based on axioms for the coefficient functions $C^k_{i_1 \ldots i_n}$ of operator product expansions.

19.8.3 Binary Planar Graphs and the Renormalization of Quantum Electrodynamics

> Quantum electrodynamics was renormalized to all orders by Dyson in 1949.[73]... It is standard to define free, bare, and renormalized propagators.
> - The free electron propagator is the 2-point Green's function for an electron without electromagnetic interaction.
> - The bare electron propagator is the 2-point Green's function (or the 2-point correlation function) for an electron with electromagnetic interaction, but without renormalization. In the perturbation expansion of the bare electron propagator, all terms (except the first one) are infinite.
> - The renormalized electron propagator is the 2-point Green's function for an electron with electromagnetic interaction, after renormalization.
>
> Similarly, we define the free, bare, and renormalized photon propagator.
> When practitioners of quantum electrodynamics calculate multi-loop contributions to renormalized propagators, they are often struck by the many *cancellations* in the calculation. These cancellations are partly due to the existence of the Ward identity, which provides a relation between self-energy and vertex counterterms. In this paper, we derive recursive equations for the renormalized electron and photon propagators that take full account of the Ward identity... In this paper, the Dyson relations between renormalized and bare photon and electron propagators are expanded over planar binary trees. This yields explicit recursive relations for the terms of the expansions. When all the trees corresponding to a given power of the expansion parameter (i.e., the fine structure constant) are summed, recursive relations are obtained for the *finite* coefficients of the renormalized electron and photon propagators. These relations significantly decrease the

[73] F. Dyson, The S-matrix in quantum electrodynamics, Phys. Rev. **75** (1949), 1736–1755.

number of integrals to carry out, as compared to the standard Feynman diagram technique.

Christian Brouder and Alessandra Frabetti, 2001

This quotation is taken from C. Brouder and A. Frabetti, Renormalization of QED with planar binary trees, Eur. Phys. J. **C19** (2001), 715–741 (reprinted with permission).

19.9 The Master Ward Identity

The following crucial question is not yet satisfactorily answered for quantum field theories: *Can the symmetries of the underlying classical theory be maintained in the process of renormalization?* The difficulties are connected with the singular character of quantized fields which forbids a straightforward transfer of the arguments valid for the classical theory.
In this paper we attack this question in a model independent way:
The Master Ward Identity (MWI) is a universal formulation of all symmetries which follow in classical field theory from the field equations. It is formulated in such a way that

- it is well defined in perturbative quantum field theory, and
- we impose it there as a renormalization condition.

If this condition is fulfilled, the validity of e.g. the following symmetries is maintained in quantization:

- the field equations,
- conservation of the energy-momentum tensor,
- charge conservation, and
- ghost number conservation (in case of a non-Abelian gauge theory).

The MWI implies also a rigorous substitute for equal-time commutation relations of quark currents, and the master BRST identity. The latter is a model independent equation which expresses the BRST symmetry of the time-ordered products. Applied to a classical BRST invariant (gauge) model, it implies the full BRST symmetry of the corresponding perturbative quantum field theory... The MWI can nearly always be satisfied... The only counter-examples we know are the usual, well-known anomalies of perturbative quantum field theory.

Michael Dütsch, 2002

We refer to the lecture given by M. Dütsch, The Master Ward Identity: a universal formulation of classical symmetries. Can they be realized in perturbative quantum field theory? 17 pages. In: F. Scheck (Ed.), Theory of Renormalization and Regularization. Lecture Notes, 2002.
Internet: `http://www.thep.physik.uni.mainz.de~scheck/Hessbg`
See also the hints (XII) for further reading on page 1036.

19.10 Trouble in Quantum Electrodynamics

19.10.1 The Landau Inconsistency Problem in Quantum Electrodynamics

In 1954 Lev Landau noted that, by calculating the renormalized electron charge using a plausible partial summation of Feynman graphs, one arrives

> at the conclusion that it vanishes in the limit of infinite cut-off.[74] He concluded that this absurd result invalidates quantum field theory which should be "buried with due honors."[75]
>
> Kerson Huang, 1998

> The investigations of the high-energy behavior of quantum electrodynamics by Källén, Landau, and especially that by Gell-Mann and Low[76] showed that the perturbative approach in quantum electrodynamics unavoidably breaks down, ironically, as a necessity of charge renormalization. Landau and his collaborators argued further that remaining within the perturbative framework would lead either to no interaction (zero renormalized electron charge) or to the occurrence of ghost states rendering the theory apparently inconsistent...
>
> The attitude of theoretical physicists toward the issue of consistency differed sharply. For most practicing physicists, consistency is just a pedantic problem. As pragmatists, they are guided only by their scientific experience and have little interest in speculating about the ultimate consistency of a theory. For Dirac, however, renormalization theory with the cut-off going to infinity was illogical and nonsensically physically.[77] In his opinion, what was required were new forms of interactions and new mathematics such as possibly the use of an indefinite metric, or of non-associative algebra, or perhaps something even more esoteric.
>
> The positions adapted by Landau and Chew were more radical and drastic. What they rejected were not merely particular forms of interactions and perturbative versions of quantum field theory, but the general framework of quantum field theory itself. For them the very concept of a local field operator and the postulation of any detailed mechanism for interactions in microscopic space-time region were unacceptable because these were too speculative to be observable, even in principle. Their position was supported by the presence of divergences in quantum field theory and by the lack of a proof of the consistency of renormalization theory, even though Landau's argument for the inconsistency of renormalized quantum electrodynamics could not claim to be conclusive.[78]
>
> Silvan Schweber, 1994

In terms of a simplified mathematical toy model, it is discussed in the monograph by V. Rivasseau, From Perturbative to Constructive Renormalization, Princeton

[74] L. Landau, On the quantum theory of fields, pp. 52–69. In: W. Pauli (Ed.), Niels Bohr and the Development of Physics, Pergamon Press, New York, 1955. The textbook version can be found in L. Landau and M. Lifshitz, Course of Theoretical Physics, Vol 4: Quantum Electrodynamics, Sect. 128. Butterworth–Heinemann, Oxford, 1982.
Landau (1908–1968) worked in Moscow. In 1962 he was awarded the Nobel prize in physics for his pioneering theories for condensed matter, especially liquid helium.

[75] K. Huang, Quantum Field Theory: From Operators to Path Integrals, Wiley, New York, 1998.

[76] M. Gell-Mann and F. Low, Quantum electrodynamics at small distances, Rev. Phys. **95**(5) (1954), 1300–1317 (see Sect. 3.2.2 of Vol. I).

[77] P. Dirac, The origin of quantum field theory, pp. 39–55. In: L. Brown and L. Hoddeson, The Birth of Particle Physics, Cambridge University, 1983.

[78] Schweber, Silvan, QED and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga, Princeton University Press, Princeton, New Jersey, 1994. Reprinted by permission of Princeton University Press.

University Press, 1991, that the Landau inconsistency problem is a global problem of the renormalization procedure called the "renormalon problem." Roughly speaking, in each order n of perturbation theory, the expressions behave consistently. The inconsistency only appears by studying the limit $n \to \infty$.

19.10.2 The Lack of Asymptotic Freedom in Quantum Electrodynamics

A quantum field theory is called asymptotically free iff the high-energy limit $E \to +\infty$ corresponds to free particles. For example, in quantum chromodynamics, quarks move freely if the energy goes to infinity. Unfortunately, quantum electrodynamics is not asymptotically free (see Sect. 3.2.3 of Vol. I). Many physicists expect that the renormalization of quantum electrodynamics to all orders leads to a trivial theory with vanishing coupling constant. If this was true, then we would have the following strange situation in quantum electrodynamics:

- Low-order perturbation theory yields nontrivial results which are in fantastic coincidence with the physical experiment.
- High-order perturbation theory yields a trivial result.

This means that quantum electrodynamics represents an incomplete (and possibly ill-posed) theory (see Sect. 2.4.4). One has to study the Standard Model in particle physics which possibly completes quantum electrodynamics.

> *The renormalization of quantum electrodynamics to all orders is not the end of the story.*

19.11 Hints for Further Reading

As an introduction to renormalization, we recommend the following two monographs:

> L. Brown, Quantum Field Theory, Cambridge University Press, New York, 1996 (generating functionals, Feynman functional integrals, dimensional regularization, applications to the scalar φ^4-model and to quantum electrodynamics).
>
> W. McComb, Renormalization Methods: A Guide for Beginners, Oxford University Press, 2007 (the renormalization group method).

Furthermore, we recommend:

> J. Collins, Renormalization: An Introduction to Renormalization, the Renormalization Group, and the Operator-Product Expansion, Cambridge University Press, 1984 (BPHZ renormalization).
>
> O. Piguet and S. Sorella, Algebraic Renormalization: Perturbative Renormalization, Symmetries, and Anomalies, Springer, Berlin, 1995 (the quantum action principle for the effective action and the BRST symmetry).
>
> T. Kugo, Gauge Field Theory, Springer, Berlin, 1997 (in German).
>
> K. Hepp, Renormalization theory. In: C. DeWitt and R. Stora (Eds.), Statistical Mechanics and Quantum Field Theory, Gordon and Breach, New York, 1971, pp. 429–500 (axiomatic approach to renormalization theory).
>
> L. Faddeev and A. Slavnov, Gauge Fields, Benjamin, Reading, Massachusetts, 1980 (Feynman's functional integrals, Faddeev–Popov ghosts,

Bogoliubov's R-iterative method and the renormalization of gauge theories).

J. Feldman, T. Hurd, L. Rosen, and J. Wright: QED (quantum electrodynamics): A Proof of Renormalizability, Springer, Berlin, 1988.

M. Veltman, Diagrammatica: the Path to Feynman Diagrams, Cambridge University Press, 1995 (dimensional renormalization).

A. Connes and M. Marcolli, Noncommutative Geometry, Quantum Fields, and Motives, Amer. Math. Soc., Providence, Rhode Island, 2008 (BPHZ renormalization, Hopf algebras of Feynman graphs and the Riemann–Hilbert method in renormalization theory, the universal motivic Galois group in renormalization group theory, noncommutative geometry and the Standard Model in particle physics – relations between the constants of the Standard Model – noncommutative geometry and the Riemann zeta function, statistical physics, KMS states, and number theory).
Internet: `http://www.math.fsu.edu/~marcolli/bookjune4.pdf`

We also recommend:

E. Manoukian, Renormalization, Academic Press, New York, 1983 (proof of a general version of Weinberg's power-counting theorem and general principles in renormalization theory; the relation between the Dyson–Salam renormalization and the BPHZ renormalization).

O. Zavialov, Renormalized Quantum Field Theory, Kluwer, Dordrecht, 1989 (BPHZ renormalization, and the equivalence between BPHZ and other methods – dimensional regularization and analytical renormalization).

V. Smirnov, Feynman Integral Calculus, Springer, Berlin, 2006 (computation of algebraic Feynman integrals).

A. Grozin, QED (Quantum Electrodynamics) and QCD (Quantum Chromodynamics): Practical Calculation and Renormalization of One-and Multi-Loop Feynman Diagrams, World Scientific Singapore, 2007.

Standard textbooks in physics are:

J. Bjorken and S. Drell, Relativistic Quantum Fields, McGraw-Hill, New York, 1965.

C. Itzykson and J. Zuber, Quantum Field Theory, MacGraw-Hill, New York, 1980.

L. Landau and E. Lifshitz, Course of Theoretical Physics, Vol 4: Quantum Electrodynamics. Edited by W. Berestetzkii, E. Lifshitz, and L. Pitaevskii, Butterworth–Heinemann, Oxford, 1982 (much material on concrete physical effects).

M. Peskin and D. Schroeder, An Introduction to Quantum Field Theory, Addison-Wesley, Reading, Massachusetts, 1995.

S. Weinberg, Quantum Field Theory, Vols. 1–3, Cambridge University Press, 1995.

S. Weinberg, Cosmology, Oxford University Press, 2008.

J. Zinn-Justin, Quantum Field Theory and Critical Phenomena, Clarendon Press, Oxford, 2003.

S. Narison, QCD (Quantum Chromodynamics) as a Theory of Hadrons: From Partons to Confinement, Cambridge University Press, 2004.

Furthermore, we recommend the following introductory material:

- Pedagogical lectures on classical approaches:
  G. Velo and A. Wightman (Eds.), Renormalization Theory, Reidel, Dordrecht, 1976.

  E. Speer, Lectures on Analytic Renormalisation, Seminars on Renormalisation Theory Volume I: Technical Report No. 73–067, Center for Theoretical Physics of the Department of Physics and Astronomy, University of Maryland, College Park, 1973.

  J. Lowenstein, Normal Product Methods in Renormalised Perturbation Theory, Seminars on Renormalisation Theory Volume II: Technical Report No. 73–068, Center for Theoretical Physics of the Department of Physics and Astronomy, University of Maryland, College Park, 1973.

  G. 't Hooft and M. Veltman, Diagrammar, Lecture Notes, CERN, 1973, Report 73/9 (renormalization of gauge theories).
  Internet: `http://doc.cern.ch/yellowrep/1973/1973-009/p1.pdf`
- Pedagogical lectures on modern approaches:

  F. Scheck (Ed.), Theory of Renormalization and Regularization. School held at the Hesselberg Academy (Germany), 2002, Lecture Notes.
  Internet: `http://www.thep.physik.uni.mainz.de~scheck/Hessbg`

  K. Sibold, Perturbative Renormalization – Quantization of Gauge Theories, Lecture Notes MPI-Ph/93-1, Max Planck Institute for Physics, Werner Heisenberg, Munich, 1993 (in German).

  R. Wulkenhaar, Euclidean quantum field theory and commutative and noncommutative spaces, Lecture Notes, pp. 59–100. In: H. Ocampo, S. Paycha, and A. Vargas (Eds.), Geometric and Topological Methods for Quantum Theory, Springer, Berlin, 2005.

  E. Kraus, Renormalization of the electroweak standard model, Lectures given at the Saalburg Summer School (Germany), 1997.
  Internet: `http://arxiv.org/hep-th/9809069`
- Poincaré Seminar:
  Poincaré Seminar 2002: Vacuum Energy – Renormalization. Edited by B. Duplantier and V. Rivasseau, Birkhäuser, Basel, 2003.

  Poincaré Seminar 2007: Quantum Spaces. Edited by B. Duplantier and V. Rivasseau, Birkhäuser, Basel, 2007 (noncommutative geometry and elementary particle physics).
- Axioms of renormalization:
  K. Hepp, Renormalization theory, pp. 429–500. In: C. DeWitt and R. Stora (Eds.), Statistical Mechanics and Quantum Field Theory, Gordon and Breach, New York, 1971.
- State of the art in rigorous quantum field theory:
  K. Fredenhagen, K. Rehren, and E. Seiler, Quantum field theory: where we are, pp. 61–87. In: Lecture Notes in Physics **721**, Springer, Berlin, 2007.

In what follows we will summarize important approaches to renormalization theory together with hints for further reading. These approaches will be studied in the volumes to come.

