Geometrical Dynamics of Complex Systems

A Unified Modelling Approach to Physics, Control, Biomechanics, Neurodynamics and Psycho-Socio-Economical Dynamics

by

VLADIMIR G. IVANCEVIC

AND TIJANA T. IVANCEVIC

MICROPROCESSOR-BASED AND INTELLIGENT SYSTEMS ENGINEERING

Springer

Geometrical Dynamics of Complex Systems

International Series on

MICROPROCESSOR-BASED AND
INTELLIGENT SYSTEMS ENGINEERING

VOLUME 31

Editor

Jacobs

Editorial Advisory Board

Geometrical Dynamics of Complex Systems

A Unified Modelling Approach to Physics, Control, Biomechanics, Neurodynamics and Psycho-Socio-Economical Dynamics

by

VLADIMIR G. IVANCEVIC
Defence Science and Technology Organisation, Adelaide, SA, Australia

and

TIJANA T. IVANCEVIC
The University of Adelaide, SA, Australia

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-4544-1 (HB)
ISBN-13 978-1-4020-4544-8 (HB)
ISBN-10 1-4020-4545-X (e-book)
ISBN-13 978-1-4020-4545-5 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

Printed in the Netherlands.

Dedicated to Nitya, Atma and Kali

Contents

Preface

Geometrical Dynamics of Complex Systems is a graduate–level monographic textbook. It represents a comprehensive introduction into rigorous geometrical dynamics of complex systems of various natures. By 'complex systems', in this book are meant high–dimensional nonlinear systems, which can be (but not necessarily are) adaptive. This monograph proposes a unified geometrical approach to dynamics of complex systems of various kinds: engineering, physical, biophysical, psychophysical, sociophysical, econophysical, etc. As their names suggest, all these multi–input multi–output (MIMO) systems have something in common: the underlying physics. However, instead of dealing with the popular 'soft complexity philosophy',[1] we rather propose a rigorous geometrical and topological approach. We believe that our rigorous approach has much greater *predictive power* than the soft one. We argue that science and technology is all about *prediction* and *control*. Observation, understanding and explanation are important in education at undergraduate level, but after that it should be all prediction and control. The main objective of this book is to show that high–dimensional nonlinear systems and processes of 'real life' can be modelled and analyzed using rigorous mathematics, which enables their complete predictability and controllability, as if they were linear systems.

It is well–known that *linear systems*, which are completely predictable and controllable by definition – live only in Euclidean spaces (of various dimensions). They are as simple as possible, mathematically elegant and fully elaborated from either scientific or engineering side. However, in nature, nothing is linear. In reality, everything has a certain degree of nonlinearity, which means: unpredictability, with subsequent uncontrollability. So, our simple and elegant linear systems, that cover almost all of our university textbooks in

[1] It is well–known that the 'soft complexity philosophy', which has been proclaimed and developed in the famous Santa Fe Institute, actually advocates 'simplicity' by means of 'reduction' under the name 'complexity'. As such, it is very different from the 'general system theory' which states that a complex system is more than a sum of its components.

mathematics, physics and engineering, do not match the nonlinear complexity of real life. We have a very common situation that best students in these three disciplines, when they come to real life, discover that their knowledge does not work in practice. Their *linear knowledge* does not match *nonlinear reality*. So, they give it up in favor of 'modern soft technologies', which *are* designed to match the nonlinear reality, but *are not* too much concerned about 'old–fashioned issues' of prediction and control.

As an illustration of this very common situation, consider a typical shock–mitigation analysis using 'linear' accelerometer data (for the car, train, aircraft, or spacecraft industry). No matter how well the shock–absorbers and suspension under the pilot/passenger seats are designed, the resulting process is nonlinear by its nature. Here, we clearly have an over–damped oscillation process – as required, but a completely different one from all the simplified process that cover our mechanical or control textbooks – as it is nonlinear. If we numerically expand the real accelerometer–data into a Taylor series around the shock–event point, we can see that a linear over–damped curve is just the first–approximation to the real process. Thus, by adding quadratic, cubic, and higher–order terms, we can get closer and closer to reality. This is the nonlinear engineering that we propose in this book.

As a more complicated example, consider human vs. humanoid (loco)motion systems. If we neglect the neuro–muscular component of human motion (which is intractable at this stage of science and technology)[2] – we still have a considerable difference in the joints of the two mechanisms. Human joints are simply more flexible than robot joints. Besides the dominant rotational degrees of freedom, which *are* correctly replicated by majority of advanced Japanese humanoid robots (Honda, Sony, Waseda), human joints also have a number of 'hidden', or 'higher–order constraint' degrees of freedom (DOFs). These 'hidden DOFs' are small translations that exist in all human joints, and are associated to all major joint rotations. These small translations *are not* replicated by contemporary humanoids, for obvious reasons of increased complexity in dynamical balance, stability and control. These 'higher–order constraints' of

[2] Recall that subjects like 'complexity', 'self–organization', 'connectionism' and 'adaptive systems' had already been studied in the 1940s and 1950s, in fields like 'cybernetics' through researchers like N. Wiener, W.R. Ashby and J. von Neumann. As they lived before the time of computers, they tackled complex systems only with mathematics, pencil and paper (e.g., von Neumann discovered cellular automata and self–reproducing systems without computers, in the same way as A. Lyapunov and H. Poincaré worked on the foundations of chaos theory). Old "C–Theories": cybernetics, catastrophe theory, chaos theory,... have the common goal to explain complex systems which consist of a large number of mutually interacting and interwoven parts. New complexity theories: cellular automata (CA), neural networks (NN), artificial intelligence (AI), and artificial life (ALife), are related fields, but they do not try to describe general complex systems. Using an intelligent combination of all these fields, we are interested in prediction and control of complex systems.

human joints can be viewed as higher–order terms of some multidimensional Taylor–series expansions of the 'first–order DOFs' currently existing in the robot joints.

Thirdly, as an apparently unexpected example, we note the striking geometrical and dynamical similarity between a typical nonlinear MIMO control system (with M input processes, N output processes, and K feedback lops), and a closed oriented superstring system (with M incoming strings, interacting through internal K loops, to produce N outgoing strings). One might argue that it is not a cost–effective job to explain a control system using a string theory, which is clearly true in case of relatively simple control systems, but nevertheless, this realization that the underlying 'stringy control' geometrical dynamics is the same in both cases – has its own scientific value that might eventually lead to discovery of new control systems.

Now, recall that *nonlinearity* can be two–fold: natural deterministic systems range from almost linear to chaotic, while natural stochastic systems range from almost linear to ergodic. In this book we will cover them all. Now, as majority of our readership, we have also passed through 'linear schools of thought', so to be able to comprehend any natural system, we will always try to find its corresponding linear system, if it exists. From geometrical point of view, nonlinear systems (of any dimension) can be viewed as such deformations of associated linear systems, which are locally topologically equivalent to the corresponding linear ones that live in Euclidean spaces. This idea leads to the fundamental concept of a *manifold*. Our geometrical machinery starts with manifolds.

On the other hand, geometrical elaboration of the above subtle idea of 'hidden DOFs', or 'higher–order constraints', in complex systems, has developed into the notion of 'higher–order tangency' of modern *jet bundles*. We believe that the most rigorous analysis of complex systems of any nature can be performed using the formalism of jet bundles, which is the pinnacle of our geometrical machinery. The question might be cost–effectiveness, but it is still good to have a fully reliable machinery, at least as a benchmark.

Our approach to dynamics of complex systems is somewhat similar to the approach to mathematical physics used at the beginning of the 20th Century by the leading two mathematicians: David Hilbert and John von Neumann – the approach of combining mathematical rigor with conceptual clarity. In addition, we put a strong accent on modern geometrical methods, so that besides physical clarity we emphasize also *geometrical intuition* that underpins it.

This geometrical intuition is based on the main work of Albert Einstein, technically known as *geometrodynamics*. It is another name for Einstein's theory of gravitation, the term proposed by his younger collaborator from Princeton, John Wheeler.[3] Aiming at a systematic identification of matter

[3] John A. Wheeler from Institute for Advanced Study at Princeton, was best known for coining the term 'Black Hole'. He was a PhD–supervisor of Richard Feynman.

with space, geometrodynamics has often been said to be an extension of the philosophy of nature as conceived by Descartes and Spinoza.

That much about history. A *modern geometrical dynamics* is a strong contemporary trend of unifying inter–scientific dynamical methods based on differential geometry, topology and Feynman's path integrals.[4]

As a final motivation, consider the vital problem of *prediction* and *control/prevention* of some natural disaster (e.g., a hurricane). The role of science in dealing with a phenomenon/treat like this can be depicted as a feedback–loop:

$$\begin{array}{ccccccc} Observation & \longrightarrow & Understanding & \longrightarrow & Prediction & \longrightarrow & Control \\ \uparrow & & & & & & \downarrow \\ & \longleftarrow & \longleftarrow & \longleftarrow & \longleftarrow & \longleftarrow & \end{array}$$

with the following four components/phases:

1. *Observation*, i.e., monitoring a phenomenon in case, using experimental sensing/measuring methods (e.g., orbital satellite imaging). This phase produces measurement data that could be fitted as graphs of analytical functions.
2. *Understanding*, in the form of geometrical pattern recognition, i.e., recognizing the turbulent patterns of spatio–temporal chaotic behavior of the approaching hurricane, in terms of geometrical objects (e.g., tensor– and spinor–fields). This phase recognizes the observation graphs as cross–sections of some jet bundles, thus representing the validity criterion for the observation phase.
3. *Prediction*: when, where and how will the hurricane strike?
   Now, common, inductive approach here means fitting a statistical model into empirical satellite data. However, we know that this works only for a very short time in the future, as extrapolation is not a valid predictive procedure, even if (adaptive) extended Kalman filter is used. Instead, we suggest a deductive approach of fitting some data into a well–defined dynamical model. This means formulating a dynamical system on configuration and phase–space manifolds, which incorporates all previously

In the 1960s, Wheeler tried to achieve Einstein's unfinished project of a *unified field theory*, under the title *Geometrodynamics Program* [Whe61, Whe62]. As we know, both the Einstein's unified theory and Wheeler's program failed, in a similar way as the famous Hilbert's Program of axiomatization of all mathematical sciences failed. However, their influences on today's developments in modern physics are as strong as Hilbert's influence on developments in modern mathematics.

[4] Recall that one of the leading mathematicians of this age, and one of the founders of modern geometrical dynamics, Vladimir I. Arnold, starts his controversial article on teaching mathematics with: "Mathematics is a part of physics. Physics is an experimental science, a part of natural science. Mathematics is the part of physics where experiments are cheap..." This is the spirit of modern geometrical dynamics. Only, it goes beyond physics, into the realms of engineering, biophysics, psychophysics, sociophysics and econophysics.

recognized turbulent patterns of the hurricane's spatio–temporal behavior. Once a valid dynamical model is formulated, the necessary empirical satellite data would include system parameters, initial and boundary conditions. So, this would be a pattern–driven modelling of the hurricane, rather than blind data–driven statistical modelling. This phase is the validity criterion for the understanding phase.

4. *Control*: this is the final stage of manipulating the hurricane to prevent the destruction. If we have already formulated a valid geometrical–pattern–based dynamical model, this task can be relatively easily accomplished, as

$$Control\ System\ =\ Dynamical\ System\ +\ Controller.$$

So, here the problem is to design a feedback controller/compensator for the dynamical model. This phase is the validity criterion for the prediction phase.

This book has two Chapters and Appendix. The first Chapter develops our geometrical machinery, in both intuitive and rigorous manner. The second Chapter applies this geometrical machinery to a number of examples of complex systems, including mechanical, physical, control, biomechanical/robotic, neurodynamical and psycho–socio–economical systems. The Appendix gives all the necessary background for comprehensive reading of this book, so that it can be used as a two–semester graduate course in engineering, physics and mathematics. Target readership includes all researchers and students of complex systems (in engineering, mathematics, physics, chemistry, biology, psychology, sociology, economics, medicine, etc.), working both in industry (i.e., clinics) and academia.

Adelaide,
November 2005

V. Ivancevic, Defence Science & Technology Organisation, Australia, e-mail: Vladimir.Ivancevic@dsto.defence.gov.au

T. Ivancevic, School of Mathematics, The University of Adelaide, e-mail: Tijana.Ivancevic@adelaide.edu.au

Acknowledgments

The authors wish to thank Land Operations Division, Defence Science & Technology Organisation, Australia, for the support in developing the *Human Biodynamics Engine* (HBE) and all the HBE–related text in this monograph.

We also express our gratitude to *Springer* book series *Microprocessor–Based and Intelligent Systems Engineering,* and especially to Editor, Professor Spyros Tzafestas.

Glossary of Frequently Used Symbols

General

– 'iff' means 'if and only if';
– 'r.h.s' means 'right hand side'; 'l.h.s' means 'l.h.s.';
– ODE means ordinary differential equation, while PDE means partial differential equation;
– *Einstein's summation convention over repeated indices* (not necessarily one up and one down) *is assumed in the whole text,* unless explicitly stated otherwise.

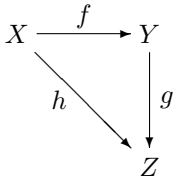
Sets

$\mathbb{N}$ – natural numbers;
$\mathbb{Z}$ – integers;
$\mathbb{R}$ – real numbers;
$\mathbb{C}$ – complex numbers;
$\mathbb{H}$ – quaternions;
$\mathbb{K}$ – number field of real numbers, complex numbers, or quaternions.

Maps

$f : A \to B$ – a function, (or map) between sets $A \equiv \operatorname{Dom} f$ and $B \equiv \operatorname{Cod} f$;

$$\begin{aligned} \operatorname{Ker} f &= f^{-1}(e_B) - \text{a kernel of } f; \\ \operatorname{Im} f &= f(A) - \text{an image of } f; \\ \operatorname{Coker} f &= \operatorname{Cod} f / \operatorname{Im} f - \text{a cokernel of } f; \\ \operatorname{Coim} f &= \operatorname{Dom} f / \operatorname{Ker} f - \text{a coimage of } f; \end{aligned}$$



– a commutative diagram, requiring $h = g \circ f$.

Derivatives

$C^\infty(A,B)$ – set of $k-$times differentiable functions between sets A to B;
$C^\infty(A,B)$ – set of *smooth* functions between sets A to B;
$C^0(A,B)$ – set of *continuous* functions between sets A to B;
$f'(x) = \frac{df(x)}{dx}$ – derivative of f with respect to x;
$\dot{x}$ – total time derivative of x;
$\partial_t \equiv \frac{\partial}{\partial t}$ – partial time derivative;
$\partial_{x^i} \equiv \partial_i \equiv \frac{\partial}{\partial x^i}$ – partial coordinate derivative;
$\dot{f} = \partial_t f + \partial_{x^i} f\, \dot{x}^i$ – total time derivative of the scalar field $f = f(t, x^i)$;
$u_t \equiv \partial_t u,\ u_x \equiv \partial_x u,\ u_{xx} \equiv \partial_{x^2} u$ – only in partial differential equations;
$L_{x^i} \equiv \partial_{x^i} L,\ L_{\dot{x}^i} \equiv \partial_{\dot{x}^i} L$ – coordinate and velocity partial derivatives of the Lagrangian function;
d – exterior derivative;
d^n – coboundary operator;
∂_n – boundary operator;
$\nabla = \nabla(g)$ – affine Levi–Civita connection on a smooth manifold M with Riemannian metric tensor $g = g_{ij}$;
Γ^i_{jk} – Christoffel symbols of the affine connection ∇;
$\nabla_X T$ – covariant derivative of the tensor–field T with respect to the vector–field X, defined by means of Γ^i_{jk};
$T_{;x^i} \equiv T_{|x^i}$ – covariant derivative of the tensor–field T with respect to the coordinate basis $\{x^i\}$;
$\dot{T} \equiv \frac{DT}{dt} \equiv \frac{\nabla T}{dt}$ – absolute (intrinsic, or Bianchi) derivative of the tensor–field T upon the parameter t; e.g., acceleration vector is the *absolute time derivative* of the velocity vector, $a^i = \dot{\vec{v}}^i \equiv \frac{Dv^i}{dt}$; note that in general, $a^i \neq \dot{v}^i$ – this is crucial for *proper definition of Newtonian force* (see Appendix);
$\mathcal{L}_X T$ – Lie derivative of the tensor–field T in direction of the vector–field X;
$[X,Y]$ – Lie bracket (commutator) of two vector–fields X and Y;
$[F,G]$, or $\{F,G\}$ – Poisson bracket, or Lie–Poisson bracket, of two functions F and G.

Smooth Manifolds, Fibre Bundles and Jet Spaces

Unless otherwise specified, all *manifolds* $M, N, ...$ are assumed $C^\infty-$smooth, real, finite–dimensional, Hausdorff, paracompact, connected and without boundary,[5] while all maps are assumed smooth (C^∞). We use the symbols $\otimes$, $\vee$, $\wedge$ and $\oplus$ for the tensor, symmetrized and exterior products, as well as the Whitney sum[6], respectively, while $\rfloor$ denotes the interior product (contraction) of

[5] The only 1D manifolds obeying these conditions are the real line $\mathbb{R}$ and the circle S^1.

[6] *Whitney sum* $\oplus$ is an analog of the direct (Cartesian) product for vector bundles. Given two vector bundles Y and Y' over the same base X, their Cartesian product is a vector bundle over $X \times X$. The *diagonal map* induces a vector bundle over X called the Whitney sum of these vector bundles and denoted by $Y \oplus Y'$.

(multi)vectors and $p-$forms, and $\hookrightarrow$ denotes a manifold imbedding (i.e., both a submanifold and a topological subspace of the codomain manifold). The symbols ∂^A_B denote partial derivatives with respect to coordinates possessing multi–indices ${}^B_A$ (e.g., $\partial_\alpha = \partial/\partial x^\alpha$);
TM – tangent bundle of the manifold M;
$\pi_M : TM \to M$ – natural projection;
T^*M – cotangent bundle of the manifold M;
$\pi : Y \to X$ – fibre bundle;
(E, π, M) – vector bundle with total space E, base M and projection π;
(Y, π, X, V) – fibre bundle with total space Y, base X, projection π and standard fibre V;
$J^k(M, N)$ – space of $k-$jets of smooth functions between manifolds M and N;
$J^k(X, Y)$ – k–jet space of a fibre bundle $Y \to X$; in particular, in mechanics we have a 1–jet space $J^1(\mathbb{R}, Q)$, with 1–jet coordinate maps $j^1_t s : t \mapsto (t, x^i, \dot{x}^i)$, as well as a 2–jet space $J^2(\mathbb{R}, Q)$, with 2–jet coordinate maps $j^2_t s : t \mapsto (t, x^i, \dot{x}^i, \ddot{x}^i)$;
$j^k_x s$ – $k-$jets of sections $s^i : X \to Y$ of a fibre bundle $Y \to X$;
We use the following kinds of *manifold maps*: immersion, imbedding, submersion, and projection. A map $f : M \to M'$ is called the *immersion* if the tangent map Tf at every point $x \in M$ is an injection (i.e., '1–1' map). When f is both an immersion and an injection, its image is said to be a submanifold of M'. A submanifold which also is a topological subspace is called imbedded submanifold. A map $f : M \to M'$ is called *submersion* if the tangent map Tf at every point $x \in M$ is a surjection (i.e., 'onto' map). If f is both a submersion and a surjection, it is called *projection* or *fibre bundle*.

Lie and (Co)Homology Groups

G – usually a general Lie group;
$GL(n)$ – general linear group with real coefficients in dimension n;
$SO(n)$ – group of rotations in dimension n;
T^n – toral (Abelian) group in dimension n;
$Sp(n)$ – symplectic group in dimension n;
$T(n)$ – group of translations in dimension n;
$SE(n)$ – Euclidean group in dimension n;
$H_n(M) = \operatorname{Ker} \partial_n / \operatorname{Im} \partial_{n-1}$ – nth homology group of the manifold M;
$H^n(M) = \operatorname{Ker} d^n / \operatorname{Im} d^{n+1}$ – nth cohomology group of the manifold M.

Other Spaces and Operators

$\mathrm{i} \equiv \sqrt{-1}$ – imaginary unit;
$C^\infty(M)$ – space of $k-$differentiable functions on the manifold M;
$\Omega^k(M)$ – space of $k-$forms on the manifold M;
$\mathfrak{g}$ – Lie algebra of a Lie group G, i.e., the tangent space of G at its identity

element;
$Ad(g)$ – adjoint endomorphism; recall that *adjoint representation* of a Lie group G is the linearized version of the action of G on itself by conjugation, i.e., for each $g \in G$, the inner automorphism $x \mapsto gxg^{-1}$ gives a linear transformation $Ad(g) : \mathfrak{g} \to \mathfrak{g}$, from the Lie algebra $\mathfrak{g}$ of G to itself;
nD space (group, system) means $n-$dimensional space (group, system), for $n \in \mathbb{N}$;
$\triangleright$ – semidirect (noncommutative) product; e.g., $SE(3) = SO(3) \triangleright \mathbb{R}^3$;
$\oint\!\!\!\!\!\!\int$ – Feynman path integral symbol, denoting integration over continuous spectrum of smooth paths and summation over discrete spectrum of Markov chains; e.g., $\oint\!\!\!\!\!\!\int \mathcal{D}[x]\, \mathrm{e}^{\mathrm{i}S[x]}$ denotes the path integral (i.e., sum–over–histories) over all possible paths $x^i = x^i(t)$ defined by the Hamiltonian action, $S[x] = \frac{1}{2}\int_{t_0}^{t_1} g_{ij}\, \dot{x}^i \dot{x}^j\, dt$, while $\oint\!\!\!\!\!\!\int \mathcal{D}[\Phi]\, \mathrm{e}^{\mathrm{i}S[\Phi]}$ denotes the path integral over all possible fields $\Phi^i = \Phi^i(x)$ defined by some field action $S[\Phi]$.

Categories

$\mathcal{S}$ – all sets as objects and all functions between them as morphisms;
$\mathcal{PS}$ – all pointed sets as objects and all functions between them preserving base point as morphisms;
$\mathcal{V}$ – all vector spaces as objects and all linear maps between them as morphisms;
$\mathcal{B}$ – Banach spaces over $\mathbb{R}$ as objects and bounded linear maps between them as morphisms;
$\mathcal{G}$ – all groups as objects, all homomorphisms between them as morphisms;
$\mathcal{A}$ – Abelian groups as objects, homomorphisms between them as morphisms;
$\mathcal{AL}$ – all algebras (over a given number field $\mathbb{K}$) as objects, all their homomorphisms between them as morphisms;
$\mathcal{T}$ – all topological spaces as objects, all continuous functions between them as morphisms;
$\mathcal{PT}$ – pointed topological spaces as objects, continuous functions between them preserving base point as morphisms;
$\mathcal{TG}$ – all topological groups as objects, their continuous homomorphisms as morphisms;
$\mathcal{M}$ – all smooth manifolds as objects, all smooth maps between them as morphisms;
$\mathcal{M}_n$ – nD manifolds as objects, their local diffeomorphisms as morphisms;
$\mathcal{LG}$ – all Lie groups as objects, all smooth homomorphisms between them as morphisms;
$\mathcal{LAL}$ – all Lie algebras (over a given field $\mathbb{K}$) as objects, all smooth homomorphisms between them as morphisms;
$\mathcal{TB}$ – all tangent bundles as objects, all smooth tangent maps between them as morphisms;

$\mathcal{T}^*\mathcal{B}$ – all cotangent bundles as objects, all smooth cotangent maps between them as morphisms;
$\mathcal{VB}$ – all smooth vector bundles as objects, all smooth homomorphisms between them as morphisms;
$\mathcal{FB}$ – all smooth fibre bundles as objects, all smooth homomorphisms between them as morphisms;
Symplec – all symplectic manifolds (i.e., physical phase–spaces), all symplectic maps (i.e., canonical transformations) between them as morphisms;
Hilbert – all Hilbert spaces and all unitary operators as morphisms.

1

Modern Geometrical Machinery

1.1 Introduction

As stated in the preface, the objective of this monograph is a *unified rigorous geometrical dynamics of complex systems of various natures*, where by 'complex systems' we mean *high–dimensional nonlinear systems* (with or without *adaptation*), predominantly with *continuous–time dynamics*. The main purpose of this book is to show that high–dimensional nonlinear systems and processes from 'real life' can be modelled and analyzed using rigorous mathematics, which enables their complete *predictability & controllability*. With this book we try to fit into the 'market' currently occupied by two 'worlds': (i) traditional 'hard linear engineering', and (ii) modern 'soft nonlinear engineering'. We believe that our new approach has advantages of both of these engineering worlds. So, before proceeding, let us quickly observe both of them.

Today, we see a number of the so–called 'soft agent simulators' of complex systems, which are supposedly 'good models of real life situations'; they can be 'simply taken off–shelf' and are 'very easy to understand and implement', both in software and in hardware. Some of these soft simulators have 'discrete–time' dynamics, while others have some strange 'dynamics' with no time at all. Regarding this soft–complexity fashion, we argue:
(i) that every complex behavior in nature is a temporal process (some are even spatio–temporal processes); and more,
(ii) that sequences of discrete events occurring in equidistant time steps do not exist in nature.[1]
In other words, we question the *predictive value* of 'discrete event simulators', and even more the *general realism* of time–independent soft simulators and multi–agent systems based on chess–like games.[2] It is true that real–life games

[1] An experimental evidence for this claim is well–known to any experimentalist: the sampling rate of the measuring device necessarily has to match the measured dynamics – otherwise we miss the valuable information.

[2] Recall that in a discrete event simulation approach to human performance modelling, the whole simulation represents a schedule–like process to be modelled,

are often decided by a 'single knock–out blow'; however, their dynamics are never instantaneous and repeatable in equidistant time steps. In such simplistic discrete models we often miss the most vital signal/system information. If we eventually need a discrete–time functional approximation, we will rather approximate the solution – not the problem, as the overall modelling error will be much smaller, and consequently the predictive value much higher.

In a word, we argue that continuous–time models are much finer predictors for real–life situations than discrete–time models, which are coarse by nature and therefore represent quick routes to artificially–chaotic behavior.[3] For example, popular Belief–Desire–Intention (BDI) agents based on Boolean logic (see e.g., [RG98], as well as subsection 2.6.8 below) – necessarily show unrealistic and 'jerky' behavior.[4]

Such discrete dynamics is most often defined by some (two–state, or Ising spin) Boolean function $f(x, y, \dots)$, while all the change that f undergoes can be represented by the associated *Boolean derivative*,

$$\frac{\partial f}{\partial x} = \begin{cases} 1 & \text{if } f(|x-1|, y, \dots) \neq f(x, y, \dots), \\ 0 & \text{otherwise,} \end{cases}$$

---

while individual activities (or, tasks) are events scheduled to occur from the queue. This approach is useful as a 'bird view' on a complex schedule including a number of unimportant activities. However, each of the included activities is modelled as a discrete event described by only two characteristics: Start and Stop. The obvious question here is: what happens between the start and stop? Where is the process of the action? It does not exist. Therefore, this approach is O.K. as long as there is no a single important human performance action that we might really want to simulate, understand, predict, and control. If there is only one important activity or task in the schedule, then this methodology crashes at that task. This is more obvious if in the schedule we have at least one physical, chemical, physiological, or cognitive process. Such a process represents a temporal (or even a spatio–temporal) system with its own input–output variables, its own initial (and possibly boundary) conditions, as well as measurable system parameters. If such a 'strange' event is actually included in the schedule, then we need a very different methodology to simulate, predict and control it.

[3] This is similar to numerical solution of differential equations versus difference equations (or, 'maps'). In realms of physical, chemical, biological, psychological and socio–economical worlds, differential equations are generally much finer models, even if they are numerically solved (providing the use of sophisticated integrators with either adaptive or very small time steps) – than the corresponding difference maps, which are both coarse approximations to reality and quick routes to chaos. Namely, the necessary condition for chaos can be satisfied (as a minimum) by a three–dimensional continuous–time dynamical system compared to a one–dimensional discrete–time map. And usually, this apparent chaos exists only in the model – not in the real system described by the model.

[4] This problem can be solved by using either *temporal fuzzy–logic* [Kos92], or *temporal logic of actions* [Lam94], instead of simplistic and 'jerky' Boolean logic.

which represents a measure of sensitivity of a function f with respect to its arguments $x, y, \ldots$. For example, this kind of sequential dynamics is used in cellular automata [BFR04].

Contrasted to this simplistic Boolean dynamics, within our geometrical machinery we propose a multitude of high–dimensional functions, with their changes defined by a variety of sophisticated derivatives, including: covariant, absolute, exterior, and Lie–derivatives.

However, this does not mean that we will not use modern soft techniques in our modelling of complex adaptive systems. On the contrary, they can be very useful for 'hybrid adaptive hard–soft simulators'. We are only concerned with their *weak predictive power*, which we intend to build–up using our rigorous geometrical machinery. For example, we will describe the activation dynamics of a neural network using a smooth manifold, and add on the top of it a discrete–time (Hebbian or reinforcement) learning dynamics.

On the other hand, recall that linear systems are, by definition, fully predictable and controllable. They live only in Euclidean spaces of various dimensions. Unfortunately, nothing is linear in nature – everything has some degree of nonlinearity. Natural deterministic systems range from almost linear to chaotic, while natural stochastic systems range from almost linear to ergodic. In general, nonlinear systems can be viewed as such deformations of associated linear systems, which are *locally topologically equivalent* to the corresponding linear ones.

In our opinion, a strong predictive model of a complex system, suitable for real–life applications, is a nonlinear, multiple–input multiple–output (MIMO) system,[5] predominantly continuous–time and smooth, but at the same time

[5] It is well–known that linear MIMO control systems can always be put into Kalman's canonical (modular) state–space form of order n, with m inputs and k outputs (see [KFA69]). In case of continuous–time systems we have the state and output equation of the form

$$\begin{aligned} d\mathbf{x}/dt &= \mathbf{A}(t)\,\mathbf{x}(t) + \mathbf{B}(t)\,\mathbf{u}(t), \\ \mathbf{y}(t) &= \mathbf{C}(t)\,\mathbf{x}(t) + \mathbf{D}(t)\,\mathbf{u}(t), \end{aligned} \tag{1.1}$$

while in case of discrete time systems we have the state and output equation of the form

$$\begin{aligned} \mathbf{x}(n+1) &= \mathbf{A}(n)\,\mathbf{x}(n) + \mathbf{B}(n)\,\mathbf{u}(n), \\ \mathbf{y}(n) &= \mathbf{C}(n)\,\mathbf{x}(n) + \mathbf{D}(n)\,\mathbf{u}(n). \end{aligned} \tag{1.2}$$

Both in (1.1) and in (1.2) the variables have the following meaning:

$\mathbf{x}(t) \in \mathbb{X}$ is an $n-$vector of state variables belonging to the *vector state–space* $\mathbb{X} \subset \mathbb{R}^n$;

$\mathbf{u}(t) \in \mathbb{U}$ is an $m-$vector of inputs belonging to the *vector input space* $\mathbb{U} \subset \mathbb{R}^m$;

$\mathbf{y}(t) \in \mathbb{Y}$ is a $k-$vector of outputs belonging to the *vector output space* $\mathbb{Y} \subset \mathbb{R}^k$;

$\mathbf{A}(t) : \mathbb{X} \to \mathbb{X}$ is an $n \times n$ matrix of state dynamics;

$\mathbf{B}(t) : \mathbb{U} \to \mathbb{X}$ is an $n \times m$ matrix of input map;

allowing discrete–time control and adaptation. It resembles a real–life situation (possibly including noise, uncertainty, imprecision and misinformation). At the same time, to be a good predictor, it needs full observability, controllability and stability. On the top of these three essential criteria, we might put the forth one, adaptability, as required by complexity theory (see e.g., [B-Y97]). In this book we propose a unified geometrical approach, somewhat similar to the popular one presented in [AS92, CD98], as well as to the more serious one presented in [Arn92, Arn93] – for design of such models.

As Einstein said: "Nature is simple only when analyzed locally. Why? Because, locally any system appears to be linear, and therefore fully predictable and controllable. Geometrical elaboration of this fundamental idea has produced the concept of a *manifold*, a topological space which locally looks like Euclidean $\mathbb{R}^n-$spaces, but globally can be totally different. In addition, to be able to use calculus on our manifolds, in much the same way as in ordinary $\mathbb{R}^n-$spaces, the manifolds need to be smooth (i.e., differentiable as many times as required, technically denoted by C^∞).

Consider a classical example, comparing a surface of an apple with a Euclidean plane. A small neighborhood of every point on the surface of an apple (excluding its stem) looks like a Euclidean plane (denoted by $\mathbb{R}^2$), with its *local geodesics* appearing like straight lines. In other words, a smooth surface is *locally topologically equivalent* to the Euclidean plane. This same concept of nonlinear geometry holds in any dimension. If dimension is high, we are dealing with complex systems. Therefore, while continuous–time linear systems live in Euclidean $\mathbb{R}^n-$spaces, continuous–time complex systems live in nD smooth manifolds, usually denoted by M.

Finally, note that there are two dynamical paradigms of smooth manifolds:
(i) Einstein's 4D *space–time manifold*, historically the first one, and
(ii) nD *configuration manifold*, which is our *core* geometrical concept.

As the Einstein space–time manifold is both simpler to comprehend and consequently much more elaborated, we start our geometrical machinery with it, keeping in mind that the same fundamental dynamics holds for all smooth manifolds, regardless of their dimension.

Throughout the book we will try to follow the Hilbertian pedagogical method of development: (i) intuitively introduce a new geometrical concept; (ii) rigorously define it; (iii) apply it to solve a real–world problem.

Intuition Behind Einstein's Geometrodynamics

Briefly, Einstein–Wheeler geometrodynamics can be summarized as:

$\mathbf{C}(t) : \mathbb{X} \to \mathbb{Y}$ is a $k \times n$ matrix of output map;
$\mathbf{D}(t) : \mathbb{U} \to \mathbb{Y}$ is a $k \times m$ matrix of input–output transform.
Input $\mathbf{u}(t) \in \mathbb{U}$ can be empirically determined by trial and error; it is properly defined by quadratic optimization process called *Kalman regulator*, or more generally (in the presence of noise), by (extended) *Kalman filter* [Kal60].

1. Gravity is not a Newtonian force, but an aspect of the geometry of space–time.
2. Space is not an absolute invariant entity, but is influenced by the distribution of mass and energy in the Universe. The fundamental Geometrodynamics Principle states:
   Space tells matter how to move, while matter tells space how to curve.
3. Large masses introduce a strong curvature in space–time. Light and matter are forced to move according to this metric. Since all the matter is in motion, the geometry of space is constantly changing.

The celebrated *Einstein equation* relates the curvature of space–time to the mass/energy density. It reads (in the so–called 'normal' units: $c = 8\pi G = 1$):

$$\mathbf{G} = \mathbf{T}, \qquad \text{or, in components,} \qquad G_{\alpha\beta} = T_{\alpha\beta}, \tag{1.3}$$

where $\mathbf{G} = G_{\alpha\beta}$ is the Einstein curvature tensor, representing space–time geometry, while $\mathbf{T} = T_{\alpha\beta}$ is the stress–energy–momentum tensor, the 'mystical' SEM–tensor, representing matter; the 4D indices $\alpha, \beta = (0, 1, 2, 3)$ label respectively the four space–time directions: (t, x, y, z).

To grasp the intuition behind the Einstein equation (1.3), we need to consider a ball filled with test particles that are all initially at rest relative to each other. Let $V = V(t)$ be the volume of the ball after a proper time t has elapsed, as measured by the particle at the center of the ball. Then the Einstein equation says:

$$\frac{\ddot{V}}{V}\Big|_{t=0} = -\frac{1}{2}\begin{pmatrix} \text{flow of } t\text{−momentum in } \ t-\text{direction} + \\ \text{flow of } x\text{−momentum in } x-\text{direction} + \\ \text{flow of } y\text{−momentum in } y-\text{direction} + \\ \text{flow of } z\text{−momentum in } z-\text{direction} \end{pmatrix},$$

where these flows are measured at the center of the ball at time $t = 0$, using local inertial coordinates. These flows are the diagonal components of the SEM–tensor $\mathbf{T}$. Its components $T_{\alpha\beta}$ tell us how much momentum in the α−direction is flowing in the β−direction through a given point of space–time. The flow of t−momentum in the t−direction is just the energy density, $T_{00} = \rho$. The flow of x−momentum in the x−direction is the 'pressure in the x−direction', $T_{11} = P_1 \equiv P_x$, and similarly for y and z.

In any event, we may summarize the Einstein equation (1.3) as

$$\frac{\ddot{V}}{V}\Big|_{t=0} = -\frac{1}{2}(\rho + P_x + P_y + P_z) \equiv -\frac{1}{2}(T_{00} + T_{11} + T_{22} + T_{33}). \tag{1.4}$$

This new equation tells us that positive energy density and positive pressure curve space–time in a way that makes a freely falling ball of point particles tend to shrink. Since $E = mc^2$ and we are working in normal units, ordinary mass density counts as a form of energy density. Thus a massive object will make a swarm of freely falling particles at rest around it start to shrink. In short, (1.4) tells us that *gravity attracts* (see e.g., [MTW73, Bae01]).

To see why equation (1.4) is equivalent to the Einstein equation (1.3), we need to understand the *Riemann curvature tensor* and the *geodesic deviation equation*. Namely, when space–time is curved, the result of *parallel transport* depends on the path taken. To quantify this notion, pick two vectors u and v at a point p in space–time. In the limit where $\epsilon \longrightarrow 0$, we can approximately speak of a 'parallelogram' with sides ϵu and ϵv. Take another vector w at p and parallel transport it first along ϵv and then along ϵu to the opposite corner of this parallelogram. The result is some vector w_1. Alternatively, parallel transport w first along ϵu and then along ϵv. The result is a slightly different vector, w_2. The limit

$$\lim_{\epsilon \longrightarrow 0} \frac{w_2 - w_1}{\epsilon^2} = R(u, v)w \tag{1.5}$$

is well–defined, and it measures the curvature of space–time at the point p. In local coordinates, we can write it as

$$R(u, v)w = R^{\alpha}_{\beta\gamma\delta} u^{\beta} v^{\gamma} w^{\delta}.$$

The quantity $R^{\alpha}_{\beta\gamma\delta}$ is called the *Riemann curvature tensor*. We can use this tensor to calculate the relative acceleration of nearby particles in free fall if they are initially at rest relative to one another. Consider two freely falling particles at nearby points p and q. Let v be the velocity of the particle at p, and let ϵu be the vector from p to q. Since the two particles start out at rest relative to one other, the velocity of the particle at q is obtained by parallel transporting v along ϵu.

Now let us wait a short while. Both particles trace out geodesics as time passes, and at time ϵ they will be at new points, say p' and q'. The point p' is displaced from p by an amount ϵv, so we get a little parallelogram, exactly as in the definition of the Riemann curvature:

Next let us calculate the new relative velocity of the two particles. To compare vectors we must carry one to another using parallel transport. Let v_1 be the vector we get by taking the velocity vector of the particle at p' and parallel transporting it to q' along the top edge of our parallelogram. Let v_2 be the velocity of the particle at q'. The difference $v_2 - v_1$ is the new relative velocity. It follows that over this passage of time, the average relative acceleration of the two particles is $a = (v_2 - v_1)/\epsilon$. By equation (1.5),

$$\lim_{\epsilon \to 0} \frac{v_2 - v_1}{\epsilon^2} = R(u, v)v, \qquad \text{therefore} \qquad \lim_{\epsilon \to 0} \frac{a}{\epsilon} = R(u, v)v.$$

This is the simplified form of the *geodesic deviation equation*. From the definition of the Riemann curvature it is easy to see that $R(u, v)w = -R(v, u)w$, so we can also write this equation as

$$\lim_{\epsilon \longrightarrow 0} \frac{a^{\alpha}}{\epsilon} = -R^{\alpha}_{\beta\gamma\delta} v^{\beta} u^{\gamma} v^{\delta}. \tag{1.6}$$

Using geodesic deviation equation (1.6) we can work out the second time derivative of the volume $V(t)$ of a small ball of test particles that start out at

rest relative to each other. For this we must let u range over an orthonormal basis of tangent vectors, and sum the 'outwards' component of acceleration for each one of these. By equation (1.6) this gives

$$\lim_{V \longrightarrow 0} \frac{\ddot{V}}{V}\Big|_{t=0} = -R^{\alpha}_{\beta\alpha\delta} v^{\beta} v^{\delta}.$$

In terms of the so–called *Ricci tensor*, which is a contracted Riemann tensor,

$$R_{\beta\delta} = R^{\alpha}_{\beta\alpha\delta}\,,$$

we may write the above expression as

$$\lim_{V \longrightarrow 0} \frac{\ddot{V}}{V}\Big|_{t=0} = -R_{\beta\delta} v^{\beta} v^{\delta}.$$

In local inertial coordinates, where the ball starts out at rest, we have $v = (1, 0, 0, 0)$, so

$$\lim_{V \longrightarrow 0} \frac{\ddot{V}}{V}\Big|_{t=0} = -R_{00}. \tag{1.7}$$

In short, the Ricci tensor says how our ball of freely falling test particles starts changing in volume. The Ricci tensor only captures some of the information in the *Riemann curvature tensor*. The rest is captured by the so–called the *Weyl tensor* (see e.g., [Pen89, Pen94, Pen97]), which says how any such ball starts changing in shape. The Weyl tensor describes tidal forces, gravitational waves and the like.

Now, the Einstein equation in its usual form says

$$G_{\alpha\beta} = T_{\alpha\beta}.$$

Here the right side is the stress-energy tensor, while the left side, the 'Einstein tensor', is just an abbreviation for a quantity constructed from the Ricci tensor:

$$G_{\alpha\beta} = R_{\alpha\beta} - \frac{1}{2} g_{\alpha\beta} R^{\gamma}_{\gamma}.$$

Thus the Einstein equation really says

$$R_{\alpha\beta} - \frac{1}{2} g_{\alpha\beta} R^{\gamma}_{\gamma} = T_{\alpha\beta}. \tag{1.8}$$

This implies

$$R^{\alpha}_{\alpha} - \frac{1}{2} g^{\alpha}_{\alpha} R^{\gamma}_{\gamma} = T^{\alpha}_{\alpha},$$

but $g^{\alpha}_{\alpha} = 4$, so

$$-R^{\alpha}_{\alpha} = T^{\alpha}_{\alpha}.$$

Substituting this into equation (1.8), we get

$$R_{\alpha\beta} = T_{\alpha\beta} - \frac{1}{2} g_{\alpha\beta} T^{\gamma}_{\gamma}. \tag{1.9}$$

This is an equivalent version of the Einstein equation, but with the roles of R and T switched [Bae01]. This is a formula for the Ricci tensor, which has a simple geometrical meaning.

Equation (1.9) will be true if any one component holds in all local inertial coordinate systems. This is a bit like the observation that all of Maxwell's equations are contained in Gauss's law and and $\nabla \cdot B = 0$. Clearly, this is only true if we know how the fields transform under change of coordinates. Here we assume that the transformation laws are known. Given this, the Einstein equation (1.3) is equivalent to the fact that

$$R_{00} = T_{00} - \frac{1}{2} g_{00} T^{\gamma}_{\gamma} \tag{1.10}$$

in every local inertial coordinate system about every point. In such coordinates we have

$$g = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.11}$$

so $g_{00} = -1$, as well as

$$T^{\gamma}_{\gamma} = -T_{00} + T_{11} + T_{22} + T_{33}.$$

Equation (1.10) thus says that

$$R_{00} = \frac{1}{2}(T_{00} + T_{11} + T_{22} + T_{33}).$$

By equation (1.7), this is equivalent to the required

$$\lim_{V \to 0} \frac{\ddot{V}}{V}\Big|_{t=0} = -\frac{1}{2}(T_{00} + T_{11} + T_{22} + T_{33}).$$

Quick Definition of Einstein's Geometrodynamics

As a final introductory motivation, we give an 'express–flight bird–view' on *derivation of the Einstein equation* from the *Hilbert action principle*, starting from the Einstein space–time manifold M. For all technical details, see [MTW73], which is still, after 33 years, the core textbook on the subject.

$$
\begin{aligned}
& M \qquad \text{... the } \textbf{space–time manifold } M \\
g_{ij} &= g_{ij}(x^i) \in T_x M \qquad \text{... metric tensor on } M \\
g^{ij} &= (g_{ij})^{-1} \qquad \text{... inverse metric tensor on } M \\
\Gamma_{ijk} &= \frac{1}{2}(\partial_{x^k} g_{ij} + \partial_{x^j} g_{ki} - \partial_{x^i} g_{jk}) \qquad \text{... 1–order Christoffel symbols} \\
\Gamma^k_{ij} &= g^{kl}\Gamma_{ijl} \qquad \text{... 2–order Christoffel symbols (Levi–Civita connection)} \\
R^l_{ijk} &= \partial_{x^j}\Gamma^l_{ik} - \partial_{x^k}\Gamma^l_{ij} + \Gamma^l_{rj}\Gamma^r_{ik} - \Gamma^l_{rk}\Gamma^r_{ij} \quad \text{... Riemann curvature tensor} \\
R_{ij} &= R^l_{ijl} \qquad \text{... Ricci tensor is the trace of Riemann} \\
R &= g^{ij}R_{ij} \qquad \text{... scalar curvature is the trace of Ricci} \\
G_{ij} &= R_{ij} - \frac{1}{2}Rg_{ij} \qquad \text{... Einstein tensor is the trace–reversed Ricci} \\
T_{ij} &= -2\frac{\delta L_{Hilb}}{\delta g^{ij}} + g_{ij}L_{Hilb} \quad \text{... stress–energy–momentum (SEM) tensor} \\
L_{Hilb} &= \frac{1}{16\pi}g^{ij}R_{ij}(-g)^{1/2} \qquad \text{... is derived from the Hilbert Lagrangian} \\
\delta S &= \delta \int L_{Hilb}(-g)^{1/2}d^4x = 0 \qquad \text{... the Hilbert action principle gives} \\
G_{ij} &= 8\pi T_{ij} \qquad \text{... the } \textbf{Einstein equation}.
\end{aligned}
$$

We will continue Einstein's geometrodynamics in subsection 1.5.4 below.

1.2 Smooth Manifolds

1.2.1 Intuition Behind a Smooth Manifold

As we have already got the initial feeling, in the heart of geometrical dynamics is the concept of a *manifold* (see [Rha84]). To get some *dynamical intuition* behind this concept, let us consider a simple 3DOF mechanical system determined by three generalized coordinates, $q^i = \{q^1, q^2, q^3\}$. There is a unique way to represent this system as a 3D manifold, such that to each point of the manifold there corresponds a definite configuration of the mechanical system with coordinates q^i; therefore, we have a geometrical representation of the configurations of our mechanical system, called the *configuration manifold*. If the mechanical system moves in any way, its coordinates are given as the functions of the time. Thus, the motion is given by equations of the form: $q^i = q^i(t)$. As t varies (i.e., $t \in \mathbb{R}$), we observe that the system's *representative point* in the configuration manifold describes a *curve* and $q^i = q^i(t)$ are the equations of this curve.

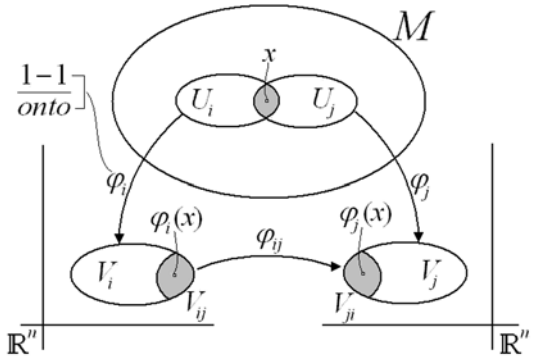


Fig. 1.1. An intuitive geometrical picture behind the manifold concept (see text).

On the other hand, to get some *geometrical intuition* behind the concept of a manifold, consider a set M (see Figure 1.1) which is a *candidate* for a manifold. Any point $x \in M$[6] has its *Euclidean chart*, given by a 1–1 and *onto* map $\varphi_i : M \to \mathbb{R}^n$, with its *Euclidean image* $V_i = \varphi_i(U_i)$. More precisely, a chart φ_i is defined by

$$\varphi_i : M \supset U_i \ni x \mapsto \varphi_i(x) \in V_i \subset \mathbb{R}^n,$$

where $U_i \subset M$ and $V_i \subset \mathbb{R}^n$ are open sets (see [Arn78, Rha84]).

Clearly, any point $x \in M$ can have several different charts (see Figure 1.1). Consider a case of two charts, $\varphi_i, \varphi_j : M \to \mathbb{R}^n$, having in their images two open sets, $V_{ij} = \varphi_i(U_i \cap U_j)$ and $V_{ji} = \varphi_j(U_i \cap U_j)$. Then we have *transition functions* φ_{ij} between them,

$$\varphi_{ij} = \varphi_j \circ \varphi_i^{-1} : V_{ij} \to V_{ji}, \qquad \text{locally given by} \qquad \varphi_{ij}(x) = \varphi_j(\varphi_i^{-1}(x)).$$

If transition functions φ_{ij} exist, then we say that two charts, φ_i and φ_j are *compatible*. Transition functions represent a general (nonlinear) *transformations of coordinates*, which are the core of classical *tensor calculus* (Appendix).

A set of compatible charts $\varphi_i : M \to \mathbb{R}^n$, such that each point $x \in M$ has its Euclidean image in at least one chart, is called an *atlas*. Two atlases are *equivalent* iff all their charts are compatible (i.e., transition functions exist between them), so their union is also an atlas. A *manifold structure* is a class of equivalent atlases.

Finally, as charts $\varphi_i : M \to \mathbb{R}^n$ were supposed to be 1-1 and onto maps, they can be either *homeomorphisms*, in which case we have a *topological* (C^0) manifold, or *diffeomorphisms*, in which case we have a *smooth* (C^∞) manifold.

[6] Note that sometimes we will denote the point in a manifold M by m, and sometimes by x (thus implicitly assuming the existence of coordinates $x = (x^i)$).

Slightly more precisely, a topological (respectively smooth) manifold is a separable space M which is locally homeomorphic (resp. diffeomorphic) to Euclidean space $\mathbb{R}^n$, having the following properties (reflected in Figure 1.1):

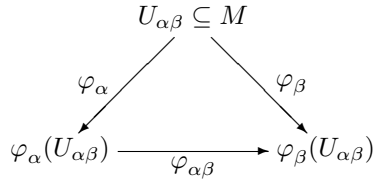
1. M is a *Hausdorff space*: For every pair of points $x_1, x_2 \in M$, there are disjoint open subsets $U_1, U_2 \subset M$ such that $x_1 \in U_1$ and $x_2 \in U_2$.
2. M is *second–countable space*: There exists a countable basis for the topology of M.
3. M *is locally Euclidean of dimension* n: Every point of M has a neighborhood that is homeomorphic (resp. diffeomorphic) to an open subset of $\mathbb{R}^n$.

This implies that for any point $x \in M$ there is a homeomorphism (resp. diffeomorphism) $\varphi : U \to \varphi(U) \subseteq \mathbb{R}^n$, where U is an open neighborhood of x in M and $\varphi(U)$ is an open subset in $\mathbb{R}^n$. The pair (U, φ) is called a *coordinate chart* at a point $x \in M$, etc.

1.2.2 Definition of a Smooth Manifold

Given a chart (U, φ), we call the set U a *coordinate domain*, or a coordinate neighborhood of each of its points. If in addition $\varphi(U)$ is an open ball in $\mathbb{R}^n$, then U is called a *coordinate ball*. The map φ is called a (*local*) *coordinate map*, and the component functions $(x^1, ..., x^n)$ of φ, defined by $\varphi(m) = (x^1(m), ..., x^n(m))$, are called *local coordinates* on U.

Two charts (U_1, φ_1) and (U_2, φ_2) such that $U_1 \cap U_2 \neq \varnothing$ are called *compatible* if $\varphi_1(U_1 \cap U_2)$ and $\varphi_2(U_2 \cap U_1)$ are open subsets of $\mathbb{R}^n$. A family $(U_\alpha, \varphi_\alpha)_{\alpha \in A}$ of compatible charts on M such that the U_α form a *covering* of M is called an *atlas*. The maps $\varphi_{\alpha\beta} = \varphi_\beta \circ \varphi_\alpha^{-1} : \varphi_\alpha(U_{\alpha\beta}) \to \varphi_\beta(U_{\alpha\beta})$ are called the *transition maps*, for the atlas $(U_\alpha, \varphi_\alpha)_{\alpha \in A}$, where $U_{\alpha\beta} = U_\alpha \cap U_\beta$, so that we have a commutative triangle:



An atlas $(U_\alpha, \varphi_\alpha)_{\alpha \in A}$ for a manifold M is said to be a $C^\infty-$*atlas*, if all transition maps $\varphi_{\alpha\beta} : \varphi_\alpha(U_{\alpha\beta}) \to \varphi_\beta(U_{\alpha\beta})$ are of class C^∞. Two C^∞ atlases are called $C^\infty-$*equivalent*, if their union is again a $C^\infty-$atlas for M. An equivalence class of $C^\infty-$atlases is called a $C^\infty-$*structure* on M. In other words, a smooth structure on M is a *maximal* smooth atlas on M, i.e., such an atlas that is not contained in any strictly larger smooth atlas. By a $C^\infty-$*manifold* M, we mean a topological manifold together with a $C^\infty-$structure and a chart on M will be a chart belonging to some atlas of the $C^\infty-$structure. Smooth

manifold means $C^\infty-$manifold, and the word '*smooth*' is used synonymously for C^∞ [Rha84].

Sometimes the terms 'local coordinate system' or 'parametrization' are used instead of charts. That M is not defined with any particular atlas, but with an equivalence class of atlases, is a mathematical formulation of the *general covariance* principle. Every suitable coordinate system is equally good. A Euclidean chart may well suffice for an open subset of $\mathbb{R}^n$, but this coordinate system is not to be preferred to the others, which may require many charts (as with polar coordinates), but are more convenient in other respects.

For example, the atlas of an $n-$sphere S^n has two charts. If $N = (1, 0, ..., 0)$ and $S = (-1, ..., 0, 0)$ are the north and south poles of S^n respectively, then the two charts are given by the stereographic projections from N and S:

$$\varphi_1 : S^n\backslash\{N\} \to \mathbb{R}^n, \varphi_1(x^1, ..., x^{n+1}) = (x^2/(1-x^1), \ldots, x^{n+1}/(1-x^1)), \text{ and}$$
$$\varphi_2 : S^n\backslash\{S\} \to \mathbb{R}^n, \varphi_2(x^1, ..., x^{n+1}) = (x^2/(1+x^1), \ldots, x^{n+1}/(1+x^1)),$$

while the overlap map $\varphi_2 \circ \varphi_1^{-1} : \mathbb{R}^n\backslash\{0\} \to \mathbb{R}^n\backslash\{0\}$ is given by the diffeomorphism $(\varphi_2 \circ \varphi_1^{-1})(z) = z/||z||^2$, for z in $\mathbb{R}^n\backslash\{0\}$, from $\mathbb{R}^n\backslash\{0\}$ to itself.

Various *additional structures* can be imposed on $\mathbb{R}^n$, and the corresponding manifold M will inherit them through its covering by charts. For example, if a covering by charts takes their values in a *Banach space* E, then E is called the *model space* and M is referred to as a $C^\infty-$*Banach manifold* modelled on E. Similarly, if a covering by charts takes their values in a *Hilbert space* $\mathcal{H}$, then $\mathcal{H}$ is called the *model space* and M is referred to as a $C^\infty-$*Hilbert manifold* modelled on $\mathcal{H}$. If not otherwise specified, we will consider M to be an Euclidean manifold, with its covering by charts taking their values in $\mathbb{R}^n$.

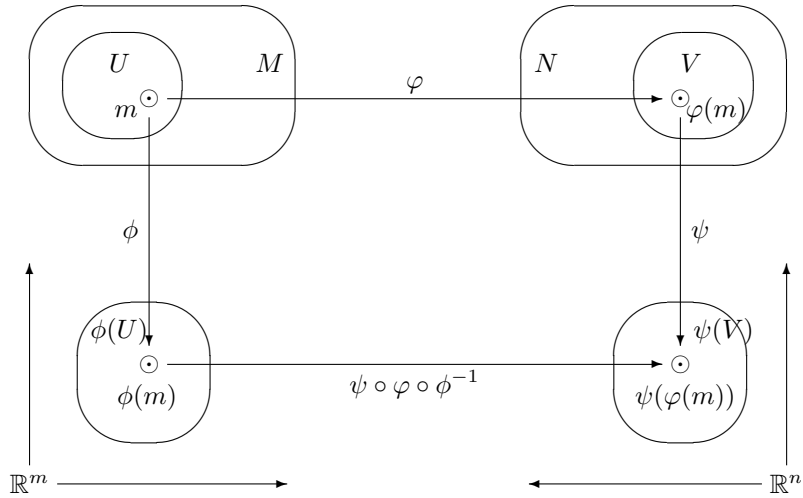
For a Hausdorff $C^\infty-$manifold the following properties are equivalent [KMS93]: (i) it is paracompact; (ii) it is metrizable; (iii) it admits a Riemannian metric;[7] (iv) each connected component is separable.

1.2.3 Smooth Maps Between Manifolds

A map $\varphi : M \to N$ between two manifolds M and N, with $M \ni m \mapsto \varphi(m) \in N$, is called a *smooth map*, or $C^\infty-$map, if we have the following charting:

[7] Recall the corresponding properties of a *Euclidean metric* d. For any three points $x, y, z \in \mathbb{R}^n$, the following axioms are valid:

$$M_1 : d(x,y) > 0, \quad \text{for} \quad x \neq y; \qquad \text{and} \qquad d(x,y) = 0, \quad \text{for} \quad x = y;$$
$$M_2 : d(x,y) = d(y,x); \qquad M_3 : d(x,y) \leq d(x,z) + d(z,y).$$



This means that for each $m \in M$ and each chart (V, ψ) on N with $\varphi(m) \in V$ there is a chart (U, ϕ) on M with $m \in U, \varphi(U) \subseteq V$, and $\Phi = \psi \circ \varphi \circ \phi^{-1}$ is C^∞, that is, the following diagram commutes:

$$\begin{array}{ccc} M \supseteq U & \xrightarrow{\varphi} & V \subseteq N \\ \phi \downarrow & & \downarrow \psi \\ \phi(U) & \xrightarrow[\Phi]{} & \psi(V) \end{array}$$

Let M and N be smooth manifolds and let $\varphi : M \to N$ be a smooth map. The map φ is called a *covering*, or equivalently, M is said to *cover* N, if φ is surjective and each point $n \in N$ admits an open neighborhood V such that $\varphi^{-1}(V)$ is a union of disjoint open sets, each diffeomorphic via φ to V.

A $C^\infty-$map $\varphi : M \to N$ is called a $C^\infty-$*diffeomorphism* if φ is a bijection, $\varphi^{-1} : N \to M$ exists and is also C^∞. Two manifolds are called diffeomorphic if there exists a diffeomorphism between them. All smooth manifolds and smooth maps between them form the category $\mathcal{M}$.

Intuition Behind Topological Invariants of Manifolds

Now, restricting to the topology of nD compact (i.e., closed and bounded) and connected manifolds, the only cases in which we have a complete understanding of topology are $n = 0, 1, 2$. The only compact and connected 0D manifold is a point. A 1D compact and connected manifold can either be a line element or a circle, and it is intuitively clear (and can easily be proven) that these two spaces are topologically different. In 2D, there is already an infinite number of different topologies: a 2D compact and connected surface

can have an arbitrary number of handles and boundaries, and can either be orientable or non–orientable (see figure 1.2). Again, it is intuitively quite clear that two surfaces are not homeomorphic if they differ in one of these respects. On the other hand, it can be proven that any two surfaces for which these data are the same can be continuously mapped to one another, and hence this gives a complete classification of the possible topologies of such surfaces.

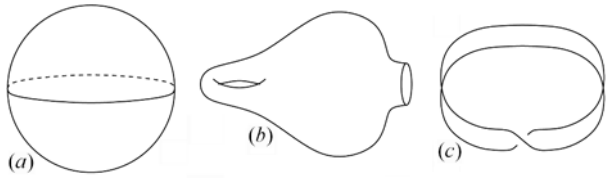


Fig. 1.2. Three examples of 2D manifolds: (*a*) The sphere S^2 is an orientable manifold without handles or boundaries. (*b*) An orientable manifold with one boundary and one handle. (*c*) The Möbius strip is an unorientable manifold with one boundary and no handles.

A quantity such as the number of boundaries of a surface is called a *topological invariant*. A topological invariant is a number, or more generally any type of structure, which one can associate to a topological space, and which does not change under continuous mappings. Topological invariants can be used to distinguish between topological spaces: if two surfaces have a different number of boundaries, they can certainly not be topologically equivalent. On the other hand, the knowledge of a topological invariant is in general not enough to decide whether two spaces are homeomorphic: a torus and a sphere have the same number of boundaries (zero), but are clearly not homeomorphic. Only when one has some *complete set* of topological invariants, such as the number of handles and boundaries in the 2D case, is it possible to determine whether or not two topological spaces are homeomorphic. In more than 2D, many topological invariants are known, but for no dimension larger than two has a complete set of topological invariants been found. In 3D, it is generally believed that a finite number of countable invariants would suffice for compact manifolds, but this is not rigorously proven, and in particular there is at present no generally accepted construction of a complete set. A very interesting and intimately related problem is the famous *Poincaré conjecture*, stating that if a 3D manifold has a certain set of topological invariants called its 'homotopy groups' equal to those of the 3–sphere S^3, it is actually homeomorphic to the three-sphere. In four or more dimensions, a complete set of topological invariants would consist of an uncountably infinite number of invariants! A general classification of topologies is therefore very hard to get, but even without such a general classification, each new invariant that can be constructed gives us a lot of interesting new information. For this rea-

son, the construction of topological invariants of manifolds is one of the most important issues in topology.

1.2.4 (Co)Tangent Bundles of a Smooth Manifold

Intuition Behind a Tangent Bundle

In mechanics, to each nD *configuration manifold* M there is associated its $2n$D *velocity phase–space manifold*, denoted by TM and called the tangent bundle of M (see Figure 1.3). The original smooth manifold M is called the *base* of TM. There is an onto map $\pi : TM \longrightarrow M$, called the *projection*. Above each point $x \in M$ there is a *tangent space* $T_xM = \pi^{-1}(x)$ to M at x, which is called a *fibre*. The fibre $T_xM \subset TM$ is the subset of TM, such that the total tangent bundle, $TM = \bigsqcup_{m \in M} T_xM$, is a *disjoint union* of tangent spaces T_xM to M for all points $x \in M$. From dynamical perspective, the most important quantity in the tangent bundle concept is the smooth map $v : M \longrightarrow TM$, which is an inverse to the projection π, i.e, $\pi \circ v = \mathrm{Id}_M$, $\pi(v(x)) = x$. It is called the *velocity vector–field*. Its graph $(x, v(x))$ represents the *cross–section* of the tangent bundle TM. This explains the dynamical term velocity phase–space, given to the tangent bundle TM of the manifold M.

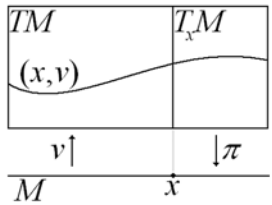


Fig. 1.3. A sketch of a tangent bundle TM of a smooth manifold M (see text for explanation).

Definition of a Tangent Bundle

Recall that if $[a, b]$ is a closed interval, a C^0-map $\gamma : [a, b] \to M$ is said to be *differentiable* at the endpoint a if there is a chart (U, ϕ) at $\gamma(a)$ such that the following limit exists and is finite [AMR88]:

$$\frac{d}{dt}(\phi \circ \gamma)(a) \equiv (\phi \circ \gamma)'(a) = \lim_{t \to a} \frac{(\phi \circ \gamma)(t) - (\phi \circ \gamma)(a)}{t - a}. \tag{1.12}$$

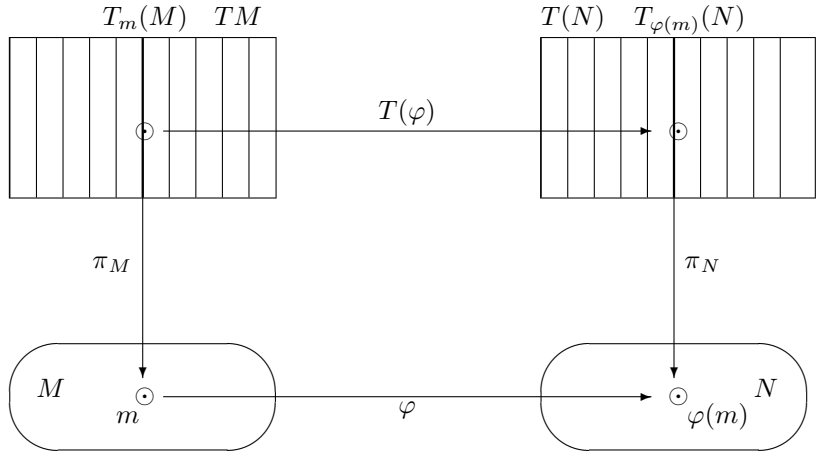
Generalizing (1.12), we get the notion of the *curve on a manifold*. For a smooth manifold M and a point $m \in M$ a curve at m is a C^0-map $\gamma : I \to M$ from an interval $I \subset \mathbb{R}$ into M with $0 \in I$ and $\gamma(0) = m$.

Two curves γ_1 and γ_2 passing though a point $m \in U$ are *tangent at* m with respect to the chart (U, ϕ) if $(\phi \circ \gamma_1)'(0) = (\phi \circ \gamma_2)'(0)$. Thus, two curves are tangent if they have identical tangent vectors (same direction and speed) in a local chart on a manifold.

For a smooth manifold M and a point $m \in M$, the *tangent space* T_mM to M at m is the *set of equivalence classes* of curves at m:

$$T_mM = \{[\gamma]_m : \gamma \text{ is a curve at a point } m \in M\}.$$

A $C^\infty-$map $\varphi : M \ni m \mapsto \varphi(m) \in N$ between two manifolds M and N induces a linear map $T_m\varphi : T_mM \to T_{\varphi(m)}N$ for each point $m \in M$, called a *tangent map*, if we have:



i.e., the following diagram commutes:

$$\begin{array}{ccc} T_mM & \xrightarrow{T_m\varphi} & T_{\varphi(m)}N \\ \downarrow{\scriptstyle \pi_M} & & \downarrow{\scriptstyle \pi_N} \\ M \ni m & \xrightarrow[\varphi]{} & \varphi(m) \in N \end{array}$$

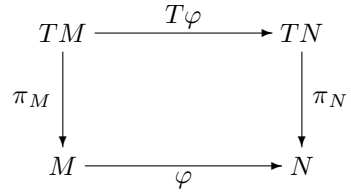
with the *natural projection* $\pi_M : TM \to M$, given by $\pi_M(T_mM) = m$, that takes a tangent vector v to the point $m \in M$ at which the vector v is attached i.e., $v \in T_mM$.

For an nD smooth manifold M, its nD *tangent bundle* TM is the disjoint union of all its tangent spaces T_mM at all points $m \in M$, $TM = \bigsqcup_{m \in M} T_mM$.

To define the smooth structure on TM, we need to specify how to construct local coordinates on TM. To do this, let $(x^1(m), ..., x^n(m))$ be local coordinates of a point m on M and let $(v^1(m), ..., v^n(m))$ be components of a tangent vector in this coordinate system. Then the $2n$ numbers $(x^1(m), ..., x^n(m), v^1(m), ..., v^n(m))$ give a *local coordinate system* on TM.

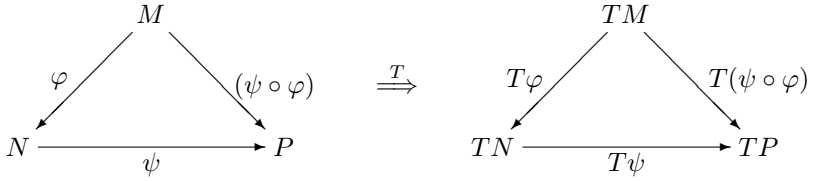
$TM = \bigsqcup_{m \in M} T_m M$ defines a family of vector spaces parameterized by M. The inverse image $\pi_M^{-1}(m)$ of a point $m \in M$ under the natural projection π_M is the tangent space $T_m M$. This space is called the *fibre* of the tangent bundle over the point $m \in M$ [Ste72].

A $C^\infty-$map $\varphi : M \to N$ between two manifolds M and N induces a linear *tangent map* $T\varphi : TM \to TN$ between their tangent bundles, i.e., the following diagram commutes:



All tangent bundles and their tangent maps form the category $\mathcal{TB}$. The category $\mathcal{TB}$ is the natural framework for *Lagrangian dynamics*.

Now, we can formulate the *global version of the chain rule*. If $\varphi : M \to N$ and $\psi : N \to P$ are two smooth maps, then we have $T(\psi \circ \varphi) = T\psi \circ T\varphi$ (see [KMS93]). In other words, we have a functor $T : \mathcal{M} \Rightarrow \mathcal{TB}$ from the category $\mathcal{M}$ of smooth manifolds to the category $\mathcal{TB}$ of their tangent bundles:



Definition of a Cotangent Bundle

A *dual* notion to the tangent space $T_m M$ to a smooth manifold M at a point m is its *cotangent space* $T_m^* M$ at the same point m. Similarly to the tangent bundle, for a smooth manifold M of dimension n, its *cotangent bundle* T^*M is the disjoint union of all its cotangent spaces $T_m^* M$ at all points $m \in M$, i.e., $T^*M = \bigsqcup_{m \in M} T_m^* M$. Therefore, the cotangent bundle of an $n-$manifold M is the vector bundle $T^*M = (TM)^*$, the (real) dual of the tangent bundle TM.

If M is an $n-$manifold, then T^*M is a $2n-$manifold. To define the smooth structure on T^*M, we need to specify how to construct local coordinates on T^*M. To do this, let $(x^1(m), ..., x^n(m))$ be local coordinates of a point m on M and let $(p_1(m), ..., p_n(m))$ be components of a covector in this coordinate system. Then the $2n$ numbers $(x^1(m), ..., x^n(m), p_1(m), ..., p_n(m))$ give a local coordinate system on T^*M. This is the basic idea one uses to prove that indeed T^*M is a $2n-$manifold.

$T^*M = \bigsqcup_{m\in M} T^*_m M$ defines a family of vector spaces parameterized by M, with the *conatural projection*, $\pi^*_M : T^*M \to M$, given by $\pi^*_M(T^*_m M) = m$, that takes a covector p to the point $m \in M$ at which the covector p is attached i.e., $p \in T^*_m M$. The inverse image $\pi^{-1}_M(m)$ of a point $m \in M$ under the conatural projection π^*_M is the cotangent space $T^*_m M$. This space is called the *fibre* of the cotangent bundle over the point $m \in M$.

In a similar way, a $C^\infty-$map $\varphi : M \to N$ between two manifolds M and N induces a linear *cotangent map* $T^*\varphi : T^*M \to T^*N$ between their cotangent bundles, i.e., the following diagram commutes:

$$\begin{array}{ccc} T^*M & \xrightarrow{T^*\varphi} & T^*N \\ {\scriptstyle \pi^*_M}\downarrow & & \downarrow{\scriptstyle \pi^*_N} \\ M & \xrightarrow[\varphi]{} & N \end{array}$$

All cotangent bundles and their cotangent maps form the category $\mathcal{T}^*\mathcal{B}$. The category $\mathcal{T}^*\mathcal{B}$ is the natural stage for *Hamiltonian dynamics*.

Now, we can formulate the *dual version of the global chain rule*. If $\varphi : M \to N$ and $\psi : N \to P$ are two smooth maps, then we have $T^*(\psi\circ\varphi) = T^*\psi \circ T^*\varphi$. In other words, we have a cofunctor $T^* : \mathcal{M} \Rightarrow \mathcal{T}^*\mathcal{B}$ from the category $\mathcal{M}$ of smooth manifolds to the category $\mathcal{T}^*\mathcal{B}$ of their cotangent bundles:

$$\begin{array}{ccc} & M & \\ {\scriptstyle\varphi}\swarrow & & \searrow{\scriptstyle(\psi\circ\varphi)} \\ N & \xrightarrow[\psi]{} & P \end{array} \quad \overset{T^*}{\Longrightarrow} \quad \begin{array}{ccc} & T^*M & \\ {\scriptstyle T^*\varphi}\nearrow & & \nwarrow{\scriptstyle T^*(\psi\circ\varphi)} \\ T^*N & \xleftarrow[T^*\psi]{} & T^*P \end{array}$$

1.2.5 Tensor Fields and Bundles of a Smooth Manifold

A *tensor bundle* $\mathcal{T}$ associated to a smooth $n-$manifold M is defined as a *tensor product* of tangent and cotangent bundles:

$$\mathcal{T} = \bigotimes^q T^*M \otimes \bigotimes^p TM = \overbrace{TM \otimes ... \otimes TM}^{p \text{ times}} \otimes \overbrace{T^*M \otimes ... \otimes T^*M}^{q \text{ times}}.$$

Tensor bundles are special case of more general fibre bundles (see section 1.3 below).

A *tensor–field* of type (p, q) (see Appendix) on a smooth $n-$manifold M is defined as a *smooth section* $\tau : M \to \mathcal{T}$ of the tensor bundle $\mathcal{T}$. The coefficients of the tensor–field τ are smooth (C^∞) functions with p indices up and q indices down. The classical position of indices can be explained in

modern terms as follows. If (U, ϕ) is a chart at a point $m \in M$ with local coordinates $(x^1, ..., x^n)$, we have the *holonomous frame field*

$$\partial_{x^{i_1}} \otimes \partial_{x^{i_2}} \otimes ... \otimes \partial_{x^{i_p}} \otimes dx^{j_1} \otimes dx^{j_2} ... \otimes dx^{j_q},$$

for $i \in \{1, ..., n\}^p$, $j = \{1, ..., n\}^q$, over U of this tensor bundle, and for any $(p, q)-$tensor–field τ we have

$$\tau | U = \tau^{i_1 ... i_p}_{j_1 ... j_q} \partial_{x^{i_1}} \otimes \partial_{x^{i_2}} \otimes ... \otimes \partial_{x^{i_p}} \otimes dx^{j_1} \otimes dx^{j_2} ... \otimes dx^{j_q}.$$

For such tensor–fields the *Lie derivative* along any vector–field is defined (see subsection 1.2.6 below), and it is a *derivation* (i.e., both linearity and Leibniz rules hold) with respect to the tensor product. Tensor bundle $\mathcal{T}$ admits many natural transformations (see [KMS93]). For example, a 'contraction' like the trace $T^*M \otimes TM = L(TM, TM) \to M \times \mathbb{R}$, but applied just to one specified factor of type T^*M and another one of type TM, is a natural transformation. And any 'permutation of the same kind of factors' is a natural transformation.

The tangent bundle $\pi_M : TM \to M$ of a manifold M (introduced above) is a special tensor bundle over M such that, given an atlas $\{(U_\alpha, \varphi_\alpha)\}$ of M, TM has the *holonomic atlas*

$$\Psi = \{(U_\alpha, \varphi_\alpha = T\varphi_\alpha)\}.$$

The associated linear bundle coordinates are the induced coordinates $(\dot{x}^\lambda)$ at a point $m \in M$ with respect to the *holonomic frames* $\{\partial_\lambda\}$ in tangent spaces $T_m M$. Their transition functions read (see Appendix)

$$\dot{x}'^\lambda = \frac{\partial x'^\lambda}{\partial x^\mu} \dot{x}^\mu.$$

Technically, the tangent bundle TM is a tensor bundle with the structure Lie group $GL(\dim M, \mathbb{R})$ (see section 1.2.7 below).

Recall that the cotangent bundle of M is the dual T^*M of TM. It is equipped with the induced coordinates $(\dot{x}_\lambda)$ at a point $m \in M$ with respect to *holonomic coframes* $\{dx^\lambda\}$ dual of $\{\partial_\lambda\}$. Their transition functions read

$$\dot{x}'_\lambda = \frac{\partial x'^\mu}{\partial x^\lambda} \dot{x}_\mu.$$

The Pull–Back and Push–Forward

In this subsection we define two important operations, following [AMR88], which will be used in the further text.

Let $\varphi : M \to N$ be a C^∞ map of manifolds and $f \in C^\infty(N, \mathbb{R})$. Define the *pull–back* of f by φ by

$$\varphi^* f = f \circ \varphi \in C^\infty(M, \mathbb{R}).$$

If f is a C^∞ diffeomorphism and $X \in \mathcal{X}^k(M)$, the *push–forward* of X by φ is defined by

$$\varphi_* X = T\varphi \circ X \circ \varphi^{-1} \in \mathcal{X}^k(N).$$

If x^i are local coordinates on M and y^j local coordinates on N, the preceding formula gives the components of $\varphi_* X$ by

$$(\varphi_* X)^j(y) = \frac{\partial \varphi^j}{\partial x^i}(x)\, X^i(x), \qquad \text{where} \quad y = \varphi(x).$$

We can interchange pull–back and push–forward by changing φ to φ^{-1}, that is, defining φ_* (resp. φ^*) by $\varphi_* = (\varphi^{-1})^*$ (resp. $\varphi^* = (\varphi^{-1})_*$). Thus the *push–forward* of a function f on M is $\varphi_* f = f \circ \varphi^{-1}$ and the *pull–back* of a vector–field Y on N is $\varphi^* Y = (T\varphi)^{-1} \circ Y \circ \varphi$.

Notice that φ must be a diffeomorphism in order that the pull–back and push–forward operations make sense, the only exception being pull–back of functions. Thus vector–fields can only be pulled back and pushed forward by diffeomorphisms. However, even when φ is not a diffeomorphism we can talk about *$\varphi-$related vector–fields* as follows.

Let $\varphi : M \to N$ be a C^∞ map of manifolds. The vector–fields $X \in \mathcal{X}^{k-1}(M)$ and $\mathrm{Y} \in \mathcal{X}^{k-1}(N)$ are called *$\varphi-$related*, denoted $X \sim_\varphi Y$, if $T\varphi \circ X = Y \circ \varphi$.

Note that if φ is diffeomorphism and X and Y are $\varphi-$related, then $Y = \varphi_* X$. However, in general, X can be $\varphi-$related to more than one vector–field on N. $\varphi-$relatedness means that the following diagram commutes:

$$\begin{array}{ccc} TM & \xrightarrow{\;T\varphi\;} & TN \\ {\scriptstyle X}\uparrow & & \uparrow{\scriptstyle Y} \\ M & \xrightarrow[\;\varphi\;]{} & N \end{array}$$

The behavior of flows under these operations is as follows: Let $\varphi : M \to N$ be a $C^\infty-$map of manifolds, $X \in \mathcal{X}^k(M)$ and $Y \in \mathcal{X}^k(N)$. Let F_t and G_t denote the flows of X and Y respectively. Then $X \sim_\varphi Y$ iff $\varphi \circ F_t = G_t \circ \varphi$. In particular, if φ is a diffeomorphism, then the equality $Y = \varphi_* X$ holds iff the flow of Y is $\varphi \circ F_t \circ \varphi^{-1}$ (This is called the push–forward of F_t by φ since it is the natural way to construct a diffeomorphism on N out of one on M). In particular, $(F_t)_* X = X$. Therefore, the flow of the push–forward of a vector–field is the push–forward of its flow.

Dynamical Evolution and Flow

As a motivational example, consider a mechanical system that is capable of assuming various states described by points in a set U. For example, U might be $\mathbb{R}^3 \times \mathbb{R}^3$ and a state might be the positions and momenta (x^i, p_i) of a

particle moving under the influence of the central force field, with $i = 1, 2, 3$. As time passes, the state evolves. If the state is $\gamma_0 \in U$ at time s and this changes to γ at a later time t, we set

$$F_{t,s}(\gamma_0) = \gamma,$$

and call $F_{t,s}$ the *evolution operator*; it maps a state at time s to what the state would be at time t; that is, after time $t - s$. has elapsed. Determinism is expressed by the *Chapman–Kolmogorov law* [AMR88]:

$$F_{\tau,t} \circ F_{t,s} = F_{\tau,s}, \qquad F_{t,t} = \text{identity}. \tag{1.13}$$

The evolution laws are called *time independent*, or *autonomous*, when $F_{t,s}$ depends only on $t-s$. In this case the preceding law (1.13) becomes the *group property*:

$$F_t \circ F_s = F_{t+s}, \qquad F_0 = \text{identity}. \tag{1.14}$$

We call such an F_t a *flow* and $F_{t,s}$ a *time–dependent flow*, or an evolution operator. If the system is irreversible, that is, defined only for $t \geq s$, we speak of a *semi–flow* [AMR88].

Usually, instead of $F_{t,s}$ the *laws of motion* are given in the form of ODEs that we must solve to find the flow. These equations of motion have the form:

$$\dot{\gamma} = X(\gamma), \qquad \gamma(0) = \gamma_0,$$

where X is a (possibly time–dependent) vector–field on U.

As a continuation of the previous example, consider the motion of a particle of mass m under the influence of the central force field (like gravity, or Coulombic potential) F^i $(i = 1, 2, 3)$, described by the *Newtonian equation of motion*:

$$m\ddot{x}^i = F^i(x). \tag{1.15}$$

By introducing momenta $p_i = m\dot{x}^i$, equation (1.15) splits into two Hamiltonian equations:

$$\dot{x}^i = p_i/m, \qquad \dot{p}_i = F_i(x). \tag{1.16}$$

Note that in Euclidean space we can freely interchange subscripts and superscripts. However, in general case of a Riemannian manifold, $p_i = mg_{ij}\dot{x}^j$ and (1.16) properly reads

$$\dot{x}^i = g^{ij}p_j/m, \qquad \dot{p}_i = F_i(x). \tag{1.17}$$

The phase–space here is the *Riemannian manifold* $(\mathbb{R}^3\backslash\{0\}) \times \mathbb{R}^3$, that is, the cotangent bundle of $\mathbb{R}^3\backslash\{0\}$, which is itself a smooth manifold for the central force field. The r.h.s of equations (1.17) define a Hamiltonian vector–field on this 6D manifold by

$$X(x, p) = \big((x^i, p_i), (p_i/m, F_i(x))\big). \tag{1.18}$$

Integration of equations (1.17) produces trajectories (in this particular case, planar conic sections). These trajectories comprise the flow F_t of the vector–field $X(x, p)$ defined in (1.18).

Vector–Fields and Their Flows

Vector–Fields on M

A *vector–field* X on U, where U is an open chart in $n-$manifold M, is a *smooth function* from U to M assigning to each point $m \in U$ a vector at that point, i.e., $X(m) = (m, X(m))$. If $X(m)$ is tangent to M for each $m \in M$, X is said to be a *tangent vector–field* on M. If $X(m)$ is orthogonal to M (i.e., $X(p) \in M_m^{\perp}$) for each $X(m) \in M$, X is said to be a *normal vector–field* on M.

In other words, let M be a $C^{\infty}-$manifold. A $C^{\infty}-$vector–field on M is a $C^{\infty}-$section of the tangent bundle TM of M. Thus a vector–field X on a manifold M is a $C^{\infty}-$map $X : M \to TM$ such that $X(m) \in T_m M$ for all points $m \in M$,and $\pi_M \circ X = Id_M$. Therefore, a vector–field assigns to each point m of M a vector based (i.e., bound) at that point. The set of all C^{∞} vector–fields on M is denoted by $\mathcal{X}^k(M)$.

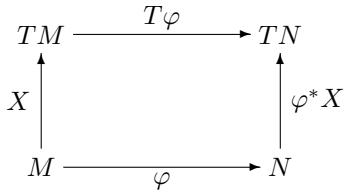
A vector–field $X \in \mathcal{X}^k(M)$ represents a field of direction indicators [Thi79]: to every point m of M it assigns a vector in the tangent space $T_m M$ at that point. If X is a vector–field on M and (U, ϕ) is a chart on M and $m \in U$, then we have $X(m) = X(m)\, \phi^i \frac{\partial}{\partial \phi^i}$. Following [KMS93], we write $X|_U = X\, \phi^i \frac{\partial}{\partial \phi^i}$.

Let M be a connected $n-$manifold, and let $f : U \to \mathbb{R}$ (U an open set in M) and $c \in \mathbb{R}$ be such that $M = f^{-1}(c)$ (i.e., M is the *level set* of the function f at *height* c) and $\nabla f(m) \neq 0$ for all $m \in M$. Then there exist on M exactly two smooth unit normal vector–fields $N_{1,2}(m) = \pm \frac{\nabla f(m)}{|\nabla f(m)|}$ (here $|X| = (X \cdot X)^{1/2}$ denotes the norm or length of a vector X, and $(\cdot)$ denotes the scalar product on M) for all $m \in M$, called *orientations* on M.

Let $\varphi : M \to N$ be a smooth map. Recall that two vector–fields $X \in \mathcal{X}^k(M)$ and $Y \in \mathcal{X}(N)$ are called $\varphi-$related, if $T\varphi \circ X = Y \circ \varphi$ holds, i.e., if the following diagram commutes:

$$\begin{array}{ccc} TM & \xrightarrow{\quad T\varphi \quad} & TN \\ {\scriptstyle X}\uparrow & & \uparrow{\scriptstyle Y} \\ M & \xrightarrow[\quad \varphi \quad]{} & N \end{array}$$

In particular, a diffeomorphism $\varphi : M \to N$ induces a linear map between vector–fields on two manifolds, $\varphi^* : \mathcal{X}^k(M) \to \mathcal{X}(N)$, such that $\varphi^* X = T\varphi \circ X \circ \varphi^{-1} : N \to TN$, i.e., the following diagram commutes:



The correspondences $M \to TM$ and $\varphi \to T\varphi$ obviously define a functor $T : \mathcal{M} \Rightarrow \mathcal{M}$ from the category of smooth manifolds to itself. T is another special case of the vector bundle functor (1.3.3), and closely related to the tangent bundle functor (1.2.4).

A C^∞ *time–dependent vector–field* is a $C^\infty-$map $X : \mathbb{R} \times M \to TM$ such that $X(t, m) \in T_m M$ for all $(t, m) \in \mathbb{R} \times M$, i.e., $X_t(m) = X(t, m)$.

Integral Curves as Dynamical Trajectories

Recall (1.2.4) that a curve γ at a point m of an $n-$manifold M is a C^0-map from an open interval $I \subset \mathbb{R}$ into M such that $0 \in I$ and $\gamma(0) = m$. For such a curve we may assign a tangent vector at each point $\gamma(t)$, $t \in I$, by $\dot{\gamma}(t) = T_t\gamma(1)$.

Let X be a smooth tangent vector–field on the smooth $n-$manifold M, and let $m \in M$. Then there exists an open interval $I \subset \mathbb{R}$ containing 0 and a parameterized curve $\gamma : I \to M$ such that:

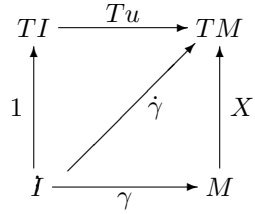
1. $\gamma(0) = m$;
2. $\dot{\gamma}(t) = X(\gamma(t))$ for all $t \in I$; and
3. If $\beta : \tilde{I} \to M$ is any other parameterized curve in M satisfying (1) and (2), then $\tilde{I} \subset I$ and $\beta(t) = \gamma(t)$ for all $t \in \tilde{I}$.

A parameterized curve $\gamma : I \to M$ satisfying condition (2) is called an *integral curve* of the tangent vector–field X. The unique γ satisfying conditions (1)–(3) is the *maximal integral curve* of X through $m \in M$.

In other words, let $\gamma : I \to M$, $t \mapsto \gamma(t)$ be a smooth curve in a manifold M defined on an interval $I \subseteq \mathbb{R}$. $\dot{\gamma}(t) = \frac{d}{dt}\gamma(t)$ defines a smooth vector–field along γ since we have $\pi_M \circ \dot{\gamma} = \gamma$. Curve γ is called an *integral curve* or *flow line* of a vector–field $X \in \mathcal{X}^k(M)$ if the tangent vector determined by γ equals X at every point $m \in M$, i.e.,

$$\dot{\gamma} = X \circ \gamma,$$

or, if the following diagram commutes:



On a chart (U, ϕ) with coordinates $\phi(m) = \left(x^1(m), ..., x^n(m)\right)$, for which $\varphi \circ \gamma : t \mapsto \gamma_i(t)$ and $T\varphi \circ X \circ \varphi^{-1} : x^i \mapsto \left(x^i, X_i(m)\right)$, this is written

$$\dot{\gamma}_i(t) = X_i(\gamma(t)), \text{ for all } t \in I \subseteq \mathbb{R}, \tag{1.19}$$

which is an ordinary differential equation of first–order in n dimensions.

The *velocity* $\dot{\gamma}$ of the parameterized curve $\gamma(t)$ is a vector–field along γ defined by

$$\dot{\gamma}(t) = (\gamma(t), \dot{x}^1(t), \dots \dot{x}^n(t)).$$

Its length $|\dot{\gamma}| : I \to \mathbb{R}$, defined by $|\dot{\gamma}|(t) = |\dot{\gamma}(t)|$ for all $t \in I$, is a function along α. $|\dot{\gamma}|$ is called *speed* of γ [Arn89].

Each vector–field X along γ is of the form $X(t) = (\gamma(t), X_1(t), \dots, X_n(t))$, where each component X_i is a function along γ. X is *smooth* if each $X_i : I \to M$ is smooth. The *derivative* of a smooth vector–field X along a curve $\gamma(t)$ is the vector–field $\dot{X}$ along γ defined by

$$\dot{X}(t) = (\gamma(t), \dot{X}_1(t), \dots \dot{X}_n(t)).$$

$\dot{X}(t)$ measures the *rate of change of the vector part* $(X_1(t), \dots X_n(t))$ of $X(t)$ *along* γ. Thus, the *acceleration* $\ddot{\gamma}(t)$ of a parameterized curve $\gamma(t)$ is the vector–field along γ get by differentiating the velocity field $\dot{\gamma}(t)$.

Differentiation of vector–fields along parameterized curves has the following properties. For X and Y smooth vector–fields on M along the parameterized curve $\gamma : I \to M$ and f a smooth function along γ, we have:

1. $\frac{d}{dt}(X + Y) = \dot{X} + \dot{Y}$;
2. $\frac{d}{dt}(fX) = \dot{f}X + f\dot{X}$; and
3. $\frac{d}{dt}(X \cdot Y) = \dot{X}Y + X\dot{Y}$.

A *geodesic* in M is a parameterized curve $\gamma : I \to M$ whose acceleration $\ddot{\gamma}$ is everywhere orthogonal to M; that is, $\ddot{\gamma}(t) \in M^{\perp}_{\alpha(t)}$ for all $t \in I \subset \mathbb{R}$. Thus a geodesic is a curve in M which always goes 'straight ahead' in the surface. Its acceleration serves only to keep it in the surface. It has no component of acceleration tangent to the surface. Therefore, it also has a constant speed $\dot{\gamma}(t)$.

Let $v \in M_m$ be a vector on M. Then there exists an open interval $I \subset \mathbb{R}$ containing 0 and a geodesic $\gamma : I \to M$ such that:

1. $\gamma(0) = m$ and $\dot{\gamma}(0) = v$; and
2. If $\beta : \tilde{I} \to M$ is any other geodesic in M with $\beta(0) = m$ and $\dot{\beta}(0) = v$, then $\tilde{I} \subset I$ and $\beta(t) = \gamma(t)$ for all $t \in \tilde{I}$.

The geodesic γ is now called the *maximal geodesic* in M passing through m with initial velocity v.

By definition, a parameterized curve $\gamma : I \to M$ is a geodesic of M iff its acceleration is everywhere perpendicular to M, i.e., iff $\ddot{\gamma}(t)$ is a multiple of the

orientation $N(\gamma(t))$ for all $t \in I$, i.e., $\ddot{\gamma}(t) = g(t)\, N(\gamma(t))$, where $g : I \to \mathbb{R}$. Taking the scalar product of both sides of this equation with $N(\gamma(t))$ we find $g = -\dot{\gamma}\dot{N}(\gamma(t))$. Thus $\gamma : I \to M$ is geodesic iff it satisfies the differential equation

$$\ddot{\gamma}(t) + \dot{N}(\gamma(t))\, N(\gamma(t)) = 0.$$

This vector equation represents the system of second–order component ODEs

$$\ddot{x}^i + N_i(x+1, \ldots, x^n) \frac{\partial N_j}{\partial x^k}(x+1, \ldots, x^n)\, \dot{x}^j \dot{x}^k = 0.$$

The substitution $u^i = \dot{x}^i$ reduces this second–order differential system (in n variables x^i) to the first–order differential system

$$\dot{x}^i = u^i, \qquad \dot{u}^i = -N_i(x+1, \ldots, x^n) \frac{\partial N_j}{\partial x^k}(x+1, \ldots, x^n)\, \dot{x}^j \dot{x}^k$$

(in $2n$ variables x^i and u^i). This first–order system is just the differential equation for the integral curves of the vector–field X in $U \times \mathbb{R}$ (U open chart in M), in which case X is called a *geodesic spray*.

Now, when an integral curve $\gamma(t)$ is the path a mechanical system Ξ follows, i.e., the solution of the equations of motion, it is called a *trajectory*. In this case the parameter t represents time, so that (1.19) describes motion of the system Ξ on its configuration manifold M.

If $X_i\,(m)$ is C^0 the existence of a local solution is guaranteed, and a *Lipschitz condition* would imply that it is unique. Therefore, exactly one integral curve passes through every point, and different integral curves can never cross. As $X \in \mathcal{X}^k(M)$ is C^∞, the following statement about the solution with arbitrary initial conditions holds [Thi79, Arn89]:

Theorem. Given a vector–field $X \in \mathcal{X}(M)$, for all points $p \in M$, there exist $\eta > 0$, a neighborhood V of p, and a function $\gamma : (-\eta, \eta) \times V \to M$, $\big(t, x^i\,(0)\big) \mapsto \gamma\,\big(t, x^i\,(0)\big)$ such that

$$\dot{\gamma} = X \circ \gamma, \qquad \gamma\,\big(0, x^i\,(0)\big) = x^i\,(0) \qquad \text{for all } \; x^i\,(0) \in V \subseteq M.$$

For all $|t| < \eta$, the map $x^i\,(0) \mapsto \gamma\,\big(t, x^i\,(0)\big)$ is a diffeomorphism f_t^X between V and some open set of M. For proof, see [Die69], I, 10.7.4 and 10.8.

This theorem states that trajectories that are near neighbors cannot suddenly be separated. There is a well–known estimate (see [Die69], I, 10.5) according to which points cannot diverge faster than exponentially in time if the derivative of X is uniformly bounded.

An integral curve $\gamma\,(t)$ is said to be *maximal* if it is not a restriction of an integral curve defined on a larger interval $I \subseteq \mathbb{R}$. It follows from the existence and uniqueness theorems for ODEs with smooth r.h.s and from elementary properties of Hausdorff spaces that for any point $m \in M$ there exists a maximal integral curve γ_m of X, passing for $t = 0$ through point m, i.e., $\gamma(0) = m$.

Theorem (Local Existence, Uniqueness, and Smoothness) [AMR88]. Let E be a Banach space, $U \subset E$ be open, and suppose $X : U \subset E \to E$ is of class C^∞, $k \geq 1$. Then

1. For each $x_0 \in U$, there is a curve $\gamma : I \to U$ at x_0 such that $\dot{\gamma}(t) = X(\gamma(t))$ for all $t \in I$.

2. Any two such curves are equal on the intersection of their domains.

3. There is a neighborhood U_0 of the point $x_0 \in U$, a real number $a > 0$, and a C^∞ map $F : U_0 \times I \to E$, where I is the open interval $]-a, a[$, such that the curve $\gamma_u : I \to E$, defined by $\gamma_u(t) = F(u,t)$ is a curve at $u \in E$ satisfying the ODEs $\dot{\gamma}_u(t) = X(\gamma_u(t))$ for all $t \in I$.

Proposition (Global Uniqueness). Suppose γ_1 and γ_2 are two integral curves of a vector–field X at a point $m \in M$. Then $\gamma_1 = \gamma_2$ on the intersection of their domains [AMR88].

If for every point $m \in M$ the curve γ_m is defined on the entire real axis $\mathbb{R}$, then the vector–field X is said to be *complete*.

The *support of a vector–field* X defined on a manifold M is defined to be the closure of the set $\{m \in M | X(m) = 0\}$. A C^∞ vector–field with compact support on a manifold M is complete. In particular, a C^∞ vector–field on a compact manifold is complete. Completeness corresponds to well–defined dynamics persisting eternally.

Now, following [AMR88], for the *derivative* of a C^∞ function $f : E \to \mathbb{R}$ in the direction X we use the notation $X[f] = df \cdot X$, where df stands for the *derivative map*. In standard coordinates on $\mathbb{R}^n$ this is a standard *gradient*

$$df(x) = \nabla f = (\partial_{x^1} f, ..., \partial_{x^n} f), \qquad \text{and} \qquad X[f] = X^i \partial_{x^i} f.$$

Let F_t be the flow of X. Then $f(F_t(x)) = f(F_s(x))$ if $t \geq s$.

For example, Newtonian equations for a moving particle of mass m in a potential field V in $\mathbb{R}^n$ are given by $\ddot{q}^i(t) = -(1/m)\nabla V\left(q^i(t)\right)$, for a smooth function $V : \mathbb{R}^n \to \mathbb{R}$. If there are constants $a, b \in \mathbb{R}$, $b \geq 0$ such that $(1/m)V(q^i) \geq a - b\left\|q^i\right\|^2$, then *every solution exists for all time*. To show this, rewrite the second–order equations as a first–order system $\dot{q}^i = (1/m)\, p_i$, $\dot{p}_i = -V(q^i)$ and note that the energy $E(q^i, p_i) = (1/2m)\left\| p_i\right\|^2 + V(q)$ is a *first integral* of the motion. Thus, for any solution $\left(q^i(t), p_i(t)\right)$ we have $E\left(q^i(t), p_i(t)\right) = E\left(q^i(0), p_i(0)\right) = V(q(0))$.

Let X_t be a C^∞ time–dependent vector–field on an $n-$manifold M, $k \geq 1$, and let m_0 be an *equilibrium* of X_t, that is, $X_t(m_0) = 0$ for all t. Then for any T there exists a neighborhood V of m_0 such that any $m \in V$ has integral curve existing for time $t \in [-T, T]$.

Dynamical Flows on M

Recall (1.2.5) that the *flow* F_t of a C^∞ vector–field $X \in \mathcal{X}^k(M)$ is the *one–parameter group of diffeomorphisms* $F_t : M \to M$ such that $t \mapsto F_t(m)$ is the integral curve of X with initial condition m for all $m \in M$ and $t \in I \subseteq \mathbb{R}$. The flow $F_t(m)$ is C^∞ by induction on k. It is defined as [AMR88]:

$$\frac{d}{dt}F_t(x) = X(F_t(x)).$$

Existence and uniqueness theorems for ODEs guarantee that F_t is smooth in m and t. From uniqueness, we get the *flow property*:

$$F_{t+s} = F_t \circ F_s$$

along with the initial conditions $F_0 =$ identity. The flow property generalizes the situation where $M = V$ is a linear space, $X(x) = A\,x$ for a (bounded) linear operator A, and where $F_t(x) = e^{tA}x$ – to the nonlinear case. Therefore, the flow $F_t(m)$ can be defined as a *formal exponential*

$$F_t(m) = \exp(t\,X) = (I + t\,X + \frac{t^2}{2}X^2 + ...) = \sum_{k=0}^{\infty} \frac{X^k t^k}{k!}.$$

recall that a *time–dependent vector–field* is a map $X : M \times \mathbb{R} \to TM$ such that $X(m,t) \in T_mM$ for each point $m \in M$ and $t \in \mathbb{R}$. An *integral curve* of X is a curve $\gamma(t)$ in M such that

$$\dot{\gamma}(t) = X\left(\gamma\left(t\right), t\right), \qquad \text{for all} \quad t \in I \subseteq \mathbb{R}.$$

In this case, the flow is the one–parameter group of diffeomorphisms $F_{t,s} : M \to M$ such that $t \mapsto F_{t,s}\left(m\right)$ is the integral curve $\gamma(t)$ with initial condition $\gamma(s) = m$ at $t = s$. Again, the existence and uniqueness theorem from ODE–theory applies here, and in particular, uniqueness gives the time–dependent flow property, i.e., the *Chapman–Kolmogorov law*

$$F_{t,r} = F_{t,s} \circ F_{s,r}.$$

If X happens to be time independent, the two notions of flows are related by $F_{t,s} = F_{t-s}$ (see [MR99]).

Categories of ODEs

Ordinary differential equations are naturally organized into their categories (see [Koc81]). First order ODEs are organized into a category ODE_1. A first–order ODE on a manifold–like object M is a vector–field $X : M \to TM$, and a *morphism of vector–fields* $(M_1, X_1) \to (M_2, X_2)$ is a map $f : M_1 \to M_2$ such that the following diagram commutes

$$\begin{array}{ccc} TM_1 & \xrightarrow{\quad Tf \quad} & TM_2 \\ {\scriptstyle X_1}\uparrow & & \uparrow{\scriptstyle X_2} \\ M_1 & \xrightarrow[\quad f \quad]{} & M_2 \end{array}$$

A *global solution* of the differential equation (M, X), or a *flow line* of a vector–field X, is a morphism from $\left(\mathbb{R}, \frac{\partial}{\partial x}\right)$ to (M, X).

Similarly, second–order ODEs are organized into a category ODE_2. A second–order ODE on M is usually constructed as a vector–field on TM, $\xi : TM \to TTM$, and a morphism of vector–fields $(M_1, \xi_1) \to (M_2, \xi_2)$ is a map $f : M_1 \to M_2$ such that the following diagram commutes

$$\begin{array}{ccc} TTM_1 & \xrightarrow{TTf} & TTM_2 \\ \uparrow{\scriptstyle \xi_1} & & \uparrow{\scriptstyle \xi_2} \\ TM_1 & \xrightarrow[Tf]{} & TM_2 \end{array}$$

Unlike solutions for first–order ODEs, solutions for second–order ODEs are not in general homomorphisms from $\mathbb{R}$, unless the second–order ODE is a *spray* [KR03].

Differential Forms on Smooth Manifolds

Recall (see Appendix, subsection 3.1.4) that *exterior differential forms* are a special kind of antisymmetrical covariant tensors, that formally occur as *integrands* under ordinary integral signs in $\mathbb{R}^3$. To give a more precise exposition, here we start with $1-$forms, which are dual to vector–fields, and after that introduce general $k-$forms.

$1-$Forms on M

Dual to the notion of a C^∞ vector–field X on an $n-$manifold M is a C^∞ *covector–field*, or a C^∞ $1-$form α, which is defined as a $C^\infty-$section of the cotangent bundle T^*M, i.e., $\alpha : M \to T^*M$ is smooth and $\pi^*_M \circ X = Id_M$. We denote the set of all C^∞ $1-$forms by $\Omega^1(M)$. A basic example of a $1-$form is the differential df of a real–valued function $f \in C^\infty(M, \mathbb{R})$. With point wise addition and scalar multiplication $\Omega^1(M)$ becomes a vector space.

In other words, a C^∞ $1-$form α on a C^∞ manifold M is a real–valued function on the set of all tangent vectors to M, i.e., $\alpha : TM \to \mathbb{R}$ with the following properties:

1. α is linear on the tangent space T_mM for each $m \in M$;
2. For any C^∞ vector–field $X \in \mathcal{X}^k(M)$, the function $f : M \to \mathbb{R}$ is C^∞.

Given a $1-$form α, for each point $m \in M$ the map $\alpha(m) : T_mM \to \mathbb{R}$ is an element of the dual space T^*_mM. Therefore, the space of $1-$forms $\Omega^1(M)$ is dual to the space of vector–fields $\mathcal{X}^k(M)$.

In particular, the *coordinate $1-$forms* $dx^1, ..., dx^n$ are locally defined at any point $m \in M$ by the property that for any vector–field $X = \left(X^1, ..., X^n\right) \in \mathcal{X}^k(M)$,

$$dx^i(X) = X^i.$$

The dx^i's form a *basis* for the 1−forms at any point $m \in M$, with local coordinates $\left(x^1, ..., x^n\right)$, so any 1−form α may be expressed in the form

$$\alpha = f_i(m)\, dx^i.$$

If a vector–field X on M has the form $X(m) = \left(X^1(m), ..., X^n(m)\right)$, then at any point $m \in M$,

$$\alpha_m(X) = f_i(m)\, X^i(m),$$

where $f \in C^\infty(M, \mathbb{R})$.

Suppose we have a 1D closed curve $\gamma = \gamma(t)$ inside a smooth manifold M. Using a simplified 'physical' notation, a 1–form $\alpha(x)$ defined at a point $x \in M$, given by

$$\alpha(x) = \alpha_i(x)\, dx^i, \tag{1.20}$$

can be unambiguously integrated over a curve $\gamma \in M$, as follows. Parameterize γ by a parameter t, so that its coordinates are given by $x^i(t)$. At time t, the velocity $\dot{x} = \dot{x}(t)$ is a tangent vector to M at $x(t)$. One can insert this tangent vector into the linear map $\alpha(x)$ to get a real number. By definition, inserting the vector $\dot{x}(t)$ into the linear map dx^i gives the component $\dot{x}^i = \dot{x}^i(t)$. Doing this for every t, we can then integrate over t,

$$\int \left(\alpha_i(x(t))\dot{x}^i\right) dt. \tag{1.21}$$

Note that this expression is independent of the parametrization in terms of t. Moreover, from the way that tangent vectors transform, one can deduce how the linear maps dx^i should transform, and from this how the coefficients $\alpha_i(x)$ should transform. Doing this, one sees that the above expression is also invariant under changes of coordinates on M. Therefore, a 1–form can be unambiguously integrated over a curve in M. We write such an integral as

$$\int_\gamma \alpha_i(x)\, dx^i, \qquad \text{or, even shorter, as} \qquad \int_\gamma \alpha.$$

Clearly, when M is itself a 1D manifold, (1.21) gives precisely the ordinary integration of a function $\alpha(x)$ over x, so the above notation is indeed natural.

The 1−forms on M are part of an algebra, called the *exterior algebra*, or *Grassmann algebra* on M. The multiplication $\wedge$ in this algebra is called *wedge product* (see (1.23) below), and it is skew–symmetric,

$$dx^i \wedge dx^j = -dx^j \wedge dx^i.$$

One consequence of this is that $dx^i \wedge dx^i = 0$.

$k-$Forms on M

A differential form, or an exterior form α of degree k, or a $k-$form for short, is a section of the vector bundle $\Lambda^k T^*M$, i.e., $\alpha : M \to \Lambda^k T^*M$. In other words, $\alpha(m) : T_mM \times ... \times T_mM \to \mathbb{R}$ (with k factors T_mM) is a function that assigns to each point $m \in M$ a skew–symmetric $k-$multilinear map on the tangent space T_mM to M at m. Without the skew–symmetry assumption, α would be called a $(0,k)-$tensor–field. The space of all $k-$forms is denoted by $\Omega^k(M)$. It may also be viewed as the space of all skew symmetric $(0,k)-$tensor–fields, the space of all maps

$$\Phi : \mathcal{X}^k(M) \times ... \times \mathcal{X}^k(M) \to C^\infty(M, \mathbb{R}),$$

which are $k-$linear and skew–symmetric (see (1.23) below). We put $\Omega^k(M) = C^\infty(M, \mathbb{R})$.

In particular, a $2-$form ω on an $n-$manifold M is a section of the vector bundle $\Lambda^2 T^*M$. If (U, ϕ) is a chart at a point $m \in M$ with local coordinates $\left(x^1, ..., x^n\right)$ let $\{e_1, ..., e_n\} = \{\partial_{x^1}, ..., \partial_{x^n}\}$ – be the corresponding basis for T_mM, and let $\left\{e^1, ..., e^n\right\} = \left\{dx^1, ..., dx^n\right\}$ – be the dual basis for T^*_mM. Then at each point $m \in M$, we can write a $2-$form ω as

$$\omega_m(v, u) = \omega_{ij}(m)\, v^i u^j, \quad \text{where } \omega_{ij}(m) = \omega_m(\partial_{x^i}, \partial_{x^j}).$$

Similarly to the case of a 1–form α (1.20), one would like to define a 2–form ω as something which can naturally be integrated over a 2D surface Σ within a smooth manifold M. At a specific point $x \in M$, the tangent plane to such a surface is spanned by a pair of tangent vectors, $(\dot{x}^1, \dot{x}^2)$. So, to generalize the construction of a 1–form, we should give a bilinear map from such a pair to $\mathbb{R}$. The most general form of such a map is

$$\omega_{ij}(x)\, dx^i \otimes dx^j, \tag{1.22}$$

where the tensor product of two cotangent vectors acts on a pair of vectors as,

$$dx^i \otimes dx^j\, (\dot{x}^1, \dot{x}^2) = dx^i(\dot{x}^1)\, dx^j(\dot{x}^2).$$

On the r.h.s. of this equation, one multiplies two ordinary numbers got by letting the linear map dx^i act on $\dot{x}^1$, and dx^j on $\dot{x}^2$.

However, the bilinear map (1.22) is slightly too general to give a good integration procedure. The reason is that we would like the integral to change sign if we change the orientation of integration, just like in the 1D case. In 2D, changing the orientation means exchanging $\dot{x}^1$ and $\dot{x}^2$, so we want our bilinear map to be antisymmetric under this exchange. This is achieved by defining a 2–form to be

$$\omega = \omega_{ij}(x)\, \left(dx^i \otimes dx^j - dx^j \otimes dx^i\right) \equiv \omega_{ij}(x)\, dx^i \wedge dx^j$$

We now see why a2–form corresponds to an *antisymmetric* tensor field: the symmetric part of ω_{ij} would give a vanishing contribution to ω. Now, parameterizing a surface Σ in M with two coordinates t_1 and t_2, and reasoning exactly like we did in the case of a 1–form, one can show that the integration of a 2–form over such a surface is indeed well–defined, and independent of the parametrization of both Σ and M.

If each summand of a differential form $\alpha \in \Omega^k(M)$ contains k basis $1-$forms dx^i's, the form is called a $k-$form. Functions $f \in C^\infty(M, \mathbb{R})$ are considered to be $0-$forms, and any form on an $n-$manifold M of degree $k > n$ must be zero due to the skew–symmetry.

Any $k-$form $\alpha \in \Omega^k(M)$ may be expressed in the form

$$\alpha = f_I \, dx^{i_1} \wedge ... \wedge dx^{i_k} = f_I \, dx^I,$$

where I is a *multiindex* $I = (i_1, ..., i_k)$ of length k, and $\wedge$ is the wedge product which is associative, bilinear and anticommutative.

Just as $1-$forms act on vector–fields to give real–valued functions, so $k-$forms act on $k-$tuples of vector–fields to give real–valued functions.

The *wedge product* of two differential forms, a $k-$form $\alpha \in \Omega^k(M)$ and an $l-$form $\beta \in \Omega^l(M)$ is a $(k+l)-$form $\alpha \wedge \beta$ defined as:

$$\alpha \wedge \beta = \frac{(k+l)!}{k!l!} \mathbf{A}(\alpha \otimes \beta), \tag{1.23}$$

where $\mathbf{A} : \Omega^k(M) \to \Omega^k(M)$, $\mathbf{A}\tau(e_1, ..., e_k) = \frac{1}{k!} \sum_{\sigma \in S_k} (\operatorname{sign} \sigma) \, \tau(e_{\sigma(1)}, ..., e_{\sigma(k)})$, where S_k is the permutation group on k elements consisting of all bijections $\sigma : \{1, ..., k\} \to \{1, ..., k\}$.

For any $k-$form $\alpha \in \Omega^k(M)$ and $l-$form $\beta \in \Omega^l(M)$, the wedge product is defined fiberwise, i.e., $(\alpha \wedge \beta)_m = \alpha_x \wedge \beta_m$ for each point $m \in M$. It is also associative, i.e., $(\alpha \wedge \beta) \wedge \gamma = \alpha \wedge (\beta \wedge \gamma)$, and graded commutative, i.e., $\alpha \wedge \beta = (-1)^{kl} \beta \wedge \alpha$. These properties are proved in multilinear algebra. So $M \Longrightarrow \Omega^k(M)$ is a contravariant functor from the category $\mathcal{M}$ into the category of real graded commutative algebras [KMS93].

Let M be an $n-$manifold, $X \in \mathcal{X}^k(M)$, and $\alpha \in \Omega^{k+1}(M)$. The *interior product*, or *contraction*, $i_X \alpha = X \rfloor \alpha \in \Omega^k(M)$ of X and α (with *insertion operator* i_X) is defined as

$$i_X \alpha(X^1, ..., X^k) = \alpha(X, X^1, ..., X^k).$$

Insertion operator i_X of a vector–field $X \in \mathcal{X}^k(M)$ is natural with respect to the pull–back F^* of a diffeomorphism $F : M \to N$ between two manifolds, i.e., the following diagram commutes:

$$\begin{array}{ccc} \Omega^k(N) & \xrightarrow{F^*} & \Omega^k(M) \\ \Big\downarrow{\scriptstyle i_X} & & \Big\downarrow{\scriptstyle i_{F^*X}} \\ \Omega^{k-1}(N) & \xrightarrow[F^*]{} & \Omega^{k-1}(M) \end{array}$$

Similarly, insertion operator i_X of a vector–field $X \in \mathcal{Y}^k(M)$ is natural with respect to the push–forward F_* of a diffeomorphism $F : M \to N$, i.e., the following diagram commutes:

$$\begin{array}{ccc} \Omega^k(M) & \xrightarrow{F_*} & \Omega^k(N) \\ \Big\downarrow{\scriptstyle i_Y} & & \Big\downarrow{\scriptstyle i_{F_*Y}} \\ \Omega^{k-1}(M) & \xrightarrow[F_*]{} & \Omega^{k-1}(N) \end{array}$$

In case of *Riemannian manifolds* there is another exterior operation. Let M be a smooth $n-$manifold with Riemannian metric $g = \langle , \rangle$ and the corresponding volume element μ. The *Hodge star operator* $* : \Omega^k(M) \to \Omega^{n-k}(M)$ on M is defined as

$$\alpha \wedge *\beta = \langle \alpha, \beta \rangle \, \mu \quad \text{for } \alpha, \beta \in \Omega^k(M).$$

The Hodge star operator satisfies the following properties for $\alpha, \beta \in \Omega^k(M)$ [AMR88]:

1. $\alpha \wedge *\beta = \langle \alpha, \beta \rangle \, \mu = \beta \wedge *\alpha$;
2. $*1 = \mu, \quad *\mu = (-1)^{Ind(g)}$;
3. $** \alpha = (-1)^{Ind(g)} (-1)^{k(n-k)} \alpha$;
4. $\langle \alpha, \beta \rangle = (-1)^{Ind(g)} \, \langle *\alpha, *\beta \rangle$, where $Ind(g)$ is the *index* of the metric g.

Exterior Differential Systems

Here we give an informal introduction to *exterior differential systems* (EDS, for short), which are expressions involving differential forms related to any manifold M. Later, when we fully develop the necessary differential geometrical as well as variational machinery (see (1.4.7) below), we will give a more precise definition of EDS.

Central in the language of EDS is the notion of *coframing*, which is a real finite–dimensional smooth manifold M with a given global *cobasis* and coordinates, but without requirement for a proper topological and differential structures. For example, $M = \mathbb{R}^3$ is a coframing with cobasis $\{dx, dy, dz\}$ and coordinates $\{x, y, z\}$. In addition to the cobasis and coordinates, a coframing can be given structure equations (1.2.9) and restrictions. For example, $M = \mathbb{R}^2 \backslash \{0\}$ is a coframing with cobasis $\{e^1, e^2\}$, a single coordinate $\{r\}$, structure equations $\{dr = e^1, \, de^1 = 0, \, de^2 = e^1 \wedge e^2 / r\}$ and restrictions $\{r \neq 0\}$.

A *system* S on M in EDS terminology is a list of expressions including differential forms (e.g., $S = \{dz - ydx\}$).

Now, a simple EDS is a triple (S, Ω, M), where S is a system on M, and Ω is an *independence condition*: either a decomposable $k-$form or a system of $k-$forms on M. An EDS is a list of simple EDS objects where the various coframings are all disjoint.

An *integral element* of an exterior system (S, Ω, M) is a subspace $P \subset T_m M$ of the tangent space at some point $m \in M$ such that all forms in S vanish when evaluated on vectors from P. Alternatively, an integral element $P \subset T_m M$ can be represented by its annihilator $P^\perp \subset T_m^* M$, comprising those $1-$forms at m which annul every vector in P. For example, with $M = \mathbb{R}^3 = \{(x, y, z)\}$, $S = \{dx \wedge dz\}$ and $\Omega = \{dx, dz\}$, the integral element $P = \{\partial_x + \partial_z, \partial_y\}$ is equally determined by its annihilator $P^\perp = \{dz - dx\}$. Again, for $S = \{dz - y dx\}$ and $\Omega = \{dx\}$, the integral element $P = \{\partial_x + y\partial_z\}$ can be specified as $\{dy\}$.

Exterior Derivative on a Smooth Manifold

The *exterior derivative* is an operation that takes $k-$forms to $(k+1)-$forms on a smooth manifold M. It defines a unique family of maps $d : \Omega^k(U) \to \Omega^{k+1}(U)$, U open in M, such that (see [AMR88]):

1. d is a $\wedge-$*antiderivation*; that is, d is $\mathbb{R}-$linear and for two forms $\alpha \in \Omega^k(U)$, $\beta \in \Omega^l(U)$,
$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^k \alpha \wedge d\beta.$$
2. If $f \in C^\infty(U, \mathbb{R})$ is a function on M, then $df = \frac{\partial f}{\partial x^i} dx^i : M \to T^*M$ is the differential of f, such that $df(X) = i_X df = \mathcal{L}_X f - d i_X f = \mathcal{L}_X f = X[f]$ for any $X \in \mathcal{X}^k(M)$.
3. $d^2 = d \circ d = 0$ (that is, $d^{k+1}(U) \circ d^k(U) = 0$).
4. d is natural with respect to restrictions $|U$; that is, if $U \subset V \subset M$ are open and $\alpha \in \Omega^k(V)$, then $d(\alpha|U) = (d\alpha)|U$, or the following diagram commutes:
$$\begin{array}{ccc} \Omega^k(V) & \xrightarrow{|U} & \Omega^k(U) \\ \Big\downarrow d & & \Big\downarrow d \\ \Omega^{k+1}(V) & \xrightarrow[|U]{} & \Omega^{k+1}(U) \end{array}$$
5. d is natural with respect to the Lie derivative $\mathcal{L}_X$ (1.3.3) along any vector–field $X \in \mathcal{X}^k(M)$; that is, for $\omega \in \Omega^k(M)$ we have $\mathcal{L}_X \omega \in \Omega^k(M)$ and $d\mathcal{L}_X \omega = \mathcal{L}_X d\omega$, or the following diagram commutes:
$$\begin{array}{ccc} \Omega^k(M) & \xrightarrow{\mathcal{L}_X} & \Omega^k(M) \\ \Big\downarrow d & & \Big\downarrow d \\ \Omega^{k+1}(M) & \xrightarrow[\mathcal{L}_X]{} & \Omega^{k+1}(M) \end{array}$$

6. Let $\varphi : M \to N$ be a C^∞ map of manifolds. Then $\varphi^* : \Omega^k(N) \to \Omega^k(M)$ is a homomorphism of differential algebras (with $\wedge$ and d) and d is natural with respect to $\varphi^* = F^*$; that is, $\varphi^* d\omega = d\varphi^* \omega$, or the following diagram commutes:

$$\begin{array}{ccc} \Omega^k(N) & \xrightarrow{\varphi^*} & \Omega^k(M) \\ \Big\downarrow d & & \Big\downarrow d \\ \Omega^{k+1}(N) & \xrightarrow[\varphi^*]{} & \Omega^{k+1}(M) \end{array}$$

7. Analogously, d is natural with respect to diffeomorphism $\varphi_* = (F^*)^{-1}$; that is, $\varphi_* d\omega = d\varphi_* \omega$, or the following diagram commutes:

$$\begin{array}{ccc} \Omega^k(N) & \xrightarrow{\varphi_*} & \Omega^k(M) \\ \Big\downarrow d & & \Big\downarrow d \\ \Omega^{k+1}(N) & \xrightarrow[\varphi_*]{} & \Omega^{k+1}(M) \end{array}$$

8. $\mathcal{L}_X = i_X \circ d + d \circ i_X$ for any $X \in \mathcal{X}^k(M)$ (the Cartan 'magic' formula).
9. $\mathcal{L}_X \circ d = d \circ \mathcal{L}_X$, i.e., $[\mathcal{L}_X, d] = 0$ for any $X \in \mathcal{X}^k(M)$.
10. $[\mathcal{L}_X, i_Y] = i_{[x,y]}$; in particular, $i_X \circ \mathcal{L}_X = \mathcal{L}_X \circ i_X$ for all $X, Y \in \mathcal{X}^k(M)$.

Given a $k-$form $\alpha = f_I \, dx^I \in \Omega^k(M)$, the exterior derivative is defined in local coordinates $\left(x^1, ..., x^n\right)$ of a point $m \in M$ as

$$d\alpha = d\left(f_I \, dx^I\right) = \frac{\partial f_I}{\partial x^{i_k}} dx^{i_k} \wedge dx^I = df_I \wedge dx^{i_1} \wedge ... \wedge dx^{i_k}.$$

In particular, the exterior derivative of a function $f \in C^\infty(M, \mathbb{R})$ is a $1-$form $df \in \Omega^1(M)$, with the property that for any $m \in M$, and $X \in \mathcal{X}^k(M)$,

$$df_m(X) = X(f),$$

i.e., $df_m(X)$ is a Lie derivative of f at m in the direction of X. Therefore, in local coordinates $\left(x^1, ..., x^n\right)$ of a point $m \in M$ we have

$$df = \frac{\partial f}{\partial x^i} dx^i.$$

For any two functions $f, g \in C^\infty(M, \mathbb{R})$, exterior derivative obeys the *Leibniz rule*:

$$d(fg) = g \, df + f \, dg,$$

and the *chain rule*:

$$d\left(g(f)\right) = g'(f) \, df.$$

A $k-$form $\alpha \in \Omega^k(M)$ is called *closed form* if $d\alpha = 0$, and it is called *exact form* if there exists a $(k-1)-$form $\beta \in \Omega^{k-1}(M)$ such that $\alpha = d\beta$. Since $d^2 = 0$, every exact form is closed. The converse is only partially true (Poincaré Lemma): every closed form is *locally exact*. This means that given a closed $k-$form $\alpha \in \Omega^k(M)$ on an open set $U \subset M$, any point $m \in U$ has a neighborhood on which there exists a $(k-1)-$form $\beta \in \Omega^{k-1}(U)$ such that $d\beta = \alpha|_U$.

The Poincaré lemma is a generalization and unification of two well–known facts in vector calculus:

1. If $\operatorname{curl} F = 0$, then locally $F = \operatorname{grad} f$;
2. If $\operatorname{div} F = 0$, then locally $F = \operatorname{curl} G$.

Poincaré lemma for contractible manifolds: Any closed form on a smoothly contractible manifold is exact.

Intuition Behind Cohomology

The simple formula $d^2 = 0$ leads to the important topological notion of *cohomology*. Let us try to solve the equation $d\omega = 0$ for a $p-$form ω. A trivial solution is $\omega = 0$. From the above formula, we can actually find a much larger class of trivial solutions: $\omega = d\alpha$ for a $(p-1)-$form α. More generally, if ω is any solution to $d\omega = 0$, then so is $\omega + d\alpha$. We want to consider these two solutions as equivalent:

$$\omega \sim \omega + \omega' \qquad \text{if} \qquad \omega' \in \operatorname{Im}\ d,$$

where $\operatorname{Im}\ d$ is the image of d, that is, the collection of all $p-$forms of the form $d\alpha$. (To be precise, the image of d contains $q-$forms for any $0 < q \leq n$, so we should restrict this image to the $p-$forms for the p we are interested in.) The set of all $p-$forms which satisfy $d\omega = 0$ is called the kernel of d, denoted $\operatorname{Ker} d$, so we are interested in $\operatorname{Ker} d$ up to the *equivalence classes* defined by adding elements of $\operatorname{Im}\ d$. (Again, strictly speaking, $\operatorname{Ker} d$ consists of $q-$forms for several values of q, so we should restrict it to the $p-$forms for our particular choice of p.) This set of equivalence classes is called $H^p(M)$, the $p-$th *de Rham cohomology group* of M,

$$H^p(M) = \frac{\operatorname{Ker} d}{\operatorname{Im}\ d}.$$

Clearly, $\operatorname{Ker} d$ is a group under addition: if two forms $\omega^{(1)}$ and $\omega^{(2)}$ satisfy $d\omega^{(1)} = d\omega^{(2)} = 0$, then so does $\omega^{(1)} + \omega^{(2)}$. Moreover, if we change $\omega^{(i)}$ by adding some $d\alpha^{(i)}$, the result of the addition will still be in the same cohomology class, since it differs from $\omega^{(1)} + \omega^{(2)}$ by $d(\alpha^{(1)} + \alpha^{(2)})$. Therefore, we can view this addition really as an addition of cohomology classes: $H^p(M)$ is itself an additive group. Also note that if $\omega^{(3)}$ and $\omega^{(4)}$ are in the same cohomology class (that is, their difference is of the form $d\alpha^{(3)}$), then so are $c\omega^{(3)}$ and

$c\omega^{(4)}$ for any constant factor c. In other words, we can multiply a cohomology class by a constant to obtain another cohomology class: cohomology classes actually form a *vector space*.

Intuition Behind Homology

Another operator similar to the exterior derivative d is the *boundary operator* δ, which maps compact submanifolds of a smooth manifold M to their boundary. Here, $\delta C = 0$ means that a submanifold C of M has no boundary, and $C = \delta U$ means that C is itself the boundary of some submanifold U. It is intuitively clear, and not very hard to prove, that $\delta^2 = 0$: the boundary of a compact submanifold does not have a boundary itself. That the objects on which δ acts are independent of its coordinates is also clear. So is the grading of the objects: the degree p is the dimension of the submanifold C.[8] What is less clear is that the collection of submanifolds actually forms a vector space, but one can always *define* this vector space to consist of formal linear combinations of submanifolds, and this is precisely how one proceeds. The pD elements of this vector space are called $p-$*chains*. One should think of $-C$ as C with its orientation reversed, and of the sum of two disjoint sets, C^1+C^2, as their union. The equivalence classes constructed from δ are called *homology* classes.

For example, in Figure 1.4, C^1 and C^2 both satisfy $\delta C = 0$, so they are elements of $\operatorname{Ker}\delta$. Moreover, it is clear that neither of them separately can be viewed as the boundary of another submanifold, so they are not in the trivial homology class $\operatorname{Im}\ \delta$. However, the boundary of U is $C^1 - C^2$. (The minus sign in front of C^2 is a result of the fact that C^2 itself actually has the wrong orientation to be considered a boundary of U.) This can be written as $C^1 - C^2 = \delta U$, or equivalently $C^1 = C^2 + \delta U$, showing that C^1 and C^2 are in the same homology class.

The cohomology groups for the $\delta-$operator are called *homology groups*, and denoted by $H_p(M)$, with a lower index.[9] The $p-$chains C that satisfy $\delta C = 0$ are called $p-$*cycles*. Again, the $H_p(M)$ only exist for $0 \leq p \leq n$.

There is an interesting relation between cohomology and homology groups. Note that we can construct a bilinear map from $H^p(M) \times H_p(M) \to \mathbb{R}$ by

$$([\omega], [C]) \mapsto \int_C \omega, \tag{1.24}$$

where $[\omega]$ denotes the cohomology class of a $p-$form ω, and $[\Sigma]$ the homology class of a $p-$cycle Σ. Using Stokes' theorem, it can be seen that the result does not depend on the representatives for either ω or C

[8] Note that here we have an example of an operator that maps objects of degree p to objects of degree $p-1$ instead of $p+1$.

[9] Historically, as can be seen from the terminology, homology came first and cohomology was related to it in the way we will discuss below. However, since the cohomology groups have a more natural additive structure, it is the name 'cohomology' which is actually used for generalizations.

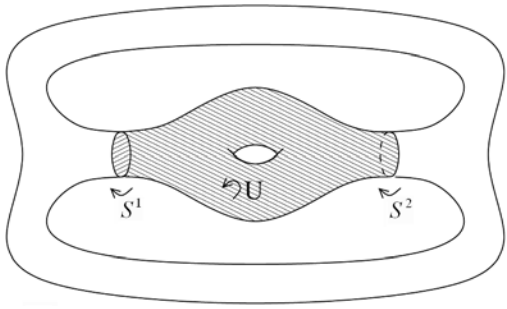


Fig. 1.4. The 1D submanifolds S^1 and S^2 represent the same homology class, since their difference is the boundary of U.

$$\begin{aligned}\int_{C+\delta U} \omega + d\alpha &= \int_C \omega + \int_C d\alpha + \int_{\delta U} \omega + d\alpha \\ &= \int_C \omega + \int_{\delta C} \alpha + \int_U d(\omega + d\alpha) = \int_C \omega,\end{aligned}$$

where we used that by the definition of (co)homology classes, $\delta C = 0$ and $d\omega = 0$. As a result, the above map is indeed well–defined on homology and cohomology classes. A very important theorem by de Rham says that this map is nondegenerate [Rha84]. This means that if we take some $[\omega]$ and we know the result of the map (1.24) for all $[C]$, this uniquely determines $[\omega]$, and similarly if we start by picking an $[C]$. This in particular means that the vector space $H^p(M)$ is the dual vector space of $H_p(M)$.

The de Rham Complex and Homotopy Operators on M

After an intuitive introduction of (co)homology ideas, we now turn to their proper definitions. Given a smooth manifold M, let $\Omega^p(M)$ denote the space of all smooth $p-$forms on M. The differential d, mapping $p-$forms to $(p+1)-$forms, serves to define the *de Rham complex* on M

$$0 \to \Omega^0(M) \xrightarrow{d^0} \Omega^1(M) \xrightarrow{d^1} ... \xrightarrow{d^{n-1}} \Omega^n(M) \to 0. \qquad (1.25)$$

Recall (from subsection (3.2.8) in the Appendix) that in general, a *complex* is defined as a sequence of vector spaces, and linear maps between successive spaces, with the property that the composition of any pair of successive maps is identically 0. In the case of the de Rham complex (1.25), this requirement is a restatement of the closure property for the exterior differential: $d \circ d = 0$.

In particular, for $n = 3$, the de Rham complex on a manifold M reads

$$0 \to \Omega^0(M) \xrightarrow{d^0} \Omega^1(M) \xrightarrow{d^1} \Omega^2(M) \xrightarrow{d^2} \Omega^3(M) \to 0. \quad (1.26)$$

If $\omega \equiv f(x,y,z) \in \Omega^0(M)$, then

$$d^0\omega \equiv d^0 f = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz = \operatorname{grad}\omega.$$

If $\omega \equiv f dx + g dy + h dz \in \Omega^1(M)$, then

$$d^1\omega \equiv \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\right) dx\wedge dy + \left(\frac{\partial h}{\partial y} - \frac{\partial g}{\partial z}\right) dy\wedge dz + \left(\frac{\partial f}{\partial z} - \frac{\partial h}{\partial x}\right) dz\wedge dx = \operatorname{curl}\omega.$$

If $\omega \equiv F dy\wedge dz + G dz\wedge dx + H dx\wedge dy \in \Omega^2(M)$, then

$$d^2\omega \equiv \frac{\partial F}{\partial x} + \frac{\partial G}{\partial y} + \frac{\partial H}{\partial z} = \operatorname{div}\omega.$$

Therefore, the de Rham complex (1.26) can be written as

$$0 \to \Omega^0(M) \xrightarrow{\text{grad}} \to \Omega^1(M) \xrightarrow{\text{curl}} \Omega^2(M) \xrightarrow{\text{div}} \Omega^3(M) \to 0.$$

Using the closure property for the exterior differential, $d \circ d = 0$, we get the standard identities from vector calculus

$$\operatorname{curl}\cdot\operatorname{grad} = 0 \qquad \text{and} \qquad \operatorname{div}\cdot\operatorname{curl} = 0.$$

The definition of the complex requires that the kernel of one of the linear maps contains the image of the preceding map. The complex is *exact* if this containment is *equality*. In the case of the de Rham complex (1.25), exactness means that a closed $p-$form ω, meaning that $d\omega = 0$, is necessarily an exact $p-$form, meaning that there exists a $(p-1)-$form θ such that $\omega = d\theta$. (For $p = 0$, it says that a smooth function f is closed, $df = 0$, iff it is constant). Clearly, any exact form is closed, but the converse need not hold. Thus the de Rham complex is *not* in general exact. The celebrated *de Rham theorem* states that the extent to which this complex fails to be exact measures purely topological information about the manifold M, its cohomology group.

On the local side, for special types of domains in Euclidean space $\mathbb{R}^m$, there is only trivial topology and we do have exactness of the de Rham complex (1.25). This result, known as the *Poincaré lemma*, holds for *star–shaped* domains $M \subset \mathbb{R}^m$: Let $M \subset \mathbb{R}^m$ be a star–shaped domain. Then the de Rham complex over M is exact.

The key to the proof of exactness of the de Rham complex lies in the construction of suitable *homotopy operators*. By definition, these are linear operators $h : \Omega^p \to \Omega^{p-1}$, taking differential $p-$forms into $(p-1)-$forms, and satisfying the basic identity [Olv86]

$$\omega = dh(\omega) + h(d\omega), \tag{1.27}$$

for all $p-$forms $\omega \in \Omega^p$. The discovery of such a set of operators immediately implies exactness of the complex. For if ω is closed, $d\omega = 0$, then (1.27) reduces to $\omega = d\theta$ where $\theta = h(\omega)$, so ω is exact.

Stokes Theorem and de Rham Cohomology of M

Stokes theorem states that if α is an $(n-1)-$form on an orientable $n-$manifold M, then the integral of $d\alpha$ over M equals the integral of α over ∂M, the boundary of M. The classical theorems of Gauss, Green, and Stokes are special cases of this result.

A *manifold with boundary* is a set M together with an atlas of charts (U, ϕ) with boundary on M. Define (see [AMR88]) the *interior* and *boundary* of M respectively as

$$\operatorname{Int} M = \bigcup_U \phi^{-1}\left(\operatorname{Int}\left(\phi(U)\right)\right), \qquad \text{and} \qquad \partial M = \bigcup_U \phi^{-1}\left(\partial\left(\phi(U)\right)\right).$$

If M is a manifold with boundary, then its interior $\operatorname{Int} M$ and its boundary ∂M are smooth manifolds without boundary. Moreover, if $f : M \to N$ is a diffeomorphism, N being another manifold with boundary, then f induces, by restriction, two diffeomorphisms

$$\operatorname{Int} f : \operatorname{Int} M \to \operatorname{Int} N, \qquad \text{and} \qquad \partial f : \partial M \to \partial N.$$

If $n = \dim M$, then $\dim(\operatorname{Int} M) = n$ and $\dim(\partial M) = n - 1$.

To integrate a differential $n-$form over an $n-$manifold M, M must be oriented. If $\operatorname{Int} M$ is oriented, we want to choose an *orientation* on ∂M compatible with it. As for manifolds without boundary a volume form on an $n-$manifold with boundary M is a nowhere vanishing $n-$form on M. Fix an orientation on $\mathbb{R}^n$. Then a chart (U, ϕ) is called *positively oriented* if the map $T_m\phi : T_mM \to \mathbb{R}^n$ is orientation preserving for all $m \in U$.

Let M be a compact, oriented kD smooth manifold with boundary ∂M. Let α be a smooth $(k-1)-$form on M. Then the classical *Stokes formula* holds

$$\int_M d\alpha = \int_{\partial M} \alpha.$$

If $\partial M = \emptyset$ then $\int_M d\alpha = 0$.

The quotient space

$$H^k(M) = \frac{\operatorname{Ker}\left(d : \Omega^k(M) \to \Omega^{k+1}(M)\right)}{\operatorname{Im}\left(d : \Omega^{k-1}(M) \to \Omega^k(M)\right)}$$

represents the kth *de Rham cohomology group* of a manifold M. recall that the *de Rham theorem* states that these Abelian groups are isomorphic to the so–called singular cohomology groups of M defined in algebraic topology in terms of simplices and that depend only on the topological structure of M and not on its differentiable structure. The isomorphism is provided by integration; the fact that the integration map drops to the preceding quotient is guaranteed by Stokes' theorem.

The exterior derivative commutes with the pull–back of differential forms. That means that the vector bundle $\Lambda^k T^*M$ is in fact the value of a functor,

which associates a bundle over M to each manifold M and a vector bundle homomorphism over φ to each (local) diffeomorphism φ between manifolds of the same dimension. This is a simple example of the concept of a natural bundle. The fact that the exterior derivative d transforms sections of $\Lambda^k T^*M$ into sections of $\Lambda^{k+1}T^*M$ for every manifold M can be expressed by saying that d is an operator from $\Lambda^k T^*M$ into $\Lambda^{k+1}T^*M$. That the exterior derivative d commutes with (local) diffeomorphisms now means, that d is a natural operator from the functor $\Lambda^k T^*$ into functor $\Lambda^{k+1}T^*$. If $k > 0$, one can show that d is the unique natural operator between these two natural bundles up to a constant. So even linearity is a consequence of naturality [KMS93].

Euler–Poincaré Characteristics of M

The *Euler–Poincaré characteristics* of a manifold M equals the sum of its *Betti numbers*

$$\chi(M) = \sum_{p=0}^{n} (-1)^p \, b_p.$$

In case of $2n$D oriented compact Riemannian manifold M (*Gauss–Bonnet theorem*) its Euler–Poincaré characteristics is equal

$$\chi(M) = \int_M \gamma,$$

where γ is a closed $2n$ form on M, given by

$$\gamma = \frac{(-1)^n}{(4\pi)^n n!} \epsilon^{1\ldots 2n}_{i_1 \ldots i_{2n}} \, \Omega^{i_1}_{i_2} \wedge \Omega^{i_{2n-1}}_{i_{2n}},$$

where Ω^i_j is the curvature $2-$form of a Riemannian connection on M.

