Introduction to the Classical Theory of Particles and Fields

Boris Kosyakov

Introduction to the Classical Theory of Particles and Fields

Springer

Dr. Boris Pavlovich Kosyakov
Russian Federal Nuclear Center
Institute of Theoretical and
Mathematical Physics
Mir Avenue 37
Sarov
607190 Nizhnii Novgorod Region
Russia
E-mail: kosyakov@vniief.ru

Library of Congress Control Number: 2006932179

ISBN-10 3-540-40933-5 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-40933-5 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media
springer.com

Typesetting: by the author and techbooks using a Springer LaTeX macro package
Cover design: eStudio Calamar S.L., F. Steinen-Broo, Pau/Girona, Spain

Printed on acid-free paper SPIN: 11801504 56/techbooks 5 4 3 2 1 0

Preface

This book is intended as an introduction to gauge field theory for advanced undergraduate and graduate students in high energy physics. The discussion is restricted to the classical (nonquantum) theory. Furthermore, general relativity is outside the limits of this book.

My initial plan was to review the self-interaction problem in classical gauge theories with particular reference to the electrodynamics of point electrons and Yang–Mills interactions of point quarks. The first impetus to summarize current affairs in this problem came to me from the late Professor Asim Barut at a conference on mathematical physics in Minsk, Belarus, during the summer of 1994. He pointed out that developing a unified approach to self-interaction in the classical context might help to illuminate the far more involved quantum version of this problem. The idea of writing a review of this kind came up again in my discussions with Professor Rudolf Haag during a physics conference at the Garda Lake, in the autumn of 1998. He advised me to extend the initial project to cover all attendant issues so that the review would meet the needs of senior students.

Self-interaction is a real challenge. Traditionally, students become aware of this problem in the course of quantum field theory. They encounter numerous divergences of the S matrix, and recognize them as a stimulus for understanding the procedure of renormalization. But even after expending the time and effort to master this procedure one may still not understand the physics of self-interaction. For example, it is difficult to elicit from textbooks whether the orders of divergence are characteristic of the interaction or are an artifact of the perturbative method used for calculations. On the other hand, the structure of self-interaction is explicit in solvable models. We will see that many classical gauge theory problems can be completely or partly integrated. In contrast, quantum field theory has defied solvability, with the exception of two- and three-dimensional models.

The self-interacting electron was of great concern to fundamental physics during the 20th century. However, classical aspects of this problem are gradually fading from the collective consciousness of theoretical physics. Textbooks which cover this topic in sufficient detail are rare. Among them, the best known is the 1965 Rohrlich's volume. This excellent review represents the state of the art in the mid-1960s. Since then many penetrating insights into this subject have been gained, in particular exact retarded solutions of the Yang–Mills–Wong theory. It is therefore timely to elaborate a unified view of the classical self-interaction problem. The present work is a contribution to this task.

The book is, rather arbitrarily, divided into two parts. The first part, which involves Chaps. 1–5 and 7, is a coherent survey of special relativity and field theory, notably the Maxwell–Lorentz and Yang–Mills–Wong theories. In addition, Mathematical Appendices cover the topics that are usually beyond the standard knowledge of advanced undergraduates: Cartan's differential forms, Lie groups and Lie algebras, γ-matrices and Dirac spinors, the conformal group, Grassmannian variables, and distributions. These appendices are meant as pragmatic reviews for a quick introduction to the subject, so that the reader will hopefully be able to read the main text without resorting to other sources.

The second part of this book, stretching over Chaps. 6 and 8–10, focuses on the self-interaction problem. The conceptual basis of this study is not entirely conventional. The discussion relies heavily on three key notions: the *rearrangement* of the initial degrees of freedom resulting in the occurrence of *dressed particles*, and *spontaneous symmetry deformation*.

We now give an outline of this book.

Chapter 1 discusses special relativity. Following the famous approach of Minkowski, we treat it as merely the geometry of four-dimensional pseudoeuclidean spacetime. Section 1.1 offers the physical motivation of this point of view, and introduces Minkowski space. Mathematical aspects of special relativity are then detailed in Sects. 1.2 through 1.6.

Chapter 2 covers the relativistic mechanics of point particles. Newton's second law is embedded in the four-dimensional geometry of Minkowski space to yield the dynamical law of relativistic particles. We define the electromagnetic field through the Lorentz force law, and the Yang–Mills field through the Wong force law. Electromagnetic field configurations are classified according to their algebraic properties. We develop a regular method of solving the equation of motion for a charged particle driven by a constant and uniform electromagnetic field. Section 2.5 reviews the Lagrangian formalism of relativistic mechanical systems. Reparametrization invariance is studied in Sect. 2.6. It is shown that a consistent dynamics is possible not only for massive, but also for massless particles. Section 2.7 explores the behavior of free spinning particles. Since the rigorous two-particle problem in electrodynamics is a formidable task, we pose a more tractable approximate problem, the so-called relativistic Kepler problem. We then analyze a binary system composed of a heavy magnetic monopole and a light charged particle. Collisions and decays of relativistic particles are briefly discussed in the final section.

Chapter 3 gives a derivation of the equation of motion for the electromagnetic field, Maxwell's equations. We show that some of the structure of Maxwell's equations is dictated by the geometrical features of our universe, in particular the fact that there are three space dimensions. The residual information translates into four assumptions: locality, linearity, the extended action-reaction principle, and lack of magnetic monopoles.

Chapter 4 covers solutions to Maxwell's equations. It begins by considering static electric and constant magnetic fields. Some general properties of solutions to Maxwell's equations are summarized in Sect. 4.2. Free electromagnetic fields are then examined in Sect. 4.3. We use the Green's function technique to solve the inhomogeneous wave equation in Sect. 4.4. The Liénard–Wiechert field appears as the retarded solution to Maxwell's equations with a point source moving along an arbitrary timelike smooth world line. A method of solving Maxwell's equations without resort to Green's functions is studied in Sect. 4.7. This method will prove useful later in solving the Yang–Mills equations. We show that the retarded electromagnetic field F generated by a single arbitrarily moving charge is invariant under local $\mathrm{SL}(2,\mathbb{R})$ transformations. This is the same as saying there is a frame in which the Liénard–Wiechert field F appears as a pure Coulomb field at each observation point. The chapter concludes with a discussion of the electromagnetic field due to a magnetic monopole.

Chapter 5 covers the Lagrangian formalism of general field theories, with emphasis on systems of charged particles interacting with the electromagnetic field. Much attention is given to symmetries and their associated conservation laws in electrodynamics. These symmetries are of utmost importance in the theory of fundamental interactions. The reader may wish to familiarize himself or herself with these concepts early in the study of field theory; the Maxwell–Lorentz theory seems to be a good testing ground. An overview of strings and branes completes this chapter. This material may be useful in its own right, and as an application of the calculus of variations to systems that combine mechanical and field-theoretic features.

Chapter 6 treats self-interaction in electrodynamics. We begin with the Goldstone and Higgs models to illustrate the mechanism of rearrangement whereby the original degrees of freedom appearing in the Lagrangian are rearranged to give new, stable modes. We then introduce the basic concept of radiation, and derive energy-momentum balance showing that mechanical and electromagnetic degrees of freedom are rearranged into dressed particles and radiation. The Lorentz–Dirac equation governing a dressed particle is discussed in Sect. 6.4. Two alternative ways of deriving this equation are given in Sect. 6.5.

The essentials of classical gauge theories are examined in Chap. 7. Section 7.1 introduces the Yang–Mills–Wong theory of point particles interacting with gauge fields, in close analogy with the Maxwell–Lorentz theory. We briefly review a Lagrangian framework for the standard model describing the three fundamental forces mediated by gauge fields: electromagnetic, weak, and strong. Section 7.3 outlines gauge field dynamics on spacetime lattices.

Exact solutions to the Yang–Mills equations are the theme of Chap. 8. It seems impossible to cover all known solutions. Many of them are omitted, partly because these solutions are of doubtful value in accounting for the subnuclear realm and partly because they are covered elsewhere. The emphasis is on exact retarded solutions to the Yang–Mills equations with the source composed of several colored point particles (quarks) moving along arbitrary timelike world lines. The existence of two classes of exact solutions distinguished by symmetry groups is interpreted as a feature of the Yang–Mills–Wong theory pertinent to the description of two phases of subnuclear matter.

Chapter 9 deals with selected issues concerning self-interaction in gauge theories. The initial degrees of freedom in the Yang–Mills–Wong theory are shown to rearrange to give dressed quarks and Yang–Mills radiation. We address the question of whether the renormalization procedure used for treating the self-interaction problem is self-consistent. A plausible explanation for the paradoxes of self-interaction in the Maxwell–Lorentz theory is suggested in Sect. 9.3.

To comprehend electrodynamics as a whole, one should view this theory from different perspectives in a wider context. For this purpose Chap. 10 generalizes the principles underlying mechanics and electrodynamics. The discussion begins with a conceivable extension of Newtonian particles (governed by the second order equation of motion) to systems whose equations of motion contain higher derivatives, so-called ‘rigid’ particles. Most if not all of such systems exhibit unstable behavior when coupled to a continuum force field. Electrodynamics in various dimensions is another line of generalizations. Two specific examples, $D + 1 = 2$ and $D + 1 = 6$, are examined in some detail. If Maxwell’s equations are preserved, then a consistent description for $D + 1 = 6$ is attained through the use of rigid particle dynamics with acceleration-dependent Lagrangians. With these observations, we revise Ehrenfest’s famous question: ‘In what way does it become manifest from the fundamental laws of physics that space has three dimensions?’ Nonlinear versions of electrodynamics, such as the Born–Infeld theory, are analyzed in Sect. 10.4. We modify the Maxwell–Lorentz theory by introducing a nonlocal form factor in the interaction term. The final section outlines the direct-action approach in which the interactions of particles are such that they simulate the electromagnetic field between them.

With rare exceptions, each section has problems to be solved. Some problems explore equations that appear in the main text without derivation, while other problems introduce additional ideas or techniques. The problems are an integral part of the book. Many of them are essential for the subsequent discussion. The reader is invited to read every problem and look for its solution. When running into difficulties with a particular problem, the reader may consult the answer or hint. References to problems are made by writing the number of the section in front of the number of the problem, for example, Problem 10.4.2 for the second problem of Sect. 10.4.

Each chapter ends with Notes where some remarks and references for further reading can be found. The reader should be warned that these Notes do not pretend to provide a complete guide to the history of the subject. The selection of the literature sources is a matter of the author's personal taste and abilities. Preference is given to the most frequently cited books, representative reviews, classical original articles, and papers that are useful in some sense. References are listed by the name of the author(s) and the year of publication.

I am indebted to many people with whom I have discussed the issues addressed in this book. I am especially thankful to Professors Irina Aref'eva, Vladislav Bagrov, Asim Barut, Iosif Buchbinder, Gariĭ Efimov, Dmitriĭ Gal'tsov, Iosif Khriplovich, Vladimir Nesterenko, Lev Okun, Valeriĭ Rubakov, and Georgiĭ Savvidi for their illuminating remarks. E-mail correspondences with Professors Terry Goldman, Matej Pavšič, Martin Rivas, and Fritz Rohrlich were of great benefit to this project.

I thank Professor Rudolf Haag for his kind encouragement, interesting comments, and letter of support to the International Science and Technology Center (ISTC) with his recommendation to allot funds for writing this book.

The financial support from ISTC, under the project # 1560, during the time the first version of this book was written, is gratefully acknowledged.

Finally, I am grateful for the assistance of Professor Richard Woodard. He was my foreign collaborator in the ISTC project. Although the role of foreign collaborators in project implementation is sometimes a formality, Richard voluntarily shouldered proofreading of this rather long manuscript. He read carefully all parts of the text, made numerous corrections of my wording, and pointed to several mathematical mistakes. His comments, criticism, and suggestions concerning the most difficult physical issues are of inestimable help. In spite of his best efforts, errors minor and major, and obscurities are solely my responsibility.

July, 2006

Boris Kosyakov
Sarov

Contents

1

Geometry of Minkowski Space

1.1 Spacetime

The essence of special relativity can be expressed as follows:

> *Space and time are fused into four-dimensional spacetime which is described by pseudoeuclidean geometry.*

Hermann Minkowski in his talk at the 80th Assembly of German Natural Scientists and Physicians on 21 September 1908 asserted: ‘Space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality’.

The mathematical aspect of this statement will be discussed in subsequent sections. But in order that the statement may have a physical meaning, it is necessary to clarify the terms *space* and *time*, in particular, to define procedures for measuring spatial and temporal coordinates of various events.

We first consider the frames of reference which are best suited to measure moving objects. Such frames are called *inertial*. In the relativistic context, they are often referred to as *Lorentz frames*. At present there is no agreement regarding the rigorous definition of inertial frames of reference, even though the general idea can be stated quite simply: every inertial frame executes a uniform motion along a straight line with respect to a fixed inertial frame. As soon as one grants the existence of a single inertial frame, the whole class of such frames is defined. For example, Newton envisioned a perfect inertial frame as that attributed to the ‘fixed stars’. Inertial frames could then proliferate in his celestial mechanics through launching laboratories at constant velocities with respect to the center of mass of the solar system. However, when viewed closely, this strategy proves untenable for it is based on the notions ‘uniform’ and ‘straight’ which defy explication in the absence of a preassigned inertial frame.

It seems advisable to reduce the level of rigor and make a heuristic argument involving the notion of states in unstable equilibrium. States of unstable equilibrium are maintained only in inertial frames, because shocks and blows associated with accelerated motions of noninertial frames prevent unstable systems from being balanced. We thus come to a simple operational criterion for distinguishing between inertial and noninertial frames based on the capability of inertial frames for preserving unstable equilibrations. For example, a spaceship which drifts freely in a region of space remote from other matter, without rotation, is a frame where balanced unstable systems (such as a magnetic needle installed halfway between north poles of two identical static magnets perpendicular to the axis along which the magnets are lined up) survive[1]. It is remarkable that this definition of inertial frames can be introduced prior to geometry, that is, using only a kind of yes-no decision.

All inertial frames are regarded as *equivalent* in the sense that the dynamical laws have the same content and form in every inertial frame. This statement represents the so-called *principle of relativity.* In his lecture to a Congress of Arts and Science at St Louis on 24 September 1904 Henri Poincaré said: 'According to the Principle of Relativity, the laws of physical phenomena must be the same for a "fixed" observer as for an observer who has a uniform motion of translation relative to him: so that we have not, and cannot possibly have, any means of discerning whether we are, or are not, carried along in such a motion.'

By contrast, the form of equations of motion changes in passing from an inertial frame to noninertial frames. Thus, among all possible frames, inertial frames could be thought of as having a privileged status. The analysis of any physical problem in this book pertains to inertial frames, unless otherwise indicated.

Another idea of primary importance in relativistic physics is the existence of *maximal* velocity of motion. Looking at waves of different nature (optical, sonic, etc.), which expand from some point of emission O, one can compare their propagation rate even without knowledge of the numerical values of their velocities. This is in the same spirit as an ancient allegory: Achilles is prompter than a tortoise, an arrow is prompter than Achilles, a thunderbolt is prompter than arrows, etc. Mathematically, we have a chain $a < b < c < \ldots$ The ordering is verified by inspection, and has no need of particular velocity scale. One may suppose that there is an upper bound for propagation rates. In fact, this supposition is well verified experimentally. A light wave runs down any mover which left O before the light emission, except for another light wave emitted in O. The existence of the highest propagation rate – universally referred to as the *speed of light*, is a central tenet of relativity. It is generally believed that the fundamental interactions of nature propagate at the speed of light. This should be compared with pre-relativistic physics in which arbitrarily high velocities of bodies are allowable, and interactions (such as Newtonian gravitation) are instantaneous. Poincaré in his 1904 lecture anticipated the advent

[1] To adapt this definition of idealized inertial frames to the real world we employ testing systems which are stable against small perturbations but unstable against perturbations above some finite threshold.

of ‘entirely new kind of dynamics, which will be characterized above all by the rule, that no velocity can exceed the velocity of light’.

The concept of maximal velocity is the rationale for integrating space and time into *spacetime*. Given a clock and a radar ranging device, we have actually everything required for probing the geometry of the real world. Indeed, we can measure the distance between two points by the time that light takes to traverse it. Taking the speed of light to be unity, we find the distance we seek to be half the round-trip span between these points.

Why is radar location preferable to using yardsticks? In practice, it is not always possible to lay out a grid of the yardsticks. Take, for example, cosmic measurements. But, what is more important, when moving, the yardstick is purported to preserve size and shape (at least in the absence of stresses and temperature variations), which has yet to be proved. By contrast, with the radar-location approach, the size and shape of a given rod can be verified by continuously sounding its ends. Furthermore, the notion of rectilinearity, stemming from geometrical optics in which light rays travel along straight lines is an integral part of the radar-location approach.

Now, armed with a well defined and controllable standard for measuring lengths and for determining what is a straight line, we are allowed the option of using either yardstick or radar for a particular measurement.

Subsequent to choosing the length scale, we can define the velocity scale. In the relativistic context, it is natural to take the convention that the speed of light in vacuum[2] is 1, and measure space and time intervals in the same units, say, in meters. We follow this convention throughout.

An experimentalist, having different types of cyclic mechanisms operating on various time at his disposal, may inquire whether there is a scale to express the physical laws in the simplest form. The desired scale t actually exists. It is defined by the condition of Galileo Galilei that all inertial frames move along straight lines at constant rates with respect to a given inertial frame,

$$\frac{dx}{dt} = V = \text{const} \,. \tag{1.1}$$

This scale, called the *standard* or *laboratory* scale, is defined up to a linear transformation, that is, up to a relabeling (say, hours may be termed minutes – which is of course a matter of convention), and a change in the zero of time.

By comparison, if we take an arbitrary scale τ, we observe that inertial frames execute nonuniform motions, that is, their velocities vary with τ:

$$\frac{dx}{d\tau} = v(\tau) \,. \tag{1.2}$$

However, proceeding from an arbitrary scale τ, the standard scale t readily regains,

[2] In the literature, the speed of light is usually symbolized by the letter c. However, this designation is unnecessary if we adopt units in which $c = 1$.

$$t = F(\tau)\ . \tag{1.3}$$

Indeed, from

$$\frac{dx}{dt} = \frac{dx}{d\tau}\,\frac{d\tau}{dt} = v\,\frac{1}{F'} \tag{1.4}$$

we obtain

$$F(\tau) = \frac{1}{V}\int_0^\tau d\sigma\, v(\sigma)\ . \tag{1.5}$$

Expressions (1.2) and (1.5) suggest the way for rearranging the given chronometer to yield the standard scale t. Note that clocks which read the Galilean time rate should not be regarded as 'true', or 'accurate', even if this time rate may be judged much simpler than other rates. At present, the closest fit to the standard time rate t show atomic clocks.

It might seem that a 'perfect periodicity' can be attained if light rays are sandwiched between two parallel mirrors. To measure time one simply counts successive reflections of this light shuttle. A virtue of this light clock is that spatial and temporal measurements are assembled into a single process of sending light flashes and receiving their echoes. But a closer look at this timekeeping reveals a serious flaw: the mirrors must be separated by a *fixed* distance, which is a problem. In the absence of reliable space control procedure different from radar sounding, this chronometry would make the definition circular.

We therefore will content ourselves with the very existence of a clock (such as atomic clocks) whose readings conform with the requirement that relative motions of inertial frames be uniform. In the subsequent discussion we assume that inertial frames are equipped with clocks of this type.

It in no way follows that the light propagation is uniform if time is read on the standard time scale. However, physical reality is so structured that such is the case. One may treat this opportunity as a separate principle. In fact, it was the original 1905 proposal by Albert Einstein that the constancy of the velocity of light, together with the principle of relativity, constitute the basic postulates of special relativity. Attention is usually drawn to the condition that the velocity of light is the same in all inertial frames. But due regard must be given to another aspect of Einstein's proposal, the uniformity of light propagation.

We now address the geometric properties of spacetime. It was pointed out by Poincaré that the geometry by itself eludes verification. What is to be verified is the totality of geometry and physics. A change of geometric axioms can be accompanied by a suitable modification of physical laws in such a way that the prediction of observed phenomena are unchanged. Nevertheless, the entire theoretical scheme is not indifferent to the choice of a particular geometry[3]. Our interest here is with the simplest physically justifiable version of spacetime.

[3] As an illustration, we refer to the problem of the field generated by a magnetic charge. This problem can be stated in two alternative geometric settings. We

Special relativity assumes that spacetime is described by the geometry of *Minkowski space*. It is generally believed that Minkowski space is the best geometric framework for the great bulk of phenomena, excluding situations in which strong gravitational fields are present. The reason for this belief is threefold:

(i) the existence of inertial frames,
(ii) the existence of a standard time scale,
(iii) the existence of the highest propagation rate (associated with the speed of light); the uniformity of light propagation.

An inertial observer, having clocks with the standard time scale, will recognize time as *homogeneous* in the sense that all instants are equivalent, and space *homogeneous* and *isotropic* in the sense that there are no privileged position in space and distinguished direction of motion. These facts are responsible for the *affine* structure of Minkowski space. As this topic is reviewed again in the following sections, we put off our discussion of it until then. For now, we turn to the *metrical* structure of spacetime. An inertial observer will recognize Minkowski space to be endowed with an indefinite metric. This geometric property is most readily visualized with the aid of the so-called *k-calculus* by Hermann Bondi.

Let $\mathcal{O}$ and $\mathcal{O}'$ be two arbitrary inertial observers. Both observers carry identical clocks with the standard time scale, and are equipped with a radar ranging set. Suppose that $\mathcal{O}$ is at rest in a given inertial frame, while $\mathcal{O}'$ is moving along the x-axis at a constant velocity V. This can be expressed in geometric terms by saying that the *world lines* of both observers are rectilinear. In the two-dimensional diagram, Fig. 1.1, the vertical axis represents time, and the horizontal axis space. Light rays are depicted as straight lines at 45° angles. The world line of $\mathcal{O}$ is a straight line parallel to the time axis, while the world line of $\mathcal{O}'$, $x = x_0 + Vt$, is tilted from the vertical axis at an angle less than 45°, because $V < 1$.

Suppose that $\mathcal{O}$ sends two flashes of light separated by an interval T on his clock. The interval between reception of these flashes on the clock of $\mathcal{O}'$ is kT, where k is the so-called *k-factor*. From the preceding discussion it follows that k is a constant which depends on the relative velocity V. By the principle of relativity, the situation is symmetric when $\mathcal{O}'$ emits light signals and $\mathcal{O}$ receives them: an interval between two emitted signals as measured by $\mathcal{O}'$ is multiplied by the same factor k to give an interval between reception of these signals as measured by $\mathcal{O}$. If $\mathcal{O}'$ reflects signals which were sent by $\mathcal{O}$, then $\mathcal{O}$ records an interval k^2T between reception of the two successive echoes.

---

will see in Sect. 4.8 that the vector potential due to a magnetic monopole A_μ as viewed in the usual Euclidean space is singular on a line that issues out of the magnetic charge, while A_μ as viewed in a manifold which is obtained by gluing together two Euclidean spaces is regular everywhere except for the point where the magnetic charge is located.

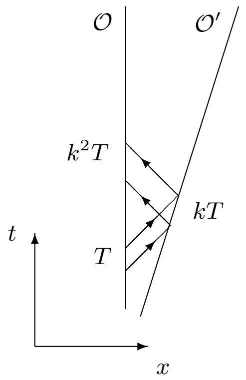


Fig. 1.1. Bondi k-calculus

To gain insight into the physical meaning of the k-factor, we first note, following Poincaré, that whether or not two spatially separated events are simultaneous is, to some extent, a matter of convention. Einstein refined this idea by the following convention (which is now called *standard synchrony*). Let $\mathcal{O}$ emit a pulse of light when his clock reads t_i. Let this pulse be reflected at a spacetime point P, and the echo is received by $\mathcal{O}$ at the instant t_f. Then event P is *simultaneous* with the event at $\mathcal{O}$ that occurred at time $\frac{1}{2}(t_i + t_f)$. We therefore assign coordinates to the event P given by

$$t = \frac{1}{2}(t_f + t_i), \quad x = \frac{1}{2}(t_f - t_i) \ . \tag{1.6}$$

In the standard synchrony $\mathcal{O}$ computes the following values for the temporal and spatial differences between the events at which $\mathcal{O}'$ reflects the first and second pulses

$$\Delta t = \frac{1}{2}\left(k^2 + 1\right) T, \quad \Delta x = \frac{1}{2}\left(k^2 - 1\right) T \ . \tag{1.7}$$

Because $\mathcal{O}'$ moves at constant velocity V between the two reflection events, we obtain a relation between k and V

$$V = \frac{\Delta x}{\Delta t} = \frac{k^2 - 1}{k^2 + 1} \ . \tag{1.8}$$

Solving for k gives

$$k = \sqrt{\frac{1 + V}{1 - V}} \ . \tag{1.9}$$

One may then imagine that $\mathcal{O}$ emits light in the form of a monochromatic wave whose frequency ν is the reciprocal of the period of oscillations T of an atomic clock. Hence, the wave carries the unit of time T adopted by $\mathcal{O}$. Our analysis is based on the assumption that $\mathcal{O}'$ sees a wave whose frequency ν' differs from ν by the factor k,

$$k = \frac{\nu}{\nu'} = \sqrt{\frac{1+V}{1-V}} . \tag{1.10}$$

For $V \ll 1$, this relation becomes

$$\nu' = \nu\,(1-V) \ , \tag{1.11}$$

which is the Doppler red shift due to the outward motion of the source, familiar from elementary physics. Thus, the k-factor is a relativistic manifestation of the Doppler shift.

We now introduce another inertial observer $\mathcal{O}''$ moving along the x-axis. Assume that a factor $k_{\mathcal{O}\mathcal{O}''}$ characterizes the relative motion of $\mathcal{O}$ and $\mathcal{O}''$, and $k_{\mathcal{O}'\mathcal{O}''}$ is the corresponding factor for $\mathcal{O}'$ and $\mathcal{O}''$. One can show (Problem 1.1.1) that

$$k_{\mathcal{O}\mathcal{O}''} = k_{\mathcal{O}\mathcal{O}'}\, k_{\mathcal{O}'\mathcal{O}''} \ . \tag{1.12}$$

Combining (1.9) and (1.12), we get the *relativistic rule for the addition of velocities*

$$V_{\mathcal{O}\mathcal{O}''} = \frac{V_{\mathcal{O}\mathcal{O}'} + V_{\mathcal{O}'\mathcal{O}''}}{1 + V_{\mathcal{O}\mathcal{O}'}\, V_{\mathcal{O}'\mathcal{O}''}} \ . \tag{1.13}$$

The value of $V_{\mathcal{O}\mathcal{O}''}$ given by (1.13) cannot exceed 1. Assuming either $V_{\mathcal{O}\mathcal{O}'}$ or $V_{\mathcal{O}'\mathcal{O}''}$, or both, equal to 1, we find $V_{\mathcal{O}\mathcal{O}''} = 1$. If $V_{\mathcal{O}\mathcal{O}'} \ll 1$, $V_{\mathcal{O}'\mathcal{O}''} \ll 1$, then (1.13) becomes

$$V_{\mathcal{O}\mathcal{O}''} = V_{\mathcal{O}\mathcal{O}'} + V_{\mathcal{O}'\mathcal{O}''} \ , \tag{1.14}$$

which is the rule for the addition of velocities in Newtonian kinematics.

Let us turn back to the situation with two inertial observers $\mathcal{O}$ and $\mathcal{O}'$. We describe an event P as viewed by both $\mathcal{O}$ or $\mathcal{O}'$. Suppose that $\mathcal{O}$ and $\mathcal{O}'$ meet at a point O, as in Fig. 1.2, and, at the instant of their meeting, both reset their clocks to 0. A light flash, which is sent by $\mathcal{O}$ at t_i to P, being reflected, returns to $\mathcal{O}$ at t_f. Likewise, $\mathcal{O}'$ locates P, the corresponding instants being t'_i and t'_f. Following the pattern shown in (1.6), $\mathcal{O}$ assigns the following coordinates to this event

$$t = \frac{1}{2}\,(t_f + t_i)\,, \quad x = \frac{1}{2}\,(t_f - t_i) \ , \tag{1.15}$$

whereas $\mathcal{O}'$ gets

$$t' = \frac{1}{2}\,(t'_f + t'_i)\,, \quad x' = \frac{1}{2}\,(t'_f - t'_i) \ . \tag{1.16}$$

Taking into account that

$$\frac{t'_i}{t_i} = \frac{t_f}{t'_f} = k \ , \tag{1.17}$$

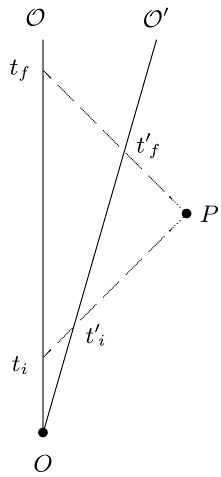


Fig. 1.2. Derivation of the Lorentz transformation

we obtain from (1.16)

$$t' = \frac{1}{2}\left(k + \frac{1}{k}\right) t - \frac{1}{2}\left(k - \frac{1}{k}\right) x, \quad x' = \frac{1}{2}\left(k + \frac{1}{k}\right) x - \frac{1}{2}\left(k - \frac{1}{k}\right) t \ . \quad (1.18)$$

We see that the coordinates (t, x) assigned to event P by $\mathcal{O}$ can be expressed in terms of the coordinates (t', x') assigned to this same event by $\mathcal{O}'$ through (1.18).

With (1.9), we rewrite (1.18) in the form

$$\begin{aligned} t' &= \gamma\,(t - Vx) \ , \\ x' &= \gamma\,(x - Vt) \ , \end{aligned} \quad (1.19)$$

where

$$\gamma = \frac{1}{\sqrt{1 - V^2}} \ . \quad (1.20)$$

Equation (1.19) is named the *Lorentz transformation* for its discoverer Hendrik Antoon Lorentz. The factor γ is referred to as *Lorentz factor*. In the current literature, (1.19) is often called the *Lorentz boost*, or simply the *boost*.

The inverse of the Lorentz transformation (1.19) is

$$\begin{aligned} t &= \gamma\,(t' + Vx') \ , \\ x &= \gamma\,(x' + Vt') \ . \end{aligned} \quad (1.21)$$

Evidently (1.21) follows from (1.19) if we replace V by $-V$, because $\mathcal{O}'$ might say that $\mathcal{O}$ moves at the velocity $-V$ along the x-axis. If $V = 0$, then (1.19) is the identity transformation, $t' = t, x' = x$. Therefore, transformations (1.19) constitute a continuous group, the *Lorentz group*. The parameter V ranges from -1 to 1,

Suppose that a rod is at rest with respect to $\mathcal{O}$, and is directed parallel to the x-axis. Let the coordinates of its ends at some instant t be x_1 and x_2. Then its length as measured by $\mathcal{O}$ is $\Delta x = x_2 - x_1$. What is the length of the rod as measured by a moving inertial observer $\mathcal{O}'$? It follows from (1.21) that

$$x_1 = \gamma\,(x'_1 + Vt')\,, \quad x_2 = \gamma\,(x'_2 + Vt')\ , \tag{1.22}$$

and so

$$\Delta x' = x'_2 - x'_1 = \gamma^{-1}\,(x_2 - x_1)\ . \tag{1.23}$$

The length of the rod at rest is called the *proper length*. Let us denote the proper length by l, and the distance between the ends of the rod as measured by a moving observer $\mathcal{O}'$ by l', then

$$l' = l\,\sqrt{1 - V^2}\ . \tag{1.24}$$

We see that the length of the rod is not identical for different inertial observers; the maximum value l is attained in the rest frame. In general, the length of a moving body is contracted in the direction of motion in comparison with its proper length. This contraction is called the *Lorentz–FitzGerald contraction*, or simply the *Lorentz contraction*.

Equation (1.19) implies

$$t'^2 - x'^2 = t^2 - x^2\ . \tag{1.25}$$

This relation shows that the quadratic form $t^2 - x^2$ is the same for any inertial observer. We therefore say that $t^2 - x^2$ is invariant under Lorentz transformations. It is reasonable to interpret this invariant quadratic form as a measure of separation between P and O in the two-dimensional diagram, a prototype of the indefinite *metric* in Minkowski space

$$t^2 - x^2 - y^2 - z^2\ . \tag{1.26}$$

Observing that the identity $\gamma^2 - \gamma^2 V^2 = 1$ is similar to the identity $(\cosh\theta)^2 - (\sinh\theta)^2 = 1$, we can rewrite the Lorentz transformation (1.19) as

$$\begin{aligned} t' &= t\cosh\theta - x\sinh\theta\ ,\\ x' &= x\cosh\theta - t\sinh\theta\ , \end{aligned} \tag{1.27}$$

where θ is related to V by

$$V = \tanh\theta\ , \tag{1.28}$$

or, in matrix notation,

$$\begin{pmatrix} t' \\ x' \end{pmatrix} = \begin{pmatrix} \cosh\theta & -\sinh\theta \\ -\sinh\theta & \cosh\theta \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix}\ . \tag{1.29}$$

The composition of two boosts (1.29) specified by respective parameters θ_1 and θ_2 is a new boost specified by

$$\theta = \theta_1 + \theta_2 \ . \tag{1.30}$$

Substituting (1.30) in (1.28) shows that the resulting boost is associated with the velocity of relative motion

$$V = \frac{V_1 + V_2}{1 + V_1 V_2} \ , \tag{1.31}$$

where $V_1 = \tanh\theta_1$, $V_2 = \tanh\theta_2$. Equation (1.31) is apparently the relativistic rule for the addition of velocities (1.13). Since (1.30) is simpler than (1.31), the use of θ, called the *rapidity*, is sometimes more advantageous than that of V. The rapidity θ runs from $-\infty$ to ∞. Negative θ are mapped by (1.28) to negative V, $\theta = 0$ corresponds to $V = 0$.

Newtonian mechanics does not have a maximum velocity. A body can travel with any velocity, and the interactions of separated bodies can be instantaneous. Therefore, the requirement to synchronize clocks is no longer relevant, because time is regarded as absolute, that is, the same for different inertial observers. Only the space coordinates assigned to some event by two inertial observers are different. These coordinates are related by the *Galilean transformation*

$$x' = x - Vt \ . \tag{1.32}$$

Note that (1.32) is not a limititing case of the Lorentz transformation (1.19) arising as $V \to 0$, until we assume that x is of order Vt.

Problem 1.1.1. Justify (1.12).

Problem 1.1.2. Consider two Lorentz observers in relative motion with velocity $\mathbf{V}$. Let the Cartesian coordinates of some event as seen by one observer be $(t, \mathbf{x})$, and those as seen by the other be $(t', \mathbf{x}')$. Introduce the decompositions $\mathbf{x} = \mathbf{x}_\parallel + \mathbf{x}_\perp$ and $\mathbf{x}' = \mathbf{x}'_\parallel + \mathbf{x}'_\perp$, where $\parallel$ stands for the projection of a given vector on $\mathbf{V}$, and $\perp$ labels the projection on a plane perpendicular to $\mathbf{V}$. Derive the Lorentz transformation of these coordinates.

Answer

$$t' = \gamma\left(t - V\mathbf{x}_\parallel\right), \quad \mathbf{x}'_\parallel = \gamma\left(\mathbf{x}_\parallel - \mathbf{V}t\right), \quad \mathbf{x}'_\perp = \mathbf{x}_\perp \ , \tag{1.33}$$

$$V = |\mathbf{V}|, \quad \gamma = \left(1 - V^2\right)^{-\frac{1}{2}} \ . \tag{1.34}$$

1.2 Affine and Metric Structures

Mathematically, Minkowski space is the set $\mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$. Elements of this set are called *points*. In order to convert an abstract set of points into a geometric space, one equips it with a geometric structure. The geometry of Minkowski space derives from *affine* and *metric* structures. We therefore begin by recalling the system of axioms (due to Hermann Weyl) which define affine spaces.

Definition 1. *Consider a set* V *whose elements, called vectors, can be added and multiplied by constants. We will call such sets vector spaces. More precisely, for any two vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *we define a vector* $\mathbf{c}$*, which is their sum,* $\mathbf{c} = \mathbf{a} + \mathbf{b}$*, and for any vector* $\mathbf{a}$ *and any real number* λ *we define a vector* $\lambda\mathbf{a}$*, the multiplication of* $\mathbf{a}$ *by* λ*. The operations of addition and multiplication by constants must fulfil the following eight rules:*

1°. $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$.
2°. $\mathbf{a} + (\mathbf{b} + \mathbf{c}) = (\mathbf{a} + \mathbf{b}) + \mathbf{c}$.
3°. There is a vector $\mathbf{0}$, the *zero vector*, such that $\mathbf{0} + \mathbf{a} = \mathbf{a}$.

Note that the vector $\mathbf{0}$ is unique. Assume that there are two zero vectors: $\mathbf{0}_1 + \mathbf{a} = \mathbf{a}$ and $\mathbf{0}_2 + \mathbf{a} = \mathbf{a}$, then $\mathbf{0}_1 = \mathbf{0}_2$. Indeed, substituting $\mathbf{a} = \mathbf{0}_1$ into $\mathbf{0}_2 + \mathbf{a} = \mathbf{a}$ gives $\mathbf{0}_2 + \mathbf{0}_1 = \mathbf{0}_1$, while substituting $\mathbf{a} = \mathbf{0}_2$ into $\mathbf{0}_1 + \mathbf{a} = \mathbf{a}$ gives $\mathbf{0}_1 + \mathbf{0}_2 = \mathbf{0}_2$, hence $\mathbf{0}_1 = \mathbf{0}_2$.

4°. For every vector $\mathbf{a}$ there is a vector $-\mathbf{a}$ such that $\mathbf{a} + (-\mathbf{a}) = \mathbf{0}$.

Note that the vector $-\mathbf{a}$ is unique in that $\mathbf{a} + \mathbf{b} = \mathbf{0}$ and $\mathbf{a} + \mathbf{c} = \mathbf{0}$ imply $\mathbf{b} = \mathbf{c}$. Indeed, $\mathbf{b} = \mathbf{0} + \mathbf{b} = (\mathbf{a} + \mathbf{c}) + \mathbf{b} = (\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{0} + \mathbf{c} = \mathbf{c}$.

5°. $1\mathbf{a} = \mathbf{a}$.
6°. $(\lambda + \mu)\,\mathbf{a} = \lambda\mathbf{a} + \mu\mathbf{a}$.
7°. $(\lambda\mu)\,\mathbf{a} = \lambda\,(\mu\mathbf{a})$.
8°. $\lambda\,(\mathbf{a} + \mathbf{b}) = \lambda\mathbf{a} + \lambda\mathbf{b}$.

If vectors are visualized as arrows, they obey the usual parallelogram law for addition. Multiplication by a positive constant alters the length of the vector, without changing its direction; for $\lambda = 0$, the result of multiplication is the zero vector, $0\mathbf{a} = \mathbf{0}$; for $\lambda < 0$, the direction of the multiplied vector is reversed, $(-1)\,\mathbf{a} = -\mathbf{a}$ (Problem 1.2.1). Hereafter, $\mathbf{b} - \mathbf{a}$ will stand for $\mathbf{b} + (-\mathbf{a})$.

If we allow for multiplication of vectors by complex numbers, assuming axioms 1°–8°, we obtain a *complex* (rather than *real*) vector space.

Now it is possible to form *linear combinations* of vectors $\lambda^1\mathbf{a}_1 + \cdots + \lambda^m\mathbf{a}_m$. A linear combination is nontrivial if not all coefficients $\lambda^1, \ldots, \lambda^m$ are vanishing. Vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$ are called *linearly dependent* if there is a nontrivial combination of these vectors which is the zero vector, $\lambda^1\mathbf{a}_1 + \cdots + \lambda^m\mathbf{a}_m = \mathbf{0}$, otherwise they are *linearly independent*. If $\mathbf{a}$ and $\mathbf{b}$ are linearly dependent, $\mathbf{a} + \lambda\mathbf{b} = \mathbf{0}$, then $\mathbf{a}$ and $\mathbf{b}$ are said to be *collinear*.

Definition 2. *Let* $\mathbf{e}_1, \ldots, \mathbf{e}_n$ *be a linearly independent family of vectors. Suppose that any vector* $\mathbf{a} \in V$ *can be represented as a linear combination of this family of vectors,*

$$\mathbf{a} = a^1\mathbf{e}_1 + \cdots + a^n\mathbf{e}_n \;, \tag{1.35}$$

then $\mathbf{e}_1, \ldots, \mathbf{e}_n$ *is called a basis of* V.

To the above axioms 1°–8° must be added one more

9°. There exists a basis involving n vectors.

With these axioms, V is defined as a vector space. This V is said to have *dimension* n. Every basis of V includes n vectors. Furthermore, any set of $n+1$ vectors is linearly dependent. We denote this vector space by V_n if we are to be more emphatic.

For a given basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$, the decomposition (1.35) is unique. To see this, we assume that there is an alternative decomposition $\mathbf{a} = \bar{a}^1\mathbf{e}_1 + \cdots + \bar{a}^n\mathbf{e}_n$. We then have $\left(a^1 - \bar{a}^1\right)\mathbf{e}_1 + \cdots + \left(a^n - \bar{a}^n\right)\mathbf{e}_n = \mathbf{0}$, which implies $a^i - \bar{a}^i = 0$ for every i. We call $a^1, \ldots, a^n$ *coordinates* of the vector $\mathbf{a}$ in the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$, and write $\mathbf{a} = (a^1, \ldots, a^n)$. The decomposition (1.35) can be recast in the brief form:

$$\mathbf{a} = a^i\mathbf{e}_i \;, \tag{1.36}$$

with the understanding that a repeated index is summed over. To be specific, indices that appear both as superscripts and subscripts are summed from 1 to n. This summation convention, proposed by Einstein, is generally accepted in the physical literature[4].

It is evident that $\mathbf{a} + \mathbf{b}$ has coordinates $a^1 + b^1, \ldots, a^n + b^n$, and $\lambda\mathbf{a}$ has coordinates $\lambda a^1, \ldots, \lambda a^n$. This provides a one-to-one mapping from V_n to $\mathbb{R}_n$, a space of all n-tuples of real numbers $(a^1, \ldots, a^n)$. One-to-one mappings which leave invariant linear operations are called *isomorphisms*. We see that every n-dimensional vector spaceis isomorphic to $\mathbb{R}_n$. Hence, all vector spaces of dimension n are isomorphic to each other.

Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and $\mathbf{e}'_1, \ldots, \mathbf{e}'_n$ be two arbitrary bases. Each vector of the latter basis can be expanded in terms of vectors of the former basis:

$$\mathbf{e}'_i = L^j{}_i\,\mathbf{e}_j \;. \tag{1.37}$$

This can be written in matrix form:

$$\begin{pmatrix} \mathbf{e}'_1 \\ \mathbf{e}'_2 \\ \vdots \\ \mathbf{e}'_n \end{pmatrix} = \begin{pmatrix} L^1{}_1 & L^2{}_1 & & L^n{}_1 \\ L^1{}_2 & L^2{}_2 & & L^n{}_2 \\ & & \ddots & \\ L^1{}_n & L^2{}_n & & L^n{}_n \end{pmatrix} \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_n \end{pmatrix} . \tag{1.38}$$

[4] In this section, we denote indices by lower-case latin letters. The notation of a repeated index (which is sometimes called an 'umbral', or 'dummy', index) may be freely changed leaving the quantity unaltered. For example, $a^i\mathbf{e}_i$ may be substituted with $a^j\mathbf{e}_j$, because these expressions are identical. We will make extensive use of this freedom for relabeling indices.

What restrictions are there on the matrix $L^j{}_i$? Since n linearly independent vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are transformed into n linearly independent vectors $\mathbf{e}'_1, \ldots, \mathbf{e}'_n$, $L^j{}_i$ must be a real $n \times n$ matrix such that

$$\det L \neq 0 \ . \tag{1.39}$$

On the other hand, $\mathbf{e}_1, \ldots, \mathbf{e}_n$ can be expanded in terms of $\mathbf{e}'_1, \ldots, \mathbf{e}'_n$,

$$\mathbf{e}_i = M^j{}_i \, \mathbf{e}'_j \ . \tag{1.40}$$

Combining (1.37) and (1.40), we obtain

$$\mathbf{e}_i = M^j{}_i \, L^k{}_j \, \mathbf{e}_k \ , \tag{1.41}$$

and

$$\mathbf{e}'_i = L^j{}_i \, M^k{}_j \, \mathbf{e}'_k \ , \tag{1.42}$$

For (1.41) and (1.42) to be identities, it is necessary that

$$M^j{}_i \, L^k{}_j = L^j{}_i \, M^k{}_j = \delta^k{}_i \ , \tag{1.43}$$

where $\delta^k{}_i$ is the *Kronecker delta*, which is zero when $k \neq i$, and unity when $k = i$. In matrix notation,

$$ML = LM = \mathbf{1} \ , \tag{1.44}$$

where $\mathbf{1}$ is the unit $n \times n$ matrix whose diagonal matrix elements equals 1, and off-diagonal matrix elements are 0. Equation (1.44) shows that M is the inverse of L,

$$M = L^{-1} \ . \tag{1.45}$$

That is why we impose the condition (1.39): if $\det L$ is nonzero, then there exists a matrix inverse of L.

Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and $\mathbf{e}'_1, \ldots, \mathbf{e}'_n$ be two bases. Any vector $\mathbf{a}$ can be written either

$$\mathbf{a} = a^i \mathbf{e}_i \tag{1.46}$$

or

$$\mathbf{a} = a'^i \mathbf{e}'_i \ . \tag{1.47}$$

Applying (1.40) to (1.46), we have

$$\mathbf{a} = a^j M^i{}_j \, \mathbf{e}'_i \ , \tag{1.48}$$

which, in view of (1.47), gives

$$a'^i = M^i{}_j \, a^j \ . \tag{1.49}$$

We see that new coordinates a'^i are obtained from old coordinates a^i by the action of a matrix, which is the inverse of transposed L.

Consider the totality of bases in V_n. One basis is related to another by a linear transformation (1.37). These transformations form a Lie group (Problem 1.2.2), the *general linear* group $\mathrm{GL}(n, \mathbb{R})$. Each element of $\mathrm{GL}(n, \mathbb{R})$ is a real $n \times n$ matrix with nonzero determinant (for an overview of Lie groups see Appendix B). Given a fixed basis, all other bases can be produced from it through (1.37) in which $L^j{}_i$ ranges over the parameter space of $\mathrm{GL}(n, \mathbb{R})$. By (1.49), the same is true for $\mathbb{R}_n$: given some n-tuple of real numbers, say, $(1, 0, \ldots, 0)$, all other n-tuples arise from it if $M^j{}_i$ ranges over the parameter space of $\mathrm{GL}(n, \mathbb{R})$. This implies a new definition of vector spaces. Indeed, $\mathbb{R}_n$ can be completely characterized by the transformation group $\mathrm{GL}(n, \mathbb{R})$. These transformations are called *automorphisms* of $\mathbb{R}_n$. Hence the vector becomes a derivative concept specified by n numbers $a^1, a^2, \ldots, a^n$ and the transformation law (1.49).

Once we have a vector space, it is natural to define an associated vector space, known as the dual space.

Definition 3. *A mapping* $\omega : V \to \mathbb{R}$ *such that*

$$\omega(\mathbf{a} + \mathbf{b}) = \omega(\mathbf{a}) + \omega(\mathbf{b}), \quad \omega(\kappa \mathbf{a}) = \kappa \omega(\mathbf{a}) \tag{1.50}$$

is called a linear functional ω*. In other words, a linear functional is a rule for writing real numbers* $\omega(\mathbf{a})$ *associated with vectors* $\mathbf{a}$ *in such a way as to obey (1.50).*

Addition and multiplication by constants are obviously defined for linear functionals,

$$(\omega_1 + \omega_2)(\mathbf{a}) = \omega_1(\mathbf{a}) + \omega_2(\mathbf{a}), \quad (\kappa\omega)(\mathbf{a}) = \kappa\, \omega(\mathbf{a})\ . \tag{1.51}$$

Thus, linear functionals form the *dual* vector space V'. If V is n-dimensional, so is V'. Indeed, let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be a basis in V. Then any $\omega \in V'$ is specified by n real numbers $\omega_1 = \omega(\mathbf{e}_1), \ldots, \omega_n = \omega(\mathbf{e}_n)$, and the value of ω on $\mathbf{a} = a^i \mathbf{e}_i$ is given by

$$\omega(\mathbf{a}) = \omega_i a^i\ . \tag{1.52}$$

We see that V' is isomorphic to V. That is why we sometimes refer to linear functionals as *covectors*. A closer look at (1.52) shows that a vector $\mathbf{a}$ can be regarded as a linear functional on V'. One can show (Problem 1.2.3) that changing the basis (1.37) implies the transformation of ω_i according to the same law:

$$\omega'_i = L^j{}_i\, \omega_j\ . \tag{1.53}$$

We will usually suppress the argument of $\omega(\mathbf{a})$, and identify ω with its components ω_i.

If dimension of V is infinite, then V' need not be isomorphic to V. Rather V' and V obey the principle of complementarity. For example, in the theory of distributions (outlined in Appendix F), in which V is a vector space of test functions, and V' is the dual space of linear continuous functionals, the larger we make the space of test functions V, the smaller is the corresponding space of distributions V'.

Similarly, a *bilinearfunctional* is defined as a mapping $\omega : V \times V \to \mathbb{R}$ such that

$$\omega(\mathbf{a}+\mathbf{x},\mathbf{b}) = \omega(\mathbf{a},\mathbf{b}) + \omega(\mathbf{x},\mathbf{b}), \quad \omega(\kappa\mathbf{a},\mathbf{b}) = \kappa\,\omega(\mathbf{a},\mathbf{b}) \; , \tag{1.54}$$

and the same for the second argument.

In a particular basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$, a bilinear functional ω is specified by its components $\omega(\mathbf{e}_i, \mathbf{e}_j) = \omega_{ij}$, and so

$$\omega(\mathbf{a},\mathbf{b}) = \omega_{ij} a^i b^j \; . \tag{1.55}$$

A bilinear functional is *symmetric* if

$$\sigma(\mathbf{a},\mathbf{b}) = \sigma(\mathbf{b},\mathbf{a}) \; , \tag{1.56}$$

and *antisymmetric* if

$$\alpha(\mathbf{a},\mathbf{b}) = -\alpha(\mathbf{b},\mathbf{a}) \; . \tag{1.57}$$

It is clear that components of a symmetric bilinear functional are symmetric,

$$\sigma_{ij} = \sigma_{ji} \; , \tag{1.58}$$

while components of an antisymmetric bilinear functional are antisymmetric,

$$\alpha_{ij} = -\alpha_{ji} \; . \tag{1.59}$$

One may define trilinear, and, generally, multilinear functionals in a similar fashion.

We now turn to point sets which are isomorphic to vector spaces.

Definition 4. *Let V be a vector space. A set $\mathcal{A}$ is said to be an affine space associated with V if there is a correspondence between any ordered pair of elements A and B of $\mathcal{A}$, called points, and a vector, indicated by $\overrightarrow{AB}$, such that:*

10°. For any point A, called the *initial* point, and an arbitrary vector $\mathbf{a}$, there exists a point B, called the *terminal* point, for which $\overrightarrow{AB} = \mathbf{a}$,

11°. For arbitrary points A, B, C, the *triangle* equation holds $\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}$.

If $A = B = C$, then the triangle equation becomes $\overrightarrow{AA} = \mathbf{0}$. If $A = C$, then $\overrightarrow{AB} = -\overrightarrow{BA}$.

An affine space $\mathcal{A}$ associated with V_n has dimension n. Affine spaces are suitable to study affine objects: points, straight lines, planes, etc. A central construction is the affine coordinate system.

Definition 5. *Let O be a fixed point in $\mathcal{A}$. For every point A there is a vector $\overrightarrow{OA}$ which is called radius vector of the point A. Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis of the associated vector space V. Coordinates $a^1, \dots, a^n$ of $\overrightarrow{OA}$ in this basis are called affine coordinates of the point A. The affine coordinate system is comprised of the point O, the origin, and the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$.*

Since $\overrightarrow{OB} = \overrightarrow{OA} + \overrightarrow{AB}$, the vector $\overrightarrow{AB}$ has coordinates $b^1 - a^1, \dots, b^n - a^n$. We see that $\mathcal{A}$ is isomorphic to $\mathbb{R}_n$ where $(0, \dots, 0)$ plays the role of the origin, and affine coordinates of any point are given by coordinates of the radius vector of this point.

Axioms 1°–11° form the basis of affine geometry. Varying the affine coordinate system engenders a transformation of affine coordinates

$$a'^i = M^i{}_j \, a^j + c^i \, . \tag{1.60}$$

Let us turn to Euclidean geometry. This geometry can be built by grafting a metric structure onto the theory of affine spaces.

Definition 6. *For any two vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *we define the scalar product, which is a real bilinear functional, denoted by* $\mathbf{a} \cdot \mathbf{b}$*, such that*

12°. $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$,
13°. $\mathbf{a} \cdot (\kappa \, \mathbf{b} + \lambda \, \mathbf{c}) = \kappa \, (\mathbf{a} \cdot \mathbf{b}) + \lambda \, (\mathbf{a} \cdot \mathbf{c})$,
14°. For every nonzero vector $\mathbf{a}$ there is a vector $\mathbf{b}$ such that $\mathbf{a} \cdot \mathbf{b} \neq 0$.

In addition, one may require that the *norm* of every nonzero vector $\mathbf{a}^2 = \mathbf{a} \cdot \mathbf{a}$ be positive:

15°. $\mathbf{a}^2 > 0$.

A vector space V with a particular scalar product satisfying axioms 12°–15° is called a *Euclidean* vector space. If axiom 15° is abandoned, then the norm is *indefinite*, and axioms 1°–14° comprise geometry of a *pseudoeuclidean* vector space.

Vectors $\mathbf{a}$ and $\mathbf{b}$ are called *orthogonal* if $\mathbf{a} \cdot \mathbf{b} = 0$. For orthogonal $\mathbf{a}$ and $\mathbf{b}$, we have

$$(\mathbf{a} + \mathbf{b})^2 = \mathbf{a}^2 + \mathbf{b}^2 \, . \tag{1.61}$$

If an affine space $\mathcal{A}$ is associated with some Euclidean vector space, then $\mathcal{A}$ is called a *Euclidean* affine space. The terms 'parallel' and 'perpendicular' are often used in place of the terms 'collinear' and 'orthogonal', respectively. Let $\mathbf{a} = \overrightarrow{OA}$ be perpendicular to $\mathbf{b} = \overrightarrow{OB}$. Then, denoting $\mathbf{b} - \mathbf{a} = \overrightarrow{AB}$, (1.61) becomes the Pythagorean theorem

$$\overrightarrow{AB}^2 = \overrightarrow{OA}^2 + \overrightarrow{OB}^2 \, . \tag{1.62}$$

In a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$, the scalar product can be written

$$g(\mathbf{a}, \mathbf{b}) = g_{ij} \, a^i b^j \, , \tag{1.63}$$

where

$$g_{ij} = g(\mathbf{e}_i, \mathbf{e}_j) = \mathbf{e}_i \cdot \mathbf{e}_j \ . \tag{1.64}$$

By axiom 12°,

$$g_{ij} = g_{ji} \ . \tag{1.65}$$

According to axiom 14°, there is no nonzero vector **a** orthogonal to all vectors. Assume that some vector **a** is orthogonal to an arbitrary vector **b**:

$$g_{ij}\, a^i b^j = 0 \ . \tag{1.66}$$

Then the coefficient of b^j is zero for every j,

$$g_{ij}\, a^i = 0 \ . \tag{1.67}$$

A nontrivial solution to this equation is ensured by

$$\det (g_{ij}) = 0 \ . \tag{1.68}$$

Thus, axiom 14° implies that

$$\det (g_{ij}) \neq 0 \ . \tag{1.69}$$

This is the same as saying that there is a matrix inverse of g_{ij},

$$g_{ij}\, g^{jk} = \delta^k_{\ i} \ . \tag{1.70}$$

Both g_{ij} and g^{ij} are collectively called the *metric*. Using g_{ij} and g^{ij}, one can turn vectors and covectors to each other,

$$a^i = g^{ij} a_j, \quad a_i = g_{ij} a^j \ . \tag{1.71}$$

If the index of a covector a_i is raised by g^{ij}, and the index of the resulting vector a^i is lowered by g_{ij}, as shown in (1.71), then we revert to the original covector a_i. Let a^j be the components of a vector **a**. Then a_i stemming from a^j may be interpreted as the scalar product of **a** and the ith basis vector $\mathbf{e}_i$:

$$a_i = g_{ij} a^j = (\mathbf{e}_i \cdot \mathbf{e}_j)\, a^j = \mathbf{e}_i \cdot (a^j \mathbf{e}_j) = \mathbf{e}_i \cdot \mathbf{a} \ . \tag{1.72}$$

We will often ignore the distinction between Euclidean vectors and covectors, calling them *vectors*. We will use the same letter for both a_i and a^i, keeping in mind as a single object, a vector **a**, which is labelled by either *covariant* (lower) or *contravariant* (upper) index.

In Euclidean space, it is convenient to use the so-called *orthonormal* basis, which corresponds to the Cartesian coordinate system. Let us convert a basis $\mathbf{f}_1, \ldots, \mathbf{f}_n$ into another basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ composed of mutually orthogonal *unit* vectors:

$$\mathbf{e}_i^2 = 1, \quad \mathbf{e}_i \cdot \mathbf{e}_j = 0 \ . \tag{1.73}$$

This can be briefly written

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij} \ , \tag{1.74}$$

where δ_{ij} is the Kronecker delta.

We first *normalize* the vector $\mathbf{f}_1$, that is, construct

$$\mathbf{e}_1 = \frac{\mathbf{f}_1}{\sqrt{\mathbf{f}_1^2}} \ , \tag{1.75}$$

which is a desired unit vector, $\mathbf{e}_1^2 = 1$.

We then identify the part of $\mathbf{f}_2$ which is orthogonal to $\mathbf{e}_1$. That is, we write $\mathbf{f}_2$ as the sum of two vectors. One of them

$$(\mathbf{e}_1 \cdot \mathbf{f}_2)\, \mathbf{e}_1 \tag{1.76}$$

is parallel to $\mathbf{e}_1$, and the other

$$\mathbf{f}_2 - (\mathbf{e}_1 \cdot \mathbf{f}_2)\, \mathbf{e}_1 \tag{1.77}$$

is orthogonal to $\mathbf{e}_1$. We eliminate $(\mathbf{e}_1 \cdot \mathbf{f}_2)\, \mathbf{e}_1$ from this sum to obtain a vector orthogonal to $\mathbf{e}_1$. We normalize the resulting vector, which gives a unit vector $\mathbf{e}_2$ orthogonal to $\mathbf{e}_1$.

In general, proceeding from an arbitrary vector $\mathbf{x}$, one can find a vector $\mathbf{x}_\perp$ orthogonal to a nonzero vector $\mathbf{a}$ with the help of the formula

$$\mathbf{x}_\perp = \mathbf{x} - \frac{\mathbf{x} \cdot \mathbf{a}}{\mathbf{a}^2}\, \mathbf{a} \ . \tag{1.78}$$

Introducing the operator $\overset{a}{\perp}$ which acts on $\mathbf{x}$ as

$$(\overset{a}{\perp} \mathbf{x})_i = \left(g_{ij} - \frac{a_i\, a_j}{a^2} \right) x^j = x_i - \frac{a_j\, x^j}{a^2}\, a_i \ , \tag{1.79}$$

we bring (1.78) to the form

$$\mathbf{x}_\perp = \overset{a}{\perp} \mathbf{x} \ . \tag{1.80}$$

With this observation, we continue the orthonormalization process by taking

$$\overset{e_1}{\perp} \overset{e_2}{\perp} \mathbf{f}_3 \tag{1.81}$$

and normalizing this vector to give a unit vector $\mathbf{e}_3$ orthogonal to $\mathbf{e}_1$ and $\mathbf{e}_2$, and so on.

The advantages of an orthonormalized basis are apparent. It is seen from (1.74) that $g_{ij} = \delta_{ij}$. Thus, for any orthonormalized basis, the metric is given by the $n \times n$ unit matrix. Covariant and contravariant coordinates of an arbitrary vector are now identical:

$$a_i = g_{ij} a^j = \delta_{ij} a^j = a^i \ . \tag{1.82}$$

The scalar product is diagonal:

$$\mathbf{a}\cdot\mathbf{b}=\delta^{ij}a_ib_j=a_1b_1+\cdots+a_nb_n\ , \tag{1.83}$$

in particular,

$$\mathbf{a}^2=a_1^2+\cdots+a_n^2\ . \tag{1.84}$$

What is the authomorphism group of the n-dimensional Euclidean affine space? We first remark that the automorphism group contains the translation group T_n as a subgroup, because $g_{ij}(a^i-b^i)(c^j-d^j)$ is invariant under translations of the origin O.

Let the origin be fixed. Consider transformations of one orthonormalized basis to another orthonormalized basis. Coordinates of a vector transform as

$$a'_i=L^j_{\ i}\,a_j\ , \tag{1.85}$$

but the norm remains invariant,

$$a'_i\,a'_j\,\delta^{ij}=L^k_{\ i}L^l_{\ j}\,\delta^{ij}a_ka_l=a_ia_j\delta^{ij}\ . \tag{1.86}$$

Therefore,

$$L^k_{\ i}L^l_{\ i}=\delta^{kl}\ . \tag{1.87}$$

It follows from this equation that the inverse of L is the transpose of this matrix:

$$L^{-1}=L^T\ . \tag{1.88}$$

Matrices satisfying (1.88) form a continuous group. To see this, we note that

$$(L_1L_2)^{-1}=L_2^{-1}L_1^{-1}=L_2^TL_1^T=(L_1L_2)^T\ . \tag{1.89}$$

This demonstrates the group property. The existence of the inverse of L, and a unit matrix are evident. We thus arrive at a $\frac{1}{2}\,(n-1)\,n$-parameter Lie group which is called the *orthogonal* group $\mathrm{O}(n)$ (see Appendix B).

By (1.87),

$$(\det L)^2=1\ . \tag{1.90}$$

Transformations which are continuously connected with identity have a unit determinant $\det L=1$. Such transformations, called *rotations*, form a subgroup $\mathrm{SO}(n)$ of the group $\mathrm{O}(n)$. Transformations whose determinant is equal to -1 are a composition of rotations and reflections

$$\mathbf{e}_i\to-\mathbf{e}_i\ . \tag{1.91}$$

We see that the authomorphism group of Euclidean affine space consists of orthogonal transformations and translations.

We now turn to pseudoeuclidean spaces. Since our interest is with the case $n=4$, we consider a basis $\mathbf{f}_0,\mathbf{f}_1,\mathbf{f}_2,\mathbf{f}_3$ (following a common practice, the subscripts begin with 0.)

In pseudoeuclidean spaces, the norm of a nonzero vector can be zero. Such vectors are called *null* vectors. It is clear that there are vectors different from null vectors. Suppose that all vectors, specifically $\mathbf{a}$ and $\mathbf{b}$, are null vectors. Then $(\mathbf{a}+\mathbf{b})^2 = 0$ and $(\mathbf{a}-\mathbf{b})^2 = 0$. Combined with $\mathbf{a}^2 = 0$ and $\mathbf{b}^2 = 0$, this results in $\mathbf{a}\cdot\mathbf{b} = 0$, contrary to axiom 14°.

Assume that $\mathbf{f}_0, \mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3$ are linearly independent vectors with finite norms. One can render these basis vectors mutually orthogonal using the operator (1.79). However, the normalization procedure is modified. If $\mathbf{f}_0^2 > 0$, then this vector can be normalized as before to yield a unit vector $\mathbf{e}_0$. If $\mathbf{f}_0^2 < 0$, then we take

$$\mathbf{e}_0 = \frac{\mathbf{f}_0}{\sqrt{-\mathbf{f}_0^2}} , \tag{1.92}$$

which is an *imaginary unit* vector

$$\mathbf{e}_0^2 = -1 . \tag{1.93}$$

One can show (Problem 1.2.5) that the vector $\overset{e_0}{\perp} \mathbf{f}_1$ has a finite norm, and hence can be normalized to give either unit or imaginary unit vector $\mathbf{e}_1$. Likewise, $\overset{e_0}{\perp}\overset{e_1}{\perp} \mathbf{f}_2$ is used to obtain a normalized vector $\mathbf{e}_2$, and $\overset{e_0}{\perp}\overset{e_1}{\perp}\overset{e_2}{\perp} \mathbf{f}_3$ is suitable to build a normalized vector $\mathbf{e}_3$.

This process culminates in k unit and $4-k$ imaginary unit basis vectors. For a given space, k is a fixed number for any choice of $\mathbf{f}_0, \mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3$. Indeed, suppose that there are two orthonormalized bases $\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ and $\bar{\mathbf{e}}_0, \bar{\mathbf{e}}_1, \bar{\mathbf{e}}_2, \bar{\mathbf{e}}_3$ such that the first basis contains k unit vectors while the second basis contains $\bar{k}$ unit vectors, and $k > \bar{k}$. Consider a set composed of k unit vectors of the first basis $\mathbf{e}_0, \ldots, \mathbf{e}_k$, and $4-\bar{k}$ imaginary unit vectors of the second basis $\bar{\mathbf{e}}_0, \ldots, \bar{\mathbf{e}}_{3-\bar{k}}$. Since the set contains $4+k-\bar{k} > 4$ vectors, these vectors must be linearly dependent:

$$\lambda_0\, \mathbf{e}_0 + \cdots + \lambda_k\, \mathbf{e}_k + \mu_0\, \bar{\mathbf{e}}_0 + \cdots + \mu_{3-\bar{k}}\, \bar{\mathbf{e}}_{3-\bar{k}} = 0 . \tag{1.94}$$

Rewrite this equation in the form

$$\lambda_0\, \mathbf{e}_0 + \cdots + \lambda_k\, \mathbf{e}_k = -\mu_0\, \bar{\mathbf{e}}_0 - \ldots - \mu_{3-\bar{k}}\, \bar{\mathbf{e}}_{3-\bar{k}} , \tag{1.95}$$

and take the square of both sides:

$$\lambda_0^2 + \cdots + \lambda_k^2 = -\mu_0^2 + \cdots - \mu_{3-\bar{k}}^2 . \tag{1.96}$$

Because $\lambda_0, \ldots, \lambda_k$ and $\mu_0, \ldots, \mu_{3-\bar{k}}$ are real coefficients, (1.96) implies

$$\lambda_0 = \ldots = \lambda_k = 0, \quad \mu_0 = \ldots = \mu_{3-\bar{k}} = 0 , \tag{1.97}$$

which runs counter to the fact that these vectors are linearly dependent. We thus have a contradiction to the initial supposition that $k > \bar{k}$.

Evidently a space with k unit and $4-k$ imaginary unit basis vectors is equivalent in metrical properties to a space with $4-k$ unit and k imaginary unit basis vectors. But the difference between the numbers of unit and imaginary unit basis vectors $\sigma = 4 - 2k$, called the *signature*, is an essential characteristic of the metric. There are only two types of four-dimensional pseudoeuclidean spaces which are characterized by either $\sigma = 2$ or $\sigma = 0$. We will denote them, respectively, as $\mathbb{R}_{1,3}$ and $\mathbb{R}_{2,2}$. The case $\sigma = 4$ (or $\sigma = -4$) corresponds to the Euclidean space $\mathbb{E}_4$. Minkowski space is a four-dimensional pseudoeuclidean space characterized by $\sigma = 2$ (or $\sigma = -2$). We define the scalar product in Minkowski space as

$$a_0 b_0 - a_1 b_1 - a_2 b_2 - a_3 b_3 \ , \tag{1.98}$$

and introduce a special designation $\mathbb{M}_4$, which will be used interchangeably with $\mathbb{R}_{1,3}$. Pseudoeuclidean spaces of the second type, $\mathbb{R}_{2,2}$, seem to be of little importance in physics.

Problem 1.2.1. Prove that $0\mathbf{a} = \mathbf{0}$, and $(-1)\,\mathbf{a} = -\mathbf{a}$.

Proof

$$0\mathbf{a} = (0+0)\mathbf{a} = 0\mathbf{a} + 0\mathbf{a} \quad \Longrightarrow \quad 0\mathbf{a} = \mathbf{0} \ , \tag{1.99}$$

$$\mathbf{a} + (-1)\mathbf{a} = (1-1)\mathbf{a} = 0\mathbf{a} = \mathbf{0} \quad \Longrightarrow \quad (-1)\,\mathbf{a} = -\mathbf{a} \ . \tag{1.100}$$

Problem 1.2.2. Prove that transformations (1.37) form a continuous group.

Hint Apart from $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and $\mathbf{e}'_1, \ldots, \mathbf{e}'_n$, consider one more basis $\mathbf{e}''_1, \ldots, \mathbf{e}''_n$ such that $\mathbf{e}''_i = K^j{}_i\, \mathbf{e}'_j$. By (1.37), $\mathbf{e}''_i = K^k{}_i\, L^j{}_k\, \mathbf{e}_j$. We thus have a transformation from $\mathbf{e}_1, \ldots, \mathbf{e}_n$ to $\mathbf{e}''_1, \ldots, \mathbf{e}''_n$ with the transformation matrix $K^k{}_i\, L^j{}_k$ which is the product of matrices $K^k{}_i$ and $L^j{}_k$. Taking $\mathbf{e}''_1, \ldots, \mathbf{e}''_n$ to be identical to $\mathbf{e}_1, \ldots, \mathbf{e}_n$, we find $K^k{}_i\, L^j{}_k = \delta^j{}_i$, which shows that there is a transformation inverse of a given transformation $L^j{}_i$. Any $n \times n$ matrix L with $\det L > 0$ can be obtained from the $n \times n$ unit matrix by a continuous variation of matrix elements.

Problem 1.2.3. Verify (1.53).

Problem 1.2.4. Let $\mathbf{a}$ and $\mathbf{b}$ be arbitrary vectors in an Euclidean space. Prove the inequality

$$(\mathbf{a} \cdot \mathbf{b})^2 \leq \mathbf{a}^2\, \mathbf{b}^2 \ , \tag{1.101}$$

which is called the *Cauchy–Schwarz–Bunyakowskiĭ inequality*. Show that (1.101) becomes equality only if the vectors $\mathbf{a}$ and $\mathbf{b}$ are collinear.

Hint Consider $f(t) = (\mathbf{a} + t\mathbf{b})^2$ which, in view of axiom 15°, is positive definite.

Problem 1.2.5. Let $\mathbf{f}_0, \mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3$ be linearly independent vectors with finite norms. Show that the orthonormalization procedure discussed in the text leads to a basis which does not contain null vectors.

1.3 Vectors, Tensors, and n-Forms

Vectors in Minkowski space are often called *four-vectors*. We denote the components of four-vectors by Greek letters. For example, the four-dimensional radius vector is denoted by x^μ, where μ runs from 0 to 3. The value $\mu = 0$ is associated with time component of a four-vector, while $1, 2, 3$ correspond to space components. We will use boldface characters to designate space components,

$$x^\mu = (x^0, x^1, x^2, x^3), \quad \text{or} \quad x^\mu = (x^0, \mathbf{x}) \,. \tag{1.102}$$

We divide vectors of Minkowski space into three groups: a vector with positive norm is *timelike*; a vector with zero norm is *lightlike* or *null*; a vector with negative norm is *spacelike*. Lorentz transformations leave timelike vectors timelike, null vectors null, and spacelike vectors spacelike. Null vectors drawn from a fixed point O comprise a surface C, called the *light cone*, which separates timelike vectors from spacelike vectors, as in Fig. 1.3. The light cone

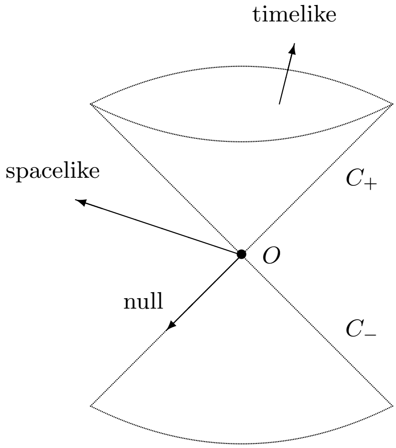


Fig. 1.3. The light cone

involves two sheets: the *forward* sheet C_+ and the *backward* sheet C_- (other names are the *future light cone* and the *past light cone*, respectively).

A timelike vector inside C_+ cannot be transformed into a timelike vector inside C_- by a continuous Lorentz transformation connected with the identity. Therefore, the distinction between timelike vectors pointing to the future and timelike vectors pointing to the past is geometrically valid. The same is true for null vectors.

Given a timelike vector k^μ, it is possible to choose a Lorentz frame where this vector is parallel to the time axis,

$$k^\mu = (k^0, 0, 0, 0) \ . \tag{1.103}$$

If k^μ points to the future, then $k^0 > 0$. In general, a vector orthogonal to a timelike vector k^μ is a linear combination of three mutually orthogonal spacelike vectors, which can be exemplified by $e_1^\mu = (0, 1, 0, 0)$, $e_2^\mu = (0, 0, 1, 0)$, and $e_3^\mu = (0, 0, 0, 1)$. There is no timelike vector orthogonal to another timelike vector.

For any spacelike vector k^μ, one can find a Lorentz frame where

$$k^\mu = (0, \mathbf{k}) \ . \tag{1.104}$$

It is seen from (1.103) and (1.104) that, for an arbitrary timelike or spacelike vector k^μ, there are three mutually orthogonal vectors e_1^μ, e_2^μ, and e_3^μ, which are also orthogonal to k^μ. Taken together these four vectors form an orthonormal basis obeying the *completeness* condition

$$\eta^{\mu\nu} = \frac{k^\mu k^\nu}{k^2} + \sum_{i=1}^{3} \frac{e_i^\mu e_i^\nu}{e_i^2} \ , \tag{1.105}$$

where

$$\eta^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{1.106}$$

is the Minkowski metric. This means that any vector X^μ can be expressed in terms of these basis vectors:

$$X^\mu = \frac{(X \cdot k)}{k^2}\, k^\mu + \sum_{i=1}^{3} \frac{(X \cdot e_i)}{e_i^2}\, e_i^\mu \ . \tag{1.107}$$

A set of vectors of this kind is known as a *vierbein*, or *tetrad*.

One may apply the operator

$$\eta^{\mu\nu} - \frac{k^\mu k^\nu}{k^2} = \sum_{i=1}^{3} \frac{e_i^\mu e_i^\nu}{e_i^2} \tag{1.108}$$

to a vector X^μ to map this vector on

$$\sum_{i=1}^{3} \frac{(X \cdot e_i)}{e_i^2}\, e_i^\mu \ . \tag{1.109}$$

We see that the operator (1.108) projects X^μ onto a three-dimensional subspace spanned by e_1^μ, e_2^μ, e_3^μ. This clarifies the reason for constructing the operator $\overset{k}{\perp}$ in the form (1.79).

Consider a null vector k^μ. There is a Lorentz frame in which

$$k^\mu = \omega\,(1,1,0,0)\ . \tag{1.110}$$

If k^μ points to the future, then $\omega > 0$.

It is clear from (1.110) that any vector q^μ orthogonal to a null vector k^μ is a linear combination of two mutually orthogonal spacelike vectors e_1^μ and e_2^μ, which are exemplified by $e_1^\mu = (0,0,1,0)$ and $e_2^\mu = (0,0,0,1)$, and the vector k^μ itself,

$$q^\mu = \alpha k^\mu + \beta e_1^\mu + \gamma e_2^\mu\ . \tag{1.111}$$

There is no timelike vector orthogonal to a null vector. If q^μ is orthogonal to a null vector k^μ, then q^μ is either spacelike or null (parallel to k^μ). We thus see that a basis may include a null vector, but this basis is not orthogonal.

A simple generalization of vectors and covectors are *tensors*. Algebraically, a tensor T of rank (m,n) is a multilinear mapping

$$T:\quad \underbrace{V'\times\ldots\times V'}_{m\ \text{times}}\times\underbrace{V\times\ldots\times V}_{n\ \text{times}}\to\mathbb{R}\ . \tag{1.112}$$

We have already encountered examples of tensors in the previous section: a scalar is a rank $(0,0)$ tensor, a vector is a rank $(1,0)$ tensor, a covector is a rank $(0,1)$ tensor, the metric g_{ij} is a rank $(0,2)$, while g^{ij} is a rank $(2,0)$ tensor, and the Kronecker delta $\delta^i{}_j$ is a rank $(1,1)$ tensor.

Just as four-vectors can be regarded as objects which transform according to the law

$$a'^\mu = \Lambda^\mu{}_\nu\, a^\nu\ , \tag{1.113}$$

where $\Lambda^\mu{}_\nu$ is the Lorentz transformation matrix relating the two frames of reference, so tensors of rank (m,n) can be described in terms of Lorentz group representations by the requirement that their transformation law be

$$T'^{\mu_1\cdots\mu_m}{}_{\nu_1\cdots\nu_n} = \Lambda^{\mu_1}{}_{\alpha_1}\ldots\Lambda^{\mu_m}{}_{\alpha_m}\Lambda^{\beta_1}{}_{\nu_1}\ldots\Lambda^{\beta_n}{}_{\nu_n}T^{\alpha_1\cdots\alpha_m}{}_{\beta_1\cdots\beta_n}\ . \tag{1.114}$$

We assign to a particular Lorentz frame a collection of numbers

$$T(\varepsilon^{\alpha_1},\ldots,\varepsilon^{\alpha_m};e_{\beta_1},\ldots,e_{\beta_n}) = T^{\alpha_1\cdots\alpha_m}{}_{\beta_1\cdots\beta_n}\ , \tag{1.115}$$

where ε^α is the αth basis covector, and e_β the βth basis vector. The $T^{\alpha_1\cdots\alpha_m}{}_{\beta_1\cdots\beta_n}$ are called the components of T in that frame. The components $T'^{\mu_1\cdots\mu_m}{}_{\nu_1\cdots\nu_n}$ in another frame are obtained from $T^{\alpha_1\cdots\alpha_m}{}_{\beta_1\cdots\beta_n}$ through the use of the transformation law (1.114).

The set of all linear combinations of tensors of a given rank is a vector space. If A is a tensor of rank (p,q) and B is a tensor of rank (k,l), then the *tensor product* $T = A\otimes B$ is defined as a tensor with components $T^{\alpha_1\cdots\alpha_p\mu_1\cdots\mu_k}{}_{\beta_1\cdots\beta_q\nu_1\cdots\nu_l} = A^{\alpha_1\cdots\alpha_p}{}_{\beta_1\cdots\beta_q}B^{\mu_1\cdots\mu_k}{}_{\nu_1\cdots\nu_l}$. It follows from (1.115) and (1.114) that the tensor product T transforms like a tensor of rank $(p+k,q+l)$. To illustrate, we refer to the metric $\eta^{\mu\nu}$ which can be written as the sum of tensor products of basis vectors, given by (1.105),

$$\eta = \frac{k \otimes k}{k^2} + \sum_{i=1}^{3} \frac{e_i \otimes e_i}{e_i^2} \; . \tag{1.116}$$

On the other hand, one can form lower rank tensors from higher rank tensors by the so-called *contraction of indices*. For example, if A is a tensor of rank $(1, 2)$, then the contraction of one upper index with one lower index according to the equation $B_\beta = A^\alpha{}_{\alpha\beta}$ defines a quantity B which transforms like a tensor of rank $(0, 1)$.

There are two important tensors that are invariant under Lorentz transformations. One of them is the Minkowski metric $\eta_{\mu\nu}$. It was argued in the previous section that the Euclidean metric g_{ij} takes a diagonal form in any orthonormalized basis. Likewise, the Minkowski metric $\eta_{\mu\nu}$ has the same components in every Lorentz frame. The contravariant metric tensor $\eta^{\mu\nu}$ (defined by $\eta^{\lambda\mu}\eta_{\mu\nu} = \delta^\lambda{}_\nu$) can be represented by a matrix, shown in (1.106), whose components are identical to those of $\eta_{\mu\nu}$.

The other invariant tensor is the completely antisymmetric rank $(4, 0)$ tensor

$$\epsilon^{\kappa\lambda\mu\nu} = \begin{cases} \;\;1 & \text{if } \kappa\lambda\mu\nu \text{ is an even permutation of } 0123, \\ -1 & \text{if } \kappa\lambda\mu\nu \text{ is an odd permutation of } 0123, \\ \;\;0 & \text{otherwise,} \end{cases} \tag{1.117}$$

which is called the *Levi-Civita tensor*. The complete antisymmetry of a tensor $A^{\kappa\lambda\mu\nu}$ means that $A^{\kappa\lambda\mu\nu}$ is antisymmetric under interchange of any two indices. A completely antisymmetric tensor $A^{\kappa\lambda\mu\nu}$, whose rank is equal to the dimension of spacetime, is reduced essentially to a single component A^{0123}. For coinciding indices, say, $\alpha = \delta$, we have $A^{\alpha\beta\gamma\alpha} = -A^{\alpha\beta\gamma\alpha}$, hence $A^{\alpha\beta\gamma\alpha} = 0$. Every component $A^{\alpha\beta\gamma\delta}$ with different α, β, γ, δ is equal to $\pm A^{0123}$ due to the complete antisymmetry of $A^{\kappa\lambda\mu\nu}$. Putting $A^{0123} = 1$ in a particular Lorentz frame, we obtain $A'^{0123} = 1$ in another Lorentz frame. Indeed, the Levi-Civita tensor $\epsilon^{\alpha\beta\gamma\delta}$ appears in the definition of the determinant of a matrix

$$\det M = M^0{}_\kappa M^1{}_\lambda M^2{}_\mu M^3{}_\nu \, \epsilon^{\kappa\lambda\mu\nu} \; . \tag{1.118}$$

We will see in Sect. 1.5 that the determinant of a Lorentz transformation matrix $\Lambda^\mu{}_\alpha$ is 1, provided that $\Lambda^\mu{}_\alpha$ does not include a spatial reflection. Thus, the components of the Levi-Civita tensor are unchanged under such Lorentz transformations,

$$\epsilon'^{0123} = \Lambda^0{}_\kappa \Lambda^1{}_\lambda \Lambda^2{}_\mu \Lambda^3{}_\nu \, \epsilon^{\kappa\lambda\mu\nu} = \det \Lambda = 1 \; . \tag{1.119}$$

In fact, the Levi-Civita tensor is not a genuine tensor, but rather a *tensor density* whose transformation law is determined (Problem 1.3.4) by the equation

$$\epsilon'^{\alpha\beta\gamma\delta} = (\det \Lambda)^{-1} \, \Lambda^\alpha{}_\kappa \Lambda^\beta{}_\lambda \Lambda^\gamma{}_\mu \Lambda^\delta{}_\nu \, \epsilon^{\kappa\lambda\mu\nu} \; . \tag{1.120}$$

In Minkowski space, vectors and covectors are related by

$$k^\mu = \eta^{\mu\nu} k_\nu, \quad k_\mu = \eta_{\mu\nu} k^\nu \; . \tag{1.121}$$

The contravariant and covariant metric tensors $\eta^{\mu\nu}$ and $\eta_{\mu\nu}$ are used to raise and lower indices on any tensor. For example, the covariant Levi-Civita tensor $\epsilon_{\kappa\lambda\mu\nu}$ is a completely antisymmetric tensor of rank $(0,4)$, which is obtained from the contravariant Levi-Civita tensor $\epsilon^{\kappa\lambda\mu\nu}$ by

$$\epsilon_{\kappa\lambda\mu\nu} = \eta_{\kappa\alpha}\eta_{\lambda\beta}\eta_{\mu\gamma}\eta_{\nu\delta}\epsilon^{\alpha\beta\gamma\delta} \; , \tag{1.122}$$

which implies

$$\epsilon_{0123} = -\epsilon^{0123} \; . \tag{1.123}$$

A special class of tensors, called *n-forms*, contains all completely antisymmetric $(0,n)$ tensors. In a four-dimensional space, there exist only five species of these objects: 0-forms, 1-forms, 2-forms, 3-forms, and 4-forms. As to n-forms with $n > 4$, their components are automatically zero by antisymmetry. One can readily recognize a 0-form as a scalar, and a 1-form as a covector. Every 4-form is proportional to the Levi-Civita tensor, and 3-forms are associated with vectors by the equation $\omega_{\lambda\mu\nu} = \epsilon_{\kappa\lambda\mu\nu}\, k^\kappa$. A simple example of a 2-form ω, which can be constructed from two 1-forms k and q, is given by

$$\omega = \frac{1}{2}\left(k \otimes q - q \otimes k\right) = k \wedge q \; . \tag{1.124}$$

This formula defines the *exterior* product of two 1-forms. One may also consider the exterior product of two vectors,

$$F^{\mu\nu} = \frac{1}{2}\left(a^\mu b^\nu - a^\nu b^\mu\right) \; , \tag{1.125}$$

which is usually called a *bivector*. In general, given a m-form α and a n-form β, we define their exterior product $\alpha \wedge \beta$ as a $(m+n)$-form with components

$$(\alpha \wedge \beta)_{\mu_1 \cdots \mu_{m+n}} = \frac{(m+n)!}{m!\, n!}\, \alpha_{[\mu_1 \cdots \mu_m}\, \beta_{\mu_{m+1} \cdots \mu_{m+n}]} \; . \tag{1.126}$$

Here, the square bracket denote *antisymmetrization*. To antisymmetrize a tensor T of rank $(0,p)$, one takes the alternating sum of all permutations of p indices and divides by the number of terms:

$$T_{[\mu_1 \cdots \mu_p]} = \frac{1}{p!} \sum_{\sigma \in S_p} \operatorname{sgn}(\sigma)\, T_{\mu_{\sigma(1)} \cdots \mu_{\sigma(p)}} \; , \tag{1.127}$$

where S_p is the set of all permutations σ of $0, 1, \ldots, p-1$, and $\operatorname{sgn}(\sigma)$ is 1 if the permutation σ is even, and -1 if the permutation σ is odd. For example,

$$T_{[\lambda\mu\nu]} = \frac{1}{6}\left(T_{\lambda\mu\nu} + T_{\nu\lambda\mu} + T_{\mu\nu\lambda} - T_{\mu\lambda\nu} - T_{\nu\mu\lambda} - T_{\lambda\nu\mu}\right) \; . \tag{1.128}$$

The exterior product is bilinear, and associative,

$$\alpha \wedge (\beta \wedge \gamma) = (\alpha \wedge \beta) \wedge \gamma \,, \tag{1.129}$$

but not commutative. Given a p-form α and a q-form β, we have

$$\alpha \wedge \beta = (-1)^{pq} \beta \wedge \alpha \,. \tag{1.130}$$

Hence odd forms anticommute, and the exterior product of identical 1-forms vanishes.

We will need also the so-called *Hodge duality* operation. Let $\Phi_{\alpha_1 \cdots \alpha_k}$ be a k-form in some d-dimensional vector space. Then the $(d-k)$-form

$$^*\Phi_{\mu_1 \cdots \mu_{d-k}} = \frac{1}{k!}\, \epsilon_{\mu_1 \cdots \mu_{d-k} \alpha_1 \cdots \alpha_k}\, \Phi^{\alpha_1 \cdots \alpha_k} \tag{1.131}$$

is dual of $\Phi_{\alpha_1 \cdots \alpha_k}$. For example, a 2-form $\Phi_{\alpha\beta}$ in a four-dimensional space is mapped by the Hodge star to its dual $^*\Phi_{\mu\nu} = \frac{1}{2}\, \epsilon_{\mu\nu\alpha\beta}\, \Phi^{\alpha\beta}$, which is again a 2-form. In Minkowski space, the recurring Hodge mapping gives $^{**}\Phi = -\Phi$.

Consider an arbitrary 2-form ω in a d-dimensional vector space:

$$\omega = \frac{1}{2} \sum_{i,j=1}^{d} \omega_{ij}\, f^i \wedge f^j \,. \tag{1.132}$$

What should be the set of 1-forms f^i (the index i labels the collection of basis 1-forms, not components of a single vector) to make the form of ω simplest? Suppose that d is even, $d = 2n$, and the rank of the matrix ω_{ij} is equal to $2n$. Then there exists a basis of 1-forms $e^1, e^2, \ldots, e^{2n}$ such that

$$\omega = e^1 \wedge e^2 + e^3 \wedge e^4 + \cdots + e^{2n-1} \wedge e^{2n} \,. \tag{1.133}$$

This writing of ω is called *canonical*. We now sketch the procedure for transforming the initial basis $f^1, f^2, \ldots, f^{2n}$ to the desired basis $e^1, e^2, \ldots, e^{2n}$ taking advantage of the multiplication rules $f^i \wedge f^j = -f^j \wedge f^i$, and $f^i \wedge f^i = 0$. Among the various terms of (1.132), we segregate all terms involving f^1 and f^2. Simple algebra gives the following expression for ω:

$$\left(\omega_{12} f^1 - \omega_{23} f^3 - \omega_{24} f^4 - \ldots - \omega_{2d} f^d\right) \wedge \left(f^2 + \frac{\omega_{13}}{\omega_{12}}\, f^3 + \cdots + \frac{\omega_{1d}}{\omega_{12}}\, f^d \right) + \omega' \,, \tag{1.134}$$

where ω' is a 2-form which involves all terms containing neither f^1 nor f^2. Let us define

$$\begin{aligned} e^1 &= \omega_{12} f^1 - \omega_{23} f^3 - \omega_{24} f^4 - \ldots - \omega_{2d} f^d \,, \\ e^2 &= f^2 + \frac{\omega_{13}}{\omega_{12}}\, f^3 + \cdots + \frac{\omega_{1d}}{\omega_{12}}\, f^d \,. \end{aligned} \tag{1.135}$$

These expressions give just the first pair of 1-forms, e^1 and e^2, appearing in (1.133). We apply a similar trick to ω' to segregate all terms involving f^3 and

f^4, which yields the next desired pair e^3 and e^4. And so on, until the initial set $f^1, f^2, \ldots, f^{2n}$ is exhausted.

For $d = 4$, the canonical decomposition becomes

$$\omega = e^1 \wedge e^2 + e^3 \wedge e^4 \, . \tag{1.136}$$

If the matrix ω_{ij} has a non-maximal rank $2s$, that is, $s < n$, then the canonical decomposition (1.133) terminates at $e^{2s-1} \wedge e^{2s}$. Specifically, for $s = 1$,

$$\omega = e^1 \wedge e^2 \, . \tag{1.137}$$

A 2-form which can be transformed to a single term, the exterior product of two 1-forms is called *decomposable.*

If the dimension is odd, $d = 2n + 1$, then the same canonical decomposition (1.133) as for $d = 2n$ will arise. In particular, any 2-form ω in three-dimensional space can be brought into the form (1.137).

We now turn to *fields.* For simplicity, we consider here only real-valued fields. In the following we will deal also with complex-valued fields, but this generalization is superfluous for the present discussion.

A *scalar field* ϕ is a mapping $\mathbb{M}_4 \to \mathbb{R}$. A scalar field ϕ takes a definite numerical value at each point of Minkowski space, no matter what coordinates are assigned to this point by different observers. Suppose that an observer $\mathcal{O}$ assigns coordinates x to some point, and that a scalar field measured by $\mathcal{O}$ at this point is $\phi(x)$. Another observer $\mathcal{O}'$ will relate the same value of ϕ to coordinates x' of this point, giving $\phi'(x')$. Therefore,

$$\phi'(x') = \phi(x) \, . \tag{1.138}$$

A scalar field ϕ is said to be *Lorentz invariant* if the functional form of ϕ is invariant under Lorentz transformations. Any function of an invariant argument, $\phi(x^2)$, provides an example. We will see in Sect. 10.5 that an important class of invariant functions can be constructed from the four-dimensional Dirac delta-function $\delta^4(x)$.

A *vector field* ϕ^μ is a mapping $\mathbb{M}_4 \to \mathbb{M}_4$. If two observers $\mathcal{O}$ and $\mathcal{O}'$ assign coordinates x and x' of the same point to components of a vector field ϕ^μ and ϕ'^μ, respectively, then

$$\phi'^\mu(x') = \Lambda^\mu{}_\nu \phi^\nu(x) \, . \tag{1.139}$$

The simplest example of vector fields is the four-dimensional gradient of a scalar field $\partial\phi/\partial x^\mu$. The differential operator

$$\partial_\mu = \frac{\partial}{\partial x^\mu} \tag{1.140}$$

transforms like a covariant vector. To see this, we use the chain rule for differentiation:

$$\frac{\partial}{\partial x^\mu} = \frac{\partial x'^\nu}{\partial x^\mu} \frac{\partial}{\partial x'^\nu} \, , \tag{1.141}$$

and note that, for linear coordinate transformations $x'^{\mu} = \Lambda^{\mu}{}_{\nu}\, x^{\nu} + a^{\nu}$,

$$\frac{\partial x'^{\mu}}{\partial x^{\nu}} = \Lambda^{\mu}{}_{\nu}\ . \tag{1.142}$$

We will always use the shorthand notation ∂_μ, and treat this differential operator as an ordinary vector. For example, from ∂_μ we obtain a second-order differential operator

$$\Box = \partial_\mu \partial^\mu\ , \tag{1.143}$$

which is a Lorentz scalar (Problem 1.3.8). In a particular Lorentz frame,

$$\partial_\mu = \left(\frac{\partial}{\partial t}, \nabla\right), \quad \partial^\mu = \left(\frac{\partial}{\partial t}, -\nabla\right), \quad \Box = \frac{\partial^2}{\partial t^2} - \nabla^2\ , \tag{1.144}$$

where ∇ is the well-known operator 'nabla' whose Cartesian coordinates are $\nabla_i = \partial/\partial x^i$.

A tensor-valued function on spacetime is called a *tensor field.* The transformation law for tensor fields is a generalization of (1.139). For example, a second rank tensor transforms according to

$$\phi'^{\lambda\mu}(x') = \Lambda^{\lambda}{}_{\alpha}\Lambda^{\mu}{}_{\beta}\phi^{\alpha\beta}(x)\ . \tag{1.145}$$

The action of ∂_μ on a rank (m, n) tensor field gives a rank $(m, n+1)$ tensor field.

Problem 1.3.1. Let c^μ be the sum of two vectors, $c^\mu = a^\mu + b^\mu$. Show that c^μ can have any norm: $c^2 > 0$, $c^2 < 0$, and $c^2 = 0$, irrespective of whether (a) both a^μ and b^μ are timelike, or (b) both a^μ and b^μ are spacelike, or (c) both a^μ and b^μ are null, or (d) a^μ is timelike, and b^μ is spacelike, or (e) a^μ is timelike, and b^μ is null, or (f) a^μ is spacelike, and b^μ is null. Give examples of a^μ and b^μ for all eighteen cases.

Problem 1.3.2. Let k^μ be an arbitrary (timelike, spacelike, or lightlike) vector, and e_1^μ and e_2^μ are imaginary-unit vectors which are orthogonal to each other and to k^μ: $e_1^2 = e_2^2 = -1$, $e_1 \cdot e_2 = e_1 \cdot k = e_2 \cdot k = 0$. Let n^μ be a timelike unit vector, $n^2 = 1$, which is orthogonal to e_1^μ and e_2^μ. Construct an orthonormalized vierbein from these vectors, and show that the completeness condition for this vierbein reads

$$\eta^{\mu\nu} = \frac{k^\mu k^\nu}{k^2} - \sum_{i=1}^{2} e_i^\mu e_i^\nu + \frac{[k^\mu - (k\cdot n)\, n^\mu]\,[k^\nu - (k\cdot n)\, n^\nu]}{k^2 - (k\cdot n)^2}\ . \tag{1.146}$$

Problem 1.3.3. Let k^μ and q^μ be two null vectors in $\mathbb{M}_4$. Show that (i) $(k+q)\cdot(k-q) = 0$, and $(k+q)^2 = -(k-q)^2$, (ii) if $k \cdot q = 0$, then $k^\mu = Cq^\mu$, where C is a constant. Let k^μ and q^μ be two arbitrary vectors such that $k^2 = q^2$. Show that $(k+q)\cdot(k-q) = 0$.

Problem 1.3.4. Show that (1.118) implies

$$\epsilon_{\alpha\beta\gamma\delta} = (\det \Lambda)^{-1} \Lambda^{\kappa}{}_{\alpha}\Lambda^{\lambda}{}_{\beta}\Lambda^{\mu}{}_{\gamma}\Lambda^{\nu}{}_{\delta}\, \epsilon_{\kappa\lambda\mu\nu} \, . \tag{1.147}$$

Problem 1.3.5. Let $\epsilon^{\kappa\lambda\mu\nu}$ be the Levi-Civita tensor in Minkowski space. Show that the rank $(3,3)$ tensor $\epsilon^{\kappa\lambda\mu\nu}\epsilon_{\kappa\gamma\beta\alpha}$ can be written as

$$-\left(\delta^{\lambda}{}_{\gamma}\delta^{\mu}{}_{\beta}\delta^{\nu}{}_{\alpha} + \delta^{\lambda}{}_{\alpha}\delta^{\mu}{}_{\gamma}\delta^{\nu}{}_{\beta} + \delta^{\lambda}{}_{\beta}\delta^{\mu}{}_{\alpha}\delta^{\nu}{}_{\gamma} - \delta^{\lambda}{}_{\gamma}\delta^{\mu}{}_{\alpha}\delta^{\nu}{}_{\beta} - \delta^{\lambda}{}_{\beta}\delta^{\mu}{}_{\gamma}\delta^{\nu}{}_{\alpha} - \delta^{\lambda}{}_{\alpha}\delta^{\mu}{}_{\beta}\delta^{\nu}{}_{\gamma}\right) \, . \tag{1.148}$$

(The first term in the parenthesis imitates the index order of $\epsilon^{\kappa\lambda\mu\nu}\epsilon_{\kappa\gamma\beta\alpha}$, the second and third terms are produced from it by cyclic permutations of indices while the three remaining are produced from it by interchanging and cyclic permutations of indices.)

Hint Contraction of the remaining indices gives

$$\epsilon^{\kappa\lambda\mu\nu}\epsilon_{\kappa\lambda\beta\alpha} = -2\left(\delta^{\mu}{}_{\beta}\delta^{\nu}{}_{\alpha} - \delta^{\mu}{}_{\alpha}\delta^{\nu}{}_{\beta}\right) \, , \tag{1.149}$$

$$\epsilon^{\kappa\lambda\mu\nu}\epsilon_{\kappa\lambda\mu\alpha} = -3!\, \delta^{\nu}{}_{\alpha} \, , \tag{1.150}$$

$$\epsilon^{\kappa\lambda\mu\nu}\epsilon_{\kappa\lambda\mu\nu} = -4! \tag{1.151}$$

The last is the anticipated result, which proves the statement.

The three-dimensional Levi-Civita symbol ϵ_{ijk} is a completely antisymmetric object, which, in a three-dimensional space, is specified by $\epsilon_{123} = 1$. The use of ϵ_{ijk} enables one to define the *cross product* of two vectors $\mathbf{a}$ and $\mathbf{b}$ by the formula $(\mathbf{a} \times \mathbf{b})_i = \epsilon_{ijk}\, a_j b_k$, where the summation over repeated (lower) indices is understood.

Prove the relations

$$\epsilon_{klm}\epsilon_{kij} = (\delta_{li}\delta_{mj} - \delta_{lj}\delta_{mi}) \, , \tag{1.152}$$

$$\epsilon_{klm}\epsilon_{kli} = 2\delta_{mi} \, , \tag{1.153}$$

$$\epsilon_{klm}\epsilon_{klm} = 3! \tag{1.154}$$

Since the metric of a Euclidean space in Cartesian coordinates is given by the Kronecker delta δ_{ij}, equations (1.152)–(1.154) form the basis for a useful technique for ordinary three-dimensional vector algebra and vector analysis.

Problem 1.3.6. Let $S_{\mu\nu}$ and $A_{\mu\nu}$ be arbitrary symmetric and antisymmetric tensors, respectively. Prove that $S_{\mu\nu}A^{\mu\nu} = 0$. Let $S_{\mu\nu}$ be some rank 2 symmetric tensor. Prove that the general solution to the equation $S_{\mu\nu}X^{\mu\nu} = 0$ is $X^{\mu\nu} = A^{\mu\nu}$ where $A^{\mu\nu}$ is an arbitrary antisymmetric tensor. Let $T_{\mu\nu}$ be an arbitrary tensor. Prove that $T_{\mu\nu}$ can be uniquely represented as $T_{\mu\nu} = S_{\mu\nu} + A_{\mu\nu}$ where $S_{\mu\nu}$ is a symmetric tensor, and $A_{\mu\nu}$ is an antisymmetric tensor.

Problem 1.3.7. Let $G = \frac{1}{2} G_{ij}\, e^i \wedge e^j$ be an arbitrary 2-form in a $2n$-dimensional vector space. Define

$$\mathrm{Pf}(G) = \epsilon^{i_1 \cdots i_{2n}} \, G_{i_1 i_2} \, G_{i_3 i_4} \ldots G_{i_{2n-1} \, i_{2n}} \; . \tag{1.155}$$

This scalar is called the *Pfaffian*. Prove the relation

$$[\mathrm{Pf}(G)]^2 = 2^{2n} (n!)^2 \det G \; . \tag{1.156}$$

Proof Let $A^i{}_j$ be the matrix which brings G to the canonical form

$$G_{ij} = A^k{}_i \, A^l{}_j \, E_{kl} \; , \tag{1.157}$$

where

$$E_{ij} = \begin{pmatrix} 0 & 1 & \\ -1 & 0 & \\ & & \ddots \end{pmatrix} \tag{1.158}$$

In view of (1.155),

$$\mathrm{Pf}(G) = \epsilon^{i_1 \cdots i_{2n}} \, A^{k_1}_{\;i_1} A^{k_2}_{\;i_2} \ldots A^{k_{2n-1}}_{\;i_{2n-1}} A^{k_{2n}}_{\;i_{2n}} \, E_{k_1 k_2} \ldots E_{k_{2n-1} k_{2n}} \; . \tag{1.159}$$

The relation

$$\epsilon^{i_1 \cdots i_{2n}} \, A^{k_1}_{\;i_1} \ldots A^{k_{2n}}_{\;i_{2n}} = \epsilon^{k_1 \cdots k_{2n}} \det A \; , \tag{1.160}$$

shows then that

$$\mathrm{Pf}(G) = \mathrm{Pf}(E) \, \det A \; . \tag{1.161}$$

Combined with

$$\det G = (\det A)^2 \, \det E \tag{1.162}$$

equation (1.161) gives

$$\det G = \left[\frac{\mathrm{Pf}(G)}{\mathrm{Pf}(E)} \right]^2 \det E \; . \tag{1.163}$$

It is clear from (1.158) that $\det E = 1$. $\mathrm{Pf}(E)$ contains $n!$ terms arising from $\epsilon^{1 \cdots 2n} \, E_{12} \ldots E_{2n-1\,2n}$ through permutations of factors E, changing the overall sign, which must be attended with 2^n interchanges of indices in each E, reversing the signs of E and ϵ simultaneously but leaving the overall sign unaltered. Therefore,

$$\mathrm{Pf}(E) = 2^n \, n! \, (-1)^n \; , \tag{1.164}$$

which, together with (1.163), proves the statement.

Problem 1.3.8. Show that the differential operator $\Box$ defined in (1.143) transforms as a Lorentz scalar.

Hint

$$\eta_{\nu\rho} \, \eta^{\lambda\mu} \, \frac{\partial x^\nu}{\partial x'^\mu} \, \frac{\partial x^\rho}{\partial x'^\lambda} = \eta_{\nu\rho} \, \eta^{\lambda\mu} \, \Lambda^\nu{}_\mu \Lambda^\rho{}_\lambda \; . \tag{1.165}$$

1.4 Lines and Surfaces

In this section we discuss some basic geometric objects in Minkowski space.

We begin with affine manifolds, that is, m-dimensional subspaces of a n-dimensional affine space. Consider the equation

$$\phi(x) = C \ , \tag{1.166}$$

where ϕ is a linear functional, and C a constant. The argument of ϕ is regarded as radius vector x^i drawn from the origin. Equation (1.166) can be written in coordinate form:

$$\phi_i \, x^i = C \ . \tag{1.167}$$

This equation defines a $(n-1)$-dimensional affine manifold Σ called a *hyperplane.*

One may then equip the affine space with a Euclidean metric by defining the scalar product. When introducing a unit covector n_i parallel to ϕ_i,

$$\phi_i = k \, n_i, \quad n^2 = 1 \ , \tag{1.168}$$

(1.167) becomes

$$n \cdot x = D \ . \tag{1.169}$$

The vector n_i is said to be *normal* to the hyperplane Σ. Equation (1.169) has a clear geometric interpretation. If $D = 0$, then Σ consists of vectors x^i perpendicular to n_i. Let $D \neq 0$. Then D measures the distance between Σ and the origin along an axis parallel to n_i. Indeed, $n \cdot x$ is the projection of the radius vectors x^i (drawn from the origin to Σ) on this axis. By (1.169), this projection takes a constant value, D, for every x^i.

We are thus led to an alternative definition of hyperplanes. A hyperplane with normal n_i is a locus of points x^i obeying the equation

$$x^i = D \, n^i \, + (\overset{n}{\perp} x)^i \ , \tag{1.170}$$

where $\overset{n}{\perp}$ is the projection operator defined in (1.79).

Let $\phi_i^1, \ldots, \phi_i^m$ be m linearly independent covectors. We define a $(n-m)$-dimensional affine manifold $\mathcal{A}_{n-m}$ by the set of m linear equations

$$\phi_j^s \, x^j = C^s, \quad s = 1, \ldots, m \ . \tag{1.171}$$

In particular, a one-dimensional affine manifold $\mathcal{A}_1$, which is actually a straight line, is defined by a set of $n-1$ equations of the type (1.171).

On the other hand, a straight line may be thought of as a linear mapping $z^i : \mathbb{R} \to \mathcal{A}_n$, or, in coordinate notation,

$$z^i(\tau) = z_0^i + V^i \tau \ , \tag{1.172}$$

where z_0^i and V^i are some constant vectors. This straight line runs through the point z_0^i in the direction of the vector V^i.

Let $V_1^i, \dots, V_m^i$ be m linearly independent vectors which span a m-parameter affine manifold $\mathcal{A}_m$ passing through a point z_0^i. This manifold is described by a linear mapping $z^i : \mathbb{R}_m \to \mathcal{A}_n$, such that

$$z^i(\tau_1, \dots, \tau_m) = z_0^i + V_1^i \tau_1 + \dots + V_m^i \tau_m \ . \tag{1.173}$$

We now turn to Minkowski space $\mathbb{M}_4$. We will distinguish between spacelike, timelike, and null hyperplanes Σ. If n^μ is timelike, then Σ is spacelike. It is clear from (1.170) that a spacelike hyperplane Σ is characterized by the fact that any two points on it, $x, y \in \Sigma$, are separated by a spacelike interval, $(x-y)^2 < 0$. Therefore, any spacelike hyperplane may be regarded as 'all of space at a given time' in some Lorentz frame. If $n^2 < 0$, then Σ is timelike. One may visualize a timelike hyperplane as a world volume of a uniformly moving membrane. If $n^2 = 0$, then Σ is a null hyperplane.

To extend these definitions to curved manifolds, we note that a smooth curved manifold $\mathcal{M}_m$ can be approximated by an affine manifold $\mathcal{A}_m$ in the vicinity of each point $x \in \mathcal{M}_m$.

A *curve* is a smooth mapping $z^\mu : \mathbb{R} \to \mathbb{M}_4$. If we write $\tau = \tau_0 + \sigma$, and consider $z^\mu(\tau)$ for a small interval σ, then this curve is approximated by a straight line

$$z_0^\mu + V^\mu \sigma \ , \tag{1.174}$$

where the point z_0^μ refers to the instant τ_0. The direction of this straight line is determined by the vector

$$V^\mu = \frac{dz^\mu}{d\tau} = \dot{z}^\mu \ . \tag{1.175}$$

The vector $\dot{z}^\mu$ given by (1.175) is said to be *tangent* to the curve $z^\mu(\tau)$ at $\tau = \tau_0$.

We define a *hypersurface* $\mathcal{M}_{n-1}$ by

$$F(x) = C \ , \tag{1.176}$$

where F is an arbitrary smooth function $\mathbb{M}_4 \to \mathbb{R}$. Differentiating (1.176) gives

$$(\partial_\mu F)\, dx^\mu = 0 \ . \tag{1.177}$$

One may view dx^μ as a covector, and $\partial_\mu F$ as a vector. Indeed, dx^μ transforms like a covector under linear coordinate transformations $x'^\mu = \Lambda^\mu{}_\nu\, x^\nu + a^\nu$,

$$dx'^\mu = \frac{\partial x'^\mu}{\partial x^\nu}\, dx^\nu = \Lambda^\mu{}_\nu\, dx^\nu \ , \tag{1.178}$$

and $\partial_\mu F$ transforms like a vector:

$$\frac{\partial F}{\partial x'^\mu} = \frac{\partial x^\nu}{\partial x'^\mu}\,\frac{\partial F}{\partial x^\nu} = \Lambda^\nu{}_\mu\, \frac{\partial F}{\partial x^\nu} \ . \tag{1.179}$$

In Minkowski space, vectors and covectors can be converted to each other according to (1.121). For this reason, we will often regard dx^μ as vectors.

We see that (1.177) describes a hyperplane Σ with normal $\partial_\mu F$. Since the hyperplane Σ is spanned by the vectors dx^μ, it would be natural to call them *tangent vectors*, and Σ the hyperplane *tangent* to the hypersurface $\mathcal{M}_{n-1}$ at the point x. A hypersurface is said to be locally spacelike, timelike, or null, according to which case $(\partial_\mu F)^2 > 0$, $(\partial_\mu F)^2 < 0$, or $(\partial_\mu F)^2 = 0$ occurs at this point.

A m-parameter manifold $\mathcal{M}_m$ generalizes the idea of a parametrized curve. It can be defined as a smooth mapping $\mathbb{R}_m \to \mathbb{M}_4$. A manifold $\mathcal{M}_m$ looks like a m-dimensional affine space in the vicinity of each point. A general manifold $\mathcal{M}_m$ is topologically nontrivial, and its complete description requires a set of overlapping coordinate patches covering $\mathcal{M}_m$.

Although we will ordinarily refer to Cartesian coordinates, curvilinear coordinates occasionally prove useful. To specify the axes of a curvilinear system, it is often convenient to introduce the so-called 'level surfaces', following the pattern seen in (1.176). Let F^α be a smooth function of spacetime. Then $F^\alpha(x) = C$ fixes a hypersurface on which a curvilinear coordinate ξ^α is constant, $\xi^\alpha = C$, and $\partial_\mu F^\alpha$ is directed towards the increase of the αth coordinate line. This construction will find use in Sect. 4.5.

We now want to define the surface elemen ton a manifold $\mathcal{M}_m$. Let us first consider a two-dimensional spacelike surface $\mathcal{M}_2$. We draw an infinitesimal parallelogram spanned by spacelike tangent vectors a^μ and b^μ at some point on this surface. It is well known that the area of a parallelogram is the product of lengths of its sides times the sine of the angle between them:

$$A = \sqrt{a^2\, b^2}\,\sin\theta \ . \tag{1.180}$$

Taking into account that $\sqrt{-b^2}\,\sin\theta$ is the projection of the vector b^μ on a direction perpendicular to the vector a^μ, we write (1.180) in an invariant form:

$$A = \sqrt{a^2\,(\overset{a}{\perp} b)^2} \ . \tag{1.181}$$

Let $\mathcal{M}_2$ be a timelike surface. This means that a^μ is timelike, and b^μ spacelike, or the other way around. Then (1.181) is modified:

$$A = \sqrt{-a^2\,(\overset{a}{\perp} b)^2} \ . \tag{1.182}$$

Examples will occur in Sects. 4.6, 4.8, and 5.6.

Combining (1.181) and (1.182), gives

$$A = \sqrt{|a^2\,(\overset{a}{\perp} b)^2|} \ , \tag{1.183}$$

which is universally applicable to spacelike and timelike parallelograms, and even to the case that either of these vectors, a^μ and b^μ, is null. However, if a^μ and b^μ are both null vectors, then (1.183) is no longer valid.

An alternative method of defining the area is as follows. Consider

$$\Omega = \epsilon_{\mu\nu} a^\mu b^\nu \ , \tag{1.184}$$

where $\epsilon_{\mu\nu}$ is the two-dimensional Levi-Civita tensor whose components can be determined from a single one $\epsilon_{12} = 1$. Substitute the identity

$$b^\mu = \frac{b \cdot a}{a^2}\, a^\mu + (\overset{a}{\perp} b)^\mu \tag{1.185}$$

to (1.184), and take into account that $\epsilon_{\mu\nu} a^\mu a^\nu = 0$. The result is

$$\Omega = \epsilon_{\mu\nu} a^\mu (\overset{a}{\perp} b)^\nu \tag{1.186}$$

which is identical to A up to overall sign (Problem 1.4.1).

Note that Ω is a more general geometric characteristic than A. Indeed, the definition (1.184) remains valid even in the case that a^μ and b^μ are both null vectors. Furthermore, Ω not only measures the area of a parallelogram, but also indicates its *orientation*. We may take the convention that the pair of spacelike vectors a^μ and b^μ is positively oriented if the shortest transition from a^μ to b^μ is given by a counterclockwise rotation, and negatively oriented otherwise. It is clear that $\Omega > 0$ in the former case, and $\Omega < 0$ in the latter case. Any continuous deformation of the parallelogram leaves the sign of Ω unchanged. We define the orientation to be the sign of Ω for general vectors a^μ and b^μ.

These definitions extend immediately to higher dimensions. For example, the volume of a three-dimensional parallelepiped spanned by three vectors a^μ, b^μ, c^μ can be defined as

$$A_3 = \left| a^2 \,(\overset{a}{\perp} b)^2\, (\overset{a}{\perp}\overset{b}{\perp} c)^2 \right|^{\frac{1}{2}} \ , \tag{1.187}$$

or, alternatively, as

$$\Omega_3 = \epsilon_{\lambda\mu\nu}\, a^\lambda b^\mu c^\nu \ . \tag{1.188}$$

We see that Ω_3 is the determinant of the components of the three vectors a^μ, b^μ, c^μ. A family of three vectors $e_1^\mu, e_2^\mu, e_3^\mu$ is said to be positively or negatively oriented according to which possibility, $\Omega_3 > 0$ or $\Omega_3 < 0$, pertains.

The oriented area of a parallelogram spanned by two vectors a^μ and b^μ may be thought of as a 2-form Ω. Its value is given by the determinant of the components of the vectors a^μ and b^μ. More generally, one may identify the oriented volume of a p-dimensional parallelepiped spanned by vectors $e_1^\mu, \ldots, e_p^\mu$ with a p-form Ω_p given by the determinant of the components of the vectors $e_1^\mu, \ldots, e_p^\mu$.

There are three types of integrals in three-dimensional Euclidean space which can be taken over: (1) a space region, (2) a surface, and (3) a curve. The volume element is a 3-form

$$df^{ijk} = dx^i \wedge dx^j \wedge dx^k \ , \tag{1.189}$$

which can be exchanged for its *dual* 0-form, defined as

$$d^3x = \frac{1}{3!}\,\epsilon_{ijk}\,df^{ijk} = dx^1 dx^2 dx^3 \ . \tag{1.190}$$

The surface element is a 2-form

$$df^{ij} = dx^i \wedge dx^j \ , \tag{1.191}$$

or, alternatively, its dual 1-form

$$dS_k = \frac{1}{2!}\,\epsilon_{ijk}\,df^{ij} \ . \tag{1.192}$$

(Because the cross product of two vectors is unchanged under space reflections $\mathbf{x} \to -\mathbf{x}$, the surface element $d\mathbf{S} = d\mathbf{u} \times d\mathbf{v}$ is in fact an *axial vector*.) The line element is a 1-form

$$dx^i \ , \tag{1.193}$$

which can be exchanged for its dual 2-form

$$dS_{ij} = \epsilon_{ijk}\,dx^k \ . \tag{1.194}$$

In numerous physical situations we are called on to evaluate volume, surface, and line integrals. There are remarkable divergence and curl theorems known, respectively, as the *Gauss–Ostrogradskiĭ theorem*, and the *Stokes theorem*, which enable one to evaluate the net outflow from any region of space, and the net circulation around any path.

Let $\mathbf{J}$ be a vector field smoothly varying in a compact region V which is enclosed by a smooth surface ∂V. Then the Gauss–Ostrogradskiĭ theorem reads

$$\oint_{\partial V} d\mathbf{S} \cdot \mathbf{J} = \int_V d^3x\ \nabla \cdot \mathbf{J} \ . \tag{1.195}$$

Here, the vector $d\mathbf{S}$ is chosen to point out of V. $\nabla \cdot \mathbf{J}$ stands for the divergence of the vector $\mathbf{J}$, that is, the rate of increase of ‘lines of flow’ per volume.

Let $\mathbf{J}$ be a vector field on a compact surface $\mathcal{S}$ which is bounded by a smooth closed curve $\mathcal{C}$. Then the Stokes theorem reads

$$\oint_{\mathcal{C}} d\mathbf{x} \cdot \mathbf{J} = \int_{\mathcal{S}} d\mathbf{S} \cdot (\nabla \times \mathbf{J}) \ . \tag{1.196}$$

Here, the circulation integral is along $\mathcal{C}$ in a clockwise direction when looking in the direction of $d\mathbf{S}$. $\nabla \times \mathbf{J}$ denotes the curl of $\mathbf{J}$, a measure of the vorticity of the field $\mathbf{J}$.

In Minkowski space, we encounter four types of integrals. The volume element is either a 4-form or its dual 0-form:

$$df^{\kappa\lambda\mu\nu} = dx^\kappa \wedge dx^\lambda \wedge dx^\mu \wedge dx^\nu, \quad d^4x = \frac{1}{4!}\,\epsilon_{\kappa\lambda\mu\nu}\,df^{\kappa\lambda\mu\nu} = dx^0 dx^1 dx^2 dx^3 \ . \tag{1.197}$$

The differential element for hypersurfaces is either a 3-form or its dual 1-form:

$$df^{\lambda\mu\nu} = dx^\lambda \wedge dx^\mu \wedge dx^\nu , \quad d\sigma_\kappa = \frac{1}{3!}\,\epsilon_{\kappa\lambda\mu\nu}\, df^{\lambda\mu\nu} \,. \tag{1.198}$$

The differential element for two-dimensional surfaces is a 2-form

$$df^{\mu\nu} = dx^\mu \wedge dx^\nu \,. \tag{1.199}$$

It would serve no purpose to consider the dual of $df^{\mu\nu}$ which is also a 2-form. The line element is a 1-form dx^ν, or, alternatively, its dual 3-form:

$$dS_{\kappa\lambda\mu} = \epsilon_{\kappa\lambda\mu\nu}\, dx^\nu \,. \tag{1.200}$$

Let J^μ be a smooth vector field in $\mathbb{M}_4$. A four-dimensional generalization of the Gauss–Ostrogradskiĭ theorem reads:

$$\int_{\partial\mathcal{U}} d\sigma_\mu J^\mu = \int_{\mathcal{U}} d^4x\ \partial_\mu J^\mu \,, \tag{1.201}$$

where $\mathcal{U}$ is a spacetime volume bounded by a smooth three-dimensional surface $\partial\mathcal{U}$ with the surface element $d\sigma_\mu$ pointing out of $\mathcal{U}$. A four-dimensional version of the Stokes theorem reads:

$$\int_{\partial\mathcal{S}} dx^\mu J_\mu = \int_{\mathcal{S}} df^{\mu\nu}\left(\partial_\mu J_\nu - \partial_\nu J_\mu\right) \,. \tag{1.202}$$

Problem 1.4.1. Show that the magnitude of Ω as given by (1.186) is identical to the magnitude of A defined in (1.183).

Problem 1.4.2. Let $x'^\mu = f^\mu(x)$ be a smooth, one-to-one mapping of Minkowski space to itself, $\mathbb{M}_4 \to \mathbb{M}_4$, and J the Jacobian of this transformation,

$$J = \det\left(\frac{\partial x'^\mu}{\partial x^\nu}\right) \,. \tag{1.203}$$

The metric transforms according to the law

$$g'_{\mu\nu} = \frac{\partial x^\alpha}{\partial x'^\mu}\,\frac{\partial x^\beta}{\partial x'^\nu}\, g_{\alpha\beta} \,. \tag{1.204}$$

Show that

$$g = \det\left(g_{\mu\nu}\right) \tag{1.205}$$

transforms as

$$g' = J^{-2} g \tag{1.206}$$

1.5 Poincaré Invariance

Linear transformations of Minkowski space

$$x'^{\mu} = \Lambda^{\mu}{}_{\nu}\, x^{\nu} + a^{\mu} \;, \tag{1.207}$$

which preserve the line element

$$dx^2 = \eta_{\mu\nu}\, dx^{\mu} dx^{\nu} \;, \tag{1.208}$$

are called *Poincaré transformations.* The set of transformations (1.207) form a group, the *Poincaré group*, which is the automorphism group of Minkowski space. Elements of this ten-parameter Lie group are labelled with a matrix $\Lambda^{\mu}{}_{\nu}$ and a vector a^{μ}.

The homogeneous part of the Poincaré group consists of *Lorentz transformations*

$$x'^{\mu} = \Lambda^{\mu}{}_{\nu}\, x^{\nu} \;. \tag{1.209}$$

The scalar product of two vectors

$$x \cdot y = \eta_{\mu\nu}\, x^{\mu} y^{\nu} = x^0 y^0 - x^1 y^1 - x^2 y^2 - x^3 y^3 \tag{1.210}$$

is invariant under these transformations. The set of all Lorentz transformations (1.209) form a group which is called the *Lorentz group*, or *pseudoorthogonal group* $\mathrm{O}(1,3)$.

Another subgroup of the Poincaré group involves *translations*

$$x'^{\mu} = x^{\mu} + a^{\mu} \;. \tag{1.211}$$

The difference between two vectors $x^{\mu} - y^{\mu}$ is invariant under translations. Therefore, the full Poincaré group leaves invariant the relative norm $(x-y)^2$, and line element dx^2.

If two points x^{μ} and y^{μ} are separated by a spacelike interval, $(x-y)^2 < 0$, then no signal moving at or below the speed of light can travel between the two points. This is a manifestation of the *causality principle* which forbids cause-effect relations between points separated by spacelike intervals. An event at a point with coordinates x^{μ} could play a role in what happens at a point y^{μ} only if $(y-x)^2 \geq 0$, and $y^0 > x^0$.

The invariance of the scalar product under Lorentz transformations

$$\eta_{\mu\nu}\, x'^{\mu} y'^{\nu} = \eta_{\mu\nu}\, x^{\mu} y^{\nu} \;, \tag{1.212}$$

implies that

$$\Lambda^{\lambda}{}_{\mu}\, \eta_{\lambda\nu}\, \Lambda^{\nu}{}_{\rho} = \eta_{\mu\rho} \;. \tag{1.213}$$

In matrix notation this relation reads

$$\Lambda^T \eta\, \Lambda = \eta \;, \tag{1.214}$$

where

$$\eta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} , \tag{1.215}$$

and Λ^T is the transpose of Λ. Observing that the inverse of η is again η, we write (1.214) in the form

$$\Lambda^{-1} = \eta \, \Lambda^T \, \eta \, . \tag{1.216}$$

Any real symmetric 4×4 matrix has 10 independent entries. Therefore, (1.213) gives 10 constraints on the 16 elements of the matrix Λ. We see that $\Lambda^{\mu}{}_{\nu}$ depends on 6 free parameters, of which 3 describe spatial rotations $\Lambda^{i}{}_{j}$, and the remaining 3 characterize Lorentz boosts $\Lambda^{0}{}_{j}$. For example, a rotation in the (x^1, x^2)-plane around the x^3-axis can be expressed in the form

$$\Lambda^{1}{}_{2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\vartheta & \sin\vartheta & 0 \\ 0 & -\sin\vartheta & \cos\vartheta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} , \tag{1.217}$$

and a boost along the x^1-axis is given by

$$\Lambda^{0}{}_{1} = \begin{pmatrix} \cosh\theta & -\sinh\theta & 0 & 0 \\ -\sinh\theta & \cosh\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{1.218}$$

Together with 4 spacetime translations a^{μ}, these 6 parameters form a set of 10 parameters of the Poincaré group.

By (1.214),

$$(\det \Lambda)^2 = 1 \, , \tag{1.219}$$

and so

$$\det \Lambda = \pm 1 \, . \tag{1.220}$$

Letting $\mu = 0$, and $\rho = 0$ in (1.213), we have

$$(\Lambda^{0}{}_{0})^2 - \sum_{i=1}^{3} (\Lambda^{i}{}_{0})^2 = 1 \, , \tag{1.221}$$

which shows that

$$\Lambda^{0}{}_{0} \geq 1 \, , \tag{1.222}$$

or

$$\Lambda^{0}{}_{0} \leq -1 \, . \tag{1.223}$$

The most important subgroup of the Lorentz group $L^{\uparrow}_{+}$ is specified by two conditions: $\det \Lambda = 1$ and $\Lambda^{0}{}_{0} \geq 1$. All transformations of this subgroup are continuously connected to the identity. This subgroup consists of *proper orthochronous Lorentz transformations*. The term ‘proper’ means $\det \Lambda = 1$, whereas ‘orthochronous’ means $\Lambda^0{}_0 \geq 1$.

An arbitrary transformation of the Lorentz group is a composition of a proper Lorentz transformation and either space reflection P,

$$Px^0 = x^0, \quad P\mathbf{x} = -\mathbf{x} \, , \tag{1.224}$$

or time reversal T,

$$Tx^0 = -x^0, \quad T\mathbf{x} = \mathbf{x} \, , \tag{1.225}$$

or spacetime reflection PT,

$$PTx^0 = -x^0, \quad PT\mathbf{x} = -\mathbf{x} \, . \tag{1.226}$$

These discrete operations can be represented as matrices. Space reflection takes the form

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \, . \tag{1.227}$$

The time reversal matrix T equals $-P$, and the spacetime reflection matrix PT is $-\mathbf{1}$, where $\mathbf{1}$ is the 4×4 unit matrix. The subgroup of the Lorentz group with $\det \Lambda = 1$, or *proper Lorentz transformations*, is denoted by $\mathrm{SO}(1,3)$.

Elements of the Lorentz group in a neighborhood of unity can be written as

$$\Lambda = \exp\left(\frac{i}{2} M^{\mu\nu}\, \omega_{\mu\nu}\right) = \mathbf{1} + \frac{i}{2} M^{\mu\nu}\, \omega_{\mu\nu} \, , \tag{1.228}$$

where $M^{\mu\nu}$ is a 4×4 matrix defined by

$$(M^{\mu\nu})^\alpha{}_\beta = i\left(\eta^{\mu\alpha}\, \delta^\nu{}_\beta - \eta^{\nu\alpha}\, \delta^\mu{}_\beta\right) \, . \tag{1.229}$$

These matrices obey the following commutation relations (Problem 1.5.1)

$$[M^{\kappa\lambda}, M^{\mu\nu}] = -i\left(\eta^{\kappa\mu} M^{\lambda\nu} + \eta^{\nu\kappa} M^{\mu\lambda} + \eta^{\lambda\nu} M^{\kappa\mu} + \eta^{\mu\lambda} M^{\nu\kappa}\right) \, . \tag{1.230}$$

The vector space spanned by linear combinations of the six generators $M^{\mu\nu}$ with real coefficients is the Lie algebra of the Lorentz group, $\mathrm{so}(1,3)$. Elements of $\mathrm{so}(1,3)$ can be realized as differential operators. To see this, we introduce four differential operators

$$P_\mu = -i\, \frac{\partial}{\partial x^\mu} \, , \tag{1.231}$$

which are commuting,

$$[P_\mu, P_\nu] = 0 \, , \tag{1.232}$$

and act on coordinates as

$$P_\mu \, x^\nu = -i \, \delta^\mu_{\ \nu} \, . \tag{1.233}$$

Using (1.232) and (1.233), one can show (Problem 1.5.2) that the six operators

$$M_{\mu\nu} = x_\mu P_\nu - x_\nu P_\mu \tag{1.234}$$

obey commutation relations (1.230).

The differential operators P_μ defined in (1.231) are generators of translations. Indeed, for an arbitrary smooth function Φ, we have

$$\exp\left(i \, P_\nu \, a^\nu\right) \Phi(x) = \Phi(x + a) \, . \tag{1.235}$$

This is readily seen from the Taylor series:

$$\exp\left(a \cdot \partial\right) \Phi(x) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(a \cdot \partial\right)^n \Phi(x) = \Phi(x + a) \, . \tag{1.236}$$

With P_λ and $M_{\mu\nu}$ given by (1.231) and (1.234), one can show (Problem 1.5.3) that

$$[M^{\kappa\lambda}, P^\mu] = -i \left(\eta^{\kappa\mu} P^\lambda - \eta^{\lambda\mu} P^\kappa\right) \, . \tag{1.237}$$

The closed set of commutation relations (1.232), (1.237), and (1.230) furnishes the Lie algebra of the Poincaré group. One takes these commutation relations as the basis for investigations of the Poincaré group.

In a particular inertial frame, the anisymmetric tensor $M^{\mu\nu}$ can be expressed in terms of two three-dimensional vectors: $M^{\mu\nu} = (\mathbf{L}, \, \mathbf{K})$,

$$K_i = M^{i0}, \quad L_i = -\frac{1}{2} \, \epsilon_{ijk} \, M^{jk} \, , \tag{1.238}$$

ϵ_{ijk} is the three-dimensional Levi-Civita symbol introduced in Problem 1.3.5. To justify (1.238), we choose t^μ for the time axis, $t^\mu = (1, 0, 0, 0)$, and take the convention that $\epsilon_{ijk} = \epsilon^{0ijk}$. Then $t_\mu M^{\mu\nu} = (0, -\mathbf{K})$ and $t_\mu {}^*M^{\mu\nu} = (0, -\mathbf{L})$.

Equations (1.238) can be written in matrix form:

$$M^{\mu\nu} = \begin{pmatrix} 0 & -K_1 & -K_2 & -K_3 \\ K_1 & 0 & -L_3 & L_2 \\ K_2 & L_3 & 0 & -L_1 \\ K_3 & -L_2 & L_1 & 0 \end{pmatrix} \, . \tag{1.239}$$

Since the matrix M^{jk} affects spatial indices, it may be interpreted as the generator of rotations in the (j, k)-plane. The operator M^{0i} is the generator of boosts acting along the ith coordinate axis. Indeed, if we specialize κ, λ, μ, and ν to be, respectively, k, l, m, and n, then (1.230) becomes

$$[L_i, L_j] = i\epsilon_{ijk} L_k \, , \tag{1.240}$$

which are commutation relations of spatial rotations so(3). The remaining commutation relations (1.230) are equivalent to

$$[L_i, K_j] = i\epsilon_{ijk}\, K_k \ , \tag{1.241}$$

$$[K_i, K_j] = -i\epsilon_{ijk}\, L_k \ . \tag{1.242}$$

Fundamental laws of physics must be invariant under Poincaré transformations. This condition is met if we write these laws as vector or tensor equations. More generally, we express equations governing physical systems in terms of irreducible representations of the Poincaré group. One would say that the formulation of a physical theory is *covariant* if the equations of this theory have the same form in different Lorentz frames[5].

In order to determine irreducible representations of the Poincaré group, we should find independent invariants commuting with generators of this group. The eigenvalues of these invariants will label irreducible representations.

There are two such invariants. One of them is P^2 (Problem 1.5.4). To construct another invariant, we define the so-called *Pauli–Lubański vector*

$$w^\kappa = \frac{1}{2}\, \epsilon^{\kappa\lambda\mu\nu} M_{\lambda\mu} p_\nu \ . \tag{1.243}$$

One can show (Problem 1.5.5) that

$$w^2 = w^\nu w_\nu = M_{\mu\kappa} M^{\nu\kappa} p^\mu p_\nu - \frac{1}{2}\, M_{\kappa\lambda} M^{\kappa\lambda} p^2 \tag{1.244}$$

is the desired second invariant commuting with all generators P_μ and $M_{\mu\nu}$.

Problem 1.5.1. Verify that the matrices (1.229) obey commutation relations (1.230).

Problem 1.5.2. Verify that the operators $M_{\mu\nu}$ defined in (1.234) obey commutation relations (1.230).

Problem 1.5.3. Derive (1.237) from (1.232), (1.234), and (1.233).

Problem 1.5.4. Using (1.237), show that P^2 commutes with all generators $M_{\mu\nu}$.

Problem 1.5.5. Show that the quantity w^2 defined in (1.244) commutes with P_μ and $M_{\mu\nu}$.

Hint Derive first the relations

$$w \cdot p = 0 \ , \tag{1.245}$$

$$[P_\mu, w_\nu] = 0, \quad [M_{\mu\nu}, w_\lambda] = -i\left(\eta_{\mu\lambda} w_\nu - \eta_{\nu\lambda} w_\mu\right). \tag{1.246}$$

[5] This requirement on the form of physical relations should not be confused with the classification of tensor components according to whether their indices are covariant (lower) or contravariant (upper).

Problem 1.5.6. Supposing that P^μ is a timelike vector, $P^2 = M^2$, show that there exists a Lorentz frame in which the Pauli–Lubański vector w^μ becomes $w^\mu = M(0, \mathbf{S})$, where components S_i obey the commutation relations

$$[S_i, S_j] = i\,\epsilon_{ijk}\,S_k\ . \tag{1.247}$$

This suggests the interpretation of the Pauli–Lubański vector as spin for particles with nonzero mass M.

Answer A frame of reference with the time axis along the vector P^μ.

Problem 1.5.7. The noncompact Lie algebra so$(1,3)$ can be converted to the compact Lie algebra so(4) if the generators K_i are exchanged for $B_i = iK_i$. This corresponds to converting boosts from imaginary rotations, as in (1.218), to real rotations, as in (1.217). Show that the commutation relations of the resulting Lie algebra so(4) are given by

$$[L_i, L_j] = i\epsilon_{ijk}L_k, \quad [L_i, B_j] = i\epsilon_{ijk}\,B_k, \quad [B_i, B_j] = i\epsilon_{ijk}\,L_k\ . \tag{1.248}$$

Note the change of sign of the last commutator in comparison with (1.242).

Denoting $I_i = \frac{1}{2}\,(L_i + B_i)$ and $J_i = \frac{1}{2}\,(L_i - B_i)$, show that the generators I_i and J_i form two independent Lie algebras so(3), which implies that the Lie algebra so(4) is isomorphic to the Lie algebra so$(3) \oplus$ so(3).

Hint Derive the commutation relations

$$[I_i, I_j] = i\epsilon_{ijk}I_k, \quad [J_i, J_j] = i\epsilon_{ijk}\,J_k, \quad [I_i, J_j] = 0\ . \tag{1.249}$$

1.6 World Lines

The passage of a particle through spacetime is depicted by its *world line*, a curve that extends from the remote past to the far future. The concept of world lines merits notice.

Intuitively, a curve in $\mathbb{E}_n$ is a continuous sequence of points. However, according to present views, the notion best suited to describe a curve is a smooth *mapping*

$$\mathbf{z}: \quad \tau \to \mathbf{z}(\tau) \tag{1.250}$$

of an interval $[\tau_1, \tau_2]$ of the real axis $\mathbb{R}$ onto $\mathbb{E}_n$, rather than a *set* of points. We therefore will use the term ‘world line’ to designate a *parametrized* curve, which is defined as a smooth function $z^\mu(\tau)$ whose argument τ ranges from $-\infty$ to $+\infty$. In order to bridge the gap between this analytical definition and the intuitive idea of curves, one may think of a world line as an *equivalence class* of parametrized curves. Whatever parametrized curve among this equivalence class is chosen, one follows along the same path, though at potentially different values of the parameter τ. We will revert to this refined description of world lines in Sect. 2.6. For now, our treatment of parametrized curves seems quite sufficient.

We are entitled to choose parametrize the world line to suit our convenience. We first identify τ with laboratory time t in a particular Lorentz frame, and keep watch of how a particle moves with respect to this frame. This gives $z^\mu(t) = (t, \mathbf{z}(t))$, where $\mathbf{z}(t)$ is the particle's *trajectory* (recall that spatial parts of four-dimensional vectors are denoted by boldface characters). We now define the ordinary three-velocity and three-acceleration:

$$\mathbf{v} = \frac{d\mathbf{z}}{dt}, \quad \mathbf{a} = \frac{d\mathbf{v}}{dt}. \tag{1.251}$$

In relativistic problems, where preference is given to an explicitly covariant description, one employs an invariant parameter of evolution τ. It is convenient to parametrize curves with the so-called *proper* time s. Suppose that a particle is equipped with a clock of the same construction as the clock of a stationary Lorentz observer (for example, both clocks are governed by a single atomic standard). The proper time is defined as the time read from this moving clock. In the instantaneously comoving Lorentz frame where the particle is at rest at a given instant, the line element dz^2 equals the squared local time interval ds^2. On the other hand, in the stationary Lorentz frame, we have

$$dz^2 = dt^2 - d\mathbf{z}^2 , \tag{1.252}$$

which implies

$$ds = dt\,\sqrt{1 - \mathbf{v}^2}\,. \tag{1.253}$$

With the use of the Lorentz factor

$$\gamma = \frac{1}{\sqrt{1 - \mathbf{v}^2}}, \tag{1.254}$$

(1.253) becomes

$$dt = \gamma\, ds\,. \tag{1.255}$$

The proper time can be interpreted geometrically as the length of the world line,

$$s = \int_0^t dt\,\sqrt{1 - \mathbf{v}^2}\,. \tag{1.256}$$

The *time dilation* for moving clocks is readily apparent from (1.256). Historically, this effect was a major prediction of special relativity.

One can express s in terms of any other parameter of evolution τ by noting that

$$ds = \sqrt{dz^2} = \sqrt{\dot z^\mu\, \dot z_\mu}\, d\tau\,, \tag{1.257}$$

where $\dot z^\mu = dz^\mu/d\tau$.

We now define the *four-velocity* v^μ and the *four-acceleration* a^μ as

$$v^\mu = \frac{dz^\mu}{ds} \ , \tag{1.258}$$

$$a^\mu = \frac{dv^\mu}{ds} \ . \tag{1.259}$$

Taking into account that $dz^2 = ds^2$, we deduce from (1.258) the identity

$$v^2 = 1 \ . \tag{1.260}$$

Differentiating this identity with respect to s gives

$$v \cdot a = 0 \ . \tag{1.261}$$

The four-velocity v^μ is by definition a tangent vector. We see from (1.260), that v^μ is a unit timelike vector. Furthermore, (1.261) shows that the four-acceleration is always perpendicular to the four-velocity. Because any vector perpendicular to a timelike vector is spacelike, a^μ is spacelike,

$$a^2 < 0 \ . \tag{1.262}$$

Note that these arguments are valid only for particles moving at velocities lower than the velocity of light. The world lines of such particles are called *timelike*. A curve is said to be timelike if the tangent vector is always timelike.

Similarly, we define *spacelike* and *lightlike* (*null*) curves by the conditions that the tangent vector be, respectively, spacelike or null.

Using (1.253), one can show (Problem 1.6.1) that

$$v^\mu = (\gamma, \, \gamma \, \mathbf{v}) \ , \tag{1.263}$$

and

$$a^\mu = \left((\mathbf{a} \cdot \mathbf{v}) \, \gamma^4, \, \mathbf{a} \, \gamma^2 + \mathbf{v} \, (\mathbf{a} \cdot \mathbf{v}) \, \gamma^4 \right) \ . \tag{1.264}$$

In the instantaneously comoving inertial frame where the particle is at rest,

$$v^\mu = (1, \, \mathbf{0}) \ , \tag{1.265}$$

$$a^\mu = (0, \, \mathbf{a}) \ . \tag{1.266}$$

By (1.266),

$$a^2 = -\mathbf{a}^2 \ . \tag{1.267}$$

Therefore, the magnitude of three-acceleration as viewed by a comoving Lorentz observer is an invariant quantity.

Let a particle be moving along a straight line, say, the x-axis. Comparing (1.260) with

$$(\cosh \alpha)^2 - (\sinh \alpha)^2 = 1 \ , \tag{1.268}$$

we can write

$$v^\mu = (\cosh \alpha, \, \sinh \alpha, \, 0, \, 0) \ , \tag{1.269}$$

where α is an arbitrary function of s. Differentiation of (1.269) with respect to s gives

$$a^{\mu} = \dot{\alpha}\,(\sinh\alpha,\, \cosh\alpha,\, 0,\, 0)\ , \tag{1.270}$$

where the dot stands for a derivative with respect to s. Squaring (1.270) gives

$$a^2 = -\dot{\alpha}^2\ . \tag{1.271}$$

Let us take a closer look at the case that

$$\ddot{\alpha} = 0\ . \tag{1.272}$$

This equation can be immediately integrated to give

$$\alpha = \alpha_0 + ws\ , \tag{1.273}$$

where α_0 and w are constants. By (1.271),

$$a^2 = -w^2\ . \tag{1.274}$$

Comparison of (1.274) and (1.267) shows that (1.273) describes *uniform acceleration*. For reference, the motion is called uniformly accelerated if acceleration **a** viewed by all instantaneously comoving Lorentz observers is constant. If $\alpha_0 = 0$, then (1.269) becomes

$$v^{\mu}(s) = (\cosh ws,\, \sinh ws,\, 0,\, 0)\ , \tag{1.275}$$

which, when integrated, yields the world line

$$z^{\mu}(s) = z^{\mu}(0) + \frac{1}{w}\,(\sinh ws,\, \cosh ws,\, 0,\, 0)\ . \tag{1.276}$$

The dependence of t and x upon the proper time s is clearly revealed from this expression:

$$t - t_0 = \frac{1}{w}\sinh ws, \quad x - x_0 = \frac{1}{w}\cosh ws\ . \tag{1.277}$$

It follows that

$$x - x_0 = \frac{1}{w}\sqrt{1 + w^2(t - t_0)^2}\ . \tag{1.278}$$

The two-dimensional plot of this curve is a hyperbola approaching asymptotically to the light cone rays $t + x = 0$ as $t \to -\infty$ and $t - x = 0$ as $t \to \infty$. Hence the name *hyperbolic* motion can be used synonymously with the term ‘uniformly accelerated’ motion.

The three-velocity is

$$|\mathbf{v}| = \frac{w\,(t - t_0)}{\sqrt{1 + w^2(t - t_0)^2}}\ . \tag{1.279}$$

It follows that $|\mathbf{v}| \ll 1$ over a period of time that $|t - t_0| \ll w^{-1}$. Within this period, the particle moves slowly, and the well-known formulas for uniformly accelerated motion in Newtonian mechanics $x - x_0 = \frac{1}{2}\, w\, (t - t_0)^2$ and $v = w\,(t - t_0)$ are recovered from (1.278) and (1.279).

We now direct our attention to the question: what is the class of all allowable world lines? It is reasonable to assume that smooth timelike curves belong to this class. Most relativistic expressions involve the Lorentz factor (1.254) which becomes infinite at $\mathbf{v}^2 = 1$. This precludes acceleration of particles to velocities greater than that of light. No particle can be accelerated past the light barrier. Hence, world lines with continuously joined timelike and spacelike fragments must be eliminated from consideration.

Of even greater concern is the causality argument. If faster-than-light particles exist, then it is possible for an observer to send signals to his own past. A sequence of events of this kind is called a *causal cycle*. The admission of superluminal signals would result in a patently absurd process in which an observer could cause his own destruction which, in turn prevents the destructive signal from being sent. This seems to forbid spacelike world lines from classical relativistic theory.

However, there are particles which travel at luminal velocities, even though they were never accelerated from a slower speed. These particles begin moving with the speed of light immediately upon creation and remain in this state. World lines of these particles are depicted by null curves. The existence of such objects presents no problems with causality.

We thus see that the class of allowable world lines could possibly accommodate both timelike and null smooth world lines.

Problem 1.6.1. Derive (1.263) and (1.264).

Problem 1.6.2. Find components of $\dot a^\mu$ in a particular Lorentz frame.

Answer

$$\dot a^0 = \left[(\dot{\mathbf{a}}\cdot\mathbf{v}) + \mathbf{a}^2\right]\gamma^5 + 4\,(\mathbf{a}\cdot\mathbf{v})^2\,\gamma^7\,,$$
$$\dot a^i = \dot{\mathbf{a}}\,\gamma^3 + \left[3\mathbf{a}\,(\mathbf{a}\cdot\mathbf{v}) + \mathbf{v}\,(\dot{\mathbf{a}}\cdot\mathbf{v}) + \mathbf{v}\,\mathbf{a}^2\right]\gamma^5 + 4\mathbf{v}\,(\mathbf{a}\cdot\mathbf{v})^2\,\gamma^7\,, \quad (1.280)$$

Problem 1.6.3. Find $\dot a^\mu$ for motion along a straight line.

Answer

$$\dot a^\mu = \left(\ddot\alpha\sinh\alpha + \dot\alpha^2\cosh\alpha,\ \ddot\alpha\cosh\alpha + \dot\alpha^2\sinh\alpha,\ 0,\ 0\right)\,. \quad (1.281)$$

Problem 1.6.4. Show that

$$v\cdot\dot a = -a^2, \quad v\cdot\ddot a = -3\,(a\cdot\dot a)\,. \quad (1.282)$$

Problem 1.6.5. Denote $\dot z^\mu = dz^\mu/d\tau$, where τ is an arbitrary parameter of evolution, and introduce the generalized Lorentz factor

$$\gamma = \frac{1}{\sqrt{\dot{z} \cdot \dot{z}}} \ . \tag{1.283}$$

Taking τ to be laboratory time in a particular Lorentz frame t, we have $\dot{z}^0 = 1$, $\dot{\mathbf{z}} = \mathbf{v}$, and (1.283) becomes identical to (1.254).

Show that the four-acceleration is given by

$$a^\mu = \gamma \frac{d}{d\tau} (\gamma \, \dot{z}^\mu) \ . \tag{1.284}$$

Problem 1.6.6. Let O and P be two points separated by a timelike interval, and Γ a timelike curve connecting these points. Various curves Γ have different lengths s, defined in (1.256). Prove that s is *maximal* for a straight line. In other words, the longest proper time is measured by a clock that moves along a straight line between O and P.

Hint Approximate a smooth curve by polygonal lines with small rectilinear fragments. Consider a triangle ABC whose sides AC, AB, and BC are given by timelike intervals. Show that the sum of two sides AB and BC is smaller than the third side AC.

Notes

1. Lorentz (1904*b*); Poincaré (1898, 1904, 1905, 1906), Einstein , (1905*a* 1905*b*), and Minkowski (1909) laid the foundation of special relativity. The findings by Einstein and Minkowski had an especially strong hold on the minds of the scientific community in the early 20th century. The term 'special relativity' was introduced by Einstein. For historical details and further references see Whittaker (1953), and Pais (1982). There are many texts on special relativity (which is also treated at some length in most of the books on general relativity). The following have enjoyed popularity for years: Weyl (1918), Pauli (1958), and Synge (1956).

2. *Section 1.1.* Analyzing the concepts of space and time, and guided by experiment, Poincaré (1898, 1902, 1904) concluded that no motion with respect to the aether is detectable. He formulated the principle of relativity, which, coupled with the idea of a maximum signal velocity in nature, offered a clear view that the simultaneity of spatially separated events is a matter of convention. Einstein (1905*a*) discarded the concept of an aether, and proceeded from two postulates: (1) the laws of physics are the same in all inertial frames, and (2) the speed of light in vacuum is constant. He proposed a convention for simultaneity of separated points, now known as the standard synchrony, which provided a means of deriving the Lorentz transformation (1.19). This transformation was earlier derived by Larmor (1900) and Lorentz (1904*b*) on the hypothesis that any effect of the motion through the aether is unobservable. Poincaré (1905) showed that the set of all transformations (1.19),

combined with space rotations, constitutes a group, to which he gave the name the Lorentz group. The spacetime metric (1.26) was first introduced by Poincaré (1906). Minkowski (1909) unified space and time into an indivisible four-dimensional entity which he called 'the world'.

Neumann (1870) noted that Newton's first and second laws assume their simplest form if the standard time scale is used, otherwise these laws become more complicated. Grünbaum (1963) raised the question of whether a transported yardstick is self-congruent, at least in the absence of stresses and temperature variations.

Bondi (1967) proposed a pictorial treatment of special relativity, called the k-calculus, which is used here to make the introduction to the subject as simple as possible.

For those who wish to read more widely on the physical problems outlined in this section, the following references may be of assistance: Weyl (1918), Pauli (1958), Synge (1956), Misner et al. (1973). The foundations of special relativity and its epistemological implications are discussed by Jammer (1954), Reichenbach (1958), and Grünbaum (1963).

3. *Section 1.2.* This section is to provide a quick introduction to the Weyl axioms of affine and Euclidean spaces. Klein (1872) developed the view of geometry as the invariant theory of some definite group (nowadays called the automorphism group).

4. *Section 1.3.* The material of this section is quite standard. For an extended discussion see Misner et al. (1973), Spivak (1974), and Dubrovin, Fomenko & Novikov (1992). Useful exercises can be found in Lightman et al. (1975).

5. *Section 1.4.* The reader should be familiar with elements of vector analysis, including notions of curves and surfaces, vector differential operations (gradient, divergence, and curl, and results of their integration (the theorems by Gauss–Ostrogradskiĭ, and Stokes). A general reference is Morse & Feshbach (1953). The task of this section is to translate this standard knowledge into an invariant geometric language.

The theorem that a volume integral of a divergence can be written as a surface integral was proved by Ostrogradskiĭ. This proof, reported to the St Petersbourg Academy of Science by Ostrogradsky[6] in 1828, was published in 1831. Gauss (1813) introduced the notion of surface integrals, and derived some special cases of the theorem. Green (1828) established the relation

$$\int_V d^3x \left(v\nabla^2 u - u\nabla^2 v\right) = \oint_{\partial V} dS \left[v(\mathbf{n}\cdot\nabla)u - u(\mathbf{n}\cdot\nabla)v\right] \,, \tag{1.285}$$

which is referred to by his name. The extension of the Gauss–Ostrogradskiĭ theorem from $n = 3$ to any integer n was made by Ostrogradskiĭ in 1834.

[6] Ostrogradskiĭ is the same as Ostrogradsky but Ostrogradskiĭ is closer to the Russian transcription.

The significance of this theorem was understood later. Maxwell (1873) appreciated its utility in electrodynamics and emphasized Ostrogradskiĭ's priority. Stokes (1849) proved the theorem expressed by (1.196). For the history of the Stokes theorem see Spivak (1965). A modern formulation and derivation of this theorem (which would be properly termed the Newton–Leibnitz–Gauss–Green–Ostrogradskiĭ–Stokes–Poincaré theorem) can be found in many books, for example: De Rham (1955), Spivak (1965, 1974), Cartan (1967), Schwartz (1967), and Dubrovin, Fomenko & Novikov (1992).

6. *Section 1.5.* Wigner (1939) demonstrated the significance of irreducible representations of the Poincaré group. We briefly review the most important features of the Poincaré group in this section. A general reference is Bogoliubov, Logunov, Oksak & Todorov (1990), Naimark (1964), Gel'fand et al. (1963), and Barut & Rączka (1977).

7. *Section 1.6.* The relation of relativity and causality is discussed in many places. For an extensive discussion of this subject see Frank (1932), and Bunge (1959). A more sophisticated version of causality involving hypothetical superluminal signals is advocated by Bilaniuk & Sudarshan (1969).

The relativistic kinematics outlined here is discussed in all books on this theory. A detailed account can be found in Synge (1956).

2

Relativistic Mechanics

Newtonian mechanics rests on three axioms of motion:

I. *Every particle continues in its state of rest or uniform motion in a straight line unless it is acted upon by some exterior force.*
II. *The rate of change of momentum of a particle is proportional to the force impressed upon it, and is in the direction in which the force is acting.*
III. *To every action there is an equal and oppositely directed reaction.*

In relativistic mechanics, aimed at an adequate description of particles moving at speeds comparable with the speed of light, these statements are no longer valid when taken literally. Nevertheless, their role is every bit as important as in Newtonian mechanics. They cease to be regarded as strict dynamical laws of universal applicability, rather they are elevated to the status of guiding principles.

We will see in Sect. 2.1 that Newton's second law requires neither modification nor generalization. It should be only *embedded* in the four-dimensional geometry of Minkowski space. The key observation is that the dynamical law, as it was originally formulated by Newton, holds in any instantaneously comoving inertial frame. Such a treatment of Newton's second law begets equation (2.7) governing the behavior of relativistic particles.

It will transpire that Newton's first law is entirely valid only for a certain class of mechanical objects, *Galilean particles*. In the absence of external forces, Galilean particles move along straight world lines. However, such an evolution law need not be the case for *non-Galilean* particles. We will gain some insight into the behavior of these particles in Problem 2.1.3, and Sect. 2.7; other examples of non-Galilean dynamics will be adduced in later chapters.

Newton's third law seems to be violated in relativistic mechanics where the influence of one particle on another propagates at a finite speed, and the response arises with some retardation. An important exception is contact interactions, in which one particle acts upon another and experiences back reaction at the same point. That is the reason for the explicit conservation of linear momentum in collisions and decays (Sect. 2.10).

Meanwhile Newton's third law may be understood in a broader sense. As will be shown in Sect. 3.2, the extended action–reaction principle enables us to view the electric charge of a particle as both a measure of the influence of electromagnetic fields on the particle and of the particle's strength as a field source. The Maxwell–Lorentz theory inherits the action–reaction principle as an element of field dynamics.

The basic consequence of the action–reaction principle for mechanics is momentum conservation. We learn from Noether's first theorem (Sect. 2.5) that linear momentum is conserved due to invariance of the Lagrangian under spatial translations. Likewise, angular momentum conservation results from invariance of the Lagrangian under spatial rotations. These conservation laws are exact only for closed systems 'particles plus fields'. One may imagine the situation that these conservation laws are violated for mechanical degrees of freedom, when the field contribution is overlooked. A breakdown of angular momentum conservation in the system of two particles, one with an electric charge and the other with a magnetic charge, is an example discussed in Sect. 2.9.

2.1 Dynamical Law for Relativistic Particles

The best known form of Newton's second law is given by an ordinary differential equation

$$\frac{d\mathbf{p}}{dt} = \mathbf{f} \,, \tag{2.1}$$

where $\mathbf{p}$ is the Newtonian momentum. In Newtonian mechanics, $\mathbf{p}$ is proportional to the velocity $\mathbf{v}$ of the body under discussion,

$$\mathbf{p} = m\mathbf{v} \,. \tag{2.2}$$

The coefficient of proportionality m measures inertia of the body. We will call m the *Newtonian mass*. Newton envisioned the mass as the quantity of matter that the body contains. In today's parlance, this might be interpreted as the number of identical atoms in the body. However, we are concerned here with idealized particles of no size rather than real extended bodies. So for our purposes, each point particle is simply endowed with a positive number m. We regard this number m as a primary notion, which is associated with inertia of the particle, and may or may not reflect the Newtonian idea of the quantity of matter.

Before proceeding to a discussion of relativistic dynamics, let us clarify the present status of Newton's second law. It is still common to see the assertion that (2.1) is unsuited for the description of the relativistic particle behavior, and that a completely different equation (derived by Poincaré, and independently by Max Planck, in 1906)

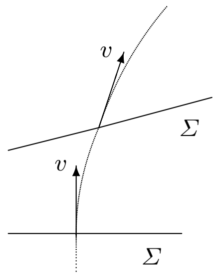


Fig. 2.1. The hyperplanes Σ perpendicular to the world line

$$\frac{d}{dt}\,\frac{m\mathbf{v}}{\sqrt{1-\mathbf{v}^2}}=\mathbf{F} \tag{2.3}$$

must be used instead. Does this mean that special relativity abolishes Newton's second law, or, in less categorical terms, generalizes (2.1) in such a way as to yield (2.3)? No. The Newtonian equation (2.1) need be neither rejected nor modified. To derive relativistic dynamics, we should *embed* the three-dimensional (2.1) into the four-dimensional geometry of Minkowski space.

The idea of the embedding is based on the fact that (2.1) becomes a *strictly accurate* law as $\mathbf{v}\to 0$. In other words, granting that an *instantaneously comoving* inertial frame is given at some instant, one can precisely predict the evolution of the particle in this frame during an ensuing evanescent time interval. In the geometric language, the vector relation (2.1) is strict on a hyperplane Σ perpendicular to the world line. Meanwhile the hyperplane Σ tilts together with its normal v^μ as one moves along the world line, Fig. 2.1. To recover the global evolution one joins together the small fragments from each of the instantaneously comoving frames. The algorithm for reconstructing the world line is the following: fix a rest frame at the initial instant $s=s_0$, evaluate a local fragment of the curve using (2.1), fix a rest frame at a nearby instant $s=s_0+\epsilon$, evaluate the next local fragment, etc. The output is reminiscent of a movie, which is actually a discrete set of pictures representing the dynamical affair at the hyperplanes Σ.

For the embedding to be made smooth, we need an operator $\overset{v}{\perp}$ that continually projects vectors of Minkowski space on hyperplanes Σ perpendicular to the world line. As is clear from the discussion of Sects. 1.2 and 1.3, the desired operator is

$$\overset{v}{\perp}_{\mu\nu}=\eta_{\mu\nu}-\frac{v_\mu v_\nu}{v^2}\,. \tag{2.4}$$

Note that the projector (2.4) is the same for any parametrization. If $v_\mu=dz_\mu/ds$ is replaced by $\dot z_\mu=dz_\mu/d\tau$, this leaves the form of $\overset{v}{\perp}$ unchanged.

Let us take a closer look at how the projector (2.4) embeds the Newtonian equations (2.1) in four dimensions. The time axis in the instantaneously comoving frame is parallel to the tangent of the world line, hence dt equals

ds. Formally, this follows from $ds = \gamma^{-1}dt$, where γ goes to 1 as $\mathbf{v} \to 0$. Accordingly, in the instantaneously comoving frame, the derivative with respect to t may be replaced by the derivative with respect to s.

Given the three-dimensional vector $\mathbf{f}$ in the hyperplane Σ, one can unambiguously construct a four-dimensional vector f^μ. Indeed, in the instantaneously comoving frame,

$$f^\mu = (0, \mathbf{f}) \, . \tag{2.5}$$

In an arbitrary inertial frame, components of f^μ can be found from (2.5) through the appropriate Lorentz boost. f^μ is called the *Minkowski force* or *four-force*.

In the rest frame, $v^\mu = (1, \mathbf{0})$, and, by (2.5), $v^\mu f_\mu = 0$. Since $v^\mu f_\mu$ is an invariant, the Minkowski force f^μ is perpendicular to the four-velocity v^μ in any Lorentz frame.

Let us define the *four-momentum* of the particle p^μ. To this end, we consider the derivative of the four-momentum with respect to the proper time dp^μ/ds, and require that its spatial components in the instantaneously comoving frame dp^i/ds be identical to the respective components of $d\mathbf{p}/dt$ in (2.1). Note, however, that dp^0/ds remains indeterminate in this frame.

The desired embedding of (2.1) in hyperplanes perpendicular to the world line is

$$\overset{v}{\perp}_{\mu\nu} \left(\frac{dp^\nu}{ds} - f^\nu \right) = 0 \, . \tag{2.6}$$

This is the basic *dynamical law* for relativistic structureless point objects. Expressed in symbolic form (2.6) is

$$\overset{v}{\perp} (\dot{p} - f) = 0 \, . \tag{2.7}$$

We thus may conclude that relativistic dynamics of a particle is essentially contained in Newtonian dynamics in its primordial formulation (2.1). Equation (2.7) is a mere geometric restatement of Newton's second law.

The presence of the projector $\overset{v}{\perp}$ in (2.7) suggests that we have three independent equations. Indeed, when contracted with v^μ, the four equations reveal a linear relation resulting from the identity

$$v^\mu \left(\eta_{\mu\nu} - \frac{v_\mu v_\nu}{v^2} \right) = 0 \, . \tag{2.8}$$

The dynamical law (2.7) is compatible with any relation between the four-momentum p^μ and kinematical variables of the particle. If a particular expression for p^μ has been used in (2.7), the resulting equation is called the *equation of motion* for the given particle.

The simplest mechanical entity is a *Galilean particle*. Such a particle is a point object possessing the four-momentum

$$p^\mu = m v^\mu \, . \tag{2.9}$$

It follows that $\dot{p}^\mu = ma^\mu$. Since $v \cdot a = 0$, the projector $\overset{v}{\perp}$ in (2.6) acts as a unit operator, and (2.6) becomes

$$ma^\mu = f^\mu \,. \tag{2.10}$$

It may be worth pointing out that (2.10) is the equation of motion governing only Galilean particles, while the general law for both Galilean and non-Galilean objects is (2.6).

In the limit $\mathbf{v} \to 0$, space part of the four-momentum (2.9) is identical to the Newtonian momentum (2.2). Hence m in (2.9) is the Newtonian mass.

If $f^\mu = 0$, equation (2.10) assumes

$$\dot{v}^\mu = 0 \,. \tag{2.11}$$

This equation has a unique solution $v^\mu = \text{const}$. Thus, a free Galilean particle moves along *straight* world lines.

It is usual to take the word 'particle' to mean a Galilean object. We follow this tradition, but we add the epithet 'Galilean' whenever the lack of emphasis may conceal some pertinent nuance.

We now fix a particular Lorentz frame. The four-momentum of a Galilean particle is

$$p^\mu = (\varepsilon, \mathbf{p}) \,. \tag{2.12}$$

Here, ε and $\mathbf{p}$ are interpreted as the total kinetic *energy* and linear *momentum* of a Galilean particle. Using the decomposition $v^\mu = \gamma\,(1, \mathbf{v})$, these quantities evaluate to

$$\varepsilon = m\gamma \,, \tag{2.13}$$

$$\mathbf{p} = m\gamma\mathbf{v} \,. \tag{2.14}$$

In the nonrelativistic region $|\mathbf{v}| \ll 1$, these expressions can be expanded in powers of $\mathbf{v}$:

$$\varepsilon = \frac{m}{\sqrt{1-\mathbf{v}^2}} = m + \frac{m\mathbf{v}^2}{2} + \cdots \,, \tag{2.15}$$

$$\mathbf{p} = \frac{m\mathbf{v}}{\sqrt{1-\mathbf{v}^2}} = m\mathbf{v} + \cdots \,. \tag{2.16}$$

The term $m\mathbf{v}$ on the right of (2.16) is the Newtonian momentum, while the two terms on the right of (2.15) differ from the conventional nonrelativistic kinetic energy by m. Recall, however, that, the energy in Newtonian mechanics is defined up to an arbitrary constant; the lowest energy level can be assigned arbitrarily. The first term on the right of (2.15) is called the *rest energy*. We see that special relativity requires that the expansion of ε should begin with m. Thus, the Newtonian mass m is not only a measure of inertia of a Galilean particle; in the absence of external forces, m is equal to the total energy of this particle in its rest frame.