(I) The Epstein–Glaser approach (also called the causal approach) based on tempered distributions: As a basic monograph, we recommend:

> G. Scharf, Finite Quantum Electrodynamics: the Causal Approach, Springer, Berlin, 1995.

Furthermore, we refer to:

J. Gracia-Bondia, The Epstein–Glaser Approach to Quantum Field Theory, Lecture Notes, AIP Conference Proceedings **809** (2005), American Institute of Physics, New York, pp. 24–43.

M. Dütsch and K. Fredenhagen, A local perturbative construction of observables in gauge theories: The example of QED, Commun. Math. Phys. **203** (1999), 71–105.

R. Brunetti and K. Fredenhagen, Micro-local analysis and interacting quantum field theories: renormalization on physical backgrounds, Commun. Math. Phys. **208** (2000), 623–661.

C. Bergbauer and D. Kreimer, The Hopf algebra of rooted trees in Epstein–Glaser renormalization, Ann. H. Poincaré **6** (2005), 343–367.

The classic paper is:

H. Epstein and V. Glaser, The role of locality in perturbation theory, Ann. Inst. Poincaré A **19**(3) (1973), 211–295.

See also (XII) page 1036 concerning the Master Ward Identity on

(II) The Bogoliubov–Parasiuk–Hepp–Zimmermann (BPHZ) renormalization approach: As an introduction, we recommend:

J. Collins, Renormalization, Cambridge University Press, 1984.

T. Kugo, Gauge Field Theory, Springer, Berlin, 1997 (in German).

Furthermore, we refer to:

N. Bogoliubov and D. Shirkov, Introduction to Quantum Field Theory, 3rd edn., Wiley, New York, 1980 (translated from the 1957 Russian edition into English).

N. Bogoliubov and D. Shirkov, Quantum Fields: Lectures held at the Moscow Lomonosov University, Reading, Massachusetts, 1983.

O. Zavialov, Renormalized Quantum Field Theory, Kluwer, Dordrecht, 1989.

The classic papers of the BPHZ approach are:

N. Bogoliubov and O. Parasiuk, On the multiplication of propagators in quantum field theory, Acta Math. **97** (1957), 227–326 (in German).

K. Hepp, Proof of the Bogoliubov–Parasiuk theorem on renormalization, Commun. Math. Phys. **2** (1966), 301–326.

W. Zimmermann, Convergence of Bogoliubov's method of renormalization in momentum space, Commun. Math. Phys. **15** (1969), 208–234.

In addition, we refer to the following classic papers:

S. Weinberg, High energy behavior in quantum field theory, Phys. Rev. **118** (1960), 838–849 (the power-counting theorem).

J. Lowenstein and W. Zimmermann, The power-counting theorem for Feynman integrals with massless propagators, Commun. Math. Phys. **44** (1975), 73.

J. Lowenstein and E. Speer, Distributional limits of renormalized integrals with zero-mass denominator, Commun. Math. Phys. **47** (1976), 43–51.

P. Breitenlohner and D. Maison, Dimensional renormalization and action principle, Commun. Math. Phys. **52** (1977), 11–39 (equivalence between BPHZ renormalization and dimensional renormalization).

P. Breitenlohner and D. Maison, Dimensional renormalized Green's functions for theories with massless particles, Commun. Math. Phys. **52**(1) (1977), 39–75.

A. Connes and D. Kreimer, Renormalization in quantum field theory and the Riemann–Hilbert problem I: The Hopf algebra structure of graphs and the main theorem. II: The beta function, diffeomorphisms, and the renormalization group. Commun. Math. Phys. **210** (2000), 249–273; **216** (2001), 215–241.

(III) Technical tools for the effective computation and renormalization of algebraic Feynman integrals: A general setting can be found in:

E. Manoukian, Renormalization, Academic Press, New York.

This concerns the following basic topics:

- the Weinberg power-counting theorem;
- the adiabatic limit of regularized algebraic Feynman integrals in the sense of tempered distributions;
- the construction of subtraction terms for algebraic Feynman integrals by means of counterterms in the Lagrangian.

Furthermore, we refer to:

V. Radanovic, Problem Book in Quantum Field Theory, Springer, New York, 2006 (e.g., exercises for the renormalization of quantum field models along with their solutions).

M. Veltman, Diagrammatica: The Path to Feynman Diagrams, Cambridge University Press, 1995 (dimensional renormalization).

N. Nakanishi, Graph Theory and Feynman Integrals, Gordon and Breach, New York, 1971.

E. Speer, Dimensional and Analytic Renormalization, pp. 25–94. In: G. Velo and A. Wightman (Eds.), Renormalization, Reidel, Dordrecht, 1976.

P. Pascual and R. Tarrach, QCD (Quantum Chromodynamics): Renormalization for Practitioners, Springer, Berlin, 1984.

A. Grozin, QED (Quantum Electrodynamics) and QCD (Quantum Chromodynamics): Practical Calculation and Renormalization of One- and Multi-Loop Feynman Diagrams, World Scientific Singapore, 2007.

V. Smirnov, Renormalization and Asymptotic Expansions, Birkhäuser, Basel, 1991.

V. Smirnov, Feynman Integral Calculus, Springer, Berlin, 2006.

See also (XX) on page 1042. Software systems for computing algebraic Feynman integrals in higher orders of perturbation theory (multi-loop computations) can be found on page 977.

The method of analytic regularization of algebraic Feynman integrals leads to the study of meromorphic functions with values in the space of tempered distributions. This can be found in:

I. Gelfand and G. Shilov, Generalized Functions, Vol. 1 (last chapter), Academic Press, New York, 1964.

N. Ortner and P. Wagner, Distribution-Valued Analytic Functions: Theory and Applications, Lecture Notes 37/2008, Max Planck Institute for Mathematics in the Sciences, Leipzig, 2008.
Internet: `http://www.mis.de/preprints`

The Epstein–Glaser approach is based on Hörmander's causal product for special classes of tempered distributions where the (causal) wave front condition is satisfied (see Sect. 12.6.2 of Vol. I). It is also possible to use the Colombeau product of distributions. Here, we refer to:

J. Colombeau, New Generalized Functions and Multiplication of Distributions, North-Holland, Amsterdam, 1984.

J. Colombeau, Multiplication of Distributions, North-Holland, Amsterdam, 1992.

H. Embacher, G. Grüble, and M. Oberguggenberger, Products of distributions in several variables and applications to zero-mass quantum electrodynamics in two-dimensional space-time, Zeitschrift für Analysis und ihre Anwendungen **11** (1992), 437–454. (This approach avoids the occurence of renormalization ambiguities from the very beginning.)

É. Charpentier, The cancellation of infinities in quantum field theory: a comparison of zeta regularization and the Colombeau interpretation, Dissertationes Mathematicae **383** (1999), 56 pages (in French).

The relation between Feynman diagrams, topology and algebraic geometry (the Picard–Lefschetz theory) is studied in:

R. Hwa and V. Teplitz, Homology and Feynman Diagrams, Benjamin, Reading, Massachusetts, 1966 (singularities of Feynman amplitudes and the Picard–Lefschetz theorem).

S. Lefschetz, Applications of Algebraic Topology: Graphs, and Networks, the Picard–Lefschetz Theory, and Feynman Algorithms, Springer, New York, 1975.

(IV) Compendium of important relations frequently used in quantum field theory:

V. Borodulin, R. Rogalyov, and S. Slabopitsky (1995), Compendium of Relations, 1995. Internet: `http://arxiv.org/hep-th/9507456`

S. Narison, QCD (Quantum Chromodynamics) as a Theory of Hadrons: From Partons to Confinement, Cambridge University Press, 2004 (appendix).

V. Smirnov, Feynman Integral Calculus, Springer, Berlin, 2006.

(V) BPHZ renormalization, Hopf algebras, and Rota–Baxter algebras:

K. Ebrahimi-Fard and D. Kreimer, The Hopf algebra approach to Feynman diagram calculations. Topical Review. J. Phys. A: Mathematical and General **38** (2005), R385–R407.

D. Kreimer, Knots and Feynman Diagrams, Cambridge University Press, 2000.

A. Connes and D. Kreimer, Renormalization in quantum field theory and the Riemann–Hilbert problem I: The Hopf algebra structure of graphs and the main theorem. II: The beta function, diffeomorphisms, and the renormalization group, Commun. Math. Phys. **210** (2000), 249–273; **216** (2001), 215–241.

B. Fauser, On the Hopf algebraic origin of Wick normal-ordering, J. Phys. A: Mathematical and General **34** (2001), 105–116.

C. Brouder, B. Fauser, A. Frabetti, and R. Oeckl, Quantum field theory and Hopf algebra cohomology, J. Phys. A: Mathematical and General **37** (2004), 5895–5927 (Hopf superalgebras).

C. Brouder, Quantum field theory meets Hopf algebra, 2006. Internet: `http://arxiv.org/hep-th/0611153`

K. Ebrahimi-Fard, L. Guo, and D. Kreimer, Spitzer's identity and the algebraic Birkhoff decomposition in perturbative quantum field theory, J. Phys. A: Mathematical and General **37** (2004), 11037–11052.

C. Bergbauer and D. Kreimer, Hopf algebras in renormalization theory: locality and Dyson–Schwinger equations from Hochschild cohomology, pp. 133–164. In: L. Nyssen (Ed.), Physics and Number Theory, European Mathematical Society, 2006.

S. Bloch, H. Esnault, and D. Kreimer, Motives associated to graph polynomials, Commun. Math. Phys. **267**(1) (2006), 181–225 (applications in number theory).

A. Connes and M. Marcolli, Noncommutative Geometry, Quantum Fields, and Motives, Amer. Math. Soc., Providence, Rhode Island, 2008 (applications to the Standard Model in particle physics and to the Riemann zeta function).

(VI) Quantum electrodynamics (QED) and Hopf algebras:

D. Volovich and I. Prokhorenko, Renormalization in quantum electrodynamics, and Hopf algebras, Trudy Mat. Inst. Steklova **245** (2004), 288–295.

W. van Suijlekom, The Hopf algebra of Feynman graphs in quantum electrodynamics, Letters in Mathematical Physics **77** (2006), 265–318.

W. van Suijlekom, Renormalization of gauge fields: the Hopf algebra approach, 2006. Internet: `http://arxiv.org/hep-th/0610137`

(VII) Quantum electrodynamics, the Dyson–Schwinger equation, and planar binary trees:

C. Brouder, Runge–Kutta methods and renormalization, Eur. J. Phys. **C12** (2000), 521–534.

C. Brouder, On the trees of quantum fields, Eur. J. Phys. **C12** (2000), 535–549.

C. Brouder and A. Frabetti, Renormalization of QED with planar binary trees, Eur. Phys. J. **C19** (2001), 715–741.

C. Brouder and A. Frabetti, QED and Hopf algebras on planar binary trees, J. Alg. **267** (2003), 298–322.

For the properties of the ground state of quantum fields (also called the vacuum), we refer to:

W. Dittrich and M. Gies, Probing the Quantum Vacuum: Effective Action Approach in Quantum Electrodynamics and its Applications, Springer, Berlin, 2003.

H. Genz, Nichts als das Nichts (The vacuum energy), Wiley–VCH, Weinheim, 2004 (in German).

(VIII) The Polchinski renormalization group approach in renormalization theory:

J. Polchinski, Renormalization and effective Lagrangians, Nucl. Phys. **B231** (1984), 269–295.

G. Keller and C. Kopper, Renormalizability proof for QED based on flow equations, Commun. Math. Phys. **176** (1996), 193–226.

G. Keller, C. Kopper, and C. Schophaus, Renormalization with flow equations in Minkowski space, Helv. Phys. Acta **70** (1997), 247–274.

M. Salmhofer, Perturbative renormalizability of φ^3 in six dimensions by renormalization group differential equations, 26 pages. In: F. Scheck, Theory of Renormalization and Regularization, 2002. Internet: `http://www.thep.physik.uni.mainz.de~ scheck/Hessbg`

M. Salmhofer, Renormalization: An Introduction, Springer, Berlin, 1999.

(IX) Perturbative quantum electrodynamics and axiomatic quantum field theory:
O. Steinmann, Perturbative Quantum Electrodynamics and Axiomatic Field Theory, Springer, Berlin, 2000.

(X) The Faddeev–Popov approach to gauge theory (Feynman functional integrals and ghosts):
L. Faddeev and A. Slavnov, Gauge Fields, Benjamin, Reading, Massachusetts, 1980.

(XI) The fundamental role of symmetry in the renormalization of quantum field theories (algebraic renormalization theory): For the BRST (Becchi, Rouet, Stora, Tyutin) approach to renormalization theory via cohomology, we refer to:
C. Becchi, A. Rouet, and R. Stora, Renormalization of the Abelian Higgs–Kible model, Commun. Math. Phys. **52** (1975), 127–162.
C. Becchi, A. Rouet, and R. Stora, Renormalization of gauge theories, Annals of Physics **98** (1976), 287–321.
O. Piguet, Renormalization in Quantum Field Theory, Lecture Notes I, Department of Theoretical Physics, University of Geneva, Switzerland, 1982 (in French).
O. Piguet, Renormalization of Gauge Theory, Lecture Notes II, Department of Theoretical Physics, University of Geneva, Switzerland, 1983 (in French).
O. Piguet and S. Sorella, Algebraic Renormalization: Perturbative Renormalization, Symmetries, and Anomalies. Springer, Berlin, 1995.
K. Sibold, Perturbative Renormalization – Quantization of Gauge Theories, Lecture Notes MPI-Ph/93-1, Max Planck Institute for Physics, Werner Heisenberg, Munich, 1993 (in German).
E. Kraus, Renormalization of the electroweak standard model to all orders, Annals of Physics **262** (1998), 155–259.

As an introduction to the connection between symmetries and generalized Ward identities (Slavnov–Taylor identities), we recommend the Lecture Notes by K. Sibold (1993) and Piguet, Sorella (1995) above. Furthermore, we refer to:
M. Böhm, A. Denner, and H. Joos, Gauge Theories of the Strong and Electroweak Interaction, Teubner, Stuttgart, 2001.
T. Kugo, Gauge Field Theory, Springer, Berlin, 1997 (in German).