Poincaré–Hopf theorem: The Euler–Poincaré characteristics $\chi(M)$ of a compact manifold M equals the sum of indices of zeros of any vector–field on M which has only isolated zeros.

Duality of Chains and Forms on M

In topology of finite–dimensional smooth (i.e., C^{p+1} with $p \geq 0$) manifolds, a fundamental notion is the *duality* between $p-$chains C and $p-$forms (i.e., $p-$cochains) ω on the smooth manifold M, or domains of integration and integrands – as an integral on M represents a bilinear functional (see [BM82, DP97])

$$\int_C \omega \equiv \langle C, \omega \rangle \,, \tag{1.28}$$

where the integral is called the *period* of ω. Period depends only on the cohomology class of ω and the homology class of C. A closed form (cocycle) is exact (coboundary) if all its periods vanish, i.e., $d\omega = 0$ implies $\omega = d\theta$. The duality (1.28) is based on the classical Stokes formula

$$\int_C d\omega = \int_{\partial C} \omega.$$

This is written in terms of scalar products on M as

$$\langle C, d\omega \rangle = \langle \partial C, \omega \rangle ,$$

where ∂C is the boundary of the $p-$chain C oriented coherently with C. While the boundary operator ∂ is a global operator, the coboundary operator, that is, the exterior derivative d, is local, and thus more suitable for applications. The main property of the exterior differential,

$$d^2 = 0 \qquad \text{implies} \qquad \partial^2 = 0,$$

can be easily proved by the use of Stokes' formula

$$\left\langle \partial^2 C, \omega \right\rangle = \langle \partial C, d\omega \rangle = \left\langle C, d^2\omega \right\rangle = 0.$$

The analysis of p–chains and p–forms on the finite–dimensional smooth manifold M is usually performed in (co)homology categories (see [DP97, Die88]) related to M.

Let $\mathcal{M}^\bullet$ denote the category of cochains, (i.e., p–forms) on the smooth manifold M. When $\mathcal{C} = \mathcal{M}^\bullet$, we have the category $\mathcal{S}^\bullet(\mathcal{M}^\bullet)$ of generalized cochain complexes $A^\bullet$ in $\mathcal{M}^\bullet$, and if $A' = 0$ for $n < 0$ we have a subcategory $\mathcal{S}^\bullet_{\mathcal{DR}}(\mathcal{M}^\bullet)$ of the de Rham differential complexes in $\mathcal{M}^\bullet$

$$\begin{aligned} A^\bullet_{DR} : \; 0 \to \Omega^0(M) \xrightarrow{\;d\;} \Omega^1(M) \xrightarrow{\;d\;} \Omega^2(M) \cdots \\ \cdots \xrightarrow{\;d\;} \Omega^n(M) \xrightarrow{\;d\;} \cdots . \end{aligned} \tag{1.29}$$

Here $A' = \Omega^n(M)$ is the vector space over $\mathbb{R}$ of all p–forms ω on M (for $p = 0$ the smooth functions on M) and $d_n = d : \Omega^{n-1}(M) \to \Omega^n(M)$ is the exterior differential. A form $\omega \in \Omega^n(M)$ such that $d\omega = 0$ is a closed form or n–cocycle. A form $\omega \in \Omega^n(M)$ such that $\omega = d\theta$, where $\theta \in \Omega^{n-1}(M)$, is an exact form or n–coboundary. Let $Z^n(M) = \mathrm{Ker}(d)$ (resp. $B^n(M) = \mathrm{Im}(d)$) denote a real vector space of cocycles (resp. coboundaries) of degree n. Since $d_{n+1}\, d_n = d^2 = 0$, we have $B^n(M) \subset Z^n(M)$. The quotient vector space

$$H^n_{DR}(M) = \mathrm{Ker}(d)/\,\mathrm{Im}(d) = Z^n(M)/B^n(M)$$

is the de Rham cohomology group. The elements of $H^n_{DR}(M)$ represent equivalence sets of cocycles. Two cocycles ω_1, ω_2 belong to the same equivalence set, or are cohomologous (written $\omega_1 \sim \omega_2$) iff they differ by a coboundary $\omega_1 - \omega_2 = d\theta$. The de Rham cohomology class of any form $\omega \in \Omega^n(M)$ is $[\omega] \in H^n_{DR}(M)$. The de Rham differential complex (1.29) can be considered as a system of second–order ODEs $d^2\theta = 0$, $\theta \in \Omega^{n-1}(M)$ having a solution represented by $Z^n(M) = \mathrm{Ker}(d)$.

Analogously let $\mathcal{M}_\bullet$ denote the category of chains on the smooth manifold M. When $\mathcal{C} = \mathcal{M}_\bullet$, we have the category $\mathcal{S}_\bullet(\mathcal{M}_\bullet)$ of generalized chain complexes $A_\bullet$ in $\mathcal{M}_\bullet$, and if $A_n = 0$ for $n < 0$ we have a subcategory $\mathcal{S}_\bullet^{\mathcal{C}}(\mathcal{M}_\bullet)$ of chain complexes in $\mathcal{M}_\bullet$

$$A_\bullet : 0 \leftarrow C^0(M) \xleftarrow{\partial} C^1(M) \xleftarrow{\partial} C^2(M) \cdots \xleftarrow{\partial} C^n(M) \xleftarrow{\partial} \cdots .$$

Here $A_n = C^n(M)$ is the vector space over $\mathbb{R}$ of all finite chains C on the manifold M and $\partial_n = \partial : C^{n+1}(M) \to C^n(M)$. A finite chain C such that $\partial C = 0$ is an $n-$cycle. A finite chain C such that $C = \partial B$ is an $n-$boundary. Let $Z_n(M) = \mathrm{Ker}(\partial)$ (resp. $B_n(M) = \mathrm{Im}(\partial)$) denote a real vector space of cycles (resp. boundaries) of degree n. Since $\partial_{n+1}\partial_n = \partial^2 = 0$, we have $B_n(M) \subset Z_n(M)$. The quotient vector space

$$H_n^C(M) = \mathrm{Ker}(\partial)/\,\mathrm{Im}(\partial) = Z_n(M)/B_n(M)$$

is the $n-$homology group. The elements of $H_n^C(M)$ are equivalence sets of cycles. Two cycles C_1, C_2 belong to the same equivalence set, or are homologous (written $C_1 \sim C_2$), iff they differ by a boundary $C_1 - C_2 = \partial B$). The homology class of a finite chain $C \in C^n(M)$ is $[C] \in H_n^C(M)$.

The dimension of the $n-$cohomology (resp. $n-$homology) group equals the nth Betti number b^n (resp. b_n) of the manifold M. *Poincaré lemma* says that on an open set $U \in M$ diffeomorphic to $\mathbb{R}^N$, all closed forms (cycles) of degree $p \geq 1$ are exact (boundaries). That is, the Betti numbers satisfy $b^p = 0$ (resp. $b_p = 0$) for $p = 1, \ldots, n$.

The *de Rham theorem* states the following. The map Φ: $H_n \times H^n \to \mathbb{R}$ given by $([C], [\omega]) \to \langle C, \omega \rangle$ for $C \in Z_n, \omega \in Z^n$ is a bilinear nondegenerate map which establishes the duality of the groups (vector spaces) H_n and H^n and the equality $b_n = b^n$.

Hodge Star Operator and Harmonic Forms

As the configuration manifold M is an oriented ND Riemannian manifold, we may select an orientation on all tangent spaces T_mM and all cotangent spaces T_m^*M, with the local coordinates $x^i = (q^i, p_i)$ at a point $m \in M$, in a consistent manner. The simplest way to do that is to choose the Euclidean orthonormal basis $\partial_1, ..., \partial_N$ of $\mathbb{R}^N$ as being positive.

Since the manifold M carries a Riemannian structure $g = \langle , \rangle$, we have a scalar product on each T_m^*M. So, we can define (as above) the linear *Hodge star operator*

$$* : \Lambda^p(T_m^*M) \to \Lambda^{N-p}(T_m^*M),$$

which is a base point preserving operator

$$* : \Omega^p(M) \to \Omega^{N-p}(M), \qquad (\Omega^p(M) = \Gamma(\Lambda^p(M)))$$

(here $\Lambda^p(V)$ denotes the p-fold exterior product of any vector space V, $\Omega^p(M)$ is a space of all $p-$forms on M, and $\Gamma(E)$ denotes the space of sections of the vector bundle E). Also,

$$** = (-1)^{p(N-p)} : \Lambda^p(T^*_x M) \to \Lambda^p(T^*_m M).$$

As the metric on $T^*_m M$ is given by $g^{ij}(x) = (g_{ij}(x))^{-1}$, we have the *volume form* defined in local coordinates as

$$*(1) \;=\; \sqrt{\det(g_{ij})}dx^1 \wedge ... \wedge dx^n, \qquad \text{and} \qquad \mathrm{vol}(M) \;=\; \int_M *(1).$$

For any to $p-$forms $\alpha, \beta \in \Omega^p(M)$ with compact support, we define the (bilinear and positive definite) L^2-product as

$$(\alpha, \beta) = \int_M \langle \alpha, \beta \rangle * (1) = \int_M \alpha \wedge *\beta.$$

We can extend the product $(\cdot, \cdot)$ to $L^2(\Omega^p(M))$; it remains bilinear and positive definite, because as usual, in the definition of L^2, functions that differ only on a set of measure zero are identified.

Using the Hodge star operator $*$, we can introduce the *codifferential* operator δ, which is formally adjoint to the exterior derivative $d : \Omega^p(M) \to \Omega^{p+1}(M)$ on $\oplus_{p=0}^N \Omega^p(M)$ w.r.t. $(\cdot, \cdot)$. This means that for $\alpha \in \Omega^{p-1}(M), \beta \in \Omega^p(M)$

$$(d\alpha, \beta) = (\alpha, \delta\beta).$$

Therefore, we have $\delta : \Omega^p(M) \to \Omega^{p-1}(M)$ and

$$\delta = (-1)^{N(p+1)+1} * d * .$$

Now, the *Laplace–Beltrami operator* (or, *Hodge Laplacian*, see [Gri83b, Voi02] as well as subsection (2.4.4) below), Δ on $\Omega^p(M)$, is defined by relation similar to (1.27) above

$$\Delta = d\delta + \delta d : \Omega^p(M) \to \Omega^p(M) \tag{1.30}$$

and an exterior differential form $\alpha \in \Omega^p(M)$ is called *harmonic* if $\Delta\alpha = 0$.

Let M be a compact, oriented Riemannian manifold, E a vector bundle with a bundle metric $\langle \cdot, \cdot \rangle$ over M,

$$D = d + A : \Omega^{p-1}(Ad_E) \to \Omega^p(Ad_E), \qquad \text{with} \;\; A \in \Omega^1(Ad_E)$$

– a tensorial and $\mathbb{R}-$linear metric connection on E with curvature $F_D \in \Omega^2(Ad_E)$ (Here by $\Omega^p(Ad_E)$ we denote the space of those elements of $\Omega^p(End_E)$ for which the endomorphism of each fibre is skew symmetric; End_E denotes the space of linear endomorphisms of the fibers of E).

1.2.6 Lie Derivative on a Smooth Manifold

Lie derivative is popularly called '*fisherman's derivative*'. In continuum mechanics it is called *Liouville operator*. This is a central differential operator in modern differential geometry and its physical and control applications.

Lie Derivative Operating on Functions

To define how vector–fields operate on functions on an $m-$manifold M, we will use the *Lie derivative*. Let $f : M \to \mathbb{R}$ so $Tf : TM \to T\mathbb{R} = \mathbb{R} \times \mathbb{R}$. Following [AMR88] we write Tf acting on a vector $v \in T_m M$ in the form

$$Tf \cdot v = (f(m), df(m) \cdot v)\,.$$

This defines, for each point $m \in M$, the element $df(m) \in T_m^* M$. Thus df is a section of the cotangent bundle T^*M, i.e., a $1-$form. The $1-$form $df : M \to T^*M$ defined this way is called the *differential* of f. If f is C^∞, then df is C^{k-1}.

If $\phi : U \subset M \to V \subset E$ is a local chart for M, then the local representative of $f \in C^\infty(M, \mathbb{R})$ is the map $f : V \to \mathbb{R}$ defined by $f = f \circ \phi^{-1}$. The local representative of Tf is the tangent map for local manifolds,

$$Tf(x, v) = (f(x), Df(x) \cdot v)\,.$$

Thus the local representative of df is the derivative of the local representative of f. In particular, if $(x^1, ..., x^n)$ are local coordinates on M, then the local components of df are

$$(df)^i = \partial_{x^i} f.$$

The introduction of df leads to the following definition of the Lie derivative. The *directional* or *Lie derivative* $\mathcal{L}_X : C^\infty(M, \mathbb{R}) \to C^{k-1}(M, \mathbb{R})$ of a function $f \in C^\infty(M, \mathbb{R})$ along a vector–field X is defined by

$$\mathcal{L}_X f(m) = X[f](m) = df(m) \cdot X(m),$$

for any $m \in M$. Denote by $X[f] = df(X)$ the map $M \ni m \mapsto X[f](m) \in \mathbb{R}$. If f is $F-$valued, the same definition is used, but now $X[f]$ is $F-$valued.

If a local chart (U, ϕ) on an $n-$manifold M has local coordinates $(x^1, ..., x^n)$, the local representative of $X[f]$ is given by the function

$$\mathcal{L}_X f = X[f] = X^i\, \partial_{x^i} f.$$

Evidently if f is C^∞ and X is C^{k-1} then $X[f]$ is C^{k-1}.

Let $\varphi : M \to N$ be a diffeomorphism. Then $\mathcal{L}_X$ is natural with respect to push–forward by φ. That is, for each $f \in C^\infty(M, \mathbb{R})$,

$$\mathcal{L}_{\varphi_* X}(\varphi_* f) = \varphi_* \mathcal{L}_X f,$$

i.e., the following diagram commutes:

$$\begin{array}{ccc} C^\infty(M, \mathbb{R}) & \xrightarrow{\varphi_*} & C^\infty(N, \mathbb{R}) \\ \Big\downarrow{\scriptstyle \mathcal{L}_X} & & \Big\downarrow{\scriptstyle \mathcal{L}_{\varphi_* X}} \\ C^\infty(M, \mathbb{R}) & \xrightarrow[\varphi_*]{} & C^\infty(N, \mathbb{R}) \end{array}$$

Also, $\mathcal{L}_X$ is natural with respect to restrictions. That is, for U open in M and $f \in C^\infty(M, \mathbb{R})$,

$$\mathcal{L}_{X|U}(f|U) = (\mathcal{L}_X f)|U,$$

where $|U : C^\infty(M, \mathbb{R}) \to C^\infty(U, \mathbb{R})$ denotes restriction to U, i.e., the following diagram commutes:

$$\begin{array}{ccc} C^\infty(M,\mathbb{R}) & \xrightarrow{|U} & C^\infty(U,\mathbb{R}) \\ \Big\downarrow{\scriptstyle \mathcal{L}_X} & & \Big\downarrow{\scriptstyle \mathcal{L}_{X|U}} \\ C^\infty(M,\mathbb{R}) & \xrightarrow[|U]{} & C^\infty(U,\mathbb{R}) \end{array}$$

Since $\varphi^* = (\varphi^{-1})_*$ the Lie derivative is also natural with respect to pull–back by φ. This has a generalization to $\varphi-$related vector–fields as follows: Let $\varphi : M \to N$ be a $C^\infty-$map, $X \in \mathcal{X}^{k-1}(M)$ and $Y \in \mathcal{X}^{k-1}(N)$, $k \geq 1$. If $X \sim_\varphi Y$, then

$$\mathcal{L}_X(\varphi^* f) = \varphi^* \mathcal{L}_Y f$$

for all $f \in C^\infty(N, \mathbb{R})$, i.e., the following diagram commutes:

$$\begin{array}{ccc} C^\infty(N,\mathbb{R}) & \xrightarrow{\varphi^*} & C^\infty(M,\mathbb{R}) \\ \Big\downarrow{\scriptstyle \mathcal{L}_Y} & & \Big\downarrow{\scriptstyle \mathcal{L}_X} \\ C^\infty(N,\mathbb{R}) & \xrightarrow[\varphi^*]{} & C^\infty(M,\mathbb{R}) \end{array}$$

The Lie derivative map $\mathcal{L}_X : C^\infty(M, \mathbb{R}) \to C^{k-1}(M, \mathbb{R})$ is a *derivation*, i.e., for two functions $f, g \in C^\infty(M, \mathbb{R})$ the *Leibniz rule* is satisfied

$$\mathcal{L}_X(fg) = g\mathcal{L}_X f + f\mathcal{L}_X g;$$

Also, Lie derivative of a constant function is zero, $\mathcal{L}_X(\text{const}) = 0$.

The connection between the Lie derivative $\mathcal{L}_X f$ of a function $f \in C^\infty(M, \mathbb{R})$ and the flow F_t of a vector–field $X \in \mathcal{X}^{k-1}(M)$ is given as:

$$\frac{d}{dt}\left(F_t^* f\right) = F_t^*\left(\mathcal{L}_X f\right).$$

Lie Derivative of Vector Fields

If $X, Y \in \mathcal{X}^k(M)$, $k \geq 1$ are two vector–fields on M, then

$$[\mathcal{L}_X, \mathcal{L}_Y] = \mathcal{L}_X \circ \mathcal{L}_Y - \mathcal{L}_Y \circ \mathcal{L}_X$$

is a derivation map from $C^{k+1}(M, \mathbb{R})$ to $C^{k-1}(M, \mathbb{R})$. Then there is a unique vector–field, $[X, Y] \in \mathcal{X}^k(M)$ of X and Y such that $\mathcal{L}_{[X,Y]} = [\mathcal{L}_X, \mathcal{L}_Y]$ and

$[X,Y](f) = X\left(Y(f)\right) - Y\left(X(f)\right)$ holds for all functions $f \in C^\infty(M,\mathbb{R})$. This vector–field is also denoted $\mathcal{L}_X Y$ and is called the Lie derivative of Y with respect to X, or the *Lie bracket* of X and Y. In a local chart (U,ϕ) at a point $m \in M$ with coordinates $(x^1,...,x^n)$, for $X|_U = X^i\partial_{x^i}$ and $Y|_U = Y^i\partial_{x^i}$ we have

$$\left[X^i\partial_{x^i}, Y^j\partial_{x^j}\right] = \left(X^i\left(\partial_{x^i}Y^j\right) - Y^i\left(\partial_{x^i}X^j\right)\right)\partial_{x^j},$$

since second partials commute. If, also X has flow F_t, then [AMR88]

$$\frac{d}{dt}\left(F_t^* Y\right) = F_t^*\left(\mathcal{L}_X Y\right).$$

In particular, if $t = 0$, this formula becomes

$$\frac{d}{dt}\,|_{t=0}\left(F_t^* Y\right) = \mathcal{L}_X Y.$$

Then the unique C^{k-1} vector–field $\mathcal{L}_X Y = [X,Y]$ on M defined by

$$[X,Y] = \frac{d}{dt}\,|_{t=0}\left(F_t^* Y\right),$$

is called the Lie derivative of Y with respect to X, or the Lie bracket of X and Y, and can be interpreted as the leading order term that results from the sequence of flows

$$F_t^{-Y} \circ F_t^{-X} \circ F_t^{Y} \circ F_t^{-X}(m) = \epsilon^2[X,Y](m) + \mathcal{O}(\epsilon^3), \tag{1.31}$$

for some real $\epsilon > 0$. Therefore a Lie bracket can be interpreted as a 'new direction' in which the system can flow, by executing the sequence of flows (1.31).

Lie bracket satisfies the following property:

$$[X,Y][f] = X[Y[f]] - Y[X[f]],$$

for all $f \in C^{k+1}(U,\mathbb{R})$, where U is open in M.

An important relationship between flows of vector–fields is given by the *Campbell–Baker–Hausdorff formula*:

$$F_t^Y \circ F_t^X = F_t^{X+Y+\frac{1}{2}[X,Y]+\frac{1}{12}([X,[X,Y]]-[Y,[X,Y]])+\ldots} \tag{1.32}$$

Essentially, if given the composition of multiple flows along multiple vector–fields, this formula gives the one flow along one vector–field which results in the same net flow. One way to prove the Campbell–Baker–Hausdorff formula (1.32) is to expand the product of two formal exponentials and equate terms in the resulting formal power series.

Lie bracket is the $\mathbb{R}-$bilinear map $[,] : \mathcal{X}^k(M) \times \mathcal{X}^k(M) \to \mathcal{X}^k(M)$ with the following properties:

1. $[X,Y] = -[Y,X]$, i.e., $\mathcal{L}_X Y = -\mathcal{L}_Y X$ for all $X,Y \in \mathcal{X}^k(M)$ – skew–symmetry;
2. $[X,X] = 0$ for all $X \in \mathcal{X}^k(M)$;
3. $[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0$ for all $X,Y,Z \in \mathcal{X}^k(M)$ – the Jacobi identity;
4. $[fX,Y] = f[X,Y] - (Yf)X$, i.e., $\mathcal{L}_{fX}(Y) = f(\mathcal{L}_X Y) - (\mathcal{L}_Y f)X$ for all $X,Y \in \mathcal{X}^k(M)$ and $f \in C^\infty(M,\mathbb{R})$;
5. $[X,fY] = f[X,Y] + (Xf)Y$, i.e., $\mathcal{L}_X(fY) = f(\mathcal{L}_X Y) + (\mathcal{L}_X f)Y$ for all $X,Y \in \mathcal{X}^k(M)$ and $f \in C^\infty(M,\mathbb{R})$;
6. $[\mathcal{L}_X, \mathcal{L}_Y] = \mathcal{L}_{[x,y]}$ for all $X,Y \in \mathcal{X}^k(M)$.

The pair $(\mathcal{X}^k(M),[,])$ is the prototype of a *Lie algebra* [KMS93]. In more general case of a *general linear Lie algebra* $\mathfrak{gl}(n)$, which is the Lie algebra associated to the Lie group $GL(n)$, Lie bracket is given by a *matrix commutator*

$$[A,B] = AB - BA,$$

for any two matrices $A,B \in \mathfrak{gl}(n)$.

Let $\varphi : M \to N$ be a diffeomorphism. Then $\mathcal{L}_X : \mathcal{X}^k(M) \to \mathcal{X}^k(M)$ is natural with respect to push–forward by φ. That is, for each $f \in C^\infty(M,\mathbb{R})$,

$$\mathcal{L}_{\varphi_* X}(\varphi_* Y) = \varphi_* \mathcal{L}_X Y,$$

i.e., the following diagram commutes:

$$\begin{array}{ccc} \mathcal{X}^k(M) & \xrightarrow{\varphi_*} & \mathcal{X}^k(N) \\ \Big\downarrow{\scriptstyle \mathcal{L}_X} & & \Big\downarrow{\scriptstyle \mathcal{L}_{\varphi_* X}} \\ \mathcal{X}^k(M) & \xrightarrow[\varphi_*]{} & \mathcal{X}^k(N) \end{array}$$

Also, $\mathcal{L}_X$ is natural with respect to restrictions. That is, for U open in M and $f \in C^\infty(M,\mathbb{R})$,

$$[X|U, Y|U] = [X,Y]|U,$$

where $U : C^\infty(M,\mathbb{R}) \to C^\infty(U,\mathbb{R})$ denotes restriction to U, i.e., the following diagram commutes [AMR88]:

$$\begin{array}{ccc} \mathcal{X}^k(M) & \xrightarrow{|U} & \mathcal{X}^k(U) \\ \Big\downarrow{\scriptstyle \mathcal{L}_X} & & \Big\downarrow{\scriptstyle \mathcal{L}_{X|U}} \\ \mathcal{X}^k(M) & \xrightarrow[|U]{} & \mathcal{X}^k(U) \end{array}$$

If a local chart (U,ϕ) on an $n-$manifold M has local coordinates $(x^1,...,x^n)$, then the local components of a Lie bracket are

$$[X,Y]^j = X^i\,\partial_{x^i}Y^j - Y^i\,\partial_{x^i}X^j,$$

that is, $[X,Y] = (X\cdot\nabla)Y - (Y\cdot\nabla)X$.

Let $\varphi : M \to N$ be a $C^\infty-$map, $X \in \mathcal{X}^{k-1}(M)$ and $Y \in \mathcal{X}^{k-1}(N)$, $k \geq 1$. Then $X \sim_\varphi Y$, iff

$$(Y[f]) \circ \varphi = X[f \circ \varphi]$$

for all $f \in C^\infty(V,\mathbb{R})$, where V is open in N.

For every $X \in \mathcal{X}^k(M)$, the operator $\mathcal{L}_X$ is a derivation on $\big(C^\infty(M,\mathbb{R}), \mathcal{X}^k(M)\big)$, i.e., $\mathcal{L}_X$ is $\mathbb{R}-$linear.

For any two vector–fields $X \in \mathcal{X}^k(M)$ and $Y \in \mathcal{X}^k(N)$, $k \geq 1$ with flows F_t and G_t, respectively, if $[X,Y] = 0$ then $F_t^*Y = Y$ and $G_t^*X = X$.

Derivative of the Evolution Operator

Recall that the time–dependent flow or evolution operator $F_{t,s}$ of a vector–field $X \in \mathcal{X}^k(M)$ is defined by the requirement that $t \mapsto F_{t,s}(m)$ be the integral curve of X starting at a point $m \in M$ at time $t = s$, i.e.,

$$\frac{d}{dt}F_{t,s}(m) = X\left(t, F_{t,s}(m)\right) \qquad \text{and} \qquad F_{t,t}(m) = m.$$

By uniqueness of integral curves we have $F_{t,s} \circ F_{s,r} = F_{t,r}$ (replacing the flow property $F_{t+s} = F_t + F_s$) and $F_{t,t} =$ identity.

Let $X_t \in \mathcal{X}^k(M)$, $k \geq 1$ for each t and suppose $X(t,m)$ is continuous in $(t,m) \in \mathbb{R}\times M$. Then $F_{t,s}$ is of class C^∞ and for $f \in C^{k+1}(M,\mathbb{R})$ [AMR88], and $Y \in \mathcal{X}^k(M)$, we have

1. $\frac{d}{dt}F_{t,s}^*\,f = F_{t,s}^*\left(\mathcal{L}_{X_t}\,f\right)$, and
2. $\frac{d}{dt}F_{t,s}^*\,f = F_{t,s}^*([X_t,Y]) = F_{t,s}^*\left(\mathcal{L}_{X_t}\,Y\right)$.

From the above theorem, the following identity holds:

$$\frac{d}{dt}F_{t,s}^*\,f = -X_t\left[F_{t,s}^*\,f\right].$$

Lie Derivative of Differential Forms

Since $\mathcal{F} : M \Longrightarrow \Lambda^k T^*M$ is a vector bundle functor on $\mathcal{M}$, the Lie derivative (1.3.3) of a $k-$form $\alpha \in \Omega^k(M)$ along a vector–field $X \in \mathcal{X}^k(M)$ is defined by

$$\mathcal{L}_X\alpha = \frac{d}{dt}\,|_{t=0}\; F_t^*\alpha.$$

It has the following properties:

1. $\mathcal{L}_X(\alpha\wedge\beta) = \mathcal{L}_X\,\alpha\wedge\beta + \alpha\wedge\mathcal{L}_X\,\beta$, so $\mathcal{L}_X$ is a derivation.
2. $[\mathcal{L}_X,\mathcal{L}_Y]\;\alpha = \mathcal{L}_{[X,Y]}\,\alpha$.
3. $\frac{d}{dt}F_t^*\alpha = F_t^*\mathcal{L}_X\alpha = \mathcal{L}_X\left(F_t^*\alpha\right)$.

Formula (3) holds also for time–dependent vector–fields in the sense that $\frac{d}{dt}F_{t,s}^{*}\alpha = F_{t,s}^{*}\mathcal{L}_X\alpha = \mathcal{L}_X\left(F_{t,s}^{*}\alpha\right)$ and in the expression $\mathcal{L}_X\alpha$ the vector–field X is evaluated at time t.

The famous *Cartan magic formula* (see [MR99]) states: the Lie derivative of a $k-$form $\alpha \in \Omega^k(M)$ along a vector–field $X \in \mathcal{X}^k(M)$ on a smooth manifold M is defined as

$$\mathcal{L}_X\alpha = di_X\alpha + i_X d\alpha = d(X\rfloor\alpha) + X\rfloor d\alpha.$$

Also, the following identities hold [MR99, KMS93]:

1. $\mathcal{L}_{fX}\alpha = f\mathcal{L}_X\alpha + df \wedge i_x\alpha$.
2. $\mathcal{L}_{[X,Y]}\alpha = \mathcal{L}_X\mathcal{L}_Y\alpha - \mathcal{L}_Y\mathcal{L}_X\alpha$.
3. $i_{[X,Y]}\alpha = \mathcal{L}_X i_Y\alpha - i_Y\mathcal{L}_X\alpha$.
4. $\mathcal{L}_X d\alpha = d\mathcal{L}_X\alpha$, i.e., $[\mathcal{L}_X, d] = 0$.
5. $\mathcal{L}_X i_X\alpha = i_X\mathcal{L}_X\alpha$, i.e., $[\mathcal{L}_X, i_X] = 0$.
6. $\mathcal{L}_X(\alpha\wedge\beta) = \mathcal{L}_X\alpha\wedge\beta + \alpha\wedge\mathcal{L}_X\beta$.

Lie Derivative of Various Tensor Fields

In this subsection, we use local coordinates x^i $(i = 1, ..., n)$ on a biomechanical $n-$manifold M, to calculate the Lie derivative $\mathcal{L}_{X^i}$ with respect to a generic vector–field X^i. (As always, $\partial_{x^i} \equiv \frac{\partial}{\partial x^i}$).

Lie Derivative of a Scalar Field

Given the scalar field ϕ, its Lie derivative $\mathcal{L}_{X^i}\phi$ is given as

$$\mathcal{L}_{X^i}\phi = X^i\partial_{x^i}\phi = X^1\partial_{x^1}\phi + X^2\partial_{x^2}\phi + ... + X^n\partial_{x^n}\phi.$$

Lie Derivative of Vector and Covector–Fields

Given a contravariant vector–field V^i, its Lie derivative $\mathcal{L}_{X^i}V^i$ is given as

$$\mathcal{L}_{X^i}V^i = X^k\partial_{x^k}V^i - V^k\partial_{x^k}X^i \equiv [X^i, V^i] - \text{the Lie bracket.}$$

Given a covariant vector–field (i.e., a one–form) ω_i, its Lie derivative $\mathcal{L}_{X^i}\omega_i$ is given as

$$\mathcal{L}_{X^i}\omega_i = X^k\partial_{x^k}\omega_i + \omega_k\partial_{x^i}X^k.$$

Lie Derivative of a Second–Order Tensor–Field

Given a $(2,0)$ tensor–field S^{ij}, its Lie derivative $\mathcal{L}_{X^i}S^{ij}$ is given as

$$\mathcal{L}_{X^i}S^{ij} = X^i\partial_{x^i}S^{ij} - S^{ij}\partial_{x^i}X^i - S^{ii}\partial_{x^i}X^j.$$

Given a $(1,1)$ tensor–field S^i_j, its Lie derivative $\mathcal{L}_{X^i}S^i_j$ is given as

$$\mathcal{L}_{X^i}S^i_j = X^i\partial_{x^i}S^i_j - S^i_j\partial_{x^i}X^i + S^i_i\partial_{x^j}X^i.$$

Given a $(0,2)$ tensor–field S_{ij}, its Lie derivative $\mathcal{L}_{X^i}S_{ij}$ is given as

$$\mathcal{L}_{X^i}S_{ij} = X^i\partial_{x^i}S_{ij} + S_{ij}\partial_{x^i}X^i + S_{ii}\partial_{x^j}X^i.$$

Lie Derivative of a Third–Order Tensor–Field

Given a $(3,0)$ tensor–field T^{ijk}, its Lie derivative $\mathcal{L}_{X^i}T^{ijk}$ is given as

$$\mathcal{L}_{X^i}T^{ijk} = X^i\partial_{x^i}T^{ijk} - T^{ijk}\partial_{x^i}X^i - T^{iik}\partial_{x^i}X^j - T^{iji}\partial_{x^i}X^k.$$

Given a $(2,1)$ tensor–field T^{ij}_k, its Lie derivative $\mathcal{L}_{X^i}T^{ij}_k$ is given as

$$\mathcal{L}_{X^i}T^{ij}_k = X^i\partial_{x^i}T^{ij}_k - T^{ij}_k\partial_{x^i}X^i + T^{ij}_i\partial_{x^k}X^i - T^{ii}_k\partial_{x^i}X^j.$$

Given a $(1,2)$ tensor–field T^i_{jk}, its Lie derivative $\mathcal{L}_{X^i}T^i_{jk}$ is given as

$$\mathcal{L}_{X^i}T^i_{jk} = X^i\partial_{x^i}T^i_{jk} - T^i_{jk}\partial_{x^i}X^i + T^i_{ik}\partial_{x^j}X^i + T^i_{ji}\partial_{x^k}X^i.$$

Given a $(0,3)$ tensor–field T_{ijk}, its Lie derivative $\mathcal{L}_{X^i}T_{ijk}$ is given as

$$\mathcal{L}_{X^i}T_{ijk} = X^i\partial_{x^i}T_{ijk} + T_{ijk}\partial_{x^i}X^i + T_{iik}\partial_{x^j}X^i + T_{iji}\partial_{x^k}X^i.$$

Lie Derivative of a Fourth–Order Tensor–Field

Given a $(4,0)$ tensor–field R^{ijkl}, its Lie derivative $\mathcal{L}_{X^i}R^{ijkl}$ is given as

$$\mathcal{L}_{X^i}R^{ijkl} = X^i\partial_{x^i}R^{ijkl} - R^{ijkl}\partial_{x^i}X^i - R^{iikl}\partial_{x^i}X^j - R^{ijil}\partial_{x^i}X^k - R^{ijki}\partial_{x^i}X^l.$$

Given a $(3,1)$ tensor–field R^{ijk}_l, its Lie derivative $\mathcal{L}_{X^i}R^{ijk}_l$ is given as

$$\mathcal{L}_{X^i}R^{ijk}_l = X^i\partial_{x^i}R^{ijk}_l - R^{ijk}_l\partial_{x^i}X^i + R^{ijk}_i\partial_{x^l}X^i - R^{iik}_l\partial_{x^i}X^j - R^{iji}_l\partial_{x^i}X^k.$$

Given a $(2,2)$ tensor–field R^{ij}_{kl}, its Lie derivative $\mathcal{L}_{X^i}R^{ij}_{kl}$ is given as

$$\mathcal{L}_{X^i}R^{ij}_{kl} = X^i\partial_{x^i}R^{ij}_{kl} - R^{ij}_{kl}\partial_{x^i}X^i + R^{ij}_{il}\partial_{x^k}X^i + R^{ij}_{ki}\partial_{x^l}X^i - R^{ii}_{kl}\partial_{x^i}X^j.$$

Given a $(1,3)$ tensor–field R^i_{jkl}, its Lie derivative $\mathcal{L}_{X^i}R^i_{jkl}$ is given as

$$\mathcal{L}_{X^i}R^i_{jkl} = X^i\partial_{x^i}R^i_{jkl} - R^i_{jkl}\partial_{x^i}X^i + R^i_{ikl}\partial_{x^j}X^i + R^i_{jil}\partial_{x^k}X^i + R^i_{jki}\partial_{x^l}X^i.$$

Given a $(0,4)$ tensor–field R_{ijkl}, its Lie derivative $\mathcal{L}_{X^i}R_{ijkl}$ is given as

$$\mathcal{L}_{X^i}R_{ijkl} = X^i\partial_{x^i}R_{ijkl} + R_{ijkl}\partial_{x^i}X^i + R_{iikl}\partial_{x^j}X^i + R_{ijil}\partial_{x^k}X^i + R_{ijki}\partial_{x^l}X^i.$$

Finally, recall that a *spinor* is a two–component complex column vector. Physically, spinors can describe both bosons and fermions, while tensors can describe only bosons. The Lie derivative of a spinor ϕ is defined by

$$\mathcal{L}_X\phi(x) = \lim_{t\to 0}\frac{\bar{\phi}_t(x) - \phi(x)}{t},$$

where $\bar{\phi}_t$ is the image of ϕ by a one–parameter group of isometries with X its generator. For a vector–field X^a and a covariant derivative ∇_a, the Lie derivative of ϕ is given explicitly by

$$\mathcal{L}_X\phi = X^a\nabla_a\phi - \frac{1}{8}(\nabla_a X_b - \nabla_b X_a)\,\gamma^a\gamma^b\phi,$$

where γ^a and γ^b are *Dirac matrices* (see, e.g., [BM00]).

Lie Algebras

Recall from Introduction that an *algebra* A is a vector space with a product. The product must have the property that

$$a(uv) = (au)v = u(av),$$

for every $a \in \mathbb{R}$ and $u, v \in A$. A map $\phi : A \to A'$ between algebras is called an *algebra homomorphism* if $\phi(u \cdot v) = \phi(u) \cdot \phi(v)$. A vector subspace $\mathfrak{I}$ of an algebra A is called a *left ideal* (resp. *right ideal*) if it is closed under algebra multiplication and if $u \in A$ and $i \in \mathfrak{I}$ implies that $ui \in \mathfrak{I}$ (resp. $iu \in \mathfrak{I}$). A subspace $\mathfrak{I}$ is said to be a *two–sided ideal* if it is both a left and right ideal. An ideal may not be an algebra itself, but the quotient of an algebra by a two–sided ideal inherits an algebra structure from A.

A *Lie algebra* is an algebra A where the multiplication, i.e., the *Lie bracket* $(u, v) \mapsto [u, v]$, has the following properties:

LA 1. $[u, u] = 0$ for every $u \in A$, and

LA 2. $[u, [v, w]] + [w, [u, v]] + [v, w, u]] = 0$ for all $u, v, w \in A$.

The condition LA 2 is usually called *Jacobi identity*. A subspace $E \subset A$ of a Lie algebra is called a *Lie subalgebra* if $[u, v] \in E$ for every $u, v \in E$. A map $\phi : A \to A'$ between Lie algebras is called a *Lie algebra homomorphism* if $\phi([u, v]) = [\phi(u), \phi(v)]$ for each $u, v \in A$.

All Lie algebras (over a given field $\mathbb{K}$) and all smooth homomorphisms between them form the category $\mathcal{LAL}$, which is itself a complete subcategory of the category $\mathcal{AL}$ of all algebras and their homomorphisms.

1.2.7 Lie Groups and Associated Lie Algebras

In the middle of the 19th Century S. Lie made a far reaching discovery that techniques designed to solve particular unrelated types of ODEs, such as separable, homogeneous and exact equations, were in fact all special cases of a general form of integration procedure based on the invariance of the differential equation under a continuous group of symmetries. Roughly speaking a symmetry group of a system of differential equations is a group that transforms solutions of the system to other solutions. Once the symmetry group has been identified a number of techniques to solve and classify these differential equations becomes possible. In the classical framework of Lie, these groups were local groups and arose locally as groups of transformations on some Euclidean space. The passage from the local Lie group to the present day definition using manifolds was accomplished by E. Cartan at the end of the 19th Century, whose work is a striking synthesis of Lie theory, classical geometry, differential geometry and topology.

These continuous groups, which originally appeared as symmetry groups of differential equations, have over the years had a profound impact on diverse areas such as algebraic topology, differential geometry, numerical analysis, control theory, classical mechanics, quantum mechanics etc. They are now universally known as Lie groups.

Definition of a Lie Group

A *Lie group* is a smooth (Banach) manifold M that has at the same time a group $G-$structure consistent with its manifold $M-$structure in the sense that *group multiplication*

$$\mu : G \times G \to G, \qquad (g, h) \mapsto gh \tag{1.33}$$

and the *group inversion*

$$\nu : G \to G, \qquad g \mapsto g^{-1} \tag{1.34}$$

are $C^\infty-$maps [Che55, AMR88, MR99, Put93]. A point $e \in G$ is called the *group identity element*.

For example, any nD Banach vector space V is an Abelian Lie group with group operations $\mu : V \times V \to V$, $\mu(x, y) = x + y$, and $\nu : V \to V$, $\nu(x) = -x$. The identity is just the zero vector. We call such a Lie group a *vector group*.

Let G and H be two Lie groups. A map $G \to H$ is said to be a *morphism* of Lie groups (or their *smooth homomorphism*) if it is their homomorphism as abstract groups and their smooth map as manifolds [Pos86].

All Lie groups and all their morphisms form the category $\mathcal{LG}$ (more precisely, there is a countable family of categories $\mathcal{LG}$ depending on C^k-smoothness of the corresponding manifolds).

Similarly, a group G which is at the same time a topological space is said to be a *topological group* if maps (1.33–1.34) are continuous, i.e., C^0-maps for it. The homomorphism $G \to H$ of topological groups is said to be continuous if it is a continuous map. Topological groups and their continuous homomorphisms form the category $\mathcal{TG}$.

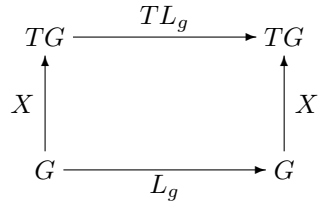
A topological group (as well as a smooth manifold) is not necessarily Hausdorff. A topological group G is Hausdorff iff its identity is closed. As a corollary we have that every Lie group is a Hausdorff topological group (see [Pos86]).

For every g in a Lie group G, the two maps,

$$\begin{aligned} L_g &: G \to G, \qquad h \mapsto gh, \qquad \text{and} \\ R_h &: G \to G, \qquad g \mapsto gh, \end{aligned}$$

are called *left* and *right translation* maps. Since $L_g \circ L_h = L_{gh}$, and $R_g \circ R_h = R_{gh}$, it follows that $(L_g)^{-1} = L_{g^{-1}}$ and $(R_g)^{-1} = R_{g^{-1}}$, so both L_g and R_g are diffeomorphisms. Moreover $L_g \circ R_h = R_h \circ L_g$, i.e., left and right translation commute.

A vector–field X on G is called *left invariant vector–field* if for every $g \in G$, $L_g^* X = X$, that is, if $(T_h L_g)X(h) = X(gh)$ for all $h \in G$, i.e., the following diagram commutes:



The correspondences $G \to TG$ and $L_g \to TL_g$ obviously define a functor $\mathcal{F} : \mathcal{LG} \Rightarrow \mathcal{LG}$ from the category G of Lie groups to itself. $\mathcal{F}$ is a special case of the vector bundle functor (see (1.3.3) below).

Let $\mathcal{X}_L(G)$ denote the set of left invariant vector–fields on G; it is a Lie subalgebra of $\mathcal{X}(G)$, the set of all vector–fields on G, since $L_g^*[X,Y] = [L_g^*X, L_g^*Y] = [X,Y]$, so the Lie bracket $[X,Y] \in \mathcal{X}_L(G)$.

Let e be the identity element of G. Then for each ξ on the tangent space T_eG we define a vector–field X_ξ on G by

$$X_\xi(g) = T_eL_g(\xi).$$

$\mathcal{X}_L(G)$ and T_eG are isomorphic as vector spaces. Define the Lie bracket on T_eG by

$$[\xi,\eta] = [X_\xi, X_\eta]\,(e),$$

for all $\xi, \eta \in T_eG$. This makes T_eG into a Lie algebra. Also, by construction, we have

$$[X_\xi, X_\eta] = X_{[\xi,\eta]},$$

this defines a bracket in T_eG via *left extension*. The vector space T_eG with the above algebra structure is called the Lie algebra of the Lie group G and is denoted $\mathfrak{g}$.

For example, let V be a nD vector space. Then $T_eV \simeq V$ and the left invariant vector–field defined by $\xi \in T_eV$ is the constant vector–field $X_\xi(\eta) = \xi$, for all $\eta \in V$. The Lie algebra of V is V itself.

Since any two elements of an Abelian Lie group G commute, it follows that all adjoint operators Ad_g, $g \in G$, equal the identity. Therefore, the Lie algebra g is Abelian; that is, $[\xi,\eta] = 0$ for all $\xi, \eta \in \mathfrak{g}$ [MR99].

Recall (1.2.6) that Lie algebras and their smooth homomorphisms form the category $\mathcal{LAL}$. We can now introduce the fundamental *Lie functor*, $\mathcal{F} : \mathcal{LG} \Rightarrow \mathcal{LAL}$, from the category of Lie groups to the category of Lie algebras [Pos86].

Let X_ξ be a left invariant vector–field on G corresponding to ξ in $\mathfrak{g}$. Then there is a unique integral curve $\gamma_\xi : \mathbb{R} \to G$ of X_ξ starting at e, i.e.,

$$\dot{\gamma}_\xi(t) = X_\xi\left(\gamma_\xi(t)\right), \qquad \gamma_\xi(0) = e.$$

$\gamma_\xi(t)$ is a smooth *one parameter subgroup* of G, i.e.,

$$\gamma_\xi(t+s) = \gamma_\xi(t) \cdot \gamma_\xi(s),$$

since, as functions of t both sides equal $\gamma_\xi(s)$ at $t = 0$ and both satisfy differential equation

$$\dot{\gamma}(t) = X_\xi\left(\gamma_\xi(t)\right)$$

by left invariance of X_ξ, so they are equal. Left invariance can be also used to show that $\gamma_\xi(t)$ is defined for all $t \in \mathbb{R}$. Moreover, if $\phi : \mathbb{R} \to G$ is a one parameter subgroup of G, i.e., a *smooth homomorphism* of the additive group $\mathbb{R}$ into G, then $\phi = \gamma_\xi$ with $\xi = \dot{\phi}(0)$, since taking derivative at $s = 0$ in the relation

$$\phi(t+s) = \phi(t)\cdot\phi(s) \qquad \text{gives} \qquad \dot{\phi}(t) = X_{\dot{\phi}(0)}\left(\phi(t)\right),$$

so $\phi = \gamma_\xi$ since both equal e at $t = 0$. Therefore, all one parameter subgroups of G are of the form $\gamma_\xi(t)$ for some $\xi \in \mathfrak{g}$.

The map $\exp : \mathfrak{g} \to G$, given by

$$\exp(\xi) = \gamma_\xi(1), \qquad \exp(0) = e, \tag{1.35}$$

is called the *exponential map* of the Lie algebra $\mathfrak{g}$ of G into G. exp is a C^∞–map, similar to the projection π of tangent and cotangent bundles; exp is locally a diffeomorphism from a neighborhood of zero in $\mathfrak{g}$ onto a neighborhood of e in G; if $f : G \to H$ is a smooth homomorphism of Lie groups, then

$$f \circ \exp_G = \exp_H \circ T_e f\,.$$

Also, in this case (see [Che55, MR99, Pos86])

$$\exp(s\xi) = \gamma_\xi(s).$$

Indeed, for fixed $s \in \mathbb{R}$, the curve $t \mapsto \gamma_\xi(ts)$, which at $t = 0$ passes through e, satisfies the differential equation

$$\frac{d}{dt}\gamma_\xi(ts) = sX_\xi\left(\gamma_\xi(ts)\right) = X_{s\xi}\left(\gamma_\xi(ts)\right).$$

Since $\gamma_{s\xi}(t)$ satisfies the same differential equation and passes through e at $t = 0$, it follows that $\gamma_{s\xi}(t) = \gamma_\xi(st)$. Putting $t = 1$ induces $\exp(s\xi) = \gamma_\xi(s)$ [MR99].

Hence exp maps the line $s\xi$ in $\mathfrak{g}$ onto the one–parameter subgroup $\gamma_\xi(s)$ of G, which is tangent to ξ at e. It follows from left invariance that the flow F_t^ξ of X satisfies $F_t^\xi(g) = g\exp(s\xi)$.

Globally, the exponential map exp, as given by (1.35), is a natural operation, i.e., for any morphism $\varphi : G \to H$ of Lie groups G and H and a Lie functor $\mathcal{F}$, the following diagram commutes [Pos86]:

$$\begin{array}{ccc} \mathcal{F}(G) & \xrightarrow{\mathcal{F}(\varphi)} & \mathcal{F}(H) \\ \exp\downarrow & & \downarrow\exp \\ G & \xrightarrow[\varphi]{} & H \end{array}$$

Let G_1 and G_2 be Lie groups with Lie algebras $\mathfrak{g}_1$ and $\mathfrak{g}_2$. Then $G_1 \times G_2$ is a Lie group with Lie algebra $\mathfrak{g}_1 \times \mathfrak{g}_2$, and the exponential map is given by [MR99].

$$\exp : \mathfrak{g}_1 \times \mathfrak{g}_2 \to G_1 \times G_2, \qquad (\xi_1, \xi_2) \mapsto (\exp_1(\xi_1), \exp_2(\xi_2)) .$$

For example, in case of a nD vector space, or infinite–dimensional Banach space, the exponential map is the identity.

The unit circle in the complex plane $S^1 = \{z \in \mathbb{C} : |z| = 1\}$ is an Abelian Lie group under multiplication. The tangent space $T_e S^1$ is the imaginary axis, and we identify $\mathbb{R}$ with $T_e S^1$ by $t \mapsto 2\pi i t$. With this identification, the exponential map $\exp : \mathbb{R} \to S^1$ is given by $\exp(t) = e^{2\pi i t}$.

The nD torus $T^n = S^1 \times \cdots \times S^1$ (n times) is an Abelian Lie group. The exponential map $\exp : \mathbb{R}^n \to T^n$ is given by

$$\exp(t_1, ..., t_n) = (e^{2\pi i t_1}, ..., e^{2\pi i t_n}).$$

Since $S^1 = \mathbb{R}/\mathbb{Z}$, it follows that

$$T^n = \mathbb{R}^n / \mathbb{Z}^n,$$

the projection $\mathbb{R}^n \to T^n$ being given by the exp map (see [MR99, Pos86]).

For every $g \in G$, the map

$$Ad_g = T_e \left(R_{g^{-1}} \circ L_g \right) : \mathfrak{g} \to \mathfrak{g}$$

is called the *adjoint map* (or *operator*) associated with g.

For each $\xi \in \mathfrak{g}$ and $g \in G$ we have

$$\exp \left(Ad_g \xi \right) = g \left(\exp \xi \right) g^{-1}.$$

The relation between the adjoint map and the Lie bracket is the following: For all $\xi, \eta \in \mathfrak{g}$ we have

$$\left. \frac{d}{dt} \right|_{t=0} Ad_{\exp(t\xi)} \eta = [\xi, \eta].$$

A Lie subgroup H of G is a subgroup H of G which is also a submanifold of G. Then $\mathfrak{h}$ is a Lie subalgebra of $\mathfrak{g}$ and moreover $\mathfrak{h} = \{\xi \in \mathfrak{g} | \exp(t\xi) \in H,$ for all $t \in \mathbb{R}\}$.

Recall that one can characterize *Lebesgue measure* up to a multiplicative constant on $\mathbb{R}^n$ by its invariance under translations. Similarly, on a locally compact group there is a unique (up to a nonzero multiplicative constant) left–invariant measure, called *Haar measure*. For Lie groups the existence of such measures is especially simple [MR99]: Let G be a Lie group. Then there is a volume form $Ub5$, unique up to nonzero multiplicative constants, that is left invariant. If G is compact, $Ub5$ is right invariant as well.

Actions of Lie Groups on Smooth Manifolds

Let M be a smooth manifold. An *action of a Lie group* G (with the unit element e) on M is a smooth map $\phi : G \times M \to M$, such that for all $x \in M$ and $g, h \in G$, (i) $\phi(e, x) = x$ and (ii) $\phi(g, \phi(h, x)) = \phi(gh, x)$. In other words, letting $\phi_g : x \in M \mapsto \phi_g(x) = \phi(g, x) \in M$, we have (i') $\phi_e = id_M$ and (ii') $\phi_g \circ \phi_h = \phi_{gh}$. ϕ_g is a diffeomorphism, since $(\phi_g)^{-1} = \phi_{g^{-1}}$. We say that the map $g \in G \mapsto \phi_g \in Diff(M)$ is a homomorphism of G into the group of diffeomorphisms of M. In case that M is a vector space and each ϕ_g is a linear operator, the function of G on M is called a representation of G on M [Put93]

An action ϕ of G on M is said to be *transitive group action*, if for every $x, y \in M$, there is $g \in G$ such that $\phi(g, x) = y$; *effective group action*, if $\phi_g = id_M$ implies $g = e$, that is $g \mapsto \phi_g$ is 1–1; and *free group action*, if for each $x \in M$, $g \mapsto \phi_g(x)$ is 1–1.

For example,

1. $G = \mathbb{R}$ acts on $M = \mathbb{R}$ by translations; explicitly,

$$\phi : G \times M \to M, \qquad \phi(s, x) = x + s.$$

   Then for $x \in \mathbb{R}$, $O_x = \mathbb{R}$. Hence M/G is a single point, and the action is transitive and free.
2. A complete flow ϕ_t of a vector–field X on M gives an action of $\mathbb{R}$ on M, namely

$$(t, x) \in \mathbb{R} \times M \mapsto \phi_t(x) \in M.$$

3. Left translation $L_g : G \to G$ defines an effective action of G on itself. It is also transitive.
4. The coadjoint action of G on $\mathfrak{g}^*$ is given by

$$Ad^* : (g, \alpha) \in G \times \mathfrak{g}^* \mapsto Ad^*_{g^{-1}}(\alpha) = \left(T_e(R_{g^{-1}} \circ L_g)\right)^* \alpha \in \mathfrak{g}^*.$$

Let ϕ be an action of G on M. For $x \in M$ the *orbit* of x is defined by

$$O_x = \{\phi_g(x) | g \in G\} \subset M$$

and the *isotropy group* of ϕ at x is given by

$$G_x = \{g \in G | \phi(g, x) = x\} \subset G.$$

An action ϕ of G on a manifold M defines an equivalence relation on M by the relation belonging to the same orbit; explicitly, for $x, y \in M$, we write $x \sim y$ if there exists a $g \in G$ such that $\phi(g, x) = y$, that is, if $y \in O_x$. The set of all orbits M/G is called the *group orbit space*.

For example, let $M = \mathbb{R}^2 \backslash \{0\}$, $G = SO(2)$, the group of rotations in plane, and the action of G on M given by

$$\left(\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, (x,y) \right) \longmapsto (x\cos\theta - y\sin\theta,\ x\sin\theta + y\cos\theta).$$

The action is always free and effective, and the orbits are concentric circles, thus the orbit space is $M/G \simeq \mathbb{R}^*_+$.

A crucial concept in mechanics is the infinitesimal description of an action. Let $\phi : G \times M \to M$ be an action of a Lie group G on a smooth manifold M. For each $\xi \in \mathfrak{g}$,

$$\phi_\xi : \mathbb{R} \times M \to M, \qquad \phi_\xi(t,x) = \phi\left(\exp(t\xi), x\right)$$

is an $\mathbb{R}$–action on M. Therefore, $\phi_{\exp(t\xi)} : M \to M$ is a flow on M; the corresponding vector–field on M, given by

$$\xi_M(x) = \left.\frac{d}{dt}\right|_{t=0} \phi_{\exp(t\xi)}(x)$$

is called the infinitesimal generator of the action, corresponding to ξ in $\mathfrak{g}$.

The tangent space at x to an orbit O_x is given by

$$T_x O_x = \{\xi_M(x) | \xi \in \mathfrak{g}\}.$$

Let $\phi : G \times M \to M$ be a smooth $G-$-action. For all $g \in G$, all $\xi, \eta \in \mathfrak{g}$ and all $\alpha, \beta \in \mathbb{R}$, we have:

$(Ad_g\xi)_M = \phi^*_{g^{-1}}\xi_M$, $[\xi_M, \eta_M] = -\left[\xi,\eta\right]_M$, and $(\alpha\xi + \beta\eta)_M = \alpha\xi_M + \beta\eta_M$.

Let M be a smooth manifold, G a Lie group and $\phi : G \times M \to M$ a $G-$action on M. We say that a smooth map $f : M \to M$ is with respect to this action if for all $g \in G$,

$$f \circ \phi_g = \phi_g \circ f.$$

Let $f : M \to M$ be an equivariant smooth map. Then for any $\xi \in \mathfrak{g}$ we have

$$Tf \circ \xi_M = \xi_M \circ f.$$

Basic Dynamical Groups

Here we give the first two examples of Lie groups, namely Galilei group and general linear group. Further examples will be given in association with particular dynamical systems.

Galilei Group

The *Galilei group* is the group of transformations in space and time that connect those Cartesian systems that are termed 'inertial frames' in Newtonian mechanics. The most general relationship between two such frames is the following. The origin of the time scale in the inertial frame S' may be shifted compared with that in S; the orientation of the Cartesian axes in S' may

be different from that in S; the origin O of the Cartesian frame in S' may be moving relative to the origin O in S at a uniform velocity. The transition from S to S' involves ten parameters; thus the Galilei group is a ten parameter group. The basic assumption inherent in Galilei–Newtonian relativity is that there is an absolute time scale, so that the only way in which the time variables used by two different 'inertial observers' could possibly differ is that the zero of time for one of them may be shifted relative to the zero of time for the other.

Galilei space–time structure involves the following three elements:

1. *World*, as a 4D affine space A^4. The points of A^4 are called *world points* or *events*. The parallel transitions of the world A^4 form a linear (i.e., Euclidean) space $\mathbb{R}^4$.
2. *Time*, as a linear map $t : \mathbb{R}^4 \to \mathbb{R}$ of the linear space of the world parallel transitions onto the real 'time axes'. Time interval from the event $a \in A^4$ to $b \in A^4$ is called the number $t(b-a)$; if $t(b-a) = 0$ then the events a and b are called synchronous. The set of all mutually synchronous events consists a 3D affine space A^3, being a subspace of the world A^4. The kernel of the mapping t consists of the parallel transitions of A^4 translating arbitrary (and every) event to the synchronous one; it is a linear 3D subspace $\mathbb{R}^3$ of the space $\mathbb{R}^4$.
3. *Distance* (*metric*) between the synchronous events,

$$\rho(a, b) = \| a - b \|, \qquad \text{for all} \quad a, b \in A^3,$$

given by the scalar product in $\mathbb{R}^3$. The distance transforms arbitrary space of synchronous events into the well known 3D Euclidean space E^3.

The space A^4, with the Galilei space–time structure on it, is called Galilei space. Galilei group is the group of all possible transformations of the Galilei space, preserving its structure. The elements of the Galilei group are called Galilei transformations. Therefore, Galilei transformations are affine transformations of the world A^4 preserving the time intervals and distances between the synchronous events.

The direct product $\mathbb{R} \times \mathbb{R}^3$, of the time axes with the 3D linear space R3 with a fixed Euclidean structure, has a natural Galilei structure. It is called Galilei coordinate system.

General Linear Group

The group of linear isomorphisms of $\mathbb{R}^n$ to $\mathbb{R}^n$ is a Lie group of dimension n^2, called the *general linear group* and denoted $Gl(n, \mathbb{R})$. It is a smooth manifold, since it is a subset of the vector space $L(\mathbb{R}^n, \mathbb{R}^n)$ of all linear maps of $\mathbb{R}^n$ to $\mathbb{R}^n$, as $Gl(n, \mathbb{R})$ is the inverse image of $\mathbb{R}\backslash\{0\}$ under the continuous map $A \mapsto \det A$ of $L(\mathbb{R}^n, \mathbb{R}^n)$ to $\mathbb{R}$. The group operation is composition

$$(A, B) \in Gl(n, \mathbb{R}) \times Gl(n, \mathbb{R}) \mapsto A \circ B \in Gl(n, \mathbb{R})$$

and the inverse map is

$$A \in Gl(n,\mathbb{R}) \mapsto A^{-1} \in Gl(n,\mathbb{R}).$$

If we choose a basis in $\mathbb{R}^n$, we can represent each element $A \in Gl(n,\mathbb{R})$ by an invertible $(n \times n)$–matrix. The group operation is then matrix multiplication and the inversion is matrix inversion. The identity is the identity matrix I_n. The group operations are smooth since the formulas for the product and inverse of matrices are smooth in the matrix components.

The Lie algebra of $Gl(n,\mathbb{R})$ is $\mathfrak{gl}(n)$, the vector space $L(\mathbb{R}^n,\mathbb{R}^n)$ of all linear transformations of $\mathbb{R}^n$, with the commutator bracket

$$[A,B] = AB - BA.$$

For every $A \in L(\mathbb{R}^n,\mathbb{R}^n)$,

$$\gamma_A : t \in \mathbb{R} \mapsto \gamma_A(t) = \sum_{i=0}^{\infty} \frac{t^i}{i!} A^i \in Gl(n,\mathbb{R})$$

is a one parameter subgroup of $Gl(n,\mathbb{R})$, because

$$\gamma_A(0) = I, \qquad \text{and} \qquad \dot{\gamma}_A(t) = \sum_{i=0}^{\infty} \frac{t^{i-1}}{(i-1)!} A^i = \gamma_A(t)\, A.$$

Hence γ_A is an integral curve of the left invariant vector–field X_A. Therefore, the exponential map is given by

$$\exp : A \in L(\mathbb{R}^n,\mathbb{R}^n) \mapsto \exp(A) \equiv e^A = \gamma_A(1) = \sum_{i=0}^{\infty} \frac{A^i}{i!} \in Gl(n,\mathbb{R}).$$

For each $A \in Gl(n,\mathbb{R})$ the corresponding adjoint map

$$Ad_A : L(\mathbb{R}^n,\mathbb{R}^n) \to L(\mathbb{R}^n,\mathbb{R}^n)$$

is given by

$$Ad_A B = A \cdot B \cdot A^{-1}.$$

1.2.8 Lie Symmetries and Prolongations on Manifolds

In this subsection we continue our expose on Lie groups of symmetry, as a link to modern jet machinery, developed below.

Lie Symmetry Groups

Exponentiation of Vector Fields on M

Let $x = (x^1, ..., x^r)$ be local coordinates at a point m on a smooth $n-$manifold M. Recall that the flow generated by the vector–field

$$v = \xi^i(x)\, \partial_{x^i} \in M,$$

is a solution of the system of ODEs

$$\frac{dx^i}{d\varepsilon} = \xi^i(x^1, ..., x^m), \qquad (i = 1, ..., r).$$

The computation of the flow, or one–parameter group of diffeomorphisms, generated by a given vector–field v (i.e., solving the system of ODEs) is often referred to as *exponentiation of a vector–field*, denoted by $\exp(\varepsilon v)\, x$ (see [Olv86]).

If $v, w \in M$ are two vectors defined by

$$v = \xi^i(x)\, \partial_{x^i} \qquad \text{and} \qquad w = \eta^i(x)\, \partial_{x^i},$$

then

$$\exp(\varepsilon v)\ \exp(\theta w)\, x = \exp(\theta w) \exp(\varepsilon v)\, x,$$

for all $\epsilon, \theta \in \mathbb{R}$,$x \in M$, such that both sides are defined, iff they commute, i.e., $[v, w] = 0$ everywhere [Olv86].

A system of vector–fields $\{v_1, ..., v_r\}$ on a smooth manifold M is in *involution* if there exist smooth real–valued functions $h^k_{ij}(x)$, $x \in M$, $i, j, k = 1, ..., r$, such that for each i, j,

$$[v_i, v_j] = h^k_{ij} \cdot v_k.$$

Let $v \neq 0$ be a right–invariant vector–field on a Lie group G. Then the flow generated by v through the identity e, namely

$$g_\varepsilon = \exp(\varepsilon v)\, e \equiv \exp(\varepsilon v),$$

is defined for all $\varepsilon \in \mathbb{R}$ and forms a one–parameter subgroup of G, with

$$g_{\varepsilon+\delta} = g_\varepsilon \cdot g_\delta, \qquad g_0 = e, \qquad g_\varepsilon^{-1} = g_{-\varepsilon},$$

isomorphic to either $\mathbb{R}$ itself or the circle group $SO(2)$. Conversely, any connected 1D subgroup of G is generated by such a right–invariant vector–field in the above manner [Olv86].

For example, let $G = GL(n)$ with Lie algebra $\mathfrak{gl}(n)$, the space of all $n \times n$ matrices with commutator as the Lie bracket. If $A \in \mathfrak{gl}(n)$, then the corresponding right–invariant vector–field v_A on $GL(n)$ has the expression [Olv86]

$$v_A = a^i_k x^k_j\, \partial_{x^i_j}.$$

The one–parameter subgroup $\exp(\varepsilon v_A)\, e$ is found by integrating the system of n^2 ordinary differential equations

$$\frac{dx^i_j}{d\varepsilon} = a^i_k x^k_j, \qquad x^i_j(0) = \delta^i_j, \qquad (i,j = 1,...,n),$$

involving matrix entries of A. The solution is just the matrix exponential $X(\varepsilon) = e^{\varepsilon A}$, which is the one–parameter subgroup of $GL(n)$ generated by a matrix A in $\mathfrak{gl}(n)$.

Recall that the *exponential map* $\exp : \mathfrak{g} \to G$ is get by setting $\varepsilon = 1$ in the one–parameter subgroup generated by vector–field v :

$$\exp(v) \equiv \exp(v)\, e.$$

Its differential at 0,

$$d\exp : T\mathfrak{g}|_0 \simeq \mathfrak{g} \to TG|_e \simeq \mathfrak{g}$$

is the identity map.

Lie Symmetry Groups and General Differential Equations

Consider a system $\mathcal{S}$ of general differential equations (DEs, to be distinguished from ODEs) involving p independent variables $x = (x^1, ..., x^p)$, and q dependent variables $u = (u^1, ..., u^q)$. The solution of the system will be of the form $u = f(x)$, or, in components, $u^\alpha = f^\alpha(x^1, ..., x^p)$, $\alpha = 1, ..., q$ (so that Latin indices refer to independent variables while Greek indices refer to dependent variables). Let $X = \mathbb{R}^p$, with coordinates $x = (x^1, ..., x^p)$, be the space representing the independent variables, and let $U = \mathbb{R}^q$, with coordinates $u = (u^1, ..., u^q)$, represent dependent variables. A Lie symmetry group G of the system $\mathcal{S}$ will be a local group of transformations acting on some open subset $M \subset X \times U$ in such way that G transforms solutions of $\mathcal{S}$ to other solutions of $\mathcal{S}$ [Olv86].

More precisely, we need to explain exactly how a given transformation $g \in G$, where G is a Lie group, transforms a function $u = f(x)$. We firstly identify the function $u = f(x)$ with its graph

$$\Gamma_f \equiv \{(x, f(x)) : x \in \operatorname{dom} f \equiv \Omega\} \subset X \times U,$$

where Γ_f is a submanifold of $X \times U$. If $\Gamma_f \subset M_g \equiv \operatorname{dom} g$, then the transform of Γ_f by g is defined as

$$g \cdot \Gamma_f = \{(\tilde{x}, \tilde{u}) = g \cdot (x, u) : (x, u) \in \Gamma_f\}.$$

We write $\tilde{f} = g \cdot f$ and call the function $\tilde{f}$ the *transform* of f by g.

For example, let $p = 1$ and $q = 1$, so $X = \mathbb{R}$ with a single independent variable x, and $U = \mathbb{R}$ with a single dependent variable u, so we have a single ODE involving a single function $u = f(x)$. Let $G = SO(2)$ be the rotation group acting on $X \times U \simeq \mathbb{R}^2$. The transformations in G are given by

$$(\tilde{x}, \tilde{u}) = \theta \cdot (x, u) = (x \cos\theta - u \sin\theta,\ x \sin\theta + u \cos\theta).$$

Let $u = f(x)$ be a function whose graph is a subset $\Gamma_f \subset X \times U$. The group $SO(2)$ acts on f by rotating its graph.

In general, the procedure for finding the transformed function $\tilde{f} = g \cdot f$ is given by [Olv86]:

$$g \cdot f = [\Phi_g \circ (1 \times f)] \circ [\Xi_g \circ (1 \times f)]^{-1}, \tag{1.36}$$

where $\Xi_g = \Xi_g(x, u)$, $\Phi_g = \Phi_g(x, u)$ are smooth functions such that

$$(\tilde{x}, \tilde{u}) = g \cdot (x, u) = (\Xi_g(x, u),\ \Phi_g(x, u)),$$

while 1 denotes the identity function of X, so $1(x) = x$. Formula (1.36) holds whenever the second factor is invertible.

Let $\mathcal{S}$ be a system of DEs. A *symmetry group* of the system $\mathcal{S}$ is a local Lie group of transformations G acting on an open subset $M \subset X \times U$ of the space $X \times U$ of independent and dependent variables of the system with the property that whenever $u = f(x)$ is a solution of $\mathcal{S}$, and whenever $g \cdot f$ is defined for $g \in G$, then $u = g \cdot f(x)$ is also a solution of the system.

For example, in the case of the ODE $u_{xx} = 0$, the rotation group $SO(2)$ is obviously a symmetry group, since the solutions are all linear functions and $SO(2)$ takes any linear function to another linear function. Another easy example is given by the classical *heat equation* $u_t = u_{xx}$. Here the group of translations

$$(x, t, u) \mapsto (x + \varepsilon a,\ t + \varepsilon b, u), \qquad \varepsilon \in \mathbb{R},$$

is a symmetry group since $u = f(x - \varepsilon a,\ t - \varepsilon b)$ is a solution to the heat equation whenever $u = f(x, t)$ is.

Prolongations

Prolongations of Functions

Given a smooth real–valued function $u = f(x) = f(x^1, ..., x^p)$ of p independent variables, there is an induced function $u^{(n)} = \mathbf{pr}^{(n)} f(x)$, called the nth *prolongation* of f [Olv86], which is defined by the equations

$$u_J = \partial_J f(x) = \frac{\partial^k f(x)}{\partial x^{j_1} \partial x^{j_2 \ldots} \partial x^{j_k}},$$

where the *multi–index* $J = (j_1, ..., j_k)$ is an unordered $k-$tuple of integers, with entries $1 \le j_k \le p$ indicating which derivatives are being taken. More generally, if $f : X \to U$ is a smooth function from $X \simeq \mathbb{R}^p$ to $U \simeq \mathbb{R}^q$, so $u = f(x) = f(f^1(x), ..., f^q(x))$, there are $q \cdot p_k$ numbers

$$u^\alpha_J = \partial_J f^\alpha(x) = \frac{\partial^k f^\alpha(x)}{\partial x^{j_1} \partial x^{j_2 \ldots} \partial x^{j_k}},$$

needed to represent all the different kth order derivatives of the components of f at a point x. Thus $\mathbf{pr}^{(n)} f : X \to U^{(n)}$ is a function from X to the space $U^{(n)}$, and for each $x \in X$, $\mathbf{pr}^{(n)} f(x)$ is a vector whose $q \cdot p^{(n)}$ entries represent the values of f and al its derivatives up to order n at the point x.

For example, in the case $p = 2$, $q = 1$ we have $X \simeq \mathbb{R}^2$ with coordinates $(x^1, x^2) = (x, y)$, and $U \simeq \mathbb{R}$ with the single coordinate $u = f(x, y)$. The second prolongation $u^{(2)} = \mathbf{pr}^{(2)} f(x, y)$ is given by [Olv86]

$$(u; u_x, u_y; u_{xx}, u_{xy}, u_{yy}) = \left(f; \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}; \frac{\partial^2 f}{\partial x^2}, \frac{\partial^2 f}{\partial x \partial y}, \frac{\partial^2 f}{\partial y^2} \right), \tag{1.37}$$

all evaluated at (x, y).

The nth prolongation $\mathbf{pr}^{(n)} f(x)$ is also known as the $n-$jet of f. In other words, the nth prolongation $\mathbf{pr}^{(n)} f(x)$ represents the Taylor polynomial of degree n for f at the point x, since the derivatives of order $\leq n$ determine the Taylor polynomial and vice versa.

Prolongations of Differential Equations

A system $\mathcal{S}$ of nth order DEs in p independent and q dependent variables is given as a system of equations [Olv86]

$$\Delta_r(x, u^{(n)}) = 0, \qquad (r = 1, ..., l), \tag{1.38}$$

involving $x = (x^1, ..., x^p)$, $u = (u^1, ..., u^q)$ and the derivatives of u with respect to x up to order n. The functions $\Delta(x, u^{(n)}) = (\Delta_1(x, u^{(n)}), ..., \Delta_l(x, u^{(n)}))$ are assumed to be smooth in their arguments, so $\Delta : X \times U^{(n)} \to \mathbb{R}^l$ represents a smooth map from the *jet space* $X \times U^{(n)}$ to some lD Euclidean space (see section 1.4 below). The DEs themselves tell where the given map Δ vanishes on the jet space $X \times U^{(n)}$, and thus determine a submanifold

$$\mathcal{S}_\Delta = \left\{ (x, u^{(n)}) : \Delta(x, u^{(n)}) = 0 \right\} \subset X \times U^{(n)} \tag{1.39}$$

of the total the jet space $X \times U^{(n)}$.

We can identify the system of DEs (1.38) with its corresponding submanifold $\mathcal{S}_\Delta$ (1.39). From this point of view, a smooth *solution* of the given system of DEs is a smooth function $u = f(x)$ such that [Olv86]

$$\Delta_r(x, \mathbf{pr}^{(n)} f(x)) = 0, \qquad (r = 1, ..., l),$$

whenever x lies in the domain of f. This is just a restatement of the fact that the derivatives $\partial_J f^\alpha(x)$ of f must satisfy the algebraic constraints imposed by the system of DEs. This condition is equivalent to the statement that the graph of the prolongation $\mathbf{pr}^{(n)} f(x)$ must lie entirely within the submanifold $\mathcal{S}_\Delta$ determined by the system:

$$\Gamma_f^{(n)} \equiv \left\{ (x, \mathbf{pr}^{(n)} f(x)) \right\} \subset \mathcal{S}_\Delta = \left\{ \Delta(x, u^{(n)}) = 0 \right\}.$$

We can thus take an nth order system of DEs to be a submanifold $\mathcal{S}_\Delta$ in the $n-$jet space $X \times U^{(n)}$ and a solution to be a function $u = f(x)$ such that the graph of the nth prolongation $\mathbf{pr}^{(n)} f(x)$ is contained in the submanifold $\mathcal{S}_\Delta$.

For example, consider the case of *Laplace equation* in the plane

$$u_{xx} + u_{yy} = 0 \qquad (\text{remember}, \quad u_x \equiv \partial_x u).$$

Here $p = 2$ since there are two independent variables x and y, and $q = 1$ since there is one dependent variable u. Also $n = 2$ since the equation is second–order, so $\mathcal{S}_\Delta \subset X \times U^{(2)}$ is given by (1.37). A solution $u = f(x, y)$ must satisfy

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0$$

for all (x, y). This is the same as requiring that the graph of the second prolongation $\mathbf{pr}^{(2)} f$ lie in $\mathcal{S}_\Delta$.

Prolongations of Group Actions

Let G be a local group of transformations acting on an open subset $M \subset X \times U$ of the space of independent and dependent variables. There is an induced local action of G on the $n-$jet space $M^{(n)}$, called the nth prolongation $\mathbf{pr}^{(n)} G$ of the action of G on M. This prolongation is defined so that it transforms the derivatives of functions $u = f(x)$ into the corresponding derivatives of the transformed function $\tilde{u} = \tilde{f}(\tilde{x})$ [Olv86].

More precisely, suppose $(x_0, u_0^{(n)})$ is a given point in $M^{(n)}$. Choose any smooth function $u = f(x)$ defined in a neighborhood of x_0, whose graph Γ_f lies in M, and has the given derivatives at x_0 :

$$u_0^{(n)} = \mathbf{pr}^{(n)} f(x_0), \qquad \text{i.e.,} \qquad u^\alpha_{J0} = \partial_J f^\alpha(x_0).$$

If g is an element of G sufficiently near the identity, the transformed function $g \cdot f$ as given by (1.36) is defined in a neighborhood of the corresponding point $(\tilde{x}_0, \tilde{u}_0) = g \cdot (x_0, u_0)$, with $u_0 = f(x_0)$ being the zeroth order components of $u_0^{(n)}$. We then determine the action of the prolonged group of transformations $\mathbf{pr}^{(n)} g$ on the point $(x_0, u_0^{(n)})$ by evaluating the derivatives of the transformed function $g \cdot f$ at $\tilde{x}_0$; explicitly [Olv86]

$$\mathbf{pr}^{(n)} g \cdot (x_0, u_0^{(n)}) = (\tilde{x}_0, \tilde{u}_0^{(n)}),$$

where

$$\tilde{u}_0^{(n)} \equiv \mathbf{pr}^{(n)} (g \cdot f)(\tilde{x}_0).$$

For example, let $p = q = 1$, so $X \times U \simeq \mathbb{R}^2$, and consider the action of the rotation group $SO(2)$. To calculate its first prolongation $\mathbf{pr}^{(1)} SO(2)$, first note that $X \times U^{(1)} \simeq \mathbb{R}^3$, with coordinates (x, u, u_x). given a function $u = f(x)$, the first prolongation is [Olv86]

$$\mathbf{pr}^{(1)} f(x) = (f(x), f'(x)).$$

Now, given a point $(x^0, u^0, u^0_x) \in X \times U^{(1)}$, and a rotation in $SO(2)$ characterized by the angle θ as given above, the corresponding transformed point

$$\mathbf{pr}^{(1)}\theta \cdot (x^0, u^0, u^0_x) = (\tilde{x}^0, \tilde{u}^0, \tilde{u}^0_x)$$

(provided it exists). As for the first–order derivative, we find

$$\tilde{u}^0_x = \frac{\sin\theta + u_x \cos\theta}{\cos\theta - u_x \sin\theta}.$$

Now, applying the group transformations given above, and dropping the $0-$indices, we find that the prolonged action $\mathbf{pr}^{(1)} SO(2)$ on $X \times U^{(1)}$ is given by

$$\mathbf{pr}^{(1)}\theta \cdot (x, u, u_x) = \left(x\cos\theta - u\sin\theta,\ x\sin\theta + u\cos\theta,\ \frac{\sin\theta + u_x\cos\theta}{\cos\theta - u_x\sin\theta} \right),$$

which is defined for $|\theta| < |\operatorname{arccot} u_x|$. Note that even though $SO(2)$ is a linear, globally defined group of transformations, its first prolongation $\mathbf{pr}^{(1)} SO(2)$ is both nonlinear and only locally defined. This fact demonstrates the complexity of the operation of prolonging a group of transformations.

In general, for any Lie group G, the first prolongation $\mathbf{pr}^{(1)} G$ acts on the original variables (x, u) exactly the same way that G itself does; only the action on the derivative u_x gives an new information. Therefore, $\mathbf{pr}^{(0)} G$ agrees with G itself, acting on $M^{(0)} = M$.

Prolongations of Vector Fields

Prolongation of the infinitesimal generators of the group action turn out to be the *infinitesimal generators* of the *prolonged group action* [Olv86]. Let $M \subset X \times U$ be open and suppose v is a vector–field on M, with corresponding local one–parameter group $\exp(\varepsilon v)$. The nth prolongation of v, denoted $\mathbf{pr}^{(n)} v$, will be a vector–field on the $n-$jet space $M^{(n)}$, and is defined to be the infinitesimal generator of the corresponding prolonged on–parameter group $\mathbf{pr}^{(n)}[\exp(\varepsilon v)]$. In other words,

$$\mathbf{pr}^{(n)} v|_{(x,u^{(n)})} = \left. \frac{d}{d\varepsilon} \right|_{\varepsilon=0} \mathbf{pr}^{(n)}[\exp(\varepsilon v)](x, u^{(n)}) \tag{1.40}$$

for any $(x, u^{(n)}) \in M^{(n)}$.

For a vector–field v on M, given by

$$v = \xi^i(x,u)\frac{\partial}{\partial x^i} + \phi^\alpha(x,u)\frac{\partial}{\partial u^\alpha}, \qquad (i = 1, ..., p,\ \alpha = 1, ..., q),$$

the nth prolongation $\mathbf{pr}^{(n)} v$ is given by [Olv86]

$$\mathbf{pr}^{(n)}v = \xi^i(x,u)\frac{\partial}{\partial x^i} + \phi^\alpha_J(x,u^{(n)})\frac{\partial}{\partial u^\alpha_J},$$

with $\phi^\alpha_0 = \phi^\alpha$, and J a multiindex defined above.

For example, in the case of $SO(2)$ group, the corresponding infinitesimal generator is

$$v = -u\frac{\partial}{\partial x} + x\frac{\partial}{\partial u},$$

with

$$\exp(\varepsilon v)(x,u) = \left(x\cos\varepsilon - u\sin\varepsilon,\ x\sin\varepsilon + u\cos\varepsilon\right),$$

being the rotation through angle ε. The first prolongation takes the form

$$\mathbf{pr}^{(1)}[\exp(\varepsilon v)](x,u,u_x) = \left(x\cos\varepsilon - u\sin\varepsilon,\ x\sin\varepsilon + u\cos\varepsilon,\ \frac{\sin\varepsilon + u_x\cos\varepsilon}{\cos\varepsilon - u_x\sin\varepsilon}\right).$$

According to (1.40), the first prolongation of v is get by differentiating these expressions with respect to ε and setting $\varepsilon = 0$, which gives

$$\mathbf{pr}^{(1)}v = -u\frac{\partial}{\partial x} + x\frac{\partial}{\partial u} + (1+u_x^2)\frac{\partial}{\partial u_x}.$$

General Prolongation Formula

Let

$$v = \xi^i(x,u)\frac{\partial}{\partial x^i} + \phi^\alpha(x,u)\frac{\partial}{\partial u^\alpha}, \qquad (i = 1,...,p,\ \alpha = 1,...,q), \tag{1.41}$$

be a vector–field defined on an open subset $M \subset X \times U$. The nth prolongation of v is the vector–field [Olv86]

$$\mathbf{pr}^{(n)}v = v + \phi^\alpha_J(x,u^{(n)})\frac{\partial}{\partial u^\alpha_J}, \tag{1.42}$$

defined on the corresponding jet space $M^{(n)} \subset X \times U^{(n)}$. The coefficient functions ϕ^α_J are given by the following formula:

$$\phi^\alpha_J = D_J\left(\phi^\alpha - \xi^i u^\alpha_i\right) + \xi^i u^\alpha_{J,i}, \tag{1.43}$$

where $u^\alpha_i = \partial u^\alpha/\partial x^i$, and $u^\alpha_{J,i} = \partial u^\alpha_J/\partial x^i$. D_J is the *total derivative* with respect to the multiindex J, i.e.,

$$D_J = D_{j_1}D_{j_2}...D_{j_k},$$

while the total derivative with respect to the ordinary index, D_i, is defined as follows. Let $P(x,u^{(n)})$ be a smooth function of x,u and derivatives of u up to order n, defined on an open subset $M^{(n)} \subset X \times U^{(n)}$. the total derivative of P with respect to x^i is the unique smooth function $D_iP(x,u^{(n)})$ defined on

$M^{(n+1)}$ and depending on derivatives of u up to order $n+1$, with the *recursive property* that if $u = f(x)$ is any smooth function then

$$D_i P(x, \mathbf{pr}^{(n+1)} f(x)) = \partial_{x^i}\{P(x, \mathbf{pr}^{(n)} f(x))\}.$$

For example, in the case of $SO(2)$ group, with the infinitesimal generator

$$v = -u\frac{\partial}{\partial x} + x\frac{\partial}{\partial u},$$

the first prolongation is (as calculated above)

$$\mathbf{pr}^{(1)} v = -u\frac{\partial}{\partial x} + x\frac{\partial}{\partial u} + \phi^x \frac{\partial}{\partial u_x},$$

where

$$\phi^x = D_x(\phi - \xi u_x) + \xi u_{xx} = 1 + u_x^2.$$

Also,

$$\phi^{xx} = D_x \phi^x - u_{xx} D_x \xi = 3u_x u_{xx},$$

thus the infinitesimal generator of the second prolongation $\mathbf{pr}^{(2)} SO(2)$ acting on $X \times U^{(2)}$ is

$$\mathbf{pr}^{(2)} v = -u\frac{\partial}{\partial x} + x\frac{\partial}{\partial u} + (1 + u_x^2)\frac{\partial}{\partial u_x} + 3u_x u_{xx}\frac{\partial}{\partial u_{xx}}.$$

Let v and w be two smooth vector–fields on $M \subset X \times U$. Then their nth prolongations, $\mathbf{pr}^{(n)} v$ and $\mathbf{pr}^{(n)} w$ respectively, have the *linearity property*

$$\mathbf{pr}^{(n)}(c_1 v + c_2 w) = c_1 \mathbf{pr}^{(n)} v + c_2 \mathbf{pr}^{(n)} w, \qquad (c_1, c_2 - \text{constant}),$$

and the *Lie bracket property*

$$\mathbf{pr}^{(n)}[v, w] = [\mathbf{pr}^{(n)} v, \mathbf{pr}^{(n)} w].$$

Generalized Lie Symmetries

Consider a vector–field (1.41) defined on an open subset $M \subset X \times U$. Provided the coefficient functions ξ^i and ϕ^α depend only on x and u, v will generate a (local) one–parameter group of transformations $\exp(\varepsilon v)$ acting pointwise on the underlying space M. A significant generalization of the notion of symmetry group is get by relaxing this geometrical assumption, and allowing the coefficient functions ξ^i and ϕ^α to also depend on derivatives of u [Olv86].

A *generalized vector–field* is a (formal) expression

$$v = \xi^i[u]\frac{\partial}{\partial x^i} + \phi^\alpha[u]\frac{\partial}{\partial u^\alpha}, \qquad (i = 1, ..., p,\ \alpha = 1, ..., q), \tag{1.44}$$

in which ξ^i and ϕ^α are smooth functions. For example,

$$v = xu_x \frac{\partial}{\partial x} + u_{xx} \frac{\partial}{\partial u}$$

is a generalized vector in the case $p = q = 1$.

According to the general prolongation formula (1.42), we can define the *prolonged generalized vector–field*

$$\mathbf{pr}^{(n)} v = v + \phi_J^\alpha[u] \frac{\partial}{\partial u_J^\alpha},$$

whose coefficients are as before determined by the formula (1.43). Thus, in our previous example [Olv86],

$$\mathbf{pr}^{(n)} v = xu_x \frac{\partial}{\partial x} + u_{xx} \frac{\partial}{\partial u} + [u_{xxx} - (xu_{xx} + u_x)u_x] \frac{\partial}{\partial u_x}.$$

Given a generalized vector–field v, its *infinite prolongation* (including all the derivatives) is the formal expression

$$\mathbf{pr}\, v = \xi^i \frac{\partial}{\partial x^i} + \phi_J^\alpha \frac{\partial}{\partial u_J^\alpha}.$$

Now, a generalized vector–field v is a *generalized infinitesimal symmetry* of a system $\mathcal{S}$ of differential equations

$$\Delta_r[u] = \Delta_r(x, u^{(n)}) = 0, \qquad (r = 1, ..., l),$$

iff

$$\mathbf{pr}\, v[\Delta_r] = 0$$

for every smooth solution m $u = f(x)$ [Olv86].

For example, consider the heat equation

$$\Delta[u] = u_t - u_{xx} = 0.$$

The generalized vector–field $v = u_x \frac{\partial}{\partial u}$ has prolongation

$$\mathbf{pr}\, v = u_x \frac{\partial}{\partial u} + u_{xx} \frac{\partial}{\partial u_x} + u_{xt} \frac{\partial}{\partial u_t} + u_{xxx} \frac{\partial}{\partial u_{xx}} + ...$$

Thus

$$\mathbf{pr}\, v(\Delta) = u_{xt} - u_{xxx} = D_x(u_t - u_{xx}) = D_x \Delta,$$

and hence v is a generalized symmetry of the heat equation.

Noether Symmetries

Here we present some results about *Noether symmetries*, in particular for the first–order Lagrangians $L(q, \dot{q})$ (see [BGG89, PSS96]). We start with a *Noether–Lagrangian symmetry*,

$$\delta L = \dot{F},$$

and we will investigate the conversion of this symmetry to the Hamiltonian formalism. Defining

$$G = (\partial L/\partial \dot{q}^i)\,\delta q^i - F,$$

we can write

$$\delta_i L\,\delta q^i + \dot{G} = 0, \tag{1.45}$$

where $\delta_i L$ is the *Euler–Lagrangian functional derivative* of L,

$$\delta_i L = \alpha_i - W_{ik}\,\ddot{q}^k,$$

where

$$W_{ik} \equiv \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^k} \qquad \text{and} \qquad \alpha_i \equiv -\frac{\partial^2 L}{\partial \dot{q}^i \partial q^k}\,\dot{q}^k + \frac{\partial L}{\partial q^i}.$$

We consider the general case where the mass matrix, or *Hessian* (W_{ij}), may be a singular matrix. In this case there exists a kernel for the pull–back $\mathbb{F}L^*$ of the *Legendre map*, i.e., *fibre–derivative* $\mathbb{F}L$, from the *velocity phase–space manifold* TM (tangent bundle of the biomechanical manifold M) to the *momentum phase–space manifold* T^*M (cotangent bundle of M). This kernel is spanned by the vector–fields

$$\Gamma_\mu = \gamma^i_\mu \frac{\partial}{\partial \dot{q}^i},$$

where γ^i_μ are a basis for the null vectors of W_{ij}. The Lagrangian time–evolution differential operator can therefore be expressed as:

$$X = \partial_t + \dot{q}^k \frac{\partial}{\partial q^k} + a^k(q,\dot{q})\frac{\partial}{\partial \dot{q}^k} + \lambda^\mu \Gamma_\mu \equiv X_o + \lambda^\mu \Gamma_\mu,$$

where a^k are functions which are determined by the formalism, and λ^μ are arbitrary functions. It is not necessary to use the Hamiltonian technique to find the Γ_μ, but it does facilitate the calculation:

$$\gamma^i_\mu = \mathbb{F}L^* \left(\frac{\partial \phi_\mu}{\partial p_i} \right), \tag{1.46}$$

where the ϕ_μ are the Hamiltonian primary first class constraints.

Notice that the highest derivative in (1.45), $\ddot{q}^i$, appears linearly. Because δL is a symmetry, (1.45) is identically satisfied, and therefore the coefficient of $\ddot{q}^i$ vanishes:

$$W_{ik}\delta q^k - \frac{\partial G}{\partial \dot{q}^i} = 0. \tag{1.47}$$

We contract with a null vector γ^i_μ to find that

$$\Gamma_\mu G = 0.$$

It follows that G is projectable to a function $G_{\rm H}$ in T^*Q; that is, it is the pull–back of a function (not necessarily unique) in T^*Q:

$$G = \mathbb{F}L^*(G_{\rm H}).$$

This important property is valid for any conserved quantity associated with a Noether symmetry. Observe that $G_{\rm H}$ is determined up to the addition of linear combinations of the primary constraints. Substitution of this result in (1.47) gives

$$W_{ik}\left[\delta q^k - \mathbb{F}L^*\left(\frac{\partial G_{\rm H}}{\partial p_k}\right)\right] = 0,$$

and so the brackets enclose a null vector of W_{ik}:

$$\delta q^i - \mathbb{F}L^*\left(\frac{\partial G_{\rm H}}{\partial p_i}\right) = r^\mu \gamma^i_\mu, \tag{1.48}$$

for some $r^\mu(t, q, \dot{q})$.

We shall investigate the projectability of variations generated by diffeomorphisms in the following section. Assume that an infinitesimal transformation δq^i is projectable:

$$\Gamma_\mu \delta q^i = 0.$$

If δq^i is projectable, so must be r^μ, so that $r^\mu = \mathbb{F}L^*(r^\mu_{\rm H})$. Then, using (1.46) and (1.48), we see that

$$\delta q^i = \mathbb{F}L^*\left(\frac{\partial(G_H + r^\mu_{\rm H}\phi_\mu)}{\partial p_i}\right).$$

We now redefine $G_{\rm H}$ to absorb the piece $r^\mu_{\rm H}\phi_\mu$, and from now on we will have

$$\delta q^i = \mathbb{F}L^*\left(\frac{\partial G_{\rm H}}{\partial p_i}\right).$$

Define

$$\hat{p}_i = \frac{\partial L}{\partial \dot{q}^i};$$

after eliminating (1.47) times $\ddot{q}^i$ from (1.45), we get

$$\left(\frac{\partial L}{\partial q^i} - \dot{q}^k \frac{\partial \hat{p}_i}{\partial q^k}\right)\mathbb{F}L^*(\frac{\partial G_{\rm H}}{\partial p_i}) + \dot{q}^i \frac{\partial}{\partial q^i}\mathbb{F}L^*(G_{\rm H}) + \mathbb{F}L^*\partial_t G_{\rm H} = 0,$$

which simplifies to

$$\frac{\partial L}{\partial q^i}\mathbb{F}L^*(\frac{\partial G_{\rm H}}{\partial p_i}) + \dot{q}^i\mathbb{F}L^*(\frac{\partial G_{\rm H}}{\partial q^i}) + \mathbb{F}L^*\partial_t G_{\rm H} = 0. \tag{1.49}$$

Now let us invoke two identities [BGG89] that are at the core of the connection between the Lagrangian and the Hamiltonian equations of motion. They are

$$\dot{q}^i = \mathbb{F}L^*(\frac{\partial H}{\partial p_i}) + v^\mu(q,\dot{q})\mathbb{F}L^*(\frac{\partial \phi_\mu}{\partial p_i}),$$

and

$$\frac{\partial L}{\partial q^i} = -\mathbb{F}L^*(\frac{\partial H}{\partial q^i}) - v^\mu(q,\dot{q})\mathbb{F}L^*(\frac{\partial \phi_\mu}{\partial q^i});$$

where H is any canonical Hamiltonian, so that $\mathbb{F}L^*(H) = \dot{q}^i(\partial L/\partial \dot{q}^i) - L = \hat{E}$, the Lagrangian energy, and the functions v^μ are determined so as to render the first relation an identity. Notice the important relation

$$\Gamma_\mu v^\nu = \delta^\nu_\mu,$$

which stems from applying Γ_μ to the first identity and taking into account that

$$\Gamma_\mu \circ \mathbb{F}L^* = 0.$$

Substitution of these two identities into (1.49) induces (where $\{\,,\,\}$ denotes the *Poisson bracket*)

$$\mathbb{F}L^*\{G_{\rm H}, H\} + v^\mu \mathbb{F}L^*\{G_{\rm H}, \phi_\mu\} + \mathbb{F}L^*\partial_t G_{\rm H} = 0.$$

This result can be split through the action of Γ_μ into

$$\mathbb{F}L^*\{G_{\rm H}, H\} + \mathbb{F}L^*\partial_t G_{\rm H} = 0,$$

and

$$\mathbb{F}L^*\{G_{\rm H}, \phi_\mu\} = 0;$$

or equivalently,

$$\{G_{\rm H}, H\} + \partial_t G_{\rm H} = pc,$$

and

$$\{G_{\rm H}, \phi_\mu\} = pc,$$

where pc stands for any linear combination of primary constraints. In this way, we have arrived at a neat characterization for a generator $G_{\rm H}$ of Noether transformations in the canonical formalism.

Lie Symmetries in Biophysics

In this subsection we consider two most important equations for biophysics:

1. The heat equation, which has been analyzed in muscular mechanics since the early works of A.V. Hill ([Hil38]); and
2. The Korteveg–de Vries equation, the basic equation for solitary models of muscular excitation–contraction dynamics (see subsection (2.4.2) below).

Suppose

$$\mathcal{S} : \Delta_r(x, u^{(n)}) = 0, \qquad (r = 1, ..., l),$$

is a system of DEs of maximal rank defined over $M \subset X \times U$. If G is a local group of transformations acting on M, and

$$\mathbf{pr}^{(n)} v[\Delta_r(x, u^{(n)})] = 0, \qquad \text{whenever} \qquad \Delta(x, u^{(n)}) = 0, \tag{1.50}$$

(with $r = 1, ..., l$) for every infinitesimal generator v of G, then G is a symmetry group of the system $\mathcal{S}$ [Olv86].

The Heat Equation

Recall that the (1+1)D *heat equation* (with the thermal diffusivity normalized to unity)

$$u_t = u_{xx} \tag{1.51}$$

has two independent variables x and t, and one dependent variable u, so $p = 2$ and $q = 1$. Equation (1.51) has the second–order, $n = 2$, and can be identified with the linear submanifold $M^{(2)} \subset X \times U^{(2)}$ determined by the vanishing of $\Delta(x, t, u^{(2)}) = u_t - u_{xx}$.

Let

$$v = \xi(x, t, u)\frac{\partial}{\partial x} + \tau(x, t, u)\frac{\partial}{\partial t} + \phi(x, t, u)\frac{\partial}{\partial u}$$

be a vector–field on $X \times U$. According to (1.50) we need to now the second prolongation

$$\mathbf{pr}^{(2)} v = v + \phi^x \frac{\partial}{\partial u_x} + \phi^t \frac{\partial}{\partial u_t} + \phi^{xx} \frac{\partial}{\partial u_{xx}} + \phi^{xt} \frac{\partial}{\partial u_{xt}} + \phi^{tt} \frac{\partial}{\partial u_{tt}}$$

of v. Applying $\mathbf{pr}^{(2)} v$ to (1.51) we find the infinitesimal criterion (1.50) to be

$$\phi^t = \phi^{xx},$$

which must be satisfied whenever $u_t = u_{xx}$.

The Korteveg–De Vries Equation

Recall that the *Korteveg–de Vries equation*

$$u_t + u_{xxx} + uu_x = 0 \tag{1.52}$$

arises in physical systems in which both nonlinear and dispersive effects are relevant. A vector–field

$$v = \xi(x, t, u)\frac{\partial}{\partial x} + \tau(x, t, u)\frac{\partial}{\partial t} + \phi(x, t, u)\frac{\partial}{\partial u}$$

generates a one–parameter symmetry group iff

$$\phi^t + \phi^{xxx} + u\phi^x + u_x\phi = 0,$$

whenever u satisfies (1.52), etc.

Lie–Invariant Geometric Objects

Robot Kinematics

It is well known (see [BL92, Pry96]), that motion planning, numerically controlled machining and robotics are just a few of many areas of manufacturing automation in which the analysis and representation of *swept volumes* plays a crucial role. The swept volume modelling is also an important part of task-oriented robot motion planning. A typical *motion planning* problem consists in a collection of objects moving around obstacles from an initial to a final configuration. This may include in particular, solving the *collision detection* problem.

When a solid object undergoes a rigid motion, the totality of points through which it passed constitutes a region in space called the *swept volume*. To describe the geometrical structure of the swept volume we pose this problem as one of geometric study of some manifold swept by surface points using powerful tools from both modern differential geometry and nonlinear dynamical systems theory [Ric93, LP94, Pry96, GJ94] on manifolds. For some special cases of the Euclidean motion in the space $\mathbb{R}^3$ one can construct a very rich hydrodynamic system [BL92] modelling a sweep flow, which appears to be a *completely integrable Hamiltonian system* having a special *Lax type representation*. To describe in detail these and other properties of swept volume dynamical systems, we develop *Cartan's theory of Lie–invariant geometric objects* generated by closed ideals in the Grassmann's algebra, following [BPS98].

Let a Lie group G act on an analytical manifold Y in the transitive way, that is the action $G \times Y \xrightarrow{\rho} Y$ generates some nonlinear exact representation of the Lie group G on the manifold Y. In the frame of the Cartan's theory, the representation $G \times Y \xrightarrow{\rho} Y$ can be described by means of a system of differential 1–forms (see section 1.4.7 below)

$$\bar{\beta}^j = dy^j + \xi_i^j \bar{\omega}^i(a, da) \tag{1.53}$$

in the Grassmann algebra $\Lambda(Y \times G)$ on the product $Y \times G$, where $\bar{\omega}^i(a, da) \in T_a^*(G)$, $i = 1, ..., r = \dim G$ is a basis of left invariant Cartan's forms of the Lie group G at a point $a \in G$, $y = \{y^j : j = 1, ..., n = \dim Y\} \in Y$ and $\xi_i^j : Y \times G \to \mathbb{R}$ are some smooth real valued functions.

The following *Cartan theorem* (see [BPS98]) is basic in describing a geometric object invariant with respect to the mentioned above group action $G \times Y \xrightarrow{\rho} Y$: The system of differential forms (1.53) is a system of an invariant geometric object iff the following conditions are fulfilled:

1. The coefficients $\xi_i^j \in C^k(Y; R)$ for all $i = 1, ..., r$, $j = 1, ..., n$, are some *analytical functions* on Y; and
2. The differential system (1.53) is completely integrable within the *Frobenius–Cartan criterion*.

The Cartan's theorem actually says that the differential system (1.53) can be written down as

$$\bar{\beta}^j = dy^j + \xi^j_i(y)\bar{\omega}^i(a, da), \tag{1.54}$$

where 1–forms $\{\bar{\omega}^i(a, da) : i = 1, ..., r\}$ satisfy the standard *Maurer–Cartan equations*

$$\bar{\Omega}^j = d\bar{\omega}^j + \frac{1}{2}c^j_{ik}\bar{\omega}^i \wedge \bar{\omega}^k = 0 \tag{1.55}$$

for all $j = 1, ..., r$ on G, coefficients $c^j_{ik} \in \mathbb{R}$, $i, j, k = 1, ..., r$, being the corresponding structure constants of the Lie algebra $\mathcal{G}$ of the Lie group G.

Maurer–Cartan 1–Forms

Let be given a Lie group G with the Lie algebra $\mathcal{G} \simeq T_e(G)$, whose basis is a set $\{A_i \in \mathcal{G} : i = 1, ..., r\}$, where $r = \dim G \equiv \dim \mathcal{G}$. Let also a set $U_0 \subset \{a^i \in \mathbb{R} : i = 1, ..., r\}$ be some open neighborhood of the zero point in $\mathbb{R}^r$. The exponential mapping $\exp : U_0 \to G_0$, where by definition [BPS98]

$$\mathbb{R}^r \supset U_0 \ni (a^1, \dots, a^r) : \xrightarrow{\exp} \exp\left(a^i A_i\right) = a \in G_0 \subset G, \tag{1.56}$$

is an analytical mapping of the whole U_0 on some open neighborhood G_0 of the unity element $e \in G$. From (1.56) it is easy to find that $T_e(G) = T_e(G_0) \simeq \mathcal{G}$, where $e = \exp(0) \in G$. Define now the following left invariant $\mathcal{G}-$valued differential 1–form on $G_0 \subset G$:

$$\bar{\omega}(a, da) = a^{-1}da = \bar{\omega}^j(a, da)A_j, \tag{1.57}$$

where $A_j \in \mathcal{G}$, $\bar{\omega}^j(a, da) \in T^*_a(G)$, $a \in G_0$, $j = 1, ..., r$. To build effectively the unknown forms $\{\bar{\omega}^j(a, da) : j = 1, ..., r\}$, let us consider the following analytical one–parameter 1–form $\bar{\omega}_t(a, da) = \bar{\omega}(a_t; da_t)$ on G_0, where $a_t = \exp\left(ta^i A_i\right)$, $t \in [0, 1]$, and differentiate this form with respect to the parameter $t \in [0, 1]$. We will get [BPS98]

$$d\bar{\omega}_t/dt = -a^j A_j a_t^{-1} da_t + a_t^{-1} a_t da^j A_j + a_t^{-1} da_t a^j A_j = -a^j[A_j, \bar{\omega}_t] + A_j da_j. \tag{1.58}$$

Having used the Lie identity $[A_j, A_k] = c^i_{jk}A_i$, $j, k = 1, ..., r$, and the right hand side of (1.57) in form

$$\bar{\omega}^j(a, da) = \bar{\omega}^j_k(a)da^k, \tag{1.59}$$

we ultimately obtain that

$$\frac{d}{dt}(t\bar{\omega}^j_i(ta)) = \mathcal{A}^j_k t\bar{\omega}^k_i(ta) + \delta^j_i, \tag{1.60}$$

where the matrix $\mathcal{A}^k_i$, $i, k = 1, ..., r$, is defined as follows:

$$\mathcal{A}^k_i = c^k_{ij}a^j. \tag{1.61}$$

Thus, the matrix $W_i^j(t) = t\bar{\omega}_i^j(ta)$, $i,j = 1,...,r$, satisfies the following from (1.60) differential equation [Che55]

$$dW/dt = \mathcal{A}W + E, \qquad W|_{t=0} = 0, \tag{1.62}$$

where $E = \|\delta_i^j\|$ is the unity matrix. The solution of (1.62) is representable as

$$W(t) = \sum_{n=1}^{\infty} \frac{t^n}{n!}\mathcal{A}^{n-1} \tag{1.63}$$

for all $t \in [0,1]$. Whence, recalling the above definition of the matrix $W(t)$, we obtain easily that

$$\bar{\omega}_k^j(a) = W_k^j(t)\Big|_{t=1} = \sum_{n=1}^{\infty}(n!)^{-1}\mathcal{A}^{n-1}. \tag{1.64}$$

Therefore, the following theorem solves the problem of finding in an effective algebraic way corresponding to a Lie algebra $\mathcal{G}$ the left invariant 1–form $\bar{\omega}(a,da) \in T_a^*(G) \otimes \mathcal{G}$ at any $a \in G$: Let's be given a Lie algebra G with the structure constants $c_{ij}^k \in R$, $i,j,k = 1,...,r = \dim \mathcal{G}$, related to some basis $\{A_j \in G : j = 1,...,r\}$. Then the adjoint to G left–invariant Maurer–Cartan 1–form $\bar{\omega}(a,da)$ is built as follows [BPS98]:

$$\bar{\omega}(a,da) = A_j\bar{\omega}_k^j(a)da^k, \tag{1.65}$$

where the matrix $W = \|\bar{w}_k^j(a)\|$, $j,k = 1,...,r$, is given exactly as

$$W = \sum_{n=1}^{\infty}(n!)^{-1}\mathcal{A}^{n-1}, \qquad \mathcal{A}_k^j = c_{ki}^j a^i. \tag{1.66}$$

Below we shall try to use the experience gained above in solving an analogous problem of the theory of connections over a principal fibre bundle $P(M;G)$ as well as over associated with it a fibre bundle $P(M;Y,G)$.

General Structure of Integrable One–Forms

Given 2−forms generating a closed ideal $\mathfrak{I}(\alpha)$ in the Grassmann algebra $\Lambda(M)$, we will denote as above by $\mathfrak{I}(\alpha,\beta)$ an augmented ideal in $\Lambda(M;Y)$, where the manifold Y will be called in further the representation space of some adjoint Lie group G action: $G \times Y \xrightarrow{\rho} Y$. We can find, therefore, the determining relationships for the set of 1–forms $\{\beta\}$ and 2–forms $\{\alpha\}$

$$\begin{aligned} \{\alpha\} &= \{\alpha^j \in \Lambda^2(M) : j = 1,...,m_\alpha\}, \\ \{\beta\} &= \{\beta^j \in \Lambda^1(M \times Y) : j = 1,...,n = \dim Y\}, \end{aligned} \tag{1.67}$$

satisfying such equations [BPS98]:

$$
\begin{aligned}
d\alpha^i &= a^i_k(\alpha) \wedge \alpha^k, \\
d\beta^j &= f^j_k \alpha^k + \omega^j_s \wedge \beta^s,
\end{aligned}
\tag{1.68}
$$

where $a^i_k(\alpha) \in \Lambda^1(M)$, $f^j_k \in \Lambda^0(M \times Y)$ and $\omega^j_s \in \Lambda^1(M \times Y)$ for all $i, k = 1, ..., m_\alpha$, $j, s = 1, ..., n$. Since the identity $d^2\beta^j \equiv 0$ takes place for all $j = 1, ..., n$, from (1.68) we deduce the following relationship:

$$
\left(d\omega^j_k + \omega^j_s \wedge \omega^s_k\right) \wedge \beta^k + \left(df^j_s + \omega^j_k f^k_s + f^j_l a^l_s(\alpha)\right) \wedge \alpha^s \equiv 0. \tag{1.69}
$$

As a result of (1.69) we obtain [BPS98]

$$
\begin{aligned}
&d\omega^j_k + \omega^j_s \wedge \omega^s_k \in \mathfrak{I}(\alpha, \beta), \\
&df^j_s + \omega^j_k f^k_s + f^j_l a^l_s(\alpha) \in \mathfrak{I}(\alpha, \beta)
\end{aligned}
\tag{1.70}
$$

for all $j, k = 1, ..., n$, $s = 1, ..., m_\alpha$. The second inclusion in (1.70) gives a possibility to define the 1–forms $\theta^j_s = f^j_l a^l_s(\alpha)$ satisfying the inclusion

$$
d\theta^j_s + \omega^j_k \wedge \theta^k_s \in \mathfrak{I}(\alpha, \beta) \oplus f^j_l c^l_s(\alpha), \tag{1.71}
$$

which we obtained having used the identities $d^2\alpha^j \equiv 0$, $j = 1, ..., m_\alpha$, in the form $c^j_s(\alpha) \wedge \alpha^s \equiv 0$,

$$
c^j_s(\alpha) = da^j_s(\alpha) + a^j_l(\alpha) \wedge a^l_s(\alpha), \tag{1.72}
$$

following from (1.68). Let us suppose further that as $s = s_0$ the 2–forms $c^j_{s_0}(\alpha) \equiv 0$ for all $j = 1, ..., m_\alpha$. Then as $s = s_0$, we can define a set of 1–forms $\theta^j = \theta^j_{s_0} \in \Lambda^1(M \times Y)$, $j = 1, ..., n$, satisfying the exact inclusions:

$$
d\theta^j + \omega^j_k \wedge \theta^k = \Theta^j \in \mathfrak{I}(\alpha, \beta) \tag{1.73}
$$

together with a set of inclusions for 1–forms $\omega^j_k \in \Lambda^1(M \times Y)$

$$
d\omega^j_k + \omega^j_s \wedge \omega^s_k = \Omega^j_k \in \mathfrak{I}(\alpha, \beta) \tag{1.74}
$$

As it follows from the general theory [SW72] of connections on the fibred frame space $P(M; GL(n))$ over a base manifold M, we can interpret the equations (1.74) as the equations defining the curvature 2–forms $\Omega^j_k \in \Lambda^2(P)$, as well as interpret the equations (1.73) as those, defining the torsion 2–forms $\Theta^j \in \Lambda^2(P)$. Since $\mathfrak{I}(\alpha) = 0 = \mathfrak{I}(\alpha, \beta)$ upon the integral submanifold $\bar{M} \subset M$, the reduced fibred frame space $P(\bar{M}; GL(n))$ will have the flat curvature and be torsion free, being as a result, completely trivialized on $\bar{M} \subset M$. Consequently, we can formulate the following theorem.

Let the condition above on the ideals $\mathfrak{I}(\alpha)$ and $\mathfrak{I}(\alpha, \beta)$ be fulfilled. Then the set of 1–forms $\{\beta\}$ generates the integrable augmented ideal $\mathfrak{I}(\alpha, \beta) \subset \Lambda(M \times Y)$ iff there exists some curvature 1–form $\omega \in \Lambda^1(P) \otimes Gl(n)$ and

torsion 1–form $\theta \in \Lambda^1(P) \otimes \mathbb{R}^n$ on the adjoint fibred frame space $P(M; GL(n))$, satisfying the inclusions [BPS98]

$$\begin{aligned} &d\omega + \omega \wedge \omega \in \mathfrak{I}(\alpha, \beta) \otimes \mathcal{G}l(n), \\ &d\theta + \omega \wedge \theta \in \mathfrak{I}(\alpha, \beta) \otimes \mathbb{R}^n. \end{aligned} \tag{1.75}$$

Upon the reduced fibred frame space $P(\bar{M}; GL(n))$ the corresponding curvature and torsion are vanishing, where $\bar{M} \subset M$ is the integral submanifold of the ideal $\mathfrak{I}(\alpha) \subset \Lambda(M)$.

Lax Integrable Dynamical Systems

Consider some set $\{\beta\}$ defining a Cartan's Lie group G invariant object on a manifold $M \times Y$:

$$\beta^j = dy^j + \xi^j_k(y) b^k(z), \tag{1.76}$$

where $i = 1, ..., n = \dim Y$, $r = \dim G$. The set (1.76) defines on the manifold Y a set $\{\xi\}$ of vector–fields, compiling a representation $\rho : \mathcal{G} \to \{\xi\}$ of a given Lie algebra $\mathcal{G}$, that is vector–fields $\xi_s = \xi^j_s(y) \frac{\partial}{\partial y^j} \in \{\xi\}$, $s = 1, ..., r$, enjoy the following Lie algebra $\mathcal{G}$ relationships

$$[\xi_s, \xi_l] = c^k_{sl} \xi_k \tag{1.77}$$

for all $s, l, k = 1, ..., r$. We can now compute the differentials $d\beta^j \in \Lambda^2(M \times Y)$, $j = 1, ..., n$, using (1.76) and (1.77) as follows [BPS98]:

$$d\beta^j = \tfrac{\partial \xi^j_k(y)}{\partial y^l} \left(\beta^l - \xi^l_s(y) b^s(z) \right) \wedge b^k(z) + \xi^j_k(y) db^k(z) \tag{1.78}$$

which is equal to

$$\frac{\partial \xi^j_k(y)}{\partial y^l} \beta^l \wedge b_k(z) + \xi^j_l \left(db^l(z) + \frac{1}{2} c^l_{ks} db^k(z) \wedge db^s(z) \right),$$

where $\{\alpha\} \subset \Lambda^2(M)$ is some *a priori* given integrable system of 2–forms on M, vanishing upon the integral submanifold $\bar{M} \subset M$. It is obvious that inclusions (1.78) take place iff the following conditions are fulfilled: for all $j = 1, ..., r$

$$db^j(z) + \frac{1}{2} c^j_{ks} db^k(z) \wedge db^s(z) \in \mathfrak{I}(\alpha). \tag{1.79}$$

The inclusions (1.79) mean in particular, that upon the integral submanifold $\bar{M} \subset M$ of the ideal $\mathfrak{I}(\alpha) \subset \Lambda(M)$ the equalities

$$\mu^* \bar{\omega}^j \equiv s^* b^j \tag{1.80}$$

are true, where $\bar{\omega}^j \in T^*_e(G)$, $j = 1, ..., r$, are the left invariant Maurer–Cartan forms on the invariance Lie group G. Thus, due to inclusions (1.79) all conditions of Cartan's theorem are enjoyed, giving rise to a possibility to obtain

the set of forms $b^j(z) \in \Lambda^1(M)$ in an explicit form. To do this, let us define a $\mathcal{G}-$valued curvature 1–form $\omega \in \Lambda^1(P(M;G)) \otimes \mathcal{G}$ as follows [BPS98]

$$\omega = Ad_{a^{-1}} \left(A_j b^j\right) + \bar{\omega} \tag{1.81}$$

where $\bar{\omega} \in \mathcal{G}$ is the standard Maurer–Cartan 1–form on G. This 1–form satisfies followed by (1.79) the canonical structure inclusion for $\Gamma = A_j b^j \in \Lambda^1(M) \otimes \mathcal{G}$:

$$d\Gamma + \Gamma \wedge \Gamma \in \mathfrak{I}(\alpha) \otimes \mathcal{G}, \tag{1.82}$$

serving as a main relationships determining the form (1.81). To proceed further we need to give the set of 2–forms $\{\alpha\} \subset \Lambda^2(M)$ in explicit form. A standard example is the *Burgers dynamical system* (see equation (2.565) below).

1.2.9 Riemannian Manifolds

Local Riemannian Geometry

An important class of problems in Riemannian geometry is to understand the interaction between the curvature and topology on a smooth manifold (see [CC99]). A prime example of this interaction is the *Gauss–Bonnet formula* on a closed surface M^2, which says

$$\int_M K \, dA = 2\pi \, \chi(M), \tag{1.83}$$

where dA is the area element of a metric g on M, K is the Gaussian curvature of g, and $\chi(M)$ is the *Euler characteristic* of M.

To study the geometry of a smooth manifold we need an additional structure: the *Riemannian metric tensor*. The metric is an inner product on each of the tangent spaces and tells us how to measure angles and distances infinitesimally. In local coordinates $(x^1, x^2, \cdots, x^n)$, the metric g is given by $g_{ij}(x)\, dx^i \otimes dx^j$, where $(g_{ij}(x))$ is a positive definite symmetric matrix at each point x. For a smooth manifold one can differentiate functions. A Riemannian metric defines a natural way of differentiating vector–fields: *covariant differentiation*. In Euclidean space, one can change the order of differentiation. On a Riemannian manifold the commutator of twice covariant differentiating vector–fields is in general nonzero and is called the *Riemann curvature tensor*, which is a 4−tensor–field on the manifold.

For surfaces, the Riemann curvature tensor is equivalent to the *Gaussian curvature* K, a scalar function. In dimensions 3 or more, the Riemann curvature tensor is inherently a tensor–field. In local coordinates, it is denoted by R_{ijkl}, which is anti-symmetric in i and k and in j and l, and symmetric in the pairs $\{ij\}$ and $\{kl\}$. Thus, it can be considered as a bilinear form on 2−forms which is called the *curvature operator*. We now describe heuristically the various curvatures associated to the Riemann curvature tensor. Given a

point $x \in M^n$ and 2-plane Π in the tangent space of M at x, we can define a surface S in M to be the union of all geodesics passing through x and tangent to Π. In a neighborhood of x, S is a smooth 2D submanifold of M. We define the *sectional curvature* $K(\Pi)$ of the 2−plane to be the Gauss curvature of S at x:

$$K(\Pi) = K_S(x).$$

Thus the sectional curvature K of a Riemannian manifold associates to each 2-plane in a tangent space a real number. Given a line L in a tangent space, we can average the sectional curvatures of all planes through L to get the *Ricci tensor* $Rc(L)$. Likewise, given a point $x \in M$, we can average the Ricci curvatures of all lines in the tangent space of x to get the *scalar curvature* $R(x)$. In local coordinates, the Ricci tensor is given by $R_{ik} = g^{jl}R_{ijkl}$ and the scalar curvature is given by $R = g^{ik}R_{ik}$, where $(g^{ij}) = (g_{ij})^{-1}$ is the inverse of the metric tensor (g_{ij}).

Riemannian Metric on M

In this subsection we mainly follow [Pet99, Pet98].

Riemann in 1854 observed that around each point $m \in M$ one can pick a *special* coordinate system $(x^1, \ldots, x^n)$ such that there is a symmetric $(0,2)$−tensor–field $g_{ij}(m)$ called the *metric tensor* defined as

$$g_{ij}(m) = g(\partial_{x^i}, \partial_{x^j}) = \delta_{ij}, \qquad \partial_{x^k} g_{ij}(m) = 0.$$

Thus the metric, at the specified point $m \in M$, in the coordinates $(x^1, \ldots, x^n)$ looks like the Euclidean metric on $\mathbb{R}^n$. We emphasize that these conditions only hold at the specified point $m \in M$. When passing to different points it is necessary to pick different coordinates. If a curve γ passes through m, say, $\gamma(0) = m$, then the acceleration at 0 is defined by firstly, writing the curve out in our special coordinates

$$\gamma(t) = (\gamma^1(t), \ldots, \gamma^n(t)),$$

secondly, defining the tangent, *velocity* vector–field, as

$$\dot{\gamma} = \dot{\gamma}^i(t) \cdot \partial_{x^i},$$

and finally, the *acceleration* vector–field as

$$\ddot{\gamma}(0) = \ddot{\gamma}^i(0) \cdot \partial_{x^i}.$$

Here, the background idea is that we have a *connection*.

Recall that a *connection* on a smooth manifold M tells us how to *parallel transport a vector* at a point $x \in M$ to a vector at a point $x' \in M$ along a curve $\gamma \in M$. Roughly, to parallel transport vectors along curves, it is enough if we can define parallel transport under an infinitesimal displacement: given a vector X at x, we would like to define its parallel transported version $\tilde{X}$

after an infinitesimal displacement by ϵv, where v is a tangent vector to M at x.

More precisely, a vector–field X along a parameterized curve $\alpha : I \to M$ in M is *tangent to* M *along* α if $X(t) \in M_{\alpha(t)}$ for all for $t \in I \subset \mathbb{R}$. However, the derivative $\dot{X}$ of such a vector–field is, in general, not tangent to M. We can, nevertheless, get a vector–field tangent to M by projecting $\dot{X}(t)$ orthogonally onto $M_{\alpha(t)}$ for each $t \in I$. This process of differentiating and then projecting onto the tangent space to M defines an operation with the same properties as differentiation, except that now differentiation of vector–fields tangent to M induces vector–fields tangent to M. This operation is called *covariant differentiation*.

Let $\gamma : I \to M$ be a parameterized curve in M, and let X be a smooth vector–field tangent to M along α. The *absolute covariant derivative* of X is the vector–field $\dot{\dot{X}}$ tangent to M along α, defined by $\dot{\dot{X}} = \dot{X}(t) - [\dot{X}(t) \cdot N(\alpha(t))]\, N(\alpha(t))$, where N is an orientation on M. Note that $\dot{\dot{X}}$ is independent of the choice of N since replacing N by $-N$ has no effect on the above formula.

Lie bracket (1.2.6) defines a *symmetric affine connection* ∇ on any manifold M:

$$[X, Y] = \nabla_X Y - \nabla_Y X.$$

In case of a Riemannian manifold M, the connection ∇ is also compatible with the Riemannian metrics g on M and is called the *Levi–Civita connection* on TM.

For a function $f \in C^\infty(M, \mathbb{R})$ and a vector a vector–field $X \in \mathcal{X}^k(M)$ we always have the Lie derivative (1.2.6)

$$\mathcal{L}_X f = \nabla_X f = df(X).$$

But there is no natural definition for $\nabla_X Y$, where $Y \in \mathcal{X}^k(M)$, unless one also has a Riemannian metric. Given the tangent field $\dot{\gamma}$, the acceleration can then be computed by using a Leibniz rule on the r.h.s, if we can make sense of the derivative of ∂_{x^i} in the direction of $\dot{\gamma}$. This is exactly what the *covariant derivative* $\nabla_X Y$ does. If $Y \in T_m M$ then we can write $Y = a^i \partial_{x^i}$, and therefore

$$\nabla_X Y = \mathcal{L}_X a^i \partial_{x^i}. \tag{1.84}$$

Since there are several ways of choosing these coordinates, one must check that the definition does not depend on the choice. Note that for two vector–fields we define $(\nabla_Y X)(m) = \nabla_{Y(m)} X$. In the end we get a *connection*

$$\nabla : \mathcal{X}^k(M) \times \mathcal{X}^k(M) \to \mathcal{X}^k(M),$$

which satisfies (for all $f \in C^\infty(M, \mathbb{R})$ and $X, Y, Z \in \mathcal{X}^k(M)$):

1. $Y \to \nabla_Y X$ is tensorial, i.e., linear and $\nabla_{fY} X = f \nabla_Y X$.
2. $X \to \nabla_Y X$ is linear.
3. $\nabla_X (fY) = (\nabla_X f) Y(m) + f(m) \nabla_X Y$.

4. $\nabla_X Y - \nabla_Y X = [X, Y]$.
5. $\mathcal{L}_X g(Z, Y) = g(\nabla_X Z, Y) + g(Z, \nabla_X Y)$.

A semicolon is commonly used to denote covariant differentiation with respect to a natural basis vector. If $X = \partial_{x^i}$, then the components of $\nabla_X Y$ in (1.84) are denoted

$$Y^k_{;\,i} = \partial_{x^i} Y^k + \Gamma^k_{ij}\, Y^j, \tag{1.85}$$

where Γ^k_{ij} are *Christoffel symbols* defined in (1.86) below. Similar relations hold for higher–order tensor–fields (with as many terms with Christoffel symbols as is the tensor valence).

Therefore, no matter which coordinates we use, we can now define the acceleration of a curve in the following way:

$$\begin{aligned}
\gamma(t) &= (\gamma^1(t), \ldots, \gamma^n(t)),\\
\dot\gamma(t) &= \dot\gamma^i(t)\partial_{x^i},\\
\ddot\gamma(t) &= \ddot\gamma^i(t)\partial_{x^i} + \dot\gamma^i(t)\nabla_{\dot\gamma(t)}\partial_{x^i}.
\end{aligned}$$

We call γ a *geodesic* if $\gamma(t) = 0$. This is a second–order nonlinear ODE in a fixed coordinate system $(x^1, \ldots, x^n)$ at the specified point $m \in M$. Thus we see that given any tangent vector $X \in T_m M$, there is a unique geodesic $\gamma_X(t)$ with $\dot\gamma_X(0) = X$. If the manifold M is closed, the geodesic must exist for all time, but in case the manifold M is open this might not be so. To see this, take as M any open subset of Euclidean space with the induced metric.

Given an arbitrary vector–field $Y(t)$ along γ, i.e., $Y(t) \in T_{\gamma(t)}M$ for all t, we can also define the derivative $\dot Y \equiv \frac{dY}{dt}$ in the direction of $\dot\gamma$ by writing

$$\begin{aligned}
Y(t) &= a^i(t)\partial_{x^i},\\
\dot Y(t) &= \dot a^i(t)\partial_{x^i} + a^i(t)\nabla_{\dot\gamma(t)}\partial_{x^i}.
\end{aligned}$$

Here the derivative of the tangent field $\dot\gamma$ is the acceleration γ. The field Y is said to be *parallel* iff $\dot Y = 0$. The equation for a field to be parallel is a first–order linear ODE, so we see that for any $X \in T_{\gamma(t_0)}M$ there is a unique parallel field $Y(t)$ defined on the entire domain of γ with the property that $Y(t_0) = X$. Given two such parallel fields $Y, Z \in \mathcal{X}^k(M)$, we have that

$$\dot g(Y, Z) = D_{\dot\gamma} g(Y, Z) = g(\dot Y, Z) + g(Y, \dot Z) = 0.$$

Thus X and Y are both of constant length and form constant angles along γ. Hence, 'parallel translation' along a curve defines an orthogonal transformation between the tangent spaces to the manifold along the curve. However, in contrast to Euclidean space, this parallel translation will depend on the choice of curve.

An infinitesimal distance between the two nearby local points m and n on M is defined by an *arc–element*

$$ds^2 = g_{ij}\, dx^i dx^j,$$

and realized by the curves $x^i(s)$ of shortest distance, called *geodesics*, addressed by the *Hilbert 4th problem*. In local coordinates $(x^1(s), ..., x^n(s))$ at a point $m \in M$, the geodesic defining equation is a second–order ODE,

$$\ddot{x}^i + \Gamma^i_{jk}\, \dot{x}^j\, \dot{x}^k = 0,$$

where the overdot denotes the derivative with respect to the affine parameter s, $\dot{x}^i(s) = \frac{d}{ds}x^i(s)$ is the tangent vector to the base geodesic, while the *Christoffel symbols* $\Gamma^i_{jk} = \Gamma^i_{jk}(m)$ of the *affine Levi–Civita connection* ∇ at the point $m \in M$ are defined, in a holonomic coordinate basis e_i as

$$\Gamma^k_{ij} = g^{kl}\Gamma_{ijl}, \qquad \text{with} \qquad g^{ij} = (g_{ij})^{-1} \qquad \text{and} \qquad (1.86)$$
$$\Gamma_{ijk} = \frac{1}{2}(\partial_{x^k} g_{ij} + \partial_{x^j} g_{ki} - \partial_{x^i} g_{jk}).$$

Note that the Christoffel symbols (1.86) do not transform as tensors on the tangent bundle. They are the components of an object on the second tangent bundle, a *spray*. However, they do transform as tensors on the *jet space* (see subsection 1.4.3 below).

In *nonholonomic coordinates*, (1.86) takes the extended form

$$\Gamma^i_{kl} = \frac{1}{2} g^{im}\left(\partial_{x^l} g_{mk} + \partial_{x^k}\partial g_{ml} - \partial_{x^m}\partial g_{kl} + c_{mkl} + c_{mlk} - c_{klm}\right),$$

where $c_{klm} = g_{mp}c^p_{kl}$ are the commutation coefficients of the basis, i.e., $[e_k, e_l] = c^m_{kl}e_m$.

The *torsion* tensor–field T of the connection ∇ is the function $T : \mathcal{X}^k(M) \times \mathcal{X}^k(M) \to \mathcal{X}^k(M)$ given by

$$T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y].$$

From the skew symmetry ($[X, Y] = -[Y, X]$) of the Lie bracket, follows the skew symmetry ($T(X, Y) = -T(Y, X)$) of the torsion tensor. The mapping T is said to be $f-$bilinear since it is linear in both arguments and also satisfies $T(fX, Y) = fT(X, Y)$ for smooth functions f. Since $[\partial_{x^i}, \partial_{x^j}] = 0$ for all $1 \le i, j \le n$, it follows that

$$T(\partial_{x^i}, \partial_{x^j}) = (\Gamma^k_{ij} - \Gamma^k_{ji})\partial_{x^k}.$$

Consequently, torsion T is a $(1, 2)$ tensor–field, locally given by

$$T = T^k_{i\,j}\, dx^i \otimes \partial_{x^k} \otimes dx^j,$$

where the torsion components $T^k_{i\,j}$ are given by

$$T^k_{i\,j} = \Gamma^k_{ij} - \Gamma^k_{ji}.$$

Therefore, the torsion tensor gives a measure of the nonsymmetry of the connection coefficients. Hence, $T = 0$ if and only if these coefficients are symmetric

in their subscripts. A connection ∇ with $T = 0$ is said to be *torsion free* or *symmetric*.

The connection also enables us to define many other classical concepts from calculus in the setting of Riemannian manifolds. Suppose we have a function $f \in C^\infty(M, \mathbb{R})$. If the manifold is not equipped with a Riemannian metric, then we have the differential of f defined by $df(X) = \mathcal{L}_X f$, which is a $1-$form. The dual concept, the *gradient* of f, is supposed to be a vector–field. But we need a metric g to define it. Namely, ∇f is defined by the relationship

$$g(\nabla f, X) = df(X).$$

Having defined the gradient of a function on a Riemannian manifold, we can then use the connection to define the *Hessian* as the linear map

$$\nabla^2 f : TM \to TM, \qquad \nabla^2 f(X) = \nabla_X \nabla f.$$

The corresponding bilinear map is then defined as

$$\nabla^2 f(X, Y) = g(\nabla^2 f(X), Y).$$

One can check that this is a symmetric bilinear form. The *Laplacian* of f, Δf, is now defined as the trace of the Hessian

$$\Delta f = \mathrm{Tr}(\nabla^2 f(X)) = \mathrm{Tr}(\nabla_X \nabla f),$$

which is a linear map. It is also called the *Laplace–Beltrami operator*, since Beltrami first considered this operator on Riemannian manifolds.

Riemannian metric has the following mechanical interpretation. Let M be a closed Riemannian manifold with the *mechanical metric* $g = g_{ij} v^i v^j \equiv \langle v, v \rangle$, with $v^i = \dot{x}^i$. Consider the *Lagrangian function*

$$L : TM \to \mathbb{R}, \qquad (x, v) \mapsto \frac{1}{2}\langle v, v \rangle - U(x) \tag{1.87}$$

where $U(x)$ is a smooth function on M called the *potential*. On a fixed level of energy E, bigger than the maximum of U, the Lagrangian flow generated by (1.87) is conjugate to the geodesic flow with metric $\bar{g} = 2(e - U(x))\langle v, v \rangle$. Moreover, the reduced action of the Lagrangian is the distance for $g = \langle v, v \rangle$ [Arn89, AMR88]. Both of these statements are known as the *Maupertius action principle*.

Geodesics on M

For a $C^\infty, k \geq 2$ curve $\gamma : I \to M$, we define its *length* on I as

$$L(\gamma, I) = \int_I |\dot{\gamma}| \; dt = \int_I \sqrt{g(\dot{\gamma}, \dot{\gamma})} dt.$$

This length is independent of our parametrization of the curve γ. Thus the curve γ can be reparameterized, in such a way that it has unit velocity. The

distance between two points m_1 and m_2 on M, $d(m_1, m_2)$, can now be defined as the infimum of the lengths of all curves from m_1 to m_2, i.e.,

$$L(\gamma, I) \to \min.$$

This means that the distance measures the shortest way one can travel from m_1 to m_2.

If we take a variation $V(s,t) : (-\varepsilon, \varepsilon) \times [0, \ell] \to M$ of a smooth curve $\gamma(t) = V(0,t)$ parameterized by arc–length L and of length ℓ, then the first derivative of the arc–length function

$$\begin{aligned} L(s) &= \int_0^\ell |\dot{V}|\, dt, \qquad \text{is given by} \\ \frac{dL(0)}{ds} &\equiv \dot{L}(0) = g(\dot{\gamma}, X)\big|_0^\ell - \int_0^\ell g(\gamma, X)\, dt, \end{aligned} \tag{1.88}$$

where $X(t) = \frac{\partial V}{\partial s}(0,t)$ is the so–called *variation vector–field*. Equation (1.88) is called the *first variation formula*. Given any vector–field X along γ, one can produce a variation whose variational field is X. If the variation fixes the endpoints, $X(a) = X(b) = 0$, then the second term in the formula drops out, and we note that the length of γ can always be decreased as long as the acceleration of γ is not everywhere zero. Thus the *Euler–Lagrangian equations* for the arc–length functional are the equations for a curve to be a *geodesic*.

In local coordinates $x^i \in U$, where U is an open subset in the Riemannian manifold M, the geodesics are defined by the *geodesic equation* (see Appendix)

$$\ddot{x}^i + \Gamma^i_{jk} \dot{x}^j \dot{x}^k = 0, \tag{1.89}$$

where overdot means derivative upon the line parameter s, while Γ^i_{jk} are Christoffel symbols of the affine Levi–Civita connection ∇ on M. From (1.89) it follows that the linear *connection homotopy*,

$$\bar{\Gamma}^i_{jk} = s\Gamma^i_{jk} + (1-s)\Gamma^i_{jk}, \qquad (0 \le s \le 1),$$

determines the same geodesics as the original Γ^i_{jk}.

Riemannian Curvature on M

The *Riemann curvature tensor* is a rather ominous tensor of type $(1,3)$; i.e., it has three vector variables and its value is a vector as well. It is defined through the Lie bracket (1.2.6) as

$$R(X,Y)Z = \big(\nabla_{[X,Y]} - [\nabla_X, \nabla_Y]\big) Z = \nabla_{[X,Y]} Z - \nabla_X \nabla_Y Z + \nabla_Y \nabla_X Z.$$

This turns out to be a vector valued $(1,3)-$tensor–field in the three variables $X, Y, Z \in \mathcal{X}^k(M)$. We can then create a $(0,4)-$tensor,

$$R(X,Y,Z,W) = g\left(\nabla_{[X,Y]}Z - \nabla_X\nabla_Y Z + \nabla_Y\nabla_X Z, W\right).$$

Clearly this tensor is skew–symmetric in X and Y, and also in Z and $W \in \mathcal{X}^k(M)$. This was already known to Riemann, but there are some further, more subtle properties that were discovered a little later by Bianchi. The *Bianchi symmetry condition* reads

$$R(X,Y,Z,W) = R(Z,W,X,Y).$$

Thus the Riemann curvature tensor is a symmetric *curvature operator*

$$\mathfrak{R} : \Lambda^2 TM \to \Lambda^2 TM.$$

The *Ricci tensor* is the $(1,1)-$ or $(0,2)-$tensor defined by

$$\mathrm{Ric}(X) = R(\partial_{x^i}, X)\partial_{x^i}, \qquad \mathrm{Ric}(X,Y) = g(R(\partial_{x^i}, X)\partial_{x^i}, Y),$$

for any orthonormal basis (∂_{x^i}). In other words, the Ricci curvature is a trace of the curvature tensor. Similarly one can define the *scalar curvature* as the trace

$$\mathrm{scal}(m) = \mathrm{Tr}\,(\mathrm{Ric}) = \mathrm{Ric}(\partial_{x^i}, \partial_{x^i}).$$

When the Riemannian manifold has dimension 2, all of these curvatures are essentially the same. Since $\dim \Lambda^2 TM = 1$ and is spanned by $X \wedge Y$ where $X, Y \in \mathcal{X}^k(M)$ form an orthonormal basis for $T_m M$, we see that the curvature tensor depends only on the scalar value

$$K(m) = R(X,Y,X,Y),$$

which also turns out to be the *Gaussian curvature*. The Ricci tensor is a *homothety*

$$\mathrm{Ric}(X) = K(m)X, \qquad \mathrm{Ric}(Y) = K(m)Y,$$

and the scalar curvature is twice the Gauss curvature. In dimension 3 there are also some redundancies as $\dim TM = \dim \Lambda^2 TM = 3$. In particular, the Ricci tensor and the curvature tensor contain the same amount of information.

The *sectional curvature* is a kind of generalization of the Gauss curvature whose importance Riemann was already aware of. Given a $2-$plane $\pi \subset T_m M$ spanned by an orthonormal basis $X, Y \in \mathcal{X}^k(M)$ it is defined as

$$\mathrm{sec}(\pi) = R(X,Y,X,Y).$$

The remarkable observation by Riemann was that the *curvature operator is a homothety*, i.e., looks like $\mathfrak{R} = kI$ on $\Lambda^2 T_m M$ iff all sectional curvatures of planes in $T_m M$ are equal to k. This result is not completely trivial, as the sectional curvature is not the entire quadratic form associated to the symmetric operator $\mathfrak{R}$. In fact, it is not true that $\mathrm{sec} \geq 0$ implies that the curvature operator is nonnegative in the sense that all its eigenvalues are

nonnegative. What Riemann did was to show that our special coordinates $(x^1, \ldots, x^n)$ at m can be chosen to be *normal* at m, i.e., satisfy the condition

$$x^i = \delta^i_j x^j, \qquad (\delta^i_j x^j = g_{ij})$$

on a neighborhood of m. One can show that such coordinates are actually exponential coordinates together with a choice of an orthonormal basis for $T_m M$ so as to identify $T_m M$ with $\mathbb{R}^n$. In these coordinates one can then expand the metric as follows:

$$g_{ij} = \delta_{ij} - \frac{1}{3} R_{ikjl} x^k x^l + O\left(r^3\right).$$

Now the equations $x^i = g_{ij} x^j$ evidently give conditions on the curvatures R_{ijkl} at m.