Since the Minkowski force is orthogonal to the four-velocity, the components of f^μ are not independent but are subject to the constraint $f^0\gamma = f^i\gamma v^i$. It is convenient to separate the Lorentz factor γ as an overall factor in f^μ:

$$f^\mu = \gamma\,(\mathbf{F}\cdot\mathbf{v},\,\mathbf{F})\,. \tag{2.17}$$

With $ds = \gamma^{-1}dt$, we find that the spatial component of (2.10),

$$\frac{d}{dt}\,(m\gamma\mathbf{v}) = \mathbf{F}\,, \tag{2.18}$$

is identical to (2.3). The force $\mathbf{F}$ in (2.18) should not be confused with the Newtonian force $\mathbf{f}$. Although $\mathbf{F} = \mathbf{f}$ in the instantaneously comoving frame, such is not the case in a fixed Lorentz frame.

The time component of (2.10),

$$\frac{d}{dt}\,(m\gamma) = \mathbf{F}\cdot\mathbf{v}\,, \tag{2.19}$$

might be interpreted as the variation of the particle energy $\varepsilon = m\gamma$ due to the work performed by the force $\mathbf{F}$ in a unit time. A direct calculation (Problem 2.1.2) shows that (2.19) follows immediately from (2.18).

Two invariants can be built from vectors p^μ and v^μ,

$$M^2 = p^2 \tag{2.20}$$

and

$$m = v\cdot p\,, \tag{2.21}$$

while the third invariant $v^2 = 1$ is dynamically trivial. M and m are called the *mass* and the *rest mass*, respectively. We use the same notation for both the rest mass and Newtonian mass, because they are numerically equal for any Galilean particle. Moreover, the identity $v^2 = 1$ implies that M is coincident with m. Thus, formulas (2.20) and (2.21) represent two different definitions of the same quantity attributable to a Galilean particle.

In the subsequent discussion, we will consider also *non-Galilean* particles, for which p^μ and v^μ need not be collinear, and hence $M \neq m$. One such example is in Problem 2.1.3.

An interesting object is a particle which moves at the speed of light. We assume that the world lines of such particles are smooth lightlike curves directed to the future. The procedure of embedding the nonrelativistic equation (2.1) in Minkowski space is no longer valid, because a particle moving at the speed of light cannot be brought to rest in any frame. However, as will be seen in Sect. 2.6, a consistent dynamics of such particles is still possible.

Particles moving faster than light are called *tachyons*. It should be remarked that objects with superluminal speeds are not precluded by special relativity but they cannot be produced by accelerating subluminal objects. Assuming that a tachyon is endowed with the four-momentum $p^\alpha = m\,dz^\alpha/ds$, where m is a real coefficient and s is the Euclidean length of the tachyon world line, one obtains $p^2 = -m^2 < 0$, because $(dz/ds)^2 < 0$.

Although spacelike world lines are excluded from consideration, particles moving along timelike world lines and yet possessing spacelike four-momenta, are in principle possible. As will become evident, these particles are non-Galilean. Throughout this book we employ the term 'tachyon' to denote particles with spacelike four-momenta, $p^2 < 0$, but moving at subluminal velocities.

Problem 2.1.1. The mass of a Galilean particle is constant,

$$\dot{m} = 0 \, . \tag{2.22}$$

Show that (2.22) is consistent with the fact that f^μ is given by (2.5) in the rest frame.

Hint Write the Minkowski force in the rest frame as $f^\mu = (h, \mathbf{f})$ where h is some function of kinematical variables of the particle. Then, in addition to (2.7), we come to $\dot{m} = h$, which is in conflict with preservation of the particle self-identity (2.22).

Problem 2.1.2. Show that (2.19) follows immediately from (2.18).

Problem 2.1.3. *Discrete time and non-Galilean particles.* One may think of continuum physics as a very convenient but only approximate description of physical reality which evolves in a series of very tiny steps, and try to modify Newton's second law through the change of the differential equation $m\dot{\mathbf{v}} = \mathbf{f}$ with the difference equation

$$m \, \frac{\mathbf{v}(t+\ell) - \mathbf{v}(t)}{\ell} = \mathbf{f} \, , \tag{2.23}$$

where ℓ is a 'quantum of time'. This seemingly innocuous modification has a serious mathematical impact, because we are actually dealing with the infinite order differential equation

$$\frac{m}{\ell} \left[\exp\left(\ell \, \frac{d}{dt} \right) - 1 \right] \mathbf{v}(t) = \mathbf{f} \, , \tag{2.24}$$

in place of the first order differential equation. In (2.24) the exponential of d/dt acts as

$$\exp\left(\ell \, \frac{d}{dt} \right) \mathbf{v}(t) = \sum_{n=0}^{\infty} \frac{\ell^n}{n!} \left(\frac{d}{dt} \right)^n \mathbf{v}(t) = \mathbf{v}(t+\ell) \, . \tag{2.25}$$

Allowance for the granularity of time is thus included in the conventional calculus.

It is no great surprise that the particle governed by this modified law may show itself as a non-Galilean object. Indeed, let the particle be free, $\mathbf{f} = \mathbf{0}$.

Verify that the general solution to (2.23) is

$$\mathbf{v}(t) = \sum_{n=0}^{\infty} \mathbf{V}_n \cos\left(\frac{2\pi n}{\ell} t + \alpha_n\right), \tag{2.26}$$

where $\mathbf{V}_n$ and α_n are integration constants. This means that the free particle executes a periodic motion, the so-called *zitterbewegung*, back and forth along a straight line parallel to $\mathbf{V}_n$ for $n \geq 1$, and moves uniformly for $n = 0$.

Show that a smooth embedding of the modified equation (2.23) in Minkowski space culminates in the basic law of relativistic dynamics (2.7)

$$\overset{v}{\perp}(\dot{p} - f) = 0 \tag{2.27}$$

with

$$p^\mu(s) = m\ell^{-1}\left[z^\mu(s+\ell) - z^\mu(s)\right] . \tag{2.28}$$

Find the general solution to this equation when $f^\mu = 0$. Relying on this solution, check on whether or not the equality between M and m defined in (2.20) and (2.21) is the case.

Problem 2.1.4. Derive a covariant condition for uniform acceleration.

Answer A generalization of (1.272) reads

$$\overset{v}{\perp}\dot{a} = 0 , \tag{2.29}$$

or, in coordinate notation,

$$\dot{a}^\mu + a^2 v^\mu = 0 . \tag{2.30}$$

2.2 The Minkowski Force

In this section, we review some explicit forms of the Minkowski force. Let us first note that the four-force cannot be constant. Indeed, let $f^\mu = \text{const}$. This supposition comes into conflict with the condition that the four-force is orthogonal to the four-velocity,

$$f \cdot v = 0 . \tag{2.31}$$

More precisely, given a spacelike constant vector f^μ, one can always find a timelike vector v^μ such that (2.31) fails.

A similar consideration shows that f^μ cannot be a function of only z^μ, the particle coordinates. It is therefore imperative that f^μ be dependent on

v^μ. So, our concern in the subsequent discussion is only with $f^\mu(z, v)$, and occasionally with $f^\mu(s, z, v)$.

We now consider the case that f^μ is *linear* in v^μ. The relationship $f^\mu = \alpha v^\mu$ is impossible, because this f^μ is not orthogonal to v^μ for any $\alpha \neq 0$. Let us then put

$$f^\mu = \beta^{\mu\nu} v_\nu \ , \tag{2.32}$$

where $\beta^{\mu\nu}$ is an arbitrary tensor. By (2.31),

$$\beta^{\mu\nu} v_\mu v_\nu = 0 \ . \tag{2.33}$$

This equation is obeyed by an arbitrary v^μ provided that

$$\beta^{\mu\nu} = -\beta^{\nu\mu} \ . \tag{2.34}$$

We thus have

$$f^\mu = \beta^{\mu\nu} v_\nu \ , \tag{2.35}$$

with $\beta^{\mu\nu}$ being an antisymmetric tensor.

Is there a physical object which affects a particle through a four-force linear in the four-velocity? This object, if it proves to occur, would be distributed over all space and characterized by some $\beta^{\mu\nu}$ at each point. Such objects are collectively known as *force fields*, or simply *fields*. Note, however, that $\beta^{\mu\nu}$ contains information on both the state of the field and how it affects the particle. We should separate these concepts.

It seems natural in the first instance to take a *scalar* coupling e, implying that

$$f^\mu = e\, v_\nu F^{\mu\nu} \ . \tag{2.36}$$

This possibility is actually realized in nature. The quantity whose state at each point of Minkowski space is specified by an antisymmetric tensor $F_{\mu\nu}$ and whose influence on particles is represented by the force law (2.36) is called the *electromagnetic field*. The tensor $F_{\mu\nu}$ is referred to as the electromagnetic *field tensor*, or *field strength*.

Recall that a particle is assumed to be a point object whose nature is preserved under time evolution. In particular, the coupling of the particle and the electromagnetic field must be unchanged,

$$\dot{e} = 0 \ . \tag{2.37}$$

We will refer (provisionally) to the scalar real quantity e as the *electric charge-coupling*. If a particle may be assigned a finite e, this particle is said to be *charged*. By contrast, if the electromagnetic field does not act on the particle, we take $e = 0$, and include formally this case to our consideration. The particle is then called *neutral*. As is well known from the experiment, there is a minimal value of e, the *elementary* charge, and every charged particle carries a charge which is a multiple of this elementary charge. However, classical field theory leaves this charge quantization unexplained. So, the charge-coupling e to be considered in this book is regarded as a parameter taking any positive, zero, and negative values.

Let us fix a particular inertial frame. Two three-dimensional vectors in this frame, **E** and **B**, can be defined in terms of six components of the anti-symmetric tensor $F_{\mu\nu}$,

$$E_i = F_{0i} = F^{i0} \,, \tag{2.38}$$

$$B_k = -\frac{1}{2}\,\epsilon_{klm} F^{lm} \,, \tag{2.39}$$

where ϵ_{klm} is the three-dimensional Levi-Civita symbol. The inverse of (2.39) is

$$F^{ij} = F_{ij} = -\epsilon_{ijk} B_k \,. \tag{2.40}$$

In (2.38) and (2.40), the usual rule of raising indices of tensors in Minkowski space is applied: when a spatial index is raised or lowered, the expression is multiplied by the factor -1, whereas raising a time index do not change the sign. Notice that it would be meaningless to raise indices of the three-dimensional vectors **E** and **B**, because E_i and B_i do not present spatial components of four-dimensional vectors; **E** and **B** behave as vectors only under spatial rotations in the given Lorentz frame. Throughout this book, E_i and B_i, and ϵ_{ijk} will appear with lower indices. Although we operate only for lower case characters in expressions with E_i, B_i, ϵ_{ijk}, and δ_{ij}, the summation over repeated latin indices is understood, as is customary when tensors are Euclidean.

Equations (2.38)–(2.40) can be written in matrix form. We adopt the convention that the left index refers to the row, and the right index to the column. We then have

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & -B_3 & B_2 \\ -E_2 & B_3 & 0 & -B_1 \\ -E_3 & -B_2 & B_1 & 0 \end{pmatrix}, \quad F^{\mu\nu} = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & -B_3 & B_2 \\ E_2 & B_3 & 0 & -B_1 \\ E_3 & -B_2 & B_1 & 0 \end{pmatrix}. \tag{2.41}$$

In the language of differential forms (for notations and conventions see Appendix A), the general decomposition $F = \frac{1}{2}\, F_{\mu\nu}\, dx^\mu \wedge dx^\nu$ can be specified by analogy with (2.41):

$$F = E_i\, dx^0 \wedge dx^i - \epsilon_{ijk} B_i\, dx^j \wedge dx^k \,. \tag{2.42}$$

Consider the three-dimensional force **F** appearing in the force law (2.36). Combining the definitions of **E** and **B**, equations (2.38)–(2.40), with the relations $v_\mu = \gamma\,(1\,, -\,\mathbf{v})$ and $f_\mu = \gamma\,(\mathbf{F}\cdot\mathbf{v}\,, -\mathbf{F})$, we obtain

$$\mathbf{F} = e\,(\mathbf{E} + \mathbf{v}\times\mathbf{B}) \,. \tag{2.43}$$

This expression for the force exerted on a charged particle is due to Oliver Heaviside and Lorentz. Following historical tradition, we will call **E** and **B** the electric field intensity and the magnetic induction, respectively. Both the three-dimensional vector $\mathbf{F}$ defined in (2.43) and the four-dimensional vector f^μ defined in (2.36) will go under the general name of *Lorentz force*; it is usually quite clear from the context which of them is involved.

Consider the behavior of $\mathbf{E}$ and $\mathbf{B}$ under space reflections $\mathbf{r} \to -\mathbf{r}$. Experiment shows that the force $\mathbf{F}$ transforms under this discrete operation as

$$\mathbf{F} \to -\mathbf{F}\,. \tag{2.44}$$

The velocity $\mathbf{v}$ is also a vector,

$$\mathbf{v} \to -\mathbf{v}\,, \tag{2.45}$$

and e, by its very definition, is a scalar. It follows from (2.43)–(2.45) that $\mathbf{E}$ and $\mathbf{B}$ are polar and axial vectors, respectively,

$$\mathbf{E} \to -\mathbf{E}, \quad \mathbf{B} \to \mathbf{B}\,. \tag{2.46}$$

Likewise, if the assumption is made that the basic dynamical law (2.6) is invariant under time reversal $t \to -t$, we immediately find from (2.43) that this operation causes $\mathbf{E}$ and $\mathbf{B}$ to transform as

$$\mathbf{E} \to \mathbf{E}, \quad \mathbf{B} \to -\mathbf{B}\,. \tag{2.47}$$

We now turn to the case that the interaction of a particle and electromagnetic field is specified by a *pseudoscalar* coupling $e^\star$. We define the field ${}^*F^{\mu\nu}$ dual to $F^{\mu\nu}$ as

$${}^*F^{\mu\nu} = \frac{1}{2}\,\epsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}\,, \tag{2.48}$$

and adopt the convention that the three- and four-dimensional Levi-Civita symbols are related by

$$\epsilon_{ijk} = \epsilon^{0ijk}\,. \tag{2.49}$$

Consider the four-force

$$f^\mu = e^\star\, v_\nu\, {}^*F^{\mu\nu}\,. \tag{2.50}$$

Using (2.38)–(2.40), (2.48) and (2.49), we derive from (2.50)

$$\mathbf{F} = e^\star\,(\mathbf{B} - \mathbf{v} \times \mathbf{E})\,. \tag{2.51}$$

Taking into account (2.44), (2.45), and (2.46), we conclude that the sign of $e^\star$ changes under space reflections, and $e^\star$ is therefore a pseudoscalar.

The pseudoscalar $e^\star$ is (provisionally) called the *magnetic charge-coupling*. Particles that are affected by force $\mathbf{F}$ of the form (2.51) are referred to as *magnetic monopoles*. Doubly charged particles carrying both electric and magnetic charges e and $e^\star$, which experience the force

$$\mathbf{F} = e\,(\mathbf{E} + \mathbf{v} \times \mathbf{B}) + e^\star\,(\mathbf{B} - \mathbf{v} \times \mathbf{E})\,, \tag{2.52}$$

are called *dyons*.

We next look at a particle, possessing additional degrees of freedom, which is coupled with a field through a *vector* coupling. Let V be a vector space of dimension n, the so-called *internal* space. Borrowing nomenclature from quantum chromodynamics, we refer to V as the *color* space. Coordinates of Minkowski space are in general unrelated to coordinates of color vectors. The charge-coupling Q^a is now a color vector (rather than a scalar, or pseudoscalar), a is the color index running from 1 to n. If we define a dual space V' of linear forms on V, and assume that the field strength $G_a^{\mu\nu}$ takes values on V', then the four-force f^μ linear in v^μ is

$$f^\mu = \sum_{a=1}^{n} Q^a v_\nu G_a^{\mu\nu} \,. \tag{2.53}$$

However, this construction is too arbitrary, and additional constraints on V are called for. Let us denote elements of V by A, B, $C, \ldots$, and think of them as $n \times n$ matrices. Linear operators Ω give rise to transformations of a matrix A:

$$A' = \Omega A \Omega^{-1}, \quad \Omega\Omega^{-1} = \Omega^{-1}\Omega = 1 \,. \tag{2.54}$$

We may interpret (2.54) as the transformations of matrix elements of A induced by changes of basis of the vector space V. Given some basis, the matrix A in it is referred to as the *adjoint representation* of A.

Of primary concern to our discussion is V equipped with *Lie algebra structure*. Simply stated, A and B may not only be added, but also multiplied. We denote the product of A and B by $[A, B]$. Here, the Lie bracket $[,]$ is anticommutative and bilinear,

$$[A, B] = -[B, A] \,, \tag{2.55}$$

$$[\alpha A + \beta B, C] = \alpha\,[A, C] + \beta\,[B, C] \,, \tag{2.56}$$

α and β being any complex numbers. However, $[,]$ is not associative. $[A, [B, C]]$ is not equal to $[[A, B], C]$, and the Jacobi identity

$$[A, [B, C]] + [C, [A, B]] + [B, [C, A]] = 0 \tag{2.57}$$

is required instead.

To render V a Lie algebra $\mathfrak{a}$, a skew-symmetric bilinear map $V \times V \to V$ must be defined, namely, a rule of multiplication of any two basis elements T_a and T_b of V,

$$[T_a, T_b] = i \sum_c f^c{}_{ab}\, T_c \,, \tag{2.58}$$

must be given. Note that (2.58) is invariant under the transformation (2.54). The coefficients $f^c{}_{ab} = -f^c{}_{ba}$ are called the *structure constants* of the Lie algebra $\mathfrak{a}$. They are independent of the representation of the elements T_a and may therefore be regarded as a property of the Lie algebra $\mathfrak{a}$.

The Jacobi identity (2.57) can be expressed in terms of the structure constants,

$$f^d{}_{bc}f^e{}_{ad} + f^d{}_{ab}f^e{}_{cd} + f^d{}_{ca}f^e{}_{bd} = 0 \,. \tag{2.59}$$

The structure constants $f^c{}_{ab}$ determine the Lie algebra $\mathfrak{a}$ completely. From $f^c{}_{ab}$, one can construct a symmetric tensor $g_{ab} = -f^c{}_{ad}\, f^d{}_{bc}$ which plays the role of the metric tensor in the color space; specifically it raises and lowers color indices. If the tensor g_{ab} is not degenerate, $\det(g_{ab}) \neq 0$, the Lie algebra $\mathfrak{a}$ is semisimple[1]. We will henceforth deal with semisimple complex Lie algebras. It is always possible to choose a basis for a semisimple complex Lie algebra, called the *Cartan basis*, such that the structure constants f_{abc} are *real* and *completely antisymmetric.*

The field strength $G^{\mu\nu}_a$ which takes values on a Lie algebra is called the *Yang–Mills field*, to honor Chen Ning Yang and Robert Mills who introduced and developed this concept in 1954 by the example of the SU(2) Lie algebra.

When all the structure constants are zero, the Lie algebra is *Abelian.* As an example we refer to a Lie algebra with the basis $\{T_a\}$ containing a single element. Now, the color charge Q^aT_a is in fact a real number e, and the field strength $G^{\mu\nu}_aT^a$ is a real-valued tensor $F^{\mu\nu}$. In that case the color four-force (2.53) is reduced to the Lorentz force (2.36).

For a color particle to remain identical to itself, the coupling of this particle with the Yang–Mills field must not vary in time. This implies that the color charge of the particle Q^a is preserved in some sense. The condition $\dot{Q}^a = 0$, similar to (2.37), appears undue severe, and we assume instead that the color charge magnitude is a constant of motion,

$$Q^aQ_a = \text{const} \,. \tag{2.60}$$

Thus, the color vector Q^a shares with a top the property of precessing around some axis (this axis is yet unfixed in the color space, and may vary in the course of the particle evolution). We will see in Sect. 7.1 that (2.60) follows from the equation for the color charge evolution.

We finally address the four-force f^μ *quadratic* in the four-velocity v^μ, and select two examples of this force.

Consider a scalar field $\phi(x)$. We wish to extend the notion of the potential force $\mathbf{f} = -\nabla\Phi$, commonly used in Newtonian mechanics, to the relativistic context. Let f^μ be orthogonal to the four-velocity v^μ and contain gradients of the scalar field,

$$f_\mu = -g \overset{v}{\perp}_{\mu\nu} \partial^\nu\phi = -g\left(\partial_\mu\phi - v_\mu v^\nu\partial_\nu\phi\right) , \tag{2.61}$$

where g is the coupling constant of the particle and the field ϕ, and ∂_μ stands for $\partial/\partial z^\mu$. We see that the extension of the Newtonian potential force is f^μ quadratic in v^μ. The second term in the parenthesis of (2.61) is

$$v_\mu v^\nu\partial_\nu\phi = v_\mu \frac{dz^\nu}{ds}\,\frac{\partial\phi}{\partial z^\nu} = v_\mu \frac{d\phi}{ds} \,. \tag{2.62}$$

[1] For an overview of basic facts of Lie algebras see Appendix B.

One further term ϕa_μ might be added to (2.61) without altering the orthogonality condition (2.31). This term is exceptional in that it introduces the dependence of the force on higher derivatives of coordinates, but its presence is essential, which will be clear in Sect. 2.6. Then the equation of motion for a Galilean particle (2.10) interacting with a scalar field ϕ reads

$$\frac{d}{ds}\,(m - g\,\phi)\,v^\mu = -g\,\partial^\mu \phi\,. \tag{2.63}$$

The fact that the four-force is quadratic in the four-velocity is here implicit.

Let us turn to the general case of the quadratic dependence

$$f^\lambda = m\Gamma^\lambda_{\mu\nu} v^\mu v^\nu\,. \tag{2.64}$$

Here, $\Gamma^\lambda_{\mu\nu}$ is some tensor, which is evidently symmetric in the indices μ and ν. The orthogonality requirement (2.31) becomes

$$\Gamma_{\lambda\mu\nu} v^\lambda v^\mu v^\nu = 0\,. \tag{2.65}$$

For this equation to hold for any v^μ, it is necessary that the sign of $\Gamma_{\lambda\mu\nu}$ must change under permutation of λ and μ, and that of λ and ν. However, any tensor $\Gamma_{\lambda\mu\nu}$ which is symmetric in μ and ν, and antisymmetric in λ and μ as well as in λ and ν, is zero.

Nevertheless, expression (2.64) may be nontrivial if a *pseudo-Riemannian* metric $g_{\mu\nu}(x)$ of a curved spacetime manifold (rather than the pseudoeuclidean metric $\eta_{\mu\nu}$) is concerned. Now the line element is

$$ds^2 = g_{\mu\nu} dz^\mu dz^\nu\,. \tag{2.66}$$

Let this line element be not null. We then define a unit tangent vector $v^\mu = dz^\mu/ds$,

$$g_{\mu\nu} v^\mu v^\nu = 1\,, \tag{2.67}$$

and so

$$0 = \frac{d}{ds}\,g_{\mu\nu} v^\mu v^\nu = 2g_{\mu\nu} v^\mu a^\nu + v^\mu v^\nu\,\frac{d}{ds}\,g_{\mu\nu} = 2g_{\mu\nu} v^\mu a^\nu + v^\mu v^\nu v_\lambda \partial^\lambda g_{\mu\nu}\,. \tag{2.68}$$

Consider $\Gamma^\lambda_{\mu\nu}$, expressed in terms of $g_{\mu\nu}$ as

$$\Gamma^\lambda_{\mu\nu} = \frac{1}{2}\,g^{\lambda\sigma}\left(\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}\right)\,, \tag{2.69}$$

or, equivalently,

$$\Gamma_{\lambda\mu\nu} = \frac{1}{2}\left(\partial_\mu g_{\nu\lambda} + \partial_\nu g_{\lambda\mu} - \partial_\lambda g_{\mu\nu}\right)\,. \tag{2.70}$$

It is easy to check from (2.68) that the equation of motion for a Galilean particle

$$\frac{dv_\lambda}{ds} + \Gamma_{\lambda\mu\nu} v^\mu v^\nu = 0 \tag{2.71}$$

is orthogonal to v^λ in the sense of the pseudo-Riemannian metric (2.66). Equation (2.71) is the *geodesic* equation for the given metric. It governs the behavior of a test particle in the *gravitational field* $g_{\mu\nu}$, according to the general theory of relativity.

Problem 2.2.1. Derive the transformation law for the electromagnetic field strengths $\mathbf{E}$ and $\mathbf{B}$ in the case that a new frame moves relative to the initial one along x^1-axis with the speed V, that is, the transformation of $\mathbf{E}$ and $\mathbf{B}$ which is due to the Lorentz boost $x^0 = \gamma(x'^0 - Vx'^1)$, $x^1 = \gamma(x'^1 - Vx'^0)$, where $\gamma = (1 - V^2)^{-\frac{1}{2}}$.

Answer

$$\begin{aligned} &E_1 = E'_1, \quad E_2 = \gamma\left(E'_2 - VB'_3\right), \quad E_3 = \gamma\left(E'_3 + VB'_2\right)\,, \\ &B_1 = B'_1, \quad B_2 = \gamma\left(B'_2 + VE'_3\right), \quad B_3 = \gamma\left(B'_3 - VE'_2\right)\,. \end{aligned} \tag{2.72}$$

2.3 Invariants of the Electromagnetic Field

The state of the electromagnetic field at each point of Minkowski space is specified by six numbers $F^{\mu\nu}$. However, the tensor $F^{\mu\nu}$ carries information not only on the field by itself, but also on the frame which is used for the determination of its components. To describe the field in an intrinsic way, we need have knowledge of only two *invariants of the electromagnetic field*

$$\mathcal{S} = \frac{1}{2}\, F_{\mu\nu}F^{\mu\nu} \quad \text{and} \quad \mathcal{P} = \frac{1}{2}\, F_{\mu\nu}{}^*F^{\mu\nu}\,, \tag{2.73}$$

where ${}^*F^{\mu\nu} = \frac{1}{2}\,\epsilon^{\mu\nu\alpha\beta}F_{\alpha\beta}$. Indeed, let $F = \frac{1}{2}\,F_{\mu\nu}\,dx^\mu \wedge dx^\nu$ be the 2-form corresponding to the antisymmetric tensor $F_{\mu\nu}$. It was shown in Sect. 1.3 that, in the general case, there exists a canonical basis of vectors e_0^μ, e_1^μ, e_2^μ, e_3^μ, such that[2]

$$F = E\, e_0 \wedge e_1 - B\, e_2 \wedge e_3\,. \tag{2.74}$$

It immediately follows that the essential information on the geometric peculiarities of F is contained in two parameters E and B, or two independent functions of E and B, say, $\mathcal{S}$ and $\mathcal{P}$ (see Problem 2.3.3). The existence of only two invariants, in terms of which all scalar quantities constructed from $F^{\mu\nu}$ can be expressed, derives from the fact that spacetime is four-dimensional.

Another way of looking at the intrinsic description of the 2-form F is to to use an arbitrary basis of 1-forms dx^μ. We apply the Hodge operation* to $F = \frac{1}{2}\,F_{\mu\nu}\,dx^\mu \wedge dx^\nu$ to obtain[3]

[2] There may be exceptional cases, exemplified by a null field shown in (2.86), for which the canonical decomposition (2.74) does not hold.

[3] Note that the operator* acts on the exterior product $dx^\mu \wedge dx^\nu$ rather than on the coefficient $F_{\mu\nu}$.

$$ {}^*F = \frac{1}{2} F_{\mu\nu} \, {}^*(dx^\mu \wedge dx^\nu) = \frac{1}{4} F_{\mu\nu} \, \epsilon^{\mu\nu}{}_{\alpha\beta} \, dx^\alpha \wedge dx^\beta = \frac{1}{2} \, {}^*F_{\alpha\beta} \, dx^\alpha \wedge dx^\beta \, . \quad (2.75) $$

From F and ${}^*F$, two 4-forms can be built

$$ F \wedge F = \frac{1}{4} F_{\mu\nu} \, F_{\alpha\beta} \, dx^\mu \wedge dx^\nu \wedge dx^\alpha \wedge dx^\beta \, , \quad (2.76) $$

$$ F \wedge {}^*F = \frac{1}{4} F_{\mu\nu} \, {}^*F_{\alpha\beta} \, dx^\mu \wedge dx^\nu \wedge dx^\alpha \wedge dx^\beta \, , \quad (2.77) $$

which in turn result in two 0-forms (quantities which are independent of the basis),

$$ {}^*(F \wedge F) = \frac{1}{4} F_{\mu\nu} \, F_{\alpha\beta} \, {}^*(dx^\mu \wedge dx^\nu \wedge dx^\alpha \wedge dx^\beta) = \frac{1}{4} F_{\mu\nu} \, F_{\alpha\beta} \, \epsilon^{\mu\nu\alpha\beta} = \mathcal{P} \, , \quad (2.78) $$

$$ {}^*(F \wedge {}^*F) = \frac{1}{4} F_{\mu\nu} \, {}^*F_{\alpha\beta} \, \epsilon^{\mu\nu\alpha\beta} = \frac{1}{8} F_{\mu\nu} \, F^{\gamma\delta} \, \epsilon_{\alpha\beta\gamma\delta} \, \epsilon^{\mu\nu\alpha\beta} = -\frac{1}{2} F_{\mu\nu} \, F^{\mu\nu} = -\mathcal{S} \, . \quad (2.79) $$

It is easy to verify that ${}^*({}^*F \wedge {}^*F) = -{}^*(F \wedge F)$. Therefore, the 0-forms ${}^*(F \wedge F)$ and ${}^*(F \wedge {}^*F)$ (or, equivalently, $\mathcal{S}$ and $\mathcal{P}$) provide the desired intrinsic description of F.

We now express $\mathcal{S}$ and $\mathcal{P}$ in terms of $\mathbf{E}$ and $\mathbf{B}$. With the definitions of $\mathbf{E}$ and $\mathbf{B}$, (2.38)–(2.40), and the relation between the three- and four-dimensional Levi-Civita symbols (2.49), we find

$$ F_{\mu\nu} F^{\mu\nu} = 2(\mathbf{B}^2 - \mathbf{E}^2), \quad F_{\mu\nu} \, {}^*F^{\mu\nu} = -4\mathbf{E} \cdot \mathbf{B} \, , \quad (2.80) $$

and so

$$ \mathcal{S} = \mathbf{B}^2 - \mathbf{E}^2, \quad \mathcal{P} = -2 \, \mathbf{E} \cdot \mathbf{B} \, . \quad (2.81) $$

Because $\mathbf{E}$ and $\mathbf{B}$ behave as polar and axial vectors under space-reflections, $\mathcal{S}$ is a genuine scalar, and $\mathcal{P}$ is a pseudoscalar.

If $\mathbf{E}$ and $\mathbf{B}$ are orthogonal in some Lorentz frame, they are orthogonal in any frame, because $\mathbf{E} \cdot \mathbf{B} = 0$ is an invariant condition. If $\mathbf{E}$ and $\mathbf{B}$ have equal magnitudes in one frame of reference, $|\mathbf{E}| = |\mathbf{B}|$, they have equal magnitudes in all frames of reference, as the invariant condition $\mathbf{B}^2 - \mathbf{E}^2 = 0$ suggests. The inequality $\mathcal{S} < 0$ is also Lorentz-invariant, hence the condition $|\mathbf{E}| > |\mathbf{B}|$ in some frame of reference remains unchanged in all frames of reference. Lorentz transformations can take $\mathbf{E}$ and $\mathbf{B}$ to vectors of any magnitude and direction subject to the condition that $\mathcal{S}$ and $\mathcal{P}$ are fixed.

Thus, all states of the electromagnetic field can be divided into three classes:

(A) $\qquad \mathcal{P} = 0, \quad \mathcal{S} \neq 0.$

If $\mathcal{S} < 0$, or, equivalently, $|\mathbf{E}| > |\mathbf{B}|$, the field is said to be of *electric type.* There is a frame of reference such that $\mathbf{B} = \mathbf{0}$ and $|\mathbf{E}| = \sqrt{-\mathcal{S}}$, that is, the electromagnetic field is found to be in a pure electric state. If $\mathcal{S} > 0$, or, equivalently, $|\mathbf{B}| > |\mathbf{E}|$, we have a field of *magnetic type*. Using Lorentz transformations one can find a frame of reference such that $\mathbf{E} = 0$, and hence the field strength is purely magnetic.

$(B) \qquad \mathcal{P} = 0, \quad \mathcal{S} = 0.$

$\mathbf{E}$ and $\mathbf{B}$ are equal in magnitude and orthogonal in all frames of reference. This state is called the *null-field* state. With Lorentz transformations one can make the field amplitude $|\mathbf{E}| = |\mathbf{B}|$ to take any finite value, except that it is not possible to transform an initial nonzero field to zero (Problem 2.3.4).

$(C) \qquad \mathcal{P} \neq 0.$

By Lorentz transformations any finite $\mathbf{E}$ and $\mathbf{B}$ can be obtained consistent with given values of $\mathcal{S}$ and $\mathcal{P}$. In particular, there exists a frame of reference such that $\mathbf{E}$ and $\mathbf{B}$ are parallel.

Given $\mathcal{P} = 0$, the electromagnetic field strength can be represented as

$$F^{\mu\nu} = f^\mu g^\nu - f^\nu g^\mu \, , \tag{2.82}$$

or

$$F = f \wedge g \, , \tag{2.83}$$

with some four-vectors f^μ and g^μ. Therefore, in this case, the 2-form F is decomposable. The decomposability of F is sufficient to render $\mathcal{P}$ vanishing, as is apparent from the identity $\epsilon_{\alpha\beta\gamma\delta} f^\alpha g^\beta f^\gamma g^\delta = 0$. Furthermore, the decomposability of F is necessary for $\mathcal{P}$ to be zero. This can be proved by contradiction, assuming that E and B, appearing in the canonical decomposition (2.74), are both nonzero. Indeed, $\mathcal{P} = 0$ is equivalent to ${}^*(F \wedge F) = -2EB = 0$, which runs counter to our assumption.

Thus, classes A and B are characterized by a decomposable 2-form F. Another way of saying this is that a tetrad e_0^μ, e_1^μ, e_2^μ, e_3^μ can be found, such that the 2-form

$$F = E \, e_0 \wedge e_1 \tag{2.84}$$

corresponds to a field of electric type, the 2-form

$$F = B \, e_2 \wedge e_3 \tag{2.85}$$

corresponds to a field of magnetic type, and the 2-form

$$F = E \, (e_0 + e_2) \wedge e_1 \, . \tag{2.86}$$

corresponds to a null field.

Expressions (2.81) suggest that the combination $\mathbf{E} + i\mathbf{B}$ may be adaptable to an intrinsic description of the electromagnetic field. Consider a three-dimensional vector space $\mathbb{C}_3$ composed of complex vectors

$$\mathbf{Z} = \mathbf{E} + i\mathbf{B} \, . \tag{2.87}$$

We define a complex quadratic form on $\mathbb{C}_3$

$$\mathbf{Z}^2 = \sum_{i=1}^{3} Z_i Z_i = \mathbf{E}^2 - \mathbf{B}^2 + 2i\,\mathbf{E}\cdot\mathbf{B} \, . \tag{2.88}$$

Notice that (2.88) can not be interpreted as a metric because it is not real-valued (unlike the Hilbert scalar product $\mathbf{Z}\cdot\bar{\mathbf{Z}}$). However, this quadratic form is advantageous in constructing quantities invariant under complex orthogonal transformations of the space $\mathbb{C}_3$. These transformations leave the magnitude of the complex vector $\mathbf{Z}^2$ (together with its real and imaginary parts, $\mathrm{Re}\,\mathbf{Z}^2 = -\mathcal{S}$ and $\mathrm{Im}\,\mathbf{Z}^2 = -\mathcal{P}$) unchanged and hence are associated with linear transformations of Minkowski space which leave $\mathcal{S}$ and $\mathcal{P}$ unaltered, the Lorentz group transformations.

Let us turn to the case $\mathcal{P} \neq 0$. We take the vector $\mathbf{Z}$ to be $\mathbf{Z} = Z\mathbf{n}$, with $\mathbf{n}$ being a unit complex vector, $\mathbf{n}^2 = 1$, and Z its magnitude. Rotating $\mathbf{n}$, one can convert it to another unit vector, specifically, to that aligned with a real axis, say, $\mathbf{e}_1$. We write $Z = E_1 + iB_1$. Upon rotation of the initial vector $\mathbf{Z}$, we have $\mathbf{Z}' = \mathbf{E}' + i\mathbf{B}' = (E_1 + iB_1)\,\mathbf{e}_1$. This means that $\mathbf{E}'$ and $\mathbf{B}'$ are parallel (aligned with the x^1-axis), $\mathbf{E}' = E_1\mathbf{e}_1$, and $\mathbf{B}' = B_1\mathbf{e}_1$. Therefore, the canonical 2-form

$$F = E\,e_0 \wedge e_1 - B\,e_2 \wedge e_3 \, , \tag{2.89}$$

where vectors e_0^μ, e_1^μ, e_2^μ, e_3^μ are orthonormalized, corresponds to the field states of the class C.

Problem 2.3.1. The second rank tensor $I_{\mu\nu}$ constructed from n tensors $F_{\mu\nu}$, with all indices being sequentially contracted, except for the forward index of the first tensor and the end index of the last tensor,

$$I_{\mu\nu} = F_\mu^{\ \alpha_1} F_{\alpha_1}^{\ \alpha_2} \ldots F_{\alpha_{n-1}}^{\ \alpha_n} F_{\alpha_n\nu} \, , \tag{2.90}$$

is called the nth power monomial of the tensor $F_{\mu\nu}$. Contraction of the remaining indices gives the invariant

$$I = F_\mu^{\ \alpha_1} F_{\alpha_1}^{\ \alpha_2} \ldots F_{\alpha_{n-1}}^{\ \alpha_n} F_{\alpha_n}^{\ \mu} \, . \tag{2.91}$$

Show that any odd-power monomial of an antisymmetric tensor $F_{\mu\nu}$ is an antisymmetric tensor $I_{\mu\nu} = -I_{\nu\mu}$. The invariant I resulting from such a monomial is zero. By contrast, any even-power monomial of an antisymmetric tensor $F_{\mu\nu}$ is a symmetric tensor. Thus, nonzero invariants I correspond to even-power monomials.

Problem 2.3.2. Let $A_{\mu\nu}$ and $B_{\mu\nu}$ be antisymmetric tensors. Prove the identities

$$^*A_{\alpha\beta}\, {}^*B^{\alpha\beta} = -A_{\alpha\beta} B^{\alpha\beta} \;, \tag{2.92}$$

$$\frac{1}{2}\,\eta_{\mu\nu}\, A_{\alpha\beta} B^{\alpha\beta} = {}^*A_{\mu\alpha}\, {}^*B^{\alpha}_{\ \nu} - A_{\mu\alpha}\, B^{\alpha}_{\ \nu} \;, \tag{2.93}$$

$$^*(A_{\mu\alpha}\, B^{\alpha}_{\ \nu} - B_{\mu\alpha}\, A^{\alpha}_{\ \nu}) = {}^*A_{\mu\alpha}\, B^{\alpha}_{\ \nu} - B_{\mu\alpha}\, {}^*A^{\alpha}_{\ \nu} = A_{\mu\alpha}\, {}^*B^{\alpha}_{\ \nu} - {}^*B_{\mu\alpha}\, A^{\alpha}_{\ \nu} \;, \tag{2.94}$$

$$A^{\mu\alpha}\, B_{\alpha\beta}\, A^{\beta\nu} = -\frac{1}{2}\, A^{\mu\nu}\, (A_{\alpha\beta}\, B^{\alpha\beta}) - \frac{1}{4}\, {}^*B^{\mu\nu}\, (A_{\alpha\beta}\, {}^*A^{\alpha\beta}) \;, \tag{2.95}$$

specifically,

$$F_{\mu\alpha}\, {}^*F^{\alpha\nu} = -\frac{1}{2}\, \mathcal{P}\, \delta^{\nu}_{\mu} \;, \tag{2.96}$$

$$^*F_{\mu\alpha}\, {}^*F^{\alpha\nu} - F_{\mu\alpha}\, F^{\alpha\nu} = \mathcal{S}\, \delta^{\nu}_{\mu} \;. \tag{2.97}$$

Problem 2.3.3. Show that any invariant constructed from antisymmetric tensors $F_{\mu\nu}$ and $^*F_{\mu\nu}$ is expressible in terms of $\mathcal{S}$ and $\mathcal{P}$.

Hint Consider the 2-form F in the canonical basis (2.74). In this basis,

$$^*F = B\, e_0 \wedge e_1 + E\, e_2 \wedge e_3 \;. \tag{2.98}$$

It follows

$$^*(F \wedge F) = -2EB = \mathcal{P}, \quad {}^*(F \wedge {}^*F) = E^2 - B^2 = -\mathcal{S} \;, \tag{2.99}$$

and so

$$E^2 = \frac{1}{2}\left(\sqrt{\mathcal{S}^2 + \mathcal{P}^2} - \mathcal{S}\right), \quad B^2 = \frac{1}{2}\left(\sqrt{\mathcal{S}^2 + \mathcal{P}^2} + \mathcal{S}\right) \;. \tag{2.100}$$

Any invariant built from $F_{\mu\nu}$ and $^*F_{\mu\nu}$ can be expressed in terms of the coefficients E and B of the canonical basis, hence in terms of $\mathcal{S}$ and $\mathcal{P}$.

Problem 2.3.4. Let $\mathcal{P} = 0$ and $\mathcal{S} = 0$. Show that the amplitude of the null field $|\mathbf{E}| = |\mathbf{B}|$ might be rendered arbitrary by appropriate Lorentz transformation, with the only reservation that there is no frame of reference where $\mathbf{E} = \mathbf{B} = \mathbf{0}$.

2.4 Motion of a Charged Particle in Constant and Uniform Electromagnetic Fields

The behavior of a charged particle in a given electromagnetic field is governed by the Lorentz force equation

$$m\, \frac{dv^{\mu}}{ds} = e\, v_{\nu} F^{\mu\nu} \;. \tag{2.101}$$

Let the field be constant and homogeneous. Our concern here is only with fields of the class A. We outline a regular method of solution to the Cauchy problem for the equation of motion (2.101) with constant $F^{\mu\nu}$. The same technique can be applied to the case of constant homogeneous fields of the classes B and C (Problems 2.4.1 and 2.4.3), and even to some varying fields (Problem 2.4.2). We assume for a while that the effect of the particle on the field is neglible.

We begin with a field F of electric type defined in (2.84), or, in tensor notation,

$$F^{\mu\nu} = E\,(e_0^\mu e_1^\nu - e_0^\nu e_1^\mu) \tag{2.102}$$

where e_0^μ and e_1^μ are respectively timelike and spacelike basis vectors, $e_0^2 = 1$, $e_1^2 = -1$, $e_0 \cdot e_1 = 0$. We choose e_0^μ for the time axis; then the field is pure electric, and $\mathbf{E}$ is parallel to the x^1-axis.

At the initial instant, we put

$$v^\mu(0) = e_0^\mu\,\Gamma + e_2^\mu\,\Gamma V \tag{2.103}$$

where V is the initial velocity parallel to the x^2-axis, and Γ is the corresponding Lorentz factor

$$\Gamma = \left(1 - V^2\right)^{-\frac{1}{2}} . \tag{2.104}$$

Note that (2.103) is nothing but a covariant form of the decomposition $v^\mu = (\gamma, \gamma\mathbf{v})$.

Knowledge of the driving force and the initial velocity suggest that the particle will move in the plane of the vectors e_1^μ and e_2^μ, and thus the four-velocity at an arbitrary instant s takes the form

$$v^\mu(s) = \theta\,e_0^\mu - \eta\,e_1^\mu - \zeta\,e_2^\mu . \tag{2.105}$$

The coefficients θ, η, and ζ can be expressed as projections of the four-velocity on the basis vectors,

$$\theta = v \cdot e_0, \quad \eta = v \cdot e_1, \quad \zeta = v \cdot e_2 . \tag{2.106}$$

They are unknown functions of the proper time, $\theta(s)$, $\eta(s)$, and $\zeta(s)$. Comparing (2.105) and (2.103), we find

$$\theta(0) = \Gamma, \quad \eta(0) = 0, \quad \zeta(0) = -\Gamma V . \tag{2.107}$$

By (2.102) and (2.105), equation (2.101) becomes

$$\dot{v}^\mu = \Omega\,(\eta\,e_0^\mu - \theta\,e_1^\mu) \tag{2.108}$$

where

$$\Omega = \frac{eE}{m} . \tag{2.109}$$

We take the scalar products of (2.108) with e_0^μ, e_1^μ, and e_2^μ, to give

$$\dot{\theta} = \Omega\,\eta\,, \tag{2.110}$$

$$\dot{\eta} = \Omega\,\theta\,, \tag{2.111}$$

$$\dot{\zeta} = 0\,. \tag{2.112}$$

The general solution to (2.112) is $\zeta = \text{const}$. With the initial condition for ζ, (2.107), we obtain

$$\zeta = -\Gamma V\,. \tag{2.113}$$

Introduce new variables

$$q = \theta + \eta, \quad p = \theta - \eta\,. \tag{2.114}$$

Adding (2.110) and (2.111), we get

$$\dot{q} = \Omega\,q\,. \tag{2.115}$$

The solution to this equation, taking into account the initial data for θ and η, (2.107), is

$$q(s) = \Gamma \exp(\Omega s)\,. \tag{2.116}$$

To find $p(s)$, we note that the four-velocity at each instant is a unit vector, $v^2 = 1$, and so

$$v^2 = \theta^2 - \eta^2 - \zeta^2 = (\theta - \eta)(\theta + \eta) - \zeta^2 = pq - \Gamma^2 V^2 = p\Gamma e^{\Omega s} - \Gamma^2 + 1 = 1\,. \tag{2.117}$$

Here, we used (2.114), (2.113), (2.116), and the identity $-\Gamma^2 V^2 = 1 - \Gamma^2$. It follows that

$$p(s) = \Gamma \exp(-\Omega s)\,. \tag{2.118}$$

From (2.114), in view of (2.116) and (2.118), we have

$$\theta = \Gamma \cosh(\Omega s)\,, \tag{2.119}$$

$$\eta = \Gamma \sinh(\Omega s)\,. \tag{2.120}$$

Substituting (2.113), (2.119) and (2.120) in (2.105), we find the four-velocity

$$v^\mu(s) = \Gamma \left\{ \left[e_0^\mu \cosh(\Omega s) - e_1^\mu \sinh(\Omega s) \right] + e_2^\mu\, V \right\}\,. \tag{2.121}$$

When the initial velocity V vanishes, the solution (2.121) describes hyperbolic motion. We therefore conclude that if a charged particle enters an infinite flat capacitor with velocity parallel to lines of force, then it moves through the capacitor with a uniform acceleration. By contrast, when the initial velocity points in some transverse direction, we deal with the sum of uniform motion in this direction and hyperbolic motion in the direction of **E**.

Differentiating (2.121) and squaring the result, we find that the acceleration in the rest frame is constant, $a^2 = -\Gamma^2 \Omega^2$.

Integration of (2.121) gives the world line

$$z^\mu(s) = z^\mu(0) + \Gamma \left\{ \Omega^{-1} \left[e_0^\mu \sinh(\Omega s) - e_1^\mu \cosh(\Omega s) \right] + e_2^\mu \, V s \right\} . \qquad (2.122)$$

The dependence of the coordinates z^1 and z^2 upon the proper time could be read from this expression:

$$z^1 = -\Omega^{-1} \Gamma \cosh(\Omega s), \quad z^2 = \Gamma V s \, , \qquad (2.123)$$

where, for the sake of simplicity, the initial values of $z^1(0)$ and $z^2(0)$ are put to zero. Expressing s via z^2 and substituting it into the preceding relation, we obtain the trajectory

$$z^1 = -\Omega^{-1} \Gamma \cosh \left(\frac{\Omega}{\Gamma V} \, z^2 \right) . \qquad (2.124)$$

A distinctive feature of the Cauchy problem in relativistic mechanics, which is absent from Newtonian mechanics, is that the dynamics is always constrained. Turning to the motion under discussion, if the vector v^μ is drawn from the origin, then the endpoint of v^μ will always lie on the forward hyperboloid $v^2 = 1$, $v_0 > 0$, tracing out a trajectory across this surface. In some cases, the constraint alleviates the problem, as in the above example; every so often, it plagues the analysis; and sometimes, one manages to choose appropriate variables allowing for the constraint automatically.

We now consider the motion of a particle in a constant homogeneous field of magnetic type. This field is specified by the 2-form F defined in (2.85), or the tensor

$$F^{\mu\nu} = B \left(e_2^\mu e_3^\nu - e_2^\nu e_3^\mu \right) \qquad (2.125)$$

where e_2^μ and e_2^μ are orthonormal spacelike basis vectors, $e_2^2 = e_3^2 = -1$, $e_2 \cdot e_3 = 0$. In a frame of reference where e_0^μ is taken as the time axis, and e_2^μ and e_2^μ are space vectors, the field is pure magnetic, with the magnetic induction $\mathbf{B}$ being parallel to the x^1-axis.

Let the initial four-velocity be

$$v^\mu(0) = e_0^\mu \, \Gamma + \Gamma \left(e_1^\mu \, V_\| + e_2^\mu \, V_\perp \right) \qquad (2.126)$$

where $V_\|$ and $V_\perp$ are the initial velocity projections on the x^1- and x^2-axes, and

$$\Gamma = \left(1 - V_\|^2 - V_\perp^2 \right)^{-\frac{1}{2}} . \qquad (2.127)$$

The four-velocity at an arbitrary instant s is

$$v^\mu(s) = \theta \, e_0^\mu - \eta \, e_1^\mu - \zeta \, e_2^\mu - \xi \, e_3^\mu \qquad (2.128)$$

where

$$\theta = v \cdot e_0, \quad \eta = v \cdot e_1, \quad \zeta = v \cdot e_2, \quad \xi = v \cdot e_3 \, . \qquad (2.129)$$

Comparing (2.128) and (2.126), we find the initial data

$$\theta(0) = \Gamma, \quad \eta(0) = -\Gamma V_{\|}, \quad \zeta(0) = -\Gamma V_{\perp}, \quad \xi(0) = 0\,. \tag{2.130}$$

By (2.125) and (2.128), equation (2.101) becomes

$$\dot{v}^{\mu} = \omega\,(\xi\, e_2^{\mu} - \zeta\, e_3^{\mu}) \tag{2.131}$$

where

$$\omega = \frac{eB}{m}\,. \tag{2.132}$$

We take the scalar product of (2.131) with e_0^{μ}, e_1^{μ}, e_2^{μ}, and e_3^{μ}, to yield

$$\dot{\theta} = 0\,, \tag{2.133}$$

$$\dot{\eta} = 0\,, \tag{2.134}$$

$$\dot{\zeta} = -\omega\,\xi\,, \tag{2.135}$$

$$\dot{\xi} = \omega\,\zeta\,. \tag{2.136}$$

The solutions to (2.133) and (2.134), taking into account the initial data for θ and η, (2.130), are

$$\theta = \Gamma, \quad \eta = -\Gamma V_{\|}\,. \tag{2.137}$$

Introduce a new variable

$$x = \zeta + i\xi\,. \tag{2.138}$$

Combining (2.135) and (2.136), we arrive at

$$\dot{x} = i\omega\, x\,. \tag{2.139}$$

The solution to this equation, allowing for the initial data for ζ and ξ, (2.130), is

$$x(s) = -(\Gamma V_{\perp})\exp(i\omega s)\,. \tag{2.140}$$

Separating real and imaginary parts of (2.140), we obtain

$$\zeta = -(\Gamma V_{\perp})\cos(\omega s)\,, \tag{2.141}$$

$$\xi = -(\Gamma V_{\perp})\sin(\omega s)\,. \tag{2.142}$$

By (2.137), (2.141), and (2.142), the four-velocity (2.128) becomes

$$v^{\mu}(s) = \Gamma\left\{e_0^{\mu} + e_1^{\mu}\,V_{\|} + V_{\perp}\left[e_2^{\mu}\cos(\omega s) + e_3^{\mu}\sin(\omega s)\right]\right\}\,. \tag{2.143}$$

As pointed out above, (2.143) is a covariant expression of the decomposition $v^{\mu} = (\gamma, \gamma\mathbf{v})$. In particular, the time component γ equals the coefficient of e_0^{μ}. We see from (2.143) that γ does not vary with time; γ takes its initial value Γ. This means that energy of a charged particle moving in a constant homogeneous magnetic field is constant, $\varepsilon = m\Gamma$.

Differentiating (2.143) and squaring the result, we find that the acceleration in the rest frame is constant, $a^2 = -(\Gamma V_\perp \omega)^2$.

Integration of (2.143) gives

$$z^\mu(s) = z^\mu(0) + \left(e_0^\mu + e_1^\mu V_\|\right) \Gamma s + R\left[e_2^\mu \sin(\omega s) - e_3^\mu \cos(\omega s)\right] , \qquad (2.144)$$

where

$$R = \frac{\Gamma V_\perp}{\omega} . \qquad (2.145)$$

The world line (2.144) is an infinite timelike helix, starting in the remote past and extending to the far future. The axis of the helix is inclined to the basis vector e_0^μ at the angle α found from $\tan\alpha = V_\|$. All turns of the helix are identical. It follows that the particle moves uniformly at the velocity $V_\|$ along the magnetic field $\mathbf{B}$, and orbits in a circle of radius R with cyclic frequency ω over the plane with normal $\mathbf{B}$.

Problem 2.4.1. Determine the world line of a particle in a constant homogeneous null field defined in (2.86). At the initial instant, the particle is at rest at the origin. Find the particle energy ε. Find the acceleration squared in the rest frame.

Answer

$$z^\mu(s) = e_0^\mu \left(\frac{w^2 s^3}{6} + s\right) - e_1^\mu \frac{w\, s^2}{2} + e_2^\mu \frac{w^2 s^3}{6}, \qquad w = \frac{eE}{m};$$
$$\varepsilon = m\,(v \cdot e_0) = m\left(1 + \frac{w^2 s^2}{2}\right); \quad a^2 = -w^2 . \qquad (2.146)$$

Problem 2.4.2. Determine the world line of a charged particle in a plane-wave field. The electromagnetic field is $F = k \wedge e_1 \cos\sigma$ where k^μ is a fixed null vector, and $\sigma = k \cdot x$ (for more detail see Sect. 4.3). For simplicity, we put $k^\mu = \Omega\, e_+^\mu = \Omega(e_0^\mu + e_2^\mu)$ where $\Omega = eE/m$. At the initial instant, the particle rests at the origin. Find the particle energy ε. Find the acceleration squared in the rest frame.

Answer

$$z^\mu(s) = e_0^\mu s + \frac{1}{4}\, e_+^\mu \left[s - \frac{\sin(2\Omega s)}{2\Omega}\right] + e_1^\mu \frac{\cos(\Omega s) - 1}{\Omega};$$
$$\varepsilon = m\,(v \cdot e_0) = m\left[1 + \frac{1}{2}\sin^2(\Omega s)\right]; \quad a^2 = -\Omega^2 \cos^2(\Omega s) . \qquad (2.147)$$

Hint One can observe that $k \cdot f = 0$. It follows that $k \cdot v = \text{const}$, which, in view of the initial data, gives $k \cdot v = \Omega$, and so $\sigma = k \cdot z = \Omega s$.

Problem 2.4.3. Determine the motion of a particle in constant homogeneous parallel electric and magnetic fields as defined in (2.89). At the initial instant, the particle moves at the speed V along the x^2-axis. Find the particle energy ε. Find the acceleration squared in the rest frame.

Answer

$$z^\mu(s) = z^\mu(0) + \Gamma\Omega^{-1}[e_0^\mu \sinh(\Omega s) - e_1^\mu \cosh(\Omega s)] + R\,[e_2^\mu \sin(\omega s) + e_3^\mu \cos(\omega s)],$$

$$\varepsilon = m\,(v \cdot e_0) = m\,\Gamma \cosh(\Omega s), \quad a^2 = -\Gamma^2(\Omega^2 + V^2\omega^2)\,, \tag{2.148}$$

where Ω, ω, and R are given, respectively, by (2.109), (2.132), and (2.145).

2.5 The Principle of Least Action. Symmetries and Conservation Laws

As already noted in Sect. 2.2, particles interact with each other via intermediate agents, fields. Mathematically, fields are continuous distributions over space. Local variations of the field state propagate at a rate lower than or equal to the speed of light. Therefore, fields can be thought of as systems with infinite degrees of freedom. We will embark on a study of field dynamics in the next chapter. For now, we restrict our discussion to the Lagrangian formalism of particles, considering fields as external effects with some prescribed properties. Clearly, such a description is incomplete; we will revert to the Lagrangian description of closed systems involving both particles and fields on an equal footing in Chap. 5.

Given a mechanical system with n degrees of freedom, configurations of this system are determined by n *generalized coordinates* q_a where a runs from 1 to n. In the simplest case, we use Cartesian coordinates. However, q_a need not be Cartesian, one may utilize any curvilinear coordinates which are well suited for the constraints involved. For example, the Kepler problem is conveniently expressed by polar coordinates. One of merits of the Lagrangian formalism is that it offers a unified description for any choice of generalized coordinates. A state of a given system is specified by its position $q_a = (q_1, \dots, q_n)$, a point in the *configuration space* $\mathcal{Q}$, and its velocity $\dot{q}_a = dq_a/dt = (\dot{q}_1, \dots, \dot{q}_n)$, an element of a vector space of dimension n tangent to the manifold $\mathcal{Q}$ at the point q_a. We assume that all dynamical properties of the system are encoded into a single function $L(t, q, \dot{q})$ called the *Lagrangian*[4]. We will suppress the subscript a and denote coordinates and velocities by q and $\dot{q}$ whenever this concise notation leads to no confusion.

[4] If the system is affected by an external force, L may depend on t explicitly, not only through the functions $q(t)$ and $\dot{q}(t)$. In addition, we will see (Problem 2.5.7) that the use of t-dependent Lagrangians brings the problem of motion of damped systems within reach of methods of the calculus of variation.

One may add time to the configuration space to yield the *event space* $T\mathcal{Q}$ whose points are specified by $n+1$ coordinates (t, q). The system traces out a *world line* in the event space during the course of its evolution. Let (t_1, q_1) and (t_2, q_2) be two arbitrary points of $T\mathcal{Q}$. They can be connected by a multitude of paths. Let these paths be parametrized by time, $q = q(t)$. We define the functional called the *action*

$$S = \int_{t_1}^{t_2} dt\, L(t, q, \dot{q}) \tag{2.149}$$

on a set of smooth curves connecting the points (t_1, q_1) and (t_2, q_2).

We now proceed to discuss two central issues of the calculus of variation known as the *principle of least action* and *Noether's first theorem.* Before going into their precise formulation, we derive a general expression for small variations of the action.

The net result of our analysis should not depend on the path parametrization. In relativistic problems it will be convenient to promote time to the status of a dynamical variable and denote it by q_0. Then t in (2.149) is to be understood as an arbitrary invariant parameter of evolution, q a point in the event space with coordinates $(q_0, q_1, \ldots, q_n)$, and $\dot{q}$ a vector tangent to the world line with components $(\dot{q}_0, \dot{q}_1, \ldots, \dot{q}_n)$.

Consider a change from the given path $q(t)$ to some contiguous path

$$q'(t') = q(t) + \Delta q\ . \tag{2.150}$$

The *total coordinate variation*

$$\Delta q = q'(t') - q(t) \tag{2.151}$$

is assumed to be small but otherwise arbitrary.

In general, a passage from one world line to another is possible not only at a fixed instant t, but also at the instant t' separated from t by a short time interval,

$$t' = t + \Delta t\ . \tag{2.152}$$

We may separate the 'transverse part' of the total coordinate variation corresponding to the change from one curve to another at the same t,

$$\delta q = q'(t) - q(t)\ . \tag{2.153}$$

This quantity is usually called the *local coordinate variation.* We put

$$\delta q(t) = \epsilon\, \chi(t) \tag{2.154}$$

where ϵ is a small parameter, and $\chi(t)$ an arbitrary smooth function. Geometrically, the local variation δq results from the total variation Δq by deleting its 'longitudinal part' which is a displacement along the initial world line over the time interval Δt,

$$\delta q = \Delta q - \dot{q} \Delta t \, . \tag{2.155}$$

Indeed, rewriting (2.151) as

$$\Delta q = [q'(t') - q(t')] + [q(t') - q(t)] \, , \tag{2.156}$$

and taking into account (2.153), we find that the expressions in the square brackets are respectively δq and $\dot{q}\Delta t$ up to terms of the second order in ϵ.

It is clear from the definition (2.153) that the operation δ commutes with differentiation with respect to t,

$$\delta \frac{d}{dt} = \frac{d}{dt} \delta \, . \tag{2.157}$$

By analogy with (2.155) we can write the local variation of the velocity

$$\delta \dot{q} = \Delta \dot{q} - \ddot{q} \, \Delta t \, . \tag{2.158}$$

We impose the condition that the functional form of the Lagrangian is unchanged,

$$L'(t', q', \dot{q'}) = L(t', q', \dot{q'}) \, . \tag{2.159}$$

Owing to this condition, the total Lagrangian variation is

$$\Delta L = L'(t', q', \dot{q'}) - L(t, q, \dot{q}) = \frac{\partial L}{\partial t} \, \Delta t + \frac{\partial L}{\partial q_a} \, \Delta q_a + \frac{\partial L}{\partial \dot{q}_a} \, \Delta \dot{q}_a \, . \tag{2.160}$$

We sum over indices a (from 1 to n or $n+1$, depending on whether the configuration space $\mathcal{Q}$ or event space $T\mathcal{Q}$ is concerned) when they appear twice, without writing the summation symbol. We separate the local variation in (2.160) using (2.155) and (2.158),

$$\Delta L = \left(\frac{\partial L}{\partial t} + \frac{\partial L}{\partial q_a} \, \dot{q}_a + \frac{\partial L}{\partial \dot{q}_a} \, \ddot{q}_a \right) \Delta t + \frac{\partial L}{\partial q_a} \, \delta q_a + \frac{\partial L}{\partial \dot{q}_a} \, \delta \dot{q}_a \, . \tag{2.161}$$

With (2.157), the last term can be recast as

$$\frac{\partial L}{\partial \dot{q}} \, \delta \dot{q} = \frac{\partial L}{\partial \dot{q}} \, \frac{d}{dt} \, \delta q = \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}} \, \delta q \right) - \delta q \, \frac{d}{dt} \, \frac{\partial L}{\partial \dot{q}} \, . \tag{2.162}$$

Thus,

$$\Delta L = \frac{dL}{dt} \, \Delta t + \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_a} \, \delta q_a \right) + \left[\frac{\partial L}{\partial q_a} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_a} \right) \right] \delta q_a \, , \tag{2.163}$$

where

$$\frac{dL}{dt} = \frac{\partial L}{\partial t} + \frac{\partial L}{\partial q_a} \, \dot{q}_a + \frac{\partial L}{\partial \dot{q}_a} \, \ddot{q}_a \, . \tag{2.164}$$

The expression in the square brackets of (2.163) will play an important role in the subsequent discussion. We call it the *Eulerian* and denote

$$\mathcal{E}_a = \frac{\partial L}{\partial q_a} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_a} \ . \tag{2.165}$$

The general variation of the action, due to variations of q_a and t, is

$$\Delta S = \int_{t'_1}^{t'_2} dt' \, L(t', q', \dot{q}') - \int_{t_1}^{t_2} dt \, L(t, q, \dot{q}) \ . \tag{2.166}$$

This variation of the action may be considered as arising from two sources, namely, the Lagrangian variation, and the variation of the integration region,

$$\Delta S = \int_{t_1}^{t_2} dt \, \Delta L + \int_{t_1}^{t_2} dt \, \frac{d\Delta t}{dt} \, L \ . \tag{2.167}$$

Substitution of (2.163)–(2.165) and (2.155) in (2.167) gives

$$\Delta S = \int_{t_1}^{t_2} dt \left\{ \frac{d}{dt} \left[L\Delta t + \frac{\partial L}{\partial \dot{q}_a} \left(\Delta q_a - \dot{q}_a \Delta t \right) \right] + \mathcal{E}_a \delta q_a \right\} \ . \tag{2.168}$$

We define the *momentum conjugate to the coordinate* q_a

$$p_a = \frac{\partial L}{\partial \dot{q}_a} \ , \tag{2.169}$$

and the *Hamiltonian*

$$H = p_a \dot{q}_a - L \ . \tag{2.170}$$

With these definitions, the general variation of the action (2.168) becomes

$$\Delta S = \int_{t_1}^{t_2} dt \left[\frac{d}{dt} \left(p_a \Delta q_a - H \Delta t \right) + \mathcal{E}_a \delta q_a \right] \ . \tag{2.171}$$

Finally,

$$\Delta S = \left(p_a \Delta q_a - H \Delta t \right) \Big|_{t_1}^{t_2} + \int_{t_1}^{t_2} dt \, \mathcal{E}_a \delta q_a \ . \tag{2.172}$$

We now turn to the following problem of the calculus of variations: *find a minimum of the action S on a set of smooth curves with fixed endpoints.* The first term of (2.172) is zero since we require that variations of coordinates be vanishing at the endpoints, $\Delta q_a|_{t_1} = \Delta q_a|_{t_2} = 0$, and the integration limits fixed, $\Delta t|_{t_1} = \Delta t|_{t_2} = 0$. For S to be extremal, it is necessary that

$$\left(\frac{\partial S}{\partial \epsilon} \right)_{\epsilon = 0} = 0 \ . \tag{2.173}$$

By (2.154),

$$\int_{t_1}^{t_2} dt \, \mathcal{E}_a(q) \, \chi_a(t) = 0 \ . \tag{2.174}$$

Recall that the function $\chi_a(t)$ is arbitrary. Therefore,

$$\mathcal{E}_a(q) = 0 \, . \tag{2.175}$$

Indeed, on putting $\chi_a(t) = B_a\,\delta(t - t_*)$ where B_a is some constant vector, and $t_1 < t_* < t_2$, it follows from (2.174) that $\mathcal{E}_a(q(t_*)) = 0$. Since t_* is arbitrary, the pointwise vanishing of $\mathcal{E}_a$ can be established everywhere in the interval (t_1, t_2).

We come to *Hamilton's principle* which reads: *let the action of a given system be minimal on some world line connecting two fixed points in the event space, then this world line is described by the equations*

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_a} - \frac{\partial L}{\partial q_a} = 0 \, , \tag{2.176}$$

and the actual motion of the system proceeds along this world line.

Another name of this statement is the *principle of least action.* The behavior of the system with Lagrangian L is governed by the differential equations (2.176) which are called the *Euler–Lagrange equations.*

It is evident that (2.173) ensures an extremum, rather than minimum, of the action. In general, solutions to the Euler–Lagrange equations $q_a(t)$ show not a local minimum but only a saddle point of the action. This is the reason for the name *extremals* assigned to solutions of (2.176). Note, however, that the physical content of this principle does not amount to the condition (2.173). In fact, we endeavor to construct the Lagrangian in such a way as to ensure the availability of solutions minimizing the action.

Since the Lagrangian is a function of only q_a and $\dot{q}_a$ but is independent of higher derivatives, (2.176) comprises a set of ordinary differential equations of second order with the q's as unknown functions. To set up a Cauchy problem for these equations, initial data for the coordinates q_a and velocities $\dot{q}_a$ at some initial instant must be given. With this in mind, the state of the system is specified by $(q_1, \ldots, q_n)$ and $(\dot{q}_1, \ldots, \dot{q}_n)$.

If, on the other hand, the Lagrangian is assumed to involve higher derivatives, the system is called *rigid.* To specify states of a rigid system, the variables q_a and $\dot{q}_a$ must be supplemented with higher derivatives. Some aspect of the rigid dynamics will be reviewed in Sect. 10.1.

Using the definition of the momentum (2.169), the Euler–Lagrange equations (2.176) can be rewritten in the form reminiscent of Newton's second law,

$$\dot{p}_a = \frac{\partial L}{\partial q_a} \, . \tag{2.177}$$

In nonrelativistic mechanics, the Lagrangian is the difference between kinetic and potential energy, $L = T - U$. A close agreement of (2.177) and Newton's second law is attained when T is independent of q, and U is independent of $\dot{q}$. To illustrate, in rectilinear coordinates $q_i = \mathrm{z}_i$, the kinetic energy of a particle is $T = \frac{1}{2}m\mathbf{v}^2$, and if $U = U(\mathbf{z})$, the Euler–Lagrange equations (2.177) become

$$m\,\dot{\mathrm{v}}_i = -\nabla_i U \ . \tag{2.178}$$

In curvilinear coordinates, kinetic energy is q-dependent,

$$T(q,\dot{q}) = \frac{1}{2}\,m_{ab}(q)\dot{q}_a\dot{q}_b \ , \tag{2.179}$$

where the symmetric tensor $m_{ab} = m_{ba}$ can be treated as the metric in the configuration space $\mathcal{Q}$. In the general case, U depends not only on q, but also on $\dot{q}$, and hence can not be referred to as 'potential energy'. As we will see below, the term of the Lagrangian responsible for the electromagnetic interaction is just written as $U = U(q,\dot{q})$.

Replacing L by

$$L + \frac{d}{dt}\,f(t,q) \ , \tag{2.180}$$

where $f(t,q)$ is an arbitrary smooth function, leaves the Euler–Lagrange equations (2.176) unaffected. Indeed, this modifies S by the additional terms $f(t_2,q_2) - f(t_1,q_1)$. However, Hamilton's principle requires that the endpoints be subjected to no variation, hence the additional terms of S do not contribute to $\mathcal{E}_a(q)$ (for another argument see Problem 2.5.4).

We see that a given system is specified by the entire class of Lagrangians related to each other by the transformation (2.180) with arbitrary $f(t,q)$. If the dynamical variables q_a are subjected to a *canonical* transformation, $q_a = q_a(Q_b)$, then the new dynamical equations expressed in terms of Q's are a linear combination of the initial Euler–Lagrange equations, because

$$\frac{\delta S}{\delta Q_a} = \frac{\delta S}{\delta q_b}\,\frac{\partial q_b}{\partial Q_a} = 0 \ . \tag{2.181}$$

Thus, the Lagrangian $\Lambda(Q,\dot{Q})$ defined by the equation $\Lambda(Q,\dot{Q}) = L[q(Q),\dot{q}(Q)]$ provides an equivalent description of the given system.

If the Hamiltonian H, rather than the Lagrangian L, is taken as a basic quantity where complete dynamical information of a given system is encoded, then the canonical, or Hamiltonian, formalism must be developed.

The Lagrangian formalism can be easily arranged for the description of a relativistic particle. We identify the particle position at a given instant with a point z^μ of the event space $T\mathcal{Q}$ equipped with the Minkowski space metric. The world line $z^\mu(\tau)$ should be regarded as a function of an arbitrary time parameter τ, and $\dot{q}_a$ a vector tangent to the world line, $\dot{z}^\mu = dz^\mu/d\tau$. We take intersection points of the world line and two parallel spacelike hyperplanes as the endpoints $z^\mu(\tau_1)$ and $z^\mu(\tau_2)$ of the relativistic variational problem. The Euler–Lagrange equations (2.176) become

$$\frac{d}{d\tau}\,\frac{\partial L}{\partial \dot{z}^\mu} - \frac{\partial L}{\partial z^\mu} = 0 \ . \tag{2.182}$$

Note that we consider only infinite timelike or lightlike curves as the world lines to be tested. The relevance of this class of world lines to the least action principle is quite apparent. We abandon spacelike curves because the least action principle with endpoints (t_1, q_1) and (t_2, q_2) separated by arbitrary timelike intervals defies general formulation for such world lines. In addition, we require that the curves be smooth. Neglecting to do this would cause the occurrence of V- and Λ-shaped world lines, as in Fig. 3.3. It is clear, however, that a spacelike hyperplane may intersect a timelike V-shaped curve twice, otherwise it fails to intersect it at all. The same is true for Λ-shaped curves. Therefore, an action with endpoints separated by timelike intervals cannot be unambiguously defined for such world lines[5]. This subject is pursued further in Sect. 3.2, where an extra reason for exclusion of V- and Λ-shaped curves is adduced, *charge conservation*.

We now look into the relationship between *conserved quantities* (called alternatively *integrals of motion* or *constants of motion*) and *symmetries* of the action. Consider a p-parameter group of infinitesimal transformations G:

$$t \to t' = t + \Delta t, \quad q_a \to q'_a = q_a + \Delta q_a \ , \tag{2.183}$$

where Δt and Δq_a are linear in infinitesimal group parameters. If the action is left unchanged under (2.183), G is said to be *symmetry group*.

The Lagrangian L need not be invariant under this transformation, one may require instead that L be invariant up to a total time derivative,

$$L(t', q', \dot{q}') = L(t, q, \dot{q}) + \frac{d}{dt}\, l(q) \ , \tag{2.184}$$

but, most commonly, the Lagrangian is taken to be precisely invariant,

$$L(t', q', \dot{q}') = L(t, q, \dot{q}) \ . \tag{2.185}$$

Since our interest is with quantities which remain constant throughout the motion of the system, those are expressed in terms of solutions to the Euler–Lagrange equations (2.176). For $\mathcal{E}_a(q) = 0$, the variation of the action (2.172) becomes

$$\Delta S = \left(p_a \Delta q_a - H \Delta t\right)\Big|_{t_1}^{t_2} \ . \tag{2.186}$$

This expression makes it clear that Δq_a appearing in (2.183) are unrelated to the world line variations, and hence ΔS represents an increment of the action as a function of its endpoints $S = S(t_1, q_1; t_2, q_2)$, rather than the functional with fixed endpoints.

The invariance of S under the transformation (2.183) means that $\Delta S = 0$, and so

[5] We are not precluded from using a timelike curve with ‘harmless’ cusps which leave the orientation of timelike world lines unaltered. The least action principle for such world lines would be not a matter of concern. So, one may prefer to weaken the smoothness requirement, and regard such sectionally smooth, future oriented, curves as allowed world lines. However, this extension of the class of allowed world lines seems superfluous in the classical context.

$$(p_a \Delta q_a - H\Delta t)\,|_{t_2} = (p_a \Delta q_a - H\Delta t)\,|_{t_1} \ . \tag{2.187}$$

Since t_1 and t_2 are arbitrary, $p_a \Delta q_a - H\Delta t$ does not depend on time at all. We are now in position to formulate the first of two theorems by Emmy Noether (1918): *let the action be invariant under a continuous group of transformations (2.183), then there is a conserved quantity,*

$$\epsilon J = p_a \Delta q_a - H\Delta t \ , \tag{2.188}$$

which is built from functions $q_a(t)$ obeying the Euler–Lagrange equations (2.176). Thus, Noether's first theorem reveals explicitly a conserved quantity J which is associated with a given symmetry group[6].

To illustrate, we consider the following examples:

(i) $\Delta t = 0$, while Cartesian coordinates of particles $\mathbf{z}_I$ (index I labels particles) are translated by an arbitrary infinitesimal vector $\mathbf{c}$,

$$\mathbf{z}_I \to \mathbf{z}'_I = \mathbf{z}_I + \mathbf{c} \ , \tag{2.189}$$

that is, $\Delta \mathbf{z}_I = \mathbf{c}$. We apply (2.189) to (2.188) to give

$$\sum \mathbf{p}_I \cdot \mathbf{c} = \text{const} \ . \tag{2.190}$$

(From this point on, the summation Σ in each case extends over all the particles involved.) Since $\mathbf{c}$ is a fixed vector,

$$\mathbf{p} = \sum \mathbf{p}_I = \text{const} \ . \tag{2.191}$$

We thus recognize that the linear momentum $\mathbf{p}$ is conserved due to the invariance of S under space translations (2.189), manifesting spatial homogeneity 'all positions are equivalent'.

(ii) $\Delta q_a = 0$, while t is translated by a small amount ϵ,

$$t \to t' = t + \epsilon \ , \tag{2.192}$$

that is, $\Delta t = \epsilon$. Let the Lagrangian be explicitly time-independent. If $q_a(t)$ are extremals, the Lagrangian is invariant under (2.192) up to a total time derivative. Because dt is invariant under constant translations of t, $\Delta S = 0$. Then, by (2.188), we have

$$H = \sum \mathbf{p}_I \cdot \mathbf{v}_I - L = \text{const} \ . \tag{2.193}$$

Whenever the Hamiltonian H is conserved, it is regarded as the total energy and denoted by ε. Thus, the lack of explicit t-dependence of the

[6] It is well to bear in mind that the action S is required to be invariant under symmetry groups when it is expressed in terms of arbitrary $q_a(t)$ and $\dot{q}_a(t)$, while J in (2.188) is constant only on extremals.

Lagrangian, which can be viewed as time homogeneity 'all instants are equivalent', implies energy conservation.

It is conceivable that a Lagrangian might be invariant under translations of curvilinear coordinates q_a[7]. Generally such a translation invariance is unrelated to space homogeneity, and the corresponding conserved quantity p_a can not be interpreted as a linear momentum. For example, if the Lagrangian of a particle in a centrally symmetric field is expressed in terms of polar coordinates r, φ (Problem 2.5.2), the dependence on φ is absent from it, implying $p_\varphi = \text{const}$. This is a manifestation of spatial isotropy 'all directions are equivalent', and its associated conserved quantity is interpreted as angular momentum.

However, a direct linkage between the angular momentum conservation and spatial isotropy becomes quite evident when we use rectilinear coordinates.

(iii) $\varDelta t = 0$, while rectilinear coordinates of particles are rotated around some axis by a small angle, $\mathbf{z}_I \to \mathbf{z}'_I = \mathbf{z}_I + \varDelta\mathbf{z}_I$, where

$$(\varDelta \mathrm{z}_I)_i = \omega_{ij}\, \mathrm{z}^I_j \,. \tag{2.194}$$

Because ${\mathbf{z}'_I}^2 = \mathbf{z}^2_I$, the infinitesimal rotation parameters ω_{ij} form an antisymmetric tensor,

$$\omega_{ij} = -\omega_{ji} \,. \tag{2.195}$$

Let the Lagrangian be invariant under this rotation. Applying (2.194) and (2.195) to (2.188), we have

$$\omega_{ij} \sum p^I_i\, \mathrm{z}^I_j = \frac{1}{2}\, \omega_{ij} \sum \left(p^I_i\, \mathrm{z}^I_j - p^I_j\, \mathrm{z}^I_i\right) = \text{const} \,. \tag{2.196}$$

Thus, the conservation of the angular momentum

$$\mathrm{L}_{ij} = \sum \left(\mathrm{z}^I_i\, p^I_j - \mathrm{z}^I_j\, p^I_i\right) \tag{2.197}$$

is a consequence of rotational symmetry. In vector notation, the angular momentum takes a familiar form

$$\mathbf{L} = \sum \mathbf{r}_I \times \mathbf{p}_I \,. \tag{2.198}$$

We digress for a while and note that both the action S and the angular momentum $\mathbf{L}$ have dimension $[l][m][v]$. In quantum physics, $\mathbf{L}$ takes values which are integral multiples of Planck's constant $\hbar$. A great notational simplification has been gained by adopting a system of units in which the speed of light and Planck's constant are set equal to 1, the so-called *natural units*. With the help of these units, it is possible to cleanse the equations of quantum field theory of factors of c and $\hbar$. All quantities can be reduced to the

[7] A case in point is any Lagrangian which does not contain some coordinate q_a explicitly. For such Lagrangians, $\partial L/\partial q_a = 0$, and, by (2.177), $\dot{p}_a = 0$, hence $p_a = \text{const}$.

dimensions of a powers of length by multiplication with the requisite powers of c and $\hbar$. It is easy to check that H, $\mathbf{p}$ and m have the same dimension while $\mathbf{L}$ is dimensionless in natural units. On the other hand, since $[\mathbf{L}] = [l][m][v]$ and $[v] = 1$, mass is reciprocal to length. Choosing the unit of distance to be fm (a contraction of Fermi, or femtometer, 10^{-15} m), mass is measured in fm^{-1}. This scale of length is characteristic of subnuclear realm. Although our concern is with classical theory, we resort to natural units to facilitate dimensional arguments and to unify the nomenclature of classical and quantum descriptions.

If we assume certain constraints on the Lagrangian to be imposed, the converse of Noether's first theorem holds (Problem 2.5.5), which states that any constant of motion is associated with a symmetry.

To extend these results to the relativistic domain, we first note that space and time coordinates of a particle are fused here into a single spacetime event z_I^μ. The relativistic counterpart of the symmetry transformation (2.183) is therefore

$$z_I^\mu \to z'^\mu_I = z_I^\mu + \Delta z_I^\mu\ , \tag{2.199}$$

and the proper extension of (2.186) is

$$\Delta S = \sum \frac{\partial L}{\partial \dot z_I^\mu}\, \Delta z_I^\mu \Big|_{\tau_1^I}^{\tau_2^I}\ . \tag{2.200}$$

Invariance of the action under the transformation (2.199) implies the conservation law

$$\epsilon\, J = \sum \frac{\partial L}{\partial \dot z_I^\mu}\, \Delta z_I^\mu = \text{const}\ . \tag{2.201}$$

Let the Lagrangian L be invariant under spacetime translations[8]

$$z_I^\mu \to z'^\mu_I = z_I^\mu + \epsilon^\mu\ , \tag{2.202}$$

where ϵ^μ is an arbitrary infinitesimal vector. Then the total four-momentum

$$p_\mu = -\sum \frac{\partial L}{\partial \dot z_I^\mu} \tag{2.203}$$

is conserved. The overall minus sign in (2.203) is introduced in order to relate this formula to equation (2.169) which defines the spatial component of p_μ.

Another example is the infinitesimal Lorentz transformation

$$z_I^\mu \to z'^\mu_I = z_I^\mu + \omega^\mu{}_\nu z_I^\nu\ . \tag{2.204}$$

Because ${z'}^2 = z^2$, the infinitesimal group parameters consist of an antisymmetric tensor

[8] In general, such Lagrangians do not contain τ_I explicitly, and depend on only differences of coordinates of Ith and Jth particles $z_I^\mu - z_J^\mu$.

$$\omega_{\mu\nu} = -\omega_{\nu\mu} \, . \tag{2.205}$$

Invariance under this group implies the conservation of the total angular momentum

$$M_{\mu\nu} = \sum \left(z^I_\mu \, p^I_\nu - z^I_\nu \, p^I_\mu \right) \, . \tag{2.206}$$

The three spatial components M_{ij} are identical to the corresponding components of the tensor L_{ij} defined in (2.197). The conservation of remaining three components M_{0i} means that the velocity of the *center of mass* of the system is constant (Problem 2.5.3)

We now take a closer look at the Lagrangian description of a single relativistic charged particle. Reasoning from Noether's first theorem, we select *those actions which transform as scalars under the Poincaré group*. An appropriate action for a free particle is

$$S_{\mathrm{P}} = -m \int d\tau \, \sqrt{\dot{z}^\mu \, \dot{z}_\mu} \tag{2.207}$$

where m is some constant. The world line in (2.207) can be parametrized by any relevant time τ. We may, for instance, take τ to be laboratory time in a particular Lorentz frame t. Because

$$d\tau = \frac{d\tau}{dt} \, dt, \quad \dot{z}^\mu = \frac{dz^\mu}{d\tau} = \frac{dt}{d\tau} \, \frac{dz^\mu}{dt}, \quad \frac{dz^\mu}{dt} = (1, \mathbf{v}) \, , \tag{2.208}$$

expression (2.207) becomes

$$S_{\mathrm{P}} = -m \int dt \, \sqrt{1 - \mathbf{v}^2} \, . \tag{2.209}$$

This action is due to Poincaré and Planck.

From (2.209),

$$\mathbf{p} = \frac{\partial L}{\partial \mathbf{v}} = \frac{m\mathbf{v}}{\sqrt{1 - \mathbf{v}^2}} \, . \tag{2.210}$$

This $\mathbf{p}$ is identical to the momentum of a Galilean particle defined in (2.14), and hence m in (2.209) is a positive constant, the Newtonian mass. This justifies the overall minus sign in (2.207); otherwise the integral, which actually represents the length of the path between two points separated by a timelike interval, would be maximal rather than minimal, for a straight world line connecting those points.

The action (2.207) can be further generalized to any smooth manifold by substituting the pseudo-Riemannian metric $g_{\mu\nu}(x)$ for the Minkowski space metric $\eta_{\mu\nu}$,

$$S_{\mathrm{P}} = -m \int \sqrt{g_{\mu\nu}(z) \, dz^\mu \, dz^\nu} = -m \int ds \, . \tag{2.211}$$

The corresponding Euler–Lagrange equations are the geodesic equations (2.71) for the given metric $g_{\mu\nu}$ (Problem 2.5.6).

We next take into account the interaction of a charged particle with an electromagnetic field by appending the following term to the action

$$S_{\mathrm{S}} = -e\int d\tau\, \dot z^{\mu} A_{\mu}(z)\ . \tag{2.212}$$

Here, A_{μ} is the so-called *vector potential* of the electromagnetic field. In a particular Lorentz frame, $A^{\mu} = (\phi, \mathbf{A})$. Using (2.208), we write (2.212) as

$$S_{\mathrm{S}} = -e\int dt\,(\phi - \mathbf{v}\cdot\mathbf{A})\ . \tag{2.213}$$

The corresponding Lagrangian is

$$L_{\mathrm{S}} = -e\,\phi + e\,\mathbf{v}\cdot\mathbf{A}\ . \tag{2.214}$$

If $\mathbf{v} = 0$, the first term of (2.214) represents potential energy of a particle of the charge e in electrostatic field. Although the second term defies similar interpretation, one can envision (2.214) as the term of the Lagrangian which is responsible for the interaction between a charged particle and an electromagnetic field. It is interesting that the Lagrangian (2.214) was originally discovered by Karl Schwarzschild in 1903, before the advent of special relativity.

Consider the relativistic Euler–Lagrange equations

$$\frac{dp_{\mu}}{d\tau} = -\frac{\partial L}{\partial z^{\mu}} \tag{2.215}$$

for the Lagrangian

$$L = L_{\mathrm{P}} + L_{\mathrm{S}} = -m\sqrt{\dot z^{\mu}\,\dot z_{\mu}} - e\dot z^{\mu} A_{\mu}\ , \tag{2.216}$$

where L_{P} and L_{S} are its free and interaction parts. The four-momentum p_{μ} conjugate to the four-coordinate z^{μ} is

$$p_{\mu} = \frac{m\,\dot z_{\mu}}{\sqrt{\dot z^{\alpha}\,\dot z_{\alpha}}} + eA_{\mu}\ . \tag{2.217}$$

Because the line element ds relates to the differential $d\tau$ as

$$ds = \sqrt{\dot z^{\alpha}\,\dot z_{\alpha}}\,d\tau \tag{2.218}$$

(recall that $ds^2 = dz^{\alpha}\,dz_{\alpha}$), we have

$$\frac{\dot z^{\mu}}{\sqrt{\dot z^{\alpha}\,\dot z_{\alpha}}} = \frac{dz^{\mu}}{ds} = v^{\mu}\ , \tag{2.219}$$

and, therefore,

$$p^{\mu} = mv^{\mu} + eA^{\mu}\ . \tag{2.220}$$

Using the identity $v^2 = 1$, we obtain

$$(p - eA)^2 = m^2 \tag{2.221}$$

which is an extension of the relation $p^2 = m^2$ allowing for the electromagnetic interaction.

The τ derivative of A_μ can be expressed in terms of partial derivatives,

$$\frac{d}{d\tau} A_\mu = \frac{dz^\nu}{d\tau} \frac{\partial}{\partial z^\nu} A_\mu = \dot{z}^\nu \partial_\nu A_\mu \,. \tag{2.222}$$

This expression can be combined with

$$-\frac{\partial L}{\partial z^\mu} = e\dot{z}^\nu \partial_\mu A_\nu \tag{2.223}$$

to obtain an equation of motion involving the field strength tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \,. \tag{2.224}$$

We divide the Euler–Lagrange equations (2.215) by $\sqrt{\dot{z}^\alpha \, \dot{z}_\alpha}$, and use (2.218), (2.219), (2.222), (2.223), and (2.224), to give

$$ma_\mu = ev^\nu F_{\mu\nu} \,. \tag{2.225}$$

We arrive at the equation of motion for a charged particle (2.101) where the field strength $F_{\mu\nu}$ descended from the vector potential A_μ according to (2.224).

As the mathematical properties of the vector potential A_μ are reviewed in Chap. 4, we defer our discussion until then. The only pertinent remark is that the interaction term (2.212) is expressed in terms of an ‘auxiliary’ variable A_μ, rather than $F_{\mu\nu}$, the quantity which characterizes electromagnetic field states in a direct way.

Problem 2.5.1. Let N nonrelativistic particles experience an instantaneous mutual interaction, on the same principle as Newtonian action at a distance gravitation. Show that the sum of all forces applied to these particles is zero, in particular, for a two-particle system, the forces exerted on these particles are on the same line, equal, and oppositely directed when the potential energy U depends on the difference of particle coordinates.

Problem 2.5.2. Express the Lagrangian L for a relativistic particle in a spherically symmetric instantaneous potential $U(r)$ in terms of polar coordinates r and φ. Find from it the conjugate momenta and the Hamiltonian.

Answer

$$L = -m\sqrt{1 - \dot{r}^2 - r^2\dot{\varphi}^2} - U(r), \quad p_r = m\gamma\dot{r}, \quad p_\varphi = m\gamma r^2 \dot{\varphi} \,,$$
$$\gamma = (1 - \dot{r}^2 - r^2\dot{\varphi}^2)^{-1/2}, \quad H = \sqrt{m^2 + p_r^2 + \frac{p_\varphi^2}{r^2}} + U(r) \,. \tag{2.226}$$

Note that p_φ is identical to the angular momentum $\mathbf{L}$ defined in (2.198).

Problem 2.5.3. The total four-momentum P^μ of a closed system is a constant timelike vector pointing into the future. Taking P^μ for the time axis defines the *center-of-mass* (or *barycentric*) frame.

Show that the conservation of three components M_{0i} of the angular momentum $M_{\mu\nu}$ implies that the position of the center of mass

$$\mathbf{R} = \frac{\sum \varepsilon_I \, \mathbf{z}_I}{\varepsilon} \tag{2.227}$$

moves uniformly with the velocity

$$\mathbf{v} = \frac{\sum \mathbf{p}_I}{\varepsilon} \tag{2.228}$$

where $\varepsilon = \sum \varepsilon_I$ is the total energy of this system.

Problem 2.5.4. Verify that

$$\mathcal{L}_a = \frac{\partial}{\partial q_a} - \frac{d}{dt}\frac{\partial}{\partial \dot{q}_a} \tag{2.229}$$

annihilates identically any function which is a total derivative, that is,

$$\mathcal{L}_a \frac{d}{dt} f(t,q) = 0 \,. \tag{2.230}$$

Problem 2.5.5. Prove the converse of Noether's first theorem. The direct relation between symmetries and conservation laws is that symmetries imply conservation laws. The converse is that conservation laws imply symmetries. The direct theorem is valid for any reasonable function $L(q,\dot{q})$. However, the converse is more involved, and we must impose some additional restrictions on the Lagrangian. Consider a Lagrangian which does not depend explicitly on time, and is *nonsingular*. That is, the determinant of the Hessian matrix

$$H_{ab} = \frac{\partial^2 L}{\partial \dot{q}_a \partial \dot{q}_b} \quad (H_{ab} = H_{ba}) \tag{2.231}$$

is nonzero, $\det \|H_{ab}\| \neq 0$. Let $J(q,\dot{q})$ be a constant of motion, which means that the equation

$$\frac{d}{dt} J = 0 \tag{2.232}$$

is due to the Euler–Lagrange equations for this system. Suppose also that $J(q,\dot{q})$ can be expressed in terms of momenta

$$p_a = \frac{\partial L}{\partial \dot{q}_a} \tag{2.233}$$

and is homogeneous of first order in these p_a. [Recall that $F(t)$ is a homogeneous function of nth order if $F(kt) = k^n F(t)$ for any positive k.] The converse of Noether's theorem states: there exists an infinitesimal continuous transformation $q'_a = q_a + \Delta q_a$ corresponding to the constant of motion J such that the action S is invariant under this transformation. Find Δq_a in an explicit form. Verify that S is invariant under this transformation.

Answer

$$\Delta q_a = H^{-1}_{ab} \, \frac{\partial J}{\partial \dot{q}_b} \, \epsilon \tag{2.234}$$

where H^{-1}_{ab} is the inverse of the Hessian matrix, $H_{ab} \, H^{-1}_{bc} = \delta_{ac}$. To obtain this result, we proceed from (2.188) with $\Delta t = 0$, differentiate both sides of it with respect to $\dot{q}_b$, and apply H^{-1}_{ab}. H_{ab} is, by definition, $\partial p_a / \partial \dot{q}_b$, hence $H^{-1}_{ab} = \partial \dot{q}_b / \partial p_a$, and $\Delta q_a = (\partial J / \partial p_a)\epsilon$. The variation of the action under this transformation follows from (2.186) with $\Delta t = 0$,

$$\Delta S = (p_a \, \Delta q_a)\,\big|_{t_1}^{t_2} = \left(p_a \, \frac{\partial J}{\partial p_b} \, \epsilon \right) \Big|_{t_1}^{t_2} = \left[J(t_2) - J(t_1) \right] \epsilon = 0 \, . \tag{2.235}$$

The Euler theorem on homogeneous functions has been used, $p_a(\partial J / \partial p_a) = J$.

Problem 2.5.6. Derive the geodesic equation (2.71) for the metric $g_{\mu\nu}$ in a curved spacetime from the action (2.211).

Problem 2.5.7. *Damped oscillator.* A one-dimensional harmonic oscillator with friction is governed by the equation

$$\ddot{q} + \lambda \dot{q} + \omega^2 q = 0, \quad \lambda > 0 \, . \tag{2.236}$$

(For simplicity, we consider a particle with unit mass.) The solution to this equation is

$$q(t) = C e^{-\lambda t/2} \sin(\Omega t + \phi_0), \quad \Omega^2 = \omega^2 - \frac{\lambda^2}{4} \, . \tag{2.237}$$

This system is irreversible (that is, the equation of motion is not invariant under time reversal $t \to -t$) and nonconservative (that is, its total mechanical energy is not constant in time). Nevertheless, the equation of motion can be derived from the action

$$S = \int dt \left(\frac{1}{2} \, \dot{q}^2 - \frac{1}{2} \, \omega^2 q^2 \right) e^{\lambda t} \, . \tag{2.238}$$

Verify this.

Furthermore, there exists a canonical transformation converting this system to one of the following conservative, free of the damping, systems: a harmonic oscillator, a free particle, and a particle on a potential hill. Show that the transformation

$$Q = e^{\lambda t/2} q \tag{2.239}$$

yields a new Lagrangian

$$L' = \frac{1}{2}\dot{Q}^2 - \frac{1}{2}\Omega^2 Q^2 - \frac{\lambda}{4}\frac{d}{dt}Q^2 \ . \tag{2.240}$$

Because the last term is a total time derivative, it may be ignored. The new Lagrangian L' describes either a harmonic oscillator ($\Omega^2 > 0$), or a free particle ($\Omega^2 = 0$), or a particle on a potential hill ($\Omega^2 < 0$), depending on whether ω is higher, equal, or lower than $\lambda/2$.

An alternative Lagrangian description with the use of complex-valued coordinates is furnished by the action

$$S = \int dt \left[\dot{q}^*\dot{q} + \frac{\lambda}{2}\left(\dot{q}^* q - q^*\dot{q}\right) - \omega^2 q^* q\right] \ . \tag{2.241}$$

Find the equations for q and q^*, and solve them. Derive expressions for the momenta p and p^* conjugate to the coordinates q and q^*, and for the Hamiltonian H. Show that H is conserved, and offer a plausible explanation of this fact.

Answer

$$\ddot{q} + \lambda\dot{q} + \omega^2 q = 0, \quad \ddot{q}^* - \lambda\dot{q}^* + \omega^2 q^* = 0; \tag{2.242}$$

$$q(t) = C\, e^{-\lambda t/2}\sin(\Omega t + \phi_0), \quad q^*(t) = C^*\, e^{\lambda t/2}\sin(\Omega t + \phi_0); \tag{2.243}$$

$$p = \dot{q}^* - \lambda q^*/2, \quad p^* = \dot{q} + \lambda q/2; \tag{2.244}$$

$$H = p\,\dot{q} + p^*\dot{q}^* - L = \dot{q}^*\dot{q} + \omega^2 q^* q = (p^* - \lambda q^*/2)(p + \lambda q^*/2) + \omega^2 q^* q \ . \tag{2.245}$$

The coordinate q^* represents a 'mirror-image' oscillator with negative friction. The total energy of the aggregate system, containing both q- and q^*-subsystems, is conserved since the energy which is lost by the q-subsystem is gained by the q^*-subsystem. This clearly means that the action comprises a dissipative part coupled with an accumulative part.

2.6 Reparametrization Invariance

It was stated in Sect. 1.6 that a world line is an equivalence class of parametrized curves. By definition, a parametrized curve is a vector function z^μ of a real variable τ which takes values on some interval $\tau_1 \le \tau \le \tau_2$. Let τ be an increasing function of λ,

$$\tau = \tau(\lambda) \ , \tag{2.246}$$

obeying the conditions $\tau(\lambda_1) = \tau_1$ and $\tau(\lambda_2) = \tau_2$, then

$$y^\mu(\lambda) = z^\mu\left[\tau(\lambda)\right] \tag{2.247}$$

is another parametrization of the given curve. Equations (2.246) and (2.247) define a *reparametrization.*

What is the criterion for equivalence of parametrized curves? Two parametrized curves $z^\mu(\tau)$ and $y^\mu\,(\lambda)$ are said to be equivalent if the functional

$$S=\int_{\tau_1}^{\tau_2} d\tau\, L(z,\dot z) \tag{2.248}$$

takes identical values at $z^\mu(\tau)$ and $y^\mu\,(\lambda)$ for *any* Lagrangian L. In the modern calculus of variation, the curve $\mathcal{C}$ is defined as just such an equivalence class with reference to a certain $z^\mu(\tau)$.

However, it seems unreasonable to invoke such a refined mathematical definition. From the geometric point of view, any world line is a rather simple figure which can be smoothly straightened to the time axis. Observing that the change of variables (2.246) implies

$$d\tau=\frac{d\tau}{d\lambda}\,d\lambda\ , \tag{2.249}$$

we define a Lagrangian L as *acceptable* provided that

$$L\left[z(\tau),\dot z(\tau)\right]=\frac{d\lambda}{d\tau}\,L\left[y(\lambda),\dot y(\lambda)\right]\ . \tag{2.250}$$

Here, the dot denotes differentiation with respect to the natural argument, τ or λ. When condition (2.250) holds, the action (2.248) is called *reparametrization invariant.*

Because

$$\frac{dz^\mu}{d\tau}=\frac{d\lambda}{d\tau}\,\frac{dy^\mu}{d\lambda}\ , \tag{2.251}$$

condition (2.250) is met by Lagrangians which are homogeneous functions of the first order in $\dot z^\mu$, as exemplified by L defined in (2.216). Further examples are adduced below.

Let us discuss some implications of reparametrization invariance. It will be convenient to restrict our consideration to an infinitesimal reparametrization

$$\delta\tau=\epsilon(\tau)\ , \tag{2.252}$$

where ϵ is an arbitrary smooth function of τ close to zero, for which

$$\epsilon(\tau_1)=0,\quad \epsilon(\tau_2)=0\ . \tag{2.253}$$

Infinitesimal transformations of the form (2.252)–(2.253) constitute a group which may be looked upon as an *infinite* continuous group. Indeed, the Fourier expansion

$$\epsilon(\tau)=\sum_{n=0}^{\infty} c_n \sin\left[\frac{\pi n\,(\tau-\tau_1)}{T}\right],\quad T=\tau_2-\tau_1\ , \tag{2.254}$$

points clearly to an infinite set of parameters c_n. If the action is invariant under some infinite group, we run into the situation described by *Noether's second theorem*. We now consider the simplest version of this theorem. (For a generalization see Problem 2.6.6.)

Variation of the evolution parameter (2.252) implies a corresponding variation of the world line coordinates

$$\delta z^\mu = \dot{z}^\mu \epsilon \,. \tag{2.255}$$

Geometrically, (2.252) and (2.255) are an infinitesimal map of the given world line to the same world line. Hence, differentiation commutes with infinitesimal reparametrization,

$$\frac{d}{d\tau}\,\delta z^\mu = \delta \dot{z}^\mu \,. \tag{2.256}$$

In response to the reparametrization (2.252), the action (2.248) varies as

$$\delta S = \int d\tau \left(\frac{\partial L}{\partial z^\mu}\,\delta z^\mu + \frac{\partial L}{\partial \dot{z}^\mu}\,\delta \dot{z}^\mu \right) = \int d\tau \left(\frac{\partial L}{\partial z^\mu} - \frac{d}{d\tau}\frac{\partial L}{\partial \dot{z}^\mu} \right) \delta z^\mu = \int d\tau\, \mathcal{E}_\mu \dot{z}^\mu \epsilon \,, \tag{2.257}$$

where the second equation is obtained through integration by parts, using (2.253). Let

$$\delta S = 0 \,. \tag{2.258}$$

Assuming ϵ to be an arbitrary function of τ, one concludes that

$$\dot{z}^\mu \mathcal{E}_\mu = 0 \,. \tag{2.259}$$

This equation is a manifestation of Noether's second theorem which states: invariance of the action under the transformation group (2.252) involving an arbitrary infinitesimal function ϵ implies a linear relation between Eulerians.

The identity (2.259) suggests that $\mathcal{E}_\mu$ contains the projector $\overset{v}{\perp}$, which annihilates identically any vector parallel to $\dot{z}^\mu$. Thus, reparametrization invariance bears on the projection structure of the basic dynamical law (2.7). Recall that the projector $\overset{v}{\perp}$ arose initially from a completely different origin: the smooth embedding of Newtonian dynamics into sections of Minkowski space perpendicular to the world line.

Let us discuss some examples. Reparametrization invariance of the action

$$S = -\int d\tau \left(m\sqrt{\dot{z}\cdot\dot{z}} + e\,\dot{z}\cdot A \right) \tag{2.260}$$

is rather evident. Is it necessary to choose L as a homogeneous functions of the first order in $\dot{z}^\mu$ if we are to ensure reparametrization invariance of the action? The answer is: no. Consider the Lagrangian

$$L = -\frac{1}{2}\left(\eta\,\dot{z}^2 + \frac{m^2}{\eta} \right) \,, \tag{2.261}$$

which involves an auxiliary dynamical variable η, called the *einbein*. We assume that η transforms as

$$\delta\eta = \epsilon\,\dot\eta - \dot\epsilon\,\eta \tag{2.262}$$

in response to the infinitesimal reparametrization (2.252). Accordingly,

$$\delta\left(\frac{1}{\eta}\right) = \frac{d}{d\tau}\left(\frac{\epsilon}{\eta}\right) . \tag{2.263}$$

Combining (2.262) and (2.263) with the relation

$$\delta\dot z^\mu = \ddot z^\mu\,\epsilon + \dot z^\mu\,\dot\epsilon\,, \tag{2.264}$$

stemming from (2.255) and (2.256), one can readily show that the Lagrangian (2.261) transforms as

$$\delta L = \frac{d}{d\tau}\left(L\,\epsilon\right) . \tag{2.265}$$

It is then clear that the action

$$S = -\int d\tau\left[\frac{1}{2}\left(\eta\,\dot z^2 + \frac{m^2}{\eta}\right) + e\,\dot z\cdot A\right] \tag{2.266}$$

is invariant under transformation (2.252) subject to the boundary condition (2.253).

Varying the action (2.266) with respect to η gives the Euler–Lagrange equation

$$\dot z^2 - \eta^{-2}\,m^2 = 0\,. \tag{2.267}$$

If $m \neq 0$, then the solution to this equation is

$$\eta = m\,(\dot z\cdot\dot z)^{-1/2}\,. \tag{2.268}$$

This clarifies the physical meaning of η. If τ is chosen to be the time t in a particular Lorentz frame, the einbein proves to be identical to energy $\eta = m/\sqrt{1-\mathbf{v}^2}$, and if we take $\tau = s$, the einbein becomes mass $\eta = m$.

From (2.266), we find the four-momentum p^μ conjugate to the four-coordinate z^μ,

$$p^\mu = \eta\,\dot z^\mu + eA^\mu\,. \tag{2.269}$$

Combining (2.269) with (2.267), we arrive at the constraint

$$(p - eA)^2 - m^2 = 0\,, \tag{2.270}$$

which is identical to (2.221).

Varying the action (2.266) with respect to z^μ gives the Euler–Lagrange equation

$$\eta\,\ddot z_\mu + \dot\eta\,\dot z_\mu = e\dot z^\nu F_{\mu\nu}\,. \tag{2.271}$$

It is easy to check that (2.271), with η being given by (2.268), is identical to (2.225). Furthermore, substitution of (2.268) in (2.266), regains (2.260). The actions (2.260) and (2.266) are therefore equivalent in that they give the same Euler–Lagrange equations.

One may define a generalized Hamiltonian as

$$H = p \cdot \dot{z} + L \, . \tag{2.272}$$

This definition differs from that given by (2.170) in the overall minus sign, superimposed on the minus sign of the first term which arises from the definition of the four-momentum, (2.203). Because the quantity defined in (2.272) is a Lorentz scalar, it has nothing to do with the energy of the system transforming as the time component of the four-momentum. The generalized Hamiltonian corresponding to the action (2.266) is

$$H = \frac{1}{2\eta} \left[(p - eA)^2 - m^2 \right] \, . \tag{2.273}$$

This expression is zero due to the constraint (2.270). In the free particle case, we have the Hamiltonian

$$H = \frac{1}{2\eta} \left(p^2 - m^2 \right) \tag{2.274}$$

which is also zero. The reason why the generalized Hamiltonians (2.273) and (2.274) are zero is the reparametrization invariance of the corresponding actions. Note also that the generalized Hamiltonian resulting from the action (2.260) is identically zero.

However, the generalized Hamiltonian retains its validity as the generator of evolution. By this is meant that the *canonical equations*

$$\dot{z}^\mu = \frac{\partial H}{\partial p_\mu}, \quad \dot{p}^\mu = -\frac{\partial H}{\partial z_\mu} \tag{2.275}$$

with H given by (2.273) are converted to the equation of motion (2.271), Problem 2.6.1.

One may further define the Poisson brackets for any two quantities A and B dependent on z^μ and p^ν in the conventional way:

$$\{A, B\} = \frac{\partial A}{\partial z^\mu} \frac{\partial B}{\partial p_\mu} - \frac{\partial A}{\partial p_\mu} \frac{\partial B}{\partial z^\mu} \, . \tag{2.276}$$

This definition leads immediately to the canonical Poisson brackets

$$\{z^\alpha, z^\beta\} = 0, \quad \{p_\alpha, p_\beta\} = 0, \quad \{z^\alpha, p_\beta\} = \delta^\alpha{}_\beta \, , \tag{2.277}$$

and (2.275) becomes

$$\dot{z}^\mu = \{z^\mu, H\}, \quad \dot{p}^\mu = \{p^\mu, H\} \, . \tag{2.278}$$

We now dwell briefly on the Lagrangian description of a *massless* Galilean particle. The action (2.207) is unsuited for particles with $m = 0$. Indeed, the four-momentum p^μ derived from the action (2.207) vanishes as $m \to 0$, and the dynamics proves to be trivial.

By contrast, the action (2.266) is sound for both massive and massless particles. On putting $m = 0$, we obtain

$$S = -\int d\tau \left(\frac{1}{2}\, \eta\, \dot{z}^2 + e \dot{z} \cdot A \right) . \tag{2.279}$$

The variation of S with respect to η gives

$$\dot{z}^2 = 0 . \tag{2.280}$$

Therefore, *massless particles move along lightlike world lines.* Because the line element is zero, $dz^2 = 0$, the proper time s is no longer suitable for parametrization of such curves, and we should look at another variable τ for the parameter of evolution.

It follows from (2.280) that

$$\dot{z} \cdot \ddot{z} = 0 . \tag{2.281}$$

Since $\dot{z}^\mu$ is a null vector, $\ddot{z}^\mu$ may be either spacelike or null.

The variation of the action (2.279) with respect to z^μ gives the equation of motion for a massless particle which is identical to that for a massive particle (2.271). Thus, the only distinctive feature of the zero-mass case is the constraint (2.280). One may suspect that $\eta(\tau)$ is undetermined because this variable does not appear in (2.280). However, we are entitled to handle the reparametrization freedom making the dynamical equations as simple as possible. For some choice of the evolution variable, the einbein can be converted to a constant, $\eta = \eta_0$, Problem 2.6.2. Then (2.271) becomes

$$\eta_0 \ddot{z}_\mu = e \dot{z}^\nu F_{\mu\nu} . \tag{2.282}$$

For simple external field configurations, it is advisable to apply the technique developed in the massive case to equation (2.282), Problem 2.6.3.

Charged particles with zero mass do not appear to exist. This fact is rather strange if we remember that massless particles interact with the Yang–Mills field. For example, neutrinos interact with the SU(2) × U(1) Yang–Mills field that accounts for the W and Z vector boson dynamics. However, recent experiments suggest that neutrinos are endowed with a finite, albeit very small, mass. Were this indeed the case, the absence of massless charged particles from nature would be no great surprise.

Meanwhile the idea of a free massless particle is presently not uncommon. We now give some attention to such particles. Switching off the electromagnetic interaction in (2.279), $e = 0$, the action becomes

$$S = -\frac{1}{2} \int d\tau\, \eta\, \dot{z}^2 . \tag{2.283}$$

We indicate the four-momentum of a free massless particle by k^μ. From (2.283)

$$k^\mu = \eta\, \dot{z}^\mu \, . \tag{2.284}$$

By (2.280), k^μ is a null vector,

$$k^2 = 0 \, . \tag{2.285}$$

When this result is compared with definition of the mass (2.20), it is apparent that the action (2.283) describes an object with $M = 0$, that is, a zero-mass particle in the strict sense of the word.

Since a massless particle is free, its four-momentum k^μ is constant. It is possible to orient the axes so that we have

$$k^\mu = \omega\,(1, 1, 0, 0) \, , \tag{2.286}$$

where ω may be interpreted as the particle energy in this Lorentz frame.

We finally give a cursory glance at two reparametrization invariant actions for a particle interacting with a scalar field $\phi(x)$,

$$S = -\int d\tau \, [m - g\, \phi(z)] \sqrt{\dot{z}^\mu\, \dot{z}_\mu} \, , \tag{2.287}$$

and a symmetric tensor field $\phi_{\mu\nu}(x)$,

$$S = -\int d\tau \left[m - g\, \frac{\dot{z}^\alpha \dot{z}^\beta}{\dot{z}^2}\, \phi_{\alpha\beta}(z) \right] \sqrt{\dot{z}^\mu\, \dot{z}_\mu} \, . \tag{2.288}$$

Note that (2.287) and (2.288) have much in common with the action (2.260) for a charged particle coupled with the electromagnetic vector potential $A_\mu(x)$.

Problem 2.6.1. Show that the variation of the action

$$S = \int d\tau \, [-p \cdot \dot{z} + H(z, p)] \tag{2.289}$$

in z^μ and p^μ (assumed to be independent variables) yields the canonical equations (2.275). With the Hamiltonian (2.273), this action becomes

$$S = \int d\tau \left\{ -p \cdot \dot{z} + \frac{1}{2\eta} \left[(p - eA)^2 - m^2 \right] \right\} \, . \tag{2.290}$$

We thus conclude that η^{-1} plays the role of the Lagrange multiplier in the free variation problem corresponding to a conditional variation problem with the constraint (2.270). Verify that this S is equivalent to (2.266). Show that the canonical equations resulting from this action can be converted to the equation of motion (2.282).

Problem 2.6.2. Show that η must transform under global reparametrizations as

$$\eta(\tau) = \frac{d\tau}{d\lambda}\,\eta'(\lambda)\,. \tag{2.291}$$

Let η be a given function of the label time τ. Using the above transformation law, find the function $\lambda(\tau)$ expressing the new label time in terms of the old one such that the transformed einbein be constant, $\eta'(\lambda) = \eta_0$.

Answer

$$\lambda(\tau) = \eta_0 \int^{\tau} \frac{d\tau'}{\eta(\tau')}\,. \tag{2.292}$$

Problem 2.6.3. Determine the world lines $z^\mu(\tau)$ of a massless charged particle in a constant homogeneous field of (a) electric type, defined in (2.102), and (b) magnetic type, defined in (2.125). Specify the evolution parameter τ through the condition $\eta(\tau) = \eta_0$. At the initial instant $\tau = 0$, put: (a') $\dot z^\mu(0) = e_0^\mu + e_1^\mu$, ($a''$) $\dot z^\mu(0) = e_0^\mu + e_1^\mu \cos\alpha + e_2^\mu \sin\alpha$, and ($b$) $\dot z^\mu(0) = e_0^\mu + e_1^\mu \cos\alpha + e_2^\mu \sin\alpha\cos\beta + e_3^\mu \sin\alpha\sin\beta$. Express the evolution parameter τ in terms of the laboratory time z^0. Find the particle energy.

Determine the world line of a massless charged particle in a constant homogeneous null field, defined in (2.86), assuming that $\dot z^\mu(0) = e_0^\mu + e_2^\mu$.

Determine the motion of a massless charged particle in constant homogeneous parallel electric and magnetic fields, as defined in (2.89), assuming that $\dot z^\mu(0) = e_0^\mu + e_2^\mu$.

Determine the world line of a massless charged particle in a plane-wave field specified in Problem 2.4.2, assuming that $\dot z^\mu(0) = e_0^\mu + e_2^\mu$.

Compare these lightlike world lines of a massless charged particle with their respective timelike world lines of a massive charged particle, analyzed in Sect. 2.4.

Problem 2.6.4. Show that the equation of motion for a particle interacting with a scalar field (2.63) results from the action (2.287).

Problem 2.6.5. Derive the equation of motion for a particle interacting with a tensor field from the action (2.288).

Answer

$$\frac{d}{ds}\left[m v_\mu + g\left(\phi_{\alpha\beta} v^\alpha v^\beta v_\mu - 2\phi_{\mu\nu} v^\nu\right)\right] = -v^\alpha v^\beta \partial_\mu \phi_{\alpha\beta}\,. \tag{2.293}$$

Problem 2.6.6. *Noether's second theorem.* Consider a transformation of q_a of the form

$$\delta q_a = A_a(\tau, q, \dot q)\,\epsilon + B_a(\tau, q, \dot q)\,\dot\epsilon\,, \tag{2.294}$$

where ϵ is an infinitesimal function obeying the boundary conditions (2.253). Note that (2.294) is a transformation of the type shown in (2.262). Let the action S be invariant under (2.294). Then the Eulerians $\mathcal{E}^a$ resulting from this action satisfy the identity

$$\mathcal{E}^a A_a - \frac{d}{d\tau}\left(\mathcal{E}^a B_a\right) = 0\,. \tag{2.295}$$

Prove this statement.

To be specific, take the Lagrangian (2.261) and the transformation

$$\delta z_\mu = \dot z_\mu\,\epsilon, \quad \delta\eta = \dot\eta\,\epsilon - \eta\,\dot\epsilon\ . \tag{2.296}$$

Verify that (2.295) holds for the Eulerians $\mathcal{E}^\mu(z) = \eta\ddot z^\mu + \dot\eta\dot z^\mu$ and $\mathcal{E}(\eta) = -\frac{1}{2}\left(\dot z^2 - m^2\eta^{-2}\right)$.

2.7 Spinning Particle

So far we were concerned with particles devoid of internal degrees of freedom. Let us now turn to an object with intrinsic angular momentum, spin. We outline a model proposed by Yakov Frenkel in 1926, the first consistent description of a classical point particle with spin. Such a spinning particle may be visualized as a tiny top whose size tends to zero. For simplicity, we restrict our discussion to the free particle case.

Let a spinning particle be moving along a world line $z^\mu(s)$. Our immediate task is to find the form of this world line.

In the absence of external fields, spacetime is homogeneous and isotropic, hence the linear and angular momenta are conserved. The expression for the angular momentum of a spinless particle (2.206) should be modified as

$$M_{\mu\nu} = z_\mu p_\nu - z_\nu p_\mu + \sigma_{\mu\nu}\ , \tag{2.297}$$

where $\sigma_{\mu\nu}$ is an antisymmetric real-valued spin tensor. We call $L_{\mu\nu} = z_\mu p_\nu - z_\nu p_\mu$ the *orbital momentum*, in contrast with the total angular momentum $M_{\mu\nu}$ which may be a combination of $L_{\mu\nu}$ with the intrinsic angular momentum $\sigma_{\mu\nu}$. In a particular Lorentz frame, $\sigma_{\mu\nu} = (\mathbf{N}, \mathbf{S})$, implying $\sigma_{0i} = \mathrm{N}_i$ and $\sigma_{ij} = -\epsilon_{ijk}\,\mathrm{S}_k$, or

$$\sigma_{\mu\nu} = \begin{pmatrix} 0 & \mathrm{N}_1 & \mathrm{N}_2 & \mathrm{N}_3 \\ & 0 & -\mathrm{S}_3 & \mathrm{S}_2 \\ & & 0 & -\mathrm{S}_1 \\ & & & 0 \end{pmatrix}\ . \tag{2.298}$$

Note that it is just $\mathbf{S}$ which is associated with intrinsic spatial rotations.

We write the conservation laws in differential form:

$$\dot p^\mu = 0\ , \tag{2.299}$$

$$\dot M_{\mu\nu} = 0\ . \tag{2.300}$$

There is no reason to augment them with the addition of spin conservation $\dot\sigma_{\mu\nu} = 0$, because no additional symmetry responsible for this constant of motion is available. By (2.300), (2.297), and (2.299),

$$\dot\sigma_{\mu\nu} = p_\mu v_\nu - p_\nu v_\mu\ . \tag{2.301}$$

Frenkel imposed a further constraint on the spin tensor $\sigma_{\mu\nu}$: in the rest frame only the components of $\mathbf{S}$ are nonzero, while $\mathbf{N} = \mathbf{0}$. This is the reason to call the vector $\mathbf{S}$ spin. In the rest frame, $v^\mu = (1,0,0,0)$, hence Frenkel's constraint takes the invariant form

$$\sigma_{\mu\nu} v^\mu = 0 \,. \tag{2.302}$$

One may define

$$\mathcal{B} = \frac{1}{2}\,\sigma_{\mu\nu}\sigma^{\mu\nu} = \mathbf{S}^2 - \mathbf{N}^2 \,. \tag{2.303}$$

In the rest frame, $\mathcal{B} = \mathbf{S}^2 > 0$ (recall that $\mathbf{S}$ is an ordinary Euclidean vector). Therefore, the spin magnitude in the rest frame is a Lorentz scalar,

$$\sigma_{\mu\nu}\sigma^{\mu\nu} = 2\,\mathbf{S}^2 \,. \tag{2.304}$$

From here on, we use the symbol $\mathbf{S}$ to mean spin as viewed by a comoving observer.

One further useful relation can be easily deduced (Problem 2.7.1),

$$\sigma_{\lambda\mu}\sigma^{\mu\nu}\sigma_{\nu\rho} = -\mathbf{S}^2\sigma_{\lambda\rho} \,. \tag{2.305}$$

Let us return to the equation of spin evolution (2.301). We contract it with $\sigma^{\mu\nu}$ and use (2.302), to give

$$\sigma^{\mu\nu}\frac{d}{ds}\,\sigma_{\mu\nu} = \frac{1}{2}\,\frac{d}{ds}\,\sigma^{\mu\nu}\sigma_{\mu\nu} = \frac{d}{ds}\,\mathbf{S}^2 = 0 \,. \tag{2.306}$$

We see that the spin magnitude $|\mathbf{S}|$ is constant, while the direction of $\mathbf{S}$ may precess around some (moving) axis. That is why Frenkel's particle is sometimes referred to as a 'pure gyroscope'.

The mass M and rest mass m of a spinning particle are defined as usual by (2.20) and (2.21). Since we consider a free particle, we assume that v^μ and p^μ are timelike future directed vectors. Thus, $M^2 > 0$ and $m > 0$. By (2.299), M does not vary in time. We will see later that $m = \text{const}$ as well.

Let us define the vector quantity

$$\zeta^\mu = \sigma^{\mu\nu} p_\nu \,. \tag{2.307}$$

From (2.307) and (2.302) it follows that

$$\zeta \cdot p = 0, \quad \zeta \cdot v = 0 \,. \tag{2.308}$$

It can be shown (Problem 2.7.2) that ζ^μ is a spacelike vector,

$$\zeta^2 < 0 \,. \tag{2.309}$$

Equation (2.301) can be recast in the form

$$\dot\zeta^\mu = -M^2 v^\mu + m p^\mu \,. \tag{2.310}$$

Contraction with ζ_μ yields $\zeta^2 = \text{const}$. This means that only the direction of ζ^μ varies in time, not the magnitude.

Differentiation of (2.305), contraction with v^ρ, and making use of (2.301) leads to

$$mv^\lambda = p^\lambda + \frac{1}{\mathbf{S}^2}\,\sigma^{\lambda\mu}\zeta_\mu \;, \tag{2.311}$$

and further contraction with p_λ results in

$$m^2 = M^2 - \frac{\zeta^2}{\mathbf{S}^2} \;. \tag{2.312}$$

Thus, m is a constant of motion because such are quantities in the right hand side of (2.312). Combining (2.312) and (2.309), we conclude that

$$m^2 > M^2 \;. \tag{2.313}$$

Why $M \neq m$? The mass is not identical to the rest mass because v^μ and p^μ are not collinear. To see this, we differentiate (2.311) and take into account (2.299), (2.301), (2.308), (2.310), (2.302), and (2.307). The result is

$$\mathbf{S}^2\dot{v}^\mu = \zeta^\mu \;. \tag{2.314}$$

Further differentiation and using (2.310) leads to

$$\mathbf{S}^2\ddot{v}^\mu + M^2 v^\mu = mp^\mu \;. \tag{2.315}$$

We see that p^μ may be proportional to v^μ, provided that $\ddot{v}^\mu = 0$ (the condition which is trivially met for a free Galilean particle by virtue of the equation of motion in the absence of external forces). In general, such is not the case for a spinning particle however.

The reader will easily observe the similarity of (2.315) with the equation of a harmonic oscillator under the action of an external constant force. Thus a solution to (2.315) is

$$v^\mu(s) = \frac{m}{M^2}\,p^\mu - \alpha^\mu \sin\omega s + \beta^\mu \cos\omega s \;, \tag{2.316}$$

where p^μ, α^μ, and β^μ are arbitrary except that they are subject to the conditions $\alpha\cdot p = \beta\cdot p = \alpha\cdot\beta = 0,\ \alpha^2 = \beta^2$, and $\omega = M/|\mathbf{S}|$. One more integration gives the world line:

$$z^\mu(s) = z^\mu(0) + \frac{m}{M^2}\,p^\mu s + \frac{\alpha^\mu}{\omega}\,\cos\omega s + \frac{\beta^\mu}{\omega}\,\sin\omega s \;. \tag{2.317}$$

This is a helix, wound around a fixed axis parallel to p^μ. The particle moves uniformly along a straight line and rotates about it. On a large scale, this appears as a tiny oscillation about the uniform motion. Only for $\alpha^\mu = \beta^\mu = 0$, the world line is straight. This non-Galilean behavior of a free particle was discovered by Erwin Schrödinger in 1930; since then, this oscillatory regime of motion bears the German name 'Zitterbewegung'.

If we assume that $p^2 < 0$, then (2.317) is replaced by

$$z^\mu(s) = -\frac{m}{\mathcal{M}^2}\, p^\mu s + \frac{\alpha^\mu}{\Omega}\cosh \Omega s + \frac{\beta^\mu}{\Omega}\sinh \Omega s\,, \tag{2.318}$$

where $\mathcal{M}^2 = -p^2$, $\Omega = \mathcal{M}/|\mathbf{S}|$, and α^μ and β^μ meet the condition $\alpha^2 = -\beta^2$. This solution, describing motion across the plane spanned by two vectors p^μ and α^μ, shows a rapid increase in velocity, with the velocity of the particle tending ultimately to the velocity of light.

If we had a Galilean particle with $p^\mu = mv^\mu$, then the ban against spacelike world lines would automatically exclude spacelike four-momenta. The four-momentum of a spinning particle p^μ is expressed in terms of kinematical variables through equation (2.315), written from the right to the left. Hence the requirement that world lines are timelike does not imply that $p^2 \geq 0$. If the condition $p^2 \geq 0$ is imposed explicitly, then the zitterbewegung (2.317) is among the allowed motions, while the runaway (2.318) is not.

One may average equation (2.316) over s, that is, apply the operation

$$< F >= \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} ds\, F(s) \tag{2.319}$$

to each term. Then

$$< v^\mu >= \frac{m}{M^2}\, p^\mu\,. \tag{2.320}$$

Let us trace the motion of a point with coordinate

$$y^\mu = z^\mu + \frac{1}{M^2}\,\zeta^\mu\,. \tag{2.321}$$

By (2.310) and (2.320),

$$\dot y^\mu = \frac{m}{M^2}\, p^\mu =< v^\mu > \;. \tag{2.322}$$

A point with coordinate y^μ traces a straight world line parallel to the vector p^μ. This point is interpreted as the 'center of mass' of the oscillating gyroscope. The conserved four-momentum p^μ must be assigned to an imagined carrier that is located at the center of mass and moves along the averaged world line.

Equations (2.311) and (2.299) can readily be derived in the form of the canonical equations (2.275) or (2.278) using the Hamiltonian

$$H = \frac{1}{2\eta}\left(p^2 - m^2 - \frac{\zeta^2}{\mathbf{S}^2}\right) \tag{2.323}$$

together with the canonical Poisson brackets (2.277) and

$$\{x_\lambda, \sigma_{\mu\nu}\} = \{p_\lambda, \sigma_{\mu\nu}\} = 0,\;\; \{\sigma_{\mu\nu}, \sigma_{\rho\sigma}\} = \sigma_{\mu\rho}\eta_{\nu\sigma} + \sigma_{\nu\sigma}\eta_{\mu\rho} - \sigma_{\mu\sigma}\eta_{\nu\rho} - \sigma_{\nu\rho}\eta_{\mu\sigma}\,. \tag{2.324}$$

These Poisson brackets for spin variables bear a general resemblance to the commutation relations for generators of the Poincaré group discussed in Sect. 1.5. If we choose the parametrization of the world line such that the einbein η is fixed as $\eta = m$, then the parameter of evolution τ is the proper time s. Note that the expression in parentheses in (2.323) is the constraint (2.312), whereby $H = 0$. Hamiltonian (2.323) differs from that for a Galilean particle (2.274) by the presence of the last term, which generates evolution for the spin degrees of freedom. Equation (2.310) can also be written as the canonical equation

$$\dot{\zeta} = \{\zeta, H\} \tag{2.325}$$

if we use Hamiltonian (2.323) and Poisson brackets (2.324).

The inertia of a pure gyroscope is specified by two scalars, M and m, defined in (2.20) and (2.21). In the absence of external forces, both M and m are conserved, and $m > M$.

If we attempt to define a zero-mass gyroscope, we encounter a problem. What is meant by zero mass: $M = 0$ or $m = 0$? If we adopt $M = 0$, then the role of the positive conserved scalar quantity m is obscure. If, on the other hand, we prefer the condition $m = 0$, then $M^2 < 0$, and the free 'zero-mass particle' is actually in a tachyonic state.

Problem 2.7.1. Prove (2.305).

Hint One should first show that the 2-form σ obeying the constraint (2.302) can be represented as

$$\sigma = |\mathbf{S}|\, e_1 \wedge e_2 \;, \tag{2.326}$$

where $|\mathbf{S}|$ is the spin magnitude in the rest frame in which $v^\mu = (1, 0, 0, 0)$, $e_1^\mu = (0, 1, 0, 0)$, and $e_2^\mu = (0, 0, 1, 0)$.

Problem 2.7.2. Prove (2.309).

Problem 2.7.3. So far we used only *kinematical* rest frames in which $v^\mu = (1, 0, 0, 0)$. One may also consider *dynamical* rest frames in which $p^\mu = M\,(1, 0, 0, 0)$. Such frames are closely related to the notion of the center of mass, as is clear from (2.321) and (2.322).

Show that

$$M_{\mu\nu} = y_\mu p_\nu - y_\nu p_\mu + \Xi_{\mu\nu} \;, \tag{2.327}$$

where y_μ is the center-of-mass coordinate defined in (2.321), and

$$\Xi_{\mu\nu} = \sigma_{\mu\nu} - \frac{1}{M^2}\left(\zeta_\mu p_\nu - \zeta_\nu p_\mu\right) \;. \tag{2.328}$$

The tensor $\Xi_{\mu\nu}$ plays the same role now as did $\sigma_{\mu\nu}$. Prove the constraint

$$\Xi_{\mu\nu} p^\nu = 0 \;, \tag{2.329}$$

a counterpart of (2.302). Prove relations analogous to (2.326), (2.304), and (2.305), changing $\sigma_{\mu\nu}$ to $\Xi_{\mu\nu}$. Verify that

$$\dot{\Xi}_{\mu\nu} = 0 \tag{2.330}$$

replaces (2.301).

Problem 2.7.4. *Model with Grassmannian variables.* One would like to use real-valued odd elements of a Grassmann algebra θ^μ and θ_5 to describe spin degrees of freedom. An appropriate reparametrization invariant action is

$$\begin{aligned} S = \int_{\tau_1}^{\tau_2} d\tau \left[-p_\mu \dot{z}^\mu + \frac{1}{2\eta}\left(p^2 - m^2\right) - \frac{i}{2}\left(\dot{\theta}^\mu \theta_\mu + \dot{\theta}_5 \theta_5\right) + i\chi\left(\theta^\mu p_\mu + m\theta_5\right)\right] \\ - \frac{i}{2}\left[\theta^\mu(\tau_1)\theta_\mu(\tau_2) + \theta_5(\tau_1)\theta_5(\tau_2)\right] , \end{aligned} \tag{2.331}$$

and the endpoint variation conditions are

$$\Delta z^\mu(\tau_1) = \Delta z^\mu(\tau_2) = 0, \quad \Delta\theta^\mu(\tau_1) + \Delta\theta^\mu(\tau_2) = 0, \quad \Delta\theta_5(\tau_1) + \Delta\theta_5(\tau_2) = 0 . \tag{2.332}$$

The Grassmannian variable $\chi(\tau)$ plays the role of a Lagrange multiplier of the constraint. The even Grassmannian construction $i\theta^\mu\theta^\nu$ is similar to the spin tensor $\sigma^{\mu\nu}$.

Derive the dynamical equations

$$\dot{p}^\mu = 0 , \tag{2.333}$$

$$-\dot{z}^\mu + \eta^{-1}p^\mu + i\chi\theta^\mu = 0 , \tag{2.334}$$

$$-\dot{\theta}^\mu + \chi p^\mu = 0 , \tag{2.335}$$

$$-\dot{\theta}_5 + \chi m = 0 \tag{2.336}$$

(proper time is chosen to be the parameter of evolution: $\tau = s$) and the constraints

$$p^2 - m^2 = 0 , \tag{2.337}$$

$$\theta^\mu p_\mu + m\,\theta_5 = 0 . \tag{2.338}$$

At first glance, the dependence between four-momentum and four-velocity (2.334), is a direct analogue of (2.311), which was shown to imply the non-Galilean behavior of Frenkel's pure gyroscope. However, if we seek solutions describing a world line and four-momentum in the field of real numbers, this resemblance proves deceptive. Indeed, because $\theta^0\theta^0 = \theta^1\theta^1 = \theta^2\theta^2 = \theta^3\theta^3 = 0$, it follows from (2.334) that

$$(\dot{z}^0 - \eta^{-1}p^0)^2 = (\dot{z}^1 - \eta^{-1}p^1)^2 = (\dot{z}^2 - \eta^{-1}p^2)^2 = (\dot{z}^3 - \eta^{-1}p^3)^2 = 0 . \tag{2.339}$$

We see that p^μ is parallel to $\dot{z}^\mu$. But $\dot{z}^\mu$ is a timelike vector pointing to the future, hence, in view of (2.337), $\eta = m$. Equation (2.334) is satisfied only for $\chi = 0$. From (2.335) and (2.336) it follows that $\dot{\theta}^\mu = 0$, $\dot{\theta}_5 = 0$, while (2.333) and (2.334) imply $\dot{z}^\mu = \text{const}$. Thus, spin and configuration variables evolve independently. As to inertia of this particle, it is characterized by a single quantity, $m = M$

It is important to realize that the set of even elements of a Grassmann algebra contains a subset of real numbers (c-numbers) which is a strict embedding. Among the unknown functions appearing in (2.333)–(2.338), even variables, such as $\chi\theta^\mu$, need not be c-number quantities, and the complete collection of solutions to equations (2.333)–(2.338) is not a unique solution describing a straight world line. However, a world line $z^\mu(s)$, built from even Grassmannian variables that are not c-numbers, has no operational definition.

2.8 Relativistic Kepler Problem

Now that we have a general grasp of a single relativistic particle, we need to give some attention to two-particle problems. Let two charged particles be moving along timelike world lines as in Fig. 2.2. Because electromagnetic interactions propagate with the speed of light, particle 2 situated at a point O, receives an impulse which is sent by particle 1 from the retarded point A separated from O by a null interval. It is clear that Newton's action–reaction law fails if restricted to just the two particles. The four-force f^μ_{21} exerted on particle 2 at O need not be equal in magnitude and opposite in direction to the four-force f^μ_{12} acting on particle 1 at A because O and A are not simultaneous. Alternatively, one may take the fragment of the world line $z^\mu_1(s_1)$ bounded by A and B, the points of intersection of this world line with the light cone at O. Every point within this fragment is separated from O by a spacelike interval, hence simultaneous with O, as seen by some Lorentz observer. However, f^μ_{12} varies from one point to another, and may be equal in magnitude and opposite in direction to f^μ_{21} at O quite accidentally.

This retarded action-at-a-distance two-particle problem comprises the system of six ordinary differential equations, which are in fact differential-delay

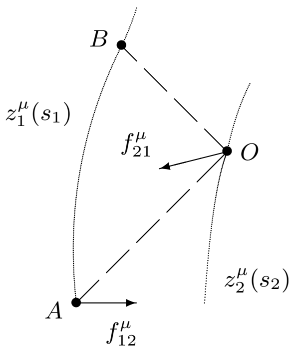


Fig. 2.2. Two-particle problem

equations with retarded arguments. The exact solution to this problem is a challenging task. (Strange though it may seem, when it comes to a realm with one time and one space dimension, this problem is readily solved. We will review this solution in Sect. 10.2.)

With simplifying assumptions, we can pose this problem in a more tractable way. A key assumption is that the retardation is a small effect which can be included by means of a perturbation expansion. Let us fix a particular Lorentz frame. To a first approximation, we take electromagnetic interactions as instantaneous, and retain only Coulomb terms. This two-particle problem is called the *relativistic Kepler problem*.

Our next assumption is rather technical: we take one particle to be much heavier than another[9]. As a first approximation, the former particle may be thought of as infinitely massive and hence motionless. Our prime concern here is with the attractive Coulomb interaction (for repulsive interactions see Problem 2.8.1). We thus come to the following single-particle problem: describe the motion of a particle of mass m and charge $-e$ in a centrally symmetric attractive field generated by a static charge Ze placed at the origin.

Anticipating the Coulomb field (which will be detailed in Sect. 4.1), the equation of motion under an attractive central force varying inversely as the square of the distance is

$$\frac{d}{dt}\, m\gamma\mathbf{v} = -\frac{Ze^2}{r^2}\,\mathbf{n}\ , \tag{2.340}$$

where $\mathbf{n}$ is a unit vector aligned with the radius vector, $\mathbf{n} = \mathbf{r}/r$. The spherical symmetry of the force implies conservation of orbital momentum $\mathbf{L} = \mathbf{r} \times \mathbf{p}$. Indeed, taking the cross product of (2.340) with $\mathbf{r}$, and using the identities $\mathbf{r} \times \mathbf{r} = 0$ and $\mathbf{v} \times \mathbf{v} = 0$, we have

$$\frac{d}{dt}\,\mathbf{r} \times m\gamma\mathbf{v} = 0\ , \tag{2.341}$$

whence

$$\mathbf{L} = \mathbf{r} \times m\gamma\mathbf{v} = \text{const}\ . \tag{2.342}$$

It follows from (2.342) that the particle moves over a plane with normal parallel to $\mathbf{L}$. (In general, motions in response to an arbitrary spherically symmetric field of force share the common property: *every trajectory is planar.*) The analysis of such problems is most conveniently performed in terms of polar coordinates r and φ. We introduce the radial and angular velocities $v_r = \dot{r} = dr/dt$ and $v_\varphi = r\dot{\varphi}$, so that

[9] A real prototype of this system is a hydrogen atom composed of a proton and an electron, because protons are 1836 times as heavy as electrons. Another case in point is a quarkonium, a meson composed of a light quark, say, u or d ($m_u \approx 5$ MeV, $m_d \approx 7$ MeV), and a heavy quark, say, c or b ($m_c \approx 1.2$ GeV, $m_b \approx 4.5$ GeV).

$$\mathbf{v}^2 = \dot{r}^2 + r^2\dot{\varphi}^2 \,. \tag{2.343}$$

One further integral of motion, the total energy of the particle in the Coulomb field, can be found if we take the scalar product of (2.340) with $\mathbf{v}$,

$$\mathbf{v} \cdot \left(\frac{d}{dt}\, m\gamma\mathbf{v} \right) = \frac{d}{dt}\, m\gamma = -Ze^2\, \frac{\mathbf{v}\cdot\mathbf{r}}{r^3} = Ze^2\, \frac{d}{dt}\left(\frac{1}{r}\right) \,. \tag{2.344}$$

This implies

$$\frac{d}{dt}\left(m\gamma - \frac{Ze^2}{r} \right) = 0 \,, \tag{2.345}$$

and

$$\varepsilon = m\gamma - \frac{Ze^2}{r} = \text{const} \,. \tag{2.346}$$

Of course, (2.346) can be obtained directly from the Lagrangian

$$L = -m\sqrt{1-\mathbf{v}^2} + \frac{Ze^2}{r} \,, \tag{2.347}$$

following the general line of Sect. 2.5.

One may define the radial and angular momenta conjugate to r and φ as $p_r = \partial L/\partial v_r$ and $p_\varphi = \partial L/\partial v_\varphi$. With (2.347) and (2.343), we have $p_r = m\gamma\dot{r}$ and

$$p_\varphi = m\gamma r^2\dot{\varphi} = \text{const} \,. \tag{2.348}$$

Note that p_φ is identical to $\mathbf{L}$. When expressed in terms of p_r and p_φ, the Hamiltonian, equal to the total energy defined in (2.346), is

$$H = \sqrt{m^2 + p_r^2 + \frac{p_\varphi^2}{r^2}} - \frac{Ze^2}{r} \,. \tag{2.349}$$

It is possible to extend our discussion to the case an arbitrary spherically symmetric attractive interaction with a negative potential energy $U(r)$ vanishing at infinity (in place of $-Ze^2/r$), and write

$$H = \sqrt{m^2 + p_r^2 + \frac{\mathbf{L}^2}{r^2}} + U(r) \tag{2.350}$$

(note that p_φ^2 is replaced by $\mathbf{L}^2$).

We put $p_r = 0$ in (2.350) to ‘switch off’ the dynamics. This gives the *effective potential*

$$\mathcal{U}(r) = \sqrt{m^2 + \frac{\mathbf{L}^2}{r^2}} + U(r) \tag{2.351}$$

which is a useful tool for analyzing the particle behavior near the origin.

There are three alternatives. First, the potential energy $U(r)$ is more singular at the origin than the centrifugal term $|\mathbf{L}|/r$. The effective potential $\mathcal{U}(r)$ in this case is shown in Fig. 2.3a. The particle can in principle orbit

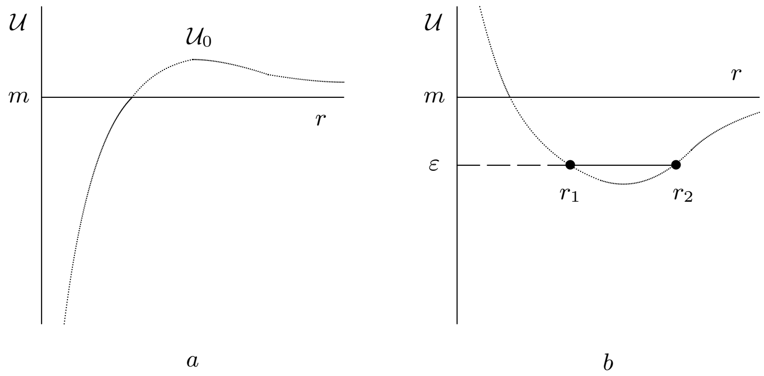


Fig. 2.3. Effective potential

in a circle, which corresponds to staying on the top of the potential hill $\mathcal{U}_0$ provided that $\varepsilon = \mathcal{U}_0$. However, such an orbit is unstable, and falling to the center or to infinity is highly probable. For both $\varepsilon > \mathcal{U}_0$ and $\varepsilon < \mathcal{U}_0$, falling one way or the other is unavoidable. One may envision it as sliding down along the curve $\mathcal{U}(r)$.

Second, $U(r)$ is less singular than $|\mathbf{L}|/r$, in particular, for $U(r) = -Ze^2/r$, this means that $Ze^2 < |\mathbf{L}|$. The shape of the curve $\mathcal{U}(r)$ for this case is depicted in Fig. 2.3*b*. If $\varepsilon < m$, then the particle executes a finite motion in the potential well. Radial displacements are allowed within a domain bounded by turning points r_A, $A = 1, 2$, which are the roots of the equation $\varepsilon = \mathcal{U}(r_A)$. (For the differential equation of the orbit see Problem 2.8.2.) If $\varepsilon > m$, then the particle executes an infinite motion, which may be interpreted as a scattering of the particle by the static center of attraction. Falling to the center is impossible, except when $\mathbf{L} = 0$ which correspond to a head-on collision.

Third, the singularities of $U(r)$ and $|\mathbf{L}|/r$ are identical. This alternative is explored in Problem 2.8.3. The particle travels in a stable orbit that passes through the center, but this does not bring the motion to a halt. Both finite and infinite motions are possible.

To summarize, falling to the center can be prevented if the attractive potential energy is less singular than the centrifugal term. Broadly speaking, the reason for keeping the particle from falling to the centre is that its kinetic energy dominates over the interaction energy of attraction.

It should be mentioned that the balancing of kinetic and interaction energies is quite subtle. To illustrate, falling to the center in the nonrelativistic limit is impossible if the potential energy $U(r)$ is less singular than $\sim 1/r^2$ (Problem 2.8.4). As noted above, the relativistic Kepler problem is based on the assumption that retardation is negligible. It is reasonable to expect that comparison between singularities of attractive and centrifugal terms is a mere provisional, approximate criterion. We will see later that this criterion fails when the effect of radiation is taken into account. Furthermore, in 1940 John Synge was able to apply a method of successive approximation to a two-particle bound system confined by the retarded Lorentz force, without regard to radiation, to show that energy disappears from the system. The orbiting particle slowly spirals in, but the rate at which this occurs is much less than that given by the formula for radiation from an accelerated charge.

Problem 2.8.1. Write the effective potential $\mathcal{U}(r)$ for a repulsive Coulomb interaction. Depict it schematically.

Answer Figure 2.4

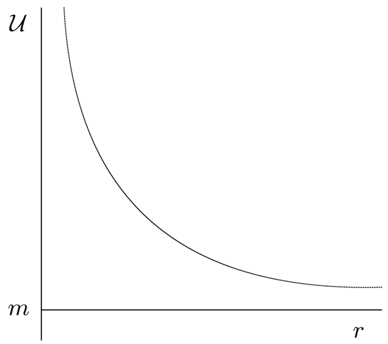


Fig. 2.4. Effective potential for a repulsive Coulomb interaction

Problem 2.8.2. Consider the case $\mathbf{L}^2 > Z^2e^4$ shown in Fig. 2.3*b*. Since the force is central, the time t can be eliminated altogether from the equation of motion, and we arrive at a differential equation of the path of the particle. With this in mind, we introduce $u = 1/r$, and consider q as a function of φ. Derive the differential equation of the orbit with u as the unknown function.

Answer

$$\frac{d^2u}{d\varphi^2} + \left(1 - \frac{Z^2e^4}{\mathbf{L}^2}\right)u = -\frac{Ze^2\varepsilon}{\mathbf{L}^2}. \tag{2.352}$$

Problem 2.8.3. Let $U(r) = -Ze^2/r$, and $Ze^2 = |\mathbf{L}|$. Depict the effective potential $\mathcal{U}(r)$ for this case.

Answer Figure 2.5.

Problem 2.8.4. Show that the nonrelativistic case corresponds to $m^2 \gg \mathbf{L}^2/r^2$, and find the nonrelativistic expression for the effective potential $\mathcal{U}(r)$ defined in (2.351). What is the criterion for the attractive potential $U(r)$ to ensure that decay to the center in the nonrelativistic Kepler problem is suppressed?

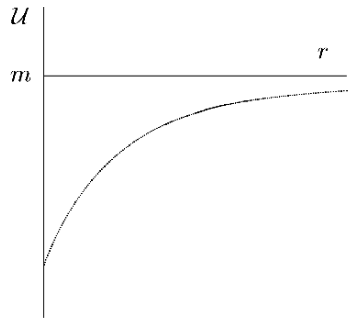


Fig. 2.5. Effective potential for the case $U(r) = -Ze^2/r$, $Ze^2 = |\mathbf{L}|$

Answer

$$\mathcal{U}(r) = m + \frac{\mathbf{L}^2}{2mr^2} - U(r) \,, \tag{2.353}$$

$U(r)$ must be less singular than the nonrelativistic centrifugal term $\mathbf{L}^2/2mr^2$.

Problem 2.8.5. While the mass of a proton and an electron is greater than that of a hydrogen atom, because energy must be supplied to break the electromagnetic bond in the atom, one may conceive a stable binary system such that its rest energy ε exceeds the sum of masses of its constituents. That is, the binding energy is positive. Consider a linearly rising interaction energy

$$U(r) = -\frac{\alpha_s}{r} + k\,r \quad (\alpha_s > 0,\, k > 0) \,, \tag{2.354}$$

and take $|\mathbf{L}| > \alpha_s$.

Depict the effective potential. What kind of motion is allowable? Show that $\mathcal{U}_{\min} > m$.

Answer Figure 2.6. A finite motion within the range from r_1 to r_2 is allowable.

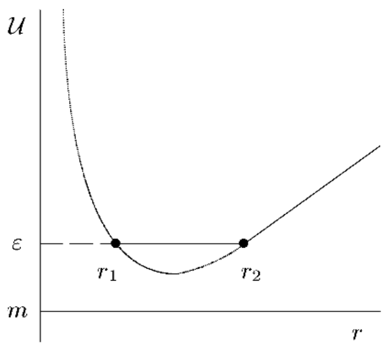


Fig. 2.6. Effective potential for a linearly rising potential energy

Problem 2.8.6. Let a 'relativistic hydrogen atom' be specified by the effective potential shown in Fig. 2.3*b*. For $\varepsilon < m$, the motion is finite, and ε may be regarded as the rest energy of this binary system minus the nuclear mass. Loosely speaking, ε represents the mass of the bound electron. When compared with a free electron, part of the mass is eaten by the Coulomb binding. The *binding energy* is defined as $\Delta m = \varepsilon - m$. The magnitude of this quantity is maximal if the electron orbits in a circle corresponding to the minimum of the effective potential $\mathcal{U}_{\rm min}$. Take $|\mathbf{L}|/Ze^2 = 1 + \epsilon$ where ϵ is a small positive number.

Evaluate $\Delta m = \mathcal{U}_{\rm min} - m$.

Answer

$$\Delta m \approx m\,(\sqrt{2\epsilon}/Ze^2 - 1)\,. \tag{2.355}$$

2.9 A Charged Particle Driven by a Magnetic Monopole

Consider a binary system composed of particle 1 with mass m_1 possessing electric charge e and particle 2 with mass m_2 possessing magnetic charge $e^\star$ (a magnetic monopole). As might be expected for reasons which are of little significance for the present discussion and hence omitted, magnetically charged particles are much heavier than their electrically charged counterparts. For simplicity, we assume that particle 2 is infinitely heavy and motionless. The two-particle problem can therefore be reduced to the problem of a single charged particle driven by a static magnetic monopole situated at the origin. We restrict our discussion to the nonrelativistic case which is sufficient to clarify essential features of this problem, in particular to provide insight into the difference between the behavior of $e - e$ and $e - e^\star$ binary systems.

We write the equation of motion for particle 1, omitting the label 1,

$$m\,\frac{d\mathbf{v}}{dt} = e\,\mathbf{v}\times\mathbf{B}\,. \tag{2.356}$$

As will be shown in Sect. 4.8, the magnetic field $\mathbf{B}$ due to a static magnetic monopole is

$$\mathbf{B} = e^\star\,\frac{\mathbf{n}}{r^2}\,, \tag{2.357}$$

where $\mathbf{n}$ is a unit vector directed along the radius vector $\mathbf{r}$. Combining (2.356) and (2.357),

$$m\,\frac{d\mathbf{v}}{dt} = e\,e^\star\,\frac{\mathbf{v}\times\mathbf{n}}{r^2}. \tag{2.358}$$

Take the scalar product of (2.358) with $\mathbf{v}$. The result is that the kinetic energy $\frac{1}{2}\,m\mathbf{v}^2$ is conserved, specifically, $|\mathbf{v}| = \text{const}$. However, the orbital angular momentum

$$\mathbf{L} = \mathbf{r}\times m\mathbf{v} \tag{2.359}$$

is not conserved as the force is not central (that is, not directed towards the origin). Indeed, take the cross product of (2.358) with $\mathbf{r}$,

$$\mathbf{r} \times \frac{d}{dt} m\mathbf{v} = \frac{d}{dt} (\mathbf{r} \times m\,\mathbf{v}) = \frac{e\,e^\star}{r^2} \mathbf{r} \times \mathbf{v} \times \mathbf{n} = \frac{e\,e^\star}{r} [\mathbf{v} - \mathbf{n}\,(\mathbf{v} \cdot \mathbf{n})] \ . \tag{2.360}$$

Observing that

$$\frac{dr}{dt} = \frac{d}{dt} \sqrt{\mathbf{r} \cdot \mathbf{r}} = \frac{\mathbf{r} \cdot \mathbf{v}}{\sqrt{\mathbf{r} \cdot \mathbf{r}}} = \mathbf{n} \cdot \mathbf{v} \ , \tag{2.361}$$

and

$$\frac{d\mathbf{n}}{dt} = \frac{d}{dt} \left(\frac{\mathbf{r}}{r} \right) = \frac{1}{r} [\mathbf{v} - \mathbf{n}\,(\mathbf{n} \cdot \mathbf{v})] \ , \tag{2.362}$$

rewrite (2.360), using (2.359), as

$$\frac{d}{dt} \mathbf{L} = e\,e^\star \frac{d}{dt} \mathbf{n} \ , \tag{2.363}$$

which shows that the vector

$$\mathbf{M} = \mathbf{L} - e\,e^\star\,\mathbf{n} \tag{2.364}$$

is conserved. We will see in Problem 5.2.10 that $\mathbf{M}$ is the total angular momentum of the system of two particles plus associated electromagnetic fields, with the field contribution being $e\,e^\star\,\mathbf{n}$, and particle contribution $\mathbf{L}$. Thus, account must be taken of the angular momentum residing in the field, which may well be comparable to $\mathbf{L}$ in magnitude.

The fact that $\mathbf{M}$ (rather than $\mathbf{L}$) is conserved, implies that the *trajectory is not planar.* Indeed,

$$\mathbf{M} \cdot \mathbf{n} = -e\,e^\star \ , \tag{2.365}$$

so that the radius vector $\mathbf{r}$ makes a constant angle ψ with $\mathbf{M}$,

$$\cos\psi = -\frac{e\,e^\star}{|\mathbf{M}|}. \tag{2.366}$$

In other words, the motion of a charged particle in the field of a static monopole is a spiral on the surface of a circular cone with semi-vertical angle $\cos^{-1}(e\,e^\star/\mathrm{M})$ and axis $-\mathbf{M}$ with its apex at the monopole, as viewed in Fig. 2.7. Because the trajectory goes to infinity, no bound states occur.

It follows from (2.364) that $\mathbf{M}^2 = \mathbf{L}^2 + (e\,e^\star)^2$. Therefore, $|\mathbf{L}|$ is a constant of motion.

One can show (Problem 2.9.1) that, if we consider the plane formed by unrolling the cone, the particle trajectory is a straight line,

$$\mathbf{r} \cdot \mathbf{b} = \text{const.} \tag{2.367}$$

Furthermore, the particle moves uniformly along this straight line,

$$r^2 = \mathbf{v}^2 t^2 + \mathbf{b}^2, \tag{2.368}$$

with $t = 0$ corresponding to the minimal separation between the particle and the center $b = |\mathbf{b}|$ (which is called the impact parameter). Note that b and $|\mathbf{v}|$ are specified arbitrarily.

Comparing (2.362) and (2.360), and taking into account (2.364) and (2.359), we find

$$\frac{d\mathbf{n}}{dt} = \frac{1}{r^2}\left(\mathbf{r} \times \mathbf{v} \times \mathbf{n}\right) = \frac{1}{mr^2}\left(\mathbf{M} \times \mathbf{n}\right) , \tag{2.369}$$

or

$$\frac{d\mathbf{n}}{dt} = \boldsymbol{\Omega} \times \mathbf{n}, \quad \boldsymbol{\Omega} = \frac{\mathbf{M}}{mr^2}. \tag{2.370}$$

This equation shows that $\mathbf{n}$ is rotating about the fixed axis $\mathbf{M}$ with angular velocity $\boldsymbol{\Omega}$.

If the Newtonian equation of motion (2.356) is replaced by

$$m\,\frac{d}{dt}\,\gamma\mathbf{v} = e\,\mathbf{v} \times \mathbf{B} , \tag{2.371}$$

this gives minor changes of final results. Indeed,

$$\mathbf{v} \cdot \left(\frac{d}{dt}\,\gamma\mathbf{v}\right) = \frac{d}{dt}\,\gamma = 0 . \tag{2.372}$$

This means that $\mathbf{v}^2 = \text{const}$, and (2.371) can be rewritten as

$$m\,\frac{d\mathbf{v}}{dt} = e\,\gamma^{-1}\,\mathbf{v} \times \mathbf{B} , \tag{2.373}$$

which differs from (2.356) by the presence of a constant factor γ^{-1} in the right hand side. Thus, the relativistic generalization of above formulas amounts to the change $e \to \gamma^{-1}e$.

Problem 2.9.1. Consider the plane formed by unrolling the cone of Fig. 2.7. Show that the locus of motion on this plane is a straight line expressed by

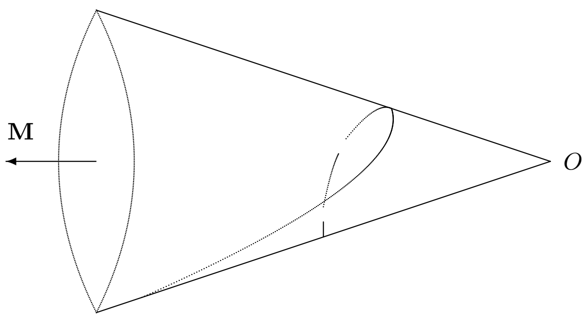


Fig. 2.7. The cone on which the trajectory lies

equation (2.367), and r varies with t according to (2.368). Find the relation between the angle ψ, defined in (2.366), and parameters b and $|\mathbf{v}|$.

Answer

$$\cos\psi = -\frac{e\,e^\star}{\sqrt{m^2b^2\mathbf{v}^2 + (e\,e^\star)^2}}\,. \tag{2.374}$$

Hint $\mathbf{r}$ and $\mathbf{v}$ belong to a plane tangent to the cone, hence both the angle between these vectors, α, and r are maintained on the plane formed by unrolling the cone. Note also that $|\mathbf{v}| = \text{const}$ and $|\mathbf{L}| = mr|\mathbf{v}|\sin\alpha = \text{const}$, which implies the equation of a straight line $r\sin\alpha = b$ on the unwrapped surface of the cone. Two legs b and $|\mathbf{v}|t$ and hypotenuse r span a right triangle. If $r = b$, then $\mathbf{v}$ is perpendicular to $\mathbf{r}$, and $|\mathbf{L}| = mb|\mathbf{v}|$, which gives the desired relation between ψ, b, and $|\mathbf{v}|$.

Problem 2.9.2. Find the relation between the maximal magnitude of angular velocity $\mathbf{\Omega}$, defined in (2.370), angle ψ, defined in (2.366), and parameters b and $|\mathbf{v}|$.

Answer

$$\mathbf{\Omega}^2_{\max} = \frac{\mathbf{v}^2}{b^2\sin^2\psi}\,. \tag{2.375}$$

2.10 Collisions and Decays

The relativistic two-particle problem is so extremely difficult in its precise setting because six mechanical degrees of freedom are supplemented with infinite degrees of freedom of the field mediating the interaction between the particles. In two preceding sections we developed approximations simplifying equations to the extent that it is possible to cope with the task. We now turn to a limiting case that particles interact *by contact*, exemplified by head-on collisions of point particles and disintegrations of particle aggregates into their constituents. In this case, the original action–reaction law retains its validity.

Before considering these processes, let us go back to the equation of motion

$$\dot{\mathbf{p}}_I = \mathbf{F}_I\,. \tag{2.376}$$

We take the index I to mean the label of a particle belonging to some N-particle system, the dot stands for the derivative with respect to time in a particular Lorentz frame. If we sum all these equations over I, we get

$$\sum \dot{\mathbf{p}}_I = \sum \mathbf{F}^\mu_I\,. \tag{2.377}$$

Suppose that the system is closed in that only the N particles exert forces on one another, and there is no external force. Let us imagine that all these N particles collide at some point of Minkowski space. Then, by the action–reaction law,

$$\sum \mathbf{F}_I^{\mu} = 0\ , \tag{2.378}$$

and (2.377) implies that contact interactions ensure conservation of the total momentum:

$$\sum \mathbf{p}_I = \text{const}\ . \tag{2.379}$$

A similar argument applies to show conservation of total energy.

Note that head-on collisions are common for particles living on a straight line, but highly improbable in higher dimensions. With the understanding, of course, that no central attractive force pulls the particles together, such events comprise a set of zero measure.

Assuming that the interaction between given particles is reasonably short-range, we may invoke only global four-momentum conservation for *incoming* and *outgoing* particles, considering them as free objects over all spacetime, except for regions of their 'quasi-local' interaction.

To be specific, consider a scattering event

$$1 + 2 \to 3 + 4\ , \tag{2.380}$$

by which is meant both an *elastic collision*, when the mass of each incoming particle equals the mass of a respective outgoing particle, and an *inelastic collision* rendering initial particles final particles of different species. Furthermore, putting the four-momentum of either incoming particle, say, particle 1, equal to zero, we cause it to drop out of the problem, and, in place of scattering, we have disintegration of a single particle into two pieces,

$$2 \to 3 + 4\ . \tag{2.381}$$

Let the four-momenta of incoming particles be p_1^{μ} and p_2^{μ}, and those of the outgoing particles p_3^{μ} and p_4^{μ}. Global four-momentum conservation reads

$$p_1 + p_2 = p_3 + p_4\ . \tag{2.382}$$

We ignore spin, if any, and regard these particles as Galilean. The mass M of each particle is identical to its rest mass m; both these quantities are indicated in this section by m. We thus have

$$p_I^2 = m_I^2, \quad I = 1, \ldots, 4\ . \tag{2.383}$$

It is conventional to refer to (2.383) as the *mass shell* constraints.

Six equations (2.382) and (2.383) are intended for finding eight unknown quantities p_3^{μ} and p_4^{μ}. The solution should therefore contain two free parameters. Note also that our description is frame-dependent. As will soon become clear, in the barycentric frame, one of the two free parameters consists of the scattering angle, and the other gives the orientation of the plane of scattering.

Besides the four squares of each four-momentum, equation (2.383), there are two independent scalar products. It is, however, convenient to use the three invariants

$$\begin{aligned}
s &= (p_1 + p_2)^2 = (p_3 + p_4)^2 \ , \\
t &= (p_1 - p_3)^2 = (p_2 - p_4)^2 \ , \\
u &= (p_1 - p_4)^2 = (p_2 - p_3)^2 \ .
\end{aligned} \tag{2.384}$$

With (2.382) and (2.383), the invariants s, t, u are linearly related through

$$s + t + u = m_1^2 + m_2^2 + m_3^2 + m_4^2 \ . \tag{2.385}$$

Indeed, applying (2.383) to (2.384), we obtain

$$(p_1 + p_2)^2 + (p_1 - p_3)^2 + (p_1 - p_4)^2 = 3m_1^2 + m_2^2 + m_3^2 + m_4^2 - 2p_1 \cdot (p_2 - p_3 - p_4) \ , \tag{2.386}$$

which, by (2.382), is

$$3m_1^2 + m_2^2 + m_3^2 + m_4^2 - 2p_1 \cdot (-p_1) = m_1^2 + m_2^2 + m_3^2 + m_4^2 \ . \tag{2.387}$$

The invariant s is the square of the total energy in the barycentric frame. Such a frame is defined (see Problem 2.5.3) by the condition that the total three-dimensional momentum is zero:

$$\mathbf{p}_1 + \mathbf{p}_2 = \mathbf{0}, \quad \mathbf{p}_3 + \mathbf{p}_4 = \mathbf{0} \ . \tag{2.388}$$

With this condition,

$$s = (\varepsilon_1 + \varepsilon_2)^2 = (\varepsilon_3 + \varepsilon_4)^2 \ . \tag{2.389}$$

The difference between four-momenta of a single incoming and outgoing particle is called its *four-momentum transfer*. If the incoming particle 1 is identical to the outgoing particle 3, the invariant t is the square of the four-momentum transfer. One may conceive, however, a process where particle 1 is identical to particle 4, the so-called exchange process. Then the invariant u is the square of the four-momentum transfer.