(XII) Symmetry, the Master Ward Identity (MWI), and renormalization: It has recently been discovered that there exists a so-called Master Ward Identity which governs the symmetries of quantized theories. In addition, the Master Ward Identity can be used as a general principle for constructing the renormalization of quantum field theories. As an introduction, we recommend:
M. Dütsch, The Master Ward Identity: a universal formulation of classical symmetries. Can they be realized in perturbative quantum field theory (pQFT)? 17 pages. In: F. Scheck (Ed.), Theory of Renormalization and Regularization. Lecture Notes, 2002.
Internet: `http://www.thep.physik.uni.mainz.de~scheck/Hessbg`
Furthermore, we refer to:
M. Dütsch and F. Boas, The Master Ward Identity, Rev. Mod. Phys. **14** (2002), 977–1049.
M. Dütsch and F. Brennecke, Removal of violations of the Master Ward Identity in perturbative quantum field theory (pQFT), 2007.
Internet: `http://www.mis.mpg.de/preprints/2007`

There exist crucial relations between the following topics:

- *retarded functions* in quantum field theory – the GLZ (Glaser, Lehmann, Zimmermann) approach as an alternative to the LSZ (Lehmann, Szymanzik, Zimmermann) approach via Green's functions,
- the retarded product of classical fields via classical retarded Green's functions,
- the Poisson–Peierls brackets in classical relativistic field theory defined by retarded products,
- the time-ordered product and the Epstein–Glaser approach,
- the Action Ward Identity (AWI),
- the Master Ward Identity (MWI),
- the quantum action principle and the Dyson–Schwinger equation,
- the BRST (Becchi, Rouet, Stora, Tyutin) approach in gauge theory,
- quantum deformation,
- renormalization group.

These relations can be found in:

M. Dütsch and K. Fredenhagen, The Master Ward Identity and the generalized Schwinger–Dyson equation in classical field theory, Commun. Math. Phys. **243** (2003), 275–314.

M. Dütsch and K. Fredenhagen, Causal perturbation theory in terms of retarded products, and a proof of the Action Ward Identity, Rev. Math. Phys. **16**(10) (2004), 1291–1348.

M. Dütsch and K. Fredenhagen, Action Ward Identity and the Stueckelberg–Petermann renormalization group, pp. 113–123. In: A. Boutet de Monvel et al. (Eds.), Rigorous Quantum Field Theory, Birkhäuser, Basel, 2006.

The classic papers are:

R. Peierls, The commutation laws of relativistic field theory, Proc. Royal Soc. London **A214** (1952), 143–157.

E. Stueckelberg and D. Rivier, Causalité et structure de la matrice S, Helv. Phys. Acta **23** (1950), 215–222 (in French).

E. Stueckelberg and T. Green, T. (1951), Elimination des constantes arbitraires dans la théorie relativiste des quanta, Helv. Phys. Acta **24** (1951), 153–174 (in French).

E. Stueckelberg and A. Petermann, La normalisation des constantes dans la théorie des quanta, Helv. Phys. Acta **26** (1953), 215–222 (in French).

H. Lehmann, K. Symanzik, and W. Zimmermann, On the formulation of quantized field theories, Nuovo Cimento **1** (1955), 205–225; **6** (1957), 319–333 (the LSZ approach via Green's functions).

V. Glaser, H. Lehmann, and W. Zimmermann, Field operators and retarded functions, Nuovo Cimento **6** (1957), 1122–1128 (the GLZ approach via retarded functions).

O. Steinmann, On the definition of retarded and time-ordered products), Helv. Phys. Acta **36** (1963), 90–112 (in German).

O. Steinmann, Perturbation Expansions in Axiomatic Field Theory, Springer, Berlin, 1971 (retarded functions).

The *quantum action principle* represents the fundamental dynamical principle in quantum field theory. This principle describes the changing of the Green's functions (correlation functions) under a change of the following quantities:

- external parameters (e.g., external sources) and
- quantum fields (e.g., symmetry transformations like global symmetries or local gauge symmetries).

As an introduction, we refer to:

K. Sibold, Perturbative Renormalization – Quantization of Gauge Theories, Lecture Notes MPI-Ph/93-1, Max Planck Institute for Physics, Werner Heisenberg, Munich, 1993 (in German).

O. Piguet and S. Sorella, Algebraic Renormalization: Perturbative Renormalization, Symmetries, and Anomalies. Springer, Berlin, 1995.

Classic papers are:

J. Schwinger, A note on the quantum dynamical principle, Phil. Mag. **44** (1953), 1171–1193.

J. Schwinger, Relativistic quantum field theory, pp. 140–154. In: Nobel Lectures 1963–1970 in Physics, World Scientific, Singapore. (Schwinger's Nobel lecture is mainly devoted to his quantum action principle.)

J. Schwinger, Quantum Mechanics, Springer, New York, 2001.

J. Lowenstein, Differential vertex operations in Lagrangian field theory, Commun. Math. Phys. **24** (1971), 1–21.

Y. Lam, Perturbation Lagrangian theory for scalar fields: Ward–Takahashi identity and current algebra, Phys. Rev. **D6** (1972), 2145–2161.

The fundamental role played by *hierarchies of functions* in quantum field theory is discussed in:

R. Haag, Local Quantum Physics: Fields, Particles, Algebras, Springer, Berlin, 1996 (Chap. II.2).

S. Narison, QCD (Quantum Chromodynamics) as a Theory of Hadrons: From Partons to Confinement, Cambridge University Press, 2004 (Chap. 4).

This concerns the following functions:

- Wightman functions (axiomatic quantum field theory),
- Green's functions (LSZ approach), and
- retarded functions (GLZ approach).

Intuitively, if $\varphi = \varphi(x)$ is a (naive) scalar quantum field (i.e, $\varphi(x)$ is a linear operator in a Hilbert space for each space-time point x), then we have the following complex-valued functions at hand, which describe the quantum field φ by means of vacuum expectation values:

- $W(x,y) := \langle 0|\ \varphi(x)\varphi(y)\ |0\rangle$ (2-point Wightman function);
- $G(x,y) := \langle 0|\mathcal{T}(\varphi(x)\varphi(y))|0\rangle$ (2-point Green's function also called the 2-point correlation function);
- $R(x,y) := -\mathrm{i}\theta(x-y)\langle 0|\ [\varphi(x),\varphi(y)]\ |0\rangle$ (2-point retarded function).

Here, $x = (\mathbf{x},t)$ and $y = (\mathbf{y},s)$ are arbitrary points of the Minkowski space (i.e., $\mathbf{x}$ and $\mathbf{y}$ are position vectors; t and s are time variables). The symbol $\mathcal{T}(\varphi(x)\varphi(y))$ denotes the time-ordered product, that is,

$$\mathcal{T}(\varphi(x)\varphi(y)) := \begin{cases} \varphi(x)\varphi(y) & \text{if} \quad t > s, \\ \varphi(y)\varphi(x) & \text{if} \quad s > t. \end{cases}$$

Furthermore, we set

$$\theta(x) := \begin{cases} 1 & \text{if} \quad t \geq 0 \\ 0 & \text{if} \quad t < 0. \end{cases}$$

Finally, we use the commutator

$$[\varphi(x), \varphi(y)] := \varphi(x)\varphi(y) - \varphi(y)\varphi(x).$$

The corresponding n-point functions are defined similarly. For example, we get:

- $W(x, y, z) := \langle 0|\ \varphi(x)\varphi(y)\varphi(z)\ |0\rangle$ (3-point Wightman function);
- $G(x, y, z) := \langle 0|\ \mathcal{T}(\varphi(x)\varphi(y)\varphi(z))\ |0\rangle$ (3-point Green's function also called 3-point correlation function);
- $R(x, y, z)$ (3-point retarded function) is equal to

$$\begin{aligned} &-\theta(x-y)\theta(x-z)\ \langle 0|\ [\,[\varphi(x), \varphi(y)], \varphi(z)]\ |0\rangle \\ &\quad -\theta(x-z)\theta(x-y)\ \langle 0|\ [\,[\varphi(x), \varphi(z)], \varphi(y)]\ |0\rangle. \end{aligned}$$

  Here, the variable x is distinguished, and we use a permutation of the remaining variables y and z.

This way we obtain the following three hierarchies of functions:

- n-point Wightman functions,
- n-point Green's functions, and
- n-point retarded functions.

Here, $n = 1, 2, \ldots$ Formulated naively, each of these hierarchies determines the quantum field φ. It turns out that this heuristic approach has to be modified by passing to tempered distributions, and so on. The elegant idea of describing a quantum field by a hierarchy of functions (i.e., vacuum expectation values of appropriate operator products) was created in the 1950s by Wightman, Glaser, Lehmann, Szymanzik, and Zimmermann. In axiomatic quantum field theory, a fixed hierarchy of functions is characterized by axioms. The goal is to prove a so-called reconstruction theorem which shows the existence of a corresponding quantum field as an operator-valued tempered distribution (see Sect. 15.6.1 of Vol. I and (IX) on page 1036 above).

(XIII) Anomalies: The violation of classical symmetries after quantization corresponds to so-called anomalies. We recommend:

A. Zee, Quantum Field Theory in a Nutshell, Princeton University Press, 2003.

O. Piguet and S. Sorella, Algebraic Renormalization: Perturbative Renormalization, Symmetries, and Anomalies. Springer, Berlin, 1995.

K. Fujikawa and H. Suzuki, Path Integrals and Quantum Anomalies, Oxford University Press, 2004.

V. Parameswaran Nair, Quantum Field Theory: A Modern Perspective, Springer, New York, 2005.

S. Adler, Perturbation theory anomalies, pp. 1–164. In: S. Deser et al. (Eds.), Lectures on Elementary Particles and Quantum Field Theory, MIT Press, Cambridge, Massachusetts, 1970.

S. Adler, Adventures in Theoretical Physics, Selected Papers with Commentaries, World Scientific, Singapore, 2006.

(XIV) Operator product expansion (OPE):

J. Collins, Renormalization: An Introduction to Renormalization, the Renormalization Group, and the Operator-Product Expansion, Cambridge University Press, 1984.

T. Kugo, Gauge Field Theory, Springer, Berlin, 1997 (in German).

M. Peskin and D. Schroeder, An Introduction to Quantum Field Theory, Addison-Wesley, Reading, Massachusetts, 1995.

The classic papers on operator product expansions are:

K. Wilson, On products of field operators at short distances, Cornell Report, 1964.

K. Wilson, Non-Lagrangian models of current algebra, Phys. Rev. **179** (1969), 1499–1512.

W. Zimmermann, Normal products and the short-distance expansion in the perturbation theory of renormalizable interactions, Annals of Physics **77** (1973), 536–601.

(XV) Renormalization in curved space-time:

R. Brunetti and K. Fredenhagen, Micro-local analysis and interacting quantum field theories: renormalization on physical backgrounds, Commun. Math. Phys. **208** (2000), 623–661.

S. Hollands and R. Wald, On the renormalization group in curved space-time, Commun. Math. Phys. **237** (2003), 123–160.

S. Hollands, The operator product expansion for perturbative quantum field theory in curved space-time, Commun. Math. Phys. **273** (2007), 1–36.

S. Hollands, Renormalized Yang–Mills fields in curved space-time, 2007, 115 pp. Internet: `http://arxiv.org/0705.3340`

The last two papers use the following crucial tool from algebraic geometry:

W. Fulton and R. MacPherson, A compactification of configuration spaces, Ann. of Math. **139** (1994), 183–225.

(XVI) Euclidean approach to renormalization:

J. Feldman, T. Hurd, L. Rosen, and J. Wright, QED: A Proof of Renormalizability, Springer, Berlin, 1988.

G. Keller and C. Kopper, Renormalizability proof for QED based on flow equations, Commun. Math. Phys. **176** (1996), 193–226.

Renormalization of many-particle systems in statistical physics can be found in:

J. Feldman, H. Knörrer, and E. Trubowitz, A two-dimensional Fermi liquid, 2003. An impressive series of ten papers, all accessible at the Internet: `http://www.math.ubc.ca/~feldman/fl.htlm`

For renormalization in noncommutative geometry, we refer to:

H. Grosse and R. Wulkenhaar, Renormalisation of φ^4-theory on noncommutative $\mathbb{R}^4$ in the matrix base, Commun. Math. Phys. **256** (2005), 305–374.

R. Wulkenhaar, Euclidean quantum field theory and commutative and noncommutative spaces, Lecture Notes. In: H. Ocampo, S. Paycha, and A. Vargas (Eds.), Geometric and Topological Methods for Quantum Theory, Springer, Berlin, 2005, pp. 59–100.

V. Rivasseau, Noncommutative renormalization. In: Poincaré Seminar 2007, Quantum Spaces. Edited by B. Duplantier and V. Rivasseau, Birkhäuser, Basel, 2007, pp. 19–108.

(XVII) Classical differential equations and renormalization:

J. Feldman and E. Trubowitz, Renormalization in classical mechanics and many-body quantum field theory, Jerusalem J. d'Analyse Mathématique **52** (1992), 213–247.

J. Bricmont and A. Kupiainen, Renormalizing partial differential equations, pp. 83–115. In: V. Rivasseau (Ed.), Constructive Physics, Springer, Berlin, 1995.

J. Bricmont, K. Gawędzki, and A. Kupiainen, Kolmogorov–Arnold–Moser (KAM) theorem and quantum field theory, Commun. Math. Phys. **201**(3) (2004), 699–727.

(XVIII) Classical partial differential equations and mathematical aspects of quantum electrodynamics:

E. Lieb, The Stability of Matter: From Atoms to Stars, Selecta of Elliott Lieb. Edited by W. Thirring, Springer, New York, 2002.

J. Fröhlich, The quantum theory of light and matter – mathematical results, pp. 69–78 (survey). In: G. Buschhorn and J. Wess (Eds.), Fundamental Physics: Heisenberg and Beyond, Springer, Berlin, 2004.

S. Gustafson and I. Sigal, Mathematical Concepts of Quantum Mechanics, Springer, Berlin, 2003.

E. Lieb and H. Siedentop, Renormalization of the regularized relativistic electron-positron field, Commun. Math. Phys. **213**(3) (2000), 673–683.

C. Hainzl and H. Siedentop, Non-perturbative mass and charge renormalization in relativistic no-photon quantum electrodynamics, Commun. Math. Phys. **243**(2) (2003), 241–260.

V. Bach, J. Fröhlich, and I. Sigal, Quantum electrodynamics of confined non-relativistic particles, Advances in Mathematics **137** (1998), 205–395.

V. Bach, J. Barbaroux, B. Helffer, and H. Siedentop, On the stability of the relativistic electron-positron field, Commun. Math. Phys. **202**(2) (1999), 445–460.

V. Bach, J. Fröhlich, and I. Sigal, Spectral analysis for systems of atoms and molecules coupled to the quantized radiation field, Commun. Math. Phys. **207**(2) (1999), 249–290.

J. Fröhlich, M. Griesemer, and B. Schlein, Asymptotic completeness for Rayleigh scattering, Ann. Henri Poincaré **3** (2002), 107–170.

J. Fröhlich, M. Griesemer, and B. Schlein, Asymptotic completeness for Compton scattering, Commun. Math. Phys. **252** (2004), 415–476.

W. Appel and M. Kiessling, Mass and spin renormalization in Lorentz electrodynamics, Annals of Physics **289** (2001), 24–83.

M. Kiessling, Electromagnetic field theory without divergence problems I, II: J. Statistical Physics **116** (2004), 1057–1122.

J. Dereziński and H. Siedentop (Eds.), Large Coulomb Systems: Lecture Notes on Mathematical Aspects of Quantum Electrodynamics, Springer, Berlin, 2006.

(XIX) Limits of physical systems:

A. Elgart, B. Erdös, B. Schlein, and H. Yau, Nonlinear Hartree equation as the mean field limit of weakly coupled fermions, J. Math. Pures Appl. **83** (2004), 1241–1273.