If $\Gamma^i_{jk}(m) = 0$, the manifold M is flat at the point m. This means that the $(1,3)$ curvature tensor, defined locally at $m \in M$ as

$$R^l_{ijk} = \partial_{x^j} \Gamma^l_{ik} - \partial_{x^k} \Gamma^l_{ij} + \Gamma^l_{rj} \Gamma^r_{ik} - \Gamma^l_{rk} \Gamma^r_{ij},$$

also vanishes at that point, i.e., $R^l_{ijk}(m) = 0$.

Now, the rate of change of a vector–field A^k on the manifold M along the curve $x^i(s)$ is properly defined by the *absolute covariant derivative*

$$\frac{D}{ds} A^k = \dot{x}^i \nabla_i A^k = \dot{x}^i \left(\partial_{x^i} A^k + \Gamma^k_{ij} A^j\right) = \dot{A}^k + \Gamma^k_{ij} \dot{x}^i A^j.$$

By applying this result to itself, we can get an expression for the second covariant derivative of the vector–field A^k along the curve $x^i(s)$:

$$\frac{D^2}{ds^2} A^k = \frac{d}{ds}\left(\dot{A}^k + \Gamma^k_{ij} \dot{x}^i A^j\right) + \Gamma^k_{ij} \dot{x}^i (\dot{A}^j + \Gamma^j_{mn} \dot{x}^m A^n).$$

In the local coordinates $(x^1(s), ..., x^n(s))$ at a point $m \in M$, if $\delta x^i = \delta x^i(s)$ denotes the *geodesic deviation*, i.e., the infinitesimal vector describing perpendicular separation between the two neighboring geodesics, passing through two neighboring points $m, n \in M$, then the *Jacobi equation of geodesic deviation* on the manifold M holds:

$$\frac{D^2 \delta x^i}{ds^2} + R^i_{jkl} \dot{x}^j \delta x^k \dot{x}^l = 0. \tag{1.90}$$

This equation describes the *relative acceleration* between two infinitesimally close facial geodesics, which is proportional to the facial curvature (measured by the Riemann tensor R^i_{jkl} at a point $m \in M$), and to the geodesic deviation δx^i. Solutions of equation (1.90) are called *Jacobi fields*.

In particular, if the manifold M is a 2D–surface in $\mathbb{R}^3$, the Riemann curvature tensor simplifies into

$$R^i_{jmn} = \frac{1}{2} R\, g^{ik}(g_{km}\, g_{jn} - g_{kn}\, g_{jm}),$$

where R denotes the *scalar Gaussian curvature*. Consequently the equation of geodesic deviation (1.90) also simplifies into

$$\frac{D^2}{ds^2}\delta x^i + \frac{R}{2}\delta x^i - \frac{R}{2}\dot{x}^i(g_{jk}\,\dot{x}^j\,\delta x^k) = 0. \tag{1.91}$$

This simplifies even more if we work in a locally Cartesian coordinate system; in this case the covariant derivative $\frac{D^2}{Ds^2}$ reduces to an ordinary derivative $\frac{d^2}{ds^2}$ and the metric tensor g_{ij} reduces to identity matrix I_{ij}, so our 2D equation of geodesic deviation (1.91) reduces into a simple second–order ODE in just two coordinates $x^i\,(i = 1, 2)$

$$\ddot{x}^i + \frac{R}{2}\delta x^i - \frac{R}{2}\dot{x}^i(I_{jk}\,\dot{x}^j\,\delta x^k) = 0.$$

Global Riemannian Geometry

The Second Variation Formula

Cartan also establishes another important property of manifolds with nonpositive curvature. First he observes that all spaces of constant zero curvature have torsion–free fundamental groups. This is because any isometry of finite order on Euclidean space must have a fixed point (the center of mass of any orbit is necessarily a fixed point). Then he notices that one can geometrically describe the L^∞ center of mass of finitely many points $\{m_1, \ldots, m_k\}$ in Euclidean space as the unique minimum for the strictly convex function

$$x \to \max_{i=1,\cdots,k} \frac{1}{2}\left\{(d\,(m_i, x))^2\right\}.$$

In other words, the center of mass is the center of the ball of smallest radius containing $\{m_1, \ldots, m_k\}$. Now Cartan's observation from above was that the exponential map is expanding and globally distance nondecreasing as a map:

$$(T_m M,\ \text{Euclidean metric}) \to (T_m M,\ \text{with pull–back metric}).$$

Thus distance functions are convex in nonpositive curvature as well as in Euclidean space. Hence the above argument can in fact be used to conclude that any Riemannian manifold of nonpositive curvature must also have torsion free fundamental group.

Now, let us set up the *second variation formula* and explain how it is used. We have already seen the first variation formula and how it can be used to characterize geodesics. Now suppose that we have a unit speed geodesic $\gamma\,(t)$ parameterized on $[0, \ell]$ and consider a variation $V\,(s, t)$, where $V\,(0, t) = \gamma\,(t)$. Synge then shows that ($\ddot{L} \equiv \frac{d^2 L}{ds^2}$)

$$\ddot{L}(0) = \int_0^\ell \{g(\dot{X}, \dot{X}) - (g(\dot{X}, \dot{\gamma}))^2 - g(R(X, \dot{\gamma})X, \dot{\gamma})\}dt + g(\dot{\gamma}, A)|_0^\ell,$$

where $X(t) = \frac{\partial V}{\partial s}(0, t)$ is the variational vector–field, $\dot{X} = \nabla_{\dot{\gamma}} X$, and $A(t) = \nabla_{\frac{\partial V}{\partial s}} X$. In the special case where the variation fixes the endpoints, i.e., $s \to V(s, a)$ and $s \to V(s, b)$ are constant, the term with A in it falls out. We can also assume that the variation is perpendicular to the geodesic and then drop the term $g\left(\dot{X}, \dot{\gamma}\right)$. Thus, we arrive at the following simple form:

$$\ddot{L}(0) = \int_0^\ell \{g(\dot{X}, \dot{X}) - g\left(R\left(X, \dot{\gamma}\right) X, \dot{\gamma}\right)\}dt = \int_0^\ell \{|\dot{X}|^2 - \sec(\dot{\gamma}, X)\,|X|^2\}dt.$$

Therefore, if the sectional curvature is nonpositive, we immediately observe that any geodesic locally minimizes length (that is, among close–by curves), even if it does not minimize globally (for instance γ could be a closed geodesic). On the other hand, in positive curvature we can see that if a geodesic is too long, then it cannot minimize even locally. The motivation for this result comes from the unit sphere, where we can consider geodesics of length $> \pi$. Globally, we know that it would be shorter to go in the opposite direction. However, if we consider a variation of γ where the variational field looks like $X = \sin\left(t \cdot \frac{\pi}{\ell}\right) E$ and E is a unit length parallel field along γ which is also perpendicular to γ, then we get

$$\begin{aligned}
\ddot{L}(0) &= \int_0^\ell \left\{ \left|\dot{X}\right|^2 - \sec\left(\dot{\gamma}, X\right)|X|^2 \right\} dt \\
&= \int_0^\ell \left\{ \left(\frac{\pi}{\ell}\right)^2 \cdot \cos^2\left(t \cdot \frac{\pi}{\ell}\right) - \sec\left(\dot{\gamma}, X\right) \sin^2\left(t \cdot \frac{\pi}{\ell}\right) \right\} dt \\
&= \int_0^\ell \left(\left(\frac{\pi}{\ell}\right)^2 \cdot \cos^2\left(t \cdot \frac{\pi}{\ell}\right) - \sin^2\left(t \cdot \frac{\pi}{\ell}\right) \right) dt = -\frac{1}{2\ell}\left(\ell^2 - \pi^2\right),
\end{aligned}$$

which is negative if the length ℓ of the geodesic is greater than π. Therefore, the variation gives a family of curves that are both close to and shorter than γ. In the general case, we can then observe that if $\sec \geq 1$, then for the same type of variation we get

$$\ddot{L}(0) \leq -\frac{1}{2\ell}\left(\ell^2 - \pi^2\right).$$

Thus we can conclude that, if the space is complete, then the diameter must be $\leq \pi$ because in this case any two points are joined by a segment, which cannot minimize if it has length $> \pi$. With some minor modifications one can now conclude that any complete Riemannian manifold (M, g) with $\sec \geq k^2 > 0$ must satisfy $\operatorname{diam}(M, g) \leq \pi \cdot k^{-1}$. In particular, M must be compact. Since the universal covering of M satisfies the same curvature hypothesis, the conclusion must also hold for this space; hence M must have compact universal covering space and finite fundamental group.

In odd dimensions all spaces of constant positive curvature must be orientable, as orientation reversing orthogonal transformation on odd–dimensional spheres have fixed points. This can now be generalized to manifolds of varying positive curvature. Synge did it in the following way: Suppose M is not simply–connected (or not orientable), and use this to find a shortest closed geodesic in a free homotopy class of curves (that reverses orientation). Now consider parallel translation around this geodesic. As the tangent field to the geodesic is itself a parallel field, we see that parallel translation preserves the orthogonal complement to the geodesic. This complement is now odd dimensional (even dimensional), and by assumption parallel translation preserves (reverses) the orientation; thus it must have a fixed point. In other words, there must exist a closed parallel field X perpendicular to the closed geodesic γ. We can now use the above second variation formula

$$\ddot{L}(0) = \int_0^\ell \{|\dot{X}|^2 - |X|^2 \sec(\dot{\gamma}, X)\} dt + g(\dot{\gamma}, A)|_0^\ell = -\int_0^\ell |X|^2 \sec(\dot{\gamma}, X)\, dt.$$

Here the boundary term drops out because the variation closes up at the endpoints, and $\dot{X} = 0$ since we used a parallel field. In case the sectional curvature is always positive we then see that the above quantity is negative. But this means that the closed geodesic has nearby closed curves which are shorter. However, this is in contradiction with the fact that the geodesic was constructed as a length minimizing curve in a free homotopy class.

In 1941 Myers generalized the diameter bound to the situation where one only has a lower bound for the Ricci curvature. The idea is that $\mathrm{Ric}(\dot{\gamma}, \dot{\gamma}) = \sum_{i=1}^{n-1} \sec(E_i, \dot{\gamma})$ for any set of vector–fields E_i along γ such that $\dot{\gamma}, E_1, \ldots, E_{n-1}$ forms an orthonormal frame. Now assume that the fields are parallel and consider the $n-1$ variations coming from the variational vector–fields $\sin\left(t \cdot \frac{\pi}{\ell}\right) E_i$. Adding up the contributions from the variational formula applied to these fields then induces

$$\begin{aligned}\sum_{i=1}^{n-1} \ddot{L}(0) &= \sum_{i=1}^{n-1} \int_0^\ell \left\{ \left(\frac{\pi}{\ell}\right)^2 \cdot \cos^2\left(t \cdot \frac{\pi}{\ell}\right) - \sec(\dot{\gamma}, E_i) \sin^2\left(t \cdot \frac{\pi}{\ell}\right) \right\} dt \\ &= \int_0^\ell \left\{ (n-1) \left(\frac{\pi}{\ell}\right)^2 \cdot \cos^2\left(t \cdot \frac{\pi}{\ell}\right) - \mathrm{Ric}(\dot{\gamma}, \dot{\gamma}) \sin^2\left(t \cdot \frac{\pi}{\ell}\right) \right\} dt.\end{aligned}$$

Therefore, if $\mathrm{Ric}(\dot{\gamma}, \dot{\gamma}) \geq (n-1)\, k^2$ (this is the Ricci curvature of S_k^n), then

$$\begin{aligned}\sum_{i=1}^{n-1} \ddot{L}(0) &\leq (n-1) \int_0^\ell \left\{ \left(\frac{\pi}{\ell}\right)^2 \cdot \cos^2\left(t \cdot \frac{\pi}{\ell}\right) - k^2 \sin^2\left(t \cdot \frac{\pi}{\ell}\right) \right\} dt \\ &= -(n-1) \frac{1}{2\ell} \left(\ell^2 k^2 - \pi^2\right),\end{aligned}$$

which is negative when $\ell > \pi \cdot k^{-1}$ (the diameter of S_k^n). Thus at least one of the contributions $\frac{d^2 L_i}{ds^2}(0)$ must be negative as well, implying that the geodesic cannot be a segment in this situation.

Gauss–Bonnet Formula

In 1926 Hopf proved that in fact there is a *Gauss–Bonnet formula* for all even–dimensional hypersurfaces $H^{2n} \subset \mathbb{R}^{2n+1}$. The idea is that the determinant of the differential of the Gauss map $G : H^{2n} \to S^{2n}$ is the Gaussian curvature of the hypersurface. Moreover, this is an intrinsically computable quantity. If we integrate this over the hypersurface, we get,

$$\frac{1}{\operatorname{vol} S^{2n}} \int_H \det(DG) = \deg(G),$$

where $\deg(G)$ is the *Brouwer degree* of the *Gauss map*. Note that this can also be done for odd–dimensional surfaces, in particular curves, but in this case the degree of the Gauss map will depend on the embedding or immersion of the hypersurface. Instead one gets the so–called winding number. Hopf then showed, as Dyck had earlier done for surfaces, that $\deg(G)$ is always half the *Euler characteristic* of H, thus yielding

$$\frac{2}{\operatorname{vol} S^{2n}} \int_H \det(DG) = \chi(H). \tag{1.92}$$

Since the l.h.s of this formula is in fact intrinsic, it is natural to conjecture that such a formula should hold for all manifolds.

Ricci Flow on M

Ricci flow, or the *parabolic Einstein equation*, was introduced by R. Hamilton in 1982 [Ham82] in the form

$$\partial_t g_{ij} = -2R_{ij}. \tag{1.93}$$

Now, because of the minus sign in the front of the Ricci tensor R_{ij} in this equation, the solution metric g_{ij} to the Ricci flow shrinks in positive Ricci curvature direction while it expands in the negative Ricci curvature direction. For example, on the 2−sphere S^2, any metric of positive Gaussian curvature will shrink to a point in finite time. Since the Ricci flow (1.93) does not preserve volume in general, one often considers the *normalized* Ricci flow defined by

$$\partial_t g_{ij} = -2R_{ij} + \frac{2}{n} r g_{ij}, \tag{1.94}$$

where $r = \int R dV \big/ \int dV$ is the average scalar curvature. Under this normalized flow, which is equivalent to the (unnormalized) Ricci flow (1.93) by reparameterizing in time t and scaling the metric in space by a function of t, the volume of the solution metric is constant in time. Also that Einstein metrics (i.e., $R_{ij} = c g_{ij}$) are fixed points of (1.94).

Hamilton [Ham82] showed that on a closed Riemannian 3−manifold M^3 with initial metric of positive Ricci curvature, the solution $g(t)$ to the normalized Ricci flow (1.94) exists for all time and the metrics $g(t)$ converge

exponentially fast, as time t tends to the infinity, to a constant positive sectional curvature metric g_∞ on M^3.

Since the Ricci flow lies in the realm of parabolic partial differential equations, where the prototype is the heat equation, here is a brief review of the *heat equation* [CC99].

Let (M^n, g) be a Riemannian manifold. Given a C^2 function $u : M \to \mathbb{R}$, its Laplacian is defined in local coordinates $\{x^i\}$ to be

$$\Delta u = \mathrm{Tr}\left(\nabla^2 u\right) = g^{ij}\nabla_i \nabla_j u,$$

where $\nabla_i = \nabla_{\partial_{x^i}}$ is its associated covariant derivative (Levi–Civita connection). We say that a C^2 function $u : M^n \times [0,T) \to \mathbb{R}$, where $T \in (0,\infty]$, is a solution to the heat equation if

$$\partial_t u = \Delta u.$$

One of the most important properties satisfied by the heat equation is the maximum principle, which says that for any smooth solution to the heat equation, whatever pointwise bounds hold at $t = 0$ also hold for $t > 0$. Let $u : M^n \times [0,T) \to \mathbb{R}$ be a C^2 solution to the heat equation on a complete Riemannian manifold. If $C_1 \leq u(x,0) \leq C_2$ for all $x \in M$, for some constants $C_1, C_2 \in \mathbb{R}$, then $C_1 \leq u(x,t) \leq C_2$ for all $x \in M$ and $t \in [0,T)$ [CC99].

Now, given a smooth manifold M, a one–parameter family of metrics $g(t)$, where $t \in [0,T)$ for some $T > 0$, is a solution to the Ricci flow if (1.93) is valid at all $x \in M$ and $t \in [0,T)$. The minus sign in the equation (1.93) makes the Ricci flow a *forward* heat equation [CC99] (with the normalization factor 2).

In local geodesic coordinates $\{x^i\}$, we have [CC99]

$$g_{ij}(x) = \delta_{ij} - \frac{1}{3}R_{ipjq}x^p x^q + O\left(|x|^3\right), \qquad \text{therefore,} \qquad \Delta g_{ij}(0) = -\frac{1}{3}R_{ij},$$

where Δ is the standard Euclidean Laplacian. Hence the Ricci flow is like the heat equation for a Riemannian metric

$$\partial_t g_{ij} = 6\Delta g_{ij}.$$

The practical study of the Ricci flow is made possible by the following short–time existence result: Given any smooth compact Riemannian manifold (M, g_o), there exists a unique smooth solution $g(t)$ to the Ricci flow defined on some time interval $t \in [0,\epsilon)$ such that $g(0) = g_o$ [CC99].

Now, given that short–time existence holds for any smooth initial metric, one of the main problems concerning the Ricci flow is to determine under what conditions the solution to the normalized equation exists for all time and converges to a constant curvature metric. Results in this direction have been established under various curvature assumptions, most of them being some sort of positive curvature. Since the Ricci flow (1.93) does not preserve volume in general, one often considers, as we mentioned in the Introduction,

the normalized Ricci flow (1.94). Under this flow, the volume of the solution $g(t)$ is independent of time.

To study the long–time existence of the normalized Ricci flow, it is important to know what kind of curvature conditions are preserved under the equation. In general, the Ricci flow tends to preserve some kind of positivity of curvatures. For example, positive scalar curvature is preserved in all dimensions. This follows from applying the maximum principle to the evolution equation for scalar curvature R, which is

$$\partial_t R = \Delta R + 2\left|R_{ij}\right|^2.$$

In dimension 3, positive Ricci curvature is preserved under the Ricci flow. This is a special feature of dimension 3 and is related to the fact that the Riemann curvature tensor may be recovered algebraically from the Ricci tensor and the metric in dimension 3. Positivity of sectional curvature is not preserved in general. However, the stronger condition of positive curvature operator is preserved under the Ricci flow.

Structure Equations on M

Let $\{X_a\}_{a=1}^m$, $\{Y_i\}_{i=1}^n$ be local orthonormal framings on M, N respectively and $\{e_i\}_{i=1}^n$ be the induced framing on E defined by $e_i = Y_i \circ \phi$, then there exist smooth local coframings $\{\omega_a\}_{a=1}^m$, $\{\eta_i\}_{i=1}^n$ and $\{\phi^*\eta_i\}_{i=1}^n$ on TM, TN and E respectively such that (locally)

$$g = \sum_{a=1}^{m} \omega_a^2 \qquad \text{and} \qquad h = \sum_{i=1}^{n} \eta_i^2.$$

The corresponding *first structure equations* are [Mus99]:

$$\begin{aligned} d\omega_a &= \omega_b \wedge \omega_{ba}, & \omega_{ab} &= -\omega_{ba}, \\ d\eta_i &= \eta_j \wedge \eta_{ji}, & \eta_{ij} &= -\eta_{ji}, \\ d(\phi^*\eta_i) &= \phi^*\eta_j \wedge \phi^*\eta_{ji}, & \phi^*\eta_{ij} &= -\phi^*\eta_{ji}, \end{aligned}$$

where the unique 1–forms ω_{ab}, η_{ij}, $\phi^*\eta_{ij}$ are the respective connection forms. The *second structure equations* are

$$\begin{aligned} d\omega_{ab} &= \omega_{ac} \wedge \omega_{cb} + \Omega_{ab}^M, \qquad d\eta_{ij} = \eta_{ik} \wedge \eta_{kj} + \Omega_{ij}^N, \\ d(\phi^*\eta_{ij}) &= \phi^*\eta_{ik} \wedge \phi^*\eta_{kj} + \phi^*\Omega_{ij}^N, \end{aligned}$$

where the curvature 2–forms are given by

$$\Omega_{ab}^M = -\frac{1}{2} R_{abcd}^M \omega_c \wedge \omega_d \qquad \text{and} \qquad \Omega_{ij}^N = -\frac{1}{2} R_{ijkl}^N \eta_k \wedge \eta_l.$$

The pull back map ϕ^* and the push forward map ϕ_* can be written as [Mus99]

$$\phi^*\eta_i = f_{ia}\omega_a$$

for unique functions f_{ia} on $U \subset M$, so that

$$\phi_* = e_i \otimes \phi^*\eta_i = f_{ia} e_i \otimes \omega_a.$$

Note that ϕ_* is a section of the vector bundle $\phi^{-1}TN \otimes T^*M$.

The covariant differential operators are represented as

$$\nabla^M X_a = \omega_{ab} \otimes X_b, \qquad \nabla^N Y_i = \eta_{ij} \otimes Y_j, \qquad \nabla^* \omega_a = -\omega_{ca} \otimes \omega_c,$$

where ∇^* is the dual connection on the cotangent bundle T^*M.

Furthermore, the induced connection ∇^ϕ on E is

$$\nabla^\phi e_i = \big(\eta_{ij}(Y_k) \circ \phi\big)\, e_j \otimes f_{ka}\omega_a.$$

The components of the Ricci tensor and scalar curvature are defined respectively by

$$R^M_{ab} = R^M_{acbc} \qquad \text{and} \qquad R^M = R^M_{aa}.$$

Given a function $f : M \to$, there exist unique functions $f_{cb} = f_{bc}$ such that

$$df_c - f_b\omega_{cb} = f_{cb}\omega_b\,, \tag{1.95}$$

where $f_c = df(X_c)$ for a local orthonormal frame $\{X_c\}_{c=1}^m$. To prove this we take the exterior derivative of $df = \sum_{c=1}^m f_c\omega_c$ and using structure equations, we have

$$0 = [df_c \wedge \omega_c + f_{bc}\omega_b \wedge \omega_{bc}] = [(df_c - f_b\omega_{cb}) \wedge \omega_c]\,.$$

Hence by *Cartan's lemma* (cf. [Wil93]), there exist unique functions $f_{cb} = f_{bc}$ such that

$$df_c - f_b\omega_{cb} = f_{cb}\omega_b.$$

The Laplacian of a function f on M is given by

$$\Delta f = -\operatorname{Tr}(\nabla df),$$

that is, negative of the usual Laplacian on functions.

Basics of Morse and (Co)Bordism Theories

Morse Theory on Smooth Manifolds

At the same time the variational formulae were discovered, a related technique, called *Morse theory*, was introduced into Riemannian geometry. This theory was developed by Morse, first for functions on manifolds in 1925, and then in 1934, for the loop space. The latter theory, as we shall see, sets up a very nice connection between the first and second variation formulae from the previous section and the topology of M. It is this relationship that we shall explore at

a general level here. In section 5 we shall then see how this theory was applied in various specific settings.

If we have a *proper function* $f : M \to \mathbb{R}$, then its Hessian (as a quadratic form) is in fact well defined at its *critical points* without specifying an underlying Riemannian metric. The *nullity* of f at a *critical point* is defined as the dimension of the *kernel* of $\nabla^2 f$, while the *index* is the number of negative eigenvalues counted with multiplicity. A function is said to be a *Morse function* if the nullity at any of its critical points is zero. Note that this guarantees in particular that all critical points are isolated. The first fundamental theorem of Morse theory is that one can determine the topological structure of a manifold from a Morse function. More specifically, if one can order the critical points $x_1, \ldots, x_k$ so that $f(x_1) < \cdots < f(x_k)$ and the index of x_i is denoted λ_i, then M has the structure of a CW complex with a cell of dimension λ_i for each i. Note that in case M is closed then x_1 must be a minimum and so $\lambda_1 = 0$, while x_k is a maximum and $\lambda_k = n$. The classical example of Milnor of this theorem in action is a torus in 3–space and f the height function.

We are now left with the problem of trying to find appropriate Morse functions. While there are always plenty of such functions, there does not seem to be a natural way of finding one. However, there are natural choices for Morse functions on the loop space to a Riemannian manifold. This is, somewhat inconveniently, infinite–dimensional. Still, one can develop Morse theory as above for suitable functions, and moreover *the loop space of a manifold determines the topology of the underlying manifold*.

If $m, p \in M$, then we denote by Ω_{mp} the space of all C^∞ paths from m to p. The first observation about this space is that

$$\pi_{i+1}(M) = \pi_i(\Omega_{mp}).$$

To see this, just fix a path from m to q and then join this path to every curve in Ω_{mp}. In this way Ω_{mp} is identified with Ω_m, *the space of loops fixed at* m. For this space the above relationship between the homotopy groups is almost self-evident.

On the space Ω_{mp} we have two naturally defined functions, the *arc–length* and *energy functionals*:

$$L(\gamma, I) = \int_I |\dot{\gamma}|\, dt, \quad \text{and} \quad E(\gamma, I) = \frac{1}{2}\int_I |\dot{\gamma}|^2\, dt.$$

While the energy functional is easier to work with, it is the arc–length functional that we are really interested in. In order to make things work out nicely for the arc–length functional, it is convenient to parameterize all curves on $[0, 1]$ and proportionally to arc–length. We shall think of Ω_{mp} as an *infinite–dimensional manifold*. For each curve $\gamma \in \Omega_{mp}$ the natural choice for the tangent space consists of the vector–fields along γ which vanish at the endpoints of γ. This is because these vector–fields are exactly the variational fields for curves through γ in Ω_{mp}, i.e., fixed endpoint variations of γ. An inner product on the tangent space is then naturally defined by

$$(X,Y) = \int_0^1 g\,(X,Y)\;dt.$$

Now the first variation formula for arc–length tells us that the gradient for L at γ is $-\nabla_{\dot\gamma}\dot\gamma$. Actually this cannot be quite right, as $-\nabla_{\dot\gamma}\dot\gamma$ does not vanish at the endpoints. The real gradient is gotten in the same way we find the gradient for a function on a surface in space, namely, by projecting it down into the correct tangent space. In any case we note that the critical points for L are exactly the geodesics from m to p. The second variation formula tells us that the Hessian of L at these critical points is given by

$$\nabla^2 L\,(X) = \ddot X + R\,(X,\dot\gamma)\;\dot\gamma,$$

at least for vector–fields X which are perpendicular to γ. Again we ignore the fact that we have the same trouble with endpoint conditions as above. We now need to impose the Morse condition that this Hessian is not allowed to have any kernel. The vector–fields J for which $\ddot J + R\,(J,\dot\gamma)\,\dot\gamma = 0$ are called *Jacobi fields*. Thus we have to Figure out whether there are any Jacobi fields which vanish at the endpoints of γ. The first observation is that Jacobi fields must always come from geodesic variations. The Jacobi fields which vanish at m can therefore be found using the exponential map $\exp_m$. If the Jacobi field also has to vanish at p, then p must be a critical value for $\exp_m$. Now Sard's theorem asserts that the set of critical values has measure zero. For given $m \in M$ it will therefore be true that the arc–length functional on Ω_{mp} is a Morse function for almost all $p \in M$. Note that it may not be possible to choose $p = m$, the simplest example being the standard sphere. We are now left with trying to decide what the *index* should be. This is the dimension of the largest subspace on which the Hessian is negative definite. It turns out that this index can also be computed using Jacobi fields and is in fact always finite. Thus one can calculate the topology of Ω_{mp}, and hence M, by finding all the geodesics from m to p and then computing their index.

In geometrical situations it is often unrealistic to suppose that one can calculate the index precisely, but as we shall see it is often possible to given lower bounds for the index. As an example, note that if M is not simply–connected, then Ω_{mp} is not connected. Each curve of minimal length in the path components is a geodesic from m to p which is a local minimum for the arc–length functional. Such geodesics evidently have index zero. In particular, if one can show that all geodesics, except for the minimal ones from m to p, have index > 0, then the manifold must be simply–connected. We will apply Morse theory in biomechanics/robotic in subsection (2.4.4) below.

(Co)Bordism Theory on Smooth Manifolds

(Co)bordism appeared as a revival of Poincaré's unsuccessful 1895 attempts to define homology using only manifolds. Smooth manifolds (without boundary) are again considered as 'negligible' when they are *boundaries* of smooth

manifolds–with–boundary. But there is a big difference, which keeps definition of 'addition' of manifolds from running into the difficulties encountered by Poincaré; it is now the disjoint union. The (unoriented) *(co)bordism relation* between two compact smooth manifolds M_1, M_2 of same dimension n means that their disjoint union $\partial W = M_1 \cup M_2$ is the boundary ∂W of an $(n+1)$D smooth manifold–with–boundary W. This is an *equivalence relation*, and the classes for that relation of nD manifolds form a *commutative group* $\mathfrak{N}_n$ in which every element has order 2. The direct sum $\mathfrak{N}_\bullet = \oplus_{n\geq 0}\mathfrak{N}_n$ is a ring for the multiplication of classes deduced from the Cartesian product of manifolds.

More precisely, a manifold M is said to be a (co)bordism from A to B if exists a diffeomorphism from a disjoint sum, $\varphi \in \mathrm{diff}(A^* \cup B, \partial M)$. Two (co)bordisms $M(\varphi)$ and $M'(\varphi')$ are equivalent if there is a $\Phi \in \mathrm{diff}(M, M')$ such that $\varphi' = \Phi \circ \varphi$. The equivalence class of (co)bordisms is denoted by $M(A, B) \in Cob(A, B)$ [Sto68].

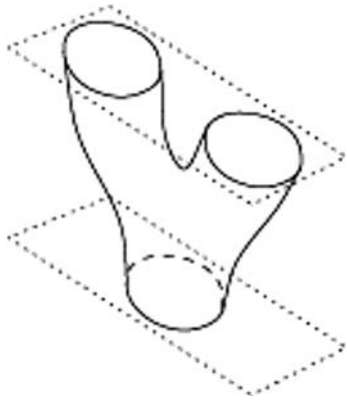
Composition c_{Cob} of (co)bordisms comes from gluing of manifolds [BD95]. Let $\varphi' \in \mathrm{diff}(C^* \cup D, \partial N)$. One can glue (co)bordism M with N by identifying B with C^*, $(\varphi')^{-1} \circ \varphi \in \mathrm{diff}(B, C^*)$. We get the glued (co)bordism $(M \circ N)(A, D)$ and a semigroup operation,

$$c(A, B, D) : Cob(A, B) \times Cob(B, D) \longrightarrow Cob(A, D).$$

A *surgery* is an operation of cutting a manifold M and gluing to cylinders. A surgery gives new (co)bordism: from $M(A, B)$ into $N(A, B)$. The disjoint sum of $M(A, B)$ with $N(C, D)$ is a (co)bordism $(M \cup N)(A \cup C, B \cup D)$. We got a 2–graph of (co)bordism Cob with $Cob_0 = Man_d$, $Cob_1 = Man_{d+1}$, whose 2–cells from Cob_2 are surgery operations.

There is an $n-$category of (co)bordisms $\mathcal{BO}$ [Lei03] with:

- 0−cells: 0−manifolds, where 'manifold' means 'compact, smooth, oriented manifold'. A typical 0−cell is • • • • .
- 1−cells: 1−manifolds with corners, i.e., (co)bordisms between 0−manifolds, such as  (this being a 1−cell from the 4−point manifold to the 2−point 0−manifold).



- 2−cells: 2−manifolds with corners, such as
- 3−cells, 4−cells,... are defined similarly;
- Composition is gluing of manifolds.

The (co)bordisms theme was taken a step further by [BD95], when when they started a programme to understand the subtle relations between certain TMFT models for manifolds of different dimensions, frequently referred to as the dimensional ladder. This programme is based on higher–dimensional algebra, a generalization of the theory of categories and functors to $n-$categories and $n-$functors. In this framework a topological quantum field theory (TMFT) becomes an $n-$functor from the $n-$category $\mathcal{BO}$ of $n-$cobordisms to the $n-$category of $n-$Hilbert spaces.

1.2.10 Finsler Manifolds

Recall that Finsler geometry is such a generalization of Riemannian geometry, that is closely related to multivariable calculus of variations.

Definition of a Finsler Manifold

Let M be a real, smooth, connected, finite–dimensional manifold. The pair (M, F) is called a *Finsler manifold* iff there exists a *fundamental function* $F : TM \to \mathbb{R}$ that satisfies the following set of axioms (see, e.g., [UN99]):

F1 $F(x, y) > 0$ for all $x \in M,\ y \neq 0$.
F2 $F(x, \lambda y) = |\lambda| F(x, y)$ for all $\lambda \in \mathbb{R}$, $(x, y) \in TM$.
F3 the *fundamental metric tensor* g_{ij} on M, given by

$$g_{ij}(x, y) = \frac{1}{2} \frac{\partial^2 F^2}{\partial y^i \partial y^j},$$

is *positive definite*.
F4 F is *smooth* (C^∞) at every point $(x, y) \in TM$ with $y \neq 0$ and *continuous* (C^0) at every $(x, 0) \in TM$. Then, the *absolute Finsler energy function* is given by

$$F^2(x, y) = g_{ij}(x, y) y^i y^j.$$

Let $c = c(t) : [a, b] \to M$ be a smooth regular curve on M. For any two vector–fields $X(t) = X^i(t)\, \frac{\partial}{\partial x^i}\big|_{c(t)}$ and $Y(t) = Y^i(t)\, \frac{\partial}{\partial x^i}\big|_{c(t)}$ along the curve $c = c(t)$, we introduce the *scalar (inner) product* [Che96]

$$g(X, Y)(c) = g_{ij}(c, \dot{c}) X^i Y^j$$

along the curve c.

In particular, if $X = Y$ then we have $\|X\| = \sqrt{g(X, X)}$. The vector–fields X and Y are *orthogonal* along the curve c, denoted by $X \perp Y$, iff $g(X, Y) = 0$.

Let $C\Gamma(N) = (L^i_{jk}, N^i_j, C^i_{jk})$ be the *Cartan canonical $N-$linear metric connection* determined by the metric tensor $g_{ij}(x, y)$. The coefficients of this connection are expressed by [UN99]

$$L^i_{jk} = \frac{1}{2} g^{im} \left(\frac{\delta g_{mk}}{\delta x^j} + \frac{\delta g_{jm}}{\delta x^k} - \frac{\delta g_{jk}}{\delta x^m} \right), \quad C^i_{jk} = \frac{1}{2} g^{im} \left(\frac{\partial g_{mk}}{\partial y^j} + \frac{\partial g_{jm}}{\partial y^k} - \frac{\partial g_{jk}}{\partial y^m} \right),$$

$$N^i_j = \frac{1}{2} \frac{\partial}{\partial y^j} \left(\Gamma^i_{kl} y^k y^l \right) = \frac{1}{2} \frac{\partial \Gamma^i_{00}}{\partial y^j}, \qquad \Gamma^i_{jk} = \frac{1}{2} g^{im} \left(\frac{\partial g_{mk}}{\partial x^j} + \frac{\partial g_{jm}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^m} \right),$$

$$\text{where} \quad \frac{\delta}{\delta x^i} = \frac{\partial}{\partial x^i} + N^j_i \frac{\partial}{\partial y^j}.$$

Let X be a vector–field along the curve c expressed locally by $X(t) = X^i(t)\, \frac{\partial}{\partial x^i}\big|_{c(t)}$. Using the Cartan $N-$linear connection, we define the *covariant derivative* $\frac{\nabla X}{dt}$ of $X(t)$ *along the curve* $c(t)$, by [UN99]

$$\frac{\nabla X}{dt} = \{ \dot{X}^i + X^m [L^i_{mk}(c, \dot{c}) \dot{c}^k + C^i_{mk}(c, \dot{c}) \frac{\delta}{\delta t} (\dot{c}^k)] \}\, \frac{\partial}{\partial x^i}\Big|_{c(t)}.$$

$$\text{Since} \quad \frac{\delta}{\delta t}(\dot{c}^k) = \ddot{c}^k + N^k_l(c, \dot{c}) \dot{c}^l,$$

$$\text{we have} \quad \frac{\nabla X}{dt} = \{ \dot{X}^i + X^m [\Gamma^i_{mk}(c, \dot{c}) \dot{c}^k + C^i_{mk}(c, \dot{c}) \ddot{c}^k] \}\, \frac{\partial}{\partial x^i}\Big|_{c(t)}, \quad (1.96)$$

$$\text{where} \quad \Gamma^i_{mk}(c, \dot{c}) = L^i_{mk}(c, \dot{c}) + C^i_{ml}(c, \dot{c}) N^l_k(c, \dot{c}).$$

In particular, c is a *geodesic* iff $\dfrac{\nabla \dot{c}}{dt} = 0$.

Since $C\Gamma(N)$ is a metric connection, we have

$$\frac{d}{dt} [g(X, Y)] = g \left(\frac{\nabla X}{dt}, Y \right) + g \left(X, \frac{\nabla Y}{dt} \right).$$

Energy Functional, Variations and Extrema

Let $x_0, x_1 \in M$ be two points not necessarily distinct. We introduce the $\Omega-$set on M, as

$$\Omega = \{c : [0,1] \to M \mid c \text{ is piecewise } C^\infty \text{ regular curve}, c(0) = x_0, c(1) = x_1\}.$$

For every $p \in \mathbb{R} - \{0\}$, we can define the *p−energy functional* on M [UN99]

$$E_p \ : \ \Omega \to \mathbb{R}_+, \qquad \text{as}$$
$$E_p(c) = \int_0^1 [g_{ij}(c,\dot{c})\dot{c}^i\dot{c}^j]^{p/2}dt = \int_0^1 [g(\dot{c},\dot{c})]^{p/2}dt = \int_0^1 \|\dot{c}\|^p dt.$$

In particular, for $p = 1$ we get the *length functional*

$$L(c) = \int_0^1 \|\dot{c}\|dt,$$

and for $p = 2$ we get the *energy functional*

$$E(c) = \int_0^1 \|\dot{c}\|^2 dt.$$

Also, for any naturally parametrized curve (i.e., $\|\dot{c}\| = \text{const}$) we have

$$E_p(c) = (L(c))^p = (E(c))^{p/2}.$$

Note that the $p-$energy of a curve is dependent of parametrization if $p \neq 1$.

For every curve $c \in \Omega$, we define the *tangent space* $T_c\Omega$ as

$T_c\Omega = \{X : [0,1] \to TM \mid X$ is continuous, piecewise $C^\infty, X(t) \in T_{c(t)}M$, for all $t \in [0,1],\ X(0) = X(1) = 0\}$.

Let $(c_s)_{s\in(-\epsilon,\epsilon)} \subset \Omega$ be a *one–parameter variation* of the curve $c \in \Omega$. We define

$$X(t) = \frac{dc_s}{ds}(0,t) \in T_c\Omega.$$

Using the equality

$$g\left(\frac{\nabla \dot{c}_s}{\partial s}, \dot{c}_s\right) = g\left(\frac{\nabla}{\partial t}\left(\frac{\partial c_s}{\partial s}\right), \dot{c}_s\right),$$

we can prove the following theorem: The first variation of the $p-$energy is

$$\frac{1}{p}\frac{dE_p(c_s)}{ds}(0) = -\sum_t g(X, \Delta_t(\|\dot{c}\|^{p-2}\dot{c}))$$
$$- \int_0^1 \|\dot{c}\|^{p-4} g\left(X, \|\dot{c}\|^2 \frac{\nabla \dot{c}}{dt} + (p-2)g\left(\frac{\nabla \dot{c}}{dt}, \dot{c}\right)\dot{c}\right) dt,$$

where $\Delta_t(\|\dot{c}\|^{p-2}\dot{c}) = (\|\dot{c}\|^{p-2}\dot{c})_{t^+} - (\|\dot{c}\|^{p-2}\dot{c})_{t^-}$ represents the jump of $\|\dot{c}\|^{p-2}\dot{c}$ at the discontinuity point $t \in (0,1)$ [UN99].

The curve c is a *critical point* of E_p iff c is a geodesic.

In particular, for $p = 1$ the curve c is a *reparametrized geodesic*.

Now, let $c \in \Omega$ be a critical point for E_p (i.e., the curve c is a geodesic). Let $(c_{s_1 s_2})_{s_1,s_2\in(-\epsilon,\epsilon)} \subset \Omega$ be a *two–parameter variation* of c. Using the notations:

$$X(t) = \frac{\partial c_{s_1 s_2}}{\partial s_1}(0,0,t) \in T_c\Omega, \qquad Y(t) = \frac{\partial c_{s_1 s_2}}{\partial s_2}(0,0,t) \in T_c\Omega,$$

$$\|\dot{c}\| = v = \text{constant}, \qquad \text{and} \qquad I_p(X,Y) = \frac{\partial^2 E_p(c_{s_1 s_2})}{\partial s_1 \partial s_2}(0,0),$$

we get the following theorem: The second variation of the $p-$energy is [UN99]

$$\frac{1}{pv^{p-4}} I_p(X,Y) = -\sum_t g\left(Y, v^2 \Delta_t\left(\frac{\nabla X}{dt}\right) + (p-2) g\left(\Delta_t\left(\frac{\nabla X}{dt}\right), \dot{c}\right)\dot{c}\right)$$

$$-\int_0^1 g\left(Y, v^2\left[\frac{\nabla}{dt}\frac{\nabla X}{dt} + R^2(X,\dot{c})\dot{c}\right] + (p-2) g\left(\left[\frac{\nabla}{dt}\frac{\nabla X}{dt} + R^2(X,\dot{c})\dot{c}\right], \dot{c}\right)\dot{c}\right) dt,$$

where $\Delta_t\left(\frac{\nabla X}{dt}\right) = \left(\frac{\nabla X}{dt}\right)_{t^+} - \left(\frac{\nabla X}{dt}\right)_{t^-}$ represents the jump of $\frac{\nabla X}{dt}$ at the discontinuity point $t \in (0,1)$; also, if $R^l_{ijk}(c,\dot{c})$ represents the components of the *Finsler curvature tensor*, then

$$R^2(X,\dot{c})\dot{c} = R^l_{ijk}(c,\dot{c})\dot{c}^i\dot{c}^j X^k \frac{\partial}{\partial x^l} = R^l_{jk}(c,\dot{c})\dot{c}^j X^k \frac{\partial}{\partial x^l}.$$

In particular, we have

$$R^i_{jk} = \frac{\delta N^i_j}{\delta x^k} - \frac{\delta N^i_k}{\delta x^j}, \quad \text{and} \quad R^i_{hjk} = \frac{\delta L^i_{hj}}{\delta x^k} - \frac{\delta L^i_{hk}}{\delta x^j} + L^s_{hj}L^i_{sk} - L^s_{hk}L^i_{sj} + C^i_{hs}R^s_{jk}.$$

Moreover, using the *Ricci identities* for the *deflection tensors*, we also have

$$R^i_{jk} = R^i_{mjk}y^m = R^i_{0jk}.$$

$I_p(X,Y) = 0$ (for all $Y \in T_c\Omega$) iff X is a *Jacobi field*, i.e.,

$$\frac{\nabla}{dt}\frac{\nabla X}{dt} + R^2(X,\dot{c})\dot{c} = 0.$$

In these conditions we have the following definition: A point $c(b)$ $(0 \le a < b < 1)$ of a geodesic $c \in \Omega$ is called a *conjugate point* of a point $c(a)$ along the curve $c(t)$, if there exists a non–zero Jacobi field which vanishes at $t \in \{a,\ b\}$.

Now, integrating by parts and using the property of metric connection, we find

$$\frac{1}{pv^{p-4}} I_p(X,Y) = \int_0^1 v^2\left[g\left(\frac{\nabla X}{dt}, \frac{\nabla Y}{dt}\right) - R^2(X,\dot{c},Y,\dot{c})\right]$$

$$+\ (p-2) g\left(\dot{c}, \frac{\nabla X}{dt}\right) g\left(\dot{c}, \frac{\nabla Y}{dt}\right) dt,$$

where $R^2(X,\dot{c},Y,\dot{c}) = g(R^2(Y,\dot{c})\dot{c},X) = R_{0i0j}(c(t),\dot{c}(t))X^iY^j$.

Let $R_{ijk} = g_{jm}R^m_{ik}$. In any Finsler space the following identity is satisfied,

$$R_{ijk} + R_{jki} + R_{kij} = 0,$$

get by the *Bianchi identities*. As $R_{0i0j} = R_{i0j} = R_{j0i} = R_{0j0i}$ we get $R^2(X,\dot{c},Y,\dot{c}) = R^2(Y,\dot{c},X,\dot{c})$.

The quadratic form associated to the *Hessian of the p−energy* is given by

$$I_p(X) = I_p(X,X) = \int_0^1 v^2 \left[\left\|\frac{\nabla X}{dt}\right\|^2 - R^2(X,\dot{c},X,\dot{c})\right] + (p-2)\left[g\left(\dot{c},\frac{\nabla X}{dt}\right)\right]^2 dt.$$

Let

$$T_c^{\perp}\Omega = \{X \in T_c\Omega \mid g(X,\dot{c}) = 0\}, \quad \text{and}$$
$$T_c^{'}\Omega = \{X \in T_c\Omega \mid X = f\dot{c}, \text{ where } f\colon [0,1] \to R \text{ is continuous, piecewise } C^\infty,\ f(0) = f(1) = 0\}.$$

Let c be a geodesic and $p \in R - \{0,1\}$. Then $I_p(T_c'\Omega) \geq 0$ for $p \in (-\infty,0) \cup (1,\infty)$, and $I_p(T_c'\Omega) \leq 0$ for $p \in (0,1)$. Moreover, in both cases: $I_p(X) = 0$ iff $X = 0$. To prove it, let $X = f\dot{c} \in T_c'\Omega$. Then we have [UN99]

$$\begin{aligned}\frac{1}{v^{p-4}} I_p(X) &= p\int_0^1 \left\{v^2\left[g(f'\dot{c},f'\dot{c}) - R^2(f\dot{c},\dot{c},f\dot{c},\dot{c})\right] + (p-2)\left[g(\dot{c},f'\dot{c})\right]^2\right\} dt \\ &= p\int_0^1 \left[v^4(f')^2 + (p-2)v^4(f')^2\right] dt = \int_0^1 p(p-1)v^4(f')^2 dt.\end{aligned}$$

Moreover, if $I_p(X) = 0$, then $f' = 0$, which means that f is constant. The conditions $f(0) = f(1) = 0$ imply that $f = 0$.

As $I_p(T_c'\Omega)$ is positive definite for $p \in (-\infty,0) \cup (1,\infty)$ and negative definite for $p \in (0,1)$, it is sufficient to study the behavior of I_p restricted to $T_c^{\perp}\Omega$. Since $X \perp \dot{c}$ and the curve c is a geodesic it follows

$$g\left(\dot{c},\frac{\nabla X}{dt}\right) = 0.$$

Hence, for all $X \in T_c^{\perp}\Omega$, we have

$$\frac{1}{pv^{p-2}} I_p(X) = \int_0^1 \left[\left\|\frac{\nabla X}{dt}\right\|^2 - R^2(X,\dot{c},X,\dot{c})\right] dt = I(X).$$

Constant Curvature Finsler Manifolds

We assume the Finsler space (M,F) is complete, of dimension $n \geq 3$ and of constant curvature $K \in R$. Hence, we have

$$H_{ijkl} = K(g_{ik}g_{jl} - g_{il}g_{jk}),$$

where H_{ijkl} are the components of the *h−curvature tensor* H of the *Berwald connection* $B\Gamma$. It follows that

$$R_{ijk} = KF\left(g_{ik}\frac{y_j}{F} - g_{ij}\frac{y_k}{F}\right),$$

where $y_j = g_{jk}y^k$. We also have

$$R_{i0k} = R_{ijk}y^j = K(g_{ik}F^2 - y_iy_k).$$

Hence, along the geodesic $c \in \Omega$, we get

$$R^2(X,\dot{c})\dot{c} = K\{\|\dot{c}\|^2X - g(X,\dot{c})\dot{c}\}.$$

This equality is also true in the case of constant $h-$curvature for the Cartan canonical connection. Following Matsumoto [Mat82] we have:
(i) If $K \leq 0$, then the geodesic c has no conjugate points to $x_0 = c(0)$.
(ii) If $K \geq 0$ and the geodesic c has conjugate points to $x_0 = c(0)$, then the number of conjugate points is finite, according to the *Morse index theorem for Finsler manifolds*. Moreover, in the case (ii), choosing an orthonormal frame of vector–fields $\{E_i\}_{i=\overline{1,n-1}} \in T_c^{\perp}\Omega$ parallel–propagated along the geodesic c, we can build a basis $\{U_i, V_i\}_{i=\overline{1,n-1}}$ in the set of Jacobi fields orthogonal to $\dot{c}$, defining

$$U_i(t) = \sin(\sqrt{K}vt)E_i, \quad \text{and} \quad V_i(t) = \cos(\sqrt{K}vt)E_i,$$

where $v = \|\dot{c}\| = \text{const}$. In conclusion, the distance between two consecutive conjugate points is $\pi/\sqrt{K}$. In these conditions we can prove the following theorem [UN99]: Let (M,F) be a Finsler space, as above, and let $c = c_p \in \Omega$ be a global extremum point for the $p-$energy functional E_p, where p is a number in $R - \{0,1\}$. In these conditions we have:
(i) If $p \in (-\infty, 0)$, then c has conjugate points, $K > 0$ and

$$\left[\frac{(m(c)+1)\pi}{\sqrt{K}}\right]^p \leq E_p(c) \leq \left[\frac{m(c)\pi}{\sqrt{K}}\right]^p,$$

where $m(c)$ is the maximal number of conjugate points to $x_0 = c(0)$ along the geodesic c.
(ii) If $p \in (0,1)$, then c has conjugate points, $K > 0$ and

$$\left[\frac{m(c)\pi}{\sqrt{K}}\right]^p \leq E_p(c) \leq \left[\frac{(m(c)+1)\pi}{\sqrt{K}}\right]^p.$$

(iii) If $p \in (1, \infty)$, then c is a minimal geodesic (i.e., it minimizes the length functional). If we denote $m = \sup\{m(c) \mid c \in \Omega,\ c-\text{geodesic}\} \in N$, we get the following corollary: If there is $c \in \Omega$ a global extremum point for the $p-$energy functional E_p, where $p \in (-\infty, 0) \cup (0, 1)$, we must have $m < \infty$ and $m(c) = m$.

For example, in the case of Riemannian unit sphere $S^n \subset R^{n+1},\ n \geq 2$, it is well known that the geodesics are precisely the great circles, that is the intersections of S^n with the hyperplanes trough the center of S^n. Moreover, two arbitrary points on S^n are conjugate along a geodesic γ if they are antipodal points. In these conditions, for any two points x_0 and x_1 on the sphere S^n, there is no geodesic trough these points which has a finite maximal number of conjugate points, because we can surround the sphere infinite times. Hence, for the unit sphere S^n, we have $m = \infty$. In conclusion, in the case $p \in (-\infty, 0) \cup (0, 1)$, the $p-$energy functional on the sphere has no global extremum points [UN99].

1.2.11 Symplectic Manifolds

Symplectic Algebra

Symplectic algebra works in the category of symplectic vector spaces V_i and linear symplectic mappings $t \in L(V_i, V_j)$ [Put93].

Let V be a nD real vector space and $L^2(V, \mathbb{R})$ the space of all bilinear maps from $V \times V$ to $\mathbb{R}$. We say that a *bilinear map* $\omega \in L^2(V, \mathbb{R})$ is *nondegenerate*, i.e., if $\omega(v_1, v_2) = 0$ for all $v_2 \in V$ implies $v_1 = 0$.

If $\{e_1, ..., e_n\}$ is a basis of V and $\{e^1, ..., e^n\}$ is the dual basis, $\omega_{ij} = \omega(e_i, e_j)$ is the matrix of ω. A bilinear map $\omega \in L^2(V, \mathbb{R})$ is nondegenerate iff its matrix ω_{ij} is nonsingular. The transpose ω^t of ω is defined by $\omega^t(e_i, e_j) = \omega(e_j, e_i)$. ω is *symmetric* if $\omega^t = \omega$, and *skew–symmetric* if $\omega^t = -\omega$.

Let $A^2(V)$ denote the space of skew–symmetric bilinear maps on V. An element $\omega \in A^2(V)$ is called a $2-$form on V. If $\omega \in A^2(V)$ is nondegenerate then in the basis $\{e_1, ..., e_n\}$ its matrix $\omega(e_i, e_j)$ has the form $J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$.

A *symplectic form* on a real vector space V of dimension $2n$ is a nondegenerate $2-$form $\omega \in A^2(V)$. The pair (V, ω) is called a symplectic vector space. If (V_1, ω_1) and (V_2, ω_2) are symplectic vector spaces, a linear map $t \in L(V_1, V_2)$ is a *symplectomorphism* (i.e., a symplectic mapping) iff $t^*\omega_2 = \omega_1$. If (V, ω) is a symplectic vector space, we have an orientation Ω_ω on V given by

$$\Omega_\omega = \frac{(-1)^{\frac{n(n-1)}{2}}}{n!}\omega^n.$$

Let (V, ω) be a $2n$D symplectic vector space and $t \in L(V, V)$ a symplectomorphism. Then t is volume preserving, i.e., $t^*(\Omega_\omega) = \Omega_\omega$, and $\det_{\Omega_\omega}(t) = 1$.

The set of all symplectomorphisms $t : V \to V$ of a $2n$D symplectic vector space (V, ω) forms a group under composition, called the *symplectic group*, denoted by $Sp(V, \omega)$.

In matrix notation, there is a basis of V in which the matrix of ω is $J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$, such that $J^{-1} = J^t = -J$, and $J^2 = -I$. For $t \in L(V, V)$ with matrix $T = [T^i_j]$ relative to this basis, the condition $t \in Sp(V, \omega)$, i.e., $t^*\omega = \omega$, becomes

$$T^t J T = J.$$

In general, by definition a matrix $A \in M_{2n \times 2n}(\mathbb{R})$ is symplectic iff $A^t J A = J$.

Let (V, ω) be a symplectic vector space, $t \in Sp(V, \omega)$ and $\lambda \in \mathbb{C}$ an eigenvalue of t. Then λ^{-1}, $\bar{\lambda}$ and $\bar{\lambda}^{-1}$ are eigenvalues of t.

Symplectic Geometry

Symplectic geometry is a globalization of symplectic algebra [Put93]; it works in the category **Symplec** of symplectic manifolds M and symplectic diffeomorphisms f. The phase–space of a conservative dynamical system is a symplectic manifold, and its time evolution is a one–parameter family of symplectic diffeomorphisms.

A *symplectic form* or a symplectic structure on a smooth (*i.e.*, C^∞) manifold M is a nondegenerate closed $2-$form ω on M, i.e., for each $x \in M$ $\omega(x)$ is nondegenerate, and $d\omega = 0$. A *symplectic manifold* is a pair (M, ω) where M is a smooth $2n$D manifold and ω is a symplectic form on it. If (M_1, ω_1) and (M_2, ω_2) are symplectic manifolds then a smooth map $f : M_1 \to M_2$ is called *symplectic map* or *canonical transformation* if $f^*\omega_2 = \omega_1$.

For example, any symplectic vector space (V, ω) is also a symplectic manifold; the requirement $d\omega = 0$ is automatically satisfied since ω is a constant map. Also, any orientable, compact surface Σ is a symplectic manifold; any nonvanishing $2-$form (volume element) ω on Σ is a symplectic form on Σ.

If (M, ω) is a symplectic manifold then it is orientable with the standard volume form

$$\Omega_\omega = \frac{(-1)^{\frac{n(n-1)}{2}}}{n!}\omega^n,$$

If $f : M \to M$ is a symplectic map, then f is volume preserving, $\det_{\Omega_\omega}(f) = 1$ and f is a local diffeomorphism.

In general, if (M, ω) is a $2n$D compact symplectic manifold then ω^n is a volume element on M, so the *de Rham cohomology class* $[\omega^n] \in H^{2n}(M, \mathbb{R})$ is nonzero. Since $[\omega^n] = [\omega]^n$, $[\omega] \in H^2(M, \mathbb{R})$ and all of its powers through the nth must be nonzero as well. The existence of such an element of $H^2(M, \mathbb{R})$ is a *necessary condition* for the compact manifold to admit a symplectic structure.

However, if M is a $2n$D compact manifold without boundary, then there does not exist any exact symplectic structure, $\omega = d\theta$ on M, as its total volume is zero (by Stokes' theorem),

$$\int_M \Omega_\omega = \frac{(-1)^{\frac{n(n-1)}{2}}}{n!} \int_M \omega^n = \frac{(-1)^{\frac{n(n-1)}{2}}}{n!} \int_M d(\theta \wedge \omega^{n-1}) = 0.$$

For example, spheres S^{2n} do not admit a symplectic structure for $n \geq 2$, since the second de Rham group vanishes, i.e., $H^2(S^{2n}, \mathbb{R}) = 0$. This argument applies to any compact manifold without boundary and having $H^2(M, \mathbb{R}) = 0$.

In mechanics, the phase–space is the cotangent bundle T^*M of a configuration space M. There is a natural symplectic structure on T^*M that is usually defined as follows. Let M be a smooth nD manifold and pick local coordinates $\{dq^1, ..., dq^n\}$. Then $\{dq^1, ..., dq^n\}$ defines a basis of the tangent space T^*_qM, and by writing $\theta \in T^*_qM$ as $\theta = p_i dq^i$ we get local coordinates $\{q^1, ..., q^n, p_1, ..., p_n\}$ on T^*M. Define the canonical symplectic form ω on T^*M by

$$\omega = dp_i \wedge dq^i.$$

This $2-$form ω is obviously independent of the choice of coordinates $\{q^1, ..., q^n\}$ and independent of the base point $\{q^1, ..., q^n, p_1, ..., p_n\} \in T^*_qM$; therefore, it is locally constant, and so $d\omega = 0$.

The canonical $1-$form θ on T^*M is the unique $1-$form with the property that, for any $1-$form β which is a section of T^*M we have $\beta^*\theta = \theta$.

Let $f : M \to M$ be a diffeomorphism. Then T^*f preserves the canonical $1-$form θ on T^*M, i.e., $(T^*f)^*\theta = \theta$. Thus T^*f is symplectic diffeomorphism.

If (M, ω) is a $2n$D symplectic manifold then about each point $x \in M$ there are local coordinates $\{q^1, ..., q^n, p_1, ..., p_n\}$ such that $\omega = dp_i \wedge dq^i$. These coordinates are called canonical or symplectic. By the Darboux theorem, ω is constant in this local chart, i.e., $d\omega = 0$.

Momentum Map and Symplectic Reduction

Let (M, ω) be a connected symplectic manifold and $\phi : G \times M \to M$ a symplectic action of the Lie group G on M, that is, for each $g \in G$ the map $\phi_g : M \to M$ is a symplectic diffeomorphism. If for each $\xi \in \mathfrak{g}$ there exists a globally defined function $\hat{J}(\xi) : M \to \mathbb{R}$ such that $\xi_M = X_{\hat{J}(\xi)}$, then the map $J : M \to \mathfrak{g}^*$, given by

$$J : x \in M \mapsto J(x) \in \mathfrak{g}^*, \qquad J(x)(\xi) = \hat{J}(\xi)(x)$$

is called the *momentum map* for ϕ [MR99, Put93].

Since ϕ is symplectic, $\phi_{\exp(t\xi)}$ is a one parameter family of canonical transformations, i.e., $\phi^*_{\exp(t\xi)}\omega = \omega$, hence ξ_M is locally Hamiltonian and not generally Hamiltonian. That is why not every symplectic action has a momentum map. $\phi : G \times M \to M$ is Hamiltonian iff $\hat{J} : \mathfrak{g} \to C^\infty(M, \mathbb{R})$ is a Lie algebra homomorphism.

Let $H : M \to \mathbb{R}$ be $G-$-invariant, that is $H\left(\phi_g(x)\right) = H(x)$ for all $x \in M$ and $g \in G$. Then $\hat{J}(\xi)$ is a constant of motion for dynamics generated by H.

Let ϕ be a symplectic action of G on (M,ω) with the momentum map J. Suppose $H : M \to \mathbb{R}$ is $G-$-invariant under this action. Then the Noether's theorem states that J is a constant of motion of H, i.e., $J \circ \phi_t = J$, where ϕ_t is the flow of X_H.

A *Hamiltonian action* is a symplectic action with an Ad^*–equivariant momentum map J, i.e.,

$$J\left(\phi_g(x)\right) = Ad^*_{g^{-1}}\left(J(x)\right),$$

for all $x \in M$ and $g \in G$.

Let ϕ be a symplectic action of a Lie group G on (M,ω). Assume that the symplectic form ω on M is exact, i.e., $\omega = d\theta$, and that the action ϕ of G on M leaves the one form $\theta \in M$ invariant. Then $J : M \to \mathfrak{g}^*$ given by $\left(J(x)\right)(\xi) = \left(i_{\xi_M}\theta\right)(x)$ is an Ad^*–equivariant momentum map of the action.

In particular, in the case of the cotangent bundle $(M = T^*M,\, \omega = d\theta)$ of a mechanical configuration manifold M, we can *lift* up an action ϕ of a Lie group G on M to get an action of G on T^*M. To perform this lift, let G act on M by transformations $\phi_g : M \to M$ and define the *lifted action* to the cotangent bundle by $(\phi_g)_* : T^*M \to T^*M$ by pushing forward one forms, $(\phi_g)_*(\alpha)\cdot v = \alpha\left(T\phi_g^{-1}v\right)$,where $\alpha \in T^*_qM$ and $v \in T_{\phi_g(q)}M$. The lifted action $(\phi_g)_*$ preserves the canonical one form θ on T^*M and the momentum map for $(\phi_g)_*$ is given by

$$J : T^*M \to \mathfrak{g}^*, \qquad J\left(\alpha_q\right)(\xi) = \alpha_q\left(\xi_M(q)\right).$$

For example, let $M = \mathbb{R}^n$, $G = \mathbb{R}^n$ and let G act on $\mathbb{R}^n$ by translations:

$$\phi : (t,q) \in \mathbb{R}^n \times \mathbb{R}^n \mapsto t + q \in \mathbb{R}^n.$$

Then $\mathfrak{g} = \mathbb{R}^n$ and for each $\xi \in \mathfrak{g}$ we have $\xi_{\mathbb{R}^n}(q) = \xi$.

In case of the group of rotations in $\mathbb{R}^3$, $M = \mathbb{R}^3$, $G = SO(3)$ and let G act on $\mathbb{R}^3$ by $\phi(A,q) = A \cdot q$. Then $\mathfrak{g} \simeq \mathbb{R}^3$ and for each $\xi \in \mathfrak{g}$ we have $\xi_{\mathbb{R}^3}(q) = \xi \times q$.

Let G act transitively on (M,ω) by a Hamiltonian action. Then $J(M) = \left\{Ad^*_{g^{-1}}\left(J(x)\right) | g \in G\right\}$ is a coadjoint orbit.

Now, let (M,ω) be a symplectic manifold, G a Lie group and $\phi : G \times M \to M$ a Hamiltonian action of G on M with Ad^*–equivariant momentum map $J : M \to \mathfrak{g}^*$. Let $\mu \in \mathfrak{g}^*$ be a regular value of J; then $J^{-1}(\mu)$ is a submanifold of M such that $\dim\left(J^{-1}(\mu)\right) = \dim(M) - \dim(G)$. Let $G_\mu = \{g \in G | Ad^*_g\mu = \mu\}$ be the isotropy subgroup of μ for the coadjoint action. By Ad^*–equivariance, if $x \in J^{-1}(\mu)$ then $\phi_g(x) = J^{-1}(\mu)$ for all $g \in G$, i.e., $J^{-1}(\mu)$ is invariant under the induced G_μ–action and we can form the quotient space $M_\mu = J^{-1}(\mu)/G_\mu$, called the *reduced phase–space* at $\mu \in \mathfrak{g}^*$.

Let (M,ω) be a symplectic $2n$D manifold and let $f_1, ..., f_k$ be k functions in involution, i.e., $\{f_i, f_j\}_\omega = 0$, $i = 1, ..., k$. Because the flow of X_{f_i} and X_{f_j} commute, we can use them to define a symplectic action of $G = \mathbb{R}^k$ on M. Here $\mu \in \mathbb{R}^k$ is in the range space of $f_1 \times ... \times f_k$ and $J = f_1 \times ... \times f_k$ is the

momentum map of this action. Assume that $\{df_1, ..., df_k\}$ are independent at each point, so μ is a regular value for J. Since G is Abelian, $G_\mu = G$ so we get a symplectic manifold $J^{-1}(\mu)/G$ of dimension $2n - 2k$. If $k = n$ we have *integrable systems*.

For example, let $G = SO(3)$ and $(M, \omega) = \left(\mathbb{R}^6, \sum_{i=1}^3 dp_i \wedge dq^i\right)$, and the action of G on $\mathbb{R}^6$ is given by $\phi : (R, (q, p)) \mapsto (R_q, R_p)$. Then the momentum map is the well known angular momentum and for each $\mu \in \mathfrak{g}^* \simeq \mathbb{R}^3 \mu \neq 0$, $G_\mu \simeq S^1$ and the reduced phase–space (M_μ, ω_μ) is $(T^*\mathbb{R},\ \omega = dp_i \wedge dq^i)$, so that $\dim(M_\mu) = \dim(M) - \dim(G) - \dim(G_\mu)$. This reduction is in celestial mechanics called by Jacobi 'the elimination of the nodes'.

The equations of motion: $\dot{f} = \{f, H\}_\omega$ on M reduce to the equations of motion: $\dot{f}_\mu = \{f_\mu, H_\mu\}_{\omega_\mu}$ on M_μ (see [MR99]).

Multisymplectic Geometry

Multisymplectic geometry constitutes the general framework for a geometrical, covariant formulation of classical field theory. Here, covariant formulation means that space–like and time–like directions on a given space–time be treated on equal footing. With this principle, one can construct a covariant form of the Legendre transformation which associates to every field variable as many conjugated momenta, the multimomenta, as there are space–time dimensions. These, together with the field variables, those of nD space–time, and an extra variable, the energy variable, span the multiphase–space [KS76]. For a recent exposition on the differential geometry of this construction, see [Got91a]. Multiphase–space, together with a closed, nondegenerate differential $(n + 1)-$form, the multisymplectic form, is an example of a multisymplectic manifold. Among the achievements of the multisymplectic approach is a geometrical formulation of the relation of infinitesimal symmetries and covariantly conserved quantities (Noether's theorem), see [LMD04] for a recent review, and [GM92, FR04] for a clarification of the improvement techniques ('Belinfante–Rosenfeld formula') of the energy–momentum tensor in classical field theory. Multisymplectic geometry also gives convenient sets of variational integrators for the numerical study of partial differential equations [MPS98].

Since in multisymplectic geometry, the symplectic two–form of classical mechanics is replaced by a closed differential form of higher tensor degree, multivector–fields and differential forms have their natural appearance. (See [PR02] for an interpretation of multivector–fields as describing solutions to field equations infinitesimally.) Multivector–fields form a graded Lie algebra with the Schouten bracket (see [Kos04] for a review and unified viewpoint). Using the multisymplectic $(n+1)-$form, one can construct a new bracket for the differential forms, the Poisson forms [FPR04], generalizing a well–known formula for the *Poisson brackets* related to a symplectic two–form. A remarkable fact is that in order to establish a Jacobi identity, the multisymplectic form has to have a potential, a condition that is not seen in symplectic geometry. Further, the admissible differential forms, the Poisson forms, are subject

to the rather strong restrictions on their dependence on the multimomentum variables [Got91b]. In particular, $(n-1)-$forms in this context can be shown to arise exactly from the covariantly conserved currents of Noether symmetries, which allows their pairing with space–like hypersurfaces to yield conserved charges in a geometrical way.

The Hamiltonian, infinite dimensional formulation of classical field theory requires the choice of a space–like hypersurface ('Cauchy surface'), which manifestly breaks the general covariance of the theory at hand. For $(n-1)-$forms, the above new bracket reduces to the Peierls–deWitt bracket after integration over the space–like hypersurface [GN80]. With the choice of a hypersurface, a constraint analysis a'la Dirac [HT92, GIM04] can be performed [Lan95]. Again, the necessary breaking of general covariance raises the need for an alternative formulation of all this [MW74]; first attempts have been made to carry out a Marsden–Weinstein reduction [MRS04] for multisymplectic manifolds with symmetries. However, not very much is known about how to quantize a multisymplectic geometry, see [BS04] for an approach using a path integral.

So far, everything was valid for field theories of first–order, i.e., where the Lagrangian depends on the fields and their first derivatives. Higher order theories can be reduced to first–order ones for the price of introducing auxiliary fields. A direct treatment would involve higher order jet bundles [Sau89]. A definition of the covariant Legendre transform and the multiphase–space has been given for this case [Got91a].

1.2.12 Complex and Kähler Manifolds

Just as a smooth manifold has enough structure to define the notion of differentiable functions, a *complex manifold* is one with enough structure to define the notion of holomorphic (or, analytic) functions $f : X \to \mathbb{C}$. Namely, if we demand that the transition functions $\phi_j \circ \phi_i^{-1}$ in the charts U_i on M (see Figure 1.5) satisfy the *Cauchy–Riemann equations*

$$\partial_x u = \partial_y v, \qquad \partial_y u = -\partial_x v,$$

then the analytic properties of f can be studied using its coordinate representative $f \circ \phi_i^{-1}$ with assurance that the conclusions drawn are patch independent. Introducing local complex coordinates in the charts U_i on M, the ϕ_i can be expressed as maps from U_i to an open set in $\mathbb{C}^{\frac{n}{2}}$, with $\phi_j \circ \phi_i^{-1}$ being a holomorphic map from $\mathbb{C}^{\frac{n}{2}}$ to $\mathbb{C}^{\frac{n}{2}}$. Clearly, n must be even for this to make sense. In local complex coordinates, we recall that a function $h : \mathbb{C}^{\frac{n}{2}} \to \mathbb{C}^{\frac{n}{2}}$ is *holomorphic* if $h(z^1, \bar{z}^1, ..., z^{\frac{n}{2}}, \bar{z}^{\frac{n}{2}})$ is actually independent of all the $\bar{z}^j$.

In a given patch on any even–dimensional manifold, we can always introduce local complex coordinates by, for instance, forming the combinations $z^j = x^j + ix^{\frac{n}{2}+j}$, where the x^j are local real coordinates on M. The real test is whether the transition functions from one patch to another — when expressed

in terms of the local complex coordinates — are holomorphic maps. If they are, we say that M is a complex manifold of complex dimension $d = n/2$. The local complex coordinates with holomorphic transition functions give M with a *complex structure* (see [Gre96]).

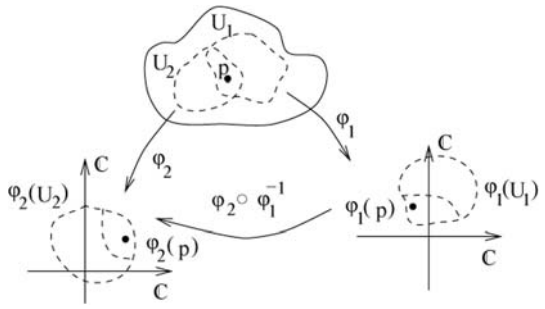


Fig. 1.5. The charts for a complex manifold M have the complex coordinates.

Given a smooth manifold with even real dimension n, it can be a difficult question to determine whether or not a complex structure exists. On the other hand, if some smooth manifold M does admit a complex structure, we are not able to decide whether it is unique, i.e., there may be numerous inequivalent ways of defining complex coordinates on M.

Now, in the same way as a *homeomorphism* defines an equivalence between topological manifolds, and a *diffeomorphism* defines an equivalence between smooth manifolds, a *biholomorphism* defines an equivalence between complex manifolds. If M and N are complex manifolds, we consider them to be equivalent if there is a map $\phi : M \to N$ which in addition to being a diffeomorphism, is also a holomorphic map. That is, when expressed in terms of the complex structures on M and N respectively, ϕ is holomorphic. It is not hard to show that this necessarily implies that ϕ^{-1} is holomorphic as well and hence ϕ is known as a biholomorphism. Such a map allows us to identify the complex structures on M and N and hence they are isomorphic as complex manifolds.

These definitions are important because there are pairs of smooth manifolds M and N which are homeomorphic but not diffeomorphic, as well as, there are complex manifolds M and N which are diffeomorphic but not biholomorphic. This means that if one simply ignored the fact that M and N admit local complex coordinates (with holomorphic transition functions), and one only worked in real coordinates, there would be no distinction between M and N. The difference between them only arises from the way in which complex coordinates have been laid down upon them.

Again, recall that a *tangent space* to a manifold M *at a point* p is the closest flat approximation to M at that point. A convenient basis for the tangent space of M at p consists of the n linearly independent partial derivatives,

$$T_pM : \{\partial_{x^1}|_p, ..., \partial_{x^n}|_p\}. \tag{1.97}$$

A vector $v \in T_pM$ can then be expressed as $v = v^\alpha \partial_{x^\alpha}|_p$.

Also, a convenient basis for the dual, *cotangent space* T_p^*M, is the basis of one–forms, which is dual to (1.97) and usually denoted by

$$T_p^*M : \{dx^1|_p, ..., dx^n|_p\}, \tag{1.98}$$

where, by definition, $dx^i : T_pM \to \mathbb{R}$ is a linear map with $dx^i_p(\partial_{x^j}|_p) = \delta^i_j$.

Now, if M is a complex manifold of complex dimension $d = n/2$, there is a notion of the *complexified tangent space* of M, denoted by $T_pM^{\mathbb{C}}$, which is the same as the real tangent space T_pM except that we allow complex coefficients to be used in the vector space manipulations. This is often denoted by writing $T_pM^{\mathbb{C}} = T_pM \otimes \mathbb{C}$. We can still take our basis to be as in (1.97) with an arbitrary vector $v \in T_pM^{\mathbb{C}}$ being expressed as $v = v^\alpha \frac{\partial}{\partial x^\alpha}|_p$, where the v^α can now be complex numbers. In fact, it is convenient to rearrange the basis vectors in (1.97) to more directly reflect the underlying complex structure. Specifically, we take the following linear combinations of basis vectors in (1.97) to be our new basis vectors:

$$\begin{aligned} T_pM^{\mathbb{C}} \quad : \quad & \{(\partial_{x^1} + i\partial_{x^{d+1}})|_p, ..., \\ & (\partial_{x^d} + i\partial_{x^{2d}})|_p, (\partial_{x^1} - i\partial_{x^{d+1}})|_p, ..., (\partial_{x^d} - i\partial_{x^{2d}})|_p\}. \end{aligned} \tag{1.99}$$

In terms of complex coordinates we can write the basis (1.99) as

$$T_pM^{\mathbb{C}} : \{\partial_{z^1}|_p, ..., \partial_{z^d}|_p, \partial_{\bar{z}^1}|_p, ..., \partial_{\bar{z}^d}|_p\}.$$

From the point of view of real vector spaces, $\partial_{x^j}|_p$ and $i\partial_{x^j}|_p$ would be considered linearly independent and hence $T_pM^{\mathbb{C}}$ has real dimension $4d$.

In exact analogy with the real case, we can define the dual to $T_pM^{\mathbb{C}}$, which we denote by $T_p^*M^{\mathbb{C}} = T_p^*M \otimes \mathbb{C}$, with the one–forms basis

$$T_p^*M^{\mathbb{C}} : \{dz^1|_p, ..., dz^d|_p, d\bar{z}^1|_p, ..., d\bar{z}^d|_p\}.$$

For certain types of complex manifolds M, it is worthwhile to refine the definition of the complexified tangent and cotangent spaces, which pulls apart the holomorphic and anti–holomorphic directions in each of these two vector spaces. That is, we can write

$$T_pM^{\mathbb{C}} = T_pM^{(1,0)} \oplus T_pM^{(0,1)},$$

where $T_pM^{(1,0)}$ is the vector space spanned by $\{\partial_{z^1}|_p, ..., \partial_{z^d}|_p\}$ and $T_pM^{(0,1)}$ is the vector space spanned by $\{\partial_{\bar{z}^1}|_p, ..., \partial_{\bar{z}^d}|_p\}$. Similarly, we can write

$$T_p^*M^{\mathbb{C}} = T_p^*M^{(1,0)} \oplus T_p^*M^{(0,1)},$$

where $T_p^*M^{(1,0)}$ is the vector space spanned by $\{dz^1|_p, ..., dz^d|_p\}$ and $T_p^*M^{(0,1)}$ is the vector space spanned by $\{d\bar{z}^1|_p, ..., d\bar{z}^d|_p\}$. We call $T_pM^{(1,0)}$ the *holomorphic tangent space*; it has complex dimension d and we call $T_p^*M^{1,0}$ the

holomorphic cotangent space. It also has complex dimension d. Their complements are known as the *anti–holomorphic* tangent and cotangent spaces respectively [Gre96].

Now, a *complex vector bundle* is a vector bundle $\pi : E \to M$ whose fiber bundle $\pi^{-1}(x)$ is a complex vector space. It is not necessarily a complex manifold, even if its base manifold M is a complex manifold. If a complex vector bundle also has the structure of a complex manifold, and is holomorphic, then it is called a holomorphic vector bundle.

A *Hermitian metric* on a complex vector bundle assigns a *Hermitian inner product* to every fiber bundle. The basic example is the trivial bundle $\pi : U \times \mathbb{C}^2 \to U$, where U is an open set in $\mathbb{R}^n$. Then a positive definite Hermitian matrix H defines a Hermitian metric by

$$\langle v, w \rangle = v^T H \bar{w},$$

where $\bar{w}$ is the complex conjugate of w. By a partition of unity, any complex vector bundle has a Hermitian metric.

In the special case of a complex manifold, the complexified tangent bundle $TM \otimes \mathbb{C}$ may have a Hermitian metric, in which case its real part is a Riemannian metric and its imaginary part is a nondegenerate alternating multilinear form ω. When ω is closed, i.e., in this case a symplectic form, then ω is called the *Kähler form*.

On a holomorphic vector bundle with a Hermitian metric h, there is a unique connection compatible with h and the complex structure. Namely, it must be $\nabla = \partial + \bar{\partial}$.

A *Kähler structure* on a complex manifold M combines a Riemannian metric on the underlying real manifold with the complex structure. Such a structure brings together geometry and complex analysis, and the main examples come from algebraic geometry. When M has n complex dimensions, then it has $2n$ real dimensions. A Kähler structure is related to the unitary group $U(n)$, which embeds in $SO(2n)$ as the orthogonal matrices that preserve the *almost complex structure* (multiplication by i). In a coordinate chart, the complex structure of M defines a multiplication by i and the metric defines orthogonality for tangent vectors. On a *Kähler manifold*, these two notions (and their derivatives) are related.

A Kähler manifold is a complex manifold for which the exterior derivative of the fundamental form ω associated with the given *Hermitian metric* vanishes, so $d\omega = 0$. In other words, it is a complex manifold with a *Kähler structure*. It has a *Kähler form*, so it is also a symplectic manifold. It has a *Kähler metric*, so it is also a Riemannian manifold.

The simplest example of a Kähler manifold is a Riemann surface, which is a complex manifold of dimension 1. In this case, the imaginary part of any Hermitian metric must be a closed form since all $2-$forms are closed on a real 2D manifold.

A *Kähler form* is a closed two–form ω on a complex manifold M which is also the negative imaginary part of a Hermitian metric $h = g - iw$ is called a

Kähler form. In this case, M is called a Kähler manifold and g, the real part of the Hermitian metric, is called a Kähler metric. The Kähler form combines the metric and the complex structure, $g(X,Y) = \omega(X, JY)$,where J is the *almost complex structure* induced by multiplication by i. Since the Kähler form comes from a Hermitian metric, it is preserved by J, since $h(X,Y) = h(JX, JY)$. The equation $d\omega = 0$ implies that the metric and the complex structure are related. It gives M a Kähler structure, and has many implications.

On $\mathbb{C}^2$, the Kähler form can be written as

$$\omega = -\frac{1}{2} i \left(dz_1 \wedge \overline{dz_1} + dz_2 \wedge \overline{dz_2} \right) = dx_1 \wedge dy_1 + dx_2 \wedge dy_2,$$

where $z_n = x_n + y_n$. In general, the Kähler form can be written in coordinates

$$\omega = g_{ij}\, dz_i \wedge \overline{dz_j},$$

where g_{ij} is a Hermitian metric, the real part of which is the Kähler metric. Locally, a Kähler form can be written as $\partial\bar{\partial} f$, where f is a function called a *Kähler potential*. The Kähler form is a real $(1,1)-$complex form. The Kähler potential is a real–valued function f on a Kähler manifold for which the Kähler form ω can be written as $\omega = i\partial\bar{\partial} f$, where,

$$\partial = \partial_{z_k} dz_k \qquad \text{and} \qquad \bar{\partial} = \partial_{\bar{z}_k} d\bar{z}_k.$$

Since the Kähler form ω is closed, it represents a cohomology class in the *de Rham cohomology*. On a compact manifold, it cannot be exact because $\omega^n/n! \neq 0$ is the volume form determined by the metric. In the special case of a projective variety, the Kähler form represents an integral cohomology class. That is, it integrates to an integer on any 1D submanifold, i.e., an algebraic curve. The Kodaira embedding theorem says that if the Kähler form represents an integral cohomology class on a compact manifold, then it must be a projective variety. There exist Kähler forms which are not projective algebraic, but it is an open question whether or not any Kähler manifold can be deformed to a projective variety (in the compact case).

A Kähler form satisfies *Wirtinger's inequality*,

$$|\omega(X,Y)| \leq |X \wedge Y|\,,$$

where the r.h.s is the volume of the parallelogram formed by the tangent vectors X and Y. Corresponding inequalities hold for the *exterior powers* of ω. Equality holds iff X and Y form a complex subspace. Therefore, there is a *calibration form*, and the complex submanifolds of a Kähler manifold are *calibrated submanifolds*. In particular, the complex submanifolds are locally volume minimizing in a Kähler manifold. For example, the graph of a holomorphic function is a locally area–minimizing surface in $\mathbb{C}^2 = \mathbb{R}^4$.

Kähler identities is a collection of identities which hold on a Kähler manifold, also called the *Hodge identities*. Let ω be a Kähler form, $d = \partial + \bar{\partial}$ be the

exterior derivative, $[A, B] = AB - BA$ be the commutator of two differential operators, and A^* denote the formal adjoint of A. The following operators also act on differential forms α on a Kähler manifold:

$$L(\alpha) = \alpha \wedge \omega, \qquad \Lambda(\alpha) = L^*(\alpha) = \alpha \rfloor \omega, \qquad d_c = -JdJ,$$

where J is the almost complex structure, $J = -I$, and $\rfloor$ denotes the interior product. Then we have

$$\begin{aligned} &[L, \bar{\partial}] = [L, \partial] = 0, \qquad [\Lambda, \bar{\partial}^*] = [\Lambda, \partial^*] = 0, \\ &[L, \bar{\partial}^*] = -i\partial, \qquad [L, \partial^*] = i\bar{\partial}, \qquad [\Lambda, \bar{\partial}] = -i\partial^*, \qquad [\Lambda, \partial] = -i\bar{\partial}. \end{aligned}$$

These identities have many implications. For example, the two operators

$$\Delta_d = dd^* + d^*d \qquad \text{and} \qquad \Delta_{\bar{\partial}} = \bar{\partial}\bar{\partial}^* + \bar{\partial}^*\bar{\partial}$$

(called Laplacians because they are elliptic operators) satisfy

$$\Delta_d = 2\Delta_{\bar{\partial}}.$$

At this point, assume that M is also a compact manifold. Along with Hodge's theorem, this equality of Laplacians proves the *Hodge decomposition*. The operators L and Λ commute with these Laplacians. By *Hodge's theorem*, they act on cohomology, which is represented by *harmonic forms*. Moreover, defining

$$H = [L, \Lambda] = \sum (p + q - n)\, \Pi^{p,q},$$

where $\Pi^{p,q}$ is projection onto the $(p, q)-$Dolbeault cohomology, they satisfy

$$[L, \Lambda] = H, \qquad [H, L] = -2L, \qquad [H, \Lambda] = 2L.$$

In other words, these operators give a group representation of the special linear Lie algebra $\mathfrak{sl}_2(\mathbb{C})$ on the complex cohomology of a compact Kähler manifold (*Lefschetz theorem*).

Dolbeault Cohomology and Hodge Numbers

A generalization of the real–valued de Rham cohomology to complex manifolds is called the *Dolbeault cohomology*. On complex mD manifolds, we have local coordinates z^i and $\bar{z}^i$. One can now study $(p, q)-$forms, which are forms containing p factors of dz^i and q factors of $d\bar{z}^j$:

$$\omega = \omega_{i_1 \ldots i_p, j_1 \ldots j_q}(z, \bar{z})\, dz^{i_1} \wedge \cdots \wedge dz^{i_p} \wedge d\bar{z}^{j_1} \wedge \cdots \wedge d\bar{z}^{j_q}.$$

Moreover, one can introduce *two* exterior derivative operators ∂ and $\bar{\partial}$, where ∂ is defined by

$$\partial\omega \equiv \frac{\partial \omega_{i_1 \ldots i_p, j_1 \ldots j_q}}{\partial z^k}\, dz^k \wedge dz^{i_1} \wedge \cdots \wedge dz^{i_p} \wedge d\bar{z}^{j_1} \wedge \cdots \wedge d\bar{z}^{j_q},$$

and $\bar{\partial}$ is defined similarly by differentiating with respect to $\bar{z}^k$ and adding a factor of $d\bar{z}^k$. Again, both of these operators square to zero. We can now construct two cohomologies – one for each of these operators – but as we will see, in the cases that we are interested in, the information contained in them is the same. Conventionally, one uses the cohomology defined by the $\bar{\partial}-$operator.

For complex manifolds, the *Hodge theorem* also holds: *each cohomology class $H^{p,q}(M)$ contains a unique harmonic form.* Here, a harmonic form ω_h is a form for which the *complex Laplacian*

$$\Delta_{\bar{\partial}} = \bar{\partial}\bar{\partial}^* + \bar{\partial}^*\bar{\partial}$$

has a zero eigenvalue: $\Delta_{\bar{\partial}}\omega_h = 0$. In general, this operator does not equal the ordinary Laplacian, but one can prove that in the case where M is a *Kähler manifold*,

$$\Delta = 2\Delta_{\bar{\partial}} = 2\Delta_{\partial}.$$

In other words, on a Kähler manifold the notion of a harmonic form is the same, independently of which exterior derivative one uses. As a first consequence, we find that the vector spaces $H^{p,q}_{\partial}(M)$ and $H^{p,q}_{\bar{\partial}}(M)$ both equal the vector space of harmonic $(p,q)-$forms, so the two cohomologies are indeed equal. Moreover, every $(p,q)-$form is a $(p+q)-$form in the de Rham cohomology, and by the above result we see that a *harmonic* $(p,q)-$form can also be viewed as a de Rham harmonic $(p+q)-$form. Conversely, any de Rham $p-$form can be written as a sum of Dolbeault forms:

$$\omega_p = \omega_{p,0} + \omega_{p-1,1} + \ldots + \omega_{0,p}. \tag{1.100}$$

Acting on this with the Laplacian, one sees that for a harmonic $p-$form,

$$\Delta\omega_p = \Delta_{\bar{\partial}}\omega_p = \Delta_{\bar{\partial}}\omega_{p,0} + \Delta_{\bar{\partial}}\omega_{p-1,1} + \ldots + \Delta_{\bar{\partial}}\omega_{0,p} = 0.$$

Since $\Delta_{\bar{\partial}}$ does not change the degree of a form, $\Delta_{\bar{\partial}}\omega_{p_1,p_2}$ is also a $(p_1,p_2)-$form. Therefore, the r.h.s. can only vanish if each term vanishes separately, so all the terms on the r.h.s. of (1.100) must be harmonic forms. Summarizing, we have shown that the vector space of harmonic de Rham $p-$forms is a direct sum of the vector spaces of harmonic Dolbeault $(p_1,p_2)-$forms with $p_1+p_2=p$. Since the harmonic forms represent the cohomology classes in a 1–1 way, we find the important result that for Kähler manifolds,

$$H^p(M) = H^{p,0}(M) \oplus H^{p-1,1}(M) \oplus \cdots \oplus H^{0,p}(M).$$

That is, the Dolbeault cohomology can be viewed as a refinement of the de Rham cohomology. In particular, we have

$$b^p = h^{p,0} + h^{p-1,1} + \ldots + h^{0,p},$$

where $h^{p,q} = \dim H^{p,q}(M)$ are called the *Hodge numbers* of M.

The Hodge numbers of a Kähler manifold give us several topological invariants, but not all of them are independent. In particular, the following two relations hold:

$$h^{p,q} = h^{q,p}, \qquad h^{p,q} = h^{m-p,m-q}. \tag{1.101}$$

The first relation immediately follows if we realize that $\omega \mapsto \overline{\omega}$ maps $\partial-$harmonic $(p,q)-$forms to $\bar{\partial}-$harmonic $(q,p)-$forms, and hence can be viewed as an invertible map between the two respective cohomologies. As we have seen, the $\partial-$cohomology and the $\bar{\partial}-$cohomology coincide on a Kähler manifold, so the first of the above two equations follows.

The second relation can be proved using the map

$$(\alpha, \omega) \mapsto \int_M \alpha \wedge \omega$$

from $H^{p,q} \times H^{m-p,m-q}$ to $\mathbb{C}$. It can be shown that this map is nondegenerate, and hence that $H^{p,q}$ and $H^{m-p,m-q}$ can be viewed as dual vector spaces. In particular, it follows that these vector spaces have the same dimension, which is the statement in the second line of (1.101).

Note that the last argument also holds for de Rham cohomology, in which case we find the relation $b^p = b^{n-p}$ between the Betti numbers. We also know that $H^{n-p}(M)$ is dual to $H_{n-p}(M)$, so combining these statements we find an identification between the vector spaces $H^p(M)$ and $H_{n-p}(M)$. Recall that this identification between $p-$form cohomology classes and $(n-p)-$cycle homology classes represents the *Poincaré duality*. Intuitively, take a certain $(n-p)-$cycle Σ representing a homology class in H_{n-p}. One can now try to define a 'delta function' $\delta(\Sigma)$ which is localized on this cycle. Locally, Σ can be parameterized by setting p coordinates equal to zero, so $\delta(\Sigma)$ is a 'pD delta function' – that is, it is an object which is naturally integrated over pD submanifolds: a $p-$form. This intuition can be made precise, and one can indeed view the cohomology class of the resulting 'delta–function' $p-$form as the Poincar é dual to Σ.

Going back to the relations (1.101), we see that the Hodge numbers of a Kähler manifold can be nicely written in a so–called *Hodge–diamond* form:

$$\begin{array}{ccccccc}
 & & & h^{0,0} & & & \\
 & & h^{1,0} & & h^{0,1} & & \\
 & \iddots & & & & \ddots & \\
h^{m,0} & & & \cdots & & & h^{0,m} \\
 & \ddots & & & & \iddots & \\
 & & h^{m,m-1} & & h^{m-1,m} & & \\
 & & & h^{m,m} & & &
\end{array}$$

The integers in this diamond are symmetrical under the reflection in its horizontal and vertical axes.

1.2.13 Conformal Killing–Riemannian Geometry

In this subsection we present some basic facts from *conformal Killing–Riemannian geometry*. In mechanics it is well–known that *symmetries of Lagrangian* or *Hamiltonian* result in *conservation laws*, that are used to deduce constants of motion for the trajectories (geodesics) on the configuration manifold M. The same constants of motion are get using geometrical language, where a *Killing vector–field* is the standard tool for the description of symmetry [MTW73]. A Killing vector–field ξ^i is a vector–field on a Riemannian manifold M with metrics g, which in coordinates $x^j \in M$ satisfies the *Killing equation*

$$\xi^{i;j} + \xi^{j;i} = \xi^{(i;j)} = 0, \qquad \text{or} \qquad \mathcal{L}_{\xi^i} g_{ij} = 0, \tag{1.102}$$

where semicolon denotes the covariant derivative on M, the indexed bracket denotes the tensor symmetry, and $\mathcal{L}$ is the Lie derivative.

The *conformal Killing vector–fields* are, by definition, infinitesimal conformal symmetries i.e., the flow of such vector–fields preserves the conformal class of the metric. The number of linearly–independent conformal Killing fields measures the degree of *conformal symmetry of the manifold*. This number is bounded by $\frac{1}{2}(n+1)(n+2)$, where n is the dimension of the manifold. It is the maximal one if the manifold is conformally flat [Bau00].

Now, to properly initialize our conformal geometry, recall that *conformal twistor spinor–fields* φ were introduced by R. Penrose into physics (see [Pen67, PR86]) as solutions of the *conformally covariant twistor equation*

$$\nabla^S_X \varphi + \frac{1}{n} X \cdot D\varphi = 0,$$

for each vector–fields X on a Riemannian manifold (M, g), where D is the *Dirac operator*. Each twistor spinor–field φ on (M, g) defines a *conformal vector–field* V_φ on M by

$$g(V_\varphi, X) = i^{k+1} \langle X \cdot \varphi, \varphi \rangle.$$

Also, each twistor spinor–field φ that satisfies the *Dirac equation* on (M, g),

$$D\varphi = \mu\varphi,$$

is called a *Killing spinor–field*. Each twistor spinor–field without zeros on (M, g) can be transformed by a conformal change of the metric g into a Killing spinor–field [Bau00].

Conformal Killing Vector–Fields and Forms on M

The space of all conformal Killing vector–fields forms the Lie algebra of the *isometry group* of a Riemannian manifold (M, g) and the number of linearly independent Killing vector–fields measures the *degree of symmetry* of M. It

is known that this number is bounded from above by the dimension of the isometry group of the standard sphere and, on compact manifolds, equality is attained if and only if the manifold M is isometric to the standard sphere or the real projective space. Slightly more generally one can consider *conformal vector–fields*, i.e., vector–fields with a flow preserving a given conformal class of metrics. There are several geometrical conditions which force a conformal vector–field to be Killing [Sem02].

A natural generalization of conformal vector–fields are the *conformal Killing forms* [Yan52], also called *twistor forms* [MS03]. These are $p-$forms α satisfying for any vector–field X on the manifold M the *Killing–Yano equation*

$$\nabla_X \alpha \; - \; \tfrac{1}{p+1} X \rfloor \, d\alpha \; + \; \tfrac{1}{n-p+1} X^* \wedge d^*\alpha \; = \; 0, \tag{1.103}$$

where n is the dimension of the manifold (M, g), ∇ denotes the covariant derivative of the Levi–Civita connection on M, X^* is $1-$form dual to X and $\rfloor$ is the operation dual to the wedge product on M. It is easy to see that a conformal Killing $1-$form is dual to a conformal vector–field. *Coclosed* conformal Killing $p-$forms are called *Killing forms*. For $p = 1$ they are dual to Killing vector–fields.

Let α be a Killing $p-$form and let γ be a geodesic on (M, g), i.e., $\nabla_{\dot\gamma}\, \dot\gamma = 0$. Then

$$\nabla_{\dot\gamma} \, (\dot\gamma \rfloor \, \alpha) \; = \; (\nabla_{\dot\gamma} \, \dot\gamma) \rfloor \, \alpha \; + \; \dot\gamma \rfloor \, \nabla_{\dot\gamma} \, \alpha \; = \; 0,$$

i.e., $\dot\gamma \rfloor \, \alpha$ is a $(p-1)$–form parallel along the geodesic γ and in particular its length is constant along γ.

The l.h.s of equation (1.103) defines a first–order elliptic differential operator T, the so–caled *twistor operator*. Equivalently one can describe a conformal Killing form as a form in the kernel of twistor operator T. From this point of view conformal Killing forms are similar to Penrose's *twistor spinors* in *Lorentzian spin geometry*. One shared property is the conformal invariance of the defining equation. In particular, any form which is parallel for some metric g, and thus a Killing form for trivial reasons, induces non–parallel conformal Killing forms for metrics conformally equivalent to g (by a non–trivial change of the metric) [Sem02].

Conformal Killing Tensors and Laplacian Symmetry on M

In an nD Riemannian manifold (M, g), a *Killing tensor–field* (of order 2) is a symmetric tensor K^{ab} satisfying (generalizing (1.102))

$$K^{(ab;c)} = 0. \tag{1.104}$$

A *conformal Killing tensor–field* (of order 2) is a symmetric tensor Q^{ab} satisfying

$$Q^{(ab;c)} = q^{(a} g^{bc)}, \qquad \text{with} \qquad q^a = (Q^{,a} + 2Q^{a;d}_{\;\;d})/(n+2), \tag{1.105}$$

where comma denotes partial derivative and $Q = Q^d_d$. When the associated *conformal vector* q^a is nonzero, the conformal Killing tensor will be called *proper* and otherwise it is a (ordinary) Killing tensor. If q^a is a Killing vector, Q^{ab} is referred to as a *homothetic Killing tensor*. If the associated conformal vector $q^a = q^{,a}$ is the gradient of some scalar field q, then Q^{ab} is called a *gradient conformal Killing tensor*. For each gradient conformal Killing tensor Q^{ab} there is an associated Killing tensor K^{ab} given by

$$K^{ab} = Q^{ab} - qg^{ab}, \tag{1.106}$$

which is defined only up to the addition of a constant multiple of the inverse metric tensor g^{ab}.

Some authors define a conformal Killing tensor as a *trace–free* tensor P^{ab} satisfying $P^{(ab;c)} = p^{(a}g^{bc)}$. Note that there is no contradiction between the two definitions: if P^{ab} is a trace–free conformal Killing tensor then for any scalar field λ, $P^{ab} + \lambda g^{ab}$ is a conformal Killing tensor and conversely if Q^{ab} is a conformal Killing tensor, its trace–free part $Q^{ab} - \frac{1}{n}Qg^{ab}$ is a trace–free Killing tensor [REB03].

Killing tensor–fields are of importance owing to their connection with quadratic first integrals of the geodesic equations: if p^a is tangent to an affinely parameterized geodesic (i.e., $p^a{}_{;b}p^b = 0$) it is easy to see that $K_{ab}p^ap^b$ is constant along the geodesic. For conformal Killing tensors $Q_{ab}p^ap^b$ is constant along *null* geodesics and here, only the trace–free part of Q_{ab} contributes to the constants of motion. Both Killing tensors and conformal Killing tensors are also of importance in connection with the separability of the Hamiltonian–Jacobi equations [CH64] (as well as other PDEs).

A Killing tensor is said to be *reducible* if it can be written as a constant linear combination of the metric and symmetrized products of Killing vectors,

$$K_{ab} = a^0 g_{ab} + a^{IJ}\xi_{I(a}\xi_{|J|b)}, \tag{1.107}$$

where ξ_I for $I = 1 \ldots N$ are the Killing vectors admitted by the manifold (M, g) and a^0 and a^{IJ} for $1 \leq I \leq J \leq N$ are constants. Generally one is interested only in Killing tensors which are not reducible since the quadratic constant of motion associated with a reducible Killing tensor is a constant linear combination of p^ap_a and of pairwise products of the linear constants of motion $\xi_{Ia}p^a$ [REB03].

More generally, any linear differential operator on a Riemannian manifold (M, g) may be written in the form [EG91, Eas02]

$$\mathcal{D} = V^{bc\cdots d}\nabla_b\nabla_c \cdots \nabla_d + \text{ lower order terms},$$

where $V^{bc\cdots d}$ is symmetric in its indices, and $\nabla_a = \partial/\partial x^a$ (differentiation in coordinates). This tensor is called the symbol of $\mathcal{D}$. We shall write $\phi^{(ab\cdots c)}$ for the symmetric part of $\phi^{ab\cdots c}$.

Now, a *conformal Killing tensor* on (M, g) is a symmetric trace–free tensor field, with s indices, satisfying

$$\text{the trace–free part of } \nabla^{(a}V^{bc\cdots d)} = 0, \tag{1.108}$$

or, equivalently,

$$\nabla^{(a}V^{bc\cdots d)} = g^{(ab}T^{c\cdots d)}, \tag{1.109}$$

for some tensor field $T^{c\cdots d}$ or, equivalently,

$$\nabla^{(a}V^{bc\cdots d)} = \tfrac{s}{n+2s-2}g^{(ab}\nabla_e V^{c\cdots d)e}, \tag{1.110}$$

where $\nabla^a = g^{ab}\nabla_b$ (the standard convention of raising and lowering indices with the metric tensor g_{ab}). When $s = 1$, these equations define a *conformal Killing vector*.

M. Eastwood proved the following theorem: any symmetry $\mathcal{D}$ of the Laplacian $\Delta = \nabla^a\nabla_a$ on a Riemannian manifold (M, g) is canonically equivalent to one whose symbol is a conformal Killing tensor [EG91, Eas02].

1.3 Fibre Bundles

1.3.1 Intuition Behind a Fibre Bundle

Recall that tangent and cotangent bundles, TM and T^*M, are special cases of a more general geometrical object called *fibre bundle*, where the word *fiber* V of a map $\pi : Y \longrightarrow X$ denotes the *preimage* $\pi^{-1}(x)$ of an element $x \in X$. It is a space which *locally* looks like a product of two spaces (similarly as a manifold locally looks like Euclidean space), but may possess a different *global* structure. To get a visual intuition behind this fundamental geometrical concept, we can say that a fibre bundle Y is a *homeomorphic generalization* of a *product space* $X \times V$ (see Figure 1.6), where X and V are called the *base* and the *fibre*, respectively. $\pi : Y \to X$ is called the *projection*, $Y_x = \pi^{-1}(x)$ denotes a fibre over a point x of the base X, while the map $f = \pi^{-1} : X \to Y$ defines the *cross–section*, producing the *graph* $(x, f(x))$ in the bundle Y (e.g., in case of a tangent bundle, $f = \dot{x}$ represents a velocity vector–field) (see [Ste72]).

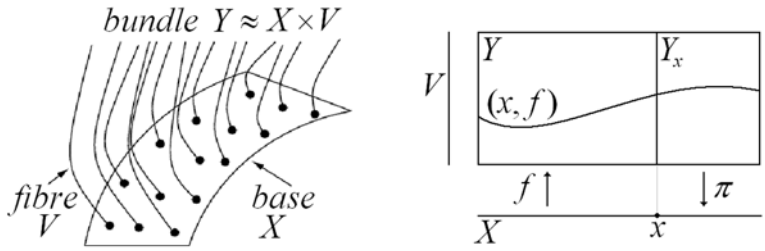


Fig. 1.6. A sketch of a fibre bundle $Y \approx X \times V$ as a generalization of a product space $X \times V$; left – main components; right – a few details (see text for explanation).

The main reason why we need to study fibre bundles is that *all dynamical objects* (including vectors, tensors, differential forms and gauge potentials) are their *cross–sections*, representing *generalizations of graphs of continuous functions*.

1.3.2 Definition of a Fibre Bundle

Let M denote an $n-$manifold with an atlas Ψ_M consisting of local coordinates $x^\alpha \in M$, $(\alpha = 1, ..., \dim M)$, given by

$$\Psi_M = \{U_\xi, \phi_\xi\}, \qquad \phi_\xi(x) = x^\alpha e_\alpha, \qquad (\text{for all } x \in U_\xi \subset M),$$

where $\{e_\alpha\}$ is a fixed basis of $\mathbb{R}^m$. Its tangent and cotangent bundles, TM and T^*M, respectively, admit atlases of induced coordinates $(x^\alpha, \dot{x}^\alpha)$ and $(x^\alpha, \dot{x}_\alpha)$, relative to the *holonomic fibre bases* $\{\partial_\alpha\}$ and $\{dx^\alpha\}$, respectively. For all elements (i.e., points) $p \in TM$ and $p^* \in T^*M$, we have (see [Sar93, Sar95, GMS97, MS00a, Sar02a])

$$p = \dot{x}^\alpha \partial_\alpha, \qquad p^* = \dot{x}_\alpha dx^\alpha, \qquad \partial_\alpha \rfloor dx^\alpha = \delta^\alpha_\alpha, \qquad (\alpha = 1, ..., \dim M).$$

Also, we will use the notation

$$\omega = dx^1 \wedge \cdots \wedge dx^n, \qquad \omega_\alpha = \partial_\alpha \rfloor \omega, \qquad \omega_{\mu\alpha} = \partial_\mu \rfloor \partial_\alpha \rfloor \omega. \tag{1.111}$$

If $f : M \to M'$ is a smooth manifold map, we define the induced *tangent map* Tf over f, given by

$$Tf : TM \longrightarrow TM', \qquad \dot{x}'^\alpha \circ Tf = \frac{\partial f^\alpha}{\partial x^\alpha} \dot{x}^\alpha. \tag{1.112}$$

Given a *manifold product* $M \times N$, π_1 and π_2 denote the *natural projections* (i.e., canonical surjections),

$$\pi_1 : M \times N \to M, \qquad \pi_2 : M \times N \to N.$$

Now, as a homeomorphic generalization of a product space, a fibre bundle can be viewed either as a topological or a geometrical (i.e., coordinate) construction. As a topological construction, a fibre bundle is a class of more general *fibrations*. To have a glimpse of this construction, let $I = [0, 1]$. A map $\pi : Y \to X$ is said to have the *homotopy lifting property* (HLP, see [Swi75]) with respect to a topological space Z if for every map $f : Z \to Y$ and homotopy $H : Z \times I \to X$ of $\pi \circ f$ there is a homotopy $V : Z \times I \to Y$ with $f = V_0$ and $\pi \circ V = H$. V is said to be a *lifting* of H. π is called a *fibration* if it has the HLP for all spaces Z and a *weak fibration* if it has the HLP for all disks D^n, $(n \geq 0)$. If $x \in X$ is the base point, then $V = \pi^{-1}(x)$ is called the *fibre* of the fibration π. The projection onto the first factor, $\pi_1 : X \times V \to X$, is clearly a fibration and is called the *trivial fibration* over X with the fibre V.

However, for the sake of applying differential and integral dynamics on fibre bundles, we will rather use Steenrod's *coordinate bundle definition* (see [Ste72]), which defines fibre bundle Y as a sextuple $(Y, X, \pi, V, G, \Psi_Y)$, with:

1. a space Y called the *total space*, *bundle space* (or simply *bundle*),
2. a space X called the *base space*,
3. a surjection $\pi : Y \longrightarrow X$ called the *projection*,
4. a space $V \subset Y$ called the *fibre*,
5. an effective topological (or Lie) transformation group G of V called the *group of the bundle*, and
6. a *bundle atlas* Ψ_Y.

Some standard examples of fibre bundles include any *Cartesian product* $X \times V \to X$ (which is a bundle over X with fibre V), the *Möbius strip* (which is a nontrivial fibre bundle over the circle S^1 with fibre given by the unit interval $I = [0, 1]$; the corresponding trivial bundle is a cylinder), the *Klein bottle* (which can be viewed as a 'twisted' circle bundle over another circle; thus, the corresponding trivial bundle is a torus, $S^1 \times S^1$), a 3*–sphere* S^3 (which is a bundle over S^2 with fibre S^1; more generally, a *sphere bundle* is a fiber bundle whose fiber is an $n-$sphere), while a *covering space* is a fiber bundle whose fiber is a discrete space.

Main properties of graphs of functions $f : X \to V$ carry over to fibre bundles. A graph of such a function, $(x, f(x))$, sits in the product space $X \times V$, or in its homeomorphic generalization bundle. A graph is always 1–1 and projects *onto* the base X.

A special class of fibre bundle is the *vector bundle*, in which the fibre is a vector space. Special cases of fibre bundles that we will use in dynamics of complex systems are: vector, affine, and principal bundles.

A fibre bundle also comes with a group G action on its fibre V, so it can also be called a $G-$bundle. This group action represents the different ways the fibre V can be viewed as equivalent (e.g., the group G might be the group of homeomorphisms (topological group) or diffeomorphisms (Lie group) of the fibre V; or, the group G on a vector bundle is the group of invertible linear maps, which reflects the equivalent descriptions of a vector space using different vector–space bases). A *principal bundle* is $G-$bundle where the fiber can be identified with the group G itself and where there is a right action of G on the bundle space which is fiber preserving.

Fibre bundles are not always used to generalize functions. Sometimes they are convenient descriptions of interesting manifolds. A common example is a torus bundle on the circle.

More specifically, a fibre bundle, or *fibre bundle* Y over an nD base X is defined as a manifold surjection

$$\pi : Y \to X, \tag{1.113}$$

where Y admits an atlas Ψ_Y of fibre coordinates

$$(x^\alpha, y^i), \qquad x^\alpha \to x'^\alpha(x^\mu), \qquad y^i \to y'^i(x^\mu, y^j), \tag{1.114}$$

compatible with the fibration (1.113), i.e., such that x^α are coordinates on the base X,

$$\pi : Y \ni (x^\alpha, y^i) \mapsto x^\alpha \in X.$$

This condition is equivalent to π being a submersion, which means that its tangent map $T\pi : TY \to TX$ is a surjection. This also implies that π is an *open map*.

A fibre bundle $Y \to X$ is said to be *trivial* if it is equivalent to the Cartesian product of manifolds, $Y \cong X \times V$, i.e., defined as $\pi_1 : X \times V \to X$.[10] A fibre bundle over a *contractible base* is always trivial [Ste72].

A fibre bundle $Y \to X$ is said to be *locally trivial* if there exists a fibred coordinate atlas Ψ_Y over an open covering $\{\pi^{-1}(U_\xi)\} \in Y$ of the bundle space Y where $\{U_\xi\} \in X$ is an open covering of the base space X. In other words, all points of the same fibre $Y_x = \pi^{-1}(x)$ of a bundle Y can be covered by the same fibred coordinate chart $\psi_\xi \in \Psi_Y$, so that we have the standard fibre–manifold V for all local bundle splittings

$$\psi_\xi : \pi^{-1}(U_\xi) \to U_\xi \times V.$$

For the purpose of our general dynamics, the most important fibre bundles are those which are at the same time smooth manifolds. A fibre bundle $Y \to X$ is said to be *smooth* (C^∞) if there exist a typical fibre–manifold V and an open covering $\{U_\xi\}$ of X such that Y is locally diffeomorphic to the splittings

$$\psi_\xi : \pi^{-1}(U_\xi) \to U_\xi \times V, \tag{1.115}$$

glued together by means of smooth *transition functions*

$$\rho_{\xi\zeta} = \psi_\xi \circ \psi_\zeta^{-1} : U_\xi \cap U_\zeta \times V \to U_\xi \cap U_\zeta \times V \tag{1.116}$$

on overlaps $U_\xi \cap U_\zeta$. It follows that fibres $Y_x = \pi^{-1}(x)$, (for all $x \in X$), of a fibre bundle are its closed imbedded submanifolds. Transition functions $\rho_{\xi\zeta}$ fulfil the *cocycle condition*

$$\rho_{\xi\zeta} \circ \rho_{\zeta\iota} = \rho_{\xi\iota} \tag{1.117}$$

on all overlaps $U_\xi \cap U_\zeta \cap U_\iota$. Trivialization charts (U_ξ, ψ_ξ) together with transition functions $\rho_{\xi\zeta}$ (1.116) constitute a bundle atlas

$$\Psi_Y = \{(U_\xi, \psi_\xi), \rho_{\xi\zeta}\} \tag{1.118}$$

of a fibre bundle $Y \to X$. Two bundle atlases are said to be *equivalent* if their union is also a bundle atlas, i.e., there exist unique transition functions between trivialization charts of different atlases. A fibre bundle $Y \to X$ is uniquely defined by a bundle atlas, and all its atlases are equivalent. Every smooth fibre bundle $Y \to X$ admits a bundle atlas Ψ_Y over a finite covering $\{U_\xi\}$ of X.

If $Y \to X$ is a fibre bundle, the fibre coordinates $(x^\alpha, y^i) \in Y$ are assumed to be bundle coordinates associated with a bundle atlas Ψ_Y, that is,

[10] A trivial fibre bundle admits different trivializations $Y \cong X \times V$ that differ from each other in surjections $Y \to V$.

$$y^i(y) = (v^i \circ \pi_2 \circ \psi_\xi)(y), \qquad (\pi(y) \in U_\xi \subset X), \tag{1.119}$$

where $v^i \in V \subset Y$ are coordinates of the standard fibre V of Y.

Maps of fibre bundles (or, *bundle maps*), by definition, *preserve their fibrations*, i.e., send a fibre to a fibre. Namely, a *bundle map* of a fibre bundle $\pi : Y \to X$ to a fibre bundle $\pi' : Y' \to X'$ is defined as a pair (Φ, f) of manifold maps such that the following diagram commutes

$$\begin{array}{ccc} Y & \xrightarrow{\Phi} & Y' \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi'} \\ X & \xrightarrow[f]{} & X' \end{array}$$

i.e., Φ is a fibrewise map over f which sends a fibre Y_x, (for all $x \in X$), to a fibre $Y'_{f(x)}$, (for all $f(x) \in X'$). A *bundle diffeomorphism* is called an *automorphism* if it is an *isomorphism* to itself. In field theory, any automorphism of a fibre bundle is treated as a *gauge transformation*.

Given a bundle $Y \to X$, every map $f : X' \to X$ induces a bundle $Y' = f^*Y$ over X' which is called the *pull–back* of the bundle Y by f, such that the following diagram commutes

$$\begin{array}{ccc} Y & \xleftarrow{f^*} & Y' \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi'} \\ X & \xleftarrow[f]{} & X' \end{array}$$

In particular, the product $Y \times Y'$ over X of bundles $\pi : Y \to X$ and $\pi' : Y' \to X$ is the pull–back

$$Y \times Y' = \pi^* Y' = \pi'^* Y.$$

Classical fields are described by *sections* of fibre bundles. A (*global*) *section* of a fibre bundle $Y \to X$ is defined as a $\pi-$inverse manifold injection $s : X \to Y$, $s(x) \mapsto Y_x$, such that $\pi \circ s = \mathrm{Id}_X$. That is, a section s sends any point $x \in X$ into the fibre $Y_x \subset Y$ over this point. A section s is an imbedding, i.e., $s(X) \subset Y$ is both a submanifold and a topological subspace of Y. It is also a closed map, which sends closed subsets of X onto closed subsets of Y. Similarly, a section of a fibre bundle $Y \to X$ over a submanifold of X is defined. Given a bundle atlas Ψ_Y and associated bundle coordinates (x^α, y^i), a section s of a fibre bundle $Y \to X$ is represented by collections of local functions $\{s^i = y^i \circ \psi_\xi \circ s\}$ on trivialization sets $U_\xi \subset X$.

A fibre bundle $Y \to X$ whose typical fibre is diffeomorphic to an Euclidean space $\mathbb{R}^m$ has a global section. More generally, its section over a closed imbedded submanifold (e.g., a point) of X is extended to a global section [Ste72].

In contrast, by a *local section* is usually meant a section over an open subset of the base X. A fibre bundle admits a local section around each point of its base, but need not have a global section.

For any $n \geq 1$ the *normal bundle* NS^n of the $n-$sphere S^n is the fibre bundle $(S^n, p', E', \mathbb{R}^1)$, where $E' = \{(x,y) \in \mathbb{R}^{n+1} \times \mathbb{R}^{n+1} : \|x\| = 1, y = \lambda x, \lambda \in \mathbb{R}^1\}$ and $p' : E' \to S^n$ is defined by $p'(x,y) = x$ [Swi75].

1.3.3 Vector and Affine Bundles

The most important fibre bundles are vector and affine bundles, which give a standard framework in both classical and quantum dynamics and field theory (e.g., matter fields are sections of vector bundles, while gauge potentials are sections of affine bundles).

Recall that both the tangent bundle (TM, π_M, M) and the cotangent bundle $(T^*M,\ \pi^*_M, M)$ are examples of a more general notion of *vector bundle* (E, π, M) of a manifold M, which consists of manifolds E (the total space) and M (the base), as well as a smooth map $\pi : E \to M$ (the projection) together with an equivalence class of vector bundle atlases (see [KMS93]). A vector bundle atlas $(U_\alpha, \phi_\alpha)_{\alpha \in A}$ for (E, π, M) is a set of pairwise compatible vector bundle charts (U_α, ϕ_α) such that $(U_\alpha)_{\alpha \in A}$ is an open cover of M. Two vector bundle atlases are called equivalent, if their union is again a vector bundle atlas.

On each fibre $E_m = \pi^{-1}(m)$ corresponding to the point $m \in M$ there is a unique structure of a real vector space, induced from any vector bundle chart (U_α, ϕ_α) with $m \in U_\alpha$. A section u of (E, π, M) is a smooth map $u : M \to E$ with $\pi \circ u = Id_M$.

Let (E, π_M, M) and (F, π_N, N) be vector bundles. A *vector bundle homomorphism* $\Phi : E \to F$ is a fibre respecting, fibre linear smooth map induced by the smooth map $\varphi : M \to N$ between the base manifolds M and N, i.e., the following diagram commutes:

$$\begin{array}{ccc} E & \xrightarrow{\ \Phi\ } & F \\ {\scriptstyle \pi_M}\downarrow & & \downarrow{\scriptstyle \pi_N} \\ M & \xrightarrow[\ \varphi\ ]{} & N \end{array}$$

We say that Φ *covers* φ. If Φ is invertible, it is called a *vector bundle isomorphism.*

All smooth vector bundles together with their homomorphisms form a category $\mathcal{VB}$.

If (E, π, M) is a vector bundle which admits a vector bundle atlas $(U_\alpha, \phi_\alpha)_{\alpha \in A}$ with the given open cover, then, we have $\phi_\alpha \circ \phi_\beta^{-1}(m, v) = \big(m, \phi_{\alpha\beta}(m)v\big)$ for $C^\infty-$transition functions $\phi_{\alpha\beta} : U_{\alpha\beta} = U_\alpha \cap U_\beta \to GL(V)$ (where we have

fixed a standard fibre V). This family of transition maps satisfies the *cocycle condition*

$$\begin{cases} \phi_{\alpha\beta}(m)\cdot\phi_{\beta\gamma}(m)=\phi_{\alpha\gamma}(m) & \text{for each } m\in U_{\alpha\beta\gamma}=U_\alpha\cap U_\beta\cap U_\gamma,\\ \phi_{\alpha\alpha}(m)=e & \text{for all } m\in U_\alpha. \end{cases}$$

The family $(\phi_{\alpha\beta})$ is called the *cocycle* of transition maps for the vector bundle atlas (U_α,ϕ_α).

Now, let us suppose that the same vector bundle (E,π,M) is described by an equivalent vector bundle atlas $(U_\alpha,\psi_\alpha)_{\alpha\in A}$ with the same open cover (U_α). Then the vector bundle charts (U_α,ϕ_α) and (U_α,ψ_α) are compatible for each α, so $\psi_\alpha\circ\phi_\beta^{-1}(m,v)=(m,\tau_\alpha(m)v)$ for some $\tau_\alpha: U_\alpha\to GL(V)$. We get

$$\tau_\alpha(m)\,\phi_{\alpha\beta}(m)=\phi_{\alpha\beta}(m)\,\tau_\beta(m)\quad\text{for all } m\in U_{\alpha\beta},$$

and we say that the two cocycles $(\phi_{\alpha\beta})$ and $(\psi_{\alpha\beta})$ of transition maps over the cover (U_α) are *cohomologous*. If $GL(V)$ is an Abelian group, i.e., if the standard fibre V is of real or complex dimension 1, then the cohomology classes of cocycles $(\phi_{\alpha\beta})$ over the open cover (U_α) form a usual cohomology group $H^1(M,GL(V))$ with coefficients in the sheaf $GL(V)$ [KMS93].

Let (E,π,M) be a vector bundle and let $\varphi: N\to M$ be a smooth map between the base manifolds N and M. Then there exists the *pull–back vector bundle* $(\varphi^*E,\varphi^*\pi,\varphi^*N)$ with the same typical fibre and a vector bundle homomorphism, given by the commutative diagram [KMS93]:

$$\begin{array}{ccc} \varphi^*E & \xrightarrow{\ \pi^*\varphi\ } & E \\ {\scriptstyle\varphi^*\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ N & \xrightarrow[\ \varphi\ ]{} & M \end{array}$$

The vector bundle $(\varphi^*E,\varphi^*\pi,\varphi^*N)$ is constructed as follows. Let $E=VB(\phi_{\alpha\beta})$ denote that E is described by a cocycle $(\phi_{\alpha\beta})$ of transition maps over an open cover (U_α) of M. Then $(\phi_{\alpha\beta}\circ\varphi)$ is a cocycle of transition maps over the open cover $\big(\varphi^{-1}(U_\alpha)\big)$ of N and the bundle is given by $\varphi^*E=VB(\phi_{\alpha\beta}\circ\varphi)$.

In other words, a vector bundle is a fibre bundle which admits an atlas of linear bundle coordinates. Typical fibres of a smooth vector bundle $\pi: Y\to X$ are *vector spaces* of some finite dimension (called the fibre dimension, $\mathrm{fdim}\,Y$ of Y), and Y admits a bundle atlas Ψ_Y (1.118) where trivialization maps $\psi_\xi(x)$ and transition functions $\rho_{\xi\zeta}(x)$ are linear isomorphisms of vector spaces. The corresponding bundle coordinates (y^i) obey a *linear* coordinate transformation law

$$y'^i=\rho^i_j(x)y^j.$$

We have the decomposition $y=y^ie_i(\pi(y))$, where

$$\{e_i(x)\}=\psi_\xi^{-1}(x)\{v_i\},\qquad x=\pi(y)\in U_\xi,$$

are fibre bases (or frames) for fibres Y_x of Y and $\{v_i\}$ is a fixed basis for the typical fibre V of Y.

There are several standard constructions of new vector bundles from old ones:

- Given two vector bundles Y and Y' over the same base X, their *Whitney sum* $Y \oplus Y'$ is a vector bundle over X whose fibres are the direct sums of those of the vector bundles Y and Y'.
- Given two vector bundles Y and Y' over the same base X, their tensor product $Y \otimes Y'$ is a vector bundle over X whose fibres are the tensor products of those of the vector bundles Y and Y'. In a similar way the exterior product $Y \wedge Y$ of vector bundles is defined, so that the *exterior bundle* of Y is defined as

$$\wedge Y = X \times \mathbb{R} \oplus Y \oplus \wedge^2 Y \oplus \cdots \oplus \wedge^m Y, \qquad (m = \operatorname{fdim} Y).$$

- Let $Y \to X$ be a vector bundle. By $Y^* \to X$ is denoted the *dual vector bundle* whose fibres are the duals of those of Y. The *interior product* (or *contraction*) of Y and Y^* is defined as a bundle map

$$\rfloor : Y \otimes Y^* \to X \times \mathbb{R}.$$

Given a linear bundle map $\Phi : Y' \to Y$ of vector bundles over X, its kernel $\operatorname{Ker} \Phi$ is defined as the inverse image $\Phi^{-1}(\widehat{0}(X))$ of the canonical zero section $\widehat{0}(X)$ of Y. If Φ is of constant rank, its kernel $\operatorname{Ker} \Phi$ and its image $\operatorname{Im} \Phi$ are subbundles of the vector bundles Y' and Y, respectively. For example, monomorphisms and epimorphisms of vector bundles fulfil this condition. If Y' is a subbundle of the vector bundle $Y \to X$, the factor bundle Y/Y' over X is defined as a vector bundle whose fibres are the quotients ${Y_x}/{Y'_x}$, $x \in X$.

Consider the *short exact sequence* of vector bundles over X,

$$0 \to Y' \xrightarrow{\ i\ } Y \xrightarrow{\ j\ } Y'' \to 0, \tag{1.120}$$

which means that i is a bundle monomorphism, j is a bundle epimorphism, and $\operatorname{Ker} j = \operatorname{Im} i$. Then Y'' is the factor bundle Y/Y'. One says that the short exact sequence (1.120) admits a *splitting* if there exists a bundle monomorphism $s : Y'' \to Y$ such that $j \circ s = \mathrm{Id}_{Y''}$, i.e.,

$$Y = i(Y') \oplus s(Y'') \cong Y' \oplus Y''.$$

Vector bundles of rank 1 are called *line bundles*.

The only two vector bundles with base space B a circle and 1D fibre F are the Mőbius band and the annulus, but the classification of all the different vector bundles over a given base space with fibre of a given dimension is quite difficult in general. For example, when the base space is a high–dimensional sphere and the dimension of the fibre is at least three, then the classification is of the same order of difficulty as the fundamental but still largely unsolved problem of computing the homotopy groups of spheres [Hat02].

Now, there is a natural direct sum operation for vector bundles over a fixed base space X, which in each fibre reduces just to direct sum of vector spaces. Using this, one can get a weaker notion of isomorphism of vector bundles by defining two vector bundles over the same base space X to be stably isomorphic if they become isomorphic after direct sum with product vector bundles $X \times \mathbb{R}^n$ for some n, perhaps different n's for the two given vector bundles. Then it turns out that the set of stable isomorphism classes of vector bundles over X forms an Abelian group under the direct sum operation, at least if X is compact Hausdorff. The traditional notation for this group is $\widetilde{KO}(X)$. In the case of spheres the groups $\widetilde{KO}(S^n)$ have the quite unexpected property of being periodic in n. This is called *Bott periodicity*, and the values of $\widetilde{KO}(S^n)$ are given by the following table [Hat02]:

$n \bmod 8$	1	2	3	4	5	6	7	8
$\widetilde{KO}(S^n)$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	0	$\mathbb{Z}$	0	0	0	$\mathbb{Z}$

For example, $\widetilde{KO}(S^1)$ is $\mathbb{Z}_2$, a cyclic group of order two, and a generator for this group is the Mőbius bundle. This has order two since the direct sum of two copies of the Mőbius bundle is the product $S^1 \times \mathbb{R}^1$, as one can see by embedding two Mőbius bands in a solid torus so that they intersect orthogonally along the common core circle of both bands, which is also the core circle of the solid torus.

The complex version of $\widetilde{KO}(X)$, called $\widetilde{K}(X)$, is constructed in the same way as $\widetilde{KO}(X)$ but using vector bundles whose fibers are vector spaces over $\mathbb{C}$ rather than $\mathbb{R}$. The complex form of Bott Periodicity asserts that $\widetilde{K}(S^n)$ is $\mathbb{Z}$ for n even and 0 for n odd, so the period is two rather than eight.

The groups $\widetilde{K}(X)$ and $\widetilde{KO}(X)$ for varying X share certain formal properties with the cohomology groups studied in classical algebraic topology. Using a more general form of Bott periodicity, it is in fact possible to extend the groups $\widetilde{K}(X)$ and $\widetilde{KO}(X)$ to a full cohomology theory, families of Abelian groups $\widetilde{K^n}(X)$ and $\widetilde{KO}^n(X)$ for $n \in \mathbb{Z}$ that are periodic in n of period two and eight, respectively. However, there is more algebraic structure here than just the additive group structure. Namely, tensor products of vector spaces give rise to tensor products of vector bundles, which in turn give product operations in both real and complex K–theory similar to cup product in ordinary cohomology. Furthermore, exterior powers of vector spaces give natural operations within K–theory [Hat02].

The Second Vector Bundle of the Manifold M

Let (E, π, M) be a vector bundle over the biomechanical manifold M with fibre addition $+_E : E \times_M E \to E$ and fibre scalar multiplication $m_t^E : E \to E$. Then (TE, π_E, E), the tangent bundle of the manifold E, is itself a vector bundle, with fibre addition denoted by $+_{TE}$ and scalar multiplication denoted by m_t^{TE}. The *second vector bundle* structure on $(TE, T\pi, TM)$, is the 'derivative' of the

original one on (E, π, M). In particular, the space $\{\Xi \in TE : T\pi.\Xi = 0 \in TM\} = (Tp)^{-1}(0)$ is denoted by VE and is called the *vertical bundle* over E. Its main characteristics are *vertical lift* and *vertical projection* (see [KMS93] for details).

All of this is valid for the *second tangent bundle* $T^2M = TTM$ of a manifold, but here we have one more natural structure at our disposal. The *canonical flip* or *involution* $\kappa_M : T^2M \to T^2M$ is defined locally by

$$(T^2\phi \circ \kappa_M \circ T^2\phi^{-1})(x, \xi; \eta, \zeta) = (x, \eta; \xi, \zeta).$$

where (U, ϕ) is a local chart on M (this definition is invariant under changes of charts). The flip κ_M has the following properties (see [KMS93]):

1. $\kappa_M \circ T^2f = T^2f \circ \kappa_M$ for each $f \in C^\infty(M, N)$;
2. $T(\pi_M) \circ \kappa_M = \pi_{TM}$;
3. $\pi_{TM} \circ \kappa_M = T(\pi_M)$;
4. $\kappa_M^{-1} = \kappa_M$;
5. κ_M is a linear isomorphism from the bundle $(TTM, T(\pi_M), TM)$ to (TTM, π_{TM}, TM), so it interchanges the two vector bundle structures on TTM;
6. κ_M is the unique smooth map $TTM \to TTM$ which, for each $\gamma : \mathbb{R} \to M$, satisfies
$$\partial_t\partial_s\gamma(t, s) = \kappa_M\partial_t\partial_s\gamma(t, s).$$

In a similar way the second cotangent bundle of a manifold M can be defined. Even more, for every manifold there is a *geometrical isomorphism* between the bundles $TT^*M = T(T^*M)$ and $T^*TM = T^*(TM)$ [MS78].

The Natural Vector Bundle

In this subsection we mainly follow [Mic01, KMS93].

A *vector bundle functor* or *natural vector bundle* is a functor $\mathcal{F}$ which associates a vector bundle $(\mathcal{F}(M), \pi_M, M)$ to each $n-$manifold M and a vector bundle homomorphism

$$\begin{array}{ccc} \mathcal{F}(M) & \xrightarrow{\mathcal{F}(\varphi)} & \mathcal{F}(N) \\ \downarrow{\scriptstyle \pi_M} & & \downarrow{\scriptstyle \pi_N} \\ M & \xrightarrow[\varphi]{} & N \end{array}$$

to each $\varphi : M \to N$ in $\mathcal{M}$, which covers φ and is fiberwise a linear isomorphism. Two common examples of the vector bundle functor $\mathcal{F}$ are tangent bundle functor T and cotangent bundle functor T^* (see subsection 1.2.4).

The space of all smooth sections of the vector bundle (E, π_M, M) is denoted by $\Gamma(E, \pi_M, M)$. Clearly, it is a vector space with fiberwise addition and scalar multiplication.

Let $\mathcal{F}$ be a vector bundle functor on $\mathcal{M}$. Let M be a smooth manifold and let $X \in \mathcal{X}(M)$ be a vector–field on M. Then the flow F_t of X for fixed t, is a diffeomorphism defined on an open subset of M. The map

$$\begin{array}{ccc} \mathcal{F}(M) & \xrightarrow{\mathcal{F}(F_t)} & \mathcal{F}(M) \\ \Big\downarrow{\scriptstyle \pi_M} & & \Big\downarrow{\scriptstyle \pi_M} \\ M & \xrightarrow[F_t]{} & M \end{array}$$

is then a vector bundle isomorphism, defined over an open subset of M.

We consider a tensor–field τ (1.2.5), which is a section $\tau \in \Gamma(\mathcal{F}(M))$ of the vector bundle $(\mathcal{F}(M), \pi_M, M)$ and we define for $t \in \mathbb{R}$

$$F_t^* \tau = \mathcal{F}(F_{-t}) \circ \tau \circ F_t,$$

a local section of the bundle $\mathcal{F}(M)$. For each point $m \in M$ the value $F_t^*\tau(x) \in \mathcal{F}(M)_m$ is defined, if t is small enough (depending on x). So, in the vector space $\mathcal{F}(M)_m$ the expression $\frac{d}{dt} |_{t=0}\, F_t^*\tau(x)$ makes sense and therefore the section

$$\mathcal{L}_X \tau = \frac{d}{dt} |_{t=0}\, F_t^* \tau$$

is globally defined and is an element of $\Gamma(\mathcal{F}(M))$. It is called the *Lie derivative of the tensor–field τ along a vector–field $X \in \mathcal{X}(M)$* (see subsection 1.2.6, for details on Lie derivative).

In this situation we have:

1. $F_t^* F_r^* \tau = F_{t+r}^* \tau$, whenever defined.
2. $\frac{d}{dt} F_t^* \tau = F_t^* \mathcal{L}_X \tau = \mathcal{L}_X (F_t^* \tau)$, so
   $[\mathcal{L}_X, F_t^*] = \mathcal{L}_X \circ F_t^* - F_t^* \circ \mathcal{L}_X = 0$, whenever defined.
3. $F_t^* \tau = \tau$ for all relevant t iff $\mathcal{L}_X \tau = 0$.

Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be two vector bundle functors on $\mathcal{M}$. Then the (fiberwise) tensor product $(\mathcal{F}_1 \otimes \mathcal{F}_2)(M) = \mathcal{F}_1(M) \otimes \mathcal{F}_2(M)$ is again a vector bundle functor and for $\tau_i \in \Gamma(\mathcal{F}_i(M))$ with $i = 1, 2$, there is a section $\tau_1 \otimes \tau_2 \in \Gamma(\mathcal{F}_1 \otimes \mathcal{F}_2)(M)$, given by the pointwise tensor product.

Also in this situation, for $X \in \mathcal{X}(M)$ we have

$$\mathcal{L}_X\, (\tau_1 \otimes \tau_2) = \mathcal{L}_X\, \tau_1 \otimes \tau_2 + \tau_1 \otimes \mathcal{L}_X\, \tau_2.$$

In particular, for $f \in C^\infty(M, \mathbb{R})$ we have $\mathcal{L}_X\, (f\, \tau) = df(X)\, \tau + f\, \mathcal{L}_X\, \tau$.

For any vector bundle functor $\mathcal{F}$ on $\mathcal{M}$ and $X, Y \in \mathcal{X}(M)$ we have:

$$[\mathcal{L}_X, \mathcal{L}_Y] = \mathcal{L}_X \circ \mathcal{L}_Y - \mathcal{L}_Y \circ \mathcal{L}_X = \mathcal{L}_{[X,Y]} : \Gamma(\mathcal{F}(M)) \to \Gamma(\mathcal{F}(M)).$$

Vertical Tangent and Cotangent Bundles

Tangent and Cotangent Bundles Revisited

Recall (from subsection 1.2.4 above) that the most important vector bundles are familiar tangent and cotangent bundles. The fibres of the tangent bundle $\pi_M : TM \to M$ of a manifold M are tangent spaces to M. The peculiarity of the tangent bundle TM in comparison with other vector bundles over M lies in the fact that, given an atlas $\Psi_M = \{(U_\xi, \phi_\xi)\}$ of a manifold M, the tangent bundle of M admits the holonomic atlas $\Psi = \{(U_\xi, \psi_\xi = T\phi_\xi)\}$, where by $T\phi_\xi$ is denoted the tangent map to ϕ_ξ. Namely, given coordinates x^α on a manifold M, the associated bundle coordinates on TM are holonomic coordinates $(\dot{x}^\alpha)$ with respect to the holonomic frames $\{\partial_\alpha\}$ for tangent spaces T_xM, $x \in M$. Their transition functions read

$$\dot{x}'^\alpha = \frac{\partial x'^\alpha}{\partial x^\mu}\dot{x}^\mu.$$

Every manifold map $f : M \to M'$ induces the linear bundle map over f of the tangent bundles (1.112).

The cotangent bundle of a manifold M is the dual $\pi_{*M} : T^*M \to M$ of the tangent bundle $TM \to M$. It is equipped with the holonomic coordinates $(x^\alpha, \dot{x}_\alpha)$ with respect to the coframes $\{dx^\alpha\}$ for T^*M which are the duals of $\{\partial_\alpha\}$. Their transition functions read

$$\dot{x}'_\alpha = \frac{\partial x^\mu}{\partial x'^\alpha}\dot{x}_\mu.$$

Recall that a tensor product of tangent and cotangent bundles over M,

$$T = (\otimes^m TM) \otimes (\otimes^k T^*M), \qquad (m, k \in \mathbb{N}), \tag{1.121}$$

is called a *tensor bundle*. Given two vector bundles Y and Y' over the same base X, their tensor product $Y \otimes Y'$ is a vector bundle over X whose fibres are the tensor products of those of the vector bundles Y and Y'.

Tangent, cotangent and tensor bundles belong to the category BUN of natural fibre bundles which admit the canonical lift of any diffeomorphism f of a base to a bundle automorphism, called the natural automorphism [KMS93]. For example, the natural automorphism of the tangent bundle TM over a diffeomorphism f of its base M is the tangent map Tf (1.112) over f. In view of the expression (1.112), natural automorphisms are also called holonomic transformations or general covariant transformations (in gravitation theory).

Let $TY \to Y$ be the tangent bundle of a bundle $Y \to X$. The following diagram commutes

$$\begin{array}{ccc} TY & \xrightarrow{T\pi} & TX \\ {\scriptstyle \pi_Y}\downarrow & & \downarrow{\scriptstyle \pi_X} \\ Y & \xrightarrow[\pi]{} & X \end{array}$$

where $T\pi : TY \to TX$ is a fibre bundle. Note that $T\pi$ is still the bundle map of the bundle $TY \to Y$ to TX over π and the fibred map of the bundle $TY \to X$ to TX over X. There is also the canonical surjection

$$\pi_T : TY \longrightarrow TX \longrightarrow Y, \qquad \text{given by} \qquad \pi_T = \pi_X \circ T\pi = \pi \circ \pi_Y.$$

Now, given the fibre coordinates (x^α, y^i) of a fibre bundle Y, the corresponding induced coordinates of TY are

$$(x^\alpha, y^i, \dot{x}^\alpha, \dot{y}^i), \qquad \dot{y}'^i = \frac{\partial y'^i}{\partial y^j}\dot{y}^j.$$

This expression shows that the tangent bundle $TY \to Y$ of a fibre bundle Y has the vector subbundle

$$VY = \operatorname{Ker} T\pi$$

where $T\pi$ is regarded as the fibred map of $TY \to X$ to TX over X. The subbundle VY consists of tangent vectors to fibres of Y. It is called the *vertical tangent bundle* of Y and provided with the induced coordinates $(x^\alpha, y^i, \dot{y}^i)$ with respect to the fibre bases $\{\partial_i\}$.