When particle 2 (the 'target particle') is at rest in some Lorentz frame (this frame is usually called the laboratory frame), we have

$$p_1^\mu = (\bar{\varepsilon}_1, \bar{\mathbf{p}}_1), \quad p_2^\mu = (m_2, \mathbf{0}), \quad p_3^\mu = (\bar{\varepsilon}_3, \bar{\mathbf{p}}_3), \quad p_4^\mu = (\bar{\varepsilon}_4, \bar{\mathbf{p}}_4) \ . \tag{2.390}$$

It follows from

$$s = (p_1 + p_2)^2 = m_1^2 + m_2^2 + 2\bar{\varepsilon}_1 m_2 \tag{2.391}$$

that

$$\bar{\varepsilon}_1 = \frac{1}{2m_2} \left(s - m_1^2 - m_2^2\right) , \tag{2.392}$$

and

$$\bar{\mathbf{p}}_1^2 = \bar{\varepsilon}_1^2 - m_1^2 = \frac{1}{4m_2^2} \left[(s - m_1^2 - m_2^2)^2 - 4m_1^2 m_2^2 \right] \ . \tag{2.393}$$

The latter formula can be rewritten as

$$|\bar{\mathbf{p}}_1| = \frac{\sqrt{\lambda(s, m_1^2, m_2^2)}}{2m_2} , \tag{2.394}$$

where λ is defined as

$$\lambda(x, y, z) = (x - y - z)^2 - 4yz . \tag{2.395}$$

Geometrically, this quadratic form is proportional to the area of the triangle with the legs $\sqrt{x}$, $\sqrt{y}$, and $\sqrt{z}$ (Problem 2.10.1).

The expression for t,

$$t = (p_2 - p_4)^2 = m_2^2 + m_4^2 - 2\bar{\varepsilon}_4 m_2 , \tag{2.396}$$

implies that

$$\bar{\varepsilon}_4 = \frac{1}{2m_2} \left(m_2^2 + m_4^2 - t\right) . \tag{2.397}$$

$\bar{\varepsilon}_3$ is found from the equation of energy conservation,

$$\bar{\varepsilon}_3 = \bar{\varepsilon}_1 + m_2 - \bar{\varepsilon}_4 = \frac{1}{2m_2} \left(s + t - m_1^2 - m_4^2\right) . \tag{2.398}$$

In view of (2.385), this can be represented as

$$\bar{\varepsilon}_3 = \frac{1}{2m_2} \left(m_2^2 + m_3^2 - u\right) . \tag{2.399}$$

By comparing (2.397) and (2.399) with (2.392), it is readily seen that the corresponding momenta $|\bar{\mathbf{p}}_4|$ and $|\bar{\mathbf{p}}_3|$ are

$$|\bar{\mathbf{p}}_4| = \frac{\sqrt{\lambda(t, m_2^2, m_4^2)}}{2m_2}, \quad |\bar{\mathbf{p}}_3| = \frac{\sqrt{\lambda(u, m_2^2, m_3^2)}}{2m_2} . \tag{2.400}$$

We call the angle $\bar{\vartheta}$ between $\bar{\mathbf{p}}_1$ and $\bar{\mathbf{p}}_3$ the scattering angle in the laboratory frame. It follows from

$$t = (p_1 - p_3)^2 = m_1^2 + m_3^2 - 2\left(\bar{\varepsilon}_1\bar{\varepsilon}_3 - |\bar{\mathbf{p}}_1|\,|\bar{\mathbf{p}}_3| \cos\bar{\vartheta}\right) \tag{2.401}$$

that

$$\cos\bar{\vartheta} = \frac{\left(s - m_1^2 - m_2^2\right)\left(m_2^2 + m_3^2 - u\right) + 2m_2^2\left(t - m_1^2 - m_3^2\right)}{\sqrt{\lambda(s, m_1^2, m_2^2)\,\lambda(u, m_2^2, m_3^2)}} . \tag{2.402}$$

In the barycentric frame, we call $\mathbf{p}$ and $\mathbf{q}$ the in and out *relative momenta*, and write

$$p_1^\mu = (\varepsilon_1, \mathbf{p}), \quad p_2^\mu = (\varepsilon_2, -\mathbf{p}), \quad p_3^\mu = (\varepsilon_3, \mathbf{q}), \quad p_4^\mu = (\varepsilon_4, -\mathbf{q}) . \tag{2.403}$$

These four-momentum decompositions are in agreement with (2.388). We have

$$\sqrt{s} = \varepsilon_1 + \varepsilon_2 \ , \tag{2.404}$$

or

$$\sqrt{s} - \sqrt{\mathbf{p}^2 + m_1^2} = \sqrt{\mathbf{p}^2 + m_2^2} \ , \tag{2.405}$$

which yields

$$\varepsilon_1 = \frac{1}{2\sqrt{s}}\,(s + m_1^2 - m_2^2), \quad \varepsilon_2 = \frac{1}{2\sqrt{s}}\,(s - m_1^2 + m_2^2) \ , \tag{2.406}$$

and

$$|\mathbf{p}| = \frac{1}{2}\sqrt{\frac{\lambda(s, m_1^2, m_2^2)}{s}} \ . \tag{2.407}$$

All the remaining pertinent quantities are readily evaluated (Problem 2.10.2),

$$\varepsilon_3 = \frac{1}{2\sqrt{s}}\,(s + m_3^2 - m_4^2), \quad \varepsilon_4 = \frac{1}{2\sqrt{s}}\,(s - m_3^2 + m_4^2) \ , \tag{2.408}$$

$$|\mathbf{q}| = \frac{1}{2}\sqrt{\frac{\lambda(s, m_3^2, m_4^2)}{s}} \ , \tag{2.409}$$

$$\cos\vartheta = \frac{s^2 + s\left(2t - m_1^2 - m_2^2 - m_3^2 - m_4^2\right) + \left(m_1^2 - m_2^2\right)\left(m_3^2 - m_4^2\right)}{\sqrt{\lambda(s, m_1^2, m_2^2)\,\lambda(s, m_2^2, m_3^2)}} \ . \tag{2.410}$$

As a simple illustration, we refer to an elastic collision of two particles of equal masses, $m_1 = m_2 = m_3 = m_4 = m$. In the barycentric frame,

$$s = 4\,(\mathbf{p}^2 + m^2), \quad t = -2\mathbf{p}^2\,(1 - \cos\vartheta), \quad u = -2\mathbf{p}^2\,(1 + \cos\vartheta) \ , \tag{2.411}$$

and so

$$\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \varepsilon_4 = \frac{\sqrt{s}}{2},\ \mathbf{p}^2 = \mathbf{q}^2 = \frac{s}{4} - m^2,\ \cos\vartheta = 1 + \frac{2t}{s - 4m^2} \ . \tag{2.412}$$

Another simple case is a decay of particle 2 into particles 3 and 4. For particle 1 to be eliminated, we put $p_1^\mu = 0$ in (2.384). Then

$$s = m_2^2, \quad t = m_3^2, \quad u = m_4^2 \ . \tag{2.413}$$

This result may be further used to evaluate the energies of the decay products,

$$\varepsilon_3 = \frac{m_2^2 + m_3^2 - m_4^2}{2m_2}, \quad \varepsilon_4 = \frac{m_2^2 - m_3^2 + m_4^2}{2m_2} \ . \tag{2.414}$$

The above formulae may seem to imply that every quantity specifying the processes (2.380) and (2.381) is uniquely determined. This impression is wrong. Take for example expressions (2.412). It is evident that the energies and momentum magnitudes of the final particles are indeed uniquely determined, but the same cannot be said of the scattering angle; notice that $\cos\vartheta$ is a function of the square of the four-momentum transfer t which takes arbitrary negative values. In addition, if we rotate $\mathbf{q}$ through any angle around the axis aligned with $\mathbf{p}$, then all the results given above remain unchanged. What this means is the scattering angle and orientation of the plane in which the scattering occurs are completely arbitrary.

The topic covered by this section is often loosely called *relativistic kinematics*. The invariants s, t, and u, defined in (2.384), are known as the *Mandelstam parameters*.

Problem 2.10.1. Show that $\lambda(x, y, z)$, defined in (2.395), can be written as

$$\lambda(x,y,z) = [x-(\sqrt{y}+\sqrt{z})^2][x-(\sqrt{y}-\sqrt{z})^2] = x^2+y^2+z^2-2xy-2yz-2xz$$
$$= (\sqrt{x}-\sqrt{y}-\sqrt{z})(\sqrt{x}+\sqrt{y}+\sqrt{z})(\sqrt{x}-\sqrt{y}+\sqrt{z})(\sqrt{x}+\sqrt{y}-\sqrt{z}) \; . \quad (2.415)$$

It is seen from the second equality that $\frac{1}{4}\sqrt{-\lambda(x,y,z)}$ represents the area of a triangle with the legs $\sqrt{x}$, $\sqrt{y}$, and $\sqrt{z}$; hence the name *triangular function* for λ. It is clear from the second line that $\lambda(x, y, z)$ is invariant under permutations of any two arguments. Show that

$$\begin{aligned} \lambda(x,y,y) &= x(x-4y) \; , \\ \lambda(x,y,0) &= (x-y)^2 \; . \end{aligned} \quad (2.416)$$

The first relation pertains to an elastic scattering of particles with identical masses, and the second one is related to an elastic scattering of a massless particle by a massive particle.

Problem 2.10.2. Verify (2.408)–(2.410).

Notes

1. The basic reference on Newtonian mechanics is Newton (1686). For a critical survey of Newtonian mechanics see Mach (1883). Key concepts of Newtonian and relativistic dynamics are discussed in Jammer (1954), (1961), (1962). Synge (1956) is an introduction to special relativity with a careful comparison of both geometric and dynamical concepts in Newtonian and relativistic frameworks. Relevant exercises are collected in Kotkin & Serbo (1971), and Lightman et al. (1975). The literature on general relativity (excluded from discussion in this book) is vast. The standard texts are Weyl (1918), Pauli (1958), Synge (1960), Landau & Lifshitz (1971), Misner et al. (1973), Weinberg (1972), Hawking & Ellis (1973), and Wald (1984).

2. *Section 2.1.* The relativistic form of Newton's second law, (2.3), was discovered by Poincaré (1906) and Planck (1906).

Poincaré (1900) suggested the identity of mass and energy of the electromagnetic field, and remarked that a charged oscillator which emits electromagnetic energy preponderantly in one direction should recoil as a gun does when it is fired. Einstein (1905*b*) derived the formula $\varepsilon = mc^2$ by kinematic arguments showing that the emission of electromagnetic radiation carrying energy ε reduces the mass of the emitting body by ε/c^2. He noted that the mass of a body is a measure of its energy content. For detail of this dramatic quest see Whittaker (1953).

Classical point particles may be classified as Galilean and non-Galilean. The present discussion entertains this classification, proposed by Kosyakov (2003). A Galilean particle is specified by the four-momentum p^μ defined in (2.9), and hence obeys the inertia law (2.11), known also as Newton's first law. A deep insight into this law owes much to the essay of Galilei (1632).

A general idea of faster-than-light particles can be had from Bilaniuk & Sudarshan (1969), for an extended discussion see Tiwari (2003).

Replacing the Newtonian differential equation of motion with a difference equation, similar to that in Problem 2.1.3, was proposed by Caldirola (1956). For the oscillatory regime of a free object see Schrödinger (1930). Such a regime, occurred initially as a visualization of solutions to the free Dirac equation and was long thought of as inherently quantum-mechanical. Its classical realization, a helical world line wound around a timelike axis, was found by Huang (1952).

3. *Section 2.2.* The idea of fields of force was advanced and developed by Michael Faraday (1839). We define the electromagnetic field through the Lorentz force law (2.43). This law was deduced by Heaviside (1889) and Lorentz (1892). The original Heaviside's paper is reprinted in the second volume of the collected articles by Heaviside (1892). A similar definition of the electromagnetic field is given in Barut (1964), and Misner et al. (1973). The virtue of this definition is that it can be extended to cover couplings with other classical gauge fields. In particular, the classical Yang–Mills field is defined through a force law of the Lorentz type. The Lie algebra structures, specifically a bracket of the form (2.58), were introduced and studied by Lie (1888, 1890, 1893). Wong (1970) obtained the force law (2.53) analyzing the classical limit of the field equations for the quantum SU(2) Yang–Mills fields.

4. *Section 2.3.* All configurations of the electromagnetic field are separated according to their algebraic properties into three classes. The class A incorporates both pure electric and pure magnetic fields. Configurations with the electric and magnetic fields of the same strength perpendicular to each other belong to the class B. As for the class C, there exists a Lorentz frame where the electric and magnetic fields are parallel to each other. The fact that $\mathcal{S}$ and $\mathcal{P}$ are Lorentz invariant was established by Poincaré (1906).

5. *Section 2.4.* Taub (1948) analyzed a general solution to the equation of motion for a charged particle in an arbitrary constant electromagnetic field F of the form

$$v(s) = \mathit{\Lambda}\, v(0) = \exp\left(\frac{e}{m}\, Fs\right) v(0)\,. \tag{2.417}$$

This result shows that the antisymmetric tensor F, describing a constant field, determines a family of Lorentz matrices, Λ, of which it is an infinitesimal generator. Taub gave a complete classification of such Lorentz transformations in terms of F as well as closed expressions for Λ. The method can be extended to the problem of a charged particle in the field of a plane wave. Salingaros (1985) employed Clifford algebraic techniques to calculate the four-velocity of a charged particle in constant electric and magnetic fields, and presented a comparison of results obtained in his and other works. The Hestenes book (1986) is an introduction to this algebraic technique.

6. *Section 2.5.* Euler (1744), Lagrange (1788), Hamilton (1834), Noether (1918), Poincaré (1906), Planck (1906), and Schwarzschild (1903) are the origins of the material presented in this section. The literature on the calculus of variations and analytical mechanics is rich and diverse. A general reference is Whittaker (1904), and Goldstein (1950). The books by Abraham & Marsden (1967), Godbillon (1969), Sudarshan & Mukunda (1973), Arnold (1978), and Marsden & Ratiu (1994) emphasize advanced mathematical topics of Lagrangian and Hamiltonian descriptions of Newtonian mechanics. Young (1969) delves into subtle aspects of the calculus of variations. Lanczos (1949) is a less formal introduction to the Lagrangian and Hamiltonian formalisms for Newtonian mechanics. Noether (1918) presented the original proof of Noether's first theorem. This theorem is described in detail by Hill (1951). For the converse of Noether's first theorem see Palmieri & Vitale (1970). The incorporation of a damped oscillator in the Lagrangian formalism was discussed by Denman (1966). In this connection, also see Sect. 3.2 of Morse & Feshbach (1953), and Van der Vaart (1967).

7. *Section 2.6.* For more on reparametrization invariance from the mathematical point of view see Young (1969), Chap. 6. Noether (1918) presented the original formulation and proof of Noether's second theorem. The reparametrization invariant action (2.266) was proposed by Brink et al. (1976).

8. *Section 2.7.* A classical theory of a point particle with intrinsic angular momentum was developed by Frenkel (1926). The present approach is loosely patterned on that in Rafanelli (1984). For a more extended discussion see Corben (1968), Hanson & Regge (1974), and Rivas (2001). Martin (1959) pioneered the use of Grassmannian variables for description of spin degrees of freedom. This model was rediscovered by Berezin & Marinov (1975), (1977), and Casalbuoni (1976*a*), (1976*b*), and refined by Galvao & Teitelboim (1980). An alternative action for a spinning particle, yet devoid of reparametrization invariance, was suggested by Barut & Zanghi (1984).

9. *Section 2.8.* The general role of the action–reaction law, as viewed by Poincaré (1900) and Planck (1908), is to afford momentum conservation. Synge (1940) attempted a direct attack on the relativistic two-particle problem, based on the assumptions (i) that particle 2 maintains particle 1 in orbit by the retarded Lorentz force, (ii) that the ratio of the masses of the two

particles is small $m_1/m_2 \ll 1$, and (iii) that the effect of radiation may be neglected. With a method of successive approximation, it was shown that radius of the orbit shrinks at a rate very much less than m_1/m_2. Additional texts of particular interest in relation to the integration of the relativistic Kepler problem are Synge (1956), and Lanczos (1949). For a discussion of the *three-body* problem within the scope of Newtonian mechanics see Birkhoff (1927), Abraham & Marsden (1967), Szebehely (1967), and Marshal (1990).

10. *Section 2.9.* The motion of a charged particle in the field of a static magnetic monopole was examined by Poincaré (1896), who showed that the typical trajectory of the particle is a geodesic on the surface of a circular cone whose apex is at the monopole, and derived the constant of motion (2.364). Dirac (1931), (1948) showed that a quantum-mechanical description of this system requires that the electric charge-coupling e and the magnetic charge-coupling $e^\star$ be related by the equation $e\,e^\star = \frac{1}{2}n$ where n is an integer. Thus, the existence of just one magnetically charged particle would provide an explanation for the quantization of electric charge. For an extended discussion of scattering of two classical dyons with arbitrary electric and magnetic charge-couplings $(e_1, e_1^\star)$ and $(e_2, e_2^\star)$ see Barut & Beker (1974).

11. *Section 2.10.* Many examples of scatterings and decays applicable to atomic and elementary particle physics are discussed at length in Synge (1956), Källen (1964), and Lanczos (1949). The convenience of invariant variables s, t, u in high energy physics was pointed out by Mandelstam (1958). The use of these variables is detailed in many textbooks, for example, in Byckling & Kajantie (1973).

3

Electromagnetic Field

In the preceding chapter we defined the electromagnetic field as a physical object that manifests itself through its influence on a particle by the four-force linear in the particle four-velocity. To be more specific, one recognizes the presence of electromagnetic field when particles experience the Lorentz force. It transpired that the *state* of electromagnetic field at each spacetime point is characterized by an antisymmetric tensor $F_{\mu\nu}$. In a particular frame of reference, this is equivalent to assigning the electric field intensity $\mathbf{E}$ and the magnetic induction $\mathbf{B}$ to each point.

In what follows, the tensor $F_{\mu\nu}$, or the associated vectors $\mathbf{E}$ and $\mathbf{B}$, will be called the electromagnetic field. Although this usage is common in theoretical physics, one should realize that the object is being substituted for the set of its states.

Our next task is to discuss the law governing the electromagnetic field behavior in space and time. This law is given by a system of partial differential equations known as *Maxwell's equations*.

We omit the history of the development of Maxwell's equations, and their current experimental status, considering these topics to be well known to the reader. Attention is centred on conceptual aspects of the subject. Our concern at first is to understand *to what extent the form of the dynamical law is ordered by geometrical features of our world*, in particular by the fact that *space has three dimensions*. The complete reconstruction of Maxwell's equations requires the adoption of four additional assumptions of non-geometric origin, which can be succinctly phrased. It would be tempting to think of them as *the principles that cover the whole physical content of Maxwell's equations*. One might be satisfied with such an understanding of Maxwell's equations, and this attitude would be quite robust. However, later on we will see, from closer inspection of symmetries peculiar to electrodynamics, that such principles may have much to do with geometry.

3.1 Geometric Contents of Maxwell's Equations

Write the evolutionary law for the electromagnetic field in the following symbolic form:

$$L\,(F) = \Im \tag{3.1}$$

where L is a differential operator that describes *local variations of the field state*, and $\Im$ is interpreted as the *source* of these variations.

Why is the differential operator favored over integral ones? Are algebraic constructions with shifted (retarded or advanced) arguments pertinent? By now difference equations play an important role in theoretical physics. To illustrate, lattice field equations, with the spacetime continuum replaced by lattices of discrete vertices, have become very popular in the past two decades. An overview of the lattice formulation of gauge theories will be given in Sect. 7.3, and we will see that the lattice dynamics is depicted by difference equations. On the other hand, we already learned in Problem 2.1.3 that a difference equation can be represented as a series of differential operators of increasing orders. Therefore, differential equations of finite orders are a special case of difference equations.

The choice of L as a differential operator relates to the fundamental theoretical idea of *local action*, by which dynamical variations of fields propagate in space from one point to all nearest with a finite velocity. Partial differential equations of the hyperbolic type are currently best suited for the mathematical expression of this idea. An alternative concept, the so-called *action at a distance*, will be outlined in Sect. 10.6.

We assume that only *first derivatives* of $F_{\mu\nu}$ appear in (3.1). This assumption may seem to contradict the situation in mechanics, where Newton's second law is given by a differential equation of second order with the particle position $\mathbf{z}$ as the unknown function. In fact, this is only an apparent contradiction because the instantaneous state of a particle is specified by the pair of variables $(\mathbf{z}, \mathbf{v})$ or, equivalently, by the pair (q_a, p_a), and the evolution of this system is given by Hamilton equations containing only first derivatives of q_a and p_a. The variables $F_{\mu\nu}$ take into complete account the state of the electromagnetic field, and they should therefore be likened to the phase space coordinates (q_a, p_a), not the configuration space coordinates q_a.

We now fix some inertial frame of reference and consider the spatial behavior of the vector functions $\mathbf{E}$ and $\mathbf{B}$. Any smooth vector function $\mathbf{V}$ can be reconstructed with the knowledge of 9 components of its gradients $\partial_j V_i$. However, to do this requires actually much less information. The tensor $\partial_j V_i$ can be written as the sum of symmetric and antisymmetric terms. In addition, a term proportional to the trace can be separated, rendering the symmetric term traceless,

$$\partial_j V_i = \frac{1}{2}\left(\partial_j V_i + \partial_i V_j - \frac{2}{3}\,\delta_{ij}\partial_k V_k\right) + \frac{1}{2}\,(\partial_j V_i - \partial_i V_j) + \frac{1}{3}\,\delta_{ij}\partial_k V_k\;, \tag{3.2}$$

where the summation over repeated indices is understood.

A remarkable feature of three-dimensional Euclidean space is that the reconstruction of $\mathbf{V}(\mathbf{x})$ requires only the knowledge of the antisymmetric term $\partial_j V_i - \partial_i V_j$, which is dual to $\nabla \times \mathbf{V}$, namely $\partial_i V_j - \partial_j V_i = \epsilon_{ijk}\, \epsilon_{klm} \partial_l V_m$, and the scalar $\partial_k V_k$, which is $\nabla \cdot \mathbf{V}$, while information on 5 components of the symmetric traceless combination $\partial_j V_i + \partial_i V_j - \frac{2}{3}\, \delta_{ij} \partial_l V_l$ is unnecessary. This statement is known as the *Helmholtz theorem*: if a smooth vector function $\mathbf{V}$ disappears at infinity, it can be reconstructed from its curl, $\mathbf{C} = \nabla \times \mathbf{V}$, and divergence, $D = \nabla \cdot \mathbf{V}$.

Indeed, the relation

$$\nabla \times (\nabla \times \mathbf{V}) = \nabla\, (\nabla \cdot \mathbf{V}) - \nabla^2\, \mathbf{V}\,, \tag{3.3}$$

familiar from any course of the vector analysis, can be rewritten as the *Poisson equation*

$$\nabla^2\, \mathbf{V} = \mathbf{S} \tag{3.4}$$

with a computable function in the right side $\mathbf{S} = \nabla\, D - \nabla \times \mathbf{C}$. In Sect. 4.1 we will see that this equation has a unique solution. This will complete the proof of the Helmholtz theorem.

An important implication of this result is that the field equation (3.1) can be expressed in terms of curls and divergences of $\mathbf{E}$ and $\mathbf{B}$. Therefore, we do not need information on all components of the spacetime derivatives $\partial_\lambda F_{\mu\nu}$; only linear combinations of components containing curls and divergences of $\mathbf{E}$ and $\mathbf{B}$ matter. We recall that $\mathbf{E}$ and $\mathbf{B}$ are related to $F_{\mu\nu}$ as

$$E_i = F_{0i} = F^{i0}\,, \tag{3.5}$$

$$F_{ij} = F^{ij} = -\epsilon_{ijk} B_k\,, \tag{3.6}$$

$$B_k = -\frac{1}{2}\, \epsilon_{klm} F^{lm}\,, \tag{3.7}$$

and that the usual rule of raising and lowering indices holds for tensors in Minkowski space. We note also that E_i, B_i, ϵ_{ijk}, and δ_{ij} occur in a particular Lorentz frame (and, by convention, the repeated latin indices in $\epsilon_{ijk} B_k$ are to be summed over $k = 1, 2, 3$).

By (3.5),

$$\mathrm{div}\, \mathbf{E} = \partial_j E_j = \partial_j F^{j0}\,. \tag{3.8}$$

Using (3.7), we find

$$(\mathrm{curl}\, \mathbf{B})_i = \epsilon_{ijk} \partial_j B_k = -\frac{1}{2}\, \epsilon_{ijk} \partial_j \epsilon_{klm} F^{lm} = \frac{1}{2} (\delta_{im}\delta_{jl} - \delta_{il}\delta_{jm}) \partial_j F^{lm} = \partial_j F^{ji}\,. \tag{3.9}$$

To express $\mathrm{div}\, \mathbf{B}$ and $\mathrm{curl}\, \mathbf{E}$ via linear combinations of $\partial_\lambda F_{\mu\nu}$, we recall the definition

$${}^*F^{\mu\nu} = \frac{1}{2}\, \epsilon^{\mu\nu\alpha\beta} F_{\alpha\beta} \tag{3.10}$$

and the convention that the three- and four-dimensional Levi-Civita symbols are related as

$$\epsilon_{ijk} = \epsilon^{0ijk} \; . \tag{3.11}$$

From (3.10), (3.11), and (3.6) it follows that

$$^*F^{i0} = \frac{1}{2}\,\epsilon^{i0\alpha\beta}F_{\alpha\beta} = -\frac{1}{2}\,\epsilon^{0ijk}F_{jk} = -\frac{1}{2}\,\epsilon_{ijk}F_{jk} = \frac{1}{2}\,\epsilon_{ijk}\epsilon_{jkl}B_l = \delta_{il}B_l = B_i \; , \tag{3.12}$$

while (3.10), (3.11), and (3.5) give

$$^*F^{ji} = \frac{1}{2}\,\epsilon^{ji\alpha\beta}\,F_{\alpha\beta} = \frac{1}{2}\,\epsilon^{ji0k}\,F_{0k} + \frac{1}{2}\,\epsilon^{jik0}\,F_{k0} = \epsilon^{0jik}\,F_{0k} = -\epsilon_{ijk}\,E_k \; . \tag{3.13}$$

Therefore,

$$\operatorname{div}\mathbf{B} = \partial_j B_j = \partial_j{}^*F^{j0} \; , \tag{3.14}$$

and

$$(\operatorname{curl}\mathbf{E})_i = \epsilon_{ijk}\partial_j E_k = -\partial_j{}^*F^{ji} \; . \tag{3.15}$$

Note that $F_{\mu\nu}$ and $\epsilon^{\alpha\beta\gamma\delta}$ are completely antisymmetric tensors defined in Minkowski space. If some index (say, α) of such a tensor is fixed to be temporal ($\alpha = 0$), then the tensor vanishes unless all the other indices are spatial ($\beta = j, \gamma = k, \delta = l$). Only one index may be 0.

Thus the desired linear combinations of derivatives are $\partial_\mu F^{\mu\nu}$ and $\partial_\mu{}^*F^{\mu\nu}$. Indeed, taking into account (3.8) and (3.9), we write

$$\begin{aligned}\partial_\mu F^{\mu\nu} &= (\partial_\mu F^{\mu 0}, \partial_\mu F^{\mu i}) = (\partial_k F^{k0}, \partial_0 F^{0i} + \partial_k F^{ki}) \\ &= (\partial_k F^{k0}, -\partial_0 F^{i0} - \partial_k \epsilon_{kij} B_j) = (\partial_k E_k, -\partial_0 E_i + \epsilon_{ikj}\partial_k B_j) \; ,\end{aligned} \tag{3.16}$$

and so

$$\partial_\mu F^{\mu\nu} = \left(\operatorname{div}\mathbf{E}, -\frac{\partial \mathbf{E}}{\partial t} + \operatorname{curl}\mathbf{B}\right) \; . \tag{3.17}$$

Likewise, using (3.14) and (3.15), we find

$$\partial_\mu{}^*F^{\mu\nu} = (\partial_j{}^*F^{j0}, \partial_0{}^*F^{0i} + \partial_j{}^*F^{ji}) = (\partial_j B_j, -\partial_0 B_i - \epsilon_{ijk}\partial_j E_k) \; , \tag{3.18}$$

and so

$$\partial_\mu{}^*F^{\mu\nu} = \left(\operatorname{div}\mathbf{B},\; -\frac{\partial \mathbf{B}}{\partial t} - \operatorname{curl}\mathbf{E}\right) \tag{3.19}$$

Finally, the symbolic field equation (3.1) can be made concrete:

$$\partial_\lambda F^{\lambda\mu} = 4\pi j^\mu \; , \tag{3.20}$$

$$\partial_\lambda{}^*F^{\lambda\mu} = 4\pi m^\mu \; , \tag{3.21}$$

where the pair of the four-vectors (j^μ, m^μ) represents $\Im$ and conceivably also part of $L(F)$. The factor 4π, fixing the so-called *Gaussian units*, will manifest its convenience later. The construction of the four-vectors j^μ and m^μ remains at this stage indeterminate.

Problem 3.1.1. Derive (3.3) using tensor calculus.

Problem 3.1.2. Show that

$$-\epsilon_{\lambda\mu\nu\sigma}\partial_\rho{}^*F^{\rho\sigma} = 2(\partial_\lambda F_{\mu\nu} + \partial_\nu F_{\lambda\mu} + \partial_\mu F_{\nu\lambda}) \; . \tag{3.22}$$

3.2 Physical Contents of Maxwell's Equations

We came to equations (3.20) and (3.21) from essentially geometric considerations. It remains to clarify what are j^μ and m^μ. To do this requires three additional assumptions which lead directly and unambiguously to Maxwell's equations. Their adoption might be motivated by reference to physical experiment where solutions to Maxwell's equations have been verified to a high degree of precision. This gives the impression that we are dealing with the net physical contents of Maxwell's equations, that is, residual information which cannot be interpreted in geometric terms.

The *first* assumption is that *the field equation* (3.1) *is linear* (an alternative, the Born and Infeld nonlinear version of electrodynamics, is reserved for Chap. 10.) For this assumption not to seem excessively technical, it can be reformulated as the so-called *superposition principle* (well established experimentally). This principle states: if sources $\Im_1$ and $\Im_2$ generate fields F_1 and F_2, respectively, then source $a\Im_1 + b\Im_2$ generates field $aF_1 + bF_2$. It follows that

$$L(aF_1 + bF_2) = aL(F_1) + bL(F_2) \tag{3.23}$$

which means that $L(F)$ is a linear operator.

Let us look more closely at the structure of equation (3.20). Linear combinations of the derivatives $\partial_\lambda F_{\mu\nu}$ are already taken into account. Therefore, only terms proportional to $g^\mu F_{\mu\nu}$ where g^μ stands for either the coordinate of Minkowski space x^μ or a fixed vector n^μ or some kinematical variable of Ith particle, say, the four-velocity at a certain point on the world line $v_I^\mu(s_I)$, are permitted. However, if it is granted that the system 'particles plus electromagnetic field' is *closed*, coefficients of all the dynamical equations, those of (3.20) and (3.21) included, must be independent of x^μ. Given some coefficient time-dependent, $C = C(t)$, this would evidence that the system suffers from an external influence which varies according to the law $C(t)$, that is, the system is not closed. The option $g^\mu = n^\mu$ is in conflict with the spacetime isotropy rendering the description not explicitly covariant under rotations or Lorentz boosts. The option $g^\mu = v_I^\mu(s_I)$ is inadmissible because the instant s_I is selected in contradiction with the time homogeneity.

Thus j^μ is independent of $F^{\mu\nu}$; it may depend only on particle characteristics, such as coupling constants e_I and world line variables $z_I^\mu(s_I)$. What are those dependences?

In order to clarify them, we digress for a while and observe the identity

$$\partial_\mu \partial_\nu F^{\mu\nu} = 0\,, \tag{3.24}$$

which is due to the antisymmetry of the tensor $F^{\mu\nu}$. Therefore, to ensure the consistency of (3.20), the relation

$$\partial_\mu j^\mu = 0 \tag{3.25}$$

must hold identically, namely, for any value of e_I and any function z_I^μ.

Consider the integral

$$\int_{\mathcal{U}} d^4x\, \partial_\mu j^\mu \tag{3.26}$$

taken over a domain of Minkowski space $\mathcal{U}$ bounded by a timelike tube T_R of large radius R and two spacelike hypersurfaces Σ_1 and Σ_1 with both normals directed towards the future, Fig. 3.1. By the Gauss–Ostrogradskiĭ theorem,

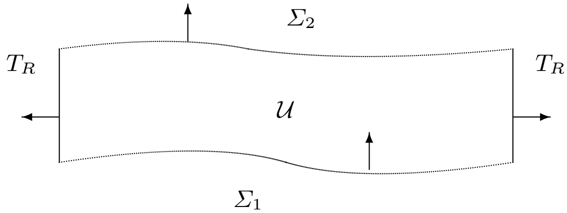


Fig. 3.1. The integration domain $\mathcal{U}$

(3.26) can be transformed to the integral over the boundary:

$$\int_{\mathcal{U}} d^4x\, \partial_\mu j^\mu = \left(\int_{T_R} + \int_{\Sigma_2} - \int_{\Sigma_1} \right) d\sigma_\mu\, j^\mu\ . \tag{3.27}$$

The minus sign of the last integral is due to the fact that the normal to the hypersurface Σ_1 is directed inward the domain $\mathcal{U}$.

Assuming that j^μ vanishes sufficiently rapidly in spacelike directions as $x^2 \to -\infty$, the integral over T_R goes to zero as $R \to \infty$. In view of (3.25), we then obtain

$$\int_{\Sigma_1} d\sigma_\mu\, j^\mu = \int_{\Sigma_2} d\sigma_\mu\, j^\mu\ . \tag{3.28}$$

Because Σ_1 and Σ_1 are arbitrary spacelike infinite hypersurfaces,

$$Q = \int_{\Sigma} d\sigma_\mu\, j^\mu = \text{const}\ . \tag{3.29}$$

Q is called the *total charge-source*. This quantity is independent of Σ. In particular, if the hypersurface of a fixed form Σ is shifted in timelike directions, Q remains invariant. Equation (3.29) is known as the *conservation of the total charge-source* in time.

The constancy of the charge-source Q would be tempting to relate to the constancy of the charge-coupling e, implied by equation (2.37). How can we do it? Let the hypersurface Σ be intersected by N world lines of charged particles. Our *second* assumption is that the *total charge-source is the sum of charge-couplings of those particles*,

$$Q = \sum_{I=1}^{N} e_I \; . \tag{3.30}$$

Imagine for a while that only a single point particle with the coupling e is in the universe, then

$$Q = e \; . \tag{3.31}$$

This equation, implying the identity of the charge-source and the charge-coupling, may be interpreted in the spirit of the extended *action–reaction* principle. Indeed, the charge-coupling measures the variation of the particle state for a given electromagnetic field state while the charge-source measures the variation of the electromagnetic field state for a given particle state. Therefore, both quantities would be reasonable to lump together as the *electric charge* or briefly the *charge*.

How could we realize (3.30) technically? The gist of the question is that the set of N *one-dimensional* objects, the world lines $z_I^\mu(s_I)$, should be *continuously* mapped onto a *four-dimensional* object, the field $j^\mu(x)$ distributed over the whole Minkowski space. This can be visualized as a bunch of N curves that should be 'smeared out' to turn into the flux of a continuous fluid. This mapping may seem to conflict with topological concepts of the dimension, and such was the widely accepted belief until the 1930s. The break-through was due to Paul Dirac who dared to identify the delta-function with the density of zero-dimensional objects. Sergeĭ Sobolev and Laurent Schwartz laid the foundation of the theory of distributions rendering the delta-function and its derivatives respectable mathematical notions. The delta-function made the formal status of the discrete equivalent to that of the continuous.

We now show that the desired mapping is

$$j^\mu(x) = \sum_{I=1}^{N} e_I \int_{-\infty}^{\infty} ds_I \, v_I^\mu(s_I) \, \delta^4\left[x - z_I(s_I)\right] \; , \tag{3.32}$$

where $v_I^\mu(s_I)$ is the four-velocity of Ith particle at proper time s_I, and $\delta^4\,(x)$ is the four-dimensional delta-function (for an overview of distribution theory see Appendix F).

Note that changes of the parametrization of the world lines, $s_I = s_I(\tau_I)$, leaves integrals unchanged. Indeed,

$$ds\, v^\mu = ds\, \frac{dz^\mu}{ds} = d\tau\, \frac{dz^\mu}{d\tau} \; . \tag{3.33}$$

In lieu of proper times s_I, we may use a laboratory time $x^0 = t$, to give

$$j^\mu = (\varrho, \mathbf{j}) \tag{3.34}$$

where

$$\varrho(\mathbf{x}, t) = \sum_{I=1}^{N} e_I \, \delta^3\left[\mathbf{x} - \mathbf{z}_I(t)\right] \; , \tag{3.35}$$

$$\mathbf{j}(\mathbf{x},t)=\sum_{I=1}^{N} e_I\,\mathbf{v}_I(t)\,\delta^3\big[\mathbf{x}-\mathbf{z}_I(t)\big]\,. \tag{3.36}$$

Here, $\mathbf{v}_I(t)$ is the three-velocity, and $\mathbf{z}_I(t)$ the three-position of Ith particle at the instant t.

Because the hypersurface Σ in (3.29) is arbitrary, we take Σ such that all the world lines are perpendicular to it at intersection points. For a small vicinity of the intersection point, we have $ds_I\, d\sigma_\mu v_I^\mu = d^4x$ where x^μ is a Cartesian coordinate in the Lorentz frame with the time axis directed along v_I^μ. Inserting (3.32) in (3.29), we arrive at (3.30).

j^μ is called the *four-current density of electric charges* or simply the *four-current.* Since the integral relation (3.29), expressing charge conservation, is shown to be equivalent to the differential equation (3.25), j^μ is said to obey the *local conservation law* (3.25).

ϱ and $\mathbf{j}$ are called respectively the *charge density* and the *charge current density.* Using the component decomposition (3.34), we rewrite (3.25) as

$$\frac{\partial\varrho}{\partial t}+\operatorname{div}\mathbf{j}=0 \tag{3.37}$$

which is known as the *equation of continuity.* The name originates from hydrodynamics and suggests the idea of a fluid with the charge density ϱ and the charge current density $\mathbf{j}=\mathbf{v}\varrho$ where $\mathbf{v}$ is the local velocity of the charged matter.

We define the charge contained in a three-dimensional domain V as

$$Q_V=\int_V d^3x\,\varrho\,. \tag{3.38}$$

When integrated over the volume of V, (3.37) becomes

$$\frac{d}{dt}\,Q_V=-\oint_{\partial V} d\mathbf{S}\cdot\mathbf{j} \tag{3.39}$$

where the Gauss–Ostrogradskiĭ theorem

$$\int_V d^3x\,\operatorname{div}\mathbf{j}=\oint_{\partial V} d\mathbf{S}\cdot\mathbf{j} \tag{3.40}$$

has been applied to derive the right hand side of (3.39). Equation (3.39) says: the increase of the charge in the domain V in a unit time is due to all contributions of elementary fluxes $d\mathbf{S}\cdot\mathbf{v}\varrho$ of positive charge flowing inward this domain through the boundary ∂V. The minus sign reflects the convention that the normal vector of the boundary surface is always outward the enclosed domain.

Note, however, that the idea of *continuous* charged matter with a reasonably steady charge distribution is inconsistent in classical electrodynamics.

We will see in Sect. 4.1 that static charged systems are unstable; any external disturbance breaks the equilibrium. A steady lump of a continuous charged fluid is unfeasible: each part of it exerts a repulsive force on every other part (because of the charge carried by the fluid), which cause the lump to become a rarefied medium. A homogeneous mixture of two oppositely charged fluids is also unstable; part of the mixture would collapse forming a neutral cluster, while the remainder possessing an uncompensated residual charge would spread.

In 1906 Poincaré conjectured that a stable existence of charged matter is ensured by the presence of *nonelectromagnetic cohesive forces.* A dramatic implication of this conjecture is that electrodynamics is *fundamentally unclosed*, that is, electromagnetic phenomena defy explanation separately from the hypothetical Poincaré forces. Although the cohesive forces should manifest themselves even on macroscopic level, the form and origin of those forces remained enigmatic for two decades.

Suddenly, in 1925, Frenkel made a striking inference that electromagnetism may be accounted for *by itself.* He assumed the electron to be a *point* in the precise geometric sense. A point particle can be envisioned as a sphere of radius r in the limit $r \to 0$. Electrostatic repulsive forces are put to *distinct points* of the sphere, and, therefore, each part of the sphere tends to move away from other parts. However, all the repulsive forces are brought into a *single point* and cancel as $r \to 0$. Therefore, a point charged particle is immune from the explosion tendency, and the stable existence of such objects has no need of the cohesive force conjecture.

Frenkel's idea was of paramount importance for the ensuing development of field theory and particle physics. Dirac was delighted with this idea, and soon afterwards, he contrived its adequate mathematical formulation through the delta-function. We will see in Chap. 6 that the model of a structureless point electron poses a problem of infinite self-energy, but such is the price for the conceptual advantage of stability. The idea of a point source was a useful guide in quantum field theory and came up with the present paradigm of *local interactions* of quantized fields.

From here on, we will keep in mind primarily a system of N particles which form the four-current (3.32). Note in passing that point particles never experience decay triggered by the electromagnetic repulsion, and, therefore, their *world lines cannot bifurcate.*

Substituting (3.35) and (3.36) in (3.39), we find

$$\Delta Q_V = -\Delta t \sum_{I=1}^{N} e_I \oint_{\partial V} d\mathbf{S} \cdot \mathbf{v}_I \,. \tag{3.41}$$

The mechanism of charge conservation is then quite simple: the charge Q_V in a domain V is constant during a time interval Δt if the total charge of particles penetrating into this domain equals the total charge of particles leaving this domain. Figure 3.2 illustrates this statement. We see one particle (left) with

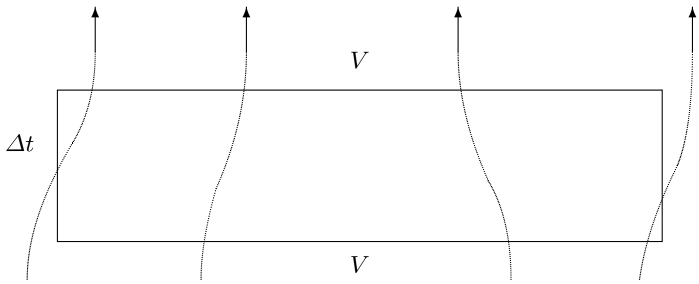


Fig. 3.2. Charge conservation in a spatial domain V

a charge e entering into V, a further particle (right) with the equal charge e departing V, and two particles (middlemost) permanently contained in V; the charge Q_V in V is thus conserved during the period Δt. Charge can appear and disappear in V only with its carrier, a point particle.

The charge conservation of each individual particle is a further reason for extending world lines infinitely. *World lines never terminate* at finite points.

Charge conservation in the quantum realm is more involved due to the availability of both particles and antiparticles together with their creation and annihilation events. Given a particle moving along a timelike world line oriented from the past to the future, its antiparticle may be thought of as an object identical to it in every respect but moving back in time. That is, the antiparticle world line is oriented from the future to the past, as in Fig. 3.3. Accordingly, the annihilation of a pair that occurs at a point A (A for annihilation) is depicted as a Λ-shaped world line of a *single particle* that runs initially from the remote past to the future up to the point A and then returns to the remote past. Likewise, the birth of a pair occurring at a point B (B for birth) is given by a V-shaped world line of a single particle that runs

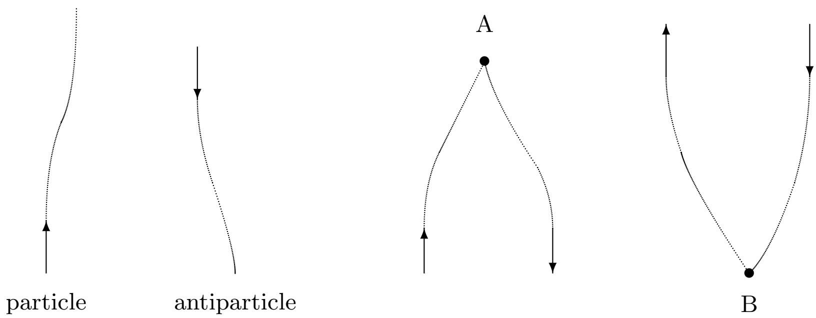


Fig. 3.3. World lines of particles and antiparticles

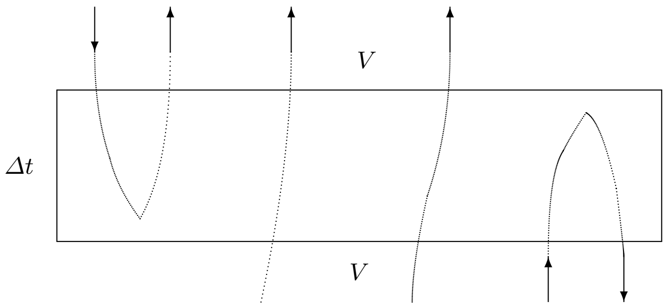


Fig. 3.4. Quantum charge conservation

initially from the far future to the past up to the point B and then returns to the far future, as in Fig. 3.3. Charge is thus conserved even though its carrier may appear or disappear, Fig. 3.4. In the quantum realm, physical quantities are more persistent than the material carriers of these quantities.

Although present considerations are within the orthodox classical framework, we should be alert to notions foreign to it. Classical theory leaves room for both normal particles which experience the proper order of events and antiparticles which follow the reverse order of events. However, creations and annihilations of pairs are banned, which precludes the occurrence of V- and Λ-shaped world lines. That is why we select infinite timelike world lines free of breaks and bifurcations. Broken curves are absent from the classical picture because the least action principle does not apply to V- and Λ-shaped world lines. (Such V- and Λ-shaped world lines would be automatically excluded if a further requirement of *smoothness* would be imposed on the allowable world lines. This requirement is well substantiated in classical theory where abrupt jumps in interparticle forces appear as a speculative, artificial construction.)

We next turn to equation (3.21). Based on the superposition principle, we reiterate mutatis mutandis the above arguments to conclude that m^μ is independent of the field variables $F^{\mu\nu}$, yet may depend on particle characteristics. The comparison between (3.17) and (3.19) shows that the roles of the electric and magnetic fields are now interchanged. Therefore, only particles possessing magnetic couplings $e^\star_I$ contribute to m^μ. In line with the extended action–reaction principle, the total magnetic charge-source $Q^\star$, defined as a conserved integral

$$Q^\star = \int d\sigma_\mu m^\mu \,, \tag{3.42}$$

equals the sum of magnetic charge-couplings,

$$Q^\star = \sum_{I=1}^{N} e^\star_I \,. \tag{3.43}$$

Accordingly, we may refer to $e^\star_I$ as the *magnetic charge* of the Ith particle.

Our *third* assumption is the *absence of magnetic charges* $e^\star_I$ *from nature*, and so

$$m^\mu = 0\,. \tag{3.44}$$

The idea of magnetic monopoles, as such, has a number of attractive theoretical aspects, some of which are discussed in this text. However, despite prodigious experimental efforts that went into searching for magnetic charges, no manifestation of them is found.

With these observations, the electromagnetic field is governed by the equations

$$\partial_\lambda F^{\lambda\mu} = 4\pi j^\mu\,, \tag{3.45}$$

$$\partial_\lambda {}^*F^{\lambda\mu} = 0\,. \tag{3.46}$$

These equations were formulated by James Clerk Maxwell in 1864, and have been named for him. The interpretation of j^μ as the current of charged particles is due to Lorentz. Dirac completed the picture by expressing j^μ according to (3.32).

To summarize, a major part of the information encoded in Maxwell's equations (3.45) and (3.46) is taken from global topological properties of spacetime, and the residual information, seemingly divorced from geometry, which represents the physical contents of these equations, translates into four assumptions:

(i) *Locality*;
(ii) *Linearity* of the dynamical equation, or the *superposition* principle;
(iii) *Identity of the charge-source and the charge-coupling*, or the extended *action–reaction* principle;
(iv) *Lack of magnetic monopoles.*

Problem 3.2.1. Consider the four-current j^μ of a single charged particle moving along an arbitrary timelike world line $z^\mu(s)$,

$$j^\mu(x) = e\int_{-\infty}^{\infty} ds\, v^\mu(s)\,\delta^4\big[x - z(s)\big]\,. \tag{3.47}$$

Show by a direct calculation that j^μ satisfies the local conservation law

$$\partial_\mu j^\mu(x) = 0 \tag{3.48}$$

for any finite point of Minkowski space x^μ.

Problem 3.2.2. Assume that the particle charge is time-dependent, $e = e(s)$, and the corresponding four-current is

$$j^\mu(x) = \int_{-\infty}^{\infty} ds\, e(s)\, v^\mu(s)\,\delta^4\big[x - z(s)\big]\,, \tag{3.49}$$

with the reservation that j^μ satisfies the relation (3.48). Show that

$$\dot{e} = 0 \, , \tag{3.50}$$

which is consistent with (2.37).

Problem 3.2.3. Let a charged particle be alone in the world. If the extended action–reaction principle is taken too literally, the source of electromagnetic field, that is, the term on the right of equation (3.45) seems required to have the form $S^\mu = j_\lambda F^{\lambda\mu}$ where j^μ is defined in (3.47). Indeed, S^μ is the density of the Lorentz force $f^\mu = e\, v_\lambda F^{\lambda\mu}$ responsible for the variation of the particle state. Why is this 'source' inappropriate?

Answer The 'source' $S^\mu = j_\lambda F^{\lambda\mu}$ does not satisfy identically the relation $\partial_\mu S^\mu = 0$, and hence equation (3.45) fails to be consistent.

3.3 Other Forms of Maxwell's Equations

Maxwell's equations in the tensor form, (3.45) and (3.46), provide a compact encoding of geometric features of our world probed by electric charges. They evidence that we live in a three-dimensional space with globally trivial (Euclidean) topology, or, to put it otherwise, in a four-dimensional pseudoeuclidean spacetime. The physical contents of the field dynamics are limited to charge properties, specifically to the constraints on allowable world lines. We will see in the subsequent text that peculiar features of electromagnetic field owe their origin to the coupling with charged matter. For now, however, the equation of motion as such should be discussed more elaborately.

With reference to Problem 3.1.2, equation (3.46) may be rearranged to give

$$\partial_\lambda F_{\mu\nu} + \partial_\nu F_{\lambda\mu} + \partial_\mu F_{\nu\lambda} = 0 \, . \tag{3.51}$$

We regard (3.46) and (3.51) as equivalent and call them collectively the *Bianchi identity.*

We next present Maxwell's equations in terms of differential forms. Given the 2-form $F = \frac{1}{2}\, F_{\mu\nu}\, dx^\mu \wedge dx^\nu$, (3.45) and (3.46) become respectively

$$d\,{}^*F = 4\pi J \, , \tag{3.52}$$

$$dF = 0 \tag{3.53}$$

where the 3-form $J = \frac{1}{6}\, J_{\lambda\mu\nu}\, dx^\lambda \wedge dx^\mu \wedge dx^\nu$ is expressed through j^μ as $J_{\lambda\mu\nu} = \epsilon_{\lambda\mu\nu\rho}\, j^\rho$.

Let us verify that (3.53) follows from (3.51). We have

$$dF = \frac{1}{2}\, (dF_{\mu\nu})\, dx^\mu \wedge dx^\nu = \frac{1}{2}\, \partial_\lambda F_{\mu\nu}\, dx^\lambda \wedge dx^\mu \wedge dx^\nu \, . \tag{3.54}$$

Because $dx^\lambda \wedge dx^\mu \wedge dx^\nu$ is completely antisymmetric, one may permute $\partial_\lambda F_{\mu\nu}$ to obtain

$$dF = \frac{1}{12}\left[(\partial_\lambda F_{\mu\nu}+\partial_\nu F_{\lambda\mu}+\partial_\mu F_{\nu\lambda})-(\partial_\lambda F_{\nu\mu}+\partial_\mu F_{\lambda\nu}+\partial_\nu F_{\mu\lambda})\right] dx^\lambda\wedge dx^\mu\wedge dx^\nu \,. \tag{3.55}$$

Any term of the second parenthesis differs from the corresponding term of the first ones by interchanging indices of the antisymmetric tensor $F_{\alpha\beta}$. Therefore, the expressions in both parentheses are equal and of opposite sign, and

$$dF = \frac{1}{6}\left(\partial_\lambda F_{\mu\nu} + \partial_\nu F_{\lambda\mu} + \partial_\mu F_{\nu\lambda}\right) dx^\lambda \wedge dx^\mu \wedge dx^\nu \,. \tag{3.56}$$

This demonstrates that the homogeneous equations (3.51) and (3.53) are equivalent.

Let us check that ${}^*d^*$ plays the role of the four-divergence, and hence (3.45) results from (3.52). The Hodge operator acts on the 2-form $dx^\mu \wedge dx^\nu$ as

$$^*(dx^\lambda \wedge dx^\mu) = \frac{1}{2}\,\epsilon_{\nu\rho}{}^{\lambda\mu}\, dx^\nu \wedge dx^\rho \,. \tag{3.57}$$

It follows that

$$^*F = \frac{1}{2}\, F_{\lambda\mu}\, {}^*(dx^\lambda \wedge dx^\mu) = \frac{1}{4}\, F_{\lambda\mu}\, \epsilon_{\nu\rho}{}^{\lambda\mu}\, dx^\nu \wedge dx^\rho = \frac{1}{2}\, {}^*F_{\nu\rho}\, dx^\nu \wedge dx^\rho \tag{3.58}$$

where ${}^*F_{\nu\rho} = \frac{1}{2}\,\epsilon_{\nu\rho\sigma\tau}\, F^{\sigma\tau}$, as defined in (3.10). Therefore,

$$d\,{}^*F = \frac{1}{6}\left(\partial_\lambda {}^*F_{\mu\nu} + \partial_\nu {}^*F_{\lambda\mu} + \partial_\mu {}^*F_{\nu\lambda}\right) dx^\lambda \wedge dx^\mu \wedge dx^\nu \,. \tag{3.59}$$

The action of the Hodge operator on the 3-form $dx^\lambda \wedge dx^\mu \wedge dx^\nu$ shows up as

$$^*(dx^\lambda \wedge dx^\mu \wedge dx^\nu) = \epsilon_\rho{}^{\lambda\mu\nu}\, dx^\rho \,. \tag{3.60}$$

By (3.59) and (3.60),

$$^*d\,{}^*F = \frac{1}{6}\left(\partial_\lambda {}^*F_{\mu\nu} + \partial_\nu {}^*F_{\lambda\mu} + \partial_\mu {}^*F_{\nu\lambda}\right) \epsilon_\rho{}^{\lambda\mu\nu}\, dx^\rho \,. \tag{3.61}$$

We restrict ourselves to the calculation of only the first term in the parenthesis,

$$\epsilon_\rho{}^{\lambda\mu\nu}\,\partial_\lambda {}^*F_{\mu\nu} = \frac{1}{2}\,\epsilon_\rho{}^{\lambda\mu\nu}\,\epsilon_{\mu\nu\alpha\beta}\,\partial_\lambda F^{\alpha\beta} = (\eta_{\rho\beta}\delta^\lambda_\alpha - \eta_{\rho\alpha}\delta^\lambda_\beta)\,\partial_\lambda F^{\alpha\beta} = 2\partial_\lambda F^\lambda{}_\rho \,, \tag{3.62}$$

since other terms may be produced by a cyclic permutation of indices.

On the other hand, applying the Hodge operator to J, in view of (3.60), we find

$$^*J = {}^*\left(\frac{1}{6}\, J_{\lambda\mu\nu}\, dx^\lambda \wedge dx^\mu \wedge dx^\nu\right) = \frac{1}{6}\,\epsilon_{\lambda\mu\nu\alpha}\, j^\alpha\, \epsilon_\rho{}^{\lambda\mu\nu}\, dx^\rho = j_\rho\, dx^\rho \,. \tag{3.63}$$

We thus arrive at the relation $\partial_\lambda F^\lambda{}_\rho = 4\pi j_\rho$ which proves the assertion.

Equations (3.52) and (3.53) represent electrodynamics in a concise and elegant form. This expression of Maxwell's equations is coordinate-free. It is applicable to any smooth four-dimensional pseudo-Riemannian manifold.

By (3.17) and (3.19), Maxwell's equations acquire the three-dimensional vector form:

$$\mathrm{div}\,\mathbf{E} = 4\pi\varrho\ , \tag{3.64}$$

$$\mathrm{curl}\,\mathbf{B} = 4\pi\mathbf{j} + \frac{\partial\mathbf{E}}{\partial t}\ , \tag{3.65}$$

$$\mathrm{div}\,\mathbf{B} = 0\ , \tag{3.66}$$

$$\mathrm{curl}\,\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t}\ . \tag{3.67}$$

Equations (3.64) and (3.66) contains no time derivatives. These equations bear no relation to the field's evolution in time, being mere *constraints*, similar to the well-known mechanical constraints that cause a particle to move across given surfaces. To set up the Cauchy problem for the evolutionary equations (3.65) and (3.67), the constraints (3.64) and (3.66) should be properly taken into account on the initial, as well as any subsequent, spacelike hyperplanes.

One may integrate (3.64) over the volume of a three-dimensional domain V enclosed by a surface ∂V, and apply the Gauss–Ostrogradskiĭ theorem

$$\int_V d^3x\,\mathrm{div}\,\mathbf{E} = \oint_{\partial V} d\mathbf{S}\cdot\mathbf{E} \tag{3.68}$$

to obtain

$$\oint_{\partial V} d\mathbf{S}\cdot\mathbf{E} = 4\pi\, Q_V\ . \tag{3.69}$$

This is the so-called *Gauss law* which states: the integral flux of the electric intensity $\mathbf{E}$ outward the domain V equals the charge Q_V contained in V times 4π.

Likewise, the integration of (3.66) gives

$$\oint_{\partial V} d\mathbf{S}\cdot\mathbf{B} = 0\ . \tag{3.70}$$

This relation tells us that the flux integral of the magnetic induction $\mathbf{B}$ outward from any domain V is zero. The *closure of magnetic lines of force*, and the lack of magnetic monopoles is thus seen.

Let a two-dimensional surface $\mathcal{S}$ be bounded by a loop $\mathcal{L}$. We calculate the flux of the vector equation (3.65) through $\mathcal{S}$ and apply the Stokes theorem

$$\int_{\mathcal{S}} d\mathbf{S}\cdot\mathrm{curl}\,\mathbf{B} = \oint_{\mathcal{L}} d\mathbf{x}\cdot\mathbf{B} \tag{3.71}$$

to get

$$\oint_{\mathcal{L}} d\mathbf{x} \cdot \mathbf{B} = \int_{\mathcal{S}} d\mathbf{S} \cdot \left(4\pi \mathbf{j} + \frac{\partial \mathbf{E}}{\partial t} \right) . \tag{3.72}$$

For stationary currents, (3.72) acquires the form

$$\oint_{\mathcal{L}} d\mathbf{x} \cdot \mathbf{B} = 4\pi \int_{\mathcal{S}} d\mathbf{S} \cdot \mathbf{j} . \tag{3.73}$$

This relation, known as *Ampère's law*, implies that the line integral of $\mathbf{B}$ around the loop $\mathcal{L}$ equals 4π of the current $\mathbf{j}$ through the surface $\mathcal{S}$.

The term $\partial \mathbf{E}/\partial t$ becomes essential in non-stationary situations. This term is called the *displacement current*. It was introduced by Maxwell for purely theoretical reasons, specifically, to make the equation of continuity (3.37) valid, see Problem 3.3.3.

In a similar fashion, the integration of (3.67) leads to

$$\oint_{\mathcal{L}} d\mathbf{x} \cdot \mathbf{E} = -\frac{d}{dt} \int_{\mathcal{S}} d\mathbf{S} \cdot \mathbf{B} . \tag{3.74}$$

The relation (3.74) is known as *Faraday's law*. It implies that the line integral of $\mathbf{E}$ around the loop $\mathcal{L}$ (called the *electromotive force* through $\mathcal{L}$) is equal to minus the time derivative of the flux of $\mathbf{B}$ through the surface $\mathcal{S}$.

The laws of electromagnetism were originally discovered experimentally in the integral forms (3.69) and (3.70), together with (3.73) and (3.74). This form of electrodynamics is convenient for applications. In particular, Gauss' law (3.69) provides a simple way for the calculation of the electric field $\mathbf{E}$ in many electrostatic problems, and Ampère's law (3.73) enables the determination of the spatial behavior of the magnetic field $\mathbf{B}$ generated by a stationary current. (See Problems 3.3.1–3.3.2.)

Problem 3.3.1. *Use of Gauss' law.* (i) Let a ball of radius R carry a charge Q which is distributed homogeneously over the volume. Using Gauss' law (3.69), find the electric field intensity $\mathbf{E}$ both inside and outside the ball.

(ii) Let a charge Q be distributed homogeneously over a sphere of radius R. Find $\mathbf{E}$ inside and outside the sphere.

(iii) Consider an infinite capacitor with flat parallel plates which carry opposite charges of magnitude σ per a unit area. Find $\mathbf{E}$ inside and outside the capacitor.

Answer (i) $\mathbf{E}$ is directed along the radius-vector,

$$E_r = Q \left[\frac{r}{R^3}\, \theta(R - r) + \frac{1}{r^2}\, \theta(r - R) \right] \tag{3.75}$$

where $\theta(R - r)$ is the Heaviside step function.

$$\text{(ii)} \quad E_r = \frac{Q}{r^2}\, \theta(r - R) . \tag{3.76}$$

(iii) $\mathbf{E}$ is perpendicular to the plates at any point inside the capacitor, $E = 4\pi\sigma$. Outside the capacitor, the field vanishes.

Problem 3.3.2. *Use of Ampère's law.* Consider a rectilinear infinite conductor of a uniform circular cross section $\mathcal{S}$ and a steady current of electrons $\mathcal{I}$ along it,

$$\mathcal{I} = \int_{\mathcal{S}} d\mathbf{S} \cdot \mathbf{j} \,. \tag{3.77}$$

Determine the magnetic induction $\mathbf{B}$ everywhere outside the conductor.

Answer Using cylindrical coordinates (z, r, and φ, with the z-axis parallel to the conductor) the magnetic induction is $\mathbf{B} = (B_z, B_r, B_\varphi)$,

$$B_z = 0, \quad B_r = 0, \quad B_\varphi = \frac{2\mathcal{I}}{r} \,. \tag{3.78}$$

Problem 3.3.3 Show that the equation of continuity (3.37) follows from two Maxwell's equations (3.64) and (3.65). Notice the role of the displacement current.

Notes

1. The basic references on the Maxwell–Lorentz electrodynamics, a classical theory of charged point particles interacting with electromagnetic field, are Maxwell (1873) and Lorentz (1909). The reader may find the two volume set by Whittaker (1910) and (1953) a useful guide in the history of classical theories of electromagnetism. The paper by Jackson & Okun (2001) narrates some little-known facts concerning the formation of the notion of gauge invariance and give further references, including historical surveys.

2. *Section 3.1.* We give a somewhat uncommon derivation of Maxwell's equations. Our line of argument rests heavily on a statement known as the Helmholtz theorem, which is in fact a corollary of results obtained in Helmholtz (1858). Most modern texts introduce Maxwell's equations in a well-established way. For example, the book by Jackson (1962) follows the inductive method, while the text by Landau & Lifshitz (1971) postulates a suitable action and invokes the principle of least action. Hehl & Obukhov (2003) derives Maxwell's equations from six assumptions, which, while having much in common with the assumptions adopted here, follow a different logical pattern.

3. *Section 3.2.* The remark that charged matter is unstable, which prompts us to search for nonelectromagnetic cohesive forces, has been a source of much speculation since the original work of Poincaré (1906). Frenkel (1925) argues that a point electron is stable, and thus dispenses with the need for the cohesive forces. Distributions as mathematical tools were promoted in Dirac

(1930), in particular the delta-function is systematically employed in this book. As early as 1899, the delta-function, together with its Fourier transform, was introduced in Heaviside (1899) under the name *impulsive function*. However, this discovery exceeded the requirements of physics at that time, and went unnoticed. The notion of distributions as functionals on some space of test functions was proposed in Sobolev (1936) and Schwartz (1950–1951). Dirac (1938) gave an account of the motion of a classical point electron through the use of the delta-function.

The electron was discovered by Thomson (1897). Analyzing the Dirac equation, Dirac (1931) assumed that its negative-energy solutions describe a particle with the same mass as the electron, but with opposite charge. The positron was discovered shortly thereafter in 1932, which gave impetus to research concerning antimatter.

The concept of magnetic monopoles was described by Dirac (1931), (1948) within the scope of quantum mechanics. For the present status of experimental searches for magnetic monopoles see Chap. 10 of Klapdor-Kleingrothaus & Zuber (1997).

4. *Section 3.3.* Maxwell's equations in vector notation (3.64)–(3.67) were brought into use in Heaviside (1892). Cartan (1924) applied the differential form calculus to represent Maxwell's equations in a coordinate-free form. For more detail on Maxwell's equations in terms of differential forms see Misner et al. (1973). Representations of Maxwell's equations are many and varied. Some of them can be found in the book by Fushchich & Nikitin (1987).

4

Solutions to Maxwell's Equations

In this chapter we study solutions to Maxwell's equations

$$\partial_\lambda {}^*F^{\lambda\mu} = 0 \, , \tag{4.1}$$

$$\partial_\lambda F^{\lambda\mu} = 4\pi j^\mu \tag{4.2}$$

for various sources j^μ. We begin with the simplest case of static electric fields and compare it with the closely similar case of constant magnetic fields. To solve Maxwell's equations in the general case, we define the four-vector potential A^μ, a key element of gauge theories. We turn to the special case $j^\mu = 0$ when Maxwell's equations reduce to the homogeneous wave equation and describe free electromagnetic fields. The inhomogeneous wave equation is used to illustrate the Green's function technique. We further discuss a method of solving Maxwell's equations with the source composed of a single arbitrarily moving point charge, without resort to Green's functions. This method will be of particular assistance in solving the Yang–Mills equations in Chap. 8. Finally we consider features of the electromagnetic field due to a magnetic monopole.

One important point to remember is that we are dealing with a system of linear partial differential equations of the hyperbolic type. A regular way for examining a *linear* differential equation with *constant* coefficients is to look at the Fourier-series or Fourier-integral expansions of the desired function which convert this differential equation to an algebraic equation. Since our interest is only with fields distributed over empty space, it is adequate to use a Fourier-integral expansion.

4.1 Statics

When charges are at rest, they generate static fields, that is, $\partial\mathbf{E}/\partial t = \partial\mathbf{B}/\partial t = 0$. The equations of electrostatics, resulting from (3.67) and (3.64), are

$$\nabla \times \mathbf{E} = 0 , \tag{4.3}$$

$$\nabla \cdot \mathbf{E} = 4\pi\varrho . \tag{4.4}$$

(In the analysis given below the symbol ∇ is preferred over the traditional notations grad, curl and div.) Our immediate task is to find solutions to (4.3) and (4.4), and check the stability of these solutions against small perturbations.

One might think that the system of equations (4.3) and (4.4) is overdetermined: four equations are intended for finding three functions E_1, E_2 and E_3. It is easy to see, however, that (4.3) is satisfied identically by the ansatz

$$\mathbf{E} = -\nabla\phi . \tag{4.5}$$

Here, ϕ is an arbitrary function of $\mathbf{r}$. This quantity is called the *scalar potential* or simply the *potential* of the electric field. Combining (4.5) and (4.4), we obtain the equation

$$\nabla^2\phi = -4\pi\varrho \tag{4.6}$$

known as the *Poisson equation*. We have arrived at the system of equations (4.5) and (4.6) with the number of equations equal to the number of desired functions.

Does relation (4.5) represent the general solution to equation (4.3)? To make sure that this is indeed the case, we write the Fourier transform

$$\mathbf{E}(\mathbf{r}) = \frac{1}{(2\pi)^3}\int d^3k\, e^{i\mathbf{k}\cdot\mathbf{r}}\, \widetilde{\mathbf{E}}(\mathbf{k}) , \tag{4.7}$$

and decompose $\widetilde{\mathbf{E}}$ into three basis vectors,

$$\widetilde{\mathbf{E}} = \mathbf{e}_1\widetilde{E}_1 + \mathbf{e}_2\widetilde{E}_2 + \mathbf{e}_3\widetilde{E}_3 , \tag{4.8}$$

where $\mathbf{e}_1$ is taken to be $\mathbf{e}_1 = \mathbf{k}/k$, $k = |\mathbf{k}|$, and $\mathbf{e}_2$ and $\mathbf{e}_3$ are arbitrary unit vectors which together with $\mathbf{e}_1$ span the orthonormalized basis,

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}, \quad i, j = 1, 2, 3 . \tag{4.9}$$

The action of ∇ on $\mathbf{E}$ amounts to the multiplication of $\widetilde{\mathbf{E}}$ by $i\mathbf{k}$. From (4.3) follows

$$\mathbf{k} \times \widetilde{\mathbf{E}} = k\,(\mathbf{e}_3\widetilde{E}_2 - \mathbf{e}_2\widetilde{E}_3) = 0 . \tag{4.10}$$

We take the scalar product of this equation with $\mathbf{e}_3$ and $\mathbf{e}_2$ using (4.9) to yield $\widetilde{E}_2 = \widetilde{E}_3 = 0$, and so

$$\widetilde{\mathbf{E}} = \mathbf{e}_1\widetilde{E}_1 = -i\mathbf{k}\,\widetilde{\phi} \tag{4.11}$$

where $\widetilde{\phi}$ is an arbitrary function of $\mathbf{k}$. Note that (4.11) is just the Fourier transform of (4.5). So, both the sense of equation (4.3), which is a constraint eliminating some Fourier modes of the electric field, and the issue of generality of its solution (4.5) become clear.

Formula (4.5) defines the potential ϕ up to adding an arbitrary constant. Indeed, $\mathbf{E}$ is unchanged by the transformation

$$\phi \to \phi' = \phi + C, \quad C = \text{const} \,. \tag{4.12}$$

In nonrelativistic mechanics where $e\phi$ plays the role of potential energy, this fact can be interpreted as the freedom in choosing the zero point of energy.

The Poisson equation (4.6) is linear in the unknown function ϕ. Therefore, the general solution is $\phi = \phi_* + \Phi$ where ϕ_* is a particular solution of this equation, and Φ the general solution of the associated homogeneous equation

$$\nabla^2 \Phi = \frac{\partial^2 \Phi}{\partial x_1^2} + \frac{\partial^2 \Phi}{\partial x_2^2} + \frac{\partial^2 \Phi}{\partial x_3^2} = 0 \tag{4.13}$$

called the *Laplace equation*. Functions Φ satisfying (4.13) are called *harmonic*.

We point out that solutions of the Laplace equation have no local maxima and minima. Indeed, at extremum points, the first derivatives of Φ vanish while the second derivatives have the same signs. However, equation (4.13) implies that the second derivatives must have different signs or simultaneously vanish, resulting in $\Phi = \text{const}$. Thus any harmonic function in some spatial domain takes its maximal and minimal values on the boundary of this domain. In particular, given a harmonic function Φ vanishing on the boundary, Φ is identically zero in this domain. (Incidentally, this is the reason for lack of magnetic field due to static electric charges. The corresponding couple of Maxwell's equations $\nabla \times \mathbf{B} = 0$, $\nabla \cdot \mathbf{B} = 0$ is equivalent to the vector Laplace equation $\nabla^2 \mathbf{B} = 0$ which admits only the trivial solution $\mathbf{B} = 0$ provided that $\mathbf{B}$ disappears at infinity.)

Let a single point charge e be placed at the origin. The potential ϕ generated by it in empty space is described by

$$\nabla^2 \phi(\mathbf{r}) = -4\pi e \, \delta^3(\mathbf{r}) \,. \tag{4.14}$$

We assume that ϕ tends to zero as $r \to \infty$. Then the solution to the Laplace equation is identically zero, and only a particular solution to equation (4.14) vanishing at infinity remains to be found. We insert the Fourier transforms of $\phi(\mathbf{r})$ and $\delta^3(\mathbf{r})$,

$$\phi(\mathbf{r}) = \frac{1}{(2\pi)^3} \int d^3k \, e^{i\mathbf{k}\cdot\mathbf{r}} \, \widetilde{\phi}(\mathbf{k}) \,, \tag{4.15}$$

$$\delta^3(\mathbf{r}) = \frac{1}{(2\pi)^3} \int d^3k \, e^{i\mathbf{k}\cdot\mathbf{r}} \,, \tag{4.16}$$

in (4.14) to obtain

$$k^2 \widetilde{\phi} = 4\pi e \,. \tag{4.17}$$

It follows that

$$\widetilde{\phi} = \frac{4\pi e}{k^2} \,. \tag{4.18}$$

Inserting (4.18) in (4.15) and using spherical coordinates in the k-space with the k_3-axis collinear to $\mathbf{r}$ gives $d^3k = k^2dk \, \sin\vartheta d\vartheta \, d\varphi$, $\mathbf{k}\cdot\mathbf{r} = kr\cos\vartheta$, and

$$\phi(\mathbf{r}) = \frac{4\pi e}{(2\pi)^3}\int_0^\infty dk \int_0^\pi d\vartheta \sin\vartheta \, e^{ikr\cos\vartheta}\int_0^{2\pi} d\varphi \, . \tag{4.19}$$

With the change of variables $\xi = kr$, we arrive at

$$\phi = \frac{e}{r}\,\beta \tag{4.20}$$

where β is some numerical factor. The calculation of β is quite elementary, we give only the net result: $\beta = 1$. Thus

$$\phi(\mathbf{r}) = \frac{e}{r} \, . \tag{4.21}$$

For the calculation of $\mathbf{E}$, we need the derivatives of Cartesian coordinates with respect to another Cartesian coordinates:

$$\frac{\partial x_j}{\partial x_i} = \delta_{ij} \, , \tag{4.22}$$

and also the derivatives of the radial distance with respect to Cartesian coordinates:

$$\frac{\partial r}{\partial x_i} = \frac{\partial}{\partial x_i}\sqrt{x_1^2 + x_2^2 + x_3^2} = \frac{x_i}{r} = n_i \tag{4.23}$$

where $\mathbf{n}$ is a unit vector directed along the radius vector. (Recall, all manipulations with Euclidean vectors and tensors are carried out using subscripts, and the repeated index summation rule is understood.) In view of (4.5),

$$\mathbf{E} = e\,\frac{\mathbf{n}}{r^2} \, . \tag{4.24}$$

The potential (4.21) is called the *Coulomb potential*, and the electric field (4.24) is called the *Coulomb field*. We point out the characteristic $1/r$ singularity of the potential and $1/r^2$ singularity of the field strength at the location of the charge.

Such a simple form of the solution is due to the adoption of Gaussian units. Were it not for the factor of 4π on the right of the Maxwell equation (3.45) there would be the excess factor $1/4\pi$ in the potential (4.21) and field strength (4.24).

Expressions (4.21) and (4.24) can be alternatively derived by observing that fields generated by a point source are spherically symmetric. That is, ϕ depends only on r. The Laplace equation holds everywhere outside the source. We may restrict our consideration to the radial part of the Laplace equation in spherical coordinates,

$$\frac{1}{r^2}\,\frac{\partial}{\partial r}\,r^2\,\frac{\partial}{\partial r}\,\phi = 0 \, . \tag{4.25}$$

The general solution to this equation is readily obtained

$$\phi = \frac{C_1}{r} + C_2 \tag{4.26}$$

where C_1 and C_2 are arbitrary constants. Because $\mathbf{E}$ is invariant under the transformation (4.12), we are entitled to choose any C_2, in particular $C_2 = 0$. Differentiating ϕ, we observe that $\mathbf{E}$ is directed along the radius vector. The flux of $\mathbf{E}$ through a sphere with its center at the origin is $4\pi r^2 E$. By Gauss' law (3.69), it must be equal to $4\pi e$, hence $C_1 = e$.

The potential generated by several static charges can be represented as a superposition of potentials generated by each individual charge:

$$\phi = \sum_{I=1}^{N} \frac{e_I}{|\mathbf{r} - \mathbf{z}_I|} \,. \tag{4.27}$$

One can then guess that a solution to equation (4.6) with an arbitrary continuous charge distribution $\varrho(\mathbf{r})$ is

$$\phi(\mathbf{r}) = \int d^3x \, \frac{\varrho(\mathbf{x})}{|\mathbf{r} - \mathbf{x}|} \,. \tag{4.28}$$

Indeed,

$$\nabla^2 \phi(\mathbf{r}) = \int d^3x \, \varrho(\mathbf{x}) \, \nabla^2 \frac{1}{|\mathbf{r} - \mathbf{x}|} = -\int d^3x \, \varrho(\mathbf{x}) \, 4\pi\delta^3(\mathbf{r} - \mathbf{x}) = -4\pi\varrho(\mathbf{r}) \tag{4.29}$$

where we have taken into account that $1/|\mathbf{r} - \mathbf{x}|$ coincides, up to a numerical factor and a shift of the origin, with the Coulomb solution (4.21) of the Poisson equation (4.14).

When $\varrho(\mathbf{r}) = \sum e_I \, \delta^3(\mathbf{r} - \mathbf{z}_I)$, the solution (4.28) regains the form (4.27).

Consider the case that a continuous distribution of charged matter is located in a compact spatial region V. Put the origin at some point of this region. How does the field behave at large distances from the source, that is, for $r \gg |\mathbf{x}|$? To answer this question, we expand $1/|\mathbf{r} - \mathbf{x}|$ in powers of $\mathbf{x}/r$:

$$\frac{1}{|\mathbf{r} - \mathbf{x}|} = \frac{1}{r} - x_i \partial_i \frac{1}{r} + \frac{1}{2} \, x_i x_j \partial_i \partial_j \frac{1}{r} + \cdots \tag{4.30}$$

We use in (4.30) relations stemming from (4.22) and (4.23)

$$\partial_i \frac{1}{r} = -\frac{n_i}{r^2}, \quad \partial_i \partial_j \frac{1}{r} = \frac{3n_i n_j - \delta_{ij}}{r^3}, \quad \delta_{ij} \partial_i \partial_j \frac{1}{r} = 0 \tag{4.31}$$

where n_i is the unit vector aligned with the radius vector, to obtain

$$\frac{1}{|\mathbf{r} - \mathbf{x}|} = \frac{1}{r} + \frac{\mathbf{x} \cdot \mathbf{n}}{r^2} + \frac{3}{2} \left(x_i x_j - \frac{1}{3} \, x^2 \delta_{ij} \right) \frac{n_i n_j}{r^3} + \cdots \tag{4.32}$$

If the total charge (zeroth-order electric moment of the source)

$$Q = \int_V d^3x\, \varrho(\mathbf{x}) \tag{4.33}$$

is finite, the term of the potential ϕ proportional to $1/r$ dominates at large r. If $Q = 0$, and yet the *electric dipole moment*

$$\mathbf{d} = \int_V d^3x\, \mathbf{x}\, \varrho(\mathbf{x}) \tag{4.34}$$

is nonzero, the potential behaves asymptotically as

$$\phi = \frac{\mathbf{d}\cdot\mathbf{n}}{r^2}\,. \tag{4.35}$$

The electric field generated by the dipole is

$$\mathbf{E} = -\nabla\,\frac{\mathbf{d}\cdot\mathbf{n}}{r^2} = \frac{3(\mathbf{d}\cdot\mathbf{n})\,\mathbf{n} - \mathbf{d}}{r^3}\,. \tag{4.36}$$

When both Q and $\mathbf{d}$ vanish, but the *electric quadrupole moment*

$$D_{ij} = \int_V d^3x\,(3x_i x_j - x^2\delta_{ij})\,\varrho(\mathbf{x}) \tag{4.37}$$

is nonzero, the leading asymptotic term of the potential becomes

$$\phi = \frac{D_{ij}\,n_i n_j}{2r^3}\,. \tag{4.38}$$

Thus neutral systems can interact by their multipoles. In materials composed of polar molecules, such as H_2O, neighboring molecular dipoles are turned to one another by opposite poles causing their attraction. Intermolecular binding of this kind is referred to as a *van der Waals interaction.* Although fields generated by multipoles fall with distance faster than the Coulomb field, the van der Waals interaction is still sufficiently strong to keep matter in a condensed state.

Are the solutions (4.27) and (4.28) stable against small perturbations of the charge density? In other words, is a stable equilibrium possible for a charge introduced in electrostatic fields? The answer is 'no'. This is due to the lack of configurations ϕ with local minima, that is, lack of points where the potential energy of the charge $e\phi(\mathbf{r})$ is minimal.

We now turn to the magnetostatic equations

$$\nabla\cdot\mathbf{B} = 0\,, \tag{4.39}$$

$$\nabla\times\mathbf{B} = 4\pi\mathbf{j} \tag{4.40}$$

derivable from (3.66) and (3.65) when $\partial\mathbf{E}/\partial t = 0$. They describe constant magnetic fields due to *stationary* electric currents. Recall, one refers to a process in a continuous medium as stationary if the density of the medium is constant in time. In the present case, the electric charge density is constant, $\partial\varrho/\partial t = 0$. The equation of continuity (3.37) takes the form

$$\nabla \cdot \mathbf{j} = 0 \,. \tag{4.41}$$

This equation is often regarded as the definition of a stationary current.

While on the subject of the current $\mathbf{j}$, one usually conceives the motion of charges along a circuit. Charges forming stationary currents execute periodic motions along closed paths. This makes clear the difference between sources in electrostatics and those in magnetostatics. The former are composed of charges which move along parallel straight world lines while the latter are composed of charges which move along regular helical world lines with parallel axes.

The system of equations (4.39) and (4.40) may also seem overdetermined. However, the ansatz

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{4.42}$$

obeys equation (4.39) identically. We call $\mathbf{A}$ the *vector potential* of the magnetic field.

Formula (4.42) defines $\mathbf{A}$ only up to addition of gradients of arbitrary scalar functions. Indeed, the magnetic induction $\mathbf{B}$ remains invariant under the transformations

$$\mathbf{A} \to \mathbf{A}' = \mathbf{A} + \nabla\chi \,, \tag{4.43}$$

where χ is an arbitrary smooth function of $\mathbf{r}$. We refer to (4.43) as *gauge transformations*. Thus (4.42) defines the equivalence class of vector-valued functions related to one other by gauge transformations (4.43), rather than a concrete function $\mathbf{A}$.

We can eliminate the gauge arbitrariness by imposing an additional constraint on $\mathbf{A}$. For example, one may require

$$\nabla \cdot \mathbf{A} = 0 \,, \tag{4.44}$$

called the *Coulomb gauge*. In fact, given a representative of the equivalence class $\mathbf{A}$, one would like to check the existence of a function χ which ensures the gauge condition $\nabla \cdot \mathbf{A}' = 0$ for some $\mathbf{A}'$. From (4.43) and (4.44) it follows that

$$\nabla^2\chi = -\nabla \cdot \mathbf{A} \,. \tag{4.45}$$

The desired function is constructed with the aid of (4.28):

$$\chi(\mathbf{r}) = \frac{1}{4\pi}\int d^3x \, \frac{\nabla \cdot \mathbf{A}(\mathbf{x})}{|\mathbf{r} - \mathbf{x}|} \,. \tag{4.46}$$

Substituting (4.42) in (4.40) and taking into account (4.44) shows

$$\nabla^2\mathbf{A} = -4\pi\mathbf{j} \,. \tag{4.47}$$

We thus come to the system of equations (4.42) and (4.47) where the number of equations equals the number of desired functions.

A solution to the vector Poisson equation (4.47) patterned after (4.28) is

$$\mathbf{A}(\mathbf{r}) = \int d^3x \, \frac{\mathbf{j}(\mathbf{x})}{|\mathbf{r}-\mathbf{x}|} \, . \tag{4.48}$$

In view of (4.42), the curl of the vector potential $\mathbf{A}$ given by (4.48) yields the magnetic induction

$$\mathbf{B}(\mathbf{r}) = \int d^3x \, \frac{\mathbf{j}(\mathbf{x}) \times (\mathbf{r}-\mathbf{x})}{|\mathbf{r}-\mathbf{x}|^3} \, . \tag{4.49}$$

This relation represents the *Biot and Savart law.* For a system of N charged particles, $\mathbf{j}$ is given by (3.36), and (4.49) becomes

$$\mathbf{B} = \sum_{I=1}^{N} e_I \, \frac{\mathbf{v}_I \times (\mathbf{r}-\mathbf{z}_I)}{|\mathbf{r}-\mathbf{z}_I|^3} \, . \tag{4.50}$$

By analogy with the electrostatics, we separate terms of the magnetic field which are generated by different magnetic multipole moments. We substitute (4.32) in (4.48) and keep only two initial terms of the expansion:

$$A_i(\mathbf{r}) = \frac{1}{r} \int_V d^3x \, j_i(\mathbf{x}) + \frac{1}{r^2} \int_V d^3x \, n_k x_k \, j_i(\mathbf{x}) + \cdots \tag{4.51}$$

The first term on the right is zero. Indeed, using the stationary state condition (4.41),

$$\partial_l \left(x_k j_l \right) = \delta_{kl} \, j_l + x_k \, \partial_l \, j_l = j_k \, . \tag{4.52}$$

Therefore, any stationary current j_i can be expressed as

$$j_i = \partial_l \left(x_i j_l \right) . \tag{4.53}$$

This suggests the use of the Gauss–Ostrogradskiĭ theorem which transforms the first term of (4.51) to the flux of $x_i j_l$ through the surface of V. But the charged matter under examination is assumed to be confined to the region V, and so this flux is zero. The vanishing of zeroth order magnetic moments reflects the conjecture that magnetic monopoles do not exist. Ampère even hypothesized that *any* magnetic field is due to the circuition of charges along closed paths (the so-called Ampère's molecular currents).

We now turn to the second term of expression (4.51). One can prove the identity

$$\mathbf{n} \times (\mathbf{x} \times \mathbf{j}) = (\mathbf{n} \cdot \mathbf{j}) \, \mathbf{x} - (\mathbf{n} \cdot \mathbf{x}) \, \mathbf{j} \, . \tag{4.54}$$

The first term on the right hand side of this identity is equal and of opposite sign to the second term plus the divergence of some quantity. Indeed, insertion of (4.53) in it gives

$$n_k j_k x_i = x_i n_k \partial_l (x_k j_l) = \partial_l (x_i n_k x_k j_l) - x_k j_l \partial_l (x_i n_k) = \partial_l (x_i n_k x_k j_l) - n_k x_k j_i \,. \tag{4.55}$$

Thus

$$\int_V d^3x \, (\mathbf{n} \cdot \mathbf{x}) \, \mathbf{j}(\mathbf{x}) = -\frac{1}{2} \, \mathbf{n} \times \left(\int_V d^3x \, \mathbf{x} \times \mathbf{j}(\mathbf{x}) \right) \tag{4.56}$$

where we discarded the surface integral of $x_i n_k x_k j_l$. If we define the *magnetic dipole moment* of the given current distribution as

$$\mathbf{m} = \frac{1}{2} \int_V d^3x \, \mathbf{x} \times \mathbf{j}(\mathbf{x}) \,, \tag{4.57}$$

we conclude that the term

$$\mathbf{A} = \frac{\mathbf{m} \times \mathbf{n}}{r^2} \tag{4.58}$$

of the expansion (4.51) dominates at large r. The associated magnetic induction is

$$\mathbf{B} = \nabla \times \left(\frac{\mathbf{m} \times \mathbf{n}}{r^2} \right) = \frac{3(\mathbf{m} \cdot \mathbf{n}) \, \mathbf{n} - \mathbf{m}}{r^3} \,. \tag{4.59}$$

Comparison of (4.59) and (4.36) shows that the magnetic induction due to a magnetic dipole behaves identically to the electric field intensity due to an electric dipole.

Despite the similarity between electrostatics and magnetostatics, the problem of the stability of the latter is not as simple as that of the former. The reasoning given above for the instability of any system of motionless charges is inapplicable to systems governed by magnetostatic laws because the vector potential $\mathbf{A}$ bears no relation to the potential energy. Unlike systems of motionless charges where equilibrium is attained due to the balance of Coulomb forces, the equilibrium of orbiting charges requires the balance of electric, magnetic, and centrifugal forces. We note also that the very stability problem in magnetostatics is rather artificial: only such perturbations of $\mathbf{j}$ are relevant that leave $\mathbf{j}$ stationary. It is reasonable to address the stability problem in the wider context of general electrodynamical systems.

Problem 4.1.1. Solve the one- and two-dimensional, $D = 1$ and $D = 2$, Poisson equation

$$\nabla^2 \phi(\mathbf{r}) = -e \, \Omega_{D-1} \, \delta^D(\mathbf{r}), \quad \Omega_{D-1} = 2 \, \frac{\pi^{D/2}}{\Gamma\,(D/2)} \,, \tag{4.60}$$

where Ω_{D-1} is the area of a $(D-1)$-dimensional unit sphere ($\Omega_0 = 2$, $\Omega_1 = 2\pi$), by the two methods presented in the text, and find the field strength $\mathbf{E}$.

Answer

$$\phi = -e \begin{cases} |\,x| & D = 1, \\ \log\,(r/l) & D = 2, \end{cases} \tag{4.61}$$

$$\mathbf{E} = e \begin{cases} \operatorname{sgn}(x) & D = 1, \\ \mathbf{n}/r & D = 2, \end{cases} \tag{4.62}$$

where l is an arbitrary parameter which has the dimensionality of length, and sgn is the signum function: $\operatorname{sgn}(x) = 1$ for $x > 0$, and $\operatorname{sgn}(x) = -1$ for $x < 0$.

Problem 4.1.2. Derive the relation $\mathbf{B} = \nabla \times \mathbf{A}$ by a direct integration of equation $\nabla \cdot \mathbf{B} = 0$ in an unbounded spatial region.

Hint The Fourier transform of this equation $\mathbf{k} \cdot \widetilde{\mathbf{B}} = 0$ means geometrically that $\widetilde{\mathbf{B}}$ belongs to a plane perpendicular to the vector $\mathbf{k}$. Any vector of this plane can be viewed as the projection of some vector $\widetilde{\mathbf{A}}$ onto this plane achieved by the operation $\mathbf{k} \times \widetilde{\mathbf{A}}$.

Problem 4.1.3. Prove the Helmholtz theorem formulated in Sect. 3.1. Does this theorem remain valid if the Euclidean metric is substituted by an indefinite metric?

Problem 4.1.4. A massive scalar field $\varPhi$ is governed by the *Klein–Gordon equation*

$$(\Box + \mu^2)\, \varPhi = J \tag{4.63}$$

where μ is the field mass (in the natural units), and J is an external source. In the static case, $\varPhi$ and J are time-independent, and we arrive at the *Helmholtz equation*

$$(\nabla^2 - \mu^2)\, \varPhi(\mathbf{x}) = -J(\mathbf{x}) \,. \tag{4.64}$$

Find the solution to the Helmholtz equation with a delta-function source

$$(\nabla^2 - \mu^2)\, \varPhi(\mathbf{x}) = -g\, \delta^3(\mathbf{x}) \tag{4.65}$$

where g is a coupling constant. This solution is known as the *Yukawa potential.* Compare the spatial behavior of the Yukawa and Coulomb potentials.

Answer

$$\varPhi = g\, \frac{e^{-\mu r}}{4\pi r} \,. \tag{4.66}$$

Problem 4.1.5. *Static magnetic monopole.* Let a static point particle with a magnetic charge $e^\star$ be at the origin $\mathbf{r} = \mathbf{0}$. It generates a magnetic field $\mathbf{B}(\mathbf{r})$ described by the equation

$$\operatorname{div} \mathbf{B} = 4\pi e^\star \delta^3(\mathbf{r}) \,. \tag{4.67}$$

The introduction of the vector potential according to the relation

$$\mathbf{B} = \operatorname{curl} \mathbf{A} \tag{4.68}$$

is ill advised because of the identity $\operatorname{div} \operatorname{curl} \mathbf{A} = 0$, valid for any smooth vector function $\mathbf{A}$. Nevertheless, Dirac proposed to use (4.68) for vector potentials

$\mathbf{A}$ singular on some line that issues out of the magnetic monopole (the so-called *Dirac string*). The idea is that the divergence of $\mathbf{B}$ vanishes almost everywhere, hence (4.68) holds almost everywhere, and yet the total flux of $\mathbf{B}$, concentrated at the singular line, is $4\pi e^\star$ to suit (4.67). The string may be conceived as a thin solenoid that carries the magnetic flux $4\pi e^\star$ to infinity.

Show that the application of the curl operator to the vector potential

$$\mathbf{A} = e^\star \, \frac{\mathbf{r} \times \hat{\mathbf{u}}}{r(r - \mathbf{r} \cdot \hat{\mathbf{u}})} \,, \tag{4.69}$$

where $\hat{\mathbf{u}}$ is an auxiliary unit vector forming the singular line $\mathbf{x}(\sigma) = \hat{\mathbf{u}}\,\sigma$, $0 \le \sigma < \infty$, gives the magnetic field $\mathbf{B}$ of the Coulomb-like form

$$\mathbf{B} = e^\star \, \frac{\mathbf{n}}{r^2} \,, \tag{4.70}$$

where $\mathbf{n}$ is a unit vector aligned with the radius vector, hence this $\mathbf{B}$ satisfies (4.67). Rewrite the expression (4.69) in spherical coordinates when (i) $\hat{\mathbf{u}}$ is aligned with the x^3-axis and (ii) opposed to it. Consider

$$\mathbf{A}' = e^\star \, \frac{(\mathbf{r} \times \hat{\mathbf{u}})\,(\mathbf{r} \cdot \hat{\mathbf{u}})}{r\,[r^2 - (\mathbf{r} \cdot \hat{\mathbf{u}})^2]} \,. \tag{4.71}$$

Is $\mathbf{A}'$ another vector potential which gives the magnetic field $\mathbf{B}$ defined in (4.70)? What is the relation between $\mathbf{A}'$ and $\mathbf{A}$?

Answer

$$\text{(i)} \quad \mathbf{A}_r = \mathbf{A}_\vartheta = 0, \quad \mathbf{A}_\varphi = -\frac{e^\star}{r \sin\vartheta}\,(1 + \cos\vartheta) \,, \tag{4.72}$$

$$\text{(ii)} \quad \mathbf{A}_r = \mathbf{A}_\vartheta = 0, \; \mathbf{A}_\varphi = \frac{e^\star(1 - \cos\vartheta)}{r \sin\vartheta} = \frac{e^\star \sin\vartheta}{r(1 + \cos\vartheta)} = \frac{e^\star}{r} \tan\frac{\vartheta}{2} \,. \tag{4.73}$$

It is clear from (4.71) that

$$\mathbf{A}' = \mathbf{A} - e^\star \, \frac{\mathbf{r} \times \hat{\mathbf{u}}}{r^2 - (\mathbf{r} \cdot \hat{\mathbf{u}})^2} \,. \tag{4.74}$$

The second term is equal to half the difference between $\mathbf{A}$ depending on some auxiliary vector $\hat{\mathbf{u}}$ and $\mathbf{A}$ depending on the oppositely directed auxiliary vector,

$$e^\star \, \frac{\mathbf{r} \times \hat{\mathbf{u}}}{r(r - \mathbf{r} \cdot \hat{\mathbf{u}})} + e^\star \, \frac{\mathbf{r} \times \hat{\mathbf{u}}}{r(r + \mathbf{r} \cdot \hat{\mathbf{u}})} = 2e^\star \, \frac{\mathbf{r} \times \hat{\mathbf{u}}}{r^2 - (\mathbf{r} \cdot \hat{\mathbf{u}})^2} \,. \tag{4.75}$$

Since both are vector potentials producing the same magnetic field, one may think of $\mathbf{A}'$ as resulting from $\mathbf{A}$ through the gauge transformation (4.43) with some gauge function χ (for the concrete form of χ see Problem 4.8.2).

4.2 Solutions to Maxwell's Equations: Some General Observations

The electromagnetic field generated by charges moving along arbitrary world lines is described by Maxwell's equations (3.64)–(3.67). We write them again:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}\,, \tag{4.76}$$

$$\nabla \cdot \mathbf{B} = 0\,, \tag{4.77}$$

$$\nabla \times \mathbf{B} = 4\pi \mathbf{j} + \frac{\partial \mathbf{E}}{\partial t}\,, \tag{4.78}$$

$$\nabla \cdot \mathbf{E} = 4\pi \varrho\,. \tag{4.79}$$

The system of equations (4.76) through (4.79) is also seemingly overdetermined: eight equations, two vector and two scalar, are intended for finding six functions **E** and **B**.

The three-dimensional vector form of Maxwell's equations (4.76)–(4.79) is not quite convenient for the subsequent analysis. One may rewrite them in terms of Cartan's differential forms, (3.53) and (3.52),

$$dF = 0\,, \tag{4.80}$$

$$d\,{}^*F = 4\pi J\,, \tag{4.81}$$

or, alternatively, in the four-dimensional tensor form, (3.46) and (3.45),

$$\partial_\lambda {}^*F^{\lambda\mu} = 0\,, \tag{4.82}$$

$$\partial_\lambda F^{\lambda\mu} = 4\pi j^\mu\,, \tag{4.83}$$

where ${}^*F^{\lambda\mu} = \frac{1}{2}\,\epsilon^{\lambda\mu\nu\rho}F_{\nu\rho}$.

The identity $dd = 0$ suggests that

$$F = dA\,, \tag{4.84}$$

with $A = A_\mu dx^\mu$ being an arbitrary 1-form, is the general solution to (4.80). The four-vector $A_\mu = A_\mu(x)$ is named the *vector potential* of the electromagnetic field. In coordinate notation, (4.84) takes the form:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu\,. \tag{4.85}$$

This construction identically satisfies (4.82) because $\epsilon^{\lambda\mu\nu\rho}\partial_\lambda\partial_\mu = 0$.

The corresponding solution to equations (4.76) and (4.77) is not as much evident. We express A^μ through its components in some Lorentz frame of reference $A^\mu = (\phi, \mathbf{A})$, or, equivalently, $A_\mu = (\phi, -\mathbf{A})$. Taking into account the definitions of the electric field $E_i = F_{0i}$ and the magnetic induction $B_i = -\frac{1}{2}\,\epsilon_{ijk}F^{jk}$ we obtain from (4.85)

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla\phi\,, \tag{4.86}$$

$$\mathbf{B} = \nabla \times \mathbf{A}\,. \tag{4.87}$$

One can then verify that (4.86) and (4.87) are solutions of equations (4.76) and (4.77) by inspection.

In order to build the 1-form A from a given 2-form F (or, in other words, to gain insight into the structure of the operator d^{-1} inverse to the exterior derivative d), we write the Fourier transform

$$F(x) = \frac{1}{(2\pi)^4}\int d^4k\, e^{-ik\cdot x}\,\widetilde{F}(k) \tag{4.88}$$

where $k \cdot x = \omega t - \mathbf{k}\cdot\mathbf{r}$. We decompose $\widetilde{F}$ into exterior products of basis vectors:

$$\widetilde{F} = \widetilde{F}^{01}e_0\wedge e_1 + \widetilde{F}^{02}e_0\wedge e_2 + \widetilde{F}^{03}e_0\wedge e_3 + \widetilde{F}^{12}e_1\wedge e_2 + \widetilde{F}^{13}e_1\wedge e_3 + \widetilde{F}^{23}e_2\wedge e_3\,. \tag{4.89}$$

Let e_0^μ be taken as $e_0^\mu = k^\mu$, while e_1^μ, e_2^μ and e_3^μ are arbitrary vectors which together with e_0^μ form a not necessarily orthonormalized basis. Then (4.80) reads

$$k \wedge \widetilde{F} = 0\,. \tag{4.90}$$

Taking into account decomposition (4.89) and the identity $k\wedge k = 0$, we obtain

$$\widetilde{F}^{12}e_0\wedge e_1\wedge e_2 + \widetilde{F}^{13}e_0\wedge e_1\wedge e_3 + \widetilde{F}^{23}e_0\wedge e_2\wedge e_3 = 0\,. \tag{4.91}$$

Exterior multiplication by e_3 yields $\widetilde{F}^{12}e_0\wedge e_1\wedge e_2\wedge e_3 = 0$, whence we see that $\widetilde{F}^{12} = 0$. Likewise, $\widetilde{F}^{13} = \widetilde{F}^{23} = 0$. Therefore,

$$\widetilde{F} = e_0 \wedge (\widetilde{F}^{01}e_1 + \widetilde{F}^{02}e_2 + \widetilde{F}^{03}e_3)\,, \tag{4.92}$$

or

$$\widetilde{F} = k \wedge \widetilde{A}\,. \tag{4.93}$$

Clearly (4.93) is just the Fourier image of (4.84). When (4.92) is compared with (4.93), it is apparent which components of $\widetilde{F}$ contribute to $\widetilde{A}$.

For other explicit forms of the operator d^{-1} see Problem 4.2.4 and Appendix A.

Formula (4.84) defines the 1-form A up to adding the external derivative of a scalar function χ. Indeed, the 2-form $F = dA$ is unaffected with respect to the substitution

$$A \to A' = A - d\chi\,, \tag{4.94}$$

or, in the coordinate language, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is invariant under the transformation

$$A_\mu \to A'_\mu = A_\mu - \partial_\mu\chi\,. \tag{4.95}$$

The Fourier transform of (4.95) is

$$\widetilde{A}_\mu \to \widetilde{A}'_\mu = \widetilde{A}_\mu + ik_\mu \widetilde{\chi} \tag{4.96}$$

where $\widetilde{\chi}$ is an arbitrary scalar function of k_μ. The transformation (4.94) [as well as (4.95) and (4.96)] is called a *gauge transformation*. Therefore, we deal with the entire equivalence class of vector potentials related to each other by gauge transformations, rather than a concrete vector function. The terms $-\partial_\mu \chi$ in (4.95) and $ik_\mu \widetilde{\chi}$ in (4.96) are called the *gauge* (or *longitudial*) modes. The evolution of these modes and the evolution of electromagnetic field are divorced from each other. Gauge modes do not contribute to $F_{\mu\nu}$, hence the dynamics of charged particles is unaffected by these modes which are missing from the Lorentz force $ev^\nu F_{\mu\nu}$. We will see presently that the inverse is also the case: the current of charged particles j^μ is not the source of gauge modes.

We insert (4.85) in (4.83) to yield

$$\Box A^\mu - \partial^\mu \partial_\nu A^\nu = 4\pi j^\mu \tag{4.97}$$

where the second-order differential operator

$$\Box = \partial^\mu \partial_\mu = \frac{\partial^2}{\partial t^2} - \nabla^2 \tag{4.98}$$

is called the *wave operator* or the *d'Alembertian*. We are led to the system of equations (4.85) and (4.97) with the number of equations equal to the number of functions sought. This amendment of the system of equations results from augmenting the field degrees of freedom by gauge variables.

Our next task is to solve equation (4.97). The Fourier transform of this equation is

$$\left(k^2 \eta^{\mu\nu} - k^\mu k^\nu\right) \widetilde{A}^\nu = -4\pi \widetilde{j}^\mu \; . \tag{4.99}$$

Let us define the matrix operator

$$\Lambda^{\mu\nu} = \left(k^2 \eta^{\mu\nu} - k^\mu k^\nu\right) . \tag{4.100}$$

If one assumes the existence of an operator Λ^{-1} inverse to Λ, a solution to (4.99) would be $\widetilde{A} = -4\pi \Lambda^{-1} \widetilde{j}$. However, it will be shown that $\det \Lambda = 0$, hence the operator Λ has no inverse.

We first consider $k^2 = 0$. In this case one can find a Lorentz frame such that $k^\mu = \omega\,(1,1,0,0)$, and so

$$\Lambda = -\omega^2 \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} . \tag{4.101}$$

Clearly, $\det \Lambda = 0$ for this matrix. Because the determinant is Lorentz invariant, our statement is true for any frame.

We now turn to $k^2 = \omega^2 > 0$. In this case there exists a Lorentz frame such that $k^\mu = \omega\,(1,0,0,0)$, and so

$$\Lambda = \omega^2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} . \tag{4.102}$$

Therefore, $\det \Lambda = 0$ in this case as well.

We finally take $k^2 = -\omega^2 < 0$. In the frame of reference where $k^\mu = \omega\,(0,1,0,0)$, we verify in the same fashion that $\det \Lambda = 0$.

This result can be obtained in an alternative way. Indeed, $\det \Lambda = \lambda_0\lambda_1\lambda_2\lambda_3$ where λ_0, λ_1, λ_2, and λ_3 are eigenvalues of the operator Λ, that is, solutions to the eigenvalue equation

$$\Lambda_{\mu\nu}\Psi^\nu_A = \lambda_A \Psi^\mu_A \tag{4.103}$$

where A runs from 0 to 3 (there is no summation over A in the right-hand side). It is clear from

$$\Lambda_{\mu\nu}k^\nu = 0 \tag{4.104}$$

that some eigenvalue (associated with the eigenvector Ψ^μ proportional to k^μ) is zero, hence $\det \Lambda = 0$.

One can readily observe that $\Lambda = k^2 \overset{k}{\perp}$ where $\overset{k}{\perp}$ is the operator projecting on directions perpendicular to the vector k^μ. This suggests that equation (4.99) governs the evolution of *transverse* modes of electromagnetic field. The source $\widetilde{j}^\mu$ generates only transverse modes. For the longitudial modes $ik_\mu\widetilde{\chi}$, we have equation (4.104), which shows that such modes are unaffected by $\widetilde{j}^\mu$ and evolve independently of the transverse modes.

Because $\det \Lambda = 0$, Λ^{-1} does not exist, and hence equation (4.97) cannot be solved directly. To tackle this problem one may take advantage of the gauge arbitrariness and impose an additional gauge fixing condition on the vector potential (for an alternative approach see Sect. 4.7). Choosing the so-called *Lorenz gauge fixing condition*[1]

$$\partial_\mu A^\mu = 0 \,, \tag{4.105}$$

(4.97) becomes the *inhomogeneous wave equation*

$$\Box A^\mu = 4\pi j^\mu \,. \tag{4.106}$$

The Fourier transform of this equation

$$k^2 \widetilde{A}^\mu = -4\pi \widetilde{j}^\mu \tag{4.107}$$

[1] The Danish physicist Ludwig Valentin Lorenz, who was the first to invoke this condition in 1867, should not be confused with the Duch physicist Hendrik Antoon Lorentz.

is solvable: $\widetilde{A}^\mu = -4\pi\widetilde{\jmath}^\mu/k^2$. The rest of the problem consists of taking the inverse Fourier transform of $\widetilde{A}^\mu(k)$, and looking for the general solution to the *homogeneous wave equation*

$$\Box A^\mu = 0\,. \tag{4.108}$$

Although the Lorenz gauge is quite convenient, it is by no means the only possible gauge choice. It is sometimes helpful to use the Coulomb gauge (4.44) or the temporal gauge $A^0 = 0$. Given a fixed timelike unit vector n^μ, these gauges can be represented in the formally covariant forms respectively as

$$[\partial_\mu - (n\cdot\partial)\,n_\mu]\,A^\mu = 0 \tag{4.109}$$

and

$$n_\mu A^\mu = 0\,. \tag{4.110}$$

In the frame of reference with the time axis parallel to n^μ, we have $n^\mu = (1,0,0,0)$, and these gauges take the familiar noncovariant form.

Introducing the vector $b_\mu = \xi\,k_\mu - (1-\xi)\,n_\mu$ where ξ is a real parameter, these conditions can be unified in a single formula:

$$b_\mu\widetilde{A}^\mu = 0\,. \tag{4.111}$$

The Lorenz and temporal gauges are regained for respectively $\xi = 1$ and $\xi = 0$, and the Coulomb gauge appears when $\xi = [1+(n\cdot k)]^{-1}$. Imposing the gauge fixing condition (4.111) ensures the solvability of (4.99).

Problem 4.2.1. Show that equation (4.99) becomes solvable when the gauge condition (4.111) is imposed.

Problem 4.2.2. Consider the world line of a charged particle

$$z^\mu(s) = \theta(-s)\,v_{\rm i}^\mu s + \theta(s)\,v_{\rm f}^\mu s \tag{4.112}$$

where $v_{\rm i}^\mu$ and $v_{\rm f}^\mu$ are fixed vectors. This world line describes a single abrupt collision at the origin. This collision should not cause the particle to turn in time, hence $v_{\rm i}\cdot v_{\rm f} > 0$.

Find the Fourier transform of the four-current of this particle.

Answer

$$\widetilde{\jmath}^\mu(k) = -ie\left(\frac{v_{\rm i}^\mu}{k\cdot v_{\rm i}} - \frac{v_{\rm f}^\mu}{k\cdot v_{\rm f}}\right)\,. \tag{4.113}$$

Problem 4.2.3. Find the expression for the vector potential A_μ corresponding to a constant field in terms of the field strength $F_{\mu\nu}$. What is the gauge condition which is therewith imposed on this A_μ?

Answer

$$A_\nu = \frac{1}{2}\,x^\lambda F_{\lambda\nu}\,, \tag{4.114}$$

$$x^\mu A_\mu = 0 \,. \tag{4.115}$$

Equation (4.115) is known as the *Fock–Schwinger gauge.*

Problem 4.2.4. Let z^μ be a smooth vector function of x^α and ξ, varying from $z^\mu = x^\mu$ at $\xi = 0$ to a point separated from x^μ by a large spacelike interval as ξ goes to $-\infty$, so that the electromagnetic field vanishes in this limit. Consider the nonlocal vector construction

$$\mathcal{A}_\mu(x) = \int_{-\infty}^{0} d\xi \, F_{\alpha\beta}(z) \, \frac{\partial z^\alpha}{\partial \xi} \, \frac{\partial z^\beta}{\partial x^\mu}, \quad \text{where} \quad F_{\alpha\beta} = \partial_\alpha A_\beta - \partial_\beta A_\alpha \,. \tag{4.116}$$

(This construction was introduced by Bryce DeWitt in 1962.) Verify the relation

$$\mathcal{A}_\mu = A_\mu - \partial_\mu \chi, \quad \text{where} \quad \chi(x) = \int_{-\infty}^{0} d\xi \, A_\alpha(z) \, \frac{\partial z^\alpha}{\partial \xi} \,, \tag{4.117}$$

which implies that $\mathcal{A}_\mu$ is related to A_μ by a gauge transformation. Therefore, $\mathcal{A}_\mu$ is a vector potential expressed in terms of the field strength $F_{\alpha\beta}$.

Hint

$$\begin{aligned}
\left(\frac{\partial A_\beta}{\partial z^\alpha} - \frac{\partial A_\alpha}{\partial z^\beta} \right) \frac{\partial z^\alpha}{\partial \xi} \, \frac{\partial z^\beta}{\partial x^\mu} &= \frac{\partial A_\beta}{\partial \xi} \, \frac{\partial z^\beta}{\partial x^\mu} - \frac{\partial A_\alpha}{\partial x^\mu} \, \frac{\partial z^\alpha}{\partial \xi} \\
= \frac{\partial}{\partial \xi} \left(A_\beta \, \frac{\partial z^\beta}{\partial x^\mu} \right) - A_\beta \, \frac{\partial}{\partial \xi} \frac{\partial z^\beta}{\partial x^\mu} &- \frac{\partial}{\partial x^\mu} \left(A_\alpha \, \frac{\partial z^\alpha}{\partial \xi} \right) + A_\alpha \, \frac{\partial}{\partial x^\mu} \frac{\partial z^\alpha}{\partial \xi} \\
&= \frac{\partial}{\partial \xi} \left(A_\beta \, \frac{\partial z^\beta}{\partial x^\mu} \right) - \frac{\partial}{\partial x^\mu} \left(A_\alpha \, \frac{\partial z^\alpha}{\partial \xi} \right) \,.
\end{aligned} \tag{4.118}$$

Integration of this equation over ξ from $-\infty$ to 0 gives

$$\mathcal{A}_\mu(x) = A_\beta(z) \, \frac{\partial z^\beta}{\partial x^\mu} \bigg|_{\xi=-\infty}^{\xi=0} - \frac{\partial}{\partial x^\mu} \int_{-\infty}^{0} d\xi \, A_\alpha(z) \, \frac{\partial z^\alpha}{\partial \xi} = A_\mu(x) - \frac{\partial}{\partial x^\mu} \, \chi(x) \,. \tag{4.119}$$

4.3 Free Electromagnetic Field

In regions free of charged matter, the electromagnetic field is governed by the equations

$$\nabla \times \mathbf{B} = \frac{\partial \mathbf{E}}{\partial t} \,, \tag{4.120}$$

$$\nabla \cdot \mathbf{B} = 0 \,, \tag{4.121}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \,, \tag{4.122}$$

$$\nabla \cdot \mathbf{E} = 0 \, . \tag{4.123}$$

Such a field is called *free* (like the free particle in the absence of forces).

At first glance, this system of equations may appear overdetermined. However, a closer inspection shows that the number of independent equations is equal to the number of unknown variables. Indeed, taking the curl of the left hand side of (4.120), we find

$$\epsilon_{ijk}\partial_j\epsilon_{klm}\partial_l B_m = \left(\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}\right)\partial_j\partial_l B_m = \partial_i\partial_j B_j - \partial_l\partial_l B_i \, . \tag{4.124}$$

By (4.121), the term $\partial_i\partial_j B_j$ may be dropped, and so

$$\nabla \times \nabla \times \mathbf{B} = -\nabla^2\mathbf{B} = \frac{\partial}{\partial t}\nabla \times \mathbf{E} \, . \tag{4.125}$$

Differentiating (4.122) with respect to t and comparing the result with (4.125), we obtain

$$\frac{\partial^2\mathbf{B}}{\partial t^2} - \nabla^2\mathbf{B} = 0 \, . \tag{4.126}$$

Likewise, equations (4.122), (4.123) and (4.120) lead to

$$\frac{\partial^2\mathbf{E}}{\partial t^2} - \nabla^2\mathbf{E} = 0 \, . \tag{4.127}$$

Thus the free field satisfies the homogeneous wave equation

$$\Box\, F_{\mu\nu} = 0 \, . \tag{4.128}$$

To confirm this statement, one may derive (4.128) directly from Maxwell's equations in the tensor form (Problem 4.3.1). We see that the system of Maxwell's equations with $j^\mu = 0$ is equivalent to the system of equations (4.128) where the number of equations equals the number of variables. Note that we came to this result without resort to the vector potential.

The Maxwell equations (4.120)–(4.123), as well as the homogeneous wave equations (4.126), (4.127), or (4.128), apply not only to regions empty of charged matter. They are suited for the description of electromagnetic fields in a world with no charges at all. We now show that the homogeneous wave equation has nontrivial solutions. This suggests that variable electromagnetic fields are possible in the world devoid of charges, unlike static fields governed by the Laplace equation. Mathematically, this distinction is due to the fact that the wave equation is hyperbolic while the Laplace equation is elliptic.

Let us consider a free field propagating in some direction, say, the x-axis. The wave equation describing it is

$$\left(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2}\right) F = 0 \, . \tag{4.129}$$

The second-order differential operator in the parenthesis admits the factorization:

$$\left(\frac{\partial}{\partial t}-\frac{\partial}{\partial x}\right)\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}\right)F=0\,. \tag{4.130}$$

We introduce new variables x_- and x_+ referred to as the *light cone variables*,

$$x_-=t-x,\quad x_+=t+x\,. \tag{4.131}$$

The transformation to the initial variables is

$$t=\frac{1}{2}\,(x_++x_-),\quad x=\frac{1}{2}\,(x_+-x_-)\,. \tag{4.132}$$

The differential of the function F can be written in two equivalent forms:

$$dF=F_t dx_0+F_x dx_1=F_- dx_-+F_+ dx_+=\frac{1}{2}F_t(dx_++dx_-)+\frac{1}{2}F_x(dx_+-dx_-)\,, \tag{4.133}$$

where subscripts of F denote partial derivatives with respect to the variables shown. Equating coefficients of dx_- and dx_+,

$$\frac{\partial}{\partial x_-}=\frac{1}{2}\left(\frac{\partial}{\partial t}-\frac{\partial}{\partial x}\right),\qquad \frac{\partial}{\partial x_+}=\frac{1}{2}\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}\right)\,. \tag{4.134}$$

Applying these expressions to (4.130) gives

$$\frac{\partial^2 F}{\partial x_-\partial x_+}=0\,. \tag{4.135}$$

The generic solution to this equation is

$$F=\Phi(x_-)+\Psi(x_+) \tag{4.136}$$

where Φ and Ψ are arbitrary functions. In terms of the initial arguments, we obtain

$$F=\Phi(t-x)+\Psi(t+x)\,. \tag{4.137}$$

We may put $\Psi=0$. Then $\Phi(t-x)$ represents the profile of the wave which is initially prescribed by $\Phi(-x)$. This profile moves as a whole along the x-axis with the speed of light. Indeed, $\Phi(\xi)$ takes equal values for any t and x when the phase $\xi=t-x$ is fixed. Therefore, fixing the phase $\xi=\xi_*$, a point of the profile traces out the world line $x=t-\xi_*$. The term $\Psi(t+x)$ in (4.137) is associated with the wave profile which moves in the direction opposite to the x-axis with the velocity $dx/dt=-1$.

The solutions $\Phi(t-x)$ and $\Psi(t+x)$ are called the *plane waves*. Noteworthy also is the nomenclature *wave packets*, or *solitons*, for sufficiently localized $\Phi(\xi)$ and $\Psi(\xi)$, like $\exp(-\xi^2)$. We reserve the name 'plane wave' for the configurations of the type $\Phi=A\cos k(x-vt)$ and $\Psi=A\cos k(x+vt)$ where A, k and v are arbitrary constants.

Given the initial data $F(0,x)=F_0(x)$ and $F_t(0,x)=V_0(x)$, the general solution to the wave equation (4.129) is

$$F(t,x) = \frac{1}{2}\left[F_0(x-t) + F_0(x+t) + \int_{x-t}^{x+t} ds\, V_0(s)\right] . \tag{4.138}$$

Plane waves can be described in a covariant manner. If the vector potential of a free field A^μ is subject to the Lorenz gauge condition

$$\partial_\mu A^\mu = 0 , \tag{4.139}$$

then, taking $j^\mu = 0$ in (4.106), we obtain the homogeneous wave equation

$$\Box A^\mu = 0 . \tag{4.140}$$

Given the electromagnetic plane wave propagating in some direction of Minkowski space, this direction can be specified by the *propagation vector* k^μ. The vector potential is

$$A_\mu(x) = \epsilon_\mu \Phi(\xi) . \tag{4.141}$$

The quantity $\xi = k \cdot x$ is called the *phase*, and ϵ_μ is called the *polarization vector* of the plane wave. Inserting (4.141) into (4.140) gives

$$k^2 \Phi'' \epsilon_\mu = 0 \tag{4.142}$$

where the prime denotes the derivative with respect to ξ. Because $\Phi(\xi)$ is assumed to be arbitrary, $k^2 = 0$. Therefore, the propagation vector of a free field k^μ is lightlike. Substitution of (4.141) in (4.139) yields

$$(k \cdot \epsilon)\, \Phi' = 0 , \tag{4.143}$$

which implies that the polarization vector is orthogonal to the propagation vector. There exists a Lorentz frame such that $k^\mu = \omega\,(1,1,0,0)$. Any vector orthogonal to k^μ can be cast as $\alpha k^\mu + \beta e_1^\mu + \gamma e_2^\mu$ where α, β and γ are arbitrary constants, $e_1^\mu = (0,0,1,0)$, and $e_2^\mu = (0,0,0,1)$. Terms proportional to k^μ are gauge modes. Therefore, the vector potential of a free field A^μ contains in fact only two polarization degrees of freedom proportional to e_1^μ and e_2^μ.

The field strength of the plane wave is

$$F_{\mu\nu} = (k_\mu \epsilon_\nu - k_\nu \epsilon_\mu)\, \Phi' . \tag{4.144}$$

It immediately follows that $F_{\mu\nu}{}^*F^{\mu\nu} = 0$. Moreover, $F_{\mu\nu}F^{\mu\nu} = 2\,[-(k\cdot\epsilon)^2 + k^2]\,{\Phi'}^2 = 0$, because $k \cdot \epsilon = 0$ and $k^2 = 0$. The vanishing of both invariants of electromagnetic field $\mathcal{P}$ and $\mathcal{S}$ for the plane wave implies that the vector $\mathbf{E}$ is equal in magnitude and orthogonal to the vector $\mathbf{B}$. Given a reference frame in which the time axis is aligned with a timelike vector n^μ, the propagation vector is $k^\mu = (|\mathbf{k}|, \mathbf{k})$, and hence the plane wave propagates in the spatial direction $\mathbf{k}$. It is clear from (4.144) that $n^\mu F_{\mu\nu}\, k^\nu = 0$ and $n^\mu\, {}^*F_{\mu\nu}\, k^\nu = 0$, which shows that $\mathbf{E} \cdot \mathbf{k} = 0$ and $\mathbf{B} \cdot \mathbf{k} = 0$. Thus $\mathbf{E}$ and $\mathbf{B}$ are orthogonal to $\mathbf{k}$. The electromagnetic field of a plane wave is said to be transverse.

One may choose $\Phi(\xi)$ to be periodic, for example, $A\cos\xi$. This choice provides an important special case of the plane wave, the so-called *monochromatic* or *simple harmonic* wave. It was seen before as a harmonic mode $\exp(-ik\cdot x)$ in the four-dimensional Fourier transform (4.88).

Let a simple harmonic wave be propagating along the x-axis. We write $k = 2\pi/\lambda$ and call λ the *wave length*. At the initial time $t = 0$, we have $\Phi(\xi) = A\cos(2\pi x/\lambda)$. The phase change $\Delta\xi = 2\pi n$, with integer n, leading to regular repetition of values of Φ, corresponds to the x-increment $\Delta x = \lambda n$, hence the name of λ. Because the phase ξ is dimensionless, λ has the dimension of length, and k has the dimension of inverse length.

We now look into the general configuration of the electromagnetic field in a world with no charges. The solution of the free field equation (4.140) is simple because each Fourier mode propagates independently of the others. Nevertheless, it is instructive to find a formal solution in closed form in terms of initial condition.

Take a spacelike hyperplane Σ_* as the initial hypersurface. We consider a Lorentz frame where the normal to Σ_* is $n^\mu = (1,0,0,0)$, and the section of Minkowski space Σ_* corresponds to the instant $x_0 = 0$. For the homogeneous wave equation (4.140), the Cauchy problem can be posed as follows: find A^λ throughout Minkowski space in terms of the Cauchy data, the field $A^\lambda|_{\Sigma_*} = A^\lambda_*$ and its normal derivative $(n\cdot\partial)A^\lambda|_{\Sigma_*} = B^\lambda_*$ on the initial surface Σ_*. Let us show that the solution to this problem is

$$A^\lambda(y) = \int d^3x \left[D(y-x)\,B^\lambda_*(x) - \frac{\partial D(y-x)}{\partial x_0}\,A^\lambda_*(x)\right] \tag{4.145}$$

or, in the covariant form, taking into account that $\partial/\partial x_0 = n^\mu\,\partial/\partial x^\mu$,

$$A^\lambda(y) = \int_{\Sigma_*} d^3x\, n^\mu D(y-x)\left(\frac{\partial}{\partial x^\mu} - \frac{\overleftarrow{\partial}}{\partial x^\mu}\right) A^\lambda(x)\;, \tag{4.146}$$

where $D(x)$ is the so-called *fundamental solution* or *Green's function* of the homogeneous wave equation:

$$\Box\, D(x) = 0\;, \tag{4.147}$$

obeying the initial-value conditions

$$D(\mathbf{x},t)|_{t=0} = 0, \qquad \left.\frac{\partial D(\mathbf{x},t)}{\partial t}\right|_{t=0} = \delta^3(\mathbf{x})\;. \tag{4.148}$$

Applying the d'Alembert operator $\Box_y$ to (4.146) and using (4.147), we find that $A^\lambda(y)$ is a solution to the homogeneous wave equation. We next must verify that (4.146) is adapted to the Cauchy data on Σ_*.

It is helpful to employ the identity

$$\left(\frac{\partial}{\partial x_\mu} + \frac{\overleftarrow{\partial}}{\partial x_\mu}\right)\left(\frac{\partial}{\partial x^\mu} - \frac{\overleftarrow{\partial}}{\partial x^\mu}\right) = \Box - \overleftarrow{\Box}\;. \tag{4.149}$$

Leibnitz's rule for differentiation of the product of two functions can be represented as

$$\frac{\partial}{\partial x}\, fg = \frac{\partial f}{\partial x}\, g + f\, \frac{\partial g}{\partial x} = f\left(\frac{\partial}{\partial x} + \frac{\overleftarrow{\partial}}{\partial x}\right) g\,. \tag{4.150}$$

With these observations,

$$\begin{aligned}
&\frac{\partial}{\partial x_\mu}\left[D(y-x)\left(\frac{\partial}{\partial x^\mu} - \frac{\overleftarrow{\partial}}{\partial x^\mu}\right) A^\lambda(x)\right] \\
&= D(y-x)\left(\frac{\partial}{\partial x^\mu} + \frac{\overleftarrow{\partial}}{\partial x^\mu}\right)\left(\frac{\partial}{\partial x^\mu} - \frac{\overleftarrow{\partial}}{\partial x^\mu}\right) A^\lambda(x) \\
&= D(y-x)\left(\Box_x - \overleftarrow{\Box}_x\right) A^\lambda(x) = 0\,,
\end{aligned} \tag{4.151}$$

because $\Box\, A = 0$ and $\Box\, D = 0$. Upon integrating (4.151) over a domain of Minkowski space bounded by two parallel spacelike hyperplanes Σ_* and Σ and applying the Gauss–Ostrogradskiĭ theorem which transforms this integral to the boundary integral, we obtain

$$\left(\int_\Sigma - \int_{\Sigma_*}\right) d^3x\; n^\mu D(y-x)\left(\frac{\partial}{\partial x^\mu} - \frac{\overleftarrow{\partial}}{\partial x^\mu}\right) A^\lambda(x) = 0\,. \tag{4.152}$$

We have assumed that $D(y-x)$ and $A^\lambda(x)$ are such that their product rapidly decreases at spatial infinity, and the integral over the infinitely distant timelike boundary vanishes. The hyperplane Σ may be chosen to contain the point y, which, in view of (4.148), gives

$$\begin{aligned}
&\int_\Sigma d^3x\, n^\mu \left[D(y-x)\,\frac{\partial A^\lambda(x)}{\partial x^\mu} - \frac{\partial D(y-x)}{\partial x^\mu}\, A^\lambda(x)\right] \\
&= \int d^3x \left[D(\mathbf{y}-\mathbf{x},0)\,\frac{\partial A^\lambda(x)}{\partial x^0} - \frac{\partial D(\mathbf{y}-\mathbf{x}, y^0 - x^0)}{\partial x^0}\Big|_{y^0=x^0} A^\lambda(x)\right] \\
&= \int d^3x\, \delta^3(\mathbf{y}-\mathbf{x}) A^\lambda(x^0,\mathbf{x})\Big|_{x^0=y^0} = A^\lambda(y^0,\mathbf{y})\,.
\end{aligned} \tag{4.153}$$

Combining this result with (4.152),

$$A^\lambda(y) = \int_{\Sigma_*} d^3x\, n^\mu \left[D(y-x)\,\frac{\partial A^\lambda(x)}{\partial x^\mu} - \frac{\partial D(y-x)}{\partial x^\mu}\, A^\lambda(x)\right]\,. \tag{4.154}$$

This completes the proof of the statement that (4.146) is a solution to the Cauchy problem.

The solution (4.145) is unique. To see this, let two solutions A_1^λ and A_2^λ be available. Consider their difference $a^\lambda = A_2^\lambda - A_1^\lambda$. It satisfies the wave equation $\Box\, a^\lambda = 0$. But the initial data a_*^λ and b_*^λ for a^λ vanish, and (4.145) shows that $a^\lambda(x) = 0$ for all x.

We now determine the explicit form of $D(x)$. We make the three-dimensional Fourier transform of this function:

$$D(\mathbf{r}, t) = \frac{1}{(2\pi)^3} \int d^3k\, e^{i\mathbf{k}\cdot\mathbf{r}}\, \widetilde{D}_{\mathbf{k}}(t)\ . \tag{4.155}$$

Then (4.147) reduces to the ordinary differential equation

$$\widetilde{D}''_{\mathbf{k}} + k^2 \widetilde{D}_{\mathbf{k}} = 0 \tag{4.156}$$

where $k = |\mathbf{k}|$, and the prime stands for the derivative with respect to t. The initial-value conditions (4.148) take the form

$$\widetilde{D}_{\mathbf{k}}(0) = 0, \quad \widetilde{D}'_{\mathbf{k}}(0) = 1\ . \tag{4.157}$$

The solution to the differential equation (4.156) with initial data (4.157) is

$$\widetilde{D}_{\mathbf{k}}(t) = \frac{\sin kt}{k}\ . \tag{4.158}$$

We insert (4.158) into (4.155). We use spherical coordinates in k-space with the k_3-axis aligned with $\mathbf{r}$, then $d^3k = k^2 dk\, \sin\vartheta\, d\vartheta\, d\varphi$ and $\mathbf{k}\cdot\mathbf{r} = kr\cos\vartheta$. The integrand is independent of φ, and so

$$D(\mathbf{r}, t) = \frac{2\pi}{(2\pi)^3} \int_0^\infty dk\, k\, \sin kt \int_0^\pi d\vartheta \sin\vartheta\, e^{ikr\cos\vartheta}\ . \tag{4.159}$$

The last integral is straightforward:

$$\int_0^\pi d\vartheta \sin\vartheta\, e^{ikr\cos\vartheta} = \int_{-1}^1 d\zeta\, e^{ikr\zeta} = 2\,\frac{\sin kr}{kr}\ . \tag{4.160}$$

To calculate the integral over k, we observe that the integrand

$$2\sin kr \sin kt = \cos k(t-r) - \cos k(t+r) \tag{4.161}$$

is an even function of k. Therefore, the integration may be extended to the negative semiaxis,

$$\begin{aligned} D(\mathbf{r}, t) &= \frac{1}{(2\pi)^2}\,\frac{1}{2r} \int_{-\infty}^\infty dk\,[\cos k(t-r) - \cos k(t+r)] \\ &= \frac{1}{2\pi}\,\frac{1}{2r}\,\frac{1}{2\pi} \int_{-\infty}^\infty dk \left(e^{ik(t-r)} - e^{ik(t+r)}\right) = \frac{1}{2\pi}\,\frac{1}{2r}\,[\delta(t-r) - \delta(t+r)]\ . \end{aligned} \tag{4.162}$$

Finally,

$$D(\mathbf{r},t) = \frac{1}{2\pi}\,\frac{1}{2r}\left[\delta(t-r) - \delta(t+r)\right] . \tag{4.163}$$

Taking into account equation $(F.21)$ of Appendix F,

$$\mathrm{sgn}(t)\,\delta(t^2 - r^2) = \frac{1}{2r}\left[\delta(t-r) - \delta(t+r)\right] , \tag{4.164}$$

where sgn(t) is the signum function

$$\mathrm{sgn}(t) = \frac{t}{|t|} = \begin{cases} 1 & \text{if } t > 0, \\ -1 & \text{if } t < 0, \end{cases} \tag{4.165}$$

we rewrite (4.163) in the Lorentz-invariant form:

$$D(x) = \frac{1}{2\pi}\,\mathrm{sgn}(x_0)\,\delta(x^2) . \tag{4.166}$$

We point out that the explicit time coordinate dependence of this expression by no means violates Lorentz invariance, because the sign of x_0 is invariant for timelike and lightlike intervals, and the presence of the delta-function ensures that the interval is lightlike.

Problem 4.3.1. Show that the tensor Maxwell equations without sources $\partial_\lambda {}^*F^{\lambda\mu} = 0$ and $\partial_\lambda F^{\lambda\mu} = 0$ are equivalent to the homogeneous wave equation $\Box\, F_{\mu\nu} = 0$.

Hint Apply the operator $\epsilon_{\mu\nu\alpha\beta}\partial^\nu$ to $\partial_\lambda {}^*F^{\lambda\mu}$, express the contraction of the Levi–Civita symbols $\epsilon_{\mu\nu\alpha\beta}\epsilon^{\lambda\mu\rho\sigma}$ in terms of the Kronecker deltas, and then, among resulting terms, drop those containing $\partial^\lambda F_{\lambda\alpha}$ or $\partial^\nu F_{\nu\beta}$.

Problem 4.3.2. Verify that the relation

$$\Box\,\frac{1}{(x-a)^2} = 0 \tag{4.167}$$

holds everywhere away from the point $x^\mu = a^\mu$.

Problem 4.3.3. Verify that

$$\Phi(t,r) = \frac{1}{r}\,\left[f(t-r) + g(t+r)\right] , \tag{4.168}$$

with $f(x)$ and $g(x)$ being arbitrary functions, is a solution to the wave equation $\Box\,\Phi = 0$ everywhere except the origin $r = 0$. Note that $f(x)$ and $g(x)$ are not necessarily smooth. They may be distributions, for example, a delta-function or one of its derivatives.

Hint Because the radial part of the Laplace operator in spherical coordinates is

$$\frac{1}{r}\,\frac{\partial^2}{\partial r^2}\,r , \tag{4.169}$$

the problem reduces to $(1+1)$-dimensional wave equations for f and g.

Problem 4.3.4. Let Φ be a free massive scalar field governed by the homogeneous Klein–Gordon equation

$$(\Box + \mu^2)\,\Phi = 0\ . \tag{4.170}$$

Show that the simple harmonic wave $\Phi = A\,\exp(-ik\cdot x) = A\,\exp[-i(\omega t - \mathbf{k}\cdot\mathbf{x})]$ with an arbitrary amplitude A and the propagation vector k^μ subjected to some constraint is a solution to this equation. Find this constraint. Determine the group velocity of propagation of this plane wave $\mathbf{v}$.

Answer

$$k^2 = \mu^2, \quad \text{or} \quad \omega = \sqrt{\mathbf{k}^2 + \mu^2}; \quad \mathbf{v} = \nabla_{\mathbf{k}}\,\omega = \frac{\mathbf{k}}{\sqrt{\mathbf{k}^2 + \mu^2}}, \quad |\mathbf{v}| < 1\ . \tag{4.171}$$

Problem 4.3.5. Verify the identity

$$(\partial_\lambda + \overleftarrow{\partial}_\lambda)\,\Gamma^\lambda(\partial, \overleftarrow{\partial}) = (\Box + \mu^2) - (\overleftarrow{\Box} + \mu^2) \tag{4.172}$$

where

$$\Gamma^\lambda(\partial, \overleftarrow{\partial}) = \partial^\lambda - \overleftarrow{\partial}^\lambda\ . \tag{4.173}$$

With this identity, show that the solution to the Cauchy problem for the Klein–Gordon equation and initial-value data for the field Φ and its normal derivative $(n\cdot\partial)\,\Phi$ on a spacelike hyperplane Σ is

$$\Phi(y) = \int_\Sigma d^3x\, n_\mu G(y - x)\,\Gamma^\mu(\partial, \overleftarrow{\partial})\,\Phi(x)\ . \tag{4.174}$$

Here, the relevant Green function $G(x)$ satisfies the equation

$$(\Box + \mu^2)\,G = 0\ . \tag{4.175}$$

and the initial-value conditions which are written in the Lorentz frame with the time axis parallel to n^μ as

$$G(\mathbf{r}, t)|_{t=0} = 0, \quad \frac{\partial}{\partial t}\,G(\mathbf{r}, t)|_{t=0} = \delta^3(\mathbf{r})\ . \tag{4.176}$$

Find $G(x)$ in the explicit form.

Answer

$$G(x) = \frac{1}{2\pi}\,\mathrm{sgn}(x_0)\left[\delta(x^2) - \frac{\mu}{2}\,\theta(x^2)\frac{J_1(\mu\sqrt{x^2})}{\sqrt{x^2}}\right] \tag{4.177}$$

where $J_1(z)$ is the Bessel function of the first kind.

Problem 4.3.6. Find the four-dimensional Fourier transform of $D(x)$ satisfying the homogeneous wave equation (4.147) and initial conditions (4.148).

Answer

$$\widetilde{D}(k) = 2\pi i \,\mathrm{sgn}(k_0)\, \delta(k^2) \ . \tag{4.178}$$

Problem 4.3.7. Let Φ be a scalar field governed by the field equation

$$(\Box + \mu^2)\, \Phi = -U'(\Phi) \tag{4.179}$$

where U is an analytic function $U(\Phi) = c_3\Phi^3 + c_4\Phi^4 + \cdots$, and the prime means the derivative with respect to Φ. Prove that a particular solution to this equation is a function of the argument $\xi = k \cdot x$. Find the relation between the phase ξ and amplitude Φ of this solution.

Answer

$$\xi + \xi_0 = \pm\sqrt{k^2} \int \frac{d\Phi}{\sqrt{2(E^2 - U) - \mu^2\Phi^2}} \tag{4.180}$$

where ξ_0 and E are arbitrary integration constants.

Problem 4.3.8. Two simple examples of the function U in Problem 4.3.7 are

$$u = \frac{1}{4}\,(\phi^2 - 1)^2 \quad \text{and} \quad u = \cos\phi - 1 \tag{4.181}$$

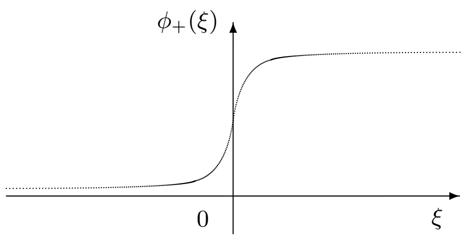
where 'proper units' $\bar{x} = \mu x$, $\bar{t} = \mu t$, $\phi = (\sqrt{\lambda}/\mu)\Phi$, $u = (\mu^4/\lambda)U$ are used. Verify that the solutions of the mentioned type are

$$\phi_\pm = \tanh\left(\pm\frac{1}{\sqrt{2}}\frac{\bar{x} - v\bar{t}}{\sqrt{1 - v^2}}\right) \quad \text{and} \quad \phi_\pm = 4\arctan\left[\exp\left(\pm\frac{\bar{x} - v\bar{t}}{\sqrt{1 - v^2}}\right)\right] . \tag{4.182}$$

These solutions provide an apt illustration of the notion of solitons. Although $\phi_\pm$ do not represent sufficiently localized functions, their derivatives do.

Find the constants k, ξ_0, and E. Depict $\phi_+(\xi)$ schematically.

Answer



4.4 The Retarded Green's Function

In this section, we examine solutions to the inhomogeneous wave equation

$$\Box A^\mu = 4\pi j^\mu \tag{4.183}$$

with the four-current j^μ of the form (3.32). A powerful tool for solving linear differential equations is the *Green's function method.* We define the Green's function $\mathcal{G}(x)$ for the wave operator as a solution to the inhomogeneous wave equation with a delta-function source,

$$\Box\, \mathcal{G}(x) = \delta^4(x)\ . \tag{4.184}$$

This 'elementary' source is nonzero within an infinitesimal domain of a single point x, and the response to such a perturbation is given by $\mathcal{G}(x)$. Since $j^\mu(x)$ may be thought of as the superposition of such elementary sources,

$$j^\mu(x) = \int d^4y\, \delta^4(y-x)\, j^\mu(y)\ , \tag{4.185}$$

it is apparent that the response to the elementary source $\delta^4(y-x)$ multiplied by the factor $4\pi j^\mu(x)$ is $4\pi\mathcal{G}(y-x)\, j^\mu(x)$, and the sum of these responses

$$A^\mu(y) = 4\pi \int d^4y\, \mathcal{G}(y-x)\, j^\mu(x) \tag{4.186}$$

is the desired solution to equation (4.183). Indeed, applying the d'Alembertian $\Box_y$ to both sides of (4.186) and taking into account (4.184), we recover (4.183).

It is clear that the Green's function $\mathcal{G}(x)$ so defined is not unique. Adding to it any solution of the homogeneous wave equation

$$\Box\, \mathcal{G}_0(x) = 0\ , \tag{4.187}$$

gives further solutions to equation (4.184).

To fix $\mathcal{G}(y-x)$, we adopt the *retarded boundary condition* by which $\mathcal{G}(y-x) = 0$ for $(y-x)^2 \geq 0$, $y_0 < x_0$ (note that the second inequality is invariant only subject to the first). The conventional interpretation of this condition is that the signal must be absent until switching on the source: cause precedes response. We will call $\mathcal{G}(x)$ satisfying this condition the *retarded Green's function* and denote it as $D_{\rm ret}(x)$.

We now establish the explicit form of $D_{\rm ret}(x)$. We insert the three-dimensional Fourier transform of this function

$$D_{\rm ret}(t,\mathbf{x}) = \frac{1}{(2\pi)^3}\int d^3k\, e^{i\mathbf{k}\cdot\mathbf{x}}\, \widetilde{D}_{\rm ret}(t,\mathbf{k}) \tag{4.188}$$

and that of the three-dimensional delta-function

$$\delta^3(\mathbf{x}) = \frac{1}{(2\pi)^3}\int d^3k\, e^{i\mathbf{k}\cdot\mathbf{x}} \tag{4.189}$$

in (4.184) to yield

$$\frac{\partial^2 \widetilde{D}_{\rm ret}(t,\mathbf{k})}{\partial t^2} + k^2 \widetilde{D}_{\rm ret}(t,\mathbf{k}) = \delta(t) \tag{4.190}$$

where $k^2 = \mathbf{k}^2$. The function $\widetilde{D}_{\rm ret}(t,\mathbf{k})$ can be represented as

$$\widetilde{D}_{\rm ret}(t,\mathbf{k}) = \theta(t)\,\widetilde{\Delta}(t,\mathbf{k}) \tag{4.191}$$

where the Heaviside step function $\theta(t)$ renders the Green's function retarded. We impose the following initial conditions

$$\widetilde{\Delta}(t,\mathbf{k})\Big|_{t=0} = 0, \qquad \frac{\partial \widetilde{\Delta}(t,\mathbf{k})}{\partial t}\Big|_{t=0} = 1\,, \tag{4.192}$$

whereby the time derivatives of $\widetilde{D}_{\rm ret}(t,\mathbf{k})$, denoted by primes, are

$$\begin{aligned} \widetilde{D}'_{\rm ret}(t,\mathbf{k}) &= \delta(t)\widetilde{\Delta}(t,\mathbf{k})+\theta(t)\widetilde{\Delta}'(t,\mathbf{k}) = \delta(t)\widetilde{\Delta}(0,\mathbf{k})+\theta(t)\widetilde{\Delta}'(t,\mathbf{k}) = \theta(t)\widetilde{\Delta}'(t,\mathbf{k}),\\ \widetilde{D}''_{\rm ret}(t,\mathbf{k}) &= \delta(t)\,\widetilde{\Delta}'(t,\mathbf{k}) + \theta(t)\,\widetilde{\Delta}''(t,\mathbf{k}) = \delta(t) + \theta(t)\,\widetilde{\Delta}''(t,\mathbf{k})\,. \end{aligned} \tag{4.193}$$

It follows that $\widetilde{\Delta}(t,\mathbf{k})$ satisfies the corresponding homogeneous equation

$$\frac{\partial^2 \widetilde{\Delta}(t,\mathbf{k})}{\partial t^2} + k^2 \widetilde{\Delta}(t,\mathbf{k}) = 0\,. \tag{4.194}$$

We already encountered a function obeying a differential equation identical to (4.194) and initial conditions similar to (4.192) [see (4.156) and (4.157)]. Thus

$$\widetilde{\Delta}(t,\mathbf{k}) = \frac{\sin kt}{k}\,. \tag{4.195}$$

Inserting (4.195) in (4.188) and working out the Fourier integral, similar to what was done in Sect. 4.3, we arrive at

$$D_{\rm ret}(t,\mathbf{x}) = \frac{\delta(t-r)}{4\pi r}\,. \tag{4.196}$$

We have omitted the term $\theta(t)\,\delta(t+r)/4\pi r$ whose support is the spacetime point $t = 0$, $r = 0$. We may ignore this part of the response to the delta-function source because it manifests itself only at the location of the source where the response is ill-defined anyway.

It is seen from (4.196) that the signal propagates along rays of the *future light cone*, $t = r$, in other words, the elementary source $\delta^4(x)$ emits a *divergent* spherical wave of infinitesimal width, Fig. 4.1. One might, however, adopt the *advanced condition* and write the factor $\theta(-t)$, in lieu of the retarded condition with the inherent factor $\theta(t)$. This would result in the *advanced Green's function*

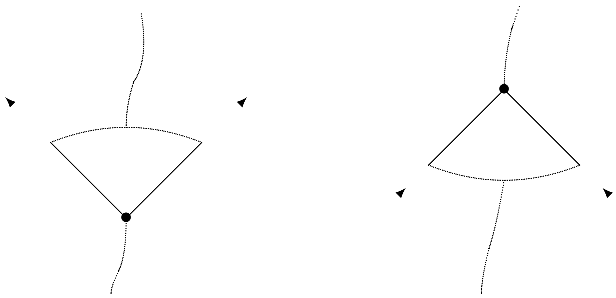


Fig. 4.1. Divergent (*left*) and convergent (*right*) waves

$$D_{\text{adv}}(t, \mathbf{x}) = \frac{\delta(t+r)}{4\pi r} \tag{4.197}$$

suggesting the signal which propagates along rays of the past light cone $t = -r$, that is, the elementary source absorbs a *convergent* spherical wave of infinitesimal width, Fig. 4.1. The picture where spherical waves converge freely to a focal point can readily be imagined: it occurs in a movie whose film shows the reversed order of events. In the normal order of events, such a picture is highly improbable. Hence retardation is not tantamount to causality as might appear at first sight. In deciding between the retarded and advanced, one usually prefers the former being guided by empirical arguments. Retardation implies not only the causal interrelationship but also the fact that *time is unidirectional.*

Using equation (F.20) of Appendix F, expressions (4.196) and (4.197) for the retarded and advanced Green's functions can be cast in a Lorentz-invariant form:

$$D_{\text{ret}}(x) = \frac{1}{2\pi}\,\theta(x_0)\,\delta(x^2)\ , \tag{4.198}$$

$$D_{\text{adv}}(x) = \frac{1}{2\pi}\,\theta(-x_0)\,\delta(x^2)\ . \tag{4.199}$$

It is clear from (4.198) and (4.199) that the retarded and advanced Green's functions obey the reciprocity relation

$$D_{\text{ret}}(-x) = D_{\text{adv}}(x)\ . \tag{4.200}$$

The support of $\delta(x^2)$ is the light cone surface $x^2 = 0$. The presence of $\theta(x_0)$ ensures that $D_{\text{ret}}(x)$ is concentrated in the forward sheet of this light cone C_+,

$$\operatorname{supp} D_{\text{ret}}(x) : x^2 = 0,\ x_0 \geq 0\ , \tag{4.201}$$

and the presence of $\theta(-x_0)$ shows that $D_{\text{adv}}(x)$ is concentrated in the backward sheet of this light cone C_-,

$$\operatorname{supp} D_{\text{adv}}(x) : x^2 = 0,\ x_0 \leq 0\ . \tag{4.202}$$

Is it possible to exploit linear combinations of $D_{\rm ret}$ and $D_{\rm adv}$? The use of the sum of the retarded and advancedfields in the relativistic action-at-a-distance electrodynamics goes back to Hugo Tetrode in 1922, Adriaan Fokker in 1929, and John Wheeler and Richard Feynman in 1945. Yet, the physical sense of any combination of $D_{\rm ret}$ and $D_{\rm adv}$ is still far from clear. Action-at-a-distance electrodynamics will receive some attention in Sect. 10.6, but until then, the retarded condition is understood throughout the text.

We complete this sketch of the Green's function method with the calculation of the retarded electromagnetic field due to a single point charge e moving along an arbitrary timelike world line $z^\mu(s)$. Let x^μ be some point outside the world line. Define

$$R^\mu = x^\mu - z^\mu(s) \tag{4.203}$$

the four-vector drawn from a point $z^\mu(s)$ on the world line, where the signal is emitted, to the point x^μ, where the signal is received. Substituting the retarded Green's function (4.198) and the four-current

$$j^\mu(x) = e\int_{-\infty}^{\infty} ds\, v^\mu(s)\,\delta^4\left[x - z(s)\right] \tag{4.204}$$

in (4.186) and taking the integral over the four-dimensional volume, we obtain

$$A^\mu(x) = 2e\int_{-\infty}^{\infty} ds\, v^\mu(s)\,\theta(R_0)\,\delta(R^2)\ . \tag{4.205}$$

The equation $R^2 = [x - z(s)]^2 = 0$ has two roots with respect to s, $z^\mu(s_{\rm ret})$ and $z^\mu(s_{\rm adv})$. The step function $\theta(X_0)$ selects the root $z^\mu(s_{\rm ret})$ corresponding to the intersection of the world line with the past light cone C_- from the point x^μ, Fig. 4.2.

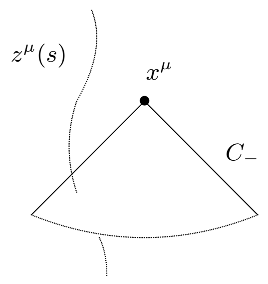


Fig. 4.2. Retarded signal received at x^μ

To integrate (4.205), we use equation (F.18) of Appendix F,

$$\delta[U(\xi)] = \sum_n \frac{1}{|U'(\xi_n)|}\,\delta(\xi - \xi_n) \tag{4.206}$$

where ξ_n are the roots of the equation $U(\xi) = 0$, and note that

$$\frac{dR^2}{ds} = -2R_\mu \frac{dz^\mu}{ds} = -2R \cdot v\ . \tag{4.207}$$

This gives

$$A^\mu = e\,\frac{v^\mu}{R \cdot v}\bigg|_{s=s_{\rm ret}}\ . \tag{4.208}$$

The relation (4.208) can be made more transparent geometrically by the following observations. Let $R^\mu = x^\mu - z^\mu(s_{\rm ret})$ be the lightlike vector drawn from the point on the world line where the signal was emitted $z^\mu(s_{\rm ret})$, to the point x^μ, where the signal was received, and let v^μ be the unit vector tangent to this curve at the point $z^\mu(s_{\rm ret})$. Consider a two-dimensional plane built out of two vectors R^μ and v^μ. Two relevant vectors, the spacelike normalized vector u^μ orthogonal to v^μ, and lightlike vector

$$c^\mu = v^\mu + u^\mu \tag{4.209}$$

can be introduced here, Fig. 4.3. All this can be expressed analytically as

$$v^2 = -u^2 = 1, \quad u \cdot v = 0, \quad c^2 = 0, \quad c \cdot v = -c \cdot u = 1. \tag{4.210}$$

We define the *invariant retarded distance* ρ between the points x^μ and $z^\mu(s_{\rm ret})$ as

$$\rho = -R \cdot u\ . \tag{4.211}$$

It is clear from (4.209) and (4.210), or from Fig. 4.3, that

$$R^\mu = \rho\, c^\mu\ , \tag{4.212}$$

and ρ is represented in another form

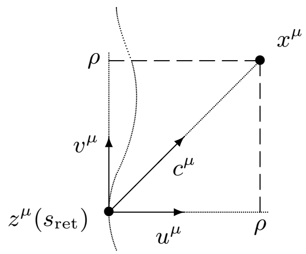


Fig. 4.3. Covariant retarded variables

$$\rho = R \cdot v \,. \tag{4.213}$$

Figure 4.3 shows also that the scalar ρ is the spatial distance between the field point and the retarded point in the instantaneously comoving Lorentz frame in which the charge is at rest at the retarded instant $s_{\rm ret}$, and u^μ is the spatial projection of c^μ.

Finally, the retarded vector potential due to a single arbitrarily moving charge e, called the *Liénard–Wiechert vector potential*, is

$$A^\mu(x) = e\,\frac{v^\mu}{\rho} \,. \tag{4.214}$$

In the general case, the source j^μ is composed of N charges e_I which move along arbitrary world lines $z_I^\mu(s_I)$. The vector potential due to this source is the superposition of contributions from individual charges:

$$A^\mu(x) = \sum_{I=1}^{N} e_I\,\frac{v_I^\mu}{\rho_I} \tag{4.215}$$

where the four-velocities v_I^μ are taken at the retarded instants $s_{I\,\rm ret}$ corresponding to the intersections of the world lines with the past light cone drawn from the observation point x^μ, and ρ_I are the invariant retarded distances between x^μ and $z_I^\mu(s_{I\,\rm ret})$.

Problem 4.4.1. Calculate the four-dimensional Fourier transforms of the retarded and advancedGreen's functions

$$\widetilde{D}_{\rm ret}(k) = \int d^4x\, e^{ikx}\, D_{\rm ret}(x), \quad \widetilde{D}_{\rm adv}(k) = \int d^4x\, e^{ikx}\, D_{\rm adv}(x) \tag{4.216}$$

where $D_{\rm ret}(x)$ and $D_{\rm adv}(x)$ are given respectively by (4.198) and (4.199). Determine positions of singularities of $\widetilde{D}_{\rm ret}(k)$ and $\widetilde{D}_{\rm adv}(k)$ at the complex k_0 plane.

Answer

$$\widetilde{D}_{\rm ret}(k) = -\frac{1}{k^2 + 2ik_0\epsilon}, \quad \widetilde{D}_{\rm adv}(k) = -\frac{1}{k^2 - 2ik_0\epsilon} \,. \tag{4.217}$$

The denominator of $\widetilde{D}_{\rm ret}(k)$ is

$$(k_0 + i\epsilon)^2 - \mathbf{k}^2 \,, \tag{4.218}$$

and both poles

$$k_0 = \pm|\mathbf{k}| - i\epsilon \tag{4.219}$$

are below the line of integration (real axis) in the complex k_0 plane. Thus, the retarded boundary condition satisfied by $D_{\rm ret}(x)$ determines the contour, in the complex k_0 plane, by which the integration is to avoid the poles of the integrand at $k^2 = 0$. By contrast, both poles of $\widetilde{D}_{\text{adv}}(k)$

$$k_0 = \pm|\mathbf{k}| + i\epsilon \tag{4.220}$$

are above the line of integration.

Problem 4.4.2. Prove the relations

$$D_{\text{ret}}(x) - D_{\text{adv}}(x) = D(x), \ D_{\text{ret}}(x) = \theta(x_0)D(x), \ D_{\text{adv}}(x) = -\theta(-x_0)D(x) \tag{4.221}$$

where $D(x)$ is the Green function for the homogeneous wave equation defined in (4.166).

Problem 4.4.3. Find the retarded Green's function $G_{\text{ret}}(x)$ and its Fourier transform $\widetilde{G}_{\text{ret}}(k)$ for the Klein–Gordon operator

$$(\Box + \mu^2)\, G_{\text{ret}}(x) = \delta^4(x) \ . \tag{4.222}$$

Specify the behavior of $G_{\text{ret}}(x)$ in the vicinity of the light cone. Determine the support of $G_{\text{ret}}(x)$ and compare it with the support of the retarded Green's function for the wave operator $D_{\text{ret}}(x)$, equation (4.201).

Answer

$$G_{\text{ret}}(x) = \frac{1}{(2\pi)^4}\int d^4k\, e^{-ikx}\, \widetilde{G}_{\text{ret}}(k), \quad \widetilde{G}_{\text{ret}}(k) = \frac{1}{\mu^2 - k^2 - ik_0\epsilon} \ . \tag{4.223}$$

$$G_{\text{ret}}(x) = \frac{1}{2\pi}\,\theta(x_0)\left[\delta(x^2) - \frac{\mu}{2}\,\theta(x^2)\frac{J_1(\mu\sqrt{x^2})}{\sqrt{x^2}}\right] \ . \tag{4.224}$$

$$G_{\text{ret}}(x) \to \frac{1}{2\pi}\,\theta(x_0)\left[\delta(x^2) - \frac{\mu^2}{4}\,\theta(x^2)\right], \quad x^2 \to 0 \ . \tag{4.225}$$

The leading singularity of $G_{\text{ret}}(x)$ is identical to that of the retarded Green's function in the massless case $\mu = 0$, that is, the retarded Green's function for the wave operator $D_{\text{ret}}(x)$.

$$\operatorname{supp} G_{\text{ret}}(x) : x^2 \geq 0, x_0 \geq 0, \qquad \operatorname{supp} D_{\text{ret}}(x) : x^2 = 0, x_0 \geq 0 \ . \tag{4.226}$$

Problem 4.4.4. We are already aware of the fact that a delta-function source brings into existence a singular field: a point charge is responsible for a simple pole of the Coulomb potential, and the singularity of the Yukawa potential in Problem 4.1.4 is due to the source proportional to $\delta^3(\mathbf{r})$. This raises the question whether the reverse is true, that is, whether any field singularity necessitates the source proportional to $\delta^3(\mathbf{r})$. Consider a spherically symmetric Klein–Gordon field

$$\phi(\mathbf{r}, t) = \theta(r)\,\Phi(r, t) \ . \tag{4.227}$$

Here, Φ is a regular function of r and t, and the step function θ emphasizes that r is positive definite. Note that $\theta(r)$ represents a weak singularity at $r = 0$ because differentiations of $\theta(r)$ give singular distributions. Does this $\theta(r)$ evidence that the mere spherical symmetry of ϕ is sufficient to ensure the existence of its source of the form $\delta^3(\mathbf{r})$?

Answer No.

Hint Recall that the radial part of ∇^2 is

$$\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} \;, \tag{4.228}$$

and compare

$$\theta''(r) = \delta'(r) = -\frac{\delta(r)}{r} \quad \text{and} \quad \frac{2}{r}\,\theta'(r) = \frac{2\delta(r)}{r} \tag{4.229}$$

with $\delta^3(\mathbf{r})$, expressed in terms of spherical coordinates,

$$\delta^3(\mathbf{r}) = \delta(\varphi)\,\frac{1}{\sin\vartheta}\,\delta(\vartheta)\,\frac{1}{r^2}\,\delta(r) \;. \tag{4.230}$$

An additional singular factor $1/r$ is seen in $\delta^3(\mathbf{r})$. Therefore, Φ must be as singular as the Coulomb potential if a $\delta^3(\mathbf{r})$ is to occur. For example, $\Phi = \ln r$ is not singular enough to have its origin in a point source.

Problem 4.4.5. Determine the retarded Green's function $D_{\rm ret}(x)$ for the wave operator in a world with one temporal and one spatial dimension,

$$\Box\, D_{\rm ret}(x) = \delta^2(x) \;. \tag{4.231}$$

Calculate the vector potential A^μ due to an arbitrarily moving charge e with the aid of this $D_{\rm ret}(x)$.

Answer

$$D_{\rm ret}(t,x) = \frac{1}{2}\,\theta(t - |x|), \quad A^\mu = -e\,R^\mu \;. \tag{4.232}$$

4.5 Covariant Retarded Variables

The point $z^\mu(s_{\rm ret})$ is rigidly linked to x^μ through the constraint

$$R^2 = 0 \;. \tag{4.233}$$

Small variations of x^μ entail certain small variations of the instant $s_{\rm ret}$. This is clear from Fig. 4.2: as the vertex x^μ shifts, the past light cone C_- moves along the world line varying the intersection point $z^\mu(s_{ret})$. To derive the precise relation between these variations, we differentiate (4.233) with respect to x^μ:

$$2R_\lambda\, \partial_\mu R^\lambda = 0\,. \tag{4.234}$$

We drop the subscript 'ret' and use the formulas

$$\frac{\partial x^\lambda}{\partial x^\mu} = \delta^\lambda_{\ \mu}, \quad \frac{\partial z^\lambda}{\partial x^\mu} = \frac{dz^\lambda}{ds}\,\frac{\partial s}{\partial x^\mu} \tag{4.235}$$

to yield

$$\partial_\mu R^\lambda = \delta^\lambda_{\ \mu} - v^\lambda \partial_\mu s\,. \tag{4.236}$$

Combining (4.236) with (4.213) and (4.212), we obtain from (4.234)

$$\partial_\mu s = c_\mu\,. \tag{4.237}$$

Relations (4.236) and (4.237) enable one to calculate derivatives of any kinematic quantities, for example,

$$\partial^\mu v^\nu = c^\mu a^\nu\,. \tag{4.238}$$

The derivatives of the retarded distance $\rho = R\cdot v$ are of particular interest. It is easy to see that

$$\partial^\mu \rho = v^\mu + (a\cdot R - 1)\, c^\mu\,. \tag{4.239}$$

We now define a further invariant retarded variable

$$\lambda = a\cdot R - 1\,, \tag{4.240}$$

and so

$$\partial^\mu \rho = v^\mu + \lambda\, c^\mu\,. \tag{4.241}$$

Further differential relations are given in Problems 4.5.1 and 4.5.2.