L. Erdös, B. Schlein, and H. Yau, Derivation of the Gross–Pitaevskii hierarchy for the dynamics of a Bose–Einstein condensate, Comm. Pure Appl. Math. **59**(12) (2006), 1659–1741.

L. Erdös, B. Schlein, and H. Yau, Derivation of the cubic non-linear Schrödinger equation from quantum dynamics of many-body systems, Invent. Math. **167**(3) (2007), 515–614.

(XX) Numerical mathematics, computer algebra, and renormalization:

J. Butcher, An algebraic theory of integration methods, Math. Comp. **26** (1972), 79–106 (new approach to the Runge–Kutta methods for the numerical solution of ordinary differential equations).

C. Brouder, Runge–Kutta methods and renormalization, Eur. J. Phys. **C12** (2000), 521–534.

G. Battle, Wavelets and Renormalization, World Scientific, Singapore, 1999.

D. Broadhurst and D. Kreimer, Renormalization automated by Hopf algebra, J. Symb. Comp. **27** (1999), 581–600.

D. Broadhurst and D. Kreimer, Combinatorial explosion of renormalization tamed by Hopf algebra. 30-loop Padé–Borel resummation, Phys. Lett. **B475** (1999), 63–70.

B. Fauser and R. Ablamowicz, Clifford and Grassmann Hopf algebras via the BIGEBRA package for MAPLE, Computer Physics Communications **170**(2) (2005), 115–130.

In algebraic geometry and computer algebra, the method of Gröbner basis plays a fundamental role:

D. Cox, J. Little, and D. O'Shea, Using Algebraic Geometry, Springer, New York, 1998.

The method of Gröbner basis can be used in order to compute algebraic Feynman integrals on the computer. We refer to:

V. Smirnov, Feynman Integral Calculus, Springer, Berlin.

See also Sect. 18.4 on page 977 about automated multi-loop computations.

(XXI) Lattice approximation in quantum field theory:

M. Creutz, Quarks, Gluons, and Lattices, Cambridge, University Press, 1983.

I. Montvay and G. Münster, Quantum Fields on a Lattice, Cambridge University Press, 1997.

(XXII) Renormalization of supersymmetric models:

O. Piguet and K. Sibold, Renormalized Supersymmetry, Birkhäuser, Boston, 1986.

W. Hollik, E. Kraus, M. Roth, C. Rupp, K. Sibold, and D. Stöckinger, Renormalization of the minimal supersymmetric standard model, Nuclear Physics **B639** (2002), 3–65.

M. Peskin, Supersymmetry: the next spectroscopy, 99–134. In: G. Buschhorn and J. Wess (Eds.), Fundamental Physics: Heisenberg and Beyond, Springer, Berlin, 2004.

(XXIII) Quantum groups and the Jackson integral for q-deformed structures:

H. Wachter, q-Integration on quantum spaces, Eur. Phys. J. **C32**, 281–297.

H. Wachter, Towards a q-deformed quantum field theory, 25 pages. In: B. Fauser, J. Tolksdorf, and E. Zeidler (Eds.), Quantum Field Theory – Competitive Models, Birkhäuser, Basel, 2008.

A. Schmidt, Towards a q-deformed supersymmetric field theory, 18 pages. In: B. Fauser, J. Tolksdorf, and E. Zeidler (Eds.) (2008) quoted above.

(XXIV) Rigorous non-perturbative approach:

B. Simon, The $P(\varphi)_2$-Euclidean Quantum Field Theory (in two dimensions), Princeton University Press, 1974.

J. Glimm and A. Jaffe, Mathematical Methods of Quantum Physics: A Functional Integral Point of View, Springer, New York, 1981.

J. Glimm and A. Jaffe, Quantum Field Theory and Statistical Mechanics: Expositions, Birkhäuser, Boston, 1985.

V. Rivasseau, From Perturbative to Constructive Renormalization, Princeton University Press, 1991.

See also the references given on page 871ff of Vol. I.

(XXV) The global Landau inconsistency problem in quantum electrodynamics (Landau pole):

L. Landau and E. Lifshitz, Course of Theoretical Physics, Vol 4: Quantum Electrodynamics, Sect. 128. Butterworth–Heinemann, Oxford, 1982.

L. Landau, Collected Papers, Pergamon Press, Oxford, 1965.

V. Rivasseau, From Perturbative to Constructive Renormalization, Princeton University Press, 1991.

S. Narison, QCD (Quantum Chromodynamics) as a Theory of Hadrons: From Partons to Confinement, Cambridge University Press, 2004.

(XXVI) History of renormalization theory:

S. Schweber, QED and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga. Princeton University Press, 1994.

L. Brown, (Ed.), Renormalization: From Lorentz to Landau and Beyond, Springer, New York, 1993.

L. Brown and L. Hoddeson, The Birth of Particle Physics, Cambridge University Press, 1983.

L. Brown, M. Dresden, L. Hoddeson, and M. Riordan (Eds.) (1995), The Rise of the Standard Model, Cambridge University Press, Cambridge, United Kingdom.

P. Dirac, The origin of quantum field theory, pp. 39–55. In: L. Brown and L. Hodeson (Eds.) (1983) (see above).

J. Schwinger, Renormalization theory of quantum electrodynamics, pp. 329–353. In: L. Brown and L. Hoddeson (1983) (see above).

A. Miller, Early Quantum Electrodynamics: A Source Book, Cambridge University Press, 1994.

S. Weinberg, The Quantum Theory of Fields, Vol. I: Foundations. Chap. 1: Historical Introduction. Cambridge University Press, 1995.

D. Kaiser, Drawing Theories Apart: The Dispersion of Feynman Diagrams in Postwar Physics, University of Chicago Press, 2005.

V. Rivasseau, An introduction to renormalization. In: Poincaré Seminar 2002, Vacuum Energy – Renormalization, pp. 213–240. Edited by B. Duplantier and V. Rivasseau, Birkhäuser, Basel, 2003.

V. Rivasseau, Non-commutative renormalization. In: Poincaré Seminar 2007, pp. 19–108. Edited by B. Duplantier and V. Rivasseau, Birkhäuser, Basel, 2007.

A. Connes, Symmétries galoisiennes et renormalisation. In: Poincaré Seminar 2002 (see above), pp. 241–264.

M. Veltman, Facts and Mysteries in Elementary Particle Physics, World Scientific, Singapore, 2003.

Epilogue

> The organic unit of mathematics is inherent in the nature of this science, for mathematics is the foundation of all exact knowledge of natural phenomena.
>
> David Hilbert, 1900

In 1900 Hilbert formulated twenty three open problems. The quotation is taken from the end of his speech at the Second International Congress of Mathematicians in Paris. Most of the problems were solved in the 20th century. This is discussed in B. Yandell, The Honors Class: Hilbert's Problems and Their Solvers, Peters, Natick, Massachusetts, 2001.

> A completely new branch of science has taken shape before our eyes in recent years, a branch that should properly be called modern mathematical physics.
> It has the same genetic origins as classical mathematical physics. But whereas the theory of partial differential equations was generated by problems of classical physics (potential theory, theory of propagation of electromagnetic waves and such), it turns out that modern theoretical physics – quantum field theory with an *infinite* number of degrees of freedom – requires different, more abstract and modern mathematical methods. These methods consist primarily of the theory of distributions (generalized functions), functional analysis and operator theory, the representation theory of groups and algebras, topological algebra, and the like.
> The solution of the new physical problems of quantum field theory was first sought through perfecting the usual methods of quantum mechanics. At that time physicists managed to realize that in order to obtain reasonable answers to their questions they needed a deeper understanding of the mathematical nature of the objects they were studying, such as distributions or unbounded operators, and they needed to raise the standard of proof in their arguments.
> Subsequently, to liberate themselves from excessive and sometimes meaningless details, they began to seek out axiomatic routes for constructing theories. It then became obvious that modern mathematical methods sometimes make it possible to obtain very strong results. In this connection we might mention the theory of functions of several complex variables or the concept of weak equivalence of representations.
> We note, finally, that several specific quantum phenomena provide a direct physical illustration of the famous theorem on the existence of *inequivalent* representations in the case of an *infinite* number of of degrees of freedom.
> The examples just mentioned come from quantum electrodynamics, the theory of strong interaction at high energies, and problems of statistical physics. In particular, in the physics of strong interactions, due to the complexity of the dynamical picture, dispersion methods based on the

> general analytic properties of the amplitude of the process turned out to be especially useful. They now have immediate applications to the needs of experimental research.
> We are at the very beginning of the route. It suffices to recall that as not a single nontrivial example of quantum field theory has so far been constructed outside of perturbation theory that is in any way close to the real physical world of four dimensions.
> The attention of physicists to the methods of modern mathematics and the interest of mathematicians in the problems of quantum physics are mutually productive.
>
> Nikolai Nikolaevich Bogoliubov, 1981

This quotation is taken from an introductory speech given by Bogoliubov at the conference on quantum field theory in Alushta (Crimea/Black Sea) in 1981. See

> V. Vladimirov, Nikolai Nikolaevich Bogoliubov (1909–1992) – Mathematician by the Grace of God. In: A. Bolibruch, Yu. Osipov, and Ya. Sinai (Eds.), Mathematical Events in the 20th Century, Springer, Berlin, and PHASIS, Moscow, 2006 (reprinted with permission).

The name Bogoliubov consists of the two Russian words: *Bog* – God, and *liubov* – love. Bogoliubov was born in 1909 in Nishnii Novgorod (Russia). His father was a prominent clergyman. The Bogoliubovs soon moved to Kiev (Ukraine). After World War II, Bogoliubov moved to Moscow. From that time he was closely related with the Steklov Mathematical Institute, the Moscow State University, and the Nuclear Research Center in Dubna (near Moscow). The monograph written by N. Bogoliubov, A. Logunov, A. Orsak, and I. Todorov, General Principles of Quantum Field Theory, Kluwer 1990, is the most comprehensive presentation of mathematical quantum field theory. We also refer to N. Bogoliubov, Selected Works. Part I: Dynamical Theory. Part II: Quantum and Classical Statistical Mechanics. Part III: Nonlinear Mechanics and Pure Mathematics. Part IV: Quantum Field Theory. Gordon and Breach, London, 1990–1995.

> The organic fusion of mathematics and physics in the work of Bogoliubov enabled him to make a decisive contribution to the development of theoretical physics, and in fact to lay the foundations of modern mathematical physics, continuing the tradition of Hilbert (1962–1943), Poincaré (1845–1912), and Einstein (1879–1955).
>
> Vasilii Vladimirov, 2005

The following quotation is taken from the beautiful article by M. Atiyah, Mathematics in the 20th Century, Bulletin of the London Mathematical Society **34** (2002), 1–15 (reprinted with permission):

> The 21st century might be the era of *quantum mathematics* or, if you like, of infinite-dimensional mathematics. What could this mean? Quantum mathematics could mean, if we get that far, 'understanding properly the analysis, geometry, topology, algebra, of various non-linear function spaces', and 'by understanding properly' I mean understanding it in such a way that as to get quite rigorous proofs of all the beautiful things the physicists have been speculating about.
> One should say that, if you go at infinite dimensions in a naïve way and ask naïve questions, you usually get the wrong answers, or the answers are dull. Physical application, insight, and motivation has enabled physicists to ask intelligent questions about infinite dimensions, and to do very subtle things where sensible answers *do* come out, and therefore doing infinite-dimensional analysis in this way is by no means a simple task. You have

to go about it in the right way. We have a lot of clues. The map is laid out; this is what should be done, but there is a long way to go yet.

What else might happen in the 21st century? I would like to emphasize Connes' non-commutative differential geometry. Alain Connes has his rather magnificent unified theory. Again, it combines everything. It combines analysis, algebra, geometry, topology, physics, and number theory, all of which combine to parts of it. It is a framework which enables us to do what differential geometers normally do, including its relationship with topology, in the context of non-commutative analysis. There are good reasons for wanting to do this, applications (potential or otherwise) in number theory, discrete groups, and so on, and in physics. An interesting link with physics is just being worked out (see the monograph by Connes and Marcolli (2008)). How far this will go, what it will achieve, remains to be seen. It certainly is something which I expect will be significantly developed in the first decade at least of the next century, and it is possible it could have a link with the as-yet-undeveloped (rigorous) quantum field theory.

Moving in another direction, there is what is called 'arithmetic geometry' or Arakelov geometry, which tries to unify as much as possible algebraic geometry and parts of number theory. It is a very successful theory. It has made a nice start, but has a long way to go. Who knows?

Of course, all of these have strands in common. I expect physics to have its impact spread all the way through, even to number theory; Andrew Wiles disagrees, and only time will tell...

Sir Michael Atiyah, 2002

References

Much material including proofs quoted in this volume can be found in the following books:

Zeidler, E. (1986), Nonlinear Functional Analysis and Its Applications.
Vol. I: Fixed-Point Theory, 3rd edn. 1998; Vol. IIA: Linear Monotone Operators, 2nd edn. 1997; Vol. IIB: Nonlinear Monotone Operators; Vol. III: Variational Methods and Optimization; Vol. IV: Applications to Mathematical Physics, 2nd edn. 1995, Springer, New York.

Zeidler, E. (1995a), Applied Functional Analysis, Vol. 1: Applications to Mathematical Physics, 2nd edn. 1997, Applied Mathematical Sciences (AMS) 108, Springer, New York.

Zeidler, E. (1995b), Applied Functional Analysis, Vol. 2: Main Principles and Their Applications, Applied Mathematical Sciences (AMS) 109, Springer, New York.

Zeidler, E. (Ed.) (2004), Oxford Users' Guide to Mathematics, Oxford University Press, New York (translated from German into English).

Zeidler, E. (2007), Die Faszination der Wechselwirkungen zwischen Mathematik und Naturwissenschaften (The fascination of the interaction between mathematics and the sciences), Jahresbericht der DMV **109** (2007) (Annual report of the German Mathematical Society; special issue on the occasion of the "Year of Mathematics 2008" in Germany), pp. S19–S40 (in German).

Zeidler, E. (Ed.) (2008a), Teubner-Taschenbuch der Mathematik (Teubner Handbook in Mathematics), Teubner Wiesbaden (in German).
Vol. 1: 3rd edition (English edition: see Zeidler (2004));
Vol. 2: 9th newly written edition. (English edition in preparation.)

Zeidler, E. (2008b), Mathematics – the cosmic eye of humanity, Nonlinear Analysis **69**, 1073–1082.
Internet: `http://www.mis.mpg.de/zeidler.hmtl`

Zeidler, E. (2008c), Gedanken zur Zukunft der Mathematik (Reflections on the future of mathematics), 25pp. In: H. Wußing, 6000 Jahre Mathematik: Eine kulturgeschichtliche Zeitreise (6000 years of mathematics: a cultural journey through time), Vols. I, II, Springer, Heidelberg (in German).