The *vertical cotangent bundle* $V^*Y \to Y$ of a fibre bundle $Y \to X$ is defined as the dual of the vertical tangent bundle $VY \to Y$. Note that it is not a subbundle of the cotangent bundle T^*Y, but there is the canonical surjection

$$\zeta : T^*Y \longrightarrow V^*Y, \qquad \dot{x}_\alpha dx^\alpha + \dot{y}_i dy^i \mapsto \dot{y}_i \overline{d}y^i, \tag{1.122}$$

where $\{\overline{d}y^i\}$ are the bases for the fibres of V^*Y which are duals of the holonomic frames $\{\partial_i\}$ for the vertical tangent bundle VY.

With VY and V^*Y, we have the following short exact sequences of vector bundles over a fibre bundle $Y \to X$:

$$0 \to VY \hookrightarrow TY \to Y \times TX \to 0, \tag{1.123}$$

$$0 \to Y \times T^*X \hookrightarrow T^*Y \to V^*Y \to 0 \tag{1.124}$$

Every splitting

$$Y \times TX \hookrightarrow TY, \qquad \partial_\alpha \mapsto \partial_\alpha + \Gamma^i_\alpha(y)\partial_i,$$

of the exact sequence (1.123) and

$$V^*Y \to T^*Y, \qquad \overline{d}^i \mapsto dy^i - \Gamma^i_\alpha(y)dx^\alpha,$$

of the exact sequence (1.124), by definition, corresponds to a certain *connection* on the bundle $Y \to X$, and *vice versa*.

Let Φ be a fibred map of a bundle $Y \to X$ to a bundle $Y' \to X'$ over $f : X \to X'$. The tangent map $T\Phi : TY \to TY'$ to Φ reads

$$(\dot{x}'^\alpha, \dot{y}'^i) \circ T\Phi = (\partial_\mu f^\alpha \dot{x}^\mu, \partial_\mu \Phi^i \dot{x}^\mu + \partial_j \Phi^i \dot{y}^j). \tag{1.125}$$

It is both the linear bundle map over Φ, given by the commutativity diagram

$$\begin{array}{ccc} TY & \xrightarrow{T\Phi} & TY' \\ {\scriptstyle \pi_Y}\downarrow & & \downarrow{\scriptstyle \pi_{Y'}} \\ Y & \xrightarrow[\Phi]{} & Y' \end{array}$$

as well as the fibred map over the tangent map Tf to f, given by the commutativity diagram

$$\begin{array}{ccc} TY & \xrightarrow{T\Phi} & TY' \\ \downarrow & & \downarrow \\ TX & \xrightarrow[Tf]{} & TX' \end{array}$$

Affine Bundles

Given a vector bundle $\overline{Y} \to X$, an *affine bundle* modelled over $\overline{Y}$ is a fibre bundle $Y \to X$ whose fibres Y_x, (for all $x \in X$), are affine spaces modelled over the corresponding fibres $\overline{Y}_x$ of the vector bundle $\overline{Y}$, and Y admits a bundle atlas Ψ_Y (1.118) whose trivialization morphisms $\psi_\xi(x)$ and transition functions functions $\rho_{\xi\zeta}(x)$ are affine maps. The corresponding bundle coordinates (y^i) possess an affine coordinate transformation law

$$y'^i = \rho^i_j(x^\alpha)y^j + \rho^i(x^\alpha).$$

In other words, an affine bundle admits an atlas of affine bundle coordinates (x^α, y^i) such that

$$r : (x^\alpha, y^i) \times (x^\alpha, \overline{y}^i) \mapsto (x^\alpha, y^i + \overline{y}^i)$$

where $(x^\alpha, \overline{y}^i)$ are linear bundle coordinates of the vector bundle $\overline{Y}$. In particular, every vector bundle Y has the canonical structure of an affine bundle modelled on Y itself by the map

$$r : (y, y') \mapsto y + y'.$$

Every affine bundle has a global section.

One can define a direct sum $Y \oplus Y'$ of a vector bundle $\overline{Y}' \to X$ and an affine bundle $Y \to X$ modelled over a vector bundle $Y \to X$, as is an affine bundle modelled over the Whitney sum of vector bundles $Y' \oplus \overline{Y}$.

Affine bundles are subject to *affine bundle maps* which are affine fibrewise maps. Any affine bundle map $\Phi : Y \to Y'$ from an affine bundle Y modelled over a vector bundle $\overline{Y}$ to an affine bundle Y' modelled over a vector bundle $\overline{Y}'$, induces the linear bundle map of these vector bundles

$$\overline{\Phi} : \overline{Y} \to \overline{Y}', \qquad \overline{y}'^i \circ \overline{\Phi} = \frac{\partial \Phi^i}{\partial y^j}\overline{y}^j. \tag{1.126}$$

Distributions and Foliations on Manifolds

Let M be an nD smooth manifold. A *smooth distribution* $\mathcal{T}$ of codimension k on M is defined as a subbundle of rank $n-k$ of the tangent bundle TM. A smooth distribution $\mathcal{T}$ is called the *involutive distribution* if $[u, u']$ is a section of $\mathcal{T}$ whenever u and u' are sections of $\mathcal{T}$.

Let $\mathcal{T}$ be a $k-$codimensional distribution on M. Its annihilator $\mathcal{T}^*$ is a kD subbundle of T^*M called the *Pfaffian system*. It means that, on a neighborhood U of every point $x \in M$, there exist k linearly independent sections $s_1, \ldots, s_k$ of $\mathcal{T}^*$ such that

$$\mathcal{T}_x \mid_U = \cap_j Ker\, s_j.$$

Let $\mathcal{C}(\mathcal{T})$ be the ideal of $\wedge(M)$ generated by sections of $\mathcal{T}^*$.

A smooth distribution $\mathcal{T}$ is involutive iff the *ideal* $\mathcal{C}(\mathcal{T})$ is *differential*, that is, $d\mathcal{C}(\mathcal{T}) \subset \mathcal{C}(\mathcal{T})$.

Given an involutive $k-$codimensional distribution $\mathcal{T}$ on M, the quotient $TM/\mathcal{T}$ is a kD vector bundle called the *transversal bundle* of $\mathcal{T}$. There is the exact sequence

$$0 \to \mathcal{T} \hookrightarrow TM \to TM/\mathcal{T} \to 0. \tag{1.127}$$

Given a bundle $Y \to X$, its vertical tangent bundle VY exemplifies an involutive distribution on Y.

A submanifold N of M is called the *integral manifold* of a distribution $\mathcal{T}$ on M if the tangent spaces to N coincide with the fibres of this distribution at each point of N.

Let $\mathcal{T}$ be a smooth involutive distribution on M. For any point $x \in M$, there exists a maximal integral manifold of $\mathcal{T}$ passing through x [KT75]. In view of this fact, involutive distributions are also called *completely integrable distributions*.

Every point $x \in M$ has an open neighborhood U which is a domain of a coordinate chart $(x^1, \ldots, x^n)$ such that the restrictions of $\mathcal{T}$ and $\mathcal{T}^*$ to U are generated by the $n-k$ vector fields $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^{n-k}}$ and the k Pfaffian forms $dx^{n-k+1}, \ldots, dx^n$ respectively.

In particular, it follows that integral manifolds of an involutive distribution constitute a *foliation*. Recall that a $k-$codimensional foliation on an nD manifold M is a partition of M into *connected leaves* F_ι with the following property: every point of M has an open neighborhood U which is a domain of a coordinate chart (x^α) such that, for every leaf F_ι, the components $F_\iota \cap U$ are described by the equations $x^{n-k+1} = \text{const}, \ldots, x^n = \text{const}$ [KT75]. Note that leaves of a foliation fail to be imbedded submanifolds in general.

For example, every projection $\pi : M \to X$ defines a foliation whose leaves are the fibres $\pi^{-1}(x)$, for all $x \in X$. Also, every nowhere vanishing vector field u on a manifold M defines a 1D involutive distribution on M. Its integral manifolds are the integral curves of u. Around each point $x \in M$, there exist local coordinates $(x^1, \ldots, x^n)$ of a neighborhood of x such that u is given by $u = \frac{\partial}{\partial x^i}$.

1.3.4 Principal Bundles

Recall that a *principal bundle* is a special case of a fibre bundle where the fibre is a group G. More specifically, G is usually a Lie group. A principal bundle is a total bundle space Y along with a surjective map $\pi : Y \to X$ to a base manifold X. Any fibre $\pi^{-1}(x)$ is a space isomorphic to G. More specifically, G acts freely and transitively without fixed point on the fibers, and this makes a fibre into a homogeneous space. It follows that the orbits of the $G-$action are precisely the fibers of $\pi : Y \to X$ and the orbit space Y/G is homeomorphic the base space X. To say that G acts freely and transitively on the fibers means that the fibers take on the structure of $G-$torsors.[11]

For example, in the case of a *circle bundle* ($G = S^1 \equiv \{e^{it}\}$), the fibers are circles, which can be rotated, although no point in particular corresponds to the identity. Near every point, the fibers can be given the group structure of G in the fibers over a neighborhood by choosing an element in each fibre to be the identity element. However, the fibers cannot be given a group structure globally, except in the case of a trivial bundle.

An important principal bundle is the *frame bundle* on a Riemannian manifold. This bundle reflects the different ways to give an orthonormal basis for tangent vectors.

In general, any fibre bundle corresponds to a principal bundle where the group (of the principal bundle) is the group of isomorphisms of the fibre (of the fibre bundle). Given a principal bundle $\pi : Y \to X$ and an action of G on a space V, which could be a group representation, this can be reversed to give an *associated* fibre bundle.

A trivialization of a principal bundle, an open set U in X such that the bundle $\pi^{-1}(U)$ over U, is expressed as $U \times G$, has the property that the group G acts on the left and transition functions take values in G, acting on the fibers by right multiplication (so that the action of G on a fibre V is independent of coordinate chart).

More precisely, a principal bundle $\pi_P : P \to Q$ of a configuration manifold Q, with a *structure Lie group* G, is a general affine bundle modelled on the right on the trivial group bundle $Q \times G$ where the group G acts freely and transitively on P on the right,

$$R_G : P \times G \to P, \qquad R_g : p \mapsto pg, \qquad (p \in P,\ g \in G). \tag{1.128}$$

We call P a *principal $G-$bundle*. A typical fibre of a principal $G-$bundle is isomorphic to the group space of G, and $P/G = Q$. The structure group G acts on the typical fibre by left multiplications which do not preserve the group structure of G. Therefore, the typical fibre of a principal bundle is only a group space, but not a group. Since the left action of transition functions on the typical fibre G commutes with its right multiplications, a principal bundle admits the global right action (1.128) of the structure group.

[11] A $G-$torsor is a space which is homeomorphic to G but lacks a group structure since there is no preferred choice of an identity element.

A principal $G-$bundle P is equipped with a bundle atlas

$$\Psi_P = \{(U_\alpha, \psi^P_\alpha, \rho_{\alpha\beta})\}, \tag{1.129}$$

whose trivialization maps

$$\psi^P_\alpha : \pi_P^{-1}(U_\alpha) \to U_\alpha \times G$$

obey the *equivariance condition*

$$(\pi_2 \circ \psi^P_\alpha)(pg) = (\pi_2 \circ \psi^P_\alpha)(p)g, \qquad (g \in G,\ p \in \pi_P^{-1}(U_\alpha)). \tag{1.130}$$

Due to this property, every trivialization map ψ^P_α determines a unique local section z_α of P over U_α such that

$$\pi_2 \circ \psi^P_\alpha \circ z_\alpha = \mathbf{1},$$

where $\mathbf{1}$ is the unit element of G. The transformation rules for z_α read

$$z_\beta(q) = z_\alpha(q)\rho_{\alpha\beta}(q), \qquad (q \in U_\alpha \cap U_\beta), \tag{1.131}$$

where $\rho_{\alpha\beta}(q)$ are $G-$valued transition functions of the atlas Ψ_P. Conversely, the family $\{(U_\alpha, z_\alpha)\}$ of local sections of P with the transition functions (1.131) determines a unique bundle atlas of P. In particular, it follows that only trivial principal bundles have global sections.

Note that the pull–back of a principal bundle is also a principal bundle with the same structure group.

The quotient of the tangent bundle $TP \to P$ and that of the vertical tangent bundle VP of P by the tangent prolongation TR_G of the canonical action R_G (1.128) are vector bundles

$$T_GP = TP/G, \qquad V_GP = VP/G \tag{1.132}$$

over Q. Sections of $T_GP \to Q$ are naturally identified with G-invariant vector–fields on P, while those of $V_GP \to Q$ are G-invariant vertical vector–fields on P. Therefore, the Lie bracket of G-invariant vector–fields on P goes to the quotients (1.132), and induces the Lie brackets of their sections. Let us write these brackets in an explicit form.

Owing to the equivariance condition (1.130), any bundle atlas (1.129) of P induces the associated bundle atlases $\{U_\alpha, T\psi^P_\alpha/G)\}$ of T_GP and $\{U_\alpha, V\psi^P_\alpha/G)\}$ of V_GP. Given a basis $\{\varepsilon_p\}$ for the right Lie algebra $\mathfrak{g}_r$, let $\{\partial_\alpha, e_p\}$ and $\{e_p\}$, where $e_p = (\psi^P_\alpha/G)^{-1}(\varepsilon_p)$, be the corresponding local fibre bases for the vector bundles T_GP and V_GP, respectively. Relative to these bases, the Lie bracket of sections

$$\xi = \xi^\alpha\partial_\alpha + \xi^p e_p, \qquad \eta = \eta^\mu\partial_\mu + \eta^q e_q$$

of the vector bundle $T_GP \to Q$ reads

$$[\xi,\eta]=(\xi^\mu\partial_\mu\eta^\alpha-\eta^\mu\partial_\mu\xi^\alpha)\partial_\alpha+(\xi^\alpha\partial_\alpha\eta^r-\eta^\alpha\partial_\alpha\xi^r+c^r_{pq}\xi^p\eta^q)e_r. \quad (1.133)$$

Putting $\xi^\alpha=0$ and $\eta^\mu=0$, we get the Lie bracket

$$[\xi,\eta]=c^r_{pq}\xi^p\eta^q e_r \quad (1.134)$$

of sections of the vector bundle $V_G\to P$.

A principal bundle P is also the general affine bundle modelled on the left on the associated group bundle $\widetilde{P}$ with the standard fibre G on which the structure group G acts by the adjoint representation. The corresponding bundle map reads

$$\widetilde{P}\times P\longrightarrow P, \qquad (\widetilde{p},p)\mapsto \widetilde{p}p.$$

Note that the standard fibre of the group bundle $\widetilde{P}$ is the group G, while that of the principal bundle P is the group space of G on which the structure group G acts on the left.

A principal bundle $P\to Q$ with a structure Lie group G possesses the canonical trivial vertical splitting

$$\alpha: VP\to P\times\mathfrak{g}_l, \qquad \pi_2\circ\alpha\circ e_m=J_m,$$

where $\{J_m\}$ is a basis for the left Lie algebra $\mathfrak{g}_l$ and e_m denotes the corresponding fundamental vector–fields on P. Given a principal bundle $P\to Q$, the bundle $TP\to TQ$ is a principal bundle

$$TP\times T(Q\times G)\to TP$$

with the structure group $TG=G\times\mathfrak{g}_l$ where $\mathfrak{g}_l$ is the left Lie algebra of left–invariant vector–fields on the group G.

If $P\to Q$ is a principal bundle with a structure group G, the exact sequence (1.123) can be reduced to the exact sequence

$$0\to V^GP\hookrightarrow T^GP\to TQ\to 0, \quad (1.135)$$
$$\text{where} \qquad T^GP=TP/G, \qquad V^GP=VP/G$$

are the quotients of the tangent bundle TP of P and the vertical tangent bundle VP of P respectively by the canonical action (1.128) of G on P on the right. The bundle $V^GP\to Q$ is called the adjoint bundle. Its standard fibre is the right Lie algebra $\mathfrak{g}_r$ of the right–invariant vector–fields on the group G. The group G acts on this standard fibre by the adjoint representation.

1.3.5 Multivector–Fields and Tangent–Valued Forms

Recall that a vector–field on a manifold M is defined as a global section of the tangent bundle $TM\to M$. The set $\mathcal{V}^1(M)$ of vector–fields on M is a real Lie algebra with respect to the *Lie bracket* [Sar93, Sar95, GMS97, MS00a, Sar02a]

$$[v,u] = (v^\alpha \partial_\alpha u^\mu - u^\alpha \partial_\alpha v^\mu)\partial_\mu, \qquad v = v^\alpha \partial_\alpha, \qquad u = u^\alpha \partial_\alpha. \tag{1.136}$$

Every vector–field on a manifold M can be seen as an infinitesimal generator of a local 1–parameter Lie group of diffeomorphisms of M as follows [KN63/9]. Given an open subset $U \subset M$ and an interval $(-\epsilon, \epsilon) \in \mathbb{R}$, by a local 1–parameter group of diffeomorphisms of M defined on $(-\epsilon, \epsilon) \times U$ is denoted a map

$$G \to M, \qquad (t,x) \mapsto G_t(x)$$

such that:

1. for each $t \in (-\epsilon, \epsilon)$, the map G_t is a diffeomorphism of U onto the open subset $G_t(U) \subset M$; and
2. $G_{t+t'}(x) = (G_t \circ G_{t'})(x)$ if $t, t', t+t' \in (-\epsilon, \epsilon)$ and $G_{t'}(x), x \in U$.

Any local 1–parameter group of diffeomorphisms G on $U \subset M$ defines a local vector–field u on U by setting $u(x)$ to be the tangent vector to the curve $x(t) = G_t(x)$ at $t = 0$. Conversely, if u is a vector–field on a manifold M, there exists a unique local 1–parameter group G_u of diffeomorphisms on a neighborhood of every point $x \in M$ which defines u. We call G_u a *flow* of the vector–field u. A vector–field u on a manifold M is called *complete* if its flow is a 1–parameter group of diffeomorphisms of M. In particular, every vector–field on a compact manifold is complete [KN63/9].

A vector–field u on a fibre bundle $Y \longrightarrow X$ is an infinitesimal generator of a local 1–parameter group G_u of isomorphisms of $Y \longrightarrow X$ iff it is a projectable vector–field on Y. A vector–field u on a fibre bundle $Y \longrightarrow X$ is called projectable if it projects onto a vector–field on X, i.e., there exists a vector–field τ on X such that the following diagram commutes:

$$\begin{array}{ccc} Y & \xrightarrow{\quad u \quad} & TY \\ {\scriptstyle \pi}\downarrow & & \downarrow{\scriptstyle T\pi} \\ X & \xrightarrow[\quad \tau \quad]{} & TX \end{array}$$

A projectable vector–field has the coordinate expression

$$u = u^\alpha(x^\mu)\partial_\alpha + u^i(x^\mu, y^j)\partial_i,$$

where u^α are local functions on X. A projectable vector–field is said to be vertical if it projects onto the zero vector–field $\tau = 0$ on X, i.e., $u = u^i \partial_i$ takes its values in the vertical tangent bundle VY.

For example, in field theory, projectable vector–fields on fibre bundles play a role of infinitesimal generators of local 1–parameter groups of gauge transformations.

In general, a vector–field $\tau = \tau^\alpha \partial_\alpha$ on a base X of a fibre bundle $Y \to X$ induces a vector–field on Y by means of a *connection* on this fibre bundle.

Nevertheless, every natural fibre bundle $Y \to X$ admits the *canonical lift* $\widetilde{\tau}$ onto Y of any vector–field τ on X. For example, if Y is the tensor bundle (1.121), the above canonical lift reads

$$\widetilde{\tau} = \tau^{\mu}\partial_{\mu} + [\partial_{\nu}\tau^{\alpha_1}\dot{x}^{\nu\alpha_2\cdots\alpha_m}_{\beta_1\cdots\beta_k} + \ldots - \partial_{\beta_1}\tau^{\nu}\dot{x}^{\alpha_1\cdots\alpha_m}_{\nu\beta_2\cdots\beta_k} - \ldots]\frac{\partial}{\partial\dot{x}^{\alpha_1\cdots\alpha_m}_{\beta_1\cdots\beta_k}}. \quad (1.137)$$

In particular, we have the canonical lift onto the tangent bundle TX,

$$\widetilde{\tau} = \tau^{\mu}\partial_{\mu} + \partial_{\nu}\tau^{\alpha}\dot{x}^{\nu}\frac{\partial}{\partial\dot{x}^{\alpha}} \quad (1.138)$$

and another one onto the cotangent bundle T^*X,

$$\widetilde{\tau} = \tau^{\mu}\partial_{\mu} - \partial_{\beta}\tau^{\nu}\dot{x}_{\nu}\frac{\partial}{\partial\dot{x}_{\beta}}. \quad (1.139)$$

A *multivector–field* ϑ of degree r (or simply a r-vector–field) on a manifold M, by definition, is a global section of the bundle $\wedge^r TM \to M$. It is given by the coordinate expression

$$\vartheta = \vartheta^{\alpha_1\ldots\alpha_r}\partial_{\alpha_1} \wedge \cdots \wedge \partial_{\alpha_r}, \qquad |\vartheta| = r,$$

where summation is over all ordered collections $(\lambda_1, ..., \lambda_r)$.

Similarly, an *exterior $r-$form* on a manifold M with local coordinates x^{α}, by definition, is a global section of the skew–symmetric tensor bundle (exterior product) $\wedge^r T^*M \to M$,

$$\phi = \frac{1}{r!}\phi_{\alpha_1\ldots\alpha_r}dx^{\alpha_1} \wedge \cdots \wedge dx^{\alpha_r}, \qquad |\phi| = r.$$

The 1–forms are also called the *Pfaffian forms*.

The vector space $\mathcal{V}^r(M)$ of $r-$vector–fields on a manifold M admits the *Schouten–Nijenhuis bracket* (or, SN bracket)

$$[.,.]_{SN} : \mathcal{V}^r(M)\times\mathcal{V}^s(M) \to \mathcal{V}^{r+s-1}(M)$$

which generalizes the Lie bracket of vector–fields (1.136). The SN–bracket has the coordinate expression:

$$\begin{aligned}
&\vartheta = \vartheta^{\alpha_1\ldots\alpha_r}\partial_{\alpha_1} \wedge \cdots \wedge \partial_{\alpha_r}, \qquad \upsilon = \upsilon^{\alpha_1\ldots\alpha_s}\partial_{\alpha_1} \wedge \cdots \wedge \partial_{\alpha_s},\\
&[\vartheta, \upsilon]_{SN} = \vartheta \star \upsilon + (-1)^{|\vartheta||\upsilon|}\upsilon \star \vartheta, \qquad \text{where}\\
&\vartheta \star \upsilon = \vartheta^{\mu\alpha_1\ldots\alpha_{r-1}}\partial_{\mu}\upsilon^{\alpha_1\ldots\alpha_s}\partial_{\alpha_1} \wedge \cdots \wedge \partial_{\alpha_{r-1}} \wedge \partial_{\alpha_1} \wedge \cdots \wedge \partial_{\alpha_s}.
\end{aligned}$$

The following relations hold for the SN–bracket:

$$\begin{aligned}
&[\vartheta, \upsilon]_{SN} = (-1)^{|\vartheta||\upsilon|}[\upsilon, \vartheta]_{SN},\\
&[\nu, \vartheta \wedge \upsilon]_{SN} = [\nu, \vartheta]_{SN} \wedge \upsilon + (-1)^{|\nu||\vartheta|+|\vartheta|}\vartheta \wedge [\nu, \upsilon]_{SN},\\
&(-1)^{|\nu||\vartheta|+|\nu|}[\nu, \vartheta \wedge \upsilon]_{SN} + (-1)^{|\vartheta||\nu|+|\vartheta|}[\vartheta, \upsilon \wedge \nu]_{SN}\\
&+ (-1)^{|\upsilon||\vartheta|+|\upsilon|}[\upsilon, \nu \wedge \vartheta]_{SN} = 0.
\end{aligned}$$

In particular, let $w = w^{\mu\nu}\partial_\mu \wedge \partial_\nu$ be a *bivector–field*. We have

$$[w, w]_{SN} = w^{\mu\alpha_1}\partial_\mu w^{\alpha_2\alpha_3}\partial_{\alpha_1} \wedge \partial_{\alpha_2} \wedge \partial_{\alpha_3}. \tag{1.140}$$

Every bivector–field w on a manifold M induces the 'sharp' bundle map $w^\sharp : T^*M \to TM$ defined by

$$w^\sharp(p)\rfloor q := w(x)(p, q), \qquad w^\sharp(p) = w^{\mu\nu}(x)p_\mu\partial_\nu, \qquad (p, q \in T^*_xM). \tag{1.141}$$

A bivector–field w whose bracket (1.140) vanishes is called the *Poisson bivector–field*.

Let $\wedge^r(M)$ denote the vector space of exterior $r-$forms on a manifold M. By definition, $\wedge^0(M) = C^\infty(M)$ is the ring of smooth real functions on M. All exterior forms on M constitute the $\mathbb{N}$-graded exterior algebra $\wedge^*(M)$ of global sections of the exterior bundle $\wedge T^*M$ with respect to the exterior product $\wedge$. This algebra admits the *exterior differential*

$$\begin{aligned} &d : \wedge^r(M) \to \wedge^{r+1}(M), \\ &d\phi = dx^\mu \wedge \partial_\mu\phi = \frac{1}{r!}\partial_\mu\phi_{\alpha_1\dots\alpha_r}dx^\mu \wedge dx^{\alpha_1} \wedge \cdots dx^{\alpha_r}, \end{aligned}$$

which is *nilpotent*, i.e., $d \circ d = 0$, and obeys the relation

$$d(\phi \wedge \sigma) = d(\phi) \wedge \sigma + (-1)^{|\phi|}\phi \wedge d(\sigma).$$

The *interior product* (or, *contraction*) of a vector–field $u = u^\mu\partial_\mu$ and an exterior $r-$form ϕ on a manifold M is given by the coordinate expression

$$\begin{aligned} u\rfloor\phi &= \sum_{k=1}^{r} \frac{(-1)^{k-1}}{r!}u^{\alpha_k}\phi_{\alpha_1\dots\alpha_k\dots\alpha_r}dx^{\alpha_1} \wedge \cdots \wedge \widehat{dx}^{\alpha_k} \wedge \cdots \wedge dx^{\alpha} \qquad (1.142) \\ &= \frac{1}{(r-1)!}u^\mu\phi_{\mu\alpha_2\dots\alpha_r}dx^{\alpha_2} \wedge \cdots \wedge dx^{\alpha_r}, \end{aligned}$$

where the caret $\widehat{\cdot}$ denotes omission. The following relations hold:

$$\phi(u_1, \dots, u_r) = u_r\rfloor \cdots u_1\rfloor\phi, \tag{1.143}$$

$$u\rfloor(\phi \wedge \sigma) = u\rfloor\phi \wedge \sigma + (-1)^{|\phi|}\phi \wedge u\rfloor\sigma, \tag{1.144}$$

$$[u, u']\rfloor\phi = u\rfloor d(u'\rfloor\phi) - u'\rfloor d(u\rfloor\phi) - u'\rfloor u\rfloor d\phi, \qquad (\phi \in \wedge^1(M)) \tag{1.145}$$

Recall from section 1.2.6 above, that the *Lie derivative* $\mathfrak{L}_u\sigma$ of an exterior form σ along a vector–field u is defined by the *Cartan relation*

$$\mathfrak{L}_u\sigma = u\rfloor d\sigma + d(u\rfloor\sigma).$$

It satisfies the relation

$$\mathfrak{L}_u(\phi \wedge \sigma) = \mathfrak{L}_u\phi \wedge \sigma + \phi \wedge \mathfrak{L}_u\sigma.$$

In particular, if f is a function, then

$$\mathfrak{L}_u f = u(f) = u\rfloor df.$$

It is important for dynamical applications that an exterior form ϕ is invariant under a local 1–parameter group of diffeomorphisms G_t of M (i.e., $G_t^*\phi = \phi$) iff its Lie derivative $\mathfrak{L}_u\phi$ along the vector–field u, generating G_t, vanishes.

Let Ω be a two–form on M. It defines the 'flat' bundle map $\Omega^\flat$, as

$$\Omega^\flat : TM \to T^*M, \qquad \Omega^\flat(v) = -v\rfloor\Omega(x), \qquad (v \in T_xM). \tag{1.146}$$

In coordinates, if $\Omega = \Omega_{\mu\nu}dx^\mu \wedge dx^\nu$ and $v = v^\mu\partial_\mu$, then

$$\Omega^\flat(v) = -\Omega_{\mu\nu}v^\mu dx^\nu.$$

One says that Ω is of *constant rank* k if the corresponding map (1.146) is of constant rank k (i.e., k is the greatest integer n such that Ω^n is not the zero form). The rank of a *nondegenerate* two–form is equal to $\dim M$. A nondegenerate closed two–form is called the *symplectic form*.

Given a manifold map $f : M \to M'$, any exterior k-form ϕ on M' induces the pull–back exterior form $f^*\phi$ on M by the condition

$$f^*\phi(v^1, \ldots, v^k)(x) = \phi(Tf(v^1), \ldots, Tf(v^k))(f(x))$$

for an arbitrary collection of tangent vectors $v^1, \cdots, v^k \in T_xM$. The following relations hold:

$$f^*(\phi \wedge \sigma) = f^*\phi \wedge f^*\sigma, \qquad df^*\phi = f^*(d\phi).$$

In particular, given a fibre bundle $\pi : Y \to X$, the pull–back onto Y of exterior forms on X by π gives the monomorphism of exterior algebras

$$\pi^* : \wedge^*(X) \to \wedge^*(Y).$$

Elements of its image $\pi^* \wedge^* (X)$ are called basic forms. Exterior forms on Y such that $u\rfloor\phi = 0$ for an arbitrary vertical vector–field u on Y are said to be *horizontal forms*. They are generated by horizontal 1–forms $\{dx^\alpha\}$. For example, basic forms are horizontal forms with coefficients in $C^\infty(X) \subset C^\infty(Y)$. A horizontal form of degree $n = \dim X$ is called a *density*. For example, Lagrangians in field theory are densities.

Elements of the tensor product $\wedge^r(M) \otimes \mathcal{V}^1(M)$ are called the *tangent–valued $r-$forms* on M. They are sections

$$\phi = \frac{1}{r!}\phi^\mu_{\alpha_1\ldots\alpha_r} dx^{\alpha_1} \wedge \cdots \wedge dx^{\alpha_r} \otimes \partial_\mu$$

of the tensor bundle

$$\wedge^r T^*M \otimes TM \to M.$$

Tangent-valued 1–forms are usually called the (1,1) *tensor fields*.

In particular, there is the 1–1 correspondence between the tangent–valued 1–forms on M and the linear bundle maps over M,

$$\phi : TM \to TM, \qquad \phi : T_xM \ni v \mapsto v \rfloor \phi(x) \in T_xM. \tag{1.147}$$

In particular, the *canonical tangent–valued one–form* $\theta_M = dx^\alpha \otimes \partial_\alpha$ defines the identity map of TM.

Tangent-valued forms play a prominent role in jet formalism and theory of connections on fibre bundles. In particular, tangent–valued 0-forms are vector–fields on M. Also, there is 1–1 correspondence between the tangent–valued 1–forms ϕ on a manifold M and the linear bundle endomorphisms

$$\widehat{\phi} : TM \to TM, \qquad \widehat{\phi} : T_xM \ni v \mapsto v \rfloor \phi(x) \in T_xM, \tag{1.148}$$

$$\widehat{\phi}^* : T^*M \to T^*M, \qquad \widehat{\phi}^* : T^*_xM \ni v^* \mapsto \phi(x) \rfloor v^* \in T^*_xM, \tag{1.149}$$

over M. For example, the *canonical tangent–valued 1–form* on M,

$$\theta_M = dx^\alpha \otimes \partial_\alpha \,, \tag{1.150}$$

corresponds to the identity maps (1.148) and (1.149).

We shall deal with the following particular types of vector–fields and differential forms on a bundle $Y \longrightarrow X$ [Sar93, Sar95, GMS97, MS00a]:

- a *projectable vector–field* on Y,

  $$u = u^\mu(x)\partial_\mu + u^i(y)\partial_i,$$

  which covers a vector–field $\tau_u = u^\mu(x)\partial_\mu$ on the base X such that the following diagram commutes:

  $$\begin{array}{ccc} Y & \xrightarrow{\quad u \quad} & TY \\ {\scriptstyle \pi}\downarrow & & \downarrow{\scriptstyle T\pi} \\ X & \xrightarrow[\quad \tau_u \quad]{} & TX \end{array}$$

- a *vertical vector–field*, $u : Y \to VY$, given by $u = u^i(y)\partial_i$, is a projectable vector–field which covers $\tau_u = 0$;
- an *exterior horizontal form*, $\phi : Y \to \wedge^r T^*X$, given by

  $$\phi = \frac{1}{r!}\phi_{\alpha_1 \ldots \alpha_r}(y) dx^{\alpha_1} \wedge \cdots \wedge dx^{\alpha_r};$$

- a *tangent–valued horizontal form*, $\phi : Y \to \wedge^r T^*X \otimes TY$, given by

  $$\phi = \frac{1}{r!} dx^{\alpha_1} \wedge \cdots \wedge dx^{\alpha_r} \otimes [\phi^\mu_{\alpha_1 \ldots \alpha_r}(y)\partial_\mu + \phi^i_{\alpha_1 \ldots \alpha_r}(y)\partial_i];$$

- a *vertical–valued horizontal form*, $\phi : Y \to \wedge^r T^*X \otimes VY$, given by

$$\phi = \frac{1}{r!}\phi^i_{\alpha_1 \dots \alpha_r}(y) dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_r} \otimes \partial_i.$$

- a vertical-valued *soldering form*, $\sigma : Y \to T^*X \otimes VY$, given by

$$\sigma = \sigma^i_\alpha(y) dx^\alpha \otimes \partial_i \tag{1.151}$$

and, in particular, the *canonical soldering form* on TX,

$$\theta_X = dx^\alpha \otimes \partial_\alpha.$$

The *pull–back–valued forms* on a bundle $Y \to X$ are the following two maps:[12]

$$\begin{aligned} &Y \to \wedge^r T^*Y \otimes TX, \qquad \phi = \frac{1}{r!}\phi^\mu_{\alpha_1 \dots \alpha_r}(y) dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_r} \otimes \partial_\mu, \\ &\qquad \text{and} \\ &Y \to \wedge^r T^*Y \otimes V^*X, \qquad \phi = \frac{1}{r!}\phi_{\alpha_1 \dots \alpha_r i}(y) dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_r} \otimes \overline{d}y^i. \end{aligned} \tag{1.152}$$

The pull–back-valued forms (1.152) are exemplified by the canonical bundle monomorphism

$$\wedge^n T^*X \otimes V^*Y \hookrightarrow \wedge^{n+1} T^*Y, \qquad \omega \otimes \overline{d}y^i \mapsto \omega \wedge dy^i.$$

All horizontal $n-$forms on a bundle $Y \longrightarrow X$ are called *horizontal densities*.

For any vector–field τ on X, we can define its pull–back on Y,

$$\pi^*\tau = \tau \circ \pi : Y \longrightarrow TX.$$

This is not a vector–field on Y, for the tangent bundle TX of X fails to be a subbundle of the tangent bundle TY of Y. One needs a connection on $Y \longrightarrow X$ in order to set the imbedding $TX \hookrightarrow TY$.

The space $\wedge^*(M) \otimes \mathcal{V}^1(M)$ of tangent–valued forms admits the *Frölicher–Nijenhuis bracket* (or, FN bracket)

[12] The forms (1.152) are not tangent–valued forms. The pull–backs

$$\phi = \frac{1}{r!}\phi^\mu_{\alpha_1 \dots \alpha_r}(x) dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_r} \otimes \partial_\mu$$

of tangent–valued forms on X onto Y by π exemplify the pull–back-valued forms (1.152). In particular, we shall refer to the pull–back $\pi^*\theta_X$ of the canonical form θ_X on the base X onto Y by π. This is a pull–back-valued horizontal one–form on Y which we denote by the same symbol

$$\theta_X : Y \to T^*X \otimes TX, \qquad \theta_X = dx^\alpha \otimes \partial_\alpha.$$

$$
\begin{aligned}
&[.,.]_{FN} : \wedge^r(M) \otimes \mathcal{V}^1(M) \times \wedge^s(M) \otimes \mathcal{V}^1(M) \to \wedge^{r+s}(M) \otimes \mathcal{V}^1(M),\\
&[\phi,\sigma]_{FN} = \frac{1}{r!s!}(\phi^{\nu}_{\alpha_1\dots\alpha_r}\partial_\nu\sigma^{\mu}_{\alpha_{r+1}\dots\alpha_{r+s}} - \sigma^{\nu}_{\alpha_{r+1}\dots\alpha_{r+s}}\partial_\nu\phi^{\mu}_{\alpha_1\dots\alpha_r} - \\
&r\phi^{\mu}_{\alpha_1\dots\alpha_{r-1}\nu}\partial_{\alpha_r}\sigma^{\nu}_{\alpha_{r+1}\dots\alpha_{r+s}} + s\sigma^{\mu}_{\nu\alpha_{r+2}\dots\alpha_{r+s}}\partial_{\alpha_{r+1}}\phi^{\nu}_{\alpha_1\dots\alpha_r})dx^{\alpha_1}\\
&\wedge\dots\wedge dx^{\alpha_{r+s}}\otimes\partial_\mu.
\end{aligned}
\tag{1.153}
$$

The following relations hold for the FN–bracket:

$$
\begin{aligned}
&[\phi,\psi]_{FN} = (-1)^{|\phi||\psi|+1}[\psi,\phi]_{FN},\\
&[\phi,[\psi,\theta]_{FN}]_{FN} = [[\phi,\psi]_{FN},\theta]_{FN} + (-1)^{|\phi||\psi|}[\psi,[\phi,\theta]_{FN}]_{FN}.
\end{aligned}
\tag{1.154}
$$

Given a tangent–valued form θ, the *Nijenhuis differential*, $d_\theta\sigma$, along θ on $\wedge^*(M)\otimes\mathcal{V}^1(M)$ is defined as

$$
d_\theta\sigma = [\theta,\sigma]_{FN}. \tag{1.155}
$$

By virtue of the relation (1.154), it has the property

$$
d_\phi[\psi,\theta]_{FN} = [d_\phi\psi,\theta]_{FN} + (-1)^{|\phi||\psi|}[\psi,d_\phi\theta]_{FN}.
$$

In particular, if $\theta = u$ is a vector–field, the Nijenhuis differential becomes the Lie derivative of tangent–valued forms

$$
\begin{aligned}
\mathfrak{L}_u\sigma = \ d_u\sigma = [u,\sigma]_{FN} &= (u^\nu\partial_\nu\sigma^{\mu}_{\alpha_1\dots\alpha_s} - \sigma^{\nu}_{\alpha_1\dots\alpha_s}\partial_\nu u^\mu \\
&+ s\sigma^{\mu}_{\nu\alpha_2\dots\alpha_s}\partial_{\alpha_1}u^\nu)dx^{\alpha_1}\wedge\dots\wedge dx^{\alpha_s}\otimes\partial_\mu, \qquad (\sigma\in\wedge^s(M)\otimes\mathcal{V}(M)).
\end{aligned}
\tag{1.156}
$$

1.4 Jet Spaces

Modern formulation of generalized Lagrangian and Hamiltonian dynamics on fibre bundles is developed in the language of *jet spaces*, or *jet manifolds* (see [KMS93, Sau89, Gri83a, BCG91, BGG03, GMS97, MOS99, MS00a, Sau89, Sar93, Sar95, Sar02a]).

Roughly speaking, given two smooth manifolds M and N, the two smooth maps $f, g : M \to N$ between them are said to determine the same $k-jet$ at a point $x \in M$, if they have the kth order contact (or, the kth order *tangency*) at x [KMS93, Arn88]. A set of all $k-$jets from M to N is a jet space $J^k(M,N)$. It is a generalization of a tangent bundle that makes a new smooth fiber bundle out of a given smooth fiber bundle – following the recursive $n-$categorical process. It makes it possible to write differential equations on sections of a fiber bundle in an invariant form. Historically, jet spaces are attributed to C. Ehresmann, and were an advance on the method of prolongation of E. Cartan, of dealing geometrically with higher derivatives, by imposing differential form conditions on newly–introduced formal variables.

1.4.1 Intuition Behind a Jet Space

The concept of *jet space* is based on the idea of *higher–order tangency*, or *higher–order contact*, at some designated point on a smooth manifold (see [Arn88, KMS93]). Namely, a pair of smooth manifold maps (see Figure 1.7),

$$f_1, f_2 : M \to N$$

are said to be $k-$*tangent* (or *tangent of order* k, or have a kth *order contact*) at a point x on a domain manifold M, denoted by $f_1 \sim f_2$, iff

$$\begin{aligned} f_1(x) &= f_2(x) \qquad \text{called} \quad 0-\text{tangent}, \\ \partial_x f_1(x) &= \partial_x f_2(x), \qquad \text{called} \quad 1-\text{tangent}, \\ \partial_{xx} f_1(x) &= \partial_{xx} f_2(x), \qquad \text{called} \quad 2-\text{tangent}, \\ &\ldots \qquad \text{etc. to the order } k. \end{aligned}$$

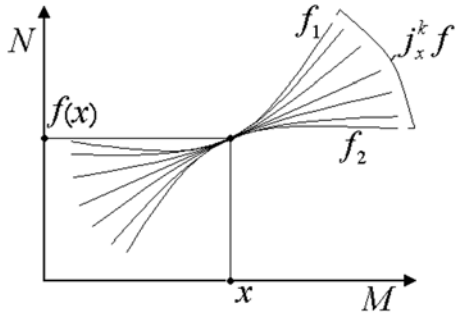


Fig. 1.7. An intuitive geometrical picture behind the $k-$jet concept, based on the idea of higher–order tangency or contact (see text for explanation).

In this way defined $k-$*tangency* is an *equivalence relation*, i.e.,

$$f_1 \sim f_2 \Rightarrow f_2 \sim f_1, \qquad f_1 \sim f_2 \sim f_3 \Rightarrow f_1 \sim f_3, \qquad f_1 \sim f_1.$$

Now a $k-$*jet* (or, a *jet of order* k), denoted by $j^k_x f$, of a smooth map $f : M \to N$ at a point $x \in M$ (see Figure 1.7), is defined as an *equivalence class* of $k-$tangent maps at x,

$$j^k_x f = \{f' : f' \text{ is } k-\text{tangent to } f \text{ at } x\}.$$

The point x is called the *source* and the point $f(x)$ is the *target* of the $k-$jet $j^k_x f$.

We choose local coordinates on M and N in the neighborhood of the points x and $f(x)$, respectively. Then the $k-$jet $j^k_x f$ of any map close to f, at

any point close to x, can be given by its Taylor–series expansion at x, with coefficients up to degree k. Therefore, in a fixed coordinate chart, a $k-$jet can be identified with the collection of Taylor coefficients up to degree k.

The set of all $k-$jets of smooth maps from M to N is called the *$k-$jet space* and denoted by $J^k(M,N)$. It has a natural smooth manifold structure. Also, a map from a $k-$jet space $J^k(M,N)$ to a smooth manifold M or N is called a *jet bundle* (we will make this notion more precise later).

For example, consider a simple function $f : X \to Y, x \mapsto y = f(x)$, mapping the $X-$axis into the $Y-$axis. In this case, $M = X$ is a domain and $N = Y$ is a codomain. A $0-$jet at a point $x \in X$ is given by its graph $(x, f(x))$. A $1-$jet is given by a triple $(x, f(x), f'(x))$, a $2-$jet is given by a quadruple $(x, f(x), f'(x), f''(x))$, and so on up to the order k (where $f'(x) = \frac{df(x)}{dx}$, etc.). The set of all $k-$jets from X to Y is called the $k-$jet space $J^k(X,Y)$.

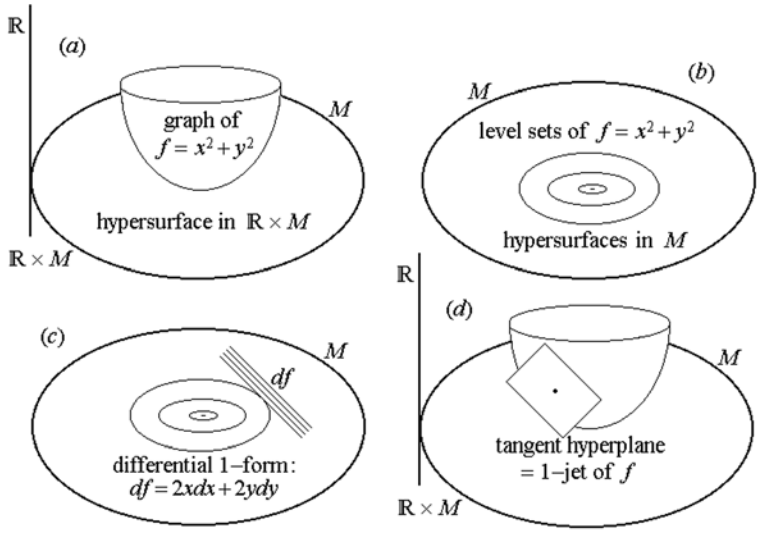


Fig. 1.8. Common spaces associated with a function f on a smooth manifold M (modified and adapted from [Omo86]; see text for explanation).

In case of a function of two variables, $f(x,y)$, the common spaces related to f, including its 1–jet $j^1 f$, are depicted in Figure 1.8. Recall that a *hypersurface* is a *codimension–1 submanifold*. Given a sample function $f(x,y) = x^2 + y^2$ in $M = \mathbb{R}^2$, then: (a) shows its graph as a hypersurface in $\mathbb{R} \times M$; (b) shows its level sets in M; (c) shows its differential form $df = 2xdx + 2ydy$ in the cotangent bundle T^*M; and (d) shows a tangent hyperplane at a point $(x_0, y_y) \in M$ to its graph in $\mathbb{R} \times M$, which is a 1–jet $j^1_{(x_0,y_0)} f$ to f at (x_0, y_y). Note that $j^1_{(x_0,y_0)} f$ is parallel to df, which means that its 1–jet space $J^1(\mathbb{R}, M)$ is an $(n+1)$D extension of the cotangent bundle T^*M.

In mechanics we will consider a pair of maps $f_1, f_2 : \mathbb{R} \to M$ from the real line $\mathbb{R}$, representing the *time* $t-$axis, into a smooth nD (configuration) manifold M. We say that the maps $f_1 = f_1(t)$ and $f_2 = f_2(t)$ have the same $k-$jet $j^k_x f$ at a specified time instant $t_0 \in \mathbb{R}$, iff:

1. $f_1(t) = f_2(t)$ at $t_0 \in \mathbb{R}$, and also
2. the first k terms of their Taylor–series expansion around $t_0 \in \mathbb{R}$ are equal.

The $k-$jet space $J^k(\mathbb{R}, M)$ is the set of all $k-$jets $j^k_x f$ from $\mathbb{R}$ to M.

Now, the fundamental geometrical construct in *time–dependent mechanics* is its *configuration fibre bundle* (see subsection 1.4.6 below). Given a configuration fibre bundle $M \to \mathbb{R}$ over the time axis $\mathbb{R}$, we say that the $1-$*jet space* $J^1(\mathbb{R}, M)$ is the set of equivalence classes $j^1_t s$ of sections $s^i : \mathbb{R} \to M$ of the bundle $M \to \mathbb{R}$, which are identified by their values $s^i(t)$, and by the values of their partial derivatives $\partial_t s^i = \partial_t s^i(t)$ at time points $t \in \mathbb{R}$. The 1–jet space $J^1(\mathbb{R}, M)$ is coordinated by $(t, x^i, \dot{x}^i)$, so the 1–jets are local coordinate maps

$$j^1_t s : t \mapsto (t, x^i, \dot{x}^i).$$

Similarly, the $2-$*jet space* $J^2(\mathbb{R}, M)$ is the set of equivalence classes $j^2_t s$ of sections $s^i : \mathbb{R} \to M$ of the bundle $M \to \mathbb{R}$, which are identified by their values $s^i(t)$, as well as the values of their first and second partial derivatives, $\partial_t s^i$ and $\partial_{tt} s^i$, at time points $t \in \mathbb{R}$. The 2–jet space $J^2(\mathbb{R}, M)$ is coordinated by $(t, x^i, \dot{x}^i, \ddot{x}^i)$, so the 2–jets are local coordinate maps

$$j^2_t s : t \mapsto (t, x^i, \dot{x}^i, \ddot{x}^i).$$

Generalization to the $k-$*jet space* $J^k(\mathbb{R}, M)$ is obvious. This mechanical jet formalism will be initiated in subsection 1.4.6 below and further developed in section 2.1 below.

More generally, in a *physical field* context, instead of the mechanical configuration bundle over the time axis $\mathbb{R}$, we have some general physical fibre bundle $Y \to X$ over some smooth manifold (base) X. In this general context, the $k-$jet space $J^k(X, Y)$ of a bundle $Y \to X$ is the set of equivalence classes $j^k_x s$ of sections $s^i : X \to Y$, which are identified by their values $s^i(x)$, as well as the first k terms of their Taylor–series expansion at points $x \in X$. This has two important physical consequences:

1. The $k-$jet space of sections $s^i : X \to Y$ of a fibre bundle $Y \to X$ is itself an nD smooth manifold, and
2. A kth–order differential operator on sections $s^i(x)$ of a fibre bundle $Y \to X$ can be described as a map of $J^k(X, Y)$ to a vector bundle over the base X.

A map from a $k-$jet space $J^k(X, Y)$ to a smooth manifold Y or X is called a *jet bundle*.

As a consequence, the dynamics of mechanical and physical field systems is played out on nD configuration and phase manifolds. Moreover, this dynamics

can be phrased in geometrical terms due to the 1–1 correspondence between sections of the jet bundle $J^1(X,Y) \to Y$ and connections on the fibre bundle $Y \to X$.

In the framework of the standard *first–order Lagrangian formalism*, the nD *configuration space of sections* $s^i : X \to Y$ of a fibre bundle $Y \to X$ is the 1–jet space $J^1(X,Y)$, coordinated by $(x^\alpha, y^i, y^i_\alpha)$, where (x^α, y^i) are fibre coordinates of Y, while y^i_α are the so–called 'derivative coordinates' or 'velocities'. A first–order *Lagrangian density*[13] on the *configuration manifold* $J^1(X,Y)$ is given by an exterior one–form (the so–called *horizontal density*)

$$L = \mathcal{L}(x^\alpha, y^i, y^i_\alpha)\omega, \qquad \text{with} \qquad \omega = dx^1 \wedge ... \wedge dx^n.$$

This physical jet formalism will be developed in section 2.2 below.

1.4.2 Definition of a 1–Jet Space

As introduced above, a 1–jet is defined as an equivalence class of functions having the same value and the same first derivatives at some designated point of the domain manifold (see Figure 1.7). Recall that in mechanical settings, the 1–jets are local coordinate maps

$$j^1_t s : t \mapsto (t, x^i, \dot{x}^i).$$

More generally, given a fibre bundle $Y \to X$ with bundle coordinates (x^α, y^i), consider the equivalence classes $j^1_x s$ of its sections $s^i : X \to Y$, which are identified by their values $s^i(x)$ and the values of their first–order derivatives $\partial_\alpha s^i = \partial_\alpha s^i(x)$ at a point x on the domain (base) manifold X. They are called the 1–jets of sections s^i at $x \in X$. One can justify that the definition of jets is coordinate–independent by observing that the set $J^1(X,Y)$ of 1–jets $j^1_x s$ is a smooth manifold with respect to the adapted coordinates $(x^\alpha, y^i, y^i_\alpha)$, such that [Sar93, Sar95, GMS97, MS00a, Sar02a]

$$y^i_\alpha(j^1_x s) = \partial_\alpha s^i(x), \qquad y'^i_\alpha = \frac{\partial x^\mu}{\partial x'^\alpha}(\partial_\mu + y^j_\mu \partial_j) y'^i.$$

$J^1(X,Y)$ is called the $1-$*jet space* of the fibre bundle $Y \to X$.

[13] Recall that in classical field theory, a distinction is made between the Lagrangian L, of which the *action* is the time integral $S[x^i] = \int L[x^i, \dot{x}^i]dt$ and the Lagrangian density $\mathcal{L}$, which one integrates over all space–time to get the action $S[\varphi^k] = \int \mathcal{L}[\varphi^k[x^i]]d^4x$. The Lagrangian is then the spatial integral of the Lagrangian density. However, $\mathcal{L}$ is also frequently simply called the Lagrangian, especially in modern use; it is far more useful in relativistic theories since it is a locally defined, Lorentz scalar field. Both definitions of the Lagrangian can be seen as special cases of the general form, depending on whether the spatial variable x^i is incorporated into the index i or the parameters s in $\varphi^k[x^i]$. Quantum field theories are usually described in terms of $\mathcal{L}$, and the terms in this form of the Lagrangian translate quickly to the rules used in evaluating Feynman diagrams.

In other words, the 1–jets $j^1_x s : x^\alpha \mapsto (x^\alpha, y^i, y^i_\alpha)$, which are first–order equivalence classes of sections of the fibre bundle $Y \to X$, can be identified with their codomain set of adapted coordinates on $J^1(X,Y)$,

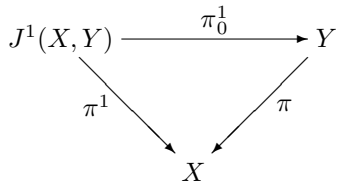
$$j^1_x s \equiv (x^\alpha, y^i, y^i_\alpha).$$

Note that in a subsection 2.1.4 below, the mechanical 1–jet space $J^1(\mathbb{R}, M) \equiv \mathbb{R} \times TM$ will be regarded as a fibre bundle over the base product–manifold $\mathbb{R} \times M$ (see [NU00a, Udr00, Nea02, NU00b, Nea00] for technical details).

The jet space $J^1(X,Y)$ admits the natural fibrations

$$\pi^1 : J^1(X,Y) \ni j^1_x s \mapsto x \in X, \qquad \text{and} \tag{1.157}$$
$$\pi^1_0 : J^1(X,Y) \ni j^1_x s \mapsto s(x) \in Y, \tag{1.158}$$

which form the commutative triangle:



It is convenient to call π^1 (1.157) the *jet bundle*, while π^1_0 (1.158) is called the *affine jet bundle*. Note that, if $Y \to X$ is a vector or an affine bundle, it also holds for the jet bundle π_1 (1.157) [Sar93, Sar95, GMS97, MS00a, Sar02a].

There exist several equivalent ways in order to give the 1–jet space $J^1(X,Y)$ with the smooth manifold structure. Let $Y \to X$ be a fibre bundle with fibred coordinate atlases (1.114). The 1–jet space $J^1(X,Y)$ of the bundle $Y \to X$ admits the adapted coordinate atlases

$$(x^\alpha, y^i, y^i_\alpha), \qquad (x^\alpha, y^i, y^i_\alpha)(j^1_x s) = (x^\alpha, s^i(x), \partial_\alpha s^i), \tag{1.159}$$
$$y'^i_\alpha = (\frac{\partial y'^i}{\partial y^j} y^j_\mu + \frac{\partial y'^i}{\partial x^\mu}) \frac{\partial x^\mu}{\partial x'^\alpha}, \tag{1.160}$$

and thus satisfies the conditions which are required of a manifold. The surjection (1.157) is a bundle. The surjection (1.158) is a bundle. If $Y \to X$ is a bundle, so is the surjection (1.157).

The transformation law (1.160) shows that the jet bundle $J^1(X,Y) \to Y$ is an affine bundle. It is modelled on the vector bundle $T^*X \otimes VY \to Y$. In particular, if Y is the trivial bundle

$$\pi_2 : V \times \mathbb{R}^m \longrightarrow \mathbb{R}^m,$$

the corresponding jet bundle $J^1(X,Y) \longrightarrow \mathbb{R}^m$ (1.157) is a trivial bundle.

There exist the following two canonical bundle monomorphisms of the jet bundle $J^1(X,Y) \longrightarrow Y$ [Sar93, Sar95, GMS97, MS00a, Sar02a]:

- the *contact map*

$$\lambda : J^1(X,Y) \hookrightarrow T^*X \otimes TY, \qquad \lambda = dx^\alpha \otimes \widehat{\partial}_\alpha = dx^\alpha \otimes (\partial_\alpha + y^i_\alpha \partial_i), \quad (1.161)$$

- the *complementary map*

$$\theta : J^1(X,Y) \hookrightarrow T^*Y \otimes VY, \qquad \theta = \widehat{dy}^i \otimes \partial_i = (dy^i - y^i_\alpha dx^\alpha) \otimes \partial_i. \quad (1.162)$$

These canonical maps enable us to express the jet–space machinery in terms of tangent–valued differential forms (see subsection 1.3.5 above).

The operators

$$\widehat{\partial}_\alpha = \partial_\alpha + y^i_\alpha \partial_i$$

are usually called the *total derivatives*, or the *formal derivatives*, while the forms

$$\widehat{dy}^i = dy^i - y^i_\alpha dx^\alpha$$

are conventionally called the *contact forms*.

Identifying the 1–jet space $J^1(X,Y)$ to its images under the canonical maps (1.161) and (1.162), one can represent 1–jets $j^1_x s \equiv (x^\alpha, y^i, y^i_\alpha)$ by tangent–valued forms

$$dx^\alpha \otimes (\partial_\alpha + y^i_\alpha \partial_i), \qquad \text{and} \qquad (dy^i - y^i_\alpha dx^\alpha) \otimes \partial_i. \quad (1.163)$$

There exists a *jet functor* $\mathfrak{J} : Bun \to Jet$, from the category Bun of fibre bundles to the category Jet of jet spaces. It implies the natural prolongation of maps of bundles to maps of jet spaces.

Every bundle map $\Phi : Y \longrightarrow Y'$ over a diffeomorphism f of X has the 1–*jet prolongation* to the bundle map $j^1\Phi : J^1(X,Y) \longrightarrow J^1(X,Y)'$, given by

$$j^1\Phi \ : \ j^1_x s \mapsto j^1_{f(x)}(\Phi \circ s \circ f^{-1}), \quad (1.164)$$
$$y'^i_\alpha \circ j^1\Phi = \partial_\alpha(\Phi^i \circ f^{-1}) + \partial_j(\Phi^i y^j_\alpha \circ f^{-1}).$$

It is both an affine bundle map over Φ and a fibred map over the diffeomorphism f. The 1–jet prolongations (1.164) of fibred maps satisfy the chain rules

$$j^1(\Phi \circ \Phi') = j^1\Phi \circ j^1\Phi', \qquad j^1(\mathrm{Id}_Y) = \mathrm{Id}_{J^1(X,Y)}.$$

If Φ is a surjection (resp. an injection), so is $j^1\Phi$.

In particular, every section s of a bundle $Y \to X$ admits the 1–jet prolongation to the section $j^1_x s$ of the jet bundle $J^1(X,Y) \to X$, given by

$$(y^i, y^i_\alpha) \circ j^1_x s = (s^i(x), \partial_\alpha s^i).$$

We have

$$\lambda \circ j^1_x s = Ts,$$

where λ is the contact map (1.161).

Every projectable vector–field u on a fibre bundle $Y \to X$,

$$u = u^\alpha(x)\partial_\alpha + u^i(y)\partial_i$$

has the 1−*jet lift* to the projectable vector–field j^1u on the 1–jet space $J^1(X,Y)$, given by

$$\begin{aligned} j^1u \equiv \overline{u} &= r_1 \circ j^1u : J^1(X,Y) \to TJ^1(X,Y), \\ j^1u \equiv \overline{u} &= u^\alpha\partial_\alpha + u^i\partial_i + (d_\alpha u^i - y^i_\mu\partial_\alpha u^\mu)\partial^\alpha_i. \end{aligned} \tag{1.165}$$

Geometrical applications of jet spaces are based on the canonical map over $J^1(X,Y)$,

$$J^1(X,Y) \times TX \to J^1(X,Y) \times TY,$$

which means the canonical horizontal splitting of the tangent bundle TY determined over $J^1(X,Y)$ as follows [Sar93, Sar95, GMS97, MS00a, Sar02a].

The canonical maps (1.161) and (1.162) induce the bundle monomorphisms

$$\widehat{\lambda} : J^1(X,Y) \times TX \to J^1(X,Y) \times TY, \qquad \partial_\alpha \mapsto \widehat{\partial}_\alpha = \partial_\alpha \rfloor \lambda \tag{1.166}$$

$$\widehat{\theta} : J^1(X,Y) \times V^*Y \to J^1(X,Y) \times T^*Y, \qquad dy^i \mapsto \widehat{d}y^i = \theta \rfloor dy^i \tag{1.167}$$

The map (1.166) determines the canonical horizontal splitting of the pull–back

$$\begin{aligned} J^1(X,Y) \times TY &= \widehat{\lambda}(TX) \oplus VY, \\ \dot{x}^\alpha\partial_\alpha + \dot{y}^i\partial_i &= \dot{x}^\alpha(\partial_\alpha + y^i_\alpha\partial_i) + (\dot{y}^i - \dot{x}^\alpha y^i_\alpha)\partial_i. \end{aligned} \tag{1.168}$$

Similarly, the map (1.167) induces the dual canonical horizontal splitting of the pull–back

$$\begin{aligned} J^1(X,Y) \times T^*Y &= T^*X \oplus \widehat{\theta}(V^*Y), \\ \dot{x}_\alpha dx^\alpha + \dot{y}_i dy^i &= (\dot{x}_\alpha + \dot{y}_i y^i_\alpha)dx^\alpha + \dot{y}_i(dy^i - y^i_\alpha dx^\alpha). \end{aligned} \tag{1.169}$$

Building on the canonical splittings (1.168) and (1.169), one gets the following canonical horizontal splittings of

- a projectable vector–field on a fibre bundle $Y \to X$,

$$u = u^\alpha\partial_\alpha + u^i\partial_i = u_H + u_V = u^\alpha(\partial_\alpha + y^i_\alpha\partial_i) + (u^i - u^\alpha y^i_\alpha)\partial_i, \tag{1.170}$$

- an exterior 1–form

$$\sigma = \sigma_\alpha dx^\alpha + \sigma_i dy^i = (\sigma_\alpha + y^i_\alpha\sigma_i)dx^\alpha + \sigma_i(dy^i - y^i_\alpha dx^\alpha),$$

- a tangent–valued projectable horizontal form

$$\begin{aligned} \phi &= dx^{\alpha_1} \wedge \cdots \wedge dx^{\alpha_r} \otimes (\phi^\mu_{\alpha_1\ldots\alpha_r}\partial_\mu + \phi^i_{\alpha_1\ldots\alpha_r}\partial_i) \\ &= dx^{\alpha_1} \wedge \cdots \wedge dx^{\alpha_r} \otimes [\phi^\mu_{\alpha_1\ldots\alpha_r}(\partial_\mu + y^i_\mu\partial_i) + (\phi^i_{\alpha_1\ldots\alpha_r} - \phi^\mu_{\alpha_1\ldots\alpha_r}y^i_\mu)\partial_i] \end{aligned}$$

and, e.g., the canonical 1–form

$$\begin{aligned} \theta_Y &= dx^\alpha \otimes \partial_\alpha + dy^i \otimes \partial_i = \alpha + \theta = dx^\alpha \otimes \widehat{\partial}_\alpha + \widehat{d}y^i \otimes \partial_i \\ &= dx^\alpha \otimes (\partial_\alpha + y^i_\alpha\partial_i) + (dy^i - y^i_\alpha dx^\alpha) \otimes \partial_i. \end{aligned} \tag{1.171}$$

The splitting (1.171) implies the canonical horizontal splitting of the exterior differential

$$d = d_{\theta_Y} = d_H + d_V = d_\alpha + d_\theta. \tag{1.172}$$

Its components d_H and d_V act on the pull–backs

$$\phi_{\alpha_1 \dots \alpha_r}(y) dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_r}$$

of horizontal exterior forms on a bundle $Y \to X$ onto $J^1(X,Y)$ by π_{01}. In this case, d_H makes the sense of the total differential

$$d_H \phi_{\alpha_1 \dots \alpha_r}(y) dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_r} = (\partial_\mu + y^i_\mu \partial_i) \phi_{\alpha_1 \dots \alpha_r}(y) dx^\mu \wedge dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_r},$$

whereas d_V is the vertical differential

$$d_V \phi_{\alpha_1 \dots \alpha_r}(y) dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_r} = \partial_i \phi_{\alpha_1 \dots \alpha_r}(y) (dy^i - y^i_\mu dx^\mu) \wedge dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_r}.$$

If $\phi = \widetilde{\phi}\omega$ is an exterior horizontal density on $Y \to X$, we have

$$d\phi = d_V \phi = \partial_i \widetilde{\phi} dy^i \wedge \omega.$$

1.4.3 Connections as Jet Fields

Recall that one can introduce the notion of *connections* on a general fibre bundle $Y \longrightarrow X$ in several equivalent ways. In this subsection, following [GMS97, KMS93, MS00a, Sau89], we start from the traditional definition of a connection as a *horizontal splitting* of the tangent space to Y at every point $y \in Y$.

A connection on a fibre bundle $Y \to X$ is usually defined as a linear bundle monomorphism

$$\Gamma : Y \times TX \to TY, \qquad \Gamma : \dot{x}^\alpha \partial_\alpha \mapsto \dot{x}^\alpha (\partial_\alpha + \Gamma^i_\alpha(y) \partial_i), \tag{1.173}$$

which splits the exact sequence (1.123), i.e.,

$$\pi_T \circ \Gamma = \mathrm{Id}_{Y \times TX}\,.$$

The image HY of $Y \times TX$ by a connection Γ is called the *horizontal distribution*. It splits the tangent bundle TY as

$$\begin{aligned} &TY = HY \oplus VY, \qquad \text{giving} \\ &\dot{x}^\alpha \partial_\alpha + \dot{y}^i \partial_i = \dot{x}^\alpha (\partial_\alpha + \Gamma^i_\alpha \partial_i) + (\dot{y}^i - \dot{x}^\alpha \Gamma^i_\alpha) \partial_i. \end{aligned} \tag{1.174}$$

Similarly, horizontal splitting of the cotangent bundle,

$$\begin{aligned} &T^*Y = T^*X \oplus \Gamma(V^*X), \qquad \text{gives} \\ &\dot{x}_\alpha dx^\alpha + \dot{y}_i dy^i = (\dot{x}_\alpha + \Gamma^i_\alpha \dot{y}_i) dx^\alpha + \dot{y}_i (dy^i - \Gamma^i_\alpha dx^\alpha). \end{aligned}$$

Alternatively, a connection on a fibre bundle $Y \to X$ can be defined as a *jet field*, i.e., a *section* of the affine jet bundle $J^1(X,Y) \to Y$. This connection is called the *Ehresmann connection*, and historically it was a primary reason for C. Ehresmann to develop the concept of jet spaces.

Due to the theorem that says [Hir66]: Every exact sequence of vector bundles (1.120) is split, a jet–field connection on a fibre bundle always exists.

A connection on a fibre bundle $Y \to X$ is defined to be a tangent–valued projectable horizontal one–form Γ on Y such that $\Gamma \rfloor \phi = \phi$ for all exterior horizontal 1–forms ϕ on Y. It is given by the coordinate expression

$$\Gamma = dx^\alpha \otimes (\partial_\alpha + \Gamma^i_\alpha(y)\partial_i), \qquad \Gamma'^i_\alpha = (\frac{\partial y'^i}{\partial y^j}\Gamma^j_\mu + \frac{\partial y'^i}{\partial x^\mu})\frac{\partial x^\mu}{\partial x'^\alpha}, \tag{1.175}$$

such that $\Gamma(\partial_\alpha) = \partial_\alpha \rfloor \Gamma$. Conversely, every horizontal tangent–valued 1–form on a fibre bundle $Y \to X$ which projects onto the canonical tangent–valued form θ_X (1.150) on X defines a connection on $Y \to X$.

In an equivalent way, the horizontal splitting (1.174) is given by the vertical–valued form

$$\Gamma = (dy^i - \Gamma^i_\alpha dx^\alpha) \otimes \partial_i, \tag{1.176}$$

which determines the epimorphism

$$\Gamma : TY \to VY, \qquad \dot{x}^\alpha\partial_\alpha + \dot{y}^i\partial_i \mapsto (\dot{x}^\alpha\partial_\alpha + \dot{y}^i\partial_i)\rfloor\Gamma = (\dot{y}^i - \dot{x}^\alpha\Gamma^i_\alpha)\partial_i.$$

Let $Y \to X$ be a vector bundle. A *linear connection* on Y reads

$$\Gamma = dx^\alpha \otimes [\partial_\alpha - \Gamma^i{}_{j\alpha}(x)y^j\partial_i]. \tag{1.177}$$

Let $Y \to X$ be an affine bundle modelled on a vector bundle $\overline{Y} \to X$. An *affine connection* on Y reads

$$\begin{aligned} &\Gamma = dx^\alpha \otimes [\partial_\alpha + (-\Gamma^i{}_{j\alpha}(x)y^j + \Gamma^i{}_\alpha(x))\partial_i], \qquad \text{where} \\ &\overline{\Gamma} = dx^\alpha \otimes [\partial_\alpha - \Gamma^i{}_{j\alpha}(x)\overline{y}^j\partial_i] \qquad \text{is a linear connection on } \overline{Y}. \end{aligned}$$

Since the affine jet bundle $J^1(X,Y) \to Y$ is modelled on the vector bundle $Y \to X$, Ehresmann connections on $Y \to X$ constitute an affine space modelled on the linear space of soldering forms on Y. If Γ is a connection and σ is a soldering form (1.151) on Y, its sum

$$\Gamma + \sigma = dx^\alpha \otimes [\partial_\alpha + (\Gamma^i_\alpha + \sigma^i_\alpha)\partial_i]$$

is a connection on Y. Conversely, if Γ and Γ' are connections on Y, then

$$\Gamma - \Gamma' = (\Gamma^i_\alpha - \Gamma'^i_\alpha)dx^\alpha \otimes \partial_i$$

is a soldering form.

Given a connection Γ, a vector–field u on a fibre bundle $Y \to X$ is called horizontal if it lives in the horizontal distribution HY, i.e., takes the form

$$u = u^\alpha(y)(\partial_\alpha + \Gamma^i_\alpha(y)\partial_i). \tag{1.178}$$

Any vector–field τ on the base X of a fibre bundle $Y \to X$ admits the horizontal lift

$$\Gamma\tau = \tau\rfloor\Gamma = \tau^\alpha(\partial_\alpha + \Gamma^i_\alpha\partial_i) \tag{1.179}$$

onto Y by means of a connection Γ on $Y \to X$.

Given the splitting (1.173), the dual splitting of the exact sequence (1.124) is

$$\Gamma : V^*Y \to T^*Y, \qquad \overline{d}y^i \mapsto \Gamma\rfloor\overline{d}y^i = dy^i - \Gamma^i_\alpha dx^\alpha, \tag{1.180}$$

where Γ is the vertical-valued form (1.176).

There is 1–1 correspondence between the connections on a fibre bundle $Y \to X$ and the jet fields, i.e., global sections of the affine jet bundle $J^1(X,Y) \to Y$. Indeed, given a global section Γ of $J^1(X,Y) \to Y$, the tangent–valued form

$$\lambda \circ \Gamma = dx^\alpha \otimes (\partial_\alpha + \Gamma^i_\alpha\partial_i)$$

gives the horizontal splitting (1.174) of TY. Therefore, the vertical–valued form

$$\theta \circ \Gamma = (dy^i - \Gamma^i_\alpha dx^\alpha)\otimes\partial_i$$

leads to the dual splitting (1.180).

It follows immediately from this definition that connections on a fibre bundle $Y \to X$ constitute an affine space modelled over the vector space of soldering forms σ (1.151). They obey the *coordinate transformation law* [GMS97, Sau89]

$$\Gamma'^i_\alpha = \frac{\partial x^\mu}{\partial x'^\alpha}(\partial_\mu + \Gamma^j_\mu\partial_j)y'^i.$$

In particular, a *linear connection* K on the tangent bundle TX of a manifold X and the *dual connection* K^* to K on the cotangent bundle T^*X are given by the coordinate expressions

$$K^\alpha_\beta = -K^\alpha{}_{\nu\beta}(x)\dot{x}^\nu, \qquad K^*_{\alpha\beta} = K^\nu{}_{\alpha\beta}(x)\dot{x}_\nu. \tag{1.181}$$

Also, given a connection Γ on $Y \to X$, the vertical tangent map $V\Gamma : VY \to J^1(\mathbb{R}, V)Y$ induces the *vertical connection*

$$V\Gamma = dx^\alpha \otimes (\partial_\alpha + \Gamma^i_\alpha\partial_{y^i} + \partial_V\Gamma^i_\alpha\partial_{\dot{y}^i}), \qquad \partial_V\Gamma^i_\alpha = \dot{y}^j\partial_j\Gamma^i_\alpha, \tag{1.182}$$

on the bundle $VY \to X$. The connection $V\Gamma$ is projectable to the connection Γ on Y, and it is a linear bundle map over Γ. The dual *covertical connection* on the bundle $V^*Y \to X$ reads

$$V^*\Gamma = dx^\alpha \otimes (\partial_\alpha + \Gamma^i_\alpha\partial_{y^i} - \partial_j\Gamma^i_\alpha\dot{y}_i\partial_{\dot{y}^i}). \tag{1.183}$$

Connections on a bundle $Y \to X$ constitute the affine space modelled on the linear space of soldering 1–forms on Y. It means that, if Γ is a connection and σ is a soldering form on a bundle Y, its sum

$$\Gamma + \sigma = dx^\alpha \otimes [\partial_\alpha + (\Gamma^i_\alpha + \sigma^i_\alpha)\partial_i]$$

is a connection on Y. Conversely, if Γ and Γ' are connections on a bundle Y, then their difference

$$\Gamma - \Gamma' = (\Gamma^i_\alpha - \Gamma'^i_\alpha)dx^\alpha \otimes \partial_i$$

is a soldering form on Y.

Given a fibre bundle $Y \to X$, let $f : X' \to X$ be a manifold map and f^*Y the pull–back of Y over X'. Any connection Γ (1.176) on $Y \to X$ induces the pull–back connection

$$f^*\Gamma = (dy^i - \Gamma^i_\alpha(f^\mu(x'\nu), y^j)\frac{\partial f^\alpha}{\partial x'^\mu}dx'^\mu) \otimes \partial_i \tag{1.184}$$

on the pull–back fibre bundle $f^*Y \to X'$.

Since the affine jet bundle $J^1(X,Y) \longrightarrow Y$ is modelled on the vector bundle $Y \longrightarrow X$, connections on a fibre bundle Y constitute the affine space modelled on the linear space of soldering forms on Y. It follows that, if Γ is a connection and

$$\sigma = \sigma^i_\alpha dx^\alpha \otimes \partial_i$$

is a soldering form on a fibre bundle Y, its sum

$$\Gamma + \sigma = dx^\alpha \otimes [\partial_\alpha + (\Gamma^i_\alpha + \sigma^i_\alpha)\partial_i]$$

is a connection on Y. Conversely, if Γ and Γ' are connections on a fibre bundle Y, then

$$\Gamma - \Gamma' = (\Gamma^i_\alpha - \Gamma'^i_\alpha)dx^\alpha \otimes \partial_i$$

is a soldering form on Y.

The key point for physical applications lies in the fact that every connection Γ on a fibre bundle $Y \longrightarrow X$ induces the first–order differential operator

$$D^\Gamma : J^1(X,Y) \to T^*X \otimes VY, \qquad D^\Gamma = \lambda - \Gamma \circ \pi^1_0 = (y^i_\alpha - \Gamma^i_\alpha)dx^\alpha \otimes \partial_i, \tag{1.185}$$

called the *covariant differential* relative to the connection Γ. If $s : X \to Y$ is a section, one defines its covariant differential

$$\nabla^\Gamma s = D_\Gamma \circ j^1 s = (\partial_\alpha s^i - \Gamma^i_\alpha \circ s)dx^\alpha \otimes \partial_i \tag{1.186}$$

and its *covariant derivative*

$$\nabla^\Gamma_\tau s = \tau \rfloor \nabla^\Gamma s \tag{1.187}$$

along a vector–field τ on X. A (local) section s of $Y \to X$ is said to be an *integral section* of a connection Γ (or parallel with respect to Γ) if s obeys the equivalent conditions

$$\nabla^\Gamma s = 0 \qquad \text{or} \qquad j^1 s = \Gamma \circ s. \tag{1.188}$$

Furthermore, if $s : X \to Y$ is a global section, there exists a connection Γ such that s is an integral section of Γ. This connection is defined as an extension of the local section $s(x) \mapsto j^1 s(x)$ of the affine jet bundle $J^1(X,Y) \to Y$ over the closed imbedded submanifold $s(X) \subset Y$.

Note that every connection Γ on the bundle $Y \longrightarrow X$ defines a system of first–order differential equations on Y (in the spirit of [BCG91, KLV85, Pom78]) which is an imbedded subbundle $\Gamma(Y) = \operatorname{Ker} D_\Gamma$ of the jet bundle $J^1(X,Y) \longrightarrow Y$. It is given by the coordinate relations

$$y^i_\alpha = \Gamma^i(y). \tag{1.189}$$

Integral sections for Γ are local solutions of (1.189), and *vice versa*.

We can introduce the following basic forms involving a connection Γ and a soldering form σ:

- the *curvature of a connection* Γ is given by the horizontal vertical-valued two–form:

$$\begin{aligned} R &= \frac{1}{2} d_\Gamma \Gamma = \frac{1}{2} R^i_{\alpha\mu} dx^\alpha \wedge dx^\mu \otimes \partial_i, \\ R^i_{\alpha\mu} &= \partial_\alpha \Gamma^i_\mu - \partial_\mu \Gamma^i_\alpha + \Gamma^j_\alpha \partial_j \Gamma^i_\mu - \Gamma^j_\mu \partial_j \Gamma^i_\alpha; \end{aligned} \tag{1.190}$$

- the *torsion of a connection* Γ with respect to σ:

$$\begin{aligned} \Omega = d_\sigma \Gamma = d_\Gamma \sigma &= \frac{1}{2} \Omega^i_{\alpha\mu} dx^\alpha \wedge dx^\mu \otimes \partial_i \\ &= (\partial_\alpha \sigma^i_\mu + \Gamma^j_\alpha \partial_j \sigma^i_\mu - \partial_j \Gamma^i_\alpha \sigma^j_\mu) dx^\alpha \wedge dx^\mu \otimes \partial_i; \end{aligned} \tag{1.191}$$

- the *soldering curvature* of σ:

$$\begin{aligned} \varepsilon = \frac{1}{2} d_\sigma \sigma &= \frac{1}{2} \varepsilon^i_{\alpha\mu} dx^\alpha \wedge dx^\mu \otimes \partial_i \\ &= \frac{1}{2} (\sigma^j_\alpha \partial_j \sigma^i_\mu - \sigma^j_\mu \partial_j \sigma^i_l) dx^\alpha \wedge dx^\mu \otimes \partial_i. \end{aligned} \tag{1.192}$$

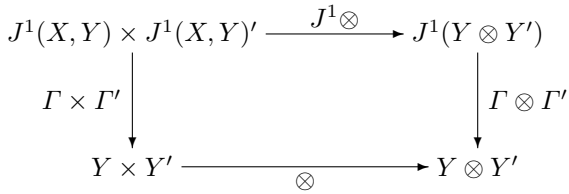
They satisfy the following relations:

$$\Gamma' = \Gamma + \sigma, \qquad R' = R + \varepsilon + \Omega, \qquad \Omega' = \Omega + 2\varepsilon.$$

In particular, the curvature (1.190) of the linear connection (1.177) reads

$$\begin{aligned} R^i_{\alpha\mu}(y) &= -R^i{}_{j\alpha\mu}(x) y^j, \\ R^i{}_{j\alpha\mu} &= \partial_\alpha \Gamma^i{}_{j\mu} - \partial_\mu \Gamma^i{}_{j\alpha} + \Gamma^k{}_{j\mu} \Gamma^i{}_{k\alpha} - \Gamma^k{}_{j\alpha} \Gamma^i{}_{k\mu}. \end{aligned}$$

Let Y and Y' be vector bundles over X. Given linear connections Γ and Γ' on Y and Y' respectively, there is the unique linear connection $\Gamma \otimes \Gamma'$ on the tensor product $Y \otimes Y' \to X$, such that the following diagram commutes:



It is called the *tensor–product connection* and has the coordinate expression

$$(\Gamma\otimes\Gamma')^{ik}_{\alpha}=\Gamma^{i}{}_{j\alpha}y^{jk}+\Gamma'^{k}{}_{j\alpha}y^{ij}.$$

Every connection Γ on $Y\to X$, by definition, induces the *horizontal distribution* on Y,

$$\Gamma: TX\hookrightarrow TY, \qquad \text{locally given by} \qquad \partial_\alpha\mapsto\partial_\alpha+\Gamma^i_\alpha(y)\partial_i.$$

It is generated by horizontal lifts

$$\tau_\Gamma=\tau^\alpha(\partial_\alpha+\Gamma^i_\alpha\partial_i)$$

onto Y of vector–fields $\tau=\tau^\alpha\partial_\alpha$ on X. The associated Pfaffian system is locally generated by the forms $(dy^i-\Gamma^i_\alpha dx^\alpha)$.

The horizontal distribution $\Gamma(TX)$ is involutive iff Γ is a *curvature–free connection*. As a proof, straightforward calculations show that $[\tau_\Gamma,\tau'_\Gamma]=([\tau,\tau'])_\Gamma$ iff the curvature R (1.190) of Γ vanishes everywhere.

Not every bundle admits a curvature–free connection. If a principal bundle over a simply–connected base (i.e., its first homotopy group is trivial) admits a curvature–free connection, this bundle is trivializable [KN63/9].

The horizontal distribution defined by a curvature–free connection is completely integrable. The corresponding foliation on Y is transversal to the foliation defined by the fibration $\pi: Y\longrightarrow X$. It is called the *horizontal foliation*. Its leaf through a point $y\in Y$ is defined locally by the integral section s_y of the connection Γ through y. Conversely, let Y admits a horizontal foliation such that, for each point $y\in Y$, the leaf of this foliation through y is locally defined by some section s_y of $Y\longrightarrow X$ through y. Then, the following map is well defined

$$\Gamma: Y\longrightarrow J^1(X,Y), \qquad \Gamma(y)=j^1_s s_y, \qquad \pi(y)=x.$$

This is a curvature–free connection on Y. There is the 1–1 correspondence between the curvature–free connections and the horizontal foliations on a bundle $Y\longrightarrow X$.

Given a horizontal foliation on $Y\longrightarrow X$, there exists the associated atlas of bundle coordinates (x^α,y^i) of Y such that (i) every leaf of this foliation is local generated by the equations $y^i=\text{const}$, and (ii) the transition functions $y^i\longrightarrow y'^i(y^j)$ are independent on the coordinates x^α of the base X [KT75]. It is called the *atlas of constant local trivializations*. Two such atlases are said

to be equivalent if their union also is an atlas of constant local trivializations. They are associated with the same horizontal foliation.

There is the 1–1 correspondence between the curvature–free connections Γ on a bundle $Y \longrightarrow X$ and the equivalence classes of atlases Ψ_c of constant local trivializations of Y such that $\Gamma^i_\alpha = 0$ relative to the coordinates of the corresponding atlas Ψ_c [Can86].

Connections on a bundle over a 1D base X^1 are curvature–free connections.

In particular, let $Y \longrightarrow X^1$ be such a bundle ($X^1 = \mathbb{R}$ or $X^1 = S^1$). It is coordinated by (t, y^i), where t is either the canonical parameter of $\mathbb{R}$ or the standard local coordinate of S^1 together with the transition functions $t' = t+$const. Relative to this coordinate, the base X^1 admits the standard vector–field ∂_t and the standard one–form dt. Let Γ be a connection on $Y \longrightarrow X^1$. Such a connection defines a horizontal foliation on $Y \longrightarrow X^1$. Its leaves are the integral curves of the horizontal lift

$$\tau_\Gamma = \partial_t + \Gamma^i \partial_i \tag{1.193}$$

of ∂_t by Γ. The corresponding Pfaffian system is locally generated by the forms $(dy^i - \Gamma^i dt)$. There exists an atlas of constant local trivializations (t, y^i) such that $\Gamma^i = 0$ and $\tau_\Gamma = \partial_t$ relative to these coordinates.

A connection Γ on $Y \to X^1$ is called *complete* if the horizontal vector–field (1.193) is complete. Every trivialization of $Y \to \mathbb{R}$ defines a complete connection. Conversely, every complete connection on $Y \to \mathbb{R}$ defines a trivialization $Y \simeq \mathbb{R}\times M$. The vector–field (1.193) becomes the vector–field ∂_t on $\mathbb{R}\times M$. As a proof, every trivialization of $Y \to \mathbb{R}$ defines a 1–parameter group of isomorphisms of $Y \to \mathbb{R}$ over $\mathrm{Id}_\mathbb{R}$, and hence a complete connection. Conversely, let Γ be a complete connection on $Y \to \mathbb{R}$. The vector–field τ_Γ (1.193) is the generator of a 1–parameter group G_Γ which acts freely on Y. The orbits of this action are the integral sections of τ_Γ. Hence we get a projection $Y \to M = Y/G_\Gamma$ which, together with the projection $Y \to \mathbb{R}$, defines a trivialization $Y \simeq \mathbb{R}\times M$.

Let us consider a bundle $\pi : Y \to X$ which admits a *composite fibration*

$$Y \to \Sigma \to X, \tag{1.194}$$

where $Y \to \Sigma$ and $\Sigma \to X$ are bundles. It is equipped with the bundle coordinates $(x^\alpha, \sigma^m, y^i)$ together with the transition functions

$$x^\alpha \to x'^\alpha(x^\mu), \qquad \sigma^m \to \sigma'^m(x^\mu, \sigma^n), \qquad y^i \to y'^i(x^\mu, \sigma^n, y^j),$$

where (x^μ, σ^m) are bundle coordinates of $\Sigma \to X$. For example, we have the composite bundles

$$TY \to Y \to X, \qquad VY \to Y \to X, \qquad J^1(X,Y) \to Y \to X.$$

Let

$$A = dx^\alpha \otimes (\partial_\alpha + A^i_\alpha \partial_i) + d\sigma^m \otimes (\partial_m + A^i_m \partial_i) \tag{1.195}$$

be a connection on the bundle $Y \longrightarrow \Sigma$ and

$$\Gamma = dx^\alpha \otimes (\partial_\alpha + \Gamma^m_\alpha \partial_m)$$

a connection on the bundle $\Sigma \longrightarrow X$. Given a vector–field τ on X, let us consider its horizontal lift τ_Γ onto Σ by Γ and then the horizontal lift $(\tau_\Gamma)_A$ of τ_Γ onto Y by the connection (1.195).

There exists the connection

$$\gamma = dx^\alpha \otimes [\partial_\alpha + \Gamma^m_\alpha \partial_m + (A^i_m \Gamma^m_\alpha + A^i_\alpha)\partial_i]. \tag{1.196}$$

on $Y \to X$ such that the horizontal lift τ_γ onto Y of any vector–field τ on X consists with the above lift $(\tau_\Gamma)_A$ [Sar93, Sar95]. It is called the *composite connection*.

Given a composite bundle Y (1.194), the exact sequence

$$0 \to VY_\Sigma \hookrightarrow VY \to Y \times V\Sigma \to 0$$

over Y take place, where VY_Σ is the vertical tangent bundle of $Y \to \Sigma$. Every connection (1.195) on the bundle $Y \to \Sigma$ induces the splitting

$$\begin{aligned} &VY = VY_\Sigma \oplus (Y \times V\Sigma), \qquad \text{given by} \\ &\dot y^i \partial_i + \dot\sigma^m \partial_m = (\dot y^i - A^i_m \dot\sigma^m)\partial_i + \dot\sigma^m(\partial_m + A^i_m \partial_i). \end{aligned}$$

Due to this splitting, one can construct the first–order differential operator

$$\begin{aligned} &\widetilde D = \pi_1 \circ D_\gamma : J^1(X,Y) \to T^*X \otimes VY \to T^*X \otimes VY_\Sigma, \\ &\widetilde D = dx^\alpha \otimes (y^i_\alpha - A^i_\alpha - A^i_m \sigma^m_\alpha)\partial_i, \end{aligned} \tag{1.197}$$

on the composite manifold Y, where D_γ is the covariant differential (1.185) relative to the composite connection (1.196). We call $\widetilde D$ the *vertical covariant differential*.

Principal Connections

The above general approach to *connections as jet fields* is suitable to formulate the classical concept of *principal connections*. In this subsection, a structure group G of a principal bundle is assumed to be a real finite–dimensional Lie group (of positive dimension $\dim G > 0$).

A principal connection A on a principal bundle $P \to Q$ is defined to be a $G-$equivariant global jet field on P such that

$$j^1 R_g \circ A = A \circ R_g$$

for each canonical map (1.128). We have

$$\begin{aligned} &A \circ R_g = j^1 R_g \circ A, \qquad (g \in G), \\ &A = dq^\alpha \otimes (\partial_\alpha + A^m_\alpha(p) e_m), \qquad (p \in P), \\ &A^m_\alpha(qg) = A^m_\alpha(p) a d g^{-1}(e_m). \end{aligned} \tag{1.198}$$

A principal connection A determines splitting $TQ \hookrightarrow T^G P$ of the exact sequence (1.135). We will refer to

$$\mathbf{A} = A - \theta_Q = A^m_\alpha dq^\alpha \otimes e_m \tag{1.199}$$

as a *local connection form*.

Let $J^1(Q,P)$ be the 1–jet space of a principal bundle $P \to Q$ with a structure Lie group G. The jet prolongation

$$J^1(Q,P) \times J^1(Q \times G) \to J^1(Q,P)$$

of the canonical action (1.128) brings the jet bundle $J^1(Q,P) \to Q$ into a general affine bundle modelled on the group bundle

$$J^1(Q \times G) = G \times (T^*Q \otimes \mathfrak{g}_l) \tag{1.200}$$

over Q. However, the jet bundle $J^1(Q,P) \to Q$ fails to be a principal bundle since the group bundle (1.200) is not a trivial bundle over Q in general. At the same time, $J^1(Q,P)$ is the G principal bundle $C \times P \to C$ over the quotient

$$C = J^1(Q,P)/G \tag{1.201}$$

of the jet bundle $J^1(Q,P) \to P$ by the 1–jet prolongations of the canonical maps (1.128).

Let $J^1(Q,P)$ be the 1–jet space of a principal $G-$bundle $P \to Q$. Its quotient (1.201) by the jet prolongation of the canonical action R_G (1.128) is a fibre bundle over Q.

Given a bundle atlas of P and the associated bundle atlas of $V_G P$, the affine bundle C admits affine bundle coordinates (t, q^i, a^q_α), and its elements are represented by $T_G P-$valued 1–forms

$$a = dq^\alpha \otimes (\partial_\alpha + a^q_\alpha e_q) \tag{1.202}$$

on Q. One calls C (1.201) the connection bundle because its sections are naturally identified with principal connections on the principal bundle $P \to Q$ as follows.

There is the 1–1 correspondence between the principal connections on a principal bundle $P \to Q$ and the global sections of the quotient bundle

$$C = J^1(Q,P)/G \to Q.$$

We shall call C the principal connection bundle. It is an affine bundle modelled on the vector bundle

$$\overline{C} = T^*Q \otimes V^G P, \tag{1.203}$$

and there is the canonical vertical splitting

$$VC = C \times \overline{C}.$$

Given a bundle atlas Ψ^P of P, the principal connection bundle C admits the fibre coordinates (q^μ, k^m_μ) so that

$$(k^m_\mu \circ A)(q) = A^m_\mu(q)$$

are coefficients of the local connection one–form (1.199). The 1–jet space $J^1(Q,C)$ of C is with the adapted coordinates

$$(q^\mu, k^m_\mu, k^m_{\mu\lambda}). \tag{1.204}$$

The affine jet bundle $J^1(Q,C) \to C$ is modelled on the vector bundle

$$T^*Q \otimes (C \times T^*Q \otimes V^G P).$$

There exists the canonical splitting

$$J^1(Q,C) = (J^2P/G) \otimes (\wedge^2 T^*Q \otimes V^G P) \tag{1.205}$$

over C where

$$C_- = C \times \wedge^2 T^*Q \otimes V^G P$$

and $C_+ \to C$ is the affine bundle modelled on the vector bundle

$$\overline{C}_+ = \wedge^2 T^*Q \otimes V^G P.$$

In the coordinates (1.204), the splitting (1.205) reads

$$k^m_{\mu\lambda} = \frac{1}{2}(k^m_{\mu\lambda} + k^m_{\lambda\mu} + c^m_{nl}k^n_\alpha k^l_\mu) + \frac{1}{2}(k^m_{\mu\lambda} - k^m_{\lambda\mu} - c^m_{nl}k^n_\alpha k^l_\mu)$$

where c^k_{mn} are structure constants of the Lie algebra $\mathfrak{g}_r$ with respect to its basis $\{I_m\}$.

There are the corresponding canonical projections given by

$$\mathcal{S} = \pi_1 : J^1(Q,C) \to C_+, \qquad \mathcal{S}^m_{\lambda\mu} = k^m_{\mu\lambda} + k^m_{\lambda\mu} + c^m_{nl}k^n_\alpha k^l_\mu,$$
$$\text{and} \quad \mathcal{F} = \pi_2 : J^1(Q,C) \to C_-, \qquad \text{with}$$
$$\mathcal{F} = \frac{1}{2}\mathcal{F}^m_{\lambda\mu} dq^\alpha \wedge dq^\mu \otimes I_m, \qquad \mathcal{F}^m_{\lambda\mu} = k^m_{\mu\lambda} - k^m_{\lambda\mu} - c^m_{nl}k^n_\alpha k^l_\mu.$$

For every principal connection A, we observe that

$$\mathcal{F} \circ j^1 A = F, \qquad F = \frac{1}{2}F^m_{\lambda\mu} dq^\alpha \wedge dq^\mu \otimes I_m, \qquad F^m_{\lambda\mu} = \partial_\alpha A^m_\mu - \partial_\mu A^m_\alpha - c^m_{nk}A^n_\alpha A^k_\mu,$$

is the strength of A.

Given a symmetric linear connection K^* on the cotangent bundle T^*Q of Q, every principal connection A on a principal bundle P induces the connection

$$S_A : C \to C_+, \qquad S_A \circ A = \mathcal{S} \circ j^1 A,$$

on the principal connection bundle C. In the coordinates (1.204), the connection S_A reads

$$S_A{}^m_{\mu\lambda} = \frac{1}{2}[c^m_{nl}k^n_\alpha k^l_\mu + \partial_\mu A^m_\alpha + \partial_\alpha A^m_\mu - c^m_{nl}(k^n_\mu A^l_\alpha + k^n_\alpha A^l_\mu)] - K^\beta{}_{\mu\lambda}(A^m_\beta - k^m_\beta). \tag{1.206}$$

The $P-$associated bundle Y admits atlases $\Psi = \{U_\xi, \psi_\xi\}$ associated with atlases $\Psi^P = \{U_\xi, z_\xi\}$ of the principal bundle P as follows:

$$\psi^{-1}_\xi(q \times V) = [z_\xi(q)]_V(V), \qquad (q \in U_\xi),$$

where by $[p]_V$ is denoted the restriction of the canonical map $P \times V \to Y$ to the subset $p \times V$.

Every principal connection A on a principal bundle P induces the associated connection Γ on a $P-$associated bundle Y such that the following diagram commutes:

$$\begin{array}{ccc} J^1(Q,P) \times V & \longrightarrow & J^1(X,Y) \\ \Big\downarrow{\scriptstyle A \times \mathrm{Id}_V} & & \Big\downarrow{\scriptstyle \Gamma} \\ P \times V & \longrightarrow & Y \end{array}$$

We call it the associated principal connection. With respect to the associated atlases Ψ of Y and Ψ^P of P, this connection is written

$$\Gamma = dq^\alpha \otimes [\partial_\alpha + A^m_\mu(q) I_m{}^i{}_j y^j \partial_i] \tag{1.207}$$

where $A^m_\mu(q)$ are coefficients of the local connection one–form (1.199) and I_m are generators of the structure group G on the standard fibre V of the bundle Y. The curvature of the connection (1.207) reads

$$R^i_{\lambda\mu} = F^m_{\lambda\mu} I_m{}^i{}_j y^i.$$

1.4.4 Definition of a 2–Jet Space

As introduced above, a 2−jet is defined as a second–order equivalence class of functions having the same value and the same first derivatives at some designated point of the domain manifold. Recall that in mechanical settings, the 2–jets are local coordinate maps

$$j^2_t s : t \mapsto (t, x^i, \dot{x}^i, \ddot{x}^i).$$

In general, if we recursively apply the jet functor $\mathfrak{J} : Bun \to Jet$ to the jet bundles, we come to the higher order jet spaces (see [KMS93, Sar93, Sar95, GMS97, MS00a, Sar02a]).

In particular, taking the 1–jet space of the 1–jet bundle $J^1(X,Y) \longrightarrow X$, we get the *repeated jet space* $J^1(X, J^1(X,Y))$, which admits the adapted coordinates

$$(x^\alpha, y^i, y^i_\alpha, \widehat{y}^i_\mu, y^i_{\mu\alpha})$$

with transition functions

$$\widehat{y}'^i_\alpha = \frac{\partial x^\alpha}{\partial x'^\alpha} d_\alpha y'^i, \qquad y'^i_{\mu\alpha} = \frac{\partial x^\alpha}{\partial x'^\mu} d_\alpha y'^i_\alpha, \qquad d_\alpha = \partial_\alpha + \widehat{y}^j_\alpha \partial_j + y^j_{\nu\alpha}\partial^\nu_j.$$

The 2*−jet space* $J^2(X,Y)$ of a fibre bundle $Y \to X$ is coordinated by $(x^\alpha, y^i, y^i_\alpha, y^i_{\alpha\mu})$, with the local symmetry condition $y^i_{\alpha\mu} = y^i_{\mu\alpha}$. The manifold $J^2(X,Y)$ is defined as the set of equivalence classes $j^2_x s$ of sections $s^i : X \to Y$ of the bundle $Y \to X$, which are identified by their values $s^i(x)$ and the values of their first and second–order partial derivatives at points $x \in X$, respectively,

$$y^i_\alpha(j^2_x s) = \partial_\alpha s^i(x), \qquad y^i_{\alpha\mu}(j^2_x s) = \partial_\alpha \partial_\mu s^i(x).$$

In other words, the 2–jets $j^2_x s : x^\alpha \mapsto (x^\alpha, y^i, y^i_\alpha, y^i_{\alpha\mu})$, which are second–order equivalence classes of sections of the fibre bundle $Y \to X$, can be identified with their codomain set of adapted coordinates on $J^2(X,Y)$,

$$j^2_x s \equiv (x^\alpha, y^i, y^i_\alpha, y^i_{\alpha\mu}).$$

Let s be a section of a fibre bundle $Y \to X$, and let $j^1 s$ be its 1–jet prolongation to a section of the jet bundle $J^1(X,Y) \to X$. The latter induces the section $j^1 j^1 s$ of the repeated jet bundle $J^1(X, J^1(X,Y)) \to X$. This section takes its values into the 2–jet space $J^2(X,Y)$. It is called the 2–*jet prolongation* of the section s, and is denoted by $j^2 s$.

We have the following affine bundle monomorphisms

$$J^2(X,Y) \hookrightarrow \widehat{J}^2(X,Y)(X,Y) \hookrightarrow J^1(X, J^1(X,Y))$$

over $J^1(X,Y)$ and the canonical splitting

$$\widehat{J}^2(X,Y)(X,Y) = J^2(X,Y) \oplus (\wedge^2 T^*X \otimes VY), \qquad \text{given locally by}$$
$$y^i_{\alpha\mu} = \frac{1}{2}(y^i_{\alpha\mu} + y^i_{\mu\alpha}) + (\frac{1}{2}(y^i_{\alpha\mu} - y^i_{\mu\alpha}).$$

In particular, the repeated jet prolongation $j^1 j^1 s$ of a section $s : X \to Y$ of the fibre bundle $Y \to X$ is a section of the jet bundle $J^1(X, J^1(X,Y)) \to X$. It takes its values into $J^2(X,Y)$ and consists with the 2–jet prolongation $j^2 s$ of s:

$$j^1 j^1 s(x) = j^2 s(x) = j^2_x s.$$

Given a 2–jet space $J^2(X,Y)$ of the fibre bundle $Y \to X$, we have
(i) the fibred map $r_2 : J^2(Y,TY) \to TJ^2(X,Y)$, given locally by

$$(\dot{y}^i_\alpha, \dot{y}^i_{\alpha\mu}) \circ r_2 = ((\dot{y}^i)_\alpha - y^i_\mu \dot{x}^\mu_\alpha, (\dot{y}^i)_{\alpha\mu} - y^i_\mu \dot{x}^\mu_{\alpha\mu} - y^i_{\alpha\mu}\dot{x}^\mu_\alpha),$$

where $J^2(Y,TY)$ is the 2–jet space of the tangent bundle TY, and
(ii) the canonical isomorphism $VJ^2(X,Y) = J^2(X,VY)$, where $VJ^2(X,Y)$ is the vertical tangent bundle of the fibre bundle $J^2(X,Y) \to X$, and $J^2(X,VY)$ is the 2–jet space of the fibre bundle $VY \to X$.

As a consequence, every vector–field u on a fibre bundle $Y \to X$ admits the 2*−jet lift* to the projectable vector–field

$$j^2 u = r_2 \circ j^2 u : J^2(X,Y) \to TJ^2(X,Y).$$

In particular, if $u = u^\alpha \partial_\alpha + u^i \partial_i$ is a projectable vector–field on Y, its 2–jet lift reads

$$\begin{aligned} j^2 u = u^\alpha \partial_\alpha + u^i \partial_i + (\partial_\alpha u^i + y^j_\alpha \partial_j u^i - y^i_\mu \partial_\alpha u^\mu)\partial^\alpha_i \\ + [(\partial_\alpha + y^j_\alpha \partial_j + y^j_{\beta\alpha}\partial^\beta_j)(\partial_\alpha + y^k_\alpha \partial_k)u^i - y^i_\mu \dot{x}^\mu_{\alpha\beta} - y^i_{\mu\beta}\dot{x}^\mu_\alpha]\partial^{\alpha\beta}_i. \end{aligned} \tag{1.208}$$

Generalizations of the contact and complementary maps (1.161–1.162) to the 2–jet space $J^2(X,Y)$ read

$$\begin{aligned} &\lambda \ : \ J^2(X,Y) \to T^*X \otimes TJ^1(X,Y) \qquad \text{is locally given by} \\ &\lambda = dx^\alpha \otimes \widehat{\partial}_\alpha = dx^\alpha \otimes (\partial_\alpha + y^i_\alpha \partial_i + y^i_{\mu\alpha}\partial^\mu_i), \qquad \text{while} \end{aligned} \tag{1.209}$$

$$\begin{aligned} &\theta \ : \ J^2(X,Y) \to T^*J^1(X,Y) \otimes VJ^1(X,Y) \qquad \text{is locally given by} \\ &\theta = (dy^i - y^i_\alpha dx^\alpha) \otimes \partial_i + (dy^i_\mu - y^i_{\mu\alpha}dx^\alpha) \otimes \partial^\mu_i. \end{aligned} \tag{1.210}$$

The contact map (1.209) defines the canonical horizontal splitting of the exact sequence

$$0 \to VJ^1(X,Y) \hookrightarrow TJ^1(X,Y) \to J^1(X,Y) \times TX \to 0.$$

Hence, we get the canonical horizontal splitting of a projectable vector–field $j^1 u \equiv \overline{u}$ on $J^1(X,Y)$ over $J^2(X,Y)$:

$$j^1 u = u_H + u_V = u^\alpha[\partial_\alpha + y^i_\alpha + y^i_{\mu\alpha}] + [(u^i - y^i_\alpha u^\alpha)\partial_i + (u^i_\mu - y^i_{\mu\alpha}u^\alpha)\partial^\mu_i].$$

Building on the maps (1.209) and (1.210), one can get the horizontal splittings of the canonical tangent–valued 1–form on $J^1(X,Y)$,

$$\theta_{J^1(X,Y)} = dx^\alpha \otimes \partial_\alpha + dy^i \otimes \partial_i + dy^i_\mu \otimes \partial^\mu_i = \alpha + \theta$$

and the associated exterior differential

$$d = d_{\theta_{J^1(X,Y)}} = d_\alpha + d_\theta = d_H + d_V. \tag{1.211}$$

They are similar to the horizontal splittings (1.171) and (1.172).

A 2–jet field (resp. a 2–connection) $\overline{\Gamma}$ on a fibre bundle $Y \to X$ is defined to be a 1–jet field (resp. a 1–connection) on the jet bundle $J^1(X,Y) \to X$, i.e., $\overline{\Gamma}$ is a section (resp. a global section) of the bundle $J^1(X, J^1(X,Y)) \to J^1(X,Y)$.

In the coordinates $(y^i_\alpha, y^i_{(\mu)}, y^i_{\alpha\mu})$ of the repeated jet space $J^1(X, J^1(X,Y))$, a 2–jet field $\overline{\Gamma}$ is given by the expression

$$(y^i_\alpha, y^i_{(\mu)}, y^i_{\alpha\mu}) \circ \overline{\Gamma} = (y^i_\alpha, \overline{\Gamma}^i_{(\mu)}, \overline{\Gamma}^i_{\alpha\mu}).$$

Using the contact map (1.209), one can represent it by the tangent–valued horizontal 1–form on the jet bundle $J^1(X,Y) \to X$,

$$\overline{\Gamma} = dx^\mu \otimes (\partial_\mu + \overline{\Gamma}^i_{(\mu)}\partial_i + \overline{\Gamma}^i_{\alpha\mu}\partial^\alpha_i). \tag{1.212}$$

A 2–jet field $\overline{\Gamma}$ on a fibre bundle $Y \to X$ is called a *sesquiholonomic* (resp. *holonomic*) 2–jet field if it takes its values into the subbundle $\widehat{J}^2(X,Y)$ (resp. $J^2(X,Y)$) of $J^1(X, J^1(X,Y))$. We have the coordinate equality $\overline{\Gamma}^i_{(\mu)} = y^i_\mu$ for a sesquiholonomic 2–jet field and the additional equality $\overline{\Gamma}^i_{\alpha\mu} = \overline{\Gamma}^i_{\mu\alpha}$ for a holonomic 2–jet field.

Given a symmetric connection K on the cotangent bundle T^*X, every connection Γ on a fibre bundle $Y \to X$ induces the connection

$$j\Gamma = dx^\mu \otimes [\partial_\mu + \Gamma^i_\mu\partial_i + (\partial_\alpha\Gamma^i_\mu + \partial_j\Gamma^i_\mu y^j_\alpha - K^\alpha_{\alpha\mu}(y^i_\alpha - \Gamma^i_\alpha))\partial^\alpha_i]$$

on the jet bundle $J^1(X,Y) \to X$. Note that the curvature R of a connection Γ on a fibre bundle $Y \to X$ induces the soldering form $\overline{\sigma}_R$ on $J^1(X,Y) \to X$,

$$\overline{\sigma}_R = R^i_{\alpha\mu}dx^\mu \otimes \partial^\alpha_i.$$

1.4.5 Higher–Order Jet Spaces

The notion of 1– and 2–jet spaces is naturally extended to higher–order jet spaces. The *k–jet space* $J^k(X,Y)$ of a fibre bundle $Y \to X$ is defined as the disjoint union of the equivalence classes $j^k_x s$ of sections $s^i : X \to Y$ of the fibre bundle $Y \to X$, identified by their values and the values of the first k terms of their Taylor–series expansion at points x^i in the domain (base) manifold X. $J^k(X,Y)$ is a smooth manifold with the adapted coordinates $(x^\alpha, y^i_{\alpha_k\dots\alpha_1})$, where

$$y^i_{\alpha_k\cdots\alpha_1}(j^k_x s) = \partial_{\alpha_k}\cdots\partial_{\alpha_1}s^i(x), \qquad (0 \le k \le k).$$

The transformation law of these coordinates reads

$$y'^i_{\alpha+\alpha_k\dots\alpha_1} = \frac{\partial x^\mu}{\partial' x^\alpha} d_\mu y'^i_{\alpha_k\dots\alpha_1}, \tag{1.213}$$

where $\alpha + \alpha_k \dots \alpha_1 = (\alpha\alpha_k \dots \alpha_1)$ and

$$d_\alpha = \partial_\alpha + \sum_{|\alpha_k \dots \alpha_1| < k} y^i_{\alpha+\alpha_k \dots \alpha_1} \partial_i^{\alpha_k \dots \alpha_1} = \partial_\alpha + y^i_\alpha \partial_i + y^i_{\alpha\mu} \partial_i^\mu + \cdots$$

are higher–order total derivatives. These derivatives act on exterior forms on $J^k(X,Y)$ and obey the relations

$$d_\alpha(\phi \wedge \sigma) = d_\alpha(\phi) \wedge \sigma + \phi \wedge d_\alpha(\sigma), \qquad d_\alpha(d\phi) = d(d_\alpha(\phi)).$$

For example,

$$d_\alpha(dx^\mu) = 0, \qquad d_\alpha(dy^i_{\alpha_k \dots \alpha_1}) = dy^i_{\alpha+\alpha_k \dots \alpha_1}.$$

Let us also mention the following two operations: the horizontal projection h_0 given by the relations

$$h_0(dx^\alpha) = dx^\alpha, \qquad h_0(dy^i_{\alpha_k \cdots \alpha_1}) = y^i_{\mu\alpha_k \dots \alpha_1} dx^\mu, \tag{1.214}$$

and the horizontal differential

$$d_H(\phi) = dx^\alpha \wedge d_\alpha(\phi), \qquad d_H \circ d_H = 0, \qquad h_0 \circ d = d_H \circ h_0.$$

In a similar way, one can describe the *infinite–order jet space*, $J^\infty(X,Y)$, which can be coordinated by $(x^\alpha, y^i, \dots, y^i_{\alpha_1 \dots \alpha_r}, \dots)$, where $\alpha_1 \dots \alpha_r$ are collections of numbers modulo rearrangements, but it fails to be a well–behaved manifold in general. At the same time, one can introduce the sheaf of smooth functions on $J^\infty(X,Y)$ and define the differential calculus on $J^\infty(X,Y)$, with suitable notation for vector–fields, derivatives and differential forms just as like as in the finite order case (see [Sar93, Sar95, GMS97, MS00a, Sar02a]).

A vector–field u_k on the k–jet space $J^k(X,Y)$ is called *projectable vector–field* if for any $l < k$ there exists a vector–field u_k on $J^l(X,Y) \to X$ such that

$$u_l \circ \pi^k_l = T\pi^k_l \circ u_k.$$

The tangent map $T\pi^k_l$ sends projectable vector–fields on $J^k(X,Y)$ onto the projectable vector–fields on $J^l(X,Y)$.

Now consider projectable vector–fields u_k which are extension to the higher–order jet spaces of infinitesimal transformations of the fibre bundle $Y \to X$. The linear space of projectable vector–fields on $J^\infty(X,Y)$ is defined as the limit of the inverse system of projectable vector–fields on $k-$jet spaces. As a consequence, every projectable vector–field on the bundle $Y \to X$,

$$u = u^\alpha \partial_\alpha + u^i \partial_i,$$

induces a projectable vector–field $\overline{u}^\infty$ on $J^\infty(X,Y)$. We have its canonical decomposition

$$\begin{aligned} \overline{u}^\infty &= u^\infty_H + u^\infty_V, \\ u^\infty_H &= u^\alpha \widehat{\partial}^\infty_\alpha = u^\alpha(\partial_\alpha + y^i_\alpha \partial_i + \dots), \\ u^\infty_V &= \sum_{k=0}^\infty \widehat{\partial}^k_{\alpha_k} \dots \widehat{\partial}^1_{\alpha_1} {u_V}^i \partial_i^{\alpha_1 \dots \alpha_k}, \end{aligned} \tag{1.215}$$

where u_V is the vertical part of the splitting (1.170) of $\pi_0^{1*}u$. In particular, u_H^∞ is the canonical lift of the vector–field $\tau = u^\alpha \partial_\alpha$ on X onto $J^\infty(X,Y)$.

By the same limiting process we can introduce the notions of inner product of exterior forms and projectable vector–fields, the Lie bracket of projectable vector–fields and the Lie derivative of exterior forms by projectable vector–fields on $J^\infty(X,Y)$. All the usual identities are satisfied.

In particular, the notion of contact forms is extended to the forms

$$\widehat{d}y^i_{\alpha_1\dots\alpha_r} = dy^i_{\alpha_1\dots\alpha_r} - y^i_{\alpha_1\dots\alpha_r\nu}dx^\nu.$$

Let $\Omega^{r,k}$ denote the space of exterior forms on $J^\infty(X,Y)$ which are of the order r in the horizontal forms dx^ν and of the order k in the contact forms. Then, the space Ω^n of exterior $n-$forms on $J^\infty(X,Y)$ admits the unique decomposition

$$\Omega^n = \Omega^{n,0} \oplus \Omega^{n-1,1} \oplus \dots \oplus \Omega^{0,n}. \tag{1.216}$$

An exterior form is called a $k-$*contact form* if it belongs to the space $\Omega^{r,k}$. In particular, we have the $k-$*contact projection* $h_k : \Omega^n \longrightarrow \Omega^{n-k,k}$. For example, the *horizontal projection* h_0 performs the replacement $dy^i_{\alpha_1\dots\alpha_k} \longrightarrow y^i_{\alpha_1\dots\alpha_k\nu}dx^\nu$.

The exterior differential d on exterior forms on $J^\infty(X,Y)$ is decomposed into the sum

$$d = d_H + d_V \tag{1.217}$$

of the total differential operator

$$d_H\phi = \widehat{\partial}^\infty_\mu \phi_{\dots} dx^\mu \wedge \dots$$

and the vertical differential operator

$$d_V\phi = \frac{\partial \phi_{\dots}}{\partial y^i_{\alpha_1\dots\alpha_r}} \widehat{d}y^i_{\alpha_1\dots\alpha_r} \wedge \dots$$

These differentials satisfy the cohomology properties

$$d_H d_H = 0, \qquad d_V d_V = 0, \qquad d_V d_H + d_H d_V = 0.$$

Note that if σ is an exterior form on the $k-$jet space $J^k(X,Y)$, the decomposition (1.217) is reduced to

$$\begin{aligned} \pi_r^{r+1*} d\sigma &= d_H\sigma + d_V\sigma, \qquad \text{which implies} \\ h_0(d\sigma) &= d_H h_0(\sigma). \end{aligned}$$

1.4.6 Jets in Mechanics

As complex nonlinear mechanics is the most exact basis of all complex nonlinear dynamical systems considered in this book, we give here the first glimpse of mechanics on jet spaces.

Recall that in ordinary mechanics we have a *configuration manifold* M and the corresponding *velocity phase–space manifold* is its tangent bundle TM. However, in modern geometrical settings (see [GMS97, MOS99, MS00a, Sau89, Sar93, Sar95, Sar02a]), the configuration manifold of time–dependent mechanics is a fibre bundle $Q \to \mathbb{R}$, called the *configuration bundle*, coordinated by (t, q^i), where $t \in \mathbb{R}$ is a Cartesian coordinate on the time axis $\mathbb{R}$ with the transition functions $t' = t+\text{const}$. The corresponding velocity phase–space is the 1–jet space $J^1(\mathbb{R}, Q)$, which admits the adapted coordinates (t, q^i, q^i_t). It was proved in [Gia92, LM96, MS98] that every *dynamical equation* ξ defines a *connection* on the affine jet bundle $J^1(\mathbb{R}, Q) \to Q$, and *vice versa*.

Due to the canonical imbedding $J^1(\mathbb{R}, Q) \to TQ$, every dynamical connection induces a nonlinear connection on the tangent bundle $TQ \to Q$, and *vice versa*. As a consequence, every dynamical equation on Q induces an equivalent geodesic equation on the tangent bundle $TQ \to Q$ in accordance with the following proposition. Given a configuration bundle $Q \to \mathbb{R}$, coordinated by (t, q^i), and its 2–jet space $J^2(\mathbb{R}, Q)$, coordinated by $(t, q^i, q^i_t, q^i_{tt})$, any dynamical equation ξ on the configuration bundle $Q \to \mathbb{R}$,

$$q^i_{tt} = \xi^i(t, q^i, q^i_t) \tag{1.218}$$

is equivalent to the *geodesic equation* with respect to a connection $\widetilde{K}$ on the tangent bundle $TQ \to Q$,

$$\dot{t} = 1, \qquad \ddot{t} = 0, \qquad \ddot{q}^i = \widetilde{K}^i_0 + \widetilde{K}^i_j \dot{q}^j,$$

which fulfills the conditions

$$\widetilde{K}^0_\alpha = 0, \qquad \xi^i = \widetilde{K}^i_0 + q^j_t \widetilde{K}^i_j \mid_{t=1, \dot{q}^i = q^i_t}. \tag{1.219}$$

Recall that the 1–jet space $J^1(\mathbb{R}, Q)$ is defined as the set of equivalence classes $j^1_t c$ of sections $c^i : \mathbb{R} \to Q$ of the fibre bundle $Q \to \mathbb{R}$, which are identified by their values $c^i(t)$ and the values of their partial derivatives $\partial_t c^i = \partial_t c^i(t)$ at time points $t \in \mathbb{R}$. Also recall that there is the canonical imbedding

$$\lambda : J^1(\mathbb{R}, Q) \hookrightarrow TQ, \qquad \text{locally given by} \qquad \lambda = d_t = \partial_t + q^i_t \partial_i, \tag{1.220}$$

where d_t denotes the total time derivative. From now on, we will identify $J^1(\mathbb{R}, Q)$ with its image in the tangent bundle TQ. This is an affine bundle modelled over the vertical tangent bundle VQ of $Q \to \mathbb{R}$.

As a consequence of (1.220), every connection Γ on a fibre bundle $Q \to \mathbb{R}$,

$$\Gamma : Q \to J^1(\mathbb{R}, Q), \qquad \text{locally given by} \qquad \Gamma = dt \otimes (\partial_t + \Gamma^i \partial_i), \tag{1.221}$$

is identified with the nowhere vanishing vector–field on Q [MS98, MOS99],

$$\Gamma : Q \to J^1(\mathbb{R}, Q) \subset TQ, \qquad \text{locally given by} \qquad \Gamma = \partial_t + \Gamma^i \partial_i. \tag{1.222}$$

This is the horizontal lift of the standard vector–field ∂_t on $\mathbb{R}$ by means of the connection (1.221). Conversely, any vector–field Γ on Q such that $dt\rfloor\Gamma = 1$ defines a connection on $Q \to \mathbb{R}$. Therefore, the covariant differential associated with a connection Γ on $Q \to \mathbb{R}$ reads

$$D^G : J^1(\mathbb{R}, Q) \to VQ, \qquad \text{locally given by} \qquad \dot{q}^i \circ D_G = q^i_t - \Gamma^i.$$

Let $J^1(\mathbb{R}, J^1(\mathbb{R}, Q))$ denote the (repeated) 1–jet space of the jet bundle $J^1(\mathbb{R}, Q) \to \mathbb{R}$, coordinated by $(t, q^i, q^i_t, q^i_{(t)}, q^i_{tt})$. The corresponding 2–jet space $J^2(\mathbb{R}, Q)$ of the fibre bundle $Q \to \mathbb{R}$ is the holonomic subbundle $q^i_t = q^i_{(t)}$ of $J^1(\mathbb{R}, J^1(\mathbb{R}, Q))$, coordinated by $(t, q^i, q^i_t, q^i_{tt})$. There are the imbeddings

$$J^2(\mathbb{R}, Q) \xrightarrow{\mathring{\lambda}} TJ^1(\mathbb{R}, Q) \xrightarrow{T\lambda} TTQ, \qquad \text{with}$$

$$\mathring{\lambda} : (t, q^i, q^i_t, q^i_{tt}) \mapsto (t, q^i, q^i_t, \dot{t} = 1, \dot{q}^i = q^i_t, \dot{q}^i_t = q^i_{tt}). \tag{1.223}$$

$$T\lambda \circ \mathring{\lambda} : (t, q^i, q^i_t, q^i_{tt}) \mapsto (t, q^i, \dot{t} = 1, \dot{q}^i = q^i_t, \ddot{t} = 0, \ddot{q}^i = q^i_{tt}), \tag{1.224}$$

where $(t, q^i, \dot{q}^i, \ddot{q}^i)$ are holonomic coordinates on the second tangent bundle TTQ. This global geometrical structure of time–dependent mechanics is depicted in Figure 1.9.

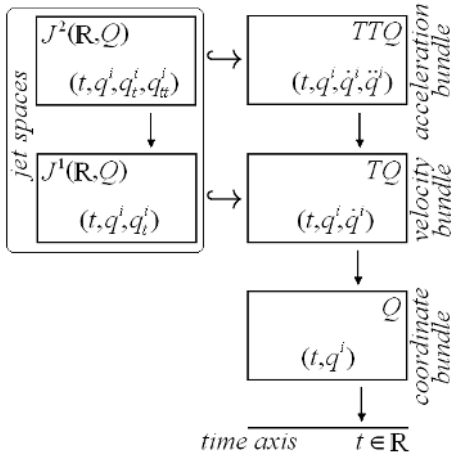


Fig. 1.9. Hierarchical geometrical structure of time–dependent mechanics. Note that (for simplicity) intermediate jet spaces, $J^1(\mathbb{R}, J^1(\mathbb{R}, Q))$ and $TJ^1(\mathbb{R}, Q)$, are not shown.

Therefore, a dynamical equation ξ on a configuration bundle $Q \to \mathbb{R}$, given in local coordinates by (1.218), is defined as the *geodesic equation* $\operatorname{Ker} D^\xi \subset J^2(\mathbb{R}, Q)$ for a holonomic connection ξ on the jet bundle $J^1(\mathbb{R}, Q) \to \mathbb{R}$. Due to the map (1.223), a holonomic connection ξ is represented by the horizontal vector–field on $J^1(\mathbb{R}, Q)$,

$$\xi = \partial_t + q^i_t\partial_i + \xi^i(q^\mu, q^i_t)\partial^t_i. \tag{1.225}$$

A dynamical equation ξ is said to be *conservative* if there exists a trivialization $Q \cong \mathbb{R}\times M$ such that the vector–field ξ (1.225) on $J^1(\mathbb{R}, Q) \cong \mathbb{R}\times TM$ is projectable onto TM. Then this projection

$$\Xi_\xi = \dot q^i\partial_i + \xi^i(q^j, \dot q^j)\dot\partial_i$$

is a second–order dynamical equation on a typical fibre M of $Q \to \mathbb{R}$,

$$\ddot q^i = \Xi^i_\xi. \tag{1.226}$$

Conversely, every second–order dynamical equation Ξ (1.226) on a manifold M can be seen as a conservative dynamical equation

$$\xi_\Xi = \partial_t + \dot q^i\partial_i + u^i\dot\partial_i$$

on the trivial fibre bundle $\mathbb{R}\times M \to \mathbb{R}$.

Now we can explore the fundamental relationship between the *holonomic connections* ξ (1.225) on the $1-$*jet bundle* $J^1(\mathbb{R}, Q) \to \mathbb{R}$ and the *dynamical connections* γ on the *affine* $1-$*jet bundle* $J^1(\mathbb{R}, Q) \to Q$, given by

$$\gamma = dq^\alpha \otimes (\partial_\alpha + \gamma^i_\alpha\partial^t_i), \qquad (q^\alpha \equiv (t, q^i),\ \partial_\alpha \equiv (\partial_t, \partial_i)). \tag{1.227}$$

Any dynamical connection γ (1.227) defines the holonomic connection ξ_γ on $J^1(\mathbb{R}, Q) \to \mathbb{R}$ [MS98]

$$\xi_\gamma = \partial_t + q^i_t\partial_i + (\gamma^i_0 + q^j_t\gamma^i_j)\partial^t_i.$$

Conversely, any holonomic connection ξ (1.225) on $J^1(\mathbb{R}, Q) \to \mathbb{R}$ defines the dynamical connection

$$\gamma_\xi = dt \otimes [\partial_t + (\xi^i - \frac{1}{2}q^j_t\partial^t_j\xi^i)\partial^t_i] + dq^j \otimes [\partial_j + \frac{1}{2}\partial^t_j\xi^i\partial^t_i]. \tag{1.228}$$

It follows that every dynamical connection γ (1.227) induces the dynamical equation (1.218) on the configuration bundle $Q \to \mathbb{R}$, rewritten here as

$$q^i_{tt} = \gamma^i_0 + q^j_t\gamma^i_j. \tag{1.229}$$

Different dynamical connections may lead to the same dynamical equation (1.229). The dynamical connection γ_ξ (1.228), associated with a dynamical equation, possesses the property

$$\gamma^k_i = \partial^t_i\gamma^k_0 + q^j_t\partial^t_i\gamma^k_j,$$

which implies the relation $\partial^t_j\gamma^k_i = \partial^t_i\gamma^k_j$. Such a dynamical connection is called symmetric. Let γ be a dynamical connection (1.227) and ξ_γ the corresponding

dynamical equation (1.4.6). Then the connection (1.228), associated with ξ_γ, takes the form

$$\gamma_{\xi_\gamma}{}^k_i = \frac{1}{2}(\gamma^k_i + \partial^t_i\gamma^k_0 + q^j_t\partial^t_i\gamma^k_j), \qquad \gamma_{\xi_\gamma}{}^k_0 = \xi^k - q^i_t\gamma_{\xi_\gamma}{}^k_i.$$

Note that $\gamma = \gamma_{\xi_\gamma}$ iff γ is symmetric.

To explore the relation between the connections γ (1.227) on the affine jet bundle $J^1(\mathbb{R}, Q) \to Q$ and the connections

$$K = dq^\alpha \otimes (\partial_\alpha + K^\beta_\alpha\dot\partial_\beta) \tag{1.230}$$

on the tangent bundle $TQ \to Q$, consider the diagram

$$\begin{array}{ccc} J^1(\mathbb{R}, J^1(\mathbb{R}, Q)) & \xrightarrow{j^1\lambda} & J^1(Q, TQ) \\ \Big\downarrow{\scriptstyle\gamma} & & \Big\downarrow{\scriptstyle K} \\ J^1(\mathbb{R}, Q) & \xrightarrow[\lambda]{} & TQ \end{array} \tag{1.231}$$

where $J^1(Q, TQ)$ is the 1–jet space of the tangent bundle $TQ \to Q$, coordinated by $(q^\alpha, \dot q^\alpha, \dot q^\alpha_\mu)$. The jet prolongation $j^1\lambda$ of the canonical imbedding λ (1.220) over Q reads

$$j^1\lambda : (t, q^i, q^i_t, q^i_{\mu t}) \mapsto (t, q^i, \dot t = 1, \dot q^i = q^i_t, \dot t_\mu = 0, \dot q^i_\mu = q^i_{\mu t}).$$

We have

$$j^1\lambda \circ \gamma : (t, q^i, q^i_t) \mapsto (t, q^i, \dot t = 1, \dot q^i = q^i_t, \dot t_\mu = 0, \dot q^i_\mu = \gamma^i_\mu),$$
$$K \circ \lambda : (t, q^i, q^i_t) \mapsto (t, q^i, \dot t = 1, \dot q^i = q^i_0, \dot t_\mu = K^0_\mu, \dot q^i_\mu = K^i_\mu).$$

It follows that the diagram (1.231) can be commutative only if the components K^0_μ of the connection K on $TQ \to Q$ vanish. Since the transition functions $t \to t'$ are independent of q^i, a connection

$$\widetilde K = dq^\alpha \otimes (\partial_\alpha + K^i_\alpha\dot\partial_i) \tag{1.232}$$

with the components $K^0_\mu = 0$ can exist on the tangent bundle $TQ \to Q$. It obeys the transformation law

$$K'^i_\alpha = (\partial_j x'^i K^j_\mu + \partial_\mu \dot x'^i)\frac{\partial q^\mu}{\partial x'^\alpha}. \tag{1.233}$$

Now the diagram (1.231) becomes commutative if the connections γ and $\widetilde K$ fulfill the relation

$$\gamma^i_\mu = K^i_\mu(t, q^i, \dot t = 1, \dot q^i = q^i_t),$$

which holds globally since the substitution of $\dot{q}^i = q^i_t$ into (1.233) restates the coordinate transformation law of γ.

Every dynamical equation (1.218) on the configuration bundle $Q \to \mathbb{R}$ can be written in the form

$$q^i_{tt} = K^i_0 \circ \lambda + q^j_t K^i_j \circ \lambda, \tag{1.234}$$

where $\widetilde{K}$ is a connection (1.232) on the tangent bundle $TQ \to Q$. Conversely, each connection $\widetilde{K}$ (1.232) on $TQ \to Q$ defines the dynamical equation (1.234) on $Q \to \mathbb{R}$.

Consider the geodesic equation (1.4.6) on TQ with respect to the connection $\widetilde{K}$. Its solution is a geodesic curve $c(t)$ which also satisfies the dynamical equation (1.218), and *vice versa*.

From the physical viewpoint, a reference frame in mechanics on a configuration bundle $Q \to \mathbb{R}$ sets a tangent vector at each point of Q which characterizes the velocity of an 'observer' at this point. Then any connection Γ on $Q \to \mathbb{R}$ is said to be such a reference frame [EMR95, MS98, MP94, Sar98].

Each connection Γ on a fibre bundle $Q \to \mathbb{R}$ defines an atlas of local constant trivializations of $Q \to \mathbb{R}$ whose transition functions are independent of t, and *vice versa*. One finds $\Gamma = \partial_t$ with respect to this atlas. In particular, there is 1–1 correspondence between the complete connections Γ (1.222) on $Q \to \mathbb{R}$ and the trivializations of this bundle.

Given a reference frame Γ, any connection K (1.230) on the tangent bundle $TQ \to Q$ defines the dynamical equation

$$\xi^i = (K^i_\alpha - \Gamma^i K^0_\alpha)\dot{q}^\alpha \mid_{t=1, \dot{q}^i = q^i_t} . \tag{1.235}$$

Given a connection Γ on the fibre bundle $Q \to \mathbb{R}$ and a connection K on the tangent bundle $TQ \to Q$, there is the connection $\widetilde{K}$ on $TQ \to Q$ with the components

$$\widetilde{K}^0_\alpha = 0, \qquad \widetilde{K}^i_\alpha = K^i_\alpha - \Gamma^i K^0_\alpha.$$

Our development of time–dependent mechanics on $k-$jet spaces is continued in section 2.1 below.

1.4.7 Jets and Action Principle

Recall that in the classical calculus of variations one studies functionals of the form

$$F_L(z) = \int_\Omega L(x, z, \nabla z)\, dx, \qquad (\text{with} \qquad \Omega \subset \mathbb{R}^n), \tag{1.236}$$

where $x = (x^1, \ldots, x^n)$, $dx = dx^1 \wedge \cdots \wedge dx^n$, $z = z(x) \in C^1(\bar{\Omega})$, and the *Lagrangian* $L = L(x, z, p)$ is a smooth function of x, z, and $p = (p_1, \ldots, p_n)$. The corresponding *Euler–Lagrangian equation*, describing functions $z(x)$ that are *stationary* for such a *functional*, is represented by the second–order PDE [BGG03]

$$\Delta z(x) = F'(z(x)).$$

For example, we may identify a function $z(x)$ with its graph $N \subset \mathbb{R}^{n+1}$, and take the Lagrangian

$$L = \sqrt{1 + ||p||^2},$$

whose associated functional $F_L(z)$ equals the area of the graph, regarded as a hypersurface in Euclidean space. The Euler–Lagrangian equation describing functions $z(x)$ stationary for this functional is $H = 0$, where H is the *mean curvature* of the graph N.

To study these Lagrangians and Euler–Lagrangian equations geometrically, we have to choose a class of admissible coordinate changes, and there are four natural candidates. In increasing order of generality, they are [BGG03]:

- Classical transformations, of the form $x' = x'(x)$, $z' = z'(z)$; in this situation, we think of (x, z, p) as coordinates on the space $J^1(\mathbb{R}^n, \mathbb{R})$ of $1-$jets of maps $\mathbb{R}^n \to \mathbb{R}$.
- Gauge transformations, of the form $x' = x'(x)$, $z' = z'(x, z)$; here, we think of (x, z, p) as coordinates on the space of $1-$jets of sections of a bundle $\mathbb{R}^{n+1} \to \mathbb{R}^n$, where $x = (x^1, \ldots, x^n)$ are coordinates on the base $\mathbb{R}^n$ and $z \in \mathbb{R}$ is a fibre coordinate.
- Point transformations, of the form $x' = x'(x, z)$, $z' = z'(x, z)$; here, we think of (x, z, p) as coordinates on the space of tangent hyperplanes

  $$\{dz - p_i dx^i\}^\perp \subset T_{(x^i, z)}(\mathbb{R}^{n+1})$$

  of the manifold $\mathbb{R}^{n+1}$ with coordinates $(x^1, \ldots, x^n, z)$.
- Contact transformations, of the form $x' = x'(x, z, p)$, $z' = z'(x, z, p)$, $p' = p'(x, z, p)$, satisfying the equation of differential $1-$forms

  $$dz' - p'_i dx^{i\prime} = f \cdot (dz - p_i dx^i)$$

  for some function $f(x, z, p) \neq 0$.

Classical calculus of variations primarily concerns the following features of the functional F_L (1.236).

The *first variation* $\delta F_L(z)$ is analogous to the derivative of a function, where $z = z(x)$ is thought of as an independent variable in an infinite–dimensional space of functions. The analog of the condition that a point be critical is the condition that $z(x)$ be stationary for all fixed–boundary variations. Formally, we write

$$\delta F_L(z) = 0,$$

which will give us a second–order scalar PDE for the unknown function $z(x)$ of the form

$$\partial_z L - \partial_{x^i}(\partial_{p_i} L) = 0, \tag{1.237}$$

namely the *Euler–Lagrangian equation* of the Lagrangian $L(x, z, p)$.

In this subsection we will study the PDE (1.237) in an invariant, geometrical setting, following [BGG03]. As a motivation for this geometrical approach, we note the fact that Lagrangian is invariant under the large class of *contact transformations*. Also, note that the Lagrangian L determines the functional F_L, but not vice versa. To see this, observe that if we add to $L(x, z, p)$ a *divergence term* and consider

$$L'(x, z, p) = L(x, z, p) + \sum (\partial_{x^i} K^i(x, z) + \partial_z K^i(x, z) p^i)$$

for functions $K^i(x, z)$, then by the *Green's theorem*, the functionals F_L and $F_{L'}$ differ by a constant depending only on values of z on $\partial\Omega$. L and L' have the same Euler–Lagrangian equations.

Also, there is a relationship between symmetries of a Lagrangian L and conservation laws for the corresponding *Euler–Lagrangian equations*, described by the *Noether theorem*. A subtlety here is that the group of symmetries of an equivalence class of Lagrangians may be strictly larger than the group of symmetries of any particular representative. We will investigate how this discrepancy is reflected in the space of conservation laws, in a manner that involves global topological issues.

Finally, one considers the second variation $\delta^2 F_L$, analogous to the Hessian of a smooth function, usually with the goal of identifying local minima of the functional. There has been a great deal of analytic work done in this area for classical variational problems, reducing the problem of local minimization to understanding the behavior of certain Jacobi operators, but the geometrical theory is not as well–developed as that of the first variation and the Euler–Lagrangian equations.

Now we turn to *multi–index notation* [Gri83a, BGG03, BM82]. An exterior differential system (EDS) is a pair (M, E) consisting of a smooth manifold M and a homogeneous, differentially closed ideal $E \subseteq \Omega^*(M)$ in the algebra of smooth differential forms on M. Some of the EDSs that we study are differentially generated by the sections of a smooth subbundle $I \subseteq T^*M$ of the cotangent bundle of M; this subbundle, and sometimes its space of sections, is called a *Pfaffian system* on M. It will be useful to use the notation $\{\alpha, \beta, \ldots\}$ for the (two–sided) *algebraic ideal* generated by forms α, β,..., and to use the notation $\{I\}$ for the algebraic ideal generated by the sections of a Pfaffian system $I \subseteq T^*M$. An *integral manifold* of an EDS (M, E) is a *submanifold immersion* $\iota : N \hookrightarrow M$ for which $\varphi_N \stackrel{def}{=} \iota^*\varphi = 0$ for all $\varphi \in E$. Integral manifolds of Pfaffian systems are defined similarly.

A differential form φ on the total space of a fibre bundle $\pi : E \to B$ is said to be *semibasic* if its contraction with any vector–field tangent to the fibers of π vanishes, or equivalently, if its value at each point $e \in E$ is the pull–back via π_e^* of some form at $\pi(e) \in B$. Some authors call such a form *horizontal*. A stronger condition is that φ be *basic*, meaning that it is locally (in open subsets of E) the pull–back via π^* of a form on the base B [BGG03].

If $(\omega^1, \ldots, \omega^n)$ is an ordered basis for a vector space V, then corresponding to a *multi–index* $I = (i_1, \ldots, i_k)$ is the $k-$vector

$$\omega^I = \omega^{i_1} \wedge \cdots \wedge \omega^{i_k} \in \wedge^k(V).$$

and for the complete multi–index we define

$$\omega = \omega^1 \wedge \cdots \wedge \omega^n.$$

Letting $(e_1, \ldots, e_n)$ be a dual basis for V^*, we also define the $(n-k)-$vector

$$\omega_{(I)} = e_I \rfloor \omega = e_{i_k} \rfloor (e_{i_{k-1}} \rfloor \cdots (e_{i_1} \rfloor \omega) \cdots).$$

This $\omega_{(I)}$ is, up to sign, just ω^{I_c}, where I_c is a multi–index complementary to I.