Invariant retarded variables are useful in constructing the so-called *retarded frame of reference* which proves more suitable for the treatment of retarded fields than the fixed rectilinear coordinate frame. Let $z^\mu(s)$ be a smooth timelike curve and x^μ an observation point off this curve. We draw the past light cone C_- from x^μ to intersect the curve at a point $z^\mu(s_{\rm ret})$, Fig. 4.2, and define the lightlike vector $R^\mu = x^\mu - z^\mu(s_{\rm ret})$. In the retarded frame of reference, the proper time s measures time evolution, R^μ plays the role of the four-dimensional radius vector drawn from the moving origin $z^\mu(s_{\rm ret})$, and the invariant retarded distance ρ offers an alternative to the radial distance. In addition, we introduce the zenith and azimuth angles ϑ and φ by the component decomposition of R^μ in the rest frame corresponding to the proper time $s_{\rm ret}$,

$$R^\mu = \rho\,(1,\, \sin\vartheta\, \cos\varphi,\, \sin\vartheta\, \sin\varphi,\, \cos\vartheta)\,. \tag{4.242}$$

Derivatives of the new coordinates s and ρ with respect to the old ones x^μ are given by formulas (4.237) and (4.241). To find derivatives of $\vartheta = \arccos\,(R^3/\rho)$ and $\varphi = \arctan\,(R^2/R^1)$, we observe that $\partial/\partial x^\mu = \partial/\partial R^\mu$, and so

$$\partial_\mu \vartheta = \frac{\partial_\mu \rho \, \cos\vartheta - \delta^3_{\ \mu}}{\rho \, \sin\vartheta}, \quad \partial_\mu \varphi = \frac{-\delta^1_{\ \mu} \, \sin\varphi + \delta^2_{\ \mu} \, \cos\varphi}{\rho \, \sin\vartheta} \, . \tag{4.243}$$

From (4.237), (4.241), (4.242) and (4.243), we obtain the Jacobian for the transformation from new coordinates to old:

$$\left| \frac{\partial \left(s, \, \rho, \, \vartheta, \, \varphi\right)}{\partial \left(x^0, \, x^1, \, x^2, \, x^3\right)} \right| = \epsilon^{\lambda\mu\nu\sigma} \, \partial_\lambda s \; \partial_\mu \rho \; \partial_\nu \vartheta \; \partial_\sigma \varphi = \frac{1}{\rho^2 \sin\vartheta} \, . \tag{4.244}$$

Thus the four-dimensional volume element is

$$d^4x = ds \, \rho^2 d\rho \, d\Omega, \quad d\Omega = \sin\vartheta \, d\vartheta \, d\varphi \, . \tag{4.245}$$

The integration over a four-dimensional domain $\mathcal{U}$ can be envisioned as the sequence of two actions: we first integrate over the future light cone C_+ with vertex at a point $z^\mu(s)$ belonging to $\mathcal{U}$, and then over the temporal coordinate s which labels the one-parameter family of light cones, each drawn from a point on the given timelike curve. A particular light cone C_+ is fixed by the equation s = const. The normal n^μ to the hypersurface C_+ is the four-dimensional gradient of this equation,

$$n^\mu = \partial^\mu s = c^\mu \, . \tag{4.246}$$

With the four-dimensional volume element (4.245), we find the surface element on the future light cone:

$$C_+ : \quad d\sigma^\mu = c^\mu \rho^2 d\rho \, d\Omega \, . \tag{4.247}$$

On the other hand, the four-dimensional integration may be performed initially over a tube T_ρ which encloses the given curve $z_\mu(s)$ being separated from it by a constant retarded distance ρ, and then over the radial coordinate ρ which labels the family of such tubes. A particular tube T_ρ is fixed by the equation ρ = const. The normal n^μ to the timelike hypersurface T_ρ is

$$n^\mu = \partial^\mu \rho = v^\mu + \lambda \, c^\mu \, . \tag{4.248}$$

The surface element on the tube T_ρ is

$$T_\rho : \quad d\sigma^\mu = \left(v^\mu + \lambda \, c^\mu\right) \rho^2 ds \, d\Omega \, . \tag{4.249}$$

One additional remark on the integration technique is in order. We will frequently deal with expressions homogeneous of some degree in u^μ, the spacelike vector directed from $z^\mu(s_{\rm ret})$ to x^μ, depicted in Fig. 4.3. The integration of such expressions over solid angle Ω is greatly simplified through the formulas

$$\int d\Omega = 4\pi \, , \tag{4.250}$$

$$\int d\Omega\, u_\mu = 0, \quad \int d\Omega\, u_\lambda u_\mu u_\nu = 0, \dots \tag{4.251}$$

$$\int d\Omega\, u_\mu u_\nu = -\frac{4\pi}{3} \overset{v}{\perp}_{\mu\nu} \ . \tag{4.252}$$

$$\int d\Omega\, u_\alpha u_\beta u_\mu u_\nu = \frac{4\pi}{3\cdot 5} \left(\overset{v}{\perp}_{\mu\nu} \overset{v}{\perp}_{\alpha\beta} + \overset{v}{\perp}_{\alpha\mu} \overset{v}{\perp}_{\beta\nu} + \overset{v}{\perp}_{\alpha\nu} \overset{v}{\perp}_{\beta\mu} \right) . \tag{4.253}$$

Consider first the case of odd number of vectors u^μ. We have

$$u^\mu = (0, \sin\vartheta\, \cos\varphi, \sin\vartheta\, \sin\varphi, \cos\vartheta) \, . \tag{4.254}$$

The vanishing of the integrals in (4.251) is explained by the fact that u^μ is orthogonal to the integration surface. Technically, this follows from either

$$\int_0^{2\pi} d\varphi\, (\sin\varphi)^j\, (\cos\varphi)^k = 0 \tag{4.255}$$

or

$$\int_0^{\pi} d\vartheta\, \sin\vartheta\, (\cos\vartheta)^{2l+1} = \int_{-1}^{1} d\zeta\, \zeta^{2l+1} = 0 \tag{4.256}$$

where j, k and l are nonnegative integers such that $j + k > 0$ and $l \geq 0$.

For the case of an even number of vectors u^μ, consider the integral

$$I_{\mu\nu} = \int d\Omega\, u_\mu u_\nu \, . \tag{4.257}$$

It is symmetric with respect to interchanging μ and ν. Note also that $v^\mu I_{\mu\nu} = v^\nu I_{\mu\nu} = 0$. Therefore, $I_{\mu\nu}$ must be proportional to the projector onto the hyperplane perpendicular to v^μ,

$$I_{\mu\nu} = C \left(\eta_{\mu\nu} - \frac{v_\mu v_\nu}{v^2} \right) = C \overset{v}{\perp}_{\mu\nu} \ . \tag{4.258}$$

We find the proportionality constant C through contraction of indices,

$$I_\mu{}^\mu = \int d\Omega\, u\cdot u = -4\pi, \quad \overset{v}{\perp}_\mu{}^\mu = \delta_\mu{}^\mu - \frac{v\cdot v}{v^2} = 4 - 1 = 3 \, . \tag{4.259}$$

It follows that $C = -4\pi/3$, and relation (4.252) is established.

Relation (4.253) can be verified in a similar way, see Problem 4.5.4.

Problem 4.5.1. Derive the following formulas

$$\partial_\mu v^\mu = \frac{\lambda + 1}{\rho} \, , \tag{4.260}$$

$$(v\cdot\partial)\, \rho = \lambda + 1, \quad (a\cdot\partial)\, \rho = \frac{\lambda\, (\lambda + 1)}{\rho}, \quad (\dot{a}\cdot\partial)\, \rho = -a^2 + \lambda\, (\dot{a}\cdot c) \, , \tag{4.261}$$

$$\partial_\mu c_\nu = \frac{1}{\rho}\left(\eta_{\mu\nu} - \lambda\, c_\mu\, c_\nu - c_\mu\, v_\nu - v_\mu\, c_\nu\right) \ , \tag{4.262}$$

$$\partial_\mu \lambda = a_\mu + \rho\, (\dot{a} \cdot c)\, c_\mu \ , \tag{4.263}$$

$$(v \cdot \partial)\, c^\nu = -\frac{\lambda + 1}{\rho}\, c^\nu \ , \tag{4.264}$$

$$(a \cdot \partial)\, c^\nu = \frac{a^\nu}{\rho} - \frac{\lambda + 1}{\rho^2}\, (v^\nu + \lambda\, c^\nu) \ , \tag{4.265}$$

$$(\dot{a} \cdot \partial)\, c^\nu = \frac{\dot{a}^\nu - (\dot{a} \cdot c)\, v^\nu + [a^2 - \lambda\, (\dot{a} \cdot c)\,]\, c^\nu}{\rho} \ , \tag{4.266}$$

$$(c \cdot \partial)\, \{s,\ c^\nu,\ v^\nu,\ a^\nu,\ \dot{a}^\nu,\ \text{etc.}\} = 0 \ , \tag{4.267}$$

$$(c \cdot \partial)\, R^\nu = c^\nu \ , \tag{4.268}$$

$$(c \cdot \partial)\, \rho = 1 \ , \tag{4.269}$$

$$(c \cdot \partial)\, \lambda = \frac{\lambda + 1}{\rho} \ , \tag{4.270}$$

$$\partial_\mu c^\mu = \frac{2}{\rho} \ . \tag{4.271}$$

Consider dimensions $D + 1$ other than 4. Is any relation D-dependent?

Problem 4.5.2. Prove

$$\Box\, s = \frac{2}{\rho} \ , \tag{4.272}$$

$$\Box^2 s = 0 \ , \tag{4.273}$$

$$\Box\, \rho = \frac{2\,(2\lambda + 1)}{\rho} \ , \tag{4.274}$$

$$\Box\, f(\rho) = (2\lambda + 1)\left(\frac{2f'}{\rho} + f''\right) \ , \tag{4.275}$$

$$\Box \left(\frac{1}{\rho}\right) = 0 \ , \tag{4.276}$$

$$\Box\, \lambda = 4\, (\dot{a} \cdot c) \ , \tag{4.277}$$

$$\Box\, (\dot{a} \cdot X) = 4(\ddot{a} \cdot c) + \frac{2a^2}{\rho} \ . \tag{4.278}$$

Problem 4.5.3. Prove (4.243) and (4.244).

Problem 4.5.4. Prove (4.253).

4.6 Electromagnetic Field Generated by a Single Charge Moving Along an Arbitrary Timelike World Line

We apply the technique of retarded covariant variables to calculate the field strength generated by a single, arbitrarily moving, charge. Differentiating the Liénard–Wiechert vector potential (4.214) and taking into account (4.237) and (4.241) results in

$$\partial^\mu A^\nu = e\left[\frac{a^\nu c^\mu}{\rho} - \frac{v^\nu(v^\mu + \lambda\, c^\mu)}{\rho^2}\right] . \tag{4.279}$$

Interchanging indices gives the field strength tensor

$$F^{\mu\nu} = c^\mu U^\nu - c^\nu U^\mu , \tag{4.280}$$

$$U^\mu = e\left(-\lambda\,\frac{v^\mu}{\rho^2} + \frac{a^\mu}{\rho}\right) . \tag{4.281}$$

Expression (4.280) remains the same if U^μ is transformed as

$$U^\mu \to U^\mu + \kappa\, c^\mu \tag{4.282}$$

with arbitrary κ. In particular, taking $\kappa = e\,(a\cdot u)/\rho$, we obtain

$$F^{\mu\nu} = \frac{e}{\rho^2}\,(c^\mu V^\nu - c^\nu V^\mu) \tag{4.283}$$

where c^μ is defined in (4.209),

$$V^\mu = v^\mu + \rho\,(\overset{u}{\perp}\, a)^\mu , \tag{4.284}$$

and the operator $\overset{u}{\perp}$ projects vectors onto the hyperplane with normal u^μ,

$$\overset{u}{\perp}_{\mu\nu} = \eta_{\mu\nu} + u_\mu u_\nu . \tag{4.285}$$

We note in passing the relation

$$V\cdot c = 1 \tag{4.286}$$

which is due to the mutual perpendicularity of the vectors v^μ, u^μ and $(\overset{u}{\perp} a)^\mu$.

It is particularly remarkable that the 2-form F describing the field of a single charge is proportional to $c\wedge V$, that is, F is decomposable. The decomposability of F stems from the fact that only three vectors c^μ, v^μ and a^μ are at our disposal for constructing bivectors $c\wedge v$, $c\wedge a$, and $v\wedge a$.

Given a decomposable 2-form F, the invariant $\mathcal{P}$ is identically zero. As for the invariant $\mathcal{S}$, using (4.286), we find $\mathcal{S} = -e^2/\rho^4$. Therefore, a single charge moving along an arbitrary timelike world line generates the retarded field $F^{\mu\nu}$ of *electric* type. In other words, whatever the motion of the charge, there is a frame of reference, special for each point x^μ, such that only electric field persists, more precisely, $|\mathbf{E}| = e/\rho^2$ and $\mathbf{B} = 0$.

When the world line is straight, $z^\mu(s) = z^\mu(0) + v^\mu s$, the field strength becomes

$$F^{\mu\nu} = \frac{e}{\rho^2}\left(c^\mu v^\nu - c^\nu v^\mu\right) = \frac{e}{\rho^2}\left(u^\mu v^\nu - u^\nu v^\mu\right) . \tag{4.287}$$

In a Lorentz frame with the time axis parallel to v^μ, the only nontrivial component of (4.287) is just the Coulomb field $F^{j0} = E_j$ which is directed along the radius vector $r_j = \rho\, u_j$ drawn from the emission point to the observation point. The charge is at rest with respect to this frame, hence the field $\mathbf{E} = e\,\mathbf{n}/r^2$ is static, and the invariant retarded distance ρ coincides with the usual laboratory distance r.

The field of an arbitrarily moving charge (4.283) can be derived from the Coulomb field (4.287) by the formal replacement of V^μ for v^μ. To see this 'miraculous' conversion, rewrite (4.283) as

$$F = \frac{e}{\rho^2}\,\varpi , \tag{4.288}$$

$$\varpi = c \wedge V . \tag{4.289}$$

A pictorial view of the bivector ϖ is the parallelogram of the vectors c^μ and V^μ, see Fig. 4.4. The area A of the parallelogram is

$$A = \sqrt{-V^2\,(\overset{V}{\perp} c)^2} = V \cdot c = 1 \tag{4.290}$$

where the last equality follows from (4.286).

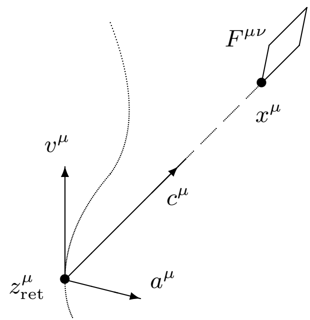


Fig. 4.4. Field $F^{\mu\nu}$ generated by a single charge

The bivector ϖ is invariant under simultaneous dilatation of c^μ and contraction of V^μ by a factor α,

$$c^\mu \to \alpha\, c^\mu, \quad V^\mu \to \alpha^{-1} V^\mu \,, \tag{4.291}$$

and also under the transformation

$$c^\mu \to c^\mu + \beta\, V^\mu, \quad V^\mu \to V^\mu + \gamma\, c^\mu \tag{4.292}$$

with arbitrary β and γ. This can be formulated slightly more rigorously. Let the vectors c^μ and V^μ be subjected to the general linear transformation

$$\begin{pmatrix} c' \\ V' \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{pmatrix} c \\ V \end{pmatrix} . \tag{4.293}$$

It induces the transformation of the bivector ϖ

$$\varpi' = (M_{11}c + M_{12}V) \wedge (M_{21}c + M_{22}V) = (M_{11}M_{22} - M_{12}M_{21})\, \varpi \,. \tag{4.294}$$

The condition $\varpi' = \varpi$ holds if the matrix M is unimodular,

$$\det M = M_{11}M_{22} - M_{12}M_{21} = 1 \,. \tag{4.295}$$

Such transformations form a group, the special linear group of real unimodular 2×2 matrices $\mathrm{SL}(2, \mathbb{R})$. They rotate and deform the initial parallelogram, converting it to parallelograms of unit area which belong to the plane spanned by the vectors c^μ and V^μ.

Thus the bivector ϖ is independent of concrete directions and magnitudes of the constituent vectors c^μ and V^μ. ϖ depends only on the parallelogram's orientation. It remains to clarify what orientations are possible. For example, is it possible for ϖ to be the exterior product of two spacelike vectors e_2^μ and e_3^μ? No. The parallelogram can be envisioned as constructed from a timelike unit vector e_0^μ and a spacelike imaginary-unit vector e_1^μ perpendicular to e_0^μ,

$$\varpi = e_0 \wedge e_1 \,. \tag{4.296}$$

Indeed, consider three versions of e_0^μ and e_1^μ constructed from V^μ and c^μ. For $V^2 > 0$,

$$e_0^\mu = \frac{V^\mu}{\sqrt{V^2}}, \quad e_1^\mu = \sqrt{V^2}\left(-c^\mu + \frac{V^\mu}{V^2}\right), \tag{4.297}$$

for $V^2 < 0$,

$$e_0^\mu = \sqrt{-V^2}\left(c^\mu - \frac{V^\mu}{V^2}\right), \quad e_1^\mu = \frac{V^\mu}{\sqrt{-V^2}} \,, \tag{4.298}$$

and for $V^2 = 0$,

$$e_0^\mu = \frac{1}{\sqrt{2}}\left(c^\mu + V^\mu\right), \quad e_1^\mu = \frac{1}{\sqrt{2}}\left(V^\mu - c^\mu\right). \tag{4.299}$$

In physical terms, this means that the choice of the Lorentz frame with the time axis parallel to the vector e_0^μ renders all components of $F^{\mu\nu}$ vanishing, except for F^{01}. The formulas (4.297)–(4.299) specify explicitly a frame in which the retarded electromagnetic field generated by a single arbitrarily moving charge appears as a pure Coulomb field at each observation point. With a curved world line, this frame is noninertial. The existence of this frame is of basic importance. It provides a picture in which any specification of F, other than the Coulombic ρ^{-2} dependence, is missing.

This inference may seem troublesome. Indeed, we are well aware of the fact that not only Coulomb fields but also magnetic fields are available in nature. Where do they come from? One can indicate at least two origins. First, the superposition principle. For the electromagnetic field F generated by several charges, the relations $\mathcal{P} = 0$, $\mathcal{S} < 0$ are generally invalid, and the 2-form F is no longer decomposable. Note that, in practice, the occurrence of magnetic fields due to the circuition of electrons around closed paths suggests a neutral system where electric fields of moving electrons and immovable nuclei mutually cancel. Second, a pure magnetic field may be related to spin and its associated magnetic dipole moment of the charged particle, as is clear from (4.59).

Decomposability is the most salient characteristic of the 2-form F associated with the field generated by a single charge. We recall that $\mathcal{P} = 0$ is the necessary and sufficient condition for the decomposability of F.

The decomposable 2-form F is invariant under the transformations (4.293) with the constraint (4.295). Since these transformations can be carried out independently at any spacetime point, we are dealing with the *local* $\mathrm{SL}(2,\mathbb{R})$ invariance. The decomposability of F and $\mathrm{SL}(2,\mathbb{R})$ invariance will henceforth be used synonymously. Note that the invariance under $\mathrm{SL}(2,\mathbb{R})$ is not pertinent to electrodynamics as a whole, hence it results in no Noether identities. Moreover, the system of a charged particle and its electromagnetic field, as such, is devoid of this invariance. It is rather a property of the *retarded solution* $F_{\rm ret}$ describing such a system. The advanced field $F_{\rm adv}$ can also be represented in a form similar to (4.283), that is, $F_{\rm adv}$ is decomposable (Problem 4.7.1), whereas combinations $F_{\rm ret} + \alpha\, F_{\rm adv}$ are not. We will see in Sect. 8.1 that the 2-form F associated with the Yang–Mills field generated by a single colored particle may be of two types, electric and magnetic, according to which phase, hot or cold, is considered. In either case, F is invariant under the local group $\mathrm{SL}(2,\mathbb{R})$. Note also that the retarded electromagnetic field F generated by a single charge in spacetimes of higher dimensions is no longer a bivector, see Problem 4.7.5.

Problem 4.6.1. Apart from (4.283)–(4.284), there are further simple expressions for the retarded field strength. Show that F can be written as

$$F = \frac{e}{\rho}\,\frac{d}{ds}\,c \wedge v\,. \tag{4.300}$$

Hint Use the relations $\dot{R}^\mu = -v^\mu$ and $\dot{\rho} = \lambda$.

Problem 4.6.2. Show that the bivector ϖ defined by (4.289) can be converted to the form (4.296) where e_0^μ and e_1^μ are given by (4.297)–(4.299). Verify that the vectors e_0^μ and e_1^μ given by formulas (4.297)–(4.299) are actually perpendicular timelike and spacelike vectors. Describe the regions of Minkowski space which correspond to the cases $V^2 < 0$, $V^2 > 0$, and $V^2 = 0$.

Answer The equation $\bar\rho = |\,a^2 + (a\cdot u)^2|^{-\frac{1}{2}}$ describes a tubular hypersurface $T_{\bar\rho}$ enclosing the world line. The case $V^2 < 0$ corresponds to the region outside $T_{\bar\rho}$, $\rho > \bar\rho$. The case $V^2 > 0$ corresponds to the region inside $T_{\bar\rho}$, $\rho < \bar\rho$. And the case $V^2 = 0$ corresponds to the hypersurface $T_{\bar\rho}$ itself, $\rho = \bar\rho$.

4.7 Another Way of Looking at Retarded Solutions

The retarded solution can be obtained without resort to the Green's function method. We turn in this section to an alternate procedure based on the technique of covariant retarded variables. (This approach has some utility in the electrodynamics of even-dimensional spacetimes, Problems 4.7.2–4.7.5. However, it will be of particular interest for the analysis of the Yang–Mills–Wong theory, Chap. 8, where the nonlinearity of the Yang–Mills equations hinders from the use of the Green's function method.)

We are searching for the retarded solution to the equation

$$\Box A^\mu - \partial^\mu \partial_\nu A^\nu = 4\pi e \int_{-\infty}^{\infty} ds\, v^\mu(s)\, \delta^4\left[x - z(s)\right] \tag{4.301}$$

where $z^\mu(s)$ is an arbitrary smooth timelike curve, the world line of a single point charge. To this end, we build an *ansatz* as follows.

As usual, let x^μ be an observation point, and let $z^\mu(s_{\rm ret})$ be the point on the world line from which the signal was sent to x^μ. We anticipate that A^μ can depend on only two vectors, $v^\mu(s_{\rm ret})$ and $R^\mu = x^\mu - z^\mu(s_{\rm ret})$. Indeed, one should bear in mind that the retarded function for the wave equation is concentrated in rays of the future light cone, and contains no derivative of the delta-function. This is the same as saying the signal carries information about the state of the source at a single point on the world line $z^\mu(s_{\rm ret})$, and this state is fully specified by the direction of the tangent v^μ at the instant $s_{\rm ret}$. The only nontrivial scalar $\rho = R\cdot v$ can be constructed from the vectors v^μ and R^μ, because $v^2 = 1$, and $R^2 = 0$. To summarize, the retarded vector potential is

$$A^\mu = v^\mu \Phi(\rho) + R^\mu \Psi(\rho) \tag{4.302}$$

where Φ and Ψ are as yet unknown functions, and v^μ is taken at the retarded instant $s_{\rm ret}$.

Now insert (4.302) in (4.301). By (4.237), (4.241), and (4.210), the left hand side of (4.301) becomes

$$\left(\Phi' + \frac{\Phi}{\rho}\right)[a^\mu + (a\cdot u)u^\mu] - \frac{1}{\rho}(\rho^2\Psi'' + 4\rho\Psi' + 2\Psi)v^\mu + (a\cdot u)(\rho^2\Psi'' + 2\rho\Psi')c^\mu \tag{4.303}$$

where the prime stands for differentiation with respect to ρ. This expression vanishes outside the world line. Equating to zero the coefficients of the three linearly independent vectors $(\overset{u}{\perp} a)^\mu = a^\mu + (a\cdot u)\,u^\mu$, v^μ, and c^μ, we arrive at an *overdetermined* system of ordinary differential equations

$$\rho\,\Phi' + \Phi = 0\ , \tag{4.304}$$

$$\rho^2\,\Psi'' + 4\,\rho\,\Psi' + 2\,\Psi = 0\ , \tag{4.305}$$

$$\rho\,\Psi'' + 2\,\Psi' = 0\ . \tag{4.306}$$

The general solution of (4.304) is

$$\Phi(\rho) = \frac{q}{\rho} \tag{4.307}$$

where q is some integration constant which will be determined below. The solution to (4.305) and (4.306) can be represented as $\Psi \propto \rho^\alpha$. It follows that either $\alpha = -1$ or $\alpha = -2$, due to (4.305). On the other hand, either $\alpha = -1$ or $\alpha = 0$, due to (4.306). Therefore, (4.305) and (4.306) are compatible when $\alpha = -1$, and their joint solution is

$$\Psi(\rho) = \frac{C}{\rho} \tag{4.308}$$

where C is another integration constant.

Taking into account (4.237), we observe that the second term of expression (4.302), with $\Psi(\rho)$ of the form (4.308), is pure gauge,

$$R^\mu\Psi(\rho) = C\,\frac{R^\mu}{\rho} = C\,\partial^\mu s\ . \tag{4.309}$$

This term does not contribute to $F_{\mu\nu}$, hence C is arbitrary.

We thus arrive at the retarded vector potential

$$A^\mu = q\,\frac{v^\mu}{\rho} \tag{4.310}$$

supplemented by the gauge term. We repeat calculations of Sect. 4.6 to obtain the field strength

$$F = \frac{q}{\rho^2}\,c\wedge V \tag{4.311}$$

with V^μ defined in (4.284). To determine q, we note that, for a straight world line, F becomes the Coulomb field. Applying Gauss' law to this case, q is identified with the charge e of the particle under study. Finally, we have the solution which coincides with the Liénard–Wiechert vector potential (4.214), up to the gauge term (4.309).

A feature of this procedure is that it does not require a gauge fixing condition for A^μ. We therefore obtain, not a unique solution, but rather the whole class of equivalent potentials A^μ related by the gauge transformations

$$\delta A^\mu = C\,\partial^\mu s\ . \tag{4.312}$$

For $C = 0$, the vector potential A^μ takes the form (4.310) which satisfies the Lorenz gauge fixing condition (4.105). For $C \neq 0$, we have solutions which do not satisfy the Lorenz condition because $\partial_\mu c^\mu \neq 0$.

If the gauge invariant operator $\eta_{\mu\nu}\,\Box - \partial_\mu\partial_\nu$ is replaced by the wave operator $\Box$, converting (4.301) to (4.106), the procedure we have just described no longer works. The ansatz (4.302) is applicable only to gauge invariant field equations.

Had we considered a massive vector field governed by the Klein–Gordon equation, rather than the Maxwell field, the support of the retarded Green's function would be the entire future light cone (see Problem 4.4.3), and our ansatz (4.302) would no longer be justified.

It is also worth mentioning that the procedure we have described rests crucially on the assumption that the world line is timelike and smooth[2]. These properties of the worldline are thus in good agreement with the retarted boundary condition.

Problem 4.7.1. Modify the procedure given above to the *advanced condition*, and find the advanced solution to (4.301). Show that the advanced field strength is

$$F = \frac{e}{\rho^2}\, c \wedge V\ , \tag{4.313}$$

$$c^\mu = -v^\mu + u^\mu\ , \tag{4.314}$$

$$V^\mu = v^\mu - \rho\,(\overset{u}{\perp} a)^\mu\ , \tag{4.315}$$

where all kinematical variables relate to the advanced instant $s_{\rm adv}$. Compare these with the corresponding expressions (4.283), (4.209), and (4.284) for the retarded field.

Problem 4.7.2. *Electrodynamics in even-dimensional spacetimes* $D+1 = 2n$. For a world with one temporal and one spatial dimension use the ansatz

$$A^\mu = R^\mu \Psi(\rho) \tag{4.316}$$

instead of (4.302). Verify that the retarded solution to Maxwell's equations is

$$A^\mu = -eR^\mu\ , \tag{4.317}$$

[2] Strictly speaking, it would suffice to rule out only cusps that render the world line Λ- or V-shaped.

and the associated field strength is

$$F = e\, c \wedge v \,. \tag{4.318}$$

Problem 4.7.3. For a world with one temporal and five spatial dimensions justify the ansatz

$$A^\mu = v^\mu\, \Phi(\rho,\, \lambda) + a^\mu\, \Omega(\rho,\, \lambda) + R^\mu\, \Psi(\rho,\, \lambda) \,, \tag{4.319}$$

where Φ, Ω and Ψ are sought functions. Verify that the retarded vector potential is

$$A^\mu = \frac{e}{3}\left(-\lambda\,\frac{v^\mu}{\rho^3} + \frac{a^\mu}{\rho^2}\right) \,, \tag{4.320}$$

modulo the gauge terms $k\,\partial^\mu \tau$ and $l\,\partial^\mu \rho$, and the field strength is

$$F = \frac{e}{3}\left(c \wedge V + \frac{a \wedge v}{\rho^3}\right) \,, \tag{4.321}$$

$$V^\mu = \frac{v^\mu}{\rho^4}\left[3\,\lambda^2 - \rho^2\,(\dot{a}\cdot c)\right] - 3\,\lambda\,\frac{a^\mu}{\rho^3} + \frac{\dot{a}^\mu}{\rho^2} \,. \tag{4.322}$$

Problem 4.7.4. The *prepotential* H_μ of the vector potential A_μ is defined as

$$A_\mu = \Box\, H_\mu \,. \tag{4.323}$$

Verify that, for $D = 1, 3$ and 5, the retarded vector potentials $A_\mu^{(D+1)}$ are related by

$$\Box\, A_\mu^{(2)} \propto (2 - D)\, A_\mu^{(4)} \,, \tag{4.324}$$

$$\Box\, A_\mu^{(4)} \propto (4 - D)\, A_\mu^{(6)} \,. \tag{4.325}$$

Prove that any $2n$-dimensional retarded vector potential $A_\mu^{(2n)}$ (up to a normalization factor) is the prepotential of the $(2n+2)$-dimensional retarded vector potential $A_\mu^{(2n+2)}$.

Problem 4.7.5. Show that the retarded electromagnetic field F generated by a single charge e in even-dimensional spacetimes, $D + 1 = 2n$, is the sum of $n - 1$ bivectors

$$F^{(2n)} = f_1 \wedge f_2 + \cdots f_{2n-3} \wedge f_{2n-2} \quad \text{with} \quad f_1 \wedge f_2 = c \wedge U^{(2n)} \,. \tag{4.326}$$

Here, $U_\mu^{(2n)}$ is related to the corresponding $(2n+2)$-dimensional vector potential $A_\mu^{(2n+2)}$ as

$$Z_{2n}\, U_\mu^{(2n)} = \rho\, A_\mu^{(2n+2)} \,, \tag{4.327}$$

where the numerical factor Z_{2n} is determined by Gauss' law.

4.8 Field Due to a Magnetic Monopole

The idea of magnetic charge is rather old. In modern times, attention to it was drawn by Dirac who argued that the coexistence of electric and magnetic charges in the quantum picture gives rise to quantization of the electric charge e according to the relation $2ee^\star = n$ where $e^\star$ is a fixed magnetic charge, and n is an integer. But we do not dwell on this subject being limited to classical physics. This section is devoted instead to features of fields generated by a classical magnetic monopole.

We begin with an imaginary world containing a single particle with a magnetic charge $e^\star$. The electromagnetic field is governed by the equations

$$\partial_\lambda F^{\lambda\mu} = 0\,, \tag{4.328}$$

$$\partial_\lambda {}^*F^{\lambda\mu} = 4\pi e^\star \int_{-\infty}^{\infty} ds\, v^\mu(s)\, \delta^4\left[x - z(s)\right]\,. \tag{4.329}$$

The formal resemblance of this system to that of a single charged particle and its field enables us to give the retarded solution to equations (4.328) and (4.329):

$${}^*F = \frac{e^\star}{\rho^2}\,\varpi \tag{4.330}$$

where the bivector ϖ is defined in (4.289). It follows that

$$F = -\frac{e^\star}{\rho^2}\,\sigma\,, \tag{4.331}$$

$$\sigma = {}^*(c \wedge V)\,. \tag{4.332}$$

The 2-form F given by (4.331) and (4.332) is decomposable. Indeed, ${}^{**}F = -F$, and, by analogy with the retarded field of a single charged particle, $\mathcal{P} = 0$.

The other field invariant is $\mathcal{S} = \frac{1}{2}\,F_{\alpha\beta}F^{\alpha\beta} = -\frac{1}{2}\,{}^*F_{\alpha\beta}{}^*F^{\alpha\beta} = (e^\star)^2/\rho^4$, which implies that we are dealing with the *field configuration of magnetic type* for any timelike world line of the source, $\mathbf{B}^2 > \mathbf{E}^2$. Thus,

$$\sigma = e_2 \wedge e_3 \tag{4.333}$$

where e_2^μ and e_3^μ are spacelike normalized vectors perpendicular to c^μ and V^μ, as depicted in Fig. 4.5. (By e_1^μ we mean the spacelike normalized vector u^μ perpendicular to e_2^μ and e_3^μ.) Vectors e_2^μ and e_3^μ can readily be constructed out of V^μ and c^μ (Problem 4.8.1).

When the world line is straight, the field strength $F^{\mu\nu}$ is constant. In the Lorentz frame with time axis parallel to v^μ, the only nontrivial component of $F^{\mu\nu}$ is F^{23} which is dual to a Coulomb-like magnetic field $B_i = -\frac{1}{2}\,\epsilon_{ijk}F^{jk}$ directed along the radius vector $r_i = \rho\, u_i$ [this is clear from (4.331) and (4.333), and the fact that u^μ, e_2^μ and e_3^μ span an orthonormal basis in the hyperplane with the normal v^μ]. Simply stated, the magnetic field $\mathbf{B}$ due to a static magnetic monopole placed at the origin is

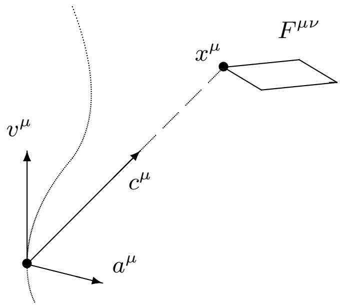


Fig. 4.5. Field due to a magnetic monopole

$$\mathbf{B} = e^\star \, \frac{\mathbf{n}}{r^2} \, , \tag{4.334}$$

where $\mathbf{n}$ is a unit vector directed along the radius vector $\mathbf{r}$.

Let us recall that the four-force f^μ acting on a magnetically charged particle is $e^\star v^\lambda F_{\mu\lambda}$. This expression, together with (4.330), shows that if we grant the existence of magnetic charges and the absence of electric ones, we may regain the conventional picture with electric charges by the merely renaming dual quantities. We should therefore be concerned with the problem of coexisting of electric and magnetic charges.

In a world where electric and magnetic charges coexist, Maxwell's equations modify:

$$\begin{aligned} \partial_\lambda F^{\lambda\mu} &= 4\pi j^\mu \, , \\ \partial_\lambda {}^*F^{\lambda\mu} &= 4\pi m^\mu \, , \end{aligned} \tag{4.335}$$

where j^μ and m^μ are the four-currents of electric and magnetic charges e_I and $e_I^\star$,

$$\begin{aligned} j^\mu(x) &= \sum_{I=1}^{N} e_I \int_{-\infty}^{\infty} ds_I \, v_I^\mu(s_I) \, \delta^4 \left[x - z_I(s_I)\right], \\ m^\mu(x) &= \sum_{I=1}^{K} e_I^\star \int_{-\infty}^{\infty} ds_I \, v_I^\mu(s_I) \, \delta^4 \left[x - z_I(s_I)\right] \, , \end{aligned} \tag{4.336}$$

obeying the local conservation laws

$$\partial_\mu j^\mu = 0, \quad \partial_\mu m^\mu = 0 \, , \tag{4.337}$$

or, in vector notation,

$$\nabla \times \mathbf{B} = 4\pi \mathbf{j} + \frac{\partial \mathbf{E}}{\partial t}, \quad \nabla \cdot \mathbf{E} = 4\pi \varrho \, , \tag{4.338}$$

$$\nabla \times \mathbf{E} = -4\pi \mathbf{j}^\star - \frac{\partial \mathbf{B}}{\partial t}, \quad \nabla \cdot \mathbf{B} = 4\pi \varrho^\star . \tag{4.339}$$

The system of equations (4.335), or, what is the same, (4.338)–(4.339), might seem overdetermined: 8 equations determine 6 unknown functions. However, two of them are constraints. We might reduce the number of equations by solving the constraints, namely expressing longitudial modes of **E** and **B** in terms of ϱ and $\varrho^\star$, but this would result in a nonlocal construction. It would be reasonable to proceed in the opposite direction increasing formally the number of unknown functions.

With the ansatz (4.85) in mind, we may, following Nicola Cabibbo and Ezio Ferrari 1962, express the tensor $F_{\mu\nu}$ in terms of *two* vector potentials A_μ and B_μ as

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - \epsilon_{\mu\nu\alpha\beta}\partial^\alpha B^\beta , \tag{4.340}$$

which is equivalent to

$${}^*F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu + \epsilon_{\mu\nu\alpha\beta}\partial^\alpha A^\beta . \tag{4.341}$$

The system of field equations (4.335) is converted to

$$\begin{aligned} \Box\, A^\mu - \partial^\mu \partial_\nu A^\nu &= 4\pi j^\mu , \\ \Box\, B^\mu - \partial^\mu \partial_\nu B^\nu &= 4\pi m^\mu , \end{aligned} \tag{4.342}$$

and we therefore arrive at the system of 8 equations with 8 unknown functions.

However, the number of field degrees of freedom still remains equal to 6. Indeed, the field strength $F_{\mu\nu}$ is invariant under the gauge transformations

$$A_\mu \to A'_\mu = A_\mu + \partial_\mu \chi, \quad B_\mu \to B'_\mu = B_\mu + \partial_\mu \omega \tag{4.343}$$

with arbitrary smooth functions χ and ω. Therefore, only transverse modes of the vector potentials A^μ and B^μ contribute to $F^{\mu\nu}$. The longitudial modes $\partial_\mu \chi$ and $\partial_\mu \omega$ are not dynamical degrees of freedom. On the other hand, among the 8 equations (4.342), only 6 are independent because the four-dimensional divergence of both sides of equations (4.342) is identically zero. These properties are most obvious for the Fourier transforms of fields. For example, the linear dependence of the first four equations in (4.342) is clearly seen from the identity

$$k_\mu \left(k^2 \widetilde{A}^\mu - k^\mu k_\nu \widetilde{A}^\nu - 4\pi \widetilde{\jmath}^\mu \right) = 0 . \tag{4.344}$$

An alternative to the two-vector-potential approach is the use of one vector potential A^μ, related to the field strength by the conventional equation

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu . \tag{4.345}$$

This equation implies $\partial_\lambda {}^*F^{\lambda\mu} = 0$, in conflict with (4.335), and is therefore apparently incorrect for smooth vector potentials A_μ mapped on the whole Minkowski space. And, nevertheless, following Dirac, we may use (4.345) for vector potentials A_μ which become singular on lines that issue from the magnetic charge (see Problem 4.1.5). The evolution of a magnetic monopole is depicted by a world sheet which the singular line sweeps out. Equation (4.345) holds almost everywhere, hence $\partial_\lambda {}^*F^{\lambda\mu}$ is zero almost everywhere, except for the singular line. Owing to this obstacle, equation (4.335) can still be valid.

It was shown by Tai Tsun Wu and Yang in 1975 that the vector potential A_μ need not be singular. Instead, we may use two (or more) maps $x_\mu \to \mathcal{A}_\mu$. To be more specific, take the region U to be all the spacetime minus the point of a magnetic charge. It is possible to divide U into two overlapping regions U_a and U_b and to define $(\mathcal{A}_\mu)_a$ and $(\mathcal{A}_\mu)_b$, each singularity free in their respective regions, so that (i) their curls are equal to the magnetic field and (ii) in the overlapping region $(\mathcal{A}_\mu)_a$ and $(\mathcal{A}_\mu)_b$ are related by a gauge transformation (for more detail see Problem 4.8.2).

Problem 4.8.1. Verify that

$$e_2^\mu = \frac{1}{|N|}\,(N^\mu - \rho\, N^2 c^\mu), \quad e_3^\mu = \frac{1}{|N|}\,\epsilon^{\mu\alpha\beta\gamma} N_\alpha\, c_\beta\, V_\gamma \quad \text{with} \quad N^\mu = (\overset{u}{\perp} a)^\mu \tag{4.346}$$

are the vectors suited for constructing the bivector σ in (4.333). Show that $\sigma = e_2 \wedge e_3$ is nonzero even for collinear u^μ and a^μ, which renders N^μ vanishing.

Hint $\sigma = 0 \iff {}^*\sigma = 0$. However, ${}^*\sigma = c \wedge V$ is nonzero for any timelike world line.

Problem 4.8.2. *The Wu and Yang vector potential.* Take the regions U_a and U_b to be

$$U_a: \quad 0 \le \vartheta < \pi/2 + \delta, \quad 0 \le \varphi < 2\pi, \quad \text{for all } t\,, \tag{4.347}$$

$$U_b: \quad \pi/2 - \delta < \vartheta \le \pi, \quad 0 \le \varphi < 2\pi, \quad \text{for all } t\,, \tag{4.348}$$

with the overlap extending throughout $\pi/2 - \delta < \vartheta < \pi/2 + \delta$. Define the vector potentials

$$(\mathcal{A}_t)_a = (\mathcal{A}_r)_a = (\mathcal{A}_\vartheta)_a = 0, \quad (\mathcal{A}_\varphi)_a = \frac{e^\star}{r\sin\vartheta}\,(1 - \cos\vartheta)\,, \tag{4.349}$$

$$(\mathcal{A}_t)_b = (\mathcal{A}_r)_b = (\mathcal{A}_\vartheta)_b = 0, \quad (\mathcal{A}_\varphi)_b = \frac{-e^\star}{r\sin\vartheta}\,(1 + \cos\vartheta)\,, \tag{4.350}$$

see Problem 4.1.5 where the vector $\hat{\mathbf{u}}$ forming the singular line should be taken to be aligned with the x^3-axis in U_b and opposed to it in U_a. Note that $(\mathcal{A}_\mu)_a$ is regular in U_a, and $(\mathcal{A}_\mu)_b$ is regular in U_b.

Verify that $(\mathcal{A}_\mu)_a$ and $(\mathcal{A}_\mu)_b$ are related by the gauge transformation

$$(\mathcal{A}_\mu)_b = (\mathcal{A}_\mu)_a - \partial_\mu \chi \quad \text{with} \quad \chi = 2e^\star \varphi \tag{4.351}$$

in the overlap of the two regions.

Hint If $(\mathcal{A}_\mu)_a$ and $(\mathcal{A}_\mu)_b$ are time-independent, then $\partial_\mu \chi = (0, \nabla\chi)$. To see that $\chi = 2e^\star\varphi$, use the gradient operator in spherical coordinates

$$\nabla = \mathbf{e}_r \frac{\partial}{\partial r} + \mathbf{e}_\vartheta \frac{1}{r}\frac{\partial}{\partial\vartheta} + \mathbf{e}_\varphi \frac{1}{r\sin\vartheta}\frac{\partial}{\partial\varphi}\,, \tag{4.352}$$

where $\mathbf{e}_r$, $\mathbf{e}_\vartheta$, $\mathbf{e}_\varphi$ are unit vectors along, respectively, the radius vector, a meridian directed toward the north pole, and a parallel of latitude in the direction of increasing longitude.

Problem 4.8.3. Consider generalizations of the Coulomb law (4.24) and the Biot and Savart law (4.50) for a nonrelativistic dyon carrying both electric and magnetic charges e and $e^\star$. Find the static magnetic and electric fields $\mathbf{B}$ and $\mathbf{E}$ due to this dyon.

Hint Use (4.24), (4.334), and (4.50), and observe that the sources $4\pi\mathbf{j}$ and $-4\pi\mathbf{j}^\star$ of equations (4.338) and (4.339) are opposite in sign.

Answer

$$\mathbf{B} = \frac{1}{r^2}\left(e^\star\,\mathbf{n} + e\,\mathbf{v}\times\mathbf{n}\right), \quad \mathbf{E} = \frac{1}{r^2}\left(e\,\mathbf{n} - e^\star\,\mathbf{v}\times\mathbf{n}\right)\,. \tag{4.353}$$

Problem 4.8.4. Consider slowly moving particle 1 possessing electric charge e_1 and magnetic charge $e_1^\star$, which interacts with static particle 2 with parameters e_2 and $e_2^\star$. Verify that, to a first approximation in $\mathbf{v}_1$, the force on particle 1 due to 2 is

$$\mathbf{F}_{12} = (e_1e_2 + e_1^\star e_2^\star)\,\frac{\mathbf{n}_{12}}{r_{12}^2} + (e_1e_2^\star - e_2e_1^\star)\,\frac{\mathbf{v}_1\times\mathbf{n}_{12}}{r_{12}^2}, \quad \mathbf{n}_{12} = \frac{\mathbf{z}_1 - \mathbf{z}_2}{|\mathbf{z}_1 - \mathbf{z}_2|}\,. \tag{4.354}$$

Problem 4.8.5. Consider a particle of charge e moving in the field of a very massive magnetic monopole of strength $e^\star$. Find the Hamiltonian of this particle.

Answer

$$H = \sqrt{p_r^2 + \frac{p_\vartheta^2}{r^2} + \frac{(p_\varphi^2 + ee^\star\cos\vartheta)^2}{r^2\sin^2\vartheta} + m^2}\,. \tag{4.355}$$

Hint Use (2.221) rewritten in the form $H - e\phi = \sqrt{(\mathbf{p} - e\mathbf{A})^2 + m^2}$. Calculate it with $\phi = 0$ and $\mathbf{A}$ given by (4.71).

Notes

1. *Section 4.1.* Additional references on exact solutions to the equations of electrostatics and magnetostatics are Morse & Feshbach (1953) and Jackson (1962). Relevant problems are collected in Batygin & Toptygin (1978).

2. *Section 4.2.* The argument that Λ^{-1} does not exist, relying on the fact that (4.104) is actually an eigenvalue equation corresponding to zero eigenvalue, has been communicated to the author by Woodard.

For more historical detail of the Lorenz gauge see Jackson & Okun (2001). A general gauge condition similar to (4.111) was discussed in Zumino (1960).

De Witt (1962) showed that the vector potential A_μ can be expressed in terms of the field strength $F_{\mu\nu}$, even though the suggested construction proves nonlocal. This result intimates that electrodynamics can be formulated in terms of $F_{\mu\nu}$. On the quantum level, A_μ is more fundamental than $F_{\mu\nu}$, as exemplified by the Aharonov–Bohm effect. For a review of this effect see Peshkin & Tonomura (1989).

3. *Section 4.3.* The identity (4.172) is sometimes referred to as the Ward–Takahashi identity. The structure of such identities is fully considered by Takahashi (1969). Solitons touched on in Problems 4.3.7 and 4.3.8 are discussed at greater length in many text, to mention just three: Whitham (1974), Ablowitz & Segur (1981), and Rajaraman (1982).

4. *Section 4.4.* The Green's function method is treated in most textbooks on mathematical physics and field theory. A general reference is Morse & Feshbach (1953), Courant (1962), Gel'fand & Shilov (1964), Iwanenko & Sokolow (1953), Jackson (1962), Barut (1964), and Rohrlich (1965). The retarded potentials for a continuous distribution of electric charge were obtained by Lorenz (1867). The retarded potentials for discrete charges (4.215) are due to Liénard (1898), and Wiechert (1900). Our treatment is close to Dirac (1938).

5. *Section 4.5.* The covariant retarded variables were defined in Dirac (1938). The paper by Synge (1970) covered many of the subject in this section. The surface element on the light cone can be found in the book by Synge (1956).

6. *Section 4.6.* Kosyakov (1994) showed that the 2-form F corresponding to the Liénard–Wiechert field is decomposable. As a consequence, the retarded electromagnetic field generated by an arbitrary moving charge is essentially the same as the Coulomb field. There exists a (noninertial) frame of reference in which the field becomes $|\mathbf{E}| = e/\rho^2$, $\mathbf{B} = 0$. It is therefore meaningless, contrary to a widespread belief, to divide F into the 'radiation field' and 'Coulomb-like field' scaling respectively as ρ^{-1} and ρ^{-2}, because the long-range ρ^{-1} field can be eliminated altogether by the mere transition to a local frame of reference determined in the appropriate way. As will be discussed in Sect. 6.2, the idea of radiation is embodied in some term of the electromagnetic stress-energy tensor.

The Coulomb field, the most salient configuration of the electromagnetic field, is due to Coulomb (1785) who invented the torsion balance for measuring the electrostatic force. The idea of a freely propagating electromagnetic wave was advanced by Lorenz (1867) and Maxwell (1873), and confirmed experimentally by Hertz (1887), (1888). For more detail see Whittaker (1910).

7. *Section 4.7.* The analysis of this section follows Kosyakov (1999).

8. *Section 4.8.* The modification of Maxwell's equations with the concurrent presence of electric and magnetic sources (4.338)–(4.339) was proposed by Heaviside (1892). Dirac (1931) and (1948) are basic references on the construction of the electromagnetic field $F_{\mu\nu}$ due to a magnetic monopole. This field is expressed in terms of a vector potential A_μ which is singular on a line called the Dirac string. Cabibbo & Ferrari (1962) constructed $F_{\mu\nu}$ in terms of two vector potentials A_μ and B_μ which are regular everywhere, except the magnetic pole. Wu & Yang (1975) expressed $F_{\mu\nu}$ in terms of a single vector potential A_μ which has no singularity on the Dirac string, but bears on a topologically involved layout.

5

Lagrangian Formalism in Electrodynamics

5.1 Action Principle. Symmetries and Conservation Laws

In this section we discuss general ideas about Lagrangian treatment of systems of particles and fields. Since the particle sector has been already developed in Sect. 2.5, it remains to see how these results can be adapted to the field sector. We then augment our analysis by the addition of terms responsible for the interaction of particles and fields.

The Lagrangian treatment of fields resembles that of particles in many respects. It is therefore expedient to recapitulate briefly the procedure of Sect. 2.5, and introduce modifications necessary for systems with infinite degrees of freedom.

Consider a system whose states are specified by n field variables $\phi_a(x)$, $a = 1, \ldots, n$. These fields may be real or complex, scalar, vector, spinor, or tensor functions of spacetime. Suppose that the behavior of the system is governed by the action

$$S = \int_{\mathcal{U}} d^4x\, \mathcal{L}(x, \phi, \partial\phi) \ , \tag{5.1}$$

where $\mathcal{U}$ is a domain bounded by spacelike surfaces Σ_1 and Σ_2 which extend to infinity. The scalar $\mathcal{L}$, called the *Lagrangian density*, or *Lagrangian* for short, is a local function of fields and their first derivatives. We assume that any explicit dependence of the Lagrangian upon spacetime is confined to external sources $J^a(x)$ which enter $\mathcal{L}$ in the form $-J^a(x)\phi_a(x)$.

By analogy with mechanics, we define the *local field variation* as

$$\delta\phi_a = {\phi'}_a(x) - \phi_a(x) \ , \tag{5.2}$$

which refers to a change from a given field configuration to a neighboring one. The *total field variation* is defined by

$$\Delta\phi_a = {\phi'}_a(x') - \phi_a(x) \ , \tag{5.3}$$

where the spacetime coordinates differ from each other by an infinitesimal amount,

$$x'^{\mu} = x^{\mu} + \Delta x^{\mu} \, . \tag{5.4}$$

We remark parenthetically that local field variations are of primary concern for the action principle, while total field variations play a decisive role in Noether's first theorem.

Clearly,

$$\Delta\phi = \delta\phi + \Delta x^{\mu}\, \partial_{\mu}\phi \, , \tag{5.5}$$

where the last term can be thought of as a deformation of the fixed field configuration in response to the change of spacetime variables (5.4).

Assume that the functional form of $\mathcal{L}$ is unchanged by these variations of ϕ and x,

$$\mathcal{L}'(x', \phi', \partial\phi') = \mathcal{L}(x', \phi', \partial\phi') \, . \tag{5.6}$$

Then the corresponding variation of the Lagrangian

$$\Delta\mathcal{L} = \mathcal{L}'(x', \phi', \partial\phi') - \mathcal{L}(x, \phi, \partial\phi) \tag{5.7}$$

takes the form

$$\Delta\mathcal{L} = (\partial_{\mu}\mathcal{L})\, \Delta x^{\mu} + \partial_{\mu}\left[\frac{\partial\mathcal{L}}{\partial(\partial_{\mu}\phi_a)}\, \delta\phi_a\right] + \mathcal{E}^a\, \delta\phi_a \, . \tag{5.8}$$

Here, $\partial_{\mu}\mathcal{L}$ is the complete partial derivative with respect to x^{μ}, that is, including the implicit x^{μ} dependence from the fields,

$$\partial_{\mu}\mathcal{L} = \frac{\partial\mathcal{L}}{\partial x^{\mu}} + \frac{\partial\mathcal{L}}{\partial\phi_a}\, \partial_{\mu}\phi_a + \frac{\partial\mathcal{L}}{\partial(\partial_{\nu}\phi_a)}\, \partial_{\mu}\partial_{\nu}\phi_a \, , \tag{5.9}$$

and $\mathcal{E}^a$ is the *Eulerian* associated with the variation with respect to ϕ_a,

$$\mathcal{E}^a = \frac{\partial\mathcal{L}}{\partial\phi_a} - \partial_{\mu}\left[\frac{\partial\mathcal{L}}{\partial(\partial_{\mu}\phi_a)}\right] \, . \tag{5.10}$$

The total variation of the action is

$$\Delta S = \int_{\mathcal{U}'} d^4x'\, \mathcal{L}'(x', \phi', \partial\phi') - \int_{\mathcal{U}} d^4x\, \mathcal{L}(x, \phi, \partial\phi) \, , \tag{5.11}$$

where

$$d^4x' = \det\left(\frac{\partial x'}{\partial x}\right) d^4x \, . \tag{5.12}$$

By (5.4),

$$\left(\frac{\partial x'}{\partial x}\right) = \begin{pmatrix} 1 + \partial_0\Delta x^0 & \partial_0\Delta x^1 & \partial_0\Delta x^2 & \partial_0\Delta x^3 \\ \partial_1\Delta x^0 & 1 + \partial_1\Delta x^1 & \partial_1\Delta x^2 & \partial_1\Delta x^3 \\ \partial_2\Delta x^0 & \partial_2\Delta x^1 & 1 + \partial_2\Delta x^2 & \partial_2\Delta x^3 \\ \partial_3\Delta x^0 & \partial_3\Delta x^1 & \partial_3\Delta x^2 & 1 + \partial_3\Delta x^3 \end{pmatrix} . \tag{5.13}$$

To the first order in Δx^μ,

$$\det\left(\frac{\partial x'}{\partial x}\right) = 1 + \partial_0 \Delta x^0 + \partial_1 \Delta x^1 + \partial_2 \Delta x^2 + \partial_3 \Delta x^3 = 1 + \partial_\mu \Delta x^\mu \; . \tag{5.14}$$

Substituting (5.12) and (5.14) into (5.11) gives

$$\Delta S = \int_{\mathcal{U}} d^4x \left(\Delta\mathcal{L} + \mathcal{L}\,\partial_\mu \Delta x^\mu\right) \; , \tag{5.15}$$

where $\Delta\mathcal{L}$ is defined in (5.7). We combine (5.8) with (5.15) and rearrange terms,

$$\Delta S = \int_{\mathcal{U}} d^4x \left[\partial_\mu \left(\mathcal{L}\Delta x^\mu + \pi^\mu_a\, \delta\phi_a\right) + \mathcal{E}^a \delta\phi_a\right] \; , \tag{5.16}$$

where π^μ_a is the *momentum* conjugate to the field variable ϕ_a,

$$\pi^\mu_a = \frac{\partial\mathcal{L}}{\partial(\partial_\mu \phi_a)} \; . \tag{5.17}$$

If the integrand falls off fast enough as $|\mathbf{x}| \to \infty$, then, applying the Gauss–Ostrogradskiĭ theorem, we obtain

$$\Delta S = Q(\Sigma_2) - Q(\Sigma_1) + \int_{\mathcal{U}} d^4x\, \mathcal{E}^a \delta\phi_a \; . \tag{5.18}$$

The surface integrals

$$Q(\Sigma_i) = \int_{\Sigma_i} d\sigma_\mu \left(\mathcal{L}\Delta x^\mu + \pi^\mu_a \delta\phi_a\right) \tag{5.19}$$

can be further transformed using (5.5),

$$Q(\Sigma_i) = \int_{\Sigma_i} d\sigma_\mu [(\delta^\mu_\nu \mathcal{L} - \pi^\mu_a\, \partial_\nu \phi_a)\Delta x^\nu + \pi^\mu_a \Delta\phi_a] = \int_{\Sigma_i} d\sigma_\mu (\pi^\mu_a \Delta\phi_a - \theta^\mu{}_\nu \Delta x^\nu) \; . \tag{5.20}$$

The quantity

$$\theta^\mu{}_\nu = \pi^\mu_a\, \partial_\nu \phi_a - \delta^\mu_\nu\, \mathcal{L} \tag{5.21}$$

is known as the *canonical stress-energy tensor*. Thus, the general expression for ΔS can be brought into close relation to that of mechanics, as is clear when (5.18) is compared with (2.172).

We now specialize these general observations to the case that the boundary surfaces Σ_1 and Σ_2 are fixed, and the total field variation vanishes at Σ_1 and Σ_2. For such variations, the surface integrals (5.19) are zero. In this context, we formulate the *action principle*: the actual field configuration makes the action extremal, $\Delta S = 0$. It follows from (5.18) that

$$\int_{\mathcal{U}} d^4x\, \mathcal{E}^a \delta\phi_a = 0 \; . \tag{5.22}$$

Taking into account that $\delta\phi_a$ are arbitrary infinitesimal functions, we conclude that the actual field configurations are solutions to the *Euler–Lagrange equations*

$$\mathcal{E}^a = \frac{\partial\mathcal{L}}{\partial\phi_a} - \partial_\mu\left[\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_a)}\right] = 0\,. \tag{5.23}$$

Fields ϕ_a obeying the Euler–Lagrange equations make the action an extremum, hence the name *extremals.*

To summarize, we have obtained partial differential equations (5.23) governing the behavior of field systems. In the field-theoretic language, these Euler–Lagrange equations are called *field equations.* All essential properties of a given system are assumed to be encoded in the Lagrangian $\mathcal{L}$. Examples of Lagrangians for the most common fields will be given below.

We now turn to Noether's first theorem. Let

$$\Delta x^\mu = \Gamma^\mu_{\ k}(x)\,\epsilon^k\,, \tag{5.24}$$

$$\Delta\phi_a = G_{ka}(x,\phi)\,\epsilon^k \tag{5.25}$$

be a group of infinitesimal transformations depending upon p parameters ϵ^k, $k = 1,\ldots,p$, with the $\Gamma^\mu_{\ k}$ being group generators and the G_{ka} being representations of these generators acting on ϕ_a. Assume that the action is invariant[1] under the infinitesimal transformations (5.24)–(5.25), which are called *symmetry transformations.* We will distinguish between *internal* and *spacetime* symmetries. Any internal symmetry is such that the coordinates x^μ are unchanged, that is, $\Delta x^\mu = 0$, and functions G_{ka} are independent of x. If the change of field variables (5.25) is instead induced by a coordinate transformation (5.24), then (5.24)–(5.25) is a spacetime transformation.

Putting $\mathcal{E}^a = 0$ in (5.18), we have

$$Q_k = \int_\Sigma d\sigma_\mu\left(\pi^\mu_a\, G_{ka} - \theta^\mu_{\ \nu}\,\Gamma^\nu_{\ k}\right) = \text{const} \tag{5.26}$$

for any spacelike Σ. In particular, Q_k is unaffected by shifts of Σ in timelike directions. Now *Noether's first theorem* reads: invariance of the action $\Delta S = 0$ under a continuous p-parameter group of transformations (5.24)–(5.25) implies p global conservation laws for the integral quantities Q_k.

Equations (5.26), expressing p global conservation laws, are equivalent to a set of p *equations of continuity*

$$\partial_\mu\mathcal{N}^\mu_{\ k} = 0, \quad k = 1,\ldots,p\,, \tag{5.27}$$

where $\mathcal{N}^\mu_{\ k}$, called the *Noether current*, is defined by

$$\mathcal{N}^\mu_{\ k} = \pi^\mu_a\, G_{ka} - \theta^\mu_{\ \nu}\,\Gamma^\nu_{\ k}\,. \tag{5.28}$$

[1] By invariance we mean that $\Delta S = 0$ for general field variables ϕ, not just for extremals.

Another name for (5.27) is *local conservation laws.*

This argument is valid even if one relaxes the condition that the functional form of the Lagrangian is invariant; rather it is sufficient to require

$$\mathcal{L}(x', \phi', \partial'\phi') = \mathcal{L}(x, \phi, \partial\phi) + \partial_\mu B^\mu \,, \tag{5.29}$$

where B^μ is an arbitrary vector function of ϕ, because total derivatives in Lagrangians contribute only to surface variations. Then the Noether's current is added by B^μ,

$$\mathcal{N}^\mu_{\ k} = \pi^\mu_a \, G_{ka} - \theta^\mu_{\ \nu} \, \Gamma^\nu_{\ k} + B^\mu \,. \tag{5.30}$$

The application of Noether's first theorem to different symmetries of electrodynamics will be discussed at greater length in the subsequent sections of this chapter.

Let us look at a few examples of Lagrangian field theories. We first take a real scalar field ϕ. Lagrangians linear in ϕ, such as $\mathcal{L} = D\phi$ where D is a constant or a differential operator, are inconsistent because the corresponding Euler–Lagrange equations are $D = 0$, whence $\mathcal{L} = 0$. The simplest Lorentz invariant Lagrangian is quadratic in ϕ,

$$\mathcal{L} = \frac{1}{2} \, (\partial_\alpha \phi)(\partial^\alpha \phi) - \frac{\mu^2}{2} \, \phi^2 \,. \tag{5.31}$$

We do not show linear terms in (5.31) since they can be absorbed by rearranging the quadratic form to the sum of diagonal terms (Problem 5.1.1).

Taking into account that partial integration of the action does not change the Euler–Lagrange equations, we can rewrite (5.31) as

$$\mathcal{L} = -\frac{1}{2} \, \phi \left(\Box + \mu^2\right) \phi \,. \tag{5.32}$$

Clearly the Euler–Lagrange equation resulting from (5.32) is the *Klein–Gordon equation*

$$\left(\Box + \mu^2\right) \phi = 0 \,. \tag{5.33}$$

We learned from Problem 4.3.4 that solutions to the Klein–Gordon equation propagate with the group velocity $|\mathbf{v}| < 1$, and hence ϕ may be regarded as a *massive* field by analogy with a massive particle. The real parameter μ in (5.33) is called the *mass* of the field ϕ.

Fields governed by *linear* equations of motion are *free*. Expression (5.31) is an example of the Lagrangian for a free field. On the other hand, if the Lagrangian involves powers of ϕ higher than quadratic, such as

$$\mathcal{L} = \frac{1}{2} \, (\partial_\alpha \phi)(\partial^\alpha \phi) - \frac{\mu^2}{2} \, \phi^2 - \frac{\lambda^2}{4} \, \phi^4 \,, \tag{5.34}$$

then the resulting Euler–Lagrange equations are *nonlinear*, and the system is referred to as *interacting*. The parameter λ^2 in (5.34) measures the strength of the quartic self-interaction.

Next in order of complexity to scalars are Dirac spinors. Let ψ be a complex spinor field, and $\bar{\psi} = \psi^* \gamma^0$ the Dirac conjugate of ψ. As indicated in Appendix C, the simplest bilinear Lorentz covariant combinations of these fields are $\bar{\psi}\psi$ and $\bar{\psi}\gamma^\mu\psi$. We employ them to construct the Lagrangian

$$\mathcal{L} = \frac{i}{2}\,\bar{\psi}\gamma^\mu \left(\partial_\mu - \overleftarrow{\partial}_\mu\right)\psi - m\,\bar{\psi}\psi\;, \tag{5.35}$$

or, dropping a divergence term,

$$\mathcal{L} = \bar{\psi}\left(i\gamma^\mu\partial_\mu - m\right)\psi\;. \tag{5.36}$$

Rather than separately vary real and imaginary parts of the complex-valued Dirac field ψ, we instead regard ψ and $\bar{\psi}$ as two independent fields. The Euler–Lagrange equations for ψ read

$$\left(i\gamma^\mu\partial_\mu - m\right)\psi = 0\;. \tag{5.37}$$

This equation is called the *Dirac equation*; it governs a free spin-$\frac{1}{2}$ field. For $\bar{\psi}$, we get

$$\bar{\psi}\left(i\gamma^\mu\overleftarrow{\partial}_\mu + m\right) = 0\;. \tag{5.38}$$

A comparison between (5.36) and (5.37), or between (5.32) and (5.33), tells us how the Lagrangian of a free field can be directly reproduced from the field equation.

Applying $i\gamma^\mu\partial_\mu + m$ to the left of (5.37) and taking into account the anticommutation relations for Dirac matrices

$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2\eta^{\mu\nu}\;, \tag{5.39}$$

we obtain

$$\left(\Box + m^2\right)\psi = 0\;. \tag{5.40}$$

From this and the preceding discussion it follows that m represents the mass of the Dirac field ψ.

Since the Lagrangian (5.35) is linear in derivatives, the Dirac equation (5.37) is a first order partial differential equation. It is possible to associate formally a similar Lagrangian with a scalar field (Problem 5.1.2), but the field equations we eventually reach still contain second derivatives.

Free spin-1, and spin-$\frac{3}{2}$ fields are explored in Problems 5.1.4 and 5.1.5.

An example of a self-interacting spinor field, governed by a nonlinear field equation, is the *Gürsey equation*

$$i\gamma^\mu\partial_\mu\psi - \sigma\left(\bar{\psi}\psi\right)^{\frac{1}{3}}\psi = 0\;, \tag{5.41}$$

resulting from the Lagrangian

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - \frac{3}{4}\,\sigma\left(\bar{\psi}\psi\right)^{\frac{4}{3}}\;. \tag{5.42}$$

Here, σ is the coupling strength to measure the Gürsey self-interaction.

There are many possibilities to build Lorentz invariant terms $\mathcal{L}_{\rm int}$ responsible for the interaction between the Dirac and Klein–Gordon fields. The simplest example is the *Yukawa term*

$$\mathcal{L}_{\rm int} = -g\,\bar{\psi}\psi\phi\ , \tag{5.43}$$

where g is the Yukawa coupling constant which measures the force between ψ and ϕ.

A simple version for the direct coupling of four Dirac fields of different species, say, those of electron ψ_e, electron neutrino ψ_{ν_e}, muon ψ_μ, and muon neutrino ψ_{ν_μ}, is given by the vector *four-fermion* interaction term

$$\mathcal{L}_{\rm int} = -\frac{G_F}{\sqrt{2}}\,J^\alpha J^\dagger_\alpha\ , \tag{5.44}$$

$$J^\alpha = \bar{\psi}_e O^\alpha \psi_{\nu_e} + \bar{\psi}_\mu O^\alpha \psi_{\nu_\mu},\ \ J^\dagger_\alpha = \bar{\psi}_{\nu_e} O_\alpha \psi_e + \bar{\psi}_{\nu_\mu} O_\alpha \psi_\mu,\ \ O^\alpha = \frac{1}{2}\,(1+\gamma_5)\,\gamma^\alpha\ . \tag{5.45}$$

G_F is the Fermi constant which measures the joining strength of this quadruple bounded Dirac field compound (the factor $1/\sqrt{2}$ is kept following a common practice).

Let us turn to electrodynamics. A suitable Lagrangian for the electromagnetic field can be built from the invariants $\mathcal{S}$ and $\mathcal{P}$ introduced in Sect. 2.3. Both are quadratic in the field strength F. To ensure that the Euler–Lagrange equations be linear in F, the Lagrangian should be a linear combination of $\mathcal{S}$ and $\mathcal{P}$. However, expressed in terms of the vector potentials, $\mathcal{P}$ is a total divergence,

$$\mathcal{P} = \frac{1}{2}F_{\mu\nu}{}^*F^{\mu\nu} = \frac{1}{4}\epsilon^{\mu\nu\alpha\beta}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial_\alpha A_\beta - \partial_\beta A_\alpha) = \partial_\mu(\epsilon^{\mu\nu\alpha\beta}A_\nu\partial_\alpha A_\beta)\ . \tag{5.46}$$

Omitting $\mathcal{P}$, we can write the Lagrangian of a free electromagnetic field as

$$\mathcal{L} = -\frac{1}{16\pi}\,F_{\mu\nu}F^{\mu\nu}\ , \tag{5.47}$$

or, in vector notation,

$$\mathcal{L} = \frac{1}{8\pi}\left(\mathbf{E}^2 - \mathbf{B}^2\right)\ . \tag{5.48}$$

The overall factor of $(16\pi)^{-1}$ in (5.47), and $(8\pi)^{-1}$ in (5.48), characterize the Gaussian units. The Lagrangian (5.48) was discovered by Joseph Larmor in 1900. By $F_{\mu\nu}$ is meant

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu\ , \tag{5.49}$$

which is equivalent to the Bianchi identity

$$\mathcal{E}_{\lambda\mu\nu} = \partial_\lambda F_{\mu\nu} + \partial_\nu F_{\lambda\mu} + \partial_\mu F_{\nu\lambda} = 0\ . \tag{5.50}$$

Substituting (5.49) in (5.47) leads to the Euler–Lagrange equations

$$(\Box\,\eta_{\mu\nu} - \partial_\mu\partial_\nu)\,A^\nu = 0\;. \tag{5.51}$$

We now return to the general line of our discussion. How can the Lagrangian formalism for a particle in an external field be combined with that for fields? A key observation is that the interaction of the electromagnetic field with an external current J^μ originates from

$$\mathcal{L}_{\rm int} = -J_\mu A^\mu\;. \tag{5.52}$$

If we add (5.47) and (5.52), then the resulting Lagrangian gives the Euler–Lagrange equations

$$(\Box\,\eta_{\mu\nu} - \partial_\mu\partial_\nu)\,A^\nu = 4\pi J_\mu\;. \tag{5.53}$$

This suggests that we choose the action to have the Poincaré–Planck term (2.207) for the particle sector, the Larmor term (5.47) for the field sector, and the Schwarzschild term (2.212) for the interaction between a charged particle and the electromagnetic field:

$$S = -m\int d\tau\,\sqrt{\dot z^\mu\,\dot z_\mu} - \frac{1}{16\pi}\int d^4x\,F_{\mu\nu}F^{\mu\nu} - e\int d\tau\,\dot z^\mu A_\mu(z)\;. \tag{5.54}$$

A remarkable fact is that the action (5.54) offers a complete description of the closed system of a charged particle and an electromagnetic field. Indeed, we may rewrite (5.54) in the form (Problem 5.1.3)

$$S = -m\int d\tau\,\sqrt{\dot z^\mu\,\dot z_\mu} - \int d^4x\left(\frac{1}{16\pi}\,F_{\mu\nu}F^{\mu\nu} + j^\mu A_\mu\right) \tag{5.55}$$

where

$$j^\mu(x) = e\int_{-\infty}^{\infty} d\tau\,v^\mu(\tau)\,\delta^4[x - z(\tau)]\;\;. \tag{5.56}$$

Varying the action (5.55) with respect to A^μ gives Maxwell's equations

$$\mathcal{E}_\mu = \frac{1}{4\pi}\,(\Box\,\eta_{\mu\nu} - \partial_\mu\partial_\nu)\,A^\nu - j_\mu = 0\;, \tag{5.57}$$

and varying the action (5.54) with respect to z^μ results in the equation of motion for a charged particle

$$\varepsilon^\lambda = m\,\ddot z^\lambda - e\dot z_\mu F^{\lambda\mu}(z) = 0\;. \tag{5.58}$$

Joint solutions to the set of equations (5.57) and (5.58) will tell us all we need to know about the behavior of this closed system of a particle and electromagnetic field.

The Lagrangian description of particles interacting with other fields follow essentially the same pattern (Problem 5.1.6).

A few words are in order concerning dimensions. Using natural units, one can define the *length dimension* of any field and parameter entering the Lagrangian. To illustrate, consider the Lagrangian (5.31). Since S is dimensionless, $\mathcal{L}$ has dimension $[l]^{-4}$. The first term of $\mathcal{L}$, containing derivatives, has dimension $[l]^{-2}[\phi]^2$, whence it follows that

$$[\phi] = [l]^{-1} \,. \tag{5.59}$$

We can then count dimensions by saying that scalar fields have dimension -1. Likewise, the vector potential A_μ, and the parameters μ and m have dimension -1, while the Dirac field ψ has dimension $-\frac{3}{2}$. Coupling constants appearing as factors of the Lagrangian interaction terms may have different dimensions. It is easy to see that the quartic, Gürsey, Yukawa, and Schwarzschild terms are distinguished among other interacting terms since λ, σ, g, and e are dimensionless. By contrast, the Fermi constant G_F has dimension 2.

Problem 5.1.1. Prove that

$$\mathcal{L} = \frac{1}{2}\,(\partial_\alpha\phi)(\partial^\alpha\phi) - \frac{\mu^2}{2}\,\phi^2 - C\phi \,, \tag{5.60}$$

where C is a parameter, is equivalent to the Lagrangian (5.31).

Proof

$$\frac{\mu^2}{2}\,\phi^2 + C\phi = \frac{\mu^2}{2}\,(\phi + C)^2 - \frac{\mu^2 C^2}{2} \,. \tag{5.61}$$

Introducing $\Phi = \phi + C$ and omitting an irrelevant constant term, we come to (5.31) where ϕ is substituted by Φ.

Problem 5.1.2. Show that

$$\mathcal{L} = \frac{1}{2}\left(\phi_\mu\partial^\mu\phi - \phi\partial^\mu\phi_\mu - \phi^\mu\phi_\mu - \mu^2\phi^2\right) \tag{5.62}$$

gives field equations equivalent to those resulting from the Lagrangian (5.31).

Problem 5.1.3. Consider a charged particle moving along a smooth timelike world line $z^\mu(s)$ which intersects spacelike hypersurfaces Σ' and Σ'' at points s' and s'', respectively. Show that

$$\int_{\Sigma'}^{\Sigma''} d^4x\, j_\mu A^\mu = e\int_{s''}^{s'} ds\, v^\mu(s)\, A_\mu(z) \,. \tag{5.63}$$

Hint

$$\int_{\Sigma'}^{\Sigma''} d^4x\, \delta^4[x - z(s)] = \theta(s'' - s)\,\theta(s - s') \,. \tag{5.64}$$

Problem 5.1.4. Let ϕ^μ be a massive, real-valued field coupled to a prescribed external current J^μ which is assumed to be conserved, $\partial_\mu J^\mu = 0$. The Lagrangian of this system is given by

$$\mathcal{L} = \frac{1}{8\pi}\,\phi_\mu \left[(\Box\,\eta_{\mu\nu} - \partial_\mu\partial_\nu) + M^2\eta_{\mu\nu}\right]\phi^\nu - J_\mu\phi^\mu \; . \tag{5.65}$$

This vector field, endowed with the mass M, is known as the *Proca field.* Determine the dimension of ϕ_μ. Derive the Euler–Lagrange equations, and show that they imply

$$\partial_\mu\phi^\mu = 0 \; . \tag{5.66}$$

Problem 5.1.5. A spin-$\frac{3}{2}$ field, the *Rarita–Schwinger field*, is described by a Dirac spinor with a Lorentz vector index, ψ^μ. The Lagrangian for a free Rarita–Schwinger field is

$$\mathcal{L} = \epsilon^{\mu\nu\rho\sigma}\bar{\psi}_\mu\gamma_5\gamma_\nu\partial_\rho\psi_\sigma + \frac{1}{2}\kappa\bar{\psi}_\mu(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu)\psi_\nu \; , \tag{5.67}$$

where κ is the mass of this field. Show that the Euler–Lagrange equations resulting from this Lagrangian is

$$\epsilon^{\mu\nu\rho\sigma}\gamma_5\gamma_\nu\partial_\rho\psi_\sigma + \frac{1}{2}\kappa(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu)\psi_\nu \; , \tag{5.68}$$

which implies

$$\left(\Box + \kappa^2\right)\psi^\mu = 0, \quad \partial_\mu\psi^\mu = 0, \quad \gamma_\mu\psi^\mu = 0 \; . \tag{5.69}$$

Determine the length dimension of ψ_μ. Show that (5.68) is equivalent to

$$\left[(i\gamma^\alpha\partial_\alpha - \kappa)\,\eta_{\mu\nu} - \frac{i}{3}\,(\gamma_\mu\partial_\nu + \gamma_\nu\partial_\mu) + \frac{1}{3}\,\gamma_\mu\,(i\gamma^\alpha\partial_\alpha + \kappa)\,\gamma_\nu\right]\psi^\nu = 0 \; . \tag{5.70}$$

Hint Use equation $(C.37)$

$$3!\gamma_5\gamma_\lambda = -i\epsilon_{\lambda\mu\nu\rho}\gamma^\mu\gamma^\nu\gamma^\rho \; , \tag{5.71}$$

and shift the field $\psi_\mu \to \psi_\mu + C\gamma_\mu(\gamma\cdot\psi)$ where C is some coefficient.

Problem 5.1.6. Consider a system of N point particles interacting with a massive scalar field. (This can be viewed as a classical version of the Yukawa model for strongly interacting nucleons which are assumed to be held together in atomic nuclei by a massive pseudoscalar pion field.) Write the action for this system.

Answer

$$S = -\sum_{I=1}^{N}\int d\tau_I\,[m_I - g_I\phi(z_I)]\,\sqrt{\dot{z}_I\cdot\dot{z}_I} + \frac{1}{2}\int d^4x\,\left(\partial_\alpha\phi\,\partial^\alpha\phi - \mu^2\phi^2\right) \; , \tag{5.72}$$

or, alternatively,

$$S = -\sum_{I=1}^{N} m_I \int d\tau_I \sqrt{\dot{z}_I^2} + \int d^4x \left[\frac{1}{2}\left(\partial_\alpha \phi\, \partial^\alpha \phi - \mu^2\phi^2\right) + j\phi\right] , \tag{5.73}$$

$$j(x) = \sum_{I=1}^{N} g_I \int_{-\infty}^{\infty} d\tau_I \sqrt{\dot{z}_I^2}\, \delta^4[x - z_I(\tau_I)] \ . \tag{5.74}$$

Problem 5.1.7. Let two particles interact through a massive scalar field. Suppose that the charges of these particles g_1 and g_2 are sufficiently small, and the motion is nonrelativistic.

Prove that the force is attractive if g_1 and g_2 are of like sign, and repulsive if g_1 and g_2 are of opposite sign. With reference to Problem 5.1.6, this will elucidate why nucleons attract each other, and are thus bound in a nucleus, and why a neutron repels an antineutron. Is this result sensitive to the sign of the coupling constant g in (5.74)? Does this statement hold for a scalar field of zero mass $\mu = 0$?

Proof Assuming that $g\phi \ll m$, the equation of motion for a particle

$$\frac{d}{ds}\left[(m - g\phi)v_\mu\right] = -g\partial_\mu \phi \tag{5.75}$$

simplifies

$$m\mathbf{a} = g\nabla\phi \tag{5.76}$$

as $\mathbf{v} \to \mathbf{0}$. The positive sign of the gradient term is due to the fact that $v_\mu = (\gamma, -\gamma\mathbf{v})$ and $\partial_\mu = (\partial/\partial t, \nabla)$. With reference to Problem 4.1.4, we give the static field equation

$$\left(\nabla^2 - \mu^2\right)\phi(\mathbf{r}) = -g\,\delta^3(\mathbf{r}) \ , \tag{5.77}$$

and its solution

$$\phi(\mathbf{r}) = g\,\frac{\exp(-\mu r)}{4\pi r} \ . \tag{5.78}$$

Let $\mathbf{r}_{12}$ be the radius vector drawn from particle 1 to particle 2. Then the force $\mathbf{f}_{12}$ acting on particle 1 is

$$\mathbf{f}_{12} = -g_1 g_2\,\frac{\exp(-\mu_2\, r_{12})}{4\pi r_{12}^3}\,(1 + \mu r)\,\mathbf{r}_{12} \ . \tag{5.79}$$

We see that $\mathbf{f}_{12}$ is opposite to $\mathbf{r}_{12}$ if $g_1 g_2 > 0$.

Problem 5.1.8. Consider a point particle interacting with the Proca field,

$$S = -\int d\tau\, [m\sqrt{\dot{z}\cdot\dot{z}} + g\dot{z}^\mu \phi_\mu(z)] - \frac{1}{16\pi}\int d^4x\, [(\partial_\mu\phi_\nu - \partial_\nu\phi_\mu)^2 - 2M^2\phi_\mu\phi^\mu] \ . \tag{5.80}$$

This is a toy model for the weak interactions which assumes that the weak forces between particles (say, electrons and neutrinos) are carried by a massive vector field, the classical realization of the W and Z bosons. Show that the force between particles 1 and 2 is attractive if their charges g_1 and g_2 are of opposite sign, and repulsive if g_1 and g_2 are of like sign. In particular, a neutrino repels neutrinos and attracts antineutrinos.

5.2 Poincaré Invariance

A major requirement for a theory specified by a Lagrangian $\mathcal{L}$ is that $\mathcal{L}$ be invariant under Poincaré transformations. This ensures relativistic invariance. To be more specific, write a coordinate transformation

$$x_\mu{}' = \Lambda_\mu{}^\nu x_\nu + c_\mu \tag{5.81}$$

obeying the pseudoorthogonality condition

$$\Lambda_\lambda{}^\mu \Lambda_\mu{}^\nu = \delta_\lambda{}^\nu \ , \tag{5.82}$$

and suppose that (5.81) induces the field transformation

$$\phi_a{}'(x') = U_a{}^b \phi_b(x) \ , \tag{5.83}$$

where $U_a{}^b$ is the matrix of some irreducible representation of the Lorentz group. We require that $\mathcal{L}$ be invariant under these transformations up to a divergence:

$$\mathcal{L}(x', \phi', \partial'\phi') = \mathcal{L}(x, \phi, \partial\phi) + \partial_\mu B^\mu \ . \tag{5.84}$$

We first consider the consequences of *translation invariance*. Letting

$$x_\mu{}' = x_\mu + \epsilon_\mu \ , \tag{5.85}$$

where ϵ_μ is a fixed infinitesimal vector, and $U_a{}^b = \delta_a{}^b$, we have

$$\phi_a{}'(x') = \phi_a(x' - \epsilon) \ . \tag{5.86}$$

To a first approximation,

$$\phi_a{}'(x') = \phi_a(x') - \epsilon^\mu \partial'_\mu \phi_a(x') \ , \tag{5.87}$$

which implies that $\delta\phi_a = \phi_a{}'(x) - \phi_a(x) = -\epsilon^\mu \partial_\mu \phi_a$, and

$$\Delta\phi_a = 0 \ . \tag{5.88}$$

The generators $\Gamma^\nu{}_\mu$ and G_{ka} appearing in (5.24) and (5.25) take the form

$$\Gamma^\nu{}_\mu = \delta^\nu{}_\mu, \quad G_{ka} = 0 \ . \tag{5.89}$$

Consequently the Noether current associated with translation invariance is identical to the canonical stress-energy tensor

$$\theta_{\mu\nu} = \frac{\partial \mathcal{L}}{\partial(\partial^\mu \phi_a)}\, \partial_\nu \phi_a - \eta_{\mu\nu}\, \mathcal{L}\;, \tag{5.90}$$

and the corresponding conserved integral quantity is the *four-momentum* of the field

$$P^\mu = \int_\Sigma d\sigma_\lambda\, \theta^{\lambda\mu}\;. \tag{5.91}$$

By Noether's first theorem [see (5.16)],

$$\partial^\mu \theta_{\mu\nu} = -\mathcal{E}^a\, \partial_\nu \phi_a\;, \tag{5.92}$$

where $\mathcal{E}^a$ is the Eulerian resulting from the variation of the action with respect to the field variable ϕ_a. If $\mathcal{E}^a = 0$, then P^μ is independent of Σ. In particular, P^μ does not vary under timelike shifts of Σ. It is this fact which is usually understood as energy-momentum conservation: P^μ is constant on extremals.

Let us turn to the consequences of invariance under proper orthochronous *Lorentz transformations*. Writing

$$x'_\mu = x_\mu + \omega_\mu{}^\nu x_\nu, \quad \omega_{\mu\nu} = -\omega_{\nu\mu}\;, \tag{5.93}$$

$$U^b{}_a = \delta^b{}_a + \frac{1}{2}\,\omega_{\alpha\beta}\,(\Gamma^{\alpha\beta})^b{}_a, \quad \Gamma^{\alpha\beta} = -\Gamma^{\beta\alpha}\;, \tag{5.94}$$

where the $\omega_{\nu\mu}$ are parameters of infinitesimal Lorentz transformations, and the $(\Gamma^{\alpha\beta})^b{}_a$ are the spin matrices in the representation of the Lorentz group according to which the field ϕ_a transforms, gives

$$\Delta x_\mu = \omega_\mu{}^\nu x_\nu\;, \tag{5.95}$$

$$\Delta \phi_a = \frac{1}{2}\,\omega_{\mu\nu}\,(\Gamma^{\mu\nu})^b{}_a\, \phi_b\;. \tag{5.96}$$

The spin matrices for common representations are

$$(\Gamma^{\mu\nu})^\alpha{}_\beta = \begin{cases} 0, & \text{if } \phi_a \text{ is a scalar field } \phi, \\ \eta^{\alpha\mu}\delta^\nu{}_\beta - \eta^{\alpha\nu}\delta^\mu{}_\beta & \text{if } \phi_a \text{ is a vector field } A_\mu, \\ \frac{1}{4}\,(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu)^\alpha{}_\beta & \text{if } \phi_a \text{ is a spinor field } \psi_\alpha. \end{cases} \tag{5.97}$$

With (5.95) and (5.96), we arrive at the pertinent Noether current

$$M_{\lambda\mu\nu} = \theta_{\lambda\mu}\, x_\nu - \theta_{\lambda\nu}\, x_\mu - \Sigma_{\lambda\mu\nu}\;, \tag{5.98}$$

$$\Sigma_{\lambda\mu\nu} = \pi^a_\lambda\,(\Gamma_{\mu\nu})^b{}_a \phi_b\;, \tag{5.99}$$

called the angular momentum density, and the corresponding conserved integral

$$M_{\mu\nu} = \int_{\Sigma} d\sigma^{\lambda}\, M_{\lambda\mu\nu} \ , \tag{5.100}$$

called the *angular momentum tensor*.

It is conventional to split $M_{\mu\nu}$ into *orbital* $L_{\mu\nu}$ and *spin* $S_{\mu\nu}$ angular momentum:

$$M_{\mu\nu} = L_{\mu\nu} + S_{\mu\nu} \ , \tag{5.101}$$

$$L_{\mu\nu} = \int_{\Sigma} d\sigma^{\lambda}\, (\theta_{\lambda\mu}\, x_{\nu} - \theta_{\lambda\nu}\, x_{\mu}) \ , \tag{5.102}$$

$$S_{\mu\nu} = - \int_{\Sigma} d\sigma^{\lambda}\, \Sigma_{\lambda\mu\nu} \ . \tag{5.103}$$

An important fact concerning $\theta_{\mu\nu}$ is that this Noether current is defined up to adding the divergence of an antisymmetric tensor. Indeed, in view of the identity

$$\partial^{\mu}\partial^{\lambda} B_{\lambda\mu\nu} = 0 \tag{5.104}$$

which is true for any $B_{\lambda\mu\nu}$ such that

$$B_{\lambda\mu\nu} = -B_{\mu\lambda\nu} \ , \tag{5.105}$$

a modified quantity

$$T_{\mu\nu} = \theta_{\mu\nu} + \partial^{\lambda} B_{\lambda\mu\nu} \tag{5.106}$$

is a further Noether current:

$$\partial^{\mu} T_{\mu\nu} = -\mathcal{E}^{a}\, \partial_{\nu}\phi_{a} \ . \tag{5.107}$$

As shown in Problem 5.2.1, replacing $\theta_{\mu\nu}$ by $T_{\mu\nu}$, leaves P^{μ} unchanged. We thus deal with an entire equivalence class of stress-energy tensors[2] related to one other by (5.106).

If the equation of continuity for some J^{μ} holds identically, then this J^{μ} is called a *strongly* conserved current, but if the equation of continuity for J^{μ} holds due to the Euler–Lagrange equations, then this J^{μ} is called a *weakly* conserved current. As an example of a strongly conserved current we refer to $\partial^{\lambda} B_{\lambda\mu\nu}$ in (5.104), while $T_{\mu\nu}$ in (5.107) provides an example of a weakly conserved current.

With a suitable choice of $B_{\lambda\mu\nu}$, it is possible to make $T_{\mu\nu}$ symmetric:

$$T_{\mu\nu} = T_{\nu\mu} \ . \tag{5.108}$$

Note that the symmetric $T_{\mu\nu}$ is not unique (Problem 5.2.2). In 1939 Federik Belinfante put forward $B_{\lambda\mu\nu}$ of the form

$$B_{\lambda\mu\nu} = \frac{1}{2} \left(\Sigma_{\lambda\mu\nu} + \Sigma_{\mu\nu\lambda} + \Sigma_{\nu\mu\lambda} \right) \ , \tag{5.109}$$

[2] The impossibility of fixing the stress-energy tensor $T_{\mu\nu}$ uniquely is partly due to the fact that the Lagrangian is itself only defined modulo divergence terms.

where $\Sigma_{\lambda\mu\nu}$ is given by (5.99). Clearly $B_{\lambda\mu\nu}$ thus constructed obeys the antisymmetry condition (5.105) because

$$\Sigma_{\lambda\mu\nu} = -\Sigma_{\lambda\nu\mu} \tag{5.110}$$

It will be left to Problem 5.2.3 to show that $T_{\mu\nu}$ of the form of (5.106) with $B_{\lambda\mu\nu}$ defined in (5.109) and (5.99) is a symmetric tensor. The weakly conserved current $T_{\mu\nu}$ constructed according to the Belinfante prescription is called the *symmetric stress-energy tensor.*

Consider

$$J_{\lambda\mu\nu} = T_{\lambda\mu}x_\nu - T_{\lambda\nu}x_\mu \tag{5.111}$$

where $T_{\mu\nu}$ is the symmetric stress-energy tensor. Because $\partial^\lambda T_{\lambda\mu} = 0$,

$$\partial^\lambda J_{\lambda\mu\nu} = T_{\nu\mu} - T_{\mu\nu} = 0 \, . \tag{5.112}$$

Therefore, $J_{\lambda\mu\nu}$ is also a weakly conserved current.

Of course, there would be little point in symmetrizing $\theta_{\mu\nu}$ for pure aesthetic reasons. We will see, however, that the canonical stress-energy tensor for the electromagnetic field is not gauge invariant, while its symmetrized modification is.

In 1915 David Hilbert proposed an alternative definition of the stress-energy tensor which is particularly attractive in that it is symmetric from the outset. Moreover, this definition is gauge invariant for gauge invariant actions.

The basic idea is to consider the variation of the action in response to a small variation of an external field, and then treat the metric as that external field. In fact, there is nothing to prevent us from using noninertial frames or curvilinear coordinates. Let us take curvilinear coordinates y^μ which are deviate slightly from the conventional Cartesian coordinates x^μ. Then tangent vectors of coordinate lines will transform as

$$dy^\mu = \frac{\partial y^\mu}{\partial x^\alpha}\, dx^\alpha = dx^\mu + \xi^\mu \, , \tag{5.113}$$

where ξ^μ is an infinitesimal vector which measures the coordinate deviation at the point x^μ. For the line element $ds^2 = \eta_{\mu\nu}\, dx^\mu dx^\nu$ to be invariant under this transformation, the metric must transform according to the law

$$g_{\alpha\beta} = \frac{\partial x^\mu}{\partial y^\alpha}\, \frac{\partial x^\nu}{\partial y^\beta}\, \eta_{\mu\nu} = \left(\delta^\mu{}_\alpha + \partial_\alpha \xi^\mu\right)\left(\delta^\nu{}_\beta + \partial_\beta \xi^\nu\right)\eta_{\mu\nu} \, . \tag{5.114}$$

Thus, the variation of the metric due to this 'external' coordinate variation is given by

$$\delta g_{\alpha\beta} = \partial_\alpha \xi_\beta + \partial_\beta \xi_\alpha \, . \tag{5.115}$$

We now refine our previous definition of the action to take the form

$$S = \int d^4x \, \sqrt{-g}\, \mathcal{L} \, , \tag{5.116}$$

where g stands for the determinant of the metric in curvilinear coordinates

$$g = \det\left(g_{\alpha\beta}\right) \ . \tag{5.117}$$

Because the line element will continue to have the signature $(1, -1, -1, -1)$, even in curvilinear coordinates where the metric tensor may not be diagonal, $\sqrt{-g}$ is always real. With reference to Problem 1.4.2, the action (5.116) turns out to be manifestly invariant under general coordinate transformations $x^\mu = f^\mu(y)$.

If $g_{\mu\nu} = \eta_{\mu\nu} + \delta g_{\mu\nu}$, then the corresponding variation of the action (5.116) is given by

$$\delta S = \frac{1}{2}\int d^4x\, \sqrt{-g}\, \mathcal{T}_{\mu\nu}\, \delta g^{\mu\nu} \ . \tag{5.118}$$

Following Hilbert, we define the *metric stress-energy tensor* by

$$\mathcal{T}_{\mu\nu} = \frac{2}{\sqrt{-g}}\, \frac{\delta S}{\delta g^{\mu\nu}} \ . \tag{5.119}$$

For practical implementation of this definition, it is convenient to use partial derivatives rather than the variational derivative. One can deduce (Problem 5.2.4) the relation

$$\delta g = -g\, g_{\alpha\beta}\, \delta g^{\alpha\beta} \tag{5.120}$$

and apply it to the variation of the action (5.116),

$$\delta S = \int d^4x \left[\delta(\sqrt{-g})\, \mathcal{L} + \sqrt{-g}\, \delta\mathcal{L}\right] = \frac{1}{2}\int d^4x\, \sqrt{-g}\left(-g_{\mu\nu}\mathcal{L} + 2\,\frac{\partial \mathcal{L}}{\partial g^{\mu\nu}}\right)\delta g^{\mu\nu} \ . \tag{5.121}$$

On putting $\sqrt{-g} = 1$, which is valid for the Minkowski metric $g_{\mu\nu} = \eta_{\mu\nu}$, we obtain

$$\mathcal{T}_{\mu\nu} = 2\,\frac{\partial \mathcal{L}}{\partial g^{\mu\nu}} - g_{\mu\nu}\, \mathcal{L} \ . \tag{5.122}$$

One additional remark needs to be made. To differentiate with respect to a *symmetric* tensor $\sigma^{\mu\nu}$, we should select *independent* components of $\sigma^{\mu\nu}$. If some scalar is expressed in terms of independent components of $\sigma^{\mu\nu}$, then each term of this scalar contributed by such components occurs twice, for example, the quantity σ_{0i} occurs twice in the expansion of $\sigma_{\mu\nu}\sigma^{\mu\nu} = -2\sigma_{0i}\sigma_{0i} + \sigma_{ij}\sigma_{ij}$. All things considered, we may use the differentiation rule

$$\frac{\partial \sigma^{\alpha\beta}}{\partial \sigma^{\mu\nu}} = \delta^{\alpha}_{\ \mu}\, \delta^{\beta}_{\ \nu} + \delta^{\alpha}_{\ \nu}\, \delta^{\beta}_{\ \mu} \ . \tag{5.123}$$

It turns out that for a large class of theories, the metric and symmetric stress-energy tensors are identical; thereafter we will use these terms interchangeably.

To illustrate the utility of the Hilbert approach, we evaluate the metric stress-energy tensor for a particle governed by the Poincaré–Planck action (2.211),

$$S_{\rm P} = -m \int \sqrt{g_{\mu\nu}(z)\, dz^\mu\, dz^\nu}\ . \tag{5.124}$$

Let us first rewrite $S_{\rm P}$ in the form

$$S_{\rm P} = -m \int_{\Sigma_1}^{\Sigma_2} d^4x \int_{-\infty}^{\infty} d\tau\, \delta^4\left[x - z(\tau)\right] \sqrt{g_{\mu\nu}(x)\, \dot z^\mu(\tau)\, \dot z^\nu(\tau)}\ , \tag{5.125}$$

where τ is an arbitrary parameter of evolution. Variation of $S_{\rm P}$ with respect to $g_{\mu\nu}$ gives

$$\delta S_{\rm P} = -\frac{m}{2} \int d^4x \int d\tau\, \delta^4\left[x - z(\tau)\right] \frac{\dot z^\mu\, \dot z^\nu}{\sqrt{g_{\alpha\beta}\, \dot z^\alpha\, \dot z^\beta}}\, \delta g_{\mu\nu}$$

$$= \frac{1}{2} \int d^4x \sqrt{-g} \left\{ m \int ds\, v^\alpha(s)\, v^\beta(s)\, \delta^4\left[x - z(s)\right] \right\} \frac{g_{\alpha\mu}\, g_{\beta\nu}}{\sqrt{-g}}\, \delta g^{\mu\nu}\ , \tag{5.126}$$

where the proper time $ds = \sqrt{g_{\alpha\beta}\, \dot z^\alpha\, \dot z^\beta}\, d\tau$ is used to parametrize the world line. We have employed the relation

$$\delta g_{\mu\nu} = -\delta g^{\alpha\beta}\, g_{\alpha\mu}\, g_{\beta\nu} \tag{5.127}$$

which results from varying

$$g_{\mu\nu} g^{\nu\lambda} = \delta_\mu^{\ \lambda}\ . \tag{5.128}$$

Comparing (5.118) and (5.126), we get

$$t_{\mu\nu} = m \int_{-\infty}^{\infty} ds\, v_\mu(s)\, v_\nu(s)\, \delta^4\left[x - z(s)\right]\ . \tag{5.129}$$

Note that the Schwarzschild action

$$S_{\rm S} = -e \int dz^\mu A_\mu \tag{5.130}$$

does not contribute to the metric stress-energy tensor. Indeed, $A_\mu\, dz^\mu$ may be thought of as the value of a 1-form A_μ on an infinitesimal tangent vector dz^μ, rather than the scalar product of two vectors A^μ and dz^μ. What this means is that $S_{\rm S}$ is metric independent. This is a special feature of the Schwarzschild interaction term which is not shared by other interactions. For example, the interaction of a particle with a scalar field (5.72) does contribute to the stress-energy tensor.

We next consider the Larmor action

$$S_{\rm L} = -\frac{1}{16\pi} \int d^4x\, (\partial^\mu A^\nu - \partial^\mu A^\nu)(\partial_\mu A_\nu - \partial_\nu A_\mu)\ . \tag{5.131}$$

The canonical stress-energy tensor corresponding to this term

$$\theta_{\mu\nu} = -\frac{1}{4\pi}\, F_\mu^{\ \lambda} \partial_\nu A_\lambda + \eta_{\mu\nu}\, \frac{1}{16\pi}\, F^{\alpha\beta} F_{\alpha\beta} \tag{5.132}$$

is not gauge invariant owing to the presence of the vector potential A_μ.

To symmetrize the stress-energy tensor, we calculate the divergence of the Belinfante term (5.109), using the middle line of (5.97),

$$\partial^\lambda B_{\lambda\mu\nu} = \frac{1}{4\pi}\,\partial^\lambda\left(F_{\mu\lambda}A_\nu\right) = \frac{1}{4\pi}\,F_{\mu\lambda}\partial^\lambda A_\nu + \frac{1}{4\pi}\,A_\nu\partial^\lambda F_{\mu\lambda}\ . \tag{5.133}$$

For a free electromagnetic field, the last term vanishes, and so

$$\partial^\lambda B_{\lambda\mu\nu} = \frac{1}{4\pi}\,F_\mu{}^\lambda\partial_\lambda A_\nu\ . \tag{5.134}$$

Combining (5.134) with (5.132), we obtain

$$\Theta_{\mu\nu} = \frac{1}{4\pi}\left(F_\mu{}^\alpha F_{\alpha\nu} + \frac{\eta_{\mu\nu}}{4}\,F^{\alpha\beta}F_{\alpha\beta}\right)\ . \tag{5.135}$$

We thus see that the Belinfante prescription makes the symmetric stress-energy tensor of a free electromagnetic field gauge invariant. It is a matter of straightforward computation (Problem 5.2.5) to check that the Hilbert definition (5.119) leads to the same expression.

Let us clarify the physical meaning of the various components of $\Theta_{\mu\nu}$. If it is granted that the quantity P^μ given by (5.91) characterizes the energy-momentum content of the electromagnetic field[3], then

$$\Theta_{00} = \frac{1}{8\pi}\left(\mathbf{E}^2 + \mathbf{B}^2\right) \tag{5.136}$$

may be interpreted as the *energy density*, and

$$\Theta_{i0} = \frac{1}{4\pi}\left(\mathbf{E}\times\mathbf{B}\right)_i \tag{5.137}$$

as the *momentum density* of the the electromagnetic field. An application of (5.136) to electrostatics is offered in Problem 5.2.6.

On the other hand, taking the equation of continuity

$$\partial_0\Theta^{00} + \partial_i\Theta^{i0} = 0\ , \tag{5.138}$$

integrating it over a spatial domain V with the boundary ∂V, and applying the Gauss–Ostrogradskiĭ theorem, gives

$$\frac{\partial}{\partial t}\int_V d^3x\,\Theta^{00} = -\int_{\partial V} dS_i\,\Theta^{i0}\ . \tag{5.139}$$

We see that the rate of change of energy of the electromagnetic field in a domain V is equal to the total flux across the boundary of this domain ∂V.

[3] An argument in support of this assumption, derived from mechanics, is that the symmetry responsible for energy and momentum conservation is spacetime homogeneity.

Consequently Θ^{i0} represents the flux of energy in the electromagnetic field. The three-vector $\mathbf{G} = (1/4\pi)\,(\mathbf{E}\times\mathbf{B})$ was introduced by John Henry Poynting in 1884, and is now referred to by his name.

Likewise, it is possible to show that the so-called Maxwell stress tensor

$$\Theta_{ij} = \frac{1}{4\pi}\left[E_iE_j + B_iB_j + \frac{1}{2}\,\delta_{ij}\left(\mathbf{E}^2 + \mathbf{B}^2\right)\right] \tag{5.140}$$

is the jth component of the flow of momentum in the electromagnetic field through a unit surface perpendicular to the x^i-axis.

Because the stress-energy tensor for the electromagnetic interaction described by the Schwarzschild Lagrangian vanishes, the total stress-energy tensor in the Maxwell–Lorentz electrodynamics reads

$$T^{\mu\nu} = t^{\mu\nu} + \Theta^{\mu\nu}\,, \tag{5.141}$$

where $t^{\mu\nu}$ and $\Theta^{\mu\nu}$ are given by (5.129) and (5.135).

To conclude this section, we state the local energy-momentum conservation law for a closed system of a charged particle and the electromagnetic field:

$$\partial_\mu T^{\lambda\mu} = \frac{1}{8\pi}\,\mathcal{E}^{\lambda\mu\nu}\,F_{\mu\nu} + \mathcal{E}_\mu\,F^{\mu\lambda} + \int ds\,\varepsilon^\lambda\,\delta^4\left[x - z(s)\right]\,, \tag{5.142}$$

where $\mathcal{E}^{\lambda\mu\nu}$, $\mathcal{E}_\mu$, and ε^λ stand for the right-hand sides of equations (5.50), (5.57), and (5.58), respectively. The proof of this identity is left to the reader (Problem 5.2.7).

Problem 5.2.1. Prove that one gets identical contributions to P^μ from $\theta_{\mu\nu}$ and $T_{\mu\nu}$ related by (5.106) where $B_{\lambda\mu\nu}$ is an arbitrary antisymmetric tensor vanishing at spatial infinity.

Proof Fixing a particular Lorentz frame in which a hyperplane Σ represents space, and choosing a domain V with spherical boundary ∂V of large radius R, we get

$$\int_\Sigma d\sigma^\mu\,\partial^\alpha B_{\alpha\mu\nu} = \int_V d^3x\,\partial^\alpha B_{\alpha 0\nu} = \int_V d^3x\,\partial^i B_{i0\nu} = \int_{\partial V} dS^i\,B_{i0\nu} \to 0 \tag{5.143}$$

as $R\to\infty$.

Problem 5.2.2. Let $\Theta_{\mu\nu}$ be the symmetric stress-energy tensor of a scalar field ϕ. Verify that

$$\tau_{\mu\nu} = \Theta_{\mu\nu} - C\left(\partial_\mu\partial_\nu - \eta_{\mu\nu}\Box\right)\phi^2 \tag{5.144}$$

is a further symmetric conserved tensor for any constant coefficient C. Show that the difference between $\tau_{\mu\nu}$ and $\Theta_{\mu\nu}$ does not affect the construction of P^μ.

Problem 5.2.3. Show that a tensor $T_{\mu\nu}$ of the form of (5.106), with the term $B_{\lambda\mu\nu}$ defined in (5.109), is a symmetric tensor.

Hint Comparing $T_{\mu\nu} - T_{\nu\mu} = \theta_{\mu\nu} - \theta_{\nu\mu} + 2\partial^\lambda B_{\lambda\mu\nu}$ with $\partial^\lambda M_{\lambda\mu\nu} = \theta_{\nu\mu} - \theta_{\mu\nu} + \partial^\lambda \Sigma_{\lambda\mu\nu} = 0$ where $M_{\lambda\mu\nu}$ is the Noether current defined in (5.98), one gets the desired result.

Problem 5.2.4. Prove (5.120).

Proof By (1.147),

$$g = \frac{1}{4!}\,\epsilon^{\kappa\lambda\mu\nu}\,\epsilon^{\alpha\beta\gamma\delta}\,g_{\kappa\alpha}\,g_{\lambda\beta}\,g_{\mu\gamma}\,g_{\nu\delta}\ , \tag{5.145}$$

which implies that

$$g\,\delta_\mu^{\ \nu} = g_{\mu\alpha}\,G^{\alpha\nu}\ , \tag{5.146}$$

where $G^{\alpha\nu}$ is the cofactor of $g_{\alpha\nu}$ given by

$$G^{\kappa\alpha} = \frac{1}{3!}\,\epsilon^{\kappa\lambda\mu\nu}\,\epsilon^{\alpha\beta\gamma\delta}\,g_{\lambda\beta}\,g_{\mu\gamma}\,g_{\nu\delta}\ . \tag{5.147}$$

We differentiate (5.145) to give

$$\frac{\partial g}{\partial g_{\mu\nu}} = G^{\mu\nu}\ , \tag{5.148}$$

which, in view of (5.146), yields

$$dg = g^{\mu\nu}dg_{\mu\nu}\ . \tag{5.149}$$

Combining this with (5.127), we arrive at (5.120).

Problem 5.2.5. Consider electrodynamics in a world of dimension $D+1$,

$$S_0 = \int d^{D+1}x\,\sqrt{-g}\,\mathcal{L}_0\ , \tag{5.150}$$

$$\mathcal{L}_0 = -\frac{1}{4\Omega_{D-1}}\,g_{\alpha\mu}\,g_{\beta\nu}\,F^{\alpha\beta}F^{\mu\nu}, \quad \Omega_{D-1} = 2\,\frac{\pi^{\frac{D}{2}}}{\Gamma(\frac{D}{2})}\ , \tag{5.151}$$

where $g_{\mu\nu}$ is the metric, and Ω_{D-1} the area of a $(D-1)$-dimensional unit sphere.

Find $\mathcal{T}_{\mu\nu}$ for the action (5.150)–(5.151).

Answer

$$\mathcal{T}_{\mu\nu} = \frac{1}{\Omega_{D-1}}\left(g^{\alpha\beta}F_{\mu\alpha}F_{\beta\nu} + \frac{g_{\mu\nu}}{4}\,F^{\alpha\beta}F_{\alpha\beta}\right)\ . \tag{5.152}$$

Problem 5.2.6. Show that the electrostatic energy is

$$\mathcal{E} = \int d^3x \, \frac{\mathbf{E}^2(\mathbf{x})}{8\pi} = \frac{1}{2} \int d^3x \, \phi(\mathbf{x}) \, \varrho(\mathbf{x}) \, . \tag{5.153}$$

Problem 5.2.7. Verify the Noether identity (5.142).

Problem 5.2.8. Let $\Theta^{\lambda\mu}$ be the symmetric stress-energy tensor of the electromagnetic field (5.135). Show that

$$\Theta^{\lambda\mu} = \frac{1}{8\pi} \left(F^{\lambda}_{\ \sigma} F^{\sigma\mu} + {}^*F^{\lambda}_{\ \sigma} {}^*F^{\sigma\mu} \right) \, . \tag{5.154}$$

Hint Use (2.97).

Problem 5.2.9. Prove that

$$\Theta^{\mu}_{\ \mu} = 0 \, , \tag{5.155}$$

$$\Theta^{\lambda\mu} \, \Theta_{\mu\nu} = \frac{1}{(8\pi)^2} \left(\mathcal{S}^2 + 4\mathcal{P}^2 \right) \delta^{\lambda}_{\ \nu} \, . \tag{5.156}$$

Proof To establish (5.155), we note that $\delta^{\mu}_{\ \mu} = 4$ if spacetime dimension is $D + 1 = 4$. Referring back to Problem 2.3.2, and equations (2.96) and (2.97),

$$\begin{gathered}(8\pi)^2 \Theta^{\lambda\mu} \, \Theta_{\mu\nu} = (F^{\lambda}_{\ \sigma} F^{\sigma\mu} + {}^*F^{\lambda}_{\ \sigma} {}^*F^{\sigma\mu})(F_{\mu\alpha} F^{\alpha}_{\ \nu} + {}^*F_{\mu\alpha} {}^*F^{\alpha}_{\ \nu}) \\ = F^{\lambda}_{\ \sigma} F^{\sigma\mu} F_{\mu\alpha} F^{\alpha}_{\ \nu} + F^{\lambda}_{\ \sigma} F^{\sigma\mu} {}^*F_{\mu\alpha} {}^*F^{\alpha}_{\ \nu} + {}^*F^{\lambda}_{\ \sigma} {}^*F^{\sigma\mu} F_{\mu\alpha} F^{\alpha}_{\ \nu} + {}^*F^{\lambda}_{\ \sigma} {}^*F^{\sigma\mu} {}^*F_{\mu\alpha} {}^*F^{\alpha}_{\ \nu} \\ = F^{\lambda}_{\ \sigma} F^{\sigma\mu} \left(-\mathcal{S} \eta_{\mu\nu} + {}^*F_{\mu\alpha} {}^*F^{\alpha}_{\ \nu} \right) + F^{\lambda}_{\ \sigma} \mathcal{P} \delta^{\sigma}_{\ \alpha} {}^*F^{\alpha}_{\ \nu} + {}^*F^{\lambda}_{\ \sigma} \mathcal{P} \delta^{\sigma}_{\ \alpha} F^{\alpha}_{\ \nu} \\ + {}^*F^{\lambda}_{\ \sigma} {}^*F^{\sigma\mu} (\mathcal{S} \eta_{\mu\nu} + F_{\mu\alpha} F^{\alpha}_{\ \nu}) = \mathcal{S}({}^*F^{\lambda}_{\ \sigma} {}^*F^{\sigma}_{\ \nu} - F^{\lambda}_{\ \sigma} F^{\sigma}_{\ \nu}) + 4\mathcal{P}^2 \delta^{\lambda}_{\ \nu} = (\mathcal{S}^2 + 4\mathcal{P}^2) \delta^{\lambda}_{\ \nu} \, .\end{gathered} \tag{5.157}$$

Problem 5.2.10. Prove that the orbital angular momentum of the electromagnetic field due to a particle of electric charge e and a static magnetic monopole of strength $e^\star$ is given by

$$\mathbf{L} = -e \, e^\star \, \mathbf{n}, \quad \mathbf{n} = \frac{\mathbf{r}}{r} \, . \tag{5.158}$$

Proof Assume that the magnetic monopole is at rest at the origin. By (4.334), it generates the magnetic field $\mathbf{B} = e^\star \, \frac{\mathbf{n}}{r^2}$. If this field is substituted in the expression for angular momentum of the electromagnetic field

$$\mathbf{L}_i = \frac{1}{2} \, \epsilon_{ijk} \int d^3x \, (\Theta_{0j} x_k - \Theta_{0k} x_j) = \frac{1}{4\pi} \int d^3x \, (\mathbf{x} \times \mathbf{E} \times \mathbf{B})_i \, , \tag{5.159}$$

then

$$\begin{gathered}\mathbf{L}_i = \frac{1}{4\pi} \int d^3x \, (\mathbf{x} \times \mathbf{E} \times \mathbf{B})_i = \frac{1}{4\pi} \int d^3x \, [\mathbf{E}_i \, (\mathbf{n} \cdot \mathbf{x}) - (\mathbf{E} \cdot \mathbf{x}) \, \mathbf{n}_i] \, \frac{e^\star}{r^2} \\ = \frac{e^\star}{4\pi} \int d^3x \, \mathbf{E}_l \, (\delta_{il} - \mathbf{n}_i \mathbf{n}_l) \, \frac{1}{r} = \frac{e^\star}{4\pi} \int d^3x \, \mathbf{E}_l \, \nabla_l \mathbf{n}_i = -\frac{e^\star}{4\pi} \int d^3x \, (\nabla \cdot \mathbf{E}) \, \mathbf{n}_i \, ,\end{gathered} \tag{5.160}$$

where the relation

$$\nabla_l \mathbf{n}_i = \frac{1}{r}\left(\delta_{il} - \mathbf{n}_i \mathbf{n}_l\right) \tag{5.161}$$

has been used. Now the desired result $\mathbf{L} = -e\, e^\star \mathbf{n}$ follows from $\nabla \cdot \mathbf{E}(\mathbf{x}) = 4\pi\delta^3(\mathbf{x} - \mathbf{z})$.

In quantum mechanics, the projection of $\mathbf{L}$ onto a fixed axis takes discrete values which are integral multiples of $\frac{1}{2}$ (in units of $\hbar$). This may serve as a heuristic derivation of Dirac's quantization condition

$$e\, e^\star = \frac{1}{2}\, n\ . \tag{5.162}$$

5.3 Conformal Invariance

Maxwell's equations are invariant not only under the 10-parameter Poincaré group of Lorentz transformations and translations, but under the larger, 15-parameter *conformal group* of spacetime transformations $\mathrm{C}(1,3)$. This fact was discovered by Harry Bateman and Ebenezer Cunningham in 1909. Since then the conformal symmetry has become a recurrent theme in field theory, at times fascinating, and often discouraging researchers.

Why does conformal symmetry create such a stir? Mathematically, the conformal group $\mathrm{C}(1,3)$ is the lowest dimensional group containing the Poincaré group[4]. Of special note is that $\mathrm{C}(1,3)$ is semisimple, even though the Poincaré group is the semidirect product of the Lorentz and translation groups. Conformal field theories are in general much more tractable than non-conformal theories differing from them by simple symmetry-breaking terms. The main physical concern with conformal invariance is due to the fact that $\mathrm{C}(1,3)$ is the largest spacetime symmetry of Maxwell's equations. Another point is that conformal invariance singles out the linear version of electromagnetism from all possible generalizations of Maxwell's theory; we will see in Sect. 10.4 that only linear equations of motion for the electromagnetic field are conformally invariant. Furthermore, Maxwell's equations enjoy the property of conformal invariance only in spacetime of dimension $D + 1 = 4$. This result, suggesting that the dimension 4 arises from conformal properties of our world, is the subject of profound philosophic speculations. Lastly, the two-dimensional conformal invariance is basic for string theory (which will be discussed in Sect. 5.6).

By the conformal group $\mathrm{C}(1,3)$ we mean the 15-parameter Lie group of nonlinear point transformations on Minkowski space which map an event x_μ into another event x'_μ such that

$$dx'^2 = e^{2\lambda(x)}\, dx^2\ . \tag{5.163}$$

[4] An overview of most generally employed mathematical properties of the conformal group $\mathrm{C}(m,n)$ acting on a pseudoeuclidean space $\mathbb{R}_{m,n}$ of dimension $D = m + n$ is given in Appendix D.

A null line element $dx^2 = 0$ is mapped onto another null line element $dx'^2 = 0$, and hence the light-cone structure of Minkowski space is left unchanged.

The infinitesimal transformations of $C(1,3)$ are described in Appendix D by formula (D.25):

$$\Delta x^\mu = \epsilon^\mu + \omega^{\mu\alpha} x_\alpha \ , \tag{5.164}$$

$$\Delta x^\mu = \gamma x^\mu \ , \tag{5.165}$$

$$\Delta x^\mu = 2\beta_\alpha x^\alpha x^\mu - x^2 \beta^\mu \ . \tag{5.166}$$

Apart from Poincaré transformations (5.164) (which are responsible for the existence of 10 conserved quantities P_μ and $M_{\mu\nu}$) there are infinitesimal *scale* transformations, or *dilatations* (5.165), and *special conformal transformations* (5.166). Combining Poincaré and scale transformations, we come to the group of *similitude* transformations.

If we require that a theory be invariant under Poincaré transformations (5.164) and special conformal transformations (5.166), then this theory is automatically scale-invariant. To see this, we refer to the fact that the commutator of generators of translations P_λ and special conformal transformations K_μ involves the generator of dilatations D [see (D.42)]. However, the converse does not follow: the Lie algebra of similitude transformations (D.33)–(D.38) is closed.

Let us begin with scale invariance. Consider a finite dilatation

$$x'_\mu = e^\gamma x_\mu \ . \tag{5.167}$$

To specify the transformation law for a field ϕ_a which is induced by (5.167), we define the *scale dimension* l_a of this field by

$$\phi'_a(x') = e^{l_a \gamma} \phi_a(x) \ . \tag{5.168}$$

An argument in favor of this transformation law is given in Problem 5.3.1.

Let γ be an infinitesimal parameter. Then

$$\Delta\phi_a = l_a \gamma \phi_a \ . \tag{5.169}$$

Referring to (5.28), we write the Noether current $\mathcal{D}_\mu$ associated with dilatations

$$\mathcal{D}_\mu = \theta_{\mu\nu}\, x^\nu - l_a \pi^a_\mu\, \phi_a \ . \tag{5.170}$$

Using (5.21) and (5.23), and the relation $\delta^\mu{}_\mu = 4$, we have

$$\partial^\mu \mathcal{D}_\mu = \theta^\mu{}_\mu - l_a \pi^a_\mu\, \partial^\mu \phi_a - l_a \frac{\partial \mathcal{L}}{\partial \phi_a}\, \phi_a = -4\mathcal{L} + (-l_a + 1)\, \pi^a_\mu\, \partial^\mu \phi_a - l_a \frac{\partial \mathcal{L}}{\partial \phi_a}\, \phi_a \ . \tag{5.171}$$

As a case in point we refer to Lagrangians of the type

$$\mathcal{L} = \frac{1}{2}\, (\partial_\mu \phi)(\partial^\mu \phi) - U(\phi) \tag{5.172}$$

which are pertinent to the description of scalar and vector fields, and those of the type

$$\mathcal{L} = \frac{i}{2}\,\bar{\psi}\gamma^\mu\left(\partial_\mu - \overleftarrow{\partial}_\mu\right)\psi - V(\psi) \tag{5.173}$$

which are appropriate to the treatment of spinor fields. It follows from (5.171) that if such theories are to be scale invariant ($\partial^\mu \mathcal{D}_\mu = 0$), then $l_\phi = -1$, and $l_\psi = -\frac{3}{2}$. (Despite the similarity between the scale dimension and the length dimension of fields in natural units, these concepts must not be confused. Indeed, all parameters in the Lagrangian, such as masses and coupling constants, are assigned zero scale dimension.) Thus, the theories of free massless scalar and spinor fields governed by the Klein–Gordon and Dirac Lagrangians (5.31) and (5.35) where the mass terms have been eliminated, as well as the free-field Maxwell's electrodynamics, are scale invariant.

A general scale invariant interacting term has scale dimension $l = -4$. This condition is clearly met for dimensionless coupling constants, exemplified by the Gürsey, Yukawa, and quartic interaction terms, shown, respectively, in (5.42), (5.43), and (5.34).

The dilatation current (5.170) can be simplified in a large class of field theories[5] through the so-called *improved* stress-energy tensor $\tau_{\mu\nu}$, which while differing from the Belinfante tensor $T_{\mu\nu}$ by extra terms, is still conserved and symmetric, and leaves P^μ unchanged. Given $\tau_{\mu\nu}$, the dilatation current becomes

$$\mathcal{D}_\mu = \tau_{\mu\nu}x^\nu \; . \tag{5.174}$$

The condition that a theory is scale invariant then reads

$$\partial^\mu \mathcal{D}_\mu = \tau^\mu_{\ \mu} = 0 \; . \tag{5.175}$$

Let us turn to conformal invariance. For theories of this class, the Noether current associated with conformal invariance is simply

$$\mathcal{K}_{\mu\nu} = \left(2x_\mu x^\alpha - x^2\delta^\alpha_{\ \mu}\right)\tau_{\alpha\nu} \; . \tag{5.176}$$

Conformal invariance leads to precisely the same consequence: the improved stress-energy tensor $\tau_{\mu\nu}$ must be traceless,

$$\partial^\nu \mathcal{K}_{\mu\nu} = 2x_\mu \tau^\nu_{\ \nu} = 0 \; . \tag{5.177}$$

We see that both scale and conformal invariance are measured by $\tau^\nu_{\ \nu}$. The breaking of scale invariance amounts to the breaking of conformal invariance, because

$$\partial^\nu \mathcal{K}_{\mu\nu} = 2x_\mu \partial^\nu \mathcal{D}_\nu \; . \tag{5.178}$$

Thus, scale invariance implies conformal invariance for the majority of Lagrangian field theories of physical interest (even though not for all).

[5] We cite this statement without proof, and refer to a particular case developed in Problem 5.3.3.

Note also that conformal invariance does not lead to new conserved quantities in addition to those arising from Poincaré invariance, P^μ and $M_{\mu\nu}$.

So far we said nothing about the transformation of fields due to conformal mappings of Minkowski space. The conventional way of obtaining representations of the conformal group from those of the Lorentz group is as follows. With $dx'^\mu = (\partial x'^\mu / \partial x^\alpha)\, dx^\alpha$, we write the defining equation for conformal transformations (5.163) in the form

$$\frac{\partial x'^\mu}{\partial x^\alpha} \frac{\partial x'^\nu}{\partial x^\beta} \eta_{\mu\nu} = e^{2\lambda(x)}\, \eta_{\alpha\beta} \; . \tag{5.179}$$

Taking the determinant

$$J^2 = \left[\det\left(\frac{\partial x'}{\partial x}\right)\right]^2 = e^{8\lambda(x)} \; , \tag{5.180}$$

(5.179) becomes

$$\frac{\partial x'^\mu}{\partial x^\alpha} \frac{\partial x'^\nu}{\partial x^\beta} \eta_{\mu\nu} = |J|^{\frac{1}{2}}\, \eta_{\alpha\beta} \; . \tag{5.181}$$

It follows that the Jacobian matrix of a conformal transformation $\partial x'^\mu / \partial x^\alpha$ is proportional to a Lorentz transformation $\Lambda^\mu{}_\alpha$, namely,

$$\Lambda^\mu{}_\alpha = |J|^{-\frac{1}{4}} \frac{\partial x'^\mu}{\partial x^\alpha} \; . \tag{5.182}$$

If ϕ transforms under the Lorentz transformation Λ as

$$\phi'(x') = D\,(\Lambda)\, \phi(x) \; , \tag{5.183}$$

then

$$\phi'(x') = |J|^{\frac{l}{4}}\, D\left[|J|^{-\frac{1}{4}} \left(\frac{\partial x'}{\partial x}\right)\right] \phi(x) \tag{5.184}$$

provides a representation of the conformal group. Here, l is the scale dimension of ϕ. For Lorentz transformations, $|J| = 1$, and (5.184) is identical to (5.183). For dilatations, we put $D\,(\Lambda) = 1$, and $J = e^{4\gamma}$; hence (5.184) becomes (5.168).

According to (5.181), the metric tensor transforms as

$$g'_{\mu\nu}(x') = |J|^{\frac{1}{2}} \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial x^\beta}{\partial x'^\nu} g_{\alpha\beta}(x) \; , \tag{5.185}$$

while the transformation law for the inverse metric $g^{\mu\nu}$ (defined by $g_{\mu\alpha} g^{\alpha\nu} = \delta^\nu{}_\mu$) is

$$g'^{\,\mu\nu}(x') = |J|^{-\frac{1}{2}} \frac{\partial x'^\mu}{\partial x^\alpha} \frac{\partial x'^\nu}{\partial x^\beta} g^{\alpha\beta}(x) \; . \tag{5.186}$$

If we suppose that the differential operator ∂_μ transforms like a covector,

$$\frac{\partial}{\partial x'^\mu} = \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial}{\partial x^\alpha} \; , \tag{5.187}$$

then (5.186) gives the transformation law for $\partial^\mu = g^{\mu\nu}\partial_\nu$:

$$\frac{\partial}{\partial {x'}_\mu} = |J|^{-\frac{1}{2}} \frac{\partial x'^\mu}{\partial x^\alpha} \frac{\partial}{\partial x_\alpha} \, . \tag{5.188}$$

We see that the transformation laws for fields and their derivatives are cumbersome. Fortunately, the analysis of the conformal properties of many theories can be simplified if we replace $\mathrm{C}(1,3)$ by a group which acts on the metric according to the rule

$$\eta_{\mu\nu} \to g_{\mu\nu}(x) = e^{2\lambda(x)}\,\eta_{\mu\nu}, \quad \eta^{\mu\nu} \to g^{\mu\nu}(x) = e^{-2\lambda(x)}\,\eta^{\mu\nu} \, , \tag{5.189}$$

but leaves coordinates x^μ unchanged. This group is called the group of *Weyl rescalings*. The transformation law for matter fields is a minor generalization of (5.168),

$$\phi_a(x) \to {\phi'}_a(x) = e^{l_a\lambda(x)}\phi_a(x) \, . \tag{5.190}$$

Comparing (5.189) and (5.190), we find that the scale dimension of $g_{\mu\nu}$ is $l = 2$, and that of $g^{\mu\nu}$ is $l = -2$. Therefore, $g = \det(g_{\mu\nu})$ has $l = 8$, and $\sqrt{-g}$ has $l = 4$. Now, if we assign $l = 0$ to the vector potential A_μ[6], then the Larmor Lagrangian

$$\mathcal{L}_{\mathrm{L}} = -\frac{1}{16\pi}\sqrt{-g}\, g^{\mu\nu}g^{\alpha\beta}F_{\mu\alpha}F_{\nu\beta}, \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{5.191}$$

is invariant under Weyl rescalings (5.189)–(5.190).

Because particle positions are assumed to be unchanged the effect of Weyl rescalings on theories of this type, specifically on the Maxwell–Lorentz theory, amounts to a rescaling of the metric alone, as indicated in (5.189).

What are implications of this symmetry? Suppose that the action S is built out of matter variables having zero scale dimension and the metric, and that S is invariant under Weyl rescalings with $\lambda(x) = \epsilon(x)$, where ϵ is an arbitrary infinitesimal function vanishing at spatial infinity. Then

$$\delta S = \frac{1}{2}\int d^4x\,\sqrt{-g}\,\mathcal{T}_{\mu\nu}\,\delta g^{\mu\nu} = \int d^4x\,\sqrt{-g}\,(\mathcal{T}_{\mu\nu}\,g^{\mu\nu})\,\epsilon = 0 \, . \tag{5.192}$$

Here, $\mathcal{T}_{\mu\nu}$ is the metric stress-energy tensor defined in (5.119), and the variation of the metric in the second equation is taken to be $\delta g^{\mu\nu} = -2\epsilon g^{\mu\nu}$ in agreement with (5.189). As ϵ is an arbitrary infinitesimal function, and $\sqrt{-g} \neq 0$, we have

$$\mathcal{T}^{\mu}_{\ \mu} = 0 \, . \tag{5.193}$$

This equation may be regarded as a criterion for selecting Weyl invariant theories. Without going into detail we simply state the theorem that if a theory is Weyl invariant, then it is invariant under $\mathrm{C}(1,3)$ as well.

[6] $A^\mu = g^{\mu\nu}A_\nu$ has scale dimension $l = -2$.

Let us use electrodynamics in an arbitrary spacetime dimension $D+1$ to illustrate the utility of (5.193). Consider the total stress-energy tensor $\mathcal{T}_{\mu\nu}$ defined in (5.141), (5.129), and (5.152). The metric stress-energy tensor of the electromagnetic field (5.152),

$$\Theta_{\mu\nu} = \frac{1}{\Omega_{D-1}} \left(F_{\mu}{}^{\alpha} F_{\alpha\nu} + \frac{\eta_{\mu\nu}}{4} F_{\mu\nu} F^{\mu\nu} \right) , \tag{5.194}$$

and the fact that $\delta^{\mu}{}_{\mu} = D+1$ imply,

$$\Theta^{\mu}{}_{\mu} = \frac{1}{4\,\Omega_{D-1}} (D-3)\, F_{\alpha\beta} F^{\alpha\beta} . \tag{5.195}$$

Hence $\Theta^{\mu}{}_{\mu} = 0$ if $D+1 = 4$. Remembering that the Schwarzschild action does not contribute to $\mathcal{T}_{\mu\nu}$, we conclude that the electromagnetic sector of the Maxwell–Lorentz theory is conformal invariant only for $D+1 = 4$. Therefore, Maxwell's equations (which owe their origin to this sector) are invariant under C(1, 3).

The stress-energy tensor $t_{\mu\nu}$ for a particle governed by the Poincaré–Planck action,

$$t_{\mu\nu}(x) = m \int_{-\infty}^{\infty} ds\, v_{\mu}(s)\, v_{\nu}(s)\, \delta^4 \left[x - z(s) \right] , \tag{5.196}$$

gives

$$t^{\mu}{}_{\mu}(x) = m \int_{-\infty}^{\infty} ds\, \delta^4 \left[x - z(s) \right] , \tag{5.197}$$

which is nonzero for $m \neq 0$. Therefore, the equation of motion for a massive particle (2.225) is not conformally invariant.

A common objection to any physical relevance for conformal symmetry is that the conformal group *violates causality.* In 1960 Julius Wess pointed out that special conformal transformations

$$x'^{\mu} = \frac{x^{\mu} - b^{\mu} x^2}{\sigma(x)} \tag{5.198}$$

where $\sigma(x) = 1 - 2b \cdot x + b^2 x^2$, can convert a timelike vector into spacelike and vice versa, because

$$x'^2 = \frac{x^2}{\sigma(x)} . \tag{5.199}$$

If b^{μ} is such that $\sigma(x) < 0$ then x^2 and x'^2 are opposite in sign.

There are two widely accepted interpretations of spacetime transformations, 'passive' and 'active'. In the 'passive' interpretation, spacetime is viewed by two observers $\mathcal{O}$ and $\mathcal{O}'$ who assign different coordinates x_{μ} and x'_{μ} to the same point and use a coordinate transformation $x'_{\mu} = f_{\mu}(x)$ to reconcile their views on the same geometry. In the 'active' interpretation, spacetime is transformed with respect to a fixed frame of reference, each spacetime point is mapped into another point. Since the infinitesimal line element dx^2 is rescaled, the metric is not preserved. Therefore, such $\mathrm{C}(1,3)$ mappings do not form an automorphism group of Minkowski space. Whatever interpretation, the conclusion is the same: the causal relation of events separated by a finite interval is incompatible with conformal invariance.

On the other hand, the infinitesimal line element transforms to,

$$dx'^2 = \frac{dx^2}{\sigma^2(x)} \ . \tag{5.200}$$

This implies that the sign of the line element dx^2 is invariant. In particular, timelike and spacelike have an invariant meaning for tangent vectors. Conformal transformations always map a timelike curve into another timelike curve. Some people argue that only the infinitesimal line element is required for physics, which makes conformally invariant theories less problematic.

Another way of looking at conformal invariance was advanced by Joe Rosen in 1968. He assumed that the conformal group leaves both spacetime and the frame of reference unaffected, and acts only on world lines and fields generated by the sources which move along these world lines. Note that conformal transformations render a straight world line curved, and, therefore, a static field becomes a varying field. Furthermore, a single curve can be converted into a two-branched curve, and hence the number of particles is not preserved by these transformations.

To be more specific, consider a straight world line aligned with the time axis, $\mathbf{z}(t) = \mathbf{0}$. If we perform a special conformal transformation (5.198) characterized by $b^\mu = (0,0,0,g)$, we get a two-branched curve in the (t,z) plane

$$t' = \frac{t}{1-g^2t^2}, \quad z' = \frac{-gt^2}{1-g^2t^2} \ , \tag{5.201}$$

shown in Fig. 5.1. This curve is readily recognized as a hyperbola

$$\left(z' - \frac{1}{2g}\right)^2 - t'^2 = \frac{1}{4g^2} \ . \tag{5.202}$$

Physically, this conformal transformation maps the history of a single particle at rest into another history of two uniformly accelerated particles. One of these particles move along the left branch of the hyperbola BOC oriented from the past to the future, and the other (being actually an antiparticle) travels back in time along the right $BADC$-branch.

The same observer will observe both initial and transformed processes which occur in the same spacetime background $\mathbb{M}_4$, but under different conditions elsewhere in the world, with the participation of other particles and antiparticles. It is then clear that causal relations of the initial setting are unrelated to the transformed ones, because conformal transformations switch between two quite different physical states realized in a conformal theory. This is Rosen's resolution of the causality problem.

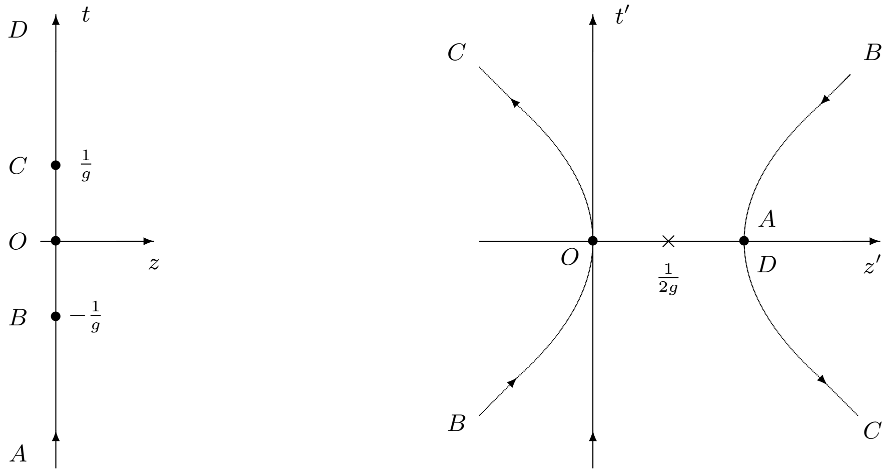


Fig. 5.1. A vertical straight line $ABOCD$ is mapped into a two-branched hyperbola BOC & $BADC$ by a conformal transformation

It seems plausible that electrodynamics of massless charged particles is conformal invariant in the sense that every physically valid state described by an exact simultaneous solution to Maxwell's equations and equations of motion for massless charged particles can be obtained via a conformal transformation of a single fixed state. However, such a theory, if any, defies Lagrangian formulation. Indeed, a particular Lagrangian refers to a definite number of particles; applying a conformal transformation to the system governed by this Lagrangian, we come to a system with another particle content.

Problem 5.3.1. Consider a free massless scalar field in a $(D+1)$-dimensional world governed by the action

$$S = \frac{1}{2}\int d^{D+1}x\,(\partial_\mu\phi)\,(\partial^\mu\phi)\ , \tag{5.203}$$

and suppose that dilatations (5.167) cause ϕ to transform according to the law

$$\phi'(x') = \phi(x)\ . \tag{5.204}$$

Show that this action is scale invariant only for $D = 1$. Show that if ϕ instead transforms according to (5.168), then scale invariance of the action (5.203) follows for

$$l_\phi = \frac{1}{2}\,(D-1)\ . \tag{5.205}$$

Problem 5.3.2. Show that the dilatation current $\mathcal{D}_\mu$ defined in (5.170) can be written in the form

$$\mathcal{D}_\mu = T_{\mu\nu}\,x^\nu + l_a\pi^a_\mu\,\phi_a + \pi^\nu_a\Sigma_{\mu\nu}\phi^a - \partial^\lambda(x^\nu B_{\lambda\mu\nu})\ , \tag{5.206}$$

where $T_{\mu\nu}$ and $B_{\lambda\mu\nu}$ are the symmetric stress-energy tensor and the Belinfante term defined respectively in (5.106) and (5.109).

Problem 5.3.3. Let a scalar field be governed by a Lagrangian of the type of (5.172). An improved stress-energy tensor is

$$\tau_{\mu\nu} = -\mathcal{L}\,\eta_{\mu\nu} + (\partial_\mu\phi)(\partial_\nu\phi) - \frac{1}{6}\left(\partial_\mu\partial_\nu - \eta_{\mu\nu}\Box\right)\phi^2 \ . \tag{5.207}$$

Show that D_μ can be brought to the form

$$\mathcal{D}_\mu = \tau_{\mu\nu}x^\nu + \frac{1}{6}\,\partial^\nu\left[\left(x_\mu\partial_\nu - x_\nu\partial_\mu\right)\phi^2\right] \ . \tag{5.208}$$

Since the divergence of an antisymmetric tensor does not contribute to $\partial^\mu D_\mu$ and P_μ, the last term in (5.208) may be omitted. This result is due to Curtis Callan, Sidney Coleman, and Roman Jackiw (1970).

Problem 5.3.4. Let $x \to x'$ be a special conformal transformation (5.198). Show that the volume element d^4x is transformed as

$$d^4x' = \sigma^{-4}\,(x)\,d^4x \ . \tag{5.209}$$

Assuming that P_μ transforms in the same way as ∂_μ, verify that the mass M (defined by $M^2 = P_\mu P^\mu$) is transformed as

$$M' = \sigma\,(x)\,M \ . \tag{5.210}$$

Hint Use (5.180), and the fact that $d^4x' = |J|\,d^4x$.

Problem 5.3.5. Find the stress-energy tensor for a particle governed by the action (2.266). Show that the equations of motion for a massless particle (2.282) and (2.280) are conformally invariant.

Answer

$$t_{\mu\nu}\,(x) = \int_{-\infty}^{\infty} d\tau\,\eta\,(\tau)\,\dot{z}_\mu\,(\tau)\,\dot{z}_\nu\,(\tau)\,\delta^4\left[x - z(\tau)\right] \ . \tag{5.211}$$

For $\dot{z}^2 = 0$ we have

$$t^\mu{}_\mu = \int_{-\infty}^{\infty} d\tau\,\eta\,(\tau)\,\dot{z}^2\,(\tau)\,\delta^4\left[x - z(\tau)\right] = 0 \ . \tag{5.212}$$

Problem 5.3.6. Consider the following construction

$$h_{\mu\nu}(x-y) = (x-y)^2\,\frac{\partial}{\partial x^\mu}\frac{\partial}{\partial y^\nu}\ln\,(x-y)^2 = \eta_{\mu\nu} - \frac{(x-y)_\mu\,(x-y)_\nu}{(x-y)^2} \ . \tag{5.213}$$

Prove that $h_{\mu\nu}$ is a conformal tensor whose index μ transforms like a covector at the point x while the index ν transforms like a covector at the point y:

$$h'_{\mu\nu}(x'-y') = \frac{1}{\sigma(x)\sigma(y)} \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial x^\beta}{\partial y'^\nu} h_{\alpha\beta}(x-y) \ . \tag{5.214}$$

This construction is due to David Boulware, Lowell Brown, and Roberto Peccei (1970).

Proof Let us note that $\ln{(x-y)}^2$ transforms additively under the special conformal operation on ${(x-y)}^2$ defined in equation $(D.9)$ of Appendix D,

$$\ln{(x'-y')}^2 = \ln{(x-y)}^2 - \ln\sigma(x) - \ln\sigma(y) \ . \tag{5.215}$$

The last two terms are annihilated by differentiation with respect to both coordinates, so the logarithm behaves effectively as a scalar. With reference to (5.187), we get (5.214).

5.4 Duality Invariance

The source-free Maxwell equations

$$\partial_\mu F^{\mu\nu} = 0, \quad \partial_\mu {}^*F^{\mu\nu} = 0 \tag{5.216}$$

are invariant under the following *duality transformation*

$$\begin{aligned} F'_{\mu\nu} &= F_{\mu\nu}\cos\theta + {}^*F_{\mu\nu}\sin\theta \ , \\ {}^*F'_{\mu\nu} &= -F_{\mu\nu}\sin\theta + {}^*F_{\mu\nu}\cos\theta \ . \end{aligned} \tag{5.217}$$

This transformation consists of a rotation between the electric and the magnetic fields,

$$\begin{aligned} \mathbf{E}' &= \mathbf{E}\cos\theta + \mathbf{B}\sin\theta \ , \\ \mathbf{B}' &= -\mathbf{E}\sin\theta + \mathbf{B}\cos\theta \ . \end{aligned} \tag{5.218}$$

At first sight it would seem that the duality transformation is an ordinary internal symmetry of the source-free Maxwell theory. An infinitesimal duality transformation

$$\begin{aligned} \delta F_{\mu\nu} &= {}^*F_{\mu\nu}\,\delta\theta \ , \\ \delta^* F_{\mu\nu} &= -F_{\mu\nu}\,\delta\theta \end{aligned} \tag{5.219}$$

leaves the Larmor action unchanged up to a divergence term:

$$\delta S = \frac{1}{8\pi}\int d^4x\, F^{\mu\nu}\delta F_{\mu\nu} = \frac{\delta\theta}{8\pi}\int d^4x\, F^{\mu\nu}\, {}^*F_{\mu\nu} = \frac{\delta\theta}{4\pi}\int d^4x\, \partial_\mu(\epsilon^{\mu\nu\alpha\beta}A_\nu\partial_\alpha A_\beta) \ . \tag{5.220}$$

The symmetric stress-energy tensor $\Theta_{\mu\nu}$ is invariant under duality transformations (5.217). This is evident from writing $\Theta_{\mu\nu}$ in a manifestly duality-symmetric form (5.154),

$$\Theta^{\lambda\mu} = \frac{1}{8\pi}\left(F^{\lambda}_{\ \sigma}F^{\sigma\mu} + {}^{*}F^{\lambda}_{\ \sigma}{}^{*}F^{\sigma\mu}\right) . \tag{5.221}$$

However, the invariants of the electromagnetic field $\mathcal{S}$ and $\mathcal{P}$ do change,

$$\begin{aligned} \mathcal{S}' &= \mathcal{S}\cos 2\theta + \mathcal{P}\sin 2\theta , \\ \mathcal{P}' &= -\mathcal{S}\sin 2\theta + \mathcal{P}\cos 2\theta . \end{aligned} \tag{5.222}$$

We see that a finite duality transformation rescales the Larmor action by a factor of $\cos 2\theta$ (plus an inessential divergence term). Therefore, electric-magnetic duality is a symmetry between the equation of motion and the Bianchi identity, rather than an invariance of the Larmor action.

Now we consider the case that sources are present. If we suppose the coexistence of electric and magnetic charges, then Maxwell's equations become

$$\partial_\lambda F^{\lambda\mu} = 4\pi j^\mu, \quad \partial_\lambda {}^{*}F^{\lambda\mu} = 4\pi m^\mu . \tag{5.223}$$

Denoting

$$\mathcal{F}^{\mu\nu} = F^{\mu\nu} - i\,{}^{*}F^{\mu\nu}, \quad \mathcal{J}^\mu = j^\mu - i m^\mu , \tag{5.224}$$

we can re-express the two real equations (5.223) as the single complex equation

$$\partial_\lambda \mathcal{F}^{\lambda\mu} = 4\pi \mathcal{J}^\mu . \tag{5.225}$$

This complex equation is invariant under the duality transformation

$$\mathcal{F}'_{\mu\nu} = \mathcal{F}_{\mu\nu}\, e^{i\theta}, \quad \mathcal{J}'_\mu = \mathcal{J}_\mu\, e^{i\theta} . \tag{5.226}$$

Although this theory exhibits an extra symmetry (5.226), it is more complicated in comparison with the Maxwell–Lorentz electrodynamics[7]. It will suffice to mention the fact (fully considered in Sect. 2.9) that the orbital momentum of a binary system of electric change e and magnetic charge $e^\star$ is not conserved. Only the total angular momentum, including the field contribution $e\,e^\star\mathbf{n}$, is conserved. Note that the magnitude of this compensating term is independent of interparticle separation.

We learned in Sect. 4.8 that solutions to equations (5.223) are no longer one-to-one smooth mappings A_μ of Minkowski space into another copy of this space $\mathbb{M}_4 \to \mathbb{M}_4$. There is instead a smooth mapping of $\mathbb{M}_4$ into $\mathbb{M}_4 \times \mathbb{M}_4$ through the Cabibbo–Ferrari pair of vector potentials:

$$A_\mu, B_\mu : \quad \mathbb{M}_4 \to \mathbb{M}_4 \times \mathbb{M}_4 . \tag{5.227}$$

[7] If the ratio of electric to magnetic charge is fixed for all particles, then performing transformation (5.226) with $\theta = \arctan(e^\star/e)$ makes $\mathcal{J}_\mu = j_\mu$, and (5.225) becomes the ordinary Maxwell equations $\partial_\lambda F^{\lambda\mu} = 4\pi j^\mu$, $\partial_\lambda {}^{*}F^{\lambda\mu} = 0$. For such a reduction to be avoided we must have dyons with differing ratios of electric to magnetic charge.

Alternatively, one can smoothly map $\mathbb{M}_4$ into a four-dimensional manifold with somewhat involved topology $\mathcal{M}_4$ using the Wu–Yang vector potential $\mathcal{A}_\mu$:

$$\mathcal{A}_\mu : \quad \mathbb{M}_4 \to \mathcal{M}_4 \, . \tag{5.228}$$

Neither (5.227) nor (5.228) is an isomorphism of $\mathbb{M}_4$.

A Lagrangian which engenders equations (5.223) was found by Daniel Zwanziger in 1971. The construction is rather sophisticated and will be approached through a series of exercises (Problems 5.4.1–5.4.4).

Problem 5.4.1. Show that the general solution to the first of equations (5.223) can be written

$$F_{\mu\nu} = -{}^*(\partial_\mu B_\nu - \partial_\nu B_\mu) + (n \cdot \partial)^{-1} \left(n_\mu j_\nu - n_\nu j_\mu\right) , \tag{5.229}$$

and the general solution to the second equation is

$${}^*F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - (n \cdot \partial)^{-1}\, {}^*(n_\mu m_\nu - n_\nu m_\mu) , \tag{5.230}$$

where A_μ and B_μ are vector potentials, n_μ an arbitrarily fixed four-vector, and $(n \cdot \partial)^{-1}$ an integral operator with kernel $(n \cdot \partial)^{-1}(x-y)$ satisfying $n \cdot \partial\, (n \cdot \partial)^{-1}(x) = \delta^4(x)$.

Problem 5.4.2. Verify that any antisymmetric tensor $G^{\mu\nu}$ obeys the identity

$$n^2 G^{\mu\nu} = \left[n^\mu \left(n_\alpha G^{\alpha\nu}\right) - n^\nu \left(n_\alpha G^{\alpha\mu}\right)\right] - {}^*\!\left[n^\mu \left(n_\alpha\, {}^*G^{\alpha\nu}\right) - n^\nu \left(n_\alpha\, {}^*G^{\alpha\mu}\right)\right] . \tag{5.231}$$

Using this identity, show that (5.229) and (5.230) can be brought to the form

$$\begin{aligned} n^2 F^{\mu\nu} &= n^\mu \left[n_\alpha(\partial^\alpha A^\nu - \partial^\nu A^\alpha)\right] - n^\nu \left[n_\alpha(\partial^\alpha A^\mu - \partial^\mu A^\alpha)\right] \\ &\quad -{}^*\{n^\mu \left[n_\alpha(\partial^\alpha B^\nu - \partial^\nu B^\alpha)\right] - n^\nu \left[n_\alpha(\partial^\alpha B^\mu - \partial^\mu B^\alpha)\right]\} , \end{aligned} \tag{5.232}$$

$$\begin{aligned} n^2\, {}^*F^{\mu\nu} &= {}^*\{n^\mu \left[n_\alpha(\partial^\alpha A^\nu - \partial^\nu A^\alpha)\right] - n^\nu \left[n_\alpha(\partial^\alpha A^\mu - \partial^\mu A^\alpha)\right]\} \\ &\quad + n^\mu \left[n_\alpha(\partial^\alpha B^\nu - \partial^\nu B^\alpha)\right] - n^\nu \left[n_\alpha(\partial^\alpha B^\mu - \partial^\mu B^\alpha)\right] . \end{aligned} \tag{5.233}$$

Problem 5.4.3. Verify that applying (5.232) and (5.233) to (5.223) gives

$$\begin{aligned} &(n \cdot \partial)^2 A^\mu - (n \cdot \partial)\, \partial^\mu (n \cdot A) - n^\mu\, (n \cdot \partial)\, (\partial \cdot A) \\ &+ n^\mu\, \Box\, (n \cdot A) - (n \cdot \partial)\, \epsilon^{\mu\nu\alpha\beta}\, n_\nu \partial_\alpha B_\beta = 4\pi\, n^2\, j^\mu , \end{aligned} \tag{5.234}$$

$$\begin{aligned} &(n \cdot \partial)^2 B^\mu - (n \cdot \partial)\, \partial^\mu (n \cdot B) - n^\mu\, (n \cdot \partial)\, (\partial \cdot B) \\ &+ n^\mu\, \Box\, (n \cdot B) + (n \cdot \partial)\, \epsilon^{\mu\nu\alpha\beta}\, n_\nu \partial_\alpha A_\beta = 4\pi\, n^2\, m^\mu . \end{aligned} \tag{5.235}$$

Problem 5.4.4. Show that (5.234) and (5.235) follow from the Lagrangian $\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_{\rm int}$, where

$$\begin{aligned} 8\pi\, n^2\, \mathcal{L}_0 = &- \left[n_\alpha(\partial^\alpha A^\mu - \partial^\mu A^\alpha)\right] \left[n_\beta\, {}^*(\partial^\beta B^\mu - \partial^\mu B^\beta)\right] \\ &+ \left[n_\alpha(\partial^\alpha B^\mu - \partial^\mu B^\alpha)\right] \left[n_\beta\, {}^*(\partial^\beta A^\mu - \partial^\mu A^\beta)\right] \\ &- \left[n_\alpha(\partial^\alpha A^\mu - \partial^\mu A^\alpha)\right]^2 - \left[n_\alpha(\partial^\alpha B^\mu - \partial^\mu B^\alpha)\right]^2 , \end{aligned} \tag{5.236}$$

$$\mathcal{L}_{\rm int} = -j^\mu A_\mu - m^\mu B_\mu \, . \tag{5.237}$$

5.5 Gauge Invariance

Expressing the field strength in terms of the vector potential

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{5.238}$$

is standard practice to solve Maxwell's equations. Thus, the vector potential is a useful tool for analyzing field configurations in electrodynamics. Furthermore, the Schwarzschild action (2.212) responsible for the interaction of a charged particle and the electromagnetic field depends upon A_μ, rather than upon $F_{\mu\nu}$. One may be inclined to promote A_μ to the status of a *fundamental* variable. However, the field strength does not determine the vector potential uniquely. Both A_μ and

$$A'_\mu = A_\mu - \partial_\mu \chi \tag{5.239}$$

give identical $F_{\mu\nu}$. In other words, A_μ involves *redundant* degrees of freedom, gauge modes. Only gauge invariant quantities have direct experimental significance.

Consider for example the Schwarzschild action

$$S_{\rm S} = -e \int_{\tau_1}^{\tau_2} d\tau\, \dot{z}^\mu A_\mu(z)\ . \tag{5.240}$$

If A_μ is changed for A'_μ according to (5.239), then the transformed action differs from the original action in

$$e \int_{\tau_1}^{\tau_2} d\tau\, \frac{dz^\mu}{d\tau}\, \frac{\partial \chi}{\partial z^\mu} = e \int_{\tau_1}^{\tau_2} d\tau\, \frac{d\chi}{d\tau} = e\chi[z(\tau_2)] - e\chi[z(\tau_1)]\ . \tag{5.241}$$

For arbitrary χ (with $\chi = 0$ at the endpoints) the variation of $S_{\rm S}$ induced by the gauge transformation (5.239) is zero. Hence, (5.240) is gauge invariant.

We now digress to discuss the interaction of an electromagnetic field with a scalar field ϕ which represents continuously distributed charged matter. We will see that charge conservation is related to the fact that ϕ is *complex-valued*. So our immediate task is to consider the Lagrangian formulation of complex-valued fields.

Let us begin with the case that ϕ is free. We may regard ϕ and its complex conjugate ϕ^* as two independent variables. The action

$$S_0 = \int d^4x \left[(\partial_\mu \phi)^* \, (\partial^\mu \phi) - \mu^2\, \phi^* \phi \right]\ . \tag{5.242}$$

leads to the Klein–Gordon equation for both ϕ and ϕ^*:

$$\mathcal{E}(\phi) = \frac{\delta S_0}{\delta \phi} = -(\Box + \mu^2)\phi^* = 0, \quad \mathcal{E}(\phi^*) = \frac{\delta S_0}{\delta \phi^*} = -(\Box + \mu^2)\phi = 0\ . \tag{5.243}$$

Because of the juxtaposition of ϕ and ϕ^*, the action (5.242) is invariant under phase transformations

$$\phi \to e^{ie\lambda}\,\phi, \quad \phi^* \to e^{-ie\lambda}\,\phi^* \,. \tag{5.244}$$

We will see that the set of transformations (5.244), which constitute the internal symmetry group U(1), underlies charge conservation.

Write (5.244) in the infinitesimal form:

$$\delta\phi = ie\lambda\phi, \quad \delta\phi^* = -ie\lambda\phi^* \,, \tag{5.245}$$

where λ is a small real parameter. The Noether current[8] associated with this symmetry is

$$j_\mu = -\left[\frac{\partial\mathcal{L}_0}{\partial(\partial^\mu\phi)}\,(ie\phi) + \frac{\partial\mathcal{L}_0}{\partial(\partial^\mu\phi^*)}\,(-ie\phi^*)\right] \,. \tag{5.246}$$

By (5.242),

$$j_\mu = ie\left[\phi^*\partial_\mu\phi - (\partial_\mu\phi^*)\phi\right] = ie\phi^*\left(\partial_\mu - \overleftarrow{\partial}_\mu\right)\phi \,. \tag{5.247}$$

The Noether identity for j^μ reads

$$\partial_\mu j^\mu = ie\left[\phi\,\mathcal{E}(\phi) - \phi^*\mathcal{E}(\phi^*)\right] \,, \tag{5.248}$$

where $\mathcal{E}(\phi)$ and $\mathcal{E}(\phi^*)$ are the Eulerians defined in (5.243).

If the complex field ϕ is expressed in terms of two real variables

$$\phi = \frac{1}{\sqrt{2}}\,(a + ib) \,, \tag{5.249}$$

then a and b decouple in the action:

$$S_0 = \frac{1}{2}\int d^4x\left[(\partial^\mu a)\,(\partial_\mu a) - \mu^2\,a^2 + (\partial^\mu b)\,(\partial_\mu b) - \mu^2\,b^2\right] \,. \tag{5.250}$$

This action is invariant under a SO(2) rotation of the fields a and b through an angle $e\lambda$,

$$a' = a\cos(e\lambda) + b\sin(e\lambda), \quad b' = -a\sin(e\lambda) + b\cos(e\lambda) \,. \tag{5.251}$$

Phase transformations (5.244) are locally isomorphic to field rotations (5.251).

One may require that λ be a function of spacetime. Transformations of this kind are called *local*. The term 'local' emphasizes the fact that the group of transformations acts at each point x^μ, quite apart from its action in other points. The action S_0 is not invariant under transformations (5.245) with local arguments $\lambda(x)$. Indeed, the variation of S_0 is

$$\delta S_0 = -\int d^4x\,j_\mu\,\partial^\mu\lambda \,, \tag{5.252}$$

[8] The overall minus sign in (5.246) was introduced in order to relate this formula to the other definition of current, $j_\mu = -\partial\mathcal{L}/\partial A^\mu$, which follows from $\mathcal{L}_{\text{int}} = -j_\mu A^\mu$.

which is nonzero unless $\lambda(x)$ is constant. Here, j_μ is the current defined in (5.247).

Invariance is regained by appropriately coupling ϕ to the vector potential. We replace all partial derivatives ∂_μ acting on complex fields ϕ by the so-called *covariant derivatives* D_μ,

$$\partial_\mu \phi \to D_\mu \phi = (\partial_\mu + ieA_\mu)\,\phi\,. \tag{5.253}$$

Then (5.242) becomes

$$\int d^4x \left[(D_\mu\phi)^*\,(D^\mu\phi) - \mu^2\,\phi^*\phi\right]\,. \tag{5.254}$$

Combining (5.254) and (5.47), we obtain the action for the system of interacting charged and electromagnetic fields

$$S = \int d^4x \left\{ \left[(\partial_\mu + ieA_\mu)\phi\right]^*(\partial_\mu + ieA_\mu)\phi - \mu^2\,\phi^*\phi - \frac{1}{16\pi}\,F^{\mu\nu}F_{\mu\nu} \right\}\,. \tag{5.255}$$

The action (5.255) is invariant under local transformations

$$\phi(x) \to e^{ie\lambda(x)}\,\phi(x), \quad \phi^*(x) \to e^{-ie\lambda(x)}\,\phi^*(x), \quad A_\mu(x) \to A_\mu(x) - \partial_\mu\lambda(x)\,. \tag{5.256}$$

The set of these transformations form an infinite group whose parameter λ depends of spacetime. This group is called the *gauge group* of electrodynamics.

To eliminate the gauge freedom, we impose the Lorenz condition $\partial_\mu A^\mu = 0$. In this gauge the Euler–Lagrange equations read

$$\mathcal{E}(\phi) = \left[-(\Box + \mu^2) + 2ie(A\cdot\partial) + e^2A^2\right]\phi^* = 0\,, \tag{5.257}$$

$$\mathcal{E}(\phi^*) = \left[-(\Box + \mu^2) - 2ie(A\cdot\partial) + e^2A^2\right]\phi = 0\,, \tag{5.258}$$

$$\mathcal{E}^\mu(A) = \frac{1}{4\pi}\,\Box\,A^\mu - j^\mu = 0\,, \tag{5.259}$$

where

$$j_\mu = ie\phi^*\left(\partial_\mu - \overleftarrow{\partial}_\mu\right)\phi - 2e^2A_\mu\,\phi^*\phi\,. \tag{5.260}$$

Since (5.259) is identical to Maxwell's equations, j_μ can be regarded as the charge current.

If $\lambda(x)$ in (5.256) is specialized to a constant, we revert to global phase transformations (5.244). Invariance under such transformations implies conservation of a Noether current. We leave as an easy exercise to check that the pertinent Noether current j_μ is given by (5.260), and that the Noether identity holds

$$\partial_\mu j^\mu = ie\phi\,\mathcal{E}(\phi) - ie\phi^*\mathcal{E}(\phi^*)\,, \tag{5.261}$$

where $\mathcal{E}(\phi)$ and $\mathcal{E}(\phi^*)$ are defined in (5.257) and (5.258).

Taking a closer look at (5.260), one finds a surprising thing: j_μ depends of A_μ. One can readily verify (Problem 5.5.1) that j_μ is invariant under the gauge transformation (5.256). On the other hand, this dependence makes it appear that A_μ is as much a charge carrier as ϕ. The Noether identity (5.261) enables escaping this wrong impression: only $\mathcal{E}(\phi)$ and $\mathcal{E}(\phi^*)$ contribute to local charge conservation, not $\mathcal{E}^\mu(A)$. This means that the actual charge flow occurs regardless of the vector field dynamics encoded in $\mathcal{E}^\mu(A)$.

The relation between a local symmetry (5.256) and current conservation (5.261) is a special case of Noether's second theorem. To formulate this theorem in a general setting, we consider a system of n interacting fields ϕ_a, and suppose that the action is invariant under infinitesimal transformations of field variables

$$\delta\phi_a = Q_a^{\,k}\,\epsilon_k + (R_a^{\,k})^\mu\,\partial_\mu\epsilon_k\ , \tag{5.262}$$

where Q_a^k and $(R_a^k)^\mu$ are some functions of ϕ_a and $\partial_\mu\phi_a$, and ϵ_k are infinitesimal *localized* parameters, $\epsilon_k = \epsilon_k(x)$. We assume that the set of infinitesimal transformations (5.262) can be integrated to give a finite $\mathcal{N}$-parameter Lie group with parameters depending on spacetime. Then, with reference to Problem 2.6.6, we write a linear relation between Eulerians and their derivatives

$$Q_a^{\,k}\,\mathcal{E}^a - \partial_\mu\left[(R_a^{\,k})^\mu\,\mathcal{E}^a\right] = 0\ , \tag{5.263}$$

where $\mathcal{E}^a$ is the Eulerian corresponding to the variation of the action with respect to ϕ_a.

Turning back to the system of fields ϕ and A_μ governed by the action (5.255), we have the local U(1) symmetry with

$$Q(\phi) = ie\phi,\quad Q(\phi^*) = -ie\phi^*,\quad R_\nu^{\,\mu}(A) = -\delta_\nu^\mu\ . \tag{5.264}$$

The Noether identity (5.263) takes the form

$$ie\phi\,\mathcal{E}(\phi) - ie\phi^*\mathcal{E}(\phi^*) + \partial_\mu\mathcal{E}^\mu(A) = 0\ , \tag{5.265}$$

which, in view of the identity

$$\partial_\mu\partial_\lambda F^{\lambda\mu} = 0\ , \tag{5.266}$$

reduces to (5.261).

Gauge invariance in electrodynamics (5.256) was discovered by Vladimir Fock in 1926, and elevated to a fundamental dynamical principle by Weyl in 1929. Following this principle, the interaction of charged matter with the electromagnetic field is introduced by replacing ∂_μ with D_μ, which is known as *minimal coupling*. One may think of this description of charged matter by a complex-valued scalar field[9] as exemplifying the idea of gauge invariance.

[9] The minimal coupling prescription can be readily applied to the Dirac spinor field describing spin-$\frac{1}{2}$ charged fluids (see Problems 5.5.2 and 5.5.8). However, the

Meanwhile it is possible to modify the Maxwell–Lorentz electrodynamics in such a way as to obtain a gauge theory. For this purpose, we introduce, quite formally, an internal 'charge space' for a point particle possessing the electric charge e. Let us specify the position of the particle in this internal space by a coordinate ζ related to e as

$$\zeta = i\sqrt{e}\,, \tag{5.267}$$

and write the action

$$S = -\int d\tau \left[\sqrt{\dot{z}_\alpha\,\dot{z}^\alpha} - \zeta^* \left(\frac{d}{d\tau} - \dot{z}^\alpha A_\alpha\right)\zeta\right] - \frac{1}{16\pi}\int d^4x\, F_{\mu\nu}F^{\mu\nu}\,. \tag{5.268}$$

This action is invariant under the local transformation

$$\zeta \to e^{i\Lambda}\,\zeta, \quad \zeta^* \to e^{-i\Lambda}\,\zeta^*, \quad A_\mu \to A_\mu + i\,\partial_\mu\Lambda\,, \tag{5.269}$$

because the variation of the term involving the vector potential

$$-i\,\frac{dz^\alpha}{d\tau}\,\frac{\partial\Lambda}{\partial z^\alpha} = -i\,\frac{d\Lambda}{d\tau} \tag{5.270}$$

cancels the derivative of the phase factor coming from $\dot{\zeta}$.

The Euler–Lagrange equations for ζ and ζ^*

$$\dot{\zeta} = (v\cdot A)\,\zeta, \quad \dot{\zeta}^* = -(v\cdot A)\,\zeta^* \tag{5.271}$$

suggest that ζ and ζ^* are internal degrees of freedom which vary in time under the influence of A_μ. However, using (5.271),

$$\frac{d}{ds}\,(\zeta^*\zeta) = \dot{\zeta}^*\zeta + \zeta^*\dot{\zeta} = -(v\cdot A)\,\zeta^*\zeta + (v\cdot A)\,\zeta^*\zeta = 0\,. \tag{5.272}$$

Therefore, $\zeta^*\zeta$ is a constant which is, in fact, the particle's charge, $\zeta^*\zeta = e$. With this note, the term responsible for the interaction of a charged particle and electromagnetic field in (5.268) is of the conventional form $-e\,v^\alpha A_\alpha$. Hence, all basic equations of the Maxwell–Lorentz theory have been regained.