We also recommend the following monographs written by my colleague Jürgen Jost:

Jost, J. (1984), Harmonic Maps between Riemannian Manifolds, ANU-Press, Canberra (Australia).

Jost, J. (1991a), Nonlinear Methods in Riemannian and Kählerian Geometry, Birkhäuser, Basel.

Jost, J. (1991b), Two-Dimensional Geometric Variational Problems, Wiley, New York.

Jost, J. (1997b), Nonpositive Curvature: Geometric and Analytic Aspects, Birkhäuser, Basel.

Jost, J., Xianqing Li-Jost (1998), Calculus of Variations, Cambridge University Press, Cambridge, United Kingdom.
Jost, J. (2001), The Bosonic String: A Mathematical Treatment, International Press, Boston.
Jost. J. (2002), Partial Differential Equations, Springer, New York.
Jost, J. (2005a), Postmodern Analysis, 3rd edn., Springer, Berlin.
Jost, J. (2005b), Dynamical Systems: Examples of Complex Behavior, Springer Berlin.
Jost, J. (2006), Compact Riemann Surfaces: An Introduction to Contemporary Mathematics, 3rd edn., Springer, Berlin.
Jost, J., Eschenburg, J., (2007), Differentialgeometrie und Minimalflächen (Differential geometry and minimal surfaces), 2nd enlarged edn., Springer, Berlin (in German).
Jost, J. (2008a), Riemannian Geometry and Geometric Analysis, 5th edn., Springer, Berlin.
Jost, J. (2008b), Geometry and Physics, Springer, Berlin.

For a summary of recent developments in quantum field theory we refer to:

Buschhorn, G., Wess, J. (Eds.) (2004), Fundamental Physics: Heisenberg and Beyond, Springer, Berlin.
Fauser, B., Tolksdorf, J., Zeidler, E. (Eds.) (2006), Quantum Gravitation: Mathematical Models and Experimental Bounds, Birkhäuser, Basel.
Fauser, B., Tolksdorf, J., Zeidler, E. (Eds.) (2008), Quantum Field Theory – Competitive Methods, Birkhäuser, Basel.

We also recommend the Lectures of the Poincaré Seminar "Bourbaphy" organized by l'Institut Poincaré in Paris. These lectures reflect the state of the art in modern physics:

Poincaré Seminar (2002), Vacuum Energy – Renormalization. Edited by B. Duplantier and V. Rivasseau, Birkhäuser, Basel, 2003.
Poincaré Seminar (2003), Bose–Einstein Condensation – Entropy. Edited by J. Dalibard, B. Duplantier, and V. Rivasseau. Birkhäuser, Basel, 2005.
Poincaré Seminar (2004a), The Quantum Hall–Effect. Edited by B. Douçot, B. Duplantier, V. Rivasseau, and V. Pasquier, Birkhäuser, Basel, 2005.
Poincaré Seminar (2004b), String Theory (to appear).
Internet: La théorie des cordes. `http://www.bourbaphy.fr`
Poincaré Seminar (2005a), Einstein 1905–2005. Edited by T. Damour, O. Darrigol, B. Duplantier, and V. Rivasseau, Birkhäuser, Basel, 2006.
Poincaré Seminar (2005b), Quantum Decoherence. Edited by B. Duplantier, J. Raimond, and V. Rivasseau, Birkhäuser, Basel, 2006.
Poincaré Seminar (2006), Gravitation and Experience (to appear).
Internet: Gravitation et expérience. `http://www.bourbaphy.fr`
Poincaré Seminar (2007a): Quantum Spaces. Edited by B. Duplantier and V. Rivasseau. Birkhäuser, Basel, 2007.
Poincaré Seminar (2007b): Spin (to appear).

The state of the art in modern mathematics can be found in:

Encyclopedia of Mathematical Sciences, 142 vol., Springer, Berlin, 1987ff.

Most of these books are translations from the original Russian version into English.

Collected Works

> It appears to me that if one wants to make progress in mathematics, one should study the masters and not the pupils.
>
> Niels Hendrik Abel

The best way of learning deep mathematics and physics is to look at the "Collected Works" of great scientists. Interestingly enough, it turns out that, quite often, deep mathematics is (at least implicitly) related to the fascinating question about the description of the fundamental forces in nature. We recommend to have a look at the following jewels of mathematical and physical literature:

Niels Hendrik Abel (1802–1829), Oeuvres, 2 vol., edited by L. Sylow and S. Lie, Christiana, Norge (Norway), 1881.

Stephen Adler, Adventures in Theoretical Physics, Selected Papers with Commentaries, World Scientific, Singapore, 2006.

Heinz-Wilhelm Alten et al., 4000 Years of Algebra: History, Human Cultures, People, Springer, Heidelberg, 2003 (in German).

Archimedes (ca. 287–212 B.C.), The Works of Archimedes, edited by T. Heath, Dover, 1953.

Aristotle (384–324 B.C.), Aristotle's Philosophy of Mathematics, by H. Apostle, University of Chicago Press, Chicago, Illinois, 1952.

Vladimir Igorevich Arnold, From Hilbert's superposition problem to dynamical systems, pp. 19–47. In: A. Bolibruch, Yu. Osipov, and Ya. Sinai (Eds.) (2006).

Michael Atiyah, Collected Works, 5 vol., Cambridge University Press, Cambridge, United Kingdom, 1988.

Michael Atiyah, Mathematics in the 20th Century, Bull. London Math. Soc. **34** (2002), 1–15.

Michael Atiyah and Daniel Iagolnitzer (Eds.), Fields Medallists' Lectures, World Scientific, Singapore, 2003.

Roger Bacon (1214–1294), A Biography, by F. Woodroff, London, 1938.

Francis Bacon (1561–1626), The Works of Francis Bacon, 14 vol., edited by J. Spedding, R. Ellis, and D. Heath, London, 1857/74 (see the philosophical remarks by Freeman Dyson on page 11).

Stefan Banach (1892–1945), Oeuvres, 2 vol., PWN (Polish Scientific Publisher), Warszawa, 1967/79.

Jakob Bernoulli (1654–1705), Opera omnia, 2 vol., Genf 1744 (Switzerland).

Johann Bernoulli (1667–1748), Opera omnia, 4 vol., Lausanne 1742 (Switzerland).

Hans Bethe (1906–2005), Basic Bethe: Seminal Articles on Nuclear Physics 1936–1937, edited by R. Bacher and M. Livingstone, American Institute of Physics, New York, 1986.

George David Birkhoff (1884–1944), Collected Mathematical Papers, 3 vol., American Mathematical Society, New York, 1950.

Garrett Birkhoff (1911–1986), Selected Papers on Topology and Algebra, Birkhäuser, Boston, 1987.

Nikolai Nikolaevich Bogolyubov (Bogoliubov) (1909–1992), Selected Works. Part I: Dynamical Theory; Part II: Quantum and Classical Statistical Mechanics; Part III: Nonlinear Mechanics and Pure Mathematics; Part IV: Quantum Field Theory; Gordon and Breach, London, 1990/95.

Niels Bohr (1885–1962), Collected Works, 11 vol., North-Holland, Amsterdam, 1972.

Niels Bohr: A Centenary Volume. Edited by A. French and J. Kennedey, Harvard University Press, Cambridge, Massachusetts, 1985.

Andrei Andreevich Bolibruch, Yurii Sergeevich Osipov, and Yakov Grigor'evich Sinai (Eds.), Mathematical Events in the 20th Century, Springer, Berlin/Phasis, Moscow, 2006.
Ludwig Boltzmann (1844–1906), Wissenschaftliche Abhandlungen, edited by F. Hasenöhrl, Wien, 1909.
Armand Borel (1923–2003), Collected Papers, 4 vol., Springer, New York, 1983.
Max Born (1882–1970), My Life: Recollection's of a Nobel Laureate, Sribner, New York, 1977.
Max Born, Physics in My Generation, Springer, 1969.
Roul Bott (1923–2006), Collected Papers, 4 vol., Birkhäuser, Boston, 1994.
Nicolas Bourbaki, Eléments d'histoire des mathématiques, Hermann, Paris, 1969.
Richard Brennan, Heisenberg Probably Slept Here: The Lives, Times, and Ideas of the Great Physicists of the 20th Century, Wiley, New York, 1997.

The Cambridge Dictionary of Philosophy. Edited by R. Audi, Cambridge University Press, Cambridge, United Kingdom, 2005.
Georg Cantor (1845–1918), Gesammelte Abhandlungen, Springer, Berlin, 1932.
Constantin Carathéodory (1873–1950), Gesammelte mathematische Schriften, 5 vol., Beck, München, 1954.
Élie Cartan (1869–1951), Oeuvres complètes, 3 vol., Gauthier-Villars, Paris, 1952.
Henri Cartan, Oeuvres, 3 vol., Springer, New York, 1978.
Augustin-Louis Cauchy (1789–1857), Oeuvres complètes. 26 vol. (2 séries), Gauthier-Villars, 1888/1958.
Arthur Cayley (1821–1895), Collected Mathematical Papers, 14 vol., Cambridge, United Kingdom, 1889/98.
Subrahmanyan Chandrasekhar (1910–1995), Selected Papers, 7 vol., University of Chicago Press, Chicago, 1989/96.
Subrahmanyan Chandrasekhar, Newton's Principia for the Common Reader, Oxford University Press, Oxford, 1997.
Pafnuti L'vovich Chebyshev (1821–1894), Oeuvres, 2 vol., Saint Petersbourgh, 1899/1907.
Shiing-Shen Chern (1911–2004), A Mathematician and His Mathematical World: Selected Papers of S. Chern. Edited by S. Cheng, P. Li, and G. Tian, World Scientific, Singapore, 2000.
Shiing-Shen Chern and Friedrich Hirzebruch (Eds.), Wolf Prize in Mathematics, 2 vol., World Scientific, Singapore, 2001.

Richard Dedekind (1831–1916), Gesammelte mathematische Werke, 3 vol., Vieweg, Braunschweig, 1930.
René Descartes (Cartesius in Latin) (1596–1650), Oeuvres de Descartes, 12 vol., edited by C. Adam and P. Tannery, Paris, 1897/1913 (see the philosophical remarks by Freeman Dyson on page 11).
Keith Devlin, The Millennium Problems: The Seven Greatest Unsolved Mathematical Puzzles of Our Time, Basic Books, New York, 2002.
Jean Dieudonné (1906–1992), Choix d'oeuvres mathemématiques, 2 vol., Hermann, Paris, 1981.
Jean Dieudonné, History of Functional Analysis, 1900–1975, North-Holland, Amsterdam, 1983.
Jean Dieudonné, History of Algebraic Geometry, 400 B.C.-1985 A.C., Chapman, New York, 1985.
Jean Dieudonné, A History of Algebraic and Differential Topology, 1900–1960, Birkhäuser, Boston, 1989.
Diophantus (ca 250 B.C.), Diophantus of Alexandria, by T. Heath, Dover, 1964.

Paul Adrien Maurice Dirac (1902–1984), The Collected Works of P. A. M. Dirac. Edited by R. Dalitz, Cambridge University Press, Cambridge, United Kingdom, 1995.
Paul Dirac, The Principles of Quantum Mechanics, Clarendon Press, Oxford, 1930 (the classic monograph on quantum mechanics).
Paul Dirac, The Development of Quantum Mechanics, Gordon and Breach, New York, 1970.
Freeman Dyson, Selected Papers of Freeman Dyson, American Mathematical Society, Providence, Rhode Island, 1996 (see also the books and essays written by Freeman Dyson quoted on page 7).

Albert Einstein (1879–1955), Collected Papers, 3 vol., Princeton University Press, Princeton, New Jersey, 1993.
Albert Einstein (1905), Einstein's Miraculous Year 1905: Five Papers that Changed the Universe, Princeton University Press, Princeton, New Jersey, 1998. English translation of Einstein's five papers edited with an introduction by J. Stachel.
Einstein: A Centenary Volume. Edited by A. French, Harvard University Press, Cambridge, Massachusetts, 1979.
Encyclopedia Britannica, 32 vol., Chicago, 1987ff.
Paul Erdös (1913–1996), Paul Erdös and His Mathematics, 2 vol., edited by G. Halasz, L. Lovasz, M. Simonovits, and V. Sos, Springer, Heidelberg, 2002.
Euclid (ca. 360–290 B.C.), The Thirteen Books of Euclid's Elements, edited by T. Heath, 3 vol., Dover, 1956.
Leonhard Euler (1707–1783), Opera omnia, 72 vol., Leipzig/Berlin, later Basel–Zürich, 1911ff. There will also appear 15 volumes containing letters. See also Varadarajan (2006).

Ludwig Dimitrievich Faddeev, 40 Years in Mathematical Physics, World Scientific, Singapore, 1995.
Michael Faraday (1791-1867), The Selected Correspondence of Michael Faraday, edited by L. Williams, Cambridge University Press, New York, 1971.
Enrico Fermi (1901–1954), Collected Papers, 2 vol., Chicago University Press, Chicago, 1962.
Richard Feynman (1918–1988), Selected Papers of Richard Feynman, edited by L. Brown, World Scientific, Singapore, 2000.
Vladimir Aleksandrovich Fok (Fock) (1898–1974), Selected Works: Quantum Mechanics and Quantum Field Theory, edited by L. Faddeev et al., Chapman, Boca Raton, Florida, 2004.
Jean-Baptiste-Joseph Fourier (1768–1830), Oeuvres, 2 vol., Paris, 1888/90.
Ivar Fredholm (1866–1927), Oeuvres complètes publiées par l'Institut Mittag-Leffler, Malmö (Sweden), 1955.
Kurt-Otto Friedrichs (1901–1982), Selecta, 2 vol., edited by Cathleen S. Morawetz, Birkhäuser, Boston, 1986.
Harald Fritzsch, Quarks, Penguin, London, 1992.
Georg Ferdinand Frobenius (1849–1917), Gesammelte Abhandlungen, 3 vol., Springer, Berlin, 1968.
Jörg Fröhlich, Non-Perturbative Quantum Field Theory: Mathematical Aspects and Applications, Selected Papers, World Scientific, Singapore, 1992.

Galileo Galilei (1564–1642), Le opere di Galileo, 20 vol., edited by A. Favaro, Firenze, Italia, 1890.
Evariste Galois (1811–1832), Ecrits et mémoires mathématiques d'Evariste Galois. Edition critiques intégrale de ses manuskrits et publications par R. Bourgne et J. Azra, Paris, 1962.
Carl Friedrich Gauß (1777–1855), Werke, 12 vol., Göttingen, 1863–1929.