Recall that a *contact manifold* (M, I) is a smooth manifold M of dimension $2n+1$, with a distinguished line subbundle $I \subset T^*M$ of the cotangent bundle which is *non–degenerate* in the sense that for any local $1-$form θ generating I,

$$\theta \wedge (d\theta)^n \neq 0.$$

For example, A $1-$jet is an equivalence class of functions having the same value and the same first derivatives at some designated point of the domain. On the space $J^1(\mathbb{R}^n, \mathbb{R})$ of $1-$jets of functions, we can take coordinates (x^i, z, p_i) corresponding to the jet at $(x^i) \in \mathbb{R}^n$ of the linear function $f(\bar{x}) = z + p_i(\bar{x}^i - x^i)$. Then we define the *contact form*

$$\theta = dz - p_i dx^i,$$

for which

$$d\theta = -dp_i \wedge dx^i,$$

so the non–degeneracy condition $\theta \wedge (d\theta)^n \neq 0$ is apparent. In fact, the *Pfaff theorem* [BGG03] implies that every contact manifold is locally isomorphic to this example; that is, every contact manifold (M, I) has local coordinates (x^i, z, p_i) for which the form $\theta = dz - p_i dx^i$ generates I.

Let (M, I) be a contact manifold of dimension $2n + 1$, and assume that I is generated by a global, non–vanishing section $\theta \in \Gamma(I)$; this assumption only simplifies our notation, and would in any case hold on a double–cover of M. Sections of I generate the *contact differential ideal*

$$I = \{\theta, d\theta\} \subset \Omega^*(M)$$

in the exterior algebra of differential forms on M. A *Legendre submanifold* of M is an immersion $\iota : N \hookrightarrow M$ of an nD submanifold N such that $\iota^*\theta = 0$ for any contact form $\theta \in \Gamma(I)$; in this case $\iota^* d\theta = 0$ as well, so a Legendre submanifold is the same thing as an integral manifold of the differential ideal I. In Pfaff coordinates with $\theta = dz - p_i dx^i$, one such integral manifold is

$N_0 = \{z = p_i = 0\}$. To see other Legendre submanifolds 'near' this one, note than any submanifold C^1-close to N_0 satisfies the *independence condition* [BGG03]

$$dx^1 \wedge \cdots \wedge dx^n \neq 0,$$

and can therefore be described locally as a graph

$$N = \{(x^i, z(x), p_i(x))\}.$$

In this case, we have

$$\theta|_N = 0 \qquad \text{iff} \qquad p_i(x) = \partial_{x^i} z(x).$$

Therefore, N is determined by the function $z(x)$, and conversely, every function $z(x)$ determines such an N; we informally say that 'the generic Legendre submanifold depends locally on one arbitrary function of n variables'. Legendre submanifolds of this form, with $dx|_N \neq 0$, are called *transverse*.

Now, we are interested in functionals given by triples (M, I, Λ), where (M, I) is a $(2n+1)$D contact manifold, and $\Lambda \in \Omega^n(M)$ is a differential form of degree n on M; such a Λ will be referred to as a *Lagrangian* on (M, I) [BGG03]. We then define a functional on the set of smooth, compact Legendre submanifolds $N \subset M$, possibly with boundary ∂N, by

$$F_\Lambda(N) = \int_N \Lambda.$$

The classical variational problems described above may be recovered from this notion by taking $M = J^1(\mathbb{R}^n, \mathbb{R}) \cong \mathbb{R}^{2n+1}$ with coordinates (x^i, z, p_i), I generated by $\theta = dz - p_i dx^i$, and $\Lambda = L(x^i, z, p_i)dx$. This formulation also admits certain functionals depending on second derivatives of $z(x)$, because there may be dp_i-terms in Λ. Later, we will restrict attention to a class of functionals which, possibly after a contact transformation, can be expressed without second derivatives.

Suppose given a Lagrangian $\Lambda \in \Omega^n(M)$ on a *contact manifold* (M, I), and a fixed–boundary variation of Legendre submanifold $F : N \times [0, 1] \to M$; we wish to calculate $\frac{d}{dt}(\int_{N_t} \Lambda)$.

To do this, first recall the calculation of the *Poincaré–Cartan form* for the equivalence class $[\Lambda] \in \bar{H}^n$. Because $I^{n+1} = \Omega^{n+1}(M)$, we can write [BGG03]

$$d\Lambda = \theta \wedge \alpha + d\theta \wedge \beta = \theta \wedge (\alpha + d\beta) + d(\theta \wedge \beta),$$

and then

$$\Pi = \theta \wedge (\alpha + d\beta) = d(\Lambda - \theta \wedge \beta). \tag{1.238}$$

We are looking for conditions on a *Legendre submanifold* $f : N \hookrightarrow M$ to be *stationary* for $[\Lambda]$ under all fixed–boundary variations, in the sense that $\frac{d}{dt}\big|_{t=0} (\int_{N_t} \Lambda) = 0$ whenever $F|_{t=0} = f$. We calculate

$$\partial_t \int_{N_t} \Lambda = \partial_t \int_{N_t} (\Lambda - \theta \wedge \beta) = \int_{N_t} L_{\partial_t}(\Lambda - \theta \wedge \beta) = \int_{N_t} \partial_t \rfloor \Pi.$$

One might express this result as

$$\delta(F_\Lambda)_N(v) = \int_N v \rfloor f^* \Pi,$$

where the variational vector–field v, lying in the space $\Gamma_0(f^*TM)$ of sections of f^*TM vanishing along ∂N, plays the role of ∂_t. The condition $\Pi \equiv 0(\mathrm{mod}\{I\})$ allows us to write $\Pi = \theta \wedge \Psi$ for some $n-$form Ψ, not uniquely determined, and we have [BGG03]

$$\left.\frac{d}{dt}\right|_{t=0} \int_{N_t} \Lambda = \int_N g\, f^* \Psi,$$

where $g = (\partial_t \rfloor F^*\theta)|_{t=0}$. It was shown previously that this g could locally be chosen arbitrarily in the interior N^o, so the necessary and sufficient condition for a Legendre submanifold $f : N \hookrightarrow M$ to be stationary for F_Λ is that $f^*\Psi = 0$.

In the particular classical situation where $M = \{(x^i, z, p_i)\}$, $\theta = dz - p_i dx^i$, and $\Lambda = L(x, z, p)dx$, we have

$$d\Lambda = L_z \theta \wedge dx + L_{p_i} dp_i \wedge dx = \theta \wedge L_z dx - d\theta \wedge L_{p_i} dx_{(i)},$$

so referring to (1.238),

$$\Pi = \theta \wedge (L_z dx - d(L_{p_i} dx_{(i)})) = \theta \wedge \Psi.$$

Now, for a transverse Legendre submanifold $N = \{(x^i, z(x), z_{x^i}(x))\}$, we have $\Psi|_N = 0$ iff (1.237) is valid along N.

Later, (see section 2.2.4 below) we will extend the jet–action formalism presented here – to the rigorous (and elegant) jet formulation of path–integrals in physical field systems.

1.5 Path Integrals: Extending Smooth Geometrical Machinery

The machinery of geometrical dynamics, as presented so far, is: (i) rigorous, (ii) elegant, and (iii) powerful – as a tool for understanding, prediction and control of complex nonlinear systems. However, due to its *smooth nature*, it is *limited* to modelling of *deterministic* and *continuous–time* dynamical systems only. Naturally, the question arises: is it possible to extend this smooth machinery so to be able to effectively deal also with *probabilistic* and *discrete–time* dynamical systems, like e.g., Markov chains? And the answer is: Yes. Namely, in the very core of the XX Century geometrodynamics, there is a powerful

conceptual and computational tool that is 'by default' used as a starting point for virtually every new physical theory – the celebrated *Feynman path integral*. In the path–integral formalism, we first formulate the specific classical *action* of a new theory, and subsequently perform its quantization by means of the associated *amplitude*. This *action–amplitude picture* is the core structure in any new physical theory. Unlike mathematical manifolds, bundles and jets, the path integral is an invention of the physical mind of Richard (Dick) Feynman. Its *virtual paths* are in general neither deterministic not smooth, although they include bundles and jets of deterministic and smooth paths, as well as Markov chains. Yet, it is essentially a (broader) geometrical dynamics, with its Riemannian and symplectic versions, among many others. At the beginning, it worked only for *conservative* physical systems. Today it includes also *dissipative structures*, as well as various *sources and sinks*. Its smooth part reveals all celebrated equations of the 20th Century, both classical and quantum. It is the core of modern quantum gravity and string theory. It is arguably the most important construct of mathematical physics. At the edge of a new millennium, if you asked a typical theoretical physicist: what will be your main research tool in the new millennium, he/she would most probably say: path integral. And today, we see it moving out from physics, into the realm of social sciences. Finally, since Feynman's fairly intuitive invention of the path integral [Fey51], a lot of research has been done to make it mathematically rigorous (see e.g., [Loo99, Loo00, AFH86, Kla97, SK98a, Kla00]).

1.5.1 Intuition Behind a Path Integral

Classical Probability Concept

Recall that a *random variable* X is defined by its *distribution function* $f(x)$. Its *probabilistic description* is based on the following rules: (i) $P(X = x_i)$ is the probability that $X = x_i$; and (ii) $P(a \leq X \leq b)$ is the probability that X lies in a closed interval $[a, b]$. Its statistical description is based on: (i) μ_X or $E(X)$ is the mean or expectation of X; and (ii) σ_X is the standard deviation of X. There are two cases of random variables: discrete and continuous, each having its own probability (and statistics) theory.

Discrete Random Variable

A discrete random variable X has only a countable number of values $\{x_i\}$. Its distribution function $f(x_i)$ has the following properties:

$$P(X = x_i) = f(x_i), \qquad f(x_i) \geq 0, \qquad \sum_i f(x_i)\, dx = 1.$$

Statistical description of X is based on its discrete mean value μ_X and standard deviation σ_X, given respectively by

$$\mu_X = E(X) = \sum_i x_i f(x_i), \qquad \sigma_X = \sqrt{E(X^2) - \mu_X^2}.$$

Continuous Random Variable

Here $f(x)$ is a piecewise continuous function such that:

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx, \qquad f(x) \geq 0, \qquad \int_{-\infty}^{\infty} f(x)\,dx = \int_{\mathbb{R}} f(x)\,dx = 1.$$

Statistical description of X is based on its continuous mean μ_X and standard deviation σ_X, given respectively by

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x f(x)\,dx, \qquad \sigma_X = \sqrt{E(X^2) - \mu_X^2}.$$

Now, let us observe the similarity between the two descriptions. The same kind of similarity between discrete and continuous quantum spectrum stroke Dirac when he suggested the combined integral approach, that he denoted by $\int\!\!\!\!\!\!\sum$ – meaning 'both integral and sum at once': summing over discrete spectrum and integration over continuous spectrum.

To emphasize this similarity even further, as well as to set–up the stage for the path integral, recall the notion of a *cumulative distribution function* of a random variable X, that is a function $F : \mathbb{R} \longrightarrow \mathbb{R}$, defined by

$$F(a) = P(X) \leq a.$$

In particular, suppose that $f(x)$ is the distribution function of X. Then

$$F(x) = \sum_{x_i \leq x} f(x_i), \qquad \text{or} \qquad F(x) = \int_{-\infty}^{\infty} f(t)\,dt,$$

according to as x is a discrete or continuous random variable. In either case, $F(a) \leq F(b)$ whenever $a \leq b$. Also,

$$\lim_{x \longrightarrow -\infty} F(x) = 0 \qquad \text{and} \qquad \lim_{x \longrightarrow \infty} F(x) = 1,$$

that is, $F(x)$ is monotonic and its limit to the left is 0 and the limit to the right is 1. Furthermore, its cumulative probability is given by

$$P(a \leq X \leq b) = F(b) - F(a),$$

and the Fundamental Theorem of Calculus tells us that, in the continuum case,

$$f(x) = \partial_x F(x).$$

General Markov Stochastic Dynamics

Recall that *Markov stochastic process* is a random process characterized by a *lack of memory*, i.e., the statistical properties of the immediate future are uniquely determined by the present, regardless of the past [Gar85].

For example, a *random walk* is an example of the *Markov chain*, i.e., a discrete–time Markov process, such that the motion of the system in consideration is viewed as a sequence of states, in which the transition from one state to another depends only on the preceding one, or the probability of the system being in state k depends only on the previous state $k-1$. The property of a Markov chain of prime importance in biomechanics is the existence of an *invariant distribution of states*: we start with an initial state x_0 whose absolute probability is 1. Ultimately the states should be distributed according to a specified distribution.

Between the pure deterministic dynamics, in which all DOF of the system in consideration are explicitly taken into account, leading to classical dynamical equations, for example in Hamiltonian form (1.17), i.e.,

$$\dot{q}^i = \partial_{p_i} H, \qquad \dot{p}_i = -\partial_{q^i} H$$

– and pure stochastic dynamics (Markov process), there is so–called *hybrid dynamics*, particularly *Brownian dynamics*, in which some of DOF are represented only through their *stochastic influence* on others. As an example, suppose a system of particles interacts with a viscous medium. Instead of specifying a detailed interaction of each particle with the particles of the viscous medium, we represent the medium as a *stochastic force* acting on the particle. The stochastic force *reduces the dimensionally* of the dynamics.

Recall that the Brownian dynamics represents the phase–space trajectories of a collection of particles that individually obey *Langevin rate equations* in the field of force (i.e., the particles interact with each other via some deterministic force). For a free particle, the Langevin equation reads [Gar85]:

$$m\dot{v} \;=\; R(t) \;-\; \beta v,$$

where m denotes the mass of the particle and v its velocity. The right–hand side represent the coupling to a *heat bath*; the effect of the random force $R(t)$ is to heat the particle. To balance overheating (on the average), the particle is subjected to *friction* β. In humanoid dynamics this is performed with the Rayleigh–Van der Pol's *dissipation*. Formally, the solution to the Langevin equation can be written as

$$v(t) \;=\; v(0)\,\exp\left(-\frac{\beta}{m}t\right) \;+\; \frac{1}{m}\int_0^t \exp[-(t-\tau)\beta/m]\,R(\tau)\,d\tau,$$

where the integral on the right–hand side is a *stochastic integral* and the solution $v(t)$ is a random variable. The stochastic properties of the solution depend significantly on the stochastic properties of the random force $R(t)$. In

the Brownian dynamics the random force $R(t)$ is Gaussian distributed. Then the problem boils down to finding the solution to the Langevin stochastic differential equation with the supplementary condition (mean zero and variance)

$$< R(t) >= 0, \qquad < R(t)\, R(0) >= 2\beta k_B T \delta(t),$$

where $< . >$ denotes the mean value, T is temperature, k_B-*equipartition* (i.e., uniform distribution of energy) coefficient, Dirac $\delta(t)-$function.

Algorithm for computer simulation of the Brownian dynamics (for a single particle) can be written as [Hee90]:

1. Assign an initial position and velocity.
2. Draw a random number from a Gaussian distribution with mean zero and variance.
3. Integrate the velocity to get v^{n+1}.
4. Add the random component to the velocity.

Another approach to taking account the coupling of the system to a heat bath is to subject the particles to collisions with *virtual particles* [Hee90]. Such collisions are imagined to affect only momenta of the particles, hence they affect the kinetic energy and introduce fluctuations in the total energy. Each stochastic collision is assumed to be an instantaneous event affecting only one particle.

The collision–coupling idea is incorporated into the Hamiltonian model of dynamics (1.17) by adding a stochastic force $R_i = R_i(t)$ to the $\dot{p}$ equation

$$\dot{q}^i = \partial_{p_i} H, \qquad \dot{p}_i = -\partial_{q^i} H + R_i(t).$$

On the other hand, the so–called *Ito stochastic integral* represents a kind of classical Riemann–Stieltjes integral from linear functional analysis, which is (in 1D case) for an arbitrary time–function $G(t)$ defined as the *mean square limit*

$$\int_{t_0}^{t} G(t) dW(t) = ms \lim_{n\to\infty} \{\sum_{i=1}^{n} G(t_{i-1}[W(t_i) - W(t_{i-1}]\}.$$

Now, the general ND Markov process can be defined by *Ito* stochastic differential equation (SDE),

$$\begin{aligned} dx_i(t) &= A_i[x^i(t), t]dt + B_{ij}[x^i(t), t]\, dW^j(t), \\ x^i(0) &= x_{i0}, \qquad (i, j = 1, \ldots, N) \end{aligned}$$

or corresponding *Ito stochastic integral equation*

$$x^i(t) = x^i(0) + \int_0^t ds\, A_i[x^i(s), s] + \int_0^t dW^j(s)\, B_{ij}[x^i(s), s],$$

in which $x^i(t)$ is the variable of interest, the vector $A_i[x(t), t]$ denotes deterministic *drift*, the matrix $B_{ij}[x(t), t]$ represents continuous stochastic *diffusion fluctuations*, and $W^j(t)$ is an N-variable *Wiener process* (i.e., generalized Brownian motion) [Wie61], and $dW^j(t) = W^j(t + dt) - W^j(t)$.

Now, there are three well–known special cases of the *Chapman–Kolmogorov equation* (see [Gar85]):

1. When both $B_{ij}[x(t),t]$ and $W(t)$ are zero, i.e., in the case of pure deterministic motion, it reduces to the *Liouville equation*

$$\partial_t P(x',t'|x'',t'') = -\sum_i \frac{\partial}{\partial x^i}\left\{A_i[x(t),t]\, P(x',t'|x'',t'')\right\}.$$

2. When only $W(t)$ is zero, it reduces to the *Fokker–Planck equation*

$$\begin{aligned}\partial_t P(x',t'|x'',t'') &= -\sum_i \frac{\partial}{\partial x^i}\left\{A_i[x(t),t]\, P(x',t'|x'',t'')\right\} \\ &\quad + \frac{1}{2}\sum_{ij} \frac{\partial^2}{\partial x^i \partial x^j}\left\{B_{ij}[x(t),t]\, P(x',t'|x'',t'')\right\}.\end{aligned}$$

3. When both $A_i[x(t),t]$ and $B_{ij}[x(t),t]$ are zero, i.e., the state–space consists of integers only, it reduces to the *Master equation* of discontinuous jumps

$$\begin{aligned}&\partial_t P(x',t'|x'',t'') \qquad = \\ &\int dx\left\{W(x'|x'',t)\, P(x',t'|x'',t'') - W(x''|x',t)\, P(x',t'|x'',t'')\right\}.\end{aligned}$$

The *Markov assumption* can now be formulated in terms of the conditional probabilities $P(x^i,t_i)$: if the times t_i increase from right to left, the conditional probability is determined entirely by the knowledge of the most recent condition. Markov process is generated by a set of conditional probabilities whose probability–density $P = P(x',t'|x'',t'')$ evolution obeys the general *Chapman–Kolmogorov integro–differential equation*

$$\begin{aligned}\partial_t P &= -\sum_i \frac{\partial}{\partial x^i}\left\{A_i[x(t),t]\, P\right\} \\ &\quad + \frac{1}{2}\sum_{ij} \frac{\partial^2}{\partial x^i \partial x^j}\left\{B_{ij}[x(t),t]\, P\right\} + \int dx\left\{W(x'|x'',t)\, P - W(x''|x',t)\, P\right\}\end{aligned}$$

including *deterministic drift*, *diffusion fluctuations* and *discontinuous jumps* (given respectively in the first, second and third terms on the r.h.s.).

It is this general Chapman–Kolmogorov integro–differential equation, with its conditional probability density evolution, $P = P(x',t'|x'',t'')$, that we are going to model by various forms of the Feynman path integral, providing us with the physical insight behind the abstract (conditional) probability densities.

Quantum Probability Concept

An alternative concept of probability, the so–called *quantum probability*, is based on the following physical facts (elaborated in detail in this section):

1. *The time–dependent Schrödinger equation* represents a *complex–valued generalization* of the real–valued *Fokker–Planck equation* for describing the spatio–temporal *probability density function* for the system exhibiting *continuous–time Markov stochastic process.*
2. The *Feynman path integral* $\oint$ is a generalization of the time–dependent Schrödinger equation, including both continuous–time and discrete–time Markov stochastic processes.
3. Both Schrödinger equation and path integral give 'physical description' of any system they are modelling in terms of its physical energy, instead of an abstract probabilistic description of the Fokker–Planck equation.

Therefore, the *Feynman path integral* $\oint$, as a generalization of the time–dependent Schrödinger equation, gives a unique physical description for the general Markov stochastic process, in terms of the physically based generalized probability density functions, valid both for continuous–time and discrete–time Markov systems.

Basic consequence: a different way for calculating probabilities. The difference is rooted in the fact that *sum of squares is different from the square of sums*, as is explained in the following text.

Namely, in Dirac–Feynman quantum formalism, each possible route from the initial system state A to the final system state B is called a *history.* This history comprises any kind of a route (see Figure 1.10), ranging from continuous and smooth deterministic (mechanical–like) paths to completely discontinues and random Markov chains (see, e.g., [Gar85]). Each history (labelled by index i) is quantitatively described by a *complex number*[14] z_i called the 'individual transition amplitude'. Its absolute square, $|z_i|^2$, is called the *individual transition probability.* Now, the *total transition amplitude* is the sum of all individual transition amplitudes, $\sum_i z_i$, called the *sum–over–histories.* The absolute square of this sum–over–histories, $|\sum_i z_i|^2$, is the *total transition probability.*

In this way, the overall probability of the system's transition from some initial state A to some final state B is given *not* by adding up the probabilities for each history–route, but by 'head–to–tail' adding up the sequence of amplitudes making–up each route first (i.e., performing the sum–over–histories) – to get the total amplitude as a 'resultant vector', and then squaring the total amplitude to get the overall transition probability.

[14] Recall that a *complex number* $z = x + \mathrm{i}y$, where $\mathrm{i} = \sqrt{-1}$ is the *imaginary unit*, x is the *real part* and y is the *imaginary part*, can be represented also in its *polar form*, $z = r(\cos\theta + \mathrm{i}\sin\theta)$, where the radius vector in the complex plane, $r = |z| = \sqrt{x^2 + y^2}$, is the *modulus* or *amplitude*, and angle θ is the *phase*; as well as in its exponential form $z = r\mathrm{e}^{\mathrm{i}\theta}$. In this way, complex numbers actually represent 2D vectors with usual vector 'head–to–tail' addition rule.

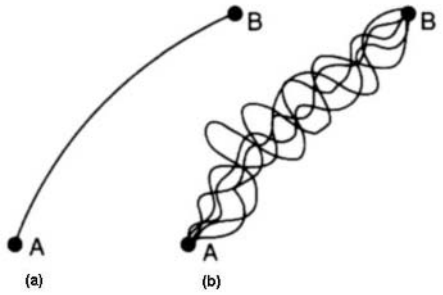


Fig. 1.10. Two ways of physical *transition* from an *initial state* A to the corresponding *final state* B. (a) Classical physics proposes *a single deterministic trajectory*, minimizing the total system's energy. (b) Quantum physics proposes *a family of Markov stochastic histories*, namely *all possible routes* from A to B, both continuous–time and discrete–time Markov chains, each giving an equal contribution to the total *transition probability*.

Quantum Coherent States

Recall that a *quantum coherent state* is a specific kind of quantum state of the quantum harmonic oscillator whose dynamics most closely resemble the oscillating behavior of a classical harmonic oscillator. It was the first example of quantum dynamics when Erwin Schrödinger derived it in 1926 while searching for solutions of the *Schrödinger equation* that satisfy the *correspondence principle*. The quantum harmonic oscillator and hence, the coherent state, arise in the quantum theory of a wide range of physical systems. For instance, a coherent state describes the oscillating motion of the particle in a quadratic potential well. In the quantum electrodynamics and other bosonic quantum field theories they were introduced by the 2005 Nobel Prize winning work of Roy Glauber in 1963 [Gla63a, Gla63b]. Here the coherent state of a field describes an oscillating field, the closest quantum state to a classical sinusoidal wave such as a continuous laser wave.

In classical optics, light is thought of as electromagnetic waves radiating from a source. Specifically, coherent light is thought of as light that is emitted by many such sources that are in phase. For instance, a light bulb radiates light that is the result of waves being emitted at all the points along the filament. Such light is incoherent because the process is highly random in space and time. On the other hand, in a laser, light is emitted by a carefully controlled system in processes that are not random but interconnected by stimulation and the resulting light is highly ordered, or coherent. Therefore a coherent state corresponds closely to the quantum state of light emitted by an ideal laser. Semi–classically we describe such a state by an electric field oscillating as a stable wave. Contrary to the coherent state, which is the most wave–like quantum state, the *Fock state* (e.g., a single photon) is the most particle–like state. It is indivisible and contains only one quanta of energy.

These two states are examples of the opposite extremes in the concept of *wave–particle duality*. A coherent state distributes its quantum–mechanical uncertainty equally, which means that the phase and amplitude uncertainty are approximately equal. Conversely, in a single–particle state the phase is completely uncertain.

Formally, the coherent state $|\alpha\rangle$ is defined to be the eigenstate of the annihilation operator a, i.e., $a|\alpha\rangle = \alpha|\alpha\rangle$. Note that since a is not Hermitian, $\alpha = |\alpha|e^{i\theta}$ is complex. $|\alpha|$ and θ are called the *amplitude* and *phase* of the state.

Physically, $a|\alpha\rangle = \alpha|\alpha\rangle$ means that a coherent state is left unchanged by the detection (or annihilation) of a particle. Consequently, in a coherent state, one has exactly the same probability to detect a second particle. Note, this condition is necessary for the coherent state's *Poisson detection statistics*. Compare this to a single–particle's Fock state: Once one particle is detected, we have zero probability of detecting another.

Now, recall that a *Bose–Einstein condensate* (BEC) is a collection of boson atoms that are all in the same quantum state. An approximate theoretical description of its properties can be derived by assuming the BEC is in a coherent state. However, unlike photons, atoms interact with each other so it now appears that it is more likely to be one of the *squeezed coherent states* (see [BSM97]). In quantum field theory and string theory, a generalization of coherent states to the case of infinitely many degrees of freedom is used to define a *vacuum state* with a different vacuum expectation value from the original vacuum.

Dirac's $< bra \,|\, ket >$ Transition Amplitude

Now, we are ready to move–on into the realm of quantum mechanics. Recall that P. Dirac [Dir82] described behavior of quantum systems in terms of complex–valued *ket–vectors* $|A>$ living in the Hilbert space $\mathcal{H}$, and their duals, *bra–covectors* $< B|$ (i.e., 1–forms) living in the *dual* Hilbert space $\mathcal{H}^*$.[15] The *Hermitian inner product* of kets and bras, the *bra–ket* $< B|A >$, is a

[15] Recall that a *norm* on a complex vector space $\mathcal{H}$ is a mapping from $\mathcal{H}$ into the complex numbers, $\|\cdot\| : \mathcal{H} \to \mathbb{C}$; $h \mapsto \|h\|$, such that the following set of norm–axioms hold:

(N1) $\|h\| \geq 0$ for all $h \in \mathcal{H}$ and $\|h\| = 0$ implies $h = 0$ (positive definiteness);

(N2) $\|\lambda\, h\| = |\lambda|\, \|h\|$ for all $h \in \mathcal{H}$ and $\lambda \in \mathbb{C}$ (homogeneity); and

(N3) $\|h_1 + h_2\| \leq \|h_1\| + \|h_2\|$ for all $h_1, h_2 \in \mathcal{H}$ (triangle inequality). The pair $(\mathcal{H}, \|\cdot\|)$ is called a *normed space*.

A *Hermitian inner product* on a complex vector space $\mathcal{H}$ is a mapping $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ such that the following set of inner–product–axioms hold:

(IP1) $\langle h\, h_1 + h_2 \rangle = \langle h\, h_1 + h\, h_2 \rangle$;

(IP2) $\langle \alpha\, h, h_1 \rangle = \underline{\alpha}\, \langle h, h_1 \rangle$;

(IP3) $\langle h_1, h_2 \rangle = \overline{\langle h_1, h_2 \rangle}$ (so $\langle h, h \rangle$ is real);

(IP4) $\langle h, h \rangle \geq 0$ and $\langle h, h \rangle = 0$ provided $h = 0$.

complex number, which is the evaluation of the ket $|A>$ by the bra $<B|$. This complex number, say $re^{i\theta}$ represents the system's *transition amplitude*[16] from its *initial state* A to its *final state* B[17], i.e.,

$$Transition\,Amplitude =< B|A >= re^{i\theta}.$$

That is, there is a process that can mediate a transition of a system from initial state A to the final state B and the amplitude for this transition equals $< B|A >= re^{i\theta}$. The absolute square of the amplitude, $| < B|A > |^2$ represents the *transition probability*. Therefore, the probability of a transition event equals the absolute square of a complex number, i.e.,

$$Transition\,Probability = | < B|A > |^2 = |re^{i\theta}|^2.$$

These complex amplitudes obey the usual *laws of probability*: when a transition event can happen in alternative ways then we add the complex numbers,

$$< B_1|A_1 > + < B_2|A_2 >= r_1e^{i\theta_1} + r_2e^{i\theta_2},$$

and when it can happen only as a succession of intermediate steps then we multiply the complex numbers,

$$< B|A >=< B|c >< c|A >= (r_1e^{i\theta_1})(r_2e^{i\theta_2}) = r_1r_2e^{i(\theta_1+\theta_2)}.$$

In general,

1. The amplitude for n *mutually alternative processes* equals the *sum* $\sum_{k=1}^{n} r_ke^{i\theta_k}$ of the amplitudes for the alternatives; and
2. If transition from A to B occurs in a *sequence of* m *steps*, then the total transition amplitude equals the *product* $\prod_{j=1}^{m} r_je^{i\theta_j}$ of the amplitudes of the steps.

Formally, we have the so–called *expansion principle*, including both products and sums,[18]

$$< B|A >= \sum_{i=1}^{n} < B|c_i >< c_i|A > . \tag{1.239}$$

The *standard inner product* on the product space $\mathbb{C}^n = \mathbb{C} \times \cdots \times \mathbb{C}$ is defined by $\langle z, w\rangle = \sum_{i=1}^{n} z_iw^i$, and axioms (IP1)–(IP4) are readily checked. Also $\mathbb{C}^n$ is a normed space with $\|z\|^2 = \sum_{i=1}^{n} |z_i|^2$. The pair $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ is called an *inner product space*.

Let $(\mathcal{H}, \|\cdot\|)$ be a normed space. If the corresponding metric d is complete, we say $(\mathcal{H}, \|\cdot\|)$ is a Banach space. If $(\mathcal{H}, \|\cdot\|)$ is an inner product space whose corresponding metric is complete, we say $(\mathcal{H}, \|\cdot\|)$ is a *Hilbert space*.

[16] Transition amplitude is otherwise called *probability amplitude*, or just *amplitude*.

[17] Recall that in quantum mechanics, complex numbers are regarded as the *vacuum–state*, or the *ground–state*, and the entire amplitude $< b|a >$ is a *vacuum–to–vacuum amplitude* for a process that includes the *creation* of the state a, its *transition* to b, and the *annihilation* of b to the vacuum once more.

[18] In Dirac's language, the *completeness* of intermediate states becomes the statement that a certain sum of projectors is equal to the identity. Namely, suppose

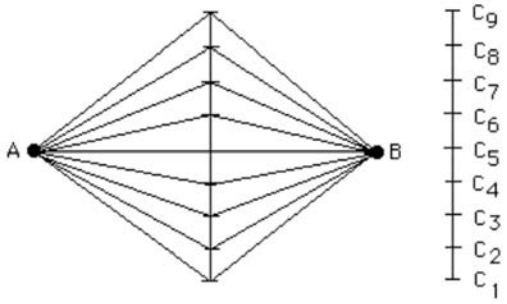


Fig. 1.11. Analysis of all possible routes from the source A to the detector B is simplified to include only double straight lines (in a plane).

Feynman's Sum–over–Histories

Now, *iterating* the Dirac's expansion principle (1.239) over a *complete set of all possible states* of the system, leads to the simplest form of the *Feynman path integral*, or, *sum–over–histories*. Imagine that the *initial* and *final* states, A and B, are points on the vertical lines $x = 0$ and $x = n + 1$, respectively, in the $x - y$ plane, and that $(c(k)_{i(k)}, k)$ is a given point on the line $x = k$ for $0 < i(k) < m$ (see Figure 1.11). Suppose that the sum of projectors for each intermediate state is complete[19] Applying the completeness iteratively, we get the following expression for the transition amplitude:

$$< B|A >= \sum \sum ... \sum < B|c(1)_{i(1)} >< c(1)_{i(1)}|c(2)_{i(2)} > ... < c(n)_{i(n)}|A >,$$

where the sum is taken over all $i(k)$ ranging between 1 and m, and k ranging between 1 and n. Each term in this sum can be construed as a *combinatorial route* from A to B in the 2D space of the $x - y$ plane. Thus the transition amplitude for the system going from some initial state A to some final state B is seen as a summation of contributions from *all the routes* connecting A to B.

Feynman used this description to produce his celebrated *path integral* expression for a transition amplitude (see, e.g., [GS98, Sch81]). His path integral takes the form

---

that $\sum_i |c_i >< c_i| = 1$ with $< c_i|c_i >= 1$ for each i. Then

$$< b|a >=< b||a >=< b| \sum_i |c_i >< c_i||a >= \sum_i < b|c_i >< c_i|a > .$$

[19] We assume that following sum is equal to one, for each k from 1 to $n - 1$:

$$|c(k)_1 >< c(k)_1| + ... + |c(k)_m >< c(k)_m| = 1.$$

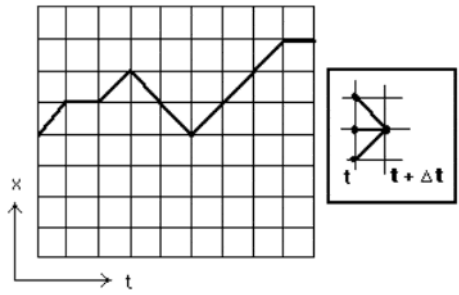


Fig. 1.12. Random walk (a particular case of Markov chain) on the $x-$axis.

$$Transition\ Amplitude =< B|A >= \oint \mathcal{D}[x]\, \mathrm{e}^{\mathrm{i}\mathcal{S}[x]}, \tag{1.240}$$

where the sum–integral $\oint$ is taken over all possible routes $x = x(t)$ from the initial point $A = A(t_{ini})$ to the final point $B = B(t_{fin})$, and $\mathcal{S} = \mathcal{S}[x]$ is the classical *action* for a particle to travel from A to B along a given extremal path x. In this way, Feynman took seriously Dirac's conjecture interpreting the exponential of the classical *action functional* ($\mathcal{D}\mathrm{e}^{\mathrm{i}\mathcal{S}}$), resembling a complex number ($r\mathrm{e}^{\mathrm{i}\theta}$), as an *elementary amplitude*. By integrating this elementary amplitude, $\mathcal{D}\mathrm{e}^{\mathrm{i}\mathcal{S}}$, over the infinitude of all possible histories, we get the total system's transition amplitude.[20]

[20] For the quantum physics associated with a classical (Newtonian) particle the action S is given by the integral along the given route from a to b of the difference $T - V$ where T is the classical kinetic energy and V is the classical potential energy of the particle.
The beauty of Feynman's approach to quantum physics is that it shows the relationship between the classical and the quantum in a particularly transparent manner. Classical motion corresponds to those regions where all nearby routes contribute constructively to the summation. This classical path occurs when the *variation of the action* is null. To ask for those paths where the variation of the action is zero is a problem in the calculus of variations, and it leads directly to Newton's equations of motion (derived using the Euler–Lagrangian equations). Thus with the appropriate choice of action, classical and quantum points of view are unified.
Also, a discretization of the Schrodinger equation

$$i\hbar\frac{d\psi}{dt} = -\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + V\psi,$$

leads to a sum–over–histories that has a discrete path integral as its solution. Therefore, the transition amplitude is equivalent to the wave ψ. The particle travelling on the $x-$axis is executing a one–step *random walk*, see Figure 1.12.

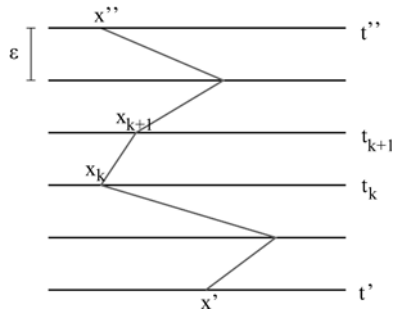


Fig. 1.13. A piecewise linear particle path contributing to the discrete Feynman propagator.

The Basic Form of a Path Integral

In Feynman's version of non–relativistic quantum mechanics, the time evolution $\psi(x',t') \mapsto \psi(x'',t'')$ of the wave function $\psi = \psi(x,t)$ of the elementary 1D particle may be described by the integral equation [GS98]

$$\psi(x'',t'') = \int_{\mathbb{R}} K(x'',x';t'',t')\,\psi(x',t'), \tag{1.241}$$

where the *propagator* or *Feynman kernel* $K = K(x'',x';t'',t')$ is defined through a limiting procedure,

$$K(x'',x';t'',t') = \lim_{\epsilon\to 0} A^{-N} \prod_{k=1}^{N-1} \int dx_k \; \mathrm{e}^{\mathrm{i}\sum_{j=0}^{N-1} \epsilon L(x_{j+1},(x_{j+1}-x_j)/\epsilon)}. \tag{1.242}$$

The time interval $t''-t'$ has been discretized into N steps of length $\epsilon = (t''-t')/N$, and the r.h.s. of (1.242) represents an integral over all piecewise linear paths $x(t)$ of a 'virtual' particle propagating from x' to x'', illustrated in Figure 1.13.

The prefactor A^{-N} is a normalization and L denotes the Lagrangian function of the particle. Knowing the propagator G is tantamount to having solved the quantum dynamics. This is the simplest instance of a path integral, and is often written schematically as

$$K(x',t';x'',t'') = \int\!\!\!\!\!\!\int \mathcal{D}[x(t)]\; \mathrm{e}^{\mathrm{i}S[x(t)]},$$

where $\mathcal{D}[x(t)]$ is a functional measure on the 'space of all paths', and the exponential weight depends on the classical action $S[x(t)]$ of a path. Recall also that this procedure can be defined in a mathematically clean way if we Wick–rotate the time variable t to imaginary values $t \mapsto \tau = \mathrm{i}t$, thereby making all integrals real [RS75].

Adaptive Path Integral

Now, we can extend the Feynman sum–over–histories (1.240), by adding the synaptic–like weights $w^i = w^i(t)$ into the measure $\mathcal{D}[x]$, to get the *adaptive path integral*:

$$Adaptive\, Transition\, Amplitude =< B|A >_w = \oint \mathcal{D}[w,x]\, \mathrm{e}^{\mathrm{i}S[x]}, \qquad (1.243)$$

where the *adaptive measure* $\mathcal{D}[w,x]$ is defined by the weighted product (of discrete time steps)

$$\mathcal{D}[w,x] = \lim_{n\longrightarrow\infty} \prod_{t=1}^{n} w^i(t)\, dx^i(t). \qquad (1.244)$$

In (1.244) the *synaptic weights* $w^i = w^i(t)$ are updated by the unsupervised *Hebbian–like learning* rule [Heb49]:

$$w^i(t+1) = w^i(t) + \frac{\sigma}{\eta}(w^i_d(t) - w^i_a(t)), \qquad (1.245)$$

where $\sigma = \sigma(t)$, $\eta = \eta(t)$ represent local *signal* and *noise* amplitudes, respectively, while superscripts d and a denote *desired* and *achieved* system states, respectively. Theoretically, equations (1.243–1.245) define an $\infty-$*dimensional complex–valued neural network*.[21] Practically, in a computer simulation we can use $10^7 \leq n \leq 10^8$, approaching the number of neurons in the brain. Such equations are usually solved using Markov–Chain Monte–Carlo methods on parallel (cluster) computers (see, e.g., [WW83a, WW83b]).

1.5.2 Path Integral History

Extract from Feynman's Nobel Lecture

In his Nobel Lecture, December 11, 1965, Richard (Dick) Feynman said that he and his PhD supervisor, John Wheeler, had found the *action* $A = A[x; t_i, t_j]$, directly involving the *motions of the charges only*,[22]

$$\begin{aligned} A[x;t_i,t_j] &= m_i \int (\dot{x}^i_\mu \dot{x}^i_\mu)^{\frac{1}{2}} dt_i + \frac{1}{2} e_i e_j \int\int \delta(I^2_{ij})\, \dot{x}^i_\mu(t_i) \dot{x}^j_\mu(t_j)\, dt_i dt_j \\ &\text{with } (i \neq j) \\ I^2_{ij} &= \left[x^i_\mu(t_i) - x^j_\mu(t_j)\right] \left[x^i_\mu(t_i) - x^j_\mu(t_j)\right], \end{aligned} \qquad (1.246)$$

[21] For details on complex–valued neural networks, see e.g., complex–domain extension of the standard backpropagation learning algorithm [GK92, BP02].

[22] *Wheeler–Feynman Idea* [WF49] "The energy tensor can be regarded only as a provisional means of representing matter. In reality, *matter consists of electrically charged particles*."

where $x^i_\mu = x^i_\mu(t_i)$ is the four–vector *position* of the ith particle as a function of the proper time t_i, while $\dot{x}^i_\mu(t_i) = dx^i_\mu(t_i)/dt_i$ is the *velocity* four–vector.

The first term in the action $A[x; t_i, t_j]$ (1.246) is the integral of the proper time t_i, the *ordinary action of relativistic mechanics of free particles of mass* m_i (summation over μ). The second term in the action $A[x; t_i, t_j]$ (1.246) represents the *electrical interaction of the charges*. It is summed over each pair of charges (the factor $\frac{1}{2}$ is to count each pair once, the term $i = j$ is omitted to avoid self–action). The *interaction is a double integral over a delta function of the square of space–time interval* I^2 *between two points on the paths*. Thus, interaction occurs only when this interval vanishes, that is, *along light cones* (see [WF49]).

Feynman comments here: "The fact that the interaction is exactly one–half advanced and half–retarded meant that we could write such a principle of least action, whereas interaction via retarded waves alone cannot be written in such a way. So, all of classical electrodynamics was contained in this very simple form."

"...The problem is only to make a quantum theory, which has as its classical analog, this expression (1.246). Now, there is no unique way to make a quantum theory from classical mechanics, although all the textbooks make believe there is. What they would tell you to do, was find the momentum variables and replace them by $(\hbar/i)(\partial/\partial x)$, but I couldn't find a momentum variable, as there wasn't any."

"The character of quantum mechanics of the day was to write things in the famous *Hamiltonian way* (in the form of Schrödinger equation), which described how the wave function changes from instant to instant, and in terms of the Hamiltonian operator H. If the classical physics could be reduced to a Hamiltonian form, everything was all right. Now, least action does not imply a Hamiltonian form if the action is a function of anything more than positions and velocities at the same moment. If the action is of the form of the integral of the Lagrangian $L = L(\dot{x}, x)$, a function of the velocities and positions at the same time t,

$$S[x] = \int L(\dot{x}, x)\, dt, \tag{1.247}$$

then you can start with the Lagrangian L and then create a Hamiltonian H and work out the quantum mechanics, more or less uniquely. But the action $A[x; t_i, t_j]$ (1.246) involves the key variables, positions (and velocities), at two different times t_i and t_j and therefore, it was not obvious what to do to make the quantum–mechanical analogue..."

So, Feynman was looking for the action integral in quantum mechanics. He says: "...I simply turned to Professor Jehle and said, 'Listen, do you know any way of doing quantum mechanics, starting with action – where the action integral comes into the quantum mechanics?" 'No", he said, 'but Dirac has a paper in which the Lagrangian, at least, comes into quantum mechanics." What Dirac said was the following: There is in quantum mechanics a very important quantity which carries the wave function from one time to another,

besides the differential equation but equivalent to it, a kind of a kernel, which we might call $K(x', x)$, which carries the wave function $\psi(x)$ known at time t, to the wave function $\psi(x')$ at time $t + \varepsilon$,

$$\psi(x', t + \varepsilon) = \int K(x', x)\, \psi(x, t)\, dx.$$

Dirac points out that this function K was analogous to the quantity in classical mechanics that you would calculate if you took the exponential of [$\mathrm{i}\varepsilon$ multiplied by the Lagrangian $L(\dot{x}, x)$], imagining that these two positions x, x' corresponded to t and $t + \varepsilon$. In other words,

$$K(x', x) \quad \text{is analogous to} \quad \mathrm{e}^{\mathrm{i}\varepsilon L(\frac{x'-x}{\varepsilon}, x)/\hbar}.$$

So, Feynman continues: "What does he mean, they are analogous; what does that mean, *analogous*? What is the use of that?" Professor Jehle said, 'You Americans! You always want to find a use for everything!" I said that I thought that Dirac must mean that they were *equal*. 'No", he explained, 'he doesn't mean they are equal." 'Well", I said, 'Let's see what happens if we make them equal."

"So, I simply put them equal, taking the simplest example where the Lagrangian is

$$L = \frac{1}{2} M \dot{x}^2 - V(x),$$

but soon found I had to put a constant of proportionality N in, suitably adjusted. When I substituted for K to get

$$\psi(x', t + \varepsilon) = \int N \exp\left[\frac{\mathrm{i}\varepsilon}{\hbar} L(\frac{x' - x}{\varepsilon}, x)\right] \psi(x, t)\, dx \tag{1.248}$$

and just calculated things out by Taylor series expansion, *out came the Schrödinger equation*. So, I turned to Professor Jehle, not really understanding, and said, 'Well, you see, Dirac meant that they were proportional." Professor Jehle's eyes were bugging out – he had taken out a little notebook and was rapidly copying it down from the blackboard, and said, 'No, no, this is an important discovery. You Americans are always trying to find out how something can be used. That's a good way to discover things!" So, I thought I was finding out what Dirac meant, but, as a matter of fact, had made the discovery that what Dirac thought was analogous, was, in fact, equal. I had then, at least, the connection between the Lagrangian and quantum mechanics, but still with wave functions and infinitesimal times."

"It must have been a day or so later when I was lying in bed thinking about these things, that I imagined what would happen if I wanted to calculate the wave function at a finite interval later. I would put one of these factors $\mathrm{e}^{\mathrm{i}\varepsilon L}$ in here, and that would give me the wave functions the next moment, $t + \varepsilon$, and then I could substitute that back into (1.248) to get another factor of $\mathrm{e}^{\mathrm{i}\varepsilon L}$ and

give me the wave function the next moment, $t + 2\varepsilon$, and so on and so on. In that way I found myself thinking of a large number of integrals, one after the other in sequence. In the integrand was the product of the exponentials, which was the exponential of the sum of terms like εL. Now, L is the Lagrangian and ε is like the time interval dt, so that if you took a sum of such terms, that's exactly like an integral. That's like Riemann's formula for the integral $\int L dt$, you just take the value at each point and add them together. We are to take the limit as $\varepsilon \to 0$. Therefore, the connection between the wave function of one instant and the wave function of another instant a finite time later could be get by an infinite number of integrals (because ε goes to zero), of exponential where S is the action expression (1.247). At last, I had succeeded in representing quantum mechanics directly in terms of the action $S[x]$."

Fully satisfied, Feynman comments: "This led later on to the idea of the *transition amplitude* for a path: that for each possible way that the particle can go from one point to another in space–time, there's an amplitude. That amplitude is **e** to the power of [i/$\hbar$ times the action $S[x]$ for the path], i.e., $\mathrm{e}^{\mathrm{i}S[x]/\hbar}$. Amplitudes from various paths superpose by addition. This then is another, a third way, of describing quantum mechanics, which looks quite different from that of Schrödinger or Heisenberg, but which is equivalent to them."

"...Now immediately after making a few checks on this thing, what we wanted to do, was to substitute the action $A[x; t_i, t_j]$ (1.246) for the other $S[x]$ (1.247). The first trouble was that I could not get the thing to work with the relativistic case of spin one–half. However, although I could deal with the matter only nonrelativistically, I could deal with the light or the photon interactions perfectly well by just putting the interaction terms of (1.246) into any action, replacing the mass terms by the non–relativistic $Ldt = \frac{1}{2}M\dot{x}^2 dt$,

$$A[x; t_i, t_j] = \frac{1}{2}\sum_i m_i \int (\dot{x}^i_\mu)^2 dt_i + \frac{1}{2}\sum_{i,j(i\neq j)} e_i e_j \int\int \delta(I^2_{ij})\, \dot{x}^i_\mu(t_i)\dot{x}^j_\mu(t_j)\, dt_i dt_j.$$

When the action has a delay, as it now had, and involved more than one time, I had to lose the idea of a wave function. That is, I could no longer describe the program as: given the amplitude for all positions at a certain time to calculate the amplitude at another time. However, that didn't cause very much trouble. It just meant developing a new idea. *Instead of wave functions we could talk about this: that if a source of a certain kind emits a particle, and a detector is there to receive it, we can give the amplitude that the source will emit and the detector receive,* $\mathrm{e}^{\mathrm{i}A[x;t_i,t_j]/\hbar}$. We do this without specifying the exact instant that the *source* emits or the exact instant that any *detector* receives, without trying to specify the state of anything at any particular time in between, but by just finding the *amplitude for the complete experiment.* And, then we could discuss how that amplitude would change if you had a scattering sample in between, as you rotated and changed angles, and so on, without really having any wave functions...It was also possible to discover what the old concepts

of energy and momentum would mean with this generalized action. And, so I believed that I had a quantum theory of classical electrodynamics – or rather of this new classical electrodynamics described by the action $A[x; t_i, t_j]$ (1.246)..."

Lagrangian Path Integral

Dirac and Feynman first developed the lagrangian approach to functional integration. To review this approach, we start with the time–dependent *Schrödinger equation*

$$\mathrm{i}\hbar\,\partial_t\psi(x,t) = -\partial_{x^2}\psi(x,t) + V(x)\,\psi(x,t)$$

appropriate to a particle of mass m moving in a potential $V(x)$, $x \in \mathbb{R}$. A solution to this equation can be written as an integral (see e.g., [Kla97, Kla00]),

$$\psi(x'',t'') = \int K(x'',t'';x',t')\,\psi(x',t')\,dx' \ ,$$

which represents the wave function $\psi(x'',t'')$ at time t'' as a linear superposition over the wave function $\psi(x',t')$ at the initial time t', $t' < t''$. The integral kernel $K(x'',t'';x',t')$ is known as the *propagator*, and according to Feynman [Fey48] it may be given by

$$K(x'',t'';x',t') = \mathcal{N}\int \mathcal{D}[x]\ \mathrm{e}^{(\mathrm{i}/\hbar)\int[(m/2)\,\dot{x}^2(t)-V(x(t))]\,dt},$$

which is a formal expression symbolizing an integral over a suitable set of paths. This integral is supposed to run over all continuous paths $x(t)$, $t' \leq t \leq t''$, where $x(t'') = x''$ and $x(t') = x'$ are fixed end points for all paths. Note that the integrand involves the *classical Lagrangian* for the system.

To overcome the convergence problems, Feynman adopted a *lattice regularization* as a procedure to yield well–defined integrals which was then followed by a limit as the lattice spacing goes to zero called the continuum limit. With $\varepsilon > 0$ denoting the lattice spacing, the details regarding the lattice regularization procedure are given by

$$K(x'',t'';x',t') = \lim_{\varepsilon\to 0}\,(m/2\pi\mathrm{i}\hbar\varepsilon)^{(N+1)/2}\int\cdots$$
$$\cdots\int \exp\{(\mathrm{i}/\hbar)\sum_{l=0}^{N}\,[(m/2\varepsilon)(x_{l+1}-x_l)^2 - \varepsilon\,V(x_l)\,]\}\,\prod_{l=1}^{N} dx_l \ ,$$

where $x_{N+1} = x''$, $x_0 = x'$, and $\varepsilon \equiv (t''-t')/(N+1)$, $N \in \{1,2,3,\dots\}$. In this version, at least, we have an expression that has a reasonable chance of being well defined, provided, that one interprets the conditionally convergent integrals involved in an appropriate manner. One common and fully acceptable

interpretation adds a convergence factor to the exponent of the preceding integral in the form $-(\varepsilon^2/2\hbar)\sum_{l=1}^{N} x_l^2$, which is a term that formally makes no contribution to the final result in the continuum limit save for ensuring that the integrals involved are now rendered absolutely convergent.

Hamiltonian Path Integral

It is necessary to retrace history at this point to recall the introduction of the *phase–space path integral* by Feynman [Fey51, GS98]. In Appendix B to this article, Feynman introduced a formal expression for the configuration or $q-$space propagator given by (see e.g., [Kla97, Kla00])

$$K(q'', t''; q', t') = \mathcal{M} \int \mathcal{D}[p]\, \mathcal{D}[q]\ \exp\{(\mathrm{i}/\hbar) \int [\, p\, \dot{q} - H(p, q)\,]\, dt\}.$$

In this equation one is instructed to integrate over all paths $q(t)$, $t' \leq t \leq t''$, with $q(t'') \equiv q''$ and $q(t') \equiv q'$ held fixed, as well as to integrate over all paths $p(t)$, $t' \leq t \leq t''$, without restriction.

It is widely appreciated that the phase–space path integral is more generally applicable than the original, Lagrangian, version of the path integral. For example, the original configuration space path integral is satisfactory for Lagrangians of the general form

$$L(x) = \frac{1}{2}\, m\dot{x}^2 + A(x)\, \dot{x} - V(x)\ ,$$

but it is unsuitable, for example, for the case of a relativistic particle with the Lagrangian

$$L(x) = -m\, qrt1 - \dot{x}^2$$

expressed in units where the speed of light is unity. For such a system – as well as many more general expressions – the phase–space form of the path integral is to be preferred. In particular, for the relativistic free particle, the phase–space path integral

$$\mathcal{M} \int \mathcal{D}[p]\, \mathcal{D}[q]\ \exp\{(\mathrm{i}/\hbar) \int [\, p\, \dot{q} - \ qrtp^2 + m^2\,]\, dt\},$$

is readily evaluated and induces the correct propagator.

Feynman–Kac Formula

Through his own research, M. Kac was fully aware of *Wiener's theory of Brownian motion* and the *associated diffusion equation* that describes the corresponding *distribution function*. Therefore, it is not surprising that he was well prepared to give a path integral expression in the sense of Feynman for an equation similar to the time–dependent Schrödinger equation save for

a rotation of the time variable by $-\pi/2$ in the complex plane, namely, by the change $t \longrightarrow -it$ (see e.g., [Kla97, Kla00]). In particular, Kac [Kac51] considered the equation

$$\partial_t \rho(x,t) = \partial_{x^2} \rho(x,t) - V(x)\ \rho(x,t). \tag{1.249}$$

This equation is analogous to Schrödinger equation but differs from it in certain details. Besides certain constants which are different, and the change $t \longrightarrow -it$, the nature of the dependent variable function $\rho(x,t)$ is quite different from the normal quantum mechanical wave function. For one thing, if the function ρ is initially real it will remain real as time proceeds. Less obvious is the fact that if $\rho(x,t) \geq 0$ for all x at some time t, then the function will continue to be nonnegative for all time t. Thus we can interpret $\rho(x,t)$ more like a probability density; in fact in the special case that $V(x) = 0$, then $\rho(x,t)$ is the probability density for a Brownian particle which underlies the *Wiener measure*. In this regard, ν is called the diffusion constant.

The fundamental solution of (1.249) with $V(x) = 0$ is readily given as

$$W(x,T;y,0) = \frac{1}{qrt2\pi\nu T}\ \exp\left(-\frac{(x-y)^2}{2\nu T}\right),$$

which describes the solution to the diffusion equation subject to the initial condition

$$\lim_{T\to 0^+}\ W(x,T;y,0) = \delta(x-y)\ .$$

Moreover, it follows that the solution of the diffusion equation for a general initial condition is given by

$$\rho(x'',t'') = \int W(x'',t'';x',t')\ \rho(x',t')\,dx'\ .$$

Iteration of this equation N times, with $\epsilon = (t''-t')/(N+1)$, leads to the equation

$$\rho(x'',t'') = N' \int \cdots \int \mathrm{e}^{-(1/2\nu\epsilon)\sum_{l=0}^{N}(x_{l+1}-x_l)^2} \prod_{l=1}^{N} dx_l\ \rho(x',t')\,dx',$$

where $x_{N+1} \equiv x''$ and $x_0 \equiv x'$. This equation features the imaginary time propagator for a free particle of unit mass as given formally as

$$W(x'',t'';x',t') = \mathcal{N} \int \mathcal{D}[x]\ \mathrm{e}^{-(1/2\nu)\int \dot{x}^2\,dt},$$

where $\mathcal{N}$ denotes a formal normalization factor.

The similarity of this expression with the Feynman path integral [for $V(x) = 0$] is clear, but there is a profound difference between these equations. In the former (Feynman) case the underlying measure is only *finitely*

additive, while in the latter (Wiener) case the continuum limit actually defines a genuine measure, i.e., a *countably additive measure* on paths, which is a version of the famous *Wiener measure*. In particular,

$$W(x'', t''; x', t') = \int d\mu_W^\nu(x),$$

where μ_W^ν denotes a measure on continuous paths $x(t)$, $t' \leq t \leq t''$, for which $x(t'') \equiv x''$ and $x(t') \equiv x'$. Such a measure is said to be a *pinned* Wiener measure, since it specifies its path values at two time points, i.e., at $t = t'$ and at $t = t'' > t'$.

We note that Brownian motion paths have the property that with probability one they are concentrated on continuous paths. However, it is also true that the time derivative of a Brownian path is almost nowhere defined, which means that, with probability one, $\dot{x}(t) = \pm\infty$ for all t.

When the potential $V(x) \neq 0$ the propagator associated with (1.249) is formally given by

$$W(x'', t''; x', t') = \mathcal{N} \int \mathcal{D}[x] \mathrm{e}^{-(1/2\nu)\int \dot{x}^2 \, dt - \int V(x) \, dt},$$

an expression which is well defined if $V(x) \geq c$, $-\infty < c < \infty$. A mathematically improved expression makes use of the Wiener measure and reads

$$W(x'', t''; x', t') = \int \mathrm{e}^{-\int V(x(t)) \, dt} \, d\mu_W^\nu(x).$$

This is an elegant relation in that it represents a solution to the differential equation (1.249) in the form of an integral over Brownian motion paths suitably weighted by the potential V. Incidentally, since the propagator is evidently a strictly positive function, it follows that the solution of the differential equation (1.249) is nonnegative for all time t provided it is nonnegative for any particular time value.

Itô Formula

Itô [Ito60] proposed another version of a *continuous–time regularization* that resolved some of the troublesome issues. In essence, the proposal of Itô takes the form given by

$$\lim_{\nu \to \infty} \mathcal{N}_\nu \int \mathcal{D}[x] \; \exp\{(\mathrm{i}/\hbar) \int [\frac{1}{2} m\dot{x}^2 - V(x)] \, dt\} \exp\{-(1/2\nu) \int [\ddot{x}^2 + \dot{x}^2] \, dt\}.$$

Note well the alternative form of the auxiliary factor introduced as a regulator. The additional term $\ddot{x}^2$, the square of the second derivative of x, acts to smooth out the paths sufficiently well so that in the case of (21) both $x(t)$ and $\dot{x}(t)$ are continuous functions, leaving $\ddot{x}(t)$ as the term which does not exist. However, since only x and $\dot{x}$ appear in the rest of the integrand, the indicated path integral can be well defined; this is already a positive contribution all by itself (see e.g., [Kla97, Kla00]).

1.5.3 Standard Path–Integral Quantization

Canonical versus Path–Integral Quantization

Recall that in the usual, *canonical formulation* of quantum mechanics, the system's phase–space coordinates, q, and momenta, p, are replaced by the corresponding Hermitian operators in the Hilbert space, with real measurable eigenvalues, which obey *Heisenberg commutation relations*.

The *path–integral quantization* is instead based directly on the notion of a propagator $K(q_f, t_f; q_i, t_i)$ which is defined such that (see [Ryd96, CL84, Gun03])

$$\psi(q_f, t_f) = \int K(q_f, t_f; q_i, t_i)\, \psi(q_i, t_i)\, dq_i, \tag{1.250}$$

i.e., the wave function $\psi(q_f, t_f)$ at final time t_f is given by a Huygens principle in terms of the wave function $\psi(q_i, t_i)$ at an initial time t_i, where we have to integrate over all the points q_i since all can, in principle, send out little wavelets that would influence the value of the wave function at q_f at the later time t_f. This equation is very general and is an expression of causality. We use the normal units with $\hbar = 1$.

According to the usual interpretation of quantum mechanics, $\psi(q_f, t_f)$ is the *probability amplitude* that the particle is at the point q_f and the time t_f, which means that $K(q_f, t_f; q_i, t_i)$ is the probability amplitude for a transition from q_i and t_i to q_f and t_f. The probability that the particle is observed at q_f at time t_f if it began at q_i at time t_i is

$$P(q_f, t_f; q_i, t_i) = |K(q_f, t_f; q_i, t_i)|^2 .$$

Let us now divide the time interval between t_i and t_f into two, with t as the intermediate time, and q the intermediate point in space. Repeated application of (1.250) gives

$$\psi(q_f, t_f) = \int\int K(q_f, t_f; q, t)\, dq\, K(q, t; q_i, t_i)\, \psi(q_i, t_i)\, dq_i,$$

from which it follows that

$$K(q_f, t_f; q_i, t_i) = \int dq\, K(q_f, t_f; q, t)\, K(q, t; q_i, t_i).$$

This equation says that the transition from (q_i, t_i) to (q_f, t_f) may be regarded as the result of the transition from (q_i, t_i) to all available intermediate points q followed by a transition from (q, t) to (q_f, t_f). This notion of *all possible paths* is crucial in the *path–integral formulation* of quantum mechanics.

Now, recall that the *state vector* $|\psi, t\rangle_S$ in the *Schrödinger picture* is related to that in the *Heisenberg picture* $|\psi\rangle_H$ by

$$|\psi, t\rangle_S = \mathrm{e}^{-\mathrm{i}Ht}\, |\psi\rangle_H ,$$

or, equivalently,

$$|\psi\rangle_H = \mathrm{e}^{\mathrm{i}Ht}\,|\psi,t\rangle_S\,.$$

We also define the vector

$$|q,t\rangle_H = \mathrm{e}^{\mathrm{i}Ht}\,|q\rangle_S\,,$$

which is the Heisenberg version of the Schrödinger state $|q\rangle$. Then, we can equally well write

$$\psi(q,t) = \langle q,t\,|\psi\rangle_H\,. \tag{1.251}$$

By completeness of states we can now write

$$\langle q_f,t_f\,|\psi\rangle_H = \int \langle q_f,t_f\,|q_i,t_i\rangle_H\;\langle q_i,t_i\,|\psi\rangle_H\;dq_i,$$

which with the definition of (1.251) becomes

$$\psi(q_f,t_f) = \int \langle q_f,t_f\,|q_i,t_i\rangle_H\;\psi(q_i,t_i)\,dq_i.$$

Comparing with (1.250), we get

$$K(q_f,t_f;q_i,t_i) = \langle q_f,t_f\,|q_i,t_i\rangle_H\,.$$

Now, let us calculate the *quantum–mechanics propagator*

$$\langle q',t'\,|q,t\rangle_H = \left\langle q'|\mathrm{e}^{-\mathrm{i}H(t-t')}\,|q\right\rangle$$

using the *path–integral formalism* that will incorporate the direct quantization of the coordinates, without Hilbert space and Hermitian operators.

The first step is to divide up the time interval into $n+1$ tiny pieces: $t_l = l\varepsilon + t$ with $t' = (n+1)\varepsilon + t$. Then, by completeness, we can write (dropping the Heisenberg picture index H from now on)

$$\begin{aligned}\langle q',t'\,|q,t\rangle = \int dq_1(t_1)...\int dq_n(t_n)\;\langle q',t'\,|q_n,t_n\rangle\\ \times\,\langle q_n,t_n\,|q_{n-1},t_{n-1}\rangle\,...\,\langle q_1,t_1\,|q,t\rangle\,.\end{aligned} \tag{1.252}$$

The integral $\int dq_1(t_1)...dq_n(t_n)$ is an *integral over all possible paths*, which are not trajectories in the normal sense, since there is no requirement of continuity, but rather *Markov chains*.

Now, for small ε we can write

$$\langle q',\varepsilon\,|q,0\rangle = \left\langle q'|\mathrm{e}^{-\mathrm{i}\varepsilon H(P,Q)}\,|q\right\rangle = \delta(q'-q) - \mathrm{i}\varepsilon\,\langle q'|H(P,Q)\,|q\rangle\,,$$

where $H(P,Q)$ is the Hamiltonian (e.g., $H(P,Q) = \frac{1}{2}P^2 + V(Q)$, where P,Q are the momentum and coordinate operators). Then we have (see [Ryd96, CL84, Gun03])

$$\langle q'|H(P,Q)\,|q\rangle = \int \frac{dp}{2\pi}\,\mathrm{e}^{\mathrm{i}p(q'-q)} H\left(p, \frac{1}{2}(q'+q)\right).$$

Putting this into our earlier form we get

$$\langle q', \varepsilon\,|q, 0\rangle \simeq \int \frac{dp}{2\pi} \exp\left[\mathrm{i}\left\{p(q'-q) - \varepsilon H\left(p, \frac{1}{2}(q'+q)\right)\right\}\right],$$

where the 0th order in $\varepsilon \to \delta(q'-q)$ and the 1st order in $\varepsilon \to -\mathrm{i}\varepsilon\,\langle q'|H(P,Q)\,|q\rangle$. If we now substitute many such forms into (1.252) we finally get

$$\langle q', t'\,|q, t\rangle = \lim_{n\to\infty} \int \prod_{i=1}^{n} dq_i \prod_{k=1}^{n+1} \frac{dp_k}{2\pi} \tag{1.253}$$

$$\times \exp\left\{\mathrm{i}\sum_{j=1}^{n+1}\left[p_j(q_j - q_{j-1})\right] - H\left(p_j, \frac{1}{2}(q_j + q_{j+1})\right)(t_j - t_{j-1})\right]\right\},$$

with $q_0 = q$ and $q_{n+1} = q'$. Roughly, the above formula says to *integrate over all possible momenta and coordinate values associated with a small interval*, weighted by something that is going to turn into the *exponential of the action* $\mathrm{e}^{\mathrm{i}S}$ in the limit where $\varepsilon \to 0$. It should be stressed that the different q_i and p_k integrals are independent, which implies that p_k for one interval can be completely different from the $p_{k'}$ for some other interval (including the neighboring intervals). In principle, the integral (1.253) should be defined by *analytic continuation into the complex plane* of, for example, the p_k integrals.

Now, if we go to the differential limit where we call $t_j - t_{j-1} \equiv d\tau$ and write $\frac{(q_j - q_{j-1})}{(t_j - t_{j-1})} \equiv \dot{q}$, then the above formula takes the form

$$\langle q', t'\,|q, t\rangle = \int \mathcal{D}[p]\mathcal{D}[q] \exp\left\{\mathrm{i}\int_t^{t'} [p\dot{q} - H(p,q)]\,d\tau\right\},$$

where we have used the shorthand notation

$$\int \mathcal{D}[p]\mathcal{D}[q] \equiv \int \prod_{\tau} \frac{dq(\tau)dp(\tau)}{2\pi}.$$

Note that the above integration is an integration over the p and q values at every time τ. This is what we call a *functional integral*. We can think of a given set of choices for all the $p(\tau)$ and $q(\tau)$ as defining a *path in the* $6D$ *phase–space*. The most important point of the above result is that we have get an expression for a *quantum–mechanical transition amplitude* in terms of an integral involving only pure complex numbers, without operators.

We can actually perform the above integral for Hamiltonians of the type $H = H(P,Q)$. We use square completion in the exponential for this, defining the integral in the complex p plane and continuing to the physical situation. In particular, we have

$$\int_{-\infty}^{\infty} \frac{dp}{2\pi} \exp\left\{\mathrm{i}\varepsilon(p\dot{q} - \frac{1}{2}p^2]\right\} = \frac{1}{\sqrt{2\pi\mathrm{i}\varepsilon}} \exp\left[\frac{1}{2}\mathrm{i}\varepsilon\dot{q}^2\right],$$

(see [Ryd96, CL84, Gun03]) which, substituting into (1.253) gives

$$\langle q', t' \mid q, t\rangle = \lim_{n\to\infty} \int \prod_i \frac{dq_i}{\sqrt{2\pi\mathrm{i}\varepsilon}} \exp\{\mathrm{i}\varepsilon \sum_{j=1}^{n+1} [\frac{1}{2}(\frac{q_j - q_{j-1}}{\varepsilon})^2 - V(\frac{q_j + q_{j+1}}{2})]\}.$$

This can be formally written as

$$\langle q', t' \mid q, t\rangle = \int \mathcal{D}[q]\, \mathrm{e}^{\mathrm{i}S[q]},$$

where

$$\int \mathcal{D}[q] \equiv \int \prod_i \frac{dq_i}{\sqrt{2\pi\mathrm{i}\varepsilon}},$$

while

$$S[q] = \int_t^{t'} L(q, \dot{q})\, d\tau$$

is the *standard action* with the *Lagrangian*

$$L = \frac{1}{2}\dot{q}^2 - V(q).$$

Generalization to many degrees of freedom is straightforward:

$$\langle {q_1}' ... {q_N}', t' | q_1 ... q_N, t\rangle = \int \mathcal{D}[p]\mathcal{D}[q] \exp\left\{\mathrm{i}\int_t^{t'} \left[\sum_{n=1}^{N} p_n\dot{q}_n - H(p_n, q_n)\right] d\tau\right\},$$

$$\text{with} \quad \int \mathcal{D}[p]\mathcal{D}[q] = \int \prod_{n=1}^{N} \frac{dq_n dp_n}{2\pi}.$$

Here, $q_n(t) = q_n$ and $q_n(t') = {q_n}'$ for all $n = 1, ..., N$, and we are allowing for the full Hamiltonian of the system to depend upon all the N momenta and coordinates collectively.

Elementary Applications

(i) Consider first

$$\begin{aligned} &\langle q', t' | Q(t_0) | q, t\rangle \\ &= \int \prod dq_i(t_i)\, \langle q', t' | q_n, t_n\rangle \, ... \, \langle q_{i0}, t_{i0} | Q(t_0) | q_{i-1}, t_{i-1}\rangle \, ... \, \langle q_1, t_1 | q, t\rangle \,, \end{aligned}$$

where we choose one of the time interval ends to coincide with t_0, i.e., $t_{i0} = t_0$. If we operate $Q(t_0)$ to the left, then it is replaced by its eigenvalue $q_{i0} = q(t_0)$.

Aside from this one addition, everything else is evaluated just as before and we will obviously get

$$\langle q', t'|Q(t_0)|q, t\rangle = \int \mathcal{D}[p]\mathcal{D}[q]\, q(t_0) \exp\left\{ \mathrm{i} \int_t^{t'} [p\dot{q} - H(p,q)]d\tau \right\}.$$

(ii) Next, suppose we want a *path–integral expression* for $\langle q', t'|Q(t_1)Q(t_2)|q, t\rangle$ in the case where $t_1 > t_2$. For this, we have to insert as intermediate states $|q_{i1}, t_{i1}\rangle \langle q_{i1}, t_{i1}|$ with $t_{i1} = t_1$ and $|q_{i2}, t_{i2}\rangle \langle q_{i2}, t_{i2}|$ with $t_{i2} = t_2$ and since we have ordered the times at which we do the insertions we must have the first insertion to the left of the 2nd insertion when $t_1 > t_2$. Once these insertions are done, we evaluate $\langle q_{i1}, t_{i1}|\, Q(t_1) = \langle q_{i1}, t_{i1}|\, q(t_1)$ and $\langle q_{i2}, t_{i2}|\, Q(t_2) = \langle q_{i2}, t_{i2}|\, q(t_2)$ and then proceed as before and get

$$\langle q', t'|Q(t_1)Q(t_2)|q, t\rangle = \int \mathcal{D}[p]\mathcal{D}[q]\, q(t_1)\, q(t_2) \exp\left\{ \mathrm{i} \int_t^{t'} [p\dot{q} - H(p,q)]d\tau \right\}.$$

Now, let us ask what the above integral is equal to if $t_2 > t_1$? It is obvious that what we get for the above integral is $\langle q', t'|Q(t_2)Q(t_1)|q, t\rangle$. Clearly, this generalizes to an arbitrary number of Q operators.

(iii) When we enter into quantum field theory, the Q's will be replaced by fields, since it is the fields that play the role of coordinates in the 2nd quantization conditions.

Sources

The *source* is represented by modifying the Lagrangian:

$$L \to L + J(t)q(t).$$

Let us define $|0, t\rangle^J$ as the ground state (vacuum) vector (in the moving frame, i.e., with the e^{iHt} included) in the presence of the source. The required *transition amplitude* is

$$Z[J] \propto \langle 0, +\infty|0, -\infty\rangle^J,$$

where the source $J = J(t)$ plays a role analogous to that of an electromagnetic current, which acts as a source of the electromagnetic field. In other words, we can think of the scalar product $J_\mu A^\mu$, where J_μ is the current from a scalar (or Dirac) field acting as a source of the potential A^μ. In the same way, we can always define a current J that acts as the source for some arbitrary field ϕ. $Z[J]$ (otherwise denoted by $W[J]$) is a functional of the current J, defined as (see [Ryd96, CL84, Gun03])

$$Z[J] \propto \int \mathcal{D}[p]\mathcal{D}[q] \exp\left\{ \mathrm{i} \int_t^{t'} [p(\tau)\dot{q}(\tau) - H(p,q) + J(\tau)q(\tau)]d\tau \right\},$$

with the *normalization condition* $Z[J=0]=1$. Here, the argument of the exponential depends upon the functions $q(\tau)$ and $p(\tau)$ and we then integrate over all possible forms of these two functions. So the exponential is a functional that maps a choice for these two functions into a number. For example, for a quadratically completable $H(p,q)$, the p integral can be performed as a q integral

$$Z[J] \propto \int \mathcal{D}[q] \exp\left\{ \mathrm{i} \int_{-\infty}^{+\infty} \left(L + Jq + \frac{1}{2} \mathrm{i}\varepsilon q^2 \right) d\tau \right\},$$

where the addittion to H was chosen in the form of a *convergence factor* $-\frac{1}{2}\mathrm{i}\varepsilon q^2$.

Fields

Let us now treat the *abstract scalar field* $\phi(x)$ as a coordinate in the sense that we imagine dividing space up into many little cubes and the average value of the field $\phi(x)$ in that cube is treated as a coordinate for that little cube. Then, we go through the multi–coordinate analogue of the procedure we just considered above and take the continuum limit. The final result is

$$Z[J] \propto \int \mathcal{D}[\phi] \exp\left\{ \mathrm{i} \int d^4x \left(\mathcal{L}\left(\phi(x)\right) + J(x)\phi(x) + \frac{1}{2}\mathrm{i}\varepsilon\phi^2 \right) \right\},$$

where for $\mathcal{L}$ we would employ the *Klein–Gordon Lagrangian* form. In the above, the dx_0 integral is the same as $d\tau$, while the $d^3\mathbf{x}$ integral is summing over the sub–Lagrangians of all the different little cubes of space and then taking the continuum limit. $\mathcal{L}$ is the *Lagrangian density* describing the Lagrangian for each little cube after taking the many–cube limit (see [Ryd96, CL84, Gun03]) for the full derivation).

We can now introduce *interactions,* $\mathcal{L}_I$. Assuming the simple form of the Hamiltonian, we have

$$Z[J] \propto \int \mathcal{D}[\phi] \exp\left\{ \mathrm{i} \int d^4x \left(\mathcal{L}\left(\phi(x)\right) + \mathcal{L}_I\left(\phi(x)\right) + J(x)\phi(x) \right) \right\},$$

again using the normalization factor required for $Z[J=0]=1$.

For example of Klein Gordon theory, we would use

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_I, \qquad \mathcal{L}_0 \frac{1}{2}[\partial_\mu\phi\partial^\mu\phi - \mu^2\phi^2], \qquad \mathcal{L}_I = \mathcal{L}_I(\phi),$$

where $\partial_\mu \equiv \partial_{x^\mu}$ and we can freely manipulate indices, as we are working in Euclidean space $\mathbb{R}^3$. In order to define the above $Z[J]$, we have to include a convergence factor $\mathrm{i}\varepsilon\phi^2$,

$$\mathcal{L}_0 \to \frac{1}{2}[\partial_\mu\phi\partial^\mu\phi - \mu^2\phi^2 + \mathrm{i}\varepsilon\phi^2], \qquad \text{so that}$$

$$Z[J] \propto \int \mathcal{D}[\phi] \exp\{\mathrm{i} \int d^4x (\frac{1}{2}[\partial_\mu\phi\partial^\mu\phi - \mu^2\phi^2 + \mathrm{i}\varepsilon\phi^2] + \mathcal{L}_I(\phi(x)) + J(x)\phi(x))\}$$

is the appropriate *generating function* in the free field theory case.

Gauges

In the path integral approach to quantization of the *gauge theory*, we implement *gauge fixing* by restricting in some manner or other the path integral over gauge fields $\int \mathcal{D}[A_\mu]$. In other words we will write instead

$$Z[J] \propto \int \mathcal{D}[A_\mu]\, \delta\,(\text{some gauge fixing condition})\exp\{\mathrm{i}\int d^4x \mathcal{L}\,(A_\mu)\}.$$

A common approach would be to start with the *gauge condition*

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2}(\partial^\mu A_\mu)^2$$

where the electrodynamic field tensor is given by $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, and calculate

$$Z[J] \propto \int \mathcal{D}[A_\mu] \exp\left\{\mathrm{i}\int d^4x\,[\mathcal{L}(A_\mu(x)) + J_\mu(x)A^\mu(x)]\right\}$$

as the *generating function* for the *vacuum expectation* values of *time ordered products* of the A_μ fields. Note that J_μ should be conserved ($\partial^\mu J_\mu = 0$) in order for the full expression $\mathcal{L}(A_\mu) + J_\mu A^\mu$ to be *gauge–invariant* under the integral sign when $A_\mu \to A_\mu + \partial^\mu \Lambda$. For a proper approach, see [Ryd96, CL84, Gun03].

Riemannian–Symplectic Geometries

In this subsection, following [SK98b], we describe path integral quantization on Riemannian–symplectic manifolds. Let $\hat{q}^j$ be a set of Cartesian coordinate canonical operators satisfying the Heisenberg commutation relations $[\hat{q}^j, \hat{q}^k] = \mathrm{i}\omega^{jk}$. Here $\omega^{jk} = -\omega^{kj}$ is the canonical symplectic structure. We introduce the canonical coherent states as $|q\rangle \equiv \mathrm{e}^{\mathrm{i}q^j\omega_{jk}\hat{q}^k}|0\rangle$, where $\omega_{jn}\omega^{nk} = \delta^k_j$, and $|0\rangle$ is the ground state of a harmonic oscillator with unit angular frequency. Any state $|\psi\rangle$ is given as a function on phase–space in this representation by $\langle q|\psi\rangle = \psi(q)$. A general operator $\hat{A}$ can be represented in the form $\hat{A} = \int dq\, a(q)|q\rangle\langle q|$, where $a(q)$ is the lower symbol of the operator and dq is a properly normalized form of the Liouville measure. The function $A(q,q') = \langle q|\hat{A}|q'\rangle$ is the *kernel of the operator*.

The main object of the path integral formalism is the integral kernel of the evolution operator

$$K_t(q,q') = \langle q|\mathrm{e}^{-\mathrm{i}t\hat{H}}|q'\rangle = \int\limits_{q(0)=q'}^{q(t)=q} \mathcal{D}[q]\ \mathrm{e}^{\mathrm{i}\int_0^t d\tau\left(\frac{1}{2}q^j\omega_{jk}\dot{q}^k - h\right)}\ . \tag{1.254}$$

Here $\hat{H}$ is the Hamiltonian, and $h(q)$ its symbol. The measure formally implies a sum over all phase-space paths pinned at the initial and final points, and a Wiener measure regularization implies the following replacement

$$\mathcal{D}[q] \to \mathcal{D}[\mu_\nu(q)] = \mathcal{D}[q]\, \mathrm{e}^{-\frac{1}{2\nu}\int_0^t d\tau\, \dot{q}^2} = N_\nu(t)\, d\mu_W^\nu(q)\ . \tag{1.255}$$

The factor $N_\nu(t)$ equals $2\pi \mathrm{e}^{\nu t/2}$ for every degree of freedom, $d\mu_W^\nu(q)$ stands for the Wiener measure, and ν denotes the diffusion constant. We denote by $K_t^\nu(q,q')$ the integral kernel of the evolution operator for a finite ν. The Wiener measure determines a stochastic process on the *flat* phase–space. The integral of the symplectic 1–form $\int q\omega dq$ is a stochastic integral that is interpreted in the Stratonovich sense. Under general coordinate transformations $q = q(\bar{q})$, the Wiener measure describes *the same* stochastic process on *flat* space in the curvilinear coordinates $dq^2 = d\sigma(\bar{q})^2$, so that the value of the integral is not changed apart from a possible phase term. After the calculation of the integral, the evolution operator kernel is get by taking the limit $\nu \to \infty$. The existence of this limit, and also the covariance under general phase-space coordinate transformations, can be proved through the *operator* formalism for the regularized kernel $K_t^\nu(q,q')$.

Note that the integral (1.254) with the Wiener measure inserted can be regarded as an ordinary Lagrangian path integral with a complex action, where the configuration space is the original phase–space and the Hamiltonian $h(q)$ serves as a potential. Making use of this observation it is not hard to derive the corresponding Schrödinger–like equation

$$\partial_t K_t^\nu(q,q') = \left[\frac{\nu}{2}\left(\partial_{q^j} + \frac{\mathrm{i}}{2}\omega_{jk}q^k\right)^2 - \mathrm{i}h(q)\right] K_t^\nu(q,q')\ , \tag{1.256}$$

subject to the initial condition $K_{t=0}^\nu(q,q') = \delta(q-q'), 0 < \nu < \infty$. One can show that $\hat{K}_t^\nu \to \hat{K}_t$ as $\nu \to \infty$ for all $t > 0$. The covariance under general coordinate transformations follows from the covariance of the "kinetic" energy of the Schrödinger operator in (1.256): The Laplace operator is replaced by the Laplace–Beltrami operator in the new curvilinear coordinates $q = q(\bar{q})$, so the solution is not changed, but written in the new coordinates. This is similar to the covariance of the ordinary Schrödinger equation and the corresponding *Lagrangian* path integral relative to general coordinate transformations on the configuration space: The kinetic energy operator (the Laplace operator) in the ordinary Schrödinger equation gives a term *quadratic* in time derivatives in the path integral measure which is sufficient for the general coordinate covariance. We remark that the regularization procedure based on the modified Schrödinger equation (1.256) applies to far more general Hamiltonians than those quadratic in canonical momenta and leading to the conventional *Lagrangian* path integral.

1.5.4 Sum over Geometries and Topologies

Recall that the term *quantum gravity* (or *quantum geometrodynamics*, or *quantum geometry*), is usually understood as a consistent fundamental quantum description of gravitational space–time geometry whose classical limit is Einstein's general relativity. Among the possible ramifications of such a theory are

a model for the structure of space–time near the Planck scale, a consistent calculational scheme to calculate gravitational effects at all energies, a description of quantum geometry near space–time singularities and a non–perturbative quantum description of 4D black holes. It might also help us in understanding cosmological issues about the beginning and end of the universe, i.e., the so–called *'big bang'* and *'big crunch'* (see e.g., [Pen89, Pen94, Pen97]).

From what we know about the quantum dynamics of other fundamental interactions it seems eminently plausible that also the gravitational excitations should at very short scales be governed by quantum laws. Now, conventional *perturbative path integral expansions of gravity*, as well as perturbative expansion in the string coupling in the case of unified approaches, both have difficulty in finding any direct or indirect evidence for quantum gravitational effects, be they experimental or observational, which could give a feedback for model building. The outstanding problems mentioned above require a non–perturbative treatment; it is not sufficient to know the first few terms of a perturbation series. The real goal is to search for a *non–perturbative* definition of such a theory, where the initial input of any fixed 'background metric' is inessential (or even undesirable), and where 'space–time' is determined *dynamically*. Whether or not such an approach necessarily requires the inclusion of higher dimensions and fundamental supersymmetry is currently unknown (see [AK93, AL98, AJL00a, AJL00b, AJL01a, AJL01b, AJL01c, DL01]).

Such a non–perturbative viewpoint is very much in line with how one proceeds in classical geometrodynamics (see introductory subsection 1.1), where a metric space–time $(M, g_{\mu\nu})$ (+ matter) emerges only as a *solution* to the familiar *Einstein equation*

$$G_{\mu\nu}[g] \equiv R_{\mu\nu}[g] - \frac{1}{2} g_{\mu\nu} R[g] = -8\pi T_{\mu\nu}[\Phi], \tag{1.257}$$

which define the classical dynamics of fields $\Phi = \Phi^{\mu\nu}$ on the space $\mathcal{M}(M)$, the space of all metrics $g = g_{\mu\nu}$ on a given smooth manifold M. The analogous question we want to address in the quantum theory is: Can we get 'quantum space–time' as a solution to a set of non–perturbative quantum equations of motion on a suitable quantum analogue of $\mathcal{M}(M)$ or rather, of the space of geometries, $Geom(M) = \mathcal{M}(M)/Diff(M)$?

Now, this is not a completely straightforward task. Whichever way we want to proceed non–perturbatively, if we give up the privileged role of a flat, Minkowskian background space–time on which the quantization is to take place, we also have to abandon the central role usually played by the Poincaré group, and with it most standard quantum field–theoretic tools for regularization and renormalization. If one works in a continuum metric formulation of gravity, the symmetry group of the *Einstein–Hilbert action* is instead the group $Diff(M)$ of diffeomorphisms on M, which in terms of local charts are the smooth invertible coordinate transformations $x^\mu \mapsto y^\mu(x^\mu)$.

In the following, we will describe a non–perturbative path integral approach to quantum gravity, defined on the *space of all geometries*, without

distinguishing any background metric structure [Lol01]. This is closely related in spirit with the canonical approach of loop quantum gravity [Rov98] and its more recent incarnations using so–called *spin networks* (see, e.g., [Ori01]). 'Non–perturbative' here means in a covariant context that the path sum or integral will have to be performed explicitly, and not just evaluated around its stationary points, which can only be achieved in an appropriate regularization. The method we will employ uses a discrete lattice regularization as an intermediate step in the construction of the quantum theory.

Simplicial Quantum Geometry

In this section we will explain how one may construct a theory of quantum gravity from a non–perturbative path integral, using the method of Lorentzian dynamical triangulations. The method is minimal in the sense of employing standard tools from quantum field theory and the theory of critical phenomena and adapting them to the case of *generally covariant systems*, without invoking any symmetries beyond those of the classical theory. At an intermediate stage of the construction, we use a regularization in terms of simplicial *Regge geometries*, that is, piecewise linear manifolds. In this approach, 'computing the path integral' amounts to a conceptually simple and geometrically transparent 'counting of geometries', with additional weight factors which are determined by the EH action. This is done first of all at a regularized level. Subsequently, one searches for interesting continuum limits of these discrete models which are possible candidates for theories of quantum gravity, a step that will always involve a renormalization. From the point of view of statistical mechanics, one may think of Lorentzian dynamical triangulations as a new class of statistical models of Lorentzian random surfaces in various dimensions, whose building blocks are flat simplices which carry a 'time arrow', and whose dynamics is entirely governed by their intrinsic geometric properties.

Before describing the details of the construction, it may be helpful to recall the path integral representation for a 1D non–relativistic particle (see previous subsection). The time evolution of the particle's wave function ψ may be described by the integral equation (1.241) above, where the *propagator*, or the *Feynman kernel* G, is defined through a limiting procedure (1.242). The time interval $t''-t'$ has been discretized into N steps of length $\epsilon = (t''-t')/N$, and the r.h.s. of (1.242) represents an integral over all piecewise linear paths $x(t)$ of a 'virtual' particle propagating from x' to x'', illustrated in Figure 1.13 above.

The prefactor A^{-N} is a normalization and L denotes the Lagrangian function of the particle. Knowing the propagator G is tantamount to having solved the quantum dynamics. This is the simplest instance of a path integral, and is often written schematically as

$$G(x',t';x'',t'') = \int\!\!\!\!\!\!\int \mathcal{D}[x(t)]\ \mathrm{e}^{\mathrm{i}S[x(t)]}, \tag{1.258}$$

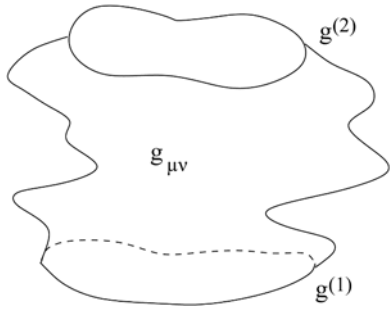


Fig. 1.14. The time–honoured way [HE79] of illustrating the gravitational path integral as the propagator from an initial to a final spatial boundary geometry.

where $\mathcal{D}[x(t)]$ is a functional measure on the 'space of all paths', and the exponential weight depends on the classical action $S[x(t)]$ of a path. Recall also that this procedure can be defined in a mathematically clean way if we Wick–rotate the time variable t to imaginary values $t \mapsto \tau = \mathrm{i}t$, thereby making all integrals real [RS75].

Can a similar strategy work for the case of Einstein geometrodynamics? As an analogue of the particle's position we can take the geometry $[g_{ij}(x)]$ (i.e., an equivalence class of spatial metrics) of a constant–time slice. Can one then define a gravitational propagator

$$G([g'_{ij}], [g''_{ij}]) = \oint\!\!\!\!\!\!\int_{Geom(M)} \mathcal{D}[g_{\mu\nu}]\ \mathrm{e}^{\mathrm{i}S^{\mathrm{EH}}[g_{\mu\nu}]} \tag{1.259}$$

from an initial geometry $[g']$ to a final geometry $[g'']$ (Figure 1.14) as a limit of some discrete construction analogous to that of the non-relativistic particle (1.242)? And crucially, what would be a suitable class of 'paths', that is, space–times $[g_{\mu\nu}]$ to sum over?

Now, to be able to perform the integration $\oint\!\!\!\!\!\!\int \mathcal{D}[g_{\mu\nu}]$ in a meaningful way, the strategy we will be following starts from a regularized version of the space $Geom(M)$ of all geometries. A regularized path integral $G(a)$ can be defined which depends on an ultraviolet cutoff a and is *convergent* in a non–trivial region of the space of coupling constants. Taking the continuum limit corresponds to letting $a \to 0$. The resulting continuum theory – if it can be shown to exist – is then investigated with regard to its geometric properties and in particular its semiclassical limit.

Discrete Gravitational Path Integrals

Trying to construct non–perturbative path integrals for gravity from sums over discretized geometries, using approach of *Lorentzian dynamical triangulations*, is not a new idea. Inspired by the successes of lattice gauge theory, attempts to describe quantum gravity by similar methods have been popular

on and off since the late 70's. Initially the emphasis was on gauge–theoretic, first–order formulations of gravity, usually based on (compactified versions of) the Lorentz group, followed in the 80's by 'quantum Regge calculus', an attempt to represent the gravitational path integral as an integral over certain piecewise linear geometries (see [Wil97] and references therein), which had first made an appearance in approximate descriptions of *classical* solutions of the Einstein equations. A variant of this approach by the name of 'dynamical triangulation(s)' attracted a lot of interest during the 90's, partly because it had proved a powerful tool in describing 2D quantum gravity (see the textbook [ADJ97] and lecture notes [AJL00a] for more details).

The problem is that none of these attempts have so far come up with convincing evidence for the existence of an underlying continuum theory of 4D quantum gravity. This conclusion is drawn largely on the basis of numerical simulations, so it is by no means water–tight, although one can make an argument that the 'symptoms' of failure are related in the various approaches [Lol98]. What goes wrong generically seems to be a dominance in the continuum limit of highly degenerate geometries, whose precise form depends on the approach chosen. One would expect that non–smooth geometries play a decisive role, in the same way as it can be shown in the particle case that the support of the measure in the continuum limit is on a set of nowhere differentiable paths. However, what seems to happen in the case of the path integral for 4–geometries is that the structures get are *too* wild, in the sense of not generating, even at coarse–grained scales, an effective geometry whose dimension is anywhere near four.

The schematic phase diagram of Euclidean dynamical triangulations shown in Figure 1.15 gives an example of what can happen. The picture turns out to be essentially the same in both three and four dimensions: the model possesses infinite-volume limits everywhere along the critical line $k_3^{\rm crit}(k_0)$, which fixes the bare cosmological constant as a function of the inverse Newton constant $k_0 \sim G_N^{-1}$. Along this line, there is a critical point $k_0^{\rm crit}$ (which we now know to be of first–order in $d = 3, 4$) below which geometries generically have a very large effective or Hausdorff dimension.[23] Above $k_0^{\rm crit}$ we find the opposite phenomenon of 'polymerization': a typical element contributing to the state sum is a thin branched polymer, with one or more dimensions 'curled up' such that its effective dimension is around two.

This problem has to do with the fact that the gravitational action is unbounded below, causing potential havoc in Euclidean versions of the path integral. Namely, what all the above-mentioned approaches have in common is that they work from the outset with *Euclidean* geometries, and associated Boltzmann-type weights $\exp(-S^{\rm eu})$ in the path integral. In other words, they integrate over 'space–times' which know nothing about time, light cones and

[23] In terms of geometry, this means that there are a few vertices at which the entire space–time 'condenses' in the sense that almost every other vertex in the simplicial space–time is about one link-distance away from them.

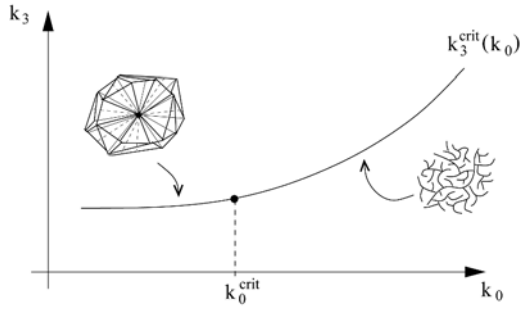


Fig. 1.15. The phase diagram of 3D and 4D Euclidean dynamical triangulations (adapted from [AJL00b, AJL01a]).

causality. This is done mainly for technical reasons, since it is difficult to set up simulations with complex weights and since until recently a suitable Wick rotation was not known.

'Lorentzian dynamical triangulations', first proposed in [AL98] and further elaborated in [AJL00b, AJL01a] tries to establish a logical connection between the fact that non–perturbative path integrals were constructed for Euclidean instead of Lorentzian geometries and their apparent failure to lead to an interesting continuum theory.

Regge Calculus

The use of simplicial methods in general relativity goes back to the pioneering work of Regge [Reg61]. In classical applications one tries to approximate a classical space–time geometry by a triangulation, that is, a piecewise linear space get by gluing together flat simplicial building blocks, which in dimension d are dD generalizations of triangles. By 'flat' we mean that they are isometric to a subspace of dD Euclidean or Minkowski space. We will only be interested in gluings leading to genuine manifolds, which therefore look locally like an R^d. A nice feature of such simplicial manifolds is that their geometric properties are completely described by the discrete set $\{l_i^2\}$ of the squared lengths of their edges. Note that this amounts to a description of geometry *without the use of coordinates*. There is nothing to prevent us from re–introducing coordinate patches covering the piecewise linear manifold, for example, on each individual simplex, with suitable transition functions between patches. In such a coordinate system the metric tensor will then assume a definite form. However, for the purposes of formulating the path integral we will not be interested in doing this, but rather work with the edge lengths, which constitute a direct, regularized parametrization of the space $Geom(M)$ of geometries.

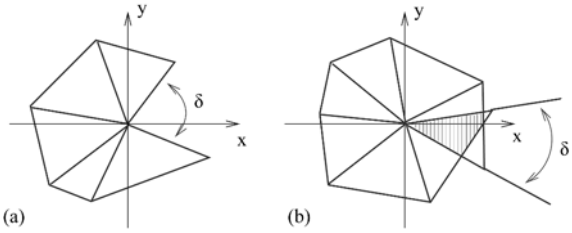


Fig. 1.16. Positive (a) and negative (b) space–like deficit angles δ (adapted from [Lol01, Lol98]).

How precisely is the intrinsic geometry of a simplicial space, most importantly, its curvature, encoded in its edge lengths? A useful example to keep in mind is the case of dimension two, which can easily be visualized. A 2d piecewise linear space is a triangulation, and its scalar curvature $R(x)$ coincides with the *Gaussian curvature* (see subsection 1.2.9 above). One way of measuring this curvature is by parallel–transporting a vector around closed curves in the manifold. In our piecewise–flat manifold such a vector will always return to its original orientation *unless* it has surrounded lattice vertices v at which the surrounding angles did not add up to 2π, but $\sum_{i\supset v} \alpha_i = 2\pi - \delta$, for $\delta \neq 0$, see Figure 1.16. The so–called deficit angle δ is precisely the rotation angle picked up by the vector and is a direct measure for the scalar curvature at the vertex. The operational description to get the scalar curvature in higher dimensions is very similar, one basically has to sum in each point over the Gaussian curvatures of all 2D submanifolds. This explains why in Regge calculus the curvature part of the EH action is given by a sum over building blocks of dimension $(d-2)$ which are the objects dual to those local 2d submanifolds. More precisely, the continuum curvature and volume terms of the action become

$$\frac{1}{2}\int_{\mathcal{R}} d^d x\ \sqrt{|\det g|}^{(d)} R \longrightarrow \sum_{i\in\mathcal{R}} Vol(i^{th}\ (d-2)-\text{simplex})\ \delta_i \quad (1.260)$$

$$\int_{\mathcal{R}} d^d x\ \sqrt{|\det g|} \longrightarrow \sum_{i\in\mathcal{R}} Vol(i^{th}\ d-\text{simplex}) \quad (1.261)$$

in the simplicial discretization. It is then a simple exercise in trigonometry to express the volumes and angles appearing in these formulas as functions of the edge lengths l_i, both in the Euclidean and the Minkowskian case.

The approach of dynamical triangulations uses a certain class of such simplicial space–times as an explicit, regularized realization of the space $Geom(M)$. For a given volume N_d, this class consists of all gluings of manifold–type of a set of N_d simplicial building blocks of top–dimension d whose edge lengths are restricted to take either one or one out of two values. In the Euclidean case we set $l_i^2 = a^2$ for all i, and in the Lorentzian case we allow for both space- and time–like links with $l_i^2 \in \{-a^2, a^2\}$, where the geodesic

distance a serves as a short-distance cutoff, which will be taken to zero later. Coming from the classical theory this may seem a grave restriction at first, but this is indeed not the case. Firstly, keep in mind that for the purposes of the quantum theory we want to sample the space of geometries 'ergodically' at a coarse-grained scale of order a. This should be contrasted with the classical theory where the objective is usually to approximate a given, *fixed* space–time to within a length scale a. In the latter case one typically requires a much finer topology on the space of metrics or geometries. It is also straightforward to see that no local curvature degrees of freedom are suppressed by fixing the edge lengths; deficit angles in all directions are still present, although they take on only a discretized set of values. In this sense, in dynamical triangulations all geometry is in the gluing of the fundamental building blocks. This is dual to how quantum Regge calculus is set up, where one usually fixes a triangulation T and then 'scans' the space of geometries by letting the l_i's run continuously over all values compatible with the triangular inequalities.

In a nutshell, Lorentzian dynamical triangulations give a definite meaning to the 'integral over geometries', namely, as a sum over inequivalent Lorentzian gluings T over any number N_d of $d-$simplices,

$$\oint_{Geom(M)} \mathcal{D}[g_{\mu\nu}]\, \mathrm{e}^{\mathrm{i}S[g_{\mu\nu}]} \quad \stackrel{\mathrm{LDT}}{\longrightarrow} \quad \sum_{T\in\mathcal{T}} \frac{1}{C_T}\, \mathrm{e}^{\mathrm{i}S^{\mathrm{Reg}}(T)}, \tag{1.262}$$

where the symmetry factor $C_T = |\mathrm{Aut}(T)|$ on the r.h.s. is the order of the automorphism group of the triangulation, consisting of all maps of T onto itself which preserve the connectivity of the simplicial lattice. We will specify below what precise class $\mathcal{T}$ of triangulations should appear in the summation.

It follows from the above that in this formulation all curvatures and volumes contributing to the *Regge simplicial action* come in discrete units. This can be illustrated by the case of a 2D triangulation with Euclidean signature, which according to the prescription of dynamical triangulations consists of *equilateral* triangles with squared edge lengths $+a^2$. All interior angles of such a triangle are equal to $\pi/3$, which implies that the deficit angle at any vertex v can take the values $2\pi - k_v\pi/3$, where k_v is the number of triangles meeting at v. As a consequence, the Einstein–Regge action S^{Reg} assumes the simple form

$$S^{\mathrm{Reg}}(T) = \kappa_{d-2}N_{d-2} - \kappa_d N_d, \tag{1.263}$$

where the coupling constants $\kappa_i = \kappa_i(\lambda, G_N)$ are simple functions of the bare cosmological and Newton constants in d dimensions. Substituting this into the path sum in (1.262) leads to

$$Z(\kappa_{d-2},\kappa_d) = \sum_{N_d} \mathrm{e}^{-\mathrm{i}\kappa_d N_d} \sum_{N_{d-2}} \mathrm{e}^{\mathrm{i}\kappa_{d-2}N_{d-2}} \sum_{T|_{N_d,N_{d-2}}} \frac{1}{C_T}, \tag{1.264}$$

The point of taking separate sums over the numbers of $d-$ and $(d-2)-$simplices in (1.264) is to make explicit that 'doing the sum' is tantamount to the combinatorial problem of *counting* triangulations of a given volume and number of

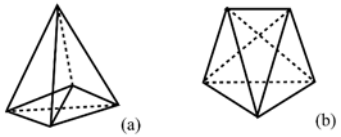


Fig. 1.17. Two types of Minkowskian 4–simplices in 4D (adapted from [Lol01, Lol98]).

simplices of codimension 2 (corresponding to the last summation in (1.264)).[24] It turns out that at least in two space–time dimensions the counting of geometries can be done completely explicitly, turning both Lorentzian and Euclidean quantum gravity into exactly soluble statistical models.

Lorentzian Path Integral

Now, the simplicial building blocks of the models are taken to be pieces of Minkowski space, and their edges have squared lengths $+a^2$ or $-a^2$. For example, the two types of 4–simplices that are used in Lorentzian dynamical triangulations in dimension four are shown in Figure 1.17. The first of them has four time–like and six space–like links (and therefore contains 4 time–like and 1 space–like tetrahedron), whereas the second one has six time–like and four space–like links (and contains 5 time–like tetrahedra). Since both are subspaces of flat space with signature $(-+++)$, they possess well–defined light–cone structures everywhere [Lol01, Lol98].

In general, gluings between pairs of $d-$simplices are only possible when the metric properties of their $(d-1)-$faces match. Having local light cones implies causal relations between pairs of points in local neighborhoods. Creating closed time–like curves will be avoided by requiring that all space–times contributing to the path sum possess a global 'time' function t. In terms of the triangulation this means that the $d-$simplices are arranged such that their space–like links all lie in slices of constant integer t, and their time–like links interpolate between adjacent spatial slices t and $t+1$. Moreover, with respect to this time, we will not allow for any *spatial* topology changes[25].

This latter condition is always satisfied in classical applications, where 'trouser points' like the one depicted in Figure 1.18 are ruled out by the requirement of having a non–degenerate Lorentzian metric defined everywhere on M (it is geometrically obvious that the light cone and hence $g_{\mu\nu}$ must degenerate in at least one point along the 'crotch'). Another way of thinking about such configurations (and their time–reversed counterparts) is that the causal past (future) of an observer changes discontinuously as her world–line

[24] The symmetry factor C_T is almost always equal to 1 for large triangulations.

[25] Note that if we were in the continuum and had introduced coordinates on space–time, such a statement would actually be diffeomorphism–invariant.

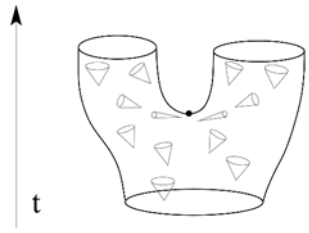


Fig. 1.18. At a branching point associated with a spatial topology change, light-cones get 'squeezed' [Lol01, Lol98].

passes near the singular point (see [Dow02] and references therein for related discussions about the issue of topology change in quantum gravity).

There is no *a priori* reason in the quantum theory to not relax some of these classical causality constraints. After all, as we stressed right at the outset, path integral histories are not in general classical solutions, nor can we attribute any other direct physical meaning to them individually. It might well be that one can construct models whose path integral configurations violate causality in this strict sense, but where this notion is somehow recovered in the resulting continuum theory. What the approach of Lorentzian dynamical triangulations has demonstrated is that *imposing causality constraints will in general lead to a different continuum theory.* This is in contrast with the intuition one may have that 'including a few isolated singular points will not make any difference'. On the contrary, tampering with causality in this way is not innocent at all, as was already anticipated by Teitelboim many years ago [Tei83].

We want to point out that one cannot conclude from the above that spatial topology changes or even *fluctuations in the space–time topology* cannot be treated in the formulation of dynamical triangulations. However, if one insists on including geometries of variable topology in a Lorentzian discrete context, one has to come up with a prescription of how to weigh these singular points in the path integral, both before and after the Wick rotation [Das02]. Maybe this can be done along the lines suggested in [LS97]; this is clearly an interesting issue for further research.

Having said this, we next have to address the question of the *Wick rotation*, in other words, of how to get rid of the factor of i in the exponent of (1.264). Without it, this expression is an infinite sum (since the volume can become arbitrarily large) of complex terms whose convergence properties will be very difficult to establish. In this situation, a Wick rotation is simply a technical tool which – in the best of all worlds – enables us to perform the state sum and determine its continuum limit. The end result will have to be Wick–rotated back to Lorentzian signature.

Fortunately, Lorentzian dynamical triangulations come with a natural notion of Wick rotation, and the strategy we just outlined can be carried out

explicitly in two space–time dimensions, leading to a unitary theory. In higher dimensions we do not yet have sufficient analytical control of the continuum theories to make specific statements about the *inverse* Wick rotation. Since we use the Wick rotation at an intermediate step, one can ask whether other Wick rotations would lead to the same result. Currently this is a somewhat academic question, since it is in practice difficult to find such alternatives. In fact, it is quite miraculous we have found a single prescription for Wick–rotating in our regularized setting, and it does not seem to have a direct continuum analogue (for more comments on this issue, see [DL01, Das02]).

Our Wick rotation W in any dimension is an injective map from Lorentzian– to Euclidean–signature simplicial space–times. Using the notation $\mathtt{T}$ for a simplicial manifold together with length assignments l_s^2 and l_t^2 to its space– and time–like links, it is defined by

$$\mathtt{T}^{\rm lor} = (T, \{l_s^2 = a^2, l_t^2 = -a^2\}) \quad \stackrel{W}{\longmapsto} \quad \mathtt{T}^{\rm eu} = (T, \{l_s^2 = a^2, l_t^2 = a^2\}). \quad (1.265)$$

Note that we have not touched the connectivity of the simplicial manifold T, but only its metric properties, by mapping all time–like links of T into space–like ones, resulting in a Euclidean 'space–time' of equilateral building blocks. It can be shown [AJL01a] that at the level of the corresponding weight factors in the path integral this Wick rotation[26] has precisely the desired effect of rotating to the exponentiated Regge action of the 'Euclideanized' geometry,

$$\mathrm{e}^{\mathrm{i}S(\mathcal{T}^{\rm lor})} \quad \stackrel{W}{\longmapsto} \quad \mathrm{e}^{-S(\mathcal{T}^{\rm eu})}. \quad (1.266)$$

The Euclideanized path sum after the Wick rotation has the form

$$\begin{aligned} Z^{\rm eu}(\kappa_{d-2}, \kappa_d) &= \sum_T \frac{1}{C_T}\, \mathrm{e}^{-\kappa_d N_d(T) + \kappa_{d-2} N_{d-2}(T)} \\ &= \sum_{N_d} \mathrm{e}^{-\kappa_d N_d} \sum_{T|_{N_d}} \frac{1}{C_T}\, \mathrm{e}^{\kappa_{d-2} N_{d-2}(T)} \\ &= \sum_{N_d} \mathrm{e}^{-\kappa_d N_d}\, \mathrm{e}^{\kappa_d^{\rm crit}(\kappa_{d-2}) N_d} \times \text{subleading}(N_d). \end{aligned} \quad (1.267)$$

In the last equality we have used that the number of Lorentzian triangulations of discrete volume N_d to leading order scales exponentially with N_d for large volumes. This can be shown explicitly in space–time dimension 2 and 3. For $d = 4$, there is strong (numerical) evidence for such an exponential bound for *Euclidean triangulations*, from which the desired result for the Lorentzian case follows (since W maps to a strict subset of all Euclidean simplicial manifolds).

[26] To get a genuine Wick rotation and not just a discrete map, one introduces a complex parameter α in $l_t^2 = -\alpha a^2$. The proper prescription leading to (1.266) is then an analytic continuation of α from 1 to -1 through the lower–half complex plane.

From the functional form of the last line of (1.267) one can immediately read off some qualitative features of the phase diagram, an example of which appeared already earlier in Figure 1.15. Namely, the sum over geometries Z^{eu} converges for values $\kappa_d > \kappa_d^{\mathrm{crit}}$ of the bare cosmological constant, and diverges (ie. is not defined) below this critical line. Generically, for all models of dynamical triangulations the infinite–volume limit is attained by approaching the critical line $\kappa_d^{\mathrm{crit}}(\kappa_{d-2})$ from above, ie. from inside the region of convergence of Z^{eu}. In the process of taking $N_d \to \infty$ and the cutoff $a \to 0$, one gets a renormalized cosmological constant Λ through

$$(\kappa_d - \kappa_d^{\mathrm{crit}}) = a^{\mu}\Lambda + O(a^{\mu+1}). \tag{1.268}$$

If the scaling is canonical (which means that the dimensionality of the renormalized coupling constant is the one expected from the classical theory), the exponent is given by $\mu = d$. Note that this construction requires a positive *bare* cosmological constant in order to make the state sum converge. Moreover, by virtue of relation (1.268) also the *renormalized* cosmological constant must be positive. Other than that, its numerical value is not determined by this argument, but by comparing observables of the theory which depend on Λ with actual physical measurements.[27] Another interesting observation is that the inclusion of a sum over topologies in the discretized sum (1.267) would lead to a super–exponential growth of at least $\propto N_d!$ of the number of triangulations with the volume N_d. Such a divergence of the path integral cannot be compensated by an additive renormalization of the cosmological constant of the kind outlined above.

There are ways in which one can sum divergent series of this type, for example, by performing a Borel sum. The problem with these stems from the fact that two different functions can share the same asymptotic expansion. Therefore, the series in itself is *not* sufficient to define the underlying theory uniquely. The non–uniqueness arises because of non–perturbative contributions to the path integral which are not represented in the perturbative expansion.[28] In order to fix these uniquely, an independent, non–perturbative definition of the theory is necessary. Unfortunately, for dynamically triangulated models of quantum gravity, no such definitions have been found so far. In the context of 2D (Euclidean) quantum gravity this difficulty is known as the 'absence of a physically motivated double-scaling limit' [AK93].

Lastly, getting an interesting continuum limit may or may not require an additional fine–tuning of the inverse gravitational coupling κ_{d-2}, depending on the dimension d. In four dimensions, one would expect to find a second-order transition along the critical line, corresponding to local gravitonic excitations. The situation in $d = 3$ is less clear, but results get so far indicate that no fine–tuning of Newton's constant is necessary [AJL01b, AJL01c].

[27] The non–negativity of the renormalized cosmological coupling may be taken as a first 'prediction' of our construction, which in the physical case of four dimensions is indeed in agreement with current observations.

[28] A field–theoretic example would be instantons and renormalons in QCD.

Before delving into the details, let me summarize briefly the results that have been get so far in the approach of Lorentzian dynamical triangulations. At the regularized level, that is, in the presence of a finite cutoff a for the edge lengths and an infrared cutoff for large space–time volume, they are well–defined statistical models of Lorentzian random geometries in $d = 2, 3, 4$. In particular, they obey a suitable notion of reflection-positivity and possess self–adjoint Hamiltonians.

The crucial questions are then to what extent the underlying combinatorial problems of counting all dD geometries with certain causal properties can be solved, whether continuum theories with non–trivial dynamics exist and how their bare coupling constants get renormalized in the process. What we know about Lorentzian dynamical triangulations so far is that they lead to continuum theories of quantum gravity in dimension 2 and 3. In $d = 2$, there is a complete analytic solution, which is distinct from the continuum theory produced by Euclidean dynamical triangulations. Also the matter–coupled model has been studied. In $d = 3$, there are numerical and partial analytical results which show that both a continuum theory exists and that it again differs from its Euclidean counterpart. Work on a more complete analytic solution which would give details about the geometric properties of the quantum theory is under way. In $d = 4$, the first numerical simulations are currently being set up. The challenge here is to do this for sufficiently large lattices, to be able to perform meaningful measurements. So far, we cannot make any statements about the existence and properties of a continuum theory in this physically most interesting case.

1.5.5 TQFT and Stringy Path Integrals

The *string theory* is a special case of a *quantum field theory* (QFT). Any QFT deals with smooth maps $\gamma : \Sigma \to M$ of Riemannian manifolds Σ and M such that the dimension of Σ is the dimension of the theory. On the set $Map(\Sigma, M)$ of all smooth maps $\gamma = \gamma(\phi)$, we also have defined an *action function* $S[\phi]$ of the *field variables* ϕ. A *non–relativistic* QFT studies *real–valued* (Euclidean) *path integrals* of the form

$$\int_{Map(\Sigma,M)} V(\phi)\mathcal{D}[\phi]\, \mathrm{e}^{-S[\phi]/\hbar}, \tag{1.269}$$

where $\mathcal{D}[\phi]$ represents some measure on the space of paths, is the *Planck constant* and $V : Map(\Sigma, M) \to \mathbb{R}$ is an *insertion function*. The number $\mathrm{e}^{-S[\phi]/\hbar}$ should be interpreted as the *probability amplitude* of the contribution of the map $\gamma : \Sigma \to M$ to the path integral. The associated integral

$$Z^E = \int_{Map(\Sigma,M)} d\phi\, \mathrm{e}^{-S[\phi]/\hbar},$$

is the *partition function* of the theory. In a *relativistic* QFT, the space Σ has a Lorentzian metric of signature $(-, +, ..., +)$. The first coordinate is reserved

for time, the rest are for space. In this case, the real–valued path integral (1.269) is replaced with the *complex–valued path integral*

$$Z^M = \int_{Map(\Sigma,M)} V(\phi)\mathcal{D}[\phi]\, \mathrm{e}^{\mathrm{i}S[\phi]/\hbar}.$$

Witten's Topological Quantum Field Theory

Before we come to (super)strings, we give a brief on *topological quantum field theory* (TQFT), as developed by Fields Medalist Edward Witten, from his original path integral point of view (see [Wit88, LL98]). TQFT originated in 1982, when Ed Witten rewrote classical *Morse theory* (see subsection 1.2.9 above, as well as subsection 2.4.4 below) in Dick Feynman's language of quantum field theory [Wit82]. Witten's arguments made use of Feynman's path integrals and consequently, at first, they were regarded as mathematically non–rigorous. However, a few years later, A. Floer reformulated a rigorous Morse–Witten theory [Flo87] (that won a Fields medal for Witten). This trend in which some mathematical structure is first constructed by quantum field theory methods and then reformulated in a rigorous mathematical ground constitutes one of the tendencies in modern physics.

In TQFT our basic topological space is an nD Riemannian manifold M with a metric $g_{\mu\nu}$. Let us consider on it a set of fields $\{\phi_i\}$, and let $S[\phi_i]$ be a real functional of these fields which is regarded as the *action* of the theory. We consider 'operators', $O_\alpha(\phi_i)$, which are in general arbitrary functionals of the fields. In TQFT these functionals are real functionals labelled by some set of indices α carrying topological or group–theoretical data. The *vacuum expectation value* (VEV) of a product of these operators is defined as

$$\langle O_{\alpha_1} O_{\alpha_2} \cdots O_{\alpha_p} \rangle = \int [D\phi_i] O_{\alpha_1}(\phi_i) O_{\alpha_2}(\phi_i) \cdots O_{\alpha_p}(\phi_i) \exp\left(-S[\phi_i]\right).$$

A quantum field theory is considered *topological* if the following relation is satisfied:

$$\frac{\delta}{\delta g^{\mu\nu}} \langle O_{\alpha_1} O_{\alpha_2} \cdots O_{\alpha_p} \rangle = 0, \tag{1.270}$$

i.e., if the VEV of some set of selected operators is independent of the metric $g_{\mu\nu}$ on M. If such is the case those operators are called 'observables'.

There are two ways to guarantee, at least formally, that condition (1.270) is satisfied. The first one corresponds to the situation in which both, the action $S[\phi_i]$, as well as the operators O_{α_i} are *metric independent*. These TQFTs are called of *Schwarz* type. The most important representative is *Chern–Simons gauge theory*. The second one corresponds to the case in which there exist a *symmetry*, whose infinitesimal form is denoted by δ, satisfying the following properties:

$$\delta O_{\alpha_i} = 0, \qquad T_{\mu\nu} = \delta G_{\mu\nu}, \tag{1.271}$$

where $T_{\mu\nu}$ is the SEM–tensor of the theory, i.e.,

$$T_{\mu\nu}(\phi_i) = \frac{\delta}{\delta g^{\mu\nu}} S[\phi_i]. \tag{1.272}$$

The fact that δ in (1.271) is a symmetry of the theory implies that the transformations $\delta\phi_i$ of the fields are such that both $\delta A[\phi_i] = 0$ and $\delta O_{\alpha_i}(\phi_i) = 0$. Conditions (1.271) lead, at least formally, to the following relation for VEVs:

$$\begin{aligned}\frac{\delta}{\delta g^{\mu\nu}}\langle O_{\alpha_1} O_{\alpha_2} \cdots O_{\alpha_p}\rangle &= -\int [D\phi_i] O_{\alpha_1}(\phi_i) O_{\alpha_2}(\phi_i) \cdots O_{\alpha_p}(\phi_i) T_{\mu\nu} \mathrm{e}^{-S[\phi_i]} \\ &= -\int [D\phi_i] \delta\left(O_{\alpha_1}(\phi_i) O_{\alpha_2}(\phi_i) \cdots O_{\alpha_p}(\phi_i) G_{\mu\nu} \exp\left(-S[\phi_i]\right)\right) = 0, \end{aligned} \tag{1.273}$$

which implies that the quantum field theory can be regarded as topological. This second type of TQFTs are called of *Witten type*. One of its main representatives is the theory related to Donaldson invariants, which is a twisted version of $N = 2$ supersymmetric *Yang–Mills gauge theory*. It is important to remark that the symmetry δ must be a scalar symmetry, i.e., that its symmetry parameter must be a scalar. The reason is that, being a global symmetry, this parameter must be covariantly constant and for arbitrary manifolds this property, if it is satisfied at all, implies strong restrictions unless the parameter is a scalar.

Most of the TQFTs of *cohomological type* satisfy the relation:

$$S[\phi_i] = \delta\Lambda(\phi_i), \tag{1.274}$$

for some functional $\Lambda(\phi_i)$. This has far–reaching consequences, for it means that the topological observables of the theory, in particular the partition function, (path integral) itself are independent of the value of the coupling constant. Indeed, let us consider for example the VEV:

$$\langle O_{\alpha_1} O_{\alpha_2} \cdots O_{\alpha_p}\rangle = \int [D\phi_i] O_{\alpha_1}(\phi_i) O_{\alpha_2}(\phi_i) \cdots O_{\alpha_p}(\phi_i)\, \mathrm{e}^{-\frac{1}{g^2} S[\phi_i]}. \tag{1.275}$$

Under a change in the coupling constant, $1/g^2 \to 1/g^2 - \Delta$, one has (assuming that the observables do not depend on the coupling), up to first–order in Δ:

$$\begin{aligned} &\langle O_{\alpha_1} O_{\alpha_2} \cdots O_{\alpha_p}\rangle \longrightarrow \langle O_{\alpha_1} O_{\alpha_2} \cdots O_{\alpha_p}\rangle \\ &+ \Delta \int [D\phi_i] \delta\left[O_{\alpha_1}(\phi_i) O_{\alpha_2}(\phi_i) \cdots O_{\alpha_p}(\phi_i) \Lambda(\phi_i) \exp\left(-\frac{1}{g^2} S[\phi_i]\right)\right] \\ &= \langle O_{\alpha_1} O_{\alpha_2} \cdots O_{\alpha_p}\rangle. \end{aligned} \tag{1.276}$$

Hence, observables can be computed either in the *weak coupling limit*, $g \to 0$, or in the *strong coupling limit*, $g \to \infty$.

So far we have presented a rather general definition of TQFT and made a series of elementary remarks. Now we will analyze some aspects of its structure. We begin pointing out that given a theory in which (1.271) holds one

can build *correlators* which correspond to *topological invariants* (in the sense that they are invariant under deformations of the metric $g_{\mu\nu}$) just by considering the operators of the theory which are *invariant under the symmetry*. We will call these operators *observables*. In virtue of (1.273), if one of these operators can be written as a symmetry transformation of another operator, its presence in a correlation function will make it vanish. Thus we may identify operators satisfying (1.271) which differ by an operator which corresponds to a symmetry transformation of another operator. Let us denote the set of the resulting classes by $\{\Phi\}$. By restricting the analysis to the appropriate set of operators, one has that in fact,

$$\delta^2 = 0. \tag{1.277}$$

Property (1.277) has consequences on the features of TQFT. First, the *symmetry must be odd* which implies the presence in the theory of *commuting* and *anticommuting fields*. For example, the tensor $G_{\mu\nu}$ in (1.271) must be anticommuting. This is the first appearance of an odd non–spinorial field in TQFT. Those kinds of objects are standard features of *cohomological* TQFTs. Second, if we denote by Q the operator which implements this symmetry, the observables of the theory can be described as the cohomology classes of Q:

$$\{\Phi\} = \frac{\operatorname{Ker} Q}{\operatorname{Im} Q}, \qquad Q^2 = 0. \tag{1.278}$$

Equation (1.271) means that in addition to the *Poincare group* the theory possesses a symmetry generated by an odd version of the Poincare group. The corresponding odd generators are constructed out of the tensor $G_{\mu\nu}$ in much the same way as the ordinary Poincare generators are built out of $T_{\mu\nu}$. For example, if P_μ represents the ordinary momentum operator, there exists a corresponding odd one G_μ such that

$$P_\mu = \{Q, G_\mu\}. \tag{1.279}$$

Now, let us discuss the structure of the Hilbert space of the theory in virtue of the symmetries that we have just described. The states of this space must correspond to representations of the algebra generated by the operators in the Poincare groups and by Q. Furthermore, as follows from our analysis of operators leading to (1.278), if one is interested only in states $|\Psi\rangle$ leading to topological invariants one must consider states which satisfy

$$Q|\Psi\rangle = 0, \tag{1.280}$$

and two states which differ by a $Q-$exact state must be identified. The odd Poincare group can be used to generate descendant states out of a state satisfying (1.280). The operators G_μ act non–trivially on the states and in fact, out of a state satisfying (1.280) we can build additional states using this generator. The simplest case consists of

$$\int_{\gamma_1} G_\mu |\Psi\rangle,$$

where γ_1 is a 1–cycle. One can verify using (1.271) that this new state satisfies (1.280):

$$Q \int_{\gamma_1} G_\mu |\Psi\rangle = \int_{\gamma_1} \{Q, G_\mu\} |\Psi\rangle = \int_{\gamma_1} P_\mu |\Psi\rangle = 0.$$

Similarly, one may construct other invariants tensoring n operators G_μ and integrating over $n-$cycles γ_n:

$$\int_{\gamma_n} G_{\mu_1} G_{\mu_2} ... G_{\mu_n} |\Psi\rangle. \tag{1.281}$$

Notice that since the operator G_μ is odd and its algebra is Poincare–like the integrand in this expression is an exterior differential $n-$form. These states also satisfy condition (1.280). Therefore, starting from a state $|\Psi\rangle \in \ker Q$ we have built a set of partners or descendants giving rise to a topological multiplet. The members of a multiplet have well defined *ghost* number. If one assigns ghost number -1 to the operator G_μ the state in (1.281) has ghost number $-n$ plus the ghost number of $|\Psi\rangle$. Now, n is bounded by the dimension of the manifold X. Among the states constructed in this way there may be many which are related via another state which is $Q-$exact, i.e., which can be written as Q acting on some other state. Let us try to single out representatives at each level of ghost number in a given topological multiplet.

Consider an $(n-1)-$cycle which is the boundary of an nD surface, $\gamma_{n-1} = \partial S_n$. If one builds a state taking such a cycle one finds ($P_\mu = -\mathrm{i}\partial_\mu$),

$$\begin{aligned} \int_{\gamma_{n-1}} G_{\mu_1} G_{\mu_2} ... G_{\mu_{n-1}} |\Psi\rangle &= \mathrm{i} \int_{S_n} P_{[\mu_1} G_{\mu_2} G_{\mu_3} ... G_{\mu_n]} |\Psi\rangle \\ &= \mathrm{i} Q \int_{S_n} G_{\mu_1} G_{\mu_2} ... G_{\mu_n} |\Psi\rangle, \end{aligned} \tag{1.282}$$

i.e., it is $Q-$exact. The square–bracketed subscripts in (1.282) denote that all indices between them must by antisymmetrized. In (1.282) use has been made of (1.279). This result tells us that the representatives we are looking for are built out of the homology cycles of the manifold X. Given a manifold X, the homology cycles are equivalence classes among cycles, the equivalence relation being that two $n-$cycles are equivalent if they differ by a cycle which is the boundary of an $n+1$ surface. Thus, knowledge on the homology of the manifold on which the TQFT is defined allows us to classify the representatives among the operators (1.281). Let us assume that X has dimension d and that its homology cycles are γ_{i_n}, $(i_n = 1, ..., d_n,\ n = 0, ..., d)$, where d_n is the dimension of the $n-$homology group, and d the dimension of X. Then, the non–trivial partners or descendants of a given $|\Psi\rangle$ *highest–ghost–number state* are labelled in the following way:

$$\int_{\gamma_{i_n}} G_{\mu_1} G_{\mu_2} ... G_{\mu_n} |\Psi\rangle, \qquad (i_n = 1, ..., d_n, \;\; n = 0, ..., d).$$

A similar construction to the one just described can be made for fields. Starting with a field $\phi(x)$ which satisfies,

$$[Q, \phi(x)] = 0, \tag{1.283}$$

one can construct other fields using the operators G_μ. These fields, which we call *partners* are antisymmetric tensors defined as,

$$\phi^{(n)}_{\mu_1 \mu_2 ... \mu_n}(x) = \frac{1}{n!} [G_{\mu_1}, [G_{\mu_2} ... [G_{\mu_n}, \phi(x)\}...\}\}, \qquad (n = 1, ..., d).$$

Using (1.279) and (1.283) one finds that these fields satisfy the so–called *topological descent equations*:

$$d\phi^{(n)} = \mathrm{i}[Q, \phi^{(n+1)}\},$$

where the subindices of the forms have been suppressed for simplicity, and the highest–ghost–number field $\phi(x)$ has been denoted as $\phi^{(0)}(x)$. These equations enclose all the relevant properties of the observables which are constructed out of them. They constitute a very useful tool to build the observables of the theory.

Stringy Actions and Amplitudes

Now we give a brief review of modern path–integral methods in (super)string theory (mainly following [DEF99]). Recall that the fundamental quantities in quantum field theory (QFT) are the transition amplitudes $Amp : IN \Longrightarrow OUT$, describing processes in which a number IN of incoming particles scatter to produce a number OUT of outgoing particles. The square modulus of the transition amplitude yields the *probability* for this process to take place.

Strings

Recall that in string theory, *elementary particles* are not described as 0–dimensional *points*, but instead as 1D *strings*. If M_s and $M(\sim \mathbb{R} \times M_s)$ denote the 3D space and 4D space–time manifolds respectively, then we picture strings as in Figure 1.19.

While the point–particle sweeps out a 1D *world–line*, the string sweeps out a *world–sheet*, i.e., a 2D real surface. For a *free string*, the topology of the world–sheet is a *cylinder* (in the case of a closed string) or a *sheet* (for an open string).

Roughly, different elementary particles correspond to different *vibration modes* of the string just as different minimal notes correspond to different vibrational modes of musical string instruments.

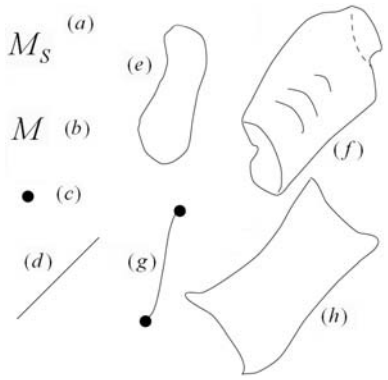


Fig. 1.19. Basic geometrical objects of string theory: (a) a space with fixed time; (b) a space–time picture; (c) a point–particle; (d) a world–line of a point–particle; (e) a closed string; (f) a world–sheet of a closed string; (g) an open string; (h) a world–sheet of an open string.

It turns out that the physical size of strings is set by *gravity*, more precisely the *Planck length* $\ell_P \sim 10^{-33}$ cm. This scale is so small that we effectively only see point–particles at our distance scales. Thus, for length scales much larger than ℓ_P, we expect to recover a QFT–description of point–particles, plus typical *string corrections* that represent physics at the Planck scale.

Interactions

While the string itself is an extended 1D object, the fundamental string interactions are *local*, just as for point–particles. The interaction takes place when strings overlap in space at the same time. In case of *closed string theories* the interactions have a form depicted in Figure 1.20, while in case of *open string theories* the interactions have a form depicted in Figure 1.21. Other interactions result from combining the interactions defined above.

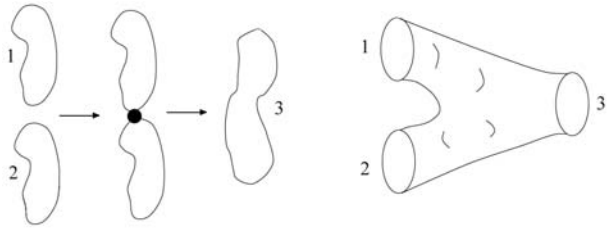


Fig. 1.20. Interactions in closed string theories (left 2D–picture and right 3D–picture).

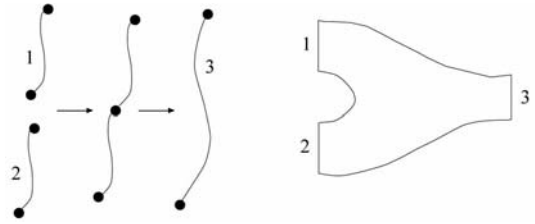


Fig. 1.21. Interactions in open string theories (left 2D–picture and right 3D–picture).

In point–particle theories, the fundamental interactions are read off from the QFT–Lagrangian. An interaction occurs at a *geometrical point*, where the world–lines join and cease to be a manifold. In *Lorentz–invariant theories* (where manifold M is a flat Minkowski space–time), the interaction point is Lorentz–invariant. To specify how the point–particles interact, additional data must be supplied at the interaction point, giving rise to *many possible* distinct quantum field theories.

In string theory, the interaction point *depends* upon the Lorentz frame chosen to observe the process. In the Figure above, equal time slices are indicated from the point of view of two different Lorentz frames, schematically indicated by t and t'. The closed string interaction, as seen from frames t and t', occurs at times t_2 and t'_2 and at (distinct) points P and P' respectively.

Lorentz invariance of interaction forbids that any point on the world–sheet be singled out as interaction point. Instead, the interaction results purely from the joining and splitting of strings. While free closed strings are characterized by their topology being that of a cylinder, interacting strings are characterized by the fact that their associated world–sheet is connected to at least 3 strings, incoming and/or outgoing.

As a result, the free string determines the nature of the interactions completely, leaving only the string coupling constant undetermined.

The *orientation* is an additional structure of closed strings, dividing them into two categories: (i) *oriented strings*, in which all world–sheets are assumed to be orientable; and (ii) non–oriented strings, in which world–sheets are non–orientable, such as the Möbius strip, Klein bottle, etc.

Loop Expansion – Topology of Closed Surfaces

For simplicity, here we consider closed oriented strings only, so that the associated world–sheet is also oriented. A general string configuration describing the process in which M incoming strings interact and produce N outgoing strings looks at the topological level like a closed surface with $M + N = E$ *boundary components* and any number of *handles* (see Figure 1.22). This picture is a kind of topological generalization of nonlinear control MIMO–systems with M inputs, N outputs X states (see subsection 2.3.1 below).

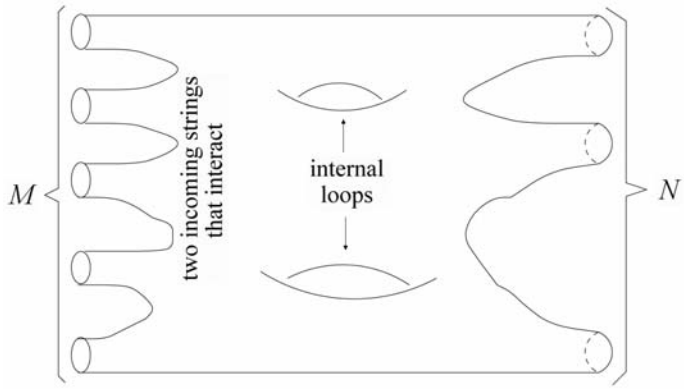


Fig. 1.22. Boundary components and handles of closed oriented system of M incoming strings, interacting through internal loops, to produce N outgoing strings. Note the striking similarity with MIMO–systems of *nonlinear control theory*, with M input processes and N output processes(see subsection 2.3.1 below).

The internal loops may arise when virtual particle pairs are produced, just as in quantum field theory. For example, a *Feynman diagram* in quantum field theory that involves a loop is shown in Figure 1.23 together with the corresponding string diagram.

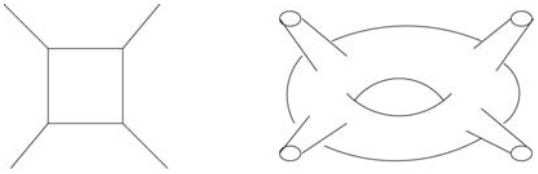


Fig. 1.23. A QFT Feynman diagram that involves an internal loop (left), with the corresponding string diagram (right).

Surfaces associated with closed oriented strings have two topological invariants: (i) the *number of boundary components* $E = M + N$ (which may be shrunk to punctures, under certain conditions), and (ii) the number h of *handles* on the surface, which equals the surface *genus*.

When $E = 0$, we just have the topological classification of compact oriented surfaces without boundary. Rendering $E > 0$ is achieved by removing E discs from the surface.

Recall that in QFT, an expansion in powers of Planck's constant $\hbar$ yields an expansion in the number of loops of the associated Feynman diagram, for a given number of external states:

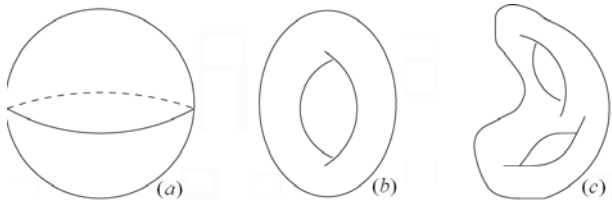


Fig. 1.24. Number h of handles on the surface of closed oriented strings, which equals the string–surface genus: (a) $h = 0$ for sphere S^2; (b) $h = 1$ for torus T^2; (c) $h = 2$ for string–surfaces with higher genus, etc.

$$\hbar^{E+h-1} = \begin{cases} \hbar & \text{for every propagator} \\ \hbar^{-1} & \text{for every vertex} \\ -1 & \text{for overall momentum conservation} \end{cases}$$

Thus, in string theory we expect that, for a given number of external strings E, the topological expansion genus by genus should correspond to a *loop* expansion as well.

Recall that in QFT, there are in general many Feynman diagrams that correspond to an amplitude with a given number of external particles and a given number of loops. For example, for $E = 4$ external particles and $h = 1$ loop in ϕ^3 theory are given in Figure 1.25, together with the same process in string theory (for closed oriented strings), where it is described by just a single diagram (right).

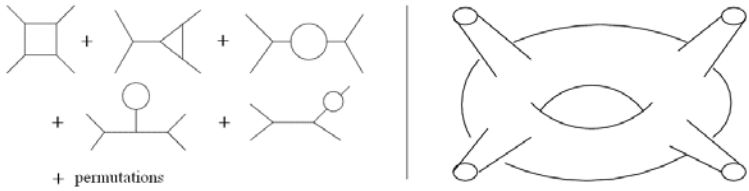


Fig. 1.25. Feynman QFT–diagrams for ϕ^3 theory with $E = 4$ external particles and $h = 1$ loop (left), and a single corresponding string diagram (right). In this way the usual Feynman diagrams of quantum field theory are generalized by arbitrary Riemannian surfaces.

Much of recent interest has been focused on the so–called $D-$branes. A $D-$*brane* is a submanifold of space–time with the property that strings can *end* or *begin* on it.

Transition Amplitudes for Strings

The only way we have today to define string theory is by giving a *rule* for the evaluation of transition amplitudes, order by order in the loop expansion, i.e.,

genus by genus. The rule is to assign a relative weight to a given configuration and then to sum over all configurations [DEF99]. To make this more precise, we first describe the system's configuration manifold M (see Figure 1.26).

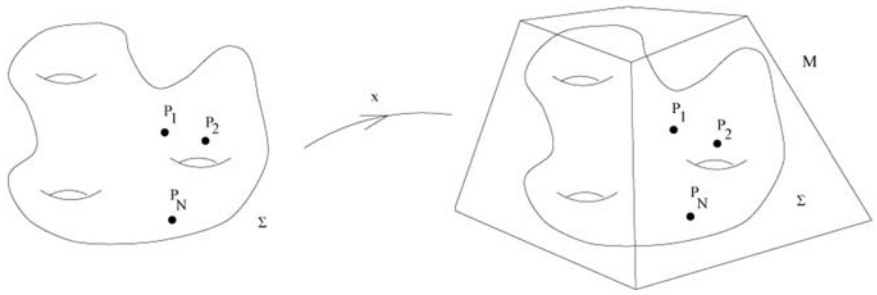


Fig. 1.26. The embedding map x from the reference surface Σ into the pseudo–Riemannian configuration manifold M (adapted from [DEF99]).