Problem 5.5.1. Show that the current j_μ defined in (5.260) is gauge invariant.

Hint Rewrite j_μ using covariant derivatives:

$$j_\mu = ie\left[\phi^* D_\mu\phi - (D_\mu\phi)^*\phi\right]\,. \tag{5.273}$$

---

minimal coupling of a higher-spin complex field with the electromagnetic field seems to be physically inconsistent. For example, the minimal coupling of the Rarita–Schwinger spin-$\frac{3}{2}$ field with the electromagnetic field yields a hyperbolic equation whose solutions propagate at velocities exceeding the speed of light, which violates causality.

Problem 5.5.2. Apply the minimal coupling prescription to the free Dirac Lagrangian $\mathcal{L}_0 = \bar{\psi}\,(i\gamma^\mu\partial_\mu - m)\,\psi$. Find: (i) Lagrangian for the Dirac field ψ interacting with A_μ, (ii) field equations for ψ and $\bar{\psi}$, (iii) gauge transformations, (iv) Noether current, and (v) Noether identity expressing charge conservation.

Answer

$$\mathcal{L} = \bar{\psi}\,(i\gamma^\mu\partial_\mu - m - e\gamma^\mu A_\mu)\,\psi\ , \tag{5.274}$$

$$\mathcal{E}_{\bar{\psi}} = (i\gamma^\mu\partial_\mu - m - e\gamma^\mu A_\mu)\,\psi = 0, \quad \mathcal{E}_\psi = \bar{\psi}(i\gamma^\mu \overleftarrow{\partial}_\mu + m + e\gamma^\mu A_\mu) = 0\ , \tag{5.275}$$

$$\psi \to e^{ie\lambda}\,\psi, \quad \bar{\psi} \to e^{-ie\lambda}\,\bar{\psi}, \quad A_\mu \to A_\mu - \partial_\mu\lambda\ , \tag{5.276}$$

$$j_\mu = e\,\bar{\psi}\gamma_\mu\psi\ , \tag{5.277}$$

$$\partial_\mu j^\mu = e\left(\bar{\psi}\,\mathcal{E}_{\bar{\psi}} + \mathcal{E}_\psi\psi\right)\ . \tag{5.278}$$

The current given by (5.277) is independent of A_μ, as opposed to that of the scalar field, (5.260).

Problem 5.5.3. Assuming the Lorenz condition $\partial_\mu A^\mu = 0$, show that the conservation law (5.261) in which j_μ is expressed as (5.260) is equivalent to the Ward–Takahashi identity

$$\phi^*(\partial_\mu + \overleftarrow{\partial}_\mu)\Gamma^\mu\phi = \phi^*\left[(\Box + \mu^2) + 2ieA^\mu\partial_\mu - e^2A^2\right] - \left[(\overleftarrow{\Box} + \mu^2) - 2ie\overleftarrow{\partial}_\mu A^\mu - e^2A^2\right]\phi\ , \tag{5.279}$$

$$\Gamma_\mu = (\partial_\mu + ieA_\mu) - (\overleftarrow{\partial}_\mu - ieA_\mu)\ . \tag{5.280}$$

Compare (5.279) with the Ward–Takahashi identity for a free scalar field in Problem 4.3.5

Problem 5.5.4. Consider a gauge transformation

$$\phi \to \phi + \epsilon\,U, \quad \partial^\mu\phi \to \partial^\mu\phi + (\partial^\mu\epsilon)\,U + \epsilon\,(\partial^\mu U)\ , \tag{5.281}$$

where ϵ is an arbitrary infinitesimal function of x, and U may depend on ϕ. Do not assume that (5.281) necessarily leaves the action S invariant.

Using the Euler–Lagrange equations, verify that

$$\partial^\mu\left[\frac{\delta S}{\delta(\partial^\mu\epsilon)}\right] = \frac{\delta S}{\delta\epsilon}\ , \tag{5.282}$$

which shows that if S is invariant under this transformation, $\delta S/\delta\epsilon = 0$, there exist a conserved current

$$j_\mu = \frac{\delta S}{\delta(\partial^\mu\epsilon)}\ . \tag{5.283}$$

We thus arrived at a combination of Noether's first and second theorems which is known as the *Gell-Mann–Levy result*.

Problem 5.5.5. Let ψ be a free spinor field governed by $\mathcal{L} = \bar{\psi}\,(i\gamma^\mu\partial_\mu - m)\,\psi$. Consider the transformation

$$\psi \to \exp(ig\chi\,\gamma_5)\,\psi, \quad \bar{\psi} \to \bar{\psi}\exp(ig\chi\,\gamma_5)\;. \tag{5.284}$$

where χ is a continuous parameter. (The plus sign of the phase of the exponential applied to $\bar{\psi}$ derives from the fact that $\bar{\psi} = \psi^*\gamma^0$, and $\gamma_5\gamma^0\gamma^\mu = \gamma^0\gamma^\mu\gamma_5$ for any γ^μ.) Clearly, $\mathcal{L}$ is not invariant under the transformation (5.284), due to the presence of the mass term.

Calculate from (5.283) the corresponding current j^5_μ (which is called the *axial current*), and find its four-divergence.

Answer

$$j^5_\mu = g\bar{\psi}\,\gamma_5\gamma^\mu\,\psi, \quad \partial^\mu j^5_\mu = -2igm\,\bar{\psi}\,\gamma_5\,\psi\;. \tag{5.285}$$

Problem 5.5.6. Find $T_{\mu\nu}$ for the system with the action (5.255).

Answer $T_{\mu\nu}$ is given by

$$(D_\mu\phi)^*D_\nu\phi + (D_\nu\phi)^*D_\mu\phi - \eta_{\mu\nu}\left[(D_\alpha\phi)^*D^\alpha\phi - \mu^2\,\phi^*\phi\right]$$
$$+\frac{1}{4\pi}\left(F_\mu{}^\alpha F_{\alpha\nu} + \frac{\eta_{\mu\nu}}{4}\,F^{\alpha\beta}F_{\alpha\beta}\right)\;. \tag{5.286}$$

Problem 5.5.7. Apply the minimal coupling prescription to a system with the first-order Lagrangian $\mathcal{L}_0 = 2\phi^*_\mu\phi^\mu - \phi^*_\mu\partial^\mu\phi + (\partial^\mu\phi^*_\mu)\phi - (\partial^\mu\phi^*)\phi^\mu + \phi^*\partial_\mu\phi^\mu - 2\mu^2\phi^*\phi$, where ϕ and ϕ^μ are free complex-valued scalar and vector fields. Verify that that the resulting theory is equivalent to that described by the action (5.254).

Problem 5.5.8. Find the canonical and symmetric (Belinfante) stress-energy tensors $\theta_{\mu\nu}$ and $T_{\mu\nu}$ for the Maxwell–Dirac theory with the Lagrangian

$$\mathcal{L} = -\frac{1}{16\pi}\,F^{\mu\nu}F_{\mu\nu} + \bar{\psi}\,(i\gamma^\mu D_\mu - m)\,\psi, \quad D_\mu = \partial_\mu + ieA_\mu\;. \tag{5.287}$$

Answer

$$\theta_{\mu\nu} = \frac{i}{2}\,\bar{\psi}\,\gamma_\mu\,(\partial_\nu - \overleftarrow{\partial}_\nu)\,\psi - \frac{1}{4\pi}\,F_{\mu\alpha}\partial_\nu A^\alpha - \eta_{\mu\nu}\,\mathcal{L}\;, \tag{5.288}$$

$$T_{\mu\nu} = \frac{i}{2}\left(\bar{\psi}\,\gamma_\mu\,D_\nu\,\psi + \bar{\psi}\,\gamma_\nu\,D_\mu\,\psi\right) - \eta_{\mu\nu}\,\bar{\psi}\,(i\gamma^\alpha D_\alpha - m)\,\psi$$
$$+\frac{1}{4\pi}\left(F_\mu{}^\alpha F_{\alpha\nu} + \frac{\eta_{\mu\nu}}{4}\,F^{\alpha\beta}F_{\alpha\beta}\right)\;. \tag{5.289}$$

Problem 5.5.9. Is the Lagrangian (5.65) invariant under transformations $\phi'_\mu = \phi_\mu - \partial_\mu\chi$?

5.6 Strings and Branes

We normally think of particles as fundamental. However, one may equally well accord fundamental status to extended objects such as strings, membranes, and p-branes. In fact, these objects, collectively known as *branes*, have been a central preoccupation of many high-energy theorists in the last two decades. The Lagrangian formalism for strings (and, generally, for branes) can be obtained following essentially the same principles as those behind the particle dynamics. On the other hand, a string is a system with infinite degrees of freedom, and its dynamics share many features of field theory.

To begin let us imagine a one-dimensional flexible object whose form and length vary arbitrarily with time. Such a system is referred to as a *relativistic string*, or *string* for short. There are closed and open strings. The points of the string are specified by spacetime coordinates X^μ. During its motion, the string sweeps out a two-dimensional surface in Minkowski space, called the *world sheet*,

$$X^\mu = X^\mu(\sigma, \tau) \ . \tag{5.290}$$

The coordinates τ and σ parameterize the world sheet: σ labels the position of a point on the string and τ measures its time evolution.

We assume that all world sheets are *timelike*, smooth surfaces, which means that a two-dimensional plane tangent to the world sheet is spanned by a timelike and a spacelike vectors. If we set

$$\dot{X}_\mu = \frac{\partial X_\mu}{\partial \tau}, \quad X'_\mu = \frac{\partial X_\mu}{\partial \sigma} \ , \tag{5.291}$$

then $\dot{X}_\mu$ and X'_μ may be regarded as timelike and spacelike tangent vectors.

Note that the coordinate lines $\sigma = \text{const}$ and $\tau = \text{const}$ need not be orthogonal, $\dot{X} \cdot X' \neq 0$. The area of the parallelogram formed by two infinitesimal displacements $\dot{X}_\mu \, d\tau$ and $X'_\mu \, d\sigma$ is

$$A = \sqrt{(\dot{X} \cdot X')^2 - \dot{X}^2 X'^2} \, d\tau \, d\sigma \ . \tag{5.292}$$

For the world sheet to remain timelike, the quantity under the radical sign must be positive (Problem 5.6.1).

By analogy with the action for a particle, which is proportional to the length of the world line, we take the action for a string proportional to the area of the world sheet:

$$S = -\frac{T}{2\pi} \int_{\tau_1}^{\tau_2} d\tau \int_0^l d\sigma \sqrt{(\dot{X} \cdot X')^2 - \dot{X}^2 X'^2} \ . \tag{5.293}$$

This expression is known as the *Nambu action*. The constant T is necessary to make the action dimensionless. It has length dimension -2 and is called the *string tension*.

Expression (5.293) can be written more compactly. Denote $u^a = (\tau, \sigma)$, and observe that the metric on the world sheet $h_{ab}(u)$ induced by the Minkowski metric $\eta_{\mu\nu}$ is

$$h_{ab}(u) = \frac{\partial X^\mu}{\partial u^a}\frac{\partial X^\nu}{\partial u^b}\eta_{\mu\nu}, \quad a, b = 0, 1 \, . \tag{5.294}$$

Then (5.293) becomes

$$S = -\frac{T}{2\pi}\int d^2u \sqrt{-h} \, , \tag{5.295}$$

where $h = \det(h_{ab})$ (Problem 5.6.2).

When using the Lagrangian

$$\mathcal{L}(\dot{X}, X') = -\frac{T}{2\pi}\sqrt{-h} \, , \tag{5.296}$$

the action principle says: the string moves so as to minimize the area of the world sheet, with initial and final positions of the string being fixed,

$$\delta S = -\int_{\tau_1}^{\tau_2} d\tau \int_0^l d\sigma \left(\frac{\partial}{\partial \tau}\frac{\partial \mathcal{L}}{\partial \dot{X}_\mu} + \frac{\partial}{\partial \sigma}\frac{\partial \mathcal{L}}{\partial X'_\mu}\right)\delta X_\mu + \int_{\tau_1}^{\tau_2} d\tau \left(\frac{\partial \mathcal{L}}{\partial X'_\mu}\delta X_\mu\right)\Big|_{\sigma=0}^{\sigma=l} = 0 \, . \tag{5.297}$$

Taking the boundary conditions

$$\frac{\partial \mathcal{L}}{\partial X'_\mu} = 0 \quad \text{at} \quad \sigma = 0, l \, , \tag{5.298}$$

we arrive at the Euler–Lagrange equations

$$\mathcal{E}^\mu = -\frac{\partial}{\partial \tau}\frac{\partial \mathcal{L}}{\partial \dot{X}_\mu} - \frac{\partial}{\partial \sigma}\frac{\partial \mathcal{L}}{\partial X'_\mu} = 0 \, . \tag{5.299}$$

Substituting

$$\pi_1^\mu = \frac{\partial \mathcal{L}}{\partial \dot{X}_\mu} = -\frac{T}{2\pi}\frac{X'^\mu(\dot{X}\cdot X') - \dot{X}^\mu X'^2}{\sqrt{(\dot{X}\cdot X')^2 - \dot{X}^2 X'^2}} \, , \tag{5.300}$$

$$\pi_2^\mu = \frac{\partial \mathcal{L}}{\partial X'_\mu} = -\frac{T}{2\pi}\frac{\dot{X}^\mu(\dot{X}\cdot X') - X'^\mu \dot{X}^2}{\sqrt{(\dot{X}\cdot X')^2 - \dot{X}^2 X'^2}} \tag{5.301}$$

in (5.298) and (5.299) gives what seem to be highly nonlinear equations of motion. Note, however, that the action (5.293) is reparametrization invariant. Indeed, the change of variables

$$\tau = F(\tau', \sigma'), \quad \sigma = G(\tau', \sigma') \, , \tag{5.302}$$

where F and G are smooth functions, leaves the action (5.293) invariant. Transformations (5.302) are called *world sheet diffeomorphisms*. They form an infinite group, the *gauge* group of the Nambu string. Since the gauge group involves two parameters, there are two Noether's identities for the Euler–Lagrange equations (5.299) (Problem 5.6.4).

To eliminate the gauge freedom of the Nambu string, we may impose two gauge fixing conditions. A convenient choice is

$$\dot{X} \cdot X' = 0, \quad \dot{X}^2 + X'^2 = 0 , \tag{5.303}$$

whose geometrical significance is that the coordinate lines τ = const and σ = const are orthogonal and uniformly parametrized. Assuming (5.303), one comes to the string in *orthonormal* gauge.

When using the gauge (5.303), the Euler–Lagrange equations (5.299) simplify

$$X''_\mu - \ddot{X}_\mu = 0 . \tag{5.304}$$

String coordinates in the orthonormal gauge obey the wave equation. The boundary conditions (5.298) become

$$X'_\mu(\tau, 0) = X'_\mu(\tau, l) = 0 . \tag{5.305}$$

These are *Neumann boundary conditions*. By squaring expression (5.301), and using (5.303) and (5.298), we obtain

$$\dot{X}^2 = 0 \quad \text{at} \quad \sigma = 0, l . \tag{5.306}$$

End points of strings obeying Neumann boundary conditions move at the speed of light.

Alternatively, one may adopt *Dirichlet boundary conditions*

$$X_\mu(\tau, 0) = X_\mu(\tau, l) = 0 , \tag{5.307}$$

which imply that $\delta X^\mu = 0$ in the last term of (5.297). With these conditions, we are led to *Dirichlet branes*, or *D*-branes. We will return to this issue at the close of this section.

The surface term in (5.297) vanishes if we impose periodic boundary conditions

$$X_\mu(\tau, 0) = X_\mu(\tau, l) . \tag{5.308}$$

These relations are suitable for *closed* strings in the orthonormal gauge.

The Lagrangian (5.296) is explicitly Poincaré invariant. Invariance under infinitesimal transformations of string coordinates

$$\delta X^\mu = \omega^\mu{}_\nu X^\nu + \epsilon^\mu, \quad \omega_{\mu\nu} = \omega_{\nu\mu} \tag{5.309}$$

results in conservation of the total energy-momentum

$$p^\mu = \int_0^l d\sigma \, \pi_1^\mu \tag{5.310}$$

and the total angular momentum

$$M^{\mu\nu} = \int_0^l d\sigma \,(X^\mu \,\pi_1^\nu - X^\nu \,\pi_1^\mu) \tag{5.311}$$

for X^μ obeying the Euler–Lagrange equations (5.299). In the orthonormal gauge,

$$p^\mu = \frac{T}{2\pi} \int_0^l d\sigma \,\dot{X}^\mu \ , \tag{5.312}$$

$$M_{\mu\nu} = \frac{T}{2\pi} \int_0^l d\sigma \left(X_\mu \,\dot{X}_\nu - X_\nu \,\dot{X}_\mu \right) \ . \tag{5.313}$$

In view of (5.312), the mass M of the string in the orthonormal gauge is given by

$$M^2 = p^2 = \frac{T}{(2\pi)^2} \left(\int_0^l d\sigma \,\dot{X}^\mu \right)^2 \ . \tag{5.314}$$

The Nambu string possesses energy and mass, unlike the Dirac string of a magnetic monopole. Note that energy-momentum is locally conserved inside the Nambu string, and no energy-momentum flows into or out of the ends of open strings which obey the Neumann boundary conditions (Problem 5.6.5).

Just as there are alternative forms, (2.260) and (2.266), for the action of a point particle, so too the string action can be defined both in the Nambu form (5.295) and in the form

$$S = -\frac{T}{4\pi} \int d^2u \,\sqrt{-g}\, g^{ab} \,\partial_a X^\mu \,\partial_b X_\mu \ . \tag{5.315}$$

Here, g_{ab} is a field on the world sheet to be independently varied, and $g = \det(g_{ab})$. A symmetric tensor g_{ab} is interpreted as an intrinsic world sheet metric. Expression (5.315) is often called the *Polyakov action*. The actions (5.295) and (5.315) are equivalent in that the solution to the Euler–Lagrange equations is $g_{ab} = h_{ab}$ (Problem 5.6.6).

The Polyakov string is reparametrization invariant,

$$u^a = f^a(u') \ , \tag{5.316}$$

provided that the intrinsic metric g_{ab} is transformed according to the law

$$g_{ab}(u) = \frac{\partial u'^c}{\partial u^a} \frac{\partial u'^d}{\partial u^b} \,g'_{cd}(u') \ . \tag{5.317}$$

Furthermore, the action (5.315) exhibits invariance under Weyl rescalings

$$g'_{ab} = e^{2\lambda(u)} \,g_{ab}, \quad g'^{ab} = e^{-2\lambda(u)} \,g^{ab} \ . \tag{5.318}$$

This extra symmetry owes its origin to the presence of an additional field variable g_{ab}.

Because $g_{ab} = h_{ab}$, a manifold with the intrinsic metric g_{ab} is geometrically equivalent to a world sheet with a metric h_{ab} inherited from the metric $\eta_{\mu\nu}$ on ambient spacetime. We may treat the action (5.315) as that describing a massless scalar (spacetime-valued) field $X^\mu(u)$ on this manifold. Whereas the metric $\eta_{\mu\nu}$ is fixed, the metric h_{ab} is dynamic.

The gauge freedom can be used to render the metric g_{ab} flat:

$$\sqrt{-g}\, g^{ab} = \eta^{ab}, \quad \eta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{5.319}$$

In fact, conditions (5.303) and (5.319) are equivalent (Problem 5.6.7). In the gauge (5.319), the action (5.315) simplifies

$$S = -\frac{T}{4\pi} \int d^2u\, \eta^{ab}\, \partial_a X^\mu\, \partial_b X_\mu \, , \tag{5.320}$$

and the Euler–Lagrange equations become the wave equations (5.304).

Note that the covariant gauge (5.319) is Weyl invariant, because $\sqrt{-g}$ is rescaled by the factor $e^{2\lambda}$ under Weyl transformations. This residual gauge freedom corresponds to the fact that there exist an infinity of orthonormal coordinate systems on the world sheet. We can specify completely the variables τ and σ by introducing the light cone coordinates $X^\mu = (X^+, X^-, X^i)$, $X^\pm = (X^0 \pm X^1)/\sqrt{2}$, and choosing τ to be proportional to X^+,

$$X^+ = \tau. \tag{5.321}$$

We see that X^+ is independent of σ. The other light cone variable X^- can be expressed in terms of X^i using the constraint $g_{01} = \dot{X} \cdot X' = 0$. Thus, X^+ and X^- are eliminated from this noncovariant gauge as independent string variables, and only X^i, called *transverse* coordinates, remain dynamically significant.

One may wonder how strings evolve. For simplicity we consider only open strings; some insight into the behavior of closed strings can be gained from Problems 5.6.8 and 5.6.9. Assuming l to be a finite parameter, it is natural to look for solutions to the wave equation (5.304) as a Fourier series in σ. It is easy to check that the series

$$X^i(\tau,\sigma) = q^i + \frac{p^i}{p^+}\tau + i\sqrt{\frac{2}{T}} \sum_{n\neq 0} \frac{1}{n}\, \alpha_n^i \, \cos\left(\frac{\pi n \sigma}{l}\right) \exp\left(-\frac{i\pi n \tau}{l}\right) \tag{5.322}$$

satisfies both the field equation (5.304) and boundary conditions (5.305). Here, $p^+ = (p^0 + p^1)/\sqrt{2}$. The summation Σ is over all positive and negative integers n, except $n = 0$. q^i is the center-of-mass position of the string,

$$q^i = \frac{1}{l} \int_0^l d\sigma\, X^i(0,\sigma) \, , \tag{5.323}$$

and p^i is its conjugate momentum (5.310). The amplitudes of the normal modes are arbitrary complex-valued coefficients subject to the condition $\alpha^i_n = (\alpha^i_{-n})^*$ to keep X^i real-valued.

It follows that the string moves as a whole with the constant velocity p^i/p^+. On the other hand, by (5.322), the angular momentum of the string (5.313) is $M^{ij} = L^{ij} + S^{ij}$ where

$$L^{ij} = q^i\, p^j - q^j\, p^i, \quad S^{ij} = -i \sum_{n=1}^{\infty} \frac{1}{n} \left(\alpha^i_{-n}\, \alpha^j_n - \alpha^j_{-n}\, \alpha^i_n \right) . \tag{5.324}$$

In the center-of-mass rest frame, $q^i = 0$, and $L^{ij} = 0$, but S^{ij} need not be zero. Hence, the string may rotate about its center-of-mass position as prescribed by the α^i_n.

The gauge conditions (5.303) can be expressed in terms of the Fourier amplitudes α^μ_n:

$$L_n = \frac{T}{2} \sum_{m=-\infty}^{\infty} \alpha^\mu_{n-m} \alpha_{m\mu} = 0, \quad n = 0, \pm 1, \dots , \tag{5.325}$$

where $\alpha^\mu_0 = p^\mu/T$. In particular, $L_0 = 0$ gives

$$M^2 = p^2 = -\frac{T}{2} \sum_{m=-\infty}^{\infty} (\alpha_m{}^\mu)^* \alpha_{m\mu} . \tag{5.326}$$

The mass of the string M contains contributions from all oscillatory modes which are proportional to the square of their amplitudes. It may seem that M^2 is indefinite, but this impression is wrong. If we use the light-cone gauge, M^2 becomes

$$M^2 = T \sum_{m=1}^{\infty} (\alpha_m{}^i)^* \alpha^i_m , \tag{5.327}$$

which is explicitly positive definite.

More complicated string models contain Grassmannian degrees of freedom along with transverse modes in their vibration spectra. *Superstrings* extend reparametrization and Weyl invariance to incorporate supersymmetry on the world sheet. The *Green–Schwarz string* enjoys also the property of supersymmetry in spacetime, which is believed to be a key ingredient in fundamental theory. We now review very briefly the Green–Schwarz string without going into supersymmetry except to note that there are equal numbers of transverse and Grassmannian degrees of freedom. A consistent quantum description of this string requires a spacetime of dimension $D+1 = 10$. The supermanifold is specified by 10 spacetime coordinates X^μ, and 32 anticommuting real-valued spinor components θ. Adopting the noncovariant light cone gauge for X^μ, and imposing a constraint on θ of the form $(\gamma_0 + \gamma_1)\,\theta = 0$, where γ_μ are the 32×32 Dirac matrices in 10 dimensions, we reduce the number of components in θ to 16. With these conditions the Green–Schwartz action takes the form:

$$S = -\frac{T}{2\pi} \int d^2u \left(\eta^{ab}\partial_a X^i\, \partial_b X_i + i\,\theta\,\rho^a \partial_a\,\theta\right) \,. \tag{5.328}$$

Here, ρ^a are two-dimensional Dirac matrices, expressed in terms of the Pauli matrices $\rho^0 = i\sigma_2$, $\rho^1 = \sigma_1$, which act on $\theta = (\theta^1, \theta^2)$. The Euler–Lagrange equations read

$$\Box\, X^i = 0 \,, \tag{5.329}$$

$$(\partial_\tau + \partial_\sigma)\,\theta^1 = 0, \quad (\partial_\tau - \partial_\sigma)\,\theta^2 = 0 \,. \tag{5.330}$$

Equations (5.330) reduce the number of Grassmannian variables θ to 8. Thus, the number of independent components of θ is 8, which is just the number of transverse coordinates X^i.

We do not consider the manifestly supersymmetric formulation of this string. We only remark that a supersymmetric action is available, and that the Euler–Lagrange equations resulting from it are highly nonlinear. In the light cone gauge, the dynamical equations for strings and superstrings, (5.304) and (5.329)–(5.330), are linear.

To couple a free open string to an external electromagnetic field $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, we add an interaction term to the free action. This term should be chosen in a form preserving most, or, better still, all symmetries of the free action. The only Poincaré and reparametrization invariant expression is

$$S_{\text{int}} = \frac{e}{\pi} \int_{\tau_1}^{\tau_2} d\tau \int_0^l d\sigma\, \dot{X}_\mu\, X'_\nu\, F^{\mu\nu}(X) \,, \tag{5.331}$$

where e stands for the electric charge of the string. Because

$$\dot{X}_\mu \frac{\partial}{\partial X_\mu} A^\nu = \frac{\partial}{\partial \tau} A^\nu, \quad X'_\nu \frac{\partial}{\partial X_\nu} A^\mu = \frac{\partial}{\partial \sigma} A^\mu,$$

(5.331) equals

$$S_{\text{int}} = -\frac{e}{\pi} \int_{\tau_1}^{\tau_2} d\tau\, \dot{X}_\mu(\tau)\, A^\mu(X)\Big|_{\sigma=0}^{\sigma=l} \,, \tag{5.332}$$

plus two terms at $\tau = \tau_1$ and $\tau = \tau_2$, which do not contribute to the Euler–Lagrange equations. It is clear from (5.332) that the charge of an open string is located at its ends.

Adding (5.332) to the free action leaves the Euler–Lagrange equations unchanged, but the Neumann boundary conditions (5.305) become

$$2e\, \dot{X}^\nu F_{\mu\nu}(X) = T X'_\mu, \quad \text{at} \quad \sigma = 0, l \,. \tag{5.333}$$

This example helps to illuminate a peculiar feature of string interactions: open strings interact with each other at their ends. On the quantum level, strings interact locally, without mediation of long-range fields. Open strings may join when their ends contact, a single open string may spontaneously split into two pieces, or become closed, or emit a closed string, etc. Joining and splitting are the basic interactions of strings. This form of interaction respects all symmetries of free strings.

We finally consider *p-branes*, extended objects with p spatial dimensions. In three-dimensional space, it is possible to realize only membranes, in additional to particles (0-branes) and strings (1-branes). But in a more general world with one temporal and k spatial dimensions, we can define all kinds of p-branes with $p < k$. The points of a p-brane in some ambient spacetime are specified by $X^\mu(u^0, u^1, \ldots, u^p)$. The action is proportional to the *world volume* swept out by the p-brane,

$$S = -T_p \int d^{p+1}u \sqrt{-h} \, . \tag{5.334}$$

Here, T_p is the constant of dimension $-p-1$,

$$h_{ab} = \partial_a X^\mu \, \partial_b X_\mu, \quad a, b = 0, 1, \ldots, p \, , \tag{5.335}$$

is a metric on the world volume induced by the Lorentz metric of the ambient spacetime, and $h = \det(h_{ab})$. The action (5.334) is invariant under world-volume reparametrizations

$$u^a = f^a(u'), \quad h_{ab}(u) = \frac{\partial u'^c}{\partial u^a} \frac{\partial u'^d}{\partial u^b} \, h'_{cd}(u') \, , \tag{5.336}$$

where f^a are arbitrary smooth functions. These world-volume diffeomorphisms form an infinite group, the gauge group of the p-brane.

Varying X^μ in (5.334) gives the Euler–Lagrange equations

$$\frac{1}{\sqrt{-h}} \frac{\partial}{\partial u^a} \left(\sqrt{-h} \, h^{ab} \frac{\partial}{\partial u^b} \right) X^\mu = 0 \, , \tag{5.337}$$

which are highly nonlinear. A cure for this difficulty in the case $p = 1$ is to invoke the redundancy of degrees of freedom, and linearize the equations of motion using gauge fixing conditions. Does this trick apply to $p > 1$? Were the induced metric h_{ab} convertible to the form

$$\sqrt{-h} \, h_{ab} = \Omega \, \eta_{ab}, \quad \eta_{ab} = \mathrm{diag}\,(1, -1, \ldots, -1) \, , \tag{5.338}$$

where Ω is a constant, one may hope that (5.337) linearizes. It can be shown, however, that this procedure is valid only for $p = 1$, that is, for strings (Problem 5.6.10).

Another way of looking at p-branes is to interpret certain states of string theory as manifestations of extended configurations. Some of these states appear in the low energy limit of string theory as soliton solutions. In particular, if h_{ab} defined in (5.335) is extended to $h_{ab} = g_{ab} + F_{ab}$, where g_{ab} and F_{ab} are, respectively, symmetric and antisymmetric fields on the world volume, then we encounter a nonlinear theory of the Born–Infeld type, which possesses soliton solutions (see Sect. 10.4). Other solutions describe submanifolds of the ambient spacetime on which open strings terminate. These submanifolds are dynamical objects, Dirichlet branes. Excitations of an open string imply vibrations of its associated D-brane.

Problem 5.6.1. Show that if the expression under the radical sign in (5.293) is positive, then the world sheet is timelike.

Hint By (1.182),

$$\dot{X}^2 X'^2 - (\dot{X}\cdot X')^2 = \dot{X}^2 \left(\overset{\dot{X}}{\perp} X'\right)^2 , \tag{5.339}$$

which is negative if $\dot{X}_\mu$ is timelike while X'_μ is spacelike.

Problem 5.6.2. Verify that (5.295) is identical to (5.293).

Hint By (5.294),

$$\det\left(h_{ab}\right) = \det\begin{pmatrix} \dot{X}^2 & \dot{X}\cdot X' \\ \dot{X}\cdot X' & X'^2 \end{pmatrix} = \dot{X}^2 X'^2 - (\dot{X}\cdot X')^2 \, . \tag{5.340}$$

Problem 5.6.3. Prove that the action (5.295) is invariant under transformations (5.302).

Proof The gauge transformation $u^a = f^a(u')$ implies

$$h_{ab} = \frac{\partial u'^c}{\partial u^a}\frac{\partial u'^d}{\partial u^b}\, h'_{cd}, \quad \det h = \left[\det\left(\frac{\partial u'}{\partial u}\right)\right]^2 \det h' = \frac{1}{J^2}\det h \, , \tag{5.341}$$

where J is the Jacobian of this transformation. Since the volume element d^2u transforms as $d^2u = |J|\, d^2u'$, the action (5.295) is gauge invariant.

Problem 5.6.4. Show that Noether's second theorem implies the identities

$$\dot{X}_\mu \mathcal{E}^\mu = X'_\mu \mathcal{E}^\mu = 0 \, , \tag{5.342}$$

where $\mathcal{E}^\mu$ is defined in (5.299). Compare (5.342) with (2.259).

Problem 5.6.5. Show that energy momentum is locally conserved inside the string, and no energy momentum flows into or out of the ends of open strings which obey Neumann boundary conditions (5.298). Is this statement valid if one adopts Dirichlet boundary conditions?

Problem 5.6.6. Show that the string actions (5.295) and (5.315) are equivalent.

Proof Varying g_{ab} in (5.315) gives the Euler–Lagrange equations

$$\partial_a X^\mu \partial_b X_\mu = \frac{1}{2}\, g_{ab} g^{cd} \partial_c X^\mu \partial_d X_\mu \, . \tag{5.343}$$

We have used the formula for the variation of determinants (5.120). In view of $g_{ab}g^{bc} = \delta^c_{\ a}$, and $\delta^a_{\ a} = 2$, the general solution to (5.343) is

$$g_{ab} = \Omega\, \partial_a X^\mu \partial_b X_\mu \ , \tag{5.344}$$

where Ω is an arbitrary function. The form of Ω is of no concern because of the Weyl invariance (5.318). Hence,

$$g_{ab} = h_{ab} \ . \tag{5.345}$$

Applying (5.344) and (5.345) to (5.315) gives (5.295).

Problem 5.6.7. Show that gauge conditions (5.303) and (5.319) are equivalent.

Hint With reference to Problems 5.6.2 and 5.6.6, one finds that following (5.303) results in $g_{ab} = \dot{X}^2\eta_{ab}$, hence $\sqrt{-g}\, g^{ab} = \eta^{ab}$. On the other hand, proceeding from (5.319), and taking into account that $g_{ab} = h_{ab}$, we get $\partial_a X^\mu \partial_b X_\mu = \eta_{ab}$, which is just (5.303).

Problem 5.6.8. Suppose that τ is laboratory time in a particular Lorentz frame $\tau = t$. Consider a closed string which forms a circle at $t = 0$,

$$X_0(0,\sigma) = 0,\ X_1(0,\sigma) = L\cos(\sigma/L),\ X_2(0,\sigma) = L\sin(\sigma/L),\ X_3(0,\sigma) = 0\ , \tag{5.346}$$

$$\dot{X}_0 = 1, \quad \dot{X}^i = 0, \quad i = 1,2,3\ . \tag{5.347}$$

Verify that these initial data obey the gauge conditions (5.303). Solve the Cauchy problem for the wave equation (5.304) with these initial data.

Solution: Using the d'Alembert formula (4.138), we have

$$X_0 = t, \quad X_1 = L\cos(t/L)\cos(\sigma/L), \quad X_2 = L\cos(t/L)\sin(\sigma/L), \quad X_3 = 0\ . \tag{5.348}$$

This is a pulsating circle which squeezes to a point at $t = \frac{1}{2}\,\pi(2n+1)L,\ n = 0, 1, \ldots$

Problem 5.6.9. Solve equation (5.304) using a Fourier series for a closed string.

Answer

$$X^i(\tau,\sigma) = q^\mu + \frac{p^i}{p^+}\,\tau + \frac{i}{\sqrt{2T}}\sum \frac{1}{n}\left[\alpha^i_n\, e^{\frac{2i\pi n(\sigma-\tau)}{l}} + \beta^i_n\, e^{\frac{2i\pi n(\tau+\sigma)}{l}}\right]\ . \tag{5.349}$$

Again, the summation Σ is over all positive and negative integers n except for $n = 0$, q^i is the center-of-mass position of the string, p^i its energy-momentum. However, the coefficients α^i_n and β^i_n are independent. They are the amplitudes of left-moving and right-moving waves along the closed string, subject to the conditions $\alpha^i_n = (\alpha^i_{-n})^*$ and $\beta^i_n = (\beta^i_{-n})^*$.

Problem 5.6.10. Prove that (5.337) can be linearized only for $p = 1$.

Proof To linearize (5.337), one should bring h_{ab} to a flat form (5.338). Since the gauge group (5.336) contains $p+1$ arbitrary functions, we can enforce $p+1$ arbitrary conditions. On the other hand, enforcing (5.338) represents as many conditions as there are independent entries of the symmetric $(p+1) \times (p+1)$ matrix h_{ab}: $\frac{1}{2}\left[(p+1)^2 + p + 1\right]$ minus 1, to count the overall factor Ω in (5.338). The required number of conditions, $\frac{1}{2}\left(p^2 + 3p\right)$, equals the number of gauge degrees of freedom, $p+1$, only for $p = 1$.

Notes

1. *Section 5.1.* We discuss only elementary aspects of the Lagrangian formalism in field theory. We omit the complications which arise when the Lagrangian describes a constrained system, leading to functional dependences between canonical momenta (5.17). For a close examination of constraint systems see Dirac (1964), Gitman & Tyutin (1990), and Hanson, Regge & Teitelboim (1976).

The equation of motion for a scalar field (5.33) appeared in the papers by Klein (1926), Fock (1926), Kudar (1926), and Gordon (1926). The equation of motion for a spin-$\frac{1}{2}$ field (5.37) was discovered by Dirac (1928). The Lagrangian for a massive vector field (5.65) was studied by Lanczos (1929), and Proca (1936). Rarita & Schwinger (1941) proposed the equation of motion for a spin-$\frac{3}{2}$ field (5.70). The Lagrangians governing relativistic fields are discussed in most textbooks on field theory. A general reference is Schweber (1961). Since Wigner's (1939) work, the equation of motion for a free field ϕ_a is obtained from the condition that ϕ_a transforms according to some irreducible representation of the Poincaré group. A simple generalization of this idea to interacting fields can be found in Takahashi (1969).

The field sector of the Maxwell–Lorentz theory was proposed by Larmor (1900); the particle sector is due to Poincaré (1906) and Planck (1906); the term responsible for the electromagnetic interaction was discovered by Schwarzschild (1903).

2. *Section 5.2.* Belinfante (1939), 1940 and Rosenfeld (1940) developed a procedure for making the canonical stress-energy tensor symmetric. The metric stress-energy tensor was defined by Hilbert (1915).

Umow (1874) introduced the notions of energy density u (energy per unit volume at a given point) and energy flow per unit area $\mathbf{G}$ in his treatment of elastic solids and fluids. He discovered the energy balance equation in a region free of sources and sinks of energy

$$\frac{\partial u}{\partial t} + \operatorname{div} \mathbf{G} = 0 \,. \tag{5.350}$$

Similar results were obtained by Poynting (1884*a*), (1884*b*) for the electromagnetic field. He showed that the electromagnetic energy stored in unit volume is $u = (1/8\pi)\left(\mathbf{E}^2 + \mathbf{B}^2\right)$, and the flux of energy is $\mathbf{G} = (1/4\pi)\,(\mathbf{E} \times \mathbf{B})$. The commonly accepted derivation of this statement is as follows: one looks for the rate at which the Lorentz force per unit volume $\varrho\mathbf{E} + \mathbf{j} \times \mathbf{B}$ does mechanical work on the charge distribution in a domain of unit volume. The result is $\mathbf{j} \cdot \mathbf{E}$, which can be transformed by means of Maxwell's equations to $-\,(1/8\pi)\,\partial/\partial t\left(\mathbf{E}^2 + \mathbf{B}^2\right) - (1/4\pi)\,\mathrm{div}\,(\mathbf{E} \times \mathbf{B})$. Thus the rate of change of the sum of the mechanical energy and $(1/8\pi)\left(\mathbf{E}^2 + \mathbf{B}^2\right)$ is given by the surface integral of $(1/4\pi)\,(\mathbf{E} \times \mathbf{B})$. One recognizes this as a manifestation of energy conservation if $(1/8\pi)\left(\mathbf{E}^2 + \mathbf{B}^2\right)$ is taken to be the electromagnetic energy density, and $(1/4\pi)\,(\mathbf{E} \times \mathbf{B})$ is the rate of flow of energy out of the domain through the boundary.

This view of energy balance between charged matter and electromagnetic field will be somewhat refined Chap. 6. It will be shown that the electromagnetic interaction drastically rearranges mechanical and field degrees of freedom to yield two new entities: a dressed particle and radiation.

Minkowski (1909) proposed the currently accepted interpretation for all components of the symmetric stress-energy tensor $\Theta_{\mu\nu}$ of the electromagnetic field. Following Minkowski, one thinks of a general stress-energy tensor $T_{\mu\nu}$ as the flux of the μth component of four-momentum through a unit surface perpendicular to the νth axis.

3. *Section 5.3.* Bateman (1910*a*), 1910*b*, and Cunningham (1910) showed that Maxwell's equations are invariant under inversions $x^\mu \to x^\mu/x^2$, which amounts to their invariance under special conformal transformations. In effect, the Bateman's argument suggested that the conformal group $\mathrm{C}(1,3)$ is the largest Lie group of spacetime symmetries which leaves Maxwell's equations invariant. Weyl (1918) reasoned that the length of a vector has no absolute geometric meaning; the local Poincaré symmetry inherent in small regions of the pseudo-Riemannian manifold should be extended to the local conformal symmetry of spacetime to give the geometric framework for a unified theory of electromagnetism and gravitation. Dirac (1936) used the isomorphism between the conformal group in Minkowski space $\mathrm{C}(1,3)$ and the group of pseudoorthogonal transformations $\mathrm{SO}(2,4)$ to derive electrodynamics in six dimensions. Barut & Haugen (1972) put forward the view that a six-dimensional spacetime can be taken to be true physical spacetime (the fifth and sixth dimensions being related to the scale and the change of scale from point to point) and that the physical laws take their simplest form in this six-dimensional formulation. For a historical survey and further references see Schouten (1949), Kastrup (1962), and Fulton, Rohrlich & Witten (1962). Scale invariance is reviewed by Coleman (1985). The relation between conformal and Weyl symmetries is discussed in Fulton, Rohrlich & Witten (1962).

Wess (1960) observed that conformal transformations violate causality. Rosen (1968) proposed a new interpretation of conformal transformations (other than the commonly used 'active' and 'passive' interpretations) to provide a way of avoiding causality violation.

The idea of an improved stress-energy tensor simplifying the analysis of conformal theories was advanced and developed by Callan, Coleman & Jackiw (1970). Boulware, Brown & Peccei (1970) constructed the conformal metric tensor (5.213).

4. *Section 5.4.* The idea that electric and magnetic fields appear symmetrically in Maxwell's theory goes back to Heaviside (1892). Larmor (1900) noted that the source-free Maxwell's equations are invariant under a discrete transformation $\mathbf{E} \to \mathbf{B}$, $\mathbf{B} \to -\mathbf{E}$. Rainich (1925) generalized this transformation to the duality rotation (5.218). Zwanziger (1971) proposed a local Lagrangian for electrodynamics involving electric and magnetic charges. For a review of the Dirac monopole problem see Strazhev & Tomil'chik (1973), Goddard & Olive (1978), Coleman (1983), and Blagoević & Senjanović (1988).

5. *Section 5.5.* Historically, the vector potential $\mathbf{A}$ was not always viewed as an auxiliary variable to simplify the analysis of electrodynamics. Maxwell (1873) deemed the vector potential as a fundamental field; he called it the *electrotonic intensity.* $\mathbf{E}$ and $\mathbf{B}$ did not come into their own as primary concepts of electrodynamics until Heaviside (1892).

The gauge principle was developed by Fock (1926), London (1927), and Weyl (1929). For historical detail and further references see Jackson & Okun (2001). Gell-Mann & Levy (1960) proposed a convenient technique for the Noether current associated with gauge invariance.

Velo & Zwanziger (1969) showed that the minimal coupling between a charged Rarita–Schwinger field and the electromagnetic field gives rise to acausality in the sense that solutions to the equations of motion for this system propagate faster than light.

6. *Section 5.6.* Systematic studies of string theory can be found in many books. General references are Green, Schwarz & Witten (1987), and Polchinski (1998). The texts by Kiritsis (1998) and Siegel (1999) are available at the web. A major part of string theory is outlined in qualitative terms by Greene (1999).

Nambu (1970*a*), (1970*b*), Goto (1971), and Hara (1971) suggested a relativistic string to reproduce the physics of dual resonance models in the theory of strong interactions. For historical detail and further references see Green, Schwarz & Witten (1987). A short time later, Scherk & Schwarz (1974) and Yoneya (1974) observed that the limit $T \to \infty$ of closed string scattering amplitudes reproduces the results of a variant of general relativity, while open string amplitudes approximate Maxwell's electrodynamics. The subsequent development of string theory is reviewed in Green, Schwarz & Witten (1987), and Greene (1999). This theory is now regarded as a framework for the unification of all fundamental interactions, and the value of the string tension T

is assumed to be comparable with the square of the Planck mass. This is the natural mass scale of quantum gravity, $M_{\mathrm{P}} = \sqrt{\hbar c/G} = 1.22 \times 10^{19}$ GeV, where $\hbar$ is Planck's constant, c velocity of light, and G Newton's gravitational constant. An extended discussion of branes is found in Polchinski (1998), and Johnson (2003), which can also be consulted for additional references.

6

Self-Interaction in Electrodynamics

Maxwell–Lorentz electrodynamics is formulated as a Lagrangian theory for a system of charged particles and electromagnetic field. The interaction between mechanical and electromagnetic degrees of freedom rearranges dynamical variables into *dressed particles* and *radiation*.

This *rearrangement* of degrees of freedom is common to a wide variety of interacting systems. We will see in Sect. 6.1 that dynamical variables can be rearranged even in the absence of point sources and singular fields. For example, the initial spectrum of the Goldstone model involves tachyon modes of a scalar field ϕ. The rearranged system exhibits massive and massless oscillatory modes of ϕ descended from the tachyon modes.

In Sect. 6.2 we analyze the concept of radiation in Maxwell–Lorentz theory. In Sect. 6.3, we trace the rearrangement of mechanical and electromagnetic degrees of freedom through the use of energy-momentum balance. The Lorentz–Dirac equation governing a dressed charged particle is discussed in Sect. 6.4. Two alternative ways for deriving this equation are reviewed in Sect. 6.5.

6.1 Rearrangement of Degrees of Freedom

The demarcation line between interacting and free systems is often fuzzy. As indicated in Sect. 5.1, a field is regarded as a *free* system if its behavior in Minkowski space is governed by a *linear* equation with constant coefficients. In other words, a free field ϕ comes from a Lagrangian quadratic in ϕ. An example is a scalar field ϕ obeying the Klein–Gordon equation

$$(\Box + \mu^2)\,\phi = 0\,, \tag{6.1}$$

which is derived from

$$\mathcal{L}_0 = \frac{1}{2}\,(\partial_\mu\phi\,\partial^\mu\phi) - \frac{\mu^2}{2}\,\phi^2\,. \tag{6.2}$$

Note, however, that linearity is sometimes a matter of convention, and may be abandoned as the need arises. To illustrate, the equation of motion for a free string becomes either linear or nonlinear according to which gauge condition is adopted. Furthermore, the equation of motion for a free p-brane, $p \geq 2$, is always nonlinear.

A field ϕ governed by Lagrangian (6.2) is in many respects similar to a harmonic oscillator. Let us suppose that ϕ depends on spacetime through the phase $\xi = k \cdot x$,

$$\phi = \phi(\xi) \, . \tag{6.3}$$

Partial derivatives become $\partial_\mu \phi = k_\mu \phi'$, where the prime denotes the derivative with respect to ξ. Therefore,

$$\mathcal{L}_0 = \frac{k^2}{2}\,\phi'^2 - \frac{\mu^2}{2}\,\phi^2 \, . \tag{6.4}$$

If k_μ is timelike, then ξ may be interpreted as an evolution parameter, and (6.4) is the Lagrangian for a one-dimensional harmonic oscillator with kinetic energy $T = (k^2/2)\,\phi'^2$, and potential energy

$$U = \frac{\mu^2}{2}\,\phi^2 \, . \tag{6.5}$$

Substituting the Fourier transform

$$\phi(x) = \frac{1}{(2\pi)^4} \int d^4k\, e^{-ik\cdot x}\, \tilde{\phi}(k) \tag{6.6}$$

in (6.1) gives

$$k^2 = \mu^2 \, . \tag{6.7}$$

We arrive at the so-called *dispersion law* for this system. One may think of k^μ as the four-momentum[1] of a plane wave $e^{-ik\cdot x}$. Then, by (6.7), any Fourier mode of ϕ manifests itself as a particle with mass μ. If $\mu = 0$, then the particle is massless.

However, there are free fields whose behavior is much different from that of a harmonic oscillator. Let us change the sign of the mass term in (6.2),

$$\mathcal{L}_0 = \frac{1}{2}\,(\partial_\mu\phi\,\partial^\mu\phi) + \frac{\mu^2}{2}\,\phi^2 \, . \tag{6.8}$$

The field equation becomes

$$(\Box - \mu^2)\,\phi = 0 \, . \tag{6.9}$$

Taking ϕ in the form (6.3), we have

[1] In fact, Louis de Broglie proposed in 1923 that the propagation vector k^μ in a plane wave of any matter field ϕ can be likened to the energy-momentum carried by this wave.

$$\mathcal{L}_0 = \frac{k^2}{2}\,\phi'^{\,2} + \frac{\mu^2}{2}\,\phi^2\,, \tag{6.10}$$

which is the Lagrangian for a particle moving in a one-dimensional potential

$$U = -\frac{\mu^2}{2}\,\phi^2\,. \tag{6.11}$$

One can envision a particle which is at rest on the top of the potential hill $\phi = 0$ or moving downhill to $\phi \to \infty$ or $\phi \to -\infty$, rather than oscillating in a potential well.

By the dispersion law resulting from (6.9),

$$k^2 = -\mu^2\,, \tag{6.12}$$

a single Fourier mode of ϕ manifests itself as a tachyon. Of course, the wave front moves at a velocity lower than that of light. The Klein–Gordon equation (6.9) is hyperbolic for any value of μ, because the hyperbolicity is related to highest derivatives. The characteristic surface of equation (6.9) is the light cone $k^2 = 0$. A direct calculation (Problem 6.1.1) shows that ϕ propagates inside the light cone. Fourier modes of ϕ are similar to tachyons only in that their momenta k^μ are spacelike.

We now turn our attention to interacting systems. We begin with a scalar field ϕ governed by the Lagrangian (5.34) which involves a quartic self-interaction,

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu\phi\,\partial^\mu\phi\right) - \frac{\mu^2}{2}\,\phi^2 - \frac{\lambda^2}{4}\,\phi^4\,. \tag{6.13}$$

Taking ϕ in the form (6.3), we obtain a one-dimensional anharmonic oscillator,

$$\mathcal{L} = \frac{k^2}{2}\,\phi'^{\,2} - \frac{\mu^2}{2}\,\phi^2 - \frac{\lambda^2}{4}\,\phi^4\,. \tag{6.14}$$

The motion of a particle in the potential

$$U = \frac{\mu^2}{2}\,\phi^2 + \frac{\lambda^2}{4}\,\phi^4 \tag{6.15}$$

is qualitatively the same as that in the potential (6.5). The only difference is that the period of a harmonic oscillation in the potential (6.5) is independent of its amplitude, while the period of oscillations in the potential (6.15) is amplitude-dependent.

Meanwhile there are systems whose behavior changes drastically by switching-on the interaction. Consider a scalar real field ϕ which is governed by

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu\phi\,\partial^\mu\phi\right) + \frac{\mu^2}{2}\,\phi^2 - \frac{\lambda^2}{4}\,\phi^4 - U_0\,, \tag{6.16}$$

where

$$U_0 = \frac{\mu^4}{4\lambda^2} \tag{6.17}$$

is a constant which is suitable for the subsequent analysis. The Lagrangian (6.16) can be written in a compact form:

$$\mathcal{L} = \frac{1}{2}(\partial\phi)^2 - U \ , \tag{6.18}$$

where

$$U = \frac{\lambda^2}{4}\left(\phi^2 - \phi_0^2\right)^2 \ , \tag{6.19}$$

$$\phi_0 = \frac{\mu}{\lambda} \ . \tag{6.20}$$

Note that the Lagrangian (6.18) is invariant under reflection $\phi \to -\phi$.

We now look for a state of minimal energy, which is called the *ground state*. The energy of the system is

$$E = \int d^3x \left[\frac{1}{2}(\partial_0\phi)^2 + \frac{1}{2}(\nabla\phi)^2 + U\right] \ . \tag{6.21}$$

The derivative terms are minimized when ϕ is a constant. If this constant corresponds to the minimum of $U(\phi)$, then it is seen from (6.19) and (6.20) that the ground state is associated with either of two points

$$\phi = \pm\phi_0 \ . \tag{6.22}$$

If the system is in one of these points, reflection invariance ($\phi \to -\phi$) is broken, as is evident from Fig. 6.1.

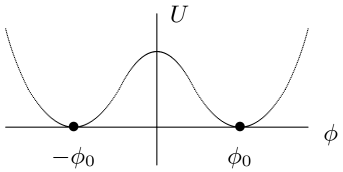


Fig. 6.1. $U(\phi)$ defined in (6.19)

Assume that the ground state is realized at $\phi = \phi_0$. Taking

$$\phi = \phi_0 + \varphi \ , \tag{6.23}$$

where φ is a small perturbation about ϕ_0, we have

$$\phi^2 = \phi_0^2 + 2\phi_0\varphi + \varphi^2 \ , \tag{6.24}$$

and

$$U = \mu^2\varphi^2 + \lambda\mu\,\varphi^3 + \frac{\lambda^2}{4}\,\varphi^4 \ . \tag{6.25}$$

The Lagrangian

$$\mathcal{L} = \frac{1}{2}\,(\partial_\mu\varphi\,\partial^\mu\varphi) - \mu^2\varphi^2 - \lambda\mu\,\varphi^3 - \frac{\lambda^2}{4}\,\varphi^4 \tag{6.26}$$

governs a system which exhibits an oscillatory mode with mass

$$m = \sqrt{2}\mu \ , \tag{6.27}$$

instead of the initial tachyon mode.

The initial system was in the state of unstable equilibrium. Upon rearranging degrees of freedom, as shown explicitly in (6.23), the system comes to a stable equilibrium. The price for this stability is that the rearranged Lagrangian (6.26) is not invariant under reflection $\varphi \to -\varphi$. This phenomenon is called *spontaneous symmetry breaking*.

We now turn to a more involved model, which is referred to as the *Goldstone model*,

$$\mathcal{L} = \frac{1}{2}\,(\partial_\mu\phi)^*\,\partial^\mu\phi + \frac{\mu^2}{2}\,\phi^*\,\phi - \frac{\lambda^2}{4}\,(\phi^*\,\phi)^2 - \frac{\mu^4}{4\lambda^2} \ . \tag{6.28}$$

Here, ϕ is a complex-valued field. Writing

$$\phi = A + iB \ , \tag{6.29}$$

we obtain a system of two interacting real fields A and B:

$$\begin{aligned}\mathcal{L} = {} & \frac{1}{2}\,(\partial_\mu A\,\partial^\mu A)^2 + \frac{\mu^2}{2}\,A^2 + \frac{1}{2}\,(\partial_\mu B\,\partial^\mu B)^2 \\ & + \frac{\mu^2}{2}\,B^2 - \frac{\lambda^2}{4}\,(A^2+B^2)^2 - \frac{\mu^4}{4\lambda^2} \ .\end{aligned} \tag{6.30}$$

The Lagrangian (6.28) is invariant under global U(1) transformations

$$\phi \to \phi' = e^{ie\omega}\phi \ . \tag{6.31}$$

The corresponding SO(2) invariance of the Lagrangian (6.30) is given by

$$\begin{aligned} A \to A' &= A\cos{(e\omega)} + B\sin{(e\omega)}\,, \\ B \to B' &= -A\sin{(e\omega)} + B\cos{(e\omega)} \ . \end{aligned} \tag{6.32}$$

The total energy of the system is

$$E = \int d^3x\,\left[\frac{1}{2}\,(\partial_0 A)^2 + \frac{1}{2}\,(\nabla A)^2 + \frac{1}{2}\,(\partial_0 B)^2 + \frac{1}{2}\,(\nabla B)^2 + U\right] \tag{6.33}$$

where

$$U = \frac{\lambda^2}{4}\left(A^2 + B^2 - \frac{\mu^2}{\lambda^2}\right)^2 . \tag{6.34}$$

Equation (6.33) suggests that the ground state is attained for constant fields A and B determined by the condition that U is minimized. By (6.34), this condition is met on the circle

$$A^2 + B^2 = \frac{\mu^2}{\lambda^2} . \tag{6.35}$$

As with the system (6.16), switching-on interaction rearranges degrees of freedom in this system. Tachyon modes disappear leaving behind oscillatory modes. A new feature of the Goldstone model is that the resulting oscillatory modes include not only massive modes but also *massless* ones. The system also acquires stability at the cost of spontaneous symmetry breakdown. The occurrence of massless modes is due to the fact that the spontaneously broken group is continuous. This statement is known as the *Goldstone theorem*. The resulting massless modes are called *Goldstone modes*.

The Goldstone theorem is most easily proved if ϕ is expressed in terms of two real fields ρ and θ,

$$\phi = \rho\, e^{i\theta} . \tag{6.36}$$

Then (6.28) becomes

$$\mathcal{L} = \frac{1}{2}\,\rho^2\,(\partial_\mu\theta\,\partial^\mu\theta) + \frac{1}{2}\,(\partial_\mu\rho\,\partial^\mu\rho) - U , \tag{6.37}$$

where

$$U = \frac{\lambda^2}{4}\left(\rho^2 - \rho_0^2\right)^2 , \tag{6.38}$$

$$\rho_0 = \frac{\mu}{\lambda} . \tag{6.39}$$

By (6.38), the ground state is given by $\rho = \rho_0$ and arbitrary θ. For definiteness, we fix θ to be a constant θ_0. Letting

$$\rho = \rho_0 + \varrho, \quad \theta = \theta_0 + \vartheta , \tag{6.40}$$

where ϱ and ϑ are perturbations about the equilibrium, we bring (6.37) to the form

$$\mathcal{L} = \frac{1}{2}\,(\rho_0 + \varrho)^2\,(\partial_\mu\vartheta\,\partial^\mu\vartheta) + \frac{1}{2}\,(\partial_\mu\varrho\,\partial^\mu\varrho) - \mu^2\,\varrho^2 - \lambda\mu\,\varrho^3 - \frac{\lambda^2}{4}\,\varrho^4 . \tag{6.41}$$

We see that the free part of the Lagrangian

$$\mathcal{L}_0 = \frac{1}{2}\,(\partial_\mu\varrho\,\partial^\mu\varrho) - \mu^2\,\varrho^2 + \frac{1}{2}\left(\rho_0^2\,\partial_\mu\vartheta\,\partial^\mu\vartheta\right) \tag{6.42}$$

contains an oscillatory massive mode ϱ and a massless mode $\rho_0\vartheta$. The emergency of the massless mode could be expected from geometric arguments. For $\rho = \rho_0$, the coordinate θ circles the bottom of the potential valley (6.38). The absence of oscillations along θ is due to the fact that the curvature in this direction is zero.

Even greater rearrangement of dynamical variables occurs in the *Higgs model*

$$\mathcal{L} = -\frac{1}{16\pi} F^{\mu\nu} F_{\mu\nu} + \frac{1}{2} (D_\mu\phi)^* D^\mu\phi + \frac{\mu^2}{2} \phi^*\phi - \frac{\lambda^2}{4} (\phi^*\phi)^2 - \frac{\mu^4}{\lambda^2} \,. \quad (6.43)$$

Here, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, and $D_\mu\phi = (\partial_\mu - ieA_\mu)\,\phi$. This Lagrangian describes the interaction between the electromagnetic field and a complex scalar field ϕ. The sign of the term $(\mu^2/2)\,(\phi^*\phi)$ is such that ϕ is a tachyon. The Lagrangian (6.43) is invariant under gauge transformations

$$\begin{aligned} \phi'(x) &= e^{ie\omega(x)}\phi(x) \,, \\ A'_\mu(x) &= A_\mu(x) + \partial_\mu\omega(x) \,. \end{aligned} \quad (6.44)$$

The total energy

$$E = \int d^3x \left[\frac{1}{8\pi}(\mathbf{E}^2 + \mathbf{B}^2) + \frac{1}{2}(D_0\phi)^*\, D^0\phi + \frac{1}{2}(D_i\phi)^*\, D^i\phi + \frac{\lambda^2}{4}(\phi^*\phi - \phi_0^2)^2 \right] \quad (6.45)$$

is explicitly gauge invariant. Therefore, the ground state is infinitely degenerate. Indeed, the ground state is associated with an entire class of field configurations related by gauge transformations (6.44). We choose a particular field configuration from this class and consider perturbations about it.

The first term of (6.45) is positive definite. The minimum of this term is achieved for $\mathbf{E} = \mathbf{0}$ and $\mathbf{B} = \mathbf{0}$, which implies ,

$$A_\mu = \partial_\mu\omega \,. \quad (6.46)$$

The covariant derivative terms (6.45) are minimized for

$$D_\mu\phi = (\partial_\mu - ie\,\partial_\mu\omega)\,\phi = 0 \,, \quad (6.47)$$

whence

$$\phi = \phi_0 e^{ie\omega} \,. \quad (6.48)$$

The constant ϕ_0 is determined by the condition that the last term of (6.45) is minimized. This term is positive definite, and vanishes for

$$\phi = \phi_0 = \frac{\mu}{\lambda} \,. \quad (6.49)$$

In (6.46) and (6.48), ω is an arbitrary function, which implies that the ground state is degenerate. Putting $\omega = 0$, we select the ground state to be

$$A_\mu = 0, \quad \phi = \phi_0 \; . \tag{6.50}$$

One may perturb the scalar field about ϕ_0:

$$\phi = \phi_0 + \alpha + i\beta \; . \tag{6.51}$$

Perturbations of the vector potential and the field strength about zero values of these quantities will be denoted by the same symbols A_μ and $F_{\mu\nu}$. Keeping only terms of the first order in perturbations, the covariant derivative becomes

$$D_\mu\phi = \partial_\mu\alpha + i\,\partial_\mu\beta - ie\phi_0 A_\mu \; . \tag{6.52}$$

Since $\phi^*\phi = (\phi_0+\alpha)^2 + \beta^2$, the potential U, up to terms quadratic in α and β, is

$$U = \frac{\lambda^2}{4}\left(\phi^*\phi - \phi_0^2\right)^2 = \frac{\lambda^2}{4}\left(2\phi_0\alpha + \alpha^2 + \beta^2\right)^2 = \mu^2\alpha^2 \; , \tag{6.53}$$

where (6.49) has been taken into account. Our concern is with terms of the Lagrangian

$$\mathcal{L} = -\frac{1}{16\pi}\,F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}\left|\partial_\mu\alpha + i\partial_\mu\beta - ie\phi_0 A_\mu\right|^2 - \mu^2\alpha^2 \tag{6.54}$$

which are quadratic in perturbations:

$$-\frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}(\partial_\mu\alpha\,\partial^\mu\alpha) - \mu^2\alpha^2 + \frac{e^2\phi_0^2}{2}\left(A_\mu - \frac{1}{e\phi_0}\partial_\mu\beta\right)\left(A^\mu - \frac{1}{e\phi_0}\partial^\mu\beta\right) . \tag{6.55}$$

The mixed term $A^\mu\partial_\mu\beta$ prevents a clear interpretation of this expression. To remedy the situation, we define

$$a_\mu = A_\mu - \frac{1}{e\phi_0}\partial_\mu\beta \; . \tag{6.56}$$

With this new field, the quadratic form (6.55) is diagonal:

$$\mathcal{L}_0 = -\frac{1}{16\pi}\,f^{\mu\nu}f_{\mu\nu} + \frac{e^2\phi_0^2}{2}\,a^\mu a_\mu + \frac{1}{2}\,(\partial_\mu\alpha\,\partial^\mu\alpha) - \mu^2\alpha^2 \; , \tag{6.57}$$

where $f_{\mu\nu} = \partial_\mu a_\nu - \partial_\nu a_\mu$.

To summarize, we come to the Lagrangian (6.57) containing only a massive vector field a_μ with mass

$$M = 4\pi e\phi_0 = 4\pi\frac{e}{\lambda}\,\mu \; , \tag{6.58}$$

and a massive real scalar field α with mass

$$m = \sqrt{2}\mu \; . \tag{6.59}$$

The massless Goldstone field β disappears[2]. We see that the vector field a_μ acquires mass by absorbing the Goldstone field β.

The initial Higgs Lagrangian (6.43) was formulated in terms of a complex-valued tachyon field ϕ having two real components, and a massless vector field A_μ having two polarization degrees of freedom. Upon rearranging degrees of freedom, shown in (6.51) and (6.56), unstable modes disappear. The rearranged Higgs system involves stable quantities: one massive vector field a_μ with three polarization degrees of freedom, and one massive scalar field α. The number of field components is apparently preserved.

It is notable that the rearranged Lagrangian (6.57) is gauge invariant despite the fact that a_μ is massive. One can verify (Problem 6.1.3) that the vector field a_μ defined in (6.56) is gauge invariant by itself. The U(1) gauge symmetry is not really broken, it is merely hidden through the Higgs mechanism. This is suggested by the persistence of the chief consequence of this symmetry: the current j^μ associated with the U(1) gauge group is still conserved (Problem 6.1.4).

Problem 6.1.1. Let ϕ be a free tachyon field in a world with one temporal and one spatial dimension. The retarded Green's function for the equation

$$\left(\Box - \mu^2\right) G(t,x) = 2\delta(t)\,\delta(x) \tag{6.60}$$

obeys the retarded boundary condition: $G(t,x) = 0$ for $t < 0$. We define an auxiliary function

$$g(t,x) = e^{-\mu t}\, G(t,x)\,, \tag{6.61}$$

where $\mu > 0$. This function satisfies the equation

$$g_{tt} - g_{xx} + 2\mu\, g_t = 2\delta(t)\,\delta(x) \tag{6.62}$$

where subscripts of g denote partial derivatives with respect to the coordinates shown. To solve (6.62), we use the Fourier transform of g,

$$g(t,x) \sim \int d\omega\, dk\, \frac{e^{-i(\omega t - kx)}}{-2i\mu\,\omega - \omega^2 + k^2}\,. \tag{6.63}$$

Changing the integration variables $k_+ = \omega + k$ and $k_- = \omega - k$, and introducing the light cone coordinates $x_+ = t + x$ and $x_- = t - x$,

$$g(x_-,x_+) \sim \int dk_+ dk_-\, \frac{e^{-\frac{i}{2}(k_+x_- + k_-x_+)}}{k_+k_- + i\mu\,(k_+ + k_-)}\,. \tag{6.64}$$

Prove that the support of $g(x_-,x_+)$ is the future light cone $x_+ > 0$, $x_- > 0$, that is, $g(x_-,x_+) = 0$ for $x_+ < 0$, or for $x_- < 0$.

Proof Let $x_+ < 0$. We carry out the integration over k_- using the Cauchy theorem for the closed contour composed of the real axis and a semicircle of

[2] This can be shown not only approximately, but also exactly (Problem 6.1.2).

large radius in the upper half-plane of the complex variable $k_- = \xi + i\zeta$. Since the pole has coordinates

$$\xi = -\frac{\mu^2 k_+}{k_+^2 + \mu^2}, \quad \zeta = -\frac{\mu k_+^2}{k_+^2 + \mu^2}, \tag{6.65}$$

it does not arise for real k_+ and $\zeta > 0$, hence $g = 0$. Likewise, $g = 0$ for $x_- < 0$, because the integral (6.64) is invariant under interchanging $(x_-, k_+) \to (x_+, k_-)$.

Problem 6.1.2. What is the transformation which removes completely the Goldstone field from the Higgs Lagrangian? Write down the transformed Lagrangian.

Answer Using local gauge invariance (6.44), one can render the complex-valued field $\phi = \alpha + i\beta$ real, $\phi = \alpha$ (the so-called *unitary gauge*). With this in mind, ω is chosen such that $\tan e\omega = \beta/\alpha$, as (6.32) suggests. The transformed Lagrangian is

$$\mathcal{L} = -\frac{1}{16\pi} F^{\mu\nu} F_{\mu\nu} + \frac{1}{2} (\partial^\mu \alpha \, \partial_\mu \alpha) + \frac{e^2}{2} \alpha^2 A_\mu A^\mu - U(\alpha) \tag{6.66}$$

where U is given by (6.19).

Problem 6.1.3. Show that the massive vector field a_μ defined in (6.56) is invariant under the local gauge transformations (6.44) with infinitesimal gauge parameters ω.

Problem 6.1.4. (i) Consider the current j^μ defined in (5.273) where ϕ and A_μ are field variables of the Higgs model after the rearrangement represented by (6.51) and (6.56). Prove that j^μ is conserved. (ii) Show that this conservation implies that the action for the rearranged Higgs model is invariant under the U(1) gauge group.

Hint (i) Apply the Ward–Takahashi identity (5.279) to the rearranged Higgs system. (ii) Use the Gell-Mann–Levy identity (5.282).

6.2 Radiation

Our starting point is the action for a single charged particle and an electromagnetic field:

$$S = -m_0 \int d\tau \sqrt{\dot{z} \cdot \dot{z}} - \int d^4x \left(A_\mu j^\mu + \frac{1}{16\pi} F_{\mu\nu} F^{\mu\nu} \right) . \tag{6.67}$$

Here, m_0 stands for the *mechanical* mass of the *bare* particle, that is, an imaginary particle devoid of its surrounding electromagnetic field. The Euler–Lagrange equations read:

$$\mathcal{E}_\mu = \partial^\nu F_{\mu\nu} + 4\pi j_\mu = 0 \ , \tag{6.68}$$

$$\varepsilon^\lambda = m_0 a^\lambda - e v_\mu F^{\lambda\mu} = 0 \ . \tag{6.69}$$

The field equation (6.68) is supplemented with the Bianchi identity

$$\mathcal{E}^{\lambda\mu\nu} = \partial^\lambda F^{\mu\nu} + \partial^\nu F^{\lambda\mu} + \partial^\mu F^{\nu\lambda} = 0 \ . \tag{6.70}$$

The challenge now is to find a simultaneous solution of equations (6.68)–(6.70). We begin with (6.70) and (6.68) on the assumption that the world line of the source $z^\mu(s)$ is an arbitrary timelike smooth curve. We are already aware of the general solution to equation (6.70):

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \ . \tag{6.71}$$

Imposing the retarded boundary condition, we obtain a solution to equation (6.68) in the form of the Liénard–Wiechert vector potential

$$A^\mu = e\,\frac{v^\mu}{\rho} \ . \tag{6.72}$$

However, the Liénard–Wiechert field strength

$$F^{\mu\nu} = \frac{e}{\rho^2}\left(c^\mu V^\nu - c^\nu V^\mu\right) \ , \tag{6.73}$$

$$V^\mu = v^\mu + \rho\,(\overset{u}{\perp}\, a)^\mu \tag{6.74}$$

is singular on the world line. Hence, substituting the retarded solution of Maxwell's equations into the equation of motion of the particle (6.69) would result in a divergent expression. One may consider this divergence as a manifestation of infinite self-interaction in this system: a charged particle experiences its own electromagnetic field which is infinite at the point of origin.

A possible cure for this difficulty is to regularize the Liénard–Wiechert field. In order to analyze regularized expressions, it is convenient to use the Noether identity (5.142),

$$\partial_\mu T^{\lambda\mu} = \frac{1}{8\pi}\,\mathcal{E}^{\lambda\mu\nu} F_{\mu\nu} + \frac{1}{4\pi}\,\mathcal{E}_\mu F^{\lambda\mu} + \int_{-\infty}^{\infty} ds\, \varepsilon^\lambda(z)\, \delta^4\left[x - z(s)\right] \ . \tag{6.75}$$

Here, $T^{\mu\nu}$ is the symmetric stress-energy tensor of this system,

$$T^{\mu\nu} = \Theta^{\mu\nu} + t^{\mu\nu} \ , \tag{6.76}$$

$$\Theta^{\mu\nu} = \frac{1}{4\pi}\left(F^{\mu\alpha} F_\alpha{}^\nu + \frac{1}{4}\,\eta^{\mu\nu} F_{\alpha\beta} F^{\alpha\beta}\right) \ , \tag{6.77}$$

$$t^{\mu\nu} = m_0 \int_{-\infty}^{\infty} ds\, v^\mu(s)\, v^\nu(s)\, \delta^4\left[x - z(s)\right] \ , \tag{6.78}$$

and $\mathcal{E}^{\lambda\mu\nu}$, $\mathcal{E}_\mu$, and ε^λ are, respectively, the left-hand sides of equations (6.70), (6.68), and (6.69).

What is the reason for invoking the Noether identity (6.75)? If we would ignore the divergence, then $\partial_\mu T^{\lambda\mu} = 0$ would imply $\mathcal{E}^{\lambda\mu\nu} = 0$, $\mathcal{E}_\mu = 0$, and $\varepsilon^\lambda = 0$. Therefore, the local conservation law for the stress-energy tensor is formally equivalent to the equation of motion for a bare particle (6.69) in which a simultaneous solution to the field equations (6.70) and (6.68) is used. We will see, however, that the equation $\partial_\mu T^{\lambda\mu} = 0$ is more illuminating to describe the structure of singular self-interactions. In particular, the form of $\Theta^{\mu\nu}$ provides insight into the nature of integrable and nonintegrable singularities.

In this section, our discussion is centered around integrable singularities. We may therefore be careless of regularization.

Substituting (6.73) and (6.74) into (6.77) gives:

$$\Theta^{\mu\nu} = \frac{e^2}{4\pi\rho^4}\left(c^\mu V^\nu + c^\nu V^\mu - V^2 c^\mu c^\nu - \frac{1}{2}\,\eta^{\mu\nu}\right) . \tag{6.79}$$

Since $V^2 = 1 + \rho^2\,(\overset{u}{\perp} a)^2$, expression (6.79) splits into two parts:

$$\Theta^{\mu\nu} = \Theta^{\mu\nu}_{\mathrm{I}} + \Theta^{\mu\nu}_{\mathrm{II}} , \tag{6.80}$$

$$\Theta^{\mu\nu}_{\mathrm{I}} = \frac{e^2}{4\pi\rho^4}\left(c^\mu V^\nu + c^\nu V^\mu - c^\mu c^\nu - \frac{1}{2}\,\eta^{\mu\nu}\right) , \tag{6.81}$$

$$\Theta^{\mu\nu}_{\mathrm{II}} = -\frac{e^2}{4\pi\rho^2}\,(\overset{u}{\perp} a)^2\, c^\mu\, c^\nu . \tag{6.82}$$

The first part $\Theta^{\mu\nu}_{\mathrm{I}}$ containing terms proportional to ρ^{-3} and ρ^{-4} may be thought of as 'near' energy-momentum density, while the second part $\Theta^{\mu\nu}_{\mathrm{II}}$, which is proportional to ρ^{-2}, is interpreted as 'far' energy-momentum density. One can show (Problem 6.2.2) that the following equations hold off the world line:

$$\partial_\mu \Theta^{\mu\nu}_{\mathrm{I}} = 0, \quad \partial_\mu \Theta^{\mu\nu}_{\mathrm{II}} = 0 . \tag{6.83}$$

These local conservation laws imply that $\Theta^{\mu\nu}_{\mathrm{I}}$ and $\Theta^{\mu\nu}_{\mathrm{II}}$ are dynamically independent off the world line. This fact was discovered by Claudio Teitelboim in 1970.

Because $\Theta^{\mu\nu}_{\mathrm{II}}$ behaves like ρ^{-2} near the world line, this part of the stress-energy tensor involves integrable singularities. Following Teitelboim, we call $\Theta^{\mu\nu}_{\mathrm{II}}$ the *radiation*.

What are characteristic features of the radiation? We first note that $\Theta^{\mu\nu}_{\mathrm{II}}$ leaves the source at the speed of light. Indeed, by (4.247), the surface element of the future light cone C_+ is $d\sigma^\mu = c^\mu \rho^2 d\rho\, d\Omega$. Since c^μ is a null vector, the flux of $\Theta^{\mu\nu}_{\mathrm{II}}$ through C_+ vanishes, $d\sigma_\mu \Theta^{\mu\nu}_{\mathrm{II}} = 0$. This means that $\Theta^{\mu\nu}_{\mathrm{II}}$ propagates along rays of C_+.

Another feature of the radiation is that the energy-momentum flux of $\Theta^{\mu\nu}_{\mathrm{II}}$ varies as ρ^{-2} implying that the same amount of energy-momentum flows through spheres of different radii.

Note that $a^\mu = 0$ implies $\Theta_{\rm II}^{\mu\nu} = 0$. The radiation arises only when charged particles accelerate or decelerate.

These features of $\Theta_{\rm II}^{\mu\nu}$ form the basis for practical implementations of the radiation.

None of these features is shared by $\Theta_{\rm I}^{\mu\nu}$. Indeed, let $d\sigma^\mu$ be the surface element of the future light cone C_+, then

$$d\sigma_\mu\,\Theta_{\rm I}^{\mu\nu} = \frac{e^2}{8\pi\rho^4}\,c^\nu\,\rho^2 d\rho\,d\Omega\;. \tag{6.84}$$

The flux of $\Theta_{\rm I}^{\mu\nu}$ through C_+ is nonzero. Therefore, $\Theta_{\rm I}^{\mu\nu}$ moves slower than light. The propagation of $\Theta_{\rm II}^{\mu\nu}$ and $\Theta_{\rm I}^{\mu\nu}$ is shown in Fig. 6.2. One may conclude that $\Theta_{\rm II}^{\mu\nu}$ detaches from the source, while $\Theta_{\rm I}^{\mu\nu}$ remains bound to it. In other words, $\Theta_{\rm I}^{\mu\nu}$ represents a part of the electromagnetic energy-momentum that is dragged by the charge.

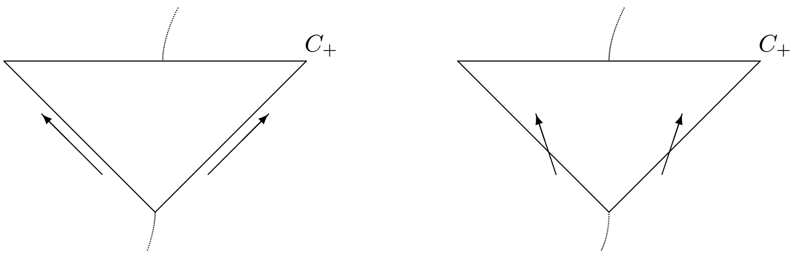


Fig. 6.2. Propagation of $\Theta_{\rm II}^{\mu\nu}$ (*left plot*) and $\Theta_{\rm I}^{\mu\nu}$ (*right plot*) with respect to the future light cone C_+

It is clear from (6.81) and (6.74) that $\Theta_{\rm I}^{\mu\nu}$ falls with distance at least as ρ^{-3}. Therefore, $\Theta_{\rm I}^{\mu\nu}$ yields the flux of energy-momentum which dies out with distance. Note also that $\Theta_{\rm I}^{\mu\nu}$ is nonvanishing for any motion of the source.

One may wonder whether it is possible to identify the degrees of freedom related to the radiation directly in the Liénard–Wiechert field. Let us split the field strength: $F = F_{\rm I} + F_{\rm II}$, where

$$F_{\rm I} = \frac{e}{\rho^2}\,c\wedge v\;, \tag{6.85}$$

$$F_{\rm II} = \frac{e}{\rho}\,c\wedge(\overset{u}{\perp} a)\;. \tag{6.86}$$

One may regard $F_{\rm I}$ as a 'generalized Coulomb field' (or 'velocity field', or 'near field'), and $F_{\rm II}$ as the 'radiation field' (other names are 'acceleration field' and 'far field'). Note that $\Theta_{\rm II}$ contains only $F_{\rm II}$. Furthermore, when built from $F_{\rm II}$, the invariants $\mathcal{P}$ and $\mathcal{S}$ are zero (Problem 6.2.4). This is the reason for denoting $F_{\rm II}$ as 'null field'.

The terms 'near zone' and 'wave zone' are still common in the literature. One might define the *wave zone* as a domain far away from the source where $F_{\rm II}$ dominates over $F_{\rm I}$. The boundary between the near zone and the wave zone is expressed in invariant terms by $F_{\rm I} = F_{\rm II}$. This boundary is given by the intersection of a tubular hypersurface $V^2 = 0$ enveloping the world line and a spacelike hyperplane Σ. It can be shown, however, that there is a direction for which this boundary is separated from the source by an infinite distance (Problem 6.2.6). This suggests that the notion of a wave zone is problematic.

One can show (Problem 6.2.5) that $F_{\rm I}$ and $F_{\rm II}$ are not dynamically independent.

Furthermore, we learned in Sect. 4.6 that there is a (noninertial) frame of reference where only $F_{\rm I}$ is nonzero, while $F_{\rm II}$ disappears, that is, $|\mathbf{E}| = e/\rho^2$ and $\mathbf{B} = 0$ at each spacetime point. Recall that $F_{\rm II}$ is eliminated by a local $\mathrm{SL}(2,\mathbb{R})$ transformation which leaves the Liénard–Wiechert field F invariant.

Therefore, the radiation is determined not only by the field F as such but also by the frame of reference in which F is measured. On the other hand, $\Theta^{\mu\nu}$ given by (6.79) is not invariant under such $\mathrm{SL}(2,\mathbb{R})$ transformations. This quantity carries information about both the field F and the frame which is used to describe F.

To sum up, we refer to a part of the stress-energy tensor $\Theta^{\mu\nu}_{\rm II}$ as radiation if

$$\text{(i)} \quad \partial_\mu \Theta^{\mu\nu}_{\rm II} = 0\,, \tag{6.87}$$

$$\text{(ii)} \quad c_\mu \Theta^{\mu\nu}_{\rm II} = 0\,, \tag{6.88}$$

$$\text{(iii)} \quad \Theta^{\mu\nu}_{\rm II} \sim \rho^{-2}\,. \tag{6.89}$$

It is conceivable that the energy flux produced by $\Theta^{\mu\nu}_{\rm II}$ is directed inward towards the field source resulting in energy gain rather than energy loss. One may regard this as the *absorption* of radiation rather than its emission. An alternate view in this case is that the emitted energy is *negative*: $\Theta^{00}_{\rm II} = v_\mu \Theta^{\mu\nu}_{\rm II} v_\nu < 0$. Two examples can be drawn from Problem 6.2.8 and Sect. 9.1. There is no universally adopted terminology that distinguishes between $\Theta^{00}_{\rm II} > 0$ and $\Theta^{00}_{\rm II} < 0$. We will reserve the term 'radiation' for the case in which the emitted energy is positive.

In Sect. 10.2, we will analyze electrodynamics in spacetimes of dimension $D+1$ other than 4. The discussion of this section will be applicable to this analysis if we replace a sphere enclosing the source by a $(D-1)$-dimensional sphere. Then condition (iii) becomes

$$\text{(iii)} \quad \Theta^{\mu\nu}_{\rm II} \sim \rho^{1-D}\,. \tag{6.90}$$

In addition, one would like to require that $\Theta^{\mu\nu}_{\rm I}$ falls more rapidly than $\Theta^{\mu\nu}_{\rm II}$:

$$\text{(iv)} \quad \Theta^{\mu\nu}_{\rm I} = o\left(\rho^{1-D}\right), \quad \rho \to \infty\,. \tag{6.91}$$

This condition ensures that $\Theta^{\mu\nu}_{\rm II}$ is distinguished asymptotically from $\Theta^{\mu\nu}_{\rm I}$. The necessity of condition (iv) is demonstrated in Problem 6.2.9.

Problem 6.2.1. Let the frame of reference be the instantaneously comoving inertial frame of a radiated charge. Consider $\Theta^{00}_{\rm II}$ in this frame. Is $\Theta^{00}_{\rm II}$ positive definite? What is the angular dependence of radiated energy?

Answer The vector $(\overset{u}{\perp} a)^\mu$ is spacelike, $(\overset{u}{\perp} a)^2 = a^2 + (a\cdot u)^2 \le 0$. We choose the z-axis to be directed along the three-vector $\mathbf{a}$, which is defined in this frame by $a^\mu = (0, \mathbf{a})$. Then angular dependence of radiated energy is proportional to $\mathbf{a}^2 \sin^2\theta$, where θ is the polar angle between $\mathbf{a}$ and radius vector from the emission point to the observation point. There is no radiation along $\mathbf{a}$. The intensity of radiated energy is maximal along rays perpendicular to $\mathbf{a}$.

Problem 6.2.2. Using the technique developed in Sect. 4.5, verify that

$$\partial_\mu \Theta^{\mu\nu}_{\rm I} = 0, \quad \partial_\mu \Theta^{\mu\nu}_{\rm II} = 0\,. \tag{6.92}$$

Problem 6.2.3. Verify the Bianchi identity for $F^{\mu\nu}_{\rm I}$ and $F^{\mu\nu}_{\rm II}$

$$\partial^\lambda F^{\mu\nu}_{\rm I,II} + \partial^\nu F^{\lambda\mu}_{\rm I,II} + \partial^\mu F^{\nu\lambda}_{\rm I,II} = 0\,. \tag{6.93}$$

It follows that $F^{\mu\nu}_{\rm I}$ and $F^{\mu\nu}_{\rm II}$ can be derived from vector potentials $A^\mu_{\rm I}$ and $A^\mu_{\rm II}$. Find $A^\mu_{\rm I}$ and $A^\mu_{\rm II}$.

Answer

$$A^\mu_{\rm I} = \frac{e}{\rho}\, c^\mu, \quad A^\mu_{\rm II} = -\frac{e}{\rho}\, u^\mu\,. \tag{6.94}$$

Problem 6.2.4. Evaluate $\mathcal{P}$ and $\mathcal{S}$ for $F_{\rm II}$. Prove that $\mathcal{P} = 0$ and $\mathcal{S} = 0$.

Problem 6.2.5. Show that

$$\partial_\mu F^{\mu\nu}_{\rm I} = -\partial_\mu F^{\mu\nu}_{\rm II} = \frac{2e\,(a\cdot c)}{\rho^2}\, c^\nu\,. \tag{6.95}$$

Verify that $\Box F^{\mu\nu}_{\rm II}$ is nonzero. These results suggest that $F_{\rm I}$ and $F_{\rm II}$ are not dynamically independent parts of the Liénard–Wiechert field.

Problem 6.2.6. The boundary between near and wave zones is defined by the condition $F_{\rm I} = F_{\rm II}$. Bring this condition to the form $V^2 = 0$. Show that there is a direction u^μ to the observation point so that the boundary is separated from the source by an infinite distance.

Answer u^μ is aligned with a^μ.

Problem 6.2.7 *Bound and emitted angular momenta.* The angular momentum density

$$M^{\lambda\mu\nu} = x^\lambda \Theta^{\mu\nu} - x^\mu \Theta^{\lambda\nu} \tag{6.96}$$

obeys the local conservation law

$$\partial_\nu M^{\lambda\mu\nu} = 0 \tag{6.97}$$

if $\Theta^{\mu\nu}$ is symmetric and conserved. Let $\Theta^{\mu\nu}$ be the symmetric stress-energy tensor of electromagnetic field expressed in terms of the Liénard–Wiechert field. The following decomposition

$$M^{\lambda\mu\nu} = M_{\rm I}^{\lambda\mu\nu} + M_{\rm II}^{\lambda\mu\nu} \,, \tag{6.98}$$

$$M_{\rm I}^{\lambda\mu\nu} = z^\lambda \Theta_{\rm I}^{\mu\nu} - z^\mu \Theta_{\rm I}^{\lambda\nu} + R^\lambda \varsigma^{\mu\nu} - R^\mu \varsigma^{\lambda\nu} \,, \tag{6.99}$$

$$M_{\rm II}^{\lambda\mu\nu} = z^\lambda \Theta_{\rm II}^{\mu\nu} - z^\mu \Theta_{\rm II}^{\lambda\nu} + R^\lambda \vartheta^{\mu\nu} - R^\mu \vartheta^{\lambda\nu} \,, \tag{6.100}$$

proposed by Carlos López and Danilo Villarroel in 1975, splits $M^{\lambda\mu\nu}$ into bound and emitted parts. Here, $\varsigma^{\lambda\mu}$ and $\vartheta^{\lambda\mu}$ are two parts of $\Theta_{\rm I}^{\lambda\mu}$,

$$\varsigma_{\mu\nu} = -\frac{e^2}{4\pi\rho^4}\left(c_\mu c_\nu - c_\mu v_\nu - v_\mu c_\nu + \frac{1}{2}\,\eta_{\mu\nu}\right) \,, \tag{6.101}$$

$$\vartheta_{\mu\nu} = \frac{e^2}{4\pi\rho^3}\left[c_\mu \,(\overset{u}{\perp} a)_\nu + c_\nu \,(\overset{u}{\perp} a)_\mu\right] \,. \tag{6.102}$$

With these definitions, the appropriate spatial behavior of $M_{\rm I}^{\lambda\mu\nu}$ and $M_{\rm II}^{\lambda\mu\nu}$ is ensured: $M_{\rm I}^{\lambda\mu\nu} = O(\rho^{-3})$, and $M_{\rm II}^{\lambda\mu\nu} \sim \rho^{-2}$, as $\rho \to \infty$. Furthermore, $M_{\rm II}^{\lambda\mu\nu}$ obeys the condition

$$c_\nu M_{\rm II}^{\lambda\mu\nu} = 0 \,, \tag{6.103}$$

which implies that the flux of $M_{\rm II}^{\lambda\mu\nu}$ through the future light cone is zero, while

$$c_\nu M_{\rm I}^{\lambda\mu\nu} = \frac{e^2}{8\pi\rho^4}\,(z^\lambda c^\mu - z^\mu c^\lambda) \,. \tag{6.104}$$

Prove that both $M_{\rm I}^{\lambda\mu\nu}$ and $M_{\rm II}^{\lambda\mu\nu}$ are separately conserved off the world line,

$$\partial_\nu M_{\rm I}^{\lambda\mu\nu} = 0 \,, \tag{6.105}$$

$$\partial_\nu M_{\rm II}^{\lambda\mu\nu} = 0 \,. \tag{6.106}$$

Hint One readily obtains

$$\partial_\nu \vartheta^{\lambda\nu} = \frac{e^2}{2\pi\rho^4}\,(a\cdot c)\,c^\lambda \,, \tag{6.107}$$

whence it follows

$$\partial_\nu \left(R^\lambda \vartheta^{\mu\nu} - R^\mu \vartheta^{\lambda\nu}\right) = 0 \,. \tag{6.108}$$

Observing that

$$\partial_\nu \left(z^\lambda \Theta_{\rm II}^{\mu\nu} - z^\mu \Theta_{\rm II}^{\lambda\nu}\right) = 0 \,, \tag{6.109}$$

one arrives at (6.106). Combining (6.106) with (6.97), one gets (6.105).

Problem 6.2.8. Let a particle be coupled to a massless scalar field Φ. Consider a closed system governed by the action

$$S = -m_0 \int ds\,\sqrt{v\cdot v} - g\int ds\,\Phi(z)\,\sqrt{v\cdot v} + \frac{1}{8\pi}\int d^4x\,\partial_\mu\Phi\,\partial^\mu\Phi\,. \tag{6.110}$$

The Euler–Lagrange equation for Φ reads:

$$\Box\,\Phi = -4\pi g\int_{-\infty}^{\infty} ds\,\sqrt{v(s)\cdot v(s)}\,\delta^4\left[x - z(s)\right] \tag{6.111}$$

A retarded solution to this equation is $\Phi = -g/\rho$. Evaluate $\Theta^{\mu\nu}_{\rm II}$ for this field.

Answer

$$\Theta^{\mu\nu}_{\rm II} = \frac{g^2}{4\pi}\,\frac{(a\cdot c)^2}{\rho^2}\,c^\mu c^\nu\,. \tag{6.112}$$

Problem 6.2.9. *Radiation in a world with one temporal and one spatial dimension.* Using results of Problem 4.7.2, show that the stress-energy tensor

$$\Theta_{\mu\nu} = \frac{1}{\Omega_{D-2}}\left(F_{\mu\alpha}F^{\alpha}_{\ \nu} + \frac{\eta_{\mu\nu}}{4}\,F_{\alpha\beta}F^{\alpha\beta}\right) \tag{6.113}$$

can be expressed in terms of the retarded field as

$$\Theta_{\mu\nu} = \frac{1}{4}\,e^2\left(c_\mu v_\nu + c_\nu v_\mu - c_\mu c_\nu\right)\,. \tag{6.114}$$

Is it correct to interpret $-\frac{1}{4}e^2 c_\mu c_\nu$ as radiation?

6.3 Energy-Momentum Balance

We return now to the Noether identity (6.75). Assume that $\mathcal{E}^{\lambda\mu\nu} = 0$ but $\mathcal{E}_\mu$ is nonzero. Then $F_{\mu\nu}$ may be regarded as a regular field vanishing sufficiently fast at spatial infinity. Substitute (6.76)–(6.78) into (6.75), and integrate this equation over a domain of spacetime bounded by two parallel spacelike hyperplanes Σ' and Σ'' with both normals directed towards the future, and a tube T_R of large radius R. Applying the Gauss–Ostrogradskiĭ theorem, we obtain

$$\left(\int_{\Sigma''} - \int_{\Sigma'} + \int_{T_R}\right) d\sigma_\mu\,\Theta^{\lambda\mu} + m_0\int_{s'}^{s''} ds\,a^\lambda = \int_{s'}^{s''} ds\,\varepsilon^\lambda + \frac{1}{4\pi}\int_{\mathcal{U}} d^4x\,\mathcal{E}_\mu\,F^{\lambda\mu}\,. \tag{6.115}$$

Here, the relation

$$v^\mu\,\frac{\partial}{\partial x^\mu}\,\delta^4\left[x - z(s)\right] = -\frac{dz^\mu}{ds}\,\frac{\partial}{\partial z^\mu}\,\delta^4\left[x - z(s)\right] = -\frac{d}{ds}\,\delta^4\left[x - z(s)\right] \tag{6.116}$$

has been used to evaluate the integral of $\partial_\mu t^{\lambda\mu}$.

However, our concern here is with the case $\mathcal{E}_\mu = 0$. If $F_{\mu\nu}$ is the Liénard–Wiechert field, then (6.115) is divergent. To proceed further with this equation, a *regularization* is essential. The singularity must be smeared out over a region bounded by a tube T_ϵ of small radius ϵ enclosing the world line. In addition, we should assume that the mechanical mass is a function of regularization: $m_0 = m_0(\epsilon)$.

Note that the regularization scheme may be arbitrary. The only requirement is that it respect the symmetries of the action (6.67).

In this section, we employ a regularization known as a *cutoff*. The cutoff prescription is to put $F_{\mu\nu} = 0$ within a tube enclosing the world line. To be more specific, we define $\mathrm{Reg}\,\varepsilon^\lambda$ by an appropriate regularization of the left-hand side of equation (6.115):

$$\left(\int_{\Sigma''(\epsilon)} - \int_{\Sigma'(\epsilon)} + \int_{T_R}\right) d\sigma_\mu\, \Theta^{\lambda\mu} + m_0(\epsilon)\int_{s'}^{s''} ds\, a^\lambda = \int_{s'}^{s''} ds\, \mathrm{Reg}\,\varepsilon^\lambda\ . \tag{6.117}$$

The cutoff prescription is achieved by perforating the hyperplanes Σ' and Σ''. Here, $\Sigma'(\epsilon)$ and $\Sigma''(\epsilon)$ are perforated hyperplanes with holes of radius ϵ around the points of intersection with the world line. The function $m_0(\epsilon)$ is chosen to make the left-hand side of (6.117) convergent in the limit $\epsilon \to 0$.

We now assume that

$$\mathrm{Reg}\,\varepsilon^\lambda = 0\ . \tag{6.118}$$

This gives

$$\lim_{\epsilon\to 0}\left[m_0(\epsilon)\, v^\lambda(s)\Big|_{s'}^{s''} + \left(\int_{\Sigma''(\epsilon)} - \int_{\Sigma'(\epsilon)} + \int_{T_R}\right) d\sigma_\mu\, \Theta^{\lambda\mu}\right] = 0\ . \tag{6.119}$$

To ensure Lorentz invariance of this cutoff procedure, we take a hyperplane Σ whose normal is directed along the world line. Suppose that a world line $z^\mu(s)$ is intersected by such a hyperplane Σ at an instant s. We define the lagging instant $\hat{s} = s - \epsilon$ with an infinitesimal time lag ϵ, and draw the future light cone C_+ from $z^\mu(\hat{s})$. We then delete the set of all points on Σ bounded by the intersection of Σ and C_+. This gives an invariant (coordinate free) hole on Σ, and renders Σ the desired perforated hyperplane $\Sigma(\epsilon)$.

The coordinates x^μ of the hole are determined from the system of equations

$$\begin{aligned} x^\mu - \hat{z}^\mu &= \rho\, \hat{c}^\mu\ , \\ v\cdot\big(x - z\big) &= 0\ , \end{aligned} \tag{6.120}$$

which describe, respectively, the future light cone C_+ drawn from $\hat{z}^\mu$, and the hyperplane Σ which is intersected along its normal by the world line at z^μ. (We denote vectors at the lagging instant $\hat{s}$ by symbols with carets.) It follows that

$$\rho = \frac{\ell}{v\cdot\hat{c}}\ , \tag{6.121}$$

where ℓ is an invariant quantity having dimension of length,

$$\ell = v \cdot (z - \hat{z}) \ . \tag{6.122}$$

Expanding $\hat{z}^\mu$ in powers of ϵ,

$$\hat{z}^\mu = z^\mu - \epsilon\, v^\mu + \frac{\epsilon^2}{2}\, a^\mu - \frac{\epsilon^3}{6}\, \dot{a}^\mu + \cdots \ , \tag{6.123}$$

and combining with

$$v^2 = 1, \quad v \cdot a = 0, \quad v \cdot \dot{a} = -a^2 \ , \tag{6.124}$$

we have

$$\ell = \epsilon + O(\epsilon^3) \ . \tag{6.125}$$

Consider the regularized four-momentum of the electromagnetic field

$$P^\lambda = \int_{\Sigma(\epsilon)} d\sigma_\mu\, \Theta^{\lambda\mu} \ . \tag{6.126}$$

The evaluation of P^λ would be simplified if the hyperplane Σ could be replaced by the surface formed by the future light cone C_+ drawn from the world line, and by a tube T_ρ which is separated from the world line by a fixed retarded distance $\rho =$ const. A close look at Fig. 6.3 shows that this is indeed possible. Let $\mathcal{V}$ be a region bounded by a perforated spacelike hyperplane $\Sigma(\epsilon)$, a truncated light cone $C_+(\epsilon)$ with vertex at $\hat{z}^\mu$ and truncation surface defined in (6.120), and a tube T_R of large radius R enveloping the world line. Since the region $\mathcal{V}$ is free of sources, we have $\partial_\mu \Theta^{\lambda\mu} = 0$, and

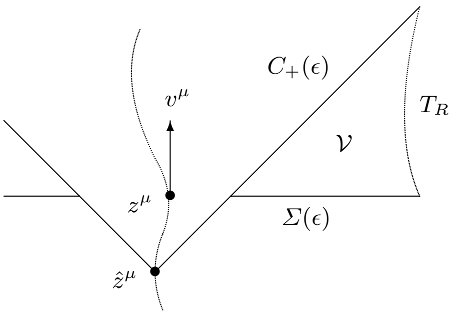


Fig. 6.3. Integration over a perforated hyperplane $\Sigma(\epsilon)$ can be changed for that over a truncated light cone $C_+(\epsilon)$ and a tube T_R

$$\int_{\mathcal{V}} d^4x\, \partial_\mu \Theta^{\lambda\mu} = \left(\int_{C_+(\epsilon)} + \int_{T_R} - \int_{\Sigma(\epsilon)} \right) d\sigma_\mu\, \Theta^{\lambda\mu} = 0\,. \tag{6.127}$$

The minus sign of the last integral signifies that normal of $\Sigma(\epsilon)$ is v^μ which is directed into the region $\mathcal{V}$. We see that the flux of $\Theta^{\lambda\mu}$ flowing into $\mathcal{V}$ through $\Sigma(\epsilon)$ equals the sum of fluxes flowing outward $\mathcal{V}$ through $C_+(\epsilon)$ and T_R.

A remarkable fact is that the integral over $C_+(\epsilon)$ is completely due to the contribution of $\Theta^{\lambda\mu}_{\mathrm{I}}$. By (6.88), $\Theta^{\lambda\mu}_{\mathrm{II}}$ yields zero flux through the future light cone. Using (6.84) we get

$$\int_{C_+(\epsilon)} d\sigma_\mu\, \Theta^{\lambda\mu} = \int_{C_+(\epsilon)} d\sigma_\mu\, \Theta^{\lambda\mu}_{\mathrm{I}} = \frac{e^2}{8\pi} \int d\Omega\, \hat{c}^\lambda \int_\rho^\infty \frac{d\varrho}{\varrho^2}\,. \tag{6.128}$$

The lower limit of the last integral is the retarded radius ρ of the truncation surface which is determined from (6.121) and (6.122). The rest of integration is done with the aid of (4.250)–(4.252). The result is

$$\frac{e^2}{8\pi\ell} \int d\Omega\, \hat{c}^\lambda\, (\hat{c}\cdot v) = \frac{e^2}{6\ell} \left[4\, \hat{v}^\lambda\, (\hat{v}\cdot v) - v^\lambda \right]\,. \tag{6.129}$$

Taking into account the expansion of $\hat{v}^\mu$ in powers of ϵ

$$\hat{v}^\mu = v^\mu - \epsilon\, a^\mu + \frac{\epsilon^2}{2}\, \dot{a}^\mu - \ldots \tag{6.130}$$

together with (6.124) and (6.125), we obtain

$$\int_{C_+(\epsilon)} d\sigma_\mu\, \Theta^{\lambda\mu} = \frac{e^2}{2\epsilon}\, v^\lambda - \frac{2}{3}\, e^2 a^\lambda + O(\epsilon)\,. \tag{6.131}$$

We see that the four-momentum associated with $\Theta^{\mu\nu}_{\mathrm{I}}$ can be recast in the form

$$P^\mu_{\mathrm{I}} = \int_{C_+(\epsilon)} d\sigma_\alpha\, \Theta^{\mu\alpha}_{\mathrm{I}}\,, \tag{6.132}$$

which is equal to that evaluated using the general definition (6.126). Teitelboim proposed to identify P^μ_{I} as *bound* four-momentum. By (6.131),

$$P^\mu_{\mathrm{I}} = \frac{e^2}{2\epsilon}\, v^\mu - \frac{2}{3}\, e^2 a^\mu\,. \tag{6.133}$$

The divergent factor

$$\delta m = \frac{e^2}{2\epsilon} \tag{6.134}$$

is called the *self-energy*. Equation (6.133) suggests that δm is a mass of electromagnetic origin. If the charge is at rest, then δm is energy of the Coulomb field (Problem 6.3.1).

A crucial step is to add up the mechanical mass m_0 and the self-energy δm, and assume that

$$m = \lim_{\epsilon \to 0} \left[m_0(\epsilon) + \delta m \right] \tag{6.135}$$

is finite and positive. This procedure is known as *mass renormalization*. The constant m defined in (6.135) is called the *renormalized mass*.

One can assemble appropriate terms in the square brackets of (6.119) to give

$$p^\mu = m v^\mu - \frac{2}{3}\, e^2 a^\mu \,. \tag{6.136}$$

This four-momentum is attributed to the *dressed particle*.

The remaining integrals in (6.119) are convergent, so that regularization is unnecessary. Let us evaluate the flux of $\Theta^{\mu\nu}_{\rm II}$ through a spacelike hyperplane Σ,

$$\mathcal{P}^\mu = \int_\Sigma d\sigma_\alpha\, \Theta^{\mu\alpha}_{\rm II} \,. \tag{6.137}$$

It is convenient to deform the surface of integration from Σ to the more geometrically motivated surface formed by combining the future light cone with a tubular hypersurface T_ρ enclosing the world line. The area of T_ρ scales as ρ^2. On the other hand, $\Theta^{\mu\nu}_{\rm II}$ behaves as ρ^{-2}. Turning back to Fig. 6.3, we see that only this part of the stress-energy tensor makes a nonvanishing contribution to the flux through T_R for large R, hence

$$\int_\Sigma d\sigma_\alpha\, \Theta^{\mu\alpha}_{\rm II} = \lim_{R\to\infty} \int_{T_R} d\sigma_\alpha\, \Theta^{\mu\alpha} \,. \tag{6.138}$$

Substitution of (4.249) and (6.82) in (6.138) gives

$$\mathcal{P}^\mu = \int_{T_R} d\sigma_\alpha\, \Theta^{\mu\alpha}_{\rm II} = -\frac{e^2}{4\pi} \int_{-\infty}^{s} d\tau \int d\Omega \left[a^2 + (a\cdot u)^2 \right] c^\mu \,. \tag{6.139}$$

The solid angle integration is made with the help of (4.250)–(4.252). The result is

$$\mathcal{P}^\mu = -\frac{2}{3}\, e^2 \int_{-\infty}^{s} d\tau\, a^2 v^\mu \,. \tag{6.140}$$

For this expression to be convergent, the integrand must fall off sufficiently rapidly as $\tau \to -\infty$. The pertinent asymptotic condition, proposed by Rudolf Haag, reads: every charge must move uniformly in the remote past,

$$\lim_{s\to -\infty} a^\mu(s) = 0 \,. \tag{6.141}$$

Differentiating (6.140) with respect to s, we obtain the four-momentum emitted by an accelerated charge per unit proper time:

$$\dot{\mathcal{P}}^\mu = -\frac{2}{3}\, e^2 a^2 v^\mu \,. \tag{6.142}$$

In a particular Lorentz frame, where $\mathcal{P}^\mu = (\mathcal{E}, \mathbf{P})$, $v^\mu = \gamma\,(1, \mathbf{v})$, and $dt = \gamma\, ds$, the rate of radiated energy is

$$\frac{d\mathcal{E}}{dt} = v \cdot \dot{\mathcal{P}} = -\frac{2}{3}\, e^2 a^2 \ . \tag{6.143}$$

This is the relativistic generalization of the *Larmor formula*

$$\frac{d\mathcal{E}}{dt} = \frac{2}{3}\, e^2\, \mathbf{a}^2 \ , \tag{6.144}$$

which gives the rate of radiated energy in an instantaneously comoving Lorentz frame.

Let us extend this analysis to a system of N charged particles governed by the action

$$S = -\sum_{I=1}^{N} m_0^I \int d\tau_I \sqrt{\dot{z}_I \cdot \dot{z}_I} - \int d^4x \left(A_\mu j^\mu + \frac{1}{16\pi}\, F_{\mu\nu} F^{\mu\nu} \right) \ . \tag{6.145}$$

We now have the retarded solution to Maxwell's equations

$$F^{\mu\nu} = \sum_{I=1}^{N} \frac{e_I}{\rho_I^2}\, (c_I^\mu\, V_I^\nu - c_I^\nu\, V_I^\mu) \ , \tag{6.146}$$

which is the sum of one-particle Liénard–Wiechert fields (6.73). For simplicity, we omit solutions to the homogeneous field equations describing a free electromagnetic field $F_0^{\mu\nu}$. If need be, this field could be taken into account in the final result as a readily calculable correction.

The stress-energy tensor becomes

$$\Theta^{\mu\nu} = \sum_I \Theta_I^{\mu\nu} + \sum_I \sum_J \Theta_{IJ}^{\mu\nu} \ , \tag{6.147}$$

where $\Theta_I^{\mu\nu}$ is comprised of the field $F_I^{\mu\nu}$ due to the Ith charge, and $\Theta_{IJ}^{\mu\nu}$ contains mixed contributions of the fields $F_I^{\mu\nu}$ and $F_J^{\mu\nu}$ generated by the Ith and the Jth charges.

One can show (Problem 6.3.2) that

$$\partial_\mu \Theta_{IJ}^{\mu\nu} = 0 \ . \tag{6.148}$$

Hence, a N-particle generalization of (6.119) reads

$$\lim_{\epsilon \to 0} \left[\sum_{I=1}^{N} m_0^I(\epsilon)\, v_I^\mu(s_I) \Big|_{s'_I}^{s''_I} + \left(\int_{\Sigma''(\epsilon)} - \int_{\Sigma'(\epsilon)} + \int_{T_R} \right) d\sigma_\alpha\, \Theta^{\mu\alpha} \right] = 0 \ , \tag{6.149}$$

where $\Theta^{\mu\nu}$ is given by (6.147).

Two modifications of the single particle procedure are required. First, the invariant cutoff prescription must be extended to the case of N singular world lines. Second, the mixed terms of the stress-energy tensor $\Theta^{\mu\nu}_{IJ}$ should be appropriately integrated.

Suppose that the world lines are all smooth timelike curves. Consider an integration domain bounded by two spacelike hypersurfaces Σ' and Σ''. Because Σ' and Σ'' are arbitrary, we may require that each world line intersect Σ' and Σ'' at right angles. Such hypersurfaces will be called *locally adjusted.* Let a smooth locally adjusted hypersurface Σ be intersected by a world line at a point z^{μ}_{I}. In the vicinity of z^{μ}_{I}, this hypersurface is approximated by a tangent hyperplane whose normal is directed along the world line. We now repeat the steps by which the perforation proceeds to conclude that the resulting small hole in $\Sigma(\epsilon)$ is an invariant construction.

To integrate the bound part of $\Theta^{\mu\nu}_{I}$, we replace the perforated hypersurface $\Sigma(\epsilon)$ by a truncated light cone $C_{+}(\epsilon)$. This gives an expression for the bound four-momentum of the type (6.133). As a consequence, N dressed particles arise, each possessing the four-momentum

$$p^{\mu}_{I} = m_I\, v^{\mu}_{I} - \frac{2}{3}\, e^2_I\, a^{\mu}_{I}\ . \tag{6.150}$$

When the Ith charge is accelerated, it radiates the electromagnetic four-momentum

$$\mathcal{P}^{\mu}_{I} = -\frac{2}{3}\, e^2_I \int_{-\infty}^{s_I} d\tau_I\, a^2_I\, v^{\mu}_{I}\ . \tag{6.151}$$

To derive this result, we need no regularization as before. A tube enclosing the Ith world line is the most appropriate integration surface.

We now turn to a mixed term $\Theta^{\mu\nu}_{IJ}$. Evidently this term is singular on both Ith and Jth world lines. Because the leading singularity is a pole ρ^{-2}, any $\Theta^{\mu\nu}_{IJ}$ is integrable. We therefore may substitute $\Sigma(\epsilon)$ by two tubes $T_I(\epsilon)$ and $T_J(\epsilon)$ of infinitesimal radius ϵ enclosing these world lines (one of these tubes is depicted in Fig. 6.4) to evaluate the four-momentum $\wp^{\mu}$ of the mixed terms. One can show (Problem 6.3.3) that

$$\wp^{\mu}_{I} = \int_{T_I(\epsilon)} d\sigma_{\alpha} \sum_{J} \Theta^{\mu\alpha}_{IJ} = -e_I \int_{-\infty}^{s_I} d\tau_I \sum_{J} F^{\mu\nu}_{IJ}(z_I)\, v^{I}_{\nu}(\tau_I)\ , \tag{6.152}$$

where $F^{\mu\nu}_{IJ}(z_I)$ is the Liénard–Wiechert field (6.146) due to the Jth charge, which is taken at the point z_I where the Ith charge is located. To interpret $\wp^{\mu}_{I}$, one observe that expression (6.152) represents the four-momentum extracted from an *external* field $F^{\mu\nu}_{IJ}(z_I)$ during the whole past history of the Ith charge prior to the instant s_I. To put it differently, $\wp^{\mu}_{I}$ is the four-momentum produced by an *external* Lorentz force

$$f^{\mu}_{I}(z_I) = e_I \sum_{J} \frac{e_J}{[v_J\cdot(z_J - z_I)]^2}\left[(V_J\cdot v_I)\, c^{\mu}_{J} - (c_J\cdot v_I)\, V^{\mu}_{J}\right] \tag{6.153}$$

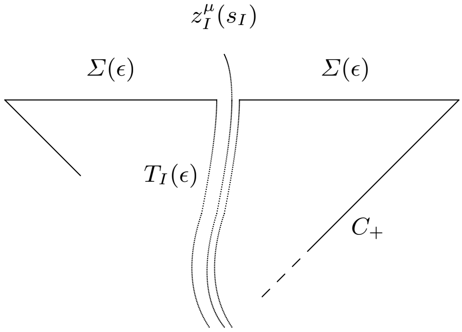


Fig. 6.4. A tube $T_I(\epsilon)$ which can be used as an integration surface

during that half-infinite period,

$$\wp_I^\mu = -\int_{-\infty}^{s_I} d\tau_I \, f_I^\mu(z_I) \, . \tag{6.154}$$

We assume that world lines are nonintersecting, and hence that the external Lorentz force $f_I^\lambda(z_I)$ defined in (6.153) is regular.

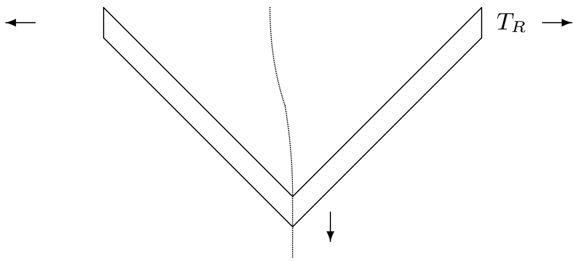


Fig. 6.5. Contribution to the flux through a tube T_R of large radius R is due to the affair in the remote past

Let us return to (6.149). If we impose the asymptotic condition (6.141), then the integral over T_R approaches zero as $R \to \infty$. Indeed, as Fig. 6.5 suggests, the flux of the radiation through T_R vanishes because world lines become straight in the remote past ($a^\mu \to 0$), while other parts of the stress-energy tensor fall more rapidly than R^{-2}, and hence their contribution to this flux disappear. Therefore, (6.149) becomes

$$\left[m_I \, v_I^\mu(s) - \frac{2}{3} \, e_I^2 \, a_I^\mu(s) \right] \Bigg|_{s'_I}^{s''_I} - \frac{2}{3} \, e_I^2 \int_{s'_I}^{s''_I} ds_I \, a_I^2 \, v_I^\mu - \int_{s'_I}^{s''_I} ds_I \, f_I^\mu = 0 \, , \tag{6.155}$$

where expressions (6.150), (6.151), and (6.154) are used, terms of identical structure are collected, and each term of the sum over I is equated to zero.

Let $s''_I = s'_I + \Delta s$ where Δs is a short period, then (6.155) becomes

$$\Delta p^\mu_I + \Delta \mathcal{P}^\mu_I = f^\mu_I \Delta s \,. \tag{6.156}$$

This is the desired *energy-momentum balance*: the four-momentum $\Delta\wp^\mu = -f^\mu \Delta s$ which is extracted from an external field during the period of time Δs is distributed between the four-momentum of a dressed particle Δp^μ and the four-momentum carried away by radiation $\Delta \mathcal{P}^\mu$.

Problem 6.3.1. Evaluate the self-energy δm of a static point charge e.

Answer

$$\delta m = \frac{1}{8\pi}\int d^3x\,\mathbf{E}^2 = \frac{1}{8\pi}\int d^3x\,\frac{e^2}{r^4} = \frac{e^2}{2}\lim_{\epsilon\to 0}\int_\epsilon^\infty \frac{dr}{r^2} = \lim_{\epsilon\to 0}\frac{e^2}{2\epsilon}\,. \tag{6.157}$$

Problem 6.3.2. Prove (6.148).

Problem 6.3.3. Prove (6.152).

Hint Use expressions (4.249) and (4.247) for the surface elements of a tube T_ρ and the future light cone C_+. Assume that world lines of the Ith and Jth particles are widely separated in the remote past. This assumption implies the vanishing of the surface integral over the future light cone C_+ drawn from a worldline point in the remote past, displayed in Fig. 6.4.

Problem 6.3.4. Evaluate the emitted four-momentum $\mathcal{P}^\mu$ for a scalar field discussed in Problem 6.2.8.

Answer

$$\mathcal{P}^\mu = -\frac{g^2}{3}\int_{-\infty}^s d\tau\, a^2 v^\mu \,. \tag{6.158}$$

Problem 6.3.5. Find the self-energy δm for a scalar field discussed in Problem 6.2.8.

Answer

$$\delta m = -\frac{g^2}{2\epsilon}\,. \tag{6.159}$$

Note that $\delta m < 0$.

Problem 6.3.6. Derive the four-momentum of a dressed particle interacting with a scalar field discussed in Problem 6.2.8.

Answer

$$p^\mu = m\,v^\mu - \frac{1}{3}\,g^2 a^\mu \,. \tag{6.160}$$

6.4 The Lorentz–Dirac Equation

Let us write (6.155) in differential form

$$ma^\mu - \frac{2}{3}\,e^2\left(\dot a^\mu + v^\mu a^2\right) = f^\mu\ . \qquad (6.161)$$

This third-order differential equation for z^μ is called the *Lorentz–Dirac equation.*

In view of identities

$$v^2 = 1, \quad v\cdot a = 0, \quad v\cdot \dot a = -a^2\ , \qquad (6.162)$$

equation (6.161) can be brought to the form

$$\overset{v}{\perp}(\dot p - f) = 0\ , \qquad (6.163)$$

where $\overset{v}{\perp}$ is the projection operator on a hyperplane with normal v^μ, p^μ the four-momentum of a dressed particle defined in (6.136), and f^μ an external four-force.

It was established in Sect. 2.1 that (6.163) is Newton's second law embedded in Minkowski space. We see that a dressed particle is an object with four-momentum p^μ defined in (6.136), whose behavior is governed by Newton's second law. The structure of (6.163) makes it clear that a dressed particle experiences only an external force f^μ. This equation contains no term through which the dressed particle interacts with itself.

The state of a dressed particle is determined by its position in Minkowski space z^μ, and by its four-momentum

$$p^\mu = m\left(v^\mu - \tau_0\, a^\mu\right)\ . \qquad (6.164)$$

Here, τ_0 stands for a characteristic time interval

$$\tau_0 = \frac{2}{3}\frac{e^2}{m}\ , \qquad (6.165)$$

which is $\tau_0 \approx 6\cdot 10^{-24}$ s if we choose e and m to be the charge and mass of a real electron.

We point out that the energy of a dressed particle p^0 is *indefinite* quantity. Indeed, in a particular Lorentz frame,

$$p^0 = m\gamma\left[1 - \tau_0\,\gamma^3\,(\mathbf{a}\cdot\mathbf{v})\right]\ . \qquad (6.166)$$

The fact that p^0 is not positive definite is scarcely surprising. Recall that p^μ is the sum of two vectors $p^\mu = m_0 v^\mu + P^\mu_{\rm I}$. The bound four-momentum $P^\mu_{\rm I}$ is a timelike future-directed vector, while the four-momentum of a bare particle $m_0 v^\mu$ is a timelike past-directed vector. This is because $m_0(\epsilon) < 0$ for small

ϵ, as (6.135) suggests. Assuming that $m_0 v^\mu + P_{\rm I}^\mu$ is a timelike vector, one recognizes that the time component of this vector can have any sign.

On the other hand, equation (6.161) merely expresses local energy-momentum balance:

$$\dot{p}^\mu + \dot{\mathcal{P}}^\mu = f^\mu \ . \tag{6.167}$$

In a particular Lorentz frame, we write $p^\mu = (p^0, \mathbf{p})$, $\mathcal{P}^\mu = (\mathcal{E}, \mathbf{P})$, $f^\mu = \gamma\,(\mathbf{F}\cdot\mathbf{v}, \mathbf{F})$, $\gamma\, ds = dt$, and $\mathbf{v}\, dt = d\mathbf{z}$. Then the time component of (6.167) becomes

$$dp^0 + d\mathcal{E} = \mathbf{F}\cdot d\mathbf{z} \ , \tag{6.168}$$

which expresses local energy balance. This equation tells us that the rate of work done by an external force is equal to the rate of change in dressed particle energy plus the rate of radiated energy. Recall that the dressed particle energy p^0 is indefinite, and hence an increase in velocity does not necessarily increase p^0.

The four-momentum of a free dressed particle p^μ need not be constant. This fact is consistent with translation invariance in the absence of external forces. Translation invariance implies that $p^\mu + \mathcal{P}^\mu = \text{const}$, rather than that $p^\mu = \text{const}$, which is immediately evident from (6.167) in the case that $f^\mu = 0$.

The rest mass

$$m = p\cdot v \tag{6.169}$$

characterizes the properties of a dressed particle at rest. This is a Lorentz invariant conserved quantity equal to the renormalized mass. By contrast, the mass M defined by

$$M^2 = p^2 \tag{6.170}$$

is not necessarily constant. Indeed, by (6.164),

$$M^2 = m^2\left(1 + \tau_0^2\, a^2\right) \ . \tag{6.171}$$

We see that M depends on acceleration. M and m agree when the world line of the dressed particle is straight.

Suppose that the acceleration of a dressed particle exceeds the critical value,

$$a^2\tau_0^2 = -1 \ . \tag{6.172}$$

Then the dressed particle becomes a *tachyon*, that is, an object whose four-momentum is spacelike, $p^2 < 0$. It should be kept in mind that this does not imply superluminal motion. The potentially tachyonic nature of a dressed particle results from the fact that the curvature of its world line can be excessively high.

The Lorentz–Dirac equation is not invariant under time reversal $s \to -s$. Indeed, this equation involves both a^μ, whose transformation law is $a^\mu \to a^\mu$, and $\dot{a}^\mu$, which transforms according to $\dot{a}^\mu \to -\dot{a}^\mu$. Hence, the dynamics of dressed particles is *irreversible*.

In the limit $\mathbf{v} \to \mathbf{0}$, only the space component of the Lorentz–Dirac equation survives,

$$m\dot{\mathbf{v}} - \frac{2}{3}\, e^2\, \ddot{\mathbf{v}} = \mathbf{f}\ , \tag{6.173}$$

where the dots indicate derivatives with respect to t. This nonrelativistic relation is known as the *Abraham–Lorentz* equation. One can write (6.173) in the form

$$\dot{\mathbf{p}} = \mathbf{f} \tag{6.174}$$

where

$$\mathbf{p} = m\mathbf{v} - \frac{2}{3}\, e^2\, \mathbf{a}\ . \tag{6.175}$$

While the Abraham–Lorentz equation (6.173) is linear, the Lorentz–Dirac equation (6.161) is nonlinear. This poses severe problems for obtaining exact solutions to (6.161). However, this equation becomes linear for straight line motion. In that case,

$$v^\mu = (\cosh\alpha, \mathbf{n}\sinh\alpha)\ , \tag{6.176}$$

where α is an unknown function of the proper time s, and $\mathbf{n}$ is a unit vector along the direction of motion. Differentiation of (6.176) gives

$$a^\mu = \dot{\alpha}\,(\sinh\alpha, \mathbf{n}\cosh\alpha)\,, \quad a^2 = -\dot{\alpha}^2\ , \tag{6.177}$$

and

$$\dot{a}^\mu = \ddot{\alpha}\,(\sinh\alpha, \mathbf{n}\cosh\alpha) + \dot{\alpha}^2\,(\cosh\alpha, \mathbf{n}\sinh\alpha)\ . \tag{6.178}$$

We then assume that

$$f^\mu = f\,(\sinh\alpha, \mathbf{n}\cosh\alpha)\ , \tag{6.179}$$

where f is a scalar function of s, z^μ, and v^μ. This four-force has the required direction and is orthogonal to the four-velocity, $f \cdot v = 0$.

Entering (6.176)–(6.179) back into (6.161), we obtain a linear equation

$$m\,\dot{\alpha} - \frac{2}{3}\, e^2\, \ddot{\alpha} = f\ . \tag{6.180}$$

This equation bears a formal similarity to the Abraham–Lorentz equation, even though it is fully relativistic.

Equation (6.180) can be integrated once to give

$$\dot{\alpha}(s) = e^{s/\tau_0}\left[C - \frac{1}{m\tau_0}\int_0^s d\sigma\, e^{-\sigma/\tau_0} f(\sigma)\right]\ . \tag{6.181}$$

If $\dot{\alpha}(s)$ is to remain finite as $s \to \infty$, one takes

$$C = \frac{1}{m\tau_0}\int_0^\infty d\sigma\, e^{-\sigma/\tau_0} f(\sigma)\ , \tag{6.182}$$

and so

$$\dot{\alpha}(s) = \frac{1}{m\tau_0} \int_0^\infty d\sigma \, e^{-\sigma/\tau_0} f(s+\sigma) \, . \tag{6.183}$$

Assuming that f vanishes as $s \to \infty$, we have $\dot{\alpha}(\infty) = 0$, and

$$\lim_{s\to\infty} a^\mu(s) = 0 \, . \tag{6.184}$$

This asymptotic condition is inherent in scattering problems.

If f is a known function of s, then (6.183) can be integrated again, and the result is a solution to the one-dimensional Lorentz–Dirac equation with this external force.

Similar arguments can be applied to the Abraham–Lorentz equation (6.173) to obtain

$$m\,\mathbf{a}(t) = \int_0^\infty du \, e^{-u} \, \mathbf{f}(t + \tau_0 u) \tag{6.185}$$

and

$$\lim_{t\to\infty} \mathbf{a}(t) = 0 \, . \tag{6.186}$$

If $\mathbf{f}$ is a function of $\mathbf{z}$ and $\mathbf{v}$, then (6.185) is an integro-differential equation stemming from the Abraham–Lorentz equation and the asymptotic condition (6.186).

Problem 6.4.1. Using the results of Problem 1.6.2, find time and space components of the second term in (6.161).

Answer

$$-\frac{2}{3}\, e^2 \gamma^5 \left[3\gamma^2 (\mathbf{v}\cdot\mathbf{a})^2 + \mathbf{v}\cdot\dot{\mathbf{a}}\right] \, ,$$

$$-\frac{2}{3}\, e^2 \gamma^3 \left\{\dot{\mathbf{a}} + 3\gamma^2 \,(\mathbf{v}\cdot\mathbf{a}) \left[\mathbf{a} + \mathbf{v}\,(\mathbf{v}\cdot\mathbf{a})\gamma^2\right] + \mathbf{v}\,\gamma^2\,(\mathbf{v}\cdot\dot{\mathbf{a}})\right\} \, . \tag{6.187}$$

These expressions (apart from the factor of γ) were derived by Max Abraham in 1904.

Problem 6.4.2. Take the scalar product of (6.161) with a_μ. This gives

$$m\left[a^2 - \tau_0\,(a\cdot\dot{a})\right] = m\left(a^2 - \frac{\tau_0}{2}\frac{d}{ds}\,a^2\right) = f\cdot a \, . \tag{6.188}$$

With this in mind, convert the Lorentz–Dirac equation (6.161) to an integro-differential equation similar to (6.185).

Answer

$$m\,a^2(s) = \int_0^\infty d\sigma \, e^{-\sigma}\,(f\cdot a)\left(s + \frac{\tau_0}{2}\,\sigma\right) \, . \tag{6.189}$$

Problem 6.4.3. Let a dressed particle be moving in a constant homogeneous magnetic field $\mathbf{B}$. Find a damped solution to the Abraham–Lorentz equation

(6.173) in an inertial frame where $\mathbf{B}$ is parallel to the x^3-axis, and the initial four-velocity is $v^\mu = \gamma\,(1, V, 0, 0)$, $V \ll 1$.

Answer

$$v^1 = V \exp\left(-\nu t\right) \cos\left(\kappa t\right), \quad v^2 = V \exp\left(-\nu t\right) \sin\left(\kappa t\right) \ , \tag{6.190}$$

$$\nu = \frac{1}{2\tau_0}\left\{\left[\frac{1}{2} + \frac{1}{2}\left(1 + 16\tau_0^2\omega^2\right)^{\frac{1}{2}}\right]^{\frac{1}{2}} - 1\right\}, \ \kappa = \frac{1}{2\tau_0}\left[-\frac{1}{2} + \frac{1}{2}(1 + 16\tau_0^2\omega^2)^{\frac{1}{2}}\right]^{\frac{1}{2}} , \tag{6.191}$$

and $\omega = eB/m$. If ω is of the same order of magnitude as $1/\tau_0$, then the dressed particle will spiral in toward the center, and lose almost all its energy in several revolutions.

Problem 6.4.4. Let a nonrelativistic dressed particle be driven by a constant external force $\mathbf{f} = -m\,\omega_0^2\,\mathbf{x}$. Find a damped solution to the Abraham–Lorentz equation (6.173) in an inertial frame where $\mathbf{f}$ is parallel to the x^1-axis, and the initial conditions are $z^\mu = (0, 0, 0, 0)$ and $v^\mu = \gamma\,(1, V, 0, 0)$, $V \ll 1$.

Answer

$$z^1 = \frac{V}{\kappa} \exp\left(-\nu t\right) \sin\left(\kappa t\right) \ , \tag{6.192}$$

$$\nu = \frac{\omega_0^2}{2}\left(M + N\right)^2, \quad \kappa = \frac{\sqrt{3}\,\omega_0^2}{2\tau_0}\left(M^2 - N^2\right) \ , \tag{6.193}$$

$$M = \left\{\frac{\tau_0}{2\omega_0^2} + \left[\left(\frac{\tau_0}{2\omega_0^2}\right)^2 + \frac{1}{27\omega_0^6}\right]^{\frac{1}{2}}\right\}^{\frac{1}{3}}, \ N = \left\{\frac{\tau_0}{2\omega_0^2} - \left[\left(\frac{\tau_0}{2\omega_0^2}\right)^2 + \frac{1}{27\omega_0^6}\right]^{\frac{1}{2}}\right\}^{\frac{1}{3}} . \tag{6.194}$$

6.5 Alternative Methods of Deriving the Equation of Motion for a Dressed Charged Particle

There are several alternative methods of deriving the Lorentz–Dirac equation. We outline two of them below.

We first consider a regularization prescription of a singular retarded field through introducing advanced fields following Dirac's original 1938 approach.

Let us begin with the equation of motion for a bare charged particle

$$m_0 a^\mu = e v_\nu F^{\mu\nu}(z) \ , \tag{6.195}$$

where $F^{\mu\nu} = F^{\mu\nu}_{\text{ret}} + F^{\mu\nu}_{\text{ext}}$, $F^{\mu\nu}_{\text{ret}}$ is the retarded self-field due to the charge in question, and $F^{\mu\nu}_{\text{ext}}$ an external field created by other charges. We can write

$$A^\mu_{\text{ret}} = \frac{1}{2}\left(A^\mu_{\text{ret}} - A^\mu_{\text{adv}}\right) + \frac{1}{2}\left(A^\mu_{\text{ret}} + A^\mu_{\text{adv}}\right) \ , \tag{6.196}$$

where $A^{\mu}_{\rm adv}$ is the corresponding advanced vector potential. In symbolic form,

$$A_{\rm ret} = A_{(-)} + A_{(+)} \, . \tag{6.197}$$

Expressions (4.196) and (4.197) indicate that the retarded vector potential generated by a delta-function source behaves similar to the advanced vector potential in the vicinity of the source. Therefore, $A_{(-)}$ is less singular than $A_{\rm ret}$ and $A_{\rm adv}$, while $A_{(+)}$ shares the singular behavior of $A_{\rm ret}$ and $A_{\rm adv}$.

With reference to Problem 4.4.2, we write the regularized part of the vector potential

$$A^{\mu}_{(-)}(x) = 2\pi \int d^4y \, D(x-y) \, j^{\mu}(y) \, , \tag{6.198}$$

where $D(x)$ is the Green's function of the homogeneous wave equation, defined in (4.166),

$$2\pi D(x) = {\rm sgn}(x_0) \, \delta(x^2) \, . \tag{6.199}$$

Let $j^{\mu}(x)$ be the current of a single point charge (4.204), then

$$A^{\mu}_{(-)}(x) = 2\pi e \int_{-\infty}^{\infty} d\tau \, v^{\mu}(\tau) \, D[x - z(\tau)] \, . \tag{6.200}$$

Denoting $R^{\mu} = x^{\mu} - z^{\mu}(\tau)$, we evaluate the regularized part of the field strength:

$$F^{\mu\nu}_{(-)}(x) = 2\pi e \int_{-\infty}^{\infty} d\tau \left(\frac{dR^2}{d\tau} \right)^{-1} \frac{d}{d\tau} \, D(R) \left[v^{\nu}(\tau) \, \frac{\partial R^2}{\partial x_{\mu}} - v^{\mu}(\tau) \, \frac{\partial R^2}{\partial x_{\nu}} \right] \, . \tag{6.201}$$

Since

$$\frac{dR^2}{d\tau} = -2R \cdot v, \quad \frac{\partial R^2}{\partial x_{\nu}} = 2R^{\nu} \, , \tag{6.202}$$

we have

$$F^{\mu\nu}_{(-)}(x) = 2\pi e \int_{-\infty}^{\infty} d\tau \, D(R) \, \frac{d}{d\tau} \left(\frac{R^{\mu} v^{\nu} - R^{\nu} v^{\mu}}{R \cdot v} \right) \, . \tag{6.203}$$

Let the observation point x^{μ} be on the world line, $x^{\mu} = z^{\mu}(s)$. All other points on the world line are separated from x^{μ} by timelike intervals. Accordingly, the delta-function in (6.199) should be understood as the limit

$$\delta(R^2) = \lim_{\epsilon \to 0} \delta(R^2 - \epsilon^2) \, . \tag{6.204}$$

Besides, we can represent the argument of the signum function in (6.199) as $R_0 = R \cdot v$.

We now write $\tau = s + \sigma$, and consider the intergand for a small interval σ. Using the expansions

$$z^\mu(s+\sigma) = z^\mu + \sigma\, v^\mu + \frac{\sigma^2}{2}\, a^\mu + \frac{\sigma^3}{6}\, \dot a^\mu + \cdots\ , \tag{6.205}$$

$$v^\mu(s+\sigma) = v^\mu + \sigma\, a^\mu + \frac{\sigma^2}{2}\, \dot a^\mu + \ldots\ , \tag{6.206}$$

where the vectors on the right-hand side refer to the instant s, we find

$$R^\mu = z^\mu - z^\mu(s+\sigma) = -\sigma\left(v^\mu + \frac{\sigma}{2}\, a^\mu + \frac{\sigma^2}{6}\, \dot a^\mu\right) + \cdots \tag{6.207}$$

It follows that

$$R^\mu v^\nu(s+\sigma) - R^\nu v^\mu(s+\sigma) = \frac{\sigma^2}{2}\,(v^\nu a^\mu - v^\mu a^\nu) + \frac{\sigma^3}{3}\,(v^\nu \dot a^\mu - v^\mu \dot a^\nu) + \cdots \tag{6.208}$$

In view of identities (6.124),

$$R\cdot v(s+\sigma) = -\sigma + O(\sigma^3)\ . \tag{6.209}$$

Substituting (6.208) and (6.209) into (6.203) and taking into account that

$$\mathrm{sgn}(R\cdot v)\,\delta(R^2-\epsilon^2) = \mathrm{sgn}(-\sigma)\,\delta(\sigma^2-\epsilon^2) = -\frac{1}{2\epsilon}\,[\delta(\sigma-\epsilon) - \delta(\sigma+\epsilon)]\ , \tag{6.210}$$

we obtain

$$F^{\mu\nu}_{(-)}(z) = \frac{2}{3}\, e\,(\dot a^\mu v^\nu - \dot a^\nu v^\mu) + O(\epsilon)\ , \tag{6.211}$$

and

$$e v_\nu F^{\mu\nu}_{(-)}(z) = \frac{2}{3}\, e^2\left(\dot a^\mu + a^2 v^\mu\right) + O(\epsilon)\ . \tag{6.212}$$

The term

$$\Gamma^\mu = \frac{2}{3}\, e^2\left(\dot a^\mu + a^2 v^\mu\right)\ , \tag{6.213}$$

again emerges. This higher-derivative term is called the *Abraham term*. In the literature, Γ^μ is often interpreted as *radiation reaction*, that is, the finite effect of the retarded Liénard–Wiechert field upon its own source. But this interpretation is wrong. It was already mentioned in the previous section that a dressed particle is acted upon by only an external force, and is free from self-interaction. We will see in Sect. 9.3 that the concept of ‘radiation reaction’ causes much confusion in understanding the rearranged Maxwell–Lorentz theory.

We now look at the symmetric part of (6.197). The corresponding Green's function is

$$D_P(R) = \frac{1}{2}\,[D_{\mathrm{ret}}(R) + D_{\mathrm{adv}}(R)] = \delta(R^2)\ . \tag{6.214}$$

We regularize this expression as follows:

$$\delta(R^2-\epsilon^2) = \delta(\sigma^2-\epsilon^2) = \frac{1}{2\epsilon}\,[\delta(\sigma-\epsilon) + \delta(\sigma+\epsilon)]\ . \tag{6.215}$$

Applying this procedure to

$$F^{\mu\nu}_{(+)}(x) = e \int_{-\infty}^{\infty} d\tau\, \delta(R^2 - \epsilon^2)\, \frac{d}{d\tau} \left(\frac{R^\mu v^\nu - R^\nu v^\mu}{R \cdot v} \right) , \tag{6.216}$$

we get

$$F^{\mu\nu}_{(+)}(z) = \frac{e}{2\epsilon} \left(v^\mu a^\nu - v^\nu a^\mu \right) + O(\epsilon) . \tag{6.217}$$

Therefore,

$$e v_\nu F^{\mu\nu}_{(+)}(z) = -\frac{e^2}{2\epsilon}\, a^\mu + O(\epsilon) . \tag{6.218}$$

We substitute (6.212) and (6.218) in (6.195) and perform mass renormalization (6.135). This gives the Lorentz–Dirac equation (6.161).

We next turn to the method proposed by Azim Barut in 1974. The key idea of this method is that the retarded field $F_{\rm ret}$ can be regularized in the vicinity of the source using a kind of analytic continuation. To be more precise, we take the field as a function of two variables $F^{\mu\nu}(x; z(s))$ and continue it analytically from null intervals between the observation point x^μ and the retarded point $z^\mu(s)$ to timelike intervals which result from assigning $x^\mu = z^\mu(s+\epsilon)$ and keeping the second variable $z^\mu(s)$ fixed.

Write the analytically continued equation of motion for a bare charged particle as

$$m_0 a^\mu(s+\epsilon) = e v_\nu(s+\epsilon) F^{\mu\nu}(z(s+\epsilon); z(s)) + f^\mu . \tag{6.219}$$

Here, $F^{\mu\nu}(z(s+\epsilon); z(s))$ is the Liénard–Wiechert field due to the charge at $z^\mu(s)$,

$$F^{\mu\nu} = \frac{e}{\rho^3} \left(R^\mu V^\nu - R^\nu V^\mu \right) , \tag{6.220}$$

$$R^\mu = z^\mu(s+\epsilon) - z^\mu(s) , \tag{6.221}$$

V^μ is defined in (4.284), and f^μ is an external four-force. The vectors v^μ and a^μ, appearing in V^μ and ρ, are referred to the 'retarded' instant s. [Note that the vectors R^μ in (6.207) and (6.221) are of opposite signs. This is because x^μ is identified with different points: $z^\mu(s)$ in the former case, and $z^\mu(s+\epsilon)$ in the latter case.] Combining (6.205) with (6.221), we obtain

$$R^\mu = \epsilon \left(v^\mu + \frac{\epsilon}{2}\, a^\mu + \frac{\epsilon^2}{6}\, \dot a^\mu \right) + O(\epsilon^4) . \tag{6.222}$$

Take the scalar products of this expression with v_μ and a_μ:

$$v \cdot R = \rho = \epsilon \left(1 - \frac{\epsilon^2}{6}\, a^2 \right) + O(\epsilon^4) , \tag{6.223}$$

$$a \cdot R = \frac{\epsilon^2}{2}\, a^2 + O(\epsilon^3) . \tag{6.224}$$

It follows

$$\frac{1}{\rho^3}\,R^\mu = \frac{1}{\epsilon^2}\,v^\mu + \frac{1}{2\epsilon}\,a^\mu + \frac{1}{6}\,\dot{a}^\mu + \frac{1}{2}\,v^\mu a^2 + O(\epsilon)\;, \tag{6.225}$$

$$V^\mu = v^\mu + \rho\, a^\mu + \left(\frac{1}{\rho}\,R^\mu - v^\mu\right)(a\cdot R) = v^\mu + \epsilon\, a^\mu + O(\epsilon^3)\;. \tag{6.226}$$

Therefore,

$$\frac{1}{\rho^3}\,(R^\mu V^\nu - R^\nu V^\mu) = \frac{1}{2\epsilon}\,(v^\mu a^\nu - a^\mu v^\nu) + \frac{1}{6}\,(\dot{a}^\mu v^\nu - \dot{a}^\nu v^\mu) + O(\epsilon)\;. \tag{6.227}$$

We then evaluate (6.219) by multiplying (6.227) with $v_\nu + \epsilon\, a_\nu$,

$$m_0\, a^\mu(s+\epsilon) = e^2\left\{-\frac{1}{2\epsilon}\,a^\mu + \frac{1}{6}\,[\dot{a}^\mu - (\dot{a}\cdot v)\,v^\mu] + \frac{1}{2}\,v^\mu a^2\right\} + f^\mu\;. \tag{6.228}$$

Adding and subtracting $\frac{1}{2}\,\dot{a}^\mu$, and using the identity $\dot{a}\cdot v = -a^2$, we find that the expression in the brace is

$$e^2\left\{-\frac{1}{2\epsilon}\,(a^\mu + \epsilon\,\dot{a}^\mu) + \frac{2}{3}\,(\dot{a}^\mu + v^\mu a^2)\right\} = e^2\left[-\frac{1}{2\epsilon}\,a^\mu(s+\epsilon) + \frac{2}{3}\,(\dot{a}^\mu + v^\mu a^2)\right]\;, \tag{6.229}$$

where terms of order $O(\epsilon)$ are omitted. The first term in the square brackets can be combined with the term in the left-hand side of (6.228) to give

$$(m_0+\delta m)\,a^\mu(s+\epsilon) = (m_0+\delta m)\,a^\mu + O(\epsilon) = \frac{2}{3}\,e^2\,(\dot{a}^\mu + v^\mu a^2) + f^\mu\;, \tag{6.230}$$

where δm is the self-energy defined in (6.134). Now after mass renormalization, we can remove the regularization $\epsilon \to 0$, and (6.230) becomes the Lorentz–Dirac equation.

We see that the Liénard–Wiechert field can be regularized in different ways. Physics at distances shorter than ϵ is altered: the field becomes finite but ϵ-dependent. This suggests that the mechanical mass should also be a function of regularization $m_0(\epsilon)$. A remarkable fact is that the mass renormalization (6.135), absorbing the self-energy divergence, makes the rearranged Maxwell–Lorentz electrodynamics a finite and unambiguous theory.

Problem 6.5.1. Derive the equation of motion for a dressed particle interacting with a massless scalar field [whose action is defined in (2.287)], using the analytic continuation prescription of Barut.

Answer

$$\frac{d}{ds}\,(m - g\phi)\,v^\mu - \frac{1}{3}\,g^2(\dot{a}^\mu + a^2 v^\mu) = -g\,\partial^\mu\phi\;, \tag{6.231}$$

where m is the renormalized mass, and ϕ is an external scalar field.

Problem 6.5.2. Derive the equation of motion for a dressed particle interacting with a massless symmetric tensor field $\phi_{\mu\nu}$ whose action is defined in (2.288).

Answer

$$ma^\mu + \frac{5}{3}\, g^2(\dot{a}^\mu + a^2 v^\mu) = g\,\frac{d}{ds}\left(2\phi^{\mu\nu}v_\nu - \phi_{\alpha\beta}v^\alpha v^\beta v^\mu\right) - g\,\partial^\mu\phi_{\alpha\beta}v^\alpha v^\beta\,, \tag{6.232}$$

where m is the renormalized mass, and $\phi_{\mu\nu}$ is an external tensor field.

Notes

1. The works by Thomson (1881), Larmor (1897), Heaviside (1902), Abraham (1903), (1904), (1905) Lorentz (1904*a*), von Laue (1909), Dirac (1938), and Teitelboim (1970) are milestones in the solving for the self-interaction problem of classical electrodynamics. The opening stage is outlined by Lorentz (1904*a*), Lorentz (1909), Abraham (1905), von Laue (1911), Schott (1912), Sommerfeld (1948), and Pauli (1958). Further progress is discussed by Eliezer (1947), Pais (1948), Iwanenko & Sokolow (1953), and Sokolov & Ternov (1986). A pedagogical account of this problem can be found in Jackson (1962), Barut (1964), Rohrlich (1965), and Parrot (1987). A historical survey on the self-interaction problem in classical electrodynamics is given by Dresden (1993).

2. *Section 6.1.* The Goldstone theorem was formulated and proved by Goldstone (1961). Higgs (1964), (1966) discovered a mechanism for avoiding Goldstone modes in gauge theories. The Higgs mechanism rearranges degrees of freedom in such a way that a massive vector field arises from the original system of a scalar field plus a gauge field. The idea that the degrees of freedom appearing in the Lagrangian can be dynamically rearranged was developed by Umezawa (1965).

3. *Section 6.2.* The concept of electromagnetic radiation grew up over a long period. For a historical account of the subject see Arzeliès (1966). Strange as it may seem, a consistent definition of the radiation in the context of classical electrodynamics was developed only by 1970. A key observation, made by Teitelboim (1970), is that the stress-energy tensor of the retarded Liénard–Wiechert field can be divided into two dynamically independent parts $\Theta^{\mu\nu}_{\mathrm{I}}$ and $\Theta^{\mu\nu}_{\mathrm{II}}$ according to (6.80)–(6.82). Further developments of this idea is discussed by Teitelboim, Villarroel & van Weert (1980). Kosyakov (1994) is a critical review of different definitions of the electromagnetic radiation. A central conclusion of this paper is that only the Teitelboim's definition can be correctly applied to the Yang–Mills–Wong theory. The decomposition of angular momentum (6.98)–(6.100) into the bound and emitted parts was discussed by López & Villarroel (1975).

4. *Section 6.3.* Larmor (1897) derived the formula for the radiation rate of an accelerated charge (6.144). Heaviside (1902) converted it to the form which, in modern notation, appears as (6.143). Abraham (1903) obtained the rate of momentum carried away from the charge by radiation. The asymptotic condition

on the space of allowable world lines (6.141) was imposed by Haag (1955). This condition means that charges move uniformly in the remote past, and hence stable bound systems, such as a hydrogen atom (which exhibits infinite helical world lines of its constituents), are excluded from classical electrodynamics according to this condition. Rohrlich (1960) proposed to integrate the stress-energy tensor of the electromagnetic field generated by a point charge over a hyperplane perpendicular to the world line of this charge. Teitelboim (1970) succeeded in deriving expressions (6.133) for the bound four-momentum, and (6.136) for the four-momentum of a dressed particle. Sorg (1974) showed that the bound part of the stress-energy tensor $\Theta_{\rm I}^{\lambda\mu}$ can be conveniently integrated over a truncated future light cone $C_{+}(\epsilon)$.

5. *Section 6.4.* Equation (6.173) was developed in the papers by Abraham (1903), (1904) and Lorentz (1904*a*) as an approximated equation of motion for a radiating electron. At that time, the electron was conceived as a finite charge distribution whose stability was explicitly assumed. For an extended treatment of this model see Lorentz (1909). Dirac (1938) attempted to construct a classical relativistic self-interaction theory for a point electron from first principles of the Maxwell–Lorentz electrodynamics, and came to equation (6.161). Various procedures for deriving the Lorentz–Dirac equation (6.161) was subsequently discussed by many authors. The treatment here is based on Teitelboim (1970). The integro-differential equation (6.185) was obtained by Iwanenko & Sokolow (1953). For relativistic generalizations of this equation see Rohrlich (1965). Solutions to the Abraham–Lorentz equations are given by Plass (1961), and Erber (1961).

6. *Section 6.5.* Dirac (1938) introduced the ansatz (6.196),

$$A^{\mu}_{\rm ret} = \frac{1}{2}\left(A^{\mu}_{\rm ret} - A^{\mu}_{\rm adv}\right) + \frac{1}{2}\left(A^{\mu}_{\rm ret} + A^{\mu}_{\rm adv}\right) \,. \tag{6.233}$$

He proposed $A^{\mu}_{\rm ret} - A^{\mu}_{\rm adv}$ to be interpreted as the radiation field, and $A^{\mu}_{\rm ret} + A^{\mu}_{\rm adv}$ as the bound field. If one would ignore the factor 2 in this definition, then the Abraham term Γ^{μ} given by (6.213) could be interpreted as the *radiation reaction*, or *radiation damping* force. We will see in Sect. 9.3 that this interpretation leads to many troubles and paradoxes. Therefore, it is best to regard (6.233) as a mere formal trick for discriminating between integrable and nonintegrable singularities of the retarded Liénard–Wiechert field.

Barut (1974) proposed to regularize the retarded Liénard–Wiechert field using analytic continuation from null intervals between the observation point and the emission point to timelike intervals. This method was extended to theories involving scalar and tensor fields by Barut & Villarroel (1975).

7

Lagrangian Formalism for Gauge Theories

We already have a general idea of the Yang–Mills field which was introduced in Sect. 2.2. In this chapter we look more closely at the dynamics of this field.

In Sect. 7.1 we retrace our steps in electrodynamics, with appropriate modifications, to yield a classical theory of point particles interacting with a gauge field, the so-called *Yang–Mills–Wong theory*, which closely resembles the Maxwell–Lorentz theory. Exact solutions to this theory will be discussed in subsequent chapters.

Three of the four fundamental forces are mediated by gauge fields: electromagnetic, weak, and strong. In Sect. 7.2 we review a Lagrangian framework for these forces, known as the *standard model*. Although the standard model is generally taken to be a quantum theory, the basic theoretical ideas underlying this theory are essentially classical and can be set forth in the classical context with reasonable clarity.

In Sect. 7.3 we briefly run through the intimate connection between gauge theories and differential geometry of *fiber bundles*. We then outline field dynamics on Euclidean *spacetime lattices*.

7.1 The Yang–Mills–Wong Theory

Consider a particle moving under the influence of the *Wong force*

$$f^{\mu} = Q^{a} G_{a}^{\mu\nu} v_{\nu} \ . \tag{7.1}$$

Here, Q^a is the non-Abelian charge carried by the particle, or the *color charge*. The index a ranges over the dimension of the *color space*. Repeated indices are summed over. The Wong force was defined in (2.53) as a close generalization of the Lorentz force. We now attempt to construct a Lagrangian governing the behavior of this color charged particle. The line of reasoning is essentially the same as that developed at the end of Sect. 5.5.

We use SU($\mathcal{N}$), the Lie group of unitary unimodular transformations in $\mathcal{N}$ dimensions as the basic paradigm for our approach. Let T_a be the generators of SU($\mathcal{N}$). This means that elements of SU($\mathcal{N}$) can be written as

$$\Omega = \exp\left(i\omega^a T_a\right) \ , \tag{7.2}$$

where T_a are $\mathcal{N}^2 - 1$ independent $\mathcal{N} \times \mathcal{N}$ matrices, and ω^a are the parameters of the transformation. For infinitesimal ω^a, (7.2) becomes

$$\Omega = \mathbf{1} + i\omega^a T_a \ . \tag{7.3}$$

If we apply unitarity ($\Omega\,\Omega^\dagger = 1$, where $\Omega^\dagger$ denotes the Hermitian adjoint of Ω) to (7.3), we obtain

$$T_a^\dagger = T_a \ . \tag{7.4}$$

Therefore, T_a is a Hermitian matrix. From the fact that Ω is unimodular, $\det\Omega = 1$, we conclude that T_a is also traceless,

$$\mathrm{tr}\,(T_a) = 0 \ . \tag{7.5}$$

We thus have a set of $\mathcal{N}^2-1$ Hermitian traceless matrices T_a which provide a basis for the color space V. This space becomes the su($\mathcal{N}$) Lie algebra associated with the SU($\mathcal{N}$) Lie group if we define matrix commutation as the multiplication rule

$$[T_a, T_b] = i f^c{}_{ab}\, T_c \ . \tag{7.6}$$

The numbers $f^c{}_{ab}$ are called the structure constants of su($\mathcal{N}$). A metric in the color space is provided by the *Killing form* $g_{ab} = -f^c{}_{ad}\, f^d{}_{bc}$. Using this metric, together with its inverse g^{ab} (defined by $g_{ab}\, g^{bc} = \delta^c{}_a$), one can raise and lower color indices. If the set of generators T_a are assembled to form the Cartan basis, then the f_{abc} are real and completely antisymmetric.

Let ψ be a vector in the color space which is invariant under SU($\mathcal{N}$):

$$\Omega\,\psi = \psi \ . \tag{7.7}$$

Using (7.3) we see that (7.7) implies $\mathcal{N}^2 - 1$ equations:

$$T_a \psi = 0 \ , \tag{7.8}$$

where $a = 1, \ldots, \mathcal{N}^2-1$. If G is a subgroup of SU($\mathcal{N}$), say, SU(n) with $n < \mathcal{N}$, then elements of G look like Ω in (7.2), but the summation over a is from 1 to $n^2 - 1$. The condition that ψ is invariant under G is again given by the set of n equations (7.8), in which $a = 1, \ldots, n^2 - 1$.

One may write the color charge Q^a and the field strength $G^a_{\mu\nu}$ in matrix notation:

$$Q = -ig\, Q^a T_a, \quad G_{\mu\nu} = \frac{i}{g}\, G^a_{\mu\nu} T_a \ , \tag{7.9}$$

where g is a real parameter which is called the *Yang–Mills coupling constant.* Recall that quantities which can be expressed as linear combinations of the

T_a, such as Q and $G_{\mu\nu}$, are said to transform according to the *adjoint* representation of SU($\mathcal{N}$). If we impose the orthonormalization condition[1]

$$\mathrm{tr}\,(T_a T_b) = \delta_{ab}\ , \tag{7.10}$$

then (7.1) becomes

$$f^\mu = v_\nu\,\mathrm{tr}\,(Q G^{\mu\nu})\ . \tag{7.11}$$

In order to describe a particle which carries color degrees of freedom, we need more variables in addition to the position z^μ. Our prescription is to replace Q as the basic color variable by what is in a rough sense a square root of the color charge Q, much as the square root of the electric charge e has been extracted in equation (5.267):

$$Q = q^a\,\eta_i^*\,(T_a)^i{}_j\,\eta^j\ . \tag{7.12}$$

The matrix $(T_a)^i{}_j$ acts on the resulting square-root quantities η_i^* and η^j as if those were $\mathcal{N}$-component row and column vectors. It is clear that η_i^* and η^j transform according to some representation of SU($\mathcal{N}$). This representation is called *fundamental.*

The action for a particle whose state is specified by spacetime coordinates z^μ and color variables η^j is

$$S = -\int d\tau \left\{ m_0\sqrt{\dot z_\mu\,\dot z^\mu} + \sum_a q^a\eta_i^*\left[\delta^i{}_j\frac{d}{d\tau} + \dot z^\lambda A^a_\lambda\,(T_a)^i{}_j\right]\eta^j\right\}\ . \tag{7.13}$$

Here, A^a_λ are color-valued vector functions of z^μ. This quantity is a generalization of the vector potential in electrodynamics. We will call A^a_λ the *gauge field* or the *Yang–Mills field.* The number of color components of A^a_λ equals the number of the group generators. For SU($\mathcal{N}$), the color index a runs from 1 to $\mathcal{N}^2-1$.

The Euler–Lagrange equations for η^i and η_j^* read

$$\begin{aligned}\dot\eta^i &= -(\dot z\cdot A^a)\,(T_a)^i{}_j\,\eta^j\ ,\\ \dot\eta_j^* &= \eta_i^*\,(\dot z\cdot A^a)\,(T_a)^i{}_j\ .\end{aligned} \tag{7.14}$$

They can be combined (Problem 7.1.1) into the *Wong equation* for the color charge,

$$\dot Q^a = -if^a{}_{bc}\left(\dot z\cdot A^b\right)Q^c\ . \tag{7.15}$$

With (7.9) and

$$A_\mu = \frac{i}{g}\,A^a_\mu\,T_a\ , \tag{7.16}$$

(7.15) becomes

[1] Note that the normalization which is used in Sects. 8.1, 8.3, 8.4 differs from that given by (7.10).

$$\dot{Q} = -ig\,[Q,\, \dot{z}^\mu A_\mu] \ , \tag{7.17}$$

Unlike spacetime coordinates, internal variables are generally associated with first order equations of motion. An example, other than (7.17), is found in the Frenkel model of a particle with spin, equation (2.325).

Since f_{abc} is completely antisymmetric, (7.15) implies

$$Q^a \dot{Q}_a = 0 \ . \tag{7.18}$$

Therefore, the color charge magnitude remains constant,

$$Q^2 = \text{const} \ . \tag{7.19}$$

Varying the action (7.13) with respect to z^μ and taking into account (7.14), one finds (Problem 7.1.2) the equation of motion for a colored particle

$$m_0 \ddot{z}^\mu = Q^a G_a^{\mu\nu} \dot{z}_\nu \ , \tag{7.20}$$

where $G_a^{\mu\nu}$ is expressed in terms of A_a^μ,

$$G_a^{\mu\nu} = \partial^\mu A_a^\nu - \partial^\nu A_a^\mu + i f_a^{\ bc} A_b^\mu\, A_c^\nu \ , \tag{7.21}$$

or

$$G_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - ig\,[A_\mu,\, A_\nu] \ . \tag{7.22}$$

By analogy with Maxwell–Lorentz electrodynamics where the auxiliary variable ζ may be acted upon by U(1) phase transformations, as discussed in Sect. 5.5, we introduce the following SU($\mathcal{N}$) phase transformations

$$\eta \to \Omega\, \eta\, , \quad \eta^* \to \eta^*\, \Omega^\dagger \ , \tag{7.23}$$

where Ω is given by (7.2). If ω^a are constant, then transformations (7.23) are called *global*, but if ω^a are functions of the particle's position z^μ, then the transformations are said to be *local*. For the action (7.13) to be invariant under local phase transformations, the gauge field must be transformed as well

$$(\dot{z} \cdot A) \to \Omega\, (\dot{z} \cdot A)\, \Omega^\dagger - \frac{i}{g}\, \dot{\Omega}\, \Omega^\dagger \ . \tag{7.24}$$

The invariance of the action (7.13) under (7.23) and (7.24) for a local transformation is ensured by the fact that

$$D_\tau = \frac{d}{d\tau} - ig\,(\dot{z} \cdot A) \tag{7.25}$$

is a one-dimensional covariant derivative. With the use of D_τ, the second term of the integrand in (7.13) can be written as

$$-\sum_a q^a \eta_i^* D_\tau \eta^j \ . \tag{7.26}$$

Our next task is to discuss the law governing the evolution of the Yang–Mills field. This law is given by a system of nonlinear partial differential equations known as the *Yang–Mills equations*. The color charge Q^a is assumed to be a source of the Yang–Mills field. Unlike the electromagnetic field, which is distributed over macroscopic regions, all fundamental forces that owe their origin to Yang–Mills fields are reputed to have very short range, characteristic of subnuclear realm. In the absence of a clear experimental view of these forces, we are guided by parallels with electrodynamics and by general symmetry principles.

Because (7.24) holds for all paths we must have

$$A_\mu \to \Omega \left(A_\mu + \frac{i}{g}\,\partial_\mu \right) \Omega^\dagger \,, \tag{7.27}$$

where Ω is defined in (7.2) and the parameters ω^a being arbitrary functions of spacetime. One can show (Problem 7.1.3) that Q and $G_{\mu\nu}$ have simple transformation properties

$$Q \to \Omega\, Q\, \Omega^\dagger, \quad G_{\mu\nu} \to \Omega\, G_{\mu\nu}\, \Omega^\dagger \tag{7.28}$$

with respect to the gauge transformation (7.23) and (7.27). The Yang–Mills field strength $G_{\mu\nu}$ is not gauge invariant (as opposed to the electromagnetic field strength $F_{\mu\nu}$ which is gauge invariant). Note that only quantities which transform as scalars with respect to gauge transformations[2] are observable. Thus, $G_{\mu\nu}$ and Q are not observable, while f^μ is.

For infinitesimal $\omega^a(x)$ the gauge transformation (7.27) becomes

$$\delta A^a_\mu = -\left(\delta^a{}_c \partial_\mu + i f^a{}_{bc} A^b_\mu\right)(i\omega^c) \,, \tag{7.29}$$

or

$$\delta A_\mu = -\frac{i}{g}\, D_\mu \Omega \,, \tag{7.30}$$

where the *covariant derivative* is

$$D_\mu = \partial_\mu - igA_\mu \,. \tag{7.31}$$

We see that the action of D_μ on any field $\phi = \phi^a T_a$, transforming according to the adjoint representation of SU($\mathcal{N}$), is given by

$$D_\mu \phi = \partial_\mu \phi - ig\,[A_\mu, \phi] \,. \tag{7.32}$$

Since $[\partial_\mu, \phi]\,\chi = \partial_\mu(\phi\chi) - \phi\,\partial_\mu\chi = (\partial_\mu\phi)\,\chi$, we have the identity $[\partial_\mu, \phi] = \partial_\mu\phi$, which, combined with (7.32), yields

$$D_\mu \phi = [D_\mu, \phi] \,. \tag{7.33}$$

[2] Such scalar quantities are one-dimensional representations of the color SU($\mathcal{N}$) group, hence the name *color singlets*.

We now suppose that the Lagrangian governing the dynamics of the Yang–Mills field is given by the simplest Lorentz and gauge invariant expression

$$\mathcal{L} = -\frac{1}{16\pi}\, G_a^{\mu\nu} G^a_{\mu\nu}\ , \tag{7.34}$$

similar to the Larmor Lagrangian in electrodynamics. It is impossible to amend this Lagrangian by addition of a mass term, proportional to $\frac{1}{2} M^2 A^a_\mu A_a^\mu$, without spoiling its gauge invariance. One can verify that $\frac{1}{2}\,\epsilon^{\mu\nu\alpha\beta} G^a_{\mu\nu} G^a_{\alpha\beta} = G^a_{\mu\nu}\,{}^*G_a^{\mu\nu}$ is a total derivative (Problem 7.1.4). This invariant makes no contribution to the Euler–Lagrange equations.

Note that the last term in (7.13), responsible for the interaction between a point particle and the Yang–Mills field, can be rewritten as

$$-\int d^4x\, j_a^\mu A^a_\mu\ , \tag{7.35}$$

where

$$j_a^\mu(x) = \int_{-\infty}^{\infty} ds\, Q_a(s)\, v^\mu\, \delta^4\left[x - z(s)\right] \tag{7.36}$$

is the color charge current carried by a point particle, analogous to the electric current. It can be shown (Problem 7.1.5) that expression (7.35) alone is gauge invariant provided that $Q^a(s)$ obeys the Wong equation (7.15). We thus see that the Yang–Mills sector is given by

$$S = -\int d^4x \left(\frac{1}{16\pi}\, G_a^{\mu\nu} G^a_{\mu\nu} + j_a^\mu A^a_\mu \right)\ . \tag{7.37}$$

We repeat mutatis mutandis the calculations of Sect. 5.1 to derive, from (7.37), the *Yang–Mills equations*

$$\partial_\mu G_a^{\mu\nu} + i f_{abc} A^b_\mu G^c_{\mu\nu} = 4\pi j_a^\nu\ , \tag{7.38}$$

or, in matrix notation,

$$[D_\mu, G^{\mu\nu}] = 4\pi j^\nu\ . \tag{7.39}$$

In contrast to Maxwell's equations, these equations are *nonlinear*. One may therefore expect new phenomena inherent in nonlinear dynamics. Indeed, we will learn in the next chapter that the Yang–Mills field is capable of realizing two phases with different symmetries, which is somewhat similar to the situation in hydrodynamics in which a fluid can be in either of two different regimes of motion, laminar and turbulent.