Izrail Moisseevich Gel'fand, Collected Papers, 3 vol., Springer, New York, 1889.
Murray Gell-Mann, The Quark and the Jaguar: The Adventures in the Simple and the Complex, Freeman, New York, 1994.
Strange Beauty: Murray Gell-Mann and the Revolution in Twentieth Century Physics, by G. Johnson, A. Knopf, New York, 2000.
Ennio de Giorgi (1928–1996), Selected Papers, edited by L. Caffarelli et al., Springer, Berlin, 2006).
James Glimm and Arthur Jaffe, Quantum Field Theory and Statistical Mechanics: Expositions, Birkhäuser, Boston, 1985.
Kurt Gödel (1906–1978), Collected Works, Oxford University Press, 1986/2003.
Hermann Grassmann (1809–1877), Mathematische und physikalische Werke, 3 vol., Leipzig, 1894/1911.
George Green (1793–1841), Mathematical Papers, Hermann, Paris, 1903.
Brian Greene, The Elegant Universe: Supersymmetric Strings, Hidden Dimensions, and the Quest for the Ultimate Theory, Norton, New York, 1999.
Alexandre Grothendieck, Récoltes et Semailles: réflexions et témoignage sur un passé de mathématicien, 1986 (ca. 1000 pages in French). (Reaping and Sowing: the life of a mathematician – reflections and bearing witness.) Internet: `http://www.fermentmagazine.org/rands/recoltes1.html`
Alexandre Grothendieck, The Grothendieck Festschrift, 3 vol., edited by P. Cartier et al., Birkhäuser, Boston, 1990.

Rudolph Haag – eighty years, by Daniel Kastler, Commun. Math. Phys. **237** (2003), 3-6.
Jaques Hadamard (1865–1963), Oeuvres, 4 vol., Paris, 1968.
William Rowan Hamilton (1805–1865), The Mathematical Papers, 3 vol., Cambridge, United Kingdom, 1931/67.
Harenberg Lexikon der Nobelpreisträger (Encyclopedia of Nobel Prize Laureates), Harenberg Lexikon Verlag, Dortmund, 2000 (in German).
Felix Hausdorff (1868–1942), Collected Mathematical Papers, 7 vol., edited by E. Brieskorn, Springer, Berlin, 2001ff.
Werner Heisenberg (1901–1976), Gesammelte Abhandlungen. Edited by W. Blum. Serie A: 3 vol., Serie B: 1 vol., Serie C: 5 vol., Springer, Berlin, 1984.
David Hilbert (1862–1943), Gesammelte Werke, 3 vol., 12th edn., Springer, Berlin, 1977.
David Hilbert (1912), Foundations of the Theory of Integral Equations, Teubner, Leipzig, 1912 (in German). (This book contains six fundamental articles written by Hilbert; it represents a supplement to his "Collected Works.")
Friedrich Hirzebruch, Gesammelte Abhandlungen, Springer, Berlin, 1987.
Hooft: See 't Hooft.
Eberhard Hopf (1902–1983), Selected Works of Eberhard Hopf: with commentaries. Edited by C. Morawetz, J. Serrin, and Ya. Sinai, Amer. Math. Soc., Providence, Rhode Island, 2002.
Heinz Hopf (1894–1971), Selecta, Springer, Berlin, 1964.
Lars Hörmander, The Analysis of Linear Partial Differential Operators, 4 vol., Springer, New York, 1983.
Lars Hörmander, by J. O'Connor and E. Robertson. Internet: `http://www-groups.dcs.st-and.ac.uk~history/Mathematicians/Hormander.html`

Carl Gustav Jakob Jacobi (1804–1851), Gesammelte Abhandlungen, 7 vol., Reimer, Berlin 1881/91.
Fritz John (1910–1994), Collected Papers, 2 vol., edited by J. Moser, Birkhäuser, Boston, 1985.

Res Jost (1918–1990), Das Märchen vom Elfenbeinernen Turm: Reden und Aufsätze. (The Fairy Tale about the Ivory Tower: essays and lectures in German and partly in English), edited by K. Hepp, W. Hunziker, and W. Kohn, Springer, Berlin, 1995.

Erich Kähler (1906–2000), Mathematical Works. Edited by R. Berndt and O. Riemenschneider, de Gruyter, Berlin, 2004.

Johannes Kepler (1571–1630), Gesammelte Werke, 22 vol., München 1938/59.

Felix Klein (1849–1925), Gesammelte mathematische Abhandlungen, 3 vol., Springer, Berlin, 1921/23.

Felix Klein, Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert (in German), 2 vol., Springer, Berlin. (English edition: Development of Mathematics in the 19th Century, with a large appendix by R. Hermann, Math. Sci. Press, New York, 1979).

Morris Kline, Mathematical Thought from Ancient to Modern Times, 3 vol., Oxford University Press, New York 1990.

Kunihiko Kodaira (1915–1997), Collected Works, Princeton University Press, Princeton, New Jersey, 1975.

Andrei Nikolaevich Kolmogorov (1903–1987), Selected Works, 3 vol., Kluwer Dordrecht, 1991/93.

Hendrik Kramers, Collected Scientific Papers, North-Holland, Amsterdam, 1956.

Leopold Kronecker (1823–1891), Werke, 5 vol., Leipzig/Berlin, 1895/1930.

Joseph Louis Lagrange (1736–1813), Oeuvres, 14 vol., Gauthier-Villars, Paris, 1867/92. Reprinted by Georg Olms, Hildesheim (Germany)/New York, 1973.

Lev Davidovich Landau (1908–1968), Collected Papers, Pergamon Press, Oxford, 1965.

Robert Langlands, The Langlands Archive. (The seminal Langlands program in number theory and representation theory is discussed in Varadarajan (2006)). Internet: www.sunsite.ubc.ca/DigitalMathArchive/Langlands/intro.htmi

Pierre Simon Laplace (1749–1827), Oeuvres, 14 vol., Gauthier-Villars, Paris, 1878/1912.

Peter Lax, Selected Papers, 2 vol., Springer, New York, 2005.

Henri Lebesgue (1875–1941), Oeuvres scientifiques, 5 vol., Enseign. Math. Genève, 1971/73.

Gottfried Wilhelm Leibniz (1646–1716), Sämtliche Schriften und Briefe (Collected works and letters), Parts I–XIII, 35 vol., Berlin 1923ff.

Gottfried Wilhelm Leibniz, Mathematische Schriften, 3 vol., Akademie-Verlag, Berlin, 1990/2003.

Leonardo da Vinci (1452–1519), Il manoscritti di Leonardo da Vinci e le loro edizioni. Edited by A. Marinoni. In: Comitato Nazionale per le orance a Leonardo da Vinci nel quinto centenario della nascita, Roma (Italia), 1952.

Jean Leray (1906–1998), Oeuvres scientifiques, 3 vol., edited by A. Borel, Springer, Berlin/Société mathématique de France, Paris, 1998.

Jacques-Louis Lions (1928–2001), Oeuvres choisies de Jacques-Louis Lions, 3 vol., EDP Sciences, Paris, 2003.

Tullio Levi-Civita (1873–1941), Opere matematiche, 4 vol., Bologna, 1954/60.

Eugenio Elia Levi (1883–1917), Opere, 2 vol., Cremonese, Roma, 1959/60.

Sophus Lie (1842–1899), Gesammelte Abhandlungen, Leipzig/Oslo, 1922/60.

Elliott Lieb, The Stability of Matter: From Atoms to Stars, Selecta of Elliott Lieb, edited by W. Thirring, Springer, New York, 2002.

Elliott Lieb, Inequalities: Selecta of Elliott Lieb, edited by M. Loss, Springer, New York, 2002.

Aleksandr Mikhailovich Lyapunov (1857–1918), Collected Works, 5 vol., Nauka, Moscow, 1954 (in Russian).

Yurii Ivanovich Manin, Selected Papers, World Scientific, Singapore, 1996.
Krysztof Maurin, The Riemann Legacy: Riemannian Ideas in Mathematics and Physics of the 20th Century, Kluwer, Dordrecht, 1997.
James Clerk Maxwell (1831–1879), The Scientific Letters and Papers of James Clerk Maxwell. Edited by P. Harmann, Cambridge University Press, New York, 1995.
John Milnor, Collected Papers, 3 vol., American Mathematical Society, Providence, Rhode Island, 2007.
Hermann Minkowski (1864–1909), Gesammelte Abhandlungen, 2 vol., Teubner, Leipzig, 1911.
Mikhail Monastirsky, Modern Mathematics in the Light of the Fields Medals, Peters, Wellersley, Massachusetts, 1997.
Marston Morse (1892–1977), Collected Papers, 6 vol., World Scientific, Singapore, 1987.
Jürgen Moser (1928–1999), Obituary on Jürgen Moser, by Peter Lax. In: Ergodic Theory and Dynamical Systems **22** (2002), 1337–1342.

Neumann: See von Neumann.
Isaac Newton (1643–1727), The Mathematical Papers of Isaac Newton. Edited by D. Whiteside, Cambridge University Press, Cambridge, United Kingdom, 1967.
Isaac Newton (1687), Philosophiae naturalis principia mathematica (Mathematical principles of natural philosophy) (in Latin). Translated into English by A. Motte, in 1729. Edited by F. Cajori, University of California Press, Berkeley, California, 1946. See also Chandrasekhar (1997).
Nobel Prize Laureates: Internet: `www.nobel.se/physics/laureates/index.html` (see also Harenberg Lexikon (2000)).
Nobel Prize Lectures, Nobel Foundation, Stockholm, 1954ff.
Amalie Emmy Noether (1882–1935), Gesammelte Abhandlungen, Springer, Berlin, 1982.

Piergiorgio Odifreddi, The Mathematical Century: The 30 Greatest Problems of the Last 100 Years, Princeton University Press, Princeton, New Jersey, 2004.

Paul Painlevé (1863–1933), Oeuvres, 3 vol., SNRS, Paris, 1976.
Wolfgang Pauli (1900–1958), Collected Scientific Papers. Edited by R. Kronig and V. Weisskopf, Interscience, New York, 1964.
Wolfgang Pauli, Scientific Correspondence with Bohr, Einstein, Heisenberg, and Others, 4 vol., Springer, New York, 1993.
Wolfgang Pauli, Writings on Physics and Philosophy, edited by C. Enzand and K. von Meyenn, Springer, New York, 1994.
Roger Penrose, Collected Works, Oxford University Press, Oxford, United Kingdom, 2008.
The Physical Review: The First Hundred Years – A Selection of Seminal Papers and Commentaries. Edited by H. Stroke, American Institute of Physics, New York, 1995 (14 survey articles on general developments, 200 fundamental articles, 800 additional articles on CD).
Max Planck (1958–1943), Physikalische Abhandlungen und Vorträge (Collected Works), 3 vol., Vieweg, Braunschweig, 1958.
Max Planck, Vorträge, Reden, Erinnerungen (Lectures, speeches, and recollections), edited by H. Roos, Springer, Berlin, 2001.

Plato (427–347 B.C.), Plato's Philosophy of Mathematics, by A. Wedberg. Almquist and Wiksell, 1955.
Henri Poincaré (1854–1912), Oeuvres, 11 vol., Paris, 1916/56.
Lev Semyonovich Pontryagin (1908–1988), Selected Works, 4 vol., Gordon and Breach, New York, 1986.

Srinivasa Ramanujan (1887–1920), Collected Papers of Srinivasa Ramanujan, edited by G. Hardy et al., Cambridge University, 1927.
Srinivasa Ramanujan, Ramanujan's Notebooks, 2 parts, edited by B. Berndt, Springer, New York, 1987.
Srinivasa Ramanujan, The Man Who Knew Infinity: A Life of the Genius Ramanujan, by R. Kanigel, Charles Scribner's Sons, New York, 1993.
Edward Regis, Who Got Einstein's Office? Eccentricity and Genius at the Institute for Advanced Study in Princeton, Addison-Wesley, Reading, Massachusetts, 1989.
Constance Reid, Hilbert (1862–1943), Springer, New York, 1970.
Constance Reid, Courant (1888–1972): the Life of an Improbable Mathematician, Springer, New York, 1976.
Bernhard Riemann (1826–1866), Collected Mathematical Works with Commentaries, Springer, New York/Teubner, Leipzig, 1990.
Frédéric (Fryges) Riesz (1880–1956), Oeuvres complètes, 2 vol., Akad. Kiadó, Budapest, 1960.
Marcel Riesz (1886–1969), Collected Papers, Springer, Berlin. 1988.
Gian-Carlo Rota (1933–1989), Gian-Carlo Rota on Combinatorics: Introductory Papers and Commentaries. Edited by J. Kung, Birkhäuser, Basel, 1995.

Abdus Salam (1926–1996), Selected Papers, World Scientific, Singapore, 1994.
Juliusz Pawel Schauder (1899–1943), Oeuvres, PWN (Polish Scientifique Publisher), Warszawa, 1978.
Erwin Schrödinger, Gesammelte Abhandlungen, 4 vol., Vieweg, Braunschweig, 1984.
Issai Schur (1875–1941), Gesammelte Abhandlungen, 3 vol., Springer, Berlin, 1973.
Laurent Schwartz (1915–2002), A Mathematician Grappling with his Century, Birkhäuser, Basel, 2001.
Hermann Amandus Schwarz (1843–1923), Gesammelte mathematische Abhandlungen, 2 vol., Berlin, 1890.
John Schwarz (Ed.), Superstrings: The First 15 Years of Superstring Theory, 2 vol., World Scientific, Singapore, 1985.
Silvan Schweber, QED and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga. Princeton University Press, Princeton, New Jersey, 1994.
Julian Schwinger (1918–1994), Selected Papers, edited by M. Flato, C. Fronsdale, and K. Milton, Reidel, Dordrecht, 1979.
Julian Schwinger (Ed.), Quantum Electrodynamics: 34 Selected Articles, Dover, New York, 1958.
Christoph Scriba and Peter Schreiber, 5000 Years of Geometry: History, Human Cultures, People, Springer, Berlin, 2005 (in German).
Jean–Piere Serre, Oeuvres, 4 vol., Springer, New York, 1986.
Francesco Severi (1879–1961), Opere matematiche, 3 vol., Accad. Nazionale dei Lincei, Roma.
Igor Rostislavovich Shafarevich, Collected Mathematical Works, Springer, Berlin, 1989.
Stephen Smale, Collected Papers of Stephen Smale, 3 vol., Singapore University Press, Singapore, 2000.
Sergei L'vovich Sobolev (1908–1989), Selected Works, Springer, New York, 2006.

Arnold Sommerfeld (1868–1951), Gesammelte Schriften, 4 vol., Vieweg, Braunschweig, 1968.

Walter Thirring, Selected Papers, American Mathematical Society, Providence, Rhode Island, 1998.
René Thom (1923–2002), 'Mathématicien et apprenti philosophe', by J. Bourguignon, Bull. Amer. Math. Soc. **41**(3) (2004), 273–274.
René Thom, The impact of Thom's cobordism theory, by M. Atiyah, Bull. Amer. Math. Soc. 41 (2004)(3), 337–340.
Sin-Itiro Tomonaga (1906–1979), Scientific Papers. Edited by T. Miyazima, Tokyo, Japan. 1971.
Gerardus 't Hooft, In Search for the Ultimate Building Blocks, Cambridge University Press, 1996.
Alan Turing (1912–1954), The Collected Works of A. M. Turing, 4 vol., North-Holland, Amsterdam.