We assume that Σ and M are smooth manifolds, of dimensions 2 and n respectively, and that x is a continuous map from Σ to M. If ξ^m, (for $m = 1, 2$), are local coordinates on Σ and x^μ, $(\mu = 1, \ldots, n)$, are local coordinates on M then the map x may be described by functions $x^\mu(\xi^m)$ which are continuous.

To each system configuration we can associate a weight $\mathrm{e}^{-S[x,\Sigma,M]}$, (for $S \in \mathbb{C}$) and the transition amplitude Amp for specified external strings (incoming and outgoing) is get by summing over all surfaces Σ and all possible maps x,

$$Amp = \sum_{surfaces\ \Sigma} \sum_{x} \mathrm{e}^{-S[x,\Sigma,M]}\,.$$

We now need to specify each of these ingredients:

(1) We assume M to be an nD Riemannian manifold, with metric g. A special case is flat Euclidean space–time $\mathbb{R}^n$. The space–time metric is assumed *fixed*.

$$ds^2 = (dx, dx)_g = g_{\mu\nu}(x) dx^\mu \otimes dx^\nu.$$

(2) The metric g on M induces a metric on Σ: $\gamma = x^*(g)$,

$$\gamma = \gamma_{mn} d\xi^m \otimes d\xi^n, \qquad \gamma_{mn} = g_{\mu\nu} \frac{\partial x^\mu}{\partial \xi^m} \frac{\partial x^\nu}{\partial \xi^n}.$$

This metric is non–negative, but depends upon x. It is advantageous to introduce an intrinsic Riemannian metric g on Σ, independently of x; in local coordinates, we have

$$g = g_{mn}(\xi) d\xi^m \otimes d\xi^n.$$

A natural intrinsic candidate for S is the area of $x(\Sigma)$, which gives the so–called *Nambu–Goto action*[29]

$$Area\,(x\,(\Sigma)) = \int_{\Sigma} d\mu_{\gamma} = \int_{\Sigma} n^2 \xi \sqrt{\det\,\gamma_{mn}}, \tag{1.284}$$

which depends only upon g and x, but not on g. However, the transition amplitudes derived from the Nambu–Goto action are *not* well–defined quantum–mechanically.

Otherwise, we can take as starting point the so–called *Polyakov action*[30]

$$S[x,g] = \kappa \int_{\Sigma} (dx, *dx)_g = \kappa \int_{\Sigma} d\mu_g g^{mn} \partial_m x^{\mu} \partial_n x^{\nu} g_{\mu\nu}(x), \tag{1.285}$$

where κ is the *string tension* (a positive constant with dimension of inverse length square). The stationary points of S with respect to g are at $g^0 = \mathrm{e}^{\phi}\gamma$ for some function ϕ on Σ, and thus $S[x,g^0] \sim Area\ (x\,(\Sigma))$.

The Polyakov action leads to *well–defined* transition amplitudes, get by integration over the space $Met(\Sigma)$ of all positive metrics on Σ for a given topology, as well as over the space of all maps $Map(\Sigma, M)$. We can define the path integral

$$Amp = \sum_{\substack{topologies \\ \Sigma}} \int_{Met(\Sigma)} \frac{1}{N(g)} \int_{Map(\Sigma,M)} \mathcal{D}[x]\, \mathrm{e}^{-S[x,g,g]},$$

where N is a normalization factor, while the measures $\mathcal{D}[g]$ and $\mathcal{D}[x]$ are constructed from $Diff^{+}(\Sigma)$ and $Diff(M)$ invariant L^2 norms on Σ and M. For fixed metric g, the action S is well–known: its stationary points are the harmonic maps $x\colon\ \Sigma \to M$ (see, e.g., [EL78]). However, g here varies and in fact is to be integrated over. For a general metric g, the action S defines a *nonlinear sigma model*, which is renormalizable because the dimension of Σ is 2. It would not in general be renormalizable in dimension higher than 2, which is usually regarded as an argument against the existence of fundamental membrane theories (see [DEF99]).

The Nambu–Goto action (1.284) and Polyakov action (1.285) represent the core of the so–called *bosonic string theory*, the original version of string theory, developed in the late 1960s. Although it has many attractive features, it also predicts a particle called the *tachyon* possessing some unsettling properties, and it has no fermions. All of its particles are *bosons*, the matter particles.

[29] Nambu–Goto action is the starting point of the analysis of string behavior, using the principles of ordinary Lagrangian mechanics. Just as the Lagrangian for a free point particle is proportional to its proper timei.e., the 'length' of its world–line, a relativistic string's Lagrangian is proportional to the area of the sheet which the string traces as it travels through space–time.

[30] The Polyakov action is the 2D action from *conformal field theory*, used in string theory to describe the world–sheet of a moving string.

The physicists have also calculated that bosonic string theory requires 26 space–time dimensions: 25 spatial dimensions and one dimension of time. In the early 1970s, *supersymmetry* was discovered in the context of string theory, and a new version of string theory called *superstring theory* (i.e., supersymmetric string theory) became the real focus, as it includes also *fermions*, the force particles. Nevertheless, bosonic string theory remains a very useful 'toy model' to understand many general features of *perturbative string theory* (see subsection 2.2.8 below).

Weyl Invariance and Vertex Operator Formulation

The action S is also invariant under *Weyl rescalings* of the metric g by a positive function on $\sigma\colon\ \Sigma \to \mathbb{R}$, given by $g \to \mathrm{e}^{2\sigma} g$. In general, Weyl invariance of the full amplitude may be spoiled by anomalies. Assuming Weyl invariance of the full amplitude, the integral defining *Amp* may be *simplified* in two ways.

1) The integration over $Met(\Sigma)$ effectively collapses to an integration over the *moduli space of surfaces*, which is finite dimensional, for each genus h.

2) The boundary components of Σ — characterizing external string states — may be mapped to regular points on an underlying compact surface without boundary by conformal transformations. The *data*, such as momenta and other quantum numbers of the external states, are mapped into *vertex operators*. The amplitudes are now given by the path integral

$$Amp = \sum_{h=0}^{\infty} \int_{Met(\Sigma)} \mathcal{D}[g] \frac{1}{N(g)} \int_{Map(\Sigma,M)} \mathcal{D}[x]\, V_1 \dots V_N\, \mathrm{e}^{-S},$$

for suitable vertex operators $V_1, \dots V_N$.

More General Actions

Generalizations of the action S given above are possible when M carries extra structure.

1) M carries a $2-$form $B \in \Omega^{(2)}(M)$. The resulting contribution to the action is also that of a 'nonlinear sigma model'

$$S_B[x, B] = \int_{\Sigma} x^*(B) = \int_{\Sigma} dx^{\mu} \wedge dx^{\nu} B_{\mu\nu}(x)$$

2) M may carry a *dilaton field* $\Phi \in \Omega^{(0)}(M)$ so that

$$S_{\Phi}[x, \Phi] = \int_{\Sigma} d\mu_g R_g \Phi(x).$$

where R_g is the Gaussian curvature of Σ for the metric g.

3) There may be a *tachyon field* $T \in \Omega^{(0)}(M)$ contributing

$$S_T[x, T] = \int_{\Sigma} d\mu_g T(x).$$

Transition Amplitude for a Single Point Particle

The transition amplitude for a single point–particle could in fact be get in a way analogous to how we prescribed string amplitudes. Let space–time be again a Riemannian manifold M, with metric g. The prescription for the transition amplitude of a particle travelling from a point $y \in M$ to a point y' to M is expressible in terms of a sum over all (continuous) paths connecting y to y':

$$Amp(y, y') = \sum_{\substack{paths \\ joining\ y\ and\ y'}} \mathrm{e}^{-S[path]}.$$

Paths may be parametrized by maps from $C = [0, 1]$ into M with $x(0) = y$, $x(1) = y'$. A simple world–line action for a massless particle is get by introducing a metric g on $[0, 1]$

$$S[x, g] = \frac{1}{2} \int_C d\tau\, g(\tau)^{-1} \dot{x}^\mu \dot{x}^\nu g_{\mu\nu}(x),$$

which is invariant under $Diff^+(C)$ and $Diff(M)$.

Recall that the analogous prescription for the point–particle transition amplitude is the path integral

$$Amp(y, y') = \int\limits_{Met(C)} \mathcal{D}[g] \frac{1}{N} \int\limits_{Map(C,M)} \mathcal{D}[x]\, \mathrm{e}^{-S[x,g]}.$$

Note that for string theory, we had a prescription for transition amplitudes valid for all topologies of the world–sheet. For point–particles, there is only the topology of the interval C, and we can only describe a single point–particle, but not interactions with other point–particles. To put those in, we would have to supply additional information.

Finally, it is very instructive to work out the amplitude Amp by carrying out the integrations. The only $Diff^+(C)$ invariant of g is the length $L = \int_0^1 d\tau\, g(\tau)$; all else is generated by $Diff^+(C)$. Defining the normalization factor to be the volume of $Diff(C)$: $N = Vol(Diff(C))$ we have $\mathcal{D}[g] = \mathcal{D}[v]\, dL$ and the transition amplitude becomes

$$Amp(y, y') = \int_0^\infty dL \int \mathcal{D}[x]\, \mathrm{e}^{-\frac{1}{2L}\int_0^1 d\tau (\dot{x}, \dot{x})_g} = \int_0^\infty dL\, \langle y' | \mathrm{e}^{-L\Delta} | y \rangle = \left\langle y' \middle| \frac{1}{\Delta} \middle| y \right\rangle.$$

Thus, the amplitude is just the Green function at (y, y') for the Laplacian Δ and corresponds to the propagation of a massless particle (see [DEF99]).

2

Dynamics of Complex Systems

2.1 Mechanical Systems

2.1.1 Autonomous Lagrangian/Hamiltonian Mechanics

Basis of Lagrangian Dynamics

Recall that Riemannian metric $g = <,>$ on the configuration manifold M is a positive–definite quadratic form $g : TM \to \mathbb{R}$, given in local coordinates $q^i \in U$ (U open in M) as

$$g_{ij} \mapsto g_{ij}(q,m)\, dq^i dq^j, \qquad \text{where} \tag{2.1}$$

$$g_{ij}(q,m) \;=\; m_\mu \delta_{rs} \frac{\partial x^r}{\partial q^i} \frac{\partial x^s}{\partial q^j} \tag{2.2}$$

is the covariant material metric tensor defining a relation between internal and external coordinates and including n segmental masses m_μ. The quantities x^r are external coordinates ($r, s = 1, \ldots, 6n$) and $i, j = 1, \ldots, N \equiv 6n - h$, where h denotes the number of holonomic constraints.

The *Lagrangian* of the system is a quadratic form $L : TM \to \mathbb{R}$ dependent on velocity v and such that $L(v) = \frac{1}{2} < v, v >$. It is locally given by

$$L(v) \;=\; \frac{1}{2} g_{ij}(q,m)\, v^i v^j.$$

On the velocity phase–space manifold TM exist:

1. a unique 1−form θ_L, defined in local coordinates q^i, $v^i = \dot{q}^i \in U_v$ (U_v open in TM) by $\theta_L = L_{v^i} dq^i$, where $L_{v^i} \equiv \partial L / \partial v^i$; and
2. a unique nondegenerate Lagrangian symplectic 2−form ω_L, which is closed ($d\omega_L = 0$) and exact ($\omega_L = d\theta_L = dL_{v^i} \wedge dq^i$).

TM is an orientable manifold, admitting the standard volume given by

$$\Omega_{\omega_L} = \frac{(-1)^{\frac{N(N+1)}{2}}}{N!}\omega_L^N,$$

in local coordinates q^i, $v^i = \dot{q}^i \in U_v$ (U_v open in TM) it is given by

$$\Omega_L = dq^1 \wedge \cdots \wedge dq^N \wedge dv^1 \wedge \cdots \wedge dv^N.$$

On the velocity phase–space manifold TM we can also define the *action* $A : TM \to \mathbb{R}$ in local coordinates q^i, $v^i = \dot{q}^i \in U_v$ (U_v open in TM) given by $A = v^i L_{v^i}$, so $E = v^i L_{v^i} - L$. The Lagrangian vector–field X_L on TM is determined by the condition $i_{X_L}\omega_L = dE$. Classically, it is given by the second–order Lagrangian equations

$$\frac{d}{dt}\frac{\partial L}{\partial v^i} = \frac{\partial L}{\partial q^i}. \tag{2.3}$$

For a Lagrangian vector–field X_L on M, there is a base integral curve $\gamma_0(t) = (q^i(t), v^i(t))$ iff $\gamma_0(t)$ is a geodesic. This is given by the contravariant velocity equation

$$\dot{q}^i = v^i, \qquad \dot{v}^i + \Gamma^i_{jk}\, v^j v^k = 0. \tag{2.4}$$

Here Γ^i_{jk} denote the *Christoffel symbols* of the Levi–Civita connection ∇ in an open chart U on M, defined on the Riemannian metric $g = <,>$ by (see Appendix, as well as section 1.2.9 above)

$$\Gamma^i_{jk} = g^{il}\Gamma_{jkl}, \qquad \Gamma_{ijk} = \frac{1}{2}(\partial_{x^i} g_{jk} + \partial_{x^j} g_{ki} + \partial_{x^k} g_{ij}). \tag{2.5}$$

The l.h.s $\dot{\bar{v}}^i = \dot{v}^i + \Gamma^i_{jk}\, v^j v^k$ in the second part of (2.4) represents the *Bianchi covariant derivative* of the velocity with respect to t. *Parallel transport* on M is defined by $\dot{\bar{v}}^i = 0$. When this applies, X_L is called the *geodesic spray* and its flow the *geodesic flow*.

For the dynamics in the gravitational potential field $V : M \to \mathbb{R}$, the Lagrangian $L : TM \to \mathbb{R}$ has an extended form

$$L(v, q) = \frac{1}{2} g_{ij} v^i v^j - V(q),$$

A Lagrangian vector–field X_L is still defined by the second–order Lagrangian equations (2.3, 2.4).

A general form of the forced, non–conservative Lagrangian equations is given as

$$\frac{d}{dt}\frac{\partial L}{\partial v^i} - \frac{\partial L}{\partial q^i} = F_i(t, q^i, v^i)).$$

Here the $F_i(t, q^i, v^i)$ represent any kind of *covariant forces* as a functions of time, coordinates and momenta. In covariant form we have

$$\dot{q}^i = v^i, \qquad g_{ij}(\dot{v}^i + \Gamma^i_{jk}\, v^j v^k) = F_j(t, q^i, v^i)).$$

Basics of Poincaré Dynamics

The Euler–Poincaré Equations.

Let G be a Lie group and let $L : TG \to \mathbb{R}$ be a *left–invariant Lagrangian*. Let $l : \mathfrak{g} \to \mathbb{R}$ be its restriction to the identity. For a curve $g(t) \in G$, let $\xi(t) = g(t)^{-1} \cdot \dot{g}(t)$; that is, $\xi(t) = T_{g(t)} L_{g(t)^{-1}} \dot{g}(t)$. Then the following are equivalent [MR99]:

1. $g(t)$ satisfies the Euler–Lagrangian equations for L on G;
2. The variational principle holds,

$$\delta \int L(g(t),\, \dot{g}(t))\, dt = 0$$

   for variations with fixed endpoints;
3. The *Euler–Poincaré equations* hold:

$$\frac{d}{dt}\frac{\partial l}{\partial \xi} = Ad^*_\xi \frac{\delta l}{\delta \xi};$$

4. The variational principle holds on $\mathfrak{g}$,

$$\delta \int l(\xi(t))\, dt = 0,$$

   using variations of the form $\delta\xi = \dot{\eta} + [\xi, \eta]$, where η vanishes at the endpoints.

The Lagrangian–Poincaré Equations.

Here we follow [MR99] and drop Euler–Lagrangian equations and variational principles from a general velocity phase–space TM to the quotient TM/G by an action of a Lie group G on M. If L is a $G-$invariant Lagrangian on TM, it induces a reduced Lagrangian l on TM/G. We introduce a connection A on the principal bundle $M \to S = M/G$, assuming that this quotient is nonsingular. This connection allows one to split the variables into a horizontal and vertical part. Let internal variables x^α be coordinates for shape–space $S = M/G$, let η^a be coordinates for the Lie algebra $\mathfrak{g}$ relative to a chosen basis, let l be the Lagrangian regarded as a function of the variables $x^\alpha, \dot{x}^\alpha, \eta^a$ and let C^a_{db} be the structure constants of the Lie algebra $\mathfrak{g}$ of G.

If one writes the Euler–Lagrangian equations on TM in a local principal bundle trivialization, with coordinates x^α on the base and η^a in the fibre, then one gets the following system of *Hamel equations*:

$$\frac{d}{dt}\frac{\partial l}{\partial \dot{x}^\alpha} = \frac{\partial l}{\partial x^\alpha}, \qquad \text{and} \qquad \frac{d}{dt}\frac{\partial l}{\partial \eta^b} = \frac{\partial l}{\partial \eta^a} C^a_{db} \eta^a.$$

However, this representation of the equations does not make global intrinsic sense. The introduction of a connection overcomes this, and one can

intrinsically and globally split the original variational principle relative to horizontal and vertical variations. One gets from one form to the other by means of the velocity shift given by replacing η^a by the vertical part relative to the *affine connection*

$$\xi^a = A^a_\alpha \dot{x}^\alpha + \eta^a.$$

Here A^a_α are the local coordinates of the connection A. This change of coordinates is motivated from the mechanical point of view, since the variables ξ^a have the interpretation of the locked angular velocity. The resulting *Lagrangian–Poincaré equations* have the following form:

$$\begin{aligned}\frac{d}{dt}\frac{\partial l}{\partial \dot{x}^\alpha} - \frac{\partial l}{\partial x^\alpha} &= \frac{\partial l}{\partial \xi^a}\left(B^a_{\alpha\beta}\dot{x}^\beta + B^a_{\alpha d}\xi^d\right),\\ \frac{d}{dt}\frac{\partial l}{\partial \xi^b} &= \frac{\partial l}{\partial \xi^a}\left(B^a_{b\alpha}\dot{x}^\alpha + C^a_{db}\xi^d\right).\end{aligned}$$

In these equations, $B^a_{\alpha\beta}$ are the coordinates of the curvature B of A,

$$B^a_{d\alpha} = C^a_{db}A^b_\alpha, \qquad \text{and} \qquad B^a_{b\alpha} = -B^a_{\alpha b}.$$

The variables ξ^a may be regarded as the rigid part of the variables on the original configuration space, while x^α are the internal variables.

Basics of Hamiltonian Mechanics

In this section we present classical Hamiltonian dynamics. Let (M, ω) be a *symplectic manifold* and $H \in C^\infty(M, \mathbb{R})$ a smooth real valued function on M. The vector–field X_H determined by the condition

$$i_{X_H}\omega + dH = 0,$$

is called *Hamiltonian vector–field* with *Hamiltonian energy function* H. A triple (M, ω, H) is called a *Hamiltonian mechanical system* [MR99, Put93].

Nondegeneracy of ω guarantees that X_H exists, but only in the nD case.

Let $\{q^1, ..., q^n, p_1, ..., p_n\}$ be *canonical coordinates* on M, i.e., $\omega = dp_i \wedge dq^i$. Then in these coordinates we have

$$X_H = \left(\frac{\partial H}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i}\frac{\partial}{\partial p_i}\right).$$

As a consequence, $\big((q^i(t)), (p_i(t))\big)$ is an integral curve of X_H (for $i = 1, ..., n$) iff *Hamiltonian equations* hold,

$$\dot{q}^i = \partial_{p_i} H, \qquad \dot{p}_i = -\partial_{q^i} H. \tag{2.6}$$

Let (M, ω, H) be a Hamiltonian mechanical system and let $\gamma(t)$ be an integral curve of X_H. Then $H(\gamma(t))$ is constant in t. Moreover, if ϕ_t is the flow of X_H, then $H \circ \phi_t = H$ for each t.

Let (M, ω, H) be a Hamiltonian mechanical system and ϕ_t be the flow of X_H. Then, by the Liouville theorem, for each t, $\phi_t^* \omega = \omega$, ($\frac{d}{dt}\phi_t^* \omega = 0$, so $\phi_t^* \omega$ is constant in t), that is, ϕ_t is symplectic, and it preserves the volume Ω_ω.

A convenient criterion for symplectomorphisms is that they preserve the form of Hamiltonian equations. More precisely, let (M, ω) be a symplectic manifold and $f : M \to M$ a diffeomorphism. Then f is symplectic iff for all $H \in C^\infty(M, \mathbb{R})$ we have $f^*(X_H) = X_{H \circ f}$.

A vector–field $X \in \mathcal{X}(M)$ on a symplectic manifold (M, ω) is called locally Hamiltonian iff $\mathcal{L}_X \omega = 0$, where L denotes the *Lie derivative*. From the equality $L_{[X,Y]}\omega = \mathcal{L}_X L_Y \omega - L_Y \mathcal{L}_X \omega$, it follows that the locally Hamiltonian vector–fields on M form a Lie subalgebra of $\mathcal{X}(M)$.

Let (M, ω) be a symplectic manifold and $f, g \in C^\infty(M, \mathbb{R})$. The *Poisson bracket* of f and g is the function

$$\{f, g\}_\omega = -\omega(X_f, X_g) = -\mathcal{L}_{X_f} g = \mathcal{L}_{X_g} f.$$

Also, for $f_0 \in C^\infty(M, \mathbb{R})$, the map $g \longmapsto \{f_0, g\}_\omega$ is a derivation. The connection between the Lie bracket and the Poisson bracket is

$$[X_f, X_g] = -X_{\{f,g\}_\omega} \qquad \Longleftrightarrow \qquad d\omega = 0.$$

The real vector space $C^\infty(M, \mathbb{R})$ together with the Poisson bracket on it forms an infinite–dimensional Lie algebra called the *algebra of classical observables*.

In canonical coordinates $\{q^1, ..., q^n, p_1, ..., p_n\}$ on (M, ω) the Poisson bracket of two functions $f, g \in C^\infty(M, \mathbb{R})$ is given by

$$\{f, g\}_\omega = \frac{\partial f}{\partial q^i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q^i}.$$

From this definition follows:

$$\{q^i, q^j\}_\omega = 0, \qquad \{p_i, p_j\}_\omega = 0, \qquad \{q^i, p_j\}_\omega = \delta^i_j.$$

Let (M, ω) be a symplectic manifold and $f : M \to M$ a diffeomorphism. Then f is symplectic iff it preserves the Poisson bracket.

Let (M, ω, H) be a Hamiltonian mechanical system and ϕ_t the flow of X_H. Then for each function $f \in C^\infty(M, \mathbb{R})$ we have the equations of motion in the Poisson bracket notation:

$$\frac{d}{dt}(f \circ \phi_t) = \{f \circ \phi_t, H\}_\omega = \{f, H\}_\omega \circ \phi_t.$$

Also, f is called a *constant of motion*, or a *first integral*, if it satisfies the following condition

$$\{f, H\}_\omega = 0.$$

If f and g are constants of motion then their Poisson bracket is also a constant of motion.

A Hamiltonian mechanical system (M, ω, H) is said to be integrable if there exists $n = \frac{1}{2}\dim(M)$ linearly–independent functions $K_1 = H, K_2, ..., K_n$ such that for each $i, j = 1, 2, ..., n$:

$$\{K_i, H\}_\omega = 0, \qquad \{K_i, K_j\}_\omega = 0.$$

Real 1*–DOF Hamiltonian Dynamics.*

A vector–field $X(t)$ on the *momentum phase–space manifold* M can be given by a system of *canonical equations of motion*

$$\dot{q} = f(q, p, t, \mu), \qquad \dot{p} = g(q, p, t, \mu), \tag{2.7}$$

where t is time, μ is a parameter, $q \in S^1$, $p \in \mathbb{R}\times S^1$ are *coordinates* and *momenta*, respectively, while f and g are smooth functions on the phase–space $\mathbb{R}\times S^1$.

If time t does not explicitly appear in the functions f and g, the vector–field X is called *autonomous*. In this case equation (2.7) simplifies as

$$\dot{q} = f(q, p, \mu), \qquad \dot{p} = g(q, p, \mu). \tag{2.8}$$

By a *solution curve* of the vector–field X we mean a map $x = (q, p)$, from some interval $I \subset \mathbb{R}$ into the phase–space manifold M, such that $t \mapsto x(t)$. The map $x(t) = (q(t), p(t))$ geometrically represents a curve in M, and equations (2.7) or (2.8) give the tangent vector at each point of the curve.

To specify an *initial condition* on the vector–field X, by

$$x(t, t_0, x_0) = (q(t, t_0, q_0),\, p(t, t_0, p_0)),$$

geometrically means to distinguish a solution curve by a particular point $x(t_0) = x_0$ in the phase–space manifold M. Similarly, it may be useful to explicitly display the parametric dependence of solution curves, as $x(t, t_0, x_0, \mu) = (q(t, t_0, q_0, \mu_q),\, p(t, t_0, p_0, \mu_p))$, where μ_q, μ_p denote $q-$dependent and $p-$dependent parameters, respectively.

The solution curve $x(t, t_0, x_0)$ of the vector–field X, may be also referred as the *phase trajectory* through the point x_0 at $t = t_0$. Its graph over t is refereed to as an *integral curve*; more precisely, *graph* $x(t, t_0, x_0) \equiv \{(x, t) \in M \times \mathbb{R} : x = x(t, t_0, x_0),\, t \in I \subset \mathbb{R}\}$.

Let $x_0 = (q_0, p_0)$ be a point on M. By the *orbit through* x_0, denoted $O(x_0)$, we mean the set of points in M that lie on a trajectory passing through x_0; more precisely, for $x_0 \in U$, U open in M, the orbit through x_0 is given by $O(x_0) = \{x \in \mathbb{R}\times S^1 : x = x(t, t_0, x_0),\, t \in I \subset \mathbb{R}\}$.

Consider a general autonomous vector–field X on the phase–space manifold M, given by equation $\dot{x} = f(x)$, $x = (q, p) \in M$. An *equilibrium solution*, *singularity*, or *fixed point* of X is a point $\bar{x} \in M$ such that $f(\bar{x}) = 0$, i.e., a solution which does not change in time.

Any solution $\bar{x}(t)$ of an autonomous vector–field X on M is *stable* if solutions starting 'close' to $\bar{x}(t)$ at a given time remain close to $\bar{x}(t)$ for all later times. It is *asymptotically stable* if nearby solutions actually converge to $\bar{x}(t)$ as $t \longrightarrow \infty$. In order to determine the stability of $\bar{x}(t)$ we must understand the nature of solutions near $\bar{x}(t)$, which is done by *linearization* of the vector–field X. The solution of the linearized vector–field Y is asymptotically stable if all eigenvalues have negative real parts. In that case the fixed point $x = \bar{x}$ of associated nonlinear vector–field X is also asymptotically stable. A fixed point $\bar{x}$ is called *hyperbolic point* if none of the eigenvalues of Y have zero real part; in that case the orbit structure near $\bar{x}$ is essentially the same for X and Y.

In the case of autonomous vector–fields on M we have also an important property of *Hamiltonian flow*. If $x(t) = (q(t), p(t))$ is a solution of $\dot{x} = f(x)$, $x \in M$, then so is $x(t+\tau)$ for any $\tau \in \mathbb{R}$. Also, for any $x_0 \in M$ there exists only one solution of an autonomous vector–field passing through this point. The autonomous vector–field

$$\dot{x} = f(x)$$

has the following properties (compare with the section (1.2.5) above):

1. $x(t, x_0)$ is C^∞;
2. $x(0, x_0) = x_0$; and
3. $x(t+s, x_0) = x(t, x(s, x_0))$.

These properties show that the solutions of an autonomous vector–field form a *one–parameter group of diffeomorphisms* of the phase–space manifold M. This is refereed to as a *phase–flow* and denoted by $\phi_t(x)$ or $\phi(t, x)$.

Consider a flow $\phi(t, x)$ generated by vector–field $\dot{x} = f(x)$. A point $x_0 = (q_0, p_0)$ on M is called an $\omega-$*limit point* of $x, = (q, p) \in M$, denoted $\omega(x)$, if there exists a sequence $\{t_i\}$, $t_i \mapsto \infty$, such that $\phi(t_i, x) \mapsto x_0$. Similarly, $\alpha-$*limit points* are defined by taking a sequence $\{t_i\}$, $t_i \mapsto -\infty$. The set of all $\omega-$limit points of a flow is called the $\omega-$*limit set*. The $\alpha-$limit set is similarly defined.

A point $x_0 = (q_0, p_0)$ on M is called *nonwandering* if for any open neighborhood $U \subset M$ of x_0, there exists some $t \neq 0$ such that $\phi(t, U) \cap U \neq 0$. The set of all nonwandering points of a flow is called the *nonwandering set* of that particular map or flow.

A closed invariant subset $A \subset M$ is called an *attracting set* if there is some open neighborhood $U \subset M$ of A such that $\phi(t, x) \in U$ and $\phi(t, x) \mapsto \infty$ for any $x \in U$ and $t \geq 0$. The *domain* or *basin of attraction* of A is given by $\cup_{t \leq 0} \phi(t, U)$. In practice, a way of locating attracting sets is to first find a *trapping region*, i.e., a closed, connected subset $V \subset M$ such that for any $t \geq 0$ $\phi(t, V) \subset V$. Then $\cap_{t \rangle 0} \phi(t, V) = A$ is an *attracting set*.

As a first example of one–DOF dynamical systems, let us consider a vector–field $x = (q, p) \in \mathbb{R} \times \mathbb{R}$ of a simple harmonic oscillator, given by equations

$$\dot{q} = p, \qquad \dot{p} = -q. \tag{2.9}$$

Here, the *solution* passing through the point $(q,p) = (1,0)$ at $t = 0$ is given by $(q(t),p(t)) = (\cos t, -\sin t)$; the *integral curve* passing through $(q,p) = (1,0)$ at $t = 0$ is given by $\{(q,p,t) \in \mathbb{R}\times\mathbb{R}\times\mathbb{R} : (q(t),p(t)) = (\cos t, -\sin t)\}$, for all $t \in \mathbb{R}$; the *orbit* passing through $(q,p) = (1,0)$ is given by the circle $q^2 + p^2 = 1$.

A one–DOF dynamical system is called *Hamiltonian system* if there exists a *first integral* or a function of the dependent variables (q,p) whose level curves give the orbits of the vector–field $X = X_H$, i.e., a total–energy *Hamiltonian function* $H = H(q,p) : U \to \mathbb{R}$, ($U$ open set on the phase–space manifold M), such that the vector–field X_H is given by *Hamiltonian canonical equations* (2.6). In (2.6), the first, $\dot{q}-$equation, is called the *velocity equation* and serves as a definition of the *momentum*, while the second, $\dot{p}-$equation is called the *force equation*, and represents the *Newtonian second law of motion.*

The simple harmonic oscillator (2.9) is a Hamiltonian system with a Hamiltonian function $H = \frac{p^2}{2} + \frac{q^2}{2}$. It has a *fixed point – center* (having purely imaginary eigenvalues) at $(q,p) = (0,0)$ and is surrounded by a one–parameter family of periodic orbits given by the Hamiltonian H.

A nice example of one–DOF dynamical system with a Hamiltonian structure is a *damped Duffing oscillator* (see, e.g., [Wig90]). This is a *plane* Hamiltonian vector–field $x = (q,p) \in \mathbb{R}^2$, given by Hamiltonian equations

$$\dot{q} = p \equiv f(q,p), \qquad \dot{p} = q - q^3 - \delta p \equiv g(q,p,\delta), \qquad \delta \geq 0. \tag{2.10}$$

For the special parameter value $\delta = 0$, we have an *undamped* Duffing oscillator with a *first integral* represented by Hamiltonian function $H = \frac{p^2}{2} - \frac{q^2}{2} + \frac{q^4}{4}$, where $\frac{p^2}{2}$ corresponds to the *kinetic energy* (with a mass scaled to unity), and $-\frac{q^2}{2} + \frac{q^4}{4} \equiv V(x)$ corresponds to the *potential energy* of the oscillator.

In general, if the first integral, i.e., a Hamiltonian function H, is defined by $H = \frac{p^2}{2} + V(x)$, then the momentum is given by $p = \pm\sqrt{2}\sqrt{H - V(x)}$. All one–DOF Hamiltonian systems are *integrable* and all the solutions lie on *level curves* of the Hamiltonian function, which are topologically equivalent with the circle S^1. This is actually a general characteristic of all $n-$DOF integrable Hamiltonian systems: their bounded motions lie on nD *invariant tori* $T^n = S^1 \times \cdots \times S^1$, or *homoclinic orbits*. The homoclinic orbit is sometimes called a *separatrix* because it is the boundary between two distinctly different types of motion.

For example, in case of a damped Duffing oscillator (2.10) with $\delta \neq 0$, we have

$$\partial_q f + \partial_p g = -\delta,$$

and according to the *Bendixon criterion* for $\delta > 0$ it has no closed orbits.

The vector–field X given by equations (2.10) has three fixed points given by $(q,p) = (0,0), (\pm 1,0)$. The *eigenvalues* $\lambda_{1,2}$ of the associated linearized vector–field are given by $\lambda_{1,2} = -\delta/2 \pm \frac{1}{2}\sqrt{\delta^2+4}$, for the fixed point $(0,0)$, and by $\lambda_{1,2} = -\delta/2 \pm \frac{1}{2}\sqrt{\delta^2-8}$, for the fixed point $(\pm 1,0)$. Hence, for $\delta > 0$,

$(0,0)$ is *unstable* and $(\pm 1,0)$ are *asymptotically stable*; for $\delta = 0$, $(\pm 1,0)$ are *stable in the linear approximation* (see, e.g., [Wig90]).

Another example of one–DOF Hamiltonian systems is a *simple pendulum* (again, all physical constants are scaled to unity), given by Hamiltonian function $H = \frac{p^2}{2} - \cos q$. This is the first integral of the *cylindrical* Hamiltonian vector–field $(q,p) \in S^1 \times \mathbb{R}$, defined by canonical equations

$$\dot{q} = p, \qquad \dot{p} = -\sin q.$$

This vector–field has fixed points at $(0,0)$, which is a center (i.e., the eigenvalues are purely imaginary), and at $(\pm\pi, 0)$, which are saddles, but since the phase–space manifold is the cylinder, these are really the same point.

The basis of human arm and leg dynamics represents the *coupling* of two uniaxial, $SO(2)-$joints. The study of two DOF Hamiltonian dynamics we shall start with the most simple case of two linearly coupled linear undamped oscillators with parameters scaled to unity. Under general conditions we can perform a change of variables to canonical coordinates (the 'normal modes') (q^i, p_i), $i = 1,2$, so that the vector–field X_H is given by

$$\dot{q}^1 = p_1, \qquad \dot{q}^2 = p_2, \qquad \dot{p}_1 = -\omega_1^2 q^1, \qquad \dot{p}_2 = -\omega_2^2 q^2.$$

This system is integrable, since we have two independent functions of (q^i, p_i), i.e., Hamiltonians

$$H_1 = \frac{p_1^2}{2} + \frac{\omega_1^2 (q^1)^2}{2}, \qquad H_2 = \frac{p_2^2}{2} + \frac{\omega_2^2 (q^2)^2}{2}.$$

The level curves of these functions are compact sets (topological circles); therefore, the orbits in the 4D phase–space $\mathbb{R}^4$ actually lie on the two–torus T^2. By making the appropriate change of variables, it can be shown (see, e.g., [Wig90]) that the whole dynamics of the two linearly coupled linear undamped oscillators is actually contained in the equations

$$\dot{\theta_1} = \omega_1, \qquad \dot{\theta_2} = \omega_2, \qquad (\theta_1, \theta_2) \in S^1 \times S^2 \equiv T^2. \tag{2.11}$$

The flow on the two–torus T^2, generated by (2.11), is simple to calculate and is given by

$$\theta_1(t) = \omega_1 t + \theta_{1_0}, \qquad \theta_1(t) = \omega_1 t + \theta_{1_0}, \qquad (\mathrm{mod}\ 2\pi),$$

and θ_1 and θ_2 are called the longitude and latitude. However, orbits under this flow will depend on how ω_1 and ω_2 are related. If ω_1 and ω_2 are *commensurate* (i.e., the equation $m\omega_1 + n\omega_2 = 0$, $(n,m) \in Z$ has solutions), then every phase curve of (2.11) is closed. However, if ω_1 and ω_2 are *incommensurate* i.e., upper equation has no solutions), then every phase curve of (2.11) is everywhere dense on T^2.

Somewhat deeper understanding of Hamiltonian dynamics is related to the method of *action–angle variables*. The easiest way to introduce this idea is to

consider again a simple harmonic oscillator (2.9). If we transform equations (2.9) into polar coordinates using $q = r\sin\theta,\ p = r\cos\theta$, then the equations of the vector–field become $\dot{r} = 0,\ \dot{\theta} = 1$, having the obvious solution $r = \text{const},\ \theta = t+\theta_0$. For this example polar coordinates work nicely because the system (2.9) is linear and, therefore, all of the periodic orbits have the same period.

For the general, nonlinear one–DOF Hamiltonian system (2.6) we will seek a coordinate transformation that has the same effect. Namely, we will seek a coordinate transformation $(q,p) \mapsto (\theta(q,p),\ I(q,p))$ with inverse transformation $(\theta, I) \mapsto (q(I,\theta),\ p(I,\theta))$ such that the vector–field (2.6) in the action–angle (θ, I) coordinates satisfies the following conditions: (i) $\dot{I} = 0$; (ii) θ changes linearly in time on the closed orbits with $\dot{\theta} = \Omega(I)$. We might even think of I and θ heuristically as 'nonlinear polar coordinates'. In such a coordinate system Hamiltonian function takes the form $H = H(I)$, and also, $\Omega(I) = \partial_I H$, i.e., specifying I specifies a periodic orbit.

The action variable $I(q,p)$ geometrically represents an area enclosed by any closed curve, which is constant in time. It is defined as an integral $I = \frac{1}{2\pi}\int_H p\,dq$, where H denotes the periodic orbit defined by $H(q,p) = H = \text{const}$. If the period of each periodic orbit defined by $H(q,p) = H = \text{const}$ is denoted by $T(H)$, the angle variable $\theta(q,p)$ is defined by

$$\theta(q,p) \;=\; \frac{2\pi}{T(H)}t(q,p),$$

where $t = t(q,p)$ represents the time taken for the solution starting from (q_0,p_0) to reach (q,p).

For the system with Hamiltonian $H = \frac{p^2}{2} + V(x)$ and momentum $p = \pm\sqrt{2}\sqrt{H - V(x)}$ the action is given by $I = \frac{\sqrt{2}}{\pi}\int_{q_{min}}^{q_{max}} \sqrt{H - V(q)}\,dq$, and the angle is given by $\theta(q,p) = \frac{2\pi}{T(H)}\int_{q_{min}}^{q_{max}} \frac{dq}{\sqrt{2}\sqrt{H-V(q)}}$.

Closely related to the action–angle variables is the *perturbation theory* (see [Nay73]). To explain the main idea of this theory, let us consider an $\epsilon-$*perturbed* vector–field periodic in t which can be in component form given as (with $(q,p) \in \mathbb{R}^2$)

$$\dot{q} \;=\; f_1(q,p) \,+\, \epsilon g_1(q,p,t,\epsilon), \qquad \dot{p} \;=\; f_2(q,p) \,+\, \epsilon g_2(q,p,t,\epsilon). \tag{2.12}$$

Setting $\epsilon = 0$ we get the *unperturbed* Hamiltonian system with a smooth scalar–valued function $H(q,p)$ for which holds

$$f_1(q,p) \;=\; \frac{\partial H(q,p)}{\partial p}, \qquad f_2(q,p) \;=\; -\frac{\partial H(q,p)}{\partial q},$$

so, the perturbed system (2.12) gets the symmetric canonical form

$$\dot{q} = \frac{\partial H(q,p)}{\partial p} + \epsilon g_1(q,p,t,\epsilon), \qquad \dot{p} = -\frac{\partial H(q,p)}{\partial q} + \epsilon g_2(q,p,t,\epsilon).$$

The perturbation (g_1, g_2) need not be Hamiltonian, although in the case where perturbation is Hamiltonian versus the case where it is not, the dynamics are very different.

Now, if we transform the coordinates of the perturbed vector–field using the action–angle transformation for the unperturbed Hamiltonian vector–field, we get

$$\begin{aligned}
\dot{I} &= \epsilon \left(\frac{\partial I}{\partial q} g_1 + \frac{\partial I}{\partial p} g_2 \right) \equiv \epsilon F(I, \theta, t, \epsilon), \qquad (2.13)\\
\dot{\theta} &= \Omega(I) + \epsilon \left(\frac{\partial \theta}{\partial q} g_1 + \frac{\partial \theta}{\partial p} g_2 \right) \equiv \Omega(I) + \epsilon G(I, \theta, t, \epsilon), \qquad \text{where}
\end{aligned}$$

$$\begin{aligned}
F(I, \theta, t, \epsilon) &= \frac{\partial I}{\partial q}(q(I, \theta), p(I, \theta))\, g_1((q(I, \theta), p(I, \theta), t, \epsilon)\\
&+ \frac{\partial I}{\partial p}(q(I, \theta), p(I, \theta))\, g_2((q(I, \theta), p(I, \theta), t, \epsilon), \qquad \text{and}\\
G(I, \theta, t, \epsilon) &= \frac{\partial \theta}{\partial q}(q(I, \theta), p(I, \theta))\, g_1((q(I, \theta), p(I, \theta), t, \epsilon)\\
&+ \frac{\partial \theta}{\partial p}(q(I, \theta), p(I, \theta))\, g_2((q(I, \theta), p(I, \theta), t, \epsilon).
\end{aligned}$$

Here, F and G are 2π periodic in θ and $T = 2\pi/\omega$ periodic in t.

Finally, we shall explain in brief the most important idea in the dynamical systems theory, the idea of *Poincaré maps*. The idea of reducing the study of continuous time systems (flows) to the study of an *associated discrete time system (map)* is due to Poincaré who first utilized it in the end of the last Century in his studies of the three body problem in celestial mechanics. Nowadays virtually any discrete time system that is associated with an ordinary differential equation is refereed to as a Poincaré map [Wig90]. This technique offers several advantages in the study of dynamical systems, including dimensional reduction, global dynamics and conceptual clarity. However, construction of a Poincaré map requires some knowledge of the phase–space of a dynamical system. One of the techniques which can be used for construction of Poincaré maps is the perturbation method.

To construct the Poincaré map for the system (2.13), we have to rewrite it as an autonomous system

$$\dot{I} = \epsilon F(I, \theta, \phi, \epsilon), \qquad \dot{\theta} = \Omega(I) + \epsilon G(I, \theta, \phi, \epsilon), \qquad \dot{\phi} = \omega, \qquad (2.14)$$

(where $(I, \theta, \phi) \in \mathbb{R}^+ \times S^1 \times S^1$. We construct a global *cross–section* Σ to this vector–field defined as $\Sigma^{\phi_0} = \{(I, \theta, \phi) | \phi = \phi_0\}$. If we denote the (I, θ) components of solutions of (2.14) by $(I_\epsilon(t), \theta_\epsilon(t))$ and the (I, θ) components of solutions of (2.14) for $\epsilon = 0$ by $(I_0, \Omega(I_0)t + \theta_0)$, then the perturbed Poincaré map is given by

$$P_\epsilon : \Sigma^{\phi_0} \to \Sigma^{\phi_0}, \qquad (I_\epsilon(0), \theta_\epsilon(0)) \mapsto (I_\epsilon(T), \theta_\epsilon(T)),$$

and the mth *iterate* of the Poincaré map is given by

$$P_\epsilon^m : \Sigma^{\phi_0} \to \Sigma^{\phi_0}, \qquad (I_\epsilon(0), \theta_\epsilon(0)) \mapsto (I_\epsilon(mT), \theta_\epsilon(mT)).$$

Now we can approximate the solutions to the perturbed problem as linear, constant–coefficient approximation

$$I_\epsilon(t) = I_0 + \epsilon I_1(t) + O(\epsilon^2), \qquad \theta_\epsilon(t) = \theta_0 + \Omega(I_0)t + \epsilon\theta_1(t) + O(\epsilon^2),$$

where we have chosen $I_\epsilon(0) = I_0$, $\theta_\epsilon(0) = \theta_0$.

As a last example of one–DOF Hamiltonian dynamics we shall analyze a *damped, forced Duffing oscillator*, given by canonical equations [Wig90]

$$\dot q = p, \qquad \dot p = q - q^3 - \delta p + \gamma\cos\omega t, \qquad \delta, \gamma, \omega \geq 0,\ (q,p) \in \mathbb{R}^2. \quad (2.15)$$

where δ, γ, and ω are real parameters physically meaning *dissipation*, *amplitude of forcing* and *frequency*, respectively.

The *perturbed* system (2.15) is given by

$$\dot q = p, \qquad \dot p = q - q^3 + \epsilon(\gamma\cos\omega t - \delta p), \tag{2.16}$$

where ϵ-perturbation is assumed small. Then the *unperturbed* system reads

$$\dot q = p, \qquad \dot p = q - q^3.$$

It is conservative with Hamiltonian function

$$H(q,p) = \frac{p^2}{2} - \frac{q^2}{2} + \frac{q^4}{4}. \tag{2.17}$$

In the unperturbed phase–space all orbits are given by the level sets of the Hamiltonian (2.17). There are three equilibrium points at the following coordinates: $(q,p) = (\pm 1, 0)$ – centers, and $(q,p) = (0,0)$ – saddle. The saddle point is connected to itself by two homoclinic orbits given by

$$q_+^0(t) = (\sqrt{2}(\cosh t)^{-1}, -\sqrt{2}(\cosh t)^{-1}\tanh t), \qquad q_-^0(t) = -q_+^0(t).$$

There are two families of *periodic orbits* $q_\pm^k(t)$, where k represents the *elliptic modulus* related to the Hamiltonian by $H(q_\pm^k(t)) \equiv H(k) = \frac{k^2-1}{(2-k^2)2}$, inside the corresponding homoclinic orbits $q_\pm^0(t)$, with the period $T(k) = 2K(k)\sqrt{2-k^2}$ ($K(k)$ is the complete elliptic integral of the first kind.

Also, there exists a family of periodic orbits outside the homoclinic orbits with the period $T(k) = 4K(k)\sqrt{k^2-1}$.

The perturbed system (2.16) can be rewritten as a third–order autonomous system

$$\dot q = p, \qquad \dot p = q - q^3 + \epsilon(\gamma\cos\phi - \delta p), \qquad \dot\phi = \omega,$$

where $(q, p, \phi) \in \mathbb{R}^2 \times S^1$, S^1 is the circle of length $2\pi/\omega$ and $\phi(t) = \omega t + \phi_0$. We form the global cross–section to the flow

$$\Sigma^{\phi_0} = \{(q, p, \phi) | \phi = \phi_0 \in [0, 2\pi/\omega]\}$$

and the associated Poincaré map is given by

$$P : \Sigma^{\phi_0} \to \Sigma^{\phi_0}, \qquad (q(0), p(0)) \mapsto (q(2\pi/\omega), p(2\pi/\omega)).$$

A detailed analysis of the perturbed Poincaré map for the damped, forced Duffing oscillator is related to the *Melnikov function* (see [Wig90]).

Complex 1–DOF Hamiltonian Dynamics

Recall that setting $z = q + ip, \quad z \in \mathbb{C}$, $i = \sqrt{-1}$, Hamiltonian equations $\dot{q} = \partial H/\partial p, \quad \dot{p} = -\partial H/\partial q$ may be written in *complex notation* as [AM78, MR99, Wig90]

$$\dot{z} = -2i \frac{\partial H}{\partial \bar{z}}. \tag{2.18}$$

Let U be an open set in the *complex phase–space manifold* M_C (i.e., manifold M modelled on $\mathbb{C}$). A C^0 function $\gamma : [a, b] \to U \subset M_C, \quad t \mapsto \gamma(t)$ represents a *solution curve* $\gamma(t) = q(t) + ip(t)$ of a complex Hamiltonian system (2.18). For example, the curve $\gamma(\theta) = \cos\theta + i \sin\theta, \quad 0 \le \theta \le 2\pi$ is the unit circle. $\gamma(t)$ is a *parameterized curve*. We call $\gamma(a)$ the *beginning point*, and $\gamma(b)$ the *end point* of the curve. By a *point on the curve* we mean a point w such that $w = \gamma(t)$ for some $t \in [a, b]$.

The derivative $\dot{\gamma}(t)$ is defined in the usual way, namely

$$\dot{\gamma}(t) = \dot{q}(t) + i\dot{p}(t),$$

so that the usual rules for the derivative of a sum, product, quotient, and chain rule are valid. The *speed* is defined as usual to be $|\dot{\gamma}(t)|$. Also, if $f : U \to M_C$ represents a *holomorphic*, or *analytic* function, then the composite $f \circ \gamma$ is differentiable (as a function of the real variable t) and $(f \circ \gamma)'(t) = f'(\gamma(t))\, \dot{\gamma}(t)$.

Recall that a *path* represents a sequence of C^1-curves,

$$\gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_n\},$$

such that the end point of γ_j, $(j = 1, \ldots, n)$ is equal to the beginning point of γ_{j+1}. If γ_j is defined on the interval $[a_j, b_j]$, this means that

$$\gamma_j(b_j) = \gamma_{j+1}(a_{j+1}).$$

We call $\gamma_1(a_1)$ the *beginning point* of γ_j, and $\gamma_n(b_n)$ the *end point* of γ_j. The path is said to *lie in an open set* $U \subset M_C$ if each curve γ_j lies in U, i.e., for each t, the point $\gamma_j(t)$ lies in U.

An open set U is *connected* if given two points α and β in U, there exists a path $\gamma = \gamma_1, \gamma_2, \ldots, \gamma_n$ in U such that α is the beginning point of γ_1 and β is the end point of γ_n; in other words, if there is a path γ in U which joins α to β. If U is a connected open set and f a holomorphic function on U such that $f' = 0$, then f is a constant. If g is a function on U such that $f' = g$, then f is called a *primitive* of g on U. Primitives can be either find out by integration or written down directly.

Let f be a C^0-function on an open set U, and suppose that γ is a curve in U, meaning that all values $\gamma(t)$ lie in U for $a \leq t \leq b$. The *integral of f along γ* is defined as

$$\int_\gamma f = \int_\gamma f(z) = \int_a^b f(\gamma(t))\,\dot{\gamma}(t)\,dt.$$

For example, let $f(z) = 1/z$, and $\gamma(\theta) = e^{i\theta}$. Then $\dot{\gamma}(\theta) = ie^{i\theta}$. We want to find the value of the integral of f over the circle, $\int_\gamma dz/z$, so $0 \leq \theta \leq 2\pi$. By definition, this integral is equal to $\int_0^{2\pi} ie^{i\theta}/e^{i\theta}\,d\theta = \mathrm{i}\int_0^{2\pi} d\theta = 2\pi i$.

The *length* $L(\gamma)$ is defined to be the integral of the speed, $L(\gamma) = \int_a^b |\dot{\gamma}(t)|\,dt$.

If $\gamma = \gamma_1, \gamma_2, \ldots, \gamma_n$ is a path, then the integral of a C^0-function f on an open set U is defined as $\int_\gamma f = \sum_{i=1}^n \int_{\gamma_i} f$, i.e., the sum of the integrals of f over each curve γ_i ($i = 1, \ldots, n$ of the path γ. The *length of a path* is defined as $L(\gamma) = \sum_{i=1}^n L(\gamma_i)$.

Let f be continuous on an open set $U \subset M_C$, and suppose that f has a primitive g, that is, g is holomorphic and $g' = f$. Let α, β be two points in U, and let γ be a path in U joining α to β. Then $\int_\gamma f = g(\beta) - g(\alpha)$; this integral is independent of the path and depends only on the beginning and end point of the path.

A *closed path* is a path whose beginning point is equal to its end point. If f is a C^0-function on an open set $U \subset M_C$ admitting a holomorphic primitive g, and γ is any closed path in U, then $\int_\gamma f = 0$.

Let γ, η be two paths defined over the same interval $[a, b]$ in an open set $U \subset M_C$. Recall (see Introduction) that γ is *homotopic* to η if there exists a C^0-function $\psi : [a, b] \times [c, d] \to U$ defined on a rectangle $[a, b] \times [c, d] \subset U$, such that $\psi(t, c) = \gamma(t)$ and $\psi(t, d) = \eta(t)$ for all $t \in [a, b]$. For each number $s \in [c, d]$ we may view the function $|psi_s(t) = \psi(t, s)$ as a continuous curve defined on $[a, b]$, and we may view the family of continuous curves ψ_s as a *deformation* of the path γ to the path η. It is said that the homotopy ψ *leaves the end points fixed* if we have $\psi(a, s) = \gamma(a)$ and $\psi(b, s) = \gamma(b)$ for all values of $s \in [c, d]$. Similarly, when we speak of a homotopy of closed paths, we assume that each path ψ_s is a closed path.

Let γ, η be paths in an open set $U \subset M_C$ having the same beginning and end points. Assume that they are homotopic in U. Let f be holomorphic on U. Then $\int_\gamma f = \int_\eta f$. The same holds for closed homotopic paths in U. In

particular, if γ is homotopic to a point in U, then $\int_\gamma f = 0$. Also, it is said that an open set $U \subset M_C$ is *simply–connected* if it is connected and if every closed path in U is homotopic to a point.

In the previous example we found that

$$\frac{1}{2\pi I} \int_\gamma \frac{1}{z}\, dz = 1,$$

if γ is a circle around the origin, oriented counterclockwise. Now we define for any closed path γ its *winding number* with respect to a point α to be

$$W(\gamma, \alpha) = \frac{1}{2\pi i} \int_\gamma \frac{1}{z - \alpha}\, dz,$$

provided the path does not pass through α. If γ is a closed path, then $W(\gamma, \alpha)$ is an integer.

A closed path $\gamma \in U \subset M_C$ is *homologous to 0 in* U if

$$\int_\gamma \frac{1}{z - \alpha}\, dz = 0,$$

for every point α not in U, or in other words, $W(\gamma, \alpha) = 0$ for every such point.

Similarly, let γ, η be closed paths in an open set $U \subset M_C$. We say that they are *homologous in* U, and write $\gamma \sim \eta$, if $W(\gamma, \alpha) = W(\eta, \alpha)$ for every point α in the complement of U. We say that γ is *homologous to* 0 *in* U, and write $\gamma \sim 0$, if $W(\gamma, \alpha) = 0$ for every point α in the complement of U.

If γ and η are closed paths in U and are homotopic, then they are homologous. If γ and η are closed paths in U and are close together, then they are homologous.

Let $\gamma_1, \ldots, \gamma_n$ be curves in an open set $U \subset M_C$, and let $m_1, \ldots, m_n$ be integers. A formal sum $\gamma = m_1\gamma_1 + \cdots + m_n\gamma_n = \sum_{i=1}^n m_i\gamma_i$ is called a *chain in* U. The chain is called *closed* if it is a finite sum of closed paths. If γ is the chain as above, then $\int_\gamma f = \sum_i m_\mathrm{i} \int_{\gamma_i} f$. If γ and η are closed chains in U, then $W(\gamma + \eta, \alpha) = W(\gamma, \alpha) + W(\eta, \alpha)$. We say that γ and η are *homologous in* U, and write $\gamma \sim \eta$, if $W(\gamma, \alpha) = W(\eta, \alpha)$ for every point α in the complement of U. We say that γ is *homologous to* 0 *in* U, and write $\gamma \sim 0$, if $W(\gamma, \alpha) = 0$ for every point α in the complement of U.

Recall that the *Cauchy theorem* states that if γ is a closed chain in an open set $U \subset M_C$, and γ is homologous to 0 in U, then $\int_\gamma f = 0$. If γ and η are closed chains in U, and $\gamma \sim \eta$ in U, then $\int_\gamma f = \int_\eta f$.

It follows from Cauchy's theorem that if γ and η are homologous, then $\int_\gamma f = \int_\eta f$ for all holomorphic functions f on U [AM78, Wig90].

Library of Basic Hamiltonian Systems

In this subsection, we present some basic Hamiltonian systems used by human–like biomechanics (for more details, see [Put93]).

1D Harmonic Oscillator

In this case we have $\{p, q\}$ as canonical coordinates on $\mathbb{R}^2$

$$M = T^*\mathbb{R} \simeq \mathbb{R}^2, \qquad \omega = dp \wedge dq,$$
$$H = \frac{1}{2}\left(p^2 + y\right), \qquad X_H = p\frac{\partial}{\partial q} - q\frac{\partial}{\partial p},$$

and Hamiltonian equations read

$$\dot{q} = p, \qquad \dot{p} = -q.$$

For each $f, g \in C^\infty(\mathbb{R}^2, \mathbb{R})$ the Poisson bracket is given by

$$\{f, g\}_\omega = \frac{\partial f}{\partial q}\frac{\partial g}{\partial p} - \frac{\partial f}{\partial p}\frac{\partial g}{\partial q}.$$

Complex Plane

Let $T^*\mathbb{R} \simeq \mathbb{R}^2$ have the canonical symplectic structure $\omega = dp \wedge dq$. Writing $z = q + ip$, we have

$$\omega = \frac{1}{2i}\, dz \wedge d\bar{z}, \qquad X_H = i\left(\frac{\partial H}{\partial z}\frac{\partial}{\partial z} - \frac{\partial H}{\partial \bar{z}}\frac{\partial}{\partial \bar{z}}\right),$$
$$\{f, g\}_\omega = \frac{\mathrm{i}}{2}\left(\frac{\partial f}{\partial z}\frac{\partial g}{\partial \bar{z}} - \frac{\partial f}{\partial \bar{z}}\frac{\partial g}{\partial z}\right),$$

so, the Hamiltonian equations, $\dot{q} = \partial_p H,\ \dot{p} = -\partial_q H$, become

$$\dot{z} = -2i\frac{\partial H}{\partial \bar{z}}.$$

2D Harmonic Oscillator

In this case we have $\{q^1, y, p_1, p_2\}$ as canonical coordinates on $\mathbb{R}^4$

$$M = T^*\mathbb{R}^2 \simeq \mathbb{R}^4, \qquad \omega = dp_1 \wedge dq^1 + dp_2 \wedge dq^2,$$
$$H = \frac{1}{2}\left[p_1^2 + p_2^2 + (q^1)^2 + (y)^2\right].$$

The functions $f = p_i p_j + q^i q^j$ and $g = p_i q^j + p_j q^i$, (for $i, j = 1, 2$), are constants of motion.

nD Harmonic Oscillator

In this case we have $(i = 1, ..., n)$

$$M = T^*\mathbb{R}^n \simeq \mathbb{R}^{2n}, \qquad \omega = dp_i \wedge dq^i,$$
$$H = \frac{1}{2}\sum_{i=1}^{n}\left[p_i^2 + (q^i)^2\right].$$

The system is integrable in an open set of $T^*\mathbb{R}^n$ with:

$$K_1 = H, \qquad K_2 = p_2^2 + (y)^2, \quad ..., \quad K_n = p_n^2 + (q^n)^2.$$

Toda Molecule

Consider three mass–points on the line with coordinates q^i, $(i = 1, 2, 3)$, and satisfying the ODEs:

$$\ddot{q}^i = -\partial_{q^i} U, \qquad \text{where} \qquad U = e^{q^1 - q^2} + e^{q^2 - q^3} - e^{q^3 - q^1}.$$

This is a Hamiltonian system with $\{q^i, p_i\}$ as canonical coordinates on $\mathbb{R}^6$,

$$M = T^*\mathbb{R}^3 \simeq \mathbb{R}^6, \qquad \omega = dp_i \wedge dq^i,$$
$$H = \frac{1}{2}\left(p_1^2 + p_2^2 + p_3^2\right) + U.$$

The *Toda molecule* (2.1.1) is an integrable Hamiltonian system in an open set of $T^*\mathbb{R}^3$ with:

$$K_1 = H, \qquad K_2 = p_1 + p_2 + p_3,$$
$$K_3 = \frac{1}{9}\left(p_1 + p_2 + p_3\right)\left(p_2 + p_3 - 2p_1\right)\left(p_3 + p_1 - 2p_2\right) - \left(p_1 + p_2 - 2p_3\right) e^{q^1 - q^2}$$
$$- \left(p_2 + p_3 - 2p_1\right) e^{q^2 - q^3} - \left(p_3 + p_1 - 2p_2\right) e^{q^3 - q^1}.$$

3–Point Vortex Problem

The motion of three–point vortices for an ideal incompressible fluid in the plane is given by the equations:

$$\dot{q}^j = -\frac{1}{2\pi} \sum_{i \neq j} \Gamma_i \left(p_j - p_i\right) / r_{ij}^2,$$
$$\dot{p}_j = \frac{1}{2\pi} \sum_{i \neq j} \Gamma_i \left(q^i - q^j\right) / r_{ij}^2,$$
$$r_{ij}^2 = \left(q^i - q^j\right)^2 + \left(p_j - p_i\right)^2,$$

where $i, j = 1, 2, 3$, and Γ_i are three nonzero constants. This mechanical system is Hamiltonian if we take:

$$M = T^*\mathbb{R}^3 \simeq \mathbb{R}^6, \qquad \omega = dp_i \wedge dq^i, \qquad (i = 1, ..., 3),$$
$$H = -\frac{1}{4\pi} \sum_{i,j=1}^{3} \Gamma_i \Gamma_i \ln\left(r_{ij}\right).$$

Moreover, it is integrable in an open set of $T^*\mathbb{R}^3$ with:

$$K_1 = H, \qquad K_2 = \sum_{i=1}^{3} \Gamma_i \left[\left(q^i\right)^2 + p_i^2\right],$$
$$K_3 = \left(\sum_{i=1}^{3} \Gamma_i q^i\right)^2 + K_2^2.$$

The Newton's Second Law as a Hamiltonian System

In the case of conservative forces, Newton's law of motion can be written on $\mathbb{R}^{3n}$ as

$$m_i \ddot{q}^i = -\partial_{q^i} U, \qquad (i = 1, 2, ..., 3n).$$

Its symplectic formulation reads:

$$M = T^*\mathbb{R}^3 \simeq \mathbb{R}^6, \qquad \omega = dp_i \wedge dq^i,$$
$$H = \sum_{i=1}^{3n} \frac{p_i^2}{2m_i} + U.$$

The Hamiltonian vector–field X_H is

$$X_H = \left(\frac{p_i}{m_i} \partial_{q^i} - \partial_{q^i} U \, \partial_{p_i} \right),$$

giving the Hamiltonian equations

$$\dot{q}^i = \frac{p_i}{m_i}, \qquad \dot{p}_i = -\partial_{q^i} U.$$

Rigid Body Fixed in a Point

The configuration space of a *rigid body fixed in a point* is $SO(3)$, the group of proper orthogonal transformations of $\mathbb{R}^3$ to itself, while the corresponding phase–space is its cotangent bundle, $T^*SO(3)$. The motion of a rigid body is a *geodesic* with respect to a *left–invariant Riemannian metric* (the inertia tensor) on $SO(3)$. The *momentum map* $J : P \to \mathbb{R}^3$ for the *left* $SO(3)-$action is *right* translation to the identity. We identify $\mathfrak{so}(3)^*$ with $\mathfrak{so}(3)$ via the *Killing form* and identify $\mathbb{R}^3$ with $\mathfrak{so}(3)$ via the map $v \mapsto \hat{v}$, where $\hat{v}(w) = v \times w$ ($\times$ being the standard cross product). Points in $\mathfrak{so}(3)^*$ are regarded as the left reduction of $T^*SO(3)$ by $G = SO(3)$ and are the angular momenta as seen from a *body–fixed frame*.

A Segment of a Human–Like Body

A *rigid body with a fixed point* is a basic model of a single segment of the human (or robot) body. This is a left–invariant Hamiltonian mechanical system on the phase–space $T^*SO(3)$. The differentiable structure on $SO(3)$ is defined using the traditional Euler angles $\{\varphi, \psi, \theta\}$. More precisely, a local chart is

$$(\varphi, \psi, \theta) \in \mathbb{R}^3 \longmapsto A \in SO(3), \qquad 0 < \varphi, \psi < 2\pi; \qquad 0 < \theta < \pi, \qquad \text{where}$$

$$A = \begin{bmatrix} \cos\psi\cos\varphi - \cos\theta\sin\varphi\sin\psi & \cos\psi\cos\varphi + \cos\theta\cos\varphi\sin\psi & \sin\theta\sin\psi \\ -\sin\psi\cos\varphi - \cos\theta\sin\varphi\sin\psi & -\sin\psi\sin\varphi + \cos\theta\cos\varphi\cos\psi & \sin\theta\cos\psi \\ \sin\theta\sin\varphi & -\sin\theta\cos\varphi & \cos\theta \end{bmatrix}$$

The corresponding conjugate momenta are denoted by $p_\varphi, p_\psi, p_\theta$, so $\{\varphi, \psi, \theta, p_\varphi, p_\psi, p_\theta\}$ is the phase–space $T^*SO(3)$. Thus, we have

$$M = T^*SO(3), \qquad \omega = dp_\varphi \wedge d\varphi + dp_\psi \wedge d\psi + dp_\theta \wedge d\theta, \qquad H = \frac{1}{2}K,$$

$$K = \frac{\left[(p_\varphi - p_\psi \cos\theta)\sin\psi + p_\theta \sin\theta\cos\psi\right]^2}{I_1 \sin^2\theta} + \frac{\left[(p_\varphi - p_\psi \cos\theta)\cos\psi - p_\theta \sin\theta\sin\psi\right]^2}{I_2 \sin^2\theta} + \frac{p_\psi^2}{I_3},$$

where I_1, I_2, I_3 are the moments of inertia, diagonalizing the inertia tensor of the body.

The Hamiltonian equations are

$$\dot\varphi = \frac{\partial H}{\partial p_\varphi}, \qquad \dot\psi = \frac{\partial H}{\partial p_\psi}, \qquad \dot\theta = \frac{\partial H}{\partial p_\theta},$$
$$\dot p_\varphi = -\frac{\partial H}{\partial \varphi}, \qquad \dot p_\psi = -\frac{\partial H}{\partial \psi}, \qquad \dot p_\theta = -\frac{\partial H}{\partial \theta}.$$

For each $f, g \in C^\infty(T^*SO(3), \mathbb{R})$ the Poisson bracket is given by

$$\{f,g\}_\omega = \frac{\partial f}{\partial \varphi}\frac{\partial g}{\partial p_\varphi} - \frac{\partial f}{\partial p_\varphi}\frac{\partial g}{\partial \varphi} + \frac{\partial f}{\partial \psi}\frac{\partial g}{\partial p_\psi} - \frac{\partial f}{\partial p_\psi}\frac{\partial g}{\partial \psi} + \frac{\partial f}{\partial \theta}\frac{\partial g}{\partial p_\theta} - \frac{\partial f}{\partial p_\theta}\frac{\partial g}{\partial \theta}.$$

The Heavy Top – Continued

Recall (see (2.4.1) above) that the heavy top is by definition a rigid body moving about a fixed point in a 3D space [Put93]. The rigidity of the top means that the distances between points of the body are fixed as the body moves. In this case we have

$$M = T^*SO(3),$$
$$\omega = dp_\varphi \wedge d\varphi + dp_\psi \wedge d\psi + dp_\theta \wedge d\theta,$$
$$H = \frac{1}{2}K + mgl\cos\theta,$$
$$K = \frac{\left[(p_\varphi - p_\psi \cos\theta)\sin\psi + p_\theta \sin\theta\cos\psi\right]^2}{I_1 \sin^2\theta} + \frac{\left[(p_\varphi - p_\psi \cos\theta)\cos\psi - p_\theta \sin\theta\sin\psi\right]^2}{I_2 \sin^2\theta} + \frac{p_\psi^2}{I_3},$$

where I_1, I_2, I_3 are the moments of inertia, m is the total mass, g is the gravitational acceleration and l is the length of the vector determining the center of mass at $t = 0$.

The Hamiltonian equations are

$$\dot\varphi = \frac{\partial H}{\partial p_\varphi}, \qquad \dot\psi = \frac{\partial H}{\partial p_\psi}, \qquad \dot\theta = \frac{\partial H}{\partial p_\theta},$$
$$\dot p_\varphi = -\frac{\partial H}{\partial \varphi}, \qquad \dot p_\psi = -\frac{\partial H}{\partial \psi}, \qquad \dot p_\theta = -\frac{\partial H}{\partial \theta}.$$

For each $f, g \in C^\infty(T^*SO(3), \mathbb{R})$ the Poisson bracket is given by

$$\{f,g\}_\omega = \frac{\partial f}{\partial \varphi}\frac{\partial g}{\partial p_\varphi} - \frac{\partial f}{\partial p_\varphi}\frac{\partial g}{\partial \varphi} + \frac{\partial f}{\partial \psi}\frac{\partial g}{\partial p_\psi} - \frac{\partial f}{\partial p_\psi}\frac{\partial g}{\partial \psi} + \frac{\partial f}{\partial \theta}\frac{\partial g}{\partial p_\theta} - \frac{\partial f}{\partial p_\theta}\frac{\partial g}{\partial \theta}.$$

The Hamiltonian H is invariant under rotations about the z–axis, *i.e.*, φ is a cyclic variable, so p_φ is a constant of motion. The momentum map for this S^1–action is $J(\varphi, \psi, \theta, p_\varphi, p_\psi, p_\theta) = p_\varphi$. The reduced phase–space $J^{-1}(p_\varphi)/S^1$ can be identified with T^*S^2 and it is parameterized by $\{\psi, \theta, p_\psi, p_\theta\}$. The equations of motion for ψ, θ are just Hamiltonian equations for H with p_φ held constant.

Two Coupled Pendula

The configuration space of the system of two coupled pendula in the plane is $T^2 = \{(\theta_1, \theta_2)\}$, where the θs are the two pendulum angles, the phase–space is T^*T^2 with its canonical symplectic structure and the Hamiltonian H is given by [Put93]

$$H = \frac{1}{2}(p_\varphi^2 + p_\psi^2) + V(\sqrt{2}\psi), \qquad \text{where}$$
$$\varphi = \frac{\theta_1 + \theta_2}{\sqrt{2}}, \qquad \psi = \frac{\theta_1 - \theta_2}{\sqrt{2}}.$$

The *circle group* S^1 acts on a torus T^2 by $\theta \cdot (\theta_1 + \theta_2) = (\theta + \theta_1, \theta + \theta_2)$ and hence the induced momentum map for the lifted action to T^*T^2 is given by $J(\varphi, \psi, p_\varphi, p_\psi) = p_\varphi$. Therefore, the reduced phase–space $J^{-1}(p_\varphi)/S^1$ is symplectically diffeomorphic to T^*S^1 with its canonical symplectic structure $\omega_\mu = dp_\psi \wedge d\psi$. The reduced Hamiltonian H_μ is $H_\mu = \frac{1}{2}p_\psi^2 + V(\sqrt{2}\psi)$, and Hamiltonian equations for H_μ are

$$\dot{\psi} = p_\psi, \qquad \dot{p}_\psi = -\sqrt{2}\dot{V}(\sqrt{2}\psi).$$

The Plane 2–Body Problem

The plane two body problem can be formulated as the triple (M, ω, H) where [Put93]

$$M = T^*\left((0,\infty) \times S^1\right), \qquad \omega = dp_r \wedge dr + dp_\theta \wedge d\theta,$$
$$H = (p_r^2 + p_\theta^2)/r^2 - 1/r.$$

The Lie group $G = SO(2) \simeq S^1$ acts on the configuration space $M = (0,\infty) \times S^1$ by rotations, i.e., if $R_\varphi \in SO(2)$ then

$$\phi : (R_\varphi, (r, \theta)) \mapsto (r, \theta + \varphi, p_r, p_\theta).$$

The corresponding momentum map is

$$J(r, \theta, p_r, p_\theta) = p_\theta.$$

The 3–Body Problem

There is a vast literature on the *restricted three–body problem* (see [MH92]). Among other things, there are investigations of the equilibriums points and their stability, investigations of the existence, stability and bifurcation of periodic orbits, and investigations of collisions and ejection orbits. The restricted problem is said to be a limit of the *full three–body problem* as one of the masses tends to zero, and so to each result for the restricted problem there should be a corresponding result for the full three–body problem.

The restricted three–body problem is a Hamiltonian system of differential equations which describes the motion of an infinitesimal particle (the *satellite*) moving under the gravitational influence of two particles of finite mass (the *primaries*) which are moving on a circular orbit of the Kepler problem [MS00].

Since the motion of the primaries is given, the restricted problem has two DOF for the planar problem and three DOF for the spatial problem. However, the full problem has six DOF in the planar case and nine DOF in the spatial case. Thus, at first the restricted problem seems too small to reflect the full complexity of the full problem; but when the symmetries of the full problem are taken into account the dimension gap narrows considerably.

The Hamiltonian of the full problem is invariant under Euclidean motions, i.e., translations and rotations, which begets the integrals of linear and angular momentum. Translations and rotations induce ignorable coordinates. Holding the integrals fixed and dropping the ignorable coordinates reduces the full problem from six to three DOF in the planar case and from nine to four DOF in the spatial case. Thus the full problem on the reduced space is only one DOF larger than the restricted problem in either the planar or the spatial case [MS00].

The full 3–body problem in 3D space has 9 DOF. By placing the center of mass at the origin and setting linear momentum equal to zero the problem reduces one with six DOF. This can be done using Jacobi coordinates. The Hamiltonian of the full 3–body problem in rotating (about the z–axis) Jacobi coordinates $(u_0, u_1, u_2, v_0, v_1, v_2)$ is

$$\begin{aligned} H = &\frac{\parallel v_0 \parallel^2}{2M_0} - u_0^T J v_0 + \frac{\parallel v_1 \parallel^2}{2M_1} - u_1^T J v_1 - \frac{m_0 m_1}{\parallel u_1 \parallel} \\ &+ \frac{\parallel v_2 \parallel^2}{2M_2} - u_2^T J v_2 - \frac{m_1 m_2}{\parallel u_2 - \alpha_0 u_1 \parallel} - \frac{m_2 m_0}{\parallel u_2 + \alpha_1 u_1 \parallel} \end{aligned}$$

where $u_i, v_i \in \mathbb{R}^3$,

$$\begin{aligned} M_0 &= m_0 + m_1 + m_2, \qquad M_1 = m_0 m_1/(m_0 + m_1), \\ M_2 &= m_2(m_0 + m_1)/(m_0 + m_1 + m_2), \\ \alpha_0 &= m_0/(m_0 + m_1), \qquad \alpha_1 = m_1/(m_0 + m_1), \end{aligned}$$

and $J = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

In these coordinates u_0 is the center of mass, v_0 is total linear momentum, and total angular momentum is: $A = u_0 \times v_0 + u_1 \times v_1 + u_2 \times v_2$. See [MH92] for further details.

n–DOF Hamiltonian Dynamics

Classically, $n-$DOF Hamiltonian dynamics combines the ideas of differential equations and variational principles (see [AM78, Arn89, MR99, Wig90]). As Hamiltonian first realized, many of the systems of mechanics and optics can be put into the special form (compare (2.6))

$$\dot{q}^i = \frac{\partial H}{\partial p_i}(q^i, p_i, t), \qquad \dot{p}_i = -\frac{\partial H}{\partial q^i}(q^i, p_i, t), \qquad (i = 1, \ldots, n),$$

or an associated variational form (summing upon the repeated index is used in the following text)

$$\delta \int (p_i dq^i - H)\, dt = 0.$$

Here the state of the system is given as a point $(q^1, \ldots, q^n, p_1, \ldots, p_n)$ in *phase–space*, the q's are the configuration coordinates, the p's are the momenta, t is time, and $H = H(q^i, p_i, t)$ is a total–energy function called Hamiltonian. The variables (q^i, p_i) are called *canonical* coordinates.

If $H = H(q^i, p_i)$ does not depend explicitly on time, the system is said to be *autonomous*. In this case, it is easy to verify that H is conserved. The search for other conserved quantities led to a new notion of solving Hamiltonian systems. Instead of finding formulae for the coordinates as a function of time, one searches for constants of the motion (*integrals*). If one can find n integrals $I_i(q^i, p_i)$ which are in *involution*:

$$[I_i, I_j] = \frac{\partial I_i}{\partial q^k}\frac{\partial I_j}{\partial p_k} - \frac{\partial I_i}{\partial p_k}\frac{\partial I_j}{\partial q^k} = 0, \qquad (i \neq j),$$

and *independent* (the vectors ∇I_i are independent 'almost everywhere'), then associated variables ϕ_i can be derived which evolve linearly in time: $\dot{\phi}^i = \frac{\partial H}{\partial I_i}(I_i)$.

Such a system is *integrable* in the sense of Liouville [Arn89]. If the sets $I = \text{const}$ are bounded, then they are nD *tori* T^n in phase–space. Choosing irreducible cycles, γ_i, on the tori, one can define a preferred set of integrals $J_i = \int_{\gamma_i} p_i dq^i$, called *action variables*, for which the corresponding ϕ_i are *angle variables* $\bmod 1$ on T^n. The quantities $\omega^i(J) = \frac{\partial H}{\partial J_i}(J_i)$ are called the *frequencies* on T^n.

Another feature of Hamiltonian systems noticed by Liouville is the preservation of phase–space volume $\int (dq)^n (dp)^n$. A more general result is that Poincaré's integral $\int p_i dq^i$ is conserved around any loop following the flow [Arn89]. This is the property that really distinguishes Hamiltonian differential equations from general ones.

The major problem with the notion of integrability is that *most systems are not integrable*. This was first appreciated when Poincaré proved that the *circular restricted three–body problem has no integral analytic in the mass ratio*. The perturbation expansions which gave excellent predictions of motion of the planets do not converge. The basic reason is that among the invariant tori of integrable systems is a dense subset on which the frequencies ω^i are *commensurate*, i.e, $m_i\omega^i = 0$ for some non–zero integer vector m_i. However, most systems have no commensurate tori, because they can be destroyed by arbitrarily small perturbation.

Poincaré went on to examine what really does happen. The key technique he used was geometrical analysis: instead of manipulating formulae for canonical transformations as Jacobi and others did, he pictured the *orbits* in phase–space. An important step in this *qualitative ODE theory* was the idea of *surface of section*. If Σ is a *codimension–one* surface (i.e., of dimension one less than that of the phase–space) transverse to a *flow*, then the sequence $\{x_j\}$ of successive intersections of an orbit with Σ gives a lot of information about that orbit. For example, if if $\{x_j\}$ is periodic then it corresponds to a periodic orbit. If $\{x_j\}$ is confined to a subset of codimension m on Σ then so is the orbit of the flow, etc.. The flow induces a mapping of Σ to itself; the map takes a point in Σ to the point at which it first returns to Σ (assuming there is on). Since the surface of section has one dimension less than the phase–space it is easier to picture the dynamics of the return map than the flow. In fact, for Hamiltonian systems one can do even better; since H is conserved, Σ decomposes into a one–parameter family of codimension two surfaces parameterized by the value of the energy, a reduction of two dimensions.

This led Poincaré to the ideas of *stable* and *unstable manifolds* for *hyperbolic* periodic orbits, which are extensions of the stable and unstable *eigenspaces* for associated linear systems, and their intersections, known as *hetero– and homo–clinic points*, whose orbits converge to one periodic orbit in the past and to another (or the same) in the future. He showed that having intersected once, the invariant manifolds must intersect infinitely often. Moreover the existence of one heteroclinic orbit implies the existence of an infinity of others.

The distance between the stable and unstable manifolds can be quantified by Melnikov's integral. This leads to a technique for proving the non–existence of integrals for a slightly perturbed, integrable Hamiltonian.

For integrable systems, nearby orbits separate linearly in time. However, dynamical systems can have exponentially separating orbits. Let δx be a tangent vector at the phase–space point x and δx_t be the evolved vector following the orbit of x. Then, recall that the average rate of exponentiation of δx_t is the *Lyapunov exponent* λ (see, e.g., [CD98])

$$\lambda(x, \delta x) = \lim_{t \longrightarrow \infty} 1/t \, \ln |\delta x_t|.$$

If λ is nonzero, then the predictions one can make will be valid for a time only logarithmic in the precision. Therefore, although deterministic in principle, a system need not be predictable in practice.

A concrete example of the complexity of behavior of typical Hamiltonian systems is provided by the 'horseshoe', a type of invariant set found near homoclinic orbits. Its points can be labelled by doubly infinite sequences of 0's and 1's corresponding to which half of a horseshoe shaped set the orbit is in at successive times. For every sequence, no matter how complicated, there is an orbit which has that symbol sequence. This implies, e.g., that a simple pendulum in a sufficiently strongly modulated time–periodic gravitational field has an initial condition such that the pendulum will turn over once each period when there is 1 in the sequence and not if there is a 0 for any sequence of 0's and 1's.

Hamilton–Poisson Mechanics

Now, instead of using symplectic structures arising in Hamiltonian mechanics, we propose the more general *Poisson manifold* $(\mathbf{g}^*, \{F, G\})$. Here $\mathbf{g}^*$ is a chosen Lie algebra with a $(\pm)$ *Lie–Poisson bracket* $\{F, G\}_{\pm}(\mu))$ and carries an abstract *Poisson evolution equation* $\dot{F} = \{F, H\}$. This approach is well–defined in both the finite– and the infinite–dimensional case. It is equivalent to the strong symplectic approach when this exists and offers a viable formulation for Poisson manifolds which are not symplectic (for technical details, see see [Wei90, AMR88, MR99, Put93, IP01a]).

Let E_1 and E_2 be Banach spaces. A continuous bilinear functional $<,>: E_1 \times E_2 \longrightarrow \mathbb{R}$ is nondegenerate if $< x, y >= 0$ implies $x = 0$ and $y = 0$ for all $x \in E_1$ and $y \in E_2$. We say E_1 and E_2 are in *duality* if there is a nondegenerate bilinear functional $<,>: E_1 \times E_2 \longrightarrow \mathbb{R}$. This functional is also referred to as an L^2-*pairing* of E_1 with E_2.

Recall that a *Lie algebra* consists of a vector space $\mathbf{g}$ (usually a Banach space) carrying a bilinear skew–symmetric operation $[,] : \mathbf{g} \times \mathbf{g} \to \mathbf{g}$, called the *commutator* or Lie bracket. This represents a pairing $[\xi, \eta] = \xi\eta - \eta\xi$ of elements $\xi, \eta \in \mathbf{g}$ and satisfies *Jacobi identity*

$$[[\xi, \eta], \mu] + [[\eta, \mu], \xi] + [[\mu, \xi], \eta] = 0.$$

Let $\mathbf{g}$ be a (finite– or infinite–dimensional) Lie algebra and $\mathbf{g}^*$ its dual Lie algebra, that is, the vector space L^2 paired with $\mathbf{g}$ *via* the inner product $<,>: \mathbf{g}^* \times \mathbf{g} \to \mathbb{R}$. If $\mathbf{g}$ is nD, this pairing reduces to the usual action (interior product) of forms on vectors. The standard way of describing any nD Lie algebra $\mathbf{g}$ is to give its n^3 *Lie structural constants* γ_{ij}^k, defined by $[\xi_i, \xi_j] = \gamma_{ij}^k \xi_k$, in some basis ξ_i, $(i = 1, \ldots, n)$.

For any two smooth functions $F : \mathbf{g}^* \to \mathbb{R}$, we define the *Fréchet derivative* D on the space $L(\mathbf{g}^*, \mathbb{R})$ of all linear diffeomorphisms from $\mathbf{g}^*$ to $\mathbb{R}$ as a map $DF : \mathbf{g}^* \to L(\mathbf{g}^*, \mathbb{R});\ \mu \mapsto DF(\mu)$. Further, we define the *functional derivative* $\delta F/\delta\mu \in \mathbf{g}$ by

$$DF(\mu) \cdot \delta\mu = < \delta\mu, \frac{\delta F}{\delta \mu} >$$

with arbitrary 'variations' $\delta\mu \in \mathbf{g}^*$.

For any two smooth functions $F, G : \mathbf{g}^* \to \mathbb{R}$, we define the $(\pm)$ *Lie–Poisson bracket* by

$$\{F, G\}_{\pm}(\mu) = \pm < \mu, \left[\frac{\delta F}{\delta \mu}, \frac{\delta G}{\delta \mu}\right] > . \tag{3.1}$$

Here $\mu \in \mathbf{g}^*$, $[\xi, \mu]$ is the Lie bracket in $\mathbf{g}$ and $\delta F/\delta\mu$, $\delta G/\delta\mu \in \mathbf{g}$ are the functional derivatives of F and G.

The $(\pm)$ Lie–Poisson bracket (3.1) is clearly a bilinear and skew–symmetric operation. It also satisfies the Jacobi identity

$$\{\{F, G\}, H\}_{\pm}(\mu) + \{\{G, H\}, F\}_{\pm}(\mu) + \{\{H, F\}, G\}_{\pm}(\mu) = 0$$

thus confirming that $\mathbf{g}^*$ is a Lie algebra, as well as Leibniz' rule

$$\{FG, H\}_{\pm}(\mu) = F\{G, H\}_{\pm}(\mu) \, + \, G\{F, H\}_{\pm}(\mu). \tag{2.19}$$

If $\mathbf{g}$ is a nD phase–space manifold with structure constants γ_{ij}^k, the $(\pm)$ Lie–Poisson bracket (2.19) becomes

$$\{F, G\}_{\pm}(\mu) = \pm \mu_k \gamma_{ij}^k \frac{\delta F}{\delta \mu_i} \frac{\delta G}{\delta \mu_j}. \tag{2.20}$$

The $(\pm)$ Lie–Poisson bracket represents a Lie–algebra generalization of the classical nD Poisson bracket $[F, G] = \omega(X_f, X_g)$ on the symplectic phase–space manifold (P, ω) for any two real–valued smooth functions $F, G : P \longrightarrow \mathbb{R}$.

As in the classical case, any two smooth functions $F, G : \mathbf{g}^* \longrightarrow \mathbb{R}$ are *in involution* if $\{F, G\}_{\pm}(\mu) = 0$.

The Lie–Poisson theorem states that a Lie algebra $\mathbf{g}^*$ with a $\pm$ Lie–Poisson bracket $\{F, G\}_{\pm}(\mu)$ represents a Poisson manifold $(\mathbf{g}^*, \{F, G\}_{\pm}(\mu))$.

Given a smooth Hamiltonian function $H : \mathbf{g}^* \to \mathbb{R}$ on the Poisson manifold $(\mathbf{g}^*, \{F, G\}_{\pm}(\mu))$, the time evolution of any smooth function $F : \mathbf{g}^* \to \mathbb{R}$ is given by the abstract *Poisson evolution equation*

$$\dot{F} = \{F, H\}. \tag{2.21}$$

Completely Integrable Hamiltonian Systems

In order to integrate a system of $2n$ ODEs, we must know $2n$ first integrals. It turns out that if we are given a canonical system of ODEs, it is often sufficient to know only n first integrals [Arn89].

Liouville Theorem on Completely Integrable Systems

Recall that a function F is a first integral of a system Ξ with Hamiltonian function H iff H and F are in involution on the system's phase–space P (which is the cotangent bundle of the system's configuration manifold T^*M), i.e., iff the Poisson bracket of H and F is identically equal to zero on P, $\{H, F\} \equiv 0$.

Liouville proved that if, in a system Ξ with n DOF (i.e., with a $2n$D phase–space $P = T^*M$), n independent first integrals in involution are known, then the system is integrable by quadratures.

Here is the exact formulation of the *Liouville theorem* [Arn89]: Suppose that we are given n functions in involution on a symplectic $2n$D manifold:

$$F_1, ..., F_n; \qquad \{F_i, F_j\} \equiv 0, \qquad (i, j = 1, ..., n).$$

Consider a level set of the functions F_i:

$$M_f = \{x : F_i(x) = f_i\}, \qquad (i = 1, ..., n).$$

Assume that the n functions F_i are independent on M_f (i.e., the n $1-$forms dF_i are linearly independent at each point of M_f). Then

1. M_f is a smooth manifold, invariant under the phase–flow with Hamiltonian function $H = F_1$.

2. If the manifold M_f is compact and connected, then it is diffeomorphic to the $n-$torus

$$T^n = \{(\varphi^1, ..., \varphi^n) \bmod 2\pi\}.$$

3. The phase–flow with Hamiltonian function H determines a conditionally periodic motion on M_f, i.e., in angular coordinates $\varphi^i = (\varphi^1, ..., \varphi^n)$ we have

$$\dot{\varphi}^i = \omega^i, \qquad \omega^i = \omega^i(f_i), \qquad (i = 1, ..., n).$$

4. The canonical equations with Hamiltonian function H can be integrated by quadratures.

For the proof of this theorem see [Arn89].

As an example with 3 DOF, we consider a heavy symmetrical Lagrangian top fixed at a point on its axis. Three first integrals are immediately obvious: H, M_z and M_3. It is easy to verify that the integrals M_z and M_3 are in involution. Furthermore, the manifold $H = h$ in the phase–space is compact. Therefore, we can say without any calculations that the motion of the top is conditionally periodic: the phase trajectories fill up the 3D torus T^3, given by: $H = c_1$, $M_z = c_2$, $M_3 = c_3$. The corresponding three frequencies are called frequencies of fundamental rotation, precession, and nutation.

Other examples arise from the following observation: *if a canonical system can be integrated by the method of Hamiltonian–Jacobi, then it has n first integrals in involution.* The method consists of a canonical transformation $(p_i, q^i) \to (P_i, Q^i)$ such that the Q^i are first integrals, while the functions Q^i and Q^i are in involution.

The Liouville theorem, as formulated above, covers all the problems of dynamics which have been integrated to the present day [Arn89].

Action–Angle Variables

Under the hypothesis of the Liouville theorem, we can find symplectic coordinates (I_i, φ^i) such that the first integrals F_i depend only on I_i and φ^i (for $i = 1, ..., n$) are angular coordinates on the $n-$torus $T^n \simeq M_f = \{x : F_i(x) = f_i\}$, which is invariant with respect to the phase–flow. We choose angular coordinates φ^i on M_f so that the phase–flow with Hamiltonian function $H = F_1$ takes an especially simple form [Arn89]:

$$\dot{\varphi}^i = \omega^i(f_i), \qquad \varphi^i(t) = \varphi^i(0) + \omega^i t.$$

Now we look at a neighborhood of the $n-$manifold $M_f = T^n$ in the system's $2n$D phase–space P.

In the coordinates (F_i, φ^i) the phase–flow with Hamiltonian function $H = F_1$ can be written in the form of the simple system of $2n$ ODEs

$$\dot{F}_i = 0, \qquad \dot{\varphi}^i = \omega^i(F_i), \qquad (i = 1, ..., n), \tag{2.22}$$

which is easily integrated: $F_i(t) = F_i(0)$, $\varphi^i(t) = \varphi^i(0) + \omega^i\,(F_i(0))\,t$.

Thus, in order to integrate explicitly the original canonical system of ODEs, it is sufficient to find the variables φ^i in explicit form. It turns out that this can be done using only quadratures. A construction of the variables φ^i is given below [Arn89].

In general, the variables (F_i, φ^i) are not symplectic coordinates. However, there are functions of F_i, which we denote by $I_i = I_i(F_i)$, $(i = 1, ..., n)$, such that the variables (I_i, φ^i) are symplectic coordinates: the original symplectic structure $dp_i \wedge dq^i$ is expressed in them as $dI_i \wedge d\varphi^i$. The variables I_i have physical dimensions of action and are called action variables; together with the angle variables φ^i they form the *action–angle system of canonical coordinates* in a neighborhood of $M_f = T^n$.

The quantities I_i are first integrals of the system with Hamiltonian function $H = F_1$, since they are functions of the first integrals F_i. In turn, the variables F_i can be expressed in terms of I_i and, in particular, $H = F_1 = H(I_i)$. In action–angle variables, the ODEs of our flow (2.22) have the form

$$\dot{I}_i = 0, \qquad \dot{\varphi}^i = \omega^i(I_i), \qquad (i = 1, ..., n).$$

A system with one DOF in the phase plane (p, q) is given by the Hamiltonian function $H(p, q)$. In order to construct the *action–angle variables*, we look for a canonical transformation $(p, q) \to (I, \varphi)$ satisfying the two conditions:

$$I = I(h), \qquad \oint_{M_h} d\varphi = 2\pi. \tag{2.23}$$

The *action variable* in the system with one DOF given by the Hamiltonian function $H(p, q)$ is the quantity

$$I(h) = \frac{1}{2\pi}\Pi(h) = \frac{1}{2\pi}\oint_{M_h} pdq,$$

which is the area bounded by the phase curve $H = h$. Arnold states the following theorem: Set $S(I,q) = \int_{q_0}^{q} pdq|_{H=h(I)}$ is a *generating function*. Then a canonical transformation $(p,q) \to (I,\varphi)$ satisfying conditions (2.23) is given by

$$p = \frac{\partial S(I,q)}{\partial q}, \qquad \varphi = \frac{\partial S(I,q)}{\partial I}, \qquad H\left(\frac{\partial S(I,q)}{\partial q}, q\right) = h(I).$$

We turn now to systems with n DOF given in $\mathbb{R}^{2n} = \{(p_i, q^i),\, i = 1, ..., n\}$ by a Hamiltonian function $H(p_i, q^i)$ and having n first integrals in involution $F_1 = H, F_2 ..., F_n$. Let $\gamma_1, ..., \gamma_n$ be a basis for the 1D cycles on the torus $M_f = T^n$ (the increase of the coordinate φ^i on the cycle γ_j is equal to 2π if $i = j$ and 0 if $i \neq j$). We set

$$I_i(f_i) = \frac{1}{2\pi}\oint_{M_h} p_i dq^i, \qquad (i = 1, ..., n). \tag{2.24}$$

The n quantities $I_i(f_i)$ given by formula (2.24) are called the *action variables* [Arn89].

We assume now that, for the given values f_i of the n integrals F_i, the n quantities I_i are independent, $\det(\partial I_i/\partial f_i)|_{f_i} \neq 0$. Then in a neighborhood of the torus $M_f = T^n$ we can take the variables I_i, φ^i as symplectic coordinates, i.e., the transformation $(p_i, q^i) \to (I_i, \varphi^i)$ is canonical, i.e.,

$$dp_i \wedge dq^i = dI_i \wedge d\varphi^i, \qquad (i = 1, ..., n).$$

Now, let m be a point on M_f, in a neighborhood of which the n variables q^i are coordinates of M_f, such that the submanifold $M_f \subset \mathbb{R}^{2n}$ is given by n equations of the form $p_i = p_i(I_i, q^i)$, $q^i(m) = q_0^i$. In a simply–connected neighborhood of the point q_0^i a single–valued function is defined,

$$S(I_i, q^i) = \int_{q_0}^{q} p_i(I_i, q^i)\, dq^i,$$

and we can use it as the generating function of a canonical transformation $(p_i, q^i) \to (I_i, \varphi^i)$:

$$p_i = \frac{\partial S}{\partial q^i}, \qquad \varphi^i = \frac{\partial S}{\partial I_i}.$$

A Universal Model for Completely Integrable Systems

A Hamiltonian system on a $2n$D symplectic manifold M is said to be *completely integrable* if it has n first integrals in involution, which are functionally independent on some open dense submanifold of M. This definition of a completely integrable system is usually found, with some minor variants, in any modern text on symplectic mechanics [Arn89, AM78, LM87, MS95, Thi79].

Starting with this definition, one uses the so–called *Liouville–Arnold theorem* to introduce action–angle variables and write the Hamiltonian system in the form

$$\dot{I}^k = 0, \qquad \dot{\phi}_k = \frac{\partial H}{\partial I_k} = \nu_k(I),$$

where $k \in \{1, \ldots, n\}$. The corresponding flow is given by

$$I^k(t) = I^k(0), \qquad \phi_k(t) = \phi_k(0) + \nu_k t. \tag{2.25}$$

The main interest in completely integrable systems relies on the fact that they can be integrated by quadratures [Arn89].

However, it is clear that even if $\nu_k dI^k$ is not an exact (or even a closed) 1–form, as long as $\dot{\nu}_k = 0$, the system can always be integrated by quadratures.

If we consider the Abelian Lie group $\mathbb{R}^n$, we can construct a Hamiltonian action of $\mathbb{R}^n$ on $T^*\mathbb{R}^n$ induced by the group addition: $\mathbb{R}^n \times T^*\mathbb{R}^n \to T^*\mathbb{R}^n$. This can be generalized to the Hamiltonian action [AGM97]

$$\mathbb{R}^n \times T^*(\mathbb{R}^k \times T^{n-k}) \to T^*(\mathbb{R}^k \times T^{n-k}),$$

of $\mathbb{R}^n$, where T^m stands for the mD torus, and reduces to $\mathbb{R}^n \times T^*T^n$ or $T^n \times T^*T^n$, when $k = 0$.

By using the standard symplectic structure on $T^*\mathbb{R}^n$, we find the momentum map $\mu : T^*\mathbb{R}^n \longrightarrow (\mathbb{R}^n)^s$, $(q, p) \mapsto p$, induced by the natural action of $\mathbb{R}^n$ on itself via translations, which is a Poisson map if $(\mathbb{R}^n)^s$ is with the (trivial) natural Poisson structure of the dual of a Lie algebra. It is now clear that any function on $(\mathbb{R}^n)^s$, when pulled back to $T^*\mathbb{R}^n$ or T^*T^n, induces a Hamiltonian system which is completely integrable (in the Liouville sense). Because the level sets of this function carry on the action of $\mathbb{R}^n$, the completely integrable system induces a 1D subgroup of the action of $\mathbb{R}^n$ on the given level set. However, the specific subgroup will depend on the particular level set, i.e., the 'frequencies' are first integrals. The property of being integrable by quadratures is captured by the fact that it is a subgroup of the $\mathbb{R}^n-$action on each level set.

It is now clear, how we can preserve this property, while giving up the requirement that our system is Hamiltonian. We can indeed consider any 1–form η on $(\mathbb{R}^n)^s$ and pull it back to $T^*\mathbb{R}^n$ or T^*T^n, then associated vector–field $\Gamma_\eta = \Lambda_0(\mu^s(\eta))$, where Λ_0 is the canonical Poisson structure in the cotangent bundle, is no more Hamiltonian, but it is still integrable by quadratures. In action–angle variables, if $\eta = \nu_k dI^k$ is the 1–form on $(\mathbb{R}^n)^s$, the associated equations of motion on T^*T^n will be [AGM97]

$$\dot{I}^k = 0, \qquad \dot{\phi}_k = \nu_k,$$

with $\dot{\nu}_k = 0$, therefore the flow will be as in (2.25), even though $\partial_{I^j}\nu_k \neq \partial_{I^k}\nu_j$.

We can now generalize this construction to any Lie group G. We consider the Hamiltonian action $G \times T^*G \to T^*G$, of G on the cotangent bundle,

induced by the right action of G on itself. The associated momentum map $\mu : T^*G \simeq \mathcal{G}^s \times G \longrightarrow \mathcal{G}^s$. It is a Poisson map with respect to the natural Poisson structure on $\mathcal{G}^s$ (see, e.g., [AGM94, LM87]).

Now, we consider any differential 1–form η on $\mathcal{G}^s$ which is annihilated by the natural Poisson structure $\Lambda_{\mathcal{G}^*}$ on $\mathcal{G}^s$ associated with the Lie bracket. Such form we call a *Casimir form*. We define the vector–field $\Gamma_\eta = \Lambda_0(\mu^s(\eta))$. Then, the corresponding dynamical system can be written as [AGM97]

$$g^{-1}\dot{g} = \eta(g,p) = \eta(p), \qquad \dot{p} = 0,$$

since $\omega_0 = d(< p, g^{-1}dg >)$ (see [AGM94]). Here we interpret the covector $\eta(p)$ on $\mathcal{G}^s$ as a vector of $\mathcal{G}$. Again, our system can be integrated by quadratures, because on each level set, get by fixing p's in $\mathcal{G}^s$, our dynamical system coincides with a one–parameter group of the action of G on that particular level set.

We give a familiar example: the rigid rotator and its generalizations [AGM97]. In the case of $G = \mathrm{SO}(3)$ the (right) momentum map

$$\mu : T^*\mathrm{SO}(3) \longrightarrow \mathfrak{so}(3)^*$$

is a Poisson map onto $\mathfrak{so}(3)^*$ with the linear Poisson structure

$$\Lambda_{\mathfrak{so}(3)^*} = \varepsilon^{ijk} p_i \partial_{p_j} \otimes \partial_{p_k}.$$

Casimir 1–forms for $\Lambda_{so(3)^*}$ read $\eta = FdH_0$, where $H_0 = \sum p_i^2/2$ is the 'free Hamiltonian' and $F = F(p)$ is an arbitrary function. Clearly, FdH_0 is not a closed form in general, but (p_i) are first integrals for the dynamical system $\Gamma_\eta = \Lambda_0(\mu^s(\eta))$. It is easy to see that

$$\Gamma_\eta = F(p)\Gamma_0 = F(p)p_i\widehat{X_i},$$

where $\widehat{X_i}$ are left–invariant vector–fields on $\mathrm{SO}(3)$, corresponding to the basis (X_i) of $\mathfrak{so}(3)$ identified with (dp_i). Here we used the identification $T^*\mathrm{SO}(3) \simeq \mathrm{SO}(3) \times \mathfrak{so}(3)^*$ given by the momentum map μ. In other words, the dynamics is given by

$$\dot{p}_i = 0, \qquad g^{-1}\dot{g} = F(p)p_iX_i \in \mathfrak{so}(3),$$

and it is completely integrable, since it reduces to left–invariant dynamics on $\mathrm{SO}(3)$ for every value of p. We recognize the usual isotropic rigid rotator, when $F(p) = 1$.

We can generalize our construction once more, replacing the cotangent bundle T^*G by its deformation, namely a group double $D(G, \Lambda_G)$ associated with a Lie–Poisson structure Λ_G on G (see e.g., [Lu90]). This double, denoted simply by D, carry on a natural *Poisson tensor–field* Λ_D^+ which is non–degenerate on the open–dense subset $D^+ = G \cdot G^s \cap G^s \cdot G$ of D (here $G^s \subset D$ is the dual group of G with respect to Λ_G). We refer to D as being

complete if $D^+ = D$. Identifying D with $G \times G^s$ if D is complete (or D^+ with an open submanifold of $G \times G^s$ in general case; we assume completeness for simplicity) via the group product, we can write Λ_D^+ in 'coordinates' $(g, u) \in G \times G^s$ in the form [AGM97]

$$\Lambda_D^+(g,u) = \Lambda_G(g) + \Lambda_{G^*}(u) - X_i^l(g) \wedge Y_i^r(u), \tag{2.26}$$

where X_i^l and Y_i^r are, respectively, the left– and right–invariant vector–fields on G and G^s relative to dual bases X_i and Y_i in the Lie algebras $\mathcal{G}$ and $\mathcal{G}^s$, and where Λ_G and Λ_{G^*} are the corresponding Lie–Poisson tensors on G and G^s (see [Lu90]). It is clear now that the projections μ_{G^*} and μ_G of (D, Λ_D^+) onto (G, Λ_G) and (G^s, Λ_{G^*}), respectively, are Poisson maps. Note that we get the cotangent bundle $(D, \Lambda_D^+) = (T^*G, \Lambda_0)$ if we put $\Lambda_G = 0$.

The group G acts on (D, Λ_D^+) by left translations which, in general are not canonical transformations. However, this is a Poisson action with respect to the inner Poisson structure Λ_G on G, which is sufficient to develop the momentum map reduction theory (see [Lu91]). For our purposes, let us take a Casimir 1–form η for Λ_{G^*}, i.e., $\Lambda_{G^*}(\eta) = 0$. By means of the momentum map
$\mu_{G^*} : D \longrightarrow G^s$, we define the vector–field on D [AGM97]:

$$\Gamma_\eta = \Lambda_D^+(\mu_{G^*}^s(\eta)).$$

In 'coordinates' (g, u), due to the fact that η is a Casimir, we get

$$\Gamma_\eta(g,u) = < Y_i^r, \eta > (u) X_i^l(g),$$

so that Γ_η is associated with the *Legendre map*

$$L_\eta : D \simeq G \times G^s \longrightarrow TG \simeq G \times \mathcal{G}, \qquad L_\eta(g,u) = < Y_i^r, \eta > (u) X_i,$$

which can be viewed also as a map $L_\eta : G^s \longrightarrow \mathcal{G}$. Thus we get the following theorem [AGM97]: The dynamics Γ_η on the group double $D(G, \Lambda_G)$, associated with a 1–form η which is a Casimir for the Lie–Poisson structure Λ_{G^*} on the dual group, is given by the system of equations

$$\dot{u} = 0, \qquad g^{-1}\dot{g} = < Y_i^r, \eta > (u) X_i \in \mathcal{G},$$

and is therefore completely integrable by quadratures.

We have seen that if we concentrate on the possibility of integrating our system by quadratures, then we can do without the requirement that the system is Hamiltonian.

By considering again the equations of motion in *action–angle variables*, we classically have

$$\dot{I}^k = 0, \qquad \dot{\phi}_k = \nu^k(I).$$

Clearly, if we have

$$\dot{I}^k = F_k(I), \qquad \dot{\phi}_k = A_k^j(I)\phi_j, \tag{2.27}$$

and we are able to integrate the first equation by quadratures, we again have the possibility to integrate by quadratures the system (2.27), if only the matrices $(A_k^j(I(t)))$ commute [AGM97]:

$$\phi(t) = \exp\left(\int_0^t A(I(s))ds\right)\phi_0.$$

Because ϕ_k are discontinues functions on the torus, we have to be more careful here. However, we show how this idea works for double groups. In the case when the 1–form η on G^s is not a Casimir 1–form for the Lie–Poisson structure Λ_{G^*}, we get, in view of (2.26),

$$\Gamma_\eta(g,u) =< Y_i^r, \eta > (u)X_i^l(g) + \Lambda_{G^*}(\eta)(u).$$

Now, the momenta evolve according to the dynamics $\Lambda_{G^*}(\eta)$ on G^s (which can be interpreted, as we will see later, as being associated with an interaction of the system with an external field) and 'control' the evolution of the field of velocities on G (being left–invariant for a fixed time) by a 'variation of constants'. Let us summarize our observations in the following theorem [AGM97]: The vector–field Γ_η on the double group $D(G,\Lambda_G)$, associated with a 1–form η on G^s, defines the following dynamics

$$\dot{u} = \Lambda_{G^*}(\eta)(u), \qquad g^{-1}\dot{g} =< Y_i^r, \alpha > (u)X_i \in \mathcal{G}, \tag{2.28}$$

and is therefore completely integrable, if only we are able to integrate the equation (2.28) and $< Y_i^r, \eta > (u(t))X_i$ lie in a commutative subalgebra of $\mathcal{G}$ for all t.

Finally, we can weaken the assumptions of the previous theorem. It is sufficient to assume [AGM97] that

$$g^{-1}\dot{g}(t) = \exp(tX)A(t)\exp(-tX),$$

for some $A(t)$, $X \in \mathcal{G}$, such that $X + A(t)$ lie in a commutative subalgebra of $\mathcal{G}$ for all t (e.g., $A(t) = \text{const}$), to assure that (2.28) is integrable by quadratures. Indeed, in the new variable

$$g_1(t) = \exp(-tX)g(t)\exp(tX),$$

the equation (2.28) reads

$$\dot{g}_1(t) = g_1(t)(X + A(t)) - Xg_1(t),$$

and, since the right– and the left–multiplications commute, we find that [AGM97]

$$g(t) = g_0\exp\left(tX + \int_0^t A(s)ds\right)\exp(-tX).$$

This procedure is similar to what is known as the Dirac interaction picture in the quantum evolution.

Killing Vector– and Tensor–Fields in Mechanics

Recall from subsection 1.2.13 above, that on a Riemannian manifold (M, g) with the system's kinetic energy metric tensor $g = (g_{ij})$, for any pair of vectors V and T, the following relation holds[1]

$$\partial_s \langle V, T\rangle = \langle \nabla_s V, T\rangle + \langle V, \nabla_s T\rangle, \tag{2.29}$$

where $\langle V, T\rangle = g_{ij} V^i T^j$. If the curve $\gamma(s)$ is a geodesic, for a generic vector X we have

$$\partial_s \langle X, \dot\gamma\rangle = \langle \nabla_s X, \dot\gamma\rangle + \langle X, \nabla_s \dot\gamma\rangle = \langle \nabla_s X, \dot\gamma\rangle \equiv \langle \nabla_{\dot\gamma} X, \dot\gamma\rangle, \tag{2.30}$$

where

$$(\nabla_{\dot\gamma} X)^i = \partial_s x^l \partial_{x^l} X^i + \Gamma^i_{jk} \partial_s x^j X^k,$$

so that in components it reads

$$\partial_s (X_i \dot x^i) = \dot x^i \nabla_i (X_j \dot x^j).$$

Using the fact that $X_j \dot x^i \nabla_i \dot x^j = X_j \nabla_{\dot\gamma} \dot\gamma^j = 0$, as well as the auto–parallelism of the geodesics, this can be rewritten as

$$\partial_s (X_i \dot x^i) = \frac{1}{2} \dot x^j \dot x^i (\nabla_i X_j + \nabla_j X_i), \qquad (i, j = 1, ..., N).$$

This means that the conservation of $X_i \dot x^i$ along a geodesic, i.e., $\partial_s (X_i \dot x^i) = 0$, is guaranteed by (see [CP02])

$$\nabla_{(i} X_{j)} \equiv \nabla_i X_j + \nabla_j X_i = 0. \tag{2.31}$$

If such a field exists on a manifold, it is the *Killing vector–field*. Recall that (2.31) is equivalent to $\mathcal{L}_X g = 0$, where $\mathcal{L}$ is the *Lie derivative*. On the biomechanical manifolds (M, g), being the unit vector $\dot q^i$ – tangent to a geodesic – proportional to the canonical momentum $p_i = \frac{\partial L}{\partial \dot q^i} = \dot q^i$, the existence of a Killing vector–field X implies that the corresponding *momentum map* (see subsection 1.2.11 above),

$$J(q, p) = X_k(q) \partial_s q^k = \frac{1}{\sqrt{2}(E - V(q))} X_k(q) \dot q^k = \frac{1}{\sqrt{2T(q)}} X_k(q) p_k, \tag{2.32}$$

is a constant of motion along the geodesic flow. Thus, for an NDOF Hamiltonian system, a physical conservation law, involving a conserved quantity linear in the canonical momenta, can always be related with a symmetry on the manifold (M, g) due to the action of a Killing vector–field on the manifold.

[1] In this subsection, the overdot denotes the derivative upon the arc–length parameter s, namely $\dot{(\,)} \equiv \partial_s \equiv d/ds$, while ∇_s is the covariant derivative along a curve $\gamma(s)$.

These are the *Noether conservation laws*. The equation (2.31) is equivalent to the vanishing of the Poisson brackets,

$$\{H, J\} = \left(\frac{\partial H}{\partial q^i} \frac{\partial J}{\partial p_i} - \frac{\partial H}{\partial p_i} \frac{\partial J}{\partial q^i} \right) = 0, \tag{2.33}$$

which is the standard definition of a constant of motion $J(q, p)$ (see, e.g., [AM78]).

However, if a 1–1 correspondence is to exist between conserved physical quantities along a Hamiltonian flow and suitable symmetries of the biomechanical manifolds (M, g), then *integrability* will be equivalent to the existence of a number of symmetries at least equal to the number of DOF, which is equal to $\dim(M)$. If a Lie group G acts on the phase–space manifold through completely canonical transformations, and there exists an associated momentum map, then every Hamiltonian having G as a symmetry group, with respect to its action, admits the momentum map as the constant of motion. These symmetries are usually referred to as *hidden symmetries* because, even though their existence is ensured by integrability, they are not easily recognizable [CP02].

Let us now extend what has been presented so far about Killing vector–fields, trying to generalize the form of the conserved quantity along a geodesic flow from $J = X_i \dot{x}^i$ to $J = K_{j_1 j_2 \dots j_r} \dot{x}^{j_1} \dot{x}^{j_2} \dots \dot{x}^{j_r}$, with $K_{j_1 j_2 \dots j_r}$ a tensor of rank r. Thus, we look for the conditions that entail

$$\partial_s (K_{j_1 j_2 \dots j_r} \dot{x}^{j_1} \dot{x}^{j_2} \dots \dot{x}^{j_r}) = \dot{x}^j \nabla_j (K_{j_1 j_2 \dots j_r} \dot{x}^{j_1} \dot{x}^{j_2} \dots \dot{x}^{j_r}) = 0. \tag{2.34}$$

In order to work out from this equation a condition for the existence of a suitable tensor $K_{j_1 j_2 \dots j_r}$, which is called a *Killing tensor–field*, let us first consider the $2r$ rank tensor $K_{j_1 j_2 \dots j_r} \dot{x}^{i_1} \dot{x}^{i_2} \dots \dot{x}^{i_r}$ and its covariant derivative along a geodesic [CP02]

$$\dot{x}^j \nabla_j (K_{j_1 j_2 \dots j_r} \dot{x}^{i_1} \dot{x}^{i_2} \dots \dot{x}^{i_r}) = \dot{x}^{i_1} \dot{x}^{i_2} \dots \dot{x}^{i_r} \dot{x}^j \nabla_j K_{j_1 j_2 \dots j_r}, \tag{2.35}$$

where we have again used $\dot{x}^j \nabla_j \dot{x}^{i_k} = 0$ along a geodesic, and a standard covariant differentiation formula (see Appendix, as well as section 1.2.9 above). Now, by contraction on the indices i_k and j_k the $2r-$rank tensor in (2.35) gives a new expression for (2.34), which reads

$$\partial_s (K_{j_1 j_2 \dots j_r} \dot{x}^{j_1} \dot{x}^{j_2} \dots \dot{x}^{j_r}) = \dot{x}^{j_1} \dot{x}^{j_2} \dots \dot{x}^{j_r} \dot{x}^j \nabla_{(j} K_{j_1 j_2 \dots j_r)}, \tag{2.36}$$

where $\nabla_{(j} K_{j_1 j_2 \dots j_r)} = \nabla_j K_{j_1 j_2 \dots j_r} + \nabla_{j_1} K_{j j_2 \dots j_r} + \dots + \nabla_{j_r} K_{j_1 j_2 \dots j_{r-1} j}$. The vanishing of (2.36), entailing the conservation of $K_{j_1 j_2 \dots j_r} \dot{x}^{j_1} \dot{x}^{j_2} \dots \dot{x}^{j_r}$ along a geodesic flow, is therefore guaranteed by the existence of a tensor–field fulfilling the conditions [CP02]

$$\nabla_{(j} K_{j_1 j_2 \dots j_r)} = 0. \tag{2.37}$$

These equations generalize (2.31) and give the definition of a Killing tensor–field on a Riemannian biomechanical manifold (M, g). These N^{r+1} equations in $(N+r-1)!/r!(N-1)!$ unknown independent components of the Killing tensor constitute an *overdetermined* system of equations. Thus, *a'priori*, we can expect that the existence of Killing tensor–fields has to be rather exceptional.

If a Killing tensor–field exists on a Riemannian manifold, then the scalar

$$K_{j_1 j_2 \dots j_r} \dot{q}^{j_1} \dot{q}^{j_2} \dots \dot{q}^{j_r}$$

is a constant of motion for the geodesic flow on the same manifold. With the only difference of a more tedious combinatorics, also in this case it turns out that the equations (2.37) are equivalent to the vanishing of the Poisson brackets of $J(q, p)$, that is

$$\{H, J\} = 0 \qquad \text{is equivalent to} \qquad \nabla_{(j} K_{j_1 j_2 \dots j_r)} = 0.$$

Thus, the existence of Killing tensor–fields, obeying (2.37), on a biomechanical manifold (M, g) give the rephrasing of integrability of Newtonian equations of motion or, equivalently, of standard Hamiltonian systems, within the Riemannian geometrical framework.

The first natural question to address concerns the existence of a Killing tensor–field, on any biomechanical manifold (M, g), to be associated with total energy conservation. Such a Killing tensor–field actually exists and coincides with the metric tensor g, in fact it satisfies by definition (2.37).

One of the simplest case of integrable system is represented by a decoupled system described by a generic Hamiltonian

$$H = \sum_{i=1}^{N} \left[\frac{p_i^2}{2} + V_i(q^i) \right] = \sum_{i=1}^{N} H_i(q^i, p_i)$$

for which all the energies E_i of the subsystems $H_i,\ i = 1, \dots, N$, are conserved. On the associated biomechanical manifold, N second–order Killing tensor–fields exist, they are given by

$$K^{(i)}_{jk} = \delta_{jk} \{ V_i(q^i)[E - V(q^i)] + \delta^i_j [E - V(q^i)]^2 \}.$$

In fact, these tensor–fields fulfil (2.37), which explicitly reads [CP02]

$$\begin{aligned} &\nabla_k K^{(i)}_{lm} + \nabla_l K^{(i)}_{mk} + \nabla_m K^{(i)}_{kl} \\ &= \partial_{q^k} K^{(i)}_{lm} + \partial_{q^l} K^{(i)}_{mk} + \partial_{q^m} K^{(i)}_{kl} - 2\Gamma^j_{kl} K^{(i)}_{jm} - 2\Gamma^j_{km} K^{(i)}_{jl} - 2\Gamma^j_{lm} K^{(i)}_{jk} = 0. \end{aligned}$$

The conserved quantities $J^{(i)}(q, p)$ are then get by saturation of the tensors $K^{(i)}$ with the velocities $\dot{q}^i$,

$$J^{(i)}(q, p) = K^{(i)}_{jk} \dot{q}^j \dot{q}^k = E_i.$$

Distributions and Nonholonomic Geometry

Let $TM = \cup_{x\in M} T_xM$, be the tangent bundle of a smooth nD mechanical manifold M. Recall (from the subsection 1.3.3 above) that sub–bundle $V = \cup_{x\in M} V_x$, where V_x is a vector subspace of T_xM, smoothly dependent on points $x \in M$, is called the *distribution*. If the manifold M is connected, $\dim V_x$ is called the dimension of the distribution. A vector–field X on M *belongs* to the distribution V if $X(x) \subset V_x$. A curve γ is *admissible relatively to* V, if the vector–field $\dot\gamma$ belongs to V. A *differential system* is a linear space of vector–fields having a structure of $C^\infty(M)$ – module. Vector–fields which belong to the distribution V form a differential system $N(V)$. A kD distribution V is *integrable* if the manifold M is foliated to kD sub–manifolds, having V_x as the tangent space at the point x. According to the Frobenius theorem, V is integrable iff the corresponding differential system $N(V)$ is *involutive*, i.e., if it is a Lie sub–algebra of the Lie algebra of vector–fields on M. The *flag* of a differential system N is a sequence of differential systems: $N_0 = N,\ N_1 = [N, N], \ldots,\ N_l = [N_{l-1}, N], \ldots$.

The differential systems N_i are not always differential systems of some distributions V_i, but if for every i, there exists V_i, such that $N_i = N(V_i)$, then there exists *a flag* of the distribution V: $V = V_0 \subset V_1 \ldots$. Such distributions, which have flags, will be called *regular*. Note that the sequence $N(V_i)$ is going to stabilize, and there exists a number r such that $N(V_{r-1}) \subset N(V_r) = N(V_{r+1})$. If there exists a number r such that $V_r = TM$, the distribution V is called *completely nonholonomic*, and minimal such r is the *degree of non–holonomicity* of the distribution V.

Now, let us see the mechanical interpretation of these geometrical objects. Consider a nonholonomic mechanical system corresponding to a Riemannian manifold (M, g), where g is a metric defined by the system's kinetic energy [DG03]. Suppose that the distribution V is defined by $(n-m)$ one–forms ω_α; in local coordinates $q = (q^1, ..., q^n)$ on M

$$\omega_\rho(q)(\dot q) = a_{\rho i}(q)\, \dot q^i = 0, \qquad (\rho = m+1, \ldots, n;\ i = 1, \ldots, n).$$

A *virtual displacement* is a vector–field X on M, such that $\omega_\rho(X) = 0$, i.e., X belongs to the differential system $N(V)$.

Differential equations of motion of a given mechanical system follow from the *D'Alambert–Lagrangian principle*: trajectory γ of the given system is a solution of the equation

$$\langle \nabla_{\dot\gamma}\dot\gamma - Q, X\rangle = 0, \tag{2.38}$$

where X is an arbitrary virtual displacement, Q a vector–field of internal forces, and ∇ is the affine *Levi–Civita connection* for the metric g.

The vector–field $R(x)$ on M, such that $R(x) \in V_x^\perp,\ V_x^\perp \oplus V_x = T_xM$, is called *reaction of ideal nonholonomic connections*. (2.38) can be rewritten as

$$\nabla_{\dot\gamma}\dot\gamma - Q = R, \qquad \omega_\alpha(\dot\gamma) = 0. \tag{2.39}$$

If the system is potential, by introducing $L = T - U$, where U is the potential energy of the system ($Q = -\operatorname{grad} U$), then in local coordinates q on M, equations (2.39) becomes the *forced Lagrangian equation*:

$$\frac{d}{dt}L_{\dot q} - L_q = \tilde R, \qquad \omega_\alpha(\dot q) = 0.$$

Now $\tilde R$ is a one–form in $(V^\perp)$, and it can be represented as a linear combination of one–forms $\omega^{m+1}, \dots, \omega^n$ which define the distribution, $\tilde R = \lambda_\alpha \omega_\alpha$.

Suppose $e_1, \dots, e_n$ are the vector–fields on M, such that $e_1(x), \dots, e_n(x)$ form a base of the vector space T_xM at every point $x \in M$, and $e_1, \dots, e_m$ generate the differential system $N(V)$. Express them through the coordinate vector–fields:

$$e_i = A_i^j(q)\partial_{q^j}, \qquad (i, j = 1, \dots, n).$$

Denote by p a projection $p : TM \to V$ orthogonal according to the metric g. Corresponding homomorphism of $C^\infty-$modules of sections of TM and V is

$$p\partial_{q^i} = p_i^a e_a, \qquad (a = 1, \dots, m,\ i = 1, \dots, n).$$

Projecting by p the equations (2.39), from $R(x) \in V^\perp(x)$, we get $p(R) = 0$, and denoting $p(Q) = \tilde Q$ we get

$$\tilde\nabla_{\dot\gamma}\dot\gamma = \tilde Q,$$

where $\tilde\nabla$ is the *projected connection* [DG03]. A relationship between standard Christoffel symbols Γ_{ij}^k and coefficients $\tilde\Gamma_{ab}^c$ of the connection $\tilde\nabla$, defined by

$$\tilde\nabla_{e_a} e_b = \tilde\Gamma_{ab}^c e_c, \qquad \text{is given by} \qquad \tilde\Gamma_{ab}^c = \Gamma_{ij}^k A_a^i A_b^j p_k^c + A_a^i\, \partial_{q^i} A_b^j\, p_j^c.$$

If the motion takes place under the inertia ($Q = \tilde Q = 0$), the trajectories of nonholonomic mechanical problem are the geodesics for $\tilde\nabla$.

Now, let V be a distribution on M. Denote a $C^\infty(M)-$module of sections on V by $\Gamma(V)$. A *nonholonomic connection* on the sub–bundle V of TM is a map $\nabla : \Gamma(V) \times \Gamma(V) \to \Gamma(V)$ with the properties:

$$\nabla_X(Y + Z) = \nabla_X Y + \nabla_X Z, \qquad \nabla_X(f \cdot Y) = X(f)Y + f\nabla_X Y,$$
$$\nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_Y Z, (X, Y, Z \in \Gamma(V);\ f, g \in C^\infty(M)).$$

Having a morphism of vector bundles $p_0 : TM \to V$, formed by the projection on V, denote by $q_0 = 1_{TM} - p_0$ the projection on W, $V \oplus W = TM$.

The tensor–field $T_\nabla : \Gamma(V) \times \Gamma(V) \to \Gamma(V)$ defined by

$$T_\nabla(X, Y) = \nabla_X Y - \nabla_Y X - p_0[X, Y], \quad (X, Y \in \Gamma(V)),$$

is called the *torsion tensor* for the connection ∇.

Suppose there is a positively defined metric tensor $g = g_{ij}$ on V. Given a distribution V, with p_0 and g, there exists a unique nonholonomic connection ∇ with the properties [DG03]

$$\nabla_X g(Y,Z) = X(g(Y,Z)) - g(\nabla_X Y, Z) - g(Y, \nabla_X Z) = 0, \qquad T_\nabla = 0.$$

These conditions can be rewritten in the form:

$$\nabla_X Y = \nabla_Y X + p_0[X,Y], \qquad Z(g(X,Y)) = g(\nabla_Z X, Y) + g(X, \nabla_Z Y).$$

By cyclic permutation of X, Y, Z and summing we get:

$$\begin{aligned} g(\nabla_X Y, Z) = \frac{1}{2}\{&X(g(Y,Z)) + Y(g(Z,X)) - Z(g(X,Y)) \\ &+ g(Z, p_0[X,Y]) + g(Y, p_0[Z,X]) - g(X, p_0[Y,Z]\}. \end{aligned} \tag{2.40}$$

Let q^i, $(i = 1, \ldots, n)$ be local coordinates on M, such that the first m coordinate vector–fields ∂_{q^j} are projected by projection p_0 into vector–fields e_a, $(a = 1, \ldots, m)$, generating the distribution V: $p_0 \partial_{q^j} = p_i^a(q) e_a$. Vector–fields e_a can be expressed in the basis ∂_{q^j} as $e_a = B_a^i \partial_{q^j}$, with $B_a^i p_i^b = \delta_a^b$. Now we give coordinate expressions for the coefficients of the connection Γ_{ab}^c, defined as $\nabla_{e_a} e_b = \Gamma_{ab}^c e_c$. From (2.40) we get

$$\Gamma_{ab}^c = \{{}^c_{ab}\} + g_{ae} g^{cd} \Omega_{bd}^e + g_{be} g^{cd} \Omega_{ad}^e - \Omega_{ab}^c,$$

where Ω is get from $p_0[e_a, e_b] = -2\Omega_{ab}^c e_c$ as

$$2\Omega_{ab}^c = p_i^c e_a(B_b^i) - p_i^c e_b(B_a^i),$$

and $\{{}^c_{ab}\} = \frac{1}{2} g^{ce}(e_a(g_{be}) + e_b(g_{ae}) - e_e(g_{ab}))$.

2.1.2 Non–Autonomous Lagrangian/Hamiltonian Mechanics

Geodesics

In this subsection we continue our study of non–autonomous, *time–dependent mechanics* on a *configuration bundle* $Q \to \mathbb{R}$, that we started in subsection 1.4.6 above. Recall that $\mathbb{R}$ is the time axis, while the corresponding *velocity phase–space manifold* is the 1–jet space $J^1(\mathbb{R}, Q)$ of sections $s : \mathbb{R} \to Q$ of the bundle $Q \to \mathbb{R}$. Also, recall that second–order dynamical equation (dynamical equation, for short) on a fibre bundle $Q \to \mathbb{R}$ is defined as a first–order dynamical equation on the jet bundle $J^1(\mathbb{R}, Q) \to \mathbb{R}$, given by a holonomic connection ξ on $J^1(\mathbb{R}, Q) \to \mathbb{R}$ which takes its values in the 2–jet space $J^2(\mathbb{R}, Q) \subset J^1(\mathbb{R}, J^1(\mathbb{R}, Q))$ (see [LM96, MS98, MP94, MS99]). The global geometrical structure of time–dependent mechanics is depicted in Figure 1.9 above.

Since a configuration bundle $Q \to \mathbb{R}$ is trivial, the existent formulations of mechanics often imply its preliminary splitting $Q = \mathbb{R} \times M$ [CF93, EMR91, LM96, MFV90]. This is not the case of mechanical systems subject to time–dependent transformations, including inertial frame transformations. Recall that different trivializations of $Q \to \mathbb{R}$ differ from each other in projections

$Q \to M$. Since a configuration bundle $Q \to \mathbb{R}$ has no canonical trivialization in general, mechanics on $Q \to \mathbb{R}$ is not a repetition of mechanics on $\mathbb{R} \times M$, but implies additionally a connection on $Q \to \mathbb{R}$ which is a reference frame [MS98, Sar98]. Considered independently on a trivialization of $Q \to \mathbb{R}$, mechanical equations make the geometrical sense of geodesic equations.

We now examine quadratic dynamical equations in details. In this case, the corresponding dynamical connection γ on $J^1(\mathbb{R}, Q) \to Q$ is *affine*, while the connection $\widetilde{K}$ (1.219) on $TQ \to Q$ is *linear*. Then the equation for Jacobi vector–fields along the geodesics of the connection $\widetilde{K}$ can be considered. This equation coincides with the existent equation for Jacobi fields of a Lagrangian system [DR94, MS98] in the case of non–degenerate quadratic Lagrangians, when they can be compared. We will consider more general case of quadratic Newtonian systems characterized by a pair (ξ, μ) of a quadratic dynamical equation ξ and a Riemannian inertia tensor μ which satisfy a certain compatibility condition. Given a reference frame, a Riemannian inertia tensor μ is extended to a Riemannian metric on the configuration space Q. Then conjugate points of solutions of the dynamical equation ξ can be examined in accordance with the well–known geometrical criteria [MS98, Sar98].

Quadratic Dynamical Equations

From the physical viewpoint, the most interesting dynamical equations are the quadratic ones, i.e.,

$$\xi^i = a^i_{jk}(q^\mu)q^j_t q^k_t + b^i_j(q^\mu)q^j_t + f^i(q^\mu). \tag{2.41}$$

This property is coordinate–independent due to the affine transformation law of coordinates q^i_t. Then, it is clear that the corresponding dynamical connection γ_ξ (1.228) is affine:

$$\gamma = dq^\alpha \otimes [\partial_\alpha + (\gamma^i_{\lambda 0}(q^\nu) + \gamma^i_{\lambda j}(q^\nu)q^j_t)\partial^t_i],$$

and *vice versa*. This connection is symmetric iff $\gamma^i_{\lambda\mu} = \gamma^i_{\mu\lambda}$.

There is 1–1 correspondence between the affine connections γ on the affine jet bundle $J^1(\mathbb{R}, Q) \to Q$ and the linear connections $\widetilde{K}$ (1.232) on the tangent bundle $TQ \to Q$.

This correspondence is given by the relation

$$\gamma^i_\mu = \gamma^i_{\mu 0} + \gamma^i_{\mu j}q^j_t, \qquad \gamma^i_{\mu\lambda} = K_\mu{}^i{}_\alpha.$$

In particular, if an affine dynamical connection γ is symmetric, so is the corresponding linear connection $\widetilde{K}$.

Any quadratic dynamical equation

$$q^i_{tt} = a^i_{jk}(q^\mu)q^j_t q^k_t + b^i_j(q^\mu)q^j_t + f^i(q^\mu) \tag{2.42}$$

is equivalent to the geodesic equation

$$\dot{t} = 1, \qquad \ddot{t} = 0,$$
$$\ddot{q}^i = a^i_{jk}(q^\mu)\dot{q}^i\dot{q}^j + b^i_j(q^\mu)\dot{q}^j\dot{t} + f^i(q^\mu)\dot{t}\dot{t}. \tag{2.43}$$

for the symmetric linear connection

$$\widetilde{K} = dq^\alpha \otimes (\partial_\alpha + K^\mu_{\alpha\nu}(t, q^i)\dot{q}^\nu\dot{\partial}_\mu)$$

on $TQ \to Q$, given by the components

$$K^0_{\alpha\nu} = 0, \quad K_0{}^i{}_0 = f^i, \quad K_0{}^i{}_j = K_j{}^i{}_0 = \frac{1}{2}b^i_j, \quad K_j{}^i{}_k = a^i_{jk}. \tag{2.44}$$

Conversely, any linear connection K on the tangent bundle $TQ \to Q$ defines the quadratic dynamical equation

$$q^i_{tt} = K_j{}^i{}_k q^j_t q^k_t + (K_0{}^i{}_j + K_j{}^i{}_0)q^j_t + K_0{}^i{}_0,$$

written with respect to a given reference frame $(t, q^i) \equiv q^\mu$.

However, the geodesic equation (2.43) is not unique for the dynamical equation (2.42). Any quadratic dynamical equation (2.41), being equivalent to the geodesic equation with respect to the linear connection $\widetilde{K}$ (2.44), is also equivalent to the geodesic equation with respect to an affine connection K' on $TQ \to Q$ which differs from $\widetilde{K}$ (2.44) in a soldering form σ on $TQ \to Q$ with the local components

$$\sigma^0_\alpha = 0, \qquad \sigma^i_k = h^i_k + (s-1)h^i_k\dot{q}^0, \qquad \sigma^i_0 = -sh^i_k\dot{q}^k - h^i_0\dot{q}^0 + h^i_0,$$

where s and h^i_α are local functions on Q.

Equation of Free–Motion

We say that the dynamical equation (1.218), that is: $q^i_{tt} = \xi^i(t, q^i, q^i_t)$, is a *free motion equation* iff there exists a reference frame $(t, \overline{q}^i)$ on the configuration bundle $Q \to \mathbb{R}$ such that this equation reads

$$\overline{q}^i_{tt} = 0. \tag{2.45}$$

With respect to arbitrary bundle coordinates (t, q^i), a free motion equation takes the form

$$q^i_{tt} = d_t\Gamma^i + \partial_j\Gamma^i(q^j_t - \Gamma^j) - \frac{\partial q^i}{\partial\overline{q}^\mu}\frac{\partial\overline{q}^\mu}{\partial q^j\partial q^k}(q^j_t - \Gamma^j)(q^k_t - \Gamma^k), \tag{2.46}$$

where $\Gamma^i = \partial_t q^i(t, \overline{q}^j)$ is the connection associated with the initial frame $(t, \overline{q}^i)$. One can think of the r.h.s. of the equation (2.46) as being the general coordinate expression of an inertial force in mechanics. The corresponding dynamical connection γ on the affine jet bundle $J^1(\mathbb{R}, Q) \to Q$ reads

$$\gamma^i_k = \partial_k\Gamma^i - \frac{\partial q^i}{\partial\overline{q}^\mu}\frac{\partial\overline{q}^\mu}{\partial q^j\partial q^k}(q^j_t - \Gamma^j), \qquad \gamma^i_0 = \partial_t\Gamma^i + \partial_j\Gamma^i q^j_t - \gamma^i_k\Gamma^k. \tag{2.47}$$

This affine dynamical connection defines a linear connection K on the tangent bundle $TQ \to Q$ whose curvature is necessarily zero. Thus, we come to the following criterion for a dynamical equation to be a free motion equation: if ξ is a free motion equation, it is quadratic and the corresponding linear symmetric connection (2.44) on the tangent bundle $TQ \to Q$ is flat. A free motion equation on a configuration bundle $Q \to \mathbb{R}$ exists iff the typical fibre M of Q admits a curvature–free linear symmetric connection.

Quadratic Lagrangian and Newtonian Systems

Recall that a Lagrangian of an nD mechanical system on $Q \to \mathbb{R}$ is defined as a function on the velocity phase–space $J^1(\mathbb{R}, Q)$. In particular, let us consider a *non–degenerate quadratic Lagrangian*

$$L = \frac{1}{2}\mu_{ij}(q^\mu)q^i_t q^j_t + k_i(q^\mu)q^i_t + f(q^\mu), \tag{2.48}$$

where μ_{ij} $(i, j = 1, ..., n)$ is a Riemannian fibre metric tensor in the vertical tangent bundle VQ, called the *inertial metric tensor*. As for quadratic dynamical equations, this property is coordinate–independent, namely one can show that any quadratic polynomial on $J^1(\mathbb{R}, Q) \subset TQ$ is extended to a bilinear form on TQ, so that the Lagrangian L (2.48) can be written as

$$L = \frac{1}{2}\gamma_{\alpha\mu}q^\alpha_t q^\mu_t, \qquad (\text{with} \qquad q^0_t = 1),$$

where γ is the (degenerate) fibre metric in the tangent bundle TQ, given by

$$\gamma_{00} = 2f, \qquad \gamma_{0i} = k_i, \qquad \gamma_{ij} = \mu_{ij}. \tag{2.49}$$

The associated Lagrangian equation takes the form

$$q^i_{tt} = (\mu^{-1})^{ik}\Gamma_{\lambda k\nu}q^\alpha_t q^\nu_t, \tag{2.50}$$
$$\text{where} \qquad \Gamma_{\lambda\mu\nu} = -\frac{1}{2}(\partial_\alpha\gamma_{\mu\nu} + \partial_\nu\gamma_{\mu\lambda} - \partial_\mu\gamma_{\lambda\nu})$$

are the *Christoffel symbols of the first–kind* of the metric $\gamma_{\alpha\mu}$ given in components by (2.49). The corresponding geodesic equation (2.43) on TQ reads

$$\ddot{q}^i = (\mu^{-1})^{ik}\Gamma_{\lambda k\nu}\dot{q}^\alpha\dot{q}^\nu, \qquad \dot{t} = 1, \qquad \ddot{t} = 0, \tag{2.51}$$

such that $\widetilde{K}$ (1.219) is a linear connection with the following components

$$\widetilde{K}^0_{\alpha^\nu} = 0, \qquad \widetilde{K}^i_{\alpha^\nu} = (\mu^{-1})^{ik}\Gamma_{\lambda k\nu}.$$

We have the relation

$$\dot{q}^\alpha(\partial_\alpha\mu_{ij} + K^i_{\alpha^\nu}\dot{q}^\nu) = 0. \tag{2.52}$$

One can show that an arbitrary Lagrangian system on a configuration bundle $Q \to \mathbb{R}$ is a particular Newtonian system on $Q \to \mathbb{R}$. The latter is defined

as a pair (ξ, μ) of a dynamical equation ξ and a (degenerate) fibre metric μ in the fibre bundle $V_Q J^1(\mathbb{R}, Q) \to J^1(\mathbb{R}, Q)$ which satisfy the symmetry condition $\partial^t_k \mu_{ij} = \partial^t_j \mu_{ik}$ and the compatibility condition [MS98, MOS99]

$$\xi \rfloor d\mu_{ij} + \mu_{ik}\gamma^k_j + \mu_{jk}\gamma^k_i = 0, \tag{2.53}$$

where γ_ξ is the dynamical connection (1.228), i.e.,

$$\gamma_\xi = dt \otimes [\partial_t + (\xi^i - \frac{1}{2} q^j_t \partial^t_j \xi^i)\partial^t_i] + dq^j \otimes [\partial_j + \frac{1}{2}\partial^t_j \xi^i \partial^t_i].$$

We restrict our consideration to non–degenerate quadratic Newtonian systems when ξ is a quadratic dynamical equation (2.41) and μ is a Riemannian fibre metric in VQ, i.e., μ is independent of q^i_t and the symmetry condition becomes trivial. In this case, the dynamical equation (2.42) is equivalent to the geodesic equation (2.43) with respect a symmetric linear connection $\widetilde{K}$ (2.44), while the compatibility condition (2.53) takes the form (2.52).