Why do we focus upon the Lagrangian quadratic in the Yang–Mills field (7.34) rather than proceed from some function of this invariant? The field equations are nonlinear anyway. The reason for this is not just simplicity. This choice makes the Yang–Mills sector (7.37) *conformally invariant*. The Yang–Mills sector is free of dimensional parameters because its coupling constant g, the color charge Q^a and the structure constants f_{abc} are all pure numbers when expressed in natural units. The metric stress-energy tensor derived from the action (7.37),

$$\Theta_{\mu\nu} = \frac{1}{4\pi}\left(G_{a\mu}{}^{\lambda} G^{a}_{\lambda\nu} + \frac{\eta_{\mu\nu}}{4}\, G_{a}^{\alpha\beta} G^{a}_{\alpha\beta}\right)\,, \tag{7.40}$$

is traceless as a consequence

$$\Theta^{\mu}{}_{\mu} = 0\,. \tag{7.41}$$

Note that the stress-energy tensor (7.40) resembles that of electrodynamics. This is not surprising. The Yang–Mills action (7.37) looks like the field action of electrodynamics, so the Hilbert definition of the stress-energy tensor leads to similar expressions for $\Theta_{\mu\nu}$.

We thus arrive at the *Yang–Mills–Wong theory*, a non-Abelian generalization of the Maxwell–Lorentz electrodynamics, that involves dynamical equations for a particle (7.15), (7.20), and for the Yang–Mills field (7.38). Our formulation of Yang–Mills–Wong theory can be extended to cover N particles interacting with the Yang–Mills field. To this end, we suppose that each particle contributes a term of the form (7.13) to the action. In the following, we will try to employ the Yang–Mills–Wong theory as a toy model of strongly interacting quarks and gluons, taking a quark as a colored Wong particle, and the gluon field as a classical non-Abelian gauge field.

The current of the color charge j^{ν} obeys the *covariant conservation law*

$$D_{\nu} j^{\nu} = 0\,. \tag{7.42}$$

To see this, note that for any gauge covariant quantity ϕ

$$\left[D_{\mu}, D_{\nu}\right]\phi = -ig\left[G_{\mu\nu}, \phi\right]\,. \tag{7.43}$$

For the case that ϕ is the Yang–Mills field strength we have

$$\left[D_{\mu}, D_{\nu}\right] G_{\mu\nu} = -ig\left[G_{\mu\nu}, G_{\mu\nu}\right] = 0\,. \tag{7.44}$$

Hence (7.39) implies

$$4\pi D_{\nu} j^{\nu} = D_{\nu} D_{\mu} G^{\mu\nu} = \frac{1}{2}\left[D_{\nu}, D_{\mu}\right] G^{\mu\nu} = 0\,. \tag{7.45}$$

We see that (7.42) is a consistency condition for the Yang–Mills equations (7.38). This condition arises – by virtue of Noether's second theorem, when one forms the invariance of the Yang–Mills action (7.37) to infinitesimal gauge transformations (7.30).

Putting $j^{\mu} = 0$, we obtain the so-called *pure Yang–Mills theory.*

By recognizing that $G_{\mu\nu}$ is expressed in terms of A_{μ}, according to (7.22), we come to a condition underlying this relation, the *Bianchi identity*,

$$\left[D_{\mu}, {}^{*}G^{\mu\nu}\right] = 0\,, \tag{7.46}$$

or, equivalently,

$$[D_\lambda, G_{\mu\nu}] + [D_\nu, G_{\lambda\mu}] + [D_\mu, G_{\nu\lambda}] = 0 \ . \tag{7.47}$$

To verify (7.47), we use the Jacobi identity

$$[D_\lambda, [D_\mu, D_\nu]] + [D_\nu, [D_\lambda, D_\mu]] + [D_\mu, [D_\nu, D_\lambda]] = 0 \ , \tag{7.48}$$

combined with (7.43).

If $\mathbf{E}_i = F_{0i}$ and $\mathbf{B}_i = -\frac{1}{2}\,\epsilon_{ijk}F^{jk}$ are interpreted as respectively electric and magnetic components of the Yang–Mills field, then the Bianchi identity (7.46) suggests that the density of magnetic charge is in general nonvanishing:

$$\nabla \cdot \mathbf{B}_a = -\frac{1}{2}\,\epsilon^{ijk}\, f_{abc} \nabla_i\, \mathbf{A}_j^b \mathbf{A}_k^c \ . \tag{7.49}$$

We will see in the next chapter that there exist field configurations ('t Hooft–Polyakov monopoles) with nonzero total magnetic charge.

When the Bianchi identity 7.46) is compared with the Yang–Mills equations (7.39), it become apparent that any Yang–Mills field which is *self-dual*

$$^*G_{\mu\nu} = \pm i\, G_{\mu\nu} \tag{7.50}$$

obeys automatically (7.39). Being a nonlinear partial differential equation of first order, the self-duality condition (7.50) is much simpler to solve than the second-order Yang–Mills equations (7.39). Some physically important field configurations, such as instantons and monopoles, obey equation (7.50).

Our discussion regarding the SU($\mathcal{N}$) gauge group pertains equally to other compact Lie groups. Such groups have finite-dimensional unitary representations Ω which can be expressed in the form (7.2). Semisimple Lie groups are best suited to the role of gauge groups because the Killing form is not degenerate. Note also that if the Killing form is positive definite, then the group is compact.

If the gauge group is simple, then all its generators T_a transform irreducibly under the action of this group, and gauge invariance implies that all A_μ^a have the same coupling constant. However, if the gauge group is a product of simple factors, such as SU(2)×U(1) to be discussed in the next section, then the generators of different factors never mix with each other under the action of the group, and the associated gauge fields can have different coupling constants.

If we turn from point particles to a continuous field, then its interaction with the Yang–Mills field is fixed by the minimal coupling prescription, namely by replacing ∂_μ with $D_\mu = \partial_\mu - igA_\mu$. To illustrate, consider a Dirac field ψ^i transforming as the fundamental representation of some gauge group. The Lagrangian governing the behavior of this field interacting with the Yang–Mills field is given by

$$\mathcal{L} = \bar{\psi}_i \left[i\gamma^\mu (D_\mu)^i{}_j - m \right] \psi^j \ , \tag{7.51}$$

where $(D_\mu)^i{}_j = \delta^i{}_j\partial_\mu - igA^a_\mu(T_a)^i{}_j$, $i, j = 1, \ldots, \mathcal{N}$. Under a gauge transformation

$$\psi \to \Omega\,\psi, \quad \bar{\psi} \to \bar{\psi}\,\Omega^\dagger, \quad A_\mu \to \Omega\,A_\mu\,\Omega^\dagger - \frac{i}{g}\,(\partial_\mu\Omega)\,\Omega^\dagger\ , \tag{7.52}$$

D_μ transforms as

$$D_\mu \to \Omega\,D_\mu\,\Omega^\dagger\ , \tag{7.53}$$

and hence the Lagrangian (7.51) is gauge invariant.

Comparing the first term of the Lagrangian (7.51) with (7.26), one can observe that the gauge interaction of the Dirac field ψ is much the same as that of the particle color variable η. One would be tempted to consider η as a kind of Dirac field in spacetime of dimension 1. However that may be, formal switching between Wong particle and Dirac field descriptions is readily available.

Problem 7.1.1. Show that (7.14) implies (7.15).

Problem 7.1.2. Derive (7.20) from (7.13).

Problem 7.1.3. Verify (7.28).

Problem 7.1.4. Verify that $G^a_{\mu\nu}\,{}^*G^{\mu\nu}_a$ (which is called the *Pontryagin density*) is a total derivative,

$$\frac{1}{2}\,G^a_{\mu\nu}\,{}^*G^{\mu\nu}_a = \partial_\lambda\mathcal{J}^\lambda\ , \tag{7.54}$$

where the vector

$$\mathcal{J}^\lambda = \epsilon^{\lambda\mu\nu\rho}\left(A_{a\mu}\partial_\nu A^a_\rho + \frac{1}{3}\,f_{abc}A^a_\mu A^b_\nu A^c_\rho\right) \tag{7.55}$$

is called the *topological Chern–Simons current.*

Problem 7.1.5. Show that the action (7.35) is gauge invariant if the color charge $Q^a(s)$ obeys the Wong equation (7.15).

Problem 7.1.6. Derive the Noether identity

$$\partial_\mu T^{\lambda\mu} = \frac{1}{4\pi}\,\mathcal{E}^a_\mu\,G^{\mu\lambda}_a + \int ds\,\varepsilon^\lambda(s)\,\delta^4\left[x - z(s)\right]\ . \tag{7.56}$$

Here, $T^{\mu\nu} = \Theta^{\mu\nu} + t^{\mu\nu}$, with $\Theta^{\mu\nu}$ and $t^{\mu\nu}$ being given by (7.40) and (6.78), respectively. $\mathcal{E}^a_\mu$ and ε^λ are the Eulerians corresponding to the Yang–Mills equations (7.38) and equation of motion for a colored particle (7.20).

Problem 7.1.7. Let $z^\mu(s)$ be a path between two spacetime points x^μ and y^μ, and ψ a field for which the covariant derivative vanishes along this path:

$$\dot{z}^\mu\,(D_\mu)_i{}^j\,\psi_j(z) = 0\ . \tag{7.57}$$

Examples include the variables η and η^* in equation (7.14), and the color charge Q in (7.15). Condition (7.57) can be viewed as a first-order partial differential equation for the unknown function ψ. Verify that (7.57) is obeyed by

$$\psi(y) = P \exp\left[ig \int_x^y ds\, \dot{z}^\mu(s) A_\mu(z)\right] \psi(x) \,, \tag{7.58}$$

where P is the *path ordering operator* which arranges factors $T_a A_\mu^a(z(s))$ in order of increasing s in each term of the power series expansion of the exponential.

7.2 The Standard Model

The *standard model* summarizes our knowledge of particles and forces in the subnuclear realm. This realm refers to short-distance structures ranging between 1 and 10^{-4} fm, which can be probed by particles accelerated to energies in the range 100 MeV – 1 TeV[3].

The standard model consists of two parts: the *Glashow–Salam–Weinberg model* that unifies weak and electromagnetic forces into a SU(2)×U(1) gauge theory, and *quantum chromodynamics* that describes strong interactions in terms of a SU(3) gauge theory. It is safe to say that these forces are dominant in the subnuclear regime; gravity is feeble in this regime and may be ignored.

The Lagrangian of electroweak interactions is invariant under SU(2)×U(1), where SU(2) is called the *weak isospin* group, and U(1) the *weak hypercharge* group. The Higgs mechanism breaks this symmetry to a U(1) subgroup of SU(2)×U(1). The three gauge fields associated with the broken generators become the massive vector particles $W_\mu^\pm$ and Z_μ^0 that mediate the weak interaction. The gauge field associated with the single unbroken generator is the massless vector field A_μ responsible for electromagnetism. By contrast, the SU(3) symmetry of quantum chromodynamics is unbroken. The strong force is mediated by the exchange of massless fields A_μ^a between *quarks* that carry an internal degree of freedom called *color*.

The full symmetry of the standard model is SU(3) × SU(2) × U(1). This symmetry was inferred from a combination of experimental observations and theoretical analyses, and has been impressively confirmed by experiment. Three parameters measure the strength of the interactions involved: the strong force coupling g_s, weak isospin coupling g_1, and weak hypercharge coupling g_2. The parameters g_1 and g_2 are expressed in terms of the electromagnetic coupling constant e:

$$g_1 = e\cos\theta_W \,, \quad g_2 = e\sin\theta_W \,, \tag{7.59}$$

where θ_W is called the *weak mixing angle* or the *Weinberg angle*.

[3] Recall that eV is a shorthand for *electron volt* – the energy acquired by an electron when accelerated through a potential difference of 1 volt; 1 MeV = 10^6 eV; 1 GeV = 10^9 eV; and 1 TeV = 10^{12} eV. Note that 1 TeV = 1.6 erg is comparable to the kinetic energy of a mosquito.

The standard model is formulated as a quantum field theory. The matter fields are represented by Dirac fields. Since these fields play the role identical to that of particles in the Yang–Mills–Wong theory, we will refer to them as 'particles', by recognizing that the concepts are closely related on the Lagrangian level.

It is customary to distinguish between the gauge particles which mediate forces – the $W^{\pm}$, Z^0 and photon – and 'matter' particles which experience these forces. Matter particles can be divided into *hadrons*, which feel the strong interaction, and *leptons* which do not. Hadrons can be subdivided into those having half-integer spin, the *baryons* – typified by protons and neutrons, and those possessing integer spin, the *mesons*–for example, the π-meson. An exhaustive list of leptons follows: the electron, the muon, the τ-lepton, and their associated neutrinos ν_e, ν_μ, and ν_τ.

Hadrons are not elementary, they are composed of *quarks*. The strong force is mediated by the exchange of gluons between quarks and binds them inside hadrons. Baryons are composed of three quarks of different colors, which are combined into color-neutral states. Mesons contain quark and antiquark pairs whose colors cancel.

At our present level of understanding, leptons and quarks are truly elementary, that is, structureless particles. There are three generations of quarks and leptons. The first generation contains two SU(2) weak isospin doublets: the lepton doublet (e, ν_e), and the quark doublet (u, d). The second and third generations are identical to the first one in every respect except for their masses; they contain, respectively, the lepton and quark doublets (μ, ν_μ), (c, s), and (τ, ν_τ), (t, b).

We now briefly review the Lagrangian of the standard model. The coupling between the matter and gauge fields is given by

$$\mathcal{L}_{\text{int}} = -A^a_\mu\, J^\mu_a - W^A_\mu J^\mu_A - B_\mu J^\mu \ , \tag{7.60}$$

where A^a_μ is the SU(3) color gauge field, W^A_μ the SU(2) weak isospin gauge field, B_μ the U(1) weak hypercharge gauge field, and the currents J^μ_a, J^μ_A and J^μ are defined by

$$j^\mu_a = -ig_s \sum_f \bar{\psi}_f\, \gamma^\mu\, t_a\, \psi_f \ , \tag{7.61}$$

$$J^\mu_A = -\frac{ig_1}{2} \sum_G \overline{\Psi}_G\, \gamma^\mu\, T_A\, (1+\gamma_5)\, \Psi_G \ , \tag{7.62}$$

$$J^\mu = ig_2 \sum_G \overline{\Psi}_G\, \gamma^\mu\, \Psi_G \ . \tag{7.63}$$

Here, f labels the quark species u, d, etc., called *flavors*, t_a and T_A are generators of SU(3) and SU(2), $\overline{\Psi}_G$ and Ψ_G stand for rows and columns with the SU(2) lepton and quark doublets of the Gth generation. For example, Ψ_1 involves

$$\begin{pmatrix} e \\ \mu_e \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} u \\ d \end{pmatrix} , \tag{7.64}$$

so that $\overline{\Psi}_1\gamma^\mu\Psi_1 = \bar{e}\gamma^\mu e + \bar{\mu}_e\gamma^\mu\mu_e + \bar{u}\gamma^\mu u + \bar{d}\gamma^\mu d$. The index G runs over the three lepton and quark generations. The presence of the projection operator

$$\frac{1}{2}(1+\gamma_5) \tag{7.65}$$

reflects the experimental fact that the weak interactions exhibit maximal parity violation. Indeed, using simple algebra one can show (Problem 7.2.1) that any mixture of γ^μ and $\gamma^\mu\gamma_5$ violates parity conservation.

The Lagrangian of free gauge and matter fields reads

$$\mathcal{L}_0 = -\frac{1}{4\pi}\, G^a_{\mu\nu}G^{\mu\nu}_a - \frac{1}{4\pi}\, H^A_{\mu\nu}H^{\mu\nu}_A - \frac{1}{4\pi}\, F_{\mu\nu}F^{\mu\nu} + \sum_G i\,\bar{\psi}_G\,\gamma^\mu\partial_\mu\psi_G , \tag{7.66}$$

where

$$G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g_s f^a{}_{bc}A^b_\mu A^c_\nu , \tag{7.67}$$

$$H^A_{\mu\nu} = \partial_\mu W^A_\nu - \partial_\nu W^A_\mu + g_1 f^A{}_{BC}W^B_\mu W^C_\nu , \tag{7.68}$$

$$F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu , \tag{7.69}$$

with f^{abc} and f^{ABC} being respectively the SU(3) and SU(2) structure constants.

Let us turn to the Higgs mechanism. We learned in Sect. 6.1 that this mechanism provides longitudinal modes to Abelian gauge fields, which then become massive. With reference to Problem 4.1.4, one may think of massive fields as mediators of short-range forces[4]. Why do we take trouble over

[4] This terminology itself need not be evidence that we are faced with a quantum phenomenon. Indeed, looking at the Yukawa potential (4.66),

$$\Phi = g\,\frac{e^{-\mu r}}{4\pi r} , \tag{7.70}$$

we see that the mass μ acts as a cutoff, and 'short-range' means that Φ is essentially vanishing at a distance about μ^{-1} from the source. The heavier the Yukawa field, the shorter the range of the Yukawa force. However, we do invoke quantum considerations (the Heisenberg uncertainty principle) when we convert explicitly from space to momentum scales. For example, in natural units, a conversion formula is

$$1\,\mathrm{fm}^{-1} = 197 \text{ MeV} . \tag{7.71}$$

Hence the π-meson field Φ, which mediates the strong interaction in Yukawa theory and has a mass of 140 MeV, is active within a radius of 1.4 fm. This is a typical distance between protons and neutrons in the atomic nucleus. The Yukawa potential (7.70) describes the nuclear force which holds the nucleons together. In fact, it is merely a phenomenological description of the residual SU(3)-color force of the standard model, much as the van der Waals interaction between neutral molecules is a vestigial effect of screened Coulomb forces.

the Higgs mechanism rather than introducing mass terms explicitly? It was already mentioned that such terms violate gauge invariance. On the other hand, the Higgs mechanism allows masses for vector fields without losing their properties under the gauge transformation. A strong reason for preserving gauge symmetry is *renormalizability.* A theory is called renormalizable if all divergences can be absorbed through redefinitions of parameters in the Lagrangian. Although the SU(2)×U(1) symmetry of electroweak interactions takes a secret form, this sector of the standard model is renormalizable. On the other hand, the SU(3) color symmetry of the strong interactions is explicit, and this sector of the standard model displays not only renormalizability, but also *asymptotic freedom.* However, these topics are beyond the scope of the present discussion.

By a similar argument, the Higgs mechanism in a non-Abelian gauge theory should endow the gauge fields with mass, so as to make the weak force short range. Consider the simplest version of the Glashow–Weinberg–Salam model involving a SU(2) doublet of Higgs scalar fields:

$$\phi = \begin{pmatrix} \phi_+ \\ \phi_0 \end{pmatrix} \tag{7.72}$$

The component ϕ^0 assumes a nonzero constant value which lies at the minimum of the Higgs potential v. Consider the Lagrangian which describes the Higgs field interacting with the gauge fields W_μ^A and B_μ

$$\begin{aligned} \mathcal{L}_{\mathrm{H}} = \frac{1}{2} \left(\partial_\mu \phi - i g_1 T_A W_\mu^A \phi + i g_2 B_\mu \phi\right)^\dagger \left(\partial^\mu \phi - i g_1 T^B W_B^\mu \phi + i g_2 B^\mu \phi\right) \\ - \frac{\lambda^2}{4} \left(v^2 - \phi^\dagger \phi\right)^2 \ , \end{aligned} \tag{7.73}$$

where $\phi^\dagger \phi = |\phi_+|^2 + |\phi_0|^2$. Now shift ϕ_0 to the minimum of the Higgs potential, $\phi_0 = v$. This gives a mass term

$$\begin{aligned} \mathcal{L}_{\mathrm{m}} &= \frac{1}{2} g_1^2 v^2 \sum_{A=1}^{2} W_A^\mu W_\mu^A + \frac{1}{2} v^2 \left(g_1 W_\mu^3 - g_2 B^\mu\right)^2 \\ &= \frac{1}{2} g_1^2 v^2 \left(W^+\right)^\mu \left(W^-\right)_\mu + \frac{1}{2} e^2 v^2 \left(W_\mu^3 \cos\theta_W - B^\mu \sin\theta_W\right)^2 \ , \end{aligned} \tag{7.74}$$

where the generator T_3 is assumed to be diagonal, more precisely, $T_3 = \mathrm{diag}\,(1, -1)$, and $W_\mu^\pm = (W_\mu^1 \pm i W_\mu^2)/\sqrt{2}$. Two charged vector fields $W_\mu^\pm$ and one neutral vector field

$$Z_\mu = -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W \tag{7.75}$$

acquire masses, while another vector field, orthogonal to Z_μ,

$$A_\mu = B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W \ , \tag{7.76}$$

remains massless, and hence may be identified with the electromagnetic field. We see that the neutral vector field Z_μ proves to be more massive than the charged ones $W_\mu^\pm$, $M_W^2 = M_Z^2 \cos\theta_W$. This is the reason for choosing the weak hypercharge gauge group U(1) to be different from the electromagnetic gauge group $\mathrm{U}(1)_{\mathrm{em}}$. Note also that leaving A_μ massless is consistent with the general condition (7.8), which now takes the form

$$(T_3 + \mathbf{1}) \begin{pmatrix} 0 \\ v \end{pmatrix} = 0 \, . \tag{7.77}$$

Therefore, the unbroken $\mathrm{U}(1)_{\mathrm{em}}$ is generated by the diagonal matrix $T_3 + \mathbf{1}$.

Lepton and quark fields can also be endowed with masses through their coupling with the Higgs field, rather than through explicit mass terms. Without going into detail we simply note that Yukawa terms $-f\bar{\psi}\psi\phi$ become $-m\bar{\psi}\psi$, with $m = fv$, when the Higgs field takes its equilibrium value $\phi_0 = v$.

Problem 7.2.1. Find the transformation laws of γ_μ and γ_5 under space reflections. Show that $\overline{\psi}\gamma_\mu\psi$ and $\bar{\psi}\gamma_\mu\gamma_5\psi$ transform as polar and axial vectors.

Hint We are looking for a space reflection operator S_P such that

$$S_P^{-1}\gamma_0 S_P = \gamma_0, \quad S_P^{-1}\gamma_i S_P = -\gamma_i \, , \tag{7.78}$$

or, in covariant notation,

$$S_P^{-1}\gamma_\mu S_P = \gamma^\mu \, . \tag{7.79}$$

Assuming that $\psi \to \psi' = S_P\psi$, equation (7.78) affords invariance of the Dirac equation

$$(i\gamma^\mu\partial_\mu - m)\,\psi = 0 \, , \tag{7.80}$$

because $(\partial_t, \nabla) \to (\partial_t, -\nabla)$. An appropriate choice of S_P is γ_0. Taking into account that $S_P = (S_P)^\dagger = S_P^{-1}$, we have $\bar{\psi}' = (\psi')^\dagger\gamma_0 = \psi^\dagger(S_P)^\dagger\gamma_0 = \psi^\dagger\gamma_0 S_P^{-1} = \bar{\psi}S_P^{-1}$, and so $\bar{\psi}'\gamma_\mu\psi' = \bar{\psi}\gamma^\mu\psi$. With $\gamma_5 S_P = -S_P\gamma_5$, the transformation law $\bar{\psi}'\gamma_\mu\gamma_5\psi' = -\bar{\psi}'\gamma^\mu\gamma_5\psi'$ can be read immediately. Hence, $\bar{\psi}\gamma_\mu\psi$ has opposite transformation properties from $\bar{\psi}\gamma_\mu\gamma_5\psi$ under space reflections. Parity violation may be achieved through the presence of both terms in the Lagrangian.

7.3 Lattice Formulation of Gauge Theories

According to the Weyl principle, discussed in Sect. 5.5, the interaction of charged matter with the electromagnetic field, arising from the U(1) local gauge invariance, can be introduced by replacing ∂_μ with $\partial_\mu + ieA_\mu$. In an alternative approach, proposed by Stanley Mandelstam in 1962, emphasis is given to the integral aspect of this principle. The *phase factor*

$$U_\Gamma(x,y) = \exp\left[ie\int_x^y d\tau\, \dot{z}^\mu(\tau)A_\mu(z)\right] \tag{7.81}$$

is adopted as a primary concept. Here, the integration is over any path Γ between x and y. Multiplication of phase factors with a coincident endpoint gives the phase factor on the composite path:

$$U_{\Gamma_1}(w,x)\,U_{\Gamma_2}(x,y) = U_{\Gamma_1+\Gamma_2}(w,y)\ . \tag{7.82}$$

Under a gauge transformation the phase factor $U(x,y)$ becomes

$$\exp\left(ie\int_x^y dz^\mu A_\mu\right) \to \exp\left[ie\chi(y)\right]\exp\left(ie\int_x^y dz^\mu A_\mu\right)\exp\left[-ie\chi(y)\right]\ . \tag{7.83}$$

Now it is possible to reformulate electrodynamics in terms of manifestly gauge invariant integral quantities. In a simply connected region, using the Stokes theorem, we obtain

$$\exp\left(ie\oint_{\mathcal{C}} dz^\mu A_\mu\right) = \exp\left(ie\int_{\mathcal{S}} dS^{\mu\nu}F_{\mu\nu}\right)\ , \tag{7.84}$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu\ . \tag{7.85}$$

This provides an alternative definition of the field strength $F_{\mu\nu}$.

This approach shows its full power in analyzing physical situations which involve topology, for example, magnetic monopoles. However, we must omit these interesting issues and instead address the following generalization of the phase factor to non-Abelian gauge theories

$$U_\Gamma(x,y) = P\exp\left(-ig\int_x^y ds\,\dot{z}^\mu(s)\,T_a A_\mu^a(z)\right)\ . \tag{7.86}$$

Here, P denotes path ordering, that is, $T_a A_\mu^a(z_1)$ is to the left of $T_a A_\mu^a(z_2)$ if, along the path Γ, z_1 is further along the path from x to y than z_2.

To elucidate the geometric meaning of $U_\Gamma(x,y)$, let us consider a color vector ψ at some point x and carry out its *parallel transport* to a close point $x+dx$. The variation of ψ, correct to first order of the infinitesimal displacement dx, is given by

$$\psi(x+dx) = \left[1 - ig\,T_a A_\mu^a(x)dx^\mu\right]\psi(x)\ . \tag{7.87}$$

Here, $A_\mu^a(x)$ is regarded as a *linear connection*, and the appropriate spacetime dimension is ensured by the parameter g. One may unite spacetime and the color space attached to every spacetime point into a quantity called the *principal fiber bundle*. The linear mapping (7.87) links the fibers attached to x and $x+dx$. The differential equation

$$d\psi(x) = -ig\,T_a A_\mu^a(x)\psi(x)\,dx^\mu \tag{7.88}$$

can be integrated (Problem 7.1.7) to give (7.86). We see that $U_\Gamma(x,y)$ implements parallel transport between points x and y separated by a finite distance.

It is instructive to derive (Problem 7.3.1) the expression for $U_{\Delta x}(0,\Delta x)$ retaining terms of order $(\Delta x)^2$,

$$U_{\Delta x}(0,\Delta x) = 1 - ig\Delta x^\mu A_\mu - \frac{ig}{2}\,\Delta x^\mu \Delta x^\nu \left(\partial_\mu A_\nu - igA_\mu A_\nu\right) \ . \tag{7.89}$$

This expression will be employed below.

Let $\psi_i(x)$ be a matter field belonging to some irreducible representation of the gauge group. The index i refers to this representation and all other indices are suppressed. Under a local gauge transformation,

$$\psi_i(x) \to \Omega_i^{\ j}(x)\,\psi_j(x), \quad T_a A^a_\mu(x) \to \Omega(x)\left[T_a A^a_\mu(x) + \frac{i}{g}\,\partial_\mu\right]\Omega^\dagger(x) \ , \tag{7.90}$$

the phase factor transforms as

$$U_\Gamma(x,y) \to \Omega(x)\,U_\Gamma(x,y)\,\Omega^\dagger(y) \ , \tag{7.91}$$

which implies that

$$\psi^\dagger(x)\,U_\Gamma(x,y)\,\psi(y) \tag{7.92}$$

is gauge invariant. Therefore, the theory can be formulated in terms of gauge invariant quantities $\psi^\dagger(x)\psi(x)$, $\psi^\dagger(x)\,U_\Gamma(x,y)\,\psi(y)$, and also $\mathrm{tr}\,[U_\mathcal{C}(x,x)]$ where $\mathcal{C}$ is a loop. The latter construction is usually written as

$$W(\mathcal{C}) = \mathrm{tr}\left[P\exp\left(-ig\oint_\mathcal{C} dz^\mu\,A_\mu\right)\right] \ , \tag{7.93}$$

and called *Wilson loop.*

We now determine $W(\mathcal{C})$ for an infinitesimal loop $\mathcal{C}$. For simplicity, we take $\mathcal{C}$ to be the border of a parallelogram with the sides Δx^μ and Δy^ν:

$$U_{\Delta x}(0,\Delta x)\,U_{\Delta y}(\Delta x,\Delta x+\Delta y)\,U_{-\Delta x}(\Delta x+\Delta y,\Delta y)\,U_{-\Delta y}(\Delta y,0) \ . \tag{7.94}$$

By (7.89),

$$U_{\Delta x}(0,\Delta x) = 1 - ig\,\Delta x^\mu A_\mu - \frac{ig}{2}\,\Delta x^\mu \Delta x^\nu \partial_\mu A_\nu - \frac{g^2}{2}\,\Delta x^\mu \Delta x^\nu A_\mu A_\nu, \tag{7.95}$$

$$\begin{aligned} U_{\Delta y}(\Delta x,\Delta x+\Delta y) &= 1-ig\,\Delta y^\mu A_\mu(\Delta x) - \frac{ig}{2}\,\Delta y^\mu \Delta y^\nu \partial_\mu A_\nu - \frac{g^2}{2}\,\Delta y^\mu \Delta y^\nu A_\mu A_\nu \\ &= 1-ig\,\Delta y^\mu A_\mu - \frac{ig}{2}\left(\Delta y^\mu \Delta y^\nu + 2\Delta x^\mu \Delta y^\nu\right)\partial_\mu A_\nu - \frac{g^2}{2}\,\Delta y^\mu \Delta y^\nu A_\mu A_\nu, \end{aligned} \tag{7.96}$$

$$\begin{aligned} U_{-\Delta x}(\Delta x+\Delta y,\Delta y) &= 1 + ig\,\Delta x^\mu A_\mu(\Delta x+\Delta y) \\ &\quad - \frac{ig}{2}\,\Delta x^\mu \Delta x^\nu \partial_\mu A_\nu - \frac{g^2}{2}\,\Delta x^\mu \Delta x^\nu A_\mu A_\nu \end{aligned}$$

$$= 1 + ig\,\Delta x^\mu A_\mu + \frac{ig}{2}\left(\Delta x^\mu \Delta x^\nu + 2\Delta x^\nu \Delta y^\mu\right)\partial_\mu A_\nu - \frac{g^2}{2}\,\Delta x^\mu \Delta x^\nu A_\mu A_\nu, \tag{7.97}$$

$$U_{-\Delta y}(\Delta y, 0) = 1 + ig\,\Delta y^\mu A_\mu(\Delta y) - \frac{ig}{2}\,\Delta y^\mu \Delta y^\nu \partial_\mu A_\nu - \frac{g^2}{2}\,\Delta y^\mu \Delta y^\nu A_\mu A_\nu$$

$$= 1 + ig\,\Delta y^\mu A_\mu + \frac{ig}{2}\,\Delta y^\mu \Delta y^\nu \partial_\mu A_\nu - \frac{g^2}{2}\,\Delta y^\mu \Delta y^\nu A_\mu A_\nu, \tag{7.98}$$

where A_μ stands for $A_\mu(0)$. Combining these expressions, we obtain

$$U_{\Delta x}(0, \Delta x)\, U_{\Delta y}(\Delta x, \Delta x + \Delta y)\, U_{-\Delta x}(\Delta x + \Delta y, \Delta y)\, U_{-\Delta y}(\Delta y, 0)$$

$$= 1 - ig\Delta x^\mu \Delta y^\nu G_{\mu\nu} + \cdots \tag{7.99}$$

This is actually just the first two terms in the expansion of an exponential. By extending the analysis to quartic order one can show that

$$W(\mathcal{C}) = \mathrm{tr}\left[\exp\left(-ig\Delta x^\mu \Delta x^\nu G_{\mu\nu}\right)\right]\ . \tag{7.100}$$

The net effect of parallel transport along an infinitesimal loop $\mathcal{C}$ is due to the field strength $G_{\mu\nu}$. Geometrically, $G_{\mu\nu}$ represents *curvature* on the principal fiber bundle which is expressed in terms of the connection A_μ by equation (7.22).

The need for an integral formulation of gauge theories becomes evident when we ask: is it possible to discretize gauge theories in such a way as to respect gauge invariance? In 1974 Kenneth Wilson gave an affirmative answer to this question. In fact, he developed a general framework for gauge theories on *Euclidean* spacetime lattices.

We now briefly describe the Wilson formulation of lattice gauge theories using the SU(N) color group as an example. To translate a continuum field theory onto a lattice we introduce a discrete grid with lattice spacing ℓ. The hypercubic lattice is labelled by a four-dimensional vector with integer components $n_\mu = (n_1, n_3, n_2, n_4)$. The continuum of field variables is replaced by a denumerable set. By enclosing the system into a box of size L, we achieve a finite number of variables. In fact, the initial quantum field system is approximated by the Gibbs ensemble of identical classical systems with $N \sim \mathcal{N}^2 (L/\ell)^4$ degrees of freedom. Large computers are now common for numerically approximating this four-dimensional statistical mechanics. Then the numerical results are extrapolated to zero spacing $\ell \to 0$, and interpreted in terms of the initial quantum field theory.

Our starting point is the manifestly gauge invariant expression (7.90). Consider first $U_\Gamma(x,y)$. The path Γ is approximated by a sequence of straight line segments joining adjacent sites. Associated with every directed bond connecting a site n with the next site $n + \hat{\mu}$ in the μth direction is an elementary phase factor

$$\mathcal{U}(n, n+\hat{\mu}) = \exp\left[-ig\ell\, T_a A^a_\mu(n)\right]\ , \tag{7.101}$$

which acts as a piece of colored string with a constant gauge field flux throughout it. There are no degrees of freedom for motion along this string from n to $n+\hat{\mu}$. The inverse of $\mathcal{U}(n,n+\hat{\mu})$ is the oppositely directed link:

$$\mathcal{U}^{-1}(n+\hat{\mu},n)=\mathcal{U}(n,n+\hat{\mu}) \ . \tag{7.102}$$

The phase factor associated with a particular discretized path is the P-ordered product of such elementary phase factors along the path. We may define the elementary Wilson loop around a 'plaquette' p (that is, around an elementary parallelogram with bonds connecting nearest neighbor nodes):

$$W(\mathrm{p})=\mathrm{tr}\left[\mathcal{U}(n,\hat{\mu})\,\mathcal{U}(n+\hat{\mu},n+\hat{\mu}+\hat{\nu})\,\mathcal{U}(n+\hat{\mu}+\hat{\nu},n+\hat{\nu})\,\mathcal{U}(n+\hat{\nu},n)\right] \ . \tag{7.103}$$

By (7.99),

$$W(\mathrm{p})=\mathrm{tr}\left\{\exp\left[-ig\ell^2\,G_{\mu\nu}(n)\right]\right\} \ , \tag{7.104}$$

where

$$G_{\mu\nu}(n)=\partial_\mu A_\nu(n)-\partial_\mu A_\nu(n)-ig\left[A_\mu(n),A_\nu(n)\right] , \tag{7.105}$$

$$\partial_\mu A_\nu(n)=\frac{1}{\ell}\left[A_\nu(n+\hat{\mu})-A_\nu(n)\right] \ . \tag{7.106}$$

Now a lattice version of the action for the gluon field may be written as

$$S=\frac{1}{8\pi g^2}\sum_n\sum_{\mathrm{p}}\left[W(\mathrm{p})-1\right] , \tag{7.107}$$

the summation is over all plaquettes and sites of the lattice. Indeed, in the limit $\ell\to 0$,

$$S\approx-\frac{1}{8\pi g^2}\sum_n\sum_{\mathrm{p}}\frac{\ell^4}{2}\,g^2\,\mathrm{tr}\left(G_{\mu\nu}G^{\mu\nu}\right)\to-\frac{1}{16\pi}\int d^4x\,\mathrm{tr}\left(G_{\mu\nu}G^{\mu\nu}\right) , \tag{7.108}$$

where use has been made of the fact that $\mathrm{tr}\left(T_a\right)=0$.

Quark degrees of freedom are described by discrete variables $\psi(n)$ defined on each lattice site n. To maintain the properties of the Dirac action under Hermitian conjugation, the derivative $\partial_\mu\psi$ is approximated by a symmetric finite difference

$$\Delta_\mu\psi(n)=\frac{1}{2\ell}\left[\psi(n+\hat{\mu})-\psi(n-\hat{\mu})\right] \ . \tag{7.109}$$

A simple version of the lattice action for the free quark field is

$$S=\ell^4\sum_n\left\{\frac{1}{2\ell}\sum_{\mu=1}^{4}\bar{\psi}(n)\,\gamma_\mu\left[\psi(n+\hat{\mu})-\psi(n-\hat{\mu})\right]-m\,\bar{\psi}(n)\,\psi(n)\right\} , \tag{7.110}$$

where γ_μ represents Euclidean Dirac matrices satisfying the anticommutation relations

$$\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu = 2\delta_{\mu\nu} \ . \tag{7.111}$$

The gauge invariant interaction between quarks and gluons is obtained by replacing $\bar{\psi}(n)\,\psi(n+\hat{\mu})$ with $\bar{\psi}(n)\,U(n,n+\hat{\mu})\,\psi(n+\hat{\mu})$ in (7.110), which is in any case necessary for gauge invariance,

$$S = \sum_n \left\{ \sum_{\mu=1}^{4} \frac{\ell^3}{2} \left[\bar{\psi}(n)\gamma_\mu U(n,n+\hat{\mu})\psi(n+\hat{\mu}) - \bar{\psi}(n)\gamma_\mu U(n,n-\hat{\mu})\psi(n-\hat{\mu}) \right] \right.$$

$$\left. - \, m\,\ell^4 \bar{\psi}(n)\,\psi(n) \right\} . \tag{7.112}$$

Combining (7.107) and (7.112), we come to a manifestly gauge invariant lattice theory.

However, the action (7.110) does not quite work. The problem is that action (7.110) possesses new and spurious degrees of freedom (the so-called *species doubling* problem, for which see Problem 7.3.2).

Problem 7.3.1. Prove (7.89).

Proof To solve the equation

$$\frac{d\psi}{d\lambda} = A\psi \ , \tag{7.113}$$

let us expand A and ψ as power series in λ:

$$\psi = \psi_0 + \lambda\psi_1 + \frac{\lambda^2}{2}\psi_2 + \cdots , \quad A = A_0 + \lambda A_1 + \cdots . \tag{7.114}$$

We have

$$\psi_1 + \lambda\psi_2 = (A_0 + \lambda A_1)(\psi_0 + \lambda\psi_1) \tag{7.115}$$

which implies

$$\psi_1 = A_0\psi_0, \quad \psi_2 = A_1\psi_0 + A_0\psi_1 = (A_1 + A_0A_0)\psi_0. \tag{7.116}$$

Substituting this in (7.113), we come to the equation

$$\frac{d\psi}{d\lambda} = [A_0 + \lambda(A_1 + A_0A_0)]\psi_0, \tag{7.117}$$

which is integrated to give

$$\psi = \psi_0 + \left[A_0\lambda + \frac{\lambda^2}{2}(A_1 + A_0A_0) \right] \psi_0 \ . \tag{7.118}$$

With the identifications

$$A_0 = -ig\, v^\mu A_\mu, \quad \lambda A_0 = -ig\, \Delta x^\mu A_\mu, \quad \lambda^2 A_0A_0 = -g^2 \Delta x^\mu \Delta x^\nu A_\mu A_\nu,$$

$$\lambda A_1 = -ig\, v^\mu \Delta x^\nu \partial_\mu A_\nu , \quad \lambda^2 A_1 = -ig\, \Delta x^\mu \Delta x^\nu \partial_\mu A_\nu , \tag{7.119}$$

we easily read (7.89) in (7.118).

Problem 7.3.2. For simplicity, assume the quark mass to be zero. Show that the lattice action for a free quark field (7.110) has a 16-fold degeneracy of quark degrees of freedom.

Hint The finite-difference equation of motion following from (7.110) is

$$\gamma_\mu \left[\psi(n+\hat{\mu}) - \psi(n-\hat{\mu}) \right] = 0 \, . \tag{7.120}$$

The Fourier transform

$$\psi(n) = \frac{1}{(2\pi)^4} \int d^4k \, \exp\left[-i(k \cdot n)\ell \right] \tilde{\psi}(k) \, , \tag{7.121}$$

where the integration is over the *Brillouin zone* $-\frac{\pi}{\ell} \leq k_\mu \leq \frac{\pi}{\ell}$, diagonalizes the action

$$S = \frac{1}{(2\pi)^4} \int d^4k \, \tilde{\psi}(-k) \, \Lambda(k) \, \psi(k), \quad \Lambda(k) = i \sum_{\mu=1}^{4} \gamma_\mu \frac{\sin(\ell k_\mu)}{\ell} \, . \tag{7.122}$$

As $\ell \to 0$ with k_μ fixed, $\Lambda(k) \to i(\gamma \cdot k)$ at $k_\mu = (0,0,0,0)$, which is the conventional limit. However, $\Lambda(k)$ has in addition 15 excess limits at $k_\mu = (\pi,0,0,0)$, $(0,\pi,0,0)$, $(0,0,\pi,0)$, $(0,0,0,\pi)$, $(\pi,\pi,0,0)$, $(\pi,0,\pi,0)$, $(\pi,0,0,\pi)$, $(0,\pi,\pi,0)$, $(0,\pi,0,\pi)$, $(0,0,\pi,\pi)$, $(\pi,\pi,\pi,0)$, $(\pi,\pi,0,\pi)$, $(\pi,0,\pi,\pi)$, $(0,\pi,\pi,\pi)$, (π,π,π,π).

Problem 7.3.3. Construct a lattice version of the action for a complex Klein–Gordon field. Show that this action presents no species doubling problem.

Answer

$$S = \ell^4 \sum_n \left\{ \frac{1}{\ell^2} \sum_{\mu=1}^{4} \phi^*(n) \left[\phi(n+\hat{\mu}) - 2\phi(n) + \phi(n-\hat{\mu}) \right] + M^2 \phi^*(n) \, \phi(n) \right\} , \tag{7.123}$$

Putting $M = 0$, and using the Fourier transform of $\phi(n)$,

$$S = \frac{1}{(2\pi)^4} \int d^4k \, \phi^*(-k) \, \Lambda(k) \, \phi(k), \quad \Lambda(k) = \frac{4}{\ell^2} \sum_{\mu=1}^{4} \sin^2 \left(\frac{\ell k_\mu}{2} \right) . \tag{7.124}$$

Within the Brillouin zone $-\frac{\pi}{\ell} \leq k_\mu \leq \frac{\pi}{\ell}$, $\Lambda(k)$ goes to a single limit k^2 as $\ell \to 0$, which implies that only the usual continuum action for a free massless scalar field is regained.

Notes

1. The theory of gauge fields is set out in many textbooks. Some which are different in subject matter and mathematical level are: De Witt (1965), Faddeev & Slavnov (1980), Konopleva & Popov (1981), Schwarz (1991), Weinberg (1996), Siegel (1999), and Rubakov (2002). The volume edited by 't Hooft (2004) is a panorama of current Yang–Mills theory.

2. *Section 7.1.* Yang & Mills (1954), and Shaw (1955) extended the Weyl principle of local gauge invariance to non-Abelian gauge groups. Klein (1938) was close to the discovery of non-Abelian gauge fields. Wong (1970) constructed classical equations of motion for a point particle with isospin degrees of freedominteracting with the SU(2) gauge field, analogous to the Maxwell–Lorentz theory. Heinz (1984) derived the equation of motion for a colored spinning particle. Balachandran, Borchardt & Stern (1978) proposed the action governing Wong particles. Balachandran, Marmo, Skagerstam & Stern (1982) is a review of fiber bundle geometry in gauge theories.

3. *Section 7.2.* It is a central tenet of high energy physics that the three fundamental forces – strong, electromagnetic and weak, owe their origin to local gauge symmetries and are mediated by the exchange of gauge fields. We have given a very brief account of these forces in the framework of the standard model omitting discussion of renormalizability, anomalies, asymptotic freedom, etc., because these issues are beyond the scope of the book. There are many special-purpose books which can help to fill the gap. Three worthy of mention are: Cheng & Li (1984), Mohapatra (1986), and Weinberg (1996). Veltman (2003) offers an insight in subnuclear physics on a popular level.

Nambu & Jona-Lasinio (1961) suggested that Dirac fields, corresponding to baryons, can gain their masses from spontaneous symmetry breaking. The Higgs mechanism was discovered by Englert & Brout (1964), Guralnik, Hagen & Kibble (1964), and Higgs (1964, 1966). A model of the weak interaction with three intermediate vector bosons $W^{\pm}_{\mu}$ and Z_{μ} was studied by Glashow (1961). The group theoretical aspects of this model were worked out by Salam & Ward (1964). Weinberg (1967), and Salam (1968) introduced vector bosons as $\mathrm{SU}(2)\times\mathrm{U}(1)$ gauge fields and applied the Higgs mechanism for generating their masses. 't Hooft & Veltman (1972) proved this model to be renormalizable. For an informal (intuitive) treatment of the renormalizability condition see Kosyakov (2001).

Gell-Mann (1964), and Zweig (1964) assumed that a fundamental SU(3)-flavor triplet of quarks, u, d, s, particles with spin $\frac{1}{2}$ and fractional values of the electric charge, are the fundamental constituents from which baryons and mesons are built up. For quark search experiments see Lions (1985). Greenberg (1964), Han & Nambu (1965), and Bogoliubov, Struminsky & Tavkhelidze (1965) introduced the SU(3)-color group as an internal symmetry group underlying the quark model. Fritzsch, Gell-Mann & Leutweyler (1973), and Weinberg (1973) formulated a SU(3)-color gauge model of strong interactions between quark and gluon fields, which was given the name quantum chromodynamics. Gross & Wilczek (1973), and Politzer (1973) showed that quantum chromodynamics offers the property of asymptotic freedom.

4. *Section 7.3.* Mandelstam (1962) used the phase factor as the basis for electrodynamics. Yang (1974) extended this approach to non-Abelian gauge theories. Wu & Yang (1975) recognized that the geometry of principal fiber bundles is merely the mathematical name of the Yang–Mills theory. For differential-geometric aspects of gauge theories see Eguchi, Gilkey & Hanson (1980), Schwarz (1991), and Dubrovin, Fomenko & Novikov (1992).

Wick (1954) showed that calculations in quantum field theory can be simplified if the wave function is analytically continued to imaginary values of t, which is tantamount to formulating quantum physics on Euclidean space-time. Schwinger (1959) transcribed quantum electrodynamics to a Euclidean metric. For a review of Euclidean quantum field theory see Jaffe & Glimm (1987), and Bogoliubov, Logunov, Oksak & Todorov (1990).

Wilson (1974) put forward a gauge invariant version of quantum chromodynamics on Euclidean lattices. For an extended discussion of this topic and further references see Creutz (1983). Lepage (2005) is a review of the present state of the art.

8

Solutions to the Yang–Mills Equations

It was shown in Chap. 4 that two solutions to Maxwell's equations, the Coulomb field and plane wave, play a key role in electrodynamics. By contrast, the situation in the Yang–Mills theory falls short of this ideal. Although a considerable body of solutions to the Yang–Mills equations exists in the literature, their physical implications are yet to be completely determined. In this chapter we review some field configurations in gauge theories and discuss their properties with an emphasis on their application in the physics of the strong interactions. Our prime interest is with solutions to the Yang–Mills–Wong theory, which seem to be well understood but have not been adequately addressed in the literature.

Recall that the theory of strong interactions, the SU(3)-color gauge field theory of quarks interacting with gluons, is defined by the Lagrangian

$$\mathcal{L} = -\frac{1}{16\pi}\, G_a^{\mu\nu} G^a_{\mu\nu} + \sum_f \left(i\,\bar{\psi}_f\,\gamma^\mu D_\mu\,\psi_f - m_f\,\bar{\psi}_f\,\psi_f \right) \ , \tag{8.1}$$

where f runs over the different flavors u, d, s, etc., and color indices of the quark field are suppressed. The quantized version of this theory is quantum chromodynamics. Quarks are assumed to be fundamental constituents of hadrons. However, all attempts to observe an isolated quark have failed. A plausible explanation is that quarks cannot escape from hadrons and do not exist as free particles. This hypothetical mechanism for keeping the quarks inside hadrons is called *confinement*. Furthermore, it is deemed that gluons are subject to the same constraint. They cannot be released from hadrons. Thus, by confinement is often meant a more general requirement: any object with color degrees of freedom must be coupled with other colored objects to make a color neutral combination, namely a singlet of the gauge group[1].

[1] In the Yang–Mills–Wong theory, which is a toy model of the full theory of strong interactions, the condition of color neutrality does not hold, which is a problem. The line of attack on this problem is open to speculation.

The most popular view of quark confinement is that a pair of a quark and an antiquark forming a meson are bounded by an attractive force whose strength does not decrease with distance. The force between the quark and antiquark is constant and the energy required to pull them apart increases linearly with their separation because the color Yang–Mills field generated by them is squeezed into a thin tube, or even into a string, joining these particles. The contraction of the field into a string with a constant tension owes its origin to the properties of the gluon vacuum determined from exact solutions of the classical Yang–Mills equations. This suggests the existence of a classical Yang–Mills field A_μ which rises linearly with distance. As is shown in Sects. 8.1, 8.3, and 8.4, this is indeed the case. Exact retarded solutions to the Yang–Mills equations with the source composed of several quarks involve such linearly rising terms, and, for some solutions, the linearly rising terms are concentrated in strings. However, we will see in Sect. 8.4 that the force due to this term is zero. Hence, the linearly rising terms of these classical Yang–Mills fields have nothing to do with the string model of quark confinement.

By now it is generally agreed that a phase transition from hadron matter to quark-gluon plasma must occur if either the temperature or the density approaches a critical point, estimated at $T_c = 200 \pm 50$ MeV, and $\rho_c = 1.5$–2.5 GeV/fm^3. This phenomenon is called *deconfinement*. It took place in the early universe at 10^{-6} s after the big bang, and may be achieved in the core of a collapsing neutron star or in laboratory experiments on relativistic heavy-ion collisions. It is reasonable to assume that the two phases are endowed with *different symmetries*. This would mean that deconfinement is a phase transition that changes not only the energy content but also the symmetry of subnuclear realm. In support of this conjecture we obtain in Sects. 8.1–8.4 two classes of exact solutions to the Yang–Mills–Wong theory which are distinguished by their groups of the gauge symmetry. In Sect. 8.5, we show that these solutions are stable against small field perturbations, and, in Sect. 8.7, we discuss the relevance of these Yang–Mills field configurations to the description of the two phases of subnuclear matter. In Sect. 8.6, we take a brief look at some exact solutions to Abelian and non-Abelian Higgs models, which are repeatedly surveyed in the literature (vortices and monopoles), to compare them with solutions to the Yang–Mills–Wong theory. However, Euclidean solutions to the pure Yang–Mills theory, such as *instantons* and *merons*, will be omitted in the present discussion since they have something to do with tunneling in Minkowski space, and their physical meaning can be fully understood only in the context of quantum field theory.

8.1 The Yang–Mills Field Generated by a Single Quark

We turn now to a study of retarded Yang–Mills fields due to colored point particles, hereafter referred to as *quarks*[2]. To begin with we consider the case that the Yang–Mills field is generated by a single quark moving along an arbitrary timelike smooth world line. We adopt the simplest non-Abelian gauge group SO(3). In the next section, we will show that the extension to larger gauge groups offers no significant changes in the main results.

When the gauge group is SO(3), the color space is a three-dimensional Euclidean space which is sometimes referred to as *isotopic-spin* space. Boldface characters will represent a triplet in this space. It is convenient to use a moving color basis spanned by $\mathbf{n}_1$, $\mathbf{n}_2$, and $\mathbf{n}_3$. The basis vector $\mathbf{n}_1$ is aligned with the quark color charge,

$$\mathbf{n}_1 = \frac{\mathbf{Q}}{|\mathbf{Q}|} \,, \tag{8.2}$$

while $\mathbf{n}_2$ and $\mathbf{n}_3$ are constrained by the conditions of orientability

$$\mathbf{n}_i \times \mathbf{n}_j = \epsilon_{ijk}\, \mathbf{n}_k \tag{8.3}$$

and orthogonality

$$\mathbf{n}_i \cdot \mathbf{n}_j = \frac{1}{2}\, \delta_{ij} \,. \tag{8.4}$$

The symbols $\times$ and $\cdot$ denote respectively the cross and scalar products of color vectors.

With these preliminaries, the Yang–Mills and Wong equations become

$$\partial_\mu \mathbf{G}^{\mu\nu} + g\mathbf{A}_\mu \times \mathbf{G}^{\mu\nu} = 4\pi g \int_{-\infty}^{\infty} ds\, \mathbf{Q}(s)\, v^\nu(s)\, \delta^4[x - z(s)] \,, \tag{8.5}$$

$$\dot{\mathbf{Q}} = -g\, v^\mu \mathbf{A}_\mu \times \mathbf{Q} \,, \tag{8.6}$$

where the field strength $\mathbf{G}_{\mu\nu}$ is expressed in terms of the vector potentials $\mathbf{A}_\mu$ as

$$\mathbf{G}_{\mu\nu} = \partial_\mu \mathbf{A}_\nu - \partial_\nu \mathbf{A}_\mu + g\mathbf{A}_\mu \times \mathbf{A}_\nu \,. \tag{8.7}$$

Equation (8.6) shows that the color charge of the source $\mathbf{Q}$ is a vector precessing around the color vector $v^\mu \mathbf{A}_\mu$ at the angular velocity $\Omega = g\,|v^\mu \mathbf{A}_\mu|$.

[2] This identification is rather conventional. Recall that the Yang–Mills–Wong particles are spinless objects. This does not necessarily mean that such particles are of little use as a model of real quarks involved in hadrons since measurements of the polarized proton structure in deep inelastic leptoproduction indicate that quarks carry only a small fraction of the spin of the nucleon. The next remark is that the color charge of a particle in the Yang–Mills–Wong theory transforms as the adjoint representation of the gauge group while the field ψ in (8.1) realizes the fundamental representation of the SU(3)-color group.

We now attempt to extend the ansatz (4.302), usefully employed in electrodynamics, to the single-quark case in Yang–Mills–Wong theory by assuming that

$$\mathbf{A}^{\mu} = \sum_{j=1}^{3} \mathbf{n}_j(s)\left[\Phi_j(\rho)\, v^{\mu}(s) + \Psi_j(\rho)\, R^{\mu}(s)\right] , \tag{8.8}$$

where s stands for the retarded instant when the signal was emitted by the quark.

We insert (8.8) in (8.5), and require that the coefficient of a^{μ} be zero off the world line. This gives three equations

$$\rho\,\Phi_j' + \Phi_j = 0 , \tag{8.9}$$

where the prime denotes the derivative with respect to ρ. They are readily integrated:

$$\Phi_j(\rho) = \frac{q_j}{\rho}, \quad q_j = \text{const} . \tag{8.10}$$

If one tries to substitute (8.8) and (8.10) into (8.6), then a divergence arises: the vector potential $\mathbf{A}_{\mu}$ is singular on the world line, which makes the color charge $\mathbf{Q}$ precess at an infinite angular velocity. This divergence cannot be absorbed into parameters entering the Wong equation (8.6) similar to that the bare mass coupled to the self-energy yields the renormalized mass in the equation of motion for a dressed particle. To see this, we recall that, apart from the dimensionless coupling constant g, there are no free parameters in (8.6). However, g is unsuitable for absorbing the term $1/\epsilon$, because, if this term is to be absorbed, it must be added to another divergent term having dimension -1.

This difficulty can be circumvented, if we take all constants of integration in (8.10) to be zero, with one exception,

$$q_2 = q_3 = 0, \quad q_1 \neq 0 , \tag{8.11}$$

and assume that $\Psi_j(\rho)$ is less singular than $\Phi_j(\rho)$, namely, $\rho\Psi_j \to 0$ as $\rho \to 0$. Then (8.6) becomes

$$\dot{\mathbf{n}}_1 = 0 . \tag{8.12}$$

We thus come to the dilemma: either $\mathbf{Q}$ precesses at infinite angular velocity, or we have a picture with the utter absence of precession. The former option seems beyond scope of physical intuition, and we adopt the latter one, $\dot{\mathbf{Q}} = 0$. Although $\mathbf{n}_1$ is fixed, two other basis vectors $\mathbf{n}_2$ and $\mathbf{n}_3$ might precess around $\mathbf{n}_1$. However, equations (8.5) and (8.6) contain no parameter of the appropriate dimension which is related to the angular velocity of this precession, and we are led to set

$$\dot{\mathbf{n}}_2 = \dot{\mathbf{n}}_3 = 0 . \tag{8.13}$$

We further define two color vectors

$$\mathbf{n}_{\pm} = \mathbf{n}_2 \pm i\mathbf{n}_3 \ , \tag{8.14}$$

which together with $\mathbf{n}_1$ span a fixed color basis. In view of (8.10), (8.11) and (8.14), the ansatz (8.8) reduces to

$$\mathbf{A}^{\mu} = q_1\, \mathbf{n}_1\, \frac{v^{\mu}}{\rho} + (\mathbf{n}_1\, \Psi_1 + \mathbf{n}_+\, \Psi_+ + \mathbf{n}_-\, \Psi_-)\, R^{\mu} \ . \tag{8.15}$$

We substitute (8.15) in (8.5) and equate to zero the coefficients of the color basis vectors $\mathbf{n}_1$ and $\mathbf{n}_{\pm}$, and those of the spacetime vectors v^{μ} and c^{μ}. In the latter case, we equate to zero separately the coefficient of $a \cdot c$ and the sum of remaining terms of the coefficient of c^{μ}. Introducing a new independent variable $\xi = \log \rho$ and denoting the derivative with respect to ξ by a prime, we find

$$\Psi_1'' + 3\Psi_1' + 2\Psi_1 = 0 \ , \tag{8.16}$$

$$\Psi_+'' + (3 - 2ig\, q_1)\, \Psi_+' + (2 - 3ig\, q_1 - g^2 q_1^2)\, \Psi_+ = 0 \ , \tag{8.17}$$

$$\Psi_1'' + \Psi_1' = 0 \ , \tag{8.18}$$

$$\Psi_+'' + (1 - ig\, q_1)\, \Psi_+' = 0 \ , \tag{8.19}$$

$$\Psi_+ \Psi_-' - \Psi_- \Psi_+' + 2ig\, q_1 \Psi_+ \Psi_- = 0 \ , \tag{8.20}$$

$$q_1 \left(\Psi_+' + 2\Psi_+\right) + e^{2\xi} \left(\Psi_+' \Psi_1 - \Psi_1' \Psi_+\right) - ig\, q_1 \left(q_1 + e^{2\xi}\, \Psi_1\right) \Psi_+ = 0 \ , \tag{8.21}$$

and three further equations derived from (8.17), (8.19), and (8.21) by complex conjugating and changing Ψ_+ for Ψ_-.

Now we are dealing with an *overdetermined* set of equations: 9 equations are used to determine 3 desired functions. It can be solved when all these equations are compatible, which is the case if integration constants are not arbitrary but take some *special* values.

Let us compare the system of equations (8.16)–(8.21) with the corresponding system of equations in electrodynamics (4.305)–(4.306). Both are overdetermined. However, equations (4.305) and (4.306) are linear, and their solution is given by equation (4.308) where C is an arbitrary integration constant. This constant serves to parametrize the family of solutions related by the gauge transformations (4.312). On the other hand, (8.16)–(8.21) represent a *nonlinear* and *overdetermined* set of equations, and unless they become linear for some reason, their solution is feasible only for exceptional integration constants.

Since equations (8.16)–(8.19) and their complex conjugate are linear, it is reasonable to look for a simultaneous solution to these equations of the form

$$\Psi_1 \propto e^{\lambda_1 \xi}, \quad \Psi_+ \propto e^{\lambda_+ \xi}, \quad \Psi_- = (\Psi_+)^* \ . \tag{8.22}$$

We find $\lambda_1 = -2$ or $\lambda_1 = -1$ from (8.16), and $\lambda_1 = 0$ or $\lambda_1 = -1$ from (8.18). Thus, (8.16) and (8.18) are compatible if

$$\Psi_1 = \eta_1\, e^{-\xi} \ . \tag{8.23}$$

We next obtain $\lambda_+ = -2 + igq_1$ or $\lambda_+ = -1 + igq_1$ from (8.17), and $\lambda_+ = 0$ or $\lambda_+ = -1 + igq_1$ from (8.19). Thus (8.17) and (8.19) are compatible if $\lambda_+ = -1 + igq_1$. Compatibility is also attained for $\lambda_+ = 0$ when $q_1 = -2i/g$ or $q_1 = -i/g$.

Let us examine the compatibility of (8.17) and (8.19) with (8.20). We first suppose that $\Psi_+\Psi_- \neq 0$. It follows from (8.20) that

$$\lambda_- - \lambda_+ + 2igq_1 = 0\ . \tag{8.24}$$

This equation is satisfied identically for $\lambda_+ = (\lambda_-)^* = -1 + igq_1$, but there is no solution for $\lambda_+ = \lambda_- = 0$. The compatibility of (8.17), (8.19), and (8.20) can also be established for $\lambda_+ = \lambda_- = 0$ provided that

$$\Psi_+\,\Psi_- = 0\ . \tag{8.25}$$

We consider finally the compatibility of (8.21) with (8.16)–(8.20). Taking $\lambda_+ = -1 + igq_1$, in combination with (8.23), we obtain $q_1\Psi_+ = 0$, while the complex conjugate equation yields $q_1\Psi_- = 0$. This implies either $q_1 = 0$, which converts the potential (8.15) to the form

$$\mathbf{A}^\mu = \mathbf{a}\,\frac{R^\mu}{\rho} \tag{8.26}$$

where $\mathbf{a}$ is a constant vector in the color space, or $\Psi_+ = \Psi_- = 0$, which results in

$$\mathbf{A}^\mu = q_1\,\mathbf{n}_1\,\frac{v^\mu}{\rho} + \eta_1\,\mathbf{n}_1\,\frac{R^\mu}{\rho}\ . \tag{8.27}$$

Recall that $R^\mu/\rho = \partial^\mu s$, hence the vector potential (8.26) as well as the second term of (8.27) are pure gauge.

For $\lambda_+ = \lambda_- = 0$, (8.21) becomes

$$\left[q_1\,(2 - igq_1) + (1 - igq_1)\,e^{2\xi}\,\Psi_1\right]\Psi_+ = 0\ . \tag{8.28}$$

Let $q_1 = -2i/g$, then (8.28) gives $\Psi_+\Psi_1 = 0$, while the complex conjugate equation is $\Psi_-\,\Psi_1 = 0$. This offers two possibilities. First, $\Psi_+ = \Psi_- = 0$, which results in $\mathbf{A}^\mu$ of the form (8.27). Second, $\Psi_1 = 0$, which is allowable for $\eta_1 = 0$. Since $\lambda_+ = 0$, we conclude from (8.22) that $\Psi_+ = \eta_+ =$ const.

For $q_1 = -i/g$, the solution to (8.28) is $\Psi_+ = 0$. With the corresponding result for the complex conjugate equation, $\Psi_- = 0$, we return to $\mathbf{A}^\mu$ of the form (8.27).

To summarize, all the equations are compatible if (8.25) holds. For $\Psi_+ = \Psi_- = 0$, there is no constraint on the parameter q_1, hence

$$\mathbf{A}^\mu = q\,\mathbf{n}_1\,\frac{v^\mu}{\rho}\ , \tag{8.29}$$

where q is an arbitrary constant. However, if we assume that only Ψ_- (or else only Ψ_+) is zero, then q_1 is found to be $-2i/g$ (or else $2i/g$), and so

$$\mathbf{A}^{\mu} = \mp \frac{2i}{g}\,\mathbf{n}_1\,\frac{v^{\mu}}{\rho} + \eta\,\mathbf{n}_{\pm}\,R^{\mu}\,, \tag{8.30}$$

where η is an arbitrary nonzero constant.

We thus have *two types* of retarded solutions: (8.29) and (8.30). It is evident that the vector potential of the type (8.29) satisfies the relations

$$\mathbf{A}_{\mu} \times \mathbf{A}_{\nu} = 0, \quad \mathbf{A}_{\mu} \times \mathbf{G}^{\mu\nu} = 0\,. \tag{8.31}$$

More generally, if $\mathbf{A}_{\mu}$ is proportional to a fixed vector in color space, then the Yang–Mills equations linearize. The solution (8.29) describes an Abelian field of the Liénard–Wiechert type. All results obtained in Maxwell–Lorentz electrodynamics are therefore extended to the Yang–Mills–Wong theory with the only replacement $e^2 \to q^2$.

By contrast, (8.30) represents a non-Abelian field configuration for which the relations (8.31) do not hold. One can readily determine the field strength:

$$\mathbf{G}^{\mu\nu} = c^{\mu}\mathbf{W}^{\nu} - c^{\nu}\mathbf{W}^{\mu}\,, \tag{8.32}$$

where

$$\mathbf{W}^{\mu} = \mp \frac{2i}{g}\,\mathbf{n}_1\,\frac{V^{\mu}}{\rho^2} + \eta\,\mathbf{n}_{\pm}\,v^{\mu}\,, \tag{8.33}$$

and V^{μ} is the same as that given by (4.284),

$$V^{\mu} = v^{\mu} + \rho(\overset{u}{\perp}\,a)^{\mu}\,. \tag{8.34}$$

Let us compare the retarded electromagnetic field (4.214) and the retarded Yang–Mills field (8.30). There are several notable distinctions.

First, the vector potential (8.30) involves a *linearly rising* term. Note that this term is inherent in the non-Abelian context. It might seem that a similar term occurs as well in electrostatics. Indeed, $\nabla^2\mathbf{r} = 0$, and so

$$\phi = \frac{e}{r} + \mathbf{E}\cdot\mathbf{r}\,, \tag{8.35}$$

with $\mathbf{E} =$ const., satisfies the Poisson equation with a delta-function source. It is clear, however, that only the first term of ϕ derives from the delta-function source, while the second term gives the uniform electric field $\mathbf{E}$ inside an infinite capacitor with flat parallel oppositely charged plates. The retarded electromagnetic vector potential of a point charge A^{μ} does not contain a term proportional to R^{μ} because

$$(\Box\,\eta_{\mu\nu} - \partial_{\mu}\partial_{\nu})\,R^{\nu} = -\frac{2v_{\mu}}{\rho}\,, \tag{8.36}$$

and there is no source like the term on the right of (8.36) in Maxwell's equations.

We further note that the coefficient η of the linearly rising term of $\mathbf{A}^{\mu}$ has dimension -2, whence it follows that *conformal symmetry is violated.* One might suspect that the linearly rising term of the solution (8.30) is pure gauge. But this impression is wrong: this term contributes to the field strength, as is clear from (8.33). Conformal invariance is therefore broken not only for $\mathbf{A}^\mu$, but also for $\mathbf{G}^{\mu\nu}$. Nevertheless, this symmetry is restored for color singlets built from $\mathbf{A}^\mu$ and $\mathbf{G}^{\mu\nu}$ defined in (8.30) and (8.32)–(8.34) where a certain sign is kept fixed (Problem 8.1.3).

The next peculiarity is that the Yang–Mills equations can determine not only the field, but also the *color charge* that generates this field

$$\mathbf{Q} = \mp\frac{2i}{g}\,\mathbf{n}_1\,. \tag{8.37}$$

The second solution (8.30) admits only a single value for the magnitude of the color charge carried by the quark $|\mathbf{Q}| = 2/g$. Recall that the electric charge of any particle in the Maxwell–Lorentz electrodynamics may be quite arbitrary. The selection of a special magnitude for the color charge of the source stems from the nonlinearity of the Yang–Mills equations.

It is natural to call the part of $\mathbf{G}^{\mu\nu}$ disappearing at spatial infinity the generalized Liénard–Wiechert term. It can be shown (Problem 8.1.4) that the color charge (8.37) is proportional to the flux of the generalized Liénard–Wiechert term through any surface enclosing the quark. Note that the linearly rising term of $\mathbf{A}^\mu$ is unrelated to the color charge content of the quark generating this field.

From (8.32)–(8.34), one obtains the field invariants

$${}^*\mathbf{G}_{\mu\nu}\cdot\mathbf{G}^{\mu\nu} = 0, \quad \mathbf{G}_{\mu\nu}\cdot\mathbf{G}^{\mu\nu} = \frac{8}{g^2\rho^4}.$$

It immediately follows that the solution (8.30) describes a field of the 'magnetic' type.

As will be seen in Sect. 8.5, the solution (8.29) is stable against small perturbations for real-valued q. Throughout the following text, we will assume that q takes arbitrary real values. Hence, the solution (8.29) represents a retarded Yang–Mills field of the 'electric' type, which bears a close similarity to the Liénard–Wiechert field in electrodynamics.

We finally notice a nonanalytic dependence of the vector potential (8.30) on g at the point $g = 0$. This is a rather general feature of nonperturbative solutions of nonlinear field equations. Some physical implications of this fact will be discussed later.

It seems advisable to rewrite our results in terms of SU(2), the double covering of SO(3). The Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{8.38}$$

are common for constructing generators of SU(2). These matrices obey the relation

$$\sigma_a \sigma_b = i\epsilon_{abc}\sigma_c + \mathbf{1}\delta_{ab} \ . \tag{8.39}$$

It is convenient to represent the color charge $\mathbf{Q}$ by a diagonal matrix. The color basis vectors are then expressed in terms of the Pauli matrices: $\mathbf{n}_1 = \tau_3$, $\mathbf{n}_2 = \tau_2$, $\mathbf{n}_3 = \tau_1$, where $\tau_i = \sigma_i/2$. One can readily see that (8.39) involves both (8.3) and (8.4). If we denote $A_\mu = A^a_\mu \, \tau_a$, then (8.29) and (8.30) become

$$A^\mu = q\, \tau_3 \, \frac{v^\mu}{\rho} \ , \tag{8.40}$$

and

$$A^\mu = \mp \frac{2i}{g} \, \tau_3 \, \frac{v^\mu}{\rho} + i\kappa \left(\tau_1 \pm i\tau_2\right) R^\mu \ , \tag{8.41}$$

where q and κ are arbitrary (nonzero) real constants.

It may appear that the two signs in the solution (8.41) are attributable to opposite color charges $2i\tau_3/g$ and $-2i\tau_3/g$. If this were so, the particles with such color charges could reasonably be identified as a quark and an antiquark. However, the appearance of opposite color charges is deceptive in the single-quark case. The sign in (8.41) can be changed (Problem 8.1.5) by a gauge transformation,

$$A^\mu_{(+)} = \Omega \, A^\mu_{(-)} \, \Omega^\dagger, \quad \Omega = \exp\left(-i\pi\tau_1\right) \ . \tag{8.42}$$

By introducing an alternative matrix basis

$$\mathcal{T}_1 = \tau_1, \quad \mathcal{T}_2 = i\tau_2, \quad \mathcal{T}_3 = \tau_3 \ , \tag{8.43}$$

we convert the solution (8.41) to the form $A_\mu = \mathcal{A}^a_\mu \, \mathcal{T}_a$ where all coefficients $\mathcal{A}^a_\mu$ are pure imaginary. Elements of this basis obey the following commutation relations

$$[\mathcal{T}_1, \mathcal{T}_2] = -\mathcal{T}_3, \quad [\mathcal{T}_2, \mathcal{T}_3] = -\mathcal{T}_1, \quad [\mathcal{T}_3, \mathcal{T}_1] = \mathcal{T}_2 \ , \tag{8.44}$$

which underlie the $\mathrm{sl}(2,\mathbb{R})$ Lie algebra (Problem 8.1.6). Thus, the gauge group of the solution (8.41) is actually $\mathrm{SL}(2,\mathbb{R})$. On the other hand, the gauge group of the solution (8.40) is the initially chosen SU(2).

Likewise, the color basis $\{\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3\}$ may be replaced by a new basis

$$\mathcal{N}_1 = i\mathbf{n}_1, \quad \mathcal{N}_2 = \mathbf{n}_2, \quad \mathcal{N}_3 = i\mathbf{n}_3 \ . \tag{8.45}$$

This renders the solution (8.30) real-valued with respect to this new basis. The color space becomes a pseudoeuclidean space with the metric $\eta_{ab} = \mathrm{diag}\,(-1, 1, -1)$. The automorphism group of this space is SO(2, 1). On the other hand, the appropriate color space for the solution (8.29) is the Euclidean space whose automorphism group is SO(3).

Should we adopt the gauge group Sp(1), rather than SO(3) or SU(2), we would come to identical results owing to equivalence between the three complex Lie algebras

$$\mathrm{sp}(1,\mathbb{C}) \sim \mathrm{sl}(2,\mathbb{C}) \sim \mathrm{so}(3,\mathbb{C})\ , \tag{8.46}$$

their real compact forms

$$\mathrm{sp}(1) \sim \mathrm{su}(2) \sim \mathrm{so}(3)\ , \tag{8.47}$$

and their real noncompact forms

$$\mathrm{sp}(1,\mathbb{R}) \sim \mathrm{sl}(2,\mathbb{R}) \sim \mathrm{su}(1,1) \sim \mathrm{so}(2,1)\ . \tag{8.48}$$

Imagine for a little that only a single quark is in the universe. Then the system quark plus the Yang–Mills field exists in two phases which are distinguished by their groups of the gauge symmetry: SU(2) and $\mathrm{SL}(2,\mathbb{R})$. These phases will be conventionally referred to as *hot* and *cold*.

A similar situation is found in the Higgs model where the presence of two types of solutions implies the existence of two alternative phases. One phase, perfectly symmetric, is unstable and rearranges into another phase possessing only part of this symmetry. The symmetric phase becomes stable at elevated temperatures, so that the full symmetry is restored. Note, however, that this symmetry restoration can be achieved only in a more general temperature-dependent theoretical framework which includes the unbroken and broken solutions of the Higgs model.

Turning to SU(2) and $\mathrm{SL}(2,\mathbb{R})$, none of them is a subgroup of the other. Where do these groups of symmetry come from? Their origin bears no relation to spontaneous symmetry breakdown. SU(2) and $\mathrm{SL}(2,\mathbb{R})$ are the *compact and noncompact real forms* of the complex group $\mathrm{SL}(2,\mathbb{C})$. Invariance of the action under SU(2) automatically entails its invariance under the complexification of this group, $\mathrm{SL}(2,\mathbb{C})$. However, a complex-valued Yang–Mills field may seem troublesome, particularly where observable quantities, such as energy, were involved. Only real forms of $\mathrm{SL}(2,\mathbb{C})$ appear to be satisfactory as gauge groups. The emergence of a solution invariant under a real form of $\mathrm{SL}(2,\mathbb{C})$ different from the initial SU(2) is a phenomenon specific to the Yang–Mills–Wong theory. We will call it *spontaneous symmetry deformation*. It follows from this discussion that the cold phase differs from the hot phase not only in its symmetry aspect, but also in physical manifestations, say, producing the Yang–Mills field of 'magnetic' rather than 'electric' type.

Problem 8.1.1. Derive (8.9) and (8.16)–(8.21).

Problem 8.1.2. Find *advanced* non-Abelian solutions to the Yang–Mills equations (8.5).

Answer

$$\begin{aligned} \mathbf{G} &= c \wedge \mathbf{W}\ , \\ c^\mu &= -v^\mu + u^\mu\ , \\ \mathbf{W}^\mu &= \pm\,\frac{2i}{g}\,\mathbf{n}_1\,\frac{V^\mu}{\rho^2} + \eta\,\mathbf{n}_\pm\,v^\mu\ , \end{aligned}$$

$$V^\mu = v^\mu - \rho\,(\overset{u}{\perp} a)^\mu\,, \tag{8.49}$$

and all the kinematical variables refer to the advanced instant $s_{\rm adv}$.

Problem 8.1.3. Let the sign of (8.30) be fixed. Prove that any color singlet built from $\mathbf{A}^\mu$ and $\mathbf{G}^{\mu\nu}$ is free of the linearly rising term contribution and hence conformally invariant.

Hint Color vectors $\mathbf{n}_\pm$, defined in (8.14), are null vectors perpendicular to $\mathbf{n}_1$: $\mathbf{n}_\pm^2 = 0$, $\mathbf{n}_\pm \cdot \mathbf{n}_1 = 0$.

Problem 8.1.4. Show that the Gauss law, as applied to (8.30), takes the following form. Let an external source of the Yang–Mills field with the color charge $\mathbf{Q}$ be contained in a domain $\mathcal{V}$ of a spacelike hyperplane having the normal v^μ. Upon integration of (8.5) over $\mathcal{V}$, the flux of the generalized Liénard–Wiechert part of the field strength through the boundary of this domain $\partial\mathcal{V}$ proves to be $4\pi\mathbf{Q}$, other terms cancel out.

Problem 8.1.5. Prove (8.42).

Hint Verify first that $\exp\left(\frac{1}{2}i\pi\sigma_1\right) = i\sigma_1$.

Problem 8.1.6. Verify that the matrices $\mathcal{T}_i$ defined in (8.43) are generators of $\mathrm{SL}(2,\mathbb{R})$.

8.2 Ansatz

Knowing the retarded solution to Maxwell's equations with the source involving a single point charge

$$A^\mu = e\,\frac{v^\mu}{\rho}\,, \tag{8.50}$$

its extension to the N-charge case follows immediately:

$$A^\mu = \sum_{I=1}^{N} e_I\,\frac{v_I^\mu}{\rho_I}\,. \tag{8.51}$$

It is clear that allowing for such linear combinations of solutions with arbitrary real coefficients is tantamount to stating that electric charges e_I take arbitrary real values.

By contrast, the superposition principle does not apply to the Yang–Mills equations

$$\Box A^\mu - \partial^\mu\partial_\nu A^\nu - ig(\partial_\nu[A^\nu, A^\mu] + [A_\nu, \partial^\nu A^\mu - \partial^\mu A^\nu]) - g^2[A_\nu, [A^\nu, A^\mu]] = 4\pi j^\mu\,, \tag{8.52}$$

$$j^\mu(x) = \sum_{I=1}^{N}\int_{-\infty}^{\infty} ds_I\, Q_I(s_I)\, v_I^\mu(s_I)\,\delta^4\left[x - z_I(s_I)\right]\,, \tag{8.53}$$

unless they become Abelian and hence linearize. This suggests that the non-Abelian single-quark solutions (8.41) are prevented from superposing, and we are forced to solve the Yang–Mills equations (8.52)–(8.53) for each N individually.

Before proceeding to the discussion of the ansatz for non-Abelian solutions, let us give some attention to Abelian solutions. To be specific, we adopt $\mathrm{SU}(\mathcal{N})$ as the gauge group. Taking into account that the Yang–Mills equations (8.52) are covariant under the gauge transformations

$$A_\mu \to \Omega \left(A_\mu + \frac{i}{g}\, \partial_\mu \right) \Omega^\dagger, \quad j_\mu \to \Omega j_\mu \Omega^\dagger \,, \tag{8.54}$$

one can always find a unitary matrix Ω to diagonalize the Hermitian matrix j_μ. Since the Lie algebra $\mathrm{su}(\mathcal{N})$ is of rank $\mathcal{N}-1$, there exist $\mathcal{N}-1$ diagonal matrices H_a, forming a *Cartan subalgebra* of commuting matrices. If we set

$$Q_I = \sum_{i=a}^{\mathcal{N}-1} e_I^a \, H_a \,, \tag{8.55}$$

where e_I^a are arbitrary coefficients, we then find that a vector potential of the form

$$A^\mu = \sum_{I=1}^{N} Q_I \, \frac{v_I^\mu}{\rho_I};, \tag{8.56}$$

represents generic retarded Abelian solutions to the Yang–Mills equations (8.52)–(8.53).

The total color charge of the system of N quarks

$$Q = \int_\Sigma d\sigma^\mu j_\mu \tag{8.57}$$

is in general surface dependent, since j^μ is not a locally conserved current. It is convenient to use a locally adjusted hypersurface of integration, which was defined in Sect. 6.3. Then Q is the sum of color charges of quarks comprising this system,

$$Q = \sum_{I=1}^{N} Q_I \,. \tag{8.58}$$

Anticipating that the field is singular on world lines, which implies the problem of an infinitely rapid precession of the color charge, we will focus on the picture where the color charge of each quark is constant

$$Q_I = \mathrm{const} \,. \tag{8.59}$$

In this picture,

$$Q = \mathrm{const} \,. \tag{8.60}$$

The Green's function method is not appropriate to solve (8.52). Note, however, that coefficients of the second derivatives in the Yang–Mills equations are identical to the respective coefficients in Maxwell's equations. Therefore, the characteristic surface of equations (8.52) is the conventional light cone.

It remains to see how the Yang–Mills field can propagate along rays of the light cone. In contrast to electrodynamics where solutions with different boundary conditions can be obtained by adding solutions of the homogeneous wave equation, the boundary condition in the Yang–Mills theory must be enforced solution-by-solution. We assume that signals of the Yang–Mills field are retarded.

Assume that the retarded Yang–Mills field is generated by N quarks moving along arbitrary world lines $z_I(s_I)$. The null vector R_I^μ drawn from the emission point on the world line of the Ith quark $z_I^\mu(s_{\rm ret})$ to the observation point x^μ is $R_I^\mu = x^\mu - z_I^\mu(s_{\rm ret})$. From vectors R_I^μ and v_I^μ, the following retarded invariants can be built:

$$\begin{aligned} \rho_I &= R_I \cdot v_I \ , \\ \beta_{IJ} &= v_I \cdot (R_I - R_J) \ , \\ \gamma_{IJ} &= v_I \cdot v_J \ , \\ \Delta_{IJ} &= (R_I - R_J)^2 = -2\, R_I \cdot R_J \ . \end{aligned} \tag{8.61}$$

We will look for retarded solutions to (8.52)–(8.53) of the following form

$$A^\mu = \sum_{I=1}^{N} \sum_{a=1}^{\mathcal{N}} T_a \left(v_I^\mu\, \Phi_I^a + R_I^\mu\, \Psi_I^a \right) \ , \tag{8.62}$$

where T_a are generators of the $\mathcal{N}$-parameter gauge group involved, and the unknown functions Φ_I^a and Ψ_I^a are assumed to be functions of ρ_I, β_{IJ}, γ_{IJ}, and Δ_{IJ}. Note that (8.62) is a natural generalization of the ansatz (8.8) to the N-quark case.

We must insert (8.62) in (8.52) and perform differentiations using (4.237), (4.241) and

$$\begin{aligned} \partial^\mu \beta_{IJ} &= \left[a^I \cdot (R_I - R_J) - 1\right] c_I^\mu + \gamma_{IJ} c_J^\mu \ , \\ \partial^\mu \gamma_{IJ} &= (a_I \cdot v_J)\, c_I^\mu + (a_J \cdot v_I)\, c_J^\mu , \\ \partial^\mu \Delta_{IJ} &= -2\left(\beta_{IJ} c_I^\mu + \beta_{JI} c_J^\mu\right) . \end{aligned} \tag{8.63}$$

This gives expressions in which it is necessary to equate to zero coefficients of the linearly independent vectors c_I^μ, v_I^μ, and a_I^μ, as well as those of color basis elements T_a. Recall that we seek solutions of the Yang–Mills equations off the world lines where the formulas of differentiation (4.237), (4.241) and (8.63) are valid. If the procedure is to be self-consistent, we must separately equate to zero coefficients of every independent scalar kinematic quantity which cannot appear in Φ_I^a and Ψ_I^a, say, scalars containing a_μ^I and $\dot a_I^\mu$. This leads us to a nonlinear and overdetermined set of equations which are similar to equations

(8.9) and (8.16)–(8.21) but may prove more complicated. The algorithm of finding their solutions is essentially the same as that detailed in the previous section.

A feature of this procedure is that no gauge fixing condition is necessary. We thus arrive at an equivalence class of solutions A^μ related by gauge transformations rather than a particular solution. On the other hand, combining the condition that the color charge of every quark is constant (8.59) with the Wong equation

$$\dot{Q}_I = -ig\,[Q_I,\, v_I^\mu A_\mu] \ , \tag{8.64}$$

we have

$$[Q_I,\, v_I^\mu A_\mu] = 0 \ , \tag{8.65}$$

which presents an additional constraint on the form of the solution (8.62).

We emphasize that the ansatz (8.62) is based crucially on the following assumptions:

(i) spacetime has *four* dimensions;
(ii) the field dynamics is *gauge invariant*;
(iii) the signals are *retarded* (or *advanced*);
(iv) the world lines are *timelike* and *smooth*.

Problem 8.2.1. Verify the formulas of differentiation (8.63).

8.3 The Yang–Mills Field Generated by Two Quarks

To look for solutions to the Yang–Mills equations with a source composed of two quarks, one should proceed from the ansatz (8.62) and pursue the strategy that was outlined in the two previous sections. We do not discuss this procedure in detail in the hope that the careful reader will go through cumbersome calculations for himself or herself. Our concern here is with features of the Yang–Mills field which are exhibited by these solutions.

We now adopt SU(3), the minimal unitary group whereby the retarded field generated by two quarks can be constructed in the general case. The reason for this enlargement of the gauge group from SU(2) to SU(3) will be elucidated below.

A commonly employed basis of the Lie algebra su(3) is that spanned by eight Gell-Mann matrices λ_a. It is more convenient for our purposes, however, to use an overcomplete basis comprised of the nonet of 3×3 Hermitian traceless matrices, including three diagonal matrices

$$H_1 = \frac{1}{2}\left(\lambda_3 + \frac{\lambda_8}{\sqrt{3}}\right) = \frac{1}{3}\begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \ , \tag{8.66}$$

$$H_2 = -\frac{1}{2}\left(\lambda_3 - \frac{\lambda_8}{\sqrt{3}}\right) = \frac{1}{3}\begin{pmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix} , \tag{8.67}$$

$$H_3 = -\frac{\lambda_8}{\sqrt{3}} = \frac{1}{3}\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix} , \tag{8.68}$$

which add up to zero matrix,

$$\sum_{n=1}^{3} H_n = 0 , \tag{8.69}$$

and the six projection matrices

$$\begin{aligned} E_{12}^{+} &= \frac{1}{2}(\lambda_1 + i\lambda_2) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \\ E_{12}^{-} = E_{21}^{+} &= \frac{1}{2}(\lambda_1 - i\lambda_2) = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \end{aligned} \tag{8.70}$$

$$\begin{aligned} E_{13}^{+} &= \frac{1}{2}(\lambda_4 + i\lambda_5) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \\ E_{13}^{-} = E_{31}^{+} &= \frac{1}{2}(\lambda_4 - i\lambda_5) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} , \end{aligned} \tag{8.71}$$

$$\begin{aligned} E_{23}^{+} &= \frac{1}{2}(\lambda_6 + i\lambda_7) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} , \\ E_{23}^{-} = E_{32}^{+} &= \frac{1}{2}(\lambda_6 - i\lambda_7) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} . \end{aligned} \tag{8.72}$$

SU(3) is a Lie group of the order 8 and rank 2. Recall that the rank of a given group is the maximal number of independent commuting elements in the Lie algebra of this group. The maximal set of commuting generators of SU(3) can be composed of two diagonal Gell-Mann matrices λ_3 and λ_8, or, alternatively, three diagonal matrices H_1, H_2, and H_3, which are defined in (8.66)–(8.68) and satisfy the constraint (8.69). The vector space spanned by the H_i represents the Cartan subalgebra of the Lie algebra su(3). The rank of the group is also the maximum number of independent polynomial invariants which can be constructed out of the generators. For su(3), quadratic and cubic invariants are

$$\frac{3}{2}\left(H_1^2 + H_2^2 + H_3^2\right) = \frac{3}{4}\left(\lambda_3^2 + \lambda_8^2\right) = \mathbf{1}\ , \tag{8.73}$$

and

$$\frac{27}{2}\, H_1 H_2 H_3 = \frac{3\sqrt{3}}{8}\left(3\lambda_3^2\lambda_8 - \lambda_8^3\right) = \mathbf{1}\ , \tag{8.74}$$

where $\mathbf{1}$ is the 3×3 unit matrix.

Using the color basis (8.66)–(8.72), one can check (Problem 8.3.1) that

$$A_\mu^{(1)} = \mp\frac{2i}{g}\left(H_1\,\frac{v_\mu^1}{\rho_1} + H_2\,\frac{v_\mu^2}{\rho_2}\right) + i\kappa\,(E_{13}^{\pm}\,R_\mu^1 + E_{23}^{\pm}\,R_\mu^2)\,\delta(R_1\cdot R_2)\ , \tag{8.75}$$

$$A_\mu^{(2)} = \mp\frac{2i}{g}\left(H_3\,\frac{v_\mu^1}{\rho_1} + H_1\,\frac{v_\mu^2}{\rho_2}\right) + i\kappa\,(E_{32}^{\pm}\,R_\mu^1 + E_{12}^{\pm}\,R_\mu^2)\,\delta(R_1\cdot R_2)\ , \tag{8.76}$$

$$A_\mu^{(3)} = \mp\frac{2i}{g}\left(H_2\,\frac{v_\mu^1}{\rho_1} + H_3\,\frac{v_\mu^2}{\rho_2}\right) + i\kappa\,(E_{21}^{\pm}\,R_\mu^1 + E_{31}^{\pm}\,R_\mu^2)\,\delta(R_1\cdot R_2) \tag{8.77}$$

are retarded solutions to the Yang–Mills equations (8.52) with source composed of two quarks, subject to the constraint (8.65). It is possible to show (Problem 8.3.2) that these three solutions are related by the gauge transformations,

$$A_\mu^{(j)} = \Omega_{1j}\,A_\mu^{(1)}\,\Omega_{1j}^{\dagger}\ , \tag{8.78}$$

where

$$\Omega_{12} = \Omega_{13}^{-1} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad \Omega_{13} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \tag{8.79}$$

Therefore, each of solutions (8.75)–(8.77) represents the same Yang–Mills field. One of them, say, (8.75), might be taken alone, discarding two others.

With reference to Problem 8.3.3, one can conclude from (8.75) that the color charge of the Ith quark is

$$Q_I = \mp\frac{2i}{g}\,H_I\ . \tag{8.80}$$

Hence there are two two-quark systems with total color charge

$$Q = \frac{2i}{g}\,(H_1 + H_2) \tag{8.81}$$

and

$$Q = -\frac{2i}{g}\,(H_1 + H_2)\ . \tag{8.82}$$

Therein lies an important distinction between the single-quark and two-quark cases. Both complex conjugate solutions (8.41) describe a Yang–Mills field generated by the same source since one is converted to another by the gauge transformation (8.42). In contrast, it is impossible to convert a solution with a fixed sign (8.75) to the complex conjugate solution. The reason is that the field invariant

$$I_3 = \mathrm{tr}\,(G_{\lambda\mu}\,G^{\mu}{}_{\nu}\,G^{\nu\lambda}) \tag{8.83}$$

is nonzero for both solutions (8.75), and I_3 changes its sign under the complex conjugation. We thus have two different field configurations generated by two sources with opposite total color charges (8.81) and (8.82).

We now turn to the spacetime dependence of the vector potential (8.75). It is seen that A_μ is independent of β_{12} and γ_{12}, while Δ_{12} appears in $\delta(R_1\cdot R_2) = 2\,\delta(\Delta_{12})$.

Since R_1^μ and R_2^μ are lightlike vectors, the equation of the support $R_1\cdot R_2 = 0$ can be satisfied only if R_1^μ is aligned with R_2^μ. The linearly rising term of A_μ, multiplied by the factor $\delta(X_1\cdot X_2)$, is therefore concentrated on the surfaces defined by the families of rays

$$\begin{aligned} x^\mu &= z_1^\mu(s) + \theta(\sigma)\,n^\mu\,\sigma, \quad n^\mu = R_1^\mu - R_2^\mu, \quad n^0 > 0, \quad n^2 = 0\,, \\ x^\mu &= z_2^\mu(s) + \theta(\sigma)\,m^\mu\,\sigma, \quad m^\mu = R_2^\mu - R_1^\mu, \quad m^0 > 0, \quad m^2 = 0\,, \end{aligned} \tag{8.84}$$

parametrized by two parameters s and σ. Here, $\theta(\sigma)$ is the Heaviside step function. The linearly rising term disappears in the region between the world lines where R_1^μ and R_2^μ can not be parallel; this term is nonzero only in the region outside this two-quark system, as viewed in Fig. 8.1. The intersection of the two-dimensional surface defined in (8.84) with a spacelike hyperplane gives a curve. However, this is not a finite string joining the quarks. We have two half-infinite curves which begin at the quarks and go to spatial infinity, Fig. 8.2.

There exists an alternative solution which is identical to that of (8.75) in every respect except that the linearly rising terms do not have the factor $\delta(R^1\cdot R^2)$,

$$A_\mu = \mp\frac{2i}{g}\left(H_1\,\frac{v^1_\mu}{\rho_1} + g\,\kappa\,E^{\pm}_{13}\,R^1_\mu\right) \mp \frac{2i}{g}\left(H_2\,\frac{v^2_\mu}{\rho_2} + g\,\kappa\,E^{\pm}_{23}\,R^2_\mu\right)\,. \tag{8.85}$$

The linearly rising terms of A^μ describe lines of force distributed over all directions in space rather than being squeezed to a string.

The solution (8.85) is essentially non-Abelian in nature since

$$[A_\mu, A_\nu] \neq 0, \quad [A_\mu, G^{\mu\nu}] \neq 0\,. \tag{8.86}$$

How is the nonlinearity of the Yang–Mills equations compatible with the fact that A_μ is the sum of two single-quark potentials? Although two such terms

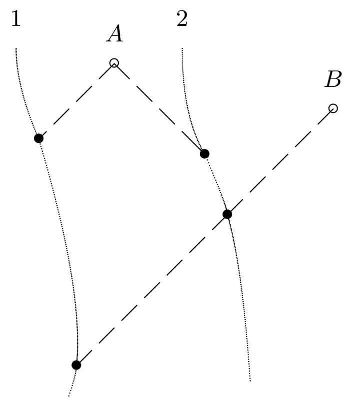


Fig. 8.1. Retarded signals arriving at a point in the interquark region A and at a point in the outer region B

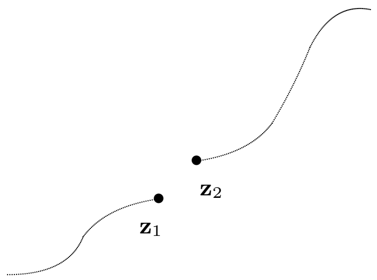


Fig. 8.2. The curves on which the linearly rising term is localized

are indeed combined in (8.85), one cannot build solution as an *arbitrary* superposition of these terms. Indeed, if either of them is multiplied by a coefficient different from 1 and added to another, no further solution arises.

The solution (8.85) describes the Yang–Mills field generated by two *bound* quarks. While on the subject of a field generated by two *free* quarks, the sign of the color charge of either quark may be chosen freely regardless of the sign of the color charge of the other quark. However, if we try to change the signs of the first term in (8.85), with the second term being fixed, we then find that the resulting expression is no longer solution. One can only change both signs simultaneously. This correlation of signs of one-quark constituents is what is generally meant by the term 'bound' in the Yang–Mills–Wong theory.

If $\kappa = 0$, then the retarded solution is a superposition of two single-quark potentials

$$A_\mu = \sum_{I=1}^{2} \sum_{n=1}^{3} q_I^n \, H_n \, \frac{v_\mu^I}{\rho_I} \;, \tag{8.87}$$

where q_I^n are arbitrary real parameters. This solution describes an Abelian Yang–Mills field generated by two free quarks.

The solutions (8.75)–(8.77) and (8.85) become imaginary-valued in the color basis

$$\begin{aligned} &\mathcal{T}_1 = \lambda_1, \quad \mathcal{T}_2 = i\lambda_2, \quad \mathcal{T}_3 = \lambda_3, \quad \mathcal{T}_4 = \lambda_4 \,, \\ &\mathcal{T}_5 = i\lambda_5, \quad \mathcal{T}_6 = \lambda_6, \quad \mathcal{T}_7 = i\lambda_7, \quad \mathcal{T}_8 = \lambda_8 \,, \end{aligned} \tag{8.88}$$

or in the basis spanned by H_n and $E^{\pm}_{mn}$. From the explicit form of H_n and $E^{\pm}_{mn}$, (8.66)–(8.72), it is clear that the $\mathcal{T}_n$ are traceless real 3×3 matrices satisfying the commutation relations of the Lie algebra $\mathrm{sl}(3, \mathbb{R})$. Thus, the gauge group of the solutions (8.75)–(8.77) and (8.85) is $\mathrm{SL}(3, \mathbb{R})$, while the gauge group of the solution (8.87) is SU(3).

Let us suppose that one of two quarks, say, quark 1, disappears. Then the solution (8.76) takes the form $A_\mu = A'_\mu + A''_\mu$ where

$$A'_\mu = \mp \frac{2i}{g} \frac{\lambda_3}{2} \frac{v_\mu}{\rho} + i\kappa\, E^{\pm}_{12}\, R_\mu, \quad A''_\mu = \mp \frac{i}{g} \frac{\lambda_8}{\sqrt{3}} \frac{v_\mu}{\rho} \,. \tag{8.89}$$

A'_μ is just the single-quark solution (8.41) while A''_μ is an Abelian term, detached from A'_μ, because λ_8 commutes with both λ_3 and $E^{\pm}_{12}$. The sufficiency of SU(2) in the single-quark case is thus confirmed; the non-Abelian piece of the solution is built out of color elements forming the Lie algebra su(2). Note that su(2) is of rank 1, and hence the cubic field invariant I_3 defined in (8.83) is zero for the single-quark solution (8.89).

The structure of (8.75)–(8.77) and (8.85) makes it clear that the Yang–Mills field due to a particular bound quark occupies individually a sort of 'elementary' $\mathrm{sl}(2, \mathbb{R})$ color cell. Neither of two such fields generated by different bound quarks can be contained in a given $\mathrm{sl}(2, \mathbb{R})$. The expert reader will find this to be similar to the Pauli blocking principle. Just as a cell of volume h^3 (where h is Planck's constant) in the phase space can be occupied by at most one fermion with a definite spin polarization, so any $\mathrm{sl}(2, \mathbb{R})$ cell is suited to the field of only one bound quark. Proceeding from $\mathrm{SO}(\mathcal{N})$ or $\mathrm{Sp}(\mathcal{N})$, rather than $\mathrm{SU}(\mathcal{N})$, one singles out the identical elementary color cell, because the Lie algebras $\mathrm{so}(2, 1)$, $\mathrm{sp}(1, \mathbb{R})$, and $\mathrm{sl}(2, \mathbb{R})$ are equivalent.

Any Lie group of the rank 2 and higher is equally good for playing the role of the gauge group of Yang–Mills fields generated by two bound quarks. Apart from SU(3), one might use three other rank-2 groups: SO(4), SO(5), and Sp(2). However, SO(4) is not semisimple, so that the Killing form is singular. On the other hand, $\mathrm{so}(5, \mathbb{C})$ is isomorphic to $\mathrm{sp}(2, \mathbb{C})$. These complex-valued Lie algebras are suitable to accommodate two elementary color cells. We thus have two alternatives for the minimal choice of the initial gauge group: SU(3) and $\mathrm{SO}(5) \sim \mathrm{Sp}(2)$.

Problem 8.3.1. Solve the Yang–Mills equations with the two-quark source, and show that (8.75)–(8.77) and (8.85) are the desired retarded solutions.

Problem 8.3.2. Show that the vector potentials $A^{(j)}_\mu$, $j = 1, 2, 3$, defined in (8.75)–(8.77) are related by gauge transformations.

Problem 8.3.3. Extend the Gauss law formulated in Problem 8.1.4 to the two-quark case using a locally adjusted hypersurface.

Problem 8.3.4. Verify that the cubic field invariant $\mathrm{tr}\,(G_{\lambda\mu}\,G^{\mu}{}_{\nu}\,G^{\nu\lambda})$ is nonzero and of different signs for complex conjugate fields defined in (8.75).

Problem 8.3.5. What kind of two-dimensional surface is defined in (8.84)?

Answer The warped surface formed by rays emanating in the n^μ and m^μ directions, respectively, from the edges $z_1^\mu(s)$ and $z_2^\mu(s)$.

Problem 8.3.6. Proceeding from the gauge group SO(5), find retarded solutions to the Yang–Mills equations with the two-quark source.

Hint Use the Cartan basis for generators of the gauge group.

8.4 The Yang–Mills Field Generated by *N* Quarks

The analysis of retarded solutions to the Yang–Mills equations with a source composed of N quarks resembles the two quark case in many ways. Anticipating that the field generated by each bound quark needs an individual elementary $\mathrm{sl}(2, \mathbb{C})$ color cell, we adopt the gauge group $\mathrm{SU}(\mathcal{N})$ with sufficiently large $\mathcal{N}$, at least $\mathcal{N} \ge N + 1$.

We first construct the *Cartan-Weyl basis* of the Lie algebra $\mathrm{su}(\mathcal{N})$. This basis consists of a set of $\mathcal{N}^2$ matrices, including $\mathcal{N}$ elements of Cartan's subalgebra

$$(H_n)_{AB} = \delta_{An}\,\delta_{Bn} - \mathcal{N}^{-1}\,\delta_{AB}\,, \tag{8.90}$$

which are related by

$$\sum_{n=1}^{\mathcal{N}} H_n = 0\,, \tag{8.91}$$

and $\mathcal{N}^2 - \mathcal{N}$ raising and lowering elements E^+_{mn} and E^-_{mn}

$$(E^+_{mn})_{AB} = \delta_{Am}\,\delta_{Bn}, \quad (E^-_{mn})_{AB} = \delta_{An}\,\delta_{Bm}\,, \tag{8.92}$$

where m, n, A, B run from 1 to $\mathcal{N}$, and the successive indices m and n are ordered such that $n > m$.

By the construction of the H_n's, these matrices are diagonal, and hence commuting. The only nontrivial commutation relations between H_m and $E^\pm_{mn}$ are

$$[H_m, E^\pm_{mn}] = \pm\, E^\pm_{mn}\,, \tag{8.93}$$

$$[E^+_{mn}, E^-_{mn}] = H_m - H_n\,, \tag{8.94}$$

$$[E^{\pm}_{kl},\, E^{\pm}_{lm}] = \pm\, E^{\pm}_{km}\ . \tag{8.95}$$

One may verify that the retarded Yang–Mills field

$$A_\mu = \mp\,\frac{2i}{g}\sum_{I=1}^{N}\left[H_I\,\frac{v^I_\mu}{\rho_I} + g\,\kappa\,E^{\pm}_{I\,N+1}\,R^I_\mu\prod_{J(\neq I)}^{N-1}\delta(R_J\cdot R_I)\right] \tag{8.96}$$

satisfies the Yang–Mills equations (8.52) with source composed of N quarks, and the constraint (8.65). There exist $C^N_{\mathcal{N}}$ solutions of this type. Indeed, the transformation ${A'}_\mu = \Omega A_\mu \Omega^\dagger$ with

$$\Omega = E^{+}_{1\,\mathcal{N}} + \sum_{I=1}^{\mathcal{N}-1} E^{-}_{I\,I+1} = \begin{pmatrix} 0 & 1 & & & & & \\ & 0 & 1 & & & & \\ & & \cdot & \cdot & & & \\ & & & \cdot & \cdot & & \\ & & & & \cdot & \cdot & \\ & & & & & 0 & 1 \\ 1 & 0 & \cdot & \cdot & \cdot & & 0 \end{pmatrix} \tag{8.97}$$

and

$$\Omega^\dagger = E^{-}_{1\,\mathcal{N}} + \sum_{I=1}^{\mathcal{N}-1} E^{+}_{I\,I+1} = \begin{pmatrix} 0 & \cdot & \cdot & \cdot & & 0 & 1 \\ 1 & 0 & & & & & \\ & 1 & 0 & & & & \\ & & \cdot & \cdot & & & \\ & & & \cdot & \cdot & & \\ & & & & \cdot & \cdot & \\ & & & & & & \end{pmatrix} \tag{8.98}$$

increases each index of H_I and $E^{\pm}_{I\,N+1}$ by one. The transformed field ${A'}_\mu$ is a further solution. All other solutions can be obtained by repetition of this procedure. A similar situation was encountered in the two-quark case where the solutions (8.75)–(8.77) are related by the gauge transformations of this sort.

The solution (8.96) describes a Yang–Mills field generated by N quarks that form a N-quark cluster. The color charge of the Ith quark is

$$Q_I = \mp\,\frac{2i}{g}\,H_I\ . \tag{8.99}$$

The total color charge of such a cluster is either

$$Q_{(+)} = \frac{2i}{g}\sum_{I=1}^{N} H_I \tag{8.100}$$

or

$$Q_{(-)} = -\frac{2i}{g}\sum_{I=1}^{N} H_I\ . \tag{8.101}$$

We see that every cluster has a nonzero color charge.

There are also solutions describing Yang–Mills fields generated by several clusters. A particular cluster is defined by the condition that the signs of the color charges are simultaneously either + or − for every quark of this cluster, whereas relative signs of total color charges of clusters held in this system are arbitrary. For example, the potential generated by two two-quark clusters is $A_\mu = A^1_\mu \pm A^2_\mu$ where A^j_μ is the potential generated by the jth cluster,

$$A^1_\mu = \pm \frac{2i}{g} \sum_{I=2}^{3} \left[H_I \, \frac{v^I_\mu}{\rho_I} + g\,\kappa\, E^{\pm}_{1\,I}\, R^I_\mu \,\delta(R_2 \cdot R_3) \right] ,$$

$$A^2_\mu = \pm \frac{2i}{g} \sum_{I=5}^{6} \left[H_I \, \frac{v^I_\mu}{\rho_I} + g\,\kappa\, E^{\pm}_{4\,I}\, R^I_\mu \,\delta(R_5 \cdot R_6) \right] . \tag{8.102}$$

Omitting the factor $\delta(R^I \cdot R^J)$ in (8.96) and (8.102) gives additional solutions.

To build the Yang–Mills field generated by free quarks, one should use color charges of the form

$$Q_I = \pm \frac{i}{g} \left(H_{I+1} - H_I \right) . \tag{8.103}$$

For example, the field of two free quarks, labeled by numbers 1 and 3, is

$$A_\mu = \pm i \left[\frac{i}{g} \left(H_2 - H_1 \right) \frac{v^1_\mu}{\rho_1} + \kappa\, E^{\pm}_{1\,2}\, R^1_\mu \right] \pm i \left[\frac{1}{g} \left(H_4 - H_3 \right) \frac{v^3_\mu}{\rho_3} + \kappa\, E^{\pm}_{3\,4}\, R^3_\mu \right] . \tag{8.104}$$

The gauge transformation $A'_\mu = \Omega A_\mu \Omega^\dagger$ with

$$\Omega = \frac{\mathcal{N}-2}{\mathcal{N}}\, \mathbf{1} + \sum_{I=3}^{\mathcal{N}} H_I + E^{+}_{1\,2} + E^{-}_{1\,2} = \begin{pmatrix} 0 & 1 & & & & & \\ 1 & 0 & & & & & \\ & & 1 & & & & \\ & & & 1 & & & \\ & & & & \cdot & & \\ & & & & & \cdot & \\ & & & & & & \cdot \end{pmatrix} \tag{8.105}$$

changes the ± signs of the first square bracket of (8.104) while the signs of the second square bracket remains invariant. It is easy to recognize this gauge transformation which complex conjugates the single-quark potential. We therefore see that the color charge of a free quark is determined modulo $\exp\,(i\pi n)$.

There are many ways to separate a given N-quark system into groups of clusters of a certain quark content and free quarks. We will call these separations *scenarios of hadronization*. The meaning of this term will be clarified in Sect. 8.7.

Let us look at the gauge symmetry of these solutions. One may define $\mathcal{N}^2$ traceless real matrices $\mathcal{H}_n$ and $\mathcal{E}^{\pm}_{mn}$ as follows:

$$\mathcal{H}_n = H_n, \quad \mathcal{E}^{\pm}_{mn} = E^{\pm}_{mn} \tag{8.106}$$

which are elements of the Lie algebra $\mathrm{sl}(\mathcal{N}, \mathbb{R})$. Thereafter, every solution given above becomes imaginary valued with respect to this basis. The solutions built from n^2 such elements, obeying a closed set of commutation relations, are invariant under $\mathrm{SL}(n, \mathbb{R})$.

In particular, the Yang–Mills field generated by a two-quark cluster (a toy meson) is invariant under $\mathrm{SL}(3, \mathbb{R})$ and that of a three-quark cluster (a toy baryon) is invariant under $\mathrm{SL}(4, \mathbb{R})$. Since $\mathrm{SL}(3, \mathbb{R})$ is a subgroup of $\mathrm{SL}(4, \mathbb{R})$, the Yang–Mills field of every toy hadron is determined by the gauge group $\mathrm{SL}(4, \mathbb{R})$. This symmetry is independent of the total number of colors $\mathcal{N}$ and survives in the limit $\mathcal{N} \to \infty$.

If $\kappa = 0$, then the Yang–Mills equations linearize, and we get an Abelian solution

$$A_\mu = \sum_{I=1}^{N} \sum_{n=1}^{\mathcal{N}} e_I^n \, H_n \, \frac{v_\mu^I}{\rho_I} \, , \tag{8.107}$$

where e_I^n are arbitrary parameters. The gauge group of this solution is $\mathrm{SU}(N)$.

We thus have two types of retarded solutions. The Yang–Mills fields described by solutions of the first type are invariant under the noncompact group $\mathrm{SL}(\mathcal{N}, \mathbb{R})$ or its subgroups, while those of the second type are invariant under the compact group $\mathrm{SU}(\mathcal{N})$.

We now turn to the color forces due to the Yang–Mills fields under study. It is clear from the expression for the Wong force acting on the Ith quark

$$f_I^\mu = v_\nu^I \, \mathrm{tr} \left[Q_I \, G^{\mu\nu}(z_I) \right] \, , \tag{8.108}$$

and the trace relations

$$\mathrm{tr} \, (H_l \, E^{\pm}_{mn}) = 0 \, , \tag{8.109}$$

valid for any l, m, n, that the linearly rising term of A_μ does not contribute to (8.108). Although it makes a nonzero contribution to the field strength $G_{\mu\nu}$, the linearly rising term of A_μ results in no force. The explanation of this result is simple. The Wong force (8.108) includes the scalar product of two color vectors $G_{\mu\nu}$ and Q_I. These vectors are not arbitrary; the exact solutions constrain them to be orthogonal to each other.

Combining (8.109) and

$$\mathrm{tr} \left(E^{\pm}_{mn} \, E^{\pm}_{mn} \right) = 0 \, , \tag{8.110}$$

one finds that the linearly rising term of A_μ does not contribute to color singlets. This is because the linearly rising term of A_μ depends upon either E^{+}_{mn} or upon E^{-}_{mn}, but not both. A more fundamental reason is that conformal invariance, which might be violated in particular field configurations due to the presence of the parameter κ having dimension -2, is unbroken in color singlets. To illustrate this, we refer to the Wong force (8.108) which is a color singlet, and hence is free of κ.

Thus, the string scheme for quark confinement finds no support from our study of classical solutions to the Yang–Mills equations.

Let us take a closer look at the interquark color forces in the cold phase. These forces are essentially of the Coulomb type. Because the color charge of the Ith bound quark is proportional to iH_I, like color charges attract and unlike ones repel. Consider these forces in the limit $\mathcal{N} \to \infty$, assuming the coupling g to be fixed. The trace relations

$$\mathrm{tr}\,(H_I)^2 = 1 - \mathcal{N}^{-1}, \quad \mathrm{tr}\,(H_I\,H_J) = -\mathcal{N}^{-1} \tag{8.111}$$

show that the color interaction between bound quarks vanishes in this limit, unless the number of quarks is of order $\mathcal{N}$.

On the other hand, it is evident from

$$\mathrm{tr}\,(H_{I+1} - H_I)^2 = 2, \quad \mathrm{tr}\,(H_{I+1} - H_I)\,H_J = 0 \tag{8.112}$$

that a free quark, while experiencing a self-interaction, does not act on other quarks.

We see that the Wong forces between quarks vanish in the limit $\mathcal{N} \to \infty$. The bound quarks are balanced in a state of neutral equilibrium[3]. This is consistent with the idea that the ground state of a hadron possesses zero orbital momentum[4]. The binding is determined by the correlation of signs of the color charges of quarks comprising a cluster.

Problem 8.4.1. With H_n and $E^{\pm}_{mn}$ defined in (8.90) and (8.92), derive commutation relations (8.93)–(8.95).

Problem 8.4.2. Verify that (8.96), (8.102), and (8.104) are exact retarded solutions to the Yang–Mills equations with appropriate sources.

Problem 8.4.3. Prove that $A'_\mu = \Omega A_\mu \Omega^\dagger$, where the unitary matrices Ω and $\Omega^\dagger$ are given respectively by (8.97) and (8.98), is a gauge transformation increasing each index of H_I and $E^{\pm}_{I\,N+1}$ of the Yang–Mills field A_μ defined in (8.96) by one.

Problem 8.4.4. Consider the unitary matrix $\Omega = \Omega^\dagger$ of relation (8.105). Prove that the gauge transformation $A'_\mu = \Omega A_\mu \Omega^\dagger$ of the vector potential

[3] If it is granted that quarks interact only through the mediation of these color Yang–Mills fields and are unaffected by other forces, then their motion in the cold phase is described by straight world lines.

[4] By contrast, in the string model, a string that ties quarks in a hadron is thought to possess a constant tension per unit length, and, therefore, the string would collapse if the quarks were stationary with respect to one another. The system can be kept in equilibrium by centrifugal force if it is made to rotate around the center of mass.

(8.104) changes only the $\pm$ signs of the first square bracket of (8.104) and leaves the second square bracket unchanged.

Problem 8.4.5. Verify (8.109)–(8.112).

Problem 8.4.6. Show that $\mathrm{sl}(4,\mathbb{R}) \sim \mathrm{so}(3,3)$. Proceeding from the gauge group Sp(3), find retarded solutions to the Yang–Mills equations with three-quark source.

Hint Use the Cartan-Weyl basis of generators for these gauge groups.

8.5 Stability

It was shown in the preceding section that quark clusters are in neutral equilibrium. Our next task is to examine whether or not the Yang–Mills field generated by these quarks A_μ is *stable*. By 'stability' one usually means that if initial data are given to be $B_\mu(0) = A_\mu(0) + b_\mu(0)$ where $b_\mu(0)$ is a small initial perturbation, then the subsequent time evolution keeps the solution near the background A_μ. More specifically, the stability requires that the equation of motion for b_μ, when linearized about A_μ, does not yield exponential growth in time. We may represent the time dependence of each mode by an exponential $\exp(i\omega t)$. If all modes oscillate harmonically with real ω's, then A_μ is stable. Complex ω's signal instability.

The equation of motion for b_μ is given by the second variation of the Yang–Mills action

$$\frac{\delta^2 S}{\delta A^a_\mu(x)\, \delta A^b_\nu(y)}\, b^a_\mu(x) = 0\ . \tag{8.113}$$

Since our prime interest is with the ground state of clusters where the quarks are at rest relative to each other, we consider the static background field A_μ generated by such quarks.

Let us begin with the Yang–Mills field due to a single quark in the cold phase, A_μ defined in (8.41). If we adopt the gauge condition

$$v^\mu b_\mu = 0\ , \tag{8.114}$$

then the color charge of the quark remains constant, $\dot{Q} = 0$, even in the presence of the perturbations b_μ. We consider the case that $z^\mu(s) = z^\mu(0) + v^\mu s$ where $v^\mu = (1, 0, 0, 0)$. Then s may be identified with laboratory time t, and the retarded distance ρ with the radial distance r. In this static case the gauge condition (8.114) becomes

$$b_0 = 0\ . \tag{8.115}$$

Let us restrict our discussion to the weak coupling regime $g \ll 1$. Taking into account that A_μ goes like g^{-1}, we may then retain in (8.113) only terms of zero order in g.

We denote spatial components of b^μ by boldface characters, $b^\mu = (b^0, \mathbf{b})$. Among all perturbations $\mathbf{b}_a$, we consider only those which are orthogonal to gauge modes,

$$\int d^3x\, \mathbf{b}_a \cdot (\nabla\lambda^a + g\epsilon^{abc}\,\mathbf{B}_b\lambda_c) = 0\;. \tag{8.116}$$

This orthogonality condition is ensured by the local equation

$$\nabla \cdot \mathbf{b}_a + g\epsilon_{abc}\,\mathbf{B}^b \cdot \mathbf{b}^c = 0\;, \tag{8.117}$$

which, in the weak coupling limit $g \to 0$, becomes

$$\nabla \cdot \mathbf{b}_a = 0\;. \tag{8.118}$$

Putting

$$\mathbf{b} = \mathbf{b}_3\,\tau_3 + \mathbf{b}_+\,(\tau_1 + i\tau_2) + \mathbf{b}_-\,(\tau_1 - i\tau_2)\;, \tag{8.119}$$

and taking into account (8.118), we obtain from (8.113)

$$\Box\,\mathbf{b}_3 = 0\;, \tag{8.120}$$

$$\left(\Box \mp \frac{4}{r}\frac{\partial}{\partial t} + \frac{4}{r^2}\right)\mathbf{b}_\pm = 0\;, \tag{8.121}$$

$$\mathbf{r} \cdot \mathbf{b}_\pm = 0\;. \tag{8.122}$$

It is clear from (8.120) that the stability of A_μ is unaffected by $\mathbf{b}_3$. The function $\mathbf{b}_-$, satisfying (8.118), (8.121), and (8.122), and possessing oscillatory behavior in time is given by

$$\begin{aligned}\mathbf{b}_-\,(t,\mathbf{r}) = \int_0^\Lambda d\omega \sum_{l,m} &\{\alpha_{lm}(\omega)\,e^{-i\omega t}\,\mathbf{Y}_{lm}\,(\theta,\phi)\,K_j\,(\omega r) \\ &+ \beta_{lm}(\omega)\,e^{i\omega t}\,[\mathbf{Y}_{lm}\,(\theta,\phi)\,K_j\,(\omega r)]^*\}\;.\end{aligned} \tag{8.123}$$

Here, Λ is an appropriate cutoff parameter, $\mathbf{Y}_{lm}(\theta,\phi)$ is a vector spherical harmonic, $K_j\,(\omega r)$ is expressed in terms of the confluent hypergeometric function,

$$K_j(x) = x^j\,e^{-ix}\,F\,(j-1,\,2j+2,\,2ix)\;, \tag{8.124}$$

and j runs through the positive roots of the equation

$$j(j+1) = l(l+1) + 4, \qquad l = 1, 2, \ldots \tag{8.125}$$

A similar solution $\mathbf{b}_+$ is obtained from (8.123) by replacing t by $-t$.

We are looking for solutions $\mathbf{b}_\pm$ in the class of functions with an appropriate falloff at infinity, and which are less singular than A_μ at $r = 0$. Every solution corresponding to a negative root of (8.125) is more singular than the background A_μ defined in (8.41), and hence should be excluded.

The function $K_j(x)$ is regular at $x = 0$ while, in the limit $x \to \infty$, it has the asymptotic form

$$K_j(x) = \left[c_j x + d_j + O\left(x^{-1}\right)\right] \exp(ix) \tag{8.126}$$

where c_j and d_j are certain known constants.

Consider

$$\mathbf{b}_-(0,\mathbf{r}) = \int_0^\Lambda d\omega \sum_{l,m} \{\alpha_{lm}(\omega)\mathbf{Y}_{lm}(\theta,\phi)K_j(\omega r) + \beta_{lm}(\omega)[\mathbf{Y}_{lm}(\theta,\phi)K_j(\omega r)]^*\} \,. \tag{8.127}$$

If $\mathbf{b}_-(0,\mathbf{r})$ is 'small' with respect to $A_\mu(0,\mathbf{r})$, then $\mathbf{b}_-(t,\mathbf{r})$ given by (8.123) meets the same requirement for smallness.

We next discuss the case that two quarks are in the cold phase and generate the retarded Yang–Mills field A^μ defined in (8.75). Let a perturbation b^μ be decomposed into color basis elements of the form (8.66)–(8.72),

$$b^\mu = \sum_{n=1}^{3} \left[b_n^\mu H_n + \sum_{k=1}^{3} \left(b_{kn-}^\mu E_{nk}^- + b_{kn+}^\mu E_{kn}^+ \right) \right] . \tag{8.128}$$

We restrict our discussion to the case that the quarks are static, and adopt the gauge condition (8.115), which ensures that the color charges of both quarks are constant. By repeating what was done in the single-quark case, we find that the $\mathbf{b}_n$ satisfy equations (8.118) and (8.120), while $\mathbf{b}_{23\pm}$ and $\mathbf{b}_{13\pm}$ satisfy equations (8.118), (8.121) and (8.122), with r playing the role of ρ_1 for $\mathbf{b}_{23\pm}$ and ρ_2 for $\mathbf{b}_{13\pm}$. From this identification, one can check that the stability of A^μ is unaffected by these perturbations.

As for $\mathbf{b}_{12\pm}$, it obeys equations (8.118) and

$$\left[\Box \mp 4 \left(\frac{1}{r_2} - \frac{1}{r_1} \right) \frac{\partial}{\partial t} + 4 \left(\frac{1}{r_2} - \frac{1}{r_1} \right)^2 \right] \mathbf{b}_{12\pm} = 0 \,, \tag{8.129}$$

where $r_I = |\mathbf{x} - \mathbf{z}_I|$, and $\Box = \partial^2/\partial t^2 - \nabla$. Let the quarks be separated by distance d. In the limit $r_I \gg d$, equation (8.129) becomes the wave equation, and its solutions have asymptotic behavior either

$$\mathbf{b}_{12\pm} \sim \text{const.} \tag{8.130}$$

or

$$\mathbf{b}_{12\pm} \sim \sum_{k,l,m} j_l(kr) \left[\mathbf{c}_{lm}^\pm(k)\, Y_{lm}(\theta,\phi)\, e^{-ikt} + \mathbf{d}_{lm}^\pm(k)\, Y_{lm}(\theta,\phi)^* e^{ikt}\right] , \tag{8.131}$$

where $j_l(kr)$ are the spherical Bessel functions

$$j_l(kr) \sim \frac{1}{kr} \sin\left(kr - \frac{\pi l}{2} \right) , \quad kr \gg l \,. \tag{8.132}$$

Hence $\mathbf{b}_{12\pm}$ does not engender an instability in A^μ.

The stability analysis of the background (8.96) generated by N-quark clusters is identical to that of the two-quark case, with $\mathbf{b}_{I\,N+1\,\pm}, I = 1, \ldots, N$ playing the role of $\mathbf{b}_{23\pm}$ and $\mathbf{b}_{13\pm}$, while $\mathbf{b}_{12\pm}$ being represented by $\mathbf{b}_{J\,L\,\pm},\ J, L = 1, \ldots, N$.

Let us go over to the problem of stability for the background A^μ associated with the hot phase. We consider the Yang–Mills fields A^μ generated by static quarks, and restrict ourselves to the single-quark case.

The field A^μ defined in (8.40) is independent of g, and so $\pm 2i/g$ must be replaced by q in all previous relations. Equation (8.121) becomes

$$\left(\Box \mp \frac{2igq}{r}\frac{\partial}{\partial t} - \frac{g^2q^2}{r^2}\right)\mathbf{b}_\pm = 0\,, \tag{8.133}$$

Equations (8.124)–(8.126) are modified to

$$K_j(x) = x^j\, e^{-ix}\, F(-igq + j + 1,\, 2j + 2,\, 2ix)\,, \tag{8.134}$$

$$j(j+1) = l(l+1) - g^2q^2, \quad l = 1, 2, \ldots \tag{8.135}$$

$$K_j(x) = O\,(x^{igq-1}\, e^{ix}), \quad x \to \infty\,. \tag{8.136}$$

It is clear from (8.136) that q must be *real* for the perturbation behave as $1/r$ at spatial infinity, as the background A^μ does. Let us compare behaviors at $r = 0$. From (8.134) it follows that $K_j(x)$ is regular at $x = 0$ if $j \geq 0$. We can write the positive solutions of (8.135) as

$$j = \frac{1}{2}\left(\sqrt{(2l+1)^2 - 4g^2q^2} - 1\right)\,. \tag{8.137}$$

Taking the minimal allowable integer $l = 1$, one finds that j is positive if

$$g^2q^2 \leq 2\,. \tag{8.138}$$

Thus, the background field in the hot phase A^μ defined in (8.40) is stable provided that q is real, and $q \leq \sqrt{2}/g$.

Problem 8.5.1. Derive equations (8.120)–(8.122).

Problem 8.5.2. Show that (8.123) satisfies (8.118), (8.121), and (8.122).

Problem 8.5.3. Derive equation (8.129).

8.6 Vortices and Monopoles

An attractive feature of the string model of hadrons, mentioned at the beginning of this chapter, is that the spectrum of a quantized string has much in common with hadron phenomenology. The set of all known hadronic states can be arranged in subsets, the so-called *Regge sequences* or *Regge trajectories*, in such a way that a linear relation between spin and mass squared

$$J = \alpha_0 + \alpha' M^2 \tag{8.139}$$

holds for every Regge trajectory. The experimental value of α' is approximately the same for all the trajectories. It has become customary to display (8.139) as straight lines of a fixed slope on the *Chew–Frautschi plot* of hadronic mass squared M^2 versus spin J. Hadrons belonging to any Regge trajectory are separated by intervals $\Delta J = 2$. It is tempting to interpret such equidistant hadronic resonances as different vibratory states of a string.

We already learned in Problem 2.8.5 that if two particles are held together by a force derived from the potential

$$U(r) = -\frac{\alpha_s}{r} + k\,r \quad (\alpha_s > 0,\, k > 0)\;, \tag{8.140}$$

then the rest energy of this binary system may be greater than the sum of masses of its constituents. The potential (8.140) arises in the string model, where k is related to the string tension. It is clear that a meson of mass M may be a system of two quarks of masses m_1 and m_2 such that $m_1 + m_2 \ll M$ if k is sufficiently large. Mass of the meson is then almost entirely due to the string tension. To illustrate, consider a π^+-meson ($M_\pi = 140$ MeV), whose constituents u and $\bar{d}$ have masses $m_u \approx 5$ MeV and $m_d \approx 7$ MeV.

Let two massless quarks be attached to a string. For simplicity, we consider a straight string of length ℓ which rotates as a rigid stick around its center of mass so that its ends move with the speed of light. The velocity v of a point separated from the center of mass by radial distance r is given by

$$v = 2r/\ell\;. \tag{8.141}$$

Assuming that the string is specified by a constant mass per unit length T, one finds the string energy

$$E = 2\int_0^{\ell/2} dr\,\frac{T}{\sqrt{1-v^2}} = \pi\ell T\;, \tag{8.142}$$

and angular momentum

$$J = 2\int_0^{\ell/2} dr\,\frac{Trv}{\sqrt{1-v^2}} = \frac{\pi\ell^2 T}{2}\;. \tag{8.143}$$

These properties of the string imply that its angular momentum is proportional to the square of its energy. Comparing this with (8.139), we conclude that the string tension T is expressed in terms of the phenomenological value of the slope α',

$$T = 2\pi\alpha' \approx 1\,\mathrm{GeV}\cdot\mathrm{fm}^{-1}\;. \tag{8.144}$$

The phenomenological theory of strong interactions was developed throughout the 1960s. Central to this theory is the idea that the scattering matrix is determined by Regge trajectory exchanges, rather than by separate hadron exchanges. Dual resonance models, in which the transition amplitudes are invariant under cyclic permutations of external momenta and have resonant poles in all channels associated with a given ordering of external momenta, are an outgrowth of this idea. The simplest example of an amplitude satisfying the duality requirement is the original 1968 proposal by Gabriele Veneziano for a process involving four external particles:

$$A(s,t,u) = V(s,t) + V(t,u) + V(u,s) \; , \tag{8.145}$$

$$V(s,t) = \int_0^1 dx\, x^{-\alpha(s)-1} \, (1-x)^{-\alpha(t)-1} \; , \tag{8.146}$$

where s, t, and u are the Mandelstam parameters, (2.384), and the Regge trajectory is

$$\alpha(s) = \alpha_0 + \alpha' s \; . \tag{8.147}$$

Shortly thereafter, it became clear that the Veneziano model represents the scattering of extended objects known as Nambu strings. The advent of string models was a step in establishing the basis for the theory of strong interactions. However, the status of such models still remained heuristic.

In 1973, Holger Nielsen and Poul Olesen showed that classical field theory with a Higgs type Lagrangian possesses *vortex-line solutions*, analogous to the vortex line in a type II superconductor. They identified these solutions with the Nambu string. Following their treatment, we turn to an Abelian version of the Higgs model whose Lagrangian is[5]

$$\mathcal{L} = -\frac{1}{4}\, F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}\, |\, (\partial_\mu + ieA_\mu)\, \phi|^2 + \frac{1}{2}\, \mu^2\, \phi^*\phi - \frac{1}{4}\, \lambda^2\, (\phi^*\phi)^2 \; . \tag{8.148}$$

The associated equations of motion are

$$\Box A_\mu - \partial_\mu \partial^\nu A_\nu = j_\mu = \frac{1}{2}\, ie\, (\phi^*\, \partial_\mu \phi - \phi\, \partial_\mu \phi^*) + e^2 A_\mu\, |\phi|^2 \; , \tag{8.149}$$

$$\Box\phi + ie\, (2A^\mu \partial_\mu \phi + \phi\, \partial_\mu A^\mu) - e^2 A^\mu A_\mu\, \phi = \frac{1}{2}\left(\mu^2 - \lambda^2\, |\phi|^2\right) \phi \; . \tag{8.150}$$

Let us write

$$\phi = \rho\, e^{-i\chi} \; , \tag{8.151}$$

where ρ and χ are real functions of spacetime. If the Higgs field is close to the equilibrium point

$$\rho_0 = \frac{\mu}{\lambda} \; , \tag{8.152}$$

[5] In this section, we switch from the Gauss units employed throughout the book to the Heaviside units which are traditionally used in quantum field theory. In these units, the factor $(4\pi)^{-1}$ moves from the Lagrangian to some solutions.

then $\rho = \rho_0$ almost everywhere. However, near some line, say, the x^3-axis, ρ may differ appreciably from ρ_0, and χ may vary by $2\pi n$ when we make a complete turn around this axis, so that ϕ remains single-valued. This is a field configuration of the vortex type.

It follows from (8.149) that

$$A_\mu = \frac{1}{e^2}\,\frac{j_\mu}{|\phi|^2} - \frac{1}{e}\,\partial_\mu\chi\ . \tag{8.153}$$

The flux of $F_{\mu\nu}$ through a three-dimensional surface $\mathcal{S}$ bounded by a loop $\mathcal{C}$ is

$$\varPhi = \int_{\mathcal{S}} dS^{\mu\nu} F_{\mu\nu} = \int_{\mathcal{C}} dx^\mu A_\mu\ . \tag{8.154}$$

We assume that there is no current j_μ through $\mathcal{S}$, and substitute (8.153) into (8.154):

$$\varPhi = -\frac{1}{e}\int_{\mathcal{C}} dx^\mu\,\partial_\mu\chi = \frac{2\pi}{e}\,n\ . \tag{8.155}$$

Thus the flux of vortex lines is quantized, to be a multiple of $2\pi/e$.

We are now looking for a vortex centred on the x^3-axis. We consider the static case, with gauge choice $A_0 = 0$. Assuming cylindrical symmetry, we can write the Higgs field in the form

$$\phi = \rho(r)\,e^{in\varphi}\ , \tag{8.156}$$

and a vector potential of the magnetic field around this vortex

$$\mathbf{A}(r) = A(r)\,\mathbf{e}_\varphi = A(r)\,\nabla\varphi\ , \tag{8.157}$$

where r and φ are polar coordinates in the equatorial plane. We put $n = 1$ in (8.156), which means that the vortex contains a single unit of quantized flux. Then equations (8.149) and (8.150) become

$$-\frac{d}{dr}\left[\frac{1}{r}\,\frac{d}{dr}\,(r\,A)\right] + \rho^2\left(e^2 A - \frac{e}{r}\right) = 0\ , \tag{8.158}$$

$$-\frac{1}{r}\,\frac{d}{dr}\left(r\,\frac{d}{dr}\,\rho\right) + \rho\left[\left(eA - \frac{1}{r}\right)^2 - \mu^2 + \lambda^2\rho^2\right] = 0\ , \tag{8.159}$$

Exact solutions to these equations have yet to be found. We assume that ρ tends to ρ_0 asymptotically as $r \to \infty$. Taking $\rho(r) = \rho_0$ renders (8.158) a modified Bessel equation which can be solved:

$$A(r) = \frac{1}{er} + \frac{C}{e}\,K_1(e\rho_0 r) \to \frac{1}{er} + \frac{C}{e}\sqrt{\frac{\pi}{2e\rho_0 r}}\,e^{-e\rho_0 r}, \quad r \to \infty\ . \tag{8.160}$$

Here, C is a constant of integration, and K_1 the Hankel function of the first kind evaluated at an imaginary argument. The magnetic induction is

$$|B(r)| = \frac{1}{2\pi r}\,\frac{d}{dr}\,\Phi(r) = \frac{1}{r}\,\frac{d}{dr}\,(rA) \to C\sqrt{\frac{\pi\rho_0}{2er}}\,e^{-e\rho_0 r}, \quad r\to\infty\,. \tag{8.161}$$

We see that the magnetic field deviates appreciably from zero only near the x^3-axis in a region with characteristic length ℓ (called the penetration length in superconductivity)

$$\ell = \frac{1}{e\rho_0} = \frac{e\lambda}{\mu}\,. \tag{8.162}$$

Taking $\rho = \rho_0 + \eta$, one obtains (Problem 8.6.2) an asymptotic form for the solution to equation (8.159)

$$\rho = \rho_0\left(1 - e^{-\xi r}\right)\,, \tag{8.163}$$

where ξ a new characteristic length

$$\xi = \frac{1}{\mu} \tag{8.164}$$

which measures the distance required for ρ to attain its asymptotic value ρ_0.

Let ℓ and ξ be of the same order. Then we have a well defined kernel in the form of a thin tube of the width $\ell \approx \xi$ which contains most of the flux lines. The magnetic field B falls exponentially outside the kernel. This magnetic flux tube is in equilibrium against the pressure of the surrounding charged superconducting Higgs field. It is clear from (8.163) that the Higgs field is pressed out from the tube.

By a suitable choice of μ, e, and λ, we can arrange that the width of the kernel ℓ is sufficiently small. Then the magnetic field B is nonzero only within the kernel, and acts as a sort of smeared out delta-function. Thus we come to a string as a mathematical idealization of this kernel in the limit $\ell \to 0$. If the string is of finite length and open, the magnetic flux must be absorbed by a magnetic monopole at each end.

Let us now substitute this vortex solution into the Lagrangian (8.148) and study how the result depends upon the string world sheet $X^\mu(\tau,\sigma)$. We denote the resulting Lagrangian by $\mathcal{L}_{\text{vortex}}$. To take proper account of an arbitrary motion of a vortex-string in the transverse direction, $\mathcal{L}_{\text{vortex}}$ should be Lorentz-contracted,

$$\mathcal{L}_{\text{vortex}} \propto \sqrt{1 - \mathbf{V}_\perp^2}\,. \tag{8.165}$$

Let $d\sigma$ be the element of length along the vortex-string. Assuming that the width of the vortex is constant and tiny little, we obtain the action

$$S_{\text{vortex}} = \int d^4x\,\mathcal{L}_{\text{vortex}} \propto \int dt\,d\sigma\,\sqrt{1 - \mathbf{V}_\perp^2}\,. \tag{8.166}$$

By definition,

$$\mathbf{V}_\perp = \dot{\mathbf{X}} - \mathbf{X}'\left(\dot{\mathbf{X}}\cdot\mathbf{X}'\right)\,, \tag{8.167}$$

where the overdot and prime stand for the derivatives with respect to t and σ,

$$S_{\text{vortex}} \propto \int dt\, d\sigma \sqrt{1 - \dot{\mathbf{X}}^2 + \left(\dot{\mathbf{X}} \cdot \mathbf{X}'\right)^2}\,. \tag{8.168}$$

One can readily show (Problem 8.6.3) that S_{vortex} is proportional to the Nambu action (5.293). The fact that the Nambu Lagrangian is derivable from the field theory Lagrangian (8.148) may be recognized as evidence that an *effective* string dynamics is intrinsic in gauge field theories with spontaneous symmetry breaking.

This line of reasoning had a profound impact on the confinement problem. However, the concrete implementation of the Nielsen–Olesen vortices, now thought to be more relevant in the confinement problem, has been revised. One speculates that the dual Meissner effect rendering chromoelectric field squeezed into a vortex is formally the same as the original one but with the roles of chromoelectricity and chromomagnetism reversed. The chromoelectric flux would be absorbed by a quark and an antiquark at the ends of the vortex. The environment would be a medium exhibiting the superconductivity property with respect to chromomagnetic charges. With vortex-like configurations of the Yang–Mills field, quarks and antiquarks would be permanently bound in pairs.

We now consider another solution, the *'t Hooft–Polyakov monopole*, which was discovered independently by Gerardus 't Hooft and Alexander Polyakov in 1974. This configuration is a finite-energy smooth solution to the SO(3) gauge theory with a Higgs triplet. The Lagrangian is:

$$\mathcal{L} = -\frac{1}{4}\, G^a_{\mu\nu} G_a^{\mu\nu} + \frac{1}{2}\, D_\mu \phi^a D^\mu \phi_a - V(\phi)\,, \tag{8.169}$$

where

$$G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g\epsilon^{abc} A^b_\mu A^c_\nu\,, \tag{8.170}$$

$$D_\mu \phi_a = \partial_\mu \phi_a + g\epsilon_{abc} A^b_\mu \phi^c\,, \tag{8.171}$$

and the Higgs potential is

$$V(\phi) = \frac{1}{4}\, \lambda^2 \left(\frac{\mu^2}{\lambda^2} - \phi^2\right)^2, \quad \phi^2 = \phi_a \phi^a\,. \tag{8.172}$$

The corresponding Euler–Lagrange equations read

$$\partial^\mu G^a_{\mu\nu} = g\epsilon^{abc} \left[(D_\nu \phi)_b\, \phi_c + A^\mu_b G^c_{\mu\nu} \right]\,, \tag{8.173}$$

$$\partial^\mu D_\mu \phi_a = g\epsilon_{abc} \left(D^\mu \phi^b\right) A^c_\mu + \mu^2 \phi_a - \lambda^2 \phi_a \phi^2\,. \tag{8.174}$$

$V(\phi)$ must approach zero at spatial infinity, which implies that the Higgs field has a nonzero limit at spatial infinity

$$\phi_a \to \frac{\mu}{\lambda}\, \hat{\phi}_a(\mathbf{n}), \quad \hat{\phi}^a \hat{\phi}_a = 1, \quad \mathbf{n}^2 = 1, \quad r \to \infty\,. \tag{8.175}$$

The boundary condition (8.175) singles out a particular isotopic axis $\hat{\phi}_a$ for each spatial direction **n**. Hence it breaks the SO(3) gauge invariance. However, solutions subject to this condition are invariant under the group of rotations about $\hat{\phi}_a$. Thus the unbroken gauge symmetry is SO(2) or, equivalently, the U(1) subgroup of SO(3).

The resulting U(1) gauge theory can be identified with Maxwell's electrodynamics of charged vector and scalar fields if generators T_a of the initial gauge group SO(3) are projected on $\hat{\phi}^a$. Then the Abelian vector potential A^μ associated with the local U(1) gauge group is

$$A^\mu = (\phi^a A^\mu_a)\,\frac{\lambda}{\mu}\ , \tag{8.176}$$

with electric charge

$$e = g\,(\phi^a T_a)\,\frac{\lambda}{\mu}\ . \tag{8.177}$$

We ask for static spherically symmetric solution to equations (8.173)–(8.174). We use the gauge condition $A^a_0 = 0$ (which gives $D_0\phi_a = 0$ and $G^a_{oi} = 0$), and introduce the ansatz

$$\phi_a = \frac{x_a\, a(r)}{gr^2}\ , \tag{8.178}$$

$$A^a_i = \frac{\epsilon_{aij}\, x_j\, [1 - b(r)]}{gr^2}\ . \tag{8.179}$$

Equations (8.173)–(8.174) reduce (Problem 8.6.4) to

$$r^2 a'' = a\left(2b^2 - \mu^2 r^2 + \frac{\lambda^2}{g^2}\, a^2\right)\ , \tag{8.180}$$

$$r^2 b'' = b\left(b^2 - 1 + a^2\right)\ , \tag{8.181}$$

where the prime stands for differentiation with respect to r.

For arbitrary μ and λ, these equations have never been solved analytically. However, in the limit $\mu \to 0$, $\lambda \to 0$, with $\mu/\lambda < \infty$, the so-called *Bogomol'ny–Prasad–Sommerfield limit*, an exact solution is given by (Problem 8.6.5)

$$a(r) = \beta r \coth(\beta r) - 1, \quad b(r) = \frac{\beta r}{\sinh(\beta r)}\ , \tag{8.182}$$

where β is an arbitrary constant.

In this limit, the Higgs field behaves as a massless long range field. It was found that two static equally charged Bogomol'ny–Prasad–Sommerfield monopoles repel each other by the inverse square law magnetic force which is equal in magnitude to the attractive force exerted by the Higgs field, and these effects exactly cancel.

One can show (Problem 8.6.6) that the solution (8.182) is self-dual. That is, it satisfies the condition

$$
{}^*G^a_{\mu\nu} = i\, G^a_{\mu\nu} \ , \tag{8.183}
$$

or equivalently,

$$
i\,\mathbf{E}^a = \mathbf{B}^a \ , \tag{8.184}
$$

where

$$
E^a_i = G^a_{0i}, \quad B^a_i = -\frac{1}{2}\,\epsilon_{ijk} G^a_{jk} \tag{8.185}
$$

are the SU(2) ‘electric’ and ‘magnetic’ Yang–Mills fields.

Any self-dual solution in Minkowski space has a vanishing stress-energy tensor. Indeed, with reference to Problem 5.2.8, we may cast the stress-energy tensor in the form

$$
\begin{aligned}
\Theta_{\mu\nu} &= \frac{1}{4\pi}\left(G_{a\mu}{}^{\lambda} G^a_{\lambda\nu} + \frac{\eta_{\mu\nu}}{4}\, G_a^{\alpha\beta} G^a_{\alpha\beta}\right) = \frac{1}{8\pi}\left(G_{a\mu}{}^{\lambda} G^a_{\lambda\nu} + {}^*G_{a\mu}{}^{\lambda}\, {}^*G^a_{\lambda\nu}\right) \\
&= \frac{1}{8\pi}\left(G_{a\mu}{}^{\lambda} + i\,{}^*G_{a\mu}{}^{\lambda}\right)\left(G^a_{\lambda\nu} - i\,{}^*G^a_{\lambda\nu}\right) \ ,
\end{aligned} \tag{8.186}
$$

which shows that $\Theta_{\mu\nu} = 0$ for ${}^*G_{\mu\nu} = \pm i\, G_{\mu\nu}$. Thus, any Bogomol’ny–Prasad–Sommerfield monopole carries zero energy.

In the general case of arbitrary μ and λ, equations (8.180) and (8.181) can be analyzed only qualitatively. Combining (8.175) with (8.178), we obtain

$$
\lim_{r\to\infty} \phi(\mathbf{r}) = \frac{\mu}{\lambda}\,\mathbf{n} \ , \tag{8.187}
$$

that is $\hat{\phi}^a = n^a$. The ansatz (8.179) is consistent with equations (8.180) and (8.181) if $b(r) = O(1)$ as $r \to \infty$. The ‘electromagnetic’ field strength $F_{\mu\nu}$ should be identified with the component of the Yang–Mills strength $G^a_{\mu\nu}$ in the direction of $\hat{\phi}^a$, which corresponds to the unbroken U(1) symmetry. Hence, $G^a_{0j} = 0$, and, far from the origin,

$$
G^a_{ij} = -\epsilon_{ijk}\, \frac{n_k n_a}{4\pi e r^2} \ , \tag{8.188}
$$

or

$$
F_{ij} = \hat{\phi}_a\, G^a_{ij} = -\frac{1}{4\pi e r^2}\, \epsilon_{ija} n_a \ . \tag{8.189}
$$

This configuration represents a radial static magnetic field

$$
B^a = -\frac{n_a}{4\pi e r^2} \ , \tag{8.190}
$$

generated by the total magnetic charge $e^\star = 1/e$.

We see that the Dirac and ’t Hooft–Polyakov monopoles differ in their structure within a core of size $\sim \lambda/\mu$. The Dirac monopole has a singularity for which a point source has to be introduced explicitly in the action, while the 't Hooft–Polyakov monopole is smooth everywhere and satisfies the field equations (8.173) and (8.174) without external sources. Outside the core these field configurations are similar, except that the 't Hooft–Polyakov monopole has no singular string.

The 't Hooft–Polyakov monopole is electrically neutral due to the combination of our gauge condition $A_0^a = 0$ and our requirement that the field configurations should be independent of time. If we abandon this gauge condition and search for static solutions of the form

$$A_0^a(r) = \frac{c(r)\, n_a}{gr^2}\ , \tag{8.191}$$

then we have nonzero G_{0i}^a. Thus modified solutions (see Problems 8.6.7 and 8.6.8) can exhibit both electric and magnetic charges.

One may assume that 't Hooft–Polyakov monopoles occur in the cold phase of the subnuclear realm, and that their condensation causes the gluon vacuum exhibit the dual Meissner effect. This raises the natural question as to what actual field plays the role of the Higgs in spontaneous symmetry breaking for quantum chromodynamics. Unfortunately, the current understanding of this issue is far from perfect.

We finally consider an interesting aspect of the behavior of a Wong particle with color charge $\frac{1}{2}\, Q^i\, \sigma_i$ in the field of a 't Hooft–Polyakov monopole. For simplicity, we assume that the monopole is heavy and stationary. Let the monopole be at the origin, and the Wong particle moving slowly in the distance from it. We already discussed such a binary system of an electrically charged particle and a magnetic monopole in Sect. 2.9. It was established that

$$\mathbf{M} = \mathbf{L} - e\, e^{\star}\, \mathbf{n} \tag{8.192}$$

is conserved. Here, $\mathbf{M}$ is the total angular momentum of this two-particle system plus associated electromagnetic field. With reference to Problem 5.2.10, we recognize $e\, e^{\star}\, \mathbf{n}$ as the electromagnetic field contribution to $\mathbf{M}$.

Let the SO(3) gauge symmetry be spontaneously broken to U(1) by virtue of the boundary condition (8.175). Then the Yang–Mills field of a monopole outside the core is projected to $\hat{\phi}_a G^a_{\mu\nu}$, which is just the magnetic field component. The Yang–Mills stress-energy tensor become identical to the electromagnetic stress-energy tensor since $G^a_{\mu\nu} G_a^{\alpha\beta} = \hat{\phi}_a G^a_{\mu\nu} \hat{\phi}^b G_b^{\alpha\beta}$. By repeating what was done in electrodynamics (Problem 5.2.10), we arrive at the field contribution

$$-|Q|\, \frac{1}{e}\, \boldsymbol{\tau}\ , \tag{8.193}$$

where $|Q| = \sqrt{Q_i Q^i}$, $e^{\star} = 1/e$, and the spatial direction $\mathbf{n}_i$ is identified with the isotopic direction $\hat{\phi}_i = \tau_i$. The occurrence of the term (8.193) is a noteworthy effect of the coupling of a Wong particle with a 't Hooft–Polyakov monopole. Following a similar line of argument, Jackiw and Claudio Rebbi, and independently Peter Hasenfratz and 't Hooft were led in 1976 to interpret the term (8.193) as 'spin originating from isospin'.

Problem 8.6.1. Verify that applying (8.156) and (8.157) to (8.149) and (8.150) gives (8.158) and (8.159).

Problem 8.6.2. Show that (8.163) is an asymptotic solution to equation (8.159) in the limit $r \to \infty$.

Problem 8.6.3. Show that the Nambu action (5.293) takes the form of equation (8.168) if σ and τ are regarded as the string length and laboratory time.

Problem 8.6.4. Show that the ansatz (8.178)–(8.179) leads to equations (8.180) and (8.181).

Problem 8.6.5. Verify that (8.182) is an exact solution to equations (8.180) and (8.181) in the limit $\mu \to 0$, $\lambda \to 0$, with $\mu/\lambda < \infty$.

Problem 8.6.6. Verify that (8.182) represents a self-dual configuration, satisfying (8.183).

Problem 8.6.7. Show that the ansatz (8.178), (8.179), and (8.191) simplifies equations (8.173)–(8.174) to give

$$r^2 a'' = a\left(2b^2 - \mu^2 r^2 + \frac{\lambda^2}{g^2}\, a^2\right), \tag{8.194}$$

$$r^2 b'' = b\left(b^2 - 1 + a^2 - c^2\right), \tag{8.195}$$

$$r^2 c'' = 2b^2 c\,. \tag{8.196}$$

Problem 8.6.8. Show that, in the limit $\mu \to 0$, $\lambda \to 0$, with $\mu/\lambda < \infty$, an exact solution to equations (8.194)–(8.196) is given by

$$a(r) = \cosh[\gamma(\beta r \coth \beta r - 1)],\ b(r) = \frac{\beta r}{\sinh(\beta r)},\ c(r) = \sinh[\gamma(\beta r \coth \beta r - 1)], \tag{8.197}$$

where β and γ are arbitrary constants.

8.7 Two Phases of the Subnuclear Realm

It is generally believed that confinement is a quantum phenomenon. The reason for this belief is simple. Quarks are confined in hadrons whose characteristic size cannot be expressed in terms of the parameters involved in the Lagrangian (8.1)[6]. The actual mechanism of quark binding gives rise to a new

[6] To illustrate, we turn once again to a π^+-meson, a system of a quark u and an antiquark $\bar{d}$. Their masses $m_u \approx 5$ MeV, $m_d \approx 7$ MeV and their associated Compton wavelengths $\lambda_u = 1/m_u \approx 40$ fm, $\lambda_d = 1/m_d \approx 30$ fm differ greatly from the corresponding quantities for π-mesons, $M_\pi = 140$ MeV and $\lambda_\pi = 1.4$ fm.

dimensional parameter specifying the confinement energy scale $\Lambda \sim 100$ MeV, which violates the conformal invariance of the Yang–Mills sector. From this discussion it follows that, although classical fields A_μ may depend on dimensional parameters, the interquark force and the general mechanism which assembles these fields into color singlets are free of these parameters. Hence conformal invariance can be assumed classically for all color-neutral constructions. The dynamical aspect of the mechanism responsible for the 'mass gap' Λ has to be explained on the quantum level.

However, distinction between the symmetry properties of the cold and hot phases of subnuclear realm seems to be amenable to the classical treatment. It has long been known that the symmetry of a physical system is determined by the properties of its ground state. A theorem by Coleman states: the invariance of the vacuum is the invariance of the world. We will not go into a discussion of this remarkable result (which would call for quantum field theory), and only note that the gluon vacuum exhibits the invariance of a classical Yang–Mills background. This background is sometimes thought of as a Bose condensate of the gluon field. Gluon quanta are excited about this condensate and inherit its properties. Hopefully, the background for both cold and hot phases can be seen in exact solutions to the Yang–Mills–Wong theory.

While on the subject of Lagrangians devoid of the Higgs fields, how is it possible for the symmetry of a solutions to be different from that of the Lagrangian? To be specific, consider the $\mathrm{SU}(\mathcal{N})$ Yang–Mills sector. The Lagrangian $\mathcal{L}$ is automatically invariant under $\mathrm{SL}(\mathcal{N},\mathbb{C})$, the complexification of $\mathrm{SU}(\mathcal{N})$. If we have no prior knowledge of the symmetry, it can be identified by the structure constants f_{abc} which appear in $\mathcal{L}$. The particular values of f_{abc} entering into the Lagrangian suggest that it is invariant under $\mathrm{SU}(\mathcal{N})$. However, for any simple complex Lie algebra, there exists a basis, the Cartan basis, such that the structure constants are found to be real, antisymmetric and identical to the structure constants of the real compact form of this Lie algebra. The basis of the Lie algebra $\mathrm{su}(\mathcal{N})$ is simultaneously the Cartan basis of its complexification $\mathrm{sl}(\mathcal{N},\mathbb{C})$. Thus, the presence of the structure constants of $\mathrm{SU}(\mathcal{N})$ in $\mathcal{L}$ need not imply that the symmetry group is $\mathrm{SU}(\mathcal{N})$. Allowing complex-valued solutions to the field equations, we enlarge the symmetry of $\mathcal{L}$ to $\mathrm{SL}(\mathcal{N},\mathbb{C})$. The only a priori constraint, stemming from the fact that the coefficients of the Yang–Mills equations are real, is that each complex solution is accompanied by the complex conjugate solution. Complex-valued Yang–Mills fields occur only as pairs of complex conjugate solutions. Whether or not complex-valued solutions actually occur depends on the structure of the total Lagrangian and choice of the boundary condition.

It was demonstrated in Sects. 8.1–8.4 that the $\mathrm{SU}(\mathcal{N})$ Yang–Mills–Wong theory indeed leaves room for complex-valued solutions, which, however, become real-valued if the complexified gauge group $\mathrm{SL}(\mathcal{N},\mathbb{C})$ is specialized to its noncompact real form $\mathrm{SL}(\mathcal{N},\mathbb{R})$. For this reason, such solutions would be more properly termed *spontaneously deformed*. It was pointed out at the end of Sect. 6.1 that gauge symmetries defy spontaneous breakdown: the symmetry is not really broken, the Higgs mechanism makes it merely hidden. However, gauge symmetries are capable of *spontaneous deformation* in the true sense of the word.

We now look more closely at the spontaneously deformed phase of subnuclear realm. We assume it to be the phase favorable for hadronization, and designate it as *cold*. We do not use the term 'confinement phase' to avoid associations with a particular mechanism of quark binding, say, the string mechanism discussed in the preceding section.

In the single-quark case, the gauge group may be chosen rather arbitrarily. Indeed, the retarded Yang–Mills field generated by a single quark in the cold phase needs an individual elementary $\mathrm{sl}(2,\mathbb{C})$ color cell. Choosing initially $\mathrm{SO}(\mathcal{N})$ or $\mathrm{Sp}(\mathcal{N})$, instead of $\mathrm{SU}(\mathcal{N})$, one singles out the same elementary color cell because $\mathrm{sl}(2,\mathbb{C}) \sim \mathrm{so}(3,\mathbb{C}) \sim \mathrm{sp}(1,\mathbb{C})$.

Turning to the Yang–Mills field generated by two- and three-quark clusters (regarded as toy mesons and toy baryons), we come respectively to $\mathrm{SL}(3,\mathbb{R})$ and $\mathrm{SL}(4,\mathbb{R})$. Since $\mathrm{SL}(3,\mathbb{R})$ is a subgroup of $\mathrm{SL}(4,\mathbb{R})$, the background of every toy hadron is specified by the gauge group $\mathrm{SL}(4,\mathbb{R})$.

If we wish to gain a more penetrating insight into the physics behind this group we should digress for a while and note that just $\mathrm{SL}(4,\mathbb{R})$, was suggested by Yuval Ne'eman and Djordje Šijački in 1985 for an exhaustive phenomenological classification of hadrons by Regge sequences. Recall that only compact groups have finite-dimensional unitary representations, while unitary representations of noncompact groups are infinite. The idea that Regge sequences of hadrons may be treated as infinite unitary multiplets of a noncompact group goes back to Yossef Dothan, Murrey Gell-Mann, and Ne'eman. In 1965, they examined $\mathrm{SL}(3,\mathbb{R})$ as a spacetime symmetry group generated by the angular momentum operators L_i and the quadrupole operators T_{ij} with the commutation relations

$$[L_i, L_j] = i\epsilon_{ijk}\, L_k\ , \tag{8.198}$$

$$[L_i, T_{jk}] = i\epsilon_{ijl}\, T_{lk} + i\epsilon_{ikl}\, T_{jl}\ , \tag{8.199}$$

$$[T_{ij}, T_{kl}] = -i\left(\delta_{ik}\epsilon_{jlm} + \delta_{il}\epsilon_{jkm} + \delta_{jl}\epsilon_{ikm}\right) L_m\ . \tag{8.200}$$

The Lie algebra given by (8.198)–(8.200) seems to be the minimal scheme capable to model two salient features of Regge trajectories: the $\Delta J = 2$ rule and the apparently infinite sequence of hadronic states. The expert reader will recognize the similarity between (8.199) and the well-known commutation relations of the quantum-mechanical number operator n and the creation and annihilation operators a and a^+ in the harmonic oscillator problem,

$$[n, a] = -a, \quad [n, a^+] = a^+\ . \tag{8.201}$$

Just as a and a^+ shift occupation numbers by one unit, so T_{ij} raises or lowers eigenvalues of the angular momentum by two units, forming the equally spaced Regge trajectories.

It was found that two infinite unitary representations belonging to the ladder series

$$D^{\mathrm{ladd}}_{\mathrm{SL}(3,\mathbb{R})}(0;\mathbb{R}): \quad J = 0, 2, 4, \ldots, \quad D^{\mathrm{ladd}}_{\mathrm{SL}(3,\mathbb{R})}(1;\mathbb{R}): \quad J = 1, 3, 5, \ldots \tag{8.202}$$

are associated with the π- and ρ-meson trajectories. In addition, there exists a unique spinorial ladder representation related to the nucleon trajectory

$$D^{\mathrm{ladd}}_{\mathrm{SL}(3,\mathbb{R})}\left(\tfrac{1}{2};\mathbb{R}\right): \quad J = \frac{1}{2}, \frac{5}{2}, \frac{9}{2}, \ldots\ , \tag{8.203}$$

while the representation starting with $J = \frac{3}{2}$ belongs to the discrete series

$$D^{\mathrm{disc}}_{\mathrm{SL}(3,\mathbb{R})}\left(\tfrac{3}{2};\mathbb{R}\right): \quad J = \frac{3}{2}, \frac{5}{2}, \frac{7^2}{2}, \frac{9^2}{2}, \frac{11^2}{2}, \ldots\ , \tag{8.204}$$

violating the rule $\Delta J = 2$. Therefore, the $\mathrm{SL}(3,\mathbb{R})$ scheme, although quite useful for understanding the Regge trajectories of mesons, turns out to be inadequate to account for those of baryons.

Matters can be improved by a simultaneous application of $\mathrm{SL}(3,\mathbb{R})$ and $\mathrm{SO}(1,3)$. The commutation relations become closed by embedding the two algebras in $\mathrm{sl}(4,\mathbb{R})$. For the mathematically inclined reader, we note that one can utilize the decomposition of the maximal compact subgroup of $\mathrm{SL}(4,\mathbb{R})$, $\mathrm{SO}(4) = \mathrm{SO}(3) \times \mathrm{SO}(3)$, as a basis with spin-parity (J^P) content:

$$J^P = (j_1 + j_2)^P,\ (j_1 + j_2 - 1)^{-P}, \ldots, (|j_1 - j_2|)^{\pm P}\ . \tag{8.205}$$

The operator T_{ij} shifts $\mathrm{SO}(4)$ multiplets in (j_1, j_2) by $\Delta j_{1,2} = 2$, and the structure of Regge sequences is reproduced by such shifting. Although this scheme is quite restrictive, it is in good agreement with the known data of hadronic spectroscopy.

A convenient basis of $\mathrm{sl}(4,\mathbb{R})$ contains 6 antisymmetric elements $M_{\mu\nu}$ and 9 symmetric traceless elements $T_{\mu\nu}$, which can be regrouped in the subsets:

$$L_i = \frac{1}{2}\epsilon_{ijk} M_{jk}, \quad K_i = M_{0i}, \quad T_{ij}, \quad N_i = T_{0i}, \quad T_{00}\ , \tag{8.206}$$

satisfying the commutation relations (8.198)–(8.200) together with

$$[K_i,\, K_j] = -i\epsilon_{ijk}\, K_k, \quad [N_i,\, N_j] = i\epsilon_{ijk}\, N_k; \tag{8.207}$$

$$[L_i,\, K_j] = i\epsilon_{ijk}\, K_k, \quad [L_i,\, N_j] = i\epsilon_{ijk}\, N_k\ , \tag{8.208}$$

$$[K_i,\, N_j] = -i\,(T_{ij} + \delta_{ij}\, T_{00}); \tag{8.209}$$

$$[K_i,\, T_{jk}] = -i(\delta_{ij}\, N_k + \delta_{ik}\, N_j)\ , \tag{8.210}$$

$$[N_i,\, T_{jk}] = -i(\delta_{ij}\, K_k + \delta_{ik}\, K_j)\ , \tag{8.211}$$

$$[L_i, T_{00}] = [T_{ij}, T_{00}] = 0, \quad [K_i, T_{00}] = -2iN_i, \quad [N_i, T_{00}] = -2iK_i \; . \tag{8.212}$$

However, the Yang–Mills–Wong particle has spin zero. Furthermore, it was mentioned in Sect. 8.4 that bound quarks move along straight world lines, and so any toy hadron should exhibit zero orbital momentum. Where does the necessary angular momentum come from? We suppose that it is largely contributed by gluon degrees of freedom[7]. Indeed, $M_{\mu\nu}$ and $T_{\mu\nu}$ can be expressed in terms of the Cartan-Weyl basis (8.90)–(8.92):

$$M_{ij} = -i\left(E^{+}_{ij} - E^{-}_{ij}\right), \quad K_j = i\left(E^{+}_{0j} - E^{-}_{0j}\right) \; , \tag{8.213}$$

$$T_{ij} = -i\left(E^{+}_{ij} + E^{-}_{ij}\right), \quad N_j = i\left(E^{+}_{0j} + E^{-}_{0j}\right) \; , \tag{8.214}$$

$$T_{00} = 2iH_0, \quad T_{jj} = -2iH_j \; . \tag{8.215}$$

It is conceivable that the color degrees of freedom of gluons in the $SL(4,\mathbb{R})$ chromagnetic background are converted to spin, much as spin appears in the isospin field of a monopole. However, there is as yet no concrete mechanism for this.