Bartel Leendert van der Waerden (Ed.) (1903–1998), Sources of Quantum Mechanics (1917–1926), Dover, New York, 1968.
Bartel Leendert van der Waerden, A History of Algebra: From al-Khwarimzi to Emmy Noether, Springer, New York, 1984.
Veravalli Varadarajan, Euler through Time: A New Look at Old Themes, American Mathematical Society, Providence, Rhode Island, 2006.
Martinus Veltman, Facts and Mysteries in Elementary Particle Physics, World Scientific, Singapore, 2003.
Vito Volterra (1860–1940), Opere matematiche, 5 vol., Acc. dei Lincei, Roma, 1954/60.
John von Neumann (1903–1957), Collected Papers, 5 vol. Pergamon Press, New York, 1961.

Waerden: See van der Waerden.
Karl Weierstraß (1815–1897), Mathematische Werke, 7 vol., Mayer and Müller, Berlin/Akademische Verlagsgesellschaft, Berlin, 1894/1927.
André Weil (1906–1996), Oeuvres scientifiques – Collected Papers, 3 vol., Springer, New York, 1979/80.
Stephen Weinberg, Quantum Field Theory, 3 vol., Cambridge University Press, 1995.
Stephen Weinberg, Cosmology, Oxford University Press, 2008.
Hermann Weyl (1885–1955), Gesammelte Werke, 4 vol., edited by K. Chandrasekharan, Springer, New York, 1968 (see also Yang (1985)).
John Archibald Wheeler (1911–2008), Geons, Black Holes, and Quantum Foam: a Life in Physics: Autobiography and Memoir, coauthored by K. Ford, Norton, New York, 1998.
Norbert Wiener (1894–1964), Collected Works, 4 vol., MIT Press, Cambridge, Massachusetts, 1976.
Eugene Wigner (1902–1995), Collected Works, 8 vol., edited by A. Wightman and J. Mehra, Springer, New York, 1993.
Eugene Wigner, Philosophical Reflections and Syntheses. Annotated by G. Emch. Edited by J. Mehra and A. Wightman, Springer, New York, 1995.
Hans Wußing, 6000 Years of Mathematics: a Cultural Journey through Time, 2 vol., Springer, Heidelberg, 2008 (in German).

Benjamin Yandell, The Honors Class: Hilbert's Problems and Their Solvers, Peters Ltd, Natick, Massachusetts, 2001.
Chen Ning Yang, Selected Papers, 1945–1980, Freeman, New York, 1983.

Chen Ning Yang, Hermann Weyl's Contributions to Physics, pp. 7–21. In: Hermann Weyl (1885–1955), Centenary Lectures, edited by K. Chandrasekharan, Springer, Berlin, 1985.

Kôsaku Yosida (1909–1990), Collected Papers, Tokyo, 1992.

Yakov Borisovich Zel'dovich (1914–1987), Selected Works, 2 vol., Princeton University Press, Princeton, New Jersey, 1991.

Many biographies of mathematicians can be found on the website of the University of St. Andrews in Scotland:

`http://www-groups.dcs.st-and.ac.uk~history/Biographies`

Summary

The complete list of all the papers quoted in this volume can be found on the author's homepage:

`http://www.mis.mpg.de/zeidler/qft.htmls`

List of Symbols

$f(x) := x^2$ (definition of f)
$f(x) \simeq g(x),\ x \to a$ (asymptotic equality); this means $\lim_{x\to a} \frac{f(x)}{g(x)} = 1$
$f(x) = o(g(x)),\ x \to a$ (Landau symbol); this means $\lim_{x\to a} \frac{f(x)}{g(x)} = 0$
$f(x) = O(g(x)),\ x \to a$, Vol. I, p. 933
$f(x) \sim \sum_{n=0}^{\infty} a_n (x-a)^n,\ x \to a$ (asymptotic expansion), 99
$f(x) \sim \sum_{n=0}^{\infty} \frac{a_n}{x^n},\ x \to +\infty$ (asymptotic expansion), 99

$\operatorname{sgn}(a)$ (sign of the real number a), $\operatorname{sgn} \pm 1 = \pm 1$, $\operatorname{sgn} 0 = 0$
$[a, b],\ ]a, b[,\ ]a, b]$ (intervals), Vol. I, p. 933
$\sum_{n=-\infty}^{\infty} b_n$, 22
δ_{ij} (Kronecker symbol), Vol. I, p. 933 $\delta_{11} := 1,\ \delta_{12} := 0$
$\delta^{ij} = \delta_{ij} = \delta^i_j$
$\delta_{\mathbf{pq}}$, Vol. I, p. 670
ε_{ij} (skew-symmetric symbol) $\varepsilon_{12} = -\varepsilon_{21} = 1,\ \varepsilon_{11} = \varepsilon_{22} = 0$, Vol. I, p. 335
$\operatorname{sgn}(\pi)$ (sign of the permutation π), 186
S_n (symmetric group), 186
A_n (alternating group), 186

x, y, z (right-handed Cartesian coordinates), 390
$\mathbf{i}, \mathbf{j}, \mathbf{k}$ (right-handed orthonormal basis)
$\mathbf{x} := x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ (position vector)
$|\mathbf{x}|$ (length of the vector $\mathbf{x}$)
$\mathbf{xy}$ (inner product of vectors), 365
$\mathbf{x} \times \mathbf{y}$, (vector product of vectors), 365
$|x| := \sqrt{|x_1|^2 + \ldots + |x_N|^2}$ (length of the vector $x = (x_1, \ldots, x_N)$ in $\mathbb{R}^N$ or in $\mathbb{C}^N$), 69
t (time)
$x^1 := x, x^2 := y, x^3 := z, x^0 := ct$ (space-time point in Minkowski space), Vol. I, p. 933
$\mu = 0, 1, 2, 3$ (indices for space-time variables in Minkowski space), Vol. I, p. 933
$j = 1, 2, 3$ (indices for spatial variables in Minkowski space), Vol. I, p. 933
$\eta_{\mu\nu}$ (Minkowski symbol), $\eta_{00} := 1,\ \eta_{11} := -1,\ \eta_{01} := 0$, Vol. I, p. 934
$\eta^{\mu\nu} = \eta_{\mu\nu} = \eta^{\nu}_{\mu}$, Vol. I, p. 933
$\epsilon^{\alpha\beta\gamma\delta}, \epsilon_{\alpha\beta\gamma\delta}$ (skew-symmetric symbol) $\epsilon_{0123} := 1,\ \epsilon_{1023} := -1$, Vol. I, p. 934
$a_\mu b^\mu := \sum_{\mu=0}^{3} a_\mu b^\mu$ (Einstein's convention in Minkowski space), Vol. 1, p. 934
$\operatorname{Ric}(g)$ (Ricci tensor), 327
$\zeta(s)$ (Riemann zeta function), 166
B_n (Bernoulli number), Vol. I, p. 278

$\theta(t)$ (Heaviside function), 20
$\theta_*(t)$ (modified Heaviside function), 456
e_λ (characteristic function of the open interval $]-\infty, \lambda[$), 497
$\delta(t)$ (Dirac delta function), Vol. I, p. 590
δ, δ_x (Dirac delta distribution), Vol. I, p. 609
δ_μ (Dirac delta function with respect to the measure μ), Vol. I, p. 602
$\delta_{\Delta t}(t)$ (standard discrete Dirac delta function), Vol. I, p. 580
$\delta_{\Delta^4 x}, \delta_T, \delta_{\text{dis}}$ (discrete Dirac delta functions), Vol. I, p. 441, p. 670
$\delta_{\mathcal{C}(L)}, \delta_{\mathcal{G}(N)}$ (discrete Dirac delta functions), 801
$\mathcal{C}(L)$ (cube in position space), 801
$\mathcal{G}(N)$ (grid in momentum space), 801
$\Delta^3 p$, 801
$\mathcal{V}$ (normalization volume), 801
$\delta(x^2 - a^2)$ (special distribution), Vol. I, p. 595

$\mathcal{P}\left(\frac{1}{x}\right)$ (special distribution), Vol. I, p. 618
$\mathcal{P}\left(\ln|x|\right)$ (special distribution), Vol. I, p. 735
$t_{\pm}^{\lambda}, |t|_{*}^{\lambda}$ (special distributions) 82
$(t \pm 0_{+}\mathrm{i})^{\lambda}$ (special distributions), 85

I, id (identity operator)
$f : X \to Y$ (map), Vol. I, p. 931
$f(U)$ (image of the set U), Vol. I, p. 931
$\mathrm{im}(f), R(f)$ (image (or range) of the map f), Vol. I, p. 931
$f^{-1}(V)$ (pre-image of the set V), Vol. I, p. 931
$\mathrm{dom}(f), D(f)$ (domain of f), Vol. I, p. 931
$f^{-1} : Y \to X$ (inverse map), Vol. I, p. 931
$\mathrm{graph}(f)$ (graph of f), 208
$\ker(f)$ (kernel of the map f), 184

$x \in U$ (the point x is an element of U)
$U \subseteq V$ (U is a subset of V)
$U \subset V$ (U is a proper subset of V), Vol. I, p. 931
$\{x : x$ has the property $\mathcal{P}\}$ (this denotes the class of all things which have the property $\mathcal{P}$)
$U \cup V$ (the union of two given sets U and V)
$U \cap V$ (the intersection of two given sets U and V)
$U \setminus V$ (the difference of two sets U and V, i.e., the set of elements of U not belonging to V)
∂U (boundary of the set U)
$\mathrm{int}(U)$ (interior of U)
$\mathrm{cl}(U) \equiv U \cup \partial U$ (closure of U)
$\emptyset$ (empty set)
2^S (power set of S), 258
$\sup(S)$ (supremum of the set S), 237
$\inf(S)$ (infimum of the set S), 237
$\bigcup_{\alpha\in\mathcal{A}} X_\alpha$ (union of a family of sets X_α)
$\bigcap_{\alpha\in\mathcal{A}} X_\alpha$ (intersection)
$\prod_{\alpha\in\mathcal{A}} X_\alpha$ (product of sets), 118
$\coprod_{\alpha\in\mathcal{A}} X_\alpha$ (coproduct of sets), 119
$\mathrm{meas}(S)$ (measure of the set S)
$\mathrm{meas}(\mathbb{S}^N)$, 69
$\mathrm{lim\,ind}_{n\to\infty} X_n$ (inductive limit), 243
$\mathrm{lim\,proj}_{n\to\infty} X_n$ (projective limit), 243
$\lim_{n\to\infty} \varphi_n = \varphi$ (limit in a Hilbert space), Vol. I, p. 337
$w - \lim \varphi_n = \varphi$ (weak limit), 501
$f(+0)$ (limit from the right), 732
$f(-0)$ (limit from the left), 732
$A - \lim_{t\to+\infty} f(t)$ (Abel limit), 754
$w - \lim_{n\to\infty} A_n = A$ (weak operator limit), 655
$s - \lim_{n\to\infty} A_n = A$ (strong operator limit), 656
$u - \lim_{n\to\infty} A_n = A$ (uniform operator limit), 656

$z = x + y\mathrm{i}$ (complex number)
$\Re(z) := x$ (real part of z)
$\Im(z) := y$ (imaginary part of z)
$z^{\dagger} := x - y\mathrm{i}$ (conjugate complex number) (synonymously, $\bar{z} \equiv z^{\dagger}$)
$|z|$ (modulus of z), Vol. I, p. 209
$\arg(z)$ (principal argument of z), $-\pi < \arg(z) \le \pi$, Vol. I, p. 209
$\arg_{*}(z)$ (argument of z), Vol. I, p. 209
$\sqrt{z} := \sqrt{|z|}\,\mathrm{e}^{\frac{1}{2}\mathrm{i}\arg(z)}$ (principal value of the square root of z), 480
$\ln z := \ln|z| + \mathrm{i}\arg(z)$ (principal value of the logarithm of z), 764
z^{λ} (principal value), 85
$\partial_x := \frac{\partial}{\partial x}, \partial_y := \frac{\partial}{\partial y}$, 318
$\partial_z = \frac{\partial}{\partial z} := \frac{1}{2}(\partial_x - \mathrm{i}\partial_y)$,
$\partial_{\bar{z}} = \frac{\partial}{\partial \bar{z}} := \frac{1}{2}(\partial_x + \mathrm{i}\partial_y)$, 318, 616
$dz, d\bar{z}$, 318
$\partial := \partial_z \cdot dz$, $\bar{\partial} := \partial_{\bar{z}} \cdot d\bar{z}$, 318, 323
$\mathrm{res}_z(f)$ (residue of the function f at the point z), Vol. I, p. 213

$\mathbb{R}$ (set of real numbers)
$\mathbb{C}$ (set of complex numbers; Gaussian plane)
$\overline{\mathbb{C}}$ (closed Gaussian plane), 195
$\mathbb{C}_{>}$ (open upper half-plane), 764
$\mathbb{C}_{\ge}$ (closed upper half-plane), 764
$\mathbb{K} = \mathbb{R}, \mathbb{C}$ (set of real or complex numbers)
$\mathbb{Z}$ (set of integers, $0, \pm1, \pm2, \ldots$)
$\mathbb{N}$ (set of natural numbers, $0, 1, 2, \ldots$)
$\mathbb{Q}$ (set of rational numbers)
$\mathbb{R}^N, \mathbb{C}^N, \mathbb{K}^N$ ($N = 1, 2, \ldots$), Vol. I, p. 328
$\mathbb{H}$ (skew-field of quaternions), 179
$\mathbb{Z}/\mathrm{mod}\ m$ (or $\mathbb{Z}_m$) (Gaussian ring of integers modulo m), 180
E^3 (3-dimensional real Hilbert space)
$\mathbb{E}^3$ (Euclidean manifold) 364
$T_P\mathbb{E}^3$ (tangent space of the Euclidean

manifold at the point P, which is isomorphic to E^3), 364
M^4 (4-dimensional real linear Minkowski space with indefinite inner product) (see Vol. III)
$\mathbb{M}^4$ (Minkowski manifold), Vol. I, p. 769
$T_P\mathbb{M}^4$ (tangent space of the Minkowski manifold at the point P, which is isomorphic to M^4)

$\mathbb{R}^\times$ (set of nonzero real numbers)
$\mathbb{N}^\times$ (set of nonzero natural numbers, $1, 2, 3, \ldots$)
$\mathbb{C}^\times$ (set of nonzero complex numbers)
$\mathbb{K}^\times$ (set of nonzero numbers in $\mathbb{K}$)
$\mathbb{R}_\geq$ (set of nonnegative real numbers, $x \geq 0$)
$\mathbb{R}_>$ (set of positive real numbers, $x > 0$)
$\mathbb{R}_\leq$ (set of non-positive real numbers, $x \leq 0$)
$\mathbb{R}_<$ (set of negative real numbers, $x < 0$)
$\mathbb{R}_+$ (additive semigroup of nonnegative real numbers)
$\mathbb{R}_+^\times$ (multiplicative group of positive real numbers)
$*\mathbb{R}$ (set of generalized real numbers), 248