Given a symmetric linear connection $\widetilde{K}$ (2.44) on the tangent bundle $TQ \to Q$, one can consider the equation for Jacobi vector–fields along geodesics of this connection, i.e., along solutions of the dynamical equation (2.42). If Q admits a Riemannian metric, the *conjugate points* of these geodesic can be investigated.

Jacobi Fields

Let us consider the quadratic dynamical equation (2.42) and the equivalent geodesic equation (2.43) with respect to the symmetric linear connection $\widetilde{K}$ (2.44). Its *Riemann curvature tensor*

$$R_{\lambda\mu}{}^\alpha_\beta = \partial_\lambda K_\mu{}^\alpha_\beta - \partial_\mu K^\alpha_{\lambda\beta} + K^\gamma_{\lambda\beta} K_\mu{}^\alpha_\gamma - K_\mu{}^\gamma{}_\beta K^\alpha_{\lambda\gamma}$$

has the temporal component

$$R_{\lambda\mu}{}^0{}_\beta = 0. \tag{2.54}$$

Then the equation for a Jacobi vector–field u along a geodesic c reads

$$\dot{q}^\beta \dot{q}^\mu (\nabla_\beta (\nabla_\mu u^\alpha) - R_{\lambda\mu}{}^\alpha_\beta u^\lambda) = 0, \qquad \nabla_\beta \dot{q}^\alpha = 0, \tag{2.55}$$

where ∇_μ denotes the covariant derivative relative to the connection $\widetilde{K}$ (see [KN63/9]). Due to the relation (2.54), the equation (2.55) for the temporal component u^0 of a Jacobi field takes the form

$$\dot{q}^\beta \dot{q}^\mu (\partial_\mu \partial_\beta u^0 + K_\mu{}^\gamma{}_\beta \partial_\gamma u^0) = 0.$$

We chose its solution

$$u^0 = 0, \tag{2.56}$$

because all geodesics obey the constraint $\dot{t} = 0$.

Note that, in the case of a quadratic Lagrangian L, the equation (2.55) coincides with the Jacobi equation

$$u^j d_0(\partial_j \dot{\partial}_i L) + d_0(\dot{u}^j \dot{\partial}_i \dot{\partial}_j L) - u^j \partial_i \partial_j L = 0$$

for a Jacobi field on solutions of the Lagrangian equations for L. This equation is the Lagrangian equation for the vertical extension L_V of the Lagrangian L (see [DR94, MS98, MOS99]).

Let us consider a quadratic Newtonian system with a Riemannian inertia tensor μ_{ij}. Given a reference frame $(t, q^i) \equiv q^\alpha$, this inertia tensor is extended to the following Riemannian metric on Q

$$\overline{g}_{00} = 1, \qquad \overline{g}_{0i} = 0, \qquad \overline{g}_{ij} = \mu_{ij}.$$

However, its covariant derivative with respect to the connection $\widetilde{K}$ (2.44) does not vanish in general. Nevertheless, due to the relations (2.52) and (2.56), we get the well–known formula for a Jacobi vector–field u along a geodesic c:

$$\int_a^b \left(\overline{g}_{\lambda\mu}(\dot{q}^\alpha \nabla_\alpha u^\lambda)(\dot{q}^\beta \nabla_\beta u^\mu) + R_{\lambda\mu\alpha\nu} u^\lambda u^\alpha \dot{q}^\mu \dot{q}^\nu\right) dt$$
$$+ \ \overline{g}_{\lambda\mu} \dot{q}^\alpha \nabla_\alpha u^\lambda u'^\mu|_{t=a} - \overline{g}_{\lambda\mu} \dot{q}^\alpha \nabla_\alpha u^\lambda u'^\mu|_{t=b} = 0.$$

Therefore, the following assertions also remain true [KN63/9]: (i) if the sectional curvature $R_{\lambda\mu\alpha\nu} u^\lambda u^\alpha \dot{q}^\mu \dot{q}^\nu$ is positive on a geodesic c, this geodesic has no conjugate points; (ii) if the sectional curvature $R_{\lambda\mu\alpha\nu} u^\lambda u^\alpha v^\mu v^\nu$, where u and v are arbitrary unit vectors on a Riemannian manifold Q less than $k < 0$, then, for every geodesic, the distance between two consecutive conjugate points is at most $\pi/\sqrt{k}$.

For example, let us consider a 1D motion described by the Lagrangian

$$L = \frac{1}{2}(\dot{q}^1)^2 - \phi(q^1),$$

where ϕ is a potential. The corresponding Lagrangian equation is equivalent to the geodesic one on the 2D Euclidean space $\mathbb{R}^2$ with respect to the connection $\widetilde{K}$ whose non–zero component is $\widetilde{K}_0{}^1{}_0 = -\partial_1 \phi$. The curvature of $\widetilde{K}$ has the non–zero component

$$R_{10}{}^1{}_0 = \partial_1 \widetilde{K}_0{}^1{}_0 = -\partial_1^2 \phi.$$

Choosing the particular Riemannian metric g with components

$$\overline{g}_{11} = 1, \qquad \overline{g}_{01} = 0, \qquad \overline{g}_{00} = 1,$$

we come to the formula

$$\int_a^b [(\dot{q}^\mu \partial_\mu u^1)^2 - \partial_1^2 \phi (u^1)^2] dt = 0$$

for a Jacobi vector–field u, which vanishes at points a and b. Then we get that, if $\partial_1^2\phi < 0$ at points of c, this motion has no conjugate points. In particular, let us consider the oscillator $\phi = k(q^1)^2/2$. In this case, the sectional curvature is $R_{0101} = -k$, while the half–period of this oscillator is exactly $\pi/\sqrt{k}$.

Constraints

Recall that symplectic and Poisson manifolds give an adequate Hamiltonian formulation of classical and quantum conservative mechanics. This is also the case of *presymplectic Hamiltonian systems*. Recall that every presymplectic form can be represented as a pull–back of a symplectic form by a coisotropic imbedding (see e.g., [Got82, MS98]), a presymplectic Hamiltonian systems can be seen as *Dirac constraint systems* [CGI95, MS98]. An autonomous Lagrangian system also exemplifies a presymplectic Hamiltonian system where a presymplectic form is the exterior differential of the Poincaré–Cartan form, while a Hamiltonian is the energy function [CR93, LM96, MS98, MR92]. A generic example of conservative Hamiltonian mechanics is a regular Poisson manifold (Z, w) where a Hamiltonian is a real function $\mathcal{H}$ on Z. Given the corresponding Hamiltonian vector–field $\vartheta_{\mathcal{H}} = w^\sharp(df)$, the closed subbundle $\vartheta_{\mathcal{H}}(Z)$ of the tangent bundle TZ is an autonomous first–order dynamical equation on a manifold Z, called the Hamiltonian equations. The evolution equation on the Poisson algebra $C^\infty(Z)$ is the *Lie derivative* $\mathfrak{L}_{\vartheta_{\mathcal{H}}} f = \{\mathcal{H}, f\}$, expressed into the *Poisson bracket* of the Hamiltonian $\mathcal{H}$ and functions f on Z. However, this description cannot be extended in a straightforward manner to time–dependent mechanics subject to time–dependent transformations.

The existent formulations of time–dependent mechanics imply usually a preliminary splitting of a configuration space $Q = \mathbb{R} \times M$ and a *momentum phase–space manifold* $\Pi = \mathbb{R} \times Z$, where Z is a Poisson manifold [CF93, CLM94, EMR91, HL84, MFV90, LM93]. From the physical viewpoint, this means that a certain reference frame is chosen. In this case, the momentum phase–space Π is with the Poisson product of the zero Poisson structure on $\mathbb{R}$ and the Poisson structure on Z. A Hamiltonian is defined as a real function $\mathcal{H}$ on Π. The corresponding Hamiltonian vector–field $\vartheta_{\mathcal{H}}$ on Π is vertical with respect to the fibration $\Pi \to \mathbb{R}$. Due to the canonical imbedding

$$\Pi \times T\mathbb{R} \to T\Pi, \tag{2.57}$$

one introduces the vector–field

$$\gamma_{\mathcal{H}} = \partial_t + \vartheta_{\mathcal{H}}, \tag{2.58}$$

where ∂_t is the standard vector–field on $\mathbb{R}$ [HL84]. The first–order dynamical equation $\gamma_{\mathcal{H}}(\Pi) \subset T\Pi$ on the manifold Π plays the role of Hamiltonian equations. The evolution equation on the Poisson algebra $C^\infty(\Pi)$ is given by the *Lie derivative*

$$\mathfrak{L}_{\gamma_{\mathcal{H}}} f = \partial_t f + \{\mathcal{H}, f\}.$$

This is not the case of mechanical systems subject to time–dependent transformations. These transformations, including canonical and inertial frame transformations, violate the splitting $\mathbb{R} \times Z$. As a consequence, there is no canonical imbedding (2.57), and the vector–field (2.58) is not well defined. At the same time, one can treat the imbedding (2.57) as a trivial connection on the bundle $\Pi \longrightarrow \mathbb{R}$, while $\gamma_{\mathcal{H}}$ (2.58) is the sum of the horizontal lift onto Π of the vector–field ∂_t by this connection and of the vertical vector–field $\vartheta_{\mathcal{H}}$.

Let $Q \to \mathbb{R}$ be a fibre bundle coordinated by (t, q^i), and $J^1(\mathbb{R}, Q)$ its 1–jet space, equipped with the adapted coordinates (t, q^i, q^i_t). Recall that there is a canonical imbedding λ given by (1.220) onto the affine subbundle of $TQ \to Q$ of elements $\upsilon \in TQ$ such that $\upsilon \rfloor dt = 1$. This subbundle is modelled over the vertical tangent bundle $VQ \to Q$. As a consequence, there is a 1–1 correspondence between the connections Γ on the fibre bundle $Q \to \mathbb{R}$, treated as sections of the affine jet bundle $\pi^1_0 : J^1(\mathbb{R}, Q) \to Q$ [MOS99], and the nowhere vanishing vector–fields $\Gamma = \partial_t + \Gamma^i \partial_i$ on Q, called horizontal vector–fields, such that $\Gamma \rfloor dt = 1$ [MS98, MOS99]. The corresponding *covariant differential* reads

$$D_\Gamma = \lambda - \Gamma : J^1(\mathbb{R}, Q) \longrightarrow VQ, \qquad q^i \circ D_\Gamma = q^i_t - \Gamma^i.$$

Let us also recall the total derivative $d_t = \partial_t + q^i_t \partial_i + \cdots$ and the exterior algebra homomorphism

$$h_0 : \phi dt + \phi_i dy^i \mapsto (\phi + \phi_i q^i_t) dt \tag{2.59}$$

which sends exterior forms on $Q \to \mathbb{R}$ onto the horizontal forms on $J^1(\mathbb{R}, Q) \to \mathbb{R}$, and vanishes on contact forms $\theta^i = dy^i - q^i_t dt$.

Lagrangian time–dependent mechanics follows directly Lagrangian field theory (see [Gia92, Kru97, LMM97, MS98, MP94], as well as subsection 2.2.2 below). This means that we have a configuration space $Q \to \mathbb{R}$ of a mechanical system, and a Lagrangian is defined as a *horizontal density* on the velocity phase–space $J^1(\mathbb{R}, Q)$,

$$L = \mathcal{L} dt, \qquad \text{with} \qquad \mathcal{L} : J^1(\mathbb{R}, Q) \to \mathbb{R}. \tag{2.60}$$

A generic *momentum phase–space manifold* of time–dependent mechanics is a fibre bundle $\Pi \longrightarrow \mathbb{R}$ with a regular Poisson structure whose characteristic distribution belongs to the vertical tangent bundle $V\Pi$ of $\Pi \longrightarrow \mathbb{R}$ [HL84]. However, such a Poisson structure cannot give dynamical equations. A first–order dynamical equation on $\Pi \longrightarrow \mathbb{R}$, by definition, is a section of the affine jet bundle $J^1(\mathbb{R}, \Pi) \longrightarrow \Pi$, i.e., a connection on $\Pi \longrightarrow \mathbb{R}$. Being a horizontal vector–field, such a connection cannot be a Hamiltonian vector–field with respect to the above Poisson structure on Π.

One can overcome this difficulty as follows. Let $Q \to \mathbb{R}$ be a configuration bundle of time–dependent mechanics. The corresponding momentum phase–space is the vertical cotangent bundle $\Pi = V^*Q \to \mathbb{R}$, called the Legendre bundle, while the cotangent bundle T^*Q is the homogeneous momentum phase–space. T^*Q admits the canonical Liouville form Ξ and the

symplectic form $d\Xi$, together with the corresponding non–degenerate Poisson bracket $\{,\}_T$ on the ring $C^\infty(T^*Q)$. Let us consider the subring of $C^\infty(T^*Q)$ which comprises the pull–backs $\zeta^* f$ onto T^*Q of functions f on the vertical cotangent bundle V^*Q by the canonical fibration

$$\zeta : T^*Q \to V^*Q. \tag{2.61}$$

This subring is closed under the Poisson bracket $\{,\}_T$, and V^*Q admits the regular Poisson structure $\{,\}_V$ such that [Vai94]

$$\zeta^*\{f,g\}_V = \{\zeta^* f, \zeta^* g\}_T.$$

Its characteristic distribution coincides with the vertical tangent bundle VV^*Q of $V^*Q \to \mathbb{R}$. Given a section h of the bundle (2.61), let us consider the pull–back forms

$$\Theta = h^*(\Xi \wedge dt), \qquad \Omega = h^*(d\Xi \wedge dt) \tag{2.62}$$

on V^*Q, but these forms are independent of a section h and are canonical exterior forms on V^*Q. The pull–backs $h^*\Xi$ are called the Hamiltonian forms. With Ω, the Hamiltonian vector–field ϑ_f for a function f on V^*Q is given by the relation

$$\vartheta_f \rfloor \Omega = -df \wedge dt,$$

while the Poisson bracket (2.1.2) is written as

$$\{f,g\}_V dt = \vartheta_g \rfloor \vartheta_f \rfloor \Omega.$$

The pair (V^*Q, Ω) is the particular *polysymplectic phase–space* of the covariant Hamiltonian field theory (see [CCI91, GMS97, Got91c, Sar95] for a survey). Following its general scheme, we can formulate the Hamiltonian time–dependent mechanics as follows [MS98, Sar98].

Recall that connection γ on the Legendre bundle $V^*Q \longrightarrow \mathbb{R}$ is called *canonical* if the corresponding horizontal vector–field is canonical for the Poisson structure on V^*Q, i.e., the form $\gamma \rfloor \Omega$ is closed. Such a form is necessarily exact. A canonical connection γ is a said to be a *Hamiltonian connection* if

$$\gamma \rfloor \Omega = dH, \tag{2.63}$$

where H is a *Hamiltonian form* on V^*Q. Every Hamiltonian form admits a unique Hamiltonian connection γ_H, while any canonical connection is locally a Hamiltonian one. Given a Hamiltonian form H, the kernel of the covariant differential D_{γ_H}, associated with the Hamiltonian connection γ_H, is a closed imbedded subbundle of the jet bundle $J^1(\mathbb{R}, V^*Q) \longrightarrow \mathbb{R}$, and so is the system of first–order PDEs on the Legendre bundle $V^*Q \longrightarrow \mathbb{R}$. These are the *Hamiltonian equations* in time–dependent mechanics, while the *Lie derivative*

$$\mathfrak{L}_{\gamma_H} f = \gamma_H \rfloor df \tag{2.64}$$

defines the evolution equation on $C^\infty(V^*Z)$. This Hamiltonian dynamics is equivalent to the Lagrangian one for hyperregular Lagrangians, while a degenerate Lagrangian involves a set of associated Hamiltonian forms in order to exhaust solutions of the Lagrangian equations [GMS97, Sar94, Sar95].

Since γ_H is not a vertical vector–field, the r.h.s. of the evolution equation (2.64) is not expressed into the Poisson bracket in a canonical way, but contains a frame–dependent term. Every connection Γ on the configuration bundle $Q \to \mathbb{R}$ is an affine section of the bundle (2.61), and defines the Hamiltonian form $H_\Gamma = \Gamma^*\Xi$ on V^*Q. The corresponding Hamiltonian connection is the canonical lift $V^*\Gamma$ of Γ onto the Legendre bundle V^*Q [GMS97, MOS99]. Then any Hamiltonian form H on V^*Q admits splittings

$$H = H_\Gamma - \widetilde{\mathcal{H}}_\Gamma dt, \qquad \text{with} \qquad \gamma_H = V^*\Gamma + \vartheta_{\widetilde{\mathcal{H}}_\Gamma}, \tag{2.65}$$

where $\vartheta_{\widetilde{\mathcal{H}}_\Gamma}$ is the vertical Hamiltonian field for the function $\widetilde{\mathcal{H}}_\Gamma$, which the energy function with respect to the reference frame Γ. With the splitting (2.65), the evolution equation (2.64) takes the form

$$\mathfrak{L}_{\gamma_H} f = V^*\Gamma \rfloor H + \{\widetilde{\mathcal{H}}_\Gamma, f\}_V. \tag{2.66}$$

Let the configuration bundle $Q \longrightarrow \mathbb{R}$ with an mD typical fibre M be coordinated by (t, q^i). Then Legendre bundle V^*Q and the cotangent bundle T^*Q admit holonomic coordinates $(t, q^i, p_i = \dot{q}_i)$ and (t, q^i, p_i, p), respectively. Relative to these coordinates, a *Hamiltonian form* H on V^*Q reads

$$H = h^*\Xi = p_i dq^i - \mathcal{H} dt. \tag{2.67}$$

It is the well–known integral *invariant of Poincaré–Cartan*, where $\mathcal{H}$ is a Hamiltonian in time–dependent mechanics. The expression (2.67) shows that $\mathcal{H}$ fails to be a scalar under time–dependent transformations. Therefore, the evolution equation (2.66) takes the local form

$$\mathfrak{L}_{\gamma_H} = \partial_t f + \{\mathcal{H}, f\}_V, \tag{2.68}$$

but one should bear in mind that the terms in its r.h.s., taken separately, are not well–behaved objects under time–dependent transformations. In particular, the equality $\{\mathcal{H}, f\}_V = 0$ is not preserved under time–dependent transformations.

Every Lagrangian L defines the *Legendre map*

$$\widehat{L} : J^1(\mathbb{R}, Q) \longrightarrow V^*Q, \qquad \text{locally given by} \qquad p_i \circ \widehat{L} = \pi_i, \tag{2.69}$$

whose image $N_L = \widehat{L}(J^1(\mathbb{R}, Q)) \subset V^*Q$ is called the *Lagrangian constraint space*. We state the comprehensive relationship between solutions of the Lagrangian equations for an almost regular Lagrangian L and solutions in N_L of the Hamiltonian equations for associated Hamiltonian forms.

Lagrangian Time–Dependent Dynamics

Given a Lagrangian L on the velocity phase–space $J^1(\mathbb{R}, Q)$, we follow the *first variational formula* of [GMS97, MS98, Sar97], which gives the canonical decomposition of the Lie derivative

$$\mathfrak{L}_{j^1 u} L = (j^1 u \rfloor \mathcal{L})\, dt \tag{2.70}$$

of L along a projectable vector–field u on $Q \longrightarrow \mathbb{R}$. We have

$$j^1 u \rfloor \mathcal{L} = u_V \rfloor \mathcal{E}_L + d_t(u \rfloor H_L),$$

where $u_V = (u \rfloor \theta^i)\partial_i$,

$$H_L = L + \pi_i \theta^i, \quad \pi_i = \partial_i^t \mathcal{L}, \tag{2.71}$$

is the Poincaré–Cartan form and

$$\mathcal{E}_L : J^2(\mathbb{R}, Q) \longrightarrow V^*Q, \qquad \mathcal{E}_L = (\partial_i - d_t \pi_i)\mathcal{L}\overline{d}q^i$$

is the *Euler–Lagrangian operator* associated with L. The kernel $\operatorname{Ker} \mathcal{E}_L \subset J^2(\mathbb{R}, Q)$ of $\mathcal{E}_L$ defines the Lagrangian equations on Q, given by the coordinate relations

$$(\partial_i - d_t \pi_i)\mathcal{L} = 0. \tag{2.72}$$

On–shell, the first variational formula (2.70) leads to the weak identity

$$\mathfrak{L}_{j^1 u} L \approx d_t(u \rfloor H_L) dt,$$

and then, if $\mathfrak{L}_{j^1 u} L = 0$, to the weak conservation law

$$d_t(u \rfloor H_L) = -d_t \mathcal{J} \approx 0$$

of the symmetry current $\mathcal{J}$, given by

$$\mathcal{J} = -(u \rfloor H_L) = -\pi_i(u^t \dot{q}^i - u^i) - u^t \mathcal{L}.$$

Being the *Lepagean equivalent* of the Lagrangian L on $J^1(\mathbb{R}, Q)$ (i.e., $L = h_0(H_L)$ where h_0 is the map (2.59), see [GMS97, MS98, Sar97]), the Poincaré–Cartan form H_L (2.71) is also the Lepagean equivalent of the Lagrangian on the repeated jet space $J^1(\mathbb{R}, J^1(\mathbb{R}, Q))$,

$$\overline{L} = \widehat{h}_0(H_L) = (\mathcal{L} + (\widehat{\dot{q}^i} - \dot{q}^i)\pi_i)dt, \qquad \widehat{h}_0(dy^i) = \widehat{q}_t^i dt,$$

coordinated by $(t, q^i, \dot{q}^i, \widehat{\dot{q}^i}, \ddot{q}^i)$. The Euler–Lagrangian operator $\mathcal{E}_{\overline{L}} : J^1(\mathbb{R}, J^1(\mathbb{R}, Q)) \to V^*J^1(\mathbb{R}, Q)$ for $\overline{L}$ is locally given by

$$\begin{aligned} &\mathcal{E}_{\overline{L}} = (\partial_i \mathcal{L} - \widehat{d}_t \pi_i + \partial_i \pi_j (\widehat{\dot{q}^j} - \dot{q}^j))\overline{d}q^i + \partial_i^t \pi_j (\widehat{\dot{q}^j} - \dot{q}^j)\overline{d}\dot{q}^i, \qquad (2.73) \\ &\text{with} \qquad \widehat{d}_t = \partial_t + \widehat{\dot{q}^i}\partial_i + \ddot{q}^i \partial_i^t. \end{aligned}$$

Its kernel $\operatorname{Ker}\mathcal{E}_{\overline{L}} \subset J^1(\mathbb{R}, J^1(\mathbb{R}, Q))$ defines the *Cartan equations*

$$\partial_i^t \pi_j(\widehat{\dot{q}^j} - \dot{q}^j) = 0, \qquad \partial_i \mathcal{L} - \widehat{d}_t \pi_i + (\widehat{\dot{q}^j} - \dot{q}^j)\partial_i \pi_j = 0. \tag{2.74}$$

Since $\mathcal{E}_{\overline{L}} \mid_{J^2(\mathbb{R},Q)} = \mathcal{E}_L$, the Cartan equations (2.74) are equivalent to the Lagrangian equations (2.72) on integrable sections $\overline{c} = \dot{c}$ of $J^1(\mathbb{R}, Q) \to \mathbb{R}$. These equations are equivalent in the case of regular Lagrangians.

On sections $\overline{c} : \mathbb{R} \longrightarrow J^1(\mathbb{R}, Q)$, the Cartan equations (2.74) are equivalent to the relation

$$\overline{c}^*(u \rfloor dH_L) = 0, \tag{2.75}$$

which is assumed to hold for all vertical vector–fields u on $J^1(\mathbb{R}, Q) \longrightarrow \mathbb{R}$.

With the Poincaré–Cartan form H_L (2.71), we have the *Legendre map*

$$\widehat{H}_L : J^1(\mathbb{R}, Q) \longrightarrow T^*Q, \qquad (p_i, p) \circ \widehat{H}_L = (\pi_i, \mathcal{L} - \pi_i \dot{q}^i).$$

Let $Z_L = \widehat{H}_L(J^1(\mathbb{R}, Q))$ be an imbedded subbundle $i_L : Z_L \hookrightarrow T^*Q$ of $T^*Q \to Q$. It admits the pull–back de Donder form $i_L^* \Xi$. We have

$$H_L = \widehat{H}_L^* \Xi_L = \widehat{H}_L^*(i_L^* \Xi).$$

By analogy with the Cartan equations (2.75), the *Hamilton–de Donder equations* for sections $\overline{r}$ of $T^*Q \to \mathbb{R}$ are written as

$$\overline{r}^*(u \rfloor d\Xi_L) = 0 \tag{2.76}$$

where u is an arbitrary vertical vector–field on $T^*Q \to \mathbb{R}$ [LM03].

Let the Legendre map $\widehat{H}_L : J^1(\mathbb{R}, Q) \longrightarrow Z_L$ be a submersion. Then a section $\overline{c}$ of $J^1(\mathbb{R}, Q) \longrightarrow \mathbb{R}$ is a solution of the Cartan equations (2.75) iff $\widehat{H}_L \circ \overline{c}$ is a solution of the Hamilton–de Donder equations (2.76), i.e., Cartan and Hamilton–de Donder equations are quasi–equivalent [Got91c, LM03].

Hamiltonian Time–Dependent Dynamics

Let the Legendre bundle $V^*Q \to \mathbb{R}$ be provided with the holonomic coordinates $(t, q^i, \dot{q}^i)$. Relative to these coordinates, the *canonical 3–form* Ω (2.62) and the *canonical Poisson structure* on V^*Q read

$$\Omega = dp_i \wedge dq^i \wedge dt, \tag{2.77}$$

$$\{f, g\}_V = \partial^i f \partial_i g - \partial^i g \partial_i f, \qquad (f, g \in C^\infty V^*Q). \tag{2.78}$$

The corresponding *symplectic foliation* coincides with the fibration $V^*Q \to \mathbb{R}$. The symplectic forms on the fibres of $V^*Q \to \mathbb{R}$ are the pull–backs $\Omega_t = dp_i \wedge dq^i$ of the canonical symplectic form on the typical fibre T^*M of the Legendre bundle $V^*Q \to \mathbb{R}$ with respect to trivialization maps [CR89, HL84, Sar98]. Given such a trivialization, the Poisson structure (2.78) is isomorphic

to the product of the zero Poisson structure on $\mathbb{R}$ and the canonical symplectic structure on T^*M.

An automorphism ρ of the Legendre bundle $V^*Q \to \mathbb{R}$ is a canonical transformation of the Poisson structure (2.78) iff it preserves the canonical 3–form $\mathbf{\Omega}$ (2.77). Let us emphasize that canonical transformations are compatible with the fibration $V^*Q \to \mathbb{R}$, but not necessarily with the fibration $\pi_Q : V^*Q \to Q$.

With respect to the Poisson bracket (2.78), the Hamiltonian vector–field ϑ_f for a function f on the *momentum phase–space manifold* V^*Q is given by

$$\vartheta_f = \partial^i f \partial_i - \partial_i f \partial^i.$$

A Hamiltonian vector–field, by definition, is canonical. Conversely, every vertical canonical vector–field on the Legendre bundle $V^*Q \longrightarrow \mathbb{R}$ is locally a Hamiltonian vector–field.

To prove this, let σ be a one–form on V^*Q. If $\sigma \wedge dt$ is closed form, it is exact. Since V^*Q is diffeomorphic to $\mathbb{R} \times T^*M$, we have the *de Rham cohomology group*

$$H^2(V^*Q) = H^0(\mathbb{R}) \otimes H^2(T^*M) \oplus H^1(\mathbb{R}) \otimes H^1(T^*M).$$

The form $\sigma \wedge dt$ belongs to its second item which is zero.

If the two–form $\sigma \wedge dt$ is exact, then $\sigma \wedge dt = dg \wedge dt$ locally [GMS97].

Let $\gamma = \partial_t + \gamma^i \partial_i + \gamma_i \partial^i$ be a canonical connection on the Legendre bundle $V^*Q \longrightarrow \mathbb{R}$. Its components obey the relations

$$\partial^i \gamma^j - \partial^j \gamma^i = 0, \qquad \partial_i \gamma_j - \partial_j \gamma_i = 0, \qquad \partial_j \gamma^i + \partial^i \gamma_j = 0.$$

Canonical connections constitute an affine space modelled over the vector space of vertical canonical vector–fields on $V^*Q \longrightarrow \mathbb{R}$.

If γ is a canonical connection, then the form $\gamma \rfloor \Omega$ is exact. Every connection Γ on $Q \to \mathbb{R}$ induces the connection on $V^*Q \to \mathbb{R}$,

$$V^*\Gamma = \partial_t + \Gamma^i \partial_i - p_i \partial_j \Gamma^i \partial^j,$$

which is a *Hamiltonian connection* for the frame *Hamiltonian form*

$$V^*\Gamma \rfloor \Omega = dH_\Gamma, \qquad H_\Gamma = p_i dq^i - p_i \Gamma^i dt. \tag{2.79}$$

Thus, every canonical connection γ on V^*Q defines an exterior one–form H modulo closed forms so that $dH = \gamma \rfloor \Omega$. Such a form is called a *locally Hamiltonian form*.

Every locally Hamiltonian form on the momentum phase–space V^*Q is locally a Hamiltonian form modulo closed forms. Given locally Hamiltonian forms H_γ and $H_{\gamma'}$, their difference $\sigma = H_\gamma - H_{\gamma'}$ is a one–form on V^*Q such that the two–form $\sigma \wedge dt$ is closed. The form $\sigma \wedge dt$ is exact and $\sigma = f dt + dg$ locally. Put $H_{\gamma'} = H_\Gamma$ where Γ is a connection on $V^*Q \longrightarrow \mathbb{R}$. Then H_γ

modulo closed forms takes the local form $H_\gamma = H_\Gamma + f dt$, and coincides with the pull–back of the Liouville form Ξ on T^*Q by the local section $p = -p_i\Gamma^i + f$ of the fibre bundle (2.61).

Conversely, each Hamiltonian form H on the momentum phase–space V^*Q admits a unique canonical connection γ_H on $V^*Q \to \mathbb{R}$ such that the relation (2.63) holds. Given a Hamiltonian form H, its exterior differential

$$dH = h^*d\Xi = (dp_i + \partial_i H dt) \wedge (dq^i - \partial^i H dt)$$

is a presymplectic form of constant rank $2m$ since the form

$$(dH)^m = (dp_i \wedge dq^i)^m - m(dp_i \wedge dq^i)^{m-1} \wedge dH \wedge dt$$

is nowhere vanishing. It is also seen that $(dH)^m \wedge dt \neq 0$. It follows that the kernel of dH is a 1D distribution. Then the desired *Hamiltonian connection*

$$\gamma_H = \partial_t + \partial^i H \partial_i - \partial_i H \partial^i \tag{2.80}$$

is a unique vector–field γ_H on V^*Q such that $\gamma_H \rfloor dH = 0$ and $\gamma_H \rfloor dt = 1$.

Hamiltonian forms constitute an affine space modelled over the vector space of horizontal densities $f dt$ on $V^*Q \to \mathbb{R}$, i.e., over $C^\infty(V^*Q)$. Therefore Hamiltonian connections γ_H form an affine space modelled over the vector space of Hamiltonian vector–fields. Every Hamiltonian form H defines the associated Hamiltonian map

$$\widehat{H} = j^1\pi_Q \circ \gamma_H : \partial_t + \partial^i\mathcal{H} : V^*Q \to J^1(\mathbb{R}, Q). \tag{2.81}$$

With the Hamiltonian map (2.81), we have another Hamiltonian form

$$H_{\widehat{H}} = -\widehat{H} \rfloor \Theta = p_i dq^i - p_i \partial^i \mathcal{H}.$$

Note that $H_{\widehat{H}} = H$ iff H is a frame Hamiltonian form.

Given a Hamiltonian connection γ_H (2.80), the corresponding Hamiltonian equations $D_{\gamma_H} = 0$ take the coordinate form

$$\dot{q}^i = \partial^i \mathcal{H}, \qquad \dot{p}_i = -\partial_i \mathcal{H}. \tag{2.82}$$

Their classical solutions are integral sections of the Hamiltonian connection γ_H, i.e., $\dot{r} = \gamma_H \circ r$. On sections r of the Legendre bundle $V^*Q \to \mathbb{R}$, the Hamiltonian equations (2.82) are equivalent to the relation

$$r^*(u \rfloor dH) = 0 \tag{2.83}$$

which is assumed to hold for any vertical vector–field u on $V^*Q \to \mathbb{R}$.

A Hamiltonian form H (2.79) is the Poincaré–Cartan form for the Lagrangian on the jet space $J^1(\mathbb{R}, V)^*Q$,

$$L_H = h_0(H) = (p_i\dot{q}^i - \mathcal{H})\omega. \tag{2.84}$$

Given a projectable vector–field u on the configuration bundle $Q \longrightarrow \mathbb{R}$ and its lift onto the Legendre bundle $V^*Q \longrightarrow \mathbb{R}$,

$$\widetilde{u} = u^t\partial_t + u^i\partial_i - \partial_i u^j p_j \partial^i, \qquad \text{we have}$$
$$\mathfrak{L}_{\widetilde{u}} H = \mathfrak{L}_{J^1\widetilde{u}} L_H. \tag{2.85}$$

Note that the Hamiltonian equations (2.82) for H are exactly the Lagrangian equations for L_H, i.e., they characterize the kernel of the Euler–Lagrangian operator

$$\mathcal{E}_H : J^1(\mathbb{R}, V)^*Q \to V^*V^*Q, \qquad \mathcal{E}_H = (\dot{q}^i - \partial^i\mathcal{H})\overline{d}p_i - (\dot{p}_i + \partial_i\mathcal{H})\overline{d}q^i$$

for the Lagrangian L_H, called the Hamiltonian operator for H.

Using the relation (2.85), let us get the *Hamiltonian conservation laws* in time–dependent mechanics. As in field theory, by gauge transformations in time–dependent mechanics are meant automorphism of the configuration bundle $Q \to \mathbb{R}$, but only over translations of the base $\mathbb{R}$. Then, projectable vector–fields on $V^*Q \to \mathbb{R}$,

$$u = u^t\partial_t + u^i\partial_i, \qquad u\rfloor dt = u^t = \text{const}, \tag{2.86}$$

can be seen as generators of local 1–parameter groups of local gauge transformations. Given a Hamiltonian form H (2.126), its Lie derivative (2.85) reads

$$\mathfrak{L}_{\widetilde{u}} H = \mathfrak{L}_{j^1\widetilde{u}} L_H = (-u^t\partial_t\mathcal{H} + p_i\partial_t u^i - u^i\partial_i\mathcal{H} + \partial_j u^i p_i\partial^j\mathcal{H})dt. \tag{2.87}$$

The first variational formula (2.70) applied to the Lagrangian L_H (2.84) leads to the weak identity $\mathfrak{L}_{\widetilde{u}} H \approx d_t(u\rfloor H)dt$. If the Lie derivative (2.87) vanishes, we have the *conserved symmetry current*

$$\mathcal{J}_u = u\rfloor dH = p_i u^i - u^t\mathcal{H}, \tag{2.88}$$

along u. Every vector–field (2.86) is a superposition of a vertical vector–field and a reference frame on $Q \to \mathbb{R}$. If u is a vertical vector–field, $\mathcal{J}_u$ is the Nöther current

$$\mathcal{J}_u(q) = u\rfloor q = p_i u^i, \qquad (q = p_i\overline{d}q^i \in V^*Q). \tag{2.89}$$

The symmetry current along a reference frame Γ, given by

$$\mathcal{J}_\Gamma = p_i\Gamma^i - \mathcal{H} = -\widetilde{\mathcal{H}}_\Gamma,$$

is the energy function with respect to the reference frame Γ, taken with the sign minus [EMR95, MS98, Sar98]. Given a Hamiltonian form H, the energy functions $\widetilde{\mathcal{H}}_\Gamma$ constitute an affine space modelled over the vector space of Nöther currents. Also, given a Hamiltonian form H, the conserved currents (2.88) form a Lie algebra with respect to the Poisson bracket

$$\{\mathcal{J}_u, \mathcal{J}_{u'}\}_V = \mathcal{J}_{[u,u']}.$$

The second of the above constructions enables us to represent the r.h.s. of the evolution equation (2.68) as a pure Poisson bracket. Given a Hamiltonian form $H = h^*\Xi$, let us consider its pull–back $\zeta^* H$ onto the cotangent bundle T^*Q. Note that the difference $\Xi - \zeta^* H$ is a horizontal one–form on $T^*Q \to \mathbb{R}$, while

$$\mathcal{H}^* = \partial_t \rfloor (\Xi - \zeta^* H) = p + \mathcal{H} \tag{2.90}$$

is a function on T^*Q. Then the relation

$$\zeta^*(\mathfrak{L}_{\gamma_H} f) = \{\mathcal{H}^*, \zeta^* f\}_T \tag{2.91}$$

holds for every function $f \in C^\infty(V^*Q)$. In particular, given a projectable vector–field u (2.86), the symmetry current $\mathcal{J}_u$ (2.88) is conserved iff

$$\{\mathcal{H}^*, \zeta^* \mathcal{J}_u\}_T = 0.$$

Moreover, let $\vartheta_{\mathcal{H}^*}$ be the Hamiltonian vector–field for the function $\mathcal{H}^*$ (2.90) with respect to the canonical Poisson structure $\{,\}_T$ on T^*Q. Then

$$T\zeta(\vartheta_{\mathcal{H}^*}) = \gamma_H.$$

Time–Dependent Constraints

The relation (2.91) enables us to extend the constraint algorithm of conservative mechanics and time–dependent mechanics on a product $\mathbb{R} \times M$ (see [CLM94, LM93]) to mechanical systems subject to time–dependent transformations.

Let H be a Hamiltonian form on the *momentum phase–space manifold* V^*Q. In accordance with the relation (2.91), a constraint $f \in I_N$ is preserved if the bracket in (2.91) vanishes. It follows that the solutions of the Hamiltonian equations (2.82) do not leave the constraint space N if

$$\{\mathcal{H}^*, \zeta^* I_N\}_T \subset \zeta^* I_N. \tag{2.92}$$

If the relation (2.92) fails to hold, let us introduce secondary constraints $\{\mathcal{H}^*, \zeta^* f\}_T$, $f \in I_N$, which belong to $\zeta^* C^\infty(V^*Q)$. If the collection of primary and secondary constraints is not closed with respect to the relation (2.92), let us add the tertiary constraints $\{\mathcal{H}^*, \{\mathcal{H}^*, \zeta^* f_a\}_T\}_T$ and so on.

Let us assume that N is a final constraint space for a Hamiltonian form H. If a Hamiltonian form H satisfies the relation (2.92), so is a Hamiltonian form

$$H_f = H - f dt, \tag{2.93}$$

where $f \in I'_N$ is a first class constraint. Though Hamiltonian forms H and H_f coincide with each other on the constraint space N, the corresponding

Hamiltonian equations have different solutions on the constraint space N because $dH|_N \neq dH_f|_N$. At the same time, $d(i_N^* H) = d(i_N^* H_f)$. Therefore, let us introduce the constrained Hamiltonian form

$$H_N = i_N^* H_f, \tag{2.94}$$

which is the same for all $f \in I'_N$. Note that H_N (2.94) is not a true Hamiltonian form on $N \longrightarrow \mathbb{R}$ in general. On sections r of the fibre bundle $N \longrightarrow \mathbb{R}$, we can write the equations

$$r^*(u_N \rfloor dH_N) = 0, \tag{2.95}$$

where u_N is an arbitrary vertical vector–field on $N \longrightarrow \mathbb{R}$. They are called the constrained Hamiltonian equations.

For any Hamiltonian form H_f (2.93), every solution of the Hamiltonian equations which lives in the constraint space N is a solution of the constrained Hamiltonian equations (2.95). The constrained Hamiltonian equations can be written as

$$r^*(u_N \rfloor di_N^* H_f) = r^*(u_N \rfloor dH_f|_N) = 0. \tag{2.96}$$

They differ from the Hamiltonian equations (2.83) for H_f restricted to N which read

$$r^*(u \rfloor dH_f|_N) = 0, \tag{2.97}$$

where r is a section of $N \to \mathbb{R}$ and u is an arbitrary vertical vector–field on $V^*Q \to \mathbb{R}$. A solution r of the equations (2.97) satisfies the weaker condition (2.96).

One can also consider the problem of constructing a generalized Hamiltonian system, similar to that for Dirac constraint system in conservative mechanics [MS98]. Let H satisfies the condition $\{\mathcal{H}^*, \zeta^* I'_N\}_T \subset I_N$, whereas $\{\mathcal{H}^*, \zeta^* I'_N\}_T \not\subset I_N$. The goal is to find a constraint $f \in I_N$ such that the modified Hamiltonian $H - fdt$ would satisfy both the conditions

$$\{\mathcal{H}^* + \zeta^* f, \zeta^* I'_N\}_T \subset \zeta^* I_N, \qquad \{\mathcal{H}^* + \zeta^* f, \zeta^* I_N\}_T \subset \zeta^* I_N.$$

The first of them is fulfilled for any $f \in I_N$, while the latter is an equation for a second–class constraint f.

Note that, in contrast with the conservative case, the Hamiltonian vector–fields ϑ_f for the first class constraints $f \in I'_N$ in time–dependent mechanics are not generators of gauge symmetries of a Hamiltonian form in general. At the same time, generators of gauge symmetries define an ideal of constraints as follows.

Lagrangian Constraints

Let us consider the Hamiltonian description of Lagrangian mechanical systems on a configuration bundle $Q \to \mathbb{R}$. If a Lagrangian is degenerate, we have the Lagrangian constraint subspace of the Legendre bundle V^*Q and a set of

Hamiltonian forms associated with the same Lagrangian. Given a Lagrangian L (2.60) on the velocity phase–space $J^1(\mathbb{R},Q)$, a Hamiltonian form H on the momentum phase–space V^*Q is said to be associated with L if H satisfies the relations

$$\widehat{L}\circ\widehat{H}\circ\widehat{L}=\widehat{L}, \qquad \text{and} \qquad H=H_{\widehat{H}}+\widehat{H}^*L \tag{2.98}$$

where $\widehat{H}$ and $\widehat{L}$ are the Hamiltonian map (2.81) and the Legendre map (2.69), respectively. Here, $\widehat{L}\circ\widehat{H}$ is the projector

$$p_i(z)=\pi_i(t,q^i,\partial^j\mathcal{H}(z)), \qquad (z\in N_L), \tag{2.99}$$

from Π onto the Lagrangian constraint space $N_L=\widehat{L}(J^1(\mathbb{R},Q))$. Therefore, $\widehat{H}\circ\widehat{L}$ is the projector from $J^1(\mathbb{R},Q)$ onto $\widehat{H}(N_L)$. A Hamiltonian form is called weakly associated with a Lagrangian L if the condition (2.98) holds on the Lagrangian constraint space N_L.

If a bundle map $\Phi: V^*Q\to J^1(\mathbb{R},Q)$ obeys the relation (2.98), then the Hamiltonian form $H=-\Phi\rfloor\Theta+\Phi^*L$ is weakly associated with the Lagrangian L. If $\Phi=\widehat{H}$, then H is associated with L [GMS97].

Any Hamiltonian form H weakly associated with a Lagrangian L obeys the relation

$$H|_{N_L}=\widehat{H}^*H_L|_{N_L}, \tag{2.100}$$

where H_L is the Poincaré–Cartan form (2.71). The relation (2.98) takes the coordinate form

$$\mathcal{H}(z)=p_i\partial^i\mathcal{H}-\mathcal{L}(t,q^i,\partial^j\mathcal{H}(z)), \qquad (z\in N_L). \tag{2.101}$$

Substituting (2.99) and (2.101) in (2.126), we get the relation (2.100).

The difference between *associated* and *weakly associated Hamiltonian forms* lies in the following. Let H be an associated Hamiltonian form, i.e., the equality (2.101) holds everywhere on V^*Q. The exterior differential of this equality leads to the relations

$$\partial_t\mathcal{H}(z)=-(\partial_t\mathcal{L})\circ\widehat{H}(z), \qquad \partial_i\mathcal{H}(z)=-(\partial_i\mathcal{L})\circ\widehat{H}(z),$$
$$(p_i-(\partial^t_i\mathcal{L})(t,q^i,\partial^j_t\mathcal{H}))\partial^i_t\partial^a_t\mathcal{H}=0, \qquad (z\in N_L).$$

The last of them shows that the Hamiltonian form is not regular outside the Lagrangian constraint space N_L. In particular, any Hamiltonian form is weakly associated with the Lagrangian $L=0$, while the associated Hamiltonian forms are only H_Γ.

Here we restrict our consideration to almost regular Lagrangians L, i.e., if: (i) the Lagrangian constraint space N_L is a closed imbedded subbundle $i_N: N_L\longrightarrow V^*Q$ of the bundle $V^*Q\longrightarrow Q$, (ii) the Legendre map $\widehat{L}: J^1(\mathbb{R},Q)\longrightarrow N_L$ is a fibre bundle, and (iii) the pre-image $\widehat{L}^{-1}(z)$ of any point $z\in N_L$ is a connected submanifold of $J^1(\mathbb{R},Q)$.

A Hamiltonian form H weakly associated with an almost regular Lagrangian L exists iff the fibre bundle $J^1(\mathbb{R}, V)^*Q \longrightarrow N_L$ admits a global section.

The condition (iii) leads to the following property [GMS97, MS98]. The Poincaré–Cartan form H_L for an almost regular Lagrangian L is constant on the connected pre–image $\widehat{L}^{-1}(z)$ of any point $z \in N_L$.

An immediate consequence of this fact is the following assertion [GMS97]. All Hamiltonian forms weakly associated with an almost regular Lagrangian L coincide with each other on the Lagrangian constraint space N_L, and the Poincaré–Cartan form H_L for L is the pull–back

$$H_L = \widehat{L}^* H, \qquad \pi_i \dot{q}^i - \mathcal{L} = \mathcal{H}(t, q^j, \pi_j),$$

of any such a Hamiltonian form H.

It follows that, given Hamiltonian forms H an H' weakly associated with an almost regular Lagrangian L, their difference is fdt, $(f \in I_N)$. Above proposition enables us to connect Lagrangian and Cartan equations for an almost regular Lagrangian L with the Hamiltonian equations for Hamiltonian forms weakly associated with L [GMS97].

Let a section r of $V^*Q \longrightarrow \mathbb{R}$ be a solution of the Hamiltonian equations (2.82) for a Hamiltonian form H weakly associated with an almost regular Lagrangian L. If r lives in the constraint space N_L, the section $c = \pi_Q \circ r$ of $Q \longrightarrow \mathbb{R}$ satisfies the Lagrangian equations (2.72), while $\overline{c} = \widehat{H} \circ r$ obeys the Cartan equations (2.74).

Given an almost regular Lagrangian L, let a section $\overline{c}$ of the jet bundle $J^1(\mathbb{R}, Q) \longrightarrow \mathbb{R}$ be a solution of the Cartan equations (2.74). Let H be a Hamiltonian form weakly associated with L, and let H satisfy the relation

$$\widehat{H} \circ \widehat{L} \circ \overline{c} = j^1(\pi_0^1 \circ \overline{c}). \tag{2.102}$$

Then, the section $r = \widehat{L} \circ \overline{c}$ of the Legendre bundle $V^*Q \longrightarrow \mathbb{R}$ is a solution of the Hamiltonian equations (2.82) for H. Since $\widehat{H} \circ \widehat{L}$ is a projection operator, the condition (2.102) implies that the solution $\overline{s}$ of the Cartan equations is actually an integrable section $\overline{c} = \dot{c}$ where c is a solution of the Lagrangian equations.

Given a Hamiltonian form H weakly associated with an almost regular Lagrangian L, let us consider the corresponding constrained Hamiltonian form H_N (2.94). H_N is the same for all Hamiltonian forms weakly associated with L, and $H_L = \widehat{L}^* H_N$.

For any Hamiltonian form H weakly associated with an almost regular Lagrangian L, every solution of the Hamiltonian equations which lives in the Lagrangian constraint space N_L is a solution of the constrained Hamiltonian equations (2.95).

Using the equality $H_L = \widehat{L}^* H_N$, one can show that the constrained Hamiltonian equations (2.95) are equivalent to the *Hamilton–de Donder equations* (2.76) and are quasi–equivalent to the Cartan equations (2.75) [GMS97, MS98, LM03].

Quadratic Degenerate Lagrangian Systems

Given a configuration bundle $Q \to \mathbb{R}$, let us consider a quadratic Lagrangian L which has the coordinate expression

$$\mathcal{L} = \frac{1}{2} a_{ij} \dot{q}^i \dot{q}^j + b_i \dot{q}^i + c, \tag{2.103}$$

where a, b and c are local functions on Q. This property is coordinate–independent due to the affine transformation law of the coordinates $\dot{q}^i$. The associated Legendre map

$$p_i \circ \widehat{L} = a_{ij} \dot{q}^j + b_i \tag{2.104}$$

is an affine map over Q. It defines the corresponding linear map

$$\overline{L} : VQ \longrightarrow V^*Q, \qquad p_i \circ \overline{L} = a_{ij} \dot{q}^j. \tag{2.105}$$

Let the Lagrangian L (2.103) be almost regular, i.e., the matrix function a_{ij} is of constant rank. Then the Lagrangian constraint space $N_L = \widehat{L}(J^1(\mathbb{R}, Q))$ is an affine subbundle of the bundle $V^*Q \longrightarrow Q$, modelled over the vector subbundle $\overline{N}_L$ (2.105) of $V^*Q \longrightarrow Q$. Hence, $N_L \longrightarrow Q$ has a global section. For the sake of simplicity, let us assume that it is the canonical zero section $\widehat{0}(Q)$ of $V^*Q \longrightarrow Q$. Then $\overline{N}_L = N_L$. Therefore, the kernel of the Legendre map (2.104) is an affine subbundle of the affine jet bundle $J^1(\mathbb{R}, Q) \longrightarrow Q$, modelled over the kernel of the linear map $\overline{L}$ (2.105). Then there exists a connection Γ on the fibre bundle $Q \longrightarrow \mathbb{R}$, given by

$$\Gamma : Q \longrightarrow \operatorname{Ker} \widehat{L} \subset J^1(\mathbb{R}, Q), \qquad \text{with} \qquad a_{ij} \Gamma^j_\mu + b_i = 0.$$

Connections Γ constitute an affine space modelled over the linear space of vertical vector–fields υ on $Q \longrightarrow \mathbb{R}$, satisfying the conditions

$$a_{ij} \upsilon^j = 0 \tag{2.106}$$

and, as a consequence, the conditions $\upsilon^i b_i = 0$. If the Lagrangian (2.103) is regular, the connection Γ is unique.

There exists a linear bundle map

$$\sigma : V^*Q \longrightarrow VQ, \qquad \dot{q}^i \circ \sigma = \sigma^{ij} p_j,$$

such that $\overline{L} \circ \sigma \circ i_N = i_N$. The map σ is a solution of the algebraic equations

$$a_{ij} \sigma^{jk} a_{kb} = a_{ib}.$$

There exist the bundle splitting

$$VQ = \operatorname{Ker} a \oplus E' \tag{2.107}$$

and a nonholonomic atlas of this bundle such that transition functions of $\mathrm{Ker}\, a$ and E' are independent. Since a is a non–degenerate fibre metric in E', there exists an atlas of E' such that a is brought into a diagonal matrix with non–vanishing components a_{AA}. Due to the splitting (2.107), we have the corresponding bundle splitting

$$V^*Q = (\mathrm{Ker}\, a)^* \oplus \mathrm{Im}\, a. \tag{2.108}$$

Then the desired map σ is represented by a direct sum $\sigma_1 \oplus \sigma_0$ of an arbitrary section σ_1 of the bundle $\vee^2 \mathrm{Ker}\, a^* \to Q$ and the section σ_0 of the bundle $\wedge^2 E' \to Q$, which has non–vanishing components $\sigma^{AA} = (a_{AA})^{-1}$ with respect to the above atlas of E'. Moreover, σ satisfies the particular relations

$$\sigma_0 = \sigma_0 \circ \overline{L} \circ \sigma_0, \quad a \circ \sigma_1 = 0, \quad \sigma_1 \circ a = 0. \tag{2.109}$$

The splitting (2.107) leads to the splitting

$$J^1(\mathbb{R}, Q) = \mathcal{S}(J^1(\mathbb{R}, Q)) \oplus \mathcal{F}(J^1(\mathbb{R}, Q)) = \mathrm{Ker}\, \widehat{L} \oplus \mathrm{Im}(\sigma \circ \widehat{L}), \tag{2.110}$$

$$\dot{q}^i = \mathcal{S}^i + \mathcal{F}^i = [\dot{q}^i - \sigma_0^{ik}(a_{kj}\dot{q}^j + b_k)] + [\sigma_0^{ik}(a_{kj}\dot{q}^j + b_k)], \tag{2.111}$$

while the splitting (2.108) can be written as

$$V^*Q = \mathcal{R}(V^*Q) \oplus \mathcal{P}(V^*Q) = \mathrm{Ker}\, \sigma_0 \oplus N_L, \tag{2.112}$$

$$p_i = \mathcal{R}_i + \mathcal{P}_i = [p_i - a_{ij}\sigma_0^{jk} p_k] + [a_{ij}\sigma_0^{jk} p_k]. \tag{2.113}$$

Note that, with respect to the coordinates $\mathcal{S}^i_\alpha$ and $\mathcal{F}^i_\alpha$ (2.111), the Lagrangian (2.103) reads

$$\mathcal{L} = \frac{1}{2} a_{ij}\mathcal{F}^i\mathcal{F}^j + c',$$

while the Lagrangian constraint space is given by the *reducible constraints*

$$\mathcal{R}_i = p_i - a_{ij}\sigma_0^{jk} p_k = 0.$$

Given the linear map σ and the connection Γ as defined above, let us consider the affine *Hamiltonian map*

$$\Phi = \widehat{\Gamma} + \sigma : V^*Q \longrightarrow J^1(\mathbb{R}, Q), \qquad \Phi^i = \Gamma^i + \sigma^{ij} p_j, \tag{2.114}$$

and the *Hamiltonian form*

$$H = H_\Phi + \Phi^* L = p_i dq^i - [p_i \Gamma^i + \frac{1}{2}\sigma_0{}^{ij} p_i p_j + \sigma_1{}^{ij} p_i p_j - c']dt \tag{2.115}$$

$$= (\mathcal{R}_i + \mathcal{P}_i)dq^i - [(\mathcal{R}_i + \mathcal{P}_i)\Gamma^i + \frac{1}{2}\sigma_0^{ij}\mathcal{P}_i\mathcal{P}_j + \sigma_1^{ij} p_i p_j - c']dt.$$

In particular, if σ_1 is non–degenerate, so is the Hamiltonian form H.

The Hamiltonian forms of the type H, parameterized by connections Γ, are weakly associated with the Lagrangian (2.103) and constitute a complete

set. Then H is weakly associated with L. Let us write the corresponding Hamiltonian equations (2.82) for a section r of the Legendre bundle $V^*Q \to \mathbb{R}$. They are

$$\dot{c} = (\widehat{\Gamma} + \sigma) \circ r, \qquad c = \pi_Q \circ r. \tag{2.116}$$

Due to the surjections $\mathcal{S}$ and $\mathcal{F}$ (2.111), the Hamiltonian equations (2.116) break in two parts

$$\mathcal{S} \circ \dot{c} = \Gamma \circ c, \qquad \dot{r}^i - \sigma^{ik}(a_{kj}\dot{r}^j + b_k) = \Gamma^i \circ c, \tag{2.117}$$

$$\mathcal{F} \circ \dot{c} = \sigma \circ r, \qquad \sigma^{ik}(a_{kj}\dot{r}^j + b_k) = \sigma^{ik} r_k. \tag{2.118}$$

Let c be an arbitrary section of $Q \to \mathbb{R}$, e.g., a solution of the Lagrangian equations. There exists a connection Γ such that the relation (2.117) holds, namely, $\Gamma = \mathcal{S} \circ \Gamma'$, where Γ' is a connection on $Q \to \mathbb{R}$ which has c as an integral section.

If $\sigma_1 = 0$, then $\Phi = \widehat{H}$ and the Hamiltonian forms H are associated with the Lagrangian (2.103). Thus, for different σ_1, we have different complete sets of Hamiltonian forms H, which differ from each other in the term $\upsilon^i \mathcal{R}_i$, where υ are vertical vector–fields (2.106). This term vanishes on the Lagrangian constraint space. The corresponding constrained Hamiltonian form $H_N = i_N^* H$ and the constrained Hamiltonian equations (2.95) can be written.

For every Hamiltonian form H, the Hamiltonian equations (2.82) and (2.118) restricted to the Lagrangian constraint space N_L are equivalent to the constrained Hamiltonian equations.

Due to the splitting (2.112), we have the corresponding splitting of the vertical tangent bundle $V_Q V^*Q$ of the bundle $V^*Q \to Q$. In particular, any vertical vector–field u on $V^*Q \to \mathbb{R}$ admits the decomposition

$$u = [u - u_{TN}] + u_{TN}, \qquad \text{with} \qquad u_{TN} = u^i \partial_i + a_{ij}\sigma_0^{jk} u_k \partial^i,$$

such that $u_N = u_{TN}|_{N_L}$ is a vertical vector–field on the Lagrangian constraint space $N_L \to \mathbb{R}$. Let us consider the equations

$$r^*(u_{TN} \rfloor dH) = 0$$

where r is a section of $V^*Q \to \mathbb{R}$ and u is an arbitrary vertical vector–field on $V^*Q \to \mathbb{R}$. They are equivalent to the pair of equations

$$r^*(a_{ij}\sigma_0^{jk}\partial^i \rfloor dH) = 0, \tag{2.119}$$

$$r^*(\partial_i \rfloor dH) = 0. \tag{2.120}$$

Restricted to the Lagrangian constraint space, the Hamiltonian equations for different Hamiltonian forms H associated with the same quadratic Lagrangian (2.103) differ from each other in the equations (2.117). These equations are independent of momenta and play the role of *gauge–type conditions*.

Time–Dependent Completely Integrable Hamiltonian Systems

Recall that the configuration space of a time–dependent mechanical system is a fibre bundle $M \to \mathbb{R}$ over the time axis $\mathbb{R}$ equipped with the bundle coordinates $q^\alpha \equiv (t, q^k)$, for $k = 1, \ldots, m$. The corresponding momentum phase–space is the vertical cotangent bundle V^*M of $M \to \mathbb{R}$ with holonomic bundle coordinates (t, q^k, p_k).

Recall that the cotangent bundle T^*M of M is coordinated by [MS98, GMS97]

$$(t, q^k, p_0 = p, p_k), \qquad p' = p + \frac{\partial q^k}{\partial t} p_k, \tag{2.121}$$

and plays the role of the homogeneous momentum phase–space of time–dependent mechanics. It admits the canonical Liouville form $\Xi = p_\alpha dq^\alpha$, the canonical symplectic form $\Omega_T = d\Xi$, and the corresponding Poisson bracket

$$\{f, f'\}_T = \partial^\alpha f \partial_\alpha f' - \partial_\alpha f \partial^\alpha f', \qquad (f, f' \in C^\infty(T^*M)). \tag{2.122}$$

There is a canonical 1D fibre bundle

$$\zeta : T^*M \to V^*M, \tag{2.123}$$

whose kernel is the annihilator of the vertical tangent bundle $VM \subset TM$. The transformation law (2.121) shows that it is a trivial affine bundle. Indeed, given a global section h of ζ, one can equip T^*M with the fibre coordinate $r = p - h$ possessing the identity transition functions.

The fibre bundle (2.123) gives the vertical cotangent bundle V^*M with the canonical Poisson structure $\{,\}_V$ such that

$$\zeta^*\{f, f'\}_V = \{\zeta^* f, \zeta^* f'\}_T, \tag{2.124}$$

$$\{f, f'\}_V = \partial^k f \partial_k f' - \partial_k f \partial^k f', \tag{2.125}$$

for all $f, f' \in C^\infty(V^*M)$. The corresponding symplectic foliation coincides with the fibration $V^*M \to \mathbb{R}$.

However, the Poisson structure (2.125) fails to give any dynamical equation on the momentum phase–space V^*M because Hamiltonian vector–fields

$$\vartheta_f = \partial^k f \partial_k - \partial_k f \partial^k, \qquad \vartheta_f \rfloor df' = \{f, f'\}_V, \qquad (f, f' \in C^\infty(V^*M)),$$

of functions on V^*M are vertical. Hamiltonian dynamics of time–dependent mechanics is described in a different way as a particular Hamiltonian dynamics on fibre bundles [MS98, GMS97].

A Hamiltonian on the momentum phase–space $V^*M \to \mathbb{R}$ of time–dependent mechanics is defined as a global section

$$h : V^*M \to T^*M, \qquad p \circ h = -\mathcal{H}(t, q^j, p_j),$$

of the affine bundle ζ (2.123). It induces the pull–back Hamiltonian form

$$H = h^*\Xi = p_k dq^k - \mathcal{H}dt, \tag{2.126}$$

on V^*M. Given H (2.126), there exists a unique vector–field γ_H on V^*M such that

$$\gamma_H \rfloor dt = 1, \qquad \gamma_H \rfloor dH = 0. \tag{2.127}$$

This vector–field reads

$$\gamma_H = \partial_t + \partial^k \mathcal{H}\partial_k - \partial_k \mathcal{H}\partial^k. \tag{2.128}$$

It defines the first–order Hamiltonian equation

$$\dot{t} = 1, \qquad \dot{q}^k = \partial^k \mathcal{H}, \qquad \dot{p}_k = -\partial_k \mathcal{H} \tag{2.129}$$

on V^*M, where $(t, q^k, p_k, \dot{t}, \dot{q}^k, \dot{p}_k)$ are holonomic coordinates on the tangent bundle TV^*M. Solutions of this equation are trajectories of the vector–field γ_H. They assemble into a (regular) foliation of V^*M.

A first integral of the Hamiltonian equation (2.129) is defined as a smooth real function F on V^*M whose Lie derivative

$$\mathcal{L}_{\gamma_H} F = \gamma_H \rfloor dF = \partial_t F + \{\mathcal{H}, F\}_V$$

along the vector–field γ_H (2.128) vanishes, i.e., the function F is constant on trajectories of the vector–field γ_H. A time–dependent Hamiltonian system (V^*M, H) on V^*M is said to be completely integrable if the Hamiltonian equation (2.129) admits m first integrals F_k which are in involution with respect to the Poisson bracket $\{,\}_V$ (2.125) and whose differentials dF_k are linearly independent almost everywhere. This system can be extended to an autonomous completely integrable Hamiltonian system on T^*M as follows.

Let us consider the pull–back $\zeta^* H$ of the Hamiltonian form $H = h^*\Xi$ onto the cotangent bundle T^*M. Note that the difference $\Xi - \zeta^* h^* \Xi$ is a horizontal 1–form on $T^*M \to \mathbb{R}$ and that

$$\mathcal{H}^* = \partial_t \rfloor (\Xi - \zeta^* h^* \Xi)) = p + \mathcal{H} \tag{2.130}$$

is a function on T^*M [Sni80]. Let us regard $\mathcal{H}^*$ (2.130) as a Hamiltonian of an autonomous Hamiltonian system on the symplectic manifold (T^*M, Ω_T). The Hamiltonian vector–field of $\mathcal{H}^*$ on T^*M reads

$$\gamma_T = \partial_t - \partial_t \mathcal{H}\partial^0 + \partial^k \mathcal{H}\partial_k - \partial_k \mathcal{H}\partial^k.$$

It is projected onto the vector–field γ_H (2.128) on V^*M, and the relation

$$\zeta^*(\mathcal{L}_{\gamma_H} f) = \{\mathcal{H}^*, \zeta^* f\}_T, \qquad (f \in C^\infty(V^*M)).$$

holds. An immediate consequence of this relation is the following.

Let $(V^*M, H; F_k)$ be a time–dependent completely integrable Hamiltonian system with first integrals $\{F_k\}$ on V^*M. Then $(T^*M; \mathcal{H}^*, \zeta^* F_k)$ is an

autonomous completely integrable Hamiltonian system on T^*M whose first integrals $\{\mathcal{H}^*, \zeta^* F_k\}$ are in involution with respect to the Poisson bracket $\{,\}_T$ (2.122). Furthermore, let N be a connected invariant manifold of the time–dependent completely integrable Hamiltonian system $(V^*M, H; F_k)$. Then $h(N) \subset T^*M$ is a connected invariant manifold of the completely integrable Hamiltonian system $(T^*M; \mathcal{H}^*, \zeta^* F_k)$ on T^*M. If N contains no critical points of first integrals F_k, then $\{\mathcal{H}^*, \zeta^* F_k\}$ have no critical points in $h(N)$.

Time–Dependent Action–Angle Coordinates

Let us introduce *time–dependent action–angle coordinates* around an invariant manifold N of a time–dependent completely integrable Hamiltonian system (V^*M, H) as those induced by the action–angle coordinates around the invariant manifold $h(N)$ of the autonomous completely integrable system $(T^*M, \mathcal{H}^*)$.

Let M' be a connected invariant manifold of an autonomous completely integrable system (F_α), $\alpha = 1, \ldots, n$, on a symplectic manifold (Z, Ω_Z), and let the Hamiltonian vector–fields of first integrals F_α on M' be complete. Let there exist a neighborhood U of M' such that F_α have no critical points in U and the submersion $\times F_\alpha : U \to \mathbb{R}^n$ is a trivial fibre bundle over a domain $V' \subset \mathbb{R}^n$. Then U is isomorphic to the *symplectic annulus*

$$W = V' \times (\mathbb{R}^{n-m} \times T^m),$$

provided with the generalized action–angle coordinates

$$(I_1, \ldots, I_n; x^1, \ldots, x^{n-m}; \phi^1, \ldots, \phi^m)$$

such that the symplectic form on W reads

$$\Omega_Z = dI_i \wedge x^i + dI_{n-m+k} \wedge d\phi^k,$$

and the first integrals F_α are functions of the action coordinates (I_α) only.

In particular, let M' be a compact invariant manifold of a completely integrable system $\{F_\alpha\}$, $\alpha = 1, \ldots, n$, on a symplectic manifold (Z, Ω_Z) which does not contain critical points of the first integrals F_α. Let the vector–field γ_H (2.128) be complete. Let a connected invariant manifold N of a time–dependent completely integrable Hamiltonian system $(V^*M, H; F_k)$ contain no critical points of first integrals F_k, and let its projection N_0 onto the fibre V_0^*M along trajectories of γ_H be compact. Then the invariant manifold $h(N)$ of the completely integrable Hamiltonian system $(T^*M; \mathcal{H}^*, \zeta^* F_k)$ has an open neighborhood U.

Now, the open neighborhood U of the invariant manifold $h(N)$ of the completely integrable Hamiltonian system $(T^*M; \mathcal{H}^*, \zeta^* F_k)$ is isomorphic to the symplectic annulus

$$W' = V' \times (\mathbb{R} \times T^m), \qquad V' = (-\varepsilon, \varepsilon) \times V, \tag{2.131}$$

provided with the generalized action–angle coordinates

$$(I_0, \ldots, I_m; t, \phi^1, \ldots, \phi^m). \tag{2.132}$$

Moreover, we find that $J_0 = r$, $a_0^\alpha = \delta_0^\alpha$ and, as a consequence,

$$a_0^0 = \frac{\partial I_0}{\partial J_0} = 1, \qquad a_i^0 = \frac{\partial I_i}{\partial J_0} = 0,$$

i.e., the action coordinate I_0 is linear in the coordinate r, while I_i are independent of r. With respect to the coordinates (2.132), the symplectic form on W' reads

$$\Omega_T = dI_0 \wedge dt + dI_k \wedge d\phi^k,$$

the Hamiltonian $\mathcal{H}^*$ is an affine function $\mathcal{H}^* = I_0 + \mathcal{H}'(I_j)$ of the action coordinate I_0, while the first integrals $\zeta^* F_k$ depends only on the action coordinates I_i. The Hamiltonian vector–field of the Hamiltonian $\mathcal{H}^*$ is

$$\gamma_T = \partial_t + \partial^i \mathcal{H}' \partial_i. \tag{2.133}$$

Since the action coordinates I_i are independent on the coordinate r, the symplectic annulus W' (2.131) inherits the composite fibration

$$W' \to V \times (\mathbb{R} \times T^m) \to \mathbb{R}. \tag{2.134}$$

Therefore, one can regard $W = V \times (\mathbb{R} \times T^m)$ as a momentum phase–space of the time–dependent Hamiltonian system in question around the invariant manifold N. It is coordinated by (I_i, t, ϕ^i), which we agree to call the time–dependent action–angle coordinates. By the relation similar to (2.124), W can be equipped with the Poisson structure

$$\{f, f'\}_W = \partial^i f \partial_i f' - \partial_i f \partial^i f',$$

while the global section $h' : W \to W'$ such that $I_0 \circ h' = -\mathcal{H}'$, of the trivial bundle ζ (2.134), gives W with the Hamiltonian form

$$H' = I_i d\phi^i - \mathcal{H}'(I_j) dt.$$

The associated vector–field γ_H (2.127) is exactly the projection onto W of the Hamiltonian vector–field γ_T (2.133), and takes the same coordinate form. It defines the Hamiltonian equation on W,

$$I_i = \text{const}, \qquad \dot{\phi}^i = \partial^i \mathcal{H}'(I_j).$$

One can think of this equation as being the Hamiltonian equation of a time–dependent Hamiltonian system around the invariant manifold N relative to time–dependent action–angle coordinates.

Lyapunov Stability

The notion of the Lyapunov stability of a dynamical equation on a smooth manifold implies that this manifold is equipped with a Riemannian metric. At the same time, no preferable Riemannian metric is associated to a first–order dynamical equation. Here, we aim to study the Lyapunov stability of first–order dynamical equations in non–autonomous mechanics with respect to different (time–dependent) Riemannian metrics.

Let us recall that a solution $s(t)$, for all $t \in \mathbb{R}$, of a first–order dynamical equation is said to be Lyapunov stable (in the positive direction) if for $t_0 \in \mathbb{R}$ and any $\varepsilon > 0$, there is $\delta > 0$ such that, if $s'(t)$ is another solution and the distance between the points $s(t_0)$ and $s'(t_0)$ is inferior to δ, then the distance between the points $s(t)$ and $s'(t)$ for all $t > t_0$ is inferior to ε. In order to formulate a criterion of the Lyapunov stability with respect to a time–dependent Riemannian metric, we introduce the notion of a covariant Lyapunov tensor as generalization of the well–known Lyapunov matrix. The latter is defined as the coefficient matrix of the variation equation [Gal83, HS74], and fails to be a tensor under coordinate transformations, unless they are linear and time–independent. On the contrary, the covariant Lyapunov tensor is a true tensor field, but it essentially depends on the choice of a Riemannian metric. The following was shown in [Sar02b]:

(i) If the covariant Lyapunov tensor is negative definite in a tubular neighborhood of a solution s at points $t \geq t_0$, this solution is Lyapunov stable.

(ii) For any first–order dynamical equation, there exists a (time–dependent) Riemannian metric such that every solution of this equation is Lyapunov stable.

(iii) Moreover, the Lyapunov exponent of any solution of a first–order dynamical equation can be made equal to any real number with respect to the appropriate (time–dependent) Riemannian metric. It follows that chaos in dynamical systems described by smooth (C^∞) first–order dynamical equations can be characterized in full by time–dependent Riemannian metrics.

First–Order Dynamical Equations

Let $\mathbb{R}$ be the time axis provided with the Cartesian coordinate t and transition functions $t' = t+\text{const}$. In geometrical terms [MS98], a (smooth) first–order dynamical equation in non–autonomous mechanics is defined as a vector–field γ on a smooth fibre bundle

$$\pi : Y \longrightarrow \mathbb{R} \tag{2.135}$$

which obeys the condition $\gamma \rfloor dt = 1$, i.e.,

$$\gamma = \partial_t + \gamma^k \partial_k. \tag{2.136}$$

The associated first–order dynamical equation takes the form

$$\dot{t} = 1, \qquad \dot{y}^k = \gamma^k(t, y^j)\partial_k, \tag{2.137}$$

where $(t, y^k, \dot{t}, \dot{y}^k)$ are holonomic coordinates on TY. Its solutions are trajectories of the vector–field γ (2.136). They assemble into a (regular) foliation $\mathcal{F}$ of Y. Equivalently, γ (2.136) is defined as a connection on the fibre bundle (2.135).

A fibre bundle Y (2.135) is trivial, but it admits different trivializations

$$Y \cong \mathbb{R} \times M, \tag{2.138}$$

distinguished by fibrations $Y \longrightarrow M$. For example, if there is a trivialization (2.138) such that, with respect to the associated coordinates, the components γ^k of the connection γ (2.136) are independent of t, one says that γ is a conservative first–order dynamical equation on M.

Hereafter, the vector–field γ (2.136) is assumed to be *complete*, i.e., there is a unique global solution of the dynamical equation γ through each point of Y. For example, if fibres of $Y \longrightarrow \mathbb{R}$ are compact, any vector–field γ (2.136) on Y is complete.

If the vector–field γ (2.136) is complete, there exists a trivialization (2.138) of Y, with an atlas $\Psi = \{(U; t, y^a)\}$ of a fibre bundle $Y \longrightarrow \mathbb{R}$ with time–independent transition functions $y'^a(y^b)$, such that any solution s of γ reads

$$s^a(t) = \text{const}, \qquad (\text{for all} \quad t \in \mathbb{R}),$$

with respect to associated bundle coordinates (t, y^a). If γ is complete, the foliation $\mathcal{F}$ of its trajectories is a fibration ζ of Y along these trajectories onto any fibre of Y, e.g., $Y_{t=0} \cong M$. This fibration induces a desired trivialization [MS98].

One can think of the coordinates (t, y^a) as being the *initial–date coordinates* because all points of the same trajectory differ from each other only in the temporal coordinate.

Let us consider the canonical lift $V\gamma$ of the vector–field γ (2.136) onto the vertical tangent bundle VY of $Y \longrightarrow \mathbb{R}$. With respect to the holonomic bundle coordinates $(t, y^k, \overline{y}^k)$ on VY, it reads

$$V\gamma = \gamma + \partial_j \gamma^k \overline{y}^j \overline{\partial}_k, \qquad \text{where} \qquad \overline{\partial}_k = \frac{\partial}{\partial \overline{y}^k}.$$

This vector–field obeys the condition $V\gamma \rfloor dt = 1$, and defines the first–order dynamical equation

$$\dot{t} = 1, \qquad \dot{y}^k = \gamma^k(t, y^i), \tag{2.139}$$

$$\dot{\overline{y}}{}^k_t = \partial_j \gamma^k(t, y^i)\overline{y}^j \tag{2.140}$$

on VY. The equation (2.139) coincides with the initial one (2.137). The equation (2.140) is the well–known variation equation. Substituting a solution s of the initial dynamical equation (2.139) into (2.140), one gets a linear dynamical equation whose solutions $\overline{s}$ are Jacobi fields of the solution s. In particular, if $Y \longrightarrow \mathbb{R}$ is a vector bundle, there are the canonical splitting $VY \cong Y \times Y$ and the map $VY \longrightarrow Y$ so that $s + \overline{s}$ obeys the initial dynamical equation (2.139) modulo the terms of order > 1 in $\overline{s}$.

Lyapunov Tensor

The collection of coefficients

$$l_j{}^k = \partial_j \gamma^k \tag{2.141}$$

of the variation equation (2.140) is called the *Lyapunov matrix*. Clearly, it is not a tensor under bundle coordinate transformations of the fibre bundle Y (2.135). Therefore, we introduce a *covariant Lyapunov tensor* as follows.

Let a fibre bundle $Y \to \mathbb{R}$ be provided with a Riemannian fibre metric g, defined as a section of the *symmetrized tensor product*

$$\vee^2 V^*Y \to Y \tag{2.142}$$

of the vertical cotangent bundle V^*Y of $Y \to \mathbb{R}$. With respect to the holonomic coordinates $(t, y^k, \overline{y}_k)$ on V^*Y, it takes the coordinate form

$$g = \frac{1}{2} g_{ij}(t, y^k) \overline{d}y^i \vee \overline{d}y^j,$$

where $\{\overline{d}y^i\}$ are the holonomic fibre bases for V^*Y.

Given a first–order dynamical equation γ, let

$$V^*\gamma = \gamma - \partial_j \gamma^k \overline{y}_k \overline{\partial}^j, \qquad \text{where} \qquad \overline{\partial}^j = \frac{\partial}{\partial \overline{y}_j}. \tag{2.143}$$

be the canonical lift of the vector–field γ (2.136) onto V^*Y. It is a connection on $V^*Y \longrightarrow \mathbb{R}$. Let us consider the Lie derivative $\mathfrak{L}_\gamma g$ of the Riemannian fibre metric g along the vector–field $V^*\gamma$ (2.143). It reads

$$L_{ij} = (\mathfrak{L}_\gamma g)_{ij} = \partial_t g_{ij} + \gamma^k \partial_k g_{ij} + \partial_i \gamma^k g_{kj} + \partial_j \gamma^k g_{ik}. \tag{2.144}$$

This is a section of the fibre bundle (2.142) and, consequently, a tensor with respect to any bundle coordinate transformation of the fibre bundle (2.135). We agree to call it the covariant Lyapunov tensor. If g is an Euclidean metric, it becomes the following symmetrization of the Lyapunov matrix (2.141),

$$L_{ij} = \partial_i \gamma^j + \partial_j \gamma^i = l_i{}^j + l_j{}^i.$$

Let us point the following two properties of the covariant Lyapunov tensor.

(i) Written with respect to the initial–date coordinates, the covariant Lyapunov tensor is given by

$$L_{ab} = \partial_t g_{ab}.$$

(ii) Given a solution s of the dynamical equation γ and a solution $\overline{s}$ of the variation equation (2.140), we have

$$L_{ij}(t, s^k(t)) \overline{s}^i \overline{s}^j = \partial_t (g_{ij}(t, s^k(t)) \overline{s}^i \overline{s}^j).$$

The definition of the covariant Lyapunov tensor (2.144) depends on the choice of a Riemannian fibre metric on the fibre bundle Y.

If the vector–field γ is complete, there is a Riemannian fibre metric on Y such that the covariant Lyapunov tensor vanishes everywhere. Let us choose the atlas of the initial–date coordinates. Using the fibration $\zeta : Y \longrightarrow Y_{t=0}$, one can give Y with a time–independent Riemannian fibre metric

$$g_{ab}(t, y^c) = h(t) g_{ab}(0, y^c) \tag{2.145}$$

where $g_{ab}(0, y^c)$ is a Riemannian metric on the fibre $Y_{t=0}$ and $h(t)$ is a positive smooth function on $\mathbb{R}$. The covariant Lyapunov tensor with respect to the metric (2.145) is given by

$$L_{ab} = \partial_t h g_{ab}.$$

Putting $h(t) = 1$, we get $L = 0$.

Lyapunov Stability of the First–Order Dynamical Equations

With the covariant Lyapunov tensor (2.144), we get the following criterion of the stability condition of Lyapunov.

Recall that, given a Riemannian fibre metric g on a fibre bundle $Y \longrightarrow \mathbb{R}$, the instantwise distance $\rho_t(s, s')$ between two solutions s and s' of a dynamical equation γ on Y at an instant t is the distance between the points $s(t)$ and $s'(t)$ in the Riemannian space $(Y_t, g(t))$.

Let s be a solution of a first–order dynamical equation γ. If there exists an open tubular neighborhood U of the trajectory s where the covariant Lyapunov tensor (2.144) is negative-definite at all instants $t \geq t_0$, then there exists an open tubular neighborhood U' of s such that

$$\lim_{t' \to \infty} [\rho_{t'}(s, s') - \rho_t(s, s')] < 0$$

for any $t > t_0$ and any solution s' crossing U'. Since the condition and the statement are coordinate–independent, let us choose the following chart of initial–date coordinates that covering the trajectory s. Put $t = 0$ without a loss of generality. There is an open neighborhood $U_0 \subset Y_0 \cap U$ of $s(0)$ in the Riemannian manifold $(Y_0, g(0))$ which can be provided with the normal coordinates (x^a) defined by the Riemannian metric $g(0)$ in Y_0 and centralized at $s(0)$. Let us consider the open tubular $U' = \zeta^{-1}(U_0)$ with the coordinates (t, x^a). It is the desired chart of initial–date coordinates. With respect to these coordinates, the solution s reads $s^a(t) = 0$. Let $s'^a(t) = u^a = \text{const}$ be another solution crossing U'. The instantwise distance $\rho_t(s, s')$, $t \geq 0$, between solutions s and s' is the distance between the points $(t, 0)$ and (t, u) in the Riemannian manifold $(Y_t, g(t))$. This distance does not exceed the length

$$\overline{\rho}_t(s, s') = \left[\int_0^1 g_{ab}(t, \tau u^c) u^a u^b d\tau \right]^{1/2} \tag{2.146}$$

of the curve

$$x^a = \tau u^a, \qquad (\tau \in [0,1]) \tag{2.147}$$

in the Riemannian space $(Y_t, g(t))$, while $\rho_0(s,s') = \overline{\rho}_0(s,s')$. The temporal derivative of the function $\overline{\rho}_t(s,s')$ (2.146) reads

$$\partial_t \overline{\rho}_t(s,s') = \frac{1}{2(\overline{\rho}_t(s,s'))^{1/2}} \int_0^1 \partial_t g_{ab}(t,\tau u^c) u^a u^b d\tau. \tag{2.148}$$

Since the bilinear form $\partial_t g_{ab} = L_{ab}$, $t \geq 0$, is negative-definite at all points of the curve (2.147), the derivative (2.148) at all points $t \geq t_0$ is also negative. Hence, we get

$$\rho_{t>0}(s,s') < \overline{\rho}_{t>0}(s,s') < \overline{\rho}_0(s,s') = \rho_0(s,s').$$

The solution s is *Lyapunov stable* with respect to the Riemannian fibre metric g. One can think of the solution s as being *isometrically Lyapunov stable*. Being Lyapunov stable with respect a Riemannian fibre metric g, a solution s need not be so with respect to another Riemannian fibre metric g', unless g' results from g by a time–independent transformation.

For any first–order dynamical equation defined by a complete vector–field γ (2.136) on a fibre bundle $Y \longrightarrow \mathbb{R}$, there exists a Riemannian fibre metric on Y such that each solution of γ is Lyapunov stable. This property obviously holds with respect to the Riemannian fibre metric (2.145) where $h = 1$.

Let λ be a real number. Given a dynamical equation γ defined by a complete vector–field γ (2.136), there is a Riemannian fibre metric on Y such that the Lyapunov spectrum of any solution of γ reduces to λ. To prove this, recall that the (upper) Lyapunov exponent of a solution s' with respect to a solution s is defined as the limit

$$K(s,s') = \overline{\lim_{t\to\infty}} \frac{1}{t} \ln(\rho_t(s,s')). \tag{2.149}$$

Let us give Y with the Riemannian fibre metric (2.145) where $h = \exp(\lambda t)$. A simple computation shows that the Laypunov exponent (2.149) with respect to this metric is exactly λ.

If the upper limit

$$\lambda = \overline{\lim_{\rho_{t=0}(s,s')\longrightarrow 0}} K(s,s')$$

is *negative*, the solution s is said to be exponentially Lyapunov stable. If there exists at least one positive Lyapunov exponent, one speaks about chaos in a dynamical system [Gut90]. This shows that chaos in smooth dynamical systems can be characterized in full by time–dependent Riemannian metrics.

Example

Here is a simple example which shows that solutions of a smooth first–order dynamical equation can be made *Lyapunov stable* at will by the choice of an appropriate time–dependent Riemannian metric.

Let $\mathbb{R}$ be the time axis provided with the Cartesian coordinate t. In geometrical terms, a (smooth) first–order dynamical equation in non–autonomous mechanics is defined as a vector–field γ on a smooth fibre bundle $Y \to \mathbb{R}$ which obeys the condition $\gamma \rfloor dt = 1$. With respect to bundle coordinates (t, y^k) on Y, this vector–field becomes (2.136). The associated first–order dynamical equation takes the form

$$\dot{y}^k = \gamma^k(t, y^j)\partial_k,$$

where $(t, y^k, \dot{t}, \dot{y}^k)$ are holonomic coordinates on the tangent bundle TY of Y. Its solutions are trajectories of the vector–field γ (2.136).

Let a fibre bundle $Y \to \mathbb{R}$ be provided with a Riemannian fibre metric g, defined as a section of the symmetrized tensor product $\vee^2 V^*Y \to Y$ of the vertical cotangent bundle V^*Y of $Y \to \mathbb{R}$. With respect to the holonomic coordinates $(t, y^k, \overline{y}_k)$ on V^*Y, it takes the coordinate form

$$g = \frac{1}{2} g_{ij}(t, y^k)\overline{d}y^i \vee \overline{d}y^j,$$

where $\{\overline{d}y^i\}$ are the holonomic fibre bases for V^*Y.

Recall that above we have proposed the following: Let λ be a real number. Given a dynamical equation defined by a complete vector–field γ (2.136), there exists a Riemannian fibre metric on Y such that the Lyapunov spectrum of any solution of γ is λ. The following example aims to illustrate this fact.

Let us consider 1D motion on the axis $\mathbb{R}$ defined by the first–order dynamical equation

$$\dot{y} = y \tag{2.150}$$

on the fibre bundle $Y = \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ coordinated by (t, y). Solutions of the equation (2.150) read

$$s(t) = c\exp(t), \qquad (\text{with } c = \text{const}). \tag{2.151}$$

Let $e_{yy} = 1$ be the standard Euclidean metric on $\mathbb{R}$. With respect to this metric, the instantwise distance between two arbitrary solutions

$$s(t) = c\exp(t), \qquad s'(t) = c'\exp(t) \tag{2.152}$$

of the equation (2.150) is

$$\rho_t(s, s')_e = |c - c'|\exp(t).$$

Hence, the Lyapunov exponent $K(s, s')$ (2.149) equals 1, and so is the Lyapunov spectrum of any solution (2.151) of the first–order dynamical equation (2.150).

Let now λ be an arbitrary real number. There exists a coordinate $y' = y\exp(-t)$ on $\mathbb{R}$ such that, written relative to this coordinate, the solutions (2.151) of the equation (2.150) read $s(t) =$const. Let us choose the Riemannian fibre metric on $Y \to \mathbb{R}$ which takes the form $g_{y'y'} = \exp(2\lambda t)$ with respect to the coordinate y'. Then relative to the coordinate y, it reads

$$g_{yy} = \frac{\partial y'}{\partial y}\frac{\partial y'}{\partial y}g_{y'y'} = \exp(2(\lambda - 1)t). \tag{2.153}$$

The instantwise distance between the solutions s and s' (2.152) with respect to the metric g (2.153) is

$$\rho_t(s, s') = [g_{yy}(s(t) - s'(t))^2]^{1/2} = |c - c'|\exp(\lambda t).$$

One at once gets that the Lyapunov spectrum of any solution of the differential equation (2.150) with respect to the metric (2.153) is λ.

2.1.3 Semi–Riemannian Geometrical Dynamics

In this subsection we develop a *Finsler–like approach* to semi–Riemannian geometrical dynamics.

Vector–Fields and Connections

Let M be an nD smooth manifold. Recall that a smooth (C^∞) vector–field X on M defines the flow

$$\dot{x} = X(x). \tag{2.154}$$

By definition, a semi–Riemannian metric g on M is a smooth symmetric tensor field of type $(0, 2)$ which assigns to each point $x \in M$ a nondegenerate inner product $g(x)$ on the tangent space T_xM of signature (r, s). The pair (M, g) is called a *semi–Riemannian manifold*.

The vector–field X and the semi–Riemannian metric g determine the *energy* $f : M \to \mathbb{R}$,given by $f = \frac{1}{2}g(X, X)$. The vector–field X (and its flow) on (M, g) is called [Udr00]:

1. *time–like*, if $f < 0$;
2. *nonspacelike or causal*, if $f \leq 0$;
3. *null or lightlike*, if $f = 0$;
4. *space–like*, if $f > 0$.

Let ∇ be the *Levi–Civita connection* of (M, g). Using the semi–Riemannian version of the covariant derivative operator (1.96), we get the *prolongation*

$$\frac{\nabla}{dt}\dot{x} = \nabla_{\dot{x}}X \tag{2.155}$$

of the differential system (2.154) or of any perturbation of the system (2.154) get adding to the second member X a parallel vector–field Y with respect

to the covariant derivative ∇. The prolongation by derivation represents the general dynamics of the flow. The vector–field Y can be used to illustrate a progression from stable to unstable flows, or converse.

The vector–field X, the metric g, and the connection ∇ determine the external $(1,1)-$tensor field

$$F = \nabla X - g^{-1} \otimes g(\nabla X), \qquad F_j{}^i = \nabla_j X^i - g^{ih} g_{kj} \nabla_h X^k,$$

(with i,j,h,k=1,...,n), which characterizes the *helicity* of vector–field X and its flow.

First we write the differential system (2.155) in the equivalent form

$$\frac{\nabla}{dt}\dot{x} = g^{-1} \otimes g(\nabla X)\,(\dot{x}) + F\,(\dot{x})\,. \tag{2.156}$$

Successively we modify the differential system (2.156) as follows [Udr00]:

$$\frac{\nabla}{dt}\dot{x} = g^{-1} \otimes g(\nabla X)(X) + F\,(\dot{x})\,, \tag{2.157}$$

$$\frac{\nabla}{dt}\dot{x} = g^{-1} \otimes g(\nabla X)\,(\dot{x}) + F(X), \tag{2.158}$$

$$\frac{\nabla}{dt}\dot{x} = g^{-1} \otimes g(\nabla X)(X) + F(X). \tag{2.159}$$

Obviously, the second–order systems (2.157), (2.158), (2.159) are prolongations of the first–order system (2.154). Each of them is connected either to the dynamics of the field X or to the dynamics of a particle which is sensitive to the vector–field X. Since

$$g^{-1} \otimes g(\nabla X)(X) = \operatorname{grad} f,$$

we shall show that the prolongation (2.157) describes a conservative dynamics of the vector–field X or of a particle which is sensitive to the vector–field X. The physical phenomenon produced by (2.158) or (2.159) was not yet studied [Udr00].

In the case $F = 0$, the kinematic system (2.154) prolongs to a potential dynamical system with n degrees of freedom, namely

$$\frac{\nabla}{dt}\dot{x} = \operatorname{grad} f. \tag{2.160}$$

In the case $F \neq 0$, the kinematic system (2.154) prolongs to a non–potential dynamical system with n degrees of freedom, namely

$$\frac{\nabla}{dt}\dot{x} = \operatorname{grad} f + F\,(\dot{x})\,. \tag{2.161}$$

Let us show that the dynamical systems (2.160) and (2.161) are conservative. To simplify the exposition we identity the tangent bundle TM with the

cotangent bundle T^*M using the semi–Riemann metric g [Udr00]. The trajectories of the dynamical system (2.160) are the extremals of the Lagrangian

$$L = \frac{1}{2} g\left(\dot{x}, \dot{x}\right) + f(x).$$

The trajectories of the dynamical system (2.161) are the extremals of the Lagrangian

$$L = \frac{1}{2} g\left(\dot{x} - X,\ \dot{x} - X\right) = \frac{1}{2} g\left(\dot{x}, \dot{x}\right) - g\left(X, \dot{x}\right) + f(x).$$

The dynamical systems (2.160) and (2.161) are conservative, the Hamiltonian being the same for both cases, namely

$$H = \frac{1}{2} g\left(\dot{x}, \dot{x}\right) - f(x).$$

The restriction of the Hamiltonian H to the flow of the vector–field X is zero.

Hamiltonian Structures on the Tangent Bundle

Let (N, ω) be a $2n$D symplectic phase–space manifold, and $H : N \to \mathbb{R}$ be a C^∞ real function. We define the Hamiltonian gradient X_H as being the vector–field which satisfies

$$\omega_p(X_H(x), v) = dH(x)(v), \qquad (\text{for all } v \in T_x N),$$

and the Hamiltonian equations as

$$\dot{x} = X_H(x).$$

Let (M, g) be a semi–Riemann nD manifold. Let X be a C^∞ vector–field on M, and $\omega = g \circ F$ the two–form associated to the tensor field $F = \nabla X - g^{-1} \otimes g(\nabla X)$ via the metric g.

The tangent bundle is usually equipped with the *Sasakian metric* G, induced by g,

$$G = g_{ij} dx^i \otimes dx^j + g_{ij} \delta y^i \otimes \delta y^j.$$

If (x^i, y^i) are the coordinates of the point $(x, y) \in TM$ and Γ^i_{jk} are the components of the connection induced by g_{ij}, then we have the following *dual frames* [Udr00]

$$\left(\frac{\delta}{\delta x^i} = \frac{\partial}{\partial x^i} - \Gamma^h_{ij} y^j \frac{\partial}{\partial y^h}, \quad \frac{\partial}{\partial y^i}\right) \subset \mathcal{X}(TM), \qquad \text{and}$$
$$\left(dx^j, \quad \delta y^j = dy^j + \Gamma^j_{hk} y^k dx^h\right) \subset \mathcal{X}^*(TM).$$

The dynamical system (2.160) lifts to TM as a Hamiltonian dynamical system with respect to the

$$\text{Hamiltonian} \qquad H = \frac{1}{2} g(\dot{x}, \dot{x}) - f(x), \qquad \text{and}$$
$$\text{symplectic two–form} \qquad \Omega_1 = g_{ij} dx^i \omega \delta y^j.$$

This can be verified by putting $\eta_1 = g_{ij} y^i dx^j$, and $d\eta_1 = -\Omega_1$.

The dynamical system (2.161) lifts to TM as a Hamiltonian dynamical system with respect to the above Hamiltonian function and the symplectic two–form

$$\Omega_2 = \frac{1}{2} \omega_{ij} dx^i \omega dx^j + g_{ij} dx^i \omega \delta y^j.$$

This can be verified by putting $\eta_2 = -g_{ij} X^i dx^j + g_{ij} y^i dx^j$, and $d\eta_2 = -\Omega_2$.

In the remainder of this subsection, we give three examples in Euclidean spaces, so we can put all indices down (still summing over repeated indices).

Pendulum Geometry

We use the Riemannian manifold $(\mathbb{R}^2, \delta_{ij})$. The small oscillations of a plane pendulum are described as solutions of the following differential system giving the plane pendulum flow,

$$\dot{x}_1 = -x_2, \qquad \dot{x}_2 = x_1. \tag{2.162}$$

In this case, the set $\{x_1(t) = 0,\ x_2(t) = 0,\ (t \in \mathbb{R})\}$ is the equilibrium point and

$$x_1(t) = c_1 \cos t + c_2 \sin t, \qquad x_2(t) = c_1 \sin t - c_2 \cos t$$

is the general solution, which is a family of circles with a common center.

$$\text{Let} \qquad X = (X_1, X_2), \qquad X_1(x_1, x_2) = -x_2, \qquad X_2(x_1, x_2) = x_1,$$
$$f(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2), \qquad \operatorname{curl} X = (0, 0, 2), \qquad \operatorname{div} X = 0.$$

The pendulum flow conserves the areas. The prolongation by derivation of the kinematic system (2.162) is [Udr00]

$$\ddot{x}_i = \frac{\partial X_i}{\partial x_j} \dot{x}_j, \qquad (i, j = 1, 2)$$
$$\text{or} \qquad \ddot{x}_1 = -\dot{x}_2, \qquad \ddot{x}_2 = \dot{x}_1.$$

This prolongation admits a family of circles as the general solution

$$x_1(t) = a_1 \cos t + a_2 \sin t + h, \qquad x_2(t) = a_1 \sin t - a_2 \cos t + k, \qquad (t \in \mathbb{R}).$$

The pendulum geometrodynamics is described by

$$\ddot{x}_i = \frac{\partial f}{\partial x_i} + \left(\frac{\partial X_i}{\partial x_j} - \frac{\partial X_j}{\partial x_i} \right) \dot{x}_j, \qquad (i, j = 1, 2),$$
$$\text{or} \qquad \ddot{x}_1 = x_1 - 2\dot{x}_2, \qquad \ddot{x}_2 = x_2 + 2\dot{x}_1, \tag{2.163}$$

with a family of spirals as the general solution

$$x_1(t) = b_1 \cos t + b_2 \sin t + b_3 t \cos t + b_4 t \sin t,$$
$$x_2(t) = b_1 \sin t - b_2 \cos t + b_3 t \sin t - b_4 t \cos t, \qquad (t \in \mathbb{R}).$$

Using

$$L = \frac{1}{2}\left[(\dot{x}_1)^2 + (\dot{x}_2)^2\right] + x_2\dot{x}_1 - x_1\dot{x}_2 + f,$$
$$H = \frac{1}{2}\left[(\dot{x}_1)^2 + (\dot{x}_2)^2\right] - f, \qquad g_{ij} = (H+f)\delta_{ij},$$
$$N_j{}^i = -F_j{}^i = -\delta^{ih}F_{jh}, \qquad F_{ij} = \frac{\partial X_j}{\partial x_i} - \frac{\partial X_i}{\partial x_j}, \qquad (i,j,h = 1,2),$$

the solutions of the differential system (2.163) are horizontal pregeodesics of the Riemann–Jacobi–Lagrangian manifold $(\mathbb{R}^2 \setminus \{0\},\ g_{ij}, N_j{}^i)$.

Geometry of the Lorenz Flow

We use the Riemannian manifold $(\mathbb{R}^3, \delta_{ij})$. The *Lorenz flow* is a first dissipative model with *chaotic behavior* discovered in numerical experiment. Its state equations are (see [Lor63, Spa82])

$$\dot{x}_1 = -\sigma x_1 + \sigma x_2, \qquad \dot{x}_2 = -x_1 x_3 + r x_1 - x_2, \qquad \dot{x}_3 = x_1 x_2 - b x_3,$$

where σ, r, b are real parameters. Usually σ, b are kept fixed whereas r is varied. At

$$r > r_0 = \frac{\sigma(\sigma + b + 3)}{\sigma - b - 1}$$

chaotic behavior is observed [Udr00].

$$\text{Let} \qquad X = (X_1, X_2, X_3), \qquad X_1(x_1,x_2,x_3) = -\sigma x_1 + \sigma x_2,$$
$$X_2(x_1,x_2,x_3) = -x_1x_3 + rx_1 - x_2, \qquad X_3(x_1,x_2,x_3) = x_1x_2 - bx_3,$$
$$f = \frac{1}{2}[(-\sigma x_1 + \sigma x_2)^2 + (-x_1x_3 + rx_1 - x_2)^2 + (x_1x_2 - bx_3)^2],$$
$$\operatorname{curl} X = (2x_1,\ -x_2,\ r - x_3 - \sigma).$$

The *Lorenz dynamics* is described by

$$\ddot{x}_i = \frac{\partial f}{\partial x_i} + \left(\frac{\partial X_i}{\partial x_j} - \frac{\partial X_j}{\partial x_i}\right)\dot{x}_j, \qquad (i,j = 1,2,3), \qquad \text{or}$$
$$\ddot{x}_1 = \frac{\partial f}{\partial x_1} + (\sigma + x_3 - r)\dot{x}_2 - x_2\dot{x}_3, \qquad \ddot{x}_2 = \frac{\partial f}{\partial x_2} + (r - x_3 - \sigma)\dot{x}_1 - 2x_1\dot{x}_3,$$
$$\ddot{x}_3 = \frac{\partial f}{\partial x_3} + x_2\dot{x}_1 + 2x_1\dot{x}_2. \qquad (2.164)$$

$$\text{Using} \qquad L = \frac{1}{2}\sum_{i=1}^{3}(\dot{x}_i)^2 - \sum_{i=1}^{3} X_i\dot{x}_i + f, \qquad H = \frac{1}{2}\sum_{i=1}^{3}(\dot{x}_i)^2 - f,$$

$$g_{ij} = (H+f)\delta_{ij}, \qquad N_j{}^i = -F_j{}^i = -\delta^{ih}F_{jh}, \qquad F_{ij} = \frac{\partial X_j}{\partial x_i} - \frac{\partial X_i}{\partial x_j},$$

$(i, j, h = 1, 2, 3)$, the solutions of the differential system (2.164) are horizontal pregeodesics of the Riemann–Jacobi–Lagrangian manifold $(\mathbb{R}^3 \setminus E,\ g_{ij},\ N_j{}^i)$, where E is the set of equilibrium points.

Geometry of the ABC Flow

We use the Riemannian manifold $(\mathbb{R}^3, \delta_{ij})$. One example of a fluid velocity that contains exponential stretching and hence instability is the ABC flow, named after *A*rnold, *B*eltrami and *C*hildress,

$$\dot{x}_1 = A\sin x_3 + C\cos x_2, \quad \dot{x}_2 = B\sin x_1 + A\cos x_3, \quad \dot{x}_3 = C\sin x_2 + B\cos x_1.$$

For nonzero values of the constants A, B, C the preceding system is not globally integrable. The topology of the flow lines is very complicated and can only be investigated numerically to reveal regions of chaotic behavior. The ABC flow conserves the volumes since the ABC field is solenoidal.

The ABC geometrodynamics is described by [Udr00]

$$\ddot{x}_i = \frac{\partial f}{\partial x_i} + \left(\frac{\partial X_i}{\partial x_j} - \frac{\partial X_j}{\partial x_i}\right)\dot{x}_j, \qquad (i, j = 1, 2, 3).$$

Since $f = \frac{1}{2}(A+B+C+2AC\sin x_3\cos x_2 + 2BA\sin x_1\cos x_3 + 2CB\sin x_2\cos x_1)$, and $\operatorname{curl} X = X$, the ABC geometrodynamics is given by the system,

$$\begin{aligned}
\ddot{x}_1 &= AB\cos x_1\cos x_3 - BC\sin x_1\sin x_2 \\
&\quad - (B\cos x_1 + C\sin x_2)\dot{x}_2 + (B\sin x_1 + A\cos x_3)\dot{x}_3, \\
\ddot{x}_2 &= -AC\sin x_2\sin x_3 + BC\cos x_1\cos x_2 \\
&\quad + (B\cos x_1 + C\sin x_2)\dot{x}_1 - (A\sin x_3 + C\cos x_2)\dot{x}_2, \\
\ddot{x}_3 &= AC\cos x_3\cos x_2 - BA\sin x_1\sin x_3 \\
&\quad - (B\sin x_1 + A\cos x_3)\dot{x}_1 + (C\cos x_2 + A\sin x_3)\dot{x}_2.
\end{aligned}$$

Using

$$L = \frac{1}{2}\dot{x}_i\dot{x}_i - X_i\dot{x}_i + f, \qquad H = \frac{1}{2}\dot{x}_i\dot{x}_i - f, \qquad (i, j, h = 1, 2, 3),$$

$$g_{ij} = (H+f)\delta_{ij}, \qquad N_j{}^i = -F_j{}^i = -\delta^{ih}F_{jh}, \qquad F_{ij} = \frac{\partial X_j}{\partial x_i} - \frac{\partial X_i}{\partial x_j},$$

the solutions of the above differential system are horizontal pregeodesics of the Riemann–Jacobi–Lagrangian manifold $(\mathbb{R}^3 \setminus E, g_{ij}, N_j{}^i)$, where E is the set of equilibrium points which is included in the surface of equation

$$\sin x_1\sin x_2\sin x_3 + \cos x_1\cos x_2\cos x_3 = 0.$$

2.1.4 Relativistic and Multi–Time Rheonomic Dynamics

Recall that a number of geometrical models in mechanics and physics are based on the notion of ordinary, autonomous Lagrangian (i.e., a smooth real function on $\mathbb{R} \times TM$). In this sense, we recall that a *Lagrangian space* $L^n = (M, L(x, y))$ is defined as a pair which consists of a real, smooth, nD manifold M with local coordinates $x^i, (i = 1, ..., n)$ and a regular Lagrangian $L : TM \to \mathbb{R}$. The geometry of Lagrangian spaces is now used in various fields to study natural phenomena where the dependence on position, velocity or momentum is involved [KO89]. Also, this geometry gives a model for both the gravitational and electromagnetic field theory, in a very natural blending of the geometrical structure of the space with the characteristic properties of the physical fields. Again, there are many problems in physics and variational calculus in which time–dependent Lagrangians are involved.

In the context exposed in [MKA88, MA94], the energy action functional E, attached to a given time–dependent Lagrangian,

$$L : \mathbb{R} \times TM \to \mathbb{R}, \qquad (t, x^i, v^i) \mapsto L(t, x^i, v^i), \qquad (i = 1, ..., n)$$

not necessarily homogenous with respect to the direction $\{v^i\}$, is of the form

$$E(c) = \int_a^b L(t, x^i(t), \dot{x}^i(t))\, dt, \tag{2.165}$$

where $[a, b] \subset \mathbb{R}$, and $c : [a, b] \to M$ is a smooth curve, locally expressed by $t \mapsto x^i(t)$, and having the velocity $\dot{x} = (\dot{x}^i(t))$. It is obvious that the non–homogeneity of the Lagrangian L, regarded as a smooth function on the product manifold $\mathbb{R} \times TM$, implies that the energy action functional E is dependent of the parametrizations of every curve c. In order to remove this difficulty, [MKA88, MA94] regard the space $\mathbb{R} \times TM$ like a fibre bundle over M. In this context, the *geometrical invariance group* of $\mathbb{R} \times TM$ is given by

$$\bar{t} = t, \qquad \bar{x}^i = \bar{x}^i(x^j), \qquad \bar{v}^i = \frac{\partial \bar{x}^i}{\partial x^j} v^j. \tag{2.166}$$

The structure of the gauge group (2.166) emphasizes the absolute character of the time t from the classical rheonomic Lagrangian mechanics. At the same time, we point out that the gauge group (2.166) is a subgroup of the *gauge group* of the configuration bundle $J^1(\mathbb{R}, M)$, given as

$$\bar{t} = \bar{t}(t), \qquad \bar{x}^i = \bar{x}^i(x^j), \qquad \bar{v}^i = \frac{\partial \bar{x}^i}{\partial x^j} \frac{dt}{d\bar{t}} v^j. \tag{2.167}$$

In other words, the gauge group (2.167) of the jet bundle $J^1(\mathbb{R}, M)$, from the relativistic rheonomic Lagrangian mechanics is more general than that used in the classical rheonomic Lagrangian mechanics, which ignores the temporal reparametrizations. A deep exposition of the physical aspects of the classical

rheonomic Lagrangian geometry is done by [Ike90], while the classical rheonomic Lagrangian mechanics is done by [Mat82].

Therefore, to remove the parametrization dependence of E, they ignore the temporal repametrizations on $\mathbb{R} \times TM$. Naturally, in these conditions, their energy functional becomes a well defined one, but their approach stands out by the 'absolute' character of the time t.

In a more general geometrical approach, [NU00a, Udr00, Nea02] tried to remove this inconvenience. Following this approach, we regard the mechanical 1–jet space $J^1(\mathbb{R}, M) \equiv \mathbb{R} \times TM$ as a fibre bundle over the base product–manifold $\mathbb{R} \times M$. The gauge group of this bundle of configurations is given by 2.167. Consequently, our gauge group does not ignore the temporal reparametrizations, hence, it stands out by the relativistic character of the time t. In these conditions, using a given semi–Riemannian metric $h_{11}(t)$ on $\mathbb{R}$, we construct the more general and natural energy action functional, setting

$$E(c) = \int_a^b L(t, x^i(t), \dot{x}^i(t)) \sqrt{|h_{11}|}\, dt. \tag{2.168}$$

Obviously, E is well defined and is independent of the curve parametrizations.

Relativistic Rheonomic Lagrangian Spaces

In order to develop the time–dependent Lagrangian geometry, following [NU00a, Udr00, Nea02, NU00b, Nea00], we consider $L : J^1(\mathbb{R}, M) \to \mathbb{R}$ to be a smooth Lagrangian function on the 1–jet bundle $J^1(\mathbb{R}, M) \to \mathbb{R}$, locally expressed by $(t, x^i, v^i) \mapsto L(t, x^i, v^i)$. The so–called *vertical fundamental metrical $d-$tensor* of L is defined by

$$G^{(1)(1)}_{(i)(j)} = \frac{1}{2} \frac{\partial^2 L}{\partial v^i \partial v^j}. \tag{2.169}$$

Let $h = (h_{11})$ be a semi–Riemannian metric on the temporal manifold $\mathbb{R}$.

A Lagrangian function $L : J^1(\mathbb{R}, M) \to \mathbb{R}$ whose vertical fundamental metrical $d-$tensor is of the form

$$G^{(1)(1)}_{(i)(j)}(t, x^k, v^k) = h^{11}(t) g_{ij}(t, x^k, v^k), \tag{2.170}$$

where $g_{ij}(t, x^k, v^k)$ is a $d-$tensor on $J^1(\mathbb{R}, M)$, symmetric, of rank n and having a constant signature on $J^1(\mathbb{R}, M)$, is called a Kronecker $h-$regular Lagrangian function, with respect to the temporal semi–Riemannian metric $h = (h_{11})$.

A pair $\mathbb{R}L^n = (J^1(\mathbb{R}, M), L)$, where $n = \dim M$, which consists of the $1-$jet space $J^1(\mathbb{R}, M)$ and a Kronecker $h-$regular Lagrangian function $L : J^1(T, M) \to \mathbb{R}$ is called a *relativistic rheonomic Lagrangian space*.

In our geometrization of the time–dependent Lagrangian function L that we

will construct, all entities with geometrical or physical meaning will be directly arisen from the vertical fundamental metrical $d-$tensor $G^{(1)(1)}_{(i)(j)}$. This fact points out the metrical character (see [GIM98]) and the naturalness of the subsequent relativistic rheonomic Lagrangian geometry.

For example, suppose that the spatial manifold M is also equipped with a semi–Riemannian metric $g = (g_{ij}(x))$. Then, the time–dependent Lagrangian function $L_1 : J^1(\mathbb{R}, M) \to \mathbb{R}$ defined by

$$L_1 = h^{11}(t) g_{ij}(x) v^i v^j \tag{2.171}$$

is a Kronecker $h-$regular time–dependent Lagrangian function. Consequently, the pair $\mathbb{R}L^n = (J^1(\mathbb{R}, M), L_1)$ is a relativistic rheonomic Lagrangian space. We underline that the Lagrangian $L_1 = L_1\sqrt{|h_{11}|}$ is exactly the energy Lagrangian whose extremals are the harmonic maps between the semi–Riemannian manifolds $(\mathbb{R}, h)$ and (M, g). At the same time, this Lagrangian is a basic object in the physical theory of bosonic strings (compare with subsection 2.2.8 below).

In above notations, taking $U^{(1)}_{(i)}(t, x)$ as a $d-$tensor field on $J^1(\mathbb{R}, M)$ and $F : \mathbb{R} \times M \to \mathbb{R}$ a smooth map, the more general Lagrangian function $L_2 : J^1(\mathbb{R}, M) \to \mathbb{R}$ defined by

$$L_2 = h^{11}(t) g_{ij}(x) v^i v^j + U^{(1)}_{(i)}(t, x) v^i + F(t, x) \tag{2.172}$$

is also a Kronecker $h-$regular Lagrangian. The relativistic rheonomic Lagrangian space $\mathbb{R}L^n = (J^1(\mathbb{R}, M), L_2)$ is called the autonomous relativistic rheonomic Lagrangian space of electrodynamics because, in the particular case $h_{11} = 1$, we recover the classical Lagrangian space of electrodynamics [MKA88, MA94] which governs the movement law of a particle placed concomitantly into a gravitational field and an electromagnetic one. From a physical point of view, the semi–Riemannian metric $h_{11}(t)$ (resp. $g_{ij}(x)$) represents the gravitational potentials of the space $\mathbb{R}$ (resp. M), the $d-$tensor $U^{(1)}_{(i)}(t, x)$ stands for the electromagnetic potentials and F is a function which is called potential function. The non-dynamical character of spatial gravitational potentials $g_{ij}(x)$ motivates us to use the term of 'autonomous'.

More general, if we consider $g_{ij}(t, x)$ a $d-$tensor field on $J^1(\mathbb{R}, M)$, symmetric, of rank n and having a constant signature on $J^1(\mathbb{R}, M)$, we can define the Kronecker $h-$regular Lagrangian function $L_3 : J^1(\mathbb{R}, M) \to \mathbb{R}$, setting

$$L_3 = h^{11}(t) g_{ij}(t, x) v^i v^j + U^{(1)}_{(i)}(t, x) v^i + F(t, x). \tag{2.173}$$

The pair $\mathbb{R}L^n = (J^1(\mathbb{R}, M), L_3)$ is a relativistic rheonomic Lagrangian space which is called the non–autonomous relativistic rheonomic Lagrangian space of electrodynamics. Physically, we remark that the gravitational potentials $g_{ij}(t, x)$ of the spatial manifold M are dependent of the temporal coordinate t, emphasizing their dynamical character.

Canonical Nonlinear Connection

Let us consider $h = (h_{11})$ a fixed semi–Riemannian metric on $\mathbb{R}$ and a rheonomic Lagrangian space $\mathbb{R}L^n = (J^1(\mathbb{R}, M), L)$, where L is a Kronecker $h-$regular Lagrangian function. Let $[a, b] \subset \mathbb{R}$ be a compact interval in the temporal manifold $\mathbb{R}$. In this context, we can define the energy action functional of $\mathbb{R}L^n$, setting

$$E : C^\infty(\mathbb{R}, M) \to \mathbb{R}, \qquad E(c) = \int_a^b L(t, x^i, v^i)\sqrt{|h|}dt,$$

where the smooth curve c is locally expressed by $(t) \to (x^i(t))$ and $v^i = \frac{dx^i}{dt}$.

The extremals of the energy functional E verifies the Euler–Lagrangian equations

$$2G^{(1)(1)}_{(i)(j)}\ddot{x}^j + \frac{\partial^2 L}{\partial x^j \partial v^i}\dot{x}^j - \frac{\partial L}{\partial x^i} + \frac{\partial^2 L}{\partial t \partial v^i} + \frac{\partial L}{\partial v^i}H^1_{11} = 0, \qquad (i = 1, ..., n), \tag{2.174}$$

where H^1_{11} are the *Christoffel symbols* of the semi–Riemannian metric h_{11}.

Taking into account the Kronecker $h-$regularity of the Lagrangian function L, it is possible to rearrange the Euler–Lagrangian equations (2.174) of the Lagrangian $L = L\sqrt{|h|}$, in the Poisson form [NU00a]

$$\begin{aligned} &\Delta_h x^i + 2G^i(t, x^i, v^i) = 0, \qquad (i = 1, ..., n), \qquad \text{where} \\ &\qquad \Delta_h x^i = h^{11}\left\{\ddot{x}^i - H^1_{11}v^i\right\}, \qquad v^i = \dot{x}^i, \\ &2G^i = \frac{g^{ii}}{2}\left\{\frac{\partial^2 L}{\partial x^j \partial v^i}v^j - \frac{\partial L}{\partial x^i} + \frac{\partial^2 L}{\partial t \partial v^i} + \frac{\partial L}{\partial v^i}H^1_{11} + 2g_{ij}h^{11}H^1_{11}v^j\right\}. \end{aligned} \tag{2.175}$$

Denoting $G^{(r)}_{(1)1} = h_{11}G^r$, the geometrical object $G = (G^{(r)}_{(1)1})$ is a spatial spray on $J^1(\mathbb{R}, M)$. By a direct calculation, we deduce that the local geometrical entities of $J^1(\mathbb{R}, M)$

$$\begin{aligned} 2S^k &= \frac{g^{ki}}{2}\left\{\frac{\partial^2 L}{\partial x^j \partial v^i}v^j - \frac{\partial L}{\partial x^i}\right\}, \\ 2H^k &= \frac{g^{ki}}{2}\left\{\frac{\partial^2 L}{\partial t \partial v^i} + \frac{\partial L}{\partial v^i}H^1_{11}\right\}, \qquad 2J^k = h^{11}H^1_{11}v^j, \end{aligned}$$

verify the following transformation rules

$$\begin{aligned} &2S^p = 2\bar{\mathcal{S}}^r\frac{\partial x^p}{\partial \bar{x}^r} + h^{11}\frac{\partial x^p}{\partial \bar{x}^l}\frac{d\bar{t}}{dt}\frac{\partial \bar{x}^l_\gamma}{\partial x^j}v^j, \qquad 2H^p = 2\bar{\mathcal{H}}^r\frac{\partial x^p}{\partial \bar{x}^r} + h^{11}\frac{\partial x^p}{\partial \bar{x}^l}\frac{d\bar{t}}{dt}\frac{\partial \bar{v}^l}{\partial t}, \\ &2J^p = 2\bar{\mathcal{J}}^r\frac{\partial x^p}{\partial \bar{x}^r} - h^{11}\frac{\partial x^p}{\partial \bar{x}^l}\frac{d\bar{t}}{dt}\frac{\partial \bar{v}^l}{\partial t}. \end{aligned}$$

Consequently, the local entities $2G^p = 2S^p + 2H^p + 2J^p$ can be modified by the transformation laws

$$2\bar{\mathcal{G}}^r = 2\mathcal{G}^p \frac{\partial \bar{x}^r}{\partial x^p} - h^{11} \frac{\partial x^p}{\partial \bar{x}^j} \frac{\partial \bar{x}^r_\mu}{\partial x^p} \bar{v}^j. \tag{2.176}$$

The extremals of the energy functional attached to a Kronecker $h-$regular Lagrangian function L on $J^1(\mathbb{R}, M)$ are harmonic curves of the time–dependent spray (H, G), with respect to the semi–Riemannian metric h, defined by the temporal components

$$H^{(i)}_{(1)1} = -\frac{1}{2} H^1_{11}(t) v^i$$

and the local spatial components

$$G^{(i)}_{(1)1} = \frac{h_{11} g^{ik}}{4} \left[\frac{\partial^2 L}{\partial x^j \partial v^k} v^j - \frac{\partial L}{\partial x^k} + \frac{\partial^2 L}{\partial t \partial v^k} + \frac{\partial L}{\partial x^k} H^1_{11} + 2h^{11} H^1_{11} g_{kl} v^l \right].$$

The time–dependent spray (H, G) constructed from the previous theorem is called the canonical time–dependent spray attached to the relativistic rheonomic Lagrangian space $\mathbb{R}L^n$.

In the particular case of an autonomous electrodynamics relativistic rheonomic Lagrangian space (i.e., $g_{ij}(t, x^k, v^k) = g_{ij}(x^k)$), the canonical spatial spray G is given by the components

$$G^{(i)}_{(1)1} = \frac{1}{2} \gamma^i_{jk} v^j v^k + \frac{h_{11} g^{li}}{4} \left[U^{(1)}_{(l)j} v^j + \frac{\partial U^{(1)}_{(l)}}{\partial t} + U^{(1)}_{(l)} H^1_{11} - \frac{\partial F}{\partial x^l} \right], \tag{2.177}$$

where $U^{(1)}_{(i)j} = \frac{\partial U^{(1)}_{(i)}}{\partial x^j} - \frac{\partial U^{(1)}_{(j)}}{\partial x^i}$.

We have the following theorem: The pair of local functions $\Gamma = (M^{(i)}_{(1)1}, N^{(i)}_{(1)j})$, which consists of the temporal components

$$M^{(i)}_{(1)1} = 2H^{(i)}_{(1)1} = -H^1_{11} v^i, \tag{2.178}$$

and the spatial components

$$N^{(i)}_{(1)j} = \frac{\partial G^i_{(1)1}}{\partial v^j}, \tag{2.179}$$

where $H^{(i)}_{(1)1}$ and $G^{(i)}_{(1)1}$ are the components of the canonical time–dependent spray of $\mathbb{R}L^n$, represents a nonlinear connection on $J^1(\mathbb{R}, M)$.

The nonlinear connection $\Gamma = (M^{(i)}_{(1)1}, N^{(i)}_{(1)j})$ from the preceding theorem is called the canonical nonlinear connection of the relativistic rheonomic Lagrangian space $\mathbb{R}L^n$.

In the case of an autonomous electrodynamics relativistic rheonomic Lagrangian space (i.e., $g_{ij}(t, x^k, v^k) = g_{ij}(x^k)$), the canonical nonlinear connection becomes $\Gamma = (M^{(i)}_{(1)1}, N^{(i)}_{(1)j})$, where

$$M^{(i)}_{(1)1} = -H^1_{11} v^i, \quad N^{(i)}_{(1)j} = \gamma^i_{jk} v^k + \frac{h_{11} g^{ik}}{4} U^{(1)}_{(k)j}. \tag{2.180}$$

Cartan's Canonical Connection

The main theorem of this paper is the theorem of existence of the Cartan canonical $h-$normal linear connection $C\Gamma$ which allow the subsequent development of the relativistic rheonomic Lagrangian geometry of physical fields, which will be exposed in the next sections.

On the relativistic rheonomic Lagrangian space $\mathbb{R}L^n = (J^1(\mathbb{R}, M), L)$ equipped with its canonical nonlinear connection Γ there is a unique $h-$normal $\Gamma-$linear connection

$$C\Gamma = (H^1_{11}, G^k_{j1}, L^i_{jk}, C^{i(1)}_{j(k)})$$

having the metrical properties:

(i) $g_{ij|k} = 0, \quad g_{ij}|^{(1)}_{(k)} = 0,$

(ii) $G^k_{j1} = \frac{g^{ki}}{2}\frac{\delta g_{ij}}{\delta t}, \quad L^k_{ij} = L^k_{ji}, \quad C^{i(1)}_{j(k)} = C^{i(1)}_{k(j)}.$

To prove this theorem, let $C\Gamma = (\bar{G}^1_{11}, G^k_{j1}, L^i_{jk}, C^{i(1)}_{j(k)})$ be a $h-$normal $\Gamma-$linear connection whose coefficients are defined by $\bar{G}^1_{11} = H^1_{11}$, $G^k_{j1} = \frac{g^{ki}}{2}\frac{\delta g_{ij}}{\delta t}$, and

$$L^i_{jk} = \frac{g^{im}}{2}\left(\frac{\delta g_{jm}}{\delta x^k} + \frac{\delta g_{km}}{\delta x^j} - \frac{\delta g_{jk}}{\delta x^m}\right), \quad C^{i(1)}_{j(k)} = \frac{g^{im}}{2}\left(\frac{\partial g_{jm}}{\partial v^k} + \frac{\partial g_{km}}{\partial v^j} - \frac{\partial g_{jk}}{\partial v^m}\right).$$

By computations, one can verify that $C\Gamma$ satisfies the conditions (i) and (ii).

Conversely, let us consider $\bar{C}\Gamma = (\bar{\bar{G}}^1_{11}, \bar{G}^k_{j1}, \bar{L}^i_{jk}, \bar{C}^{i(1)}_{j(k)})$ a $h-$normal $\Gamma-$linear connection which satisfies (i) and (ii). It follows directly that

$$\bar{\bar{G}}^1_{11} = H^1_{11}, \qquad \text{and} \qquad \bar{G}^k_{j1} = \frac{g^{ki}}{2}\frac{\delta g_{ij}}{\delta t}.$$

The condition $g_{ij|k} = 0$ is equivalent with

$$\frac{\delta g_{ij}}{\delta x^k} = g_{mj}\bar{L}^m_{ik} + g_{im}\bar{L}^m_{jk}.$$

Applying the Christoffel process to the indices $\{i, j, k\}$, we find

$$\bar{L}^i_{jk} = \frac{g^{im}}{2}\left(\frac{\delta g_{jm}}{\delta x^k} + \frac{\delta g_{km}}{\delta x^j} - \frac{\delta g_{jk}}{\delta x^m}\right).$$

By analogy, using the relations $C^{i(1)}_{j(k)} = C^{i(1)}_{k(j)}$ and $g_{ij}|^{(1)}_{(k)} = 0$, following a Christoffel process applied to the indices $\{i, j, k\}$, we get

$$\bar{C}^{i(1)}_{j(k)} = \frac{g^{im}}{2}\left(\frac{\partial g_{jm}}{\partial v^k} + \frac{\partial g_{km}}{\partial v^j} - \frac{\partial g_{jk}}{\partial v^m}\right).$$

As a rule, the Cartan canonical connection of a relativistic rheonomic Lagrangian space $\mathbb{R}L^n$ verifies also the properties

$$h_{11/1} = h_{11|k} = h_{11}|^{(1)}_{(k)} = 0 \text{ and } g_{ij/1} = 0. \tag{2.181}$$

Particularly, the coefficients of the Cartan connection of an autonomous relativistic rheonomic Lagrangian space of electrodynamics (i.e., $g_{ij}(t, x^k, v^k) = g_{ij}(x^k)$) are the same with those of the Berwald connection, namely, $C\Gamma = (H^1_{11}, 0, \gamma^i_{jk}, 0)$. Note that the Cartan connection is a $\Gamma-$linear connection, where Γ is the canonical nonlinear connection of the relativistic rheonomic Lagrangian space while the Berwald connection is a Γ_0-linear connection, Γ_0 being the canonical nonlinear connection associated to the metric pair (h_{11}, g_{ij}). Consequently, the Cartan and Berwald connections are distinct.

The torsion $d-$tensor T of the Cartan canonical connection of a relativistic rheonomic Lagrangian space is determined by only six local components, because the properties of the Cartan canonical connection imply the relations $T^m_{ij} = 0$ and $S^{(i)(1)(1)}_{(1)(j)(k)} = 0$. At the same time, we point out that the number of the curvature local $d-$tensors of the Cartan canonical connection not reduces. In conclusion, the curvature $d-$tensor R of the Cartan canonical connection is determined by five effective local $d-$tensors. The torsion and curvature $d-$tensors of the Cartan canonical connection of an $\mathbb{R}L^n$ are called the torsion and curvature of $\mathbb{R}L^n$.

All torsion $d-$tensors of an autonomous relativistic rheonomic Lagrangian space of electrodynamics vanish, except

$$R^{(m)}_{(1)1j} = -\frac{h_{11}g^{mk}}{4}\left[H^1_{11}U^{(1)}_{(k)j} + \frac{\partial U^{(1)}_{(k)j}}{\partial t}\right],$$

$$R^{(m)}_{(1)ij} = r^m_{ijk}v^k + \frac{h_{11}g^{mk}}{4}\left[U^{(1)}_{(k)i|j} + U^{(1)}_{(k)j|i}\right],$$

where r^m_{ijk} are the curvature tensors of the semi–Riemannian metric g_{ij}.

General Nonlinear Connections

Recall that a *nonlinear connection* (i.e., a supplementary–horizontal distribution of the vertical distribution of $J^1(\mathbb{R}, M)$) offers the possibility of construction of the vector or covector adapted bases on $J^1(\mathbb{R}, M)$) [NU00a]. A nonlinear connection Γ on $J^1(\mathbb{R}, M)$ is determined by a pair of local function sets $M^{(i)}_{(1)1}$ and $N^{(i)}_{(1)j}$ which modify by the transformation laws

$$\bar{M}^{(j)}_{(1)1}\frac{d\bar{t}}{dt} = M^{(k)}_{(1)1}\frac{dt}{d\bar{t}}\frac{\partial \bar{x}^j}{\partial x^k} - \frac{\partial \bar{v}^j}{\partial t}, \qquad \bar{N}^{(j)}_{(1)k}\frac{\partial \bar{x}^k}{\partial x^i} = N^{(k)}_{(1)i}\frac{dt}{d\bar{t}}\frac{\partial \bar{x}^j}{\partial x^k} - \frac{\partial \bar{v}^j}{\partial x^i}. \tag{2.182}$$

A set of local functions $M^{(i)}_{(1)1}$ (resp. $N^{(i)}_{(1)j}$) on $J^1(\mathbb{R}, M)$, which transform by the rules (2.182) is called a *temporal nonlinear connection* (resp. spatial nonlinear connection) on $J^1(\mathbb{R}, M)$.

For example, studying the transformation rules of the local components

$$M^{(i)}_{(1)1} = -H^1_{11} v^i, \qquad N^{(i)}_{(1)j} = \gamma^i_{jk} v^k,$$

where H^1_{11} (resp. γ^i_{jk}) are the Christoffel symbols of a temporal (resp. spatial) semi–Riemannian metric h (resp. φ), we conclude that $\Gamma_0 = (M^{(i)}_{(1)1}, N^{(i)}_{(1)j})$ represents a nonlinear connection on $J^1(\mathbb{R}, M)$, which is called the canonical nonlinear connection attached to the metric pair (h, φ).

If $M^{(i)}_{(1)1}$ are the components of a temporal nonlinear connection, then the components $H^{(i)}_{(1)1} = \frac{1}{2} M^{(i)}_{(1)1}$ represent a temporal spray. Conversely, if $H^{(i)}_{(1)1}$ are the components of a temporal spray, then $M^{(i)}_{(1)1} = 2H^{(i)}_{(1)1}$ are the components of a temporal nonlinear connection. If $G^{(i)}_{(1)1}$ are the components of a spatial spray, then the components $N^{(i)}_{(1)j} = \frac{\partial G^i_{(1)1}}{\partial v^j}$ represent a spatial nonlinear connection.

Conversely, the spatial nonlinear connection $N^{(i)}_{(1)j}$ induces the spatial spray $2G^{(i)}_{(1)1} = N^{(i)}_{(1)j} v^j$.

The previous theorems allow us to conclude that a time–dependent spray (H, G) induces naturally a nonlinear connection Γ on $J^1(\mathbb{R}, M)$, which is called the canonical nonlinear connection associated to the time–dependent spray (H, G). We point out that the canonical nonlinear connection Γ attached to the time–dependent spray (H, G) is a natural generalization of the canonical nonlinear connection N induced by a time–dependent spray G from the classical rheonomic Lagrangian geometry [MKA88, MA94].

Let $\Gamma = (M^{(i)}_{(1)1} N^{(i)}_{(1)j})$ be a nonlinear connection on $J^1(\mathbb{R}, M)$. Let us consider the geometrical objects,

$$\frac{\delta}{\delta t} = \frac{\partial}{\partial t} - M^{(j)}_{(1)1} \frac{\partial}{\partial v^j}, \quad \frac{\delta}{\delta x^i} = \frac{\partial}{\partial x^i} - N^{(j)}_{(1)i} \frac{\partial}{\partial v^j}, \quad \delta v^i = dy^i + M^{(i)}_{(1)1} dt + N^{(i)}_{(1)j} dx^j.$$

One can deduce that the set of vector–fields $\left\{ \frac{\delta}{\delta t}, \frac{\delta}{\delta x^i}, \frac{\partial}{\partial v^i} \right\} \subset \mathcal{X}(J^1(\mathbb{R}, M))$ and of covector–fields $\{dt, dx^i, \delta v^i\} \subset \mathcal{X}^*(J^1(\mathbb{R}, M))$ are dual bases. These are called the adapted bases on $J^1(\mathbb{R}, M)$, determined by the nonlinear connection Γ. The big advantage of the adapted bases is that the transformation laws of its elements are simple and natural. The transformation laws of the elements of the adapted bases attached to the nonlinear connection Γ are

$$\frac{\delta}{\delta t} = \frac{d\bar{t}}{dt} \frac{\delta}{\delta \bar{t}}, \qquad \frac{\delta}{\delta x^i} = \frac{\partial \bar{x}^j}{\partial x^i} \frac{\delta}{\delta \bar{x}^j}, \qquad \frac{\partial}{\partial v^i} = \frac{\partial \bar{x}^j}{\partial x^i} \frac{dt}{d\bar{t}} \frac{\delta}{\delta \bar{v}^j},$$
$$dt = \frac{dt}{d\bar{t}} d\bar{t}, \qquad dx^i = \frac{\partial x^i}{\partial \bar{x}^j} d\bar{x}^j, \qquad \delta v^i = \frac{\partial x^i}{\partial \bar{x}^j} \frac{d\bar{t}}{dt} \delta \bar{v}^j.$$

2.1.5 Geometrical Quantization

Quantization of Hamiltonian Mechanics

Recall that classical *Dirac quantization* states [Dir82]:

$$\{f, g\} = \frac{1}{i\hbar}[\hat{f}, \hat{g}],$$

which means that the quantum Poisson brackets (i.e., commutators) have the same values as the classical Poisson brackets. In other words, we can associate smooth functions defined on the symplectic phase–space manifold (M, ω) of the classical biodynamic system with operators on a Hilbert space $\mathcal{H}$ in such a way that the Poisson brackets correspond. Therefore, *there is a functor* from the category **Symplec** to the category **Hilbert**. This functor is called *prequantization*.

Let us start with the simplest symplectic manifold ($M = T^*\mathbb{R}^n$, $\omega = dp_i \wedge dq^i$) and state the *Dirac problem*: A prequantization of ($T^*\mathbb{R}^n$, $\omega = dp_i \wedge dq^i$) is a map $\delta : f \mapsto \delta_f$, taking smooth functions $f \in C^\infty(T^*\mathbb{R}^n, \mathbb{R})$ to Hermitian operators δ_f on a Hilbert space $\mathcal{H}$, satisfying the *Dirac conditions*:

1. $\delta_{f+g} = \delta_f + \delta_g$, for each $f, g \in C^\infty(T^*\mathbb{R}^n, \mathbb{R})$;
2. $\delta_{\lambda f} = \lambda \delta_f$, for each $f \in C^\infty(T^*\mathbb{R}^n, \mathbb{R})$ and $\lambda \in \mathbb{R}$;
3. $\delta_{1_{\mathbb{R}^n}} = Id_{\mathcal{H}}$; and
4. $[\delta_f, \delta_g] = (\delta_f \circ \delta_g - \delta_g \circ \delta_f) = i\hbar\delta_{\{f,g\}_\omega}$, for each $f, g \in C^\infty(T^*\mathbb{R}^n, \mathbb{R})$;

The pair $(\mathcal{H}, \delta)$, where

$$\mathcal{H} = L^2(\mathbb{R}^n, \mathbb{C}); \qquad \delta : f \in C^\infty(T^*\mathbb{R}^n, \mathbb{R}) \mapsto \delta_f : \mathcal{H} \to \mathcal{H};$$
$$\delta_f = -i\hbar X_f - \theta(X_f) + f; \qquad \theta = p_i dq^i,$$

gives a prequantization of ($T^*\mathbb{R}^n$, $dp_i \wedge dq^i$), or equivalently, the answer to the Dirac problem is affirmative [Put93].

Now, let $(M = T^*Q, \omega)$ be the cotangent bundle of an arbitrary manifold Q with its canonical symplectic structure $\omega = d\theta$. The prequantization of M is given by the pair $\left(L^2(M, \mathbb{C}), \delta^\theta\right)$, where for each $f \in C^\infty(M, \mathbb{R})$, the operator $\delta_f^\theta : L^2(M, \mathbb{C}) \to L^2(M, \mathbb{C})$ is given by

$$\delta_f^\theta = -i\hbar X_f - \theta(X_f) + f.$$

Here, *symplectic potential* θ is not uniquely determined by the condition $\omega = d\theta$; for instance $\theta' = \theta + du$ has the same property for any real function u on M. On the other hand, in the general case of an arbitrary symplectic manifold (M, ω) (not necessarily the cotangent bundle) we can find only locally a 1–form θ such that $\omega = d\theta$.

In general, a symplectic manifold $(M, \omega = d\theta)$ is quantizable (i.e., we can define the Hilbert representation space $\mathcal{H}$ and the prequantum operator δ_f

in a globally consistent way) if ω defines an *integral cohomology class*. Now, by the *construction theorem of a fiber bundle*, we can see that this condition on ω is also sufficient to guarantee the existence of a *complex line bundle* $L^\omega = (L, \pi, M)$ over M, which has $\exp(i\, u_{ji}/\hbar)$ as *gauge transformations* associated to an open cover $\mathcal{U} = \{U_i | i \in I\}$ of M such that θ_i is a *symplectic potential* defined on U_i (i.e., $d\theta_i = \omega$ and $\theta_i = \theta_i + d\, u_{ji}$ on $U_i \cap U_j$).

In particular, for exact symplectic structures ω (as in the case of cotangent bundles with their canonical symplectic structures) an integral cohomology condition is automatically satisfied, since then we have only one set $U_i = M$ and do not need any gauge transformations.

Now, for each vector–field $X \in M$ there exists an operator ∇^ω_X on the *space of sections* $\Gamma(L^\omega)$ of L^ω,

$$\nabla^\omega_X : \Gamma(L^\omega) \to \Gamma(L^\omega), \qquad \text{given by} \qquad \nabla^\omega_X f = X(f) - \frac{\mathrm{i}}{\hbar}\theta(X) f,$$

and it is easy to see that ∇^ω is a *connection* on L^ω whose curvature is $\omega/i\hbar$. In terms of this connection, the definition of δ_f becomes

$$\delta_f = -i\hbar \nabla^\omega_{X_f} + f.$$

The complex line bundle $L^\omega = (L, \pi, M)$ together with its compatible connection and Hermitian structure is usually called the *prequantum bundle* of the symplectic manifold (M, ω).

If (M, ω) is a quantizable manifold then the pair $(\mathcal{H}, \delta)$ defines its prequantization.

Examples

Each exact symplectic manifold $(M, \omega = d\theta)$ is quantizable, for the cohomology class defined by ω is zero. In particular, the cotangent bundle, with its canonical symplectic structure is always quantizable.