Let us turn to the hot phase. Deconfinement does not mean that quarks become completely free. Although they are no longer bound within hadrons, they are held together within a plasma lump. Whatever large extension of this lump may be, it is always colorless. We suppose that the total color charge of Yang–Mills–Wong particles in any plasma lump is zero,

$$\sum_{I=1}^{N} Q_I = 0 \; , \tag{8.216}$$

and the color charge squared

$$\sum_{I=1}^{N} Q_I^2 \tag{8.217}$$

is fixed. Then (8.111) shows that the most energetically advantageous field configuration is such that the color charges of quarks are lined up into a fixed direction in the color space, thereby reducing $SU(\mathcal{N})$ to $SU(2)$. This is a sort of Bose–Einstein condensation in the color space. Eventually the color space in the hot phase is effectively specified by $SU(2)$, and the color symmetry specification of quantum chromodynamics, $SU(3)$, is almost regained. One may hope that a more careful treatment of the Yang–Mills–Wong theory or its modifications can give just $SU(3)$.

[7] The experiment intimates that only about 20% of the nucleon's spin is produced by quark's spin. The missing 80% of the spin may come from gluon spins and the orbital angular momentum due to the motion of quarks and gluons within the nucleon. This adds considerable support for the assumption that Yang–Mills–Wong particles mimic real quarks in hadrons to a good approximation.

The reader will find it curious that the roles of the 'cold' and 'hot' are now interchanged in comparison with the usual order of things. We associate low-temperature phenomena, such as superfluidity or superconductivity, with quantum physics, while the cold (that is, spontaneously deformed) phase in the Yang–Mills–Wong theory is basically classical. On the other hand, we are aware of classical character of the great bulk of phenomena at room temperature, while the hot phase of the Yang–Mills–Wong theory is inherently quantum mechanical.

Problem 8.7.1. SL$(4,\mathbb{R})$ has 15 traceless generators $I_{\mu\nu}$, $\mu,\nu = 0,1,2,3$, obeying the commutation relations

$$[I_{\mu\nu}, I_{\rho\sigma}] = i\,\eta_{\nu\rho}\,I_{\mu\sigma} - i\,\eta_{\mu\sigma}\,I_{\rho\nu}\ . \tag{8.218}$$

The set of generators $I_{\mu\nu}$ can be divided into two subsets of 6 antisymmetric generators $M_{\mu\nu}$ and 9 symmetric traceless generators $T_{\mu\nu}$, which can be further regrouped according to (8.206). Verify that the commutation relations (8.198)–(8.200) and (8.207)–(8.212) follow from those of (8.218).

Problem 8.7.2. Show that the matrices (8.213)–(8.215) obey the commutation relations (8.198)–(8.200) and (8.207)–(8.212).

Problem 8.7.3. Consider the mass quadrupole operator of a N-particle cluster

$$Q_{ij} = \sum_{I=1}^{N} m_I \left(\mathbf{x}_i^I\,\mathbf{x}_j^I - \frac{1}{3}\,\mathbf{x}_I^2\,\delta_{ij}\right)\ , \tag{8.219}$$

and its time derivative

$$\dot{Q}_{ij} = \sum_{I=1}^{N} \left[\mathbf{x}_i^I\,\mathbf{p}_j^I + \mathbf{x}_j^I\,\mathbf{p}_i^I - \frac{2}{3}\,(\mathbf{x}^I\cdot\mathbf{p}^I)\,\delta_{ij}\right]\ . \tag{8.220}$$

Let us define $T_{ij} = \dot{Q}_{ij}$ and $\mathbf{L} = \sum \mathbf{r}_I \times \mathbf{p}_I$. Verify that these T_{ij} and L_i obey the commutation relations (8.198)–(8.200), taking into account the canonical commutation relations $[\mathbf{x}_i^I, \mathbf{p}_j^J] = -i\,\delta_{ij}\delta_{IJ}$.

Notes

1. The literature on exact solutions of classical Yang–Mills theories is extensive and the reader interested in this topics would do best to consult the books of Rajaraman (1982), Coleman (1985), Schwarz (1991), and Rubakov (2002). Actor (1979) is a largely complete review of classical solutions to SU(2) gauge theories that were known by the end of the 1970s. A detailed treatment of monopole solutions is given in Goddard & Olive (1978), Jaffe & Taubes

(1980), Coleman (1983), and Atiyah & Hitchin (1988). We limit our discussion primarily to retarded solutions of the Yang–Mills–Wong theory, and give a cursory glance at some other frequently cited solutions.

A general grasp of the static quark model of hadrons can be gained from the books by Close (1979), and Perkins (1972). The string mechanism of confinement, originating from Nielsen & Olesen (1973), became the dominant paradigm due to the papers by Nambu (1974, 1976), and Wilson (1974). Mandelstam (1980) and Bander (1981) cover many aspects of this mechanism. The reader will become aware that the solution of the four-dimensional quantum Yang–Mills theory, including the confinement problem as its key ingredient, has been chosen to be a 'Millennium Problem' by the Clay Mathematical Institute. The precise formulation of this problem is given by Jaffe & Witten (1999).

The current understanding of phase transitions in quantum chromodynamics, based on phenomenological models, numerical lattice simulations, and experimental data on heavy-ion collisions, is reviewed in Meyer-Ortmanns (1996).

2. *Section 8.1.* The procedure of solving the Yang–Mills–Wong theory in the single-quark case here follows a similar route as that in Kosyakov (1998); for the original derivation of the retarded field configuration (8.30) see Kosyakov (1991).

3. *Section 8.2.* The ansatz (8.8), proposed in 1991, was further generalized to yield expression (8.62) in the 1994, 1998, and 1999 papers of the author. It may well be that the technique under discussion can be applied to many more problems than those discussed in this book.

4. *Section 8.3.* The analysis of the retarded solution in the two-quark case is adopted from Kosyakov (1998); for a full-length derivation of this solution see Kosyakov (1994).

5. *Section 8.4.* The discussion of the Yang–Mills fields generated by N quarks follows by Kosyakov (1998). The fact that bound quarks forming a hadron are not affected by colorforces, and move along parallel straight world lines, as deduced from these solutions, is consistent with the intuitive idea that the ground state of the hadron should exhibit zero orbital momentum. A similar situation occurs in the dynamics of monopoles. As was shown by Manton (1977), the Bogomol'ny–Prasad–Sommerfield monopoles forming a static multi-monopole are unaffected by intermonopole forces, which exactly cancel between the repulsive Yang–Mills force and the attractive Higgs force.

6. *Section 8.5.* For the statement of the stability problem in electrodynamics and Yang–Mills theory see Mandula (1976), Jackiw & Rossi (1980), and Rajaraman (1982). The present discussion is based on results of the paper by Kosyakov (1998).

7. *Section 8.6.* Chew & Frautschi (1961) plotted hadronic mass squared versus spin to display the Regge trajectories graphically. Veneziano (1968) proposed the use of dual resonance amplitudes in phenomenological description of strong interactions. For a review of dual models see Veneziano (1974).

This section gives an overview of some exact solutions to Abelian and non-Abelian Higgs models, bearing on the confinement problem. A key observation is that spontaneous symmetry breaking renders the Higgs field similar to a superconducting medium which exhibits the Meissner effect. Weak magnetic fields are shielded, while strong fields survive in vortices penetrating the superconductor. Vortices in a type II superconductor were predicted by Abrikosov (1957) from the phenomenological theory of superconductivity by Ginzburg & Landau (1950). Nielsen & Olesen (1973) argued for the existence of vortex-line solutions in Higgs models, and identified them with the Nambu string. Based on this discovery, Creutz (1974), Nambu (1974), and Parisi (1975) suggested that, in theories where spontaneous symmetry breaking occurs, and quarks are endowed with magnetic (or, better to say, chromomagnetic) charges, vortices may provide a mechanism for quark confinement. However, presently the idea of 'electric quark' confinement (rather than the 'magnetic quark' confinement) appears to have considerable promise. In the lattice gauge theory, the chromoelectric flux is forced to lie along links of the lattice, and this bond holds quarks together.

Non-Abelian static solutions to the pure SU(2) gauge theory can be obtained using the ansatz, proposed by Wu & Yang (1969),

$$A_0^a = i\,\frac{x_a\,a(r)}{gr^2}, \quad A_i^a = \frac{\epsilon_{aij}\,x_j\,[1-b(r)]}{gr^2}\ , \tag{8.221}$$

which reduces the field equations to the form

$$r^2 a'' = 2ab^2, \quad r^2 b'' = b\left(b^2 - 1 + a^2\right)\ . \tag{8.222}$$

Various modifications of the Wu–Yang ansatz form the basis for a major part of known non-Abelian solutions. Note, however, that when one seeks a version of the Wu–Yang ansatz (in which spacetime coordinates are interwoven with color coordinates) for the equations of motion in the Yang–Mills–Wong theory, one runs into trouble with infinitely fast precession of the color charge encountered in Sect. 8.1.

Polyakov (1974) and 't Hooft (1974) analyzed the Yang–Mills–Higgs theory, using an ansatz similar to that of Wu and Yang, and discovered a field configuration possessing properties of the magnetic monopole. Prasad & Sommerfield (1975), and Bogomol'ny (1976) studied this field configuration in the limit $\mu \to 0$, $\lambda \to 0$, with $\mu/\lambda < \infty$, and obtained an exact solution, the Bogomol'ny–Prasad–Sommerfield monopole. It differs in many respects from the 't Hooft–Polyakov monopole, particularly regarding its distinctive property of self-duality. Julia & Zee (1975) modified the 't Hooft ansatz, and obtained a solution with both electric and magnetic charges, presently known as the Julia–Zee dyon. Jackiw & Rebbi (1976), and Hasenfratz & 't Hooft (1976) showed that isospinor degrees of freedom are converted into spin degrees of freedom in the field of magnetic monopoles. This issue was further developed by Jackiw & Manton (1980).

The material of this section is a slightly rearranged exposition of these pioneer papers, which are probably the best introduction to the subject.

8. *Section 8.7.* It is generally believed that understanding the Yang–Mills ground state may be of crucial importance in the confinement problem. Coleman (1966) argued that invariance of the vacuum is the invariance of world. The analogy between the Higgs model and the Ginzburg–Landau model of a superconductor was observed by Kirzhnits (1972), and Kirzhnits & Linde (1972), who pointed out that the classical Higgs field at the equilibrium state (corresponding to a minimum of the Higgs potential) can be interpreted as a Bose condensate of Higgs quanta arising below the critical point. The phenomenon of spontaneously deformed gauge symmetry was discovered by Kosyakov (1994, 1998).

Dothan, Gell-Mann & Ne'eman (1965) proposed to describe the Regge trajectories of hadrons by infinite-dimensional unitary representations of $SL(3,\mathbb{R})$. They found that two infinite unitary representations belonging to the ladder series may be associated with meson trajectories. However, Ogiyevetsky & Sokachev (1975) showed that this $SL(3,\mathbb{R})$ scheme is inadequate to account for the Regge trajectories of baryons. Ne'eman & Šijački (1985, 1988, 1993) proved that just $SL(4,\mathbb{R})$ gives an exhaustive phenomenological classification of hadrons.

The puzzle of nucleon spin is discussed comprehensively in the Stiegler's 1996 survey. Data on deep inelastic scattering of polarized electrons and muons on proton and deuteron targets evidence that about 20% of nucleon spin is carried by quark spin. The remainder of the nucleon spin is thought to be due to the quark orbital angular momenta and gluon helicity.

9

Self-Interaction in Gauge Theories

9.1 Rearrangement of the Yang–Mills–Wong Theory

Following the procedure of Sects. 6.2 and 6.3, we take, as a starting point, the action for a system of a single quark interacting with a SU(2) gauge field

$$S = -\int ds \left\{ m_0\sqrt{v\cdot v} + \sum_a q^a \eta_i^* \left[\delta^i{}_j \frac{d}{ds} + v^\mu (A^a{}_\mu T_a)^i{}_j \right] \eta^j \right\}$$

$$-\frac{1}{16\pi}\int d^4x\, G_a^{\mu\nu} G^a_{\mu\nu} \;, \tag{9.1}$$

$$G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + i f^a{}_{bs} A^b_\mu\, A^c_\nu \;. \tag{9.2}$$

Here, T_a are generators of SU(2) expressed in terms of the Pauli matrices, $T_a = \frac{1}{2}\,\sigma_a$, and $f_{abc} = \epsilon_{abc}$ are the structure constants of SU(2).

This system is governed by the equations of motion

$$\dot{Q} - ig\,[Q,\, v^\mu A_\mu] = 0 \;, \tag{9.3}$$

$$\varepsilon^\lambda = m_0 a^\lambda - \mathrm{tr}\left(Q\, G^{\lambda\mu}\right) v_\mu = 0 \;, \tag{9.4}$$

$$\mathcal{E}_\mu = D^\lambda G_{\lambda\mu} - 4\pi j_\mu = 0 \;. \tag{9.5}$$

Here, $Q = -igQ^aT_a$, $A_\mu = (i/g)A^a_\mu T_a$, etc., and

$$j_\mu(x) = \int_{-\infty}^{\infty} ds\, Q(s)\, v_\mu\, \delta^4\left[x - z(s)\right] \;. \tag{9.6}$$

We are looking for solutions to the Yang–Mills equations (9.5) satisfying the condition

$$Q(s) = \mathrm{const} \;. \tag{9.7}$$

(Abandoning this condition would pose the problem of an infinitely rapid precession of the color charge Q^a.) Let the quark be moving along an arbitrary timelike smooth world line $z^\mu(s)$. We then have two alternative solutions, Abelian and non-Abelian,

$$A^\mu = qT_3\,\frac{v^\mu}{\rho}\,, \tag{9.8}$$

and

$$A^\mu = \mp\frac{2i}{g}\,T_3\,\frac{v^\mu}{\rho} + i\kappa\,(T_1 \pm iT_2)\,R^\mu\,. \tag{9.9}$$

Here q and κ are arbitrary real nonzero parameters.

Our concern is with simultaneous solution of equations (9.4) and (9.5). However, the retarded fields (9.8) and (9.9) are singular. Their substitution into (9.4) would result in divergent expressions. Therefore, we should regularize (9.8) and (9.9). To proceed further, it is convenient to use the Noether identity (7.56),

$$\partial_\mu\left(\Theta^{\lambda\mu} + t^{\lambda\mu}\right) = \frac{1}{4\pi}\,\mathrm{tr}\left(\mathcal{E}_\mu\,G^{\mu\lambda}\right) + \int ds\,\varepsilon^\lambda(s)\,\delta^4\left[x - z(s)\right]\,. \tag{9.10}$$

Here,

$$\Theta^{\lambda\mu} = \frac{1}{4\pi}\,\mathrm{tr}\left(G^{\lambda}{}_{\alpha}G^{\alpha\mu} + \frac{\eta^{\lambda\mu}}{4}\,G^{\alpha\beta}G_{\alpha\beta}\right)\,, \tag{9.11}$$

and $t^{\lambda\mu}$ is given by (6.78). $\mathcal{E}_\mu$ and ε^λ are the left-hand sides of (9.5) and (9.4), respectively.

If we take (9.8), then all results of Maxwell–Lorentz theory are reproduced with the only replacement $e^2 \to q^2$. The degrees of freedom appearing in the action (9.1) are rearranged to give a dressed quark and Yang–Mills radiation. The basic properties of the Yang–Mills radiation are similar to those of Maxwell radiation, in particular, the four-momentum of Yang–Mills radiation is given by

$$\mathcal{P}^\mu = -\frac{2}{3}\,q^2\int_{-\infty}^{s} d\tau\,a^2 v^\mu\,. \tag{9.12}$$

A dressed quark possesses four-momentum

$$p^\mu = mv^\mu - \frac{2}{3}\,q^2 a^\mu\,. \tag{9.13}$$

Its evolution is governed by the Lorentz–Dirac equation

$$ma^\mu - \frac{2}{3}\,q^2\left(\dot{a}^\mu + v^\mu a^2\right) = f^\mu\,. \tag{9.14}$$

Here, m is the renormalized mass, and f^μ an external force.

A different situation arises with solution (9.9). We first note that this field is generated by the color charge

$$Q = \pm\frac{2i}{g}\,T_3\,. \tag{9.15}$$

Therefore, a single quark has color charge of *fixed* magnitude

$$Q^2 = -\frac{4}{g^2}\,. \tag{9.16}$$

Let us substitute (9.9) into (9.11). The linearly rising term of A_μ does not contribute to $\Theta^{\mu\nu}$. Therefore, $\Theta^{\mu\nu}$ splits into two parts: $\Theta^{\mu\nu} = \Theta^{\mu\nu}_{\rm I} + \Theta^{\mu\nu}_{\rm II}$, with $\Theta^{\mu\nu}_{\rm I}$ and $\Theta^{\mu\nu}_{\rm II}$ being expressed, respectively, by (6.81) and (6.82), in which, however, e^2 is replaced by Q^2.

It follows from (9.16) that the self-energy is negative (Problem 9.1.1),

$$\delta m = -\frac{2}{g^2\epsilon} < 0\,. \tag{9.17}$$

If m is to be finite, m_0 must be positive.

Owing to (9.16), the emitted energy is negative. This implies that an accelerated quark gains, rather than loses, energy by emitting the Yang–Mills radiation. An explicit calculation (Problem 9.1.2) shows that this is indeed the case:

$$\dot{\mathcal{P}}\cdot v = \frac{4}{g^2}\,a^2 < 0\,. \tag{9.18}$$

It was already mentioned in Sect. 6.2 that this phenomenon may be interpreted as absorbing convergent waves of positive energy rather than emitting divergent waves of negative energy.

We now turn to the N-quark case. It was concluded in Chap. 8 that a consistent Yang–Mills–Wong theory can be formulated for the gauge group $SU(\mathcal{N})$ with $\mathcal{N} \ge N+1$.

There are two classes of retarded solutions to the Yang–Mills equations, Abelian and non-Abelian. If we take the former, then all results of Sect. 6.3 extend to the N-quark case. It remains to consider the situation with non-Abelian solutions. To be specific, let us refer to the field (8.85) generated by two quarks,

$$A_\mu = \mp\frac{2i}{g}\left(H_1\,\frac{v^1_\mu}{\rho_1} + g\,\kappa\,E^{\pm}_{13}\,R^1_\mu\right) \mp \frac{2i}{g}\left(H_2\,\frac{v^2_\mu}{\rho_2} + g\,\kappa\,E^{\pm}_{23}\,R^2_\mu\right)\,. \tag{9.19}$$

We note that A_μ is the sum of two single-quark terms. Due to this feature – which is characteristic of the general N-quark case – we have

$$\Theta^{\mu\nu} = \sum_I \Theta^{\mu\nu}_I + \sum_I\sum_J \Theta^{\mu\nu}_{IJ}\,, \tag{9.20}$$

where $\Theta^{\mu\nu}_I$ is comprised of the field generated by the Ith quark, and $\Theta^{\mu\nu}_{IJ}$ contains mixed contributions of the fields due to the Ith and Jth quarks. We thus come to a stress-energy tensor $\Theta^{\mu\nu}$ which is similar in structure to that of Maxwell–Lorentz theory.

We integrate $\Theta_I^{\mu\nu}$ following the procedure of Sect. 6.3. All results obtained in the single-quark case remain unchanged, except that

$$\mathrm{tr}\,(Q_I^2) = -\frac{4}{g^2}\left(1 - \frac{1}{\mathcal{N}}\right) . \tag{9.21}$$

N dressed quarks arise possessing four-momenta

$$p_I^\mu = m_I\,(v_I^\mu + \ell_I a_I^\mu) . \tag{9.22}$$

An accelerated quark emits Yang–Mills radiation carrying four-momentum

$$\mathcal{P}_I^\mu = m_I \ell_I \int_{-\infty}^{s_I} d\tau_I\, a_I^2 v_I^\mu . \tag{9.23}$$

Here, m_I is the renormalized mass of the Ith quark, and

$$\ell_I = \frac{8}{3m_I g^2}\left(1 - \frac{1}{\mathcal{N}}\right) . \tag{9.24}$$

Because

$$\mathrm{tr}\,(H_l\, E_{mn}^{\pm}) = 0, \quad \mathrm{tr}\left(E_{kl}^{\pm}\, E_{mn}^{\pm}\right) = 0 , \tag{9.25}$$

the linearly rising term of A_μ does not contribute to $\Theta^{\mu\nu}_{IJ}$. Integration of $\Theta^{\mu\nu}_{IJ}$ gives the four-momentum produced by an external Wong force (Problem 9.1.4)

$$\wp_I^\mu = -\int_{-\infty}^{s_I} d\tau_I\, f_I^\mu(z_I) , \tag{9.26}$$

$$f^\mu_I(z_I) = \sum_J \frac{\mathrm{tr}\,(Q_I Q_J)}{[v_J\cdot(z_J - z_I)]^2}\left[(V_J\cdot v_I)\, c_J^\mu - (c_J\cdot v_I)\, V_J^\mu\right] . \tag{9.27}$$

Imposing the Haag asymptotic condition

$$\lim_{s_I\to-\infty} a_I^\mu(s_I) = 0 \tag{9.28}$$

(and omitting quark labelling), we come to the equation for local energy-momentum balance

$$\dot{p}^\mu + \dot{\mathcal{P}}^\mu = f^\mu . \tag{9.29}$$

According to this balance relation the four-momentum $d\wp^\mu = -f^\mu ds$ extracted from an external field is used for changing the four-momentum of a dressed quark dp^μ and emitting the Yang–Mills radiation four-momentum $d\mathcal{P}^\mu$. A special feature of (9.29) is that $d\mathcal{P}^0$ is associated with emanating negative-energy waves (or, absorbing positive-energy waves).

Note that (9.29) holds universally in both the Maxwell–Lorentz and Yang–Mills–Wong theories. An intuitive idea behind this equation is as follows. An external force exerted on a dressed particle causes it to emit (or, alternatively, to absorb) radiant energy $d\mathcal{P}^0$. The work done by this force $f^0 ds$ minus $d\mathcal{P}^0$ equals the change in energy of the dressed particle dp^0.

Substituting (9.22) (9.23), and (9.26) into (9.29) gives the equation of motion for a dressed quark

$$m\left[a^\mu + \ell\left(\dot{a}^\mu + v^\mu a^2\right)\right] = f^\mu\,. \tag{9.30}$$

We see that (9.30) differs from the Lorentz–Dirac equation

$$m\left[a^\mu - \tau_0\left(\dot{a}^\mu + v^\mu a^2\right)\right] = f^\mu \tag{9.31}$$

only in the overall sign of the parenthesized term.

To appreciate this difference, we consider the case $f^\mu = 0$. For any initial v^μ and a^μ, there exists a Lorentz frame in which v^μ is aligned with the t-axis, and a^μ is parallel to the z-axis. It is clear then that the world line of the free motion lies in this plane. One can readily check that the general solution to (9.30) is

$$v^\mu(s) = V^\mu \cosh(\alpha_0 + w_0\ell\, e^{-s/\ell}) + U^\mu \sinh(\alpha_0 + w_0\ell\, e^{-s/\ell})\,. \tag{9.32}$$

Here, V^μ and U^μ are constant four-vectors such that $V\!\cdot\! U = 0$, $V^2 = -U^2 = 1$, and α_0 and w_0 are arbitrary parameters. The asymptotic condition (9.28) (which is essential for the derivation of the equation of motion) can be fulfilled only for $w_0 = 0$. Therefore, a free dressed quark governed by (9.30) moves along a straight line $v^\mu(s) = \text{const}$.

On the other hand, the general solution to (9.31) is

$$v^\mu(s) = V^\mu \cosh(\alpha_0 + w_0\tau_0 e^{s/\tau_0}) + U^\mu \sinh(\alpha_0 + w_0\tau_0 e^{s/\tau_0})\,. \tag{9.33}$$

This solution (which exhibits *self-acceleration* or *runaway behavior*) obeys asymptotic condition (9.28) for any w_0. Therefore, a free a dressed particle governed by (9.31), can behave as a non-Galilean object. We reserve this issue for Sect. 9.3.

To summarize, there are two phases of the Yang–Mills–Wong theory, hot and cold. In both phases, the initial degrees of freedom are rearranged to give N dressed quarks and Yang–Mills radiation having the required properties (6.87)–(6.89). The hot phase is realized on the Cartan subgroup of the gauge group. In other words, the Yang–Mills field generated by quarks is Abelian. Hence, the rearranded degrees of freedom resemble those of the Maxwell–Lorentz theory in every detail. On the other hand, the Yang–Mills field generated by quarks in the cold phase is non-Abelian. The gauge group SU($\mathcal{N}$) is spontaneously deformed to SL($\mathcal{N},\mathbb{R}$). An accelerated quark emits Yang–Mills radiation in the form of diverging negative-energy waves. The equation of motion for a dressed quark (9.30) differs from the Lorentz–Dirac equation in the overall sign of the higher-derivative term.

Problem 9.1.1. Verify (9.17).

Problem 9.1.2. Evaluate $\dot{\mathcal{P}}^\mu$ for the Yang–Mills field (9.9). Verify (9.18).

Problem 9.1.3. Derive (9.22) and (9.23).

Problem 9.1.4. Verify (9.26)–(9.27).

Problem 9.1.5. Show that (9.32) is the general solution to equation (9.30) with $f^\mu = 0$.

9.2 Self-Consistency

In analyzing the self-interaction problem in the Maxwell–Lorentz and Yang–Mills–Wong theories, we saw that some quantities, such as δm, are *divergent.* These divergences originate from the fact that electrons and quarks are considered as *point* particles, and the interactions between such particles and fields are *local.* Consequently, electromagnetic and Yang–Mills fields are *singular* in the vicinity of their sources. Does this result signify that the idea of point particles locally coupled with fields is self-contradictory?

To gain a better insight into the divergence problem, let us look at the static case. Referring to Problem 6.3.1, we write

$$\delta m = \int d^3x \, \frac{\mathbf{E}^2(\mathbf{x})}{8\pi} \, . \tag{9.34}$$

The Fourier transform of the Coulomb field $\mathbf{E}(\mathbf{x})$ is

$$\mathbf{E}(\mathbf{x}) = \frac{1}{(2\pi)^3} \int d^3k \, e^{i\mathbf{k}\cdot\mathbf{x}} \, \widetilde{\mathbf{E}}(\mathbf{k}) \, . \tag{9.35}$$

Substituting (9.35) into (9.34) gives

$$\frac{1}{8\pi\,(2\pi)^3} \int d^3k \int d^3q \, \widetilde{\mathbf{E}}(\mathbf{k}) \cdot \widetilde{\mathbf{E}}(\mathbf{q}) \frac{1}{(2\pi)^3} \int d^3x \, e^{i(\mathbf{k}+\mathbf{q})\cdot\mathbf{x}}$$

$$= \frac{1}{8\pi\,(2\pi)^3} \int d^3k \int d^3q \, \widetilde{\mathbf{E}}(\mathbf{k})\cdot\widetilde{\mathbf{E}}(\mathbf{q}) \, \delta^3(\mathbf{k}+\mathbf{q}) = \frac{1}{8\pi\,(2\pi)^3} \int d^3k \, \widetilde{\mathbf{E}}(\mathbf{k})\cdot\widetilde{\mathbf{E}}(-\mathbf{k}) \, . \tag{9.36}$$

With

$$\widetilde{\mathbf{E}}(\mathbf{k}) = -i\mathbf{k} \, \frac{4\pi e}{k^2} \, , \tag{9.37}$$

we obtain

$$\delta m = \frac{e^2}{4\pi^2} \int \frac{d^3k}{k^2} = \lim_{\Lambda\to\infty} \frac{e^2}{\pi} \int_0^\Lambda dk \, . \tag{9.38}$$

We see that the short-distance limit $\epsilon \to 0$ corresponds to the limit of high Fourier modes $\Lambda \to \infty$. It is clear that (9.34) is divergent for any method of calculation. In field-theoretic language, expression (9.38) is said to be *ultraviolet divergent*. The name derives from the optical spectrum where 'violet' is associated with high frequencies.

To handle divergent quantities, a regularization procedure must be introduced. This means that the theory should be provisionally modified at small distances so as to make the integrals convergent. Examples of invariant regularization procedures were given in Sects. 6.3 and 6.5. In the regularized theory, we segregate finite terms from divergent ones. The latter are absorbed by redefining physical parameters of the action. Such a redefinition was discussed for mass renormalization.

The identification

$$m = \lim_{\epsilon \to 0} \left[m_0(\epsilon) + \delta m(\epsilon) \right] \tag{9.39}$$

was found to be sufficient to make the Maxwell–Lorentz and Yang–Mills–Wong theories free of divergences. In general, a theory is called *renormalizable* if it is possible to absorb all divergences by a redefinition of physical parameters in the Lagrangian. A characteristic feature of renormalizable theories is that the physics at large distances is insensitive to what happen at short distances: the entire short-distance effect is incorporated in renormalizations of free parameters.

Renormalizability became a criterion for theory selection in the late 1940s, and this state of affairs endured for about three decades. It is remarkable that this scientific strategy was crucial for the development of quantum electrodynamics and the standard model of particle physics.

However, by the early 1980s, the principle of renormalizability had been revised (mainly due to the influential work of Wilson). The search for the 'ultimate Lagrangian', which would be an accurate renormalizable description down to arbitrarily short distances, is no longer a goal of most physicists. A new paradigm of *effective field theories* has emerged instead. Any field theory is now regarded as the provisional, approximate description of some area of physical reality, useful only in a prescribed range of distance or energy. An effective theory is formulated in terms of only those degrees of freedom which are actually important within the limits of validity of this description. If we wish to go beyond these limits the theory must be replaced by a new, more accurate theory, involving other degrees of freedom. For example, quantum electrodynamics is an effective theory resulting from the standard model at distances much greater than the cutoff set by the Compton wavelength of the massive vector bosons , $\lambda_W = 1/M_W \approx 2 \cdot 10^{-16}$ cm.

The Maxwell–Lorentz theory may be understood as an effective theory resulting from quantum electrodynamics at long distances. Its validity is limited by the cutoff related to the Compton wavelength of the electron, $\lambda_e = 1/m_e = 3.86 \cdot 10^{-11}$ cm. Indeed, at these and shorter distances, the effects of pair creation become appreciable. Likewise, we may regard the Yang–Mills–Wong theory as an effective theory, and associate the cutoff with the Compton wavelength of the quark, $\lambda_q = 1/m_q$.

Historically, the nonrelativistic equation of motion for a dressed electron (6.173) was derived by Lorentz and Abraham using a model of the electron as a rigid body of finite extent. They tried to find the force of the electron on itself, that is, the net force exerted by the fields due to different parts of the charge distribution acting on one another. The electron was thought of as a charged sphere of diameter d such that the entire mass of the electron was of electromagnetic origin: $m = \delta m$. This leads to the relation $d = e^2/m$. For other charge distributions, we have similar relations, say, for a homogeneously charged ball, $d = 6e^2/5m$. The characteristic length arising in this approach

$$r_0 = \frac{e^2}{m} = 2.8 \cdot 10^{-13}\,\text{cm}\,, \tag{9.40}$$

is called the *classical radius of the electron*. The interval τ_0 defined in (6.165) may be treated as the time it takes for a light signal to travel through the 'interior of the electron'.

Of course, Maxwell–Lorentz theory ceases to be valid at distances comparable with r_0 because the cutoff adopted in this theory, λ_e, is two orders of magnitude longer than r_0. A completely different situation occurs in the cold phase of the Yang–Mills–Wong theory, where the characteristic length ℓ is given by (9.24). Unlike r_0, which is proportional to $e^2 \sim 1/137$, we see that ℓ is inversely related to g^2. For $g \sim 1$, we have $\ell \sim \lambda_q$. However, if $g \ll 1$, then $\ell \gg \lambda_q$, so that phenomena specified by ℓ are within the range of validity of this effective theory. For example, using (9.22), we get

$$p^2 = m^2\left(1 + \ell^2 a^2\right)\,. \tag{9.41}$$

Therefore, the dressed quark becomes a tachyon if its acceleration exceeds the critical value $1/\ell$. (A plausible explanation for this is that the critical point corresponds to a phase transition between cold and hot phases.)

Why should we care about the Maxwell–Lorentz and Yang–Mills–Wong theories? They are not accurate descriptions of real electrons and quarks anyway. The reason is simple. We need *solvable* four-dimensional theories. These toy-model descriptions have received much study because they offer *exact* solutions. A detailed look at these solutions may be of great importance for a deep insight into the mathematical properties of realistic gauge theories. For example, the Yang–Mills–Wong theory is the only known mathematical framework in which spontaneous symmetry deformation can be revealed explicitly. This phenomenon may be useful in exploring phase transitions of subnuclear matter.

9.3 Paradoxes

Many people have formed an impression of the Lorentz–Dirac equation

$$ma^\mu - \frac{2}{3}\,e^2\left(\dot{a}^\mu + v^\mu a^2\right) = f^\mu\ , \tag{9.42}$$

as the equation of motion for a charged particle whose four-momentum is defined by

$$p^\mu = mv^\mu\ . \tag{9.43}$$

The particle is assumed to experience both an external force f^μ and *radiation reaction* (or *radiation damping*) force

$$\Gamma^\mu = \frac{2}{3}\,e^2\left(\dot{a}^\mu + v^\mu a^2\right)\ . \tag{9.44}$$

This interpretation of the Lorentz–Dirac equation leads to many paradoxes and puzzles.

Let us consider some of them. We will show that the key to all such problems lies with Teitelboim's concept of a *dressed* particle possessing four-momentum

$$p^\mu = mv^\mu - \frac{2}{3}\,e^2 a^\mu\ . \tag{9.45}$$

We begin with the best known issue, namely, that the Lorentz–Dirac equation is inconsistent with local energy-momentum balance. Indeed, the so-called *Schott term*

$$\frac{2}{3}\,e^2\,\dot{a}^\mu \tag{9.46}$$

prevents Γ^μ from being regarded as what may be properly called the radiation reaction

$$-\dot{\mathcal{P}}^\mu = \frac{2}{3}\,e^2\,v^\mu a^2\ . \tag{9.47}$$

Naively, one expects that a radiating particle feels a recoil measured by the negative of the emission rate (9.47). However, $-\dot{\mathcal{P}}^\mu$ is not a four-force, because it is not orthogonal to v^μ. On the other hand, if we take Γ^μ (which is orthogonal to v^μ) as the radiation reaction, then energy-momentum balance becomes problematic. To see this, write the temporal component of (9.42) in the form

$$\mathbf{F}\cdot\mathbf{v} = \frac{d}{dt}\,m\gamma - \frac{2}{3}\,e^2\,a^2 - \frac{2}{3}\,e^2\,\frac{d}{dt}\,a^0\ . \tag{9.48}$$

One usually states: the rate at which the external force $\mathbf{F}$ does work on the particle is equal to the increase in the particle's kinetic energy, plus the energy radiated, plus the energy stored in the Schott term. Although the energy stored in the Schott term can be attributed to a reversible form of emission and absorption of field energy, its actual role appears mysterious.

To remedy the situation, let us impose the asymptotic condition

$$\lim_{s\to\pm\infty} a^\mu(s) = 0\ . \tag{9.49}$$

Because the Schott term is a perfect differential, (9.42) can be integrated to give

$$mv^\mu(\infty) - mv^\mu(-\infty) - \frac{2}{3}\,e^2 \int_{-\infty}^{\infty} ds\, a^2 v^\mu = \int_{-\infty}^{\infty} ds\, f^\mu \,. \tag{9.50}$$

It may appear that (9.50) provides a satisfactory solution to the problem. However, this conclusion does not stand up. Why is energy-momentum only conserved globally? There is nothing in Maxwell–Lorentz theory which suggests that electromagnetic interactions are nonlocal by their nature and, hence, that local energy-momentum balance is impossible.

Following an accurate regularization prescription, we established in Sect. 6.3 that the four-momentum of a dressed charged particle is given by (9.45) rather than by (9.43)[1]. Recall that p^μ was obtained by integrating the sum of the stress-energy tensor for a bare particle $t^{\mu\nu}$ and that for the bound part of electromagnetic field $\Theta_{\rm I}^{\mu\nu}$. The quantities $t^{\mu\nu}$, $\Theta_{\rm I}^{\mu\nu}$, and $\Theta_{\rm II}^{\mu\nu}$ are dynamically independent off the world line. On a particle's world line the energy-momentum of an external field is converted into that of the dressed particle and radiation:

$$\dot p^\mu + \dot{\mathcal{P}}^\mu = f^\mu \,. \tag{9.51}$$

As was shown in Sect. 9.1, the local balance (9.51) holds for appropriately rearranged degrees of freedom in Yang–Mills–Wong theory.

Closely related to this balance problem is the paradox of uniform acceleration. With reference to Problem 2.1.4, we invoke a covariant condition for uniform acceleration

$$\dot a^\mu + a^2 v^\mu = 0 \,. \tag{9.52}$$

Therefore, when a charged particle is uniformly accelerated, (9.42) becomes

$$m a^\mu = f^\mu \,. \tag{9.53}$$

It is clear that uniform accelerations are due to constant forces $f^\mu = \text{const.}$ This could be realized by a huge capacitor with flat parallel plates if the charge moves along the field. The paradoxial thing is that a uniformly accelerated charged particle, while emitting electromagnetic radiation, experiences no back-reaction. In addition, it seems strange that the case $f^\mu = \text{const.}$ is somewhat physically distinguished.

No paradox arises for a dressed particle. We saw in Sect. 6.4 that the Lorentz–Dirac equation can be brought to the form

$$\overset{v}{\perp}(\dot p - f) = 0 \,. \tag{9.54}$$

[1] Ascribing $p^\mu = mv^\mu$ to an object governed by the Lorentz–Dirac equation has no justification except 'to keep an analogy with mechanics'. When $p^\mu = mv^\mu$ is nonetheless insisted upon the corresponding stress-energy tensor $\mathfrak{T}^{\mu\nu}$ should also be explored. As can be readily verified, local conservation law $\partial_\mu \mathfrak{T}^{\mu\nu} = 0$ does not hold off the world line. Furthermore, regarding $T^{\mu\nu} - \mathfrak{T}^{\mu\nu}$ as the radiation part of the the stress-energy tensor would compromise the characteristic properties of the radiation (6.87) and (6.88).

The structure of this equation shows that a dressed particle experiences only an external force f^μ. Note also that (9.51) contains no 'recoil term'. The energy-momentum which is extracted from the external field is used to change the energy-momentum of the dressed particle and to emit radiation. It is impossible to identify the reactive effect of radiation in these equations for any motion, uniform acceleration is no exception.

The paradox can be formulated in another way. Suppose that neutral and charged particles, which have identical masses m, are moving along a straight line under a constant force f^μ. As before,

$$v^\mu = (\cosh\alpha, 0, 0, \sinh\alpha) \ , \tag{9.55}$$

$$f^\mu = f\,(\sinh\alpha, 0, 0, \cosh\alpha) \ . \tag{9.56}$$

For example, imagine that they fall to the surface of the Earth. Both are attracted toward the Earth by an approximately constant force $f = -mg$, where g is the acceleration of gravity. Using (9.55) and (9.56), we obtain the equation of motion for a charged particle

$$\dot\alpha - \tau_0\,\ddot\alpha = -g \ , \tag{9.57}$$

$$\tau_0 = \frac{2e^2}{3m} \ , \tag{9.58}$$

and that for a neutral particle

$$\dot\alpha = -g \ . \tag{9.59}$$

Evidently, both (9.57) and (9.59) are satisfied by

$$\alpha = -gs \ . \tag{9.60}$$

The paradox is that a given constant force causes both particles move along the same hyperbolic world line

$$z^\mu(s) = z^\mu(0) + g^{-1}\,(\sinh gs, 0, 0, -\cosh gs) \ , \tag{9.61}$$

even if accelerating charged particles radiate. Since this radiation carries off energy, the charged particle might be expected to accelerate less than the neutral one.

Recall, however, that the energy of a neutral particle is positive definite while the energy of a dressed particle is indefinite. Despite the fact that both particles execute identical motions, the energy associated with these motions is different (Problem 9.3.1).

One further concern is with the so-called counter-acceleration. One normally expects that the smallness of the radiation reaction Γ^μ results in small corrections to the essentially Newtonian behavior of a charged particle. But these expectations are not always realized.

Let a charged particle be moving along a straight line. Then, following the procedure of Sect. 6.4, we come to the equation

$$\dot{\alpha}(s) = e^{s/\tau_0} \left[C - \frac{1}{m\tau_0} \int_0^s d\sigma\, e^{-\sigma/\tau_0}\, f(\sigma) \right] , \tag{9.62}$$

where C is an arbitrary initial value of $\dot{\alpha}$ at $s = 0$. Setting $C = 0$, one finds that $\dot{\alpha}$ and f are oppositely directed.

When analyzing the behavior of a dressed particle this result presents no difficulty. Indeed, (9.54) shows that the dynamics of a dressed particle is Newtonian. However, this in no way implies that acceleration must be aligned with force; a^μ and f^μ would have the same direction only if one makes the *ad hoc* assumption that $p^\mu = mv^\mu$.

All this argumentation is (with minor modifications) translated into the Yang–Mills–Wong theory. We thus see that the term 'radiation reaction force' is an unfortunate misnomer both in electrodynamics and Yang–Mills–Wong theory.

We finally turn to the problem of runaways

$$v^\mu(s) = V^\mu \cosh(\alpha_0 + w_0\tau_0 e^{s/\tau_0}) + U^\mu \sinh(\alpha_0 + w_0\tau_0 e^{s/\tau_0}) . \tag{9.63}$$

These solutions of the Lorentz–Dirac equation are an embarrassing feature of the theory: a free charged particle moving along the world line (9.63) continually accelerates,

$$a^2(s) = -w_0^2 \exp\left(2s/\tau_0\right) , \tag{9.64}$$

and, furthermore, continually radiates. This may seem contrary to energy conservation.

With (9.51), the problem is solved immediately. Indeed, taking $f^\mu = 0$, we have

$$\dot{p}^\mu = -\dot{\mathcal{P}}^\mu . \tag{9.65}$$

This result is quite natural: the rate of change of the energy-momentum of a dressed particle is equal to the negative of the emission rate. In this instance, emitting radiation may be thought of as a reactive power source. Recall that p^0 is indefinite,

$$p^0 = m\gamma \left(1 - \tau_0\, \gamma^3\, \mathbf{a}\cdot\mathbf{v}\right) . \tag{9.66}$$

This implies that increasing $\mathbf{v}$ need not be accomplished by increasing p^0. In fact, the energy of a dressed particle executing a runaway motion decreases steadily, which exactly compensates the increase in energy of the electromagnetic field emitted (Problem 9.3.2).

It is generally believed that the runaways are unphysical. The reason for this is that a free electron with exponentially increasing acceleration has never observed experimentally.

Runaway solutions also occur in the presence of external forces. For example, the solution (9.62) describes runaway motion for every value of C, except that defined in (6.182). As in the free case, a dressed particle can only be self-accelerated for a very short period $\sim \tau_0$. Thereafter it becomes a tachyon.

Problem 9.3.1. Consider neutral and dressed charged particles moving along the same hyperbolic world line (9.61) during a finite period from s_1 to s_2. Compare the variations of their energies.

Hint The dressed particle has the energy p^0 which is obtained from (9.45) and (9.55),

$$p^0 = mv^0 - \frac{2}{3}\, e^2 a^0 = m\left[\cosh(gs) - \tau_0 g \sinh(gs)\right] \,. \tag{9.67}$$

Accordingly,

$$\Delta p^0 = p^0(s_2) - p^0(s_1) = m\left[\cosh(gs) - \tau_0 g \sinh(gs)\right]\big|_{s_1}^{s_2} \,. \tag{9.68}$$

The energy radiated during this period is

$$-\frac{2}{3}\, e^2 \int_{s_1}^{s_2} ds\, a^2 v^0 = m\tau_0 g \sinh(gs)\big|_{s_1}^{s_2} \,. \tag{9.69}$$

The sum of (9.68) and (9.69) equals the work W of the force f^μ defined in (9.56),

$$W = \int_{s_1}^{s_2} ds\, f^0 = m \cosh(gs)\big|_{s_1}^{s_2} \,, \tag{9.70}$$

as might be expected from (9.51).

For the neutral particle, $p^0 = mv^0$, and so

$$\Delta p^0 = m \cosh(gs)\big|_{s_1}^{s_2} \,. \tag{9.71}$$

This Δp^0 is equal to W.

Problem 9.3.2. Consider a self-accelerating dressed particle. Show that energy balance

$$p^0(s_2) - p^0(s_1) = -\int_{s_1}^{s_2} ds\, \dot{\mathcal{P}}^0 \tag{9.72}$$

holds for the runaway world line (9.63).

Notes

1. *Section 9.1.* The presentation of this section follows Kosyakov (1991, 1998).

2. *Section 9.2.* Dyson (1949) argued that mass and charge renormalizations are sufficient to absorb ultraviolet divergences in quantum electrodynamics. He showed further that all infinities can be absorbed into a redefinition of parameters in the Lagrangian only for a certain kind of theories. Dyson called them

renormalizable. Since then renormalizability became a criterion for fundamental theory. For historical details and further references see Brown (1993), Schweber (1994), and Cao (1999). The mathematical reason for ultraviolet divergences is that products of singular distributions with coincident points are ill-defined. This fact was pointed out by Bogoliubov & Parasiuk (1957).

Wilson (1979, 1983) developed the idea of effective theories. He noted that physical systems have many length of scales, each scale associated with its own dynamical law. For example, hydrodynamics governs liquids at the macroscopic level, Boltzmann's equations come into play at the molecular level, and so on. In Wilson's words: 'In general, events distinguished by a great disparity in size have little influence on one another; they do not communicate, and so the phenomena associated with each scale can be treated independently.'

3. *Section 9.3.* The treatment of the Lorentz–Dirac equation (9.42) as the equation of motion for a radiating electron with the four-momentum $p^\mu = mv^\mu$ goes back to Schott (1915) and Dirac (1938). It is the source of paradoxes. For an extended discussion of paradoxes of this kind and further references see Rohrlich (1965).

10

Generalizations

To grasp electrodynamics as a whole, the reader may wish to see its place in a wider context. With this in mind some general theories of this type are reviewed in this chapter.

10.1 Rigid Particle

Varying the world line of a bare particle in the action

$$S = -\int d\tau \left[m_0 \sqrt{\dot{z}\cdot\dot{z}} + e\,(\dot{z}\cdot A) \right] \,, \tag{10.1}$$

we obtain a *second-order* differential equation. However, we learned in Chap. 6 that the original degrees of freedom of the Maxwell–Lorentz theory are rearranged to yield *third-order* equations of motion, the Lorentz–Dirac equation. This suggests we explore Lagrangians which involve not only $\dot{z}^\mu$, but also $\ddot{z}^\mu$, and even greater derivatives. Because the additional geometric invariant one can construct using $\ddot{z}^\mu$ is the curvature of the world line, Lagrangians which involve $\ddot{z}^\mu$ are said to describe *rigid* particles. For simplicity, we restrict our discussion to the case of a *free* particle governed by

$$S = \int d\tau\, L(\dot{z}, \ddot{z}) \,. \tag{10.2}$$

Here, τ is an arbitrary evolution parameter.

It is worthwhile to note two caveats before commencing the discussion. First, there is no experimental evidence that higher derivative Lagrangians play any role in nature. Second, ever since the Hamiltonian formulation of these theories that was achieved by Mikhail Ostrogradskiĭ in 1850, it has been clear the Hamiltonians for higher derivative Lagrangians are almost always unbounded below. The known exceptions occur only in cases of very high symmetry, and only for the Lagrangians with a single higher derivative $\ddot{z}^\mu$.

One can enforce reparametrization invariance by writing the Lagrangian

$$L = \gamma^{-1} \varPhi(a^2) \tag{10.3}$$

in which γ is the generalized Lorentz factor[1],

$$\gamma^{-1} = \sqrt{\dot{z} \cdot \dot{z}}\ , \tag{10.4}$$

and $\varPhi$ an arbitrary function of the four-acceleration squared. The role of γ is similar to that of the einbein η in the action (2.261). Recall that the curvature squared is equal to the negative four-acceleration squared, and that, for an arbitrary τ, the four-acceleration is given by

$$a^\mu = \gamma \frac{d}{d\tau} \left(\gamma \frac{dz^\mu}{d\tau} \right) . \tag{10.5}$$

Following the line of reasoning of Sect. 2.5, the infinitesimal variation of the action (10.2) is easily evaluated (Problem 10.1.1), giving

$$\begin{aligned} \delta S = & \int_{\tau_1}^{\tau_2} d\tau \left[\frac{\partial L}{\partial z^\mu} - \frac{d}{d\tau} \left(\frac{\partial L}{\partial \dot{z}^\mu} \right) + \frac{d^2}{d\tau^2} \left(\frac{\partial L}{\partial \ddot{z}^\mu} \right) \right] \delta z^\mu \\ & + \left(H\, \varDelta\tau - p_\alpha\, \varDelta z^\alpha - \pi_\alpha\, \varDelta \dot{z}^\alpha \right) \Big|_{\tau_1}^{\tau_2} . \end{aligned} \tag{10.6}$$

Here, δz^μ is the local coordinate variation

$$\delta z^\mu = \varDelta z^\mu - \dot{z}^\mu\, \varDelta\tau\ , \tag{10.7}$$

p_μ the four-momentum conjugate to z^μ:

$$p_\mu = -\frac{\partial L}{\partial \dot{z}^\mu} + \frac{d}{d\tau} \left(\frac{\partial L}{\partial \ddot{z}^\mu} \right) , \tag{10.8}$$

π_μ the four-momentum conjugate to $\dot{z}^\mu$:

$$\pi_\mu = -\frac{\partial L}{\partial \ddot{z}^\mu}\ , \tag{10.9}$$

and H the Hamiltonian:

$$H = p \cdot \dot{z} + \pi \cdot \ddot{z} + L\ . \tag{10.10}$$

It immediately follows that the Euler–Lagrange equations

$$\frac{\partial L}{\partial z^\mu} - \frac{d}{d\tau} \left(\frac{\partial L}{\partial \dot{z}^\mu} \right) + \frac{d^2}{d\tau^2} \left(\frac{\partial L}{\partial \ddot{z}^\mu} \right) = 0 \tag{10.11}$$

[1] If τ is laboratory time t in a particular inertial frame, then γ takes the familiar form $\gamma = (1 - \mathbf{v}^2)^{-1/2}$.

are ordinary differential equations of *fourth* order with z^μ as the unknown function.

If the action is translation invariant,

$$\varDelta z^\mu = \epsilon^\mu \ , \tag{10.12}$$

then p^μ is conserved. Indeed, combining (10.8) and (10.11) for z-independent L, we have

$$\dot{p}^\mu = 0 \ . \tag{10.13}$$

However, the action cannot be invariant under velocity translations

$$\varDelta \dot{z}^\mu = \delta^\mu \ . \tag{10.14}$$

Such a symmetry would conflict with reparametrization invariance. Indeed, L has a factor of γ^{-1} to ensure reparametrization invariance, but the presence of γ^{-1} in the action eliminates (10.14) as a symmetry transformation. Thus, π_μ is not conserved.

Applying (10.5), (10.4), (10.8), and (10.9) to (10.10), one can show (Problem 10.1.2) that $H = 0$ if L is of the form (10.3). This is due to reparametrization invariance.

Let τ be the proper time s. The Euler–Lagrange equations (10.11) for a Lagrangian of the form (10.3) read

$$(\overset{v}{\perp}\, \dot{p})^\mu = 0 \ , \tag{10.15}$$

which is consistent with the basic law of relativistic mechanics (2.6) in the absence of external forces.

Let us examine the simplest case

$$\varPhi(a^2) = -\mu + \nu a^2 \ , \tag{10.16}$$

where μ and ν are real parameters. We assume that $\mu > 0$, for if we put $\nu = 0$, we would have to interpret μ as the Newtonian mass.

Substituting (10.16) into (10.8), we find

$$p^\mu = \mu v^\mu + \nu \left(2\dot{a}^\mu + 3a^2 v^\mu\right) . \tag{10.17}$$

Let this rigid particle be moving along a straight line, say, in the x^3-direction. Then

$$\begin{aligned} v^\mu &= (\cosh\alpha, 0, 0, \sinh\alpha) \ , \\ a^\mu &= \dot{\alpha}\,(\sinh\alpha, 0, 0, \cosh\alpha) \ , \end{aligned} \tag{10.18}$$

and hence $a^2 = -\dot{\alpha}^2$. Equation (10.15) becomes

$$\mu\dot{\alpha} + \nu \left(2\dddot{\alpha} - \dot{\alpha}^3\right) = 0 \ . \tag{10.19}$$

Denoting $\dot\alpha = q$ and $\mu/\nu = q_*^2$, we can write (10.19) in the form

$$\ddot q + \frac{1}{2}\, q_*^2\, q - \frac{1}{2}\, q^3 = 0\,. \tag{10.20}$$

The first integral of this equation is

$$\frac{1}{2}\,\dot q^2 + U(q) = E\,, \tag{10.21}$$

$$U(q) = -\frac{1}{8}\,(q^2 - q_*^2)^2\,. \tag{10.22}$$

Here, E is an arbitrary integration constant.

One may consider (10.20) and (10.21) as, respectively, the equation of motion and energy integral for a *fictitious* nonrelativistic particle of unit mass in the potential U. For $\nu > 0$, U is shown in the left hand side of Fig. 10.1. If $-q_*^4/8 < E < 0$, then the fictitious particle moves within the range $-q_* < q < q_*$. This means that imposing the restriction on the initial acceleration $|a^2| < \mu/\nu$, the rigid particle executes a zitterbewegung. If $E > 0$, or, alternatively, $E < -q_*^4/8$, then the fictitious particle can reach infinity. In other words, once the initial acceleration of the rigid particle has exceed $(\mu/\nu)^{1/2}$, this particle is doomed to a run away. If $E = 0$, then the fictitious particle rests at the top of the potential hill. That is, letting $a^2 = -\mu/\nu$, makes the rigid particle to execute a hyperbolic motion. However, this regime is unstable, a small perturbation converts it to a run away. If $E = -q_*^4/8$, then the fictitious particle rests on the bottom of the potential, which corresponds to uniform motion of the rigid particle, $a^2 = 0$. It is apparent that, for $\nu < 0$, or, alternatively, $\mu = 0$ (Fig. 10.1, right plot), the rigid particle always runs away. Uniform motion of the rigid particle with such μ and ν is unstable, a small perturbation will always convert it to the runaway regime.

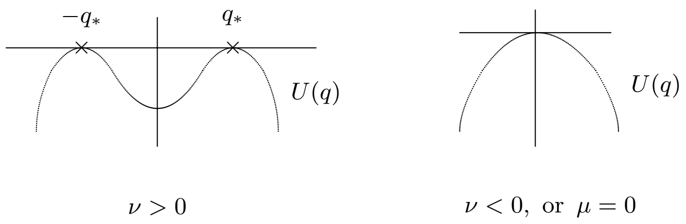


Fig. 10.1. The potential U for the fictitious particle

These results show that rigid particles are non-Galilean objects. In the absence of external forces, they are capable of not only moving uniformly along a straight line, but also of zitterbewegung, runaway, and hyperbolic evolution.

Combining (10.18) and (10.17),

$$p^\mu = (\mu - \nu\dot{\alpha}^2)(\cosh\alpha, 0, 0, \sinh\alpha) + 2\nu\ddot{\alpha}(\sinh\alpha, 0, 0, \cosh\alpha) \ . \tag{10.23}$$

Therefore,

$$p^2 = (\mu - \nu\dot{\alpha}^2)^2 - 4\nu^2\ddot{\alpha}^2 \ . \tag{10.24}$$

When compared with (10.21) and (10.22), this gives

$$p^2 = -8\nu^2 E \ . \tag{10.25}$$

We see that $p^2 < 0$ is equivalent to $E > 0$, which is a criterion for the fictitious particle to reach infinity. When a free rigid particle moving along a straight line is in a state of *spacelike* p^μ, it executes *runaway* motion. Conversely, if p^μ is *timelike*, then the only stable (against small perturbations), non-Galilean regime for it is *zitterbewegung*[2].

Rewriting (10.17) in the form

$$p^\mu = (\mu + \nu a^2)\, v^\mu + 2\nu\, (\overset{v}{\perp} \dot{a})^\mu \ , \tag{10.26}$$

$$(\overset{v}{\perp} \dot{a})^\mu = \dot{a}^\mu + a^2 v^\mu \ , \tag{10.27}$$

we get

$$M^2 = p^2 = (\mu + \nu a^2)^2 + 4\nu^2 (\overset{v}{\perp} \dot{a})^2 \ , \tag{10.28}$$

$$m = p \cdot v = \mu + \nu a^2 \ . \tag{10.29}$$

Both M and m depend on kinematical variables. Nevertheless, M is constant because p^μ is conserved. As for m, it varies with time when the rigid particle executes zitterbewegung or runaway motion. The only non-Galilean regime which leaves m unchanged, is hyperbolic motion with $a^2 = -\mu/\nu$. Note that this case is dynamically singular, $p^\mu = 0$. To see this, one should use the equation of hyperbolic motion (2.29),

$$(\overset{v}{\perp} \dot{a})^\mu = 0 \ , \tag{10.30}$$

and put $a^2 = -\mu/\nu$ in (10.26).

Of the two invariant quantities, M and m, only M is a constant of motion. Hence, it is M which characterizes the inertia of a rigid particle. This situation is in contrast to that of a dressed particle in the Maxwell–Lorentz theory.

Problem 10.1.1. Derive (10.6)–(10.10).

Problem 10.1.2. Show that $H = 0$ for any Lagrangian of the form (10.3).

[2] Note, however, that this regime is unstable against finite perturbations of curvature whose magnitude is greater than $(\mu/\nu)^{1/2}$.

10.2 Different Dimensions

In Chaps. 3–6 we studied Maxwell's equations in four-dimensional Minkowski spacetime. It is interesting to see how electrodynamics works in other dimensions, assuming that the Lagrangian

$$\mathcal{L} = -\frac{1}{4\Omega_{D-1}}\, F_{\mu\nu}F^{\mu\nu} - A_\mu j^\mu \tag{10.31}$$

is still valid. We adopt the metric $\eta_{\mu\nu} = \mathrm{diag}\,(1, -1, \ldots, -1)$. Greek letters take the values $0, 1, \ldots, D$. Repeated indices are summed over this range. The overall factor $(16\pi)^{-1}$ of the Larmor term is replaced by $(4\Omega_{D-1})^{-1}$, where

$$\Omega_{D-1} = \frac{2\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}\right)} \tag{10.32}$$

is the area of the unit $(D-1)$-sphere. For $D = 3$, we have $\Omega_{D-1} = 4\pi$, establishing correspondence with our previous conventions.

The field equation resulting from (10.31) reads

$$\partial_\lambda F^{\lambda\mu} = \Omega_{D-1} j^\mu \,. \tag{10.33}$$

Taking $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ and imposing the Lorenz gauge condition, we can write (10.33) in the form

$$\Box\, A^\mu = \Omega_{D-1}\, j^\mu \,. \tag{10.34}$$

Solutions to the wave equation (10.34) can be found with the aid of the Green's function technique. Let us show that the retarded Green's function, obeying

$$\Box\, G_{\mathrm{ret}}(x) = \delta^{D+1}(x) \,, \tag{10.35}$$

is given by

$$G_{\mathrm{ret}}(x) = \begin{cases} \frac{1}{2\pi^n}\, \theta(x_0)\, \delta^{(n-1)}(x^2) & D = 2n+1, \\ \frac{(-1)^{n-1}}{\pi^{n+\frac{1}{2}}}\, \Gamma\left(n - \frac{1}{2}\right) \theta(x_0)\, \theta(x^2) \left(x^2\right)^{\frac{1}{2}-n} & D = 2n, \end{cases} \tag{10.36}$$

where $\delta^{(n-1)}(x^2)$ is the Dirac delta-function differentiated $n-1$ times with respect to its argument.

We first consider odd-dimensional spaces $D = 2n+1$. Using the Fourier transform

$$G_{\mathrm{ret}}(t, \mathbf{x}) = \frac{1}{(2\pi)^D} \int d^D k\, e^{i\mathbf{k}\cdot\mathbf{x}}\, \widetilde{G}_{\mathrm{ret}}(t, \mathbf{k}) \tag{10.37}$$

(to maintain contact with the notation of Chap. 4, x_0 is called t), we deduce from (10.35) that

$$\widetilde{G}_{\mathrm{ret}}(t, \mathbf{k}) = \frac{\sin kt}{k} \,, \tag{10.38}$$

in a close analogy to what was done in Sect. 4.4. We choose D-dimensional spherical coordinates in k-space, and assume that the k_D-axis is aligned with $\mathbf{r}$, so that

$$d^D k = k^{2n} dk\, d\vartheta_1 \sin\vartheta_2 d\vartheta_2 \ldots \sin^{2n-1}\vartheta_{2n} d\vartheta_{2n} \;, \tag{10.39}$$

and $\mathbf{k}\cdot\mathbf{r} = kr\cos\vartheta_{2n}$. Inserting (10.38) into (10.37), and observing that the integrand does not depend on $\vartheta_1, \ldots, \vartheta_{2n-1}$,

$$G_{\mathrm{ret}}(t,\mathbf{x}) = \frac{\Omega_{2n-1}}{(2\pi)^{2n+1}} \int_0^\infty dk\, k^{2n-1}\, \sin kt \int_0^\pi d\vartheta\, \sin^{2n-1}\vartheta\, e^{ikr\cos\vartheta} \;, \tag{10.40}$$

where ϑ stands for ϑ_{2n}.

We now take advantage of two relations, well known in the theory of Bessel functions,

$$\int_0^\pi d\vartheta\, \sin^{2n-1}\vartheta\, e^{ikr\cos\vartheta} = 2^{n-\frac{1}{2}} \sqrt{\pi}\, \Gamma(n)\, (kr)^{\frac{1}{2}-n}\, J_{n-\frac{1}{2}}(kr) \;, \tag{10.41}$$

$$J_{n-\frac{1}{2}}(kr) = (-1)^{n-1} \sqrt{\frac{2}{\pi}} \left(\frac{r}{k}\right)^{n-\frac{1}{2}} \left(\frac{d}{r\,dr}\right)^{n-1} \frac{\sin kr}{r} \;. \tag{10.42}$$

Applying (10.41) and (10.42) to (10.40), we have

$$G_{\mathrm{ret}}(t,\mathbf{x}) = \frac{(-1)^{n-1}}{(2\pi)^{n-1}} \left(\frac{\partial}{r\,\partial r}\right)^{n-1} \left[\frac{1}{2\pi^2}\, \frac{1}{r} \int_0^\infty dk \sin kr \sin kt\right] \;. \tag{10.43}$$

Referring to Sects. 4.3 and 4.4, the expression in the square brackets is identified as the retarded Green's function for $D = 3$. Thus,

$$G_{\mathrm{ret}}(t,\mathbf{x}) = \frac{(-1)^{n-1}}{(2\pi)^{n-1}} \left(\frac{\partial}{r\,\partial r}\right)^{n-1} \frac{\delta(t-r)}{4\pi r} = \frac{\theta(t)}{2\pi^n} \left(\frac{\partial}{\partial t^2}\right)^{n-1} \delta(t^2 - r^2) \;, \tag{10.44}$$

which proves the first line of (10.36).

To obtain the retarded Green's function for $D = 2n$, it is sufficient to integrate (10.44) over some spatial coordinate, say, x_{D+1},

$$G^{(D)}(x) = \int_{-\infty}^{\infty} dx_{D+1}\, G^{(D+1)}(x, x_{D+1}) \;, \tag{10.45}$$

which becomes clear from the preceding discussion if we note that the Fourier transforms of these Green's functions are related by

$$\widetilde{G}^{(D)}(t,\mathbf{k}) = \widetilde{G}^{(D+1)}(t,\mathbf{k},k_{D+1})\big|_{k_{D+1}=0} \;. \tag{10.46}$$

Working out (10.45), we obtain the second line of (10.36). An alternative expression for the retarded Green's function in even space dimensions $D = 2n$ is

$$G_{\mathrm{ret}}(t,\mathbf{x}) = \frac{(-1)^{n-1}}{(2\pi)^{n-1}}\left(\frac{\partial}{r\,\partial r}\right)^{n-1}\frac{\theta(t-r)}{2\pi\sqrt{t^2-r^2}}\ . \tag{10.47}$$

A sharp distinction between wave propagation in spaces of even and odd dimensions can be understood from *Huygens' principle*. Huygens' principle means that any retarded signal carries information on the state of a point source at the instant of its emission. This idea is exemplified in the first line of (10.36): the retarded Green's function for $D = 2n + 1$ is concentrated on the forward sheet of the light cone $x^2 = 0$, $x_0 > 0$. By contrast, if $D = 2n$, then the retarded signal measured at some point x^μ derives from the entire history of the source which lies on or within the past light-cone of x^μ. If we think of the retarded signal as traveling with the speed of light then it ought to be emitted at the instant the source intersects the past light-cone of x^μ. So Huygens' principle fails in odd spacetime dimensions. The second line of (10.36) shows that the support of the retarded Green's function for $D = 2n$ is the interior of the future light cone $x^2 \geq 0$, rather than its surface.

Four-dimensional real spacetime has little in common with odd-dimensional worlds. To illustrate this, we refer to Problems 10.2.11 and 10.2.12. In the rest of this section, our main concern is with $D + 1 = 2n$. We explore electrodynamics in even-dimensional worlds with two illuminating examples: $D + 1 = 2$ and $D + 1 = 6$.

10.2.1 Two Dimensions

With reference to Problem 4.4.5, the retarded Green's function for the two-dimensional wave operator is given by $G_{\mathrm{ret}}(t,x) = \frac{1}{2}\,\theta(t - |x|)$. Although Huygens' principle fails, the procedure developed in Sect. 4.7 is still valid. Using the result of Problem 4.7.2, we can write the retarded vector potential generated by a single charge

$$A_\mu = -eR_\mu\ . \tag{10.48}$$

The field strength is

$$F_{\mu\nu} = e\,(c_\mu v_\nu - c_\nu v_\mu)\ . \tag{10.49}$$

We see that $F_{\mu\nu}$ is nonsingular, even if discontinuous, on the world line. Furthermore, $F_{\mu\nu}$ is independent of acceleration. In two-dimensional electrodynamics, fields due to charges moving along *arbitrary* timelike smooth world lines are equivalent to *static* fields.

The stress-energy tensor for the retarded field (10.49) is

$$\Theta^{\mu\nu} = \frac{1}{4}\,e^2\,(c^\mu v^\nu + c^\nu v^\mu - c^\mu c^\nu) = \frac{1}{4}\,e^2\,(v^\mu v^\nu - u^\mu u^\nu) = \frac{1}{4}\,e^2\,\eta^{\mu\nu}\ . \tag{10.50}$$

Here, we have used the completeness relation $v^\mu v^\nu - u^\mu u^\nu = \eta^{\mu\nu}$ stemming from the fact that v^μ and u^μ form a basis. It is then clear that $\partial_\mu\Theta^{\mu\nu} = 0$. We may therefore define

$$P^\mu = \int_{\mathcal{B}} d\sigma_\nu \, \Theta^{\mu\nu} \,, \tag{10.51}$$

where the integration is over a region $\mathcal{B}$ of an arbitrary surface (in fact, a curve). Taking $\mathcal{B}$ to be a large interval L of a straight line perpendicular to v^μ,

$$P^\mu = \frac{1}{4}\, e^2 \, v^\mu \, L \,, \tag{10.52}$$

which diverges as $L \to \infty$. In contrast to the four-dimensional case, where P^μ diverges due to singularities on world lines, the divergence of (10.52) originates from infinite limits of integration. In field-theoretic language, the quantity (10.52) is infrared divergent[3].

The stress-energy tensor (10.50) contains the term $-\frac{1}{4}e^2c^\mu c^\nu$. One may inquire whether it is possible to interpret it as radiation. Although this term meets three conditions (6.87), (6.88), and (6.90), the fourth condition (6.91) is violated. Indeed, $-\frac{1}{4}e^2c^\mu c^\nu$ is similar in its spatial behavior to the rest of the stress-energy tensor (10.50), contrary to the requirement that the radiation be asymptotically separated from the bound part.

There is *no radiation* in two-dimensions. The rearrangement of degrees of freedom is limited by forming a dressed particle with energy-momentum

$$p^\mu = mv^\mu \,, \tag{10.53}$$

where m is the renormalized mass. Equation (10.53) makes it clear that the dressed particle is a Galilean object.

Let two charged particles be moving under each other's influence along a line which represents space in this two-dimensional world. The equation of motion for the particle 1 reads

$$m_1 a_1^\mu = f_1^\mu = e_1 e_2 \, v_\lambda^1 \left(v_2^\lambda c_2^\mu - c_2^\lambda v_2^\mu \right) , \tag{10.54}$$

and that for particle 2 is obtained from (10.54) by the label interchange $1 \leftrightarrow 2$. We thus have a system of lag-time differential equations. A delay in the argument of the right-hand side of (10.54) is due to the fact that signals propagate at the speed of light, causing the force f_1^μ to depend on position and velocity of particle 2 at a previous time (Fig. 10.2). Surprising as it may seem, these equations are integrable.

Looking at Fig. 10.2, one can write

$$v_1^\mu = (\cosh\alpha, \sinh\alpha),\;\; u_1^\mu = (-\sinh\alpha, -\cosh\alpha),\;\; v_2^\mu = (1,0),\;\; u_2^\mu = (0,1) \,, \tag{10.55}$$

[3] We have encountered divergences before which were absorbed into the particle mass. However, those were ultraviolet divergences, associated with the behavior of the fields very near the particle. Infrared divergences typically mean that the quantity being computed is somehow unphysical. In this case the problem is that we have assumed our two dimensional spacetime contains only a single charge. Had there been more than one charge, so that the total charge vanished, then there would have been no divergence.

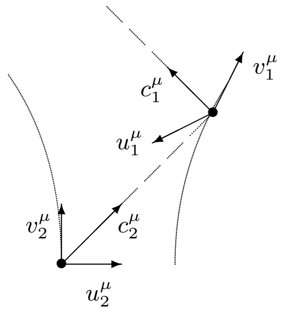


Fig. 10.2. Two-particle problem

where α is some unknown function of the proper time. We introduce two invariants

$$\Gamma = v_1 \cdot v_2, \quad \Delta = v_1 \cdot u_2 \ , \tag{10.56}$$

and use (10.55), to give $\Gamma = \cosh\alpha$, $\Delta = -\sinh\alpha$, and

$$\Gamma^2 - \Delta^2 = 1, \quad u_1 \cdot u_2 = \Gamma, \quad u_1 \cdot v_2 = \Delta \ . \tag{10.57}$$

By (10.56),

$$f_1^\mu = e_1 e_2 \left[\Gamma\, c_2^\mu - (\Gamma + \Delta)\, v_2^\mu \right] . \tag{10.58}$$

The Lorentz force f_1^μ is perpendicular to v_1^μ, and hence parallel to u_1^μ. In view of (10.57), $f_1 \cdot u_1 = e_1 e_2$. Therefore, $f_1^\mu = e_1 e_2 \,(\sinh\alpha, \cosh\alpha)$, and (10.54) becomes $m_1\dot{\alpha} = e_1 e_2$, whence $\alpha = (e_1 e_2/m_1)s_1 + \alpha_0$. We assume that $\alpha_0 = 0$ for simplicity. Integrating $v_1^\mu = (\cosh\alpha, \sinh\alpha)$, we conclude that any solution to (10.54) can be built out of fragments of the hyperbolic curve

$$z_1^\mu(s_1) = z_1^\mu(0) + a_1^{-1} \left(\sinh a_1 s_1, \cosh a_1 s_1 \right) \ , \tag{10.59}$$

where $z_1^\mu(0)$ is the particle position at the initial instant $s_1 = 0$, and $a_1 = e_1 e_2/m_1$. Particle 2 moves along a similar world line. This completely solves the two-particle problem. The N-particle problem is handled in a like manner (Problem 10.2.3).

Figure 10.3 depicts the world lines of two particles of equal charge magnitude and mass. Oppositely charged particles execute compact periodic motions, plot A. Particles of like charges execute infinite motions, as in plot B.

10.2.2 Six Dimensions

In this subsection, we restrict our discussion to a single charged particle. Looking at the first line of (10.36), we observe that the retarded Green's function for $n = 2$ involves a derivative of $\delta(x^2)$, which implies that the

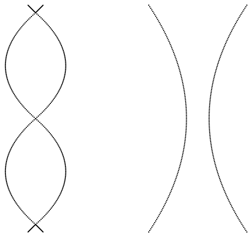


Fig. 10.3. World lines of two particles

retarded vector potential carries information on both velocity and acceleration of the source at the point of emission. Accordingly, the ansatz (4.302) alters

$$A^\mu = \Omega(\rho,\lambda)\, a^\mu + \Phi(\rho,\lambda)\, v^\mu + \Psi(\rho,\lambda)\, R^\mu \ , \tag{10.60}$$

where Ω, Φ, Ψ are unknown functions. Referring to Problem 4.7.3, we cite the solution:

$$\Omega = \frac{k}{\rho^2}, \quad \Phi = k\left(-\frac{\lambda}{\rho^3} + C_1\right), \quad \Psi = k\,\frac{\lambda C_1 + C_2}{\rho} \ , \tag{10.61}$$

where C_1 and C_2 are arbitrary integration constants, and k an overall multiplicative. Dropping pure gauge terms $C_1\,\partial^\mu\rho$ and $C_2\,\partial^\mu s$,

$$A^\mu = k\left(-\lambda\,\frac{v^\mu}{\rho^3} + \frac{a^\mu}{\rho^2}\right) \ , \tag{10.62}$$

whence

$$F^{\mu\nu} = k\left(\frac{a^\mu v^\nu - a^\nu v^\mu}{\rho^3} + c^\mu V^\nu - c^\nu V^\mu\right) \ , \tag{10.63}$$

$$V^\mu = \frac{\dot{a}^\mu}{\rho^2} - 3\lambda\,\frac{a^\mu}{\rho^3} + \frac{v^\mu}{\rho^4}\,\left[3\lambda^2 - \rho^2\,(\dot{a}\cdot c)\right] \ . \tag{10.64}$$

We apply the Gauss law to the static field $F = 3k\,(c\wedge v)\,\rho^{-4}$, to give $k = \frac{1}{3}\,e$. Thus,

$$F = \frac{e}{3}\left(\frac{a\wedge v}{\rho^3} + c\wedge V\right) \ . \tag{10.65}$$

We now turn to the problem of ultraviolet divergences. We wish to integrate $\Theta_{\mu\nu}$ over a five-dimensional surface to obtain the electromagnetic six-momentum P_μ. However, direct integration is impossible because $F_{\mu\nu}$ is singular on the world line. By (10.65),

$$F_{\mu\alpha}F^{\alpha}_{\ \ \nu}=\frac{e^2}{9}\left[-\frac{a_\mu a_\nu+a^2\,v_\mu v_\nu}{\rho^6}+(c\cdot V)\,(c_\mu V_\nu+c_\nu V_\mu)-c_\mu c_\nu\,V^2\right.$$
$$+\frac{a_\mu V_\nu+a_\nu V_\mu}{\rho^3}+(a\cdot V)\,\frac{v_\mu c_\nu+v_\nu c_\mu}{\rho^3}-(v\cdot V)\,\frac{a_\mu c_\nu+a_\nu c_\mu}{\rho^3}$$
$$\left.-\frac{\lambda+1}{\rho^4}\,(v_\nu V_\mu+v_\mu V_\nu)\right], \tag{10.66}$$

so that

$$F_{\alpha\beta}F^{\alpha\beta}=\frac{2e^2}{9}\left[\frac{a^2}{\rho^6}-(c\cdot V)^2-\frac{2}{\rho^3}\,(a\cdot V)+\frac{2}{\rho^4}\,(\lambda+1)\,(v\cdot V)\right]. \tag{10.67}$$

Since the element of measure on a five-dimensional spacelike hyperplane is proportional to $\rho^4 d\rho$, the integration of $\Theta_{\mu\nu}$ results in *cubic* and *linear* divergences. It is clear from (10.66) and (10.67) that the cubic divergence occurs even in the static case. By contrast, the linear divergence, which owes its origin to the terms scaling as ρ^{-2}, appears only for curved world lines, that is, in the case that either a^μ or $\dot a^\mu$, or both are nonzero. This implies that the Poincaré–Planck action for a bare particle (2.207) must be endowed with a term containing higher derivatives to absorb the linear divergence.

The simplest reparametrization invariant Lagrangian for a rigid bare particle is that discussed in Sect. 10.1:

$$L=-\sqrt{v\cdot v}\,\left(\mu_0-\nu_0 a^2\right)\,. \tag{10.68}$$

The corresponding six-momentum is given by (10.17),

$$p_0^\mu=\mu_0 v^\mu+\nu_0(2\dot a^\mu+3a^2 v^\mu)\,. \tag{10.69}$$

It follows on dimensional grounds (Problem 10.2.4) that the linearly divergent term arising from the integration of $\Theta_{\mu\nu}$ involves v^μ and $\dot a^\mu$ in exactly the same way as the second term of (10.69) does. Hence, the cubic and linear divergences are eliminated through the respective renormalization of μ_0 and ν_0. This suggests that ultraviole t divergences can be tamed if the particle action is curvature dependent.

A direct procedure for finding the equation of motion for a dressed charged particle would be to regularize $F_{\mu\nu}$, evaluate divergent integrals and segregate from them finite terms. For $D+1=6$, this can be accomplished similarly to what was done for $D+1=4$ in Sect. 6.5 (see Problem 10.2.5). However, we take another route. We first determine the radiation rate, and then make use of the fact that the equation of motion for a dressed particle should involve the projector $\overset{v}{\perp}$. Recall that $\overset{v}{\perp}$ results from reparametrization invariance in the Lagrangian formalism. The virtue of this procedure is that it is independent of a particular regularization prescription because the radiation rate is represented by a convergent integral.

Let us calculate the radiation rate. The radiation flux through a four-dimensional sphere enclosing the source is constant for any radius of the sphere. Therefore, the terms of $\Theta_{\mu\nu}$ responsible for this flux should scale as ρ^{-4}. It is seen from (10.66) and (10.67) that only $-c_\mu c_\nu V^2$ contains such terms. We segregate in (10.64) the term scaling as ρ^{-2},

$$b^\mu = \frac{e}{3}\left[\frac{\dot{a}^\mu}{\rho^2} - 3\,\frac{\lambda+1}{\rho^3}\,a^\mu + 3\left(\frac{\lambda+1}{\rho^2}\right)^2 v^\mu - \frac{(\dot{a}\cdot c)\,v^\mu}{\rho^2}\right]. \tag{10.70}$$

One can verify (Problem 10.2.6) that the tensor $b^2c^\mu c^\nu$ meets conditions (6.87), (6.88), (6.90), and (6.91). It is therefore natural to identify this tensor with the radiation: $b^2c^\mu c^\nu = \Theta^{\mu\nu}_{\rm II}$.

The radiated six-momentum is defined by

$$\mathcal{P}^\mu = \int_\Sigma d\sigma_\nu\,\Theta^{\mu\nu}_{\rm II}\,. \tag{10.71}$$

Since $\partial_\nu b^2c^\nu c^\mu = 0$, the surface of integration Σ in (10.71) may be chosen arbitrarily. It is convenient to deform Σ to a tubular surface T_ϵ of small radius ϵ enclosing the world line, much as was done for the radiated four-momentum in Sect. 6.3. The surface element on this five-dimensional tube is

$$d\sigma^\mu = \partial^\mu\rho\,\rho^4\,d\Omega_4\,ds = (v^\mu + \lambda c^\mu)\,\epsilon^4\,d\Omega_4\,ds\,. \tag{10.72}$$

The integral

$$\mathcal{P}^\mu = \int_{T_\epsilon} d\sigma_\nu\,b^2c^\nu c^\mu \tag{10.73}$$

becomes

$$-\frac{e^2}{9}\int d\tau\int d\Omega_4\Big\{\Big[(\overset{v}{\perp}\dot{a})^2 + 9(a\cdot u)^2a^2 + 9(a\cdot u)^4 + (\dot{a}\cdot u)^2\Big]\,v^\mu$$
$$-\left[3\,\frac{da^2}{ds}\,(a\cdot u) + 6(a\cdot u)^2(\dot{a}\cdot u)\right]u^\mu\Big\}\,. \tag{10.74}$$

To make the solid angle integration, we use the following formulas

$$\int d\Omega_4\,u_\mu u_\nu = -\frac{\Omega_4}{5}\,\overset{v}{\perp}_{\mu\nu}\,, \tag{10.75}$$

$$\int d\Omega_4\,u_\alpha u_\beta u_\mu u_\nu = \frac{\Omega_4}{5\cdot 7}\left(\overset{v}{\perp}_{\mu\nu}\overset{v}{\perp}_{\alpha\beta} + \overset{v}{\perp}_{\alpha\mu}\overset{v}{\perp}_{\beta\nu} + \overset{v}{\perp}_{\alpha\nu}\overset{v}{\perp}_{\beta\mu}\right), \tag{10.76}$$

which are readily derived. As a result we find

$$\mathcal{P}^\mu = \frac{e^2}{9}\int_{-\infty}^{s} d\tau \left\{ -\frac{4}{5}\left[\dot a^2 - \frac{16}{7}\,(a^2)^2\right] v^\mu - \frac{3}{7}\,\frac{da^2}{ds}\,a^\mu + \frac{6}{5\cdot 7}\,a^2(\overset{v}{\perp}\dot a)^\mu \right\} . \tag{10.77}$$

Therefore, the radiation rate is given by

$$\dot{\mathcal{P}}^\mu = \frac{e^2}{9}\left\{ -\frac{4}{5}\left[\dot a^2 - \frac{16}{7}\,(a^2)^2\right] v^\mu - \frac{3}{7}\,\frac{da^2}{ds}\,a^\mu + \frac{6}{5\cdot 7}\,a^2(\overset{v}{\perp}\dot a)^\mu \right\} . \tag{10.78}$$

It is easy to verify that

$$v\cdot\dot{\mathcal{P}} = -\frac{4e^2}{45}\left[\dot a^2 - \frac{16}{7}\,(a^2)^2\right] > 0\,. \tag{10.79}$$

This inequality shows that $\mathcal{P}^0$ represents positive field energy flowing outward from the source.

The bound six-momentum contains both divergent and finite terms. Similar to $\mathcal{P}^\mu$, the finite term $p^\mu_{\rm f}$ is free of dimensional parameters other than the overall factor e^2. The appropriate dimension is obtained from kinematical variables:

$$p^\mu_{\rm f} = c_1\,\ddot a^\mu + c_2\,a^2 a^\mu + c_3\,\frac{da^2}{ds}\,v^\mu\;, \tag{10.80}$$

where c_1, c_2, and c_3 are numerical coefficients. Let us suppose that the equation of motion for a dressed particle involves the projector $\overset{v}{\perp}$. Then

$$v_\mu\left(\dot p^\mu_{\rm f} + \dot{\mathcal{P}}^\mu\right) = 0\,. \tag{10.81}$$

With the identities

$$(a\cdot v) = 0,\quad (\dot a\cdot v) = -a^2,\quad (\ddot a\cdot v) = -\frac{3}{2}\,\frac{da^2}{ds},\quad (\dddot a\cdot v) = -2\,\frac{d^2a^2}{ds^2} + \dot a^2\;, \tag{10.82}$$

this gives

$$p^\mu_{\rm f} = \frac{4}{45}\,e^2\left[\ddot a^\mu + \frac{16}{7}\,a^2\,a^\mu + 2\,\frac{da^2}{ds}\,v^\mu\right] . \tag{10.83}$$

To find the divergent term, it is most convenient to perform the integration of $\Theta_{\mu\nu}$ over the surface of the future light cone. Recall, however, that the kinematic structure of the divergent six-momentum is similar to that of the bare particle six-momentum (10.69). Therefore, it is sufficient to renormalize μ_0 and ν_0 to get the combined quantity

$$\mathfrak{p}^\mu = \mu v^\mu + \nu\,(2\dot a^\mu + 3a^2\,v^\mu) + \frac{4}{45}\,e^2\left[\ddot a^\mu + \frac{16}{7}\,a^2 a^\mu + 2v^\mu\,\frac{da^2}{ds}\right] . \tag{10.84}$$

The six-momentum $\wp^\mu$ extracted from an external field $F_0^{\mu\nu}$ is found by integrating the mixed term of the stress-energy tensor $\Theta^{\mu\nu}_{\rm mix}$ over a tube T_ϵ of small radius ϵ enclosing the world line. This procedure resembles that in the four-dimensional case. So, referring to Sect. 6.3, we give the net result

$$\wp^{\mu} = \int_{T_\epsilon} d\sigma_\nu\, \Theta^{\mu\nu}_{\rm mix} = -\int_{-\infty}^{s} d\tau\, eF_0^{\mu\nu} v_\nu\ . \tag{10.85}$$

As before, $\dot{\wp}^{\mu}$ is the negative external Lorentz force, $\dot{\wp}^{\mu} = -f^{\mu} = -eF_0^{\mu\nu} v_\nu$.

By analogy with $D+1=4$, we write the local energy-momentum balance as

$$\dot{\mathfrak{p}}^{\mu} + \dot{\mathcal{P}}^{\mu} + \dot{\wp}^{\mu} = 0\ . \tag{10.86}$$

Substituting (10.84), (10.78), and (10.85) in (10.86), we arrive at the equation of motion for a dressed charged particle

$$\overset{v}{\perp}(\dot{p} - f) = 0\ , \tag{10.87}$$

where

$$p^{\mu} = \mu v^{\mu} + \nu\,(2\dot{a}^{\mu} + 3a^2 v^{\mu}) + \frac{1}{9}\,e^2\left[\frac{4}{5}\,\ddot{a}^{\mu} + 2a^2 a^{\mu} + \frac{da^2}{ds}\,v^{\mu}\right] \tag{10.88}$$

is the six-momentum of the dressed particle. Writing out this equation explicitly is left to the reader as an exercise.

It is clear from (10.88) that neither $M^2 = p^2$ nor $m = p\cdot v$ is a constant of motion. None of the scalars built from p^{μ} and derivatives of v^{μ} is conserved. Thus, a dressed particle living in a six-dimensional world has nothing which might be called its 'mass'.

Another surprising result is the occurrence of two different momenta $\mathfrak{p}^{\mu}$ and p^{μ}. Each is treated as the energy-momentum of the dressed particle in a particular context. If we take the balance equation (10.86), then the dressed particle is specified by $\mathfrak{p}^{\mu}$ shown in (10.84). On the other hand, if we invoke Newton's second law (10.87), then the dressed particle should be assigned p^{μ} given by (10.88).

Problem 10.2.1. Justify (10.59) in a heuristic way.

Hint This two-particle problem can be translated into the problem of motion of two parallel plates of a planar immense capacitor. There is only an electric field $\mathbf{E}$ between the plates, which is constant for any separation and velocity of the plates.

Problem 10.2.2. Compare the equation of motion for a dressed particle in a genuine two-dimensional realm, (10.54), and that for the case that the particle is moving along a straight line in the ambient space, (6.180). Explain the discrepancy between these results.

Problem 10.2.3. *Three-particle problem.* Classify all possible types of evolution in terms of finite and infinite motions for various charges and different relative positions of three particles on a straight line. Consider an instantaneously comoving frame for particle 1, and show that this particle is affected by the force

$$\mathbf{f}_1 = e_1 \sum_{J=2}^{3} e_J \,\mathrm{sgn}\,(x_1 - x_J) \,. \tag{10.89}$$

Can this particle be free despite the fact that all particles are charged? Verify that every particle moves along a piecewise smooth world line with constant curvature.

Problem 10.2.4. What are the dimensions of the divergent quantities e^2/ϵ^3 and e^2/ϵ appearing in the six-momentum P_μ. Reasoning from dimensional considerations, conclude that divergent terms of P_μ are

$$C_1 \frac{e^2}{\epsilon^3} v_\mu, \quad \text{and} \quad C_2 \frac{e^2}{\epsilon} \left(\dot{a}_\mu + k\, a^2 v_\mu\right) , \tag{10.90}$$

where C_1, C_2, and k are pure numbers. Assuming that the equation of motion for a dressed particle is orthogonal to v^μ, verify that $k = \frac{3}{2}$. Show that $C_1 e^2/\epsilon^3$ and $C_2 e^2/\epsilon$ can be absorbed into a redefinition of μ_0 and ν_0.

Problem 10.2.5. Using the Dirac regularization procedure discussed in Sect. 6.5, find $F^{\mu\nu}_{(-)}(z)$ and $F^{\mu\nu}_{(+)}(z)$ with the help of the relation $D_{\mathrm{ret}}(R) = \frac{1}{2\pi^2}\,\theta(R_0)\,\delta'(R^2)$, and derive the equation of motion for a dressed charged particle.

Answer

$$\mu a^\mu + \nu \left[2\ddot{a}^\mu + 3a^2 a^\mu + 3\,\frac{da^2}{ds}\, v^\mu \right] + \Gamma^\mu = f^\mu \,, \tag{10.91}$$

$$\Gamma^\mu = \frac{e^2}{9} \left\{ \frac{4}{5}\,\dddot{a}^{\,\mu} + 2a^2 \dot{a}^\mu + 3\,\frac{da^2}{ds}\, a^\mu + \left[\frac{8}{5}\,\frac{d^2 a^2}{ds^2} + 2(a^2)^2 - \frac{4}{5}\,\dot{a}^2 \right] v^\mu \right\} . \tag{10.92}$$

Problem 10.2.6. Show that $b^2 c_\mu c_\nu$ obeys (6.87)–(6.91), where b^μ is defined in (10.70).

Problem 10.2.7. Show that $\mathcal{P}^\mu$ is given by (10.74).

Problem 10.2.8. Prove (10.75) and (10.76).

Problem 10.2.9. Prove (10.77).

Problem 10.2.10. Prove (10.79).

Hint Use the inequality $(\overset{v}{\perp}\,\dot{a})^2 = \dot{a}^2 - (a^2)^2 < 0$.

Problem 10.2.11. *Electrodynamics in $D+1=3$ dimensions.* The Larmor action can be augmented by the addition of the *Chern–Simons term*:

$$S = -\frac{1}{8\pi e} \int d^3x \left(F_{\alpha\beta}F^{\alpha\beta} - \mu\, \epsilon^{\alpha\beta\gamma} A_\alpha F_{\beta\gamma} \right) \,. \tag{10.93}$$

Here, e and μ are parameters with respective dimensions $-\frac{1}{2}$ and -1. Check that this action is gauge invariant but not invariant under spacetime reflections.

This system is governed by the Euler–Lagrange equations

$$\partial_\alpha F^{\alpha\beta} + \mu\, {}^*F^\beta = 0\ , \tag{10.94}$$

where ${}^*F^\alpha = \frac{1}{2}\,\epsilon^{\alpha\beta\gamma} F_{\beta\gamma}$, together with the Bianchi identity

$$\partial_\beta {}^*F^\beta = 0\ . \tag{10.95}$$

Show that μ may be interpreted as the mass of the field A^α.

Hint Put (10.94) in the form

$$\Lambda^{\alpha\beta}(\mu)\, {}^*F_\beta = (\mu\,\eta^{\alpha\beta} + \epsilon^{\alpha\beta\gamma}\partial_\gamma)\, {}^*F_\beta = 0\ , \tag{10.96}$$

then act upon it with $\Lambda(-\mu)$, and use (10.95) to give

$$(\Box + \mu^2)\, {}^*F_\alpha = 0\ . \tag{10.97}$$

Hence A^α is massive, even though it is a gauge field. This is in sharp contrast to the Proca field which is not gauge invariant (see Problem 5.5.9).

Problem 10.2.12. *Yang–Mills theory in $D + 1 = 3$ dimensions.* The Yang–Mills action can be augmented by the addition of the non-Abelian Chern–Simons term:

$$S = -\frac{1}{8\pi g}\,\mathrm{tr}\int d^3x \left[F_{\alpha\beta}F^{\alpha\beta} - \mu\,\epsilon^{\alpha\beta\gamma} A_\alpha\left(F_{\beta\gamma} - \frac{2}{3}\,A_\beta A_\gamma\right)\right]\ . \tag{10.98}$$

Check gauge invariance. Show that μ is the mass of the field A^α.

10.3 Is the Dimension $D = 3$ Indeed Distinguished?

In 1918 Weyl pointed out that Maxwell's equations are conformally invariant only for spatial dimension $D = 3$. This property is shared by the Yang–Mills equations in pure Yang–Mills theory. This may be regarded as a hint that the $D = 3$ is singled out.

Another way of looking at this issue was proposed by Paul Ehrenfest in 1917. He raised the question: in what way does it become manifest in the fundamental laws of physics that space has three dimensions? In searching for an answer, he examined imaginary $(D+1)$-dimensional worlds assuming that the laws of mechanics and electrodynamics are encoded into the conventional action

$$S = -\sum_{I=1}^{N}\int d\tau_I\left[m_0^I\sqrt{\dot z_I\cdot\dot z_I} + e_I\,\dot z_I^\mu A_\mu(z_I)\right] - \frac{1}{4\Omega_{D-1}}\int d^{D+1}x\, F_{\mu\nu}F^{\mu\nu}\ . \tag{10.99}$$

Ehrenfest claimed that the value $D = 3$ establishes a line of demarcation between worlds where stable bound systems (such as a hydrogen atom) cannot exist from those where such systems are possible. His argument can be summarized as follows.

Consider a system of two particles with charges Ze and $-e$. Evidently the force law for these particles will vary with D. Ehrenfest supposed that the two-particle problem can be reduced to a one-particle Kepler problem. It was shown in Sect. 2.8 that the qualitative analysis of the Kepler problem is greatly facilitated if we introduce the effective potential

$$\mathcal{U}(r) = \sqrt{m^2 + \frac{\mathbf{L}^2}{r^2}} - U(r) \, . \tag{10.100}$$

Here, $U(r)$ is the potential energy between these particles, $U(r) = e\phi(\mathbf{x})$, and $\phi(\mathbf{x})$ obeys the D-dimensional Poisson equation

$$\nabla^2\phi(\mathbf{x}) = -\Omega_{D-1}\, Ze\, \delta^D(\mathbf{x}) \, . \tag{10.101}$$

With reference to Sect. 4.1, specifically to Problem 4.1.1, we write

$$\phi(\mathbf{x}) = -Ze \begin{cases} \mathrm{sgn}(2-D)|\mathbf{x}|^{2-D} & D \neq 2 \, , \\ \log|\mathbf{x}| & D = 2 \, . \end{cases} \tag{10.102}$$

For $D > 3$, the potential energy $e\phi(\mathbf{x})$ is more singular than the centrifugal term $|\mathbf{L}|/r$, and falling to the center (or, alternatively, going to infinity) is unavoidable. By contrast, for $D = 3$ and $Ze^2 \leq |\mathbf{L}|$, the potential energy is less singular than the centrifugal term, which prevents falling to the center, so that stable orbits are possible.

Of course, this inference is made using an oversimplified model with the instantaneous electrostatic interaction (10.102). If radiation of the planetary electron were taken into account, then this electron would fall to the nucleus even in the case $D = 3$. Ehrenfest invoked the Bohr quantization scheme in the hope that hydrogen atom is stable in real space due to its quantum nature. The general idea is then clear: since the electromagnetic attraction increases with D while the centrifugal effect is unchanged (on both classical and quantum levels), there is a critical dimension D_c above which planetary systems collapse. Ehrenfest believed that $D_c = 3$.

The pitfall in this reasoning is that the action (10.99) is inconsistent for $D = 4, 5, \ldots$ because divergences of the self-energy proliferate with D.

If we require that the electromagnetic sector of the action (10.99) is preserved for every D, and that all divergences of the self-energy are absorbed by redefining parameters which occur in the action, then the particle sector must be supplemented by terms with higher derivatives. For example, for the case $D = 5$ (discussed in Subsect. 10.2.2), acceleration-dependent terms are required. However, the two-particle problem in this rigid dynamics is no longer Keplerian. Further, the introduction of higher derivatives typically renders mechanical systems unstable when coupled to a continuum force field. So we see that the problem with higher dimensions is not the stability of atoms – as Ehrenfest imagined – but rather the stability of the universe. It it worth noting that this conclusion – that charged particles in higher dimensions require the inclusion of unacceptable higher derivative terms in the action – is confirmed by local quantum field theory, although the analysis would take us far beyond the scope of this classical treatment.

10.4 Nonlinear Electrodynamics

A major assumption underlying Maxwell's electrodynamics is the *superposition principle*. One may wonder whether the equations of motion for the electromagnetic field are linear in the strict sense or if this is merely a good approximation for the phenomena we have so far been able to observe.

We now extend the previous framework for description of the electromagnetic field by discarding linearity but retaining locality and the absence of magnetic monopoles. We thus proceed from the Lagrangian $\mathcal{L}$ which depends on the field strength expressed in terms of a single-valued smooth vector potential,

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \ . \tag{10.103}$$

Leaving aside higher-derivative theories, we limit our discussion to Lagrangians which are functions of the invariants $\mathcal{S}$ and $\mathcal{P}$.

The best known nonlinear theory of this type was proposed by Max Born and Leopold Infeld in 1934. It is characterized by the Lagrangian

$$\mathcal{L}_{\text{BI}} = b^2 \left(1 - \sqrt{1 + \frac{1}{b^2}\,\mathcal{S} - \frac{1}{4b^4}\,\mathcal{P}^2} \right) \ , \tag{10.104}$$

which is an elaboration of the original Born theory

$$\mathcal{L}_{\text{B}} = b^2 \left(1 - \sqrt{1 + b^{-2}\,\mathcal{S}} \right) \ . \tag{10.105}$$

Here, Heaviside units are adopted; b is a constant having dimension -2.

For weak fields, $\mathcal{L}_{\text{B}}$ approaches $-\frac{1}{2}\,\mathcal{S}$ which is the Larmor Lagrangian of Maxwell's electrodynamics[4]. The Born Lagrangian (10.105) was motivated by the so-called 'principle of finiteness' by which a satisfactory theory should avoid divergences in physical quantities. Born came to $\mathcal{L}_{\text{B}}$ following the passage from the Lagrangian of a free Newtonian particle to the corresponding relativistic expression,

$$L = \frac{1}{2}\,m\mathbf{v}^2 \to m\left(1 - \sqrt{1 - \mathbf{v}^2}\right) \ . \tag{10.106}$$

[4] Note that the factor of $1/4\pi$ is not present in Heaviside units.

Here, 1 is the upper bound for particle velocities. No particle can be accelerated past the barrier $\mathbf{v}^2 = 1$, otherwise L would be imaginary. Similarly, b in (10.105) is regarded as the *maximal* electric field. Electric fields larger than b would render all physical quantities derivable from $\mathcal{L}_\mathrm{B}$ complex-valued, and no physical significance could be attached to them.

Suppressing all powers of b, write (10.104) in vector notation:

$$\mathcal{L}_\mathrm{BI} = 1 - \sqrt{1 + \mathbf{B}^2 - \mathbf{E}^2 - (\mathbf{E} \cdot \mathbf{B})^2} \, . \tag{10.107}$$

Born and Infeld wanted a theory which is distinguished by invariance under general spacetime diffeomorphisms $x'^\mu = f^\mu(x)$. To this end, they modified the Lagrangian to

$$\mathcal{L}_\mathrm{BI} = \sqrt{-\det(g_{\mu\nu})} - \sqrt{-\det(g_{\mu\nu} + F_{\mu\nu})} \, , \tag{10.108}$$

which can be shown (Problem 10.4.2) to be identical to (10.107) in Cartesian coordinates. In fact, this reasoning is deceptive because every function of $\mathcal{S}$ and/or $\mathcal{P}$ can be arranged to give a diffeomorphism invariant action of the type (5.116).

We now turn to a general nonlinear electrodynamics with Lagrangian $\mathcal{L}(\mathcal{S}, \mathcal{P})$. We do not follow the historical tradition of viewing particles as soliton solutions of the nonlinear field equations, but instead add explicit particle terms to the action

$$S = -\int ds \left(m_0 \sqrt{v \cdot v} + e A_\mu \dot{z}^\mu \right) + \int d^4x \, \mathcal{L}(\mathcal{S}, \mathcal{P}) \ . \tag{10.109}$$

We assume that $\mathcal{L}(\mathcal{S}, \mathcal{P})$ reduces to $-\frac{1}{2}\mathcal{S}$ in the weak field limit.

Let us define the *excitation* 2-form $E = \frac{1}{2} E_{\mu\nu} \, dx^\mu \wedge dx^\nu$ with components

$$E_{\mu\nu} = \frac{\partial \mathcal{L}}{\partial F^{\mu\nu}} \ . \tag{10.110}$$

By differentiation with respect to an antisymmetric tensor we mean the conventional partial differentiation with respect to its independent components. As with symmetric tensors, we may use the rule

$$\frac{\partial F^{\alpha\beta}}{\partial F^{\mu\nu}} = \delta^\alpha_{\ \mu} \, \delta^\beta_{\ \nu} - \delta^\alpha_{\ \nu} \, \delta^\beta_{\ \mu} \ . \tag{10.111}$$

The Euler–Lagrange equations stemming from (10.109) are

$$m_0 a^\mu = e v_\nu F^{\mu\nu} \ , \tag{10.112}$$

$$\partial_\mu E^{\mu\nu} = -j^\nu \ , \tag{10.113}$$

$$j^\mu(x) = e \int_{-\infty}^{\infty} ds \, v^\mu(s) \, \delta^4 \left[x - z(s) \right] \ . \tag{10.114}$$

They should be augmented by the addition of the Bianchi identity

$$\partial_\mu {}^*F^{\mu\nu} = 0\,, \tag{10.115}$$

which is merely a restatement of (10.103), and the *constitutive equations*

$$E^{\mu\nu} = E^{\mu\nu}(F)\,, \tag{10.116}$$

following from (10.110). The constitutive equations of Maxwell's electrodynamics are linear, $E_{\mu\nu} = -F_{\mu\nu}$. In the general case, however, equation (10.110),

$$E^{\mu\nu} = 2\left(\frac{\partial\mathcal{L}}{\partial\mathcal{S}}\,F^{\mu\nu} + \frac{\partial\mathcal{L}}{\partial\mathcal{P}}\,{}^*F^{\mu\nu}\right)\,, \tag{10.117}$$

need not be linear in $F^{\mu\nu}$. Hence the name *nonlinear electrodynamics*. For example, (10.107) implies (Problem 10.4.3) the following constitutive equations

$$E^{\mu\nu} = \frac{-1}{\sqrt{1+\mathcal{S}-\frac{1}{4}\,\mathcal{P}^2}}\left(F^{\mu\nu} - \frac{\mathcal{P}}{2}\,{}^*F^{\mu\nu}\right)\,. \tag{10.118}$$

The inverse of (10.118) is

$$F^{\mu\nu} = \frac{-1}{\sqrt{1-\Sigma-\frac{1}{4}\,\Pi^2}}\left(E^{\mu\nu} + \frac{\Pi}{2}\,{}^*E^{\mu\nu}\right)\,, \tag{10.119}$$

where

$$\Sigma = \frac{1}{2}\,E_{\mu\nu}\,E^{\mu\nu}, \quad \Pi = \frac{1}{2}\,{}^*E_{\mu\nu}\,E^{\mu\nu}\,. \tag{10.120}$$

Equations (10.112)–(10.116) form a complete set of equations of a general nonlinear electrodynamics. It is clear from (10.113) and (10.112) that a point charge generates $E^{\mu\nu}$ but evokes response through $F^{\mu\nu}$. The action–reaction principle is distorted.

Just as $F^{\mu\nu}$ is associated with the electric field intensity $\mathbf{E}$ and the magnetic induction $\mathbf{B}$, so $E^{\mu\nu}$ can be related to the *electric displacement* $\mathbf{D}$ and the *magnetic field intensity* $\mathbf{H}$. Equations (10.113)–(10.116) expressed in terms of $\mathbf{E}, \mathbf{B}$, and $\mathbf{D}, \mathbf{H}$($F_{0i} = E_i, F_{ij} = -\epsilon_{ijk}B_k, E_{i0} = D_i, E_{ij} = \epsilon_{ijk}H_k$),

$$\nabla\times\mathbf{H} = \frac{\partial\mathbf{D}}{\partial t} + \mathbf{j}, \quad \nabla\cdot\mathbf{D} = \varrho\,, \tag{10.121}$$

$$\nabla\times\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t}, \quad \nabla\cdot\mathbf{B} = 0\,, \tag{10.122}$$

$$\mathbf{D} = \mathbf{D}(\mathbf{E},\mathbf{B}), \quad \mathbf{H} = \mathbf{H}(\mathbf{E},\mathbf{B})\,, \tag{10.123}$$

resemble the equations of electromagnetic field in macroscopic media. It is as if the vacuum were a dispersive medium. From the mathematical point of view, we are dealing with a system of first order partial differential equations, linear in derivatives, with coefficients which depend only on unknown variables.

A remarkable fact is that, of all nonlinear versions of electrodynamics with a reasonable weak field limit, only the Born–Infeld theory describes signal propagation without shock waves. We will accept this statement without proof.

It is an easy matter to verify (Problem 10.4.4) that the symmetric stress-energy tensor of electromagnetic field corresponding to an arbitrary Lagrangian $\mathcal{L}(\mathcal{S},\mathcal{P})$ is given by

$$\Theta^{\mu\nu} = -F^{\mu}{}_{\alpha}\, E^{\alpha\nu} - \eta^{\mu\nu}\mathcal{L}\ . \tag{10.124}$$

One can show (Problem 10.4.5) that $\Theta^{\mu\nu}$ obeys

$$\partial_\nu \Theta^{\mu\nu} = -F^{\mu\nu} j_\nu\ . \tag{10.125}$$

Thus, in a region free of electric charges,

$$\partial_\nu \Theta^{\mu\nu} = 0\ . \tag{10.126}$$

Applying (10.117) to (10.124) gives

$$\Theta^{\mu}{}_{\mu} = 4\left(\mathcal{L}_{\mathcal{S}}\,\mathcal{S} + \mathcal{L}_{\mathcal{P}}\,\mathcal{P}\right) - 4\mathcal{L}\ , \tag{10.127}$$

where $\mathcal{L}_{\mathcal{S}} = \partial\mathcal{L}/\partial\mathcal{S}$ and $\mathcal{L}_{\mathcal{P}} = \partial\mathcal{L}/\partial\mathcal{P}$. The question now arises: what should $\mathcal{L}$ be to make $\Theta^{\mu}{}_{\mu} = 0$? Denoting $l = \ln\mathcal{L}$, $s = \ln\mathcal{S}$, $p = \ln\mathcal{P}$, we can bring this equation to the form

$$l_s + l_p = 1\ . \tag{10.128}$$

Integrating this is easy,

$$l = \frac{1}{2}\,(s+p) + u(s-p)\ , \tag{10.129}$$

where u is an arbitrary function. Expressing this in terms of $\mathcal{L}$, $\mathcal{S}$ and $\mathcal{P}$ we have

$$\mathcal{L} = \sqrt{\mathcal{S}\mathcal{P}}\ U\!\left(\frac{\mathcal{S}}{\mathcal{P}}\right)\ , \tag{10.130}$$

where U is an arbitrary differentiable function. For $U = -\frac{1}{2}\sqrt{\mathcal{S}/\mathcal{P}}$, we come to the Larmor Lagrangian $\mathcal{L} = -\frac{1}{2}\,\mathcal{S}$. So Maxwell's electrodynamics is conformally invariant. However, note that $\mathcal{L}_{\mathrm{BI}}$ is outside the class of functions covered by (10.130).

The next important issue is the electric-magnetic duality. We know that Maxwell's equations without sources are invariant under duality transformation (5.217). Is it possible to extend this symmetry to nonlinear modifications of Maxwell's theory? Put $j^\mu = 0$. Then (10.113) and (10.115) are invariant under the SO(2) field rotation

$$E' = E\cos\theta + {}^*F\sin\theta, \quad {}^*F' = {}^*F\cos\theta - E\sin\theta\,. \tag{10.131}$$

This is the desired generalization of (5.217). One may find it more illuminating to show that (10.121) and (10.122) are invariant under the U(1) transformation

$$\mathbf{E}' + i\mathbf{H}' = e^{i\theta}\,(\mathbf{E} + i\mathbf{H}), \quad \mathbf{D}' + i\mathbf{B}' = e^{i\theta}\,(\mathbf{D} + i\mathbf{B})\,. \tag{10.132}$$

Unlike the equations of motion (10.113) and (10.115), the constitutive equations are in general devoid of this invariance. The criterion for duality invariance (Problem 10.4.6) is

$${}^*E_{\mu\nu}\,E^{\mu\nu} = {}^*F_{\mu\nu}\,F^{\mu\nu}\,. \tag{10.133}$$

Using (10.118), it is easy to check that $\mathcal{L}_{\rm BI}$ meets this criterion, hence the Born–Infeld theory is duality invariant.

One may think of (10.133) as a differential equation with $\mathcal{L}$ as the unknown function,

$$4\left(\mathcal{L}_{\mathcal{S}}^2 - \mathcal{L}_{\mathcal{P}}^2\right)\mathcal{P} - 8\,\mathcal{L}_{\mathcal{S}}\,\mathcal{L}_{\mathcal{P}}\,\mathcal{S} - \mathcal{P} = 0\,. \tag{10.134}$$

The Larmor and Born–Infeld Lagrangians obey this equation. The general solution to (10.134) contains an arbitrary function of one variable, but the physical meaning of such Lagrangians remains a mystery.

We now take a closer look at the static case. The second equation (10.121) becomes

$$\nabla\cdot\mathbf{D}(r) = e\,\delta^3(\mathbf{r})\,. \tag{10.135}$$

To be specific, let us turn to the Born–Infeld theory. The constitutive equations (10.118) and (10.119) read

$$\mathbf{E} = \frac{\mathbf{D}}{\sqrt{1 + b^{-2}\mathbf{D}^2}}\,, \quad \mathbf{D} = \frac{\mathbf{E}}{\sqrt{1 - b^{-2}\mathbf{E}^2}}\,, \tag{10.136}$$

where the parameter b has been restored.

The Poisson equation (10.135) is obeyed by the usual Coulomb solution

$$\mathbf{D}(r) = \frac{e}{4\pi r^2}\,\mathbf{n}\,, \tag{10.137}$$

which is singular at $r = 0$. However, the field strength $\mathbf{E}$ derived from (10.137) with the help of (10.136) is regular,

$$\mathbf{E}(r) = \frac{e}{4\pi\sqrt{r^4 + \ell^4}}\,\mathbf{n}\,. \tag{10.138}$$

Here, ℓ is a characteristic length related to the critical field b as

$$b = \frac{e}{4\pi\ell^2}\,. \tag{10.139}$$

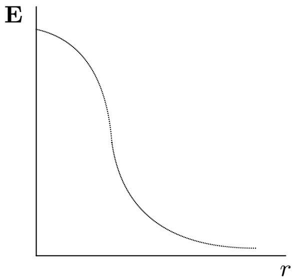


Fig. 10.4. Static solution of the Born–Infeld theory

At large r, $\mathbf{E}(r)$ approaches the Coulomb field. Note also that $\mathbf{E}(r) \to \mathbf{D}(r)$ as $\ell \to 0$. The behavior of the solution (10.138) is shown in Fig. 10.4.

The energy density is obtained from (10.124):

$$\Theta_{00} = -F_{0i}\, E_{0i} - \mathcal{L} = \mathbf{E}\cdot\mathbf{D} - \mathcal{L}\,. \tag{10.140}$$

By (10.105) and (10.138),

$$\Theta_{00} = b^2\left(\sqrt{1 - b^{-2}\mathbf{E}^2} - 1\right) + \mathbf{E}\cdot\mathbf{D} = b^2\left(\sqrt{1 + b^{-2}\mathbf{D}^2} - 1\right)\,. \tag{10.141}$$

Using (10.137) in (10.141), shows that $\Theta_{00} \sim 1/r^2$ near $r = 0$, but this singularity is integrable, and we see that the self-energy is finite,

$$\delta m = \int d^3x\, \Theta_{00} = \frac{e^2}{4\pi\ell}\int_0^\infty dy\left(\sqrt{1+y^4} - y^2\right)\,. \tag{10.142}$$

Consider the stability of a point charge in Born-Infeld electrodynamics. The self-force acting within an infinitesimal solid angle $d\Omega$ is

$$d\mathbf{f} = d\Omega\, e\mathbf{E}(r)|_{r=0} = d\Omega\,\frac{e^2}{4\pi\ell^2}\,\mathbf{n}\,. \tag{10.143}$$

This is balanced by the self-force opposite in direction. Hence, the net effect is zero. One can then explain the stability of a point charge in the Maxwell–Lorentz theory taking (10.143) as a regularized expression for the self-force, integrating this force over solid angle, and taking the limit $\ell = 0$.

The solution (10.138) for a charge at rest can be readily generalized for a moving charge by a Lorentz boost (Problem 10.4.8). However, solutions for an arbitrarily moving charge, similar to the retarded Liénard–Wiechert solution of Maxwell's electrodynamics, are unknown.

Problem 10.4.1. Show that

$$\det\left(F_{\alpha\beta}\right) = \frac{1}{4}\,\mathcal{P}^2\,. \tag{10.144}$$

Hint Use the result of Problem 1.3.7.

Problem 10.4.2. Show that (10.108) is identical to (10.104) for $g_{\mu\nu} = \eta_{\mu\nu}$.

Hint First, verify that $\det(g_{\alpha\beta} + F_{\alpha\beta}) = \det(g_{\alpha\beta} - F_{\alpha\beta})$, which implies that only terms even in F contribute to this determinant. Second, write $F = F_{01}\, dx^1 \wedge dx^2 + F_{23}\, dx^2 \wedge dx^3$, so that $\mathcal{S} = F_{23}^2 - F_{01}^2$. Third, work out $\Delta = \det(\eta_{\mu\nu} + F_{\mu\nu})$:

$$\begin{aligned}\Delta &= \epsilon^{\alpha\beta\gamma\delta}\,(\eta_{\alpha 0} + F_{\alpha 0})(\eta_{\beta 1} + F_{\beta 1})(\eta_{\gamma 2} + F_{\gamma 2})(\eta_{\delta 3} + F_{\delta 3}) \\ &= \det(\eta_{\mu\nu}) + F_{01}^2 - F_{23}^2 + \det(F_{\mu\nu})\ .\end{aligned} \tag{10.145}$$

Finally, use the result of Problem 10.4.1.

Problem 10.4.3. Derive (10.118) and (10.119).

Problem 10.4.4. Verify (10.124).

Hint Write

$$S = \int d^4x\, \sqrt{-g}\, \mathcal{L}(\mathcal{S}, \mathcal{P})\ , \tag{10.146}$$

$$\mathcal{S} = \frac{1}{2}\, g^{\mu\alpha} g^{\nu\beta} F_{\mu\nu} F_{\alpha\beta}, \quad \mathcal{P} = \frac{1}{2\sqrt{-g}}\, \epsilon^{\mu\nu\alpha\beta} F_{\mu\nu} F_{\alpha\beta}\ , \tag{10.147}$$

use the relations

$$\delta g_{\mu\nu} = -\delta g^{\alpha\beta} g_{\mu\alpha}\, g_{\nu\beta}, \quad \delta g = -g\, g_{\mu\nu}\, \delta g^{\mu\nu}\ , \tag{10.148}$$

and take into account (10.117) and (2.96).

Problem 10.4.5. Using (10.113) and (10.115), prove (10.125).

Proof

$$\begin{aligned}\partial_\nu \Theta^{\mu\nu} &= -E^{\alpha\nu} \partial_\nu F^\mu{}_\alpha - F^\mu{}_\alpha\, \partial_\nu E^{\alpha\nu} - \partial^\mu \mathcal{L} \\ &= -E_{\alpha\nu} \partial^\nu F^{\mu\alpha} - F^{\mu\alpha} j_\alpha - \frac{1}{2}\, \frac{\partial \mathcal{L}}{\partial F^{\alpha\nu}}\, \partial^\mu F^{\alpha\nu} \\ &= -\frac{1}{2}\, E_{\alpha\nu}\, (\partial^\nu F^{\mu\alpha} + \partial^\alpha F^{\nu\mu} + \partial^\mu F^{\alpha\nu}) - F^{\mu\alpha} j_\alpha = -F^{\mu\alpha} j_\alpha\ .\end{aligned} \tag{10.149}$$

Problem 10.4.6. Prove that (10.133) is the condition for duality invariance of the constitutive equations (10.116).

Proof Write (10.131) in the infinitesimal form

$$\delta F = \theta\, {}^*E, \quad \delta E = \theta\, {}^*F\ . \tag{10.150}$$

Using (10.110) and the expression for δF,

$$\delta E_{\mu\nu} = \delta F_{\sigma\tau}\,\frac{\partial}{\partial F_{\sigma\tau}}\,\frac{\partial \mathcal{L}}{\partial F^{\mu\nu}} = \theta\,{}^*E_{\sigma\tau}\,\frac{\partial^2 \mathcal{L}}{\partial F_{\sigma\tau}\partial F^{\mu\nu}} = \frac{\theta}{2}\,\epsilon_{\sigma\tau\alpha\beta}\,\frac{\partial \mathcal{L}}{\partial F_{\alpha\beta}}\,\frac{\partial^2 \mathcal{L}}{\partial F_{\sigma\tau}\partial F^{\mu\nu}}$$

$$= \frac{\theta}{4}\,\epsilon_{\sigma\tau\alpha\beta}\,\frac{\partial}{\partial F^{\mu\nu}}\left(\frac{\partial \mathcal{L}}{\partial F_{\alpha\beta}}\,\frac{\partial \mathcal{L}}{\partial F_{\sigma\tau}}\right) = \frac{\theta}{4}\,\epsilon_{\sigma\tau\alpha\beta}\,\frac{\partial}{\partial F^{\mu\nu}}\left(E^{\alpha\beta}\,E^{\sigma\tau}\right)\,. \qquad (10.151)$$

On the other hand,

$$\delta E_{\mu\nu} = \theta\,{}^*F_{\mu\nu} = \frac{\theta}{2}\,\epsilon_{\mu\nu\alpha\beta}\,F^{\alpha\beta}\,. \qquad (10.152)$$

Comparing (10.151) and (10.152), we find the equation

$$\frac{\partial}{\partial F^{\mu\nu}}\left(\frac{1}{2}\,\epsilon_{\sigma\tau\alpha\beta}\,E^{\alpha\beta}\,E^{\sigma\tau}\right) = \epsilon_{\mu\nu\alpha\beta}\,F^{\alpha\beta}\,, \qquad (10.153)$$

whose solution is

$${}^*E_{\alpha\beta}\,E^{\alpha\beta} = \frac{1}{2}\,\epsilon_{\mu\nu\alpha\beta}\,F^{\alpha\beta}\,F^{\mu\nu} = {}^*F_{\mu\nu}\,F^{\mu\nu}\,. \qquad (10.154)$$

Problem 10.4.7. Define

$$\sigma = \mathcal{S}, \quad \pi = \sqrt{\mathcal{S}^2 + \mathcal{P}^2}\,, \qquad (10.155)$$

then (10.134) becomes

$$\mathcal{L}_\sigma^2 - \mathcal{L}_\pi^2 = 1\,. \qquad (10.156)$$

Solve this equation.

Problem 10.4.8. Generalize the static solution (10.138) to the case that the charge e is moving along a straight line with a constant velocity V.

Hint Apart from (10.119) and (10.120), use

$$\begin{aligned} R^\mu &= \rho\,(v^\mu + u^\mu) = r\,(1, \mathbf{n})\,,\\ \rho &= R\cdot v = \gamma\,r\,(1 - \mathbf{n}\cdot\mathbf{V})\,,\\ E^{\mu\nu} &= \frac{e}{4\pi\rho^2}\,(v^\mu u^\nu - v^\nu u^\mu)\,,\\ v^\mu = \gamma\,(1, \mathbf{V}), \quad u^\mu &= \frac{R^\mu}{\rho} - v^\mu = \frac{(1, \mathbf{n})}{\gamma\,(1 - \mathbf{n}\cdot\mathbf{V})} - \gamma\,(1, \mathbf{V})\,,\\ D_i = E_{i0} = E^{0i}, \quad H_k &= \frac{1}{2}\,\epsilon_{klm}\,E^{lm}, \quad \epsilon^{0ijk} = \epsilon_{ijk}\,. \end{aligned} \qquad (10.157)$$

10.5 Nonlocal Interactions

Maxwell–Lorentz electrodynamics

$$S = -m_0 \int ds \sqrt{\dot{z}\cdot\dot{z}} - \int d^4x \left(A_\mu j^\mu + \frac{1}{16\pi} F^{\mu\nu} F_{\mu\nu} \right) , \tag{10.158}$$

$$j^\mu(x) = e \int_{-\infty}^{\infty} ds\, v^\mu(s)\, \delta^4 \left[x - z(s) \right] \tag{10.159}$$

is a simple example of how local interactions are arranged. It is well known that this theory suffers from infinite self-energy, and the same is true of most local field theories.

Early attempts to remedy the situation were to replace the interaction term by

$$\int d^4x \int d^4y\, A_\mu(x)\, F(x-y)\, j^\mu(y) , \tag{10.160}$$

where the *form factor* F is a smooth function of $(x-y)^2$ which looks like a sharp pulse normalized to unit area, as, for instance,

$$F(x-y) = F_0 \exp\left\{ - \left[\frac{(x-y)^2}{\ell^2} \right]^2 \right\} . \tag{10.161}$$

Interaction (10.160) can be interpreted in either of two ways. First, we may regard A_μ as a local field coupled *locally* to the *smeared*, or *effective* source

$$\mathrm{j}^\mu(y) = \int d^4x\, F(x-y)\, j^\mu(x) . \tag{10.162}$$

Under this interpretation one pictures the charge as if it were distributed over a region of size ℓ. Second, we may imagine that a *point* charge forming the current (10.159) interacts with a *nonlocal* field

$$\mathfrak{A}_\mu(x) = \int d^4y\, F(x-y)\, A_\mu(y) . \tag{10.163}$$

A central problem of the form factor theory (10.160) is causality violation. We might attempt to limit this by assuming the support of $F(x-y)$ is *compact*, and its characteristic size ℓ is small[5], keeping in mind that causality must be regained for macroscopic distances. However, this leads to problems with Lorentz invariance.

[5] The smallest length constructed from constants of nature (the velocity of light c, Planck's constant $\hbar$, and Newton's constant G) is the Planck length $l_{\mathrm{P}} = (\hbar G/c^3)^{1/2} = 1.6 \times 10^{-33}$ cm. In regions of size $\sim l_{\mathrm{P}}$, quantum fluctuations of the metric become significant, and the usual relations between cause and effect may not apply. However, this topic is beyond the scope of the present discussion.

If we would require the form factor to be a function of $(x-y)^2$, then this would conflict with compactness. The invariant region where noncausal phenomena are confined

$$(x-y)^2 < \ell^2 \tag{10.164}$$

is noncompact. Indeed, near the light cone $(x-y)^2 = 0$, the extension of this region in space is arbitrarily large.

Let q^μ be a unit vector. Introducing a positive definite quadratic form

$$d(x,y) = [q \cdot (x-y)]^2 - (x-y)^2 \ , \tag{10.165}$$

and assuming F to be a function of $d(x,y)$, enables us to limit causality violation to a compact, invariant region: $d(x,y) \le \ell^2$. The vector q^μ may be interpreted as the four-velocity of some particle v^μ. However, in the absence of particles, we are forced to use a fixed unit vector q^μ, which would distinguish a privileged frame of reference, and violate explicit Lorentz invariance. Of course, one may regard q^μ as an auxiliary unit vector, and average F over directions of q^μ, but this procedure is rather arbitrary.

Another line of attack is due to Gariĭ Efimov. Consider form factors obtained by acting an entire function of the d'Alembertian on the Dirac delta-function

$$F(x-y) = K(\Box)\, \delta^4(x-y) = \sum_{n=0}^{\infty} c_n \Box^n \delta^4(x-y) \ . \tag{10.166}$$

The relativistic invariance of $F(x-y)$ is apparent: $\delta^4(x-y)$ is invariant under Poincaré transformations, and $\Box\, \delta^4(x-y)$ shares this property (see Appendix F). With the Fourier transform

$$F(x) = \frac{1}{(2\pi)^4} \int d^4x\, e^{-ik\cdot x}\, \widetilde{F}(k) \ , \tag{10.167}$$

(10.166) becomes

$$\widetilde{F}(k) = K(-k^2) = \sum_{n=0}^{\infty} c_n k^{2n} \ . \tag{10.168}$$

The power series (10.168) represents an analytic function. The radius of convergence of this series depends on c_n. We will discuss only those power series which are convergent in the whole complex k^2-plane. In other words, $K(-k^2)$ is assumed to be an entire function.

If the decrease of the coefficients c_n is so steep that $c_n = 0$ for $n \ge N$, then $K(-k^2)$ is a polynomial. In this case we are dealing with an ordinary higher-derivative Lagrangian. This suggests that if the coefficients c_n decrease too much rapidly (even though their sequence does not terminate), then the interaction is not smeared out, but remains in in fact local. Such interactions are called *localizable*. The line of demarcation between localizable and nonlocal interactions separates entire functions $K(-k^2)$ into two classes:

$$\text{(L)} \qquad \lim_{n\to\infty} n\,|c_n|^{1/n} = 0\ ,$$
$$\text{(N)} \qquad \lim_{n\to\infty} n\,|c_n|^{1/n} = A\ . \tag{10.169}$$

With reference to Appendix F, condition (L) can be shown to be equivalent to the following bound of asymptotic growth of $K(-k^2)$ as k^2 approaches infinity in the complex plane:

$$\text{(L)} \qquad |K(-k^2)| < C\exp(\epsilon\sqrt{|k^2|})\ , \tag{10.170}$$

where ϵ is arbitrarily small. As for nonlocal interactions, we restrict our discussion to the class of form factors $K(-k^2)$ satisfying the asymptotic condition

$$\text{(N)} \qquad |K(-k^2)| < C\exp(\sigma\sqrt{|k^2|}) \tag{10.171}$$

for some fixed, positive σ.

Let us modify (10.158) by introducing a form factor of the type (10.166):

$$S = -m_0\int ds\sqrt{\dot z\cdot\dot z} - \int d^4x\left(A_\mu K(\Box)j^\mu + \frac{1}{16\pi}\,F^{\mu\nu}F_{\mu\nu}\right)\ , \tag{10.172}$$

where j^μ is defined in (10.159).

Varying the world line, we obtain the equation of motion for a bare particle

$$m_0 a^\lambda = e v_\mu K(\Box) F^{\lambda\mu}\ . \tag{10.173}$$

Varying the vector potential, we get the equation of motion for the electromagnetic field

$$\partial_\mu F^{\mu\nu} = 4\pi K(\Box) j^\nu\ . \tag{10.174}$$

Consider the static case[6]. Equation (10.174) becomes

$$\nabla^2\phi(\mathbf{r}) = -4\pi K(-\nabla^2)\,\delta^3(\mathbf{r})\ . \tag{10.175}$$

We have to see whether it is possible to make the solution $\phi(\mathbf{r})$ more regular than $1/r$ by acting $K(-\nabla^2)$ on $\delta^3(\mathbf{r})$.

To maintain the link with the Maxwell–Lorentz theory, we impose the condition

$$K(0) = 1\ . \tag{10.176}$$

Substituting

$$\phi(\mathbf{r}) = \frac{1}{(2\pi)^3}\int d^3k\, e^{i\mathbf{k}\cdot\mathbf{r}}\,\widetilde{\phi}(\mathbf{k}) \tag{10.177}$$

in (10.175) gives

[6] To simplify writing, throughout this section the electric charge e is set equal to unity.

$$\phi(\mathbf{r}) = \frac{1}{(2\pi)^3}\int d^3k\, e^{i\mathbf{k}\cdot\mathbf{r}}\,\frac{4\pi K(\mathbf{k}^2)}{\mathbf{k}^2} = \frac{2}{\pi r}\int_0^\infty dk\,\frac{\sin kr}{k}\,K(k^2)$$

$$= \frac{1}{\pi r}\,\mathrm{P}\!\int_{-\infty}^\infty dk\,\frac{K(k^2)}{k}\,\sin kr = \frac{1}{\pi r}\,\mathrm{Im}\left\{\int_{-\infty}^\infty dk\left[\frac{1}{k+i\epsilon} + i\pi\delta(k)\right]K(k^2)\,e^{ikr}\right\}$$

$$= \frac{1}{\pi r}\,\mathrm{Im}\left[\int_{-\infty}^\infty dk\,\frac{1}{k+i\epsilon}\,K(k^2)\,e^{ikr}\right] + \frac{1}{r}\ , \tag{10.178}$$

where P stands for the Cauchy principal value. Note that we have used equation (F.42) of Appendix F, and condition (10.176).

Let $K(k^2)$ be an entire function obeying (10.170). For large R, we define

$$h_R(r) = \frac{1}{\pi}\int_{-R}^R dk\,\frac{K(k^2)}{k+i\epsilon}\,e^{ikr}\ . \tag{10.179}$$

By Cauchy's theorem, integration over the real axis can be replaced by integration over a semicircle Γ_R of radius R at the upper half-plane $\mathrm{Im}\,k > 0$. Letting $k = Re^{i\vartheta}$,

$$\left|\int_{\Gamma_R}\frac{dk}{k}\,K(k^2)\,e^{ikr}\right| \le \int_{\Gamma_R}\left|\frac{dk}{k}\,K(k^2)\,e^{ikr}\right| < \int_0^\pi d\vartheta\,\exp(\epsilon R)\,\exp(-rR\sin\vartheta)\ , \tag{10.180}$$

where we have taken into account (10.170), and

$$|e^{ikr}| = \exp|iR\,(\cos\vartheta + i\sin\vartheta)r| = \exp(-rR\sin\vartheta)\ . \tag{10.181}$$

In the sector $0 < \vartheta \le \pi/2$, the following inequality is helpful,

$$\sin\vartheta \ge \frac{2}{\pi}\,\vartheta\ . \tag{10.182}$$

From which we conclude

$$\int_0^\pi d\vartheta\,e^{\epsilon R}\,e^{-rR\sin\vartheta} = 2e^{\epsilon R}\int_0^{\pi/2} d\vartheta\,e^{-rR\sin\vartheta} < 2e^{\epsilon R}\int_\delta^{\pi/2} d\vartheta\,e^{-\frac{2}{\pi}\vartheta rR}$$

$$= \frac{\pi}{rR}\left[e^{-R\left(\frac{2}{\pi}r\delta-\epsilon\right)} - e^{-R(r-\epsilon)}\right]\ . \tag{10.183}$$

For finite r and $\epsilon < 2r\delta/\pi$, this expression vanishes in the limit $R\to\infty$.

To sum up, if $K(k^2)$ obeys (10.170), then $h_R(r)\to 0$, and the potential (10.178) is just $1/r$ away from $r = 0$. We thus see that the singularity of $\phi(\mathbf{r})$ remains unaffected by applying $K(-\nabla^2)$ to $\delta^3(\mathbf{r})$, and, furthermore, the self-energy problem is not solved (Problem 10.5.2).

We next consider nonlocal interactions. Suppose that

$$|K(R^2e^{2\vartheta})| \le C\exp(\ell R\sin\vartheta) \quad \text{as} \quad R\to\infty\ . \tag{10.184}$$

Then

$$\left|\int_{\Gamma_R} \frac{dk}{k}\, K(k^2)\, e^{ikr}\right| \leq 2C \int_0^{\frac{\pi}{2}} d\vartheta\, e^{R\sin\vartheta(\ell-r)} \leq 2C \int_0^{\frac{\pi}{2}} d\vartheta\, e^{\frac{2}{\pi}R\vartheta(\ell-r)}$$

$$= \frac{\pi C}{R(\ell-r)} \left[e^{R(\ell-r)} - 1\right] . \tag{10.185}$$

For $r > \ell$, this expression vanishes in the limit $R \to \infty$. But $r < \ell$ is another matter. Let us write the potential which obeys (10.175) in the form

$$\phi(r) = \theta(r)\, \frac{\alpha(r)}{r} \tag{10.186}$$

implying that the range of r is the real axis. Here, α is a differentiable function satisfying two conditions:

$$\alpha(s) = \alpha_0 s + O(s^3), \quad s \to 0\,, \tag{10.187}$$

$$\alpha(s) = 1, \quad |s| \geq \ell\,. \tag{10.188}$$

We therefore have

$$\alpha'(s) = \frac{1}{\pi} \int_{-\infty}^{\infty} dk\, K(k^2)\, e^{iks}\,, \tag{10.189}$$

the prime being differentiation with respect to s, and the inverse relation

$$K(k^2) = \frac{1}{2} \int_{-\ell}^{\ell} ds\, \alpha'(s)\, e^{-iks}\,. \tag{10.190}$$

This formula is convenient for constructing $K(k^2)$ with the two required properties (10.176) and (10.184). One can to show (Problem 10.5.1) that if $\alpha(s)$ is a differentiable function obeying (10.187) and (10.188), then (10.190) represents an entire function $K(k^2)$ of order $\frac{1}{2}$ with indicatrix $H(\vartheta) = \ell \sin\vartheta$, which is square integrable and normalized to $K(0) = 1$.

As a simple example of functions $K(k^2)$ and $\alpha'(s)$, appearing in (10.189) and (10.190), we could take

$$K(k^2) = \frac{\sin(k\ell)}{k\ell}\,, \quad \alpha'(s) = \begin{cases} \ell^{-1} & |s| < \ell\,, \\ 0 & |s| \geq \ell. \end{cases} \tag{10.191}$$

The corresponding potential $\phi(r)$, shown in Fig. 10.5, is a truncated Coulomb potential. A similar solution was found in Maxwell's electrodynamics for a charged sphere of radius ℓ (Problem 3.3.1). However, this similarity is deceptive. If the form factor $K(k^2)$ is given by (10.190), then it is easy to verify (Problem 10.5.4) that the self-force

$$d\mathbf{f} = d\Omega \int_0^{\infty} dr\, r^2 \varrho(r)\, \mathbf{E}(r) \tag{10.192}$$

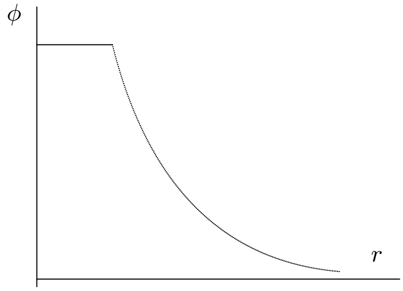


Fig. 10.5. Static potential of a nonlocal electrodynamics

is zero. Hence field configurations are stable without resort to Poincaré cohesive forces.

A further distinction between the charged sphere and smearing by the form factor (10.191) is that external forces acting of the sphere can disturb its shape and charge distribution, while the truncated potential in the nonlocal theory is always the same.

The square integrability of $K(k^2)$ implies that the self-energy is finite:

$$\delta m = \int d^3x \, \frac{\mathbf{E}^2}{8\pi} = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \, K^2(k^2) < \infty \, . \tag{10.193}$$

Consider the general case that a charge is moving along an arbitrary world line. To solve the field equation (10.174), we adopt the retarded boundary condition, and impose Lorenz gauge. The solution is given by

$$A^\mu(x) = -\frac{4\pi}{(2\pi)^4} \int d^4k \, e^{-ik\cdot x} \, \frac{K(-k^2)}{k^2 + 2ik_0\epsilon} \, \tilde{\jmath}^\mu(k) \, , \tag{10.194}$$

where

$$\tilde{\jmath}^\mu(k) = \int d^4x \, e^{ik\cdot x} \, j^\mu(x) = \int_{-\infty}^{\infty} d\tau \, e^{ik\cdot z(\tau)} \, \dot{z}^\mu(\tau) \, . \tag{10.195}$$

Looking at (10.184), we observe that $K(-k^2)$ grows exponentially when $k^2 \to \infty$. This implies that the integral (10.194) fails to converge unless $\tilde{\jmath}^\mu(k)$ falls off appropriately in timelike directions. One can demonstrate (Problem 10.5.7) the existence of world lines $z^\mu(\tau)$ for which

$$|\tilde{\jmath}^\mu(k)| < Ce^{-(\ell+\delta)\sqrt{k^2}}, \quad k^2 \to \infty \, , \tag{10.196}$$

with δ being some positive constant, so that expression (10.194) proves well defined.

Using $K(-k^2)$ given by (10.190), one would intuitively expect to get $A^\mu(x)$ identical to the Liénard–Wiechert vector potential everywhere outside a thin tube of radius $\sim\ell$ enclosing the world line (which implies that noncausal phenomena are confined to the region bounded by this tube). A rigorous proof of this statement is left to the careful reader.

The introduction of $K(\Box)$ exerts some effect on symmetries of the Maxwell–Lorentz theory. The presence of the dimensional parameter ℓ violates conformal invariance. By contrast, gauge invariance is not affected. Indeed, gauge transformations $\delta A_\mu = \partial_\mu \chi$ leave the interaction term unchanged if the local current is conserved $\partial_\mu j^\mu = 0$,

$$-\int d^4x\, j^\mu\, K(\Box)\, \partial_\mu \chi = \int d^4x\, \partial_\mu j^\mu\, K(\Box)\, \chi = 0\,. \tag{10.197}$$

A subtle point is introducing nonlocality into non-Abelian gauge theories. The lattice formulation of gauge theories, outlined in Sect. 7.3, amounts to using the form factor

$$F(x) = \delta(x^0 - \ell_0)\, \delta(x^1 - \ell_1)\, \delta(x^2 - \ell_2)\, \delta(x^3 - \ell_3)$$
$$= \exp\left(\ell_0 \frac{\partial}{\partial x^0} + \ell_1 \frac{\partial}{\partial x^1} + \ell_2 \frac{\partial}{\partial x^2} + \ell_3 \frac{\partial}{\partial x^3}\right) \delta(x_0)\, \delta(x_1)\, \delta(x_2)\, \delta(x_3)\,, \tag{10.198}$$

whose Fourier transform is an entire function $\widetilde{F}(k) = \exp(i\ell_\mu k^\mu)$. Thus, it is possible to retain gauge invariance with nonlocality at the sacrifice of Poincaré invariance.

Problem 10.5.1. Let $K(k^2)$ be an entire function of order $\frac{1}{2}$ and type ℓ. Suppose that $K(k^2)$ is square integrable on the real axis $K \in L_2$. Prove that $\alpha'(s)$ defined in (10.189) is zero for $r \geq \ell$. Conversely, let $\alpha(s)$ be a differentiable function obeying (10.187) and (10.188). Prove that (10.190) represents an entire function $K(k^2)$ of order $\frac{1}{2}$, which is square integrable on the real axis, normalized to $K(0) = 1$, and whose indicatrix is $H(\vartheta) = l \sin\vartheta$. This statement is known as the *Paley–Wiener theorem.*

Problem 10.5.2. Let $K(k^2)$ be an entire function, which is not identically zero, and such that

$$\lim_{R\to\infty} \frac{\ln|K(R^2 e^{2i\vartheta})|}{R} = 0\,. \tag{10.199}$$

Prove that $K(k^2)$ cannot be square integrable on the real axis.

Proof Assume that $K(k^2)$ is square-integrable. By the Paley–Wiener theorem, the Fourier transform of $K(k^2)$ vanishes almost everywhere outside the interval $(-\epsilon, \epsilon)$ for any ϵ. Hence $K(k^2)$ is equivalent to zero, contrary to the initial assumption.

Problem 10.5.3. Consider an entire function $K(k^2)$ described by the Paley–Wiener theorem. Show that

$$K^2(k^2) = \frac{1}{2}\int_{-2\ell}^{2\ell} ds\, \beta'(s)\, e^{-iks} \;, \tag{10.200}$$

where $\beta(s)$ is a function obeying (10.187) and (10.188) with the substitution of 2ℓ for ℓ. It follows that any power of $K(k^2)$ is an entire function of order $\frac{1}{2}$. Express $\beta(s)$ in terms of $\alpha(s)$.

Answer

$$\beta(r) = \frac{1}{2}\int_0^r ds \int_{-\ell}^{\ell} dt\, \alpha'(s-t)\, \alpha'(t) \;. \tag{10.201}$$

Problem 10.5.4. Show that the self-force (10.192) is zero for any form factor $K(k^2)$ defined by (10.190). What is $d\mathbf{f}$ for a charged sphere in Maxwell's electrodynamics?

Hint Substitute $\varrho(r) = \delta(r)/4\pi r^2$ and $\mathbf{E}(r) = -K(-\nabla^2)\nabla\phi(r) = -\mathbf{n}\,[\beta(r)/r]'$ in (10.192), where $\beta(r)$ is given by (10.201), and take into account that $\beta''(r) = O(r)$ as $r \to 0$. For a charged sphere, $\varrho(r) = \delta(r-\ell)/4\pi r^2$, $\mathbf{E}(r) = \mathbf{n}/r^2$, and $d\mathbf{f} = \mathbf{n}\, d\Omega/4\pi\ell^2$.

Problem 10.5.5. Derive (10.193).

Problem 10.5.6. Show that

$$\delta m = \frac{1}{2}\,\beta'(0) \;, \tag{10.202}$$

where $\beta'(r)$ is related to $K^2(k^2)$ by (10.200).

Problem 10.5.7. What world lines $z^\mu(s)$ ensure (10.196)?

Hint For $k^2 > 0$, we choose a Lorentz frame in which $k^\mu = (\omega, 0, 0, 0)$, and denote $z^0 = t$. Then (10.195) becomes

$$\tilde{\jmath}^\mu(\omega) = \int_{-\infty}^{\infty} dt\, e^{i\omega t}\, \frac{dz^\mu}{dt} \;. \tag{10.203}$$

Observing that $dz^\mu/dt = (1, \mathbf{v})$, we have $\tilde{\jmath}^0(\omega) = 2\pi\delta(\omega)$, and

$$\tilde{\mathbf{j}}(\omega) = \int_{-\infty}^{\infty} dt\, e^{i\omega t}\, \mathbf{v}(t) \;. \tag{10.204}$$

We write $\mathbf{v} = \mathbf{v}_+ + \mathbf{v}_- + \mathbf{v}_{\rm out}$ with $\mathbf{v}_\pm(t) = \frac{1}{2}\,[\mathbf{v}(t) \pm \mathbf{v}(-t)] - \frac{1}{2}\,(\mathbf{v}_{\rm out} \pm \mathbf{v}_{\rm in})$, and assume that $\mathbf{v}_\pm(t+is)$ are holomorphic in a strip

$$-\infty < t < \infty, \quad |s| < \ell + \varDelta \;, \tag{10.205}$$

for some $\varDelta > 0$. We also assume that these functions are integrable everywhere in this strip

$$\int_{-\infty}^{\infty} dt\, |\mathbf{v}_{\pm}(t+is)| < C \tag{10.206}$$

(in fact, these assumptions are essential only for $\mathbf{v}_{+}$). Then it follows that

$$\left|\int_{-\infty}^{\infty} dt\, e^{i\omega(t+is)}\, \mathbf{v}_{\pm}(t+is)\right| \le e^{-\omega s} \int_{-\infty}^{\infty} dt\, |\mathbf{v}_{\pm}(t+is)|\ , \tag{10.207}$$

and hence $|\tilde{\mathbf{j}}(\omega)| < Ce^{-|\omega|(\ell+\delta)}$ for $0 < \delta < \varDelta$.

Thus, the desired world lines are smooth timelike curves, subject to the conditions of analyticity in the strip (10.205) and integrability (10.206). Non-Galilean regimes do not fulfil the asymptotic conditions $a^{\mu}(s) \to 0$, $s \to \pm\infty$, and are therefore excluded from this class of allowable world lines.

Problem 10.5.8. Consider world lines for which the acceleration and its derivatives are small: $\ell^2|a^2| \ll 1$, $\ell^4|\dot{a}^2| \ll 1$, ... Show that the equation of motion for a dressed particle in nonlocal electrodynamics (10.172) is approximated by the Lorentz–Dirac equation

$$ma^{\mu} - \frac{2}{3}\, e^2 \left(\dot{a}^{\mu} + a^2 v^{\mu}\right) = f^{\mu}\ , \tag{10.208}$$

where $m = m_0 + \delta m$, with δm being represented by (10.202), and f^{μ} an external force.

Hint Note that

$$\frac{1}{k^2 + ik_0\epsilon} = \mathrm{P}\left(\frac{1}{k^2}\right) - i\pi\, \mathrm{sgn}(k_0)\, \delta(k^2)\ , \tag{10.209}$$

where $\mathrm{sgn}(k_0) = k_0/|k_0|$, and

$$2i\pi\, \mathrm{sgn}(k_0)\, \delta(k^2)\, K(k^2) = 2i\pi\, \mathrm{sgn}(k_0)\, \delta(k^2) = \widetilde{D}(k)\ , \tag{10.210}$$

and employ the procedure of Sect. 6.5.

10.6 Action at a Distance

In analyzing the energy-momentum content of the Maxwell–Lorentz theory

$$S = -\sum_{I=1}^{N} \int d\tau_I \left[m_0^I \sqrt{\dot{z}_I^2} + e_I \dot{z}_I^{\mu} A_{\mu}(z_I)\right] - \frac{1}{16\pi} \int d^4x\, F_{\mu\nu}F^{\mu\nu}\ , \tag{10.211}$$

we observed that the degrees of freedom appearing in (10.211) are rearranged into dressed particles and radiation. However, the line of argument we took is more cumbersome than that used for rearranging the Higgs model in which one merely changes variables in the Lagrangian. This raises the question:

is it possible to rearrange the Maxwell–Lorentz theory on the Lagrangian level? When the procedure of Sects. 6.2 and 6.3 is reviewed, three obstacles are immediately apparent. First, evaluating the invariant $\mathcal{S}$ for the Liénard–Wiechert field

$$F = e\,\frac{c \wedge V}{\rho^2} \tag{10.212}$$

gives

$$\mathcal{S} = -\frac{e^2}{\rho^4}\ , \tag{10.213}$$

which shows that the last term of (10.211) has a nonintegrable singularity. Substituting

$$A^\mu = e\,\frac{v^\mu}{\rho} \tag{10.214}$$

into the middlemost term leads to a further ultraviolet divergence. Second, the retarded Liénard–Wiechert field due to a single charge F reveals an extra $\mathrm{SL}(2,\mathbb{R})$ symmetry, which implies that radiation drops out of the problem. Indeed, the long-range part of the field (10.212), which goes like ρ^{-1}, makes no contribution to (10.213). Third, the equation of motion for a dressed particle is irreversible, which casts suspicion on the possibility to accomodate the rearranged dynamics to the Lagrangian formalism.

However, the divergent terms can be assembled (Problem 10.6.1) into the expression

$$-\lim_{\epsilon\to 0}\left(\frac{e_I^2}{2\epsilon}\right)\int ds_I\ . \tag{10.215}$$

After mass renormalization this results in a particle term with a finite mass m_I.

To remove the second and third obstacles, one may try to abandon the retarded field in favor of half-retarded and half-advanced fields. Since the $\mathrm{SL}(2,\mathbb{R})$ symmetry of F is now spoiled, $\mathcal{S}$ would furnish information on the radiation. Furthermore, the combination $\frac{1}{2}\left(A_{\mathrm{ret}} + A_{\mathrm{adv}}\right)$ is symmetric in time, which gives promise that the rearranged dynamics still reflects the time reversal invariance of the original Lagrangian. Indeed, omitting the self-field contribution, the Schwarzschild term of the action becomes either

$$-\frac{1}{2}\sum_I e_I \int d\tau_I \sum_{J(\neq I)}\left[A^{(J)}_{\mathrm{ret}}(z_I) + A^{(J)}_{\mathrm{adv}}(z_I)\right]\cdot \dot z_I(\tau_I) \tag{10.216}$$

or

$$-\frac{1}{2}\sum_J e_J \int d\tau_J \sum_{I(\neq J)}\left[A^{(I)}_{\mathrm{ret}}(z_J) + A^{(I)}_{\mathrm{adv}}(z_J)\right]\cdot \dot z_J(\tau_J)\ , \tag{10.217}$$

where $A^{(I)}_{\mathrm{ret}}(z_J)$ and $A^{(I)}_{\mathrm{adv}}(z_J)$ are the retarded and advanced vector potentials at z_J coming from the Ith world line. Expressions (10.216) and (10.217) are symmetric both in I and J, and under interchange of past and future. This would yield the action expressed in terms of particle variables.

The net result, suggested by Fokker in 1929, reads

$$S_{\mathrm{F}} = -\sum_I \int d\tau_I \left\{ m_I \sqrt{\dot{z}_I^2} + \frac{1}{2} \int d\tau_J \sum_{J(\neq I)} e_I e_J \dot{z}_I^\mu(\tau_I) \dot{z}_\mu^J(\tau_J)\, \delta\big[(z_I - z_J)^2\big] \right\} . \tag{10.218}$$

This is the so-called *action-at-a-distance* theory, which also bears the name of the *direct particle* theory. Owing to the delta-function, the typical points z_I and z_J on the Ith and Jth world lines can be thought of as 'interacting' if they are connected by a null interval (which is a relativistic generalization of interactions by contact occurring at zero distance). Note that the Fokker action involves retarded and advanced interactions on an equal footing. There are no unconstrained field degrees of freedom. It is as if particle I were affected by particle J *directly*, that is, without mediation of the electromagnetic field. Wheeler and Feynman assumed that radiation is completely absorbed.

What is the precise formulation of this assumption? Recall that the retarded vector potential in Maxwell–Lorentz theory can be decomposed into two terms

$$A^\mu_{\mathrm{ret}} = \frac{1}{2}\left(A^\mu_{\mathrm{ret}} + A^\mu_{\mathrm{adv}}\right) + \frac{1}{2}\left(A^\mu_{\mathrm{ret}} - A^\mu_{\mathrm{adv}}\right) = A^\mu_{(+)} + A^\mu_{(-)} \, . \tag{10.219}$$

The last term

$$A^\mu_{(-)} = \frac{1}{2}\left(A^\mu_{\mathrm{ret}} - A^\mu_{\mathrm{adv}}\right) \tag{10.220}$$

obeys the homogeneous wave equation

$$\Box\, A^\mu_{(-)} = 0 \, . \tag{10.221}$$

With zero initial data on a spacelike hyperplane Σ, $A^\mu_{(-)}|_\Sigma = 0$ and $(n\cdot\partial)\, A^\mu_{(-)}|_\Sigma = 0$, the solution to the Cauchy problem for the wave equation (10.221) is trivial

$$A^\mu_{(-)}(x) = 0 \, . \tag{10.222}$$

Another way of looking at $A^\mu_{(-)}$ is using the 'Green's function'

$$D(x) = \frac{1}{2}\left[D_{\mathrm{ret}}(x) - D_{\mathrm{adv}}(x)\right] = \mathrm{sgn}\,(x_0)\, \delta(x^2) \tag{10.223}$$

in

$$A^\mu_{(-)}(x) = 4\pi \int d^4y\, D(x-y)\, j^\mu(y) \, . \tag{10.224}$$

Note that $A^\mu_{(-)}(x)$ obeys the homogeneous wave equation (even though the source j^μ is involved in this construction) because $\Box_x D(x-y) = 0$.

Let $A^\mu_{(-)}(x)$ be the total field due to all charges in the universe. If $A^\mu_{(-)}(x)$ vanishes at one time, then it is zero at all times. Wheeler and Feynman adopted (10.222) as a supplementary constraint to the Fokker action, and interpreted it as the condition of *total absorption*. Therefore this approach is often referred to as the *absorber theory of radiation*.

Note, however, that (10.222) does not amount to the lack of radiation in the sense of the definition (6.87)–(6.89). It will be shown below that the radiation effect manifests itself in the local energy-momentum balance.

Let us turn to the dynamics underlying the Fokker action. The interactions between particles are such that they simulate the field between them. The vector potential and the field strength *adjunct* to particle I are determined by the motion of that particle, and are given by half the retarded and half the advanced solutions to Maxwell's equations:

$$A_\mu^{(I)}(x) = \int d^4y\, D_P(x-y)\, j_\mu^{(I)}(y)\;, \tag{10.225}$$

$$F_{\lambda\mu}^{(I)} = \partial_\lambda A_\mu^{(I)} - \partial_\mu A_\lambda^{(I)}\;, \tag{10.226}$$

where

$$D_P(x) = \frac{1}{2}\left[D^{\rm ret}(x) + D^{\rm adv}(x)\right] = \delta(x^2)\;, \tag{10.227}$$

and

$$j_\mu^{(I)}(x) = e_I \int d\tau_I\, \dot z_\mu^I(\tau_I)\, \delta^4\left[x - z_I(\tau_I)\right]\;. \tag{10.228}$$

These quantities identically satisfy the wave equation and the Lorenz gauge condition:

$$\Box\, A_\mu^{(I)} = 4\pi\, j_\mu^{(I)}\;, \tag{10.229}$$

$$\partial^\mu A_\mu^{(I)} = 0\;. \tag{10.230}$$

Rewrite (10.225) in the form

$$A_\mu^{(I)}(x) = e_I \int d\tau_I\, \dot z_\mu^I(\tau_I)\, \delta\left[(x - z_I)^2\right]\;, \tag{10.231}$$

and express the action (10.218) in terms of $A_\mu^{(I)}(x)$. Varying the Ith world line, we have

$$\delta S = \sum_I \int d\tau_I \left[m_I \frac{d}{d\tau_I}\left(\frac{\dot z_\lambda^I}{\sqrt{\dot z_I\cdot \dot z_I}}\right) - e_I \sum_{J(\neq I)} \left(\frac{\partial A_\mu^{(J)}}{\partial z_I^\lambda} - \frac{\partial A_\lambda^{(J)}}{\partial z_I^\mu}\right) \dot z_I^\mu\right] \delta z_I^\mu\;, \tag{10.232}$$

and so

$$m_I a_\lambda^I = e_I v_I^\mu \sum_{J(\neq I)} F_{\lambda\mu}^{(J)}\;. \tag{10.233}$$

This equation differs in two respects from the equation of motion for a bare charged particle, derivable from the action (10.211). First, m_I is the renormalized mass. Second, the Lorentz force exerted on particle I involves the symmetric combination (half-retarded plus half-advanced) of fields due to all particles, except that for particle I itself:

$$\frac{1}{2}\sum_{J(\neq I)}\left[F^{(J)}_{\rm ret}(z_I)+F^{(J)}_{\rm adv}(z_I)\right]\,. \tag{10.234}$$

Going back to the question posed at the beginning of this section, we can conclude that rearranging degrees of freedom on the Lagrangian level leads to concepts quite different from dressed particles and radiation. To fill the gap between the two results, we should invoke the Wheeler–Feynman condition (10.222). Indeed, let us write the field (10.234) in the form

$$\sum_{J(\neq I)}F^{(J)}_{\rm ret}(z_I)+\frac{1}{2}\left[F^{(I)}_{\rm ret}(z_I)-F^{(I)}_{\rm adv}(z_I)\right]-\frac{1}{2}\sum_{J}\left[F^{(J)}_{\rm ret}(z_I)-F^{(J)}_{\rm adv}(z_I)\right]\,, \tag{10.235}$$

where the last term is the sum over all particles. By (10.222),

$$\sum_{J}\left[F^{(J)}_{\rm ret}(z_I)-F^{(J)}_{\rm adv}(z_I)\right]=0 \tag{10.236}$$

at every point on the world line of particle I. Therefore, (10.234) becomes

$$\sum_{J(\neq I)}F^{(J)}_{\rm ret}(z_I)+\frac{1}{2}\left[F^{(I)}_{\rm ret}(z_I)-F^{(I)}_{\rm adv}(z_I)\right]\,. \tag{10.237}$$

The expression in the square bracket is elaborated further (as in Sect. 6.5) to give the Abraham term Γ^{μ}, and we come to the Lorentz–Dirac equation

$$m_I a^I_\lambda-\frac{2}{3}\,e_I^2\left(\dot a^I_\lambda+a_I^2 v^I_\lambda\right)=e_I v_I^\mu\sum_{J(\neq I)}F^{(J)\,\rm ret}_{\lambda\mu}(z_I)\,. \tag{10.238}$$

Thus, the dynamical equation (10.233) subject to the constraint (10.236) is equivalent to the conventional equation of motion for a dressed particle in the retarded field of all other particles. Furthermore, (10.238) represents local energy-momentum balance (6.156),

$$\dot p_I^\mu+\dot{\mathcal{P}}_I^\mu+\dot\wp_I^\mu=0\,, \tag{10.239}$$

implying that radiation effects have been incorporated in the action-at-a-distance theory.

Wheeler and Feynman assumed that the total matter in the universe behaves as a perfect absorber, and proposed (10.236) as a cosmological absorber condition. If we keep track of particle I, then the radiation of this particle is to be completely absorbed by other particles. The absorber exerts on particle I a force which is the sum of retarded forces due to other particles, and endows it with the four-momentum $p_I^\mu=m_I v^\mu-\frac{2}{3}\,e_I^2 a_I^\mu$. Recall that any point object with such a four-momentum becomes a tachyon if the magnitude of its four-acceleration exceeds $3m_I/2e_I^2$. To avoid tachyonic states, we should require that the curvature of world lines be less than the critical curvature $3m_I/2e_I^2$.

A similar procedure can be readily developed for any linear theory to convert it to a theory of direct interparticle action (see Problem 10.6.2).

Finally, we compare the symmetry properties of the Maxwell–Lorentz and Wheeler–Feynman theories. It is clear that both are manifestly Poincaré and reparametrization invariant. Of course, action-at-a-distance theories are not gauge invariant because they lack gauge fields.

The field sector of the Maxwell–Lorentz theory is conformally invariant, whereas the second term of the Fokker action (10.218) does not possess this symmetry. Recall that this term owes its origin to eliminating field degrees of freedom. This procedure requires gauge fixing. If we impose the Lorenz condition (which is the only linear Lorentz invariant gauge condition), then conformal invariance is lost. However, one can modify this term by replacing the Minkowski metric $\eta^{\mu\nu}$ with a conformal metric $h^{\mu\nu}(z_I - z_J)$ such that the index μ transforms like a vector at the point z_I while the index ν transforms like a vector at the point z_J. This restores conformal invariance (Problem 10.6.3).

But the most striking issue is time reversal. Rearranging degrees of freedom in the Maxwell–Lorentz theory leads to the Lorentz–Dirac equation which is not invariant under the discrete operation $s \to -s$. This suggests that the rearranged dynamics as a whole is irreversible. By contrast, the equations of motion (10.233) are reversible, because the adjunct field combination $\frac{1}{2}\left(F_{\rm ret} + F_{\rm adv}\right)$ is symmetric in time. If we adopt the Wheeler–Feynman condition (10.236), then (10.233) becomes (10.238), and, therefore, the action-at-a-distance theory turns out to be irreversible.

Problem 10.6.1. Let the electromagnetic field in (10.211) be represented as the Liénard–Wiechert solution. Prove that divergent terms can be assembled into (10.215).

Problem 10.6.2. Consider a system of particles interacting with a neutral scalar field whose action is given by (5.72). Formulate the corresponding action-at-a-distance theory.

Answer

$$S = -\sum_I \int d\tau_I \left[m_I \sqrt{\dot z_I^2} + \frac{1}{2} \int d\tau_J \sum_{J(\neq I)} g_I\, g_J \sqrt{\dot z_I^2}\,\sqrt{\dot z_J^2}\, G_P(z_I - z_J) \right] . \tag{10.240}$$

where G_P is the symmetric Green's function for the scalar field: $G_P = \frac{1}{2}\left(G_{\rm ret} + G_{\rm adv}\right)$,

$$\left(\Box + \mu^2\right) G_P(x) = \delta^4(x) \,. \tag{10.241}$$

Problem 10.6.3. Show that

$$dz_I^\mu \, dz_J^\nu \, h_{\mu\nu}(z_I - z_J) \, \delta\big[(z_I - z_J)^2\big] \ , \tag{10.242}$$

where $h_{\mu\nu}(x-y)$ is the Boulware–Brown–Peccei conformal metric tensor defined in (5.213), behaves like a scalar under conformal transformations.

Hint By (D.9) and (F.15),

$$\delta\big[(x'-y')^2\big] = \delta\big[\sigma^{-1}(x)\,\sigma^{-1}(y)\,(x-y)^2\big] = |\sigma(x)|\,|\sigma(y)|\,\delta\big[(x-y)^2\big] \ . \tag{10.243}$$

Combining this with (5.214) and the transformation law for vectors $dz'^\mu = (\partial z'^\mu/\partial z^\alpha)\,dz^\alpha$, we come to the desired result.

Notes

1. *Section 10.1.* Since the inception of Newtonian mechanics in the early 18th century, it became conventional to formulate evolutionary laws for physical systems in terms of second order differential equations. Ostrogradskiĭ (1850) investigated systems governed by higher-order differential equations, and developed higher derivative Lagrangian and Hamiltonian descriptions for such systems. The peculiar properties of higher derivative Lagrangians have been studied by many physicists over the years. Noteworthy papers on the subject are by Pais & Uhlenbeck (1950) and by Stelle (1977, 1978). A recent review for non-specialists has been made by Woodard (2006). For an annotated lists of studies on the rigid-particle dynamics see Pavšič & Tapia (2001).

2. *Section 10.2.* Huygens proposed a general principle which describes wave propagation as the interference of secondary wavelets arising from imaginary point sources on the existing wave front. A rigorous treatment of this principle employs the retarded Green's functions for the wave operator. Those who wish to read more widely on this subject may consult Baker & Copson (1939), Morse & Feshbach (1953), Courant (1962), and Iwanenko & Sokolow (1953).

Two-dimensional spacetimes have many unusual features in addition to those discussed in this section. One example is solitons. A solitary wave propagating along a straight line cannot be destroyed even when colliding with another such wave. Indestructible solitons are peculiar to two-dimensional partial differential equations such as the sine–Gordon equation. They do not exist in four dimensions. Solitons are discussed at length by Whitham (1974), and Ablowitz & Segur (1981).

The line of presentation in this section follows Kosyakov (1999, 2001). For more on gauge theories with the Chern–Simons term see Deser, Jackiw & Templeton (1982, 1985).

3. *Section 10.3.* Weyl (1918) noted that Maxwell's equations are conformally invariant only for $D = 3$. Ehrenfest (1917), (1920) raised the question: what

is the role of the fact that space has dimension $D = 3$ in fundamental physical laws?