$\mathbb{B}^2$ (closed unit disc)
$\mathrm{int}(\mathbb{B}^2)$ (open unit disc)
$\mathbb{S}^1 \equiv \partial\mathbb{B}^2$ (unit circle)
$\mathbb{B}^{n+1}$ ($(n+1)$-dimensional closed unit ball), 194
$\mathrm{int}(\mathbb{B}^{n+1})$ ($(n+1)$-dimensional open unit ball), 194
$\mathbb{S}^n$ (n-dimensional unit sphere), 194
$\mathbb{B}_R^{n+1}(a)$ ($(n+1)$-dimensional closed unit sphere of radius R centered at the point a), 194
$\mathbb{S}_R^n \equiv \partial\mathbb{B}_R^{n+1}(0)$ (n-dimensional sphere of radius R), 194
$\mathbb{P}^n$ (n-dimensional real projective space), 203
$\mathbb{P}_\mathbb{C}^n$ (n-dimensional complex projective space), 204
$X^\times$ (set of nonzero elements of the Hilbert space X)
$\mathbb{B}(X)$ (closed unit ball in X), 199
$\mathrm{int}(\mathbb{B}(X))$ (open unit ball in X), 199
$\mathbb{S}(X)$ (unit sphere of X), 199
$\mathbb{G}_m(X)$ (space of m-dimensional linear subspaces of X), 199
$\mathbb{P}(X)$ (projective space generated by the Hilbert space X), 199
$L^\perp$ (orthogonal complement to L in a Hilbert space), 635

$[x]$ (equivalence class of the representative x), 199
$X/\sim$ (space of equivalence classes), 199
X/G (orbit space), 197
$\mathrm{stab}(x_0)$ (stabilizer), 197

$\bar{A}$ (mean value), Vol. I, p. 351
ΔA (mean fluctuation), Vol. I, p. 351

$\dim X$ (dimension of the linear space X), Vol. I, p. 330
$\mathrm{codim}(X)$ (codimension), 189
$\mathrm{span}(S)$ (linear hull of the set S), Vol. I, p. 329
$\langle x|y\rangle$ (inner product), Vol. I, p. 336
$\langle\varphi|, |\psi\rangle$ (Dirac calculus), Vol. I, p. 359
$||\varphi||$ (norm), Vol. I, p. 336
$L(X, Y)$ (space of linear operators from X into Y)
X^d (dual space), 119
$A^\dagger$ (adjoint operator on a linear space), Vol. I, p. 341
$A^\dagger$ (formally adjoint operator on a Hilbert space), 683
A^* (adjoint operator on a Hilbert space), 683
A^d (dual operator), 119
A^{-1} (inverse operator), Vol. I, p. 931
A^c (conjugate complex operator); this means $(A^\dagger)^d$
I (unit matrix), 766
$\mathrm{diag}(a, b)$ (diagonal matrix), 766
$\mathcal{A}^\dagger$ (adjoint matrix), 766
$\mathcal{A}^d$ (dual or transposed matrix), 766
$\mathcal{A}^c$ (conjugate complex matrix), 766
$\mathcal{A}^{-1}$ (inverse matrix), 766
$[A, B]_- := AB - BA$,
$[A, B]_+ := AB + BA$,
$\mathrm{tr}(A)$ (trace), 629, 766
$\det(A)$ (determinant), 563,766
e^A (exponential function), 769
$\ln A$ (logarithmic function), 769
$\sigma(A)$ (spectrum), 766
$\varrho(A) = \mathbb{C} \setminus \sigma(A)$ (resolvent set), 766
$b^1, \ldots, b^n$ (cobasis to the basis $b_1, \ldots b_n$), 119
$dx^k \equiv b^k$
$M_k(X)$ (linear space of k-linear

$\int f(\lambda)\, dE_\lambda$ (Hilbert–von Neumann, spectral integral), Vol. I, p. 37, p. 369
$PV \int_{-\infty}^{\infty} f(x)\, dx$ (principal value), 495
$\int F(q)\, d\mu(q)$, $\int F(\varphi)\, \mathcal{D}\varphi$ (functional integral), Vol. I, p. 416, p. 442, p. 753
$\int \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}q$ (Feynman path integral in the position space), 559,
$\int \mathrm{e}^{S[q,p]/\hbar}\, \mathcal{D}q\mathcal{D}p$ (Feynman path integral in the phase space), 556
$\int_{D(A)} \mathrm{e}^{-\frac{1}{2}\langle\varphi|A\varphi\rangle}\, \mathcal{D}_G\varphi$ (normalized infinite-dimensional Gaussian integral), 572
$\int \mathrm{e}^{\mathrm{i}S[q]/\hbar}\, \mathcal{D}_G q$ (normalized infinite-dimensional Gaussian integral), 575
$\mathit{\Pi}_{\Delta t}(q_{\mathrm{in}}, q_{\mathrm{out}})$ (space of curves), Vol. I, p. 420

$\mathrm{zero}(f)$ (set of zeros of the function f), Vol. I, p. 608
$\mathrm{supp}(f)$ (support of the function f), Vol. I, p. 608
$\mathrm{supp}(F)$ (support of the distribution F), Vol. I, p. 610
$\mathrm{supp}(\mu)$ (support of the measure μ), Vol. I, p. 602
$\mathrm{sing\ supp}(F)$ (singular support of the distribution F), Vol. I, p. 705
Char(L) (characteristic set of the differential operator L), Vol. I, p. 711
$WF(G)$ (wave front set of the distribution G), Vol. I, p. 710
$\mathcal{F}g, \hat{g}$ (Fourier transform of the function g), 512, 514
$\mathcal{F}G$ (Fourier transform of the distribution G), 514
$\mathcal{F}_\hbar g, \tilde{g}$ (rescaled Fourier transform), 533
$\mathcal{F}_M g$ (Fourier–Minkowski transform of the function g), 865 (see Vol. I, p. 772)
$\mathcal{L}g$ (Laplace transform of g), Vol. I, p. 92
$f * g$ (convolution of two functions), Vol. I, p. 93, p. 534
$F * G$ (convolution of two distributions), Vol. I, p. 616
$f \otimes g$ (tensor product of two functions), Vol. I, p. 616
$F \otimes G$ (tensor product of two distributions), Vol. I, p. 616
$x \otimes y$ (tensor product of two elements of linear spaces or algebras), 121
$a * b$ (Moyal star product of two functions), 591, 607
$\mathcal{LE}(\mathbb{R}^n)$ (Hopf algebra of linear differential operators), 123
$\mathbf{1}$ (unit element of an algebra), 116
$\Delta a = \sum_a a_{(1)} \otimes a_{(2)}$ (Sweedler notation for the coproduct Δ), 129
μ (multiplication map), 128
η (unitality map), 128
ε (counitality map), 128

$C[a, b]$ (space of continuous functions), Vol. I, p. 366
$C^1[a, b]$ (space of continuously differentiable functions), Vol. I, p. 550
$C^\infty(\Omega) \equiv \mathcal{E}(\Omega)$ (space of smooth functions), Vol. I, p. 543
$C^\infty(\overline{\Omega})$, Vol. I, p. 543
$C_0^\infty(\Omega) \equiv \mathcal{D}(\Omega)$, Vol. I, p. 543
$C^\infty(\mathbb{R}^n, \mathbb{R})$ (space of smooth functions $f : \mathbb{R}^n \to \mathbb{R}$),
$C^\alpha(\Omega), C^{k,\alpha}(\Omega)$ (Hölder spaces), Vol. I, p. 554
$C^{0,1}(\Omega), C^{k,1}(\Omega)$ (Lipschitz spaces), Vol. I, p. 554
$C^k[0, 1]$ (function space), 689
$C\{s, t\}$ (function space), 549
$C_0^2[s, t]$ (function space), 550
$L_2(\mathbb{R}), L_2(\Omega)$ (Lebesgue spaces), 432 (see also Vol. I, p. 531)
$L_2(-\pi, \pi)$, Vol. I, p. 533
$L_{\mathrm{loc}}(\mathbb{R}^N)$, Vol. I, p. 609
l_2 (classical Hilbert space), Vol. I, p. 534
$L_2(\mathcal{M})$ (discrete Lebesgue space) Vol. I, p. 441
$W_2^1(\Omega)$, $\overset{\circ}{W}{}_2^1(\Omega)$ (Sobolev spaces), Vol. I, p. 557
$W_2^k(\Omega)$, Vol. I, p. 557
$W_2^{1/2}(\Omega)$ (fractional Sobolev space), Vol. I, p. 557
$W_2^1(\mathbb{R}), W_2^2(\mathbb{R})$ (Sobolev spaces), 514

$\mathcal{D}(\Omega) \equiv C_0^\infty(\Omega)$ (space of smooth test functions with compact support), Vol. I, p. 543
$\mathcal{S}(\mathbb{R}^N)$ (space of rapidly decreasing test functions), Vol. I, p. 537
$\mathcal{E}(\mathbb{R}^N) \equiv C^\infty(\mathbb{R}^N)$ (space of smooth test functions), Vol. I, p. 614
$\mathcal{D}'(\mathbb{R}^N)$ (space of distributions),

Vol. I, p. 608
$\mathcal{S}'(\mathbb{R}^N)$ (space of tempered distributions), Vol. I, p. 615
$\mathcal{E}'(\mathbb{R}^N)$ (space of distributions with compact support), Vol. I, p. 614

$\sigma^1, \sigma^2, \sigma^3$ (Pauli matrices), 793
$\gamma^0, \gamma^1, \gamma^2, \gamma^3$ (Dirac–Pauli matrices), 793
γ^5 (chiral matrix) 793
$\gamma_0 = \gamma^0$, $\gamma_j = -\gamma^j$, $j = 1, 2, 3$, 904
$\gamma_W^0, \gamma_W^1, \gamma_W^2, \gamma_W^3$ (Weyl matrices), 936
γ_W^5 (chiral Weyl matrix), 936
$\gamma_M^0, \gamma_M^1, \gamma_M^2, \gamma_M^3$ (Majorana matrices), 937
γ_M^5 (chiral Majorana matrix), 937
$\overline{\psi} \equiv \psi^\dagger \gamma^0$
$\not{\partial} \equiv \partial_\mu \gamma^\mu$, $\not{\nabla} \equiv \nabla_\mu \gamma^\mu$, $\not{A}(x) := \gamma_\mu A^\mu(x)$ (Feynman's slash symbols), 807, 840

$S[q] = \min!$ (minimum problem), 361, 374
$S[q] = \text{critical!}$ (critical point problem), 361, 374
$\delta S(q; h)$ (first variation), 362, 374
$\delta^2 S(q; h)$ (second variation), 361, 374
$\frac{\delta S[q]}{\delta q}$ (first functional derivative), 374
$\frac{\delta S[q]}{\delta q(t)}$ (local functional derivative), 374
$\frac{\delta^2 S[q]}{\delta q^2}$ (second functional derivative), 375
$\mathcal{D}(t_0, t_1)$ (or $C_0^\infty(t_0, t_1)$) (space of test functions), 373
$\sigma(H)$ (spectrum of the operator H), 504
$\sigma_{\mathrm{pp}}(H)$ (pure point spectrum), 504
$\sigma_{ac}(H), \sigma_{\mathrm{sing}}(H), \sigma_{\mathrm{c}}(H), \sigma_{\mathrm{ess}}(H)$ (absolutely continuous, singular, continuous, essential spectrum), 504
X_{bound} (space of bound states), 504
X_{scatt} (space of scattering states), 504
X_{sing} (space of singular states), 504
$\varrho(H)$ (resolvent set of the operator H), 504
$R_E = (EI - H)^{-1}$ (resolvent), 504

$P(t, s)$ (propagator), Vol. I, p. 383
$P^+(t, s)$ (retarded propagator), Vol. I, p. 384
$P^-(t, s)$ (advanced propagator), Vol. I, p. 384
ϱ (density operator), 543, 651
ζ_A (zeta function of the operator A), 563, 571

$|0\rangle, \Phi_0$ (ground state of a free quantum field)
$|0_{\mathrm{int}}\rangle, \Phi_{\mathrm{int}}$ (ground state of a quantum field under interactions)
$\mathcal{T}$ (chronological operator), 837
$:AB:$ (normal product), 823
$\mathcal{C}(AB) \equiv \langle 0|AB|0\rangle$ (contraction), 842
$\mathcal{C}_{\mathcal{T}}(AB) \equiv \langle 0|\mathcal{T}(AB)|0\rangle$ (time-ordered contraction), 850
$:\overbrace{AB}:\ \equiv \mathcal{C}(AB) \cdot I$ (paired normal product)
$:\underbrace{AB}:\ \equiv \mathcal{C}_{\mathcal{T}}(AB) \cdot I$ (time-ordered paired normal product)
$:\overbrace{ABCD}:$ (bosonic case), 848, (fermionic case), 853
$:\underbrace{AB}\underbrace{CD}:$ (bosonic case), 850, (fermionic case), 854

$\mathcal{K}(x, t; y, t_0)$ (Feynman propagator kernel), 491
$\mathcal{R}(x, y; E)$ (resolvent kernel), 492
$\mathcal{G}(x, y; E)$ (energetic Green's function), 493
$\mathsf{S}(t, t_0)$ (scattering operator with respect to the time interval $[t_0, t]$), 610, 749, 835, 839
$\mathsf{S}(T) \equiv \mathsf{S}(\frac{T}{2}, -\frac{T}{2})$,
S (scattering operator $\mathsf{S}(+\infty, -\infty)$), 749
$\mathcal{W}(T)$ (transition probability), 838
$\langle \Phi_{\mathrm{in}}|\mathsf{S}(T)\Phi_{\mathrm{out}}\rangle$ (transition amplitude), 838
$u_{\mathbf{p},s}$ (free electron function), 806
$v_{\mathbf{p},s}$ (free positron function), 806
$\overline{u}_{\mathbf{p},s} \equiv u_{\mathbf{p},s}^\dagger \gamma^0$, $\overline{v}_{\mathbf{p},s} \equiv v_{\mathbf{p},s}^\dagger \gamma^0$
e_s^μ (photon polarization 4-vector with polarization index $s = 0, 1, 2, 3$), 805
$A_{\mathbf{p},s}^\mu$ (free photon function), 828,
A_{free}^μ (free photon quantum field), 837,
ψ_{free} (free electron-positron quantum field), 837
$\overline{\psi}_{\mathrm{free}} \equiv \psi_{\mathrm{free}}^\dagger \gamma^0$ (Dirac's adjoint electron-positron quantum field), 837
$a_{\mathbf{p},s}^+$ (creation operator for a photon with momentum vector $\mathbf{p}$ and polarization number $s = 0, 1, 2, 3$), 812
$a_{\mathbf{p},s}^-$ (annihilation operator for a photon), 812
$b_{\mathbf{p},s}^+$ (creation operator for an electron with momentum vector $\mathbf{p}$ and spin number $s = \pm\frac{1}{2}$), 815
$b_{\mathbf{p},s}^-$ (annihilation operator for an

Index