4. *Section 10.4.* The history of nonlinear electrodynamics began with Mie (1912) works. The next steps in this direction was made by Born (1934), and Born & Infeld (1934). Reasoning from the 'principle of finiteness', and the (vain) attempt at constructing a genuine and unique diffeomorphism invariant theory of gravitation and electromagnetism, they proposed the Lagrangians (10.104) and (10.108). After a lapse of fifty years, Fradkin & Tseytlin (1985) vindicated the Born–Infeld action, which naturally arises as the low energy effective action of gauge fields on open strings. The literature on implications of the Born–Infeld theory and its supersymmetric extensions in brane-world scenarios and string theory is extensive; for a review see Gibbons (2003).

Blokhintsev & Orlov (1953) showed that the nonlinear system of hyperbolic equations in the Born–Infeld theory is unique in the sense that their characteristics cannot intersect, and hence no electromagnetic shock wave occurs. Boillat (1970) gave a general argument that this theory is the only version of nonlinear electrodynamics with a sensible weak field limit which is free of the birefringence (that is, signals propagate along a single characteristic cone, regardless of their polarization).

Weyl (1918) pointed out that, among all nonlinear generalizations of electrodynamics with a reasonable weak field limit, Maxwell's theory stands out as the only conformal invariant theory.

Gaillard & Zumino (1981) studied general conditions for electrodynamics to be duality invariant, and came to the criterion (10.133).

5. *Section 10.5.* Wataghin (1934) introduced a form factor in electrodynamics to get rid of ultraviolet divergences. Markov (1939) conjectured that quantum fields do not commute with spacetime variables,

$$[x^\mu, \phi(x)] \neq 0 \,. \tag{10.244}$$

Snyder (1947) considered a quantized spacetime with noncommuting coordinates,

$$[x^\mu, x^\nu] \neq 0 \,. \tag{10.245}$$

Wheeler (1957) conjectured that fluctuations of the metric in regions of size comparable with the Planck length $l_{\rm P}$ give rise to a spacetime foam. This topic is covered in Misner et al. (1973). Witten (1986), and Seiberg & Witten (1999) incorporated methods of noncommutative geometry in string theory. Madore (1995) is a review of subsequent results. Spacetimes with a p-adic order of events is discussed in the book by Vladimirov et al. (1994).

The boundary between localizable and nonlocal quantum field theories was discovered by Meiman (1964), Jaffe (1967), and Efimov (1968). A nonlocal S matrix theory, free of ultraviolet divergences, obeying the general conditions of quantum field theory: unitarity, covariance, and macroscopic causality was developed by Efimov (1970). Iofa & Fainberg (1969) extended some results of the axiomatic quantum field theory (PCT-invariance and connection between spin and statistics) to the case that vacuum expectation values reveal exponential energy growth, the feature characteristic of nonlocal interactions.

A similar technique is applied to the confinement problem by Efimov & Ivanov (1993). The absence of isolated quarks translates into the requirement that the field equation for a free quark

$$L(\partial)\psi(x) = 0 \tag{10.246}$$

has the unique solution $\psi(x) = 0$. It follows that $L(k)$ has no roots, and hence $L^{-1}(k)$ is an entire function, such as

$$L^{-1}(k) = Ce^{a(\gamma\cdot k - bk^2)} , \tag{10.247}$$

where C, a, and b are some parameters, and γ_μ are the Dirac matrices.

The presentation of this section sketches the broad outline of Kosyakov (1976). For more on entire functions see Titchmarsh (1932), and Paley & Wiener (1934). The theory of generalized functions is covered in Gel'fand & Shilov (1964, 1968).

6. *Section 10.6.* The notion of contact interaction was a prevailing view in natural philosophy until it gave way to instantaneous action-at-a-distance dynamics, originated from Newtonian gravitation theory. The theories of electricity and magnetism, arising over the 18th and the first part of the 19th centuries, were patterned after Newton's inverse square law. Faraday and Maxwell overturned this paradigm. They developed a consistent mechanism of local interactions, by which the electromagnetic field is an agent conveying an action-by-contact from one place to another. For a historical account see Whittaker (1910).

Schwarzschild (1903) used the retarded Liénard–Wiechert solution to derive a force law between two charged particles expressed in terms of particle variables (that is, with no reference to field degrees of freedom). This result was rederived by Tetrode (1922) and Fokker (1929*a*, 1929*b*), and summarized in Fokker's action (10.218). Wheeler & Feynman (1945),(1949) critically reexamined the concept of the electromagnetic field as a dynamical entity, and proposed a light-cone action-at-a-distance theory alternative to the conventional Maxwell–Lorentz electrodynamics. For an exposition of those and later developments see Pegg (1975), and Hoyle & Narlikar (1995, 1996).

Mathematical Appendices

A. Differential Forms

Differential forms are a very powerful tool for formulating physical laws on manifolds. These mathematical objects (which in fact are antisymmetric tensor fields on manifolds) have an intrinsic geometric meaning. Although coordinates may appear at intermediate stages of calculations, the use of exterior algebra can greatly reduce the length of calculations and make the final results coordinate-free.

Élie Cartan proposed to use differential coordinates dx^i as a convenient basis of 1-forms. The differentials dx^i transform like covectors under a local coordinate change,

$$dx'^j = \frac{\partial x'^j}{\partial x^i}\, dx^i \; . \tag{A.1}$$

[If the coordinate change is specialized to Euclidean transformations $x'^j = L^j{}_i\, x^i + c^j$, then $\partial x'^j / \partial x^i$ reduces to $L^j{}_i$, an orthogonal matrix with constant entries, and (A.1) becomes (1.53), the transformation law for covectors.] Furthermore, when used in the directional derivative

$$dx^i \, \frac{\partial F}{\partial x^i} \; , \tag{A.2}$$

dx^i may be viewed as a linear functional which takes real values on vectors $\partial F / \partial x^i$. The line elements dx^i are called Cartan's *differential* 1-*forms*, or simply 1-forms.

We now define the exterior product of two 1-forms dx and dy:

$$dx \wedge dy = \frac{1}{2}\,(dx \otimes dy - dy \otimes dx) = -dy \wedge dx \; . \tag{A.3}$$

The exterior product is the simplest rule for constructing 2-forms from pairs of 1-forms. In addition, we define the exterior product of three 1-forms,

$dx \wedge dy \wedge dz$, and, generally, the exterior product of p 1-forms, $dx^1 \wedge \cdots \wedge dx^p$, using the formula

$$dx^1 \wedge \cdots \wedge dx^p = \frac{1}{p!} \sum_{\sigma \in S_p} \operatorname{sgn}(\sigma)\, dx^{i_{\sigma(1)}} \cdots dx^{i_{\sigma(p)}} . \tag{A.4}$$

Here, S_p is the set of all permutations σ of $0, 1, \ldots, p-1$, and $\operatorname{sgn}(\sigma)$ is 1 if the permutation σ is even, and -1 if the permutation σ is odd. Clearly one gets zero for $p > n$.

One can show (Problem A.1) that the set of all

$$dx^{i_1} \wedge \cdots \wedge dx^{i_p}, \quad 1 \le i_1 < i_2 < \cdots < i_p \le n \tag{A.5}$$

is a basis of the vector space of all p-forms Λ^p, and that this space has dimension

$$\binom{n}{p} = \frac{n!}{p!\,(n-p)!} . \tag{A.6}$$

Therefore, any element of Λ^p can be represented as

$$\alpha = \sum_{1 \le i_1 < i_2 < \cdots < i_p \le n} \alpha_{i_1 \cdots i_p}(x)\, dx^{i_1} \wedge \cdots \wedge dx^{i_p} , \tag{A.7}$$

or more commonly, as

$$\alpha = \frac{1}{p!} \alpha_{i_1 \cdots i_p}\, dx^{i_1} \wedge \cdots \wedge dx^{i_p} , \tag{A.8}$$

where every repeated index ranges over the dimension of this manifold.

Note that for arbitrary p- and q-forms α and β,

$$\alpha \wedge \beta = (-1)^{pq} \beta \wedge \alpha . \tag{A.9}$$

We next define the *exterior differentiation* d as an operation which takes p-forms into $(p+1)$-forms so that the following rules hold:

1°. The action of d on a 0-form f (an ordinary function) gives

$$df = \frac{\partial f}{\partial x^i}\, dx^i . \tag{A.10}$$

2°. Let α and β be two p-forms, and a and b real numbers, then

$$d\,(a\alpha + b\beta) = a\, d\alpha + b\, d\beta . \tag{A.11}$$

3°. Let α be a p-form, and β a q-form, then

$$d\,(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^p\, \alpha \wedge d\beta . \tag{A.12}$$

4°. For any p-form α

$$d\,(d\alpha) = 0\ . \tag{A.13}$$

Consider the action of d on a general p-form α. By $(A.10)$, $(A.11)$, and $(A.13)$,

$$d\alpha = \sum d\alpha_{i_1 \cdots i_p} \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_p} = \sum \partial_{[j}\alpha_{i_1 \cdots i_p]}\, dx^j \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_p} \tag{A.14}$$

[summation as in $(A.7)$]. We place the new 1-form dx^j to the left of the previously existing exterior products. Note also that only the totally antisymmetric parts of the partial derivatives

$$\partial_{[i_1}\alpha_{i_2 \cdots i_{p+1}]} = \frac{1}{(p+1)!} \sum_{\sigma \in S_p} \mathrm{sgn}(\sigma)\, \partial_{i_{\sigma(1)}} \alpha_{i_{\sigma(2)} \cdots i_{\sigma(p+1)}} \tag{A.15}$$

contribute to $(A.14)$.

To illustrate, we perform the exterior differentiation of the 1-form on $\mathbb{R}_3$

$$\omega = A_1 dx^1 + A_2 dx^2 + A_3 dx^3\ . \tag{A.16}$$

A simple calculation shows that

$$d\omega = \left(\frac{\partial A_3}{\partial x^2} - \frac{\partial A_2}{\partial x^3} \right) dx^2 \wedge dx^3 + \left(\frac{\partial A_1}{\partial x^3} - \frac{\partial A_3}{\partial x^1} \right) dx^3 \wedge dx^1 + \left(\frac{\partial A_2}{\partial x^1} - \frac{\partial A_1}{\partial x^2} \right) dx^1 \wedge dx^2\ . \tag{A.17}$$

We should check that $(A.14)$ has the same form in every coordinate system. To this end, we write the inverse of $(A.1)$,

$$dx^i = \frac{\partial x^i}{\partial x'^j}\, dx'^j\ , \tag{A.18}$$

and note that the transformation law for $\alpha_{i_1 \cdots i_p}$ is given by

$$\alpha'_{j_1 \cdots j_p} = \frac{\partial x^{i_1}}{\partial x'^{j_1}} \cdots \frac{\partial x^{i_p}}{\partial x'^{j_p}}\, \alpha_{i_1 \cdots i_p}\ . \tag{A.19}$$

Using $(A.18)$ and $(A.19)$, we have

$$\begin{aligned} d\Big(\alpha'_{j_1 \cdots j_p} \wedge dx'^{j_1} \wedge \cdots \wedge dx'^{j_p}\Big) &= d\left(\frac{\partial x^{i_1}}{\partial x'^{j_1}} \cdots \frac{\partial x^{i_p}}{\partial x'^{j_p}}\, \alpha_{i_1 \cdots i_p} dx'^{j_1} \wedge \cdots \wedge dx'^{j_p} \right) \\ &= \frac{\partial x^{i_1}}{\partial x'^{j_1}} \cdots \frac{\partial x^{i_p}}{\partial x'^{j_p}}\, d\alpha_{i_1 \cdots i_p} \wedge dx'^{j_1} \wedge \cdots \wedge dx'^{j_p} = d\alpha_{i_1 \cdots i_p} \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_p}\ . \end{aligned} \tag{A.20}$$

The third equation in $(A.20)$ is obtained by observing that terms involving $\partial^2 x^i / \partial x'^j \partial x'^k$ vanish. Indeed, these partial derivatives are symmetric in j and

k whereas dx'^j and dx'^k in the exterior product are antisymmetric in these indices. Therefore,

$$d\left(\alpha'_{j_1\cdots j_p}\, dx'^{j_1} \wedge \cdots \wedge dx'^{j_p}\right) = d\left(\alpha_{i_1\cdots i_p}\, dx^{i_1} \wedge \cdots \wedge dx^{i_p}\right) \ . \qquad (A.21)$$

It will be left to Problem A.3 to show that $(A.14)$ is consistent with $(A.12)$ and $(A.13)$. These results show that the exterior differentiation d is a well-defined operation for p-forms. Condition 3° is a generalization of the usual Leibnitz rule for differentiating the product of two functions. Condition 4° is often referred to as the Poincaré lemma.

A p-form α is called *exact* if there exists a $(p-1)$-form β such that α is the exterior derivative of β: $\alpha = d\beta$. A p-form α is called *closed* if $d\alpha = 0$. According to the Poincaré lemma, every exact form is closed, $dd = 0$, but the converse is not always true. We will see below that a closed form is exact in some regions of the manifold. In general, by counting the types of closed forms which are not exact one can determine the number of 'holes' of the manifold. To make this idea more definite, the full-fledged machinery of de Rham cohomology is required. However, this mathematical development is beyond the scope of the present discussion.

Two p-forms are of basic importance in electrodynamics and Yang–Mills theory: the 1-form A (vector potential), and the 2-form F (field strength). They are related by

$$F = dA \qquad (A.22)$$

in electrodynamics, and by

$$F = dA + A \wedge A \qquad (A.23)$$

in Yang–Mills theory.

One may wonder whether it is possible to solve

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \qquad (A.24)$$

with respect to A_μ. One would write this symbolically as

$$A = d^{-1} F \ . \qquad (A.25)$$

To make this construction somewhat more concrete, we first impose a gauge condition on A_μ. For our purposes it is appropriate to employ the Fock–Schwinger gauge condition

$$x^\lambda A_\lambda = 0 \ . \qquad (A.26)$$

We differentiate $(A.26)$ to give

$$\partial_\mu (x^\lambda A_\lambda) = A_\mu + x^\lambda \partial_\mu A_\lambda = 0 \ . \qquad (A.27)$$

By $(A.24)$,

$$x^\lambda \partial_\lambda A_\mu - x^\lambda \partial_\mu A_\lambda = x^\lambda F_{\lambda\mu} \ . \qquad (A.28)$$

Combining ($A.27$) and ($A.28$), we obtain

$$A_\mu + x^\lambda \partial_\lambda A_\mu = x^\lambda F_{\lambda\mu} \ . \tag{A.29}$$

We then put $x^\mu = ty^\mu$, where t is a parameter running from 0 to 1. The left-hand side of ($A.29$) becomes

$$A_\mu(ty) + ty^\lambda \left(\frac{\partial A_\mu}{\partial x^\lambda} \right) \Bigg|_{x=ty} = \frac{d}{dt} \left[t A_\mu(ty) \right] \ . \tag{A.30}$$

Hence, integration of ($A.29$) gives

$$A_\mu(y) = \int_0^1 dt \, ty^\lambda F_{\lambda\mu}(ty) \ . \tag{A.31}$$

This is the desired expression of A in terms of F. From this discussion it follows that ($A.31$) is valid in a *star-shaped* region about the origin $y = 0$. By definition, a star-shaped region is an open set $\mathcal{U}$ of points y such that a ray drawn from the origin to the point y is contained in $\mathcal{U}$ whenever y belongs to $\mathcal{U}$.

An alternative formulation of this result is: if a 2-form F is closed in a star-shaped regions, then F is exact, $F = dA$. ($A.31$) expresses A in terms of F, up to a gauge transformation.

Equation ($A.31$) can be readily generalized to express a p-form $A_{\mu_1 \cdots \mu_p}$ in terms of its exterior derivative $F_{\lambda\mu_1 \cdots \mu_p}$ for arbitrary p (Problem A.4).

Another explicit form of d^{-1} is derived in Problem 4.2.4. Note, however, that ($A.31$) has a wider use, because it is also valid in the non-Abelian case. To be explicit, if ($A.22$) is replaced by ($A.23$), while ($A.26$) is preserved, then ($A.31$) still holds.

Problem A.1. Show that the set of p-forms defined in ($A.5$) span a basis of Λ^p. Verify that Λ^p has dimension $n!/p!\,(n-p)!$

Problem A.2. Verify ($A.9$).

Problem A.3. Show that ($A.14$) is consistent with ($A.12$) and ($A.13$).

Problem A.4. Let $F_{\lambda\mu_1 \cdots \mu_p}$ be an exact $(p+1)$-form, that is,

$$F_{\lambda\mu_1 \cdots \mu_p} = (p+1)! \, \partial_{[\lambda} A_{\mu_1 \cdots \mu_p]} \ . \tag{A.32}$$

Solve this equation with respect to $A_{\mu_1 \cdots \mu_p}$ assuming the generalized Fock–Schwinger gauge condition

$$x^{\mu_1} A_{\mu_1 \cdots \mu_p} = 0 \ . \tag{A.33}$$

Answer

$$A_{\mu_1 \cdots \mu_p}(x) = \int_0^1 dt \, t^p x^\lambda F_{\lambda\mu_1 \cdots \mu_p}(tx) \ . \tag{A.34}$$

B. Lie Groups and Lie Algebras

The reader is assumed to be familiar with the basics of the group theory, in particular with elementary concepts of Lie groups and Lie algebras. We review only those definitions and properties of Lie groups and Lie algebras which are essential for reading the main text. We will henceforth speak only of transformation groups, and will define transformations in terms of $n \times n$ matrices. Furthermore, our prime interest here is with compact groups. A Lie group G is called *compact* if its group manifold is compact. A Lie algebra $\mathfrak{g}$ is said to be compact if its associated Lie group G is compact.

We begin with the *general linear* group $\mathrm{GL}(n, \mathbb{R})$ spanned by $n \times n$ real matrices with nonzero determinant. Here, '$\mathbb{R}$' refers to real numbers. Of course, most of this discussion can be extended to cover the complex numbers $\mathbb{C}$, quaternions $\mathbb{H}$, and octonions $\mathbb{O}$[1].

If all the matrices L in the defining representation of $\mathrm{GL}(n, \mathbb{R})$ obey the constraint

$$\det(L) = 1 \; , \tag{B.1}$$

then we come to the *special linear* group $\mathrm{SL}(n, \mathbb{R})$. For an infinitesimal transformation $L = 1 + \epsilon X$, $(B.1)$ implies

$$\mathrm{tr}\,(X) = 0 \; . \tag{B.2}$$

Thus, the Lie algebra $\mathrm{sl}(n, \mathbb{R})$ associated with the special linear group involves $n^2 - 1$ independent $n \times n$ real matrices. In other words, $\mathrm{SL}(n, \mathbb{R})$ is specified by $n^2 - 1$ independent real parameters. Note that the group manifold of $\mathrm{SL}(n, \mathbb{R})$ is noncompact.

Further constraints define the metric properties of the group. Consider a $n \times n$ real matrix L obeying the condition

$$L^T L = 1 \tag{B.3}$$

where L^T is the transpose of L. Such matrices form the *orthogonal* group $\mathrm{O}(n)$. In addition, if we impose condition $(B.1)$, we get the *special orthogonal* group $\mathrm{SO}(n)$, a subgroup of $\mathrm{O}(n)$ which can be continuously deformed to the identity. As is shown in Sect. 1.2, $\mathrm{SO}(n)$ rotates vectors $\mathbf{a}$ of Euclidean space $\mathbb{E}_n$. That is, a matrix L acts on a vector $\mathbf{a}$ according to the rule

$$\mathbf{a} \to \mathbf{a}' = L\mathbf{a} \tag{B.4}$$

so that the norm of $\mathbf{a}$ is invariant, $\mathbf{a}'^2 = \mathbf{a}^2$. Such n-dimensional Euclidean vectors define the *fundamental* representation of $\mathrm{SO}(n)$.

By $(B.3)$,

[1] Recall, there are exactly four normed division algebras: the real numbers $\mathbb{R}$, complex numbers $\mathbb{C}$, quaternions $\mathbb{H}$, and octonions $\mathbb{O}$. Quaternions are noncommutative. Octonions are nonassociative.

$$(1+\epsilon X)^T (1+\epsilon X) = 1 \; , \tag{B.5}$$

which implies

$$X^T = -X \; . \tag{B.6}$$

Thus, the $\frac{1}{2}\, n(n-1)$ generators of orthogonal transformations X are $n \times n$ antisymmetric real matrices. The group manifold of $\mathrm{SO}(n)$ is a unit $(n-1)$-dimensional sphere in $\mathbb{E}_n$ [for example, the group manifold of $\mathrm{SO}(2)$ is a unit circle]. Therefore, $\mathrm{SO}(n)$ is compact. Note also that $\mathrm{SO}(n)$ is non-Abelian for $n \geq 3$.

As an illustration, we refer to the following generators of $\mathrm{SO}(3)$

$$X_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} , \quad X_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} , \quad X_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{B.7}$$

related to rotations around the Cartesian axes x^1, x^2, and x^3. $\mathrm{SO}(3)$ is specified by three independent real parameters.

Equation $(B.3)$ is equivalent to $L^T g L = g$, where g is the Euclidean metric which is proportional to the unit matrix. If this condition is replaced by

$$L^T \eta L = \eta \; , \tag{B.8}$$

where η is the diagonal matrix $\mathrm{diag}\,(1,\ldots,1,-1,\ldots,-1)$, having p positive and q negative entries, then the transformations L constitute the *pseudo-orthogonal* group $\mathrm{SO}(p,q)$. This group rotates vectors of pseudoeuclidean space with metric proportional to η.

We next consider the set of all $n \times n$ complex matrices obeying the condition

$$L^\dagger L = 1 \; , \tag{B.9}$$

where $L^\dagger = (L^T)^*$ is the Hermitian conjugate of L. Such matrices form the *unitary* group $\mathrm{U}(n)$. For example, $\mathrm{U}(1)$ is the set of all complex numbers with unit modulus $e^{i\lambda}$ whose group composition is defined by the usual complex number multiplication. If $\det(L) = 1$, then we come to the *special unitary* group $\mathrm{SU}(n)$. One may think of these matrices L as linear operators acting on n-dimensional column vectors Ψ, which gives vectors Ψ' in the same space: $\Psi' = L\Psi$. Such vectors define the fundamental representation of $\mathrm{SU}(n)$. If we define the inner product (Φ,Ψ) as the component sum $\sum \Phi^*{}_i \Psi_i$, then, in view of $(B.9)$, this quantity is unchanged under $\mathrm{SU}(n)$ transformations:

$$(\Phi',\Psi') = (L\Phi, L\Psi) = (\Phi, L^\dagger L\Psi) = (\Phi,\Psi) \; . \tag{B.10}$$

For an infinitesimal unitary transformation $L = 1 + i\epsilon X$, $(B.9)$ gives

$$X = X^\dagger \; . \tag{B.11}$$

One may readily check that generators X of $\mathrm{U}(n)$ are n^2 Hermitian $n \times n$ matrices. With the exception of the $\mathrm{U}(1)$ which is an Abelian group, $\mathrm{U}(n)$ is a compact non-Abelian group specified by n^2 independent real parameters. $\mathrm{SU}(n)$ is a subgroup of $\mathrm{U}(n)$ which is continuously connected to the identity. The number of generators required to represent it is $n^2 - 1$.

The representation defined by any $n \times n$ matrices X which serves as a basis for the Lie algebra $\mathfrak{g}$ is known as the *adjoint* representation of the group G. In particular, if a traceless Hermitian $n \times n$ matrix Φ transforms according to the rule

$$\Phi \to \Phi' = L \Phi L^\dagger , \tag{B.12}$$

where $L \in \mathrm{SU}(n)$, then Φ is said to be in the adjoint representation of $\mathrm{SU}(n)$.

If a Lie algebra $\mathfrak{g}$ is spanned by D independent elements, then this Lie algebra is of order D. The number of commuting elements of $\mathfrak{g}$ is called the *rank* of this Lie algebra. We can choose a basis in which these commuting elements are diagonal matrices H_i. The set of all H_i spans the maximal Abelian subalgebra, which is known as the *Cartan subalgebra.*

Let us consider two important examples:

su(2) is a Lie algebra of the order 3 and rank 1. The Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{B.13}$$

span the standard basis of su(2), $X_a = \frac{1}{2}\, \sigma_a$. These traceless Hermitian 2×2 matrices are orthonormalized

$$\mathrm{tr}\,(\sigma_a \sigma_b) = 2\delta_{ab} \ , \tag{B.14}$$

and satisfy the following commutation relations

$$\left[\frac{\sigma_a}{2} , \frac{\sigma_b}{2} \right] = i \epsilon_{abc} \, \frac{\sigma^c}{2} \ . \tag{B.15}$$

The only diagonal matrix is σ_3. Elements of SU(2) are given by $\exp\left(\frac{1}{2}\, i \omega^a \sigma_a\right)$. They act on two-dimensional column vectors ψ, which are called *spinors.*

su(3) is a Lie algebra of the order 8 and rank 2. A conventional basis of su(3) (proposed by Gell-Mann) is spanned by eight Hermitian traceless 3×3 matrices of which two are diagonal

$$\lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} , \tag{B.16}$$

and the other are written as

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} , \tag{B.17}$$

$$\lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} . \qquad (B.18)$$

The Gell-Mann matrices obey the orthonormalization condition

$$\mathrm{tr}\,(\lambda_a \lambda_b) = 2\delta_{ab} \ , \qquad (B.19)$$

and the commutation relations

$$\left[\frac{\lambda_a}{2}, \frac{\lambda_b}{2}\right] = i f_{abc}\, \frac{\lambda_c}{2} \ . \qquad (B.20)$$

The total antisymmetric objects f_{abc} are known as the *structure constants* of SU(3). The nonzero components are:

$$f_{123} = 2f_{147} = -2f_{156} = 2f_{246} = 2f_{257} = 2f_{345} = -2f_{367} = \frac{2f_{458}}{\sqrt{3}} = \frac{2f_{678}}{\sqrt{3}} = 1 \ . \qquad (B.21)$$

Elements of SU(3) are given by $\exp\left(\frac{1}{2}\, i\omega^a \lambda_a\right)$.

In general, a finite element of SU(n) can be represented by

$$L = \exp\left(i\alpha^a X_a\right) \ , \qquad (B.22)$$

where X_a are traceless Hermitian $n \times n$ matrices. There is a rigorous statement that any representation of a *compact* Lie group is equivalent to some finite-dimensional unitary representation. In support of this statement let us construct a unitary representation of SO(3). We first express a vector $\mathbf{a} \in \mathbb{E}_3$ in terms of Pauli matrices:

$$\mathbf{a} = a^1\sigma_1 + a^2\sigma_2 + a^3\sigma_3 = \begin{pmatrix} a^3 & a^1 - ia^2 \\ a^1 + ia^2 & -a^3 \end{pmatrix} . \qquad (B.23)$$

We define the norm of $\mathbf{a}$:

$$-\det\left(\mathbf{a}^2\right) = (a^1)^2 + (a^2)^2 + (a^3)^2. \qquad (B.24)$$

A transformation

$$\mathbf{a}' = L\mathbf{a}L^\dagger \ , \qquad (B.25)$$

where $L \in$ SU(2), leaves this norm unchanged, $\det\left(L\mathbf{a}L^\dagger L\mathbf{a}L^\dagger\right) = \det\left(\mathbf{a}^2\right)$. Furthermore, $\mathbf{a}'$ is again a traceless Hermitian matrix. Therefore, $\mathbf{a}'$ and $\mathbf{a}$ are related by a rotation. We thus see that (B.23) furnishes a unitary representation of SO(3). More precisely, (B.23) is a mapping of the fundamental representation of SO(3) onto the adjoint representation of SU(2). Note that L and $-L$ in (B.25) are associated with the same SO(3) transformation of $\mathbf{a}$. Another way of saying this is that SU(2) is the double covering of SO(3).

In contrast, *noncompact* Lie groups have no finite-dimensional unitary representations. If a noncompact group G describes a symmetry of some physical system, then we can proceed in either of two ways. First, we might use finite-dimensional nonunitary representations of G. For example, every quantity (such as those introduced in Chaps. 1–6) which transforms like a tensor under Lorentz transformations falls into this category. Second, we might employ infinite-dimensional unitary representations of G. This approach is taken in Sect. 8.7.

Let H be a subgroup of a Lie group G. For any $L \in G$ we can define another subgroup $H' = LHL^{-1}$ which is called conjugate to H. Imagine that all the conjugate subgroups are identical: that is, $H = LHL^{-1}$ for any $L \in G$. Then H is said to be a normal subgroup. A compact group which contains no normal subgroups is called *simple*. A compact group which contains no normal Abelian subgroups is called *semisimple*.

Let us choose some basis of a Lie algebra $\mathfrak{g}$. Elements of this basis X_a obey the commutation relations

$$[X_a, X_b] = if^c{}_{ab} X_c \ . \tag{B.26}$$

Whatever the representation of the X_a, the structure constants $f^c{}_{ab}$ are the same. The structure constants determine $\mathfrak{g}$ completely.

Making the identification

$$(X^a)_{bc} = if^a{}_{bc} \ , \tag{B.27}$$

the structure constants $f^c{}_{ab}$ themselves form the adjoint representation of $\mathfrak{g}$. In particular, substituting $(B.27)$ into the Jacobi identity

$$[X_a, [X_b, X_c]] + [X_c, [X_a, X_b]] + [X_b, [X_c, X_a]] = 0 \tag{B.28}$$

gives

$$f^d{}_{bc} f^e{}_{ad} + f^d{}_{ab} f^e{}_{cd} + f^d{}_{ca} f^e{}_{bd} = 0 \ , \tag{B.29}$$

which can be shown (Problem B.6) to be just the commutation relations $(B.26)$.

An important characteristic of Lie groups is the Killing form. From $f^c{}_{ab}$, one constructs a symmetric tensor

$$g_{ab} = -f^c{}_{ad} f^d{}_{bc} \tag{B.30}$$

which is called the *Killing form*. This quantity plays the role of the *metric* on the group manifold. When viewing X_a as a $n \times n$ matrix, the Killing form can be defined by

$$g_{ab} = \mathrm{tr}\,(X_a X_b) \ . \tag{B.31}$$

This definition makes it clear that the metric is invariant under the group. That is, if $X'_a = LX_aL^{-1}$ then $g'_{ab} = g_{ab}$.

One can show that a Lie algebra $\mathfrak{g}$ is *semisimple* if $\det(g_{ab}) \neq 0$. Thus, the inverse metric g^{ab} can be defined (by $g_{ab}\, g^{bc} = \delta^c{}_a$) for any semisimple Lie algebra. Using g_{ab} and g^{ab}, it is possible to raise and lower group indices, in particular to recast the structure constants in the form f_{abc}. A further important statement is that if the metric g_{ab} is *positive definite*, for example, if

$$\mathrm{tr}\,(X_a X_b) = \frac{1}{2}\,\delta_{ab}\ , \qquad (B.32)$$

then the Lie algebra is *compact.* If g_{ab} is *indefinite*, then the Lie algebra is *noncompact.*

The structure constants depend upon the choice of the basis. For simple compact groups, there is a convenient basis, called the *Cartan basis*, such that the structure constants f_{abc} are *real* and *completely antisymmetric.* It will be left to Problem B.7 to show that if the generators X_a satisfy the orthonormalization condition (B.32), then the structure constants f_{abc} are completely antisymmetric.

Killing, Cartan, and Weyl classified all simple compact Lie groups. The great bulk of them belong to the so-called *classical* Lie groups. In addition to SU(n) and SO(n), the classical groups involve the *symplectic* group Sp(n). This group is defined as the unitary group of $n \times n$ matrices over the quaternions. The remaining five simple compact Lie groups are known as *exceptional.* These groups, denoted by G_2, F_4, E_6, E_7, and E_8, can be interpreted as unitary groups over the octonions.

U(1) is locally equivalent to SO(2) because the set of complex numbers of the form $e^{i\lambda}$ can be mapped onto a unit circle. We have already learned that su(2) $\sim$ so(3). In addition, SU(2) is locally isomorphic to Sp(1). To see this, we note that Sp(1) is the set of all quaternions with unit modulus whose group composition is defined by the rule of quaternion multiplication. Recall that the Pauli matrices, which generate the fundamental representation of SU(2), also happen to obey the quaternion algebra. Finally note that SO(4) is not semisimple because so(4) = so(3) $\oplus$ so(3) (see Problem 1.5.7).

Let us give an explicit construction of the *Cartan-Weyl basis*[2] for the Lie algebra su(n). This basis consists of a set of n^2 matrices, including n elements H_a of the Cartan subalgebra

$$[H_a, H_b] = 0\ , \qquad (B.33)$$

which are related by

$$\sum_{a=1}^{n} H_a = 0\ . \qquad (B.34)$$

We choose the commuting elements H_a to be traceless, diagonal $n \times n$ matrices

$$(H_a)_{AB} = \delta_{Aa}\,\delta_{Ba} - \frac{1}{n}\,\delta_{AB} \quad \text{(no summation in } a\text{)}\ . \qquad (B.35)$$

The remaining $n^2 - n$ basis elements consist of the 'raising' and 'lowering' operators E^+_{ab} and E^-_{ab},

[2] This overcomplete basis slightly differs from that commonly used in the mathematical literature.

$$(E^{+}_{ab})_{AB} = \delta_{Aa}\,\delta_{Bb}, \quad (E^{-}_{ab})_{AB} = \delta_{Ab}\,\delta_{Ba}\ , \tag{B.36}$$

where a, b, A, B run from 1 to n, with a and b being ordered such that $b > a$. The only nontrivial commutation relations are

$$[H_a, E^{\pm}_{ab}] = \pm\, E^{\pm}_{ab}\ , \tag{B.37}$$

$$[E^{+}_{ab}, E^{-}_{ab}] = H_a - H_b\ , \tag{B.38}$$

$$[E^{\pm}_{ab},\, E^{\pm}_{bc}] = \pm\, E^{\pm}_{ac}\ . \tag{B.39}$$

This basis is extensively used in Chap. 8.

A similar realization of the Cartan-Weyl basis can be proposed for SO(n):

$$(X_{ab})^{AB} = \delta^{A}_{\ [a}\,\delta^{B}_{\ \ b]}\ , \tag{B.40}$$

and for Sp(n):

$$(X_{ab})^{AB} = \delta^{A}_{\ (a}\,\delta^{B}_{\ \ b)}\ . \tag{B.41}$$

Of special interest for our discussion is the fact that every classical group can be *complexified* by allowing the group parameters to be complex. To be specific, let us complexify SU(n), a classical group which is our paradigm. The result is SL($n, \mathbb{C}$). In fact, we need only *real forms* of SL($n, \mathbb{C}$). There is a single *compact* real form of this group, SU(n). All other real forms are *noncompact*. Among them, SL($n, \mathbb{R}$) is of great interest for our analysis of retarded solutions to the Yang–Mills equations in Chap. 8.

Finally, we review the Lorentz and Poincaré groups. We would like to show that the Lorentz group SO$(1,3)$ is locally equivalent to SL$(2,\mathbb{C})$. The method is essentially the same as that used before to establish the local equivalence SO(3) $\sim$ SU(2). Let us write a vector of Minkowski space x as a Hermitian 2×2 matrix:

$$x = x^0\mathbf{1} + x^1\sigma_1 + x^2\sigma_2 + x^3\sigma_3 = \begin{pmatrix} x^0 + x^3 & x^1 - ix^2 \\ x^1 + ix^2 & x^0 - x^3 \end{pmatrix}\ , \tag{B.42}$$

where $\mathbf{1}$ is the unit 2×2 matrix. We define the norm of x as

$$\det\left(x^2\right) = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2. \tag{B.43}$$

A linear transformation

$$x' = LxL^{\dagger}\ , \tag{B.44}$$

where L is an arbitrary complex 2×2 matrix subject to the condition

$$\det\,(L) = 1\ , \tag{B.45}$$

leaves this norm invariant:

$$\det\,(x') = \det\,(x)\ . \tag{B.46}$$

Furthermore, x retains its Hermitian property under transformations (B.44). We thus see that any vector of Minkowski space x is at the same time a rank (1,1) tensor under $\mathrm{SL}(2,\mathbb{C})$. To put it otherwise, the fundamental representation of $\mathrm{SO}(1,3)$ is the adjoint representation of $\mathrm{SL}(2,\mathbb{C})$. Note that $\mathrm{SL}(2,\mathbb{C})$ gives a double covering of $\mathrm{SO}(1,3)$ because L and $-L$ are associated with the same Lorentz transformation.

Poincaré transformations are obtained by combining spacetime translations a^μ and Lorentz transformations $\Lambda^\mu{}_\nu$:

$$x'^\mu = \Lambda^\mu{}_\nu \, x^\nu + a^\mu \; . \tag{B.47}$$

The set of all Poincaré transformations (a, Λ) form a group with the composition law

$$(a_1, \Lambda_1)\,(a_2, \Lambda_2) = (a_1 + \Lambda_1 a_2, \Lambda_1 \Lambda_2) \; . \tag{B.48}$$

Without going into detail, we simply note that, owing to (B.48), the Poincaré group is the *semidirect product* of the Lorentz group and the group of translations. We would have a direct product of these groups if, instead of (B.48), the composition rule was

$$(a_1, \Lambda_1)\,(a_2, \Lambda_2) = (a_1 + a_2, \Lambda_1 \Lambda_2) \; . \tag{B.49}$$

Problem B.1. Show that an antisymmetric $n \times n$ real matrix has $\frac{1}{2}\, n(n-1)$ independent entries.

Problem B.2. Show that a Hermitian traceless $n \times n$ matrix is specified by $n^2 - 1$ independent real parameters.

Problem B.3. Verify (B.14) and (B.15).

Problem B.4. Verify (B.19) and (B.20)–(B.21).

Problem B.5. Verify (B.24).

Problem B.6. Applying (B.27) to (B.28), show that (B.29) is equivalent to (B.26).

Problem B.7. Let elements of some Lie algebra $\mathfrak{g}$ are subject to the orthonormalization condition (B.32). Show that the structure constants f_{abc} are completely antisymmetric.

Hint Verify that $f_{cab} + f_{acb} = 0$ by writing f_{cab} as

$$f_{cab} = -2i \,\mathrm{tr}\,([X_a, X_b] X_c) = -2i \,\mathrm{tr}\,(X_a X_b X_c - X_b X_a X_c) \; . \tag{B.50}$$

C. The Gamma Matrices and Dirac Spinors

Following Dirac's original approach, we take a 'square root' of the Klein–Gordon equation

$$\left(\Box + m^2\right)\Psi = 0 \tag{C.1}$$

to give

$$(i\gamma^\mu\partial_\mu - m)\Psi = 0 \ . \tag{C.2}$$

We come back to (C.1) by applying $(i\gamma^\mu\partial_\mu + m)$ to the Dirac equation (C.2),

$$(i\gamma^\mu\partial_\mu + m)(i\gamma^\nu\partial_\nu - m)\Psi = 0 \ , \tag{C.3}$$

provided that

$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2\eta^{\mu\nu} \ . \tag{C.4}$$

A set of mathematical objects γ^μ satisfying the anticommutation relations (C.4) is called a *Clifford algebra*. To realize a Clifford algebra in Minkowski space $\mathbb{M}_4$, Dirac proposed to use 4×4 matrices with complex entries, now known as the *gamma matrices*.

Note that if a set of γ^μ obeys the anticommutation relations (C.4), then the set of

$$\gamma'^\mu = M\gamma^\mu M^{-1} \tag{C.5}$$

also obey (C.4) for any nonsingular M. Thus, condition (C.4) does not define the γ^μ uniquely. The form of the γ^μ is fixed only up to a similarity transformation (C.5). In fact, most calculations with gamma matrices can be done without referring to a particular representation. The anticommutation relations (C.4) and the requirement that the γ^μ form an irreducible set are sufficient to characterize a particular Clifford algebra.

Consider

$$\Gamma_1 = \mathbf{1} \ , \tag{C.6}$$

$$\Gamma_2 = \gamma^0, \quad \Gamma_3 = i\gamma^1, \quad \Gamma_4 = i\gamma^2, \quad \Gamma_5 = i\gamma^3, \tag{C.7}$$

$$\Gamma_6 = \gamma^0\gamma^1, \ \Gamma_7 = \gamma^0\gamma^2, \ \Gamma_8 = \gamma^0\gamma^3, \ \Gamma_9 = i\gamma^2\gamma^3, \ \Gamma_{10} = i\gamma^3\gamma^1, \ \Gamma_{11} = i\gamma^1\gamma^2 \ , \tag{C.8}$$

$$\Gamma_{12} = \gamma^1\gamma^2\gamma^3, \quad \Gamma_{13} = i\gamma^0\gamma^2\gamma^3, \quad \Gamma_{14} = i\gamma^0\gamma^3\gamma^1, \quad \Gamma_{15} = i\gamma^0\gamma^1\gamma^2, \tag{C.9}$$

$$\Gamma_{16} = \gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3 \ , \tag{C.10}$$

where γ^μ ($\mu = 0, \dots, 3$) are $n \times n$ matrices obeying the anticommutation relations (C.4), and $\mathbf{1}$ is the unit $n \times n$ matrix. With (C.4), all other products of γ^μ can be written as linear combinations of Γ_a. Furthermore, these Γ_a are linearly independent. To see this, we first note (Problem C.1) that the trace of the product of any two Γ_a's obeys, $\mathrm{tr}\,(\Gamma_a\Gamma_b) = n\delta_{ab}$. Let us suppose that the matrices Γ_a defined in (C.6)–(C.10) are linearly dependent,

$$\sum_{a=1}^{16} C_a\Gamma_a = 0 \ . \tag{C.11}$$

Multiplying (C.11) by Γ_b and taking the trace of this product, we obtain

$$\sum_{a=1}^{16} C_a \mathrm{tr}\left(\Gamma_b \Gamma_a\right) = nC_b = 0 \ . \tag{C.12}$$

It follows from $(C.12)$ that $C_b = 0$ for all b. Therefore the Γ_a are linearly independent.

It is impossible to realize the Clifford algebra in $\mathbb{M}_4$ by a set of matrices of order $n < 4$. Indeed, for $n < 4$, there are not 16 linearly independent matrices. On the other hand, four 4×4 matrices γ^μ do provide such a realization and the number of independent components in an arbitrary 4×4 matrix is 16.

Mathematically, Ψ is a *spinor*. If we write Ψ in a four-component column, then the γ^μ act upon Ψ according to the conventional matrix rules.

Pauli proved that if γ^μ and γ'^μ are two irreducible sets of 4×4 matrices obeying $(C.4)$, then there exists a nonsingular matrix M which relates γ^μ and γ'^μ by transformation $(C.5)$, and, furthermore, that M is unique except for an arbitrary multiplicative factor.

Using the Pauli matrices σ_i together with the unit 2×2 matrix 1, we can give an explicit representation of gamma matrices:

$$\gamma_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \gamma_i = \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix}, \quad \gamma_5 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \tag{C.13}$$

which is referred to as the *Dirac–Pauli*, or *standard* basis.

We now turn to the transformation properties of the matrices γ^μ and spinors Ψ under the Poincaré group. Equation $(C.2)$ is self-evidently translation invariant. We thus have to examine the Lorentz covariance of $(C.2)$. This issue can be viewed in either of two alternative ways. First, we may assume that the γ^μ transform like the components of a vector, as the index μ suggests. Then $(C.2)$ and $(C.4)$ are covariant with respect to Lorentz transformations. However, this assumption is unnatural in the context of field theory in which it is only the fields that transform. Second, we may consider the γ^μ as fixed matrices. Then, to maintain the covariance of $(C.2)$, we must attribute some transformation properties to Ψ. Consider a linear transformation law

$$\Psi'(x') = U\left(\Lambda\right)\Psi(x) \ , \tag{C.14}$$

where $U\left(\Lambda\right)$ is a nonsingular matrix which depends on the Lorentz transformation

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu \ . \tag{C.15}$$

If we require that, for any Ψ, the transformed equation

$$\left(i\gamma^\mu \frac{\partial x^\nu}{\partial x'^\mu}\frac{\partial}{\partial x^\nu} - m\right)U\Psi = \left(i\gamma^\mu \left(\Lambda^{-1}\right)^\nu{}_\mu \frac{\partial}{\partial x^\nu} - m\right)U\Psi = 0 \tag{C.16}$$

follows from $(C.2)$, or, equivalently, from

$$U\left(i\gamma^\mu\partial_\mu - m\right)\Psi = 0 \ , \tag{C.17}$$

then we must have

$$U\gamma^\mu = (\Lambda^{-1})^\mu{}_\nu \gamma^\nu U \,. \tag{C.18}$$

Let us write

$$\Lambda_{\mu\nu} = \eta_{\mu\nu} + \omega_{\mu\nu}\,, \quad (\Lambda^{-1})_{\mu\nu} = \eta_{\mu\nu} - \omega_{\mu\nu} \,, \tag{C.19}$$

and

$$U = 1 - \frac{i}{4}\, s_{\mu\nu}\, \omega^{\mu\nu} \,, \tag{C.20}$$

where the $\omega_{\mu\nu}$ are the parameters which characterize an infinitesimal Lorentz transformation ($\omega_{\mu\nu} = -\omega_{\nu\mu}$), and the $s_{\mu\nu}$ are the the Lorentz generators in the spinor representation. Then (C.18) becomes

$$[\gamma_\lambda, s_{\mu\nu}] = 2i\,(\eta_{\lambda\mu}\gamma_\nu - \eta_{\lambda\nu}\gamma_\mu) \,. \tag{C.21}$$

This equation is satisfied by

$$s_{\mu\nu} = \frac{i}{2}\,(\gamma_\mu\,\gamma_\nu - \gamma_\nu\,\gamma_\mu) \,. \tag{C.22}$$

Integrating (C.20), we get

$$U = \exp\left(-\frac{i}{4}\, s_{\mu\nu}\, \omega^{\mu\nu}\right) \,. \tag{C.23}$$

The simplest physical observables associated with the Dirac field are constructed from bilinear expressions of Ψ which transform like irreducible tensor representations of the Lorentz group. From (C.2), we obtain the equation of motion for the Hermitian conjugate of Ψ:

$$\Psi^\dagger \left[i\,(\gamma^\mu)^\dagger\, \overleftarrow{\partial}_\mu - m\right] = 0 \,. \tag{C.24}$$

Using (C.5), we can always express the $(\gamma^\mu)^\dagger$ in terms of the γ^μ. Since $(\gamma^0)^{-1} = \gamma^0$, we may write the relation

$$(\gamma^\mu)^\dagger = \gamma^0\gamma^\mu\gamma^0 \tag{C.25}$$

consistent with (C.4). It follows from (C.20) and (C.25) that

$$U^\dagger = \gamma^0 U^{-1}\gamma^0 \tag{C.26}$$

Consider the field $\overline{\Psi} = \Psi^\dagger\gamma^0$. This field obeys the equation

$$\overline{\Psi}\left(i\gamma^\mu\overleftarrow{\partial}_\mu + m\right) = 0 \,. \tag{C.27}$$

One can show (Problem C.3) that $\overline{\Psi}$ transforms according to the rule

$$\overline{\Psi}'(x') = \overline{\Psi}(x)\,\gamma^0\, U^\dagger\,\gamma^0 = \overline{\Psi}(x)\, U^{-1} \,. \tag{C.28}$$

Therefore,

$$\overline{\Psi}'(x')\, O\, \Psi'(x') = \overline{\Psi}(x)\, U^{-1}\, O\, U\, \Psi(x)\ . \qquad (C.29)$$

One can see [using (C.18)] that the bilinear form $\overline{\Psi}\gamma^\mu\Psi$ transforms like a vector:

$$\overline{\Psi}'(x')\, \gamma^\mu\, \Psi'(x') = \Lambda^\mu{}_\nu\, \overline{\Psi}(x)\, \gamma^\nu\, \Psi(x)\ , \qquad (C.30)$$

and that $\overline{\Psi}\Psi$ is a scalar with respect to Lorentz boosts and space reflections. Further examples of the transformation laws for bilinear forms are given in Problem C.4.

Problem C.1. Consider any two Γ_a's defined in (C.6) – (C.10). Show that

$$\mathrm{tr}\,(\Gamma_a \Gamma_b) = n\delta_{ab} \qquad (C.31)$$

Problem C.2. Derive the identities

$$\gamma^\mu\gamma_\mu = 4,\ \gamma^\lambda\gamma^\mu\gamma_\lambda = -2\gamma^\mu,\ \gamma^\lambda\gamma^\mu\gamma^\nu\gamma_\lambda = 4\eta^{\mu\nu},\ \gamma^\lambda\gamma^\mu\gamma^\nu\gamma^\rho\gamma_\lambda = -2\gamma^\rho\gamma^\nu\gamma^\mu\ . \qquad (C.32)$$

Problem C.3. Verify (C.28).

Hint Use (C.23).

Problem C.4. Show that the bilinear forms $\overline{\Psi}s^{\mu\nu}\Psi$, $\overline{\Psi}\gamma_5\gamma^\mu\Psi$, and $\overline{\Psi}\gamma_5\Psi$ transform, respectively, like an antisymmetric rank $(2,0)$ tensor, an axial vector, and a pseudoscalar:

$$\overline{\Psi}'(x')\, s^{\mu\nu}\, \Psi'(x') = \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta\, \overline{\Psi}(x)\, s^{\alpha\beta}\, \Psi(x)\ , \qquad (C.33)$$

$$\overline{\Psi}'(x')\, \gamma_5\gamma^\mu\, \Psi'(x') = \det(\Lambda)\, \Lambda^\mu{}_\nu\, \overline{\Psi}(x)\, \gamma_5\gamma^\nu\, \Psi(x)\ , \qquad (C.34)$$

$$\overline{\Psi}'(x')\, \gamma_5\, \Psi'(x') = \det(\Lambda)\, \overline{\Psi}(x)\, \gamma_5\, \Psi(x)\ . \qquad (C.35)$$

Hint Derive first the relations

$$\gamma_5 = -\frac{i}{4!}\, \epsilon_{\lambda\mu\nu\rho}\, \gamma^\lambda\gamma^\mu\gamma^\nu\gamma^\rho\ , \qquad (C.36)$$

$$\gamma_5\gamma_\lambda = -\frac{i}{3!}\, \epsilon_{\lambda\mu\nu\rho}\, \gamma^\mu\gamma^\nu\gamma^\rho\ . \qquad (C.37)$$

D. Conformal Transformations

Conformal transformations of Minkowski space form a group, denoted by C$(1,3)$, which consists of the 10-parameter Poincaré subgroup, the 1-parameter subgroup of *dilatations*, or *scale transformations*

$$D:\quad x'_\mu = kx_\mu, \quad k > 0\ , \qquad (D.1)$$

and the 4-parameter subgroup of *special conformal transformations*

$$C: \quad {x'}_\mu = \frac{x_\mu - b_\mu x^2}{1 - 2b\cdot x + b^2x^2} \ . \tag{D.2}$$

The special conformal transformations are closely related to the *inversion*

$$I: \quad {x'}_\mu = \frac{x_\mu}{x^2} \ . \tag{D.3}$$

More particularly, $(D.2)$ is composed of an inversion, translation, and further inversion:

$$x_\mu \to u_\mu = \frac{x_\mu}{x^2}, \quad u_\mu \to v_\mu = u_\mu - b_\mu, \quad v_\mu \to y_\mu = \frac{v_\mu}{v^2} \ . \tag{D.4}$$

Symbolically,

$$C = ITI. \tag{D.5}$$

The inversion is a *discrete* operation, while the special conformal transformations form a *continuous* subgroup continuously connected with the unit $(b^\mu = 0)$. Furthermore, it is an *Abelian* subgroup. Indeed, by $(D.5)$, the product of two subsequent transformations is $C(b_1)C(b_2) = IT(b_1)IIT(b_2)I$. Because the double application of the inversion is identity, and translations are commutative, $C(b_1)C(b_2) = C(b_2)C(b_1)$.

Rewriting $(D.5)$ as $I^{-1}TI$, one finds that the composition of two such mappings is a mapping other than translation. Thus, the translation subgroup is not an invariant subgroup of the conformal group. In fact, $\mathrm{C}(1,3)$ proves to be a *semisimple* Lie group.

Is it is possible to assemble the Poincaré transformations and the dilatations alone into a group? As will soon become clear, this is the case. Such a combination gives the group of *similitude* transformations acting on x^μ linearly,

$${x'}^\mu = k\Lambda^\mu{}_\nu x^\nu + c^\mu \ , \tag{D.6}$$

where Λ^μ_ν is a Lorentz matrix. However, the similitude group is not semisimple because it is the semidirect product of the Lorentz rotations and dilatations with translations.

One can show from $(D.2)$ that

$${x'}^2 = \frac{x^2}{\sigma(x)} \tag{D.7}$$

where

$$\sigma(x) = 1 - 2b\cdot x + b^2x^2 \ , \tag{D.8}$$

and

$$(x' - y')^2 = \frac{(x-y)^2}{\sigma(x)\,\sigma(y)} \ , \tag{D.9}$$

which implies

$$dx'^2 = \frac{dx^2}{\sigma^2(x)} \ . \tag{D.10}$$

It follows from $(D.7)$ that conformal transformations map the light cone $x^2 = 0$ onto the light cone $x'^2 = 0$, except for points $\sigma(x) = 0$ where the mapping $(D.2)$ is singular. For all points of the plane $1 - 2b \cdot x = 0$, including its intersection with the light cone $x^2 = 0$, we have $x'^2 = 1/b^2$. Hence, singular conformal mappings render lightlike vectors either spacelike or timelike, depending on whether $b^2 < 0$ or $b^2 > 0$.

Let $b^2 \neq 0$. Rewriting $(D.8)$ as

$$1 - 2b \cdot x + b^2 x^2 = b^2 \left(x - \frac{b}{b^2} \right)^2 ,$$

we see that $\sigma(x)$ may be of any sign. If $\sigma(x) < 0$, then x'^2 and x^2 are opposite in sign; a special conformal transformation can convert a timelike vector into spacelike and vice versa. By contrast, $(D.10)$ implies that the sign of the line element is invariant, in particular, spacelike and timelike have an invariant meaning for tangent vectors.

In summary, the conformal transformations are one-to-one mappings of Minkowski space on itself $x'^\mu = f^\mu(x)$ such that the resulting line element is identical to the initial one save for an overall scale factor,

$$dx'^2 = e^{2\lambda(x)}\, dx^2 \ . \tag{D.11}$$

The scale factor $e^{\lambda(x)}$ derives from the constant k in $(D.1)$, as well as from the coordinate dependent expression $|\sigma(x)|^{-1}$ in $(D.10)$.

Note that conformal transformations induce a factor of $e^{2\lambda(x)}$ not only on dx^2 but also on the scalar product of differentials dx_1^μ and dx_2^μ taken at the same point. This becomes clear from simple algebraic operations on $(dx_1' + dx_2')^2 = e^{2\lambda(x)}\,(dx_1 + dx_2)^2$. As a result, the cosine of angles between intersecting curves is invariant,

$$\cos\varphi = \frac{dx_1 \cdot dx_2}{\sqrt{dx_1^2\, dx_2^2}} = \text{const.} \tag{D.12}$$

Hence the name ‘conformal’, which indicates that the shape of any figure is unchanged by such transformations.

Because the special conformal transformations $(D.2)$ comprise an Abelian subgroup of the conformal group, one may interpret them as spacetime dependent dilatations, which bear close similarity to the *Weyl rescalings*

$$\eta_{\mu\nu} \to g_{\mu\nu} = e^{2\lambda(x)}\, \eta_{\mu\nu}, \quad \eta^{\mu\nu} \to g^{\mu\nu} = e^{-2\lambda(x)}\, \eta^{\mu\nu} \ . \tag{D.13}$$

Note that coordinates x^μ are unchanged by this transformation.

Proceeding from $(D.11)$, it is possible to reconstruct the conformal group acting on a spacetime of any dimension. Consider an infinitesimal coordinate transformation

$$x'^{\mu} = x^{\mu} + \epsilon\, \xi^{\mu}(x) \;, \tag{D.14}$$

where ξ^{μ} is an arbitrary smooth function. The line element $dx^2 = \eta_{\mu\nu} dx^{\mu} dx^{\nu}$ becomes

$$dx'^2 = dx^2 + \epsilon\, (\partial_{\mu}\xi_{\nu} + \partial_{\nu}\xi_{\mu})\, dx^{\mu} dx^{\nu} \;. \tag{D.15}$$

For $(D.14)$ to be treated as a conformal transformation, the parenthesized expression in $(D.15)$ should be proportional to $\eta_{\mu\nu}$,

$$\partial_{\mu}\xi_{\nu} + \partial_{\nu}\xi_{\mu} = B\, \eta_{\mu\nu} \;. \tag{D.16}$$

To solve this partial differential equation, we first differentiate it with respect to x^{λ},

$$\partial_{\lambda}\partial_{\mu}\xi_{\nu} + \partial_{\lambda}\partial_{\nu}\xi_{\mu} = \eta_{\mu\nu}\partial_{\lambda}B \;, \tag{D.17}$$

interchange ν and λ, and subtract the resulting equation from $(D.17)$,

$$\partial_{\lambda}\partial_{\mu}\xi_{\nu} - \partial_{\mu}\partial_{\nu}\xi_{\lambda} = \eta_{\mu\nu}\partial_{\lambda}B - \eta_{\lambda\mu}\partial_{\nu}B \;. \tag{D.18}$$

One further differentiation with respect to x^{ρ} gives

$$\partial_{\rho}\partial_{\mu}(\partial_{\lambda}\xi_{\nu} - \partial_{\nu}\xi_{\lambda}) = \eta_{\mu\nu}\partial_{\rho}\partial_{\lambda}B - \eta_{\mu\lambda}\partial_{\rho}\partial_{\nu}B \;. \tag{D.19}$$

Because the left side of $(D.19)$ is symmetric in ρ and μ, the same should also be true for the right side,

$$\eta_{\mu\nu}\partial_{\rho}\partial_{\lambda}B - \eta_{\mu\lambda}\partial_{\rho}\partial_{\nu}B = \eta_{\rho\nu}\partial_{\mu}\partial_{\lambda}B - \eta_{\rho\lambda}\partial_{\mu}\partial_{\nu}B \;, \tag{D.20}$$

which, by contraction on ρ and ν, and taking into account that $\delta^{\nu}{}_{\nu} = D+1$, gives

$$[(D-1)\,\partial_{\mu}\partial_{\lambda} + \eta_{\mu\lambda}\Box]\, B = 0 \;. \tag{D.21}$$

If $D+1=1$, every index takes a single value, and $(D.21)$ reads $0 \cdot B = 0$, which is tantamount to stating that B is an arbitrary function. If $D+1=2$, $(D.21)$ becomes $\Box B = 0$. Therefore, in a two-dimensional Euclidean space $\mathbb{R}_2$, B is an arbitrary harmonic function, while, in a pseudoeuclidean space $\mathbb{R}_{1,1}$, B is a solution to the wave equation. If $D+1>2$, then, by contraction on μ and λ in $(D.21)$, we find that $\Box B = 0$, which, in view of $(D.21)$, results in $\partial_{\mu}\partial_{\lambda}B = 0$. We thus conclude that B is linear in x,

$$B = 4\beta_{\alpha}x^{\alpha} + 2\gamma, \tag{D.22}$$

where β_{μ} and γ are arbitrary constants. We then substitute $(D.22)$ into $(D.18)$ and integrate the resulting equation with respect to x^{μ},

$$\partial_{\lambda}\xi_{\nu} - \partial_{\nu}\xi_{\lambda} = 4\,(\beta_{\lambda}x_{\nu} - \beta_{\nu}x_{\lambda}) + 2\omega_{\nu\lambda}, \tag{D.23}$$

where $\omega_{\nu\lambda}$ is a constant antisymmetric matrix. Combining $(D.16)$ and $(D.22)$ in

$$\partial_{\lambda}\xi_{\nu} + \partial_{\nu}\xi_{\lambda} = (4\beta_{\alpha}x^{\alpha} + 2\gamma)\,\eta_{\lambda\nu},$$

and adding it to $(D.23)$, we get

$$\partial_\lambda \xi_\nu = 2\left(\beta_\lambda x_\nu - \beta_\nu x_\lambda + \beta_\alpha x^\alpha \eta_{\lambda\nu}\right) + \omega_{\nu\lambda} + \gamma\, \eta_{\lambda\nu}\ . \tag{D.24}$$

Integrating $(D.24)$ with respect to x^λ we arrive at the desired solution

$$\xi_\mu = \epsilon_\mu + \omega_{\mu\alpha} x^\alpha + \gamma x_\mu + 2\beta_\alpha x^\alpha x_\mu - x^2 \beta_\mu, \quad \omega_{\mu\nu} = -\omega_{\nu\mu}\ . \tag{D.25}$$

In four dimensions, this solution contains 15 arbitrary real constants in ϵ_μ, $\omega_{\mu\nu}$, γ, and β_μ. An infinitesimal Poincaré transformation is represented by the 4 parameters in ϵ_μ and the 6 parameters in $\omega_{\mu\nu}$, while the 5 parameters in γ and β_μ are associated with infinitesimal dilatations and special conformal transformation, respectively.

Let $D+1 = 1$. The general solution to $(D.16)$ is an arbitrary smooth function. We thus see that any diffeomorphism of the real axis $\mathbb{R}$ is a conformal transformation.

The case $D+1 = 2$ is exceptional in that ξ_μ is given by either arbitrary harmonic functions (in Euclidean signature) or solutions to the wave equation (for pseudoeuclidean signature). The conformal group is infinite-dimensional.

Let $D+1 > 2$. ξ_μ is given by $(D.25)$, and the conformal group is specified by

$$(D+1) + \frac{1}{2}\, D\,(D+1) + 1 + (D+1) = \frac{1}{2}\,(D+2)\,(D+3) \tag{D.26}$$

independent parameters in ϵ_μ, $\omega_{\mu\nu}$, γ, and β_μ. This is the same number of parameters as necessary for the orthogonal group $\mathrm{SO}(D+3)$, or pseudoorthogonal groups $\mathrm{SO}(p,q)$, with $p+q = D+3$. This agreement is not accidental. We now demonstrate that the conformal group $C(m,n)$ in a pseudoeuclidean space $\mathbb{R}_{m,n}$ is isomorphic to the pseudoorthogonal group $\mathrm{SO}(m+1, n+1)$ in a pseudoeuclidean space $\mathbb{R}_{m+1,n+1}$, in particular $C(1,3)$ is isomorphic to $\mathrm{SO}(2,4)$.

Infinitesimal conformal transformations can be represented as

$$x'^\mu = \exp\left(G_k \epsilon^k\right) x^\mu = \left(1 + G_k \epsilon^k\right) x^\mu\ , \tag{D.27}$$

where G_k are generators of the conformal group. Let us list them:

translations

$$P_\mu = \partial_\mu\ , \tag{D.28}$$

rotations

$$M_{\mu\nu} = x_\mu \partial_\nu - x_\nu \partial_\mu\ , \tag{D.29}$$

dilatations

$$D = x^\alpha \partial_\alpha\ , \tag{D.30}$$

special conformal transformations

$$K_\mu = 2 x_\mu x^\alpha \partial_\alpha - x^2 \partial_\mu\ . \tag{D.31}$$

To illustrate, take ξ^μ for the special conformal transformation,

$$\xi^\mu = (\beta \cdot K)\, x^\mu = 2\beta \cdot x\, x^\mu - x^2 \beta^\mu \ . \tag{$D.32$}$$

From these expressions, the commutation relations are readily obtainable,

$$[M_{\mu\nu}, P_\lambda] = (\eta_{\nu\lambda} P_\mu - \eta_{\mu\lambda} P_\nu) \ , \tag{$D.33$}$$

$$[M_{\mu\nu}, M_{\rho\sigma}] = M_{\mu\sigma}\eta_{\nu\rho} + M_{\nu\rho}\eta_{\mu\sigma} - M_{\mu\rho}\eta_{\nu\sigma} - M_{\nu\sigma}\eta_{\mu\rho} \ , \tag{$D.34$}$$

$$[P_\mu, D] = P_\mu \ , \tag{$D.35$}$$

$$[M_{\mu\nu}, D] = 0 \ , \tag{$D.36$}$$

$$[P_\lambda, P_\mu] = 0 \ , \tag{$D.37$}$$

$$[D, D] = 0 \ , \tag{$D.38$}$$

$$[K_\lambda, K_\mu] = 0 \ , \tag{$D.39$}$$

$$[M_{\mu\nu}, K_\lambda] = (\eta_{\nu\lambda} K_\mu - \eta_{\mu\lambda} K_\nu) \ , \tag{$D.40$}$$

$$[K_\mu, D] = -K_\mu \ , \tag{$D.41$}$$

$$[P_\lambda, K_\mu] = 2\,(\eta_{\lambda\mu} D - M_{\lambda\mu}) \ . \tag{$D.42$}$$

Restricting ourselves to the commutation relations (D.33)–(D.38), we have the Lie algebra that generates the group of the similitude transformations (D.6). The generators of the Poincaré transformations and dilatations form a *closed* Lie algebra, which is a subalgebra of the full conformal algebra. Thus, on purely group-theoretical grounds, there is no reason why dilatation invariance should imply conformal invariance.

Note that commutation relations containing P_μ are similar to those containing K_μ. It is thus natural to look for a realization of the commutation relations (D.33)–(D.42) where P_μ and K_μ play interchangeable roles and take similar analytic forms.

Let us consider a six-dimensional space $\mathbb{R}_{2,4}$ of coordinates X^A, $A = 0,1,2,3,4,5$ in the following way. We first define four dimensionless coordinates

$$X^\mu = \kappa x^\mu, \quad \mu = 0,1,2,3 \ , \tag{$D.43$}$$

where κ fixes the unit of the measuring devices. A further variable λ is given by

$$\lambda = \kappa x^2 \ , \tag{$D.44$}$$

which, however, is not entirely independent but obeys the constraint

$$X^\mu X_\mu - \kappa\lambda = 0 \ . \tag{$D.45$}$$

The coordinates X^4 and X^5 are related to κ and λ by

$$X^4 = \frac{1}{2}\,(\lambda + \kappa), \quad X^5 = \frac{1}{2}\,(\lambda - \kappa) \ . \tag{$D.46$}$$

We finally introduce the metric tensor

$$\eta_{AB} = \mathrm{diag}(+1, -1, -1, -1, -1, +1) \ , \tag{D.47}$$

whereby the constraint $(D.45)$ takes the form

$$X^A X_A = 0 \ . \tag{D.48}$$

The reason for introducing the six-dimensional space is to *linearize* the nonlinear transformation $(D.2)$. Indeed, all the finite transformations of the conformal group $C(1,3)$ in the coordinates X^A turn out to be linear:

translations

$$X'^{\mu} = X^{\mu} + a^{\mu}\kappa, \quad \kappa' = \kappa, \quad \lambda' = \lambda + 2a \cdot X + a^2\kappa, \tag{D.49}$$

Lorentz transformations

$$X'^{\mu} = \Lambda^{\mu}_{\ \nu} X^{\nu}, \quad \kappa' = \kappa, \quad \lambda' = \lambda, \tag{D.50}$$

dilatations

$$X'^{\mu} = X^{\mu}, \quad \kappa' = \frac{1}{k}\,\kappa, \quad \lambda' = \lambda, \tag{D.51}$$

special conformal transformations

$$X'^{\mu} = X^{\mu} - b^{\mu}\kappa, \quad \kappa' = \kappa - 2b \cdot X + b^2\lambda, \quad \lambda' = \lambda \ . \tag{D.52}$$

Note the similar form of $(D.49)$ and $(D.52)$.

The generators of $C(1,3)$ can be expressed in terms of the new coordinates as

$$P_{\mu} = \kappa \frac{\partial}{\partial X^{\mu}} + 2X_{\mu} \frac{\partial}{\partial \lambda} = -X_4 \frac{\partial}{\partial X^{\mu}} - X_5 \frac{\partial}{\partial X^{\mu}} + X_{\mu} \frac{\partial}{\partial X^4} + X_{\mu} \frac{\partial}{\partial X^5} \ , \tag{D.53}$$

$$M_{\mu\nu} = X_{\mu} \frac{\partial}{\partial X^{\nu}} - X_{\nu} \frac{\partial}{\partial X^{\mu}} \ , \tag{D.54}$$

$$D = -\kappa \frac{\partial}{\partial \kappa} + \lambda \frac{\partial}{\partial \lambda} = X_5 \frac{\partial}{\partial X^4} - X_4 \frac{\partial}{\partial X^5} \ , \tag{D.55}$$

$$K_{\mu} = -\lambda \frac{\partial}{\partial X^{\mu}} - 2X_{\mu} \frac{\partial}{\partial \kappa} = X_4 \frac{\partial}{\partial X^{\mu}} - X_5 \frac{\partial}{\partial X^{\mu}} - X_{\mu} \frac{\partial}{\partial X^4} + X_{\mu} \frac{\partial}{\partial X^5} \ . \tag{D.56}$$

On the other hand, consider generators of rotations in $\mathbb{R}_{2,4}$,

$$L_{AB} = X_A \partial_B - X_B \partial_A, \quad A, B = 0, \ldots, 5 \ , \tag{D.57}$$

satisfying the standard commutation relations

$$[L_{AB}, L_{CD}] = \eta_{AD} L_{BC} + \eta_{CA} L_{DB} + \eta_{BC} L_{AD} + \eta_{DB} L_{CA} \ . \tag{D.58}$$

Comparing $(D.53)$–$(D.56)$ with $(D.57)$, one observes the correspondence

$$M_{\mu\nu} = L_{\mu\nu}, \quad D = L_{54}, \quad P_\mu = L_{\mu 5} + L_{\mu 4}, \quad K_\mu = L_{\mu 5} - L_{\mu 4} \ , \qquad (D.59)$$

or, in matrix form,

$$L_{AB} = \begin{pmatrix} M_{\mu\nu} & \frac{1}{2}(P_\mu - K_\mu) & \frac{1}{2}(P_\mu + K_\mu) \\ & 0 & D \\ & & 0 \end{pmatrix} , \qquad (D.60)$$

with $L_{AB} = -L_{BA}$. This correspondence makes it clear that the conformal group $C(1,3)$ is isomorphic to the pseudoorthogonal group $\mathrm{SO}(2,4)$.

This conclusion can readily be extended to the conformal group $C(m,n)$ acting on coordinates x^μ of a pseudoeuclidean space $\mathbb{R}_{m,n}$ of dimension $D + 1 = m + n$, $D > 1$. Following the same basic pattern, coordinates of this space should be supplemented by two new coordinates κ and λ, or X^{D+1} and X^{D+2}, together with the constraint $(D.45)$, to yield $\mathrm{SO}(m+1, n+1)$, a linear realization of the conformal Lie algebra commutation relations $(D.33)$–$(D.42)$. This isomorphism between $C(m,n)$ and $\mathrm{SO}(m+1, n+1)$ implies that any conformal group $C(m,n)$ is semisimple, excluding the case $D = 1$ when the conformal group is infinite-dimensional.

Problem D.1. Derive $(D.9)$.

Problem D.2. Verify $(D.33)$–$(D.42)$ using expressions $(D.28)$–$(D.31)$.

Problem D.3. Verify $(D.49)$–$(D.52)$.

Problem D.4. Prove the formulas $(D.53)$–$(D.56)$.

E. Grassmannian Variables

Consider a set $\mathcal{G}$ which consists of *even* and *odd* elements. By definition, even elements x_i are commuting numbers,

$$x_i x_j - x_j x_i = 0 \ , \qquad (E.1)$$

and odd elements θ_a are anticommuting numbers,

$$\theta_a \theta_b + \theta_b \theta_a = 0 \ . \qquad (E.2)$$

Furthermore, even elements commute with odd elements,

$$x_i \theta_a - \theta_a x_i = 0 \ . \qquad (E.3)$$

It follows from $(E.2)$ that the square of any odd element is zero,

$$\theta^2 = 0 \ . \qquad (E.4)$$

A vector space $\mathcal{G}$ with these properties is known as a *Grassmann algebra.* In general, the elements of $\mathcal{G}$ do not take numerical values; these quantities are defined by their algebraic properties. The class of even elements contains a subset of real numbers, which are called c-numbers. Only c-numbers assume (real or complex) numerical values.

A function of even and odd variables can be defined by the formal expansion,

$$F(\theta) = F_0 + F^a \theta_a + F^{ab} \theta_a \theta_b + \cdots \ , \qquad (E.5)$$

where the $F^{a_1 \cdots a_n}$ are antisymmetric c-number coefficients. This series terminates if the θ_a are finite in number. For example, if θ_a is a two-component vector, $a = 1, 2$, then the highest power is $\theta_1 \theta_2$. Thus, the $F(\theta)$ is defined by the set of its coefficients $F^{a_1 \cdots a_n}$.

Derivatives are defined by

$$dF = d\theta_a \, \frac{\partial F}{\partial \theta_a} \ , \qquad (E.6)$$

with $d\theta_a$ on the left (left derivative).

To define 'integration over Grassmannian variables', we note that the abstract measure must show translational invariance. For example, turning to the conventional c-number integral,

$$\int_{-\infty}^{\infty} dx \, F(x + a) = \int_{-\infty}^{\infty} dx \, F(x) \ . \qquad (E.7)$$

With this requirement, we come to the following basic integration rules:

$$\int d\theta_a = 0, \quad \int d\theta_a \, \theta_b = \frac{i}{\sqrt{\pi}} \, \delta_{ab} \ . \qquad (E.8)$$

Here, the coefficient of δ_{ab} may be chosen arbitrarily; the factor $i/\sqrt{\pi}$ is introduced for later convenience. Combining $(E.8)$ and $(E.5)$, one can integrate any function of Grassmannian variables.

A major application of the Grassmann algebra is *supersymmetry.* It combines fields of integer and half-integer spins (that is, bosons and fermions) into a single irreducible multiplet. Supersymmetry acts on an extended spacetime, called *superspace.* A point in superspace consists of Minkowski spacetime coordinates x^μ and additional anticommuting coordinates θ_α.

Supersymmetry is an extension of the Poincaré symmetry. Because we do not explore this idea in the main text, we will not develop it here.

Instead, we discuss the following remarkable result of Parisi and Sourlas. Let $\{x^\mu, \theta_1, \theta_2\}$ be a $(D + 2)$-dimensional superspace where the x^μ are coordinates of a D-dimensional Euclidean space and θ_1 and θ_2 are anticommuting coordinates which obey,

$$\theta_1^2 = \theta_2^2 = \theta_1 \theta_2 + \theta_2 \theta_1 = 0 \ . \qquad (E.9)$$

Suppose that the quadratic form

$$x^2 + \theta_1\theta_2 \tag{E.10}$$

is invariant under superrotations. Then this superspace is equivalent to an ordinary $(D-2)$-dimensional space. In particular, a space with only the anticommuting coordinates θ_1 and θ_2 is equivalent to an ordinary space having dimension -2 (in the sense that physical quantities are obtained by analytic continuation from positive dimensions). To see this, let us note that

$$\int d\theta_1 d\theta_2 f(\theta_1\theta_2) = -\frac{1}{\pi}\frac{df(u)}{du}\Big|_{u=0} = \lim_{n\to -2} S_n \int dr\, r^{n-1} f(r^2) = \lim_{n\to -2}\int d^n r\, f(r^2)\,, \tag{E.11}$$

where $S_n = 2\pi^{n/2}/\Gamma\left(\frac{1}{2}\,n\right)$ is the surface of the unit sphere in n dimensions.

We thus see that a D-dimensional superspace is equivalent to an ordinary space of dimension $D-2$ in the sense that

$$\int d^D x\, d\theta_1 d\theta_2\, f(x^2 + \theta_1\theta_2) = -\frac{1}{\pi}\int d^D x\, f'(x^2) = \int d^{D-2}x\, f(x^2)\,. \tag{E.12}$$

Problem E.1. Show that

$$-\frac{1}{\pi}\, f'(0) = \lim_{n\to -2} S_n \int_0^\infty dr\, r^{n-1}\, f(r^2) \tag{E.13}$$

Hint Write $n = -2(1-\delta)$. Then

$$S_n = \frac{2}{\pi\Gamma\left(-1+\delta\right)} \to -\frac{2\delta}{\pi} \tag{E.14}$$

as $\delta \to 0$. The divergent integral $(E.13)$ can be assigned a mathematical meaning through the regularization,

$$\int_0^\infty \frac{dr}{r^3}\, f(r^2) = \int_0^\infty \frac{dr}{r^3}\left[f(r^2) - f(0) - f'(0)r^2\right] + \int_0^\infty \frac{dr}{r^3}\, f(0) + \int_0^\infty \frac{dr}{r}\, f'(0)\,, \tag{E.15}$$

where

$$\int_0^\infty \frac{dr}{r^3}\, f(0) = \int_0^\infty dr\, e^{-\epsilon r} r^{n-1}\, f(0)\,, \tag{E.16}$$

and

$$\int_0^\infty \frac{dr}{r}\, f'(0) = \int_0^\infty dr\, e^{-\epsilon r} r^{n+1}\, f'(0)\,. \tag{E.17}$$

With $(E.14)$–$(E.17)$,

$$S_n \int_0^\infty \frac{dr}{r^3}\, f(0) = \frac{2}{\pi\Gamma(-1+\delta)}\,\epsilon^2\,\frac{\Gamma(-1+2\delta)}{-2+2\delta}\, f(0) \to 0\,, \tag{E.18}$$

and

$$S_n \int_0^\infty \frac{dr}{r}\, f'(0) = \frac{2}{\pi \Gamma(-1+\delta)}\, \epsilon^{2\delta}\, (-1+2\delta)\, \Gamma(-1+2\delta)\, f'(0) \to -\frac{f'(0)}{\pi} \tag{E.19}$$

as $\epsilon \to 0$ and $\delta \to 0$.

F. Distributions

There are three alternative ways of looking at the Dirac delta-function $\delta(x)$ and similar objects called *distributions* or *generalized functions*. First, a distribution f is defined as a linear continuous *functional* on a *test function space*, or *basic space* $\mathcal{K}$, namely, we associate with every *test function* $\phi \in \mathcal{K}$ a real number $\langle f, \phi \rangle$ in such a way that, for any two ϕ_1 and ϕ_2 in $\mathcal{K}$ and real numbers a_1 and a_2, we have $\langle f, a_1\phi_1 + a_2\phi_1 \rangle = a_1\langle f, \phi_1 \rangle + a_2\langle f, \phi_2 \rangle$, and any sequence of test functions ϕ_n converging to ϕ in $\mathcal{K}$ implies convergence of the numerical sequence $\langle f, \phi_n \rangle \to \langle f, \phi \rangle$. Second, a singular functional f, such as $\delta(x)$, is defined as a limit of a *sequence* of regular functionals. The third definition of generalized functions on the real axis arises in studies of *boundary values* of functions which are analytic in a complex half-plane. Any of these definitions may be found attractive for some problems, and less convenient in others; it is resonable to invoke them interchangeably.

Let $f(x)$ be an ordinary summable function. It becomes a distribution through the use of the linear functional

$$\langle f, \phi \rangle = \int_{-\infty}^{\infty} dx\, f(x)\, \phi(x)\ , \tag{F.1}$$

where ϕ are test functions disappearing at infinity. Rather than specify the function $f(x)$ by values it takes on the real axis $\mathbb{R}$, we reproduce it almost everywhere on $\mathbb{R}$ from data reported by the linear functional $(F.1)$ for all test functions $\phi \in \mathcal{K}$ when $\mathcal{K}$ is reach enough. If a distribution f is realized in terms of a conventional integral construction such as $(F.1)$, we call it *regular*, otherwise, that is, when $(F.1)$ has only a symbolic meaning, f is called *singular*. Consider, for instance, the Heaviside *step function*

$$\theta(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{F.2}$$

We have a regular distribution

$$\langle \theta, \phi \rangle = \int_0^\infty dx\, \phi(x)\ . \tag{F.3}$$

By contrast, the Dirac *delta-function* is a singular distribution whose action on test functions is given by

$$\langle \delta, \phi \rangle = \int_{-\infty}^{\infty} dx\, \delta(x)\, \phi(x) = \phi(0). \tag{F.4}$$

Letting $\phi(x) = 1$,

$$\int_{-\infty}^{\infty} dx\, \delta(x) = 1. \tag{F.5}$$

It is generally taken that test functions ϕ must be infinitely differentiable and fall rapidly as $|x| \to \infty$. This enables us to define the *derivative* of a distribution f by

$$\int_{-\infty}^{\infty} dx\, f'(x)\, \phi(x) = \int_{-\infty}^{\infty} dx\, f(x)\, \phi'(x)\ , \tag{F.6}$$

or, symbolically,

$$\langle f', \phi \rangle = -\langle f, \phi' \rangle. \tag{F.7}$$

For example,

$$\langle \theta', \phi \rangle = -\langle \theta, \phi' \rangle = -\int_{0}^{\infty} dx\, \phi'(x) = \phi(0)\ , \tag{F.8}$$

and so

$$\theta'(x) = \delta(x)\ . \tag{F.9}$$

Every distribution has derivatives of all orders. For example, the nth derivative of $\delta(x)$ is a distribution acting as

$$\langle \delta^{(n)}, \phi \rangle = (-1)^n \int_{-\infty}^{\infty} dx\, \delta(x)\, \phi^{(n)}(x) = (-1)^n\, \phi^{(n)}(0)\ . \tag{F.10}$$

A test function $\phi(x)$ is said to be *finite* if it has a finite support (which is, by definition, the closure of the set of points x where ϕ is nonzero). A basic space, designated as $\mathcal{D}$, is the space of all finite functions. In other words, $\mathcal{D}$ contains all infinitely differentiable real functions vanishing outside closed intervals. As an example, consider

$$\phi(x) = \begin{cases} \exp\left(-\frac{1}{a^2 - x^2}\right) & \text{if } |x| < a, \\ 0 & \text{elsewhere.} \end{cases} \tag{F.11}$$

This finite function is said to be *concentrated* in the compact region $|x| \leq a$. Distributions on $\mathcal{D}$ form a topological vector space denoted by $\mathcal{D}'$.

A wider basic space $\mathcal{S}$ involves all infinitely differentiable real functions that approach zero more rapidly than any inverse power of x as $|x| \to \infty$. Its associated space of distributions is denoted by $\mathcal{S}'$. Because finite test functions belong to $\mathcal{S}$, $\mathcal{D} \subset \mathcal{S}$, and $\mathcal{S}' \subset \mathcal{D}'$. Indeed, we may regard $\exp(x^2)$ as a distribution on $\mathcal{D}'$ but not on $\mathcal{S}'$.

In the nonlocal electrodynamics discussed in Sect. 10.5, the appropriate basic space $\mathcal{Z}$ contains slowly increasing *entire* functions, that is, all functions

$\phi(z)$ of the complex variable $z = x + iy$, which are analytic everywhere, except at infinity, and are subject to the conditions

$$|z^n \phi(z)| \leq C_n \exp(-a|x| + b|y|), \quad n = 0, 1, \ldots, N \; , \qquad (F.12)$$

where a, b, C_n, and N are constants (dependent on ϕ). Note that $\mathcal{Z}$ contains no finite function, because an analytic function vanishing on a finite interval is zero everywhere[3].

One may wonder of whether it is possible to change the integration variable x in (F.1). Let $U : x \to y = U(x)$ be a one-to-one smooth mapping, and $U^{-1} : y \to x = U^{-1}(y)$ its inverse. By analogy with classical analysis,

$$\int_{-\infty}^{\infty} dx\, f[U(x)]\, \phi(x) = \int_{-\infty}^{\infty} dy\, \frac{1}{|U'|}\, f(y)\, \phi[U^{-1}(y)] \; , \qquad (F.13)$$

where $U' = dy/dx$. To be specific, consider the delta-function and its derivatives.

(i) Translation $U(x) = x - a$:

$$\int_{-\infty}^{\infty} dx\, \delta(x-a)\, \phi(x) = \int_{-\infty}^{\infty} dy\, \delta(y)\, \phi(y+a) = \phi(a). \qquad (F.14)$$

(ii) Dilatation $U(x) = kx$:

$$\delta(kx) = \frac{1}{|k|}\, \delta(x) \; . \qquad (F.15)$$

This is interpreted as that $\delta(x)$ is a *homogeneous* function of order -1.

(iii) Reflection $U(x) = -x$:

$$\delta^{(n)}(-x) = (-1)^n\, \delta^{(n)}(x) \; , \qquad (F.16)$$

in particular

$$\delta(-x) = \delta(x) \; , \qquad (F.17)$$

which suggests that $\delta(x)$ may be regarded as an *even* function.

(iv) A general smooth one-to-one transformation $U(x)$:

$$\delta[U(x)] = \sum_n \frac{1}{|U'(x_n)|}\, \delta(x - x_n) \; , \qquad (F.18)$$

where x_n are ordinary roots of the equation $U(x) = 0$.

Two cases of particular interest are $\delta(x^2 - a^2)$ and $\mathrm{sgn}(x_0)\, \delta(x^2 - a^2)$, where x^2 is the squared four-dimensional radius vector $x^2 = {x_0}^2 - \mathbf{x}^2$, and

[3] It may appear that expression (F.11) provides a counterexample of this statement: $\phi(x)$ is nonzero in the region $|x| < a$, and vanishes everywhere outside it. However, $\phi(x)$ is *not* analytic (even if infinitely differentiable), since it is built from $\exp\left(-\frac{1}{a^2 - z^2}\right)$ which has essential singularities at $z = a$ and $z = -a$.

$$\operatorname{sgn}(s) = \theta(s) - \theta(-s) \tag{F.19}$$

is the *signum function*. Letting $a > 0$, the equation $U(x) = x^2 - a^2 = 0$ has two different roots $x_0 = \pm\sqrt{\mathbf{x}^2 + a^2}$, while $\partial U/\partial x|_{x_0} = 2x_0$, hence

$$\delta(x^2 - a^2) = \frac{1}{2a}\left[\delta\left(x^0 - \sqrt{\mathbf{x}^2 + a^2}\right) + \delta\left(x^0 + \sqrt{\mathbf{x}^2 + a^2}\right)\right] , \tag{F.20}$$

and

$$\operatorname{sgn}(x_0)\,\delta(x^2 - a^2) = \frac{1}{2a}\left[\delta\left(x^0 - \sqrt{\mathbf{x}^2 + a^2}\right) - \delta\left(x^0 + \sqrt{\mathbf{x}^2 + a^2}\right)\right] . \tag{F.21}$$

To gain an insight regarding distributions as limits of sequences of functions, imagine for a while that $\delta(x)$ is a regular function which looks like a sharp pulse normalized to unit area, as it must according to (F.5). Each of the following sequences of functions has this property

$$\delta_\epsilon(x) = \frac{1}{2\sqrt{\pi\epsilon}}\exp\left(-\frac{x^2}{4\epsilon}\right) , \tag{F.22}$$

$$\delta_\epsilon(x) = \frac{\epsilon}{\pi(x^2 + \epsilon^2)} , \tag{F.23}$$

$$\delta_\Lambda(x) = \frac{\sin(\Lambda x)}{\pi x} . \tag{F.24}$$

As ϵ approaches zero, or Λ goes to infinity, the peak becomes high and narrow, but normalization is preserved,

$$\frac{1}{2\sqrt{\pi\epsilon}}\int_{-\infty}^{\infty} dx\, e^{-\frac{x^2}{4\epsilon}} = \frac{\epsilon}{\pi}\int_{-\infty}^{\infty}\frac{dx}{x^2 + \epsilon^2} = \frac{1}{\pi}\int_{-\infty}^{\infty} dx\,\frac{\sin(\Lambda x)}{x} = 1. \tag{F.25}$$

With these delta sequences, it is straightforward to show that

$$\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{ikx} . \tag{F.26}$$

Indeed, by cutting off the integration region we obtain (F.24),

$$\frac{1}{2\pi}\int_{-\Lambda}^{\Lambda} dk\, e^{ikx} = \frac{\sin(\Lambda x)}{\pi x} , \tag{F.27}$$

which approximates $\delta(x)$ in the limit $\Lambda \to \infty$.

A further trick is to regularize the integrand of (F.26) by the factor $e^{-\epsilon k^2}, \epsilon > 0$, which goes to 1 as $\epsilon \to 0$. The result is the delta sequence defined in (F.22),

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{ikx - \epsilon k^2} = \frac{1}{2\sqrt{\pi\epsilon}}\exp\left(-\frac{x^2}{4\epsilon}\right) . \tag{F.28}$$

Choosing a slightly more involved regularization factor $\theta(k)e^{-\epsilon k}+\theta(-k)e^{\epsilon k}$, that approaches $\theta(k)+\theta(-k)=1$ as $\epsilon\to 0$, we obtain

$$\frac{1}{2\pi}\left(\int_{-\infty}^{0} dk\, e^{ikx+\epsilon k}+\int_{0}^{\infty} dk\, e^{ikx-\epsilon k}\right)=\frac{1}{2\pi}\left(\frac{1}{ix+\epsilon}-\frac{1}{ix-\epsilon}\right)=\frac{\epsilon}{\pi(x^2+\epsilon^2)}. \tag{F.29}$$

This expression is identical to (F.23).

In view of relation (F.26), the Fourier transform of $\delta(x)$ is 1,

$$\widetilde{\delta}(k)=\int_{-\infty}^{\infty} dx\, e^{-ikx}\,\delta(x)=1\ , \tag{F.30}$$

and the Fourier transforms of derivatives of $\delta(x)$ are monomials,

$$\widetilde{\delta}^{(n)}(k)=\int_{-\infty}^{\infty} dx\, e^{-ikx}\,\delta^{(n)}(x)=(ik)^n\ . \tag{F.31}$$

From (F.29) we see that the Fourier transform of $\theta(x)$ is

$$\widetilde{\theta}(k)=\int_{-\infty}^{\infty} dx\, e^{-ikx}\,\theta(x)=\frac{i}{-x+i\epsilon}\ . \tag{F.32}$$

Although distributions are not pointwise functions, their local properties can still be examined if the basic space contains finite functions. Indeed, a distribution f vanishes in an open domain U of a point x if $\langle f,\phi\rangle=0$ for every test function ϕ which has its support in U. The *support* of a distribution f is defined as the smallest closed set of points outside which f vanishes.

One can conclude from (F.4) that the support of $\delta(x)$ is a single point $x=0$, in other words, $\delta(x)$ is *concentrated* in the compact (one-point) set $x=0$. Equations

$$x\,\delta(x)=0\ , \tag{F.33}$$

$$x\,\delta'(x)=-\delta(x)\ , \tag{F.34}$$

$$x^n\,\delta^{(n)}(x)=(-1)^n\,n!\,\delta(x)\ , \tag{F.35}$$

derivable from (F.4) and (F.10), are immediate consequences of this fact. It is also clear that (F.34) follows from the Euler theorem on homogeneous functions.

Conversely,

$$y(x)=c\,\delta(x) \tag{F.36}$$

(where c is an arbitrary constant) is the general solution to the equation

$$x\,y(x)=0. \tag{F.37}$$

Furthermore, the general solution to the equation

$$x^N y(x)=0 \tag{F.38}$$

is

$$y(x) = \sum_{n=0}^{N-1} c_n \, \delta^{(n)}(x) \; , \tag{F.39}$$

where c_n are arbitrary coefficients. To prove this statement, write the Fourier transform of $(F.38)$,

$$\widetilde{y}^{(N)}(k) = 0. \tag{F.40}$$

This ordinary differential equation is readily solved to give the Fourier transform of $(F.39)$,

$$\widetilde{y}(k) = \sum_{n=0}^{N-1} c_n \, (ik)^n. \tag{F.41}$$

Note in passing that $\operatorname{sgn}(x_0)\,\delta(x^2 - a^2)$ is a Lorentz invariant function despite the presence of the frame dependent factor $\operatorname{sgn}(x_0)$. Indeed, as $(F.21)$ shows, this function is concentrated in two sheets of the hyperboloid $x_0 = \sqrt{\mathbf{x}^2 + a^2}$ and $x_0 = -\sqrt{\mathbf{x}^2 + a^2}$, each being a Lorentz invariant region.

We saw that the Fourier transform of distributions may be analytic functions. In general, distributions on the real axis $\mathbb{R}$ of points x are closely related to boundary values of analytic functions on the complex plane $\mathbb{C}$ of points $z = x + iy$. Such a relation can be illustrated by the formula

$$\frac{1}{x + i\epsilon} = \mathrm{P}\left(\frac{1}{x}\right) - i\pi\delta(x) \; , \tag{F.42}$$

where P stands for the Cauchy principal value. To prove it, we first note the identity

$$\frac{1}{x + i\epsilon} = \frac{1}{2}\left(\frac{1}{x + i\epsilon} + \frac{1}{x - i\epsilon}\right) + \frac{1}{2}\left(\frac{1}{x + i\epsilon} - \frac{1}{x - i\epsilon}\right) . \tag{F.43}$$

Suppose that test functions $\phi(x)$ may be continued to the complex plane. Then $(F.43)$ can be interpreted as a rule for avoiding singularities when each term, multiplied by $\phi(x)$, is integrated over the real axis. Because the singular point $z = -i\epsilon$ approaches $z = 0$ in the limit $\epsilon \to 0$, the integration contour must be slightly curved in the upper half-plane to yield a small integration semicircle Γ_+, as in Fig. .1. Likewise, the singular point $z = i\epsilon$, descending towards the real axis, makes the integration contour slightly bent in the lower half-plane forming a semicircle Γ_-. Integration of the first parenthesized term gives just what is meant by the Cauchy principal value,

$$\frac{1}{2}\int_{-\infty}^{\infty} dx \left(\frac{1}{x + i\epsilon} + \frac{1}{x - i\epsilon}\right) \phi(x) = \left(\int_{-\infty}^{-\epsilon} + \int_{\epsilon}^{\infty}\right) dx \, \frac{\phi(x)}{x} \; , \tag{F.44}$$

because the integrals along Γ_+ and Γ_- are equal but opposite in sign and cancel,

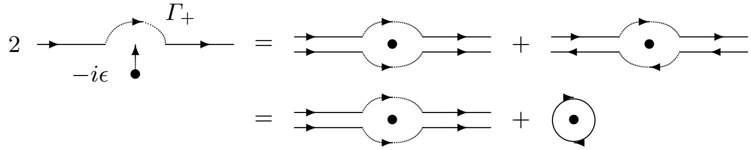


Fig. .1. The integration contours suitable for $(F.43)$

$$\int_{\Gamma_+} \frac{dz}{z}\,\phi(z) + \int_{\Gamma_-} \frac{dz}{z}\,\phi(z) = i\left(\int_\pi^0 - \int_\pi^0\right) d\vartheta\,\phi(\epsilon e^{i\vartheta}) = 0\;. \tag{F.45}$$

Thus,

$$\frac{1}{2}\int_{-\infty}^{\infty} dx\left(\frac{1}{x+i\epsilon} + \frac{1}{x-i\epsilon}\right)\phi(x) = \int_{-\infty}^{\infty} dx\,\mathrm{P}\left(\frac{1}{x}\right)\phi(x)\;. \tag{F.46}$$

Integration of the second parenthesized term gives

$$\frac{1}{2}\int_{-\infty}^{\infty} dx\,\frac{\phi(x)}{x+i\epsilon} - \frac{1}{2}\int_{-\infty}^{\infty} dx\,\frac{\phi(x)}{x-i\epsilon} = -\frac{1}{2}\oint dz\,\frac{\phi(z)}{z}\;, \tag{F.47}$$

where the closed integration contour is traversed in a counterclockwise direction with respect to the point $z = 0$. We thus arrive at the Cauchy integral formula, which, in effect, amounts to the delta-function,

$$\frac{1}{2\pi i}\oint dz\,\frac{\phi(z)}{z} = \phi(0) = \int_{-\infty}^{\infty} dx\,\delta(x)\,\phi(x)\;. \tag{F.48}$$

This completes proof of $(F.42)$. Figure 1 is a pictorial representation of this argument.

In spaces of higher dimension, the delta-function is defined as a product of expressions whose action is equivalent to that of one-dimensional delta-functions. For example, in Cartesian coordinates, the three-dimensional delta-function is

$$\delta^3(\mathbf{x}) = \delta(x_1)\,\delta(x_2)\,\delta(x_3)\;, \tag{F.49}$$

which gives

$$\int d^3x\,\delta^3(\mathbf{x})\,\phi(\mathbf{x}) = \phi(\mathbf{0}). \tag{F.50}$$

We now see that the *support* of $\delta^3(\mathbf{x})$ is a *single point* $\mathbf{x} = \mathbf{0}$.

Let A: $\mathbf{x} \to \mathbf{y} = A\mathbf{x}$ be a nonsingular linear mapping (that is, $A_{ij} = \partial \mathbf{y}_i/\partial \mathbf{x}_j$ is a constant matrix such that $\det A \neq 0$). Then $(F.13)$ extends to the three-dimensional case as follows

$$\delta^3(A\mathbf{x}) = \frac{1}{|\det A|}\,\delta^3(\mathbf{x}). \tag{F.51}$$

If A is specified to be a rotation (in which case $|\det A| = 1$), then

$$\delta^3(A\mathbf{x}) = \delta^3(\mathbf{x}). \tag{F.52}$$

Hence, $\delta^3(\mathbf{x})$ is a *rotationally invariant* distribution.

In spherical coordinates,

$$\delta^3(\mathbf{x}) = \frac{1}{r^2}\,\delta(r)\,\frac{1}{\sin\vartheta}\,\delta(\vartheta)\,\delta(\varphi). \tag{F.53}$$

Dilatations are characterized by the matrices $A_{ij} = k\,\delta_{ij}$, implying $\det A = k^3$. In lieu of (F.15),

$$\delta^3(k\mathbf{x}) = \frac{1}{|k|^3}\,\delta^3(\mathbf{x}). \tag{F.54}$$

With (F.49) and (F.26), we have

$$\delta^3(\mathbf{x}) = \frac{1}{(2\pi)^3}\int d^3k\, e^{i\mathbf{k}\cdot\mathbf{x}}\,, \tag{F.55}$$

where $\mathbf{k}\cdot\mathbf{x} = k_1x_1 + k_2x_2 + k_3x_3$, and the multiple integral is understood as usual,

$$\int d^3k = \int_{-\infty}^{\infty} dk_1 \int_{-\infty}^{\infty} dk_2 \int_{-\infty}^{\infty} dk_3. \tag{F.56}$$

Likewise, the four-dimensional delta-function is

$$\delta^4(x) = \delta(x_0)\,\delta(x_1)\,\delta(x_2)\,\delta(x_3)\,, \tag{F.57}$$

where x_0, x_1, x_2, x_3 are rectilinear coordinates in a particular Lorentz frame. One can conclude that $\delta^4(x)$ is a *Lorentz invariant* distribution concentrated at $x^\mu = 0$, and

$$\delta^4(x) = \frac{1}{(2\pi)^4}\int d^4k\, e^{-ik\cdot x}. \tag{F.58}$$

Here, $k\cdot x$ stands for the pseudoeuclidean scalar product $k_0x_0 - \mathbf{k}\cdot\mathbf{x}$.

In studies of nonlocal field theories à la Efimov, we deal with distributions of the type

$$E(x) = \sum_{n=0}^{\infty} c_n \Box^n \delta^4(x)\,, \tag{F.59}$$

where $\Box$ is the d'Alembertian. The Fourier transform of such a distribution is

$$\widetilde{E}(k^2) = \sum_{n=0}^{\infty} (-1)^n c_n k^{2n}\,. \tag{F.60}$$

Such a power series is an analytic function with radius of convergence dependent on the behavior of the c_n. Of primary concern to the present discussion are those $\widetilde{E}(k^2)$ for which the power series (F.60) are convergent everywhere in the complex k^2-plane, and hence $\widetilde{E}(k^2)$ are analytic functions everywhere except at infinity, that is, *entire* functions.

We now recall some notions of the theory of entire functions. Let F be a function of the complex variable $z = Re^{i\vartheta}$, represented by an everywhere convergent power series

$$F(z) = \sum_{n=0}^{\infty} c_n z^n \ . \tag{F.61}$$

It is imperative that the entire function has an essential singularity at $z = \infty$. This implies that, when going to $|z| = \infty$ along various directions, $|F(z)|$ may vanish asymptotically, or grow without bound, or approach any specific value. The growth of an entire function is specified by its *order* Δ and its *type* σ,

$$\Delta = \lim_{R\to\infty} \frac{\ln\ln|F(z)|}{\ln R} \ , \tag{F.62}$$

$$\sigma = \lim_{R\to\infty} \frac{\ln|F(z)|}{R^{\Delta}} \ . \tag{F.63}$$

It follows from (F.62) and (F.63) that the asymptotic growth of $|F(z)|$ obeys the bound

$$|F(z)| < C \exp\left(\sigma R^{\Delta}\right) , \quad R \to \infty \ . \tag{F.64}$$

If $0 < \sigma < \infty$, then F is said to be a function of *normal* type, and if $\sigma = 0$, then F is a function of *minimal* type.

The order and type of entire functions are related to the rate of decrease in sequence of its Taylor coefficients

$$\Delta = -\lim_{n\to\infty} \frac{n\ln n}{\ln|c_n|} \ , \tag{F.65}$$

$$(\sigma\, e\, \Delta)^{1/\Delta} = \lim_{n\to\infty} n^{1/\Delta}|c_n|^{1/n}. \tag{F.66}$$

Indeed, any term of the power series (F.61) obeys the obvious inequality

$$|c_n|\, R^n \le \max_{\vartheta} |F(Re^{i\vartheta})| \ , \tag{F.67}$$

and, by (F.64),

$$|c_n| < CR^{-n} \exp(\sigma R^{\Delta}) \ . \tag{F.68}$$

The right-hand-side of this inequality is minimized at

$$R = \left(\frac{n}{\sigma\Delta}\right)^{\frac{1}{\Delta}} \ . \tag{F.69}$$

Hence for large n we have

$$|c_n| \sim \left(\frac{e\sigma\Delta}{n}\right)^{\frac{n}{\Delta}} \ . \tag{F.70}$$

which is tantamount to stating that (F.65) and (F.66) hold.

When it comes to entire functions of order 1, (F.66) becomes

$$\sigma = \frac{1}{e} \lim_{n\to\infty} n\,|c_n|^{1/n}. \tag{F.71}$$

An important characteristic of such functions is the directrix $H(\vartheta)$ which is defined as

$$H(\vartheta) = \lim_{R\to\infty} \frac{\ln|F(Re^{i\vartheta})|}{R} \,. \tag{F.72}$$

Consider functions with

$$H(\vartheta) = \ell \sin\vartheta \,. \tag{F.73}$$

Entire functions with such a directrix behave asymptotically as

$$|F(Re^{\vartheta})| \le C \exp(\ell R \sin\vartheta), \quad R \to \infty \,. \tag{F.74}$$

Problem F.1. Show that

$$\delta(\sin x) = \sum_{n=-\infty}^{\infty} \delta(x - \pi n) \,. \tag{F.75}$$

Problem F.2. Verify

$$\lim_{\epsilon\to 0} \frac{1}{2\sqrt{\pi\epsilon}} \int_{-\infty}^{\infty} dx\, \phi(x)\, \exp\left(-\frac{x^2}{4\epsilon}\right) = \phi(0) \,, \tag{F.76}$$

and similar equations for the delta sequences (F.23) and (F.24).

Problem F.3. Show that the following sequences of regular functions

$$\theta_\epsilon(x) = \frac{1}{\pi}\left[\frac{\pi}{2} + \arctan\left(\frac{x}{\epsilon}\right)\right], \quad \theta_\Lambda(x) = \frac{1}{\exp(-\Lambda x) + 1} \tag{F.77}$$

approximate the Heaviside step function $\theta(x)$ as $\epsilon \to 0$, and $\Lambda \to \infty$.

Problem F.4. Show that (F.34) generalizes to the four-dimensional case as follows

$$(x \cdot \partial)\, \delta^4(x) = -4\, \delta^4(x) \,. \tag{F.78}$$

Notes

1. *A*. An exhaustive account of differential forms and their applications can be found in many books, for example: De Rham (1955), Spivak (1965, 1974), Cartan (1967), Schwartz (1967), and Dubrovin, Fomenko & Novikov (1992).

2. *B*. Lie (1888, 1890, 1893) laid the foundation of the theory of Lie groups and Lie algebras. Of many mathematical books devoted to this subject we

would like to mention Pontryagin (1939), and Bourbaki (1968). There are texts oriented to a general physics audience. An example is provided by Georgi (1982). The books by Gel'fand et al. (1963), Naimark (1964), and Barut & Rączka (1977) give the reader a comprehensive idea of infinite-dimensional unitary representations of the Lorentz and Poincaré groups.

3. *C.* Most of the basic properties of γ-matrices presented here is due to Pauli (1936).

4. *D.* The conformal group is discussed in many texts on mathematical physics, of which we mention two: Dubrovin, Fomenko & Novikov (1992), and Fushchich & Nikitin (1987). In this Appendix, we summarize results which occur frequently in the physics literature, following Wess (1960), Kastrup (1962), Fulton et al. (1962), and Barut & Haugen (1972). An application of dilatation invariance in particle physics can be found in the book by Coleman (1985), Chap. 3. $D + 1 = 2$ is of basic importance in string theory. In this case, the conformal group is infinite. For a review of conformal invariance in string theory see Green, Schwartz & Witten (1987), Polchinski (1998), and Siegel (1999).

5. *E.* The supersymmetry was discovered by Gol'fand & Likhtman (1971), Ramond (1971), Neveu & Schwarz (1971), Gervais & Sakita (1971), Volkov & Akulov (1972), and Wess & Zumino (1972). Parisi & Sourlas (1979) showed that field theories in a D-dimensional superspace $\{x^\mu, \theta, \bar{\theta}\}$ with D commuting coordinates x^μ and two anticommuting coordinates θ and $\bar{\theta}$ are equivalent to similar field theories in ordinary $(D-2)$-dimensional space $\{x^\mu\}$. The dimensional reduction $D \to D-2$ is attributed to the negative dimensionality of the anticommuting variables θ and $\bar{\theta}$. There is a considerable amount of pedagogical literature on supersymmetry, for example, Wess & Bagger (1983), Gates et al. (1983), and Berezin (1987). This subject is discussed to some extent in many texts on quantum field theory. Thorough reviews can be found in Mohapatra (1986), Siegel (1999), and Weinberg (1996).

6. *F.* The basic reference on the distribution theory is Schwartz (1950–1951), Gel'fand & Shilov (1964, 1968), and Bremermann (1966). For the use of distributions in quantum field theory see Bogoliubov, Logunov, Oksak & Todorov (1990). Nonlocal distributions have been studied extensively by Efimov (1968). For basic properties of entire functions the reader may consult Titchmarsh (1932), and Paley & Wiener (1934).

References

Ablowitz, M. J. & H. Segur (1981). *Solitons and the Inverse Scattering Transform*. Philadelphia: Society for Industrial and Applied Mathematics.

Abraham, M. (1903). Prinzipen der Dynamik des Elektrons. *Annalen der Physik*, **10**, 105.

Abraham, M. (1904). Zur Theorie der Strahlung. *Annalen der Physik*, **14**, 236.

Abraham, M. (1905). *Theorie der Elektrizität*. Vol. II. 5th edition 1923. Leipzig: Teubner.

Abraham, R. & J. E. Marsden (1967). *Foundations of Mechanics*. 2nd edition 1978. Reading: Benjamin-Cummings.

Abrikosov, A. A. (1957). On the magnetic properties of superconductors of the second group. *Soviet Physics – JETP*, **5**, 1174.

Actor, A. (1979). Classical solutions of SU(2) Yang–Mills theories. *Reviews of Modern Physics*, **51**, 461.

Arzeliès, H. (1966). *Rayonnement et dynamique du corpuscule chargé fortement accéléré*. Paris: Gauthier–Villars.

Arnold, V. I. (1978). *Mathematical Methods of Classical Mechanics*. Translated from the Russian edition 1974. Heidelberg: Springer.

Atiyah, M. & N. Hitchin (1988). *The Geometry and Dynamics of Magnetic Monopoles*. Princeton: Princeton University Press.

Baker, B. B. & E. T. Copson (1939). *The Mathematical Theory of Huygens' Principle*. 2nd edition 1950. Oxford: Clarendon.

Balachandran, A., S. Borchardt & A. Stern (1978). Lagrangian and Hamiltonian description of Yang–Mills particles. *Physical Review*, **D 17**, 3247.

Balachandran, A. P., G. Marmo, B. Skagerstam & A. Stern (1982). *Gauge Symmetries and Fiber Bundles - Applications to Particle Dynamics*. Lecture Notes in Physics **188**. Berlin: Springer.

Bander, M. (1981). Theories of quark confinement. *Physics Reports*, **75**, 205.

Barut, A. O. (1964). *Electrodynamics and Classical Theory of Fields and Particles*. New York: Colier–Macmillan. 2nd edition 1980. New York: Dover.

Barut, A. O. (1974). Electrodynamics in terms of retarded fields. *Physical Review*, **D 10**, 3335.

Barut, A. O. & H. Beker (1974). Classical relativistic scattering of two dyons. *Nuovo Cimento*, **A 19**, 309.

Barut, A. O. & R. Haugen (1972). Theory of conformally invariant mass. *Annals of Physics*, **71**, 519.

Barut, A. O. & R. Rączka (1977). *Theory of Group Representations and Applications*. 2nd edition 1980. Warszawa: PWN–Polish Scientific Publishers.

Barut, A. O. & D. Villarroel (1975). Radiation reaction and mass renormalization in scalar and tensor fields and linearized gravitation. *Journal of Physics*, **A 8**, 156.

Barut, A. O. & N. Zanghi (1984). Classical model of the Dirac electron. *Physical Review Letters*, **52**, 2009.

Bateman, H. (1910*a*). On the conformal transformations of a space of four dimensions and their application to geometric optics. *Proceedings of the London Mathematical Society*, **7**, 70.

Bateman, H. (1910*b*). The transformation of the electrodynamical equations. *Proceedings of the London Mathematical Society*, **8**, 223.

Batygin, V. V. & I. N. Toptygin (1978). *Problems in Electrodynamics*. Translated from the second Russian edition 1970. London: Academic.

Belinfante, F. J. (1939). On the spin angular momentum of mesons. *Physics*, **6**, 887.

Belinfante, F. J. (1940). On the current and density of the electric charge, the energy, the linear momentum and the angular momentum of arbitrary fields. *Physics*, **7**, 449.

Berezin, F. A. (1987). *Introduction to Superanalysis*. Berlin: Springer.

Berezin, F. A. & M. S. Marinov (1975). Classical spin and Grassmannian variables. *JETP Letters*, **21**, 320.

Berezin, F. A. & M. S. Marinov (1977). Particle spin dynamics as the Grassmann variant of classical mechanics. *Annals of Physics*, **104**, 336.

Bilaniuk, O.-M. & E. C. G. Sudarshan (1969). Particles beyond the light barrier. *Physics Today*, **22** (# 5), 43.

Birkhoff, G. D. (1927). *Dynamical Systems*. Revised edition 1966. Providence: American Mathematical Society.

Blagoević, M. & P. Senjanović (1988). The quantum field theory of electric and magnetic charge. *Physics Reports*, **157**, 233.

Blokhintsev, D. I. & V. V. Orlov (1953). Propagation of signals in the Born–Infeld electrodynamics. *Žhurnal Éksperimental'noi i Teoretitcheskoi Fiziki*, **25**, 513 (in Russian).

Bogoliubov, N. N., A. A. Logunov, A. I. Oksak & I. T. Todorov (1990). *General Principles of Quantum Field Theory*. Translated from the Russian edition 1987. Dordrecht: Kluwer.

Bogoliubov, N. N. & O. A. Parasiuk (1957). Über die Multiplikation der Kausalfunktionen in der Quantentheorie der Felder. *Acta Mathematica*, **97**, 227.

Bogoliubov, N. N., B. V. Struminsky & A. N. Tavkhelidze (1965). On composite models in the elementary particle theory. Preprint JINR D-1968.

Bogomol'ny, E. B. (1976). Stability of classical solutions. *Soviet Journal of Nuclear Physics*, **24**, 449.

Boillat, G. (1970). Nonlinear electrodynamics: Lagrangians and equations of motion. *Journal of Mathematical Physics*, **11**, 941.

Bondi, H. (1967). *Assumption and Myth in Physical Theory.* Cambridge: Cambridge University Press.

Born, M. (1934). On the quantum theory of the electromagnetic field. *Proceedings of the Royal Society* (London), **143**, 410.

Born, M. & L. Infeld (1934). Foundations of the new field theory. *Proceedings of the Royal Society* (London), **A 144**, 425.

Boulware, D. G., L. S. Brown & R. D. Peccei (1970). Deep-inelastic electroproduction and conformal symmetry. *Physical Review*, **D 2**, 293.

Bourbaki, N. (1968). *Éléments de Mathématique, Groupes et algèbres de Lie.* Chapitres 4, 5 et 6, Vol. 34. Paris: Hermann.

Bremermann, H. (1966). *Distributions, Complex Variables, and Fourier Transforms.* 2nd edition 1983. Reading: Addison–Wesley.

Brown, L. M. (Editor) (1993). *Renormalization: From Lorentz to Landau (and beyond).* Berlin: Springer.

Brink, L., S. Deser, B. Zumino, P. Di Vecchia & P. Howe (1976). Local supersymmetry for spinning particles. *Physics Letters*, **B 64**, 435.

Bunge, M. (1959). *Causality. The Place of the Causal Principle in Modern Science.* Cambridge: Harvard University Press.

Byckling, E. & K. Kajantie (1973). *Particle Kinematics.* New York: Wiley.

Cabibbo, N. & E. Ferrari (1962). Quantum electrodynamics with Dirac monopoles. *Nuovo Cimento*, **23**, 1147.

Caldirola, P. (1956). A new model of the classical electron. *Nuovo Cimento*, **3**, Supplimento 2, 297.

Callan, C., S. Coleman & R. Jackiw (1970). A new improved energy-momentum tensor. *Annals of Physics*, **59**, 42.

Cao, T. Y. (1999). *Conceptual Foundations of Quantum Field Theory.* Cambridge: Cambridge University Press.

Cartan, É. (1924). Sur les variétés à connexion affine et la théorie de la relativité généralisée. *Annales Scientifiques de l'École Normale Supérieure*, **41**, 1.

Cartan, H. (1967). *Calcul différentiel formes differntielles.* Paris: Hermann.

Casalbuoni, R. (1976*a*). Relatively and supersymmetries. *Physics Letters*, **B 62**, 49.

Casalbuoni, R. (1976*b*). Pseudoclassical description of spinning particle. *Nuovo Cimento*, **A 33**, 115.

Cheng, T.-P. & L.-F. Li (1984). *Gauge Theory of Elementary Particle Physics.* Oxford: Clarendon.

Chew, G. F. & S. C. Frautschi (1961). Dynamical theory for strong interactions at low momentum transfers but arbitrary energies. *Physical Review*, **123**, 1478.

Close, F. E. (1979). *An Introduction to Quarks and Partons.* New York: Academic.

Coleman, S. (1966). The invariance of the vacuum is the invariance of world. *Journal of Mathematical Physics*, **7**, 787.

Coleman, S. (1983). The magnetic monopole fifty years later. In *The Unity of the Fundamental Interactions, Proceedings of the 19th Course of the International School of Subnuclear Physics.* Edited by A. Zichichi. London: Plenum, pp. 21-117.

Coleman, S. (1985). *Aspects of Symmetry: Selected Erice Lectures.* Cambridge: Cambridge University Press.

Corben, H. C. (1968). *Classical and Quantum Theories of Spinning Particles.* San Francisco: Holden-Day.

Coulomb, Ch. A. (1785). Second mémoire sur l'électricité et le magnétisme. Oú l'on determine, suivant quelles lois le fluide magnétique, ainsi que le fluide électrique, agissent, soit par répulsion, soit par attraction. *Histoire et Mémoires de l'Académie royale des sciences*, 1785, 612.

Courant, R. (1962). *Partial Differential Equations.* Methods of Mathematical Physics. By R. Courant & D. Hilbert. Vol. II. New York: Wiley.

Creutz, M. (1974). Confinement and critical dimensionality of space-time. *Physical Review Letters*, **43**, 553.

Creutz, M. (1983). *Quarks, gluons and lattices.* Cambridge: Cambridge University Press.

Cunningham, E. (1909). The principle of relativity in electrodynamics and an extension thereof. *Proceedings of the London Mathematical Society*, **8**, 77.

Denman, H. H. (1966). On linear friction in Lagrange equation. *American Journal of Physics*, **34**, 1147.

De Rham, G. (1955). *Variétés differentiables.* Paris: Hermann.

Deser, S., R. Jackiw, & S. Templeton (1982). Three-dimensional massive gauge theories. *Physical Review Letters*, **48**, 975.

Deser, S., R. Jackiw, & S. Templeton (1985). Topologically massive gauge theories. *Annals of Physics*, **149**, 372.

De Witt, B. S., (1962). Quantum theory without electromagnetic potentials. *Physical Review*, **125**, 2189.

De Witt, B. S., (1965). *Dynamical Theory of Groups and Fields.* New York: Gordon & Breach.

Dirac, P. A. M. (1928). The quantum theory of the electron. *Proceedings of the Royal Society* (London), **117**, 610.

Dirac, P. A. M. (1930). *The Principles of Quantum Mechanics.* 4th edition 1958. Oxford: Clarendon.

Dirac, P. A. M. (1931). Quantised singularities in the electromagnetic field. *Proceedings of the Royal Society* (London), **133**, 60.

Dirac, P. A. M. (1936). Wave equations in conformal space. *Annals of Mathematics*, **37**, 429.

Dirac, P. A. M. (1938). Classical theory of radiating electron. *Proceedings of the Royal Society* (London), **A167**, 148.

Dirac, P. A. M. (1948). The theory of magnetic poles. *Physical Review*, 2nd series, **74**, 817.

Dirac, P. A. M. (1964). *Lectures on Quantum Mechanics*. New York: Belfar.

Dothan, Y. M., M. Gell-Mann & Y. Ne'eman (1965). Series of hadron energy levels as representations of non-compact group. *Physics Letters*, **17**, 148.

Dresden, M. (1993). Renormalization in historical perspective – The first stage. In *Renormalization: From Lorentz to Landau (and beyond)*. Edited by L. M. Brown. Berlin: Springer, pp. 29-55.

Dubrovin, B. A., A. T. Fomenko & S. P. Novikov (1992). *Modern Geometry – Metods and Applications*. Translated from the second Russian edition 1986. Heidelberg: Springer.

Dyson, F. J. (1949). The S matrix in quantum electrodynamics. *Physical Review*, **75**, 1736.

Efimov, G. V. (1968). On a class of relativistic invariant distributions. *Communications in Mathematical Physics*, **7**, 138.

Efimov, G. V. (1970). Nonlocal quantum scalar field theory. *Soviet Journal of Particles and Nuclei*, **1**, 255.

Efimov, G. V. & M. A. Ivanov (1993). *The Quark Confinement Model of Hadrons*. London: The Institute of Physics Publishing.

Eguchi, T, P. B. Gilkey & A. J. Hanson (1980). Gravitation, gauge theories and differential geometry. *Physics Reports*, **66**, 213.

Ehrenfest, P. (1917). In what way does it become manifest in the fundamental laws of physics that space has three dimensions? *Proceedings of the Amsterdam Academy of Sciences*, **20**, 200.

Ehrenfest, P. (1920). Welche Rolle spielt die Dreidimensionalitat des Raumes in den Grundgesetzen der Physik? *Annalen der Physik*, **61**, 440.

Einstein, A. (1905*a*). Zur Elektrodynamik der bewegter Körper. *Annalen der Physik*, **17**, 891.

Einstein, A. (1905*b*). Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig? *Annalen der Physik*, **18**, 639.

Eliezer, C. J. (1947). The interaction of electrons and an electromagnetic field. *Reviews of Modern Physics*, **19**, 147.

Englert, F. & R. Brout (1964). Broken symmetry and the mass of gauge vector mesons. *Physical Review Letters*, **13**, 321.

Erber, T. (1961). The classical theory of radiation reaction. *Fortschritte der Physik*, **9**, 343.

Euler, L. (1744). *Methodus inveniendi Lineas Curvas maximi minimive proprietate gaudentes sive solutio problematis isoperimetrici latissimo sensu accepti*. Lausannae et Geneve: Bousquet.

Faddeev, L. & A. Slavnov (1980). *Gauge Fields*. Translated from the Russian edition 1978. Reading: Benjamin-Cummings.

Faraday, M. (1839). *Experimental Researches in Electricity.* London: Taylor.

Fock, V. (1926). Über die invariante Form der Wellen- und der Bewegungsgleichungen für einen geladenen Massenpunkt. *Zeitschrift für Physik*, **39**, 226.

Fokker, A. D. (1929*a*). Ein invarianter Variationssatz für die Bewegung meherer elektrischer Massenteilchen. *Zeitschrift für Physik*, **58**, 386. English translation is available at http://www.nonloco-physics.000freehosting.com/.

Fokker, A. D. (1929*b*). Wederkeerigheid in de Werking van Geladen Deeltjes. *Physica*, **9**, 33. English translation is available at http://www.nonloco-physics.000freehosting.com/.

Fradkin, E. S., & A. A. Tseytlin (1985). Non-linear electrodynamics from quantized strings. *Physics Letters*, **B 163**, 123.

Frank, P. (1932). *Das Kausalgesetz und seine Grenzen.* Vienna: Springer.

Frenkel, J. (1925). Zur Elektrodynamik punktförmiger Elektronen. *Zeitschrift für Physik*, **32**, 518.

Frenkel, J. (1926). Die Elektrodynamik des Rotierenden Electrons. *Zeitschrift für Physik*, **37**, 243.

Fritzsch, H., M. Gell-Mann & H. Leutwyler (1973). Advantages of the color octet gluon picture. *Physics Letters*, **B 47**, 365.

Fulton, T., F. Rohrlich & L. Witten (1962). Conformal invariance in physics. *Reviews of Modern Physics*, **34**, 442.

Fushchich, W. I. & A. G. Nikitin (1987). *Symmetries of Maxwell's Equations.* Dordrecht: Reidel.

Gaillard, M. K. & B. Zumino (1981). Duality rotations for interacting fields. *Nuclear Physics*, **B 193**, 221.

Galilei, G. (1632). *Dialogo dei due massimi sistemi del mundo.* Florence: Landini. English translation: *Dialogues Concerning Two New Sciences.* Norwich: Andrew.

Galvao, C. A. P. & C. Teitelboim (1980). Classical supersymmetric particles. *Journal of Mathematical Physics*, **21**, 1863.

Gates, S. J., M. T. Grisaru, M. Roček & W. Siegel (1983). *Superspace: Or One Thousand and One Lessons in Supersymmetry.* Reading: Benjamin-Cummings. E-print hep-th/0108200.

Gauss, C. F. (1813). *Theoria attractionis corporum spheroidicorum ellipticorum homogeneorum methodo nova tractata.* Göttingen.

Gel'fand, I. M., R. A. Minlos & Z. Ya. Shapiro (1963). *Representations of the Rotational and Lorentz Groups and their Applications.* Translated from the Russian edition 1958. Oxford: Pergamon.

Gel'fand, I. M. & G. E. Shilov (1964). *Generalized Functions.* Vol. I. Translated from the second Russian edition 1959. Volume II (1968). Translated from the Russian edition 1958. New York: Academic.

Gell-Mann, M. (1964). A schematic model of baryons and mesons. *Physics Letters*, **8**, 214.

Gell-Mann, M. & M. Levy (1960). The axial vector current in beta decay. *Nuovo Cimento*, **16**, 705.

Georgi, H. (1982). *Lie Algebras in Particle Physics: From Isospin to Unified Theories*. Reading: Benjamin-Cummings.

Gervais, J.-L. & B. Sakita (1971). Generalization of dual models. *Nuclear Physics*, **B 34**, 477.

Gibbons, G. V. (2003). Aspects of Born–Infeld theory and string/M-theory. *Revista Mexicana de Fisica*, **49 S1**, 19. E-print hep-th/0106059.

Ginzburg, V. L. & L. D. Landau (1950). On the theory of superconductivity. *Zhurnal Ěksperimental'noi i Teoretitcheskoi Fiziki*, **20**, 1064 (in Russian).

Gitman, D. M. & I. V. Tyutin (1990). *Quantization of Fields with Constraints*. Translated from the Russian edition 1986. Heidelberg: Springer.

Glashow, S. (1961). Partial-symmetries of weak interactions. *Nuclear Physics*, **22**, 579.

Godbillon, C. (1969). *Geometrie différentielle et mécanique analytique*. Collection méthodes. Paris: Hermann.

Goddard, P. & D. I. Olive (1978). Magnetic monopoles in gauge field theories. *Reports on Progress in Physics*, **41**, 1357.

Goldstein, H. (1950). *Classical Mechanics*. 2nd edition 1980. Reading: Addison–Wesley.

Goldstone, J. (1961). Field theories of 'superconductor' solutions. *Nuovo Cimento*, **19**, 154.

Gol'fand, Yu. A. & E. P. Likhtman (1971). Extension of the generator algebra of the Poincaré group and breakdown of P-invariance. *JETP Letters*, **13**, 323.

Gordon, W. (1926). Der Compton Effekt nach der Schrödingerschen Theorie. *Zeitschrift für Physik*, **40**, 117.

Goto, T. (1971). Relativistic quantum mechanics of a one-dimensional mechanical continuum and subsidiary condition of dual resonance model. *Progress of Theoretical Physics*, **46**, 1560.

Green, G. (1828). An essay on the application of mathematical analysis to the theories of electricity and magnetism. Nottingham.

Green, M. B., J. H. Schwarz & E. Witten (1987). *Superstring Theory*. Vols. 1 and 2. Cambridge: Cambridge University Press.

Greenberg (1964). Spin and unitary-spin independence in a paraquark model of barions and mesons. *Physical Review Letters*, **13**, 598.

Greene, B. (1999). *The Elegant Universe*. London: Vintage.

Gross, D. & F. Wilczek (1973). Ultraviolet behavior of non-Abelian gauge theories, *Physical Review Letters*, **30**, 1343.

Grünbaum, A. (1963). *Philosophical Problems of Space and Time*. New York: Knopf. 2nd enlarged edition 1973. Dordrecht: Reidel.

Guralnik, G. C., C. R. Hagen & T. W. B. Kibble (1964). Global conservation laws and massless particles. *Physical Review Letters*, **13**, 585.

Haag, R. (1955). Die Selbstwechselwirkung des Elektrons. *Zeitschrift für Naturforschung*, **10a**, 752.

Hamilton, W. R. (1834). On a general method in dynamics, by which the study of the motions of all free systems of attracting or repelling points is reduced to the search and differentiation of the central relation or characteristic function. *Philosophical Transactions of the Royal Society*, London. **2**, 247.

Han, M. Y. & Y. Nambu (1965). Three-triplet model with double $SU(3)$ symmetry. *Physical Review*, **B 139**, 1006.

Hanson, A. J. & T. Regge (1974). The relativistic spherical top. *Annals of Physics*, **87**, 498.

Hanson, A. J., T. Regge & C. Teitelboim (1976). *Constrained Hamiltonian Systems.* Rome: Accademia Nazionale dei Lincei.

Hara, O. (1971). On origin and physical meaning of Ward-like identity in dual-resonance model. *Progress of Theoretical Physics*, **46**, 1549.

Hasenfratz, P. & G. 't Hooft (1976). Fermion-boson puzzle in a gauge theory. *Physical Review Letters*, **36**, 1119.

Hawking, S. W. & G. F. R. Ellis (1973). *The Large Scale of Spacetime.* Cambridge: Cambridge University Press.

Heaviside, O. (1889). On the electromagnetic effects due to the motion of electrification through a dielectric. *Philosophical Magazine*, **27**, 324.

Heaviside, O. (1892). *Electrical Papers.* Vols. I and II. London: Macmillan.

Heaviside, O. (1893, 1899, 1912). *Electromagnetic Theory.* Vols. I, II, and III. London: The Electrician Co. 2nd edition 1922. London: Benn. 3rd edition 1951. London: Spon.

Heaviside, O. (1902). The waste of energy from a moving electron. *Nature*, **67**, 6.

Hehl, F. W. & Y. N. Obukhov (2003). *Foundation of Classical Electrodynamics: Charge, Flux, and Metric.* Boston: Birkhäuser.

Heinz, U. (1984). A relativistic colored spinning particle in an external color field. *Physics Letters*, **B 144**, 228.

Helmholtz, H. (1858). Ueber Integrale der hydrodynamischen Gleichungen, welche den Wirbelbewegungen entsprechen. *Journal für die reine und angewandte Mathematik*, **55**, 25.

Hertz, H. (1887). Über sehr schnelle elektrische Schwingungen. *Annalen der Physik*, **31**, 421.

Hertz, H. (1888). Über die Einwirkung einer geradlinigen elektrischen Schwingung auf eine benachbarte Strombahn. *Annalen der Physik*, **34**, 155.

Hestenes, D. (1986). *New Foundation of Classical Mechanics.* 3rd edition 1990. Dordrecht: Kluwer.

Higgs, P. S. (1964). Broken symmetries, massless particles and gauge fields. *Physics Letters*, **12**, 132.

Higgs, P. S. (1966). Spontaneous symmetry breakdown without massless bosons. *Physical Review*, **145**, 1156.

Hilbert, D. (1915). Grundlagen der Physik. *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen.* Mathematisch-Physikalische Klasse, **3**, 395.

Hill, E. L. (1951). Hamilton's principle and the conservation theorems of mathematical physics. *Reviews of Modern Physics*, **23**, 253.

Hoyle, F. & J. V. Narlikar (1995). Cosmology and action-at-a-distance electrodynamics. *Reviews of Modern Physics*, **67**, 113.

Hoyle, F. & J. V. Narlikar (1996). *Lectures on Cosmology and Action at a Distance Electrodynamics*. Singapore: World Scientific.

Huang, K. (1952). On the Zitterbewegung of the Dirac electron. *American Journal of Physics*, **20**, 479.

Iofa, M. Z. & V. Ya. Fainberg (1969). Wightman formulation for nonlocalized field theory. *Soviet Physics – JETP*, **56**, 1644.

Iwanenko, D. & A. Sokolow (1953). *Die Klassische Feldtheorie*. Translated from the Russian edition 1951. Berlin: Akademie.

Jackiw, R. & N. Manton (1980). Symmetries and conservation laws in gauge theories. *Annals of Physics*, **127**, 257.

Jackiw, R. & C. Rebbi (1976). Spin from isospin in a gauge theory. *Physical Review Letters*, **36**, 1116.

Jackiw, R. & P. Rossi (1980). Stability and bifurcation in Yang–Mills theory. *Physical Review*, **D 21**, 426.

Jackson, J. D. (1962). *Classical Electrodynamics*. 2nd edition 1975, 3rd edition 1998. New York: Willey.

Jackson, J. D. & L. B. Okun (2001). Historical roots of gauge invariance. *Reviews of Modern Physics*, **73**, 663.

Jaffe, A. (1967). High energy behavior in quantum field theory, I. Strictly localizable fields. *Physical Review*, **158**, 1454.

Jaffe, A. & J. Glimm (1987). *Quantum Physics*. Heidelberg: Springer.

Jaffe, A. & C. M. Taubes (1980). *Vortices and Monopoles*. Boston: Birkhäuser.

Jaffe, A. & E. Witten (1999). Quantum Yang–Mills theory. Available at http://www.claymath.org/prize-problems/.

Jammer, M. (1954). *Concepts of Space – A History of the Theories of Space in Physics*. Cambridge: Harvard University Press. 3rd edition 1993. New York: Dover.

Jammer, M. (1961). *Concepts of Mass in Classical and Modern Physics*. Cambridge: Harvard University Press.

Jammer, M. (1962). *Concepts of Force*. New York: Harper and Brothers.

Johnson, C. V. (2003). *D-Branes*. Cambridge: Cambridge University Press.

Julia, B. & A. Zee (1975). Poles with both magnetic and electric charges in non-Abelian gauge theory. *Physical Review*, **D 11**, 2227.

Källen, G. (1964). *Elementary Particle Physics*. Reading: Addison–Wesley.

Kastrup, H. A. (1962). Zur physikalischen Deutung und darstellungstheoretischen Analyse der konformen Transformationen von Raum und Zeit. *Annalen der Physik*, **7**, 388.

Kiritsis, E. (1998). *Introduction to Superstring Theory*. Leuven: Leuven University Press. E-print hep-th/9709062.

Kirzhnits, D. A. (1972). Weinberg model and the 'hot' universe. *JETP Letters*, **15**, 745.

Kirzhnits, D. A. & A. D. Linde (1972). Macroscopic consequences of the Weinberg model. *Physics Letters*, **B 42**, 471.

Klapdor-Kleingrothaus, H. V. & K. Zuber (1997). *Teilchenastrophysik.* Stuttgart: Teubner. English translation 1999: *Particle Astrophysics.* London: The Institute of Physics Publishing.

Klein, F. (1872). *Programm zum Eintritt in die philosophische Fakultät der Universität zu Erlangen.* Erlangen: Deichert. Reprinted in *Mathematische Annalen*, **43**, 63, 1893.

Klein, O. (1926). Quantum Theorie und fünfdimensionale Relativitätstheorie. *Zeitschrift für Physik*, **37**, 895.

Klein, O. (1938). On the theory of charged fields. In *New Theories in Physics.* Conference organized by the International Union of Physics and the Polish Intellectual Co-operation Committee. Warsaw, 30 May-3 June, 1938. Paris: International Institute of Intellectual Co-operation (Scientific Collection), pp. 77-93.

Konopleva, N. & V. Popov (1981). *Gauge Fields.* Translated from the Russian edition 1972. Chur, Switzerand: Harwood Academic.

Kosyakov, B. P. (1976). Finite self-interaction of the classical electromagnetic field. *Theoretical and Mathematical Physics*, **27**, 423.

Kosyakov, B. P. (1991). Field of arbitrarily moving colored charge. *Theoretical and Mathematical Physics*, **87**, 632.

Kosyakov, B. P. (1992). Radiation in electrodynamics and the Yang–Mills theory. *Soviet Physics – Uspekhi*, **35**, 135.

Kosyakov, B. P. (1994). Exact solutions of the Yang–Mills equations with the source in the form of two point color charges. *Theoretical and Mathematical Physics*, **99**, 409.

Kosyakov, B. P. (1998). Exact solutions in the Yang–Mills–Wong theory. *Physical Review*, **D 57**, 5032. E-print hep-th/9902039.

Kosyakov, B. P. (1999). Exact solutions of classical electrodynamics and the Yang–Mills– Wong theory in even-dimensional space-time. *Theoretical and Mathematical Physics*, **119**, 493. E-print hep-th/0207217.

Kosyakov, B. P. (2001). Physical meaning of renormalizability. *Physics of Particles and Nuclei*, **32**, 488. E-print hep-th/0011235.

Kosyakov, B. P. (2003). On the inert properties of particles in classical theory. *Physics of Particles and Nuclei*, **34**, 808. E-print hep-th/0208035.

Kotkin, G. L. & V. G. Serbo (1971). *Collection of Problems in Classical Mechanics.* Translated from the Russian edition 1969. Oxford: Pergamon.

Kudar, J. (1926). Zur vierdimensionalen Formulierung der undulatorischen Mechanik. *Annalen der Physik*, **81**, 632.

Lagrange, J. L. (1788). *Mécanique Analytique.* Paris: Desaint.

Lanczos, C. (1929). Die tensoranalytischen Beziehunger der Diracschen Gleichung. *Zeitschrift für Physik*, **57**, 447.

Lanczos, C. (1949). *The Variational Principles of Mechanics.* 4nd edition 1970. Toronto: University of Toronto Press.

Landau, L. D. & E. M. Lifshitz (1971). *The Classical Theory of Fields.* Translated from the 5th Russian edition 1967. New York: Addison–Wesley.

Larmor, J. (1897). On the theory of magnetic influence on spectra; and on the radiation from moving ions. *Philosophical Magazine*, **44**, 503.

Larmor, J. (1900). *Ether and Matter.* Cambridge: Cambridge University Press.

Lepage, G. P. (2005). High-precision nonperturbative QCD. *Annals of Physics*, **315**, 193.

Lie, S. (1888, 1890, 1893). *Theorie der Transformationsgruppen.* Unter Mitwirkung von F. Engel, in three volumes. Leipzig: Teubner.

Liénard, A. (1898). Champ électrique et champ magnétique produit par une charge électrique concentrée en un point et animée d'un mouvement quelconque *L'Éclairage électrique*, **16**, 5; 53; 106.

Lightman, A. P., W. H. Press, R. H. Price & S. A. Teukolsky (1975). *Problem Book in Relativity and Gravitation.* Princeton: Princeton University Press.

Lions, L. (1985). Quark search experiments at accelerators and in cosmic rays. *Physics Reports*, **129**, 225.

London, F. (1927). Quantenmechanische Deutung der Theorie von Weyl. *Zeitschrift für Physik*, **42**, 375.

López, C. A. & D. Villarroel (1975). Bound and emitted angular momenta of a classical accelerated point charge. *Physical Review*, **D 11**, 2724.

Lorentz, H. A. (1892). La théorie électromagnétique de Maxwell et son application aux corps mouvemants. *Archives Néerlandaises des sciences exastes et naturelles*, **25**, 363.

Lorentz, H. A. (1904*a*). Weiterbildung der Maxwellischen Theorie: Elektronentheorie. *Encyklopädie der Mathetische Wissenschaften*, **2**, 145.

Lorentz, H. A. (1904*b*). Electromagnetic phenomena in a system moving with any velocity smaller than that of light. *Proceedings of the Amsterdam Academy of Sciences*, **6**, 809.

Lorentz, H. A. (1909). *The Theory of Electrons and its Applications to the Phenomena of Light and Radiant Heat.* 2nd edition 1915. Leipzig: Teubner.

Lorenz, L. V. (1867). Über die Identität der Schwingungen des Lichts mit den elektrischen Strömen. *Annalen der Physik und Chemie*, **131**, 243; On the identity of the vibrations of light with electrical currents. *Philosophical Magazine*, Series 4, **34**, 287.

Mach, E. (1883). *Die Mechanik in ihrer Entwickelung historisch-kritisch dargestellt.* 5th edition 1904. Leipzig: Brockhaus. English translation 1942: *The Science of Mechanics.* London: Open Court.

Madore, J. (1995). *An Introduction to Noncommutative Differential Geometry and its Physical Applications.* 2nd edition 1999. Cambridge: Cambridge University Press.

Mandelstam, S. (1958). Determination of the pion-nucleon scattering amplitude from dispersion relations and unitarity: General Theory. *Physical Review*, **112**, 1344.

Mandelstam, S. (1962). Quantum electrodynamics without potentials. *Annals of Physics*, **19**, 1.

Mandelstam, S. (1980). General introduction to confinement. *Physics Reports*, **67**, 109.

Mandula, J. E. (1976). Classical Yang–Mills potentials. *Physical Review*, **D 14**, 3497.

Manton, N. S. (1977). The forces between 't Hooft–Polyakov monopoles. *Nuclear Physics*, **B 126**, 525.

Markov, M. A. (1939). Note on quantum field theory. *Doklady Akademii Nauk SSSR*, **24**, 233 (in Russian).

Marsden, J. E. & T. S. Ratiu (1994). *Introduction to Mechanics and Symmetry.* 2nd edition 1998. Berlin: Springer.

Marshal, C. (1990). *The Three-Body Problem.* 2nd edition 2004. New York: Elsevier.

Martin, J. L. (1959). Generalized classical dynamics, and the 'classical analogue' of a Fermi oscillator. *Proceedings of the Royal Society* (London), **251**, 536.

Maxwell, J. C. (1873). *A Treatise on Electricity and Magnetism.* 3rd edition in two volumes, with supplementary notes by J. J. Thomson, 1891. Oxford: Clarendon.

Meiman, N. N. (1964). The causality principle and asymptitic behavior of the scattering amplitude. *Soviet Physics JETP*, **47**, 188.

Meyer-Ortmanns, H. (1996). Phase transitions in quantum chromodynamics. *Reviews of Modern Physics*, **68**, 473.

Mie, G. (1912). Grundlagen einer Theorie der Materie. *Annalen der Physik*, **37**, 511; **39**, 1.

Minkowski, H. (1909). Raum und Zeit. *Physikalische Zeitschrift*, **10**, 104.

Misner, C. W., K. S. Thorne & J. A. Wheeler (1973). *Gravitation.* San Francisco: Freeman.

Mohapatra, R. N. (1986). *Unification and Supersymmetry.* 3rd edition 2003. Berlin: Springer

Morse, P. M. & H. Feshbach (1953). *Methods of Mathematical Physics.* New York: McGraw-Hill.

Naimark, M. A., (1964). *Linear Representations of the Lorentz Group.* Translated from the Russian edition 1958. Oxford: Pergamon.

Nambu, Y. (1970*a*). Quark models and the factorization of the Veneziano amplitude. In *Symmetries in Quark Model.* Edited by R. Chand. New York: Gordon & Breach, pp. 269-279.

Nambu, Y. (1970*b*). Duality and Hydrodynamics. Lectures at the Copenhagen symposium.

Nambu, Y. (1974). Strings, monopoles and gauge fields. *Physical Review*, **D 10**, 4262.

Nambu, Y. (1976). The confinement of quarks. *Scientific American*, **235** (# 5), 48.

Nambu, Y. & G. Jona-Lasinio (1961). Dynamical model of elementary particles based on an analogy with superconductivity. I, *Physical Review.* **122**, 345.

Ne'eman, Y. & Dj. Šijački (1985). $\overline{\mathrm{SL}}(4, R)$ classification for hadrons. *Physics Letters*, **B 157**, 267.

Ne'eman, Y. & Dj. Šijački (1988). Hadrons in an $\overline{\mathrm{SL}}(4, R)$ classification. *Physical Review*, **D 37**, 3267.

Ne'eman, Y. & Dj. Šijački (1993). Hadrons in an $\overline{\mathrm{SL}}(4, R)$ classification. II. Mesons and C, P assignments. *Physical Review*, **D 47**, 4133.

Neumann, C. (1870). *Die Prinzipien der Galilei–Newton'schen Theorie.* Leipzig: Teubner.

Neveu, A. & J. H. Schwarz (1971). Factorizable dual model of pions. *Nuclear Physics*, **B 31**, 86.

Newton, I. (1686). *Philosophiæ Naturalis Principia Mathematica.* London: Streater. English translation (1962): *Mathematical Principles of Natural Philosophy.* Berkeley: University of California Press.

Nielsen, H. B. & P. Olesen (1973). Vortex-line models for dual strings. *Nuclear Physics*, **B 61**, 45.

Noether, E. (1918). Invariante Variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen.* Mathematisch-Physikalische Klasse, **2**, 235.

Ogiyevetsky, V. I. & E. Sokachev (1975). Primitive representations of the $\mathrm{sl}(3, R)$ algebra. *Theoretical and Mathematical Physics*, **23**, 845.

Ostrogradsky, M. (1831). Note sur la theórie de la chaleur. (Lu le 5 Novembre 1828). *Mémoires de l'Académie Impériale des Sciences de Saint-Pétersbourg*, VI$^{\mathrm{me}}$ série, mathematique, physique, et naturelle, **1**, 129.

Ostrogradsky, M. V. (1838). Mémoire sur le calcul des variations de intégrales multiples. (Lu le 24 janvier 1834). *Mémoires de l'Académie Impériale des Sciences de Saint-Pétersbourg*, VI$^{\mathrm{me}}$ série, mathematique, physique, et naturelle, **1**, livr. 2.

Ostrogradsky, M. V. (1850). Mémoire sur les équations différentielles relatives aux problèmes des isopérimètres. (Lu le 29 Novembre 1848). *Mémoires de l'Académie Impériale des Sciences de Saint-Pétersbourg*, VI$^{\mathrm{me}}$ série, mathematique, physique, et naturelle, **4**, 385.

Pais, A. (1948). *Developments in the Theory of the Electron.* Princeton: Princeton University Press.

Pais, A. (1982). *"Subtle is the Lord..." The Science and the Life of Albert Einstein.* Oxford: Oxford University Press.

Pais, A. & G. E. Uhlenbeck (1950). On field theories with non-localized action. *Physical Review*, **79**, 145.

Palmieri, C. & B. Vitale (1970). On the inversion of Noether's theorem in the Lagrangian formalism. *Nuovo Cimento*, **66 A**, 299.

Paley, R. E. A. C. & N. Wiener (1934). *Fourier Transforms in the Complex Domain.* New York: American Mathematical Society Publications.

Parisi, G. (1975). Quark imprisonment and vacuum repulsion. *Physical Review*, **D 11**, 970.

Parisi, G. & N. Sourlas (1979). Random magnetic fields, supersymmetry, and negative dimensions. *Physical Review Letters*, **43**, 744.

Parrot, S. (1987). *Relativistic Electrodynamics and Differential Geometry.* Berlin: Springer.

Pauli, W. (1936). Contributions mathématiques à la théorie des matrices de Dirac. *Annales de l'Institut Henri Poincaré*, **6**, 109.

Pauli, W. (1958). *Theory of Relativity.* Translated from an article 'Relativitätstheorie' in the *Enzyklopädie der Mathematischen Wissenschaften*, Band 5, Leipzig: Teubner, 1918, with supplementary notes by the author. Oxford: Pergamon.

Pavšič, M. & V. Tapia (2001). Resource letter on geometrical results for embeddings and branes. E-print gr-qc/0010045.

Pegg, D. T. (1975). Absorber theory of radiation. *Reports on Progress in Physics*, **38**, 1339.

Perkins, H. D. (1972). *Introduction to High Energy Physics.* Reading: Addison–Wesley. 4th edition 2000. Cambridge: Cambridge University Press.

Peshkin, M. & A. Tonomura (1989). *The Aharonov–Bohm Effect.* Berlin: Springer.

Planck, M. (1906). Das Prinzip der Relativität und die Grundgleichungen der Mechanik. *Verhandlungen der deutschen physikalische Gesellschaft*, **8**, 136.

Planck, M. (1908). Bemerkungen zum Prinzip der Aktion Reaktion in der allgemeinen Dynamik. *Verhandlungen der deutschen physikalische Gesellschaft*, **10**, 728.

Plass, G. N. (1961). Classical electrodynamic equations of motions with radiative reaction. *Reviews of Modern Physics*, **33**, 37.

Poincaré, H. (1896). Remarques sur une expérience de M. Birkeland. *Comptes Rendus*, **123**, 530.

Poincaré, H. (1898). La mesure du temps. *Revue de métaphysique et de morale*, **6**, 1.

Poincaré, H. (1900). La théorie de Lorentz et le principe de réaction. *Archives Néerlandaises des sciences exastes et naturelles*, **5**, 252.

Poincaré, H. (1902). *Science et hypothèse.* Paris: Flammarion.

Poincaré, H. (1904). L'etat actuel et l'avenir de la Physique mathematiqie. *Bulletin des Sciences Mathematiques*, **28**, 302.

Poincaré, H. (1905). Sur la dynamique de l'électron. *Comptes Rendus*, **140**, 1504.

Poincaré, H. (1906). Sur la dynamique de l'électron. *Rendiconti del Circolo Matematico di Palermo*, **21**, 129.

Polchinski, J. (1999). *String Theory.* Volume 1. Cambridge: Cambridge University Press.

Politzer, H. D. (1973). Reliable perturbative results for strong interactions? *Physical Review Letters*, **30**, 1346.

Polyakov, A. M. (1974). Particle spectrum in quantum field theory. *JETP Letters*, **20**, 194.

Pontryagin, L. S. (1939). *Topological Groups.* Translated from the Russian edition 1938. 2nd edition 1966. New York: Gordon & Breach.

Poynting, J. H. (1884*a*). On the transfer of energy in the electromagnetic field. *Philosophical Transactions of the Royal Society*, London. **175**, 343.

Poynting, J. H. (1884*b*). On the connexion between electric current and electric and magnetic inductions in the surrounding field. *Philosophical Transactions of the Royal Society*, London. **176**, 277.

Prasad, M. K. & C. M. Sommerfield (1975). Exact classical solution for the 't Hooft monopole and the Julia–Zee dyon. *Physical Review Letters*, **35**, 760.

Proca, A. (1936). Sur la théorie ondulatoire des électrons positifs et négatifs. *Le journal de physique et le radium.* **7**, 347.

Rafanelli, K. (1984). Theory of the relativistic spinning particle: Hamiltonian formulation and world-line invariance. *Physical Review*, **D 30**, 1707.

Rainich, G. Y. (1925). Electrodynamics in general relativity. *Transactions of the American Mathematical Society*, **27**, 106.

Rajaraman, R. (1982). *Solitons and Instantons in Quantum Field Theory.* Amsterdam: Noth-Holland.

Ramond, P. (1971). Dual theory for free fermions. *Physical Review*, **D 3**, 2415.

Rarita, W. & J. Schwinger (1941). On a theory of particles with half-integer spin. *Physical Review*, **60**, 61.

Reichenbach, H. (1958). *The Philosophy of Space and Time.* New York: Dover.

Rivas, M. (2001). *Kinematical Theory of Spinning Particles.* Dordrecht: Kluwer.

Rohrlich, F. (1960). Self energy and stability of the classical electron. *American Journal of Physics*, **28**, 639.

Rohrlich, F. (1965). *Classical Charged Particles.* 2nd edition 1990. Redwood City: Addison–Wesley.

Rosen, J. (1968). The conformal group and causality. *Annals of Physics*, **47**, 468.

Rosenfeld, L. (1940). Sur le tenseur d'impulsion-énergie. *Mémoires de l'Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique*, **18**, 1.

Rubakov, V. (2002). *Classical Theory of Gauge Fields.* Translated from the Russian edition 1999. Princeton: Princeton University Press.

Salam, A. (1968). Weak and electromagnetic interactions. In *Proceedings of the 8th Nobel Symposium.* Edited by N. Svartholm. Stockholm: Almqvist & Wiksell, pp. 367-377.

Salam, A. & J. C. Ward (1964). Electromagnetic and weak interactions. *Physics Letters*, **13**, 168.

Salingaros, N. (1985). Particle in external electromagnetic field. II. The exact velocity in a constant and uniform field. *Physical Review*, **D 31**, 3150.

Scherk, J. & J. H. Schwarz (1974). Dual models for non-hadrons. *Nuclear Physics*, **B 81**, 118.

Schott, G. A. (1912). *Electromagnetic Radiation.* Cambridge: Cambridge University Press.

Schott, G. A. (1915). On the motion of the electron. *Philosophical Magazine*, **29**, 49.

Schouten, J. A. (1949). On meson fields and conformal transformations. *Reviews of Modern Physics*, **21**, 421.

Schrödinger, E. (1930). Über die kräftfreie Bewegung in der relativistischen Quantenmechanik. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin.* Physikalish-mathematische Klasse, **24**, 418.

Schwartz, L. (1967). *Analyse Mathématique.* Cours professé à l'Ecole Polytechnique. Vol. II. Paris: Hermann.

Schwartz, L. (1950–1951). *Thèorie des distributions* I, II, nouvelle èdition 1973. Paris: Hermann.

Schwarz, A. S. (1991). *Quantum Field Theory and Topology.* Translated from the Russian edition 1989. Berlin: Springer.

Schwarzschild, K. (1903). Zur Elektrodynamik: I. Zwei Formen des Prinzips der kleinsten Action in der Elektronentheorie. II. Die elementare elektrodynamische Kraft. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen.* Mathematisch-Physikalische Klasse, **128**, 126; **128**, 132. English translation is available at http://www.nonlocophysics.000freehosting.com/.

Schweber, S. S. (1961). *An Introduction to Relativistic Quantum Field Theory.* New York: Row.

Schweber, S. S. (1994). *QED and the Men Who Made It: Dyson, Feynman, Schwinger and Tomonaga.* Princeton: Princeton University Press.

Schwinger, J. (1959). Euclidean quantum electrodynamics. *Physical Review*, **115**, 721.

Seiberg, N. & E. Witten (1999). String theory and noncommutative geometry. *Journal of High Energy Physics*, **09**, 032. E-print: hep-th/09908142.

Shaw, R. (1955). The problem of particle types and other contributions to the theory of elementary particles. Ph.D. Thesis. University of Cambridge.

Siegel, W. (1999). *Fields.* E-print: hep-th/9912205.

Snyder, H. (1947). Quantized space-time. *Physical Review*, **71**, 8.

Sobolev, S. L. (1936). Méthode nouvelle à résoudre le problème de Cauchy pour les équations linéaires hyperboliques normales. *Matematitcheskiĭ Sbornik*, **1** (43), 39.

Sokolov, A. A. & I. M. Ternov (1986). *Radiation from Relativistic Electrons.* Translated from the second Russian edition 1983. New York: AIP Translation Series.

Sommerfeld, A. (1948). *Elektrodynamik.* Vorlesungen über theoretische Physik. Band 3. Wiesbaden: Diesterich'sche Verlagsbuchhandlung.

Sorg, M. (1974). Classical, covariant finite-size model of the radiating electron. *Zeitschrift für Naturforschung*, **29a**, 1671.

Spivak, M. (1965). *Calculus on Manifolds.* New York: Benjamin.

Spivak, M. (1974). *A Comprehensive Treatise on Differential Geometry*, in five volumes. 2nd edition 1979. Boston: Publish or Perrish.

Stelle, K. S. (1977). Renormalization of higher derivative quantum gravity. *Physical Review*, **D 16**, 953.

Stelle, K. S. (1978). Classical gravity with higher derivatives. *General Relativity and Gravitation*, **9**, 353.

Stiegler, U. (1996). The nucleon spin structure. *Physics Reports*, **277**, 1.

Stokes, G. G. (1849). On some cases of fluid motion. *Transactions of the Cambridge Philosophical Society*, **8**, 105.

Strazhev, V. I. & L. M. Tomil'chik (1973). Current state of the Dirac monopole problem. *Soviet Journal of Particles and Nuclei*, **4**, 78.

Sudarshan, E. C. G. & N. Mukunda (1973). *Classical Dynamics: A Modern Perspective*. New York: Wiley.

Synge, J. L. (1940). On the electromagnetic two-body problem. *Proceedings of the Royal Society*, (London) **A 177**, 118.

Synge, J. L. (1956). *Relativity: The Special Theory*. 2nd edition 1972. Amsterdam: North Holland.

Synge, J. L. (1960). *Relativity: The General Theory*. Amsterdam: North Holland.

Synge, J. L. (1970). Point particles and energy tensor in special relativity. *Annali di Matematica Pura ed Applicata*, **84**, 33.

Szebehely, V. (1967). *Theory of Orbits. The Restricted Problem of Three Bodies*. New York: Academic.

Takahashi, Y. (1969). *An Introduction to Field Quantization*. Oxford: Pergamon.

Taub, A. H. (1948). Orbits of charged particles in constant fields. *Physical Review*, **73**, 786.

Teitelboim, C. (1970). Splitting of Maxwell tensor: Radiation reaction without advanced fields. *Physical Review*, **D 1**, 1572.

Teitelboim, C., D. Villarroel & Ch. G. van Weert (1980). Classical electrodynamics of retarded fields and point particles. *Rivista del Nuovo Cimento*, **3**, series 3, 1.

Tetrode, H. (1922). Über den Wirkungzusammenhang der Welt. Eine Erweiterung der klassischen Dynamik. *Zeitschrift für Physik*, **10**, 317. English translation is available at: www.nonloco-physics.000freehosting.com/.

Thomson, J. J. (1881). On the electric and magnetic effects produced by the motion of electrified bodies. *Philosophical Magazine*, **11**, 227.

Thomson, J. J. (1897). Cathode rays. *Philosophical Magazine*, **44**, 294.

't Hooft, G. (1974). A magnetic monopole in the Yang–Mills theory. *Nuclear Physics*, **B 79**, 276.

't Hooft, G. (Editor) (2004). *50 Years of Yang–Mills theory*. Singapore: World Scientific.

't Hooft, G. & M. Veltman (1972). Renormalization and regularization of gauge fields. *Nuclear Physics*, **B 44**, 189.

Titchmarsh, E. C. (1932). *The Theory of Functions*. 2nd edition 1939. Oxford: Oxford University Press.

Tiwari, S. C. (2003). *Superluminal Phenomena in Modern Perspective – Faster-than-light Signals: Myth or Reality?* New York: Rinton.

Umezawa, H. (1965). Dynamical rearrangement of symmetries. I. The Nambu–Heisenberg model. *Nuovo Cimento*, **A 40**, 450.

Umow, N. (1874). Ableitung der Bewegungsgleichungen der Energie in continuirlichen Körpern. *Zeitschrift der Mathematik und Physik* (Schomilch), **XIX**, 418.

Van der Vaart, H. R. (1967). Variational principles for certain nonconservative systems. *American Journal of Physics*, **35**, 419.

Velo, G. & D. Zwanziger (1969). Propagation and quantization of Rarita–Schwinger waves in an external electromagnetic potential. *Physical Review*, **186**, 1337.

Veltman, M. J. G. (2003). *Facts and Mysteries in Elementary Particle Physics.* Singapore: World Scientific.

Veneziano, G. (1968). Construction of a crossing-symmetric, Regge-behaved amplitude for linearly rising trajectories. *Nuovo Cimento*, **A 57**, 190.

Veneziano, G. (1974). An introduction to dual models of strong interactions and their physical motivations. *Physics Reports*, **C 9**, 199.

Vladimirov, V. S., I. V. Volovich & E. I. Zelenov (1994). *P-Adic Analysis and Mathematical Physics*. Singapore: World Scientific.

Volkov, D. V. & V. P. Akulov (1972). On a possible universal interaction of neutrino. *JETP Letters*, **16**, 438.

von Laue, M. (1909). Die Wellenstrahlung einer bewegten Punktladung nach dem Relativitätsprinzip. *Annalen der Physik*, **28**, 436.

von Laue, M. (1911). *Das Relativitätsprinzip.* Braunschweig: Vieweg.

Wald, R. M. (1984). *General Relativity.* Chicago: The University of Chicago Press.

Wataghin, G. (1934). Bemerkung über die Selbstenergie der Elektronen. *Zeitschrift für Physik*, **88**, 92.

Weinberg, S. (1967). A model of leptons. *Physical Review Letters*, **19**, 1264.

Weinberg, S. (1972). *Gravitation and Cosmology.* New York: Wiley

Weinberg, S. (1973). Non-Abelian gauge theories of the strong interactions. *Physical Review Letters*, **31**, 494.

Weinberg, S. (1995, 1996, 2000). *The Quantum Theory of Fields*, in three volumes: I. Foundations, II. Modern Applications, III. Supersymmetry. Cambridge: Cambridge University Press.

Wess, J. (1960). The conformal invariance in quantum field theory. *Nuovo Cimento*, **18**, 1086.

Wess, J. & J. Bagger (1983). *Supersymmetry and Supergravity.* 2nd edition 1992. Cambridge: Cambridge University Press.

Wess, J. & B. Zumino (1972). A Lagrangian model invariant under supergauge transformation. *Physics Letters*, **B 49**, 52.

Weyl, H. (1918). *Raum, Zeit, Materie.* 8th edition 1993. Berlin: Springer. *Space, Time, Matter.* Translated from the 4th German edition 1952. New York: Dover.

Weyl, H. (1929). Elektron und Gravitation. I. *Zeitschrift für Physik*, **56**, 330.

Wheeler, J. A. (1957). On the nature of quantum geometrodynamics. *Annals of Physics*, **2**, 604.

Wheeler, J. A. & R. P. Feynman (1945). Interaction with the absorber as the mechanism of radiation. *Reviews of Modern Physics*, **17**, 157.

Wheeler, J. A. & R. P. Feynman (1949). Classical electrodynamics in terms of direct interparticle action. *Reviews of Modern Physics*, **21**, 425.

Whitham, G. B. (1974). *Linear and Non-Linear Waves.* New York: Wiley.

Whittaker, E. T. (1904). *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies.* 4th edition 1959. Cambridge: Cambridge University Press.

Whittaker, E. T. (1910). *A History of the Theories of Aether and Electricity, The Classical Theories.* Revised and enlarged 1951. London: Nelson.

Whittaker, E. T. (1953). *A History of the Theories of Aether and Electricity, The Modern Theories 1900 – 1926.* London: Nelson.

Wick, G. C. (1954). Properties of Bethe–Salpeter wave functions. *Physical Review*, **96**, 1124.

Wiechert, E. (1900). Elektrodynamische Elementargesetze. *Archives Néerlandaises des sciences exastes et naturelles* (2), **5**, 549.

Wigner, E. P. (1939). On unitary representations of the inhomogeneous Lorentz group. *Annals of Mathematics*, **40**, 149.

Wilson, K. G. (1974). Confinement of quarks. *Physical Review*, **D 10**, 2445.

Wilson, K. G. (1979). Problems in physics with many scales of length. *Scientific American*, **241** (# 2), 158.

Wilson, K. G. (1983). The renormalization group and critical phenomena. *Reviews of Modern Physics*, **55**, 583.

Witten, E. (1986). Noncommutative geometry and string field theory. *Nuclear Physics*, **B 268**, 253.

Wong, S. K. (1970). Field and particle equation for the classical Yang-Mills field and particles with isotopic spin. *Nuovo Cimento*, **A 65**, 689.

Woodard, R. P. (2006). Avoiding dark energy with $1/R$ modifications of gravity. In *Proceedings of the 3rd Aegean Summer School, Chios, Greece, 26 September – 1 October, 2005.* E-print astro-ph/0601672.

Wu, T. P. & C. N. Yang (1969). Some solutions of the classical isotopic gauge field equations. In *Properties of Matter Under Unusual Conditions.* Edited by H. Mark and S. Fernbach. New York: Interscience, pp. 349-354.

Wu, T. P. & C. N. Yang (1975). Concept of nonintegrable phase factor and global formulation of gauge fields. *Physical Review*, **D 12**, 3845.

Yang, C. N. (1974). Integral formalism for gauge fields. *Physical Review Letters*, **33**, 445.

Yang, C. N. & R. Mills (1954). Conservation of isotopic spin and isotopic gauge invariance. *Physical Review*, **96**, 191.

Yoneya, T. (1974). Connection of dual models to electrodynamics and gravidynamics. *Progress of Theoretical Physics*, **51**, 1907.

Young, L. C. (1969). *Lectures on the Calculus of Variations and Optimal Control Theory.* Philadelphia: Saunders.

Zumino, B. (1960). Gauge properties of propagators in quantum electrodynamics. *Journal of Mathematical Physics*, **1**, 1.

Zwanziger, D. (1971). Local-Lagrangian quantum field theory of electric and magnetic charges. *Physical Review*, **D 3**, 880.

Zweig, G. (1964). An SU(3) model for strong interaction symmetry and its breaking I, and II. CERN preprints No. 8182/TH.401, and 8419/TH.412. Reprinted in *Developments in the Quark Theory of Hadrons*. (1980). Vol. 1, 1964–1978. Edited by D. B. Lichtenberg and S. P. Rosen. Reading: Hadronic Press.

Index