Let $(M, \omega = d\theta)$ be an exact symplectic manifold. Then it is quantizable with the prequantum bundle given by [Put93]:

$$\begin{aligned}
&L^\omega = (M \times \mathbb{C}, pr_1, M);\\
&\Gamma(L^\omega) \simeq C^\infty(M, \mathbb{C}); \qquad \nabla^\omega_X f = X(f) - \frac{\mathrm{i}}{\hbar}\theta(X) f;\\
&((x, z_1), (x, z_2))_x = \bar{z}_1 z_2; \qquad \delta_f = -i\hbar[X_f - \frac{i}{\hbar}\theta(X_f)] + f.
\end{aligned}$$

Let $(M, \omega) = (T^*\mathbb{R}, dp \wedge dq)$. It is quantizable with [Put93]:

$$\begin{aligned}
L^\omega &= (\mathbb{R}^2 \times \mathbb{C}, pr_1, \mathbb{R}^2); \qquad \Gamma(L^\omega) = C^\infty(\mathbb{R}^2, \mathbb{C});\\
\nabla^\omega_X f &= X(f) - \frac{\mathrm{i}}{\hbar} pdq(X) f; \qquad ((x, z_1), (x, z_2))_x = \bar{z}_1 z_2;\\
\delta_f &= -i\hbar \left[\frac{\partial f}{\partial p}\frac{\partial}{\partial q} - \frac{\partial f}{\partial q}\frac{\partial}{\partial p}\right] - p\frac{\partial f}{\partial p} + f.
\end{aligned}$$

Therefore,

$$\delta_q = i\hbar\frac{\partial}{\partial p} + q, \qquad \delta_p = -i\hbar\frac{\partial}{\partial q},$$

which differs from the classical result of the *Schrödinger quantization*:

$$\delta_q = q, \qquad \delta_p = -i\hbar\frac{\partial}{\partial q}.$$

Let $\mathcal{H}$ be a complex Hilbert space and $U_t : \mathcal{H} \to \mathcal{H}$ a continuous one–parameter unitary group, i.e., a homomorphism $t \mapsto U_t$ from $\mathbb{R}$ to the group of unitary operators on $\mathcal{H}$ such that for each $x \in \mathcal{H}$ the map $t \mapsto U_t(x)$ is continuous. Then we have the self–adjoint generator A of U_t, defined by

$$Ax = \frac{1}{i}\frac{d}{dt}U_t(x) = \frac{1}{i}\lim_{h\to 0}\frac{U_h(x) - x}{h}.$$

Let $\left(\mathbb{R}^2, \omega = dp \wedge dq, H = \frac{1}{2}(p^2 + q^2)\right.$ be the Hamiltonian structure of the 1D harmonic oscillator.

If we take $\theta = \frac{1}{2}(pdq - qdp)$ as the symplectic potential of ω, then the spectrum of the prequantum operator $\delta_H = i\hbar\left(q\frac{\partial}{\partial p} - p\frac{\partial}{\partial q}\right)$ is [Put93] $Spec(\delta_H) = \{..., -2\hbar, -\hbar, 0, \hbar, 2\hbar, ...\}$, where each eigenvalue occurs with infinite multiplicity.

Let $\mathfrak{g}$ be the vector space spanned by the prequantum operators $\delta_q, \delta_p, \delta_H$, given by

$$\delta_q = i\hbar\frac{\partial}{\partial p} + q, \qquad \delta_p = -i\hbar\frac{\partial}{\partial q}, \qquad \delta_H = i\hbar\left(q\frac{\partial}{\partial p} - p\frac{\partial}{\partial q}\right),$$

and Id. Then we have [Put93]:

1. $\mathfrak{g}$ is a Lie algebra called the *oscillator Lie algebra*, given by:

$$\begin{aligned}
[\delta_p, \delta_q] &= i\hbar\delta_{\{p,q\}_\omega} = i\hbar\, Id,\\
[\delta_H, \delta_q] &= i\hbar\delta_{\{H,q\}_\omega} = -i\hbar\delta_p,\\
[\delta_H, \delta_p] &= i\hbar\delta_{\{H,p\}_\omega} = i\hbar\delta_q,
\end{aligned}$$

2. $[\mathfrak{g}, \mathfrak{g}]$ is spanned by $\delta_q, \delta_p, \delta_H$ and Id, or equivalently, it is a *Heisenberg Lie algebra*.
3. The oscillator Lie algebra $\mathfrak{g}$ is *solvable*.

Quantization of Relativistic Hamiltonian Mechanics

Given a symplectic manifold (Z, Ω) and a Hamiltonian H on Z, a Dirac constraint system on a closed imbedded submanifold $i_N : N \longrightarrow Z$ of Z is defined as a Hamiltonian system on N admitting the pull–back presymplectic

form $\Omega_N = i_N^* \Omega$ and the pull–back Hamiltonian $i_N^* H$ [GNH78, MS98, MR92]. Its solution is a vector–field γ on N which fulfils the equation

$$\gamma \rfloor \Omega_N + i_N^* dH = 0.$$

Let N be coisotropic. Then a solution exists if the Poisson bracket $\{H, f\}$ vanishes on N whenever f is a function vanishing on N. It is the Hamiltonian vector–field of H on Z restricted to N [Sar03].

Recall that a configuration space of non–relativistic time–dependent mechanics (henceforth NRM) of m degrees of freedom is an $(m+1)$D smooth fibre bundle $Q \longrightarrow \mathbb{R}$ over the time axis $\mathbb{R}$ [MS98, Sar98]. It is coordinated by $(q^\alpha) = (q^0, q^i)$, where $q^0 = t$ is the standard Cartesian coordinate on $\mathbb{R}$. Let T^*Q be the cotangent bundle of Q equipped with the induced coordinates $(q^\alpha, p_\alpha = \dot{q}_\alpha)$ with respect to the holonomic coframes $\{dq^\alpha\}$. The cotangent bundle T^*Q plays the role of a homogeneous momentum phase–space of NRM, admitting the canonical symplectic form

$$\Omega = dp_\alpha \wedge dq^\alpha. \tag{2.183}$$

Its momentum phase–space is the vertical cotangent bundle V^*Q of the configuration bundle $Q \longrightarrow \mathbb{R}$, coordinated by (q^α, q^i). A Hamiltonian $\mathcal{H}$ of NRM is defined as a section $p_0 = -\mathcal{H}$ of the fibre bundle $T^*Q \longrightarrow V^*Q$. Then the momentum phase–space of NRM can be identified with the image N of $\mathcal{H}$ in T^*Q which is the one-codimensional (consequently, coisotropic) imbedded submanifold given by the constraint

$$\mathcal{H}_T = p_0 + \mathcal{H}(q^\alpha, p_k) = 0.$$

Furthermore, a solution of a non–relativistic Hamiltonian system with a Hamiltonian $\mathcal{H}$ is the restriction γ to $N \cong V^*Q$ of the Hamiltonian vector–field of $\mathcal{H}_T$ on T^*Q. It obeys the equation $\gamma \rfloor \Omega_N = 0$ [MS98, Sar98]. Moreover, one can show that geometrical quantization of V^*Q is equivalent to geometrical quantization of the cotangent bundle T^*Q where the quantum constraint $\widehat{\mathcal{H}}_T \psi = 0$ on sections ψ of the quantum bundle serves as the Schrödinger equation [Sar03].

A configuration space of *relativistic mechanics* (henceforth RM) is an oriented pseudo–Riemannian manifold (Q, g), coordinated by (t, q^i). Its momentum phase–space is the cotangent bundle T^*Q provided with the symplectic form Ω (2.183). Note that one also considers another symplectic form $\Omega + F$ where F is the strength of an electromagnetic field [Sni80]. A relativistic Hamiltonian is defined as a smooth real function H on T^*Q [MS98, Sar98]. Then a relativistic Hamiltonian system is described as a Dirac constraint system on the subspace N of T^*Q given by the equation

$$H_T = g_{\mu\nu} \partial^\mu H \partial^\nu H - 1 = 0. \tag{2.184}$$

To perform geometrical quantization of NRM, we give geometrical quantization of the cotangent bundle T^*Q and characterize a quantum relativistic Hamiltonian system by the quantum constraint

$$\widehat{H}_T\psi = 0. \tag{2.185}$$

We choose the vertical polarization on T^*Q spanned by the tangent vectors ∂^α. The corresponding *quantum algebra* $\mathcal{A} \subset C^\infty(T^*Q)$ consists of affine functions of momenta

$$f = a^\alpha(q^\mu)p_\alpha + b(q^\mu) \tag{2.186}$$

on T^*Q. They are represented by the *Schrödinger operators*

$$\widehat{f} = -ia^\alpha\partial_\alpha - \frac{\mathrm{i}}{2}\partial_\alpha a^\alpha - \frac{i}{4}a^\alpha\partial_\alpha \ln(-g) + b, \qquad (g = \det(g_{\alpha\beta})) \tag{2.187}$$

in the space $\mathbb{C}^\infty(Q)$ of smooth complex functions on Q.

Note that the function H_T (2.184) need not belong to the quantum algebra $\mathcal{A}$. Nevertheless, one can show that, if H_T is a polynomial of momenta of degree k, it can be represented as a finite composition

$$H_T = \sum_i f_{1i}\cdots f_{ki} \tag{2.188}$$

of products of affine functions (2.186), i.e., as an element of the enveloping algebra $\overline{\mathcal{A}}$ of the Lie algebra $\mathcal{A}$ [GMS02b]. Then it is quantized

$$H_T \mapsto \widehat{H}_T = \sum_i \widehat{f}_{1i}\cdots\widehat{f}_{ki} \tag{2.189}$$

as an element of $\overline{\mathcal{A}}$. However, the representation (2.188) and, consequently, the quantization (2.189) fail to be unique.

The space of relativistic velocities of RM on Q is the tangent bundle TQ of Q equipped with the induced coordinates $(t, q^i, \dot{q}^\alpha)$ with respect to the holonomic frames $\{\partial_\alpha\}$. Relativistic motion is located in the subbundle W_g of hyperboloids [MS98, MS00b]

$$g_{\mu\nu}(q)\dot{q}^\mu\dot{q}^\nu - 1 = 0 \tag{2.190}$$

of TQ. It is described by a second–order dynamical equation

$$\ddot{q}^\alpha = \Xi^\alpha(q^\mu, \dot{q}^\mu) \tag{2.191}$$

on Q which preserves the subbundle (2.190), i.e.,

$$(\dot{q}^\alpha\partial_\alpha + \Xi^\alpha\dot{\partial}_\alpha)(g_{\mu\nu}\dot{q}^\mu\dot{q}^\nu - 1) = 0, \qquad (\dot{\partial}_\alpha = \partial/\partial\dot{q}^\alpha).$$

This condition holds if the r.h.s. of the equation (2.191) takes the form

$$\Xi^\alpha = \Gamma^\alpha_{\mu\nu}\dot{q}^\mu\dot{q}^\nu + F^\alpha,$$

where $\Gamma^\alpha_{\mu\nu}$ are Christoffel symbols of a metric g, while F^α obey the relation $g_{\mu\nu}F^\mu\dot{q}^\nu = 0$. In particular, if the dynamical equation (2.191) is a geodesic equation,

$$\ddot{q}^\alpha = K^\alpha_\mu \dot{q}^\mu$$

with respect to a (non-linear) connection on the tangent bundle $TQ \to Q$,

$$K = dq^\alpha \otimes (\partial_\alpha + K^\mu_\alpha \dot{\partial}_\mu),$$

this connections splits into the sum

$$K^\alpha_\mu = \Gamma^\alpha_{\mu\nu}\dot{q}^\nu + F^\alpha_\mu \tag{2.192}$$

of the Levi–Civita connection of g and a soldering form

$$F = g^{\lambda\nu}F_{\mu\nu}dq^\mu \otimes \dot{\partial}_\alpha, \qquad F_{\mu\nu} = -F_{\nu\mu}.$$

As was mentioned above, the momentum phase–space of RM on Q is the cotangent bundle T^*Q provided with the symplectic form Ω (2.183). Let H be a smooth real function on T^*Q such that the map

$$\widetilde{H} : T^*Q \longrightarrow TQ, \qquad \dot{q}^\mu = \partial^\mu H \tag{2.193}$$

is a bundle isomorphism. Then the inverse image $N = \widetilde{H}^{-1}(W_g)$ of the subbundle of hyperboloids W_g (2.190) is a one-codimensional (consequently, coisotropic) closed imbedded subbundle of T^*Q given by the constraint $H_T = 0$ (2.184). We say that H is a relativistic Hamiltonian if the Poisson bracket $\{H, H_T\}$ vanishes on N. This means that the Hamiltonian vector–field

$$\gamma = \partial^\alpha H\partial_\alpha - \partial_\alpha H\partial^\alpha \tag{2.194}$$

of H preserves the constraint N and, restricted to N, it obeys the Hamiltonian equation

$$\gamma \rfloor \Omega_N + i^*_N dH = 0 \tag{2.195}$$

of a Dirac constraint system on N with a Hamiltonian H.

The map (2.193) sends the vector–field γ (2.194) onto the vector–field

$$\gamma_T = \dot{q}^\alpha\partial_\alpha + (\partial^\mu H\partial^\alpha\partial_\mu H - \partial_\mu H\partial^\alpha\partial^\mu H)\dot{\partial}_\alpha$$

on TQ. This vector–field defines the second–order dynamical equation

$$\ddot{q}^\alpha = \partial^\mu H\partial^\alpha\partial_\mu H - \partial_\mu H\partial^\alpha\partial^\mu H \tag{2.196}$$

on Q which preserves the subbundle of hyperboloids (2.190).

The following is a basic example of relativistic Hamiltonian systems. Put

$$H = \frac{1}{2m}g^{\mu\nu}(p_\mu - b_\mu)(p_\nu - b_\nu),$$

where m is a constant and $b_\mu dq^\mu$ is a covector–field on Q. Then $H_T = 2m^{-1}H - 1$ and $\{H, H_T\} = 0$. The constraint $H_T = 0$ defines a closed

imbedded one-codimensional subbundle N of T^*Q. The Hamiltonian equation (2.195) takes the form $\gamma \rfloor \Omega_N = 0$. Its solution (2.194) reads

$$\dot{q}^\alpha = \frac{1}{m} g^{\alpha\nu}(p_\nu - b_\nu),$$
$$\dot{p}_\alpha = -\frac{1}{2m}\partial_\alpha g^{\mu\nu}(p_\mu - b_\mu)(p_\nu - b_\nu) + \frac{1}{m} g^{\mu\nu}(p_\mu - b_\mu)\partial_\alpha b_\nu.$$

The corresponding second–order dynamical equation (2.196) on Q is

$$\ddot{q}^\alpha = \Gamma^\alpha_{\mu\nu}\dot{q}^\mu\dot{q}^\nu - \frac{1}{m} g^{\lambda\nu} F_{\mu\nu}\dot{q}^\mu, \tag{2.197}$$
$$\Gamma^\alpha_{\mu\nu} = -\frac{1}{2} g^{\lambda\beta}(\partial_\mu g_{\beta\nu} + \partial_\nu g_{\beta\mu} - \partial_\beta g_{\mu\nu}), \qquad F_{\mu\nu} = \partial_\mu b_\nu - \partial_\nu b_\mu.$$

It is a geodesic equation with respect to the affine connection

$$K^\alpha_\mu = \Gamma^\alpha_{\mu\nu}\dot{q}^\nu - \frac{1}{m} g^{\lambda\nu} F_{\mu\nu}$$

of type (2.192). For example, let g be a metric gravitational field and let $b_\mu = eA_\mu$, where A_μ is an electromagnetic potential whose gauge holds fixed. Then the equation (2.197) is the well–known equation of motion of a relativistic massive charge in the presence of these fields.

Let us now perform the quantization of RM, following the standard geometrical quantization of the cotangent bundle (see [Bla83, Sni80, Woo92]). As the canonical symplectic form Ω (2.183) on T^*Q is exact, the prequantum bundle is defined as a trivial complex line bundle C over T^*Q. Note that this bundle need no metaplectic correction since T^*X is with canonical coordinates for the symplectic form Ω. Thus, C is called the *quantum bundle*. Let its trivialization

$$C \cong T^*Q \times \mathbb{C} \tag{2.198}$$

hold fixed, and let (t, q^i, p_α, c), with $c \in \mathbb{C}$, be the associated bundle coordinates. Then one can treat sections of C (2.198) as smooth complex functions on T^*Q. Note that another trivialization of C leads to an equivalent quantization of T^*Q.

Recall that the Kostant–Souriau prequantization formula associates to each smooth real function $f \in C^\infty(T^*Q)$ on T^*Q the first–order differential operator

$$\widehat{f} = -i\nabla_{\vartheta_f} + f \tag{2.199}$$

on sections of C, where $\vartheta_f = \partial^\alpha f \partial_\alpha - \partial_\alpha f \partial^\alpha$ is the Hamiltonian vector–field of f and ∇ is the covariant differential with respect to a suitable $U(1)$-principal connection A on C. This connection preserves the *Hermitian metric* $g(c, c') = c\overline{c}'$ on C, and its curvature form obeys the prequantization condition $R = i\Omega$. For the sake of simplicity, let us assume that Q and, consequently, T^*Q is simply–connected. Then the connection A up to gauge transformations is

$$A = dp_\alpha \otimes \partial^\alpha + dq^\alpha \otimes (\partial_\alpha + icp_\alpha \partial_c), \tag{2.200}$$

and the prequantization operators (2.199) read

$$\widehat{f} = -i\vartheta_f + (f - p_\alpha \partial^\alpha f). \tag{2.201}$$

Let us choose the vertical polarization on T^*Q. It is the vertical tangent bundle VT^*Q of the fibration $\pi : T^*Q \to Q$. As was mentioned above, the corresponding quantum algebra $\mathcal{A} \subset C^\infty(T^*Q)$ consists of affine functions f (2.186) of momenta p_α. Its representation by operators (2.201) is defined in the space E of sections ρ of the quantum bundle C of compact support which obey the condition $\nabla_\vartheta \rho = 0$ for any vertical Hamiltonian vector–field ϑ on T^*Q. This condition takes the form

$$\partial_\alpha f \partial^\alpha \rho = 0, \qquad (f \in C^\infty(Q)).$$

It follows that elements of E are independent of momenta and, consequently, fail to be compactly supported, unless $\rho = 0$. This is the well–known problem of Schrödinger quantization which is solved as follows [Bla83, GMS02b].

Let $i_Q : Q \longrightarrow T^*Q$ be the canonical zero section of the cotangent bundle T^*Q. Let $C_Q = i_Q^* C$ be the pull–back of the bundle C (2.198) over Q. It is a trivial complex line bundle $C_Q = Q \times \mathbb{C}$ provided with the pull–back Hermitian metric $g(c, c') = c\overline{c}'$ and the pull–back

$$A_Q = i_Q^* A = dq^\alpha \otimes (\partial_\alpha + icp_\alpha \partial_c)$$

of the connection A (2.200) on C. Sections of C_Q are smooth complex functions on Q, but this bundle need metaplectic correction.

Let the cohomology group $H^2(Q; \mathbb{Z}_2)$ of Q be trivial. Then a metalinear bundle $\mathcal{D}$ of complex half-forms on Q is defined. It admits the canonical lift of any vector–field τ on Q such that the corresponding Lie derivative of its sections reads

$$\mathfrak{L}_\tau = \tau^\alpha \partial_\alpha + \frac{1}{2} \partial_\alpha \tau^\alpha.$$

Let us consider the tensor product $Y = C_Q \otimes \mathcal{D}$ over Q. Since the Hamiltonian vector–fields

$$\vartheta_f = a^\alpha \partial_\alpha - (p_\mu \partial_\alpha a^\mu + \partial_\alpha b)\partial^\alpha$$

of functions f (2.186) are projected onto Q, one can assign to each element f of the quantum algebra $\mathcal{A}$ the first–order differential operator

$$\widehat{f} = (-i\overline{\nabla}_{\pi\vartheta_f} + f) \otimes \mathrm{Id} + \mathrm{Id} \otimes \mathfrak{L}_{\pi\vartheta_f} = -ia^\alpha \partial_\alpha - \frac{\mathrm{i}}{2} \partial_\alpha a^\alpha + b$$

on sections ρ_Q of Y. For the sake of simplicity, let us choose a trivial metalinear bundle $\mathcal{D} \to Q$ associated to the orientation of Q. Its sections can be written in the form $\rho_Q = (-g)^{1/4}\psi$, where ψ are smooth complex functions on Q. Then the quantum algebra $\mathcal{A}$ can be represented by the operators $\widehat{f}$ (2.187)

in the space $\mathbb{C}^\infty(Q)$ of these functions. It can be justified that these operators obey the *Dirac condition*

$$[\widehat{f},\widehat{f}'] = -i\widehat{\{f,f'\}}.$$

One usually considers the subspace $E_Q \subset \mathbb{C}^\infty(Q)$ of functions of compact support. It is a pre–Hilbert space with respect to the non–degenerate Hermitian form

$$\langle\psi|\psi'\rangle = \int_Q \psi\overline{\psi}'(-g)^{1/2}d^{m+1}q$$

Note that $\widehat{f}$ (2.187) are symmetric operators $\widehat{f} = \widehat{f}^*$ in E_Q, i.e., $\langle\widehat{f}\psi|\psi'\rangle = \langle\psi|\widehat{f}\psi'\rangle$. However, the space E_Q gets no physical meaning in RM.

As was mentioned above, the function H_T (2.184) need not belong to the quantum algebra $\mathcal{A}$, but a polynomial function H_T can be quantized as an element of the enveloping algebra $\overline{\mathcal{A}}$ by operators $\widehat{H}_T$ (2.189). Then the quantum constraint (2.185) serves as a relativistic quantum equation.

Let us again consider a massive relativistic charge whose relativistic Hamiltonian is

$$H = \frac{1}{2m}g^{\mu\nu}(p_\mu - eA_\mu)(p_\nu - eA_\nu).$$

It defines the constraint

$$H_T = \frac{1}{m^2}g^{\mu\nu}(p_\mu - eA_\mu)(p_\nu - eA_\nu) - 1 = 0. \tag{2.202}$$

Let us represent the function H_T (2.202) as symmetric product of affine functions of momenta,

$$H_T = \frac{(-g)^{-1/4}}{m}(p_\mu - eA_\mu)(-g)^{1/4}g^{\mu\nu}(-g)^{1/4}(p_\nu - eA_\nu)\frac{(-g)^{-1/4}}{m} - 1.$$

It is quantized by the rule (2.189), where

$$(-g)^{1/4}\circ\widehat{\partial}_\alpha\circ(-g)^{-1/4} = -i\partial_\alpha.$$

Then the well–known relativistic quantum equation

$$(-g)^{-1/2}[(\partial_\mu - ieA_\mu)g^{\mu\nu}(-g)^{1/2}(\partial_\nu - ieA_\nu) + m^2]\psi = 0. \tag{2.203}$$

is reproduced up to the factor $(-g)^{-1/2}$.

2.2 Physical Field Systems

2.2.1 $n-$Categorical Framework

Recall that in the 19th Century Maxwell unified Faraday's electric and magnetic fields. Maxwell's theory led to Einstein's special relativity where this

unification becomes a spin–off of the unification of space end time in the form of the *Faraday tensor* [MTW73]

$$F = E \wedge dt + B,$$

where F is electromagnetic $2-$form on space–time, E is electric $1-$form on space, and B is magnetic $2-$form on space. Gauge theory considers F as secondary object to a connection–potential $1-$form A. This makes half of the *Maxwell equations* into tautologies [Bae02], i.e.,

$$F = dA \qquad \Longrightarrow \qquad dF = 0 \qquad \text{the } \textit{Bianchi relation},$$

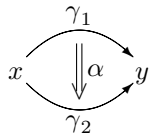
but does not imply the *dual Bianchi relation*, which is a second half of Maxwell's equations,

$$*d * F = J,$$

where $*$ is the dual *Hodge star operator* and J is current $1-$form.

To understand the deeper meaning of the connection–potential $1-$form A, we can integrate it along a path γ in space–time, $x \xrightarrow{\gamma} y$. Classically, the integral $\int_\gamma A$ represents an *action* for a charged point particle to move along the path γ. Quantum–mechanically, $\exp\left(\mathrm{i}\int_\gamma A\right)$ represents a *phase* (within the unitary group $U(1)$) by which the particle's wave–function changes as it moves along the path γ, so A is a $U(1)-$connection.

The only thing that matters here is the *difference α between two paths γ_1* and γ_2 *in the action* $\int_\gamma A$ [Bae02], which is a two–morphism (see Appendix)



To generalize this construction, consider any compact Lie group G. A connection A on a trivial $G-$bundle is a $\gamma-$valued $1-$form. A assigns a *holonomy* $P\exp\left(\mathrm{i}\int_\gamma A\right) \in G$ along any path $x \xrightarrow{\gamma} y$ and has a *curvature* F given by

$$F = dA + A \wedge A.$$

The curvature F implies the *extended Bianchi relation*

$$dF + A \wedge F = 0,$$

but does not imply the dual Bianchi relation, i.e., *Yang–Mills relation*

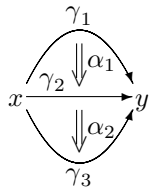
$$*(d * F + A \wedge *F) = J.$$

Further generalization is performed with string theory. Just as point particles naturally interact with a 1−form A, strings naturally interact with a 2−form B, such that [Bae02]

$$\text{action} \quad = \int_{\Sigma} B, \qquad \text{and} \qquad \text{phase} \quad = \exp\left(\mathrm{i} \int_{\Sigma} B\right).$$

This 2−form connection B has a 3−form curvature $G = dB$, which satisfies Maxwell–like equations, i.e., implies Bianchi–like relation $dG = 0$, but does not imply the dual, *current relation* $*d*G = J$, with the current 2−form J.

In this way, the *higher Yang–Mills theory* assigns holonomies to paths and also to paths of paths, so that we have a 3−morphism



allowing us to ask not merely whether holonomies along paths are equal, but whether and how they are isomorphic.

This generalization actually proposes *categorification* of the basic geometrical concepts of manifold, group, Lie group and Lie algebra [Bae02]. Replacing the words *set* and *function* by *smooth manifold* and *smooth map* we get the concept of *smooth category*. Replacing them by *group* and *homomorphism* we get the concept of 2−*group*. Replacing them by *Lie group* and *smooth homomorphism* we get the concept of *Lie* 2−*group*. Replacing them by *Lie algebra* and *Lie algebra homomorphism* we get the concept of *Lie* 2−*algebra*. Examples of the smooth categories are the following:

1. A smooth category with only identity morphisms is a smooth manifold.
2. A smooth category with one object and all morphisms invertible is a Lie group.
3. Any Lie groupoid gives a smooth category with all morphisms invertible.
4. A generalization of a vector bundle (E, M, π), where E and M are smooth manifolds and projection $\pi : E \to M$ is a smooth map, gives a vector 2−bundle (E, M, π) where E and M are smooth categories and projection $\pi : E \to M$ is a smooth functor.

2.2.2 Lagrangian Field Theory on Fibre Bundles

In this subsection we will apply the jet formalism defined in subsection 1.4 above, and already applied for development of the time–dependent mechanics in subsection 2.1.2 above, to formulate the *first–order Lagrangian field theory* on fibre bundles (see [Sar93, Sar95, GMS97, MS00a, Sar02a] for details).

Recall that the configuration space of the first–order Lagrangian field theory on a fibre bundle $Y \to X$, coordinated by $(x^\alpha, y^i, y^i_\alpha)$, is the 1–jet space $J^1(X,Y)$ of the bundle $Y \to X$, coordinated by $(x^\alpha, y^i, y^i_\alpha)$. Therefore, a first–order Lagrangian $L : J^1(X,Y) \to \wedge^n T^*X$ is defined as a *horizontal density* on $J^1(X,Y)$,

$$L = \mathcal{L}(x^\alpha, y^i, y^i_\alpha)\omega, \quad \text{with} \quad \omega = dx^1 \wedge ... \wedge dx^n, \qquad (n = \dim X). \quad (2.204)$$

Let us follow the standard formulation of the variational problem on fibre bundles where deformations of sections of a fibre bundle $Y \to X$ are induced by local 1–parameter groups of automorphisms of $Y \to X$ over X (the so-called vertical gauge transformations). Here, we will not study the calculus of variations in depth, but apply in a straightforward manner the *first variational formula* (2.70) (for technical details, see [Sar93, Sar95, GMS97, MS00a, Sar02a]).

Recall that a projectable vector–field u on a fibre bundle $Y \to X$ is an infinitesimal generator of a local 1–parameter group of gauge transformations of $Y \to X$. Therefore, one can think of its jet prolongation $j^1 u$ (1.165) as being the infinitesimal generator of gauge transformations of the configuration space $J^1(X,Y)$. Let the Lie derivative of a Lagrangian L along $j^1 u$ be given by

$$\mathfrak{L}_{j^1 u} L = [\partial_\alpha u^\alpha \mathcal{L} + (u^\alpha \partial_\alpha + u^i \partial_i + (d_\alpha u^i - y^i_\mu \partial_\alpha u^\mu)\partial^\alpha_i)\mathcal{L}]\,\omega. \quad (2.205)$$

The *first variational formula* (2.70) gives its *canonical decomposition* (in accordance with the general variational problem), which reads

$$\begin{aligned} \mathfrak{L}_{j^1 u} L &= u_V \rfloor \mathcal{E}_L + d_H h_0(u \rfloor H_L) \qquad (2.206)\\ &= (u^i - y^i_\mu u^\mu)(\partial_i - d_\alpha \partial^\alpha_i)\mathcal{L}\omega - d_\alpha[\pi^\alpha_i (u^\mu y^i_\mu - u^i) - u^\alpha \mathcal{L}]\omega. \end{aligned}$$

In the canonical decomposition (2.206), $u_V = (u \rfloor \theta^i)\partial_i$; the map

$$\mathcal{E}_L : J^2(X,Y) \to T^*Y \wedge (\wedge^n T^*X), \quad \text{given by} \quad \mathcal{E}_L = (\partial_i \mathcal{L} - d_\alpha \pi^\alpha_i)\theta^i \wedge \omega, \quad (2.207)$$

(with $\pi^\alpha_i = \partial^\alpha_i \mathcal{L}$) is the *Euler–Lagrangian operator* associated to the Lagrangian L; and the map

$$H_L : J^1(X,Y) \to M_Y = T^*Y \wedge (\wedge^{n-1} T^*X), \qquad \text{given by} \quad (2.208)$$

$$H_L = L + \pi^\alpha_i \theta^i \wedge \omega_\alpha = \pi^\alpha_i dy^i \wedge \omega_\alpha + (\mathcal{L} - \pi^\alpha_i y^i_\alpha)\omega, \quad (2.209)$$

is called the *Poincaré–Cartan form*.

The *kernel* of the Euler–Lagrangian operator $\mathcal{E}_L$ (2.207) defines the system of *second–order Euler–Lagrangian equations*, in local coordinates given by

$$(\partial_i - d_\alpha \partial^\alpha_i)\mathcal{L} = 0, \quad (2.210)$$

A solution of these equations is a section $s : X \longrightarrow Y$ of the fibre bundle $Y \longrightarrow X$, whose second–order jet prolongation $j^2 s$ lives in (2.210), i.e.,

$$\partial_i \mathcal{L} \circ s - (\partial_\alpha + \partial_\alpha s^j \partial_j + \partial_\alpha \partial_\mu s^j \partial^\mu_j)\partial^\alpha_i \mathcal{L} \circ s = 0. \tag{2.211}$$

Different Lagrangians L and L' can lead to the same Euler–Lagrangian operator $\mathcal{E}_L$ if their difference $L_0 = L - L'$ is a variationally trivial Lagrangian, whose Euler–Lagrangian operator vanishes identically. A Lagrangian L_0 is called *variationally trivial* iff

$$L_0 = h_0(\varphi), \tag{2.212}$$

where φ is a closed n–form on Y. We have at least locally $\varphi = d\xi$, and then

$$L_0 = h_0(d\xi) = d_H(h_0(\xi) = d_\alpha h_0(\xi)^\alpha \omega, \qquad h_0(\xi) = h_0(\xi)^\alpha \omega_\alpha.$$

The Poincaré–Cartan form H_L (2.208) is called a *Lepagean equivalent* of a Lagrangian L if $h_0(H_L) = L$. In contrast with other Lepagean forms (see [GMS97, MS00a]), H_L is a horizontal form on the affine jet bundle $J^1(X, Y) \to Y$.

The fibre bundle $M_Y = T^*Y \wedge (\wedge^{n-1} T^*X)$, figuring in the Poincaré–Cartan form (2.208) is called the *homogeneous Legendre bundle*. It has holonomic local coordinates $(x^\alpha, y^i, p^\alpha_i, p)$ with transition functions

$$p'^\alpha_i = \det(\frac{\partial x^\varepsilon}{\partial x'^\nu})\frac{\partial y^j}{\partial y'^i}\frac{\partial x'^\alpha}{\partial x^\mu} p^\mu_j, \qquad p' = \det(\frac{\partial x^\varepsilon}{\partial x'^\nu})(p - \frac{\partial y^j}{\partial y'^i}\frac{\partial y'^i}{\partial x^\mu} p^\mu_j). \tag{2.213}$$

Relative to these coordinates, the map (2.208) reads

$$(p^\mu_i, p) \circ H_L = (\pi^\mu_i, \mathcal{L} - \pi^\mu_i y^i_\mu).$$

The transition functions (2.213) shows that M_Y is a 1D affine bundle

$$\pi_{M\Pi} : M_Y \to \Pi \tag{2.214}$$

over the *Legendre bundle*

$$\Pi = \wedge^n T^*X \otimes V^*Y \otimes TX = V^*Y \wedge (\wedge^{n-1} T^*X), \tag{2.215}$$

with holonomic coordinates $(x^\alpha, y^i, p^\alpha_i)$. Then the composition

$$\widehat{L} = \pi_{M\Pi} \circ H_L : J^1(X, Y) \to \Pi, \qquad (x^\alpha, y^i, p^\alpha_i) \circ \widehat{L} = (x^\alpha, y^i, \pi^\alpha_i), \tag{2.216}$$

is the well–known *Legendre map*. One can think of p^α_i as being the covariant momenta of field functions, and the Legendre bundle Π (2.215) plays the role of a finite–dimensional *momentum phase–space of fields* in the covariant Hamiltonian field theory (see subsection 2.2.4 below).

The first variational formula (2.206) gives the standard procedure for the study of differential conservation laws in Lagrangian field theory as follows.

Let u be a projectable vector–field on a fibre bundle $Y \to X$ treated as the infinitesimal generator of a local 1–parameter group G_u of gauge transformations. On–shell, i.e., on the kernel (2.210) of the Euler–Lagrangian operator $\mathcal{E}_L$, the first variational formula (2.206) leads to the *weak identity*

$$\mathfrak{L}_{j^1 u} L \approx -d_\alpha \mathfrak{J}^\alpha \omega, \qquad \text{where} \tag{2.217}$$

$$\mathfrak{J} = \mathfrak{J}^\alpha \omega_\alpha, \qquad \mathfrak{J}^\alpha = \pi_i^\alpha (u^\mu y_\mu^i - u^i) - u^\alpha \mathcal{L}, \tag{2.218}$$

is the *symmetry current* along the vector–field u. Let a Lagrangian L be invariant under the gauge group G_u. This implies that the Lie derivative $\mathfrak{L}_{j^1 u} L$ (2.205) vanishes. Then we get the *weak conservation law*

$$d_\alpha \mathfrak{J}^\alpha \approx 0 \tag{2.219}$$

of the symmetry current $\mathfrak{J}$ (2.218).[2]

The weak conservation law (2.219) leads to the differential conservation law

$$\partial_\alpha (\mathfrak{J}^\alpha \circ s) = 0 \tag{2.220}$$

on solutions $s : X \longrightarrow Y$ (2.211) of the Euler–Lagrangian equations (2.210). It implies the *integral conservation law*

$$\int_{\partial N} s^* \mathfrak{J} = 0, \tag{2.221}$$

where N is a compact nD submanifold of X with the boundary ∂N.

In gauge theory, the symmetry current $\mathfrak{J}$ (2.218) takes the form

$$\mathfrak{J} = W + d_H U = (W^\alpha + d_\mu U^{\mu\alpha}) \omega_\alpha, \tag{2.222}$$

where the term W depends only on the *variational derivatives*

$$\delta_i \mathcal{L} = (\partial_i - d_\alpha \partial_i^\alpha) \mathcal{L}, \tag{2.223}$$

i.e., $W \approx 0$. The tensor–field $U = U^{\mu\alpha} \omega_{\mu\alpha} : J^1(X,Y) \to \wedge^{n-2} T^* X$ is a horizontal $(n-2)$–form on $J^1(X,Y) \to X$. Then one says that $\mathfrak{J}$ *reduces to the superpotential* U (see [FFF94, GMS97, Sar97]). On–shell, such a symmetry current reduces to a d_H–exact form (2.222). In this way, the differential conservation law (2.220) and the integral conservation law (2.221) become tautological. At the same time, the superpotential form (2.222) of $\mathfrak{J}$ implies the following integral relation

$$\int_{N^{n-1}} s^* \mathfrak{J} = \int_{\partial N^{n-1}} s^* U, \tag{2.224}$$

where N^{n-1} is a compact oriented $(n-1)$D submanifold of X with the boundary ∂N^{n-1}. One can think of this relation as being a part of the Euler–Lagrangian equations written in an integral form.

[2] The first variational formula defines the symmetry current (2.218) modulo the terms $d_\mu(c_i^{\mu\alpha}(y_\nu^i u^\nu - u^i))$, where $c_i^{\mu\alpha}$ are arbitrary skew–symmetric functions on Y [GMS97]. Here, we set aside these boundary terms which are independent of a Lagrangian L.

Let us consider conservation laws in the case of gauge transformations which preserve the Euler–Lagrangian operator $\mathcal{E}_L$, but not necessarily a Lagrangian L. Let u be a projectable vector–field on $Y \to X$, which is the infinitesimal generator of a local 1–parameter group of such transformations, i.e.,

$$\mathfrak{L}_{j^2u}\mathcal{E}_L = 0,$$

where j^2u is the second–order jet prolongation of the vector–field u. There is the useful relation [GMS97]

$$\mathfrak{L}_{j^2u}\mathcal{E}_L = \mathcal{E}_{\mathfrak{L}_{j^1u}L}. \tag{2.225}$$

Then, in accordance with (2.212), we have locally

$$\mathfrak{L}_{j^1u}L = d_\alpha h_0(\xi)^\alpha \omega. \tag{2.226}$$

In this case, the weak identity (2.217) reads

$$d_\alpha(h_0(\xi)^\alpha - \mathfrak{J}^\alpha) \approx 0, \tag{2.227}$$

where $\mathfrak{J}$ is the symmetry current (2.218).

Background fields, which do not live in the dynamical shell (2.210), violate conservation laws as follows. Let us consider the product

$$Y_{tot} = Y \times Y' \tag{2.228}$$

of a fibre bundle $Y \longrightarrow X$, coordinated by (x^α, y^i), whose sections are dynamical fields and of a fibre bundle $Y' \longrightarrow X$, coordinated by (x^α, y^A), whose sections are *background fields* that take the background values

$$y^B = \phi^B(x), \qquad \text{and} \qquad y^B_\alpha = \partial_\alpha \phi^B(x).$$

A Lagrangian L of dynamical and background fields is defined on the total configuration space $J^1(X,Y)_{tot}$. Let u be a projectable vector–field on Y_{tot} which also projects onto Y' because gauge transformations of background fields do not depend on dynamical fields. This vector–field takes the coordinate form

$$u = u^\alpha(x^\mu)\partial_\alpha + u^A(x^\mu, y^B)\partial_A + u^i(x^\mu, y^B, y^j)\partial_i. \tag{2.229}$$

Substitution of u (2.229) in the formula (2.206) leads to the first variational formula in the presence of background fields:

$$\partial_\alpha u^\alpha \mathcal{L} + [u^\alpha \partial_\alpha + u^A \partial_A + u^i \partial_i + (d_\alpha u^A - y^A_\mu \partial_\alpha u^\mu)\partial^\alpha_A \tag{2.230}$$

$$+(d_\alpha u^i - y^i_\mu \partial_\alpha u^\mu)\partial^\alpha_i]\mathcal{L} = (u^A - y^A_\alpha u^\alpha)\partial_A \mathcal{L} + \pi^\alpha_A d_\alpha(u^A - y^A_\mu u^\mu)$$

$$+(u^i - y^i_\alpha u^\alpha)\delta_i \mathcal{L} - d_\alpha[\pi^\alpha_i(u^\mu y^i_\mu - u^i) - u^\alpha \mathcal{L}]. \tag{2.231}$$

Then we have on the shell (2.210) the weak identity

$$\partial_\alpha u^\alpha \mathcal{L} + [u^\alpha \partial_\alpha + u^A \partial_A + u^i \partial_i + (d_\alpha u^A - y^A_\mu \partial_\alpha u^\mu)\partial^\alpha_A + (d_\alpha u^i - y^i_\mu \partial_\alpha u^\mu)\partial^\alpha_i]\mathcal{L}$$
$$\approx (u^A - y^A_\alpha u^\alpha)\partial_A \mathcal{L} + \pi^\alpha_A d_\alpha (u^A - y^A_\mu u^\mu) - d_\alpha[\pi^\alpha_i (u^\mu y^i_\mu - u^i) - u^\alpha \mathcal{L}].$$

If a total Lagrangian L is invariant under gauge transformations of Y_{tot}, we get the weak identity

$$(u^A - y^A_\mu u^\mu)\partial_A \mathcal{L} + \pi^\alpha_A d_\alpha (u^A - y^A_\mu u^\mu) \approx d_\alpha \mathfrak{J}^\alpha, \tag{2.232}$$

which is the transformation law of the symmetry current $\mathfrak{J}$ in the presence of background fields.

Conservation Laws

In the first–order Lagrangian field theory, we have the following differential transformation and conservation laws on solutions $s : X \longrightarrow Y$ (2.211) of the Euler–Lagrangian equations (2.210).

Recall that given fibre coordinates (x^α, y^i) of Y, the jet space $J^1(X,Y)$ is equipped with the adapted coordinates $(x^\alpha, y^i, y^i_\alpha)$, while the first–order *Lagrangian density* on $J^1(X,Y)$ is defined as the map

$$L : J^1(X,Y) \to \wedge^n T^*X, \qquad (n = \dim X),$$
$$L = \mathcal{L}(x^\alpha, y^i, y^i_\alpha)\omega, \qquad \text{with} \qquad \omega = dx^1 \wedge ... \wedge dx^n.$$

The corresponding *first–order Euler–Lagrangian equations* for sections $\overline{s} : X \longrightarrow J^1(X,Y)$ of the jet bundle $J^1(X,Y) \to X$ read

$$\partial_\alpha \overline{s}^i = \overline{s}^i_\alpha, \qquad \partial_i \mathcal{L} - (\partial_\alpha + \overline{s}^j_\alpha \partial_j + \partial_\alpha \overline{s}^j_\alpha \partial^\alpha_j)\partial^\alpha_i \mathcal{L} = 0. \tag{2.233}$$

We consider the Lie derivatives of Lagrangian densities in order to get differential conservation laws. Let

$$u = u^\alpha(x)\partial_\alpha + u^i(y)\partial_i$$

be a projectable vector–field on $Y \to X$ and $\overline{u}$ its jet lift (1.165) onto $J^1(X,Y) \to X$. Given L, let us computer the Lie derivative $\mathfrak{L}_{\overline{u}} L$. We get the identity

$$\overline{s}^* \mathfrak{L}_{\overline{u}} L \approx -\frac{d}{dx^\alpha}[\pi^\alpha_i (u^\alpha \overline{s}^i_\alpha - u^i) - u^\alpha \mathcal{L}]\omega, \qquad \pi^\alpha_i = \partial^\alpha_i \mathcal{L}, \tag{2.234}$$

modulo the Euler–Lagrangian equations (2.233).

Let L be a Lagrangian density on the jet space $J^1(X,Y)$. For the sake of simplicity, we shall denote the pull–back $\pi^{1*}_0 L$ of L onto $J^2(X,Y)$ by the same symbol L.

Let u be a projectable vector–field on $Y \longrightarrow X$ and $\overline{u}$ its jet lift (1.165) onto the configuration bundle $J^1(X,Y) \longrightarrow X$. Recall that the vector–field u is associated with some 1–parameter group of transformations of Y.

Let us calculate the Lie derivative $\mathfrak{L}_{\overline{u}}L$ of the horizontal density L when its Lepagian equivalent is chosen to be the Poincaré–Cartan form Ξ_L, given by the coordinate expression

$$\Xi_L = L\omega + \pi_i^\alpha(dy^i - y_\alpha^i dx^\alpha) \wedge \omega_\alpha. \tag{2.235}$$

In this case we recover the first variational formula (2.206) for projectable vector–fields on Y as (see [GM90, Sar97])

$$\mathfrak{L}_{\overline{u}}L = u_V \rfloor \mathcal{E}_L + h_0(d\overline{u}\rfloor \Xi_L). \tag{2.236}$$

Since the Poincaré–Cartan form Ξ_L is a horizontal form on the jet bundle $J^1(X,Y) \longrightarrow Y$, the formula (2.236) takes the form

$$\mathfrak{L}_{\overline{u}}L = u_V \rfloor \mathcal{E}_L + d_H h_0(u\rfloor \Xi_L). \tag{2.237}$$

Being restricted to the kernel

$$[\partial_i - (\partial_\alpha + y_\alpha^j\partial_j + y_{\alpha\lambda}^j\partial_j^\alpha)\partial_i^\alpha]\mathcal{L} = 0$$

of the Euler–Lagrangian operator $\mathcal{E}_L$ (2.207), the equality (2.237) reduces to the weak identity

$$\mathfrak{L}_{\overline{u}}L \approx d_H h_0(u\rfloor \Xi_L), \tag{2.238}$$

$$\partial_\alpha u^\alpha \mathcal{L} + [u^\alpha\partial_\alpha + u^i\partial_i + (\partial_\alpha u^i + y_\alpha^j\partial_j u^i - y_\alpha^i\partial_\alpha u^\alpha)\partial_i^\alpha]\mathcal{L} \approx \widehat{\partial}_\alpha[\pi_i^\alpha(u^i - u^\alpha y_\alpha^i) + u^\alpha\mathcal{L}],$$

$$\widehat{\partial}_\alpha = \partial_\alpha + y_\alpha^i\partial_i + y_{\alpha\lambda}^i\partial_i^\alpha.$$

On solutions s of the Euler–Lagrangian equations, the weak identity (2.238) becomes the weak differential transformation law

$$s^*\mathfrak{L}_{\overline{u}}L \approx d(s^*u\rfloor \Xi_L) \tag{2.239}$$

which takes the coordinate form (2.234).

Note that, in order to get the differential transformation laws on solutions s of a given system of Euler–Lagrangian equations, one can examine other Lepagian equivalents ρ_L of the Lagrangian density L, besides the Poincaré–Cartan form Ξ_L. In this case, the first variational formula (2.236) and the corresponding weak identity

$$\mathfrak{L}_{\overline{u}}L \approx h_0(d\overline{u}\rfloor \rho_L)$$

differ from relations (2.236) and (2.238) respectively in the strong identity

$$0 = h_0(d\overline{u}\rfloor\varepsilon) = d_H h_0(\overline{u}\rfloor\varepsilon), \tag{2.240}$$

where $\rho_L = \Xi_L + \varepsilon$. From the physical point of view, it means that different Lepagian equivalents result in different superpotentials $h_0(\overline{u}\rfloor\varepsilon)$.

The form ε in the identity (2.240) has the coordinate expression

$$\varepsilon = -(\widehat{\partial}_\nu c_i^{\alpha\nu}\widehat{dy}^i + c_i^{\alpha\nu}\widehat{dy}^i_\nu)\wedge\omega_\alpha + \chi.$$

It is the general local expression for Lepagian equivalents of the zero Lagrangian density. We have

$$h_0(\overline{u}\rfloor\varepsilon) = \widehat{\partial}_\nu[(u^i - y^i_\alpha u^\alpha)c_i^{\alpha\nu}]\omega_\alpha.$$

One can consider also other Lagrangian densities L' which possess the same Euler–Lagrangian operator $\mathcal{E}_L$. Then the first variational formula and the corresponding weak identity differ from relations (2.236) and (2.238) respectively in the strong identity

$$\mathfrak{L}_{\overline{u}}h_0(\epsilon) = h_0(d\overline{u}\rfloor\epsilon) \tag{2.241}$$

where ϵ is some closed exterior form on Y. However, if the form $h_0(\epsilon)$ possesses the same symmetries as the Lagrangian density L only, the contribution of the strong identity (2.241) into the weak identity (2.238) is not tautological.

Note that the weak identity (2.238) is linear in the vector–field u, and we can consider superposition of different weak identities (2.238) corresponding to different vector–fields u. For example, if u and u' are projectable vector–fields on the bundle $Y\longrightarrow X$ which are projected onto the same vector–field on the base X, their difference $u-u'$ is a vertical vector–field on $Y\longrightarrow X$. Therefore, the difference of the weak identity (2.238) with respect the vector–fields u and u' results in the weak identity (2.238) with respect to the vertical vector–field $u-u'$.

Now let us consider the case when a Lagrangian density L depends on background fields. We define such a Lagrangian density as the pull–back of the Lagrangian density L_{tot} on the total configuration space by some fixed sections $\phi(x)$ describing background fields.

Let us again consider the product (2.228), namely $Y_{tot}=Y\times Y'$, of the bundle Y whose sections are *dynamical fields* and the bundle Y' whose sections ϕ play the role of *background fields*. Let the bundles Y and Y' be coordinated by (x^α,y^i) and (x^α,y^A) respectively. The Lagrangian density L_{tot} is defined on the total configuration space $J^1(X,Y)_{tot}$.

Let u be a projectable vector–field on Y_{tot} which is also projectable with respect to projection $Y\times Y'\to Y'$. It has the coordinate form

$$u = u^\alpha(x)\partial_\alpha + u^A(x^\alpha,y^B)\partial_A + u^i(x^\alpha,y^B,y^j)\partial_i,$$

showing that transformations of background fields are independent on dynamical fields.

Calculating the Lie derivative of the Lagrangian density L_{tot} by this vector–field, we get the equality

$$\begin{aligned}&\partial_\alpha u^\alpha\mathcal{L}_{tot} + [u^\alpha\partial_\alpha + u^A\partial_A + u^i\partial_i + (\partial_\alpha u^A + y^B_\alpha\partial_B u^A - y^A_\alpha\partial_\alpha u^\alpha)\partial^\alpha_A\\ &+(\partial_\alpha u^i + y^B_\alpha\partial_B u^i + y^j_\alpha\partial_j u^i - y^i_\alpha\partial_\alpha u^\alpha)\partial^\alpha_i]\mathcal{L}_{tot} = \widehat{\partial}_\alpha[\pi^\alpha_i(u^i - u^\alpha y^i_\alpha) + u^\alpha\mathcal{L}_{tot}]\\ &+(u^i - y^i_\alpha u^\alpha)(\partial_i - \widehat{\partial}_\alpha\partial^\alpha_i)\mathcal{L}_{tot} + (u^A - y^A_\alpha u^\alpha)\partial_A\mathcal{L}_{tot} + \pi^\alpha_A\widehat{\partial}_\alpha(u^A - y^A_\alpha u^\alpha),\end{aligned}$$

which can be rewritten as

$$\partial_\alpha u^\alpha \mathcal{L}_{tot} + [u^\alpha(\partial_\alpha + y^B_\alpha \partial_B + y^B_{\alpha\lambda}\partial^\alpha_B)\mathcal{L} + u^i\partial_i + (\widehat{\partial}_\alpha u^i - y^i_\alpha \partial_\alpha u^\alpha)\partial^\alpha_i]\mathcal{L}_{tot}$$
$$= \widehat{\partial}_\alpha[\pi^\alpha_i(u^i - u^\alpha y^i_\alpha) + u^\alpha \mathcal{L}_{tot}] + (u^i - y^i_\alpha u^\alpha)(\partial_i - \widehat{\partial}_\alpha \partial^\alpha_i)\mathcal{L}_{tot}.$$

The pull–back of this equality to the bundle $Y \to X$ by sections $\phi^A(x)$ of the bundle Y' which describe the background fields results in the familiar expression (2.237) and the familiar weak identity (2.238) for the Lagrangian density $L = \phi^* L_{tot}$. Now the partial derivative ∂_α can be written as

$$\partial_\alpha = \widetilde{\partial}_\alpha + \partial_\alpha \phi^B \partial_B + \partial_\alpha \partial_\alpha \phi^B \partial^\alpha_B \,,$$

where $\widetilde{\partial}_\alpha$ denote the partial derivatives with respect to the coordinates x^α on which the Lagrangian density L_{tot} depends explicitly.

Note that Lagrangian densities of field models almost never depend explicitly on the world coordinates x^α. At the same time, almost all field models describe fields in the presence of a background world metric g on the base manifold X, except topological field theories whose classical Lagrangian densities are independent on g [BBR91] and the gravitation theory where a world metric g is a dynamical field.

By a world metric on X is denoted a nondegenerate fibre metric $g^{\alpha\nu}$ in cotangent and tangent bundles of X. In this case, the partial derivative $\partial_\alpha \mathcal{L}$ in the weak identity (2.238) contains the term $\frac{\partial \mathcal{L}}{\partial g^{\alpha\nu}}\partial_\alpha g^{\alpha\nu}$, so that the metric *stress–energy–momentum tensor* of fields (SEM–tensor, for short, see subsection 2.2.6 below)

$$t_{\alpha\nu}\sqrt{\mid g \mid} = 2\frac{\partial \mathcal{L}}{\partial g^{\alpha\nu}}, \qquad \mid g \mid = \mid \det(g_{\alpha\nu}) \mid .$$

is called into play.

The weak identity (2.238) and the weak transformation law (2.239) are basic for our analysis of differential transformation and conservation laws in field theory.

In particular, one says that an isomorphism Φ of the fibre bundle $Y \to X$ is an invariant transformation if its jet prolongation $j^1\Phi$ preserves the Lagrangian density L, i.e.,

$$j^{1*}\Phi L = L.$$

Let u be a projectable vector–field on $Y \to X$. The corresponding local 1–parameter groups of isomorphisms of Y are invariant transformations iff the *strong equality*: $\mathfrak{L}_{\overline{u}} L = 0$ holds. In this case, we have the corresponding weak conservation law

$$d(s^* u \rfloor \Xi_L) \approx 0. \tag{2.242}$$

An isomorphism Φ of the bundle $Y \to X$ is called the generalized invariant transformation if it preserves the Euler–Lagrangian operator $\mathcal{E}_L$. Let u be a projectable vector–field on $Y \to X$. The corresponding local isomorphisms of

Y are generalized invariant transformations iff $\pounds_{\overline{u}} L = h_0(\epsilon)$, where ϵ is a closed n–form on the bundle $Y \longrightarrow X$. In this case, the weak transformation law (2.239) reads

$$s^*\epsilon \approx d(s^*u \rfloor \Xi_L)$$

for every critical section s of $Y \longrightarrow X$. In particular, if $\epsilon = d\varepsilon$ is an exact form, we get the *weak conservation law*

$$d(s^*(u \rfloor \Xi_L - \varepsilon)) \approx 0.$$

In particular, gauge transformations in gauge theory on a 3D base X are the invariant transformations if L is the *Yang–Mills Lagrangian density* and they are the generalized invariant transformations if L is the *Chern–Simons Lagrangian density*.

General Covariance Condition

Now we consider the class of bundles $T \longrightarrow X$ which admit the canonical lift of vector–fields τ on X. They are called the bundles of geometrical objects. In fact, such canonical lift is the particular case of the horizontal lift of a field τ with respect to the suitable connection on the bundle $T \longrightarrow X$ [GMS05].

Let $\tau = \tau^\alpha \partial_\alpha$ be a vector–field on the manifold X. There exists the canonical lift

$$\widetilde{\tau} = T\tau = \tau^\alpha \partial_\alpha + \partial_\nu \tau^\alpha \dot{x}^\nu \frac{\partial}{\partial \dot{x}^\alpha} \tag{2.243}$$

of τ onto the tangent bundle TX of X. This lift consists with the horizontal lift of τ by means the symmetric connection K on the tangent bundle which has τ as the integral section or as the geodesic field:

$$\partial_\nu \tau^\alpha + K^\alpha{}_{\alpha\nu} \tau^\alpha = 0.$$

Generalizing the canonical lift (2.243), one can construct the canonical lifts of a vector–field τ on X onto the following bundles over X. For the sake of simplicity, we denote all these lifts by the same symbol $\widetilde{\tau}$. We have:

- the canonical lift of τ onto the cotangent bundle T^*X, given by

$$\widetilde{\tau} = \tau^\alpha \partial_\alpha - \partial_\beta \tau^\nu \dot{x}_\nu \frac{\partial}{\partial \dot{x}_\beta};$$

- the canonical lift of τ onto the tensor bundle $T^k_m X = (\otimes^m TX) \otimes (\otimes^k T^*X)$, given by

$$\widetilde{\tau} = \tau^\alpha \partial_\alpha + [\partial_\nu \tau^{\alpha_1} \dot{x}^{\nu\alpha_2\cdots\alpha_m}_{\beta_1\cdots\beta_k} + \ldots - \partial_{\beta_1} \tau^\nu \dot{x}^{\alpha_1\cdots\alpha_m}_{\nu\beta_2\cdots\beta_k} - \ldots] \frac{\partial}{\partial \dot{x}^{\alpha_1\cdots\alpha_m}_{\beta_1\cdots\beta_k}};$$

- the canonical lift of τ onto the bundle C of the linear connections on TX, given by

$$\widetilde{\tau} = \tau^\alpha \partial_\alpha + [\partial_\nu \tau^\alpha k^\nu{}_{\beta\alpha} - \partial_\beta \tau^\nu k^\alpha{}_{\nu\alpha} - \partial_\alpha \tau^\nu k^\alpha{}_{\beta\nu} - \partial_{\beta\alpha}\tau^\alpha] \frac{\partial}{\partial k^\alpha{}_{\beta\alpha}}.$$

One can think of the vector–fields $\widetilde{\tau}$ on a bundle of geometrical objects T as being the vector–fields associated with local 1–parameter groups of the holonomic isomorphisms of T induced by diffeomorphisms of its base X. In particular, if $T = TX$ they are the tangent isomorphisms. We call these isomorphisms the *general covariant transformations*.

Let T be the bundle of geometrical objects and L a Lagrangian density on the configuration space $J^1(X,T)$. Given a vector–field τ on the base X and its canonical lift $\widetilde{\tau}$ onto T, one may use the first variational formula (2.237) in order to get the corresponding SEM transformation law. The left side of this formula can be simplified if the Lagrangian density satisfies the general covariance condition.

Note that, if the Lagrangian density L depends on background fields, we should consider the corresponding total bundle (2.228) and the Lagrangian density L_{tot} on the total configuration space $J^1(X,T)_{tot}$. We say that the Lagrangian density L satisfies the general covariance condition if L_{tot} is invariant under 1–parameter groups of general covariant transformations of T_{tot} induced by diffeomorphisms of the base X. It takes place iff, for any vector–field τ on X, the Lagrangian density L_{tot} obeys the equality

$$\mathfrak{L}_{j_0^1\widetilde{\tau}} L_{tot} = 0 \tag{2.244}$$

where $\widetilde{\tau}$ is the canonical lift of τ onto T_{tot} and $j_0^1\widetilde{\tau}$ is the jet lift of $\widetilde{\tau}$ onto $J^1(X,T)_{tot}$.

If the Lagrangian density L does not depend on background fields, the equality (2.244) becomes

$$\mathfrak{L}_{j_0^1\widetilde{\tau}} L = 0. \tag{2.245}$$

Substituting it in the first variational formula (2.237), we get the week conservation law

$$d_H h_0(\widetilde{\tau} \rfloor \Xi_L) \approx 0. \tag{2.246}$$

One can show that the conserved quantity is reduced to a superpotential term.

Here, we verify this fact in case of a tensor bundle $T \longrightarrow X$. Let it be coordinated by (x^α, y^A) where the collective index A is employed. Given a vector–field τ on X, its canonical lift $\widetilde{\tau}$ on T reads

$$\widetilde{\tau} = \tau^\alpha \partial_\alpha + u^{A\beta}_{\alpha} \partial_\beta \tau^\alpha \partial_A.$$

Let a Lagrangian density L on the configuration space $J^1(X,T)$ be invariant under general covarian transformations. Then, it satisfies the equality (2.245) which takes the coordinate form

$$\partial_\alpha(\tau^\alpha\mathcal{L}) + u^{A\beta}_{\ \ \alpha}\partial_\beta\tau^\alpha\partial_A\mathcal{L} + \widehat{\partial}_\alpha(u^{A\beta}_{\ \ \alpha}\partial_\beta\tau^\alpha)\partial^\alpha_A\mathcal{L} - y^A_\alpha\partial_\beta\tau^\alpha\partial^\beta_A\mathcal{L} = 0. \quad (2.247)$$

Due to the arbitrariness of the functions τ^α, the equality (2.247) is equivalent to the system of the equalities

$$\partial_\alpha\mathcal{L} = 0,$$
$$\delta^\beta_\alpha\mathcal{L} + u^{A\beta}_{\ \ \alpha}\partial_A\mathcal{L} + \widehat{\partial}_\alpha(u^{A\beta}_{\ \ \alpha})\partial^\alpha_A\mathcal{L} - y^A_\alpha\partial^\beta_A\mathcal{L} = 0, \quad (2.248)$$
$$u^{A\beta}_{\ \ \alpha}\partial^\alpha_A\mathcal{L} + u^{A\alpha}_{\ \ \alpha}\partial^\beta_A\mathcal{L} = 0. \quad (2.249)$$

Note that the equality (2.248) can be brought into the form

$$\delta^\beta_\alpha\mathcal{L} + u^{A\beta}_{\ \ \alpha}\delta_A\mathcal{L} + \widehat{\partial}_\alpha(u^{A\beta}_{\ \ \alpha}\partial^\alpha_A\mathcal{L}) = y^A_\alpha\partial^\beta_A\mathcal{L}, \quad (2.250)$$

where $\delta_A\mathcal{L}$ are the variational derivatives of the Lagrangian density L. Substituting the relations (2.250) and (2.249) into the weak identity

$$\widehat{\partial}_\alpha[(u^{A\beta}_{\ \ \alpha}\partial_\beta\tau^\alpha - y^A_\alpha\tau^\alpha)\partial^\alpha_A\mathcal{L} + \tau^\alpha\mathcal{L}] \approx 0,$$

we get the conservation law

$$\widehat{\partial}_\alpha[-u^{A\alpha}_{\ \ \alpha}\delta_A\mathcal{L}\tau^\alpha - \widehat{\partial}_\alpha(u^{A\alpha}_{\ \ \alpha}\partial^\alpha_A\mathcal{L}\tau^\alpha)] \approx 0, \quad (2.251)$$

where the conserved current is reduced to the superpotential term

$$Q_{\widetilde{\tau}}{}^\alpha = -u^{A\alpha}_{\ \ \alpha}\delta_A\mathcal{L}\tau^\alpha - \widehat{\partial}_\alpha(u^{A\alpha}_{\ \ \alpha}\partial^\alpha_A\mathcal{L}\tau^\alpha). \quad (2.252)$$

For general field models, we have the product $T\times Y$ of a bundle $T\to X$ of geometrical objects and some other bundle $Y\to X$. The lift of a vector–field τ on the base X onto the corresponding configuration space $J^1(X,T)\times J^1(X,Y)$ reads

$$\overline{\tau} = j^1_0\widetilde{\tau} + \tau^\alpha\Gamma^i_\alpha\partial_i + (\partial_\alpha(\tau^\alpha\Gamma^i_\alpha) + \tau^\alpha y^j_\alpha\partial_j\Gamma^i_\alpha - y^i_\alpha\partial_\alpha\tau^\alpha)\partial^\alpha_i$$

where Γ is a connection on the fibre bundle $Y\to X$.

In this case, we cannot say anything about the general covariance condition independently on the invariance of a Lagrangian density with respect to the internal symmetries.

On the other hand, in *gauge theory* (see subsection 2.2.5 below), several types of gauge transformations are considered. To get the Noether conservation laws, we restrict our consideration to vertical isomorphisms of the principal bundle P. These are the $G-$equivariant isomorphism Φ of P over Id_X, that is,

$$r_g\circ\Phi = \Phi\circ r_g, \qquad (g\in G). \quad (2.253)$$

We call them the gauge isomorphisms. As is well–known, they yield the vertical isomorphisms of the bundle of principal connections C and the $P-$associated bundle E.

For example, let $P \longrightarrow X$ be a principal bundle with the structure Lie group G. Let us consider general gauge isomorphisms Φ of this principal bundle over diffeomorphisms of the base X. They satisfy the relation (2.253). We denote by u_G the projectable vector–fields on P corresponding to local 1–parameter groups of such isomorphisms. There is the 1–1 correspondence between these vector–fields and sections of the bundle $T^G P = TP/G$. They are called the *general principal vector–fields* (see [GM90]). In particular, one can show that, given a vector–field τ on the base X, its horizontal lift onto the principal bundle P by means of a principal connection on P is a general principal vector–field.

General gauge isomorphisms of the principal bundle P, as like as its vertical isomorphisms, yield the corresponding isomorphisms of the associated bundles E and the bundle of principal connections C. We denote by the same symbol u_G the corresponding general principal vector–fields on these bundles.

Consider the product $S = C \times E \times T$, where $T \to X$ is a bundle of geometrical objects. Let a Lagrangian density L on the corresponding configuration space $J^1(T, S)$ be invariant under the isomorphisms of the bundle S which are general gauge isomorphisms of $C \times E$ over diffeomorphisms of the base X and the general covariant transformations of T induced by these diffeomorphisms of X. In particular, vertical isomorphisms of S consist of vertical isomorphisms of $C \times E$ only. It should be emphasized that the general gauge isomorphisms of the bundle $C \times E$ and those of the bundle T taken separately are not the bundle isomorphisms of the product S because they must covering the same diffeomorphisms of the base X of Y. At the same time, one can say that the Lagrangian density L satisfies the general covariance condition in the sense that it is invariant under general isomorphisms of the bundle S [GM90].

This is phrased in terms of the Lie derivatives as follows. Let

$$u_G = \tau^\alpha \partial_\alpha + u^A \partial_A$$

be a general principal vector–field on the product $C \times E$ which is projected onto the vector–field $\tau = \tau^\alpha \partial_\alpha$ on the base X. The corresponding general principal vector–field on the bundle Y reads

$$\widetilde{u}_G = \widetilde{\tau} + u^A \partial_A, \tag{2.254}$$

where $\widetilde{\tau}$ is the canonical lift of τ onto the bundle of geometrical objects T. A Lagrangian density L is invariant under general isomorphisms of the bundle S iff

$$\mathfrak{L}_{j_0^1 \widetilde{u}_G} L = 0, \tag{2.255}$$

where the jet lift $j_0^1 \widetilde{u}_G$ of the vector–field $\widetilde{u}_G$ takes the coordinate form

$$j_0^1 \widetilde{u}_G = j_0^1 \widetilde{\tau} - y^A_\alpha \partial_\alpha u^A \partial^\alpha_A + u^A \partial_A + \widehat{\partial}_\alpha u^A \partial^\alpha_A.$$

There are the topological field theories, besides the gravitation theory, where we can use the condition (2.255) (see subsection 2.2.5 below).

2.2.3 Finsler–Lagrangian Field Theory

In this subsection we present *generalized Finsler–Lagrangian field theory*. The geometrical background of this theory relies on the notion of *generalized Lagrangian space*, $GL^n = (M, g_{ij}(x^k, y^k))$, which is a real nD manifold M with local coordinates $\{x^i\}$, $(i = 1, ..., n)$ and a symmetric *fundamental metric tensor–field* $g_{ij} = g_{ij}(x^k, y^k)$ of rank n and constant signature on T [MKA88, MA94].

From physical point of view, the fundamental metric tensor represents a *unified* gravitational field on TM, which consists of one *external* $(x)-$gravitational field spanned by points $\{x^i\}$, and the one *internal* $(y)-$gravitational field spanned by directions $\{y^i\}$ and equipped with some microscopic character of the space–time structure.

The field theory developed on a generalized Lagrangian space GL^n relies on a fixed *a priori* nonlinear connection $\Gamma = (N^i_j(x, y))$ on the tangent bundle TM. This plays the role of mapping operator of the internal $(y)-$field onto the external $(x)-$field, and prescribes the *interaction* between $(x)-$ and $(y)-$fields. From geometrical point of view, the nonlinear connection allows the construction of the *adapted bases* [MKA88, MA94]

$$\left\{ \frac{\delta}{\delta x^i} = \frac{\partial}{\partial x^i} - N^j_i \frac{\partial}{\partial y^j}, \frac{\partial}{\partial y^i} \right\} \subset \mathcal{X}(TM),$$
$$\{dx^i, \delta y^i = dy^i + N^i_j dx^j\} \subset \mathcal{X}^*(TM).$$

As to the spatial structure, the most important thing is to determine the *Cartan canonical connection* $C\Gamma = (L^i_{jk}, C^i_{jk})$ with respect to g_{ij}, which comes from the metric conditions

$$g_{ij|k} = \frac{\delta g_{ij}}{\delta x^k} - L^m_{ik} g_{mj} - L^m_{jk} g_{mi} = 0, \qquad g_{ij;k} = \frac{\partial g_{ij}}{\partial y^k} - C^m_{ik} g_{mj} - C^m_{jk} g_{mi} = 0,$$

where "$_{|k}$" and "$_{;k}$" are the local $h-$ and $v-$ covariant derivatives of $C\Gamma$. The importance of the Cartan connection comes from its main role played in the generalized *Finsler–Lagrangian theory* of physical fields.

Regarding the *unified* field $g_{ij}(x, y)$ of GL^n, the authors of [MKA88, MA94] constructed a *Sasakian metric* on TM,

$$G = g_{ij} dx^i \otimes dx^j + g_{ij} \delta y^i \otimes \delta y^j.$$

In this context, the *Einstein equations* for the gravitational potentials $g_{ij}(x, y)$ of a generalized Lagrangian space GL^n, $(n > 2)$, are postulated as being the Einstein equations attached to $C\Gamma$ and G,

$$R_{ij} - \frac{1}{2} R g_{ij} = KT^H_{ij}, \qquad {}'P_{ij} = KT^1_{ij},$$
$$S_{ij} - \frac{1}{2} S g_{ij} = KT^V_{ij}, \qquad {}''P_{ij} = -KT^2_{ij},$$

where $R_{ij} = R^m_{ijm}$, $S_{ij} = S^m_{ijm}$, $'P_{ij} = P^m_{ijm}$, $''P_{ij} = P^m_{imj}$ are the Ricci tensors of $C\Gamma$, $R = g^{ij}R_{ij}$ and $S = g^{ij}S_{ij}$ are the scalar curvatures, T^H_{ij}, T^V_{ij}, T^1_{ij}, T^2_{ij} are the components of the energy–momentum tensor T, and K is the Einstein constant (equal to 0 for vacuum). Moreover, the energy–momentum tensors T^H_{ij} and T^V_{ij} satisfy the *conservation laws* [MKA88, MA94]

$$KT^{H\ m}_{\ j|m} = -\frac{1}{2}(P^{hm}_{js}R^s_{hm} + 2R^s_{mj}P^m_s), \qquad KT^{V\ m}_{\ j|m} = 0.$$

The generalized Lagrangian theory of electromagnetism relies on the *canonical Liouville vector–field* $C = y^i \frac{\partial}{\partial y^i}$ and the Cartan connection $C\Gamma$ of the generalized Lagrangian space GL^n. In this context, we can introduce the *electromagnetic two—form* on TM [MA94]

$$F = F_{ij}\delta y^i \wedge dx^j + f_{ij}\delta y^i \wedge \delta y^j, \qquad \text{where}$$
$$F_{ij} = \frac{1}{2}[(g_{im}y^m)_{|j} - (g_{jm}y^m)_{|i}], \qquad f_{ij} = \frac{1}{2}[(g_{im}y^m)_{;j} - (g_{jm}y^m)_{;i}].$$

Using the Bianchi identities attached to the Cartan connection $C\Gamma$, they conclude that the electromagnetic components F_{ij} and f_{ij} are governed by the following *Maxwell–type equations*

$$F_{ij|k} + F_{jk|i} + F_{ki|j} = -[C_{imr}y^m + (g_{im}y^m)_{|r}]R^r_{jk},$$
$$F_{ij;k} + F_{jk;i} + F_{ki;j} = -(f_{ij|k} + f_{jk|i} + f_{ki|j}), \qquad f_{ij;k} + f_{jk;i} + f_{ki;j} = 0.$$

2.2.4 Hamiltonian Field Systems: Path–Integral Quantization

Recall that the Hamiltonian counterpart of the classical Lagrangian field theory (see subsection 2.2.2 above) is the *covariant Hamiltonian field theory*, in which momenta correspond to derivatives of fields with respect to all world coordinates. It is well–known that classical Lagrangian and covariant Hamiltonian field theories are equivalent in the case of a hyperregular Lagrangian, and they are quasi–equivalent if a Lagrangian is almost–regular (see [Sar93, Sar95, GMS97, GMS99, MS00a, Sar02a]). Further, in order to quantize covariant Hamiltonian field theory, one usually attempts to construct and quantize a multisymplectic generalization of the Poisson bracket. The path–integral quantization of covariant Hamiltonian field theory was recently suggested in [BS04].

Recall that the symplectic Hamiltonian technique applied to field theory leads to instantaneous Hamiltonian formalism on an infinite–dimensional phase–space coordinated by field functions at some instant of time (see [Got91c] for the strict mathematical exposition of this formalism). The true Hamiltonian counterpart of classical first–order Lagrangian field theory is covariant Hamiltonian formalism, where canonical momenta p^μ_i correspond to derivatives y^i_μ of fields y^i with respect to all world coordinates x^μ. This formalism has been developed since the 1970s in its polysymplectic, multisymplectic

and Hamilton–de Donder variants (see [GMS97, LM03]). In order to quantize covariant Hamiltonian field theory, one usually attempts to construct multisymplectic generalization of the Poisson bracket with respect to the derivatives $\partial/\partial y^i$ and $\partial/\partial p_i^\mu$ [Kan98].

We can also quantize covariant Hamiltonian field theory in path–integral terms following [BS04]. A polysymplectic Hamiltonian system with a Hamiltonian $\mathcal{H}(x^\mu, y^i, p_i^\mu)$ is equivalent to a Lagrangian system with the Lagrangian

$$\mathcal{L}_\mathcal{H}(x^\mu, y^i, p_i^\mu, y_\alpha^i) = p_i^\alpha y_\alpha^i - \mathcal{H}(x^\mu, y^i, p_i^\mu, y_\alpha^i) \tag{2.256}$$

of the variables y^i and p_i^μ. In subsection 2.2.4 below we will quantize this Lagrangian system in the framework of *perturbative quantum field theory*. Briefly, if there is no constraint and the matrix $\partial^2\mathcal{H}/\partial p_i^\mu \partial p_j^\nu$ is nondegenerate and positive–definite, this quantization is given by the generating functional

$$Z = \mathcal{N}^{-1} \int \exp\{\int (\mathcal{L}_\mathcal{H} + \Lambda + iJ_i y^i + iJ_\mu^i p_i^\mu)dx\} \prod_x [dp(x)][dy(x)] \tag{2.257}$$

of Euclidean Green functions, where Λ comes from the normalization condition

$$\int \exp\{\int (-\frac{1}{2}\partial_\mu^i \partial_\nu^j \mathcal{H} p_i^\mu p_j^\nu + \Lambda)dx\} \prod_x [dp(x)] = 1.$$

If a Hamiltonian $\mathcal{H}$ is degenerate, the Lagrangian $\mathcal{L}_\mathcal{H}$ (2.256) may admit gauge symmetries. In this case, integration of a generating functional along gauge group orbits must be finite. If there are constraints, the Lagrangian system with a Lagrangian $\mathcal{L}_\mathcal{H}$ (2.256) restricted to the constraint manifold is quantized.

In order to verify this path–integral quantization scheme, we apply it to Hamiltonian field systems associated to Lagrangian field systems with quadratic Lagrangians

$$\mathcal{L} = \frac{1}{2} a_{ij}^{\lambda\mu} y_\alpha^i y_\mu^j + b_i^\alpha y_\alpha^i + c, \tag{2.258}$$

where a, b and c are functions of world coordinates x^μ and field variables y^i. Note that, in the framework of perturbative quantum field theory, any Lagrangian is split into the sum of a quadratic Lagrangian (2.258) and an interaction term quantized as a perturbation.

For example, let the Lagrangian (2.258) be hyperregular, i.e., the matrix function a is nondegenerate. Then there exists a unique associated Hamiltonian system whose Hamiltonian $\mathcal{H}$ is quadratic in momenta p_i^μ, and so is the Lagrangian $\mathcal{L}_\mathcal{H}$ (2.256). If the matrix function a is positive–definite on an Euclidean space–time, the generating functional (2.257) is a Gaussian integral of momenta $p_i^\mu(x)$. Integrating Z with respect to $p_i^\mu(x)$, one restarts the generating functional of quantum field theory with the original Lagrangian $\mathcal{L}$ (2.258). We extend this result to field theories with almost–regular Lagrangians $\mathcal{L}$

(2.258), e.g., *Yang–Mills gauge theory*. The key point is that, though such a Lagrangian $\mathcal{L}$ induces constraints and admits different associated Hamiltonians $\mathcal{H}$, all the Lagrangians $\mathcal{L}_{\mathcal{H}}$ coincide on the constraint manifold, and we have a unique constrained Hamiltonian system which is quasi–equivalent to the original Lagrangian one [GMS97].

Covariant Hamiltonian Field Systems

To develop the covariant Hamiltonian field theory suitable for path–integral quantization, we start by following the geometrical formulation of classical field theory (see [Sar93, Sar95, GMS97, MS00a, Sar02a]), in which classical fields are represented by sections of fibre bundles. Let $Y \to X$ be a smooth fibre bundle provided with bundle coordinates (x^μ, y^i). Recall from subsection 2.2.2 above, that the configuration space of Lagrangian field theory on Y is the 1–jet space $J^1(X,Y)$ of Y. It is equipped with the bundle coordinates (x^μ, y^i, y^i_μ) compatible with the composite fibration

$$J^1(X,Y) \stackrel{\pi^1_0}{\longrightarrow} Y \stackrel{\pi}{\longrightarrow} X.$$

Any section $s : X \longrightarrow Y$ of a fibre bundle $Y \to X$ is prolonged to the section $j^1 s : X \longrightarrow J^1(X,Y)$ of the jet bundle $J^1(X,Y) \to X$, such that $y^i_\mu \circ j^1 s = \partial_\mu s^i$.

Also, recall that the first–order Lagrangian is defined as a *horizontal density*

$$L = \mathcal{L}\omega : J^1(X,Y) \to \wedge^n T^*X, \qquad (\omega = dx^1 \wedge \cdots dx^n,\ n = \dim X), \quad (2.259)$$

on the jet space $J^1(X,Y)$. The corresponding *Euler–Lagrangian equations*

$$(\partial_i - d_\alpha \partial^\alpha_i)\mathcal{L} = 0, \qquad d_\alpha = \partial_\alpha + y^i_\alpha \partial_i + y^i_{\lambda\mu}\partial^\mu_i, \quad (2.260)$$

represent the subset of the 2–jet space $J^2(X,Y)$ of Y, coordinated by $(x^\mu, y^i, y^i_\alpha, y^i_{\lambda\mu})$. A section s of $Y \to X$ is a solution of these equations if its second jet prolongation $j^2 s$ lives in the subset (2.260).

The phase–space of covariant (polysymplectic) Hamiltonian field theory on a fibre bundle $Y \longrightarrow X$ is the *Legendre bundle* (see (2.215) above)

$$\Pi = \wedge^n T^*X \otimes V^*Y \otimes TX = V^*Y \wedge (\wedge^{n-1} T^*X), \quad (2.261)$$

where V^*Y is the vertical cotangent bundle of $Y \to X$. The Legendre bundle Π is equipped with the holonomic bundle coordinates (x^α, y^i, p^μ_i) compatible with the *composite fibration*

$$\Pi \stackrel{\pi_Y}{\longrightarrow} Y \stackrel{\pi}{\longrightarrow} X, \quad (2.262)$$

admitting the canonical polysymplectic form

$$\Omega = dp_i^\alpha \wedge dy^i \wedge \omega \otimes \partial_\alpha.$$

A *covariant Hamiltonian* $\mathcal{H}$ on Π (2.262) is defined as a section $p = -\mathcal{H}$ of the trivial line bundle (i.e., 1D fibre bundle)

$$Z_Y = T^*Y \wedge (\wedge^{n-1}T^*X) \to \Pi, \tag{2.263}$$

equipped with holonomic bundle coordinates $(x^\alpha, y^i, p_i^\mu, p)$. This fibre bundle admits the canonical multisymplectic Liouville form

$$\Xi = p\omega + p_i^\alpha dy^i \wedge \omega_\alpha, \qquad \text{with} \qquad \omega_\alpha = \partial_\alpha \rfloor \omega.$$

The pull–back of Ξ onto Π by a Hamiltonian $\mathcal{H}$ gives the *Hamiltonian form*

$$H = \mathcal{H}^*\Xi_Y = p_i^\alpha dy^i \wedge \omega_\alpha - \mathcal{H}\omega \tag{2.264}$$

on Π. The corresponding *covariant Hamiltonian equations* on Π,

$$y_\alpha^i = \partial_\alpha^i \mathcal{H}, \qquad p_{\lambda i}^\alpha = -\partial_i \mathcal{H}, \tag{2.265}$$

represent the closed submanifold of the jet space $J^1(X, \Pi)$ of Π. A section r of $\Pi \to X$ is a solution of these equations if its jet prolongation $j^1 r$ lives in the submanifold (2.265).

A section r of $\Pi \longrightarrow X$ is a solution of the covariant Hamiltonian equations (2.265) iff it satisfies the condition $r^*(u \rfloor dH) = 0$ for any vertical vector–field u on $\Pi \longrightarrow X$.

Alternatively, a section r of $\Pi \longrightarrow X$ is a solution of the covariant Hamiltonian equations (2.265) iff it is a solution of the Euler–Lagrangian equations for the first–order Lagrangian $L_\mathcal{H}$ on $J^1(X, \Pi)$,

$$L_\mathcal{H} = h_0(H) = \mathcal{L}_\mathcal{H}\omega = (p_i^\alpha y_\alpha^i - \mathcal{H})\omega, \tag{2.266}$$

where h_0 sends exterior forms on Π onto horizontal exterior forms on $J^1(X, \Pi) \longrightarrow X$, using the rule $h_0(dy^i) = y_\alpha^i dx^\alpha$.

Note that, for any section r of $\Pi \longrightarrow X$, the pull–backs r^*H and $j^1 r^* L_\mathcal{H}$ coincide. This fact motivated [BS04] to quantize covariant Hamiltonian field theory with a Hamiltonian $\mathcal{H}$ on Π as a Lagrangian system with the Lagrangian $L_\mathcal{H}$ (2.266).

Furthermore, let $i_N : N \longrightarrow \Pi$ be a closed imbedded subbundle of the Legendre bundle $\Pi \longrightarrow Y$ which is regarded as a constraint space of a covariant Hamiltonian field system with a Hamiltonian $\mathcal{H}$. This Hamiltonian system is restricted to N as follows. Let $H_N = i_N^* H$ be the pull–back of the Hamiltonian form H (2.264) onto N. The constrained Hamiltonian form H_N defines the constrained Lagrangian

$$L_N = h_0(H_N) = (j^1 i_N)^* L_\mathcal{H} \tag{2.267}$$

on the jet space $J^1(X, N_L)$ of the fibre bundle $N_L \longrightarrow X$. The Euler–Lagrangian equations for this Lagrangian are called the *constrained Hamiltonian equations*.

Note that, the Lagrangian $L_{\mathcal{H}}$ (2.266) is the pull–back onto $J^1(X,\Pi)$ of the horizontal form $L_{\mathcal{H}}$ on the bundle product $\Pi \times J^1(X,Y)$ over Y by the canonical map $J^1(X,\Pi) \to \Pi \times J^1(X,Y)$. Therefore, the constrained Lagrangian L_N (2.267) is the restriction of $L_{\mathcal{H}}$ to $N \times J^1(X,Y)$.

A section r of the fibre bundle $N \to X$ is a solution of constrained Hamiltonian equations iff it satisfies the condition $r^*(u_N \rfloor dH) = 0$ for any vertical vector–field u_N on $N \to X$.

Any solution of the covariant Hamiltonian equations (2.265) which lives in the constraint manifold N is also a solution of the constrained Hamiltonian equations on N. This fact motivates us to quantize covariant Hamiltonian field theory on a constraint manifold N as a Lagrangian system with the pull–back Lagrangian L_N (2.267).

Since a constraint manifold is assumed to be a closed imbedded submanifold of Π, there exists its open neighborhood U which is a fibre bundle $U \to N$. If Π is a fibre bundle $\pi_N : \Pi \to N$ over N, it is often convenient to quantize a Lagrangian system on Π with the pull–back Lagrangian $\pi_N^* L_N$, but integration of the corresponding generating functional along the fibres of $\Pi \to N$ must be finite.

In order to verify this quantization scheme, let us associate to a Lagrangian field system on Y a covariant Hamiltonian system on Π, then let us quantize this Hamiltonian system and compare this quantization with that of an original Lagrangian system.

Associated Lagrangian and Hamiltonian Systems

In order to relate classical Lagrangian and covariant Hamiltonian field theories, let us recall that, besides the Euler–Lagrangian equations, a Lagrangian L (2.259) also induces the Cartan equations which are given by the subset

$$(\overline{y}^j_\mu - y^j_\mu)\partial^\alpha_i \partial^\mu_j \mathcal{L} = 0, \tag{2.268}$$
$$\partial_i \mathcal{L} - \overline{d}_\alpha \partial^\alpha_i \mathcal{L} + (\overline{y}^j_\mu - y^j_\mu)\partial_i \partial^\mu_j \mathcal{L} = 0, \qquad \overline{d}_\alpha = \partial_\alpha + \overline{y}^i_\alpha \partial_i + \overline{y}^i_{\lambda\mu}\partial^\mu_i,$$

of the repeated jet space $J^1J^1(X,Y)$ coordinated by $(x^\mu, y^i, y^i_\lambda, \overline{y}^i_\alpha, \overline{y}^i_{\lambda\mu})$. A solution of the Cartan equations is a section $\overline{s}$ of the jet bundle $J^1(X,Y) \to X$ whose jet prolongation $j^1\overline{s}$ lives in the subset (2.268). Every solution s of the Euler–Lagrangian equations (2.260) defines the solution $j^1 s$ of the Cartan equations (2.268). If $\overline{s}$ is a solution of the Cartan equations and $\overline{s} = j^1 s$, then s is a solution of the Euler–Lagrangian equations. If a Lagrangian L is regular, the equations (2.260) and (2.268) are equivalent.

Recall that any Lagrangian L (2.259) induces the Legendre map (2.216), i.e.,

$$\widehat{L} : J^1(X,Y) \longrightarrow \Pi, \qquad p^\alpha_i \circ \widehat{L} = \partial^\alpha_i \mathcal{L}, \tag{2.269}$$

over Id_Y whose image $N_L = \widehat{L}(J^1(X,Y))$ is called the Lagrangian constraint space. A Lagrangian L is said to be hyperregular if the Legendre map (2.269)

is a diffeomorphism. A Lagrangian L is called almost–regular if the Lagrangian constraint space is a closed imbedded subbundle $i_N : N_L \to \Pi$ of the Legendre bundle $\Pi \to Y$ and the surjection $\widehat{L} : J^1(X,Y) \to N_L$ is a submersion (i.e., a fibre bundle) whose fibres are connected. Conversely, any Hamiltonian $\mathcal{H}$ induces the Hamiltonian map

$$\widehat{H} : \Pi \to J^1(X,Y), \qquad y^i_\alpha \circ \widehat{H} = \partial^i_\alpha \mathcal{H}. \tag{2.270}$$

A Hamiltonian $\mathcal{H}$ on Π is said to be associated to a Lagrangian L on $J^1(X,Y)$ if $\mathcal{H}$ satisfies the following relations (with $(x^\mu, y^i, p^\mu_i) \in N_L$)

$$\widehat{L} \circ \widehat{H} \circ \widehat{L} = \widehat{L}, \qquad p^\mu_i = \partial^\mu_i \mathcal{L}(x^\mu, y^i, \partial^j_\alpha \mathcal{H}), \tag{2.271}$$

$$\widehat{H}^* L_{\mathcal{H}} = \widehat{H}^* L, \qquad p^\mu_i \partial^i_\mu \mathcal{H} - \mathcal{H} = \mathcal{L}(x^\mu, y^j, \partial^j_\alpha \mathcal{H}). \tag{2.272}$$

If an associated Hamiltonian $\mathcal{H}$ exists, the Lagrangian constraint space N_L is given by the coordinate relations (2.271) and $\widehat{H} \circ \widehat{L}$ is a projector from Π onto N_L.

For example, any hyperregular Lagrangian L admits a unique associated Hamiltonian $\mathcal{H}$ such that

$$\widehat{H} = \widehat{L}^{-1}, \qquad \mathcal{H} = p^\mu_i \widehat{L^i_\mu}^{-1} - \mathcal{L}(x^\alpha, y^i, \widehat{L^i_\alpha}^{-1}).$$

In this case, any solution s of the Euler–Lagrangian equations (2.260) defines the solution $r = \widehat{L} \circ j^1 s$, of the covariant Hamiltonian equations (2.265). Conversely, any solution r of these Hamiltonian equations induces the solution $s = \pi_Y \circ r$ of the Euler–Lagrangian equations (2.260).

A degenerate Lagrangian need not admit an associated Hamiltonian. If such a Hamiltonian exists, it is not necessarily unique. Let us restrict our consideration to almost–regular Lagrangians. From the physical viewpoint, the most of Lagrangian field theories is of this type. From the mathematical one, this notion of degeneracy is particularly appropriate for the study of relations between Lagrangian and covariant Hamiltonian formalisms as follows [BS04].

Let L be an almost–regular Lagrangian and $\mathcal{H}$ an associated Hamiltonian. Let a section r of $\Pi \longrightarrow X$ be a solution of the covariant Hamiltonian equations (2.265) for $\mathcal{H}$. If r lives in the constraint manifold N_L, then $s = \pi_Y \circ r$ satisfies the Euler–Lagrangian equations (2.260) for L, while $\overline{s} = \widehat{H} \circ r$ obeys the Cartan equations (2.268). Conversely, let $\overline{s}$ be a solution of the Cartan equations (2.268) for L. If $\mathcal{H}$ satisfies the relation

$$\widehat{H} \circ \widehat{L} \circ \overline{s} = j^1(\pi^1_0 \circ \overline{s}),$$

the section $r = \widehat{L} \circ \overline{s}$ of the Legendre bundle $\Pi \longrightarrow X$ is a solution of the Hamiltonian equations (2.265) for $\mathcal{H}$. If $\overline{s} = j^1 s$, we get the relation between solutions the Euler–Lagrangian equations and the covariant Hamiltonian ones.

Due to this theorem, one need a set of different associated Hamiltonians in order to recover all solutions of the Euler–Lagrangian and Cartan equations for an almost–regular Lagrangian L. We can overcome this ambiguity as follows.

Let $\mathcal{H}$, $\mathcal{H}'$ be two different Hamiltonians associated to an almost–regular Lagrangian L. Let H, H' be the corresponding Hamiltonian forms (2.264). Their pull–backs i_N^*H and i_N^*H' onto the Lagrangian constraint manifold N_L coincide with each other.

It follows that, if an almost–regular Lagrangian admits associated Hamiltonians $\mathcal{H}$, it defines a unique constrained Hamiltonian form $H_N = i_N^*H$ on the Lagrangian constraint manifold N_L and a unique constrained Lagrangian $L_N = h_0(H_N)$ (2.267) on the jet space $J^1(X, N_L)$ of the fibre bundle $N_L \to X$. For any Hamiltonian $\mathcal{H}$ associated to L, every solution r of the Hamiltonian equations which lives in the Lagrangian constraint space N_L is a solution of the constrained Hamiltonian equations for L_N.

Let an almost–regular Lagrangian L admit associated Hamiltonians. A section $\overline{s}$ of the jet bundle $J^1(X,Y) \to X$ is a solution of the Cartan equations for L iff $\widehat{L} \circ \overline{s}$ is a solution of the constrained Hamiltonian equations. In particular, any solution r of the constrained Hamiltonian equations gives the solution $\overline{s} = \widehat{H} \circ r$ of the Cartan equations. This theorem shows that the constrained Hamiltonian equations and the Cartan equations are quasi–equivalent. Thus, one can associate to an almost–regular Lagrangian L (2.259) a unique constrained Lagrangian system on the constraint Lagrangian manifold N_L. Let us compare quantizations of these Lagrangian systems on Y and $N_L \subset \Pi$ in the case of an almost–regular quadratic Lagrangian L.

Evolution Operator

Recall that the covariant Hamiltonian field theory is mainly developed in its multisymplectic and polysymplectic variants, related to the two different Legendre maps in the first–order calculus of variations on fibre bundles [Sar02c] (also, see [Got91c] for a survey of symplectic formalism).

Recall that, given a fibre bundle $Y \to X$ coordinated by (x^α, y^i), every first–order Lagrangian $L : J^1(X,Y) \to \wedge^n T^*X$ is given by $L = \mathcal{L}\omega$, $\omega = dx^1 \wedge \cdots dx^n$, $(n = \dim X)$, and induces the Legendre map of the 1–jet space $J^1(X,Y)$ to the Legendre bundle

$$\Pi = \wedge^n T^*X \otimes V^*Y \otimes TX, \tag{2.273}$$

coordinated by $(x^\alpha, y^i, p_i^\alpha)$. The Π admits the *canonical polysymplectic form*

$$\Omega_\Pi = dp_i^\alpha \wedge dy^i \wedge \omega \otimes \partial_\alpha, \tag{2.274}$$

and is regarded as the polysymplectic phase–space of fields.

The multisymplectic phase–space of fields is the homogeneous Legendre bundle

$$Z = T^*Y \wedge (\wedge^{n-1} T^*X), \tag{2.275}$$

coordinated by $(x^\alpha, y^i, p_i^\alpha, p)$ and equipped with the canonical multisymplectic form

$$\Omega = d\Xi = dp \wedge \omega + dp_i^\alpha \wedge dy^i \wedge \omega_\alpha, \qquad \text{with} \qquad \omega_\alpha = \partial_\alpha \rfloor \omega. \tag{2.276}$$

It is natural that one attempts to generalize a Poisson bracket on symplectic manifolds to polysymplectic and multisymplectic manifolds in order to get the covariant canonical quantization of field theory. Different variants of such a bracket have been suggested. However, it seems that no canonical bracket corresponds to the TX–valued polysymplectic form (2.274), unless $\dim X = 1$. On the contrary, using the exterior multisymplectic form (2.276), one can associate multivector–fields to exterior forms on the fibre bundle Z (2.275) (but not to all of them), and can introduce a desired bracket of these forms via the Schouten–Nijenhuis bracket of multivector–fields.

Note that no bracket determines the evolution operator in polysymplectic and multisymplectic Hamiltonian formalism, including the case of $\dim X = 1$ of the time–dependent mechanics (see subsection 2.1.2 above). Written as a bracket, the evolution operator can be quantized, and it determines the *Heisenberg equation*.[3]

Recall the following relationship between first–order dynamical equations, connections, multivector–fields and evolution operators on a fibre bundle.

(i) Let $\pi : Q \to X$ be a fibre bundle coordinated by (x^μ, q^a). As a section $\gamma : Q \longrightarrow J^1(X, Q)$ of the 1–jet bundle $J^1(X, Q) \to Q$, any connection

$$\gamma = dx^\mu \otimes (\partial_\mu + \gamma_\mu^a \partial_a), \tag{2.277}$$

on $Q \to X$ defines the first–order differential operator

$$D : J^1(X, Q) \to T^*X \otimes VQ, \qquad (x^\mu, q^a, q_\mu^a) \to (x^\mu, q^a, q_\mu^a - \gamma_\mu^a(x^\nu, q^b)) \tag{2.278}$$

on $Q \to X$ called the *covariant differential* with respect to γ. The kernel of this differential operator is a closed imbedded subbundle of $J^1(X, Q) \to X$, given by the first–order dynamical equation

$$q_\mu^a - \gamma_\mu^a(x^\nu, q^b) = 0 \tag{2.279}$$

on a fibre bundle $Q \to X$. Conversely, any first–order dynamical equation on $Q \to X$ is of this type.

(ii) Let $HQ \subset TQ$ be the horizontal distribution defined by a connection γ. If X is orientable, there exists a nowhere vanishing global section of the exterior product $\wedge^n HQ \to Q$. It is a locally decomposable π-transverse $n-$vector–field on Q. Conversely, every multivector–field of this type on $Q \to X$ induces

[3] Recall that in the (Lorentz–invariant) *Heisenberg quantum picture*, the state vector $|\psi >$ does not change with time, and an observable A satisfies the Heisenberg equation of motion

$$\dot{A} = (i\hbar)^{-1}[A, H] + (\partial_t A)_{\text{classic}},$$

which becomes the classical Poisson equation if we replace its commutator $[A, H]$ by the Poisson bracket A, H.

a connection and, consequently, a first–order dynamical equation on this fibre bundle [EMR98].

(iii) Given a first–order dynamical equation γ on a fibre bundle $Q \to X$, the corresponding *evolution operator* d_γ is defined as the pull–back d_γ onto the shell of the horizontal differential

$$d_H = dx^\mu(\partial_\mu + q^a_\mu \partial_a)$$

acting on smooth real functions on Q. It reads

$$d_\gamma f = (\partial_\mu + \gamma^a_\mu \partial_a) f dx^\mu, \qquad (f \in C^\infty(Q)). \tag{2.280}$$

This expression shows that d_γ is projected onto Q, and it is a first–order differential operator on functions on Q. In particular, if a function f obeys the evolution equation $d_\gamma f = 0$, it is constant on any solution of the dynamical equation (2.279).

In Hamiltonian dynamics on Q, a problem is to represent the evolution operator (2.280) as a bracket of f with some exterior form on Q.

First of all, let us study the case of fibre bundles $Y \longrightarrow X$ over a 1D orientable connected manifold X (i.e., X is either $\mathbb{R}$ or S^1). In this case, the Legendre bundle Π (2.273) is isomorphic to the vertical cotangent bundle V^*Y of $Y \longrightarrow X$ coordinated by (x, y^i, p_i), and the polysymplectic form Ω_Π (2.274) on V^*Y reads

$$\Omega_\Pi = dp_i \wedge dy^i \wedge dx \otimes \partial_x. \tag{2.281}$$

Therefore, the homogeneous Legendre bundle (2.275) is the cotangent bundle T^*Y, coordinated by (x, y^i, p, p_i), and the multisymplectic form (2.276) becomes the canonical symplectic form on T^*Y, given by

$$\Omega = dp \wedge dx + dp_i \wedge dy^i. \tag{2.282}$$

The vertical cotangent bundle V^*Y admits the canonical *Poisson bracket*

$$\{f, f'\}_V = \partial^i f \partial_i f' - \partial_i f \partial^i f', \qquad (f, f' \in C^\infty(V^*Y)), \tag{2.283}$$

given by the relation

$$\zeta^*\{f, f'\}_V = \{\zeta^* f, \zeta^* f'\},$$

where $\{,\}$ is the canonical Poisson bracket on T^*Y [MS98, Sar98]. However, the Poisson structure (2.283) fails to determine Hamiltonian dynamics on the fibre bundle $V^*Q \to X$ because all Hamiltonian vector–fields with respect to this structure are vertical. At the same time, in accordance with general polysymplectic formalism [GMS97], a section h, $p \circ h = -\mathcal{H}$, of the fibre bundle $V^*Y \longrightarrow T^*Y$ induces a polysymplectic Hamiltonian form on V^*Y,

$$H = p_i dy^i - \mathcal{H} dx. \tag{2.284}$$

The associated Hamiltonian connection on $V^*Y \longrightarrow X$ with respect to the polysymplectic form (2.281) is

$$\gamma_H = dx \otimes (\partial_x + \partial^i \mathcal{H} \partial_i - \partial_i \mathcal{H} \partial^i). \tag{2.285}$$

It defines the Hamiltonian equation on V^*Y,

$$y^i_x = \partial^i \mathcal{H}, \qquad p_{ix} = -\partial_i \mathcal{H}.$$

The corresponding evolution operator (2.280) takes the local form

$$d_\gamma f = (\partial_x f + \{\mathcal{H}, f\}_V) dx, \qquad (f \in C^\infty(V^*Q)). \tag{2.286}$$

The bracket $\{\mathcal{H}, f\}_V$ in this expression is not globally defined because $\mathcal{H}$ is not a function on V^*Q. Therefore, the evolution operator (2.286) does not reduce to the Poisson bracket (2.283).

Let us now consider the pull–back $\zeta^* H$ of the Hamiltonian form H (2.284) onto T^*Y. Then the difference

$$H^* = \Xi - \zeta^* H = (p + \mathcal{H}) dx \tag{2.287}$$

is a horizontal density on the fibre bundle $T^*Y \to X$. It is a multisymplectic Hamiltonian form. The corresponding Hamiltonian connection γ on $T^*Y \to X$ is given by the condition

$$\gamma(\Omega) = dH^*, \tag{2.288}$$

where the map

$$\gamma(\Omega) = dx \wedge [(\partial_x + \gamma^p \partial_p + \gamma^i \partial_i + \gamma_i \partial^i) \rfloor \Omega]$$

is induced by an endomorphism of T^*Y determined by the tangent–valued form γ. We get

$$\gamma = dx \otimes (\partial_x + \gamma^p \partial_p + \partial^i \mathcal{H} \partial_i - \partial_i \mathcal{H} \partial^i), \tag{2.289}$$

where the coefficient γ^p is arbitrary. Note that this connection projects to the connection γ_H (2.285) on $V^*Y \to X$. As a consequence, it defines the evolution operator whose restriction to the pull–back of functions on V^*Q is exactly the evolution operator (2.286). But now this operator locally reduces to the Poisson bracket on T^*Y,

$$d_\gamma f = \{p + \mathcal{H}, f\} dx, \qquad (f \in C^\infty(V^*Y)). \tag{2.290}$$

However, this bracket is not globally defined, too, since $p + \mathcal{H}$ is a horizontal density, but not a function on T^*Y.

Let us introduce the function $\mathcal{E} = \rho^{-1}(p + \mathcal{H})$ on T^*Y, where ρdx is some nowhere vanishing density on X. The Hamiltonian vector–field of $\mathcal{E}$ with respect to the symplectic form Ω on T^*Y reads

$$\vartheta_{\mathcal{E}} = \rho^{-1}\partial_x - \partial_x\mathcal{E}\partial^p + \partial^i\mathcal{E}\partial_i - \partial_i\mathcal{E}\partial^i.$$

This vector–field is horizontal with respect to the connection (2.289), where $\gamma_p = -\rho\partial_x\mathcal{E}$, and it determines this connection in the form

$$\gamma = dx \otimes (\partial_x - \rho\partial_x\mathcal{E}\partial^p + \rho\partial^i\mathcal{E}\partial_i - \rho\partial_i\mathcal{E}\partial^i).$$

Therefore, the evolution operator (2.290) is rewritten as

$$d_\gamma f = \rho\{\mathcal{E}, f\}dx, \tag{2.291}$$

The bracket $\{\mathcal{E}, f\}$ is well–defined, but d_γ does not equal this bracket because of the factor ρ.

The multisymplectic bracket with the horizontal density H^* (2.287) also cannot help us since there is no Hamiltonian multivector–field associated to H^* relative to the symplectic form Ω.

The manifolds $X = \mathbb{R}$ and $X = S^1$ can be equipped with coordinates x possessing transition functions $x' = x + \text{const}$, and one can always choose the density $\rho = 1$. Then the evolution operator (2.291) reduces to a Poisson bracket in full. If $X = \mathbb{R}$, this is the case of time–dependent mechanics where time reparametrization is forbidden [MS00b, GMS02a].

Now we turn to the general case of $\dim X > 1$. In the framework of polysymplectic formalism [GMS97], a polysymplectic Hamiltonian form on the Legendre bundle Π (2.273) reads

$$H = p_i^\alpha dy^i \wedge \omega_\alpha - \mathcal{H}\omega. \tag{2.292}$$

The associated Hamiltonian connection

$$\gamma_H = dx^\alpha \otimes (\partial_\alpha + \gamma_\alpha^i\partial_i + \gamma_{i\lambda}^\mu\partial_\mu^i)$$

fails to be uniquely determined, but obeys the equations

$$\gamma_\alpha^i = \partial_\alpha^i\mathcal{H}, \qquad \gamma_{i\lambda}^\alpha = -\partial_i\mathcal{H}.$$

The values of these connections assemble into a closed imbedded subbubdle

$$y_\alpha^i = \partial_\alpha^i\mathcal{H}, \qquad p_{i\lambda}^\alpha = -\partial_i\mathcal{H}$$

of the jet bundle $J^1(X, \Pi) \longrightarrow X$ which is the first–order polysymplectic Hamiltonian equation on Π. This equation is not algebraically solved for the highest order derivatives and, therefore, it is not a dynamical equation. As a consequence, the evolution operator depends on the jet coordinates $p_{i\mu}^\alpha$ and, therefore, it is not a differential operator on functions on Π. Clearly, no bracket on Π can determine such an operator.

Quadratic Degenerate Systems

Given a fibre bundle $Y \to X$, let us consider a quadratic Lagrangian L (2.258), where a, b and c are local functions on Y. This property is coordinate–independent since $J^1(X,Y) \to Y$ is an affine bundle modelled over the vector bundle $T^*X \otimes VY$, where VY denotes the vertical tangent bundle of $Y \to X$. The associated Legendre map (2.269) reads

$$p_i^\alpha \circ \widehat{L} = a_{ij}^{\alpha\mu} y_\mu^j + b_i^\alpha. \tag{2.293}$$

Let a Lagrangian L (2.258) be almost–regular, i.e., the matrix function a is a linear bundle map

$$a : T^*X \otimes VY \to \Pi, \qquad p_i^\alpha = a_{ij}^{\alpha\mu} \overline{y}_\mu^j, \tag{2.294}$$

of constant rank, where $(x^\alpha, y^i, \overline{y}_\alpha^i)$ are bundle coordinates on $T^*X \otimes VY$. Then the Lagrangian constraint space N_L (2.293) is an affine subbundle of the Legendre bundle $\Pi \to Y$ (2.273). Hence, $N_L \to Y$ has a global section. For the sake of simplicity, let us assume that it is the canonical zero section $\widehat{0}(Y)$ of $\Pi \to Y$. The kernel of the Legendre map (2.293) is also an affine subbundle of the affine jet bundle $J^1(X,Y) \to Y$. Therefore, it admits a global section

$$\Gamma : Y \to \mathrm{Ker}\ \widehat{L} \subset J^1(X,Y), \qquad a_{ij}^{\alpha\mu} \Gamma_\mu^j + b_i^\alpha = 0, \tag{2.295}$$

which is a connection on $Y \to X$. If the Lagrangian (2.258) is regular, the connection (2.295) is unique.

There exists a linear bundle map

$$\sigma : \Pi \longrightarrow T^*X \otimes VY, \qquad \overline{y}_\alpha^i \circ \sigma = \sigma_{\alpha\mu}^{ij} p_j^\mu, \tag{2.296}$$

such that

$$a \circ \sigma \circ a = a, \qquad a_{ij}^{\alpha\mu} \sigma_{\mu\alpha}^{jk} a_{kb}^{\alpha\nu} = a_{ib}^{\alpha\nu}. \tag{2.297}$$

Note that σ is not unique, but it falls into the sum $\sigma = \sigma_0 + \sigma_1$ such that

$$\sigma_0 \circ a \circ \sigma_0 = \sigma_0, \qquad a \circ \sigma_1 = \sigma_1 \circ a = 0, \tag{2.298}$$

where σ_0 is uniquely defined. For example, there exists a nondegenerate map σ (2.296).

Recall that there are the splittings

$$J^1(X,Y) = \mathcal{S}(J^1(X,Y)) \oplus \mathcal{F}(J^1(X,Y)) = \mathrm{Ker}\,\widehat{L} \oplus \mathrm{Im}(\sigma_0 \circ \widehat{L}), \tag{2.299}$$

$$\begin{aligned} y_\alpha^i &= \mathcal{S}_\alpha^i + \mathcal{F}_\alpha^i = [y_\alpha^i - \sigma_0{}_{\alpha\mu}^{ik}(a_{kj}^{\alpha\mu} y_\mu^j + b_k^\alpha)] + [\sigma_0{}_{\alpha\mu}^{ik}(a_{kj}^{\alpha\mu} y_\mu^j + b_k^\alpha)], \\ \Pi &= \mathcal{R}(\Pi) \oplus \mathcal{P}(\Pi) = \mathrm{Ker}\,\sigma_0 \oplus N_L, \\ p_i^\alpha &= \mathcal{R}_i^\alpha + \mathcal{P}_i^\alpha = [p_i^\alpha - a_{ij}^{\alpha\mu} \sigma_0{}_{\mu\alpha}^{jk} p_k^\alpha] + [a_{ij}^{\alpha\mu} \sigma_0{}_{\mu\alpha}^{jk} p_k^\alpha]. \end{aligned} \tag{2.300}$$

The relations (2.298) lead to the equalities

$$a_{ij}^{\alpha\mu}\mathcal{S}_\mu^j = 0, \qquad \sigma_0{}_{\mu\alpha}^{jk}\mathcal{R}_k^\alpha = 0, \qquad \sigma_1{}_{\mu\alpha}^{jk}\mathcal{P}_k^\alpha = 0, \qquad \mathcal{R}_i^\alpha\mathcal{F}_\alpha^i = 0. \tag{2.301}$$

Due to these equalities, the Lagrangian (2.258) takes the form

$$L = \mathcal{L}\omega, \qquad \mathcal{L} = \frac{1}{2}a_{ij}^{\alpha\mu}\mathcal{F}_\alpha^i\mathcal{F}_\mu^j + c'. \tag{2.302}$$

One can show that, this Lagrangian admits a set of associated Hamiltonians

$$\mathcal{H}_\Gamma = (\mathcal{R}_i^\alpha + \mathcal{P}_i^\alpha)\Gamma_\alpha^i + \frac{1}{2}\sigma_0{}_{\alpha\mu}^{ij}\mathcal{P}_i^\alpha\mathcal{P}_j^\mu + \frac{1}{2}\sigma_1{}_{\alpha\mu}^{ij}\mathcal{R}_i^\alpha\mathcal{R}_j^\mu - c' \tag{2.303}$$

indexed by connections Γ (2.295). Therefore, the Lagrangian constraint manifold (2.293) is given by the reducible constraints

$$\mathcal{R}_i^\alpha = p_i^\alpha - a_{ij}^{\alpha\mu}\sigma_0{}_{\mu\alpha}^{jk}p_k^\alpha = 0. \tag{2.304}$$

Given a Hamiltonian $\mathcal{H}_\Gamma$, the corresponding Lagrangian (2.266) reads

$$\mathcal{L}_{\mathcal{H}_\Gamma} = \mathcal{R}_i^\alpha(\mathcal{S}_\alpha^i - \Gamma_\alpha^i) + \mathcal{P}_i^\alpha\mathcal{F}_\alpha^i - \frac{1}{2}\sigma_0{}_{\alpha\mu}^{ij}\mathcal{P}_i^\alpha\mathcal{P}_j^\mu - \frac{1}{2}\sigma_1{}_{\alpha\mu}^{ij}\mathcal{R}_i^\alpha\mathcal{R}_j^\mu + c'. \tag{2.305}$$

Its restriction (2.267) to the Lagrangian constraint manifold N_L (2.304) is

$$L_N = \mathcal{L}_N\omega, \qquad \mathcal{L}_N = \mathcal{P}_i^\alpha\mathcal{F}_\alpha^i - \frac{1}{2}\sigma_0{}_{\alpha\mu}^{ij}\mathcal{P}_i^\alpha\mathcal{P}_j^\mu + c'. \tag{2.306}$$

It is independent of the choice of a Hamiltonian (2.303). Note that the Lagrangian $\mathcal{L}_N$ may admit gauge symmetries due to the term $\mathcal{P}_i^\alpha\mathcal{F}_\alpha^i$.

The Hamiltonian $\mathcal{H}_\Gamma$ induces the Hamiltonian map $\widehat{H}_\Gamma$ (2.270) and the projector

$$T = \widehat{L}\circ\widehat{H}_\Gamma, \qquad p_i^\alpha\circ T = T_{i\mu}^{\alpha j}p_j^\mu = a_{ik}^{\alpha\nu}\sigma_0{}_{\nu\mu}^{kj}p_j^\mu = \mathcal{P}_i^\alpha, \tag{2.307}$$

from Π onto its summand N_L in the decomposition (2.300). It obeys the relations

$$\sigma\circ T = \sigma_0, \qquad T\circ a = a. \tag{2.308}$$

The projector T (2.307) is a linear map over Id_Y. Therefore, $T : \Pi \to N_L$ is a vector bundle. Let us consider the pull–back $L_\Pi = T^*L_N$ of the constrained Lagrangian L_N (2.306) onto Π. Due to the relations (2.301), it is given by the coordinate expression

$$L_\Pi = \mathcal{L}_\Pi\omega, \qquad \mathcal{L}_\Pi = p_i^\alpha\mathcal{F}_\alpha^i - \frac{1}{2}\sigma_0{}_{\alpha\mu}^{ij}p_i^\alpha p_j^\mu + c'. \tag{2.309}$$

This Lagrangian is gauge–invariant under the subgroup of the gauge group of vertical automorphisms Φ of the affine bundle $\Pi \to Y$ such that $T\circ\Phi = T$. This subgroup coincides with the gauge group $\mathrm{Aut}\,\mathrm{Ker}\,\sigma_0$ of vertical automorphisms of the vector bundle $\mathrm{Ker}\,\sigma_0 \to Y$.

Note that the splittings (2.299) and (2.300) result from the splitting of the vector bundle

$$T^*X \otimes VY = \operatorname{Ker} a \oplus E,$$

which can be provided with the adapted coordinates $(\overline{y}^a, \overline{y}^A)$ such that a (2.294) is brought into a diagonal matrix with nonvanishing components a_{AA}. Then the Legendre bundle $\Pi \to Y$ (2.273) admits the dual (nonholonomic) coordinates (p_a, p_A), where p_A are coordinates on the Lagrangian constraint manifold N_L, given by the irreducible constraints $p_a = 0$. Written relative to these coordinates, σ_0 becomes the diagonal matrix

$$\sigma_0^{AA} = (a_{AA})^{-1}, \qquad \sigma_0^{aa} = 0, \tag{2.310}$$

while $\sigma_1^{Aa} = \sigma_1^{AB} = 0$. Moreover, one can choose the coordinates $\overline{y}^a$ (accordingly, p_a) and the map σ (2.296) such that σ_1 becomes a diagonal matrix with nonvanishing positive components $\sigma_1^{aa} = \mathcal{V}^{-1}$, where $\mathcal{V}\omega$ is a volume form on X. We further follow this choice of the adapted coordinates (p_a, p_A). Let us write

$$p_a = M_{a\alpha}^{\ \ i} p_i^\alpha, \qquad p_A = M_{A\alpha}^{\ \ i} p_i^\alpha, \tag{2.311}$$

where M are the matrix functions on Y obeying the relations

$$M_{a\alpha}^{\ \ i} a_{ij}^{\alpha\mu} = 0, \qquad M^{-1}{}_i^{\alpha a} \sigma_{0\alpha}^{\ \ i} = 0, \qquad M_{a\alpha}^{\ \ i} \mathcal{P}_i^\alpha = 0, \qquad M_{A\alpha}^{\ \ i} \mathcal{R}_i^\alpha = 0.$$

Then the Lagrangian L_N (2.306) with respect to the adapted coordinates (p_a, p_A) takes the form

$$\mathcal{L}_N = M^{-1}{}_i^{\alpha A} p_A \mathcal{F}_\alpha^i - \frac{1}{2} \sum_A (a_{AA})^{-1} (p_A)^2 + c', \tag{2.312}$$

Path Integrals and Perturbative Quantum Fields

Recall that an elegant way to make geometrical path integrals rigorous is to formulate them using the jet formalism. In this way the covariant Hamiltonian field systems were presented in [BS04]. In this subsection we give a brief review of this perturbative quantum field model.

Let us quantize a Lagrangian system with the Lagrangian L_N (2.306) on the constraint manifold N_L (2.304). In the framework of a perturbative quantum field theory, we should assume that $X = \mathbb{R}^n$ and $Y \to X$ is a trivial affine bundle. It follows that both the original coordinates $(x^\alpha, y^i, p_i^\alpha)$ and the adapted coordinates $(x^\alpha, y^i, p_a, p_A)$ on the Legendre bundle Π are global. Passing to field theory on an Euclidean space $\mathbb{R}^n$, we also assume that the matrix a in the Lagrangian L (2.302) is positive–definite, i.e., $a_{AA} > 0$.

Let us start from a Lagrangian (2.306) without gauge symmetries. Since the Lagrangian constraint space N_L can be equipped with the adapted coordinates p_A, the generating functional of Euclidean Green functions of the Lagrangian system in question reads [BS04]

$$Z = \mathcal{N}^{-1} \int \exp\{\int (\mathcal{L}_N + \frac{1}{2}\operatorname{Tr}(\ln \overline{\sigma}_0) + iJ_i y^i + iJ^A p_A)\omega\} \prod_x [dp_A(x)][dy(x)], \tag{2.313}$$

where $\mathcal{L}_N$ is given by the expression (2.312) and $\overline{\sigma}_0$ is the square matrix

$$\overline{\sigma}_0^{AB} = M^{-1}{}_i^{\alpha A} M^{-1}{}_j^{\mu B} \sigma_0{}_{\alpha\mu}^{ij} = \delta^{AB}(a_{AA})^{-1}.$$

The generating functional (2.313) a Gaussian integral of variables $p_A(x)$. Its integration with respect to $p_A(x)$ under the condition $J^A = 0$ restarts the generating functional

$$Z = \mathcal{N}^{-1} \int \exp\{\int (\mathcal{L} + iJ_i y^i)\omega\} \prod_x [dy(x)], \tag{2.314}$$

of the original Lagrangian field system on Y with the Lagrangian (2.302). However, the generating functional (2.313) cannot be rewritten with respect to the original variables p_i^μ, unless a is a nondegenerate matrix function.

In order to overcome this difficulty, let us consider a Lagrangian system on the whole Legendre manifold Π with the Lagrangian L_Π (2.309). Since this Lagrangian is constant along the fibres of the vector bundle $\Pi \to N_L$, an integration of the generating functional of this field model with respect to variables $p_a(x)$ should be finite. One can choose the generating functional in the form [BS04]

$$\begin{aligned} Z = \mathcal{N}^{-1} \int \exp\{\int (\mathcal{L}_\Pi - \frac{1}{2}\sigma_1{}_{\alpha\mu}^{ij} p_i^\alpha p_j^\mu \\ + \frac{1}{2}\operatorname{Tr}(\ln \sigma) + iJ_i y^i + iJ_\mu^i p_i^\mu)\omega\} \prod_x [dp(x)][dy(x)]. \end{aligned} \tag{2.315}$$

Its integration with respect to momenta $p_i^\alpha(x)$ restarts the generating functional (2.314) of the original Lagrangian system on Y. In order to get the generating functional (2.315), one can follow a procedure of quantization of gauge–invariant Lagrangian systems. In the case of the Lagrangian L_Π (2.309), this procedure is rather trivial, since the space of momenta variables $p_a(x)$ coincides with the translation subgroup of the gauge group $\operatorname{Aut}\operatorname{Ker}\sigma_0$.

Now let us suppose that the Lagrangian L_N (2.306) and, consequently, the Lagrangian L_Π (2.309) are invariant under some gauge group G_X of vertical automorphisms of the fibre bundle $Y \to X$ (and the induced automorphisms of $\Pi \to X$) which acts freely on the space of sections of $Y \to X$. Its infinitesimal generators are represented by vertical vector–fields $u = u^i(x^\mu, y^j)\partial_i$ on $Y \to X$ which induce the vector–fields

$$\overline{u} = u^i\partial_i - \partial_j u^i p_i^\alpha \partial_\alpha^j + d_\alpha u^i \partial_i^\alpha, \qquad d_\alpha = \partial_\alpha + y_\alpha^i \partial_i, \tag{2.316}$$

on $\Pi \times J^1(X,Y)$. Let us also assume that G_X is indexed by m parameter functions $\xi^r(x)$ such that $u = u^i(x^\alpha, y^j, \xi^r)\partial_i$, where

$$u^i(x^\alpha, y^j, \xi^r) = u^i_r(x^\alpha, y^j)\xi^r + u^{i\mu}_r(x^\alpha, y^j)\partial_\mu \xi^r \tag{2.317}$$

are linear first–order differential operators on the space of parameters $\xi^r(x)$. The vector–fields $u(\xi^r)$ must satisfy the commutation relations

$$[u(\xi^q), u(\xi'^p)] = u(c^r_{pq}\xi'^p\xi^q),$$

where c^r_{pq} are structure constants. The Lagrangian L_Π (2.309) is invariant under the above gauge transformations iff its Lie derivative $\mathfrak{L}_{\overline{u}}L_\Pi$ along vector–fields (2.316) vanishes, i.e.,

$$(u^i\partial_i - \partial_j u^i p^\alpha_i \partial^j_\alpha + d_\alpha u^i \partial^\alpha_i)\mathcal{L}_\Pi = 0. \tag{2.318}$$

Since the operator $\mathfrak{L}_{\overline{u}}$ is linear in momenta p^μ_i, the condition (2.318) falls into the independent conditions

$$(u^k\partial_k - \partial_j u^k p^\nu_k \partial^j_\nu + d_\nu u^j \partial^\nu_j)(p^\alpha_i \mathcal{F}^i_\alpha) = 0, \tag{2.319}$$

$$(u^k\partial_k - \partial_j u^k p^\nu_k \partial^j_\nu)(\sigma_0{}^{ij}_{\lambda\mu} p^\alpha_i p^\mu_j) = 0, \tag{2.320}$$

$$u^i\partial_i c' = 0. \tag{2.321}$$

It follows that the Lagrangian L_Π is gauge–invariant iff its three summands are separately gauge–invariant.

Note that, if the Lagrangian L_Π on Π is gauge–invariant, the original Lagrangian L (2.302) is also invariant under the same gauge transformations. Indeed, one gets at once from the condition (2.319) that

$$\overline{u}(\mathcal{F}^i_\mu) = \partial_j u^i \mathcal{F}^j_\mu, \tag{2.322}$$

i.e., the quantity $\mathcal{F}$ is transformed as the dual of momenta p. Then the condition (2.320) shows that the quantity $\sigma_0 p$ is transformed by the same law as $\mathcal{F}$. It follows that the term $a\mathcal{F}\mathcal{F}$ in the Lagrangian L (2.302) is transformed exactly as $a(\sigma_0 p)(\sigma_0 p) = \sigma_0 pp$, i.e., is gauge–invariant. Then this Lagrangian is gauge–invariant due to the equality (2.321).

Since $\mathcal{S}^i_\alpha = y^i_\alpha - \mathcal{F}^i_\alpha$, one can derive from the formula (2.322) the transformation law of $\mathcal{S}$,

$$\overline{u}(\mathcal{S}^i_\mu) = d_\mu u^i - \partial_j u^i \mathcal{F}^j_\mu = d_\mu u^i - \partial_j u^i (y^j_\mu - \mathcal{S}^j_\mu) = \partial_\mu u^i + \partial_j u^i \mathcal{S}^j_\mu. \tag{2.323}$$

This expression shows that the gauge group G_X acts freely on the space of sections $\mathcal{S}(x)$ of the fibre bundle $\operatorname{Ker}\widehat{L} \to Y$ in the splitting (2.299). Let the number m of parameters of the gauge group G_X do not exceed the fibre dimension of $\operatorname{Ker}\widehat{L} \to Y$. Then some combinations $b^{r\mu}_i \mathcal{S}^i_\mu$ of $\mathcal{S}^i_\mu$ can be used as the gauge condition

$$b^{r\mu}_i \mathcal{S}^i_\mu(x) - \alpha^r(x) = 0,$$

similar to the generalized Lorentz gauge in *Yang–Mills gauge theory*.

Now we turn to path–integral quantization of a Lagrangian system with the gauge–invariant Lagrangian L_Π (2.309). In accordance with the well–known

quantization procedure, let us modify the generating functional (2.315) as follows [BS04]

$$
\begin{aligned}
&Z = \mathcal{N}^{-1} \int \exp\{\int (\mathcal{L}_{\Pi} - \frac{1}{2}\sigma_1{}^{ij}_{\alpha\mu} p_i^{\alpha} p_j^{\mu} \\
&+ \frac{1}{2}\operatorname{Tr}(\ln\sigma) - \frac{1}{2} h_{rs}\alpha^r\alpha^s + iJ_i y^i + iJ^i_{\mu} p_i^{\mu})\omega\} \\
&\Delta \prod_x \times^r \delta(b^r{}_i^{\mu}\mathcal{S}^i_{\mu}(x) - \alpha^r(x))[d\alpha(x)][dp(x)][dy(x)] \\
&= \mathcal{N}'^{-1} \int \exp\{\int (\mathcal{L}_{\Pi} - \frac{1}{2}\sigma_1{}^{ij}_{\alpha\mu} p_i^{\alpha} p_j^{\mu} + \frac{1}{2}\operatorname{Tr}(\ln\sigma) \\
&- \frac{1}{2} h_{rs} b^r{}_i^{\mu} b^s{}_j^{\alpha} \mathcal{S}^i_{\mu}\mathcal{S}^j_{\alpha} + iJ_i y^i + iJ^i_{\mu} p_i^{\mu})\omega\} \Delta \prod_x [dp(x)][dy(x)], \\
&\qquad \text{where} \qquad \int \exp\{\int (-\frac{1}{2} h_{rs}\alpha^r\alpha^s)\omega\} \prod_x [d\alpha(x)]
\end{aligned}
\tag{2.324}
$$

is a Gaussian integral, while the factor Δ is defined by the condition

$$
\Delta \int \prod_x \times^r \delta(u(\xi)(b^r{}_i^{\mu}\mathcal{S}^i_{\mu}))[d\xi(x)] = 1.
$$

We have the linear second–order differential operator

$$
M^r_s \xi^s = u(\xi)(b^r{}_i^{\mu}\mathcal{S}^i_{\mu}(x)) = b^r{}_i^{\mu}(\partial_{\mu} u^i(\xi) + \partial_j u^i(\xi)\mathcal{S}^j_{\mu}) \tag{2.325}
$$

on the parameter functions $\xi(x)$, and get $\Delta = \det M$. Then the generating functional (2.325) takes the form

$$
\begin{aligned}
Z = \mathcal{N}'^{-1} \int \exp\{\int (&\mathcal{L}_{\Pi} - \frac{1}{2}\sigma_1{}^{ij}_{\alpha\mu} p_i^{\alpha} p_j^{\mu} + \frac{1}{2}\operatorname{Tr}\,\ln\sigma - \frac{1}{2} h_{rs} b^r{}_i^{\mu} b^s{}_j^{\alpha} \mathcal{S}^i_{\mu}\mathcal{S}^j_{\alpha} \\
&- \overline{c}_r M^r_s c^s + iJ_i y^i + iJ^i_{\mu} p_i^{\mu})\omega\} \prod_x [d\overline{c}][dc][dp(x)][dy(x)],
\end{aligned}
\tag{2.326}
$$

where $\overline{c}_r$, c^s are odd ghost fields. Integrating Z (2.326) with respect to momenta under the condition $J^i_{\mu} = 0$, we come to the generating functional

$$
Z = \mathcal{N}'^{-1} \int \exp\{\int (\mathcal{L} - \frac{1}{2} h_{rs} b^r{}_i^{\mu} b^s{}_j^{\alpha} \mathcal{S}^i_{\mu}\mathcal{S}^j_{\alpha} - \overline{c}_r M^r_s c^s + iJ_i y^i)\omega\} \prod_x [d\overline{c}][dc][dy(x)] \tag{2.327}
$$

of the original field model on Y with the gauge–invariant Lagrangian L (2.302).

Note that the Lagrangian

$$
\mathcal{L}' = \mathcal{L} - \frac{1}{2} h_{rs} b^r{}_i^{\mu} b^s{}_j^{\alpha} \mathcal{S}^i_{\mu}\mathcal{S}^j_{\alpha} - \overline{c}_r M^r_s c^s \tag{2.328}
$$

fails to be gauge–invariant, but it admits the so–called *BRST symmetry*[4] whose *odd operator* reads

$$\vartheta = u^i(x^\mu, y^i, c^s)\partial_i + d_\alpha u^i(x^\mu, y^i, c^s)\partial_i^\alpha + \overline{v}_r(x^\mu, y^i, y^i_\mu)\frac{\partial}{\partial \overline{c}_r} \tag{2.329}$$
$$+ v^r(x^\mu, y^i, c^s)\frac{\partial}{\partial c^r} + d_\alpha v^r(x^\mu, y^i, c^s)\frac{\partial}{\partial c^r_\alpha} + d_\mu d_\alpha v^r(x^\mu, y^i, c^s)\frac{\partial}{\partial c^r_{\mu\lambda}},$$
$$d_\alpha = \partial_\alpha + y^i_\alpha\partial_i + y^i_{\lambda\mu}\partial_i^\mu + c^r_\alpha\frac{\partial}{\partial c^r} + c^r_{\lambda\mu}\frac{\partial}{\partial c^r_\mu}.$$

Its components $u^i(x^\mu, y^i, c^s)$ are given by the expression (2.317) where parameter functions $\xi^r(x)$ are replaced with the ghosts c^r. The components $\overline{v}_r$ and v^r of the BRST operator ϑ can be derived from the condition

$$\vartheta(\mathcal{L}') = -h_{rs}M^r_q b^s{}^\alpha_j\mathcal{S}^j_\alpha c^q - \overline{v}_r M^r_q c^q + \overline{c}_r\vartheta(\vartheta(b^r{}^\alpha_j\mathcal{S}^j_\alpha)) = 0$$

of the BRST invariance of $\mathcal{L}'$. This condition falls into the two independent relations

$$h_{rs}M^r_q b^s{}^\alpha_j\mathcal{S}^j_\alpha + \overline{v}_r M^r_q = 0,$$
$$\vartheta(c^q)(\vartheta(c^p)(b^r{}^\alpha_j\mathcal{S}^j_\alpha)) = u(c^p)(u(c^q)(b^r{}^\alpha_j\mathcal{S}^j_\alpha)) + u(v^r)(b^r{}^\alpha_j\mathcal{S}^j_\alpha)$$
$$= u(\frac{1}{2}c^r_{pq}c^p c^q + v^r)(b^r{}^\alpha_j\mathcal{S}^j_\alpha) = 0.$$

Hence, we get: $\overline{v}_r = -h_{rs}b^s{}^\alpha_j\mathcal{S}^j_\alpha$, and $v^r = -\frac{1}{2}c^r_{pq}c^p c^q$.

2.2.5 Gauge Fields on Principal Connections

Connection Strength

Given a principal $G-$bundle $P \to Q$, the *Frölicher–Nijenhuis bracket* (1.153) on the space $\wedge^*(P)\otimes\mathcal{V}^1(P)$ of tangent–valued forms on P is compatible with the canonical action R_G (1.128) of G on P, and induces the Frölicher–Nijenhuis bracket on the space $\wedge^*(Q)\otimes T_G P(Q)$ of $T_G P-$valued forms on Q. Its coordinate form issues from the Lie bracket (1.133).

Then any principal connection $A \in \wedge^1(Q)\otimes T_G P(Q)$ (1.198) sets the *Nijenhuis differential*

$$d_A : \wedge^r(Q)\otimes T_G P(Q) \to \wedge^{r+1}(Q)\otimes V_G P(Q),$$
$$d_A\phi = [A, \phi]_{FN}, \quad \phi \in \wedge^r(Q)\otimes T_G P(Q), \tag{2.330}$$

[4] Recall that the BRST formalism is a method of implementing first class constraints. The letters BRST stand for Becchi, Rouet, Stora, and (independently) Tyutin who discovered this formalism. It is a rigorous method to deal with quantum theories with gauge invariance.

on the space $\wedge^*(Q)\otimes T_GP(Q)$.

The curvature R (1.190) can be equivalently defined as the Nijenhuis differential

$$R: Y\to \wedge^2T^*Q\otimes VY, \qquad \text{given by} \qquad R=\frac{1}{2}d_\Gamma\Gamma=\frac{1}{2}[\Gamma,\Gamma]_{FN}. \quad (2.331)$$

Let us define the *strength* of a principal connection A, as

$$F_A=\frac{1}{2}d_AA=\frac{1}{2}[A,A]_{FN}\in\wedge^2(Q)\otimes V_GP(Q), \quad (2.332)$$

$$F_A=\frac{1}{2}F^r_{\lambda\mu}dq^\alpha\wedge dq^\mu\otimes e_r, \qquad F^r_{\lambda\mu}=\partial_\alpha A^r_\mu-\partial_\mu A^r_\alpha+c^r_{pq}A^p_\alpha A^q_\mu. \quad (2.333)$$

It is locally given by the expression

$$F_A=d\mathbf{A}+\frac{1}{2}[\mathbf{A},\mathbf{A}]=d\mathbf{A}+\mathbf{A}\wedge\mathbf{A}, \quad (2.334)$$

where $\mathbf{A}$ is the local connection form (1.199). By definition, the strength (2.332) of a principal connection obeys the second *Bianchi identity*

$$d_AF_A=[A,F_A]_{FN}=0. \quad (2.335)$$

It should be emphasized that the strength F_A (2.332) is not the standard curvature (1.190) of a connection on P, but there are the local relations $\psi^P_\zeta F_A=z^*_\zeta\Theta$, where

$$\Theta=d\widetilde{A}+\frac{1}{2}[\widetilde{A},\widetilde{A}] \quad (2.336)$$

is the $\mathfrak{g}_l$-valued curvature form on P (see the expression (2.340) below). In particular, a principal connection is flat iff its strength vanishes.

Associated Bundles

Given a principal $G-$bundle $\pi_P: P\to Q$, let V be a manifold provided with an effective left action $G\times V\to V$, $(g,v)\mapsto gv$ of the Lie group G. Let us consider the quotient

$$Y=(P\times V)/G \quad (2.337)$$

of the product $P\times V$ by identification of elements (p,v) and $(pg,g^{-1}v)$ for all $g\in G$. We will use the notation $(pG,G^{-1}v)$ for its points. Let $[p]$ denote the restriction of the canonical surjection

$$P\times V\to (P\times V)/G \quad (2.338)$$

to the subset $\{p\}\times V$ so that $[p](v)=[pg](g^{-1}v)$. Then the map $Y\to Q$, $[p](V)\mapsto\pi_P(p)$, makes the quotient Y (2.337) to a fibre bundle over Q.

Let us note that, for any $G-$bundle, there exists an associated principal G-bundle [Ste72]. The peculiarity of the $G-$bundle Y (2.337) is that it appears

canonically associated to a principal bundle P. Indeed, every bundle atlas $\Psi_P = \{(U_\alpha, z_\alpha)\}$ of P determines a unique associated bundle atlas

$$\Psi = \{(U_\alpha, \psi_\alpha(q) = [z_\alpha(q)]^{-1})\}$$

of the quotient Y (2.337), and each automorphism of P also induces the corresponding automorphism (2.353) of Y.

Every principal connection A on a principal bundle $P \to Q$ induces a unique connection on the associated fibre bundle Y (2.337). Given the horizontal splitting of the tangent bundle TP relative to A, the tangent map to the canonical map (2.338) defines the horizontal splitting of the tangent bundle TY of Y and, consequently, a connection on $Y \to Q$ [KN63/9]. This is called the associated principal connection or a principal connection on a $P-$associated bundle $Y \to Q$. If Y is a vector bundle, this connection takes the form

$$A = dq^\alpha \otimes (\partial_\alpha - A^p_\alpha I_p{}^i_j y^j \partial_i), \tag{2.339}$$

where I_p are generators of the linear representation of the Lie algebra $\mathfrak{g}_r$ in the vector space V. The curvature (1.190) of this connection reads

$$R = -\frac{1}{2} F^p_{\lambda\mu} I_p{}^i_j y^j dq^\alpha \wedge dq^\mu \otimes \partial_i, \tag{2.340}$$

where $F^p_{\lambda\mu}$ are coefficients (2.333) of the strength of a principal connection A.

In particular, any principal connection A induces the associated linear connection on the gauge algebra bundle $V_G P \to Q$. The corresponding covariant differential $\nabla^A \xi$ (1.186) of its sections $\xi = \xi^p e_p$ reads

$$\nabla^A \xi : Q \to T^*Q \otimes V_G P, \qquad \nabla^A \xi = (\partial_\alpha \xi^r + c^r_{pq} A^p_\alpha \xi^q) dq^\alpha \otimes e_r.$$

It coincides with the Nijenhuis differential

$$d_A \xi = [A, \xi]_{FN} = \nabla^A \xi \tag{2.341}$$

of ξ seen as a $V_G P-$valued 0–form, and is given by the local expression given by the local expression

$$\nabla^A \xi = d\xi + [\mathbf{A}, \xi], \tag{2.342}$$

where $\mathbf{A}$ is the local connection form (1.199).

Classical Gauge Fields

Since gauge potentials are represented by global sections of the connection bundle $C \to Q$ (1.201), its 1–jet space $J^1(Q, C)$ plays the role of a configuration space of classical gauge theory. The key point is that the jet space $J^1(Q, C)$ admits the canonical splitting over C which leads to a unique canonical *Yang–Mills Lagrangian density* of gauge theory on $J^1(Q, C)$.

Let us describe this splitting. One can show that the principal $G-$bundle

$$J^1(Q,P) \to J^1(Q,P)/G = C \tag{2.343}$$

is canonically isomorphic to the trivial pull–back bundle

$$P_C = C \times P \to C, \tag{2.344}$$

and that the latter admits the canonical principal connection [Gar77, GMS97]

$$\mathcal{A} = dq^\alpha \otimes (\partial_\alpha + a^p_\alpha e_p) + da^r_\alpha \otimes \partial^\alpha_r \in \mathcal{O}^1(C) \otimes T_G(P_C)(C). \tag{2.345}$$

Since C (1.201) is an affine bundle modelled over the vector bundle

$$\overline{C} = T^*Q \otimes V_G P \to Q,$$

the vertical tangent bundle of C possesses the canonical trivialization

$$\begin{aligned} VC &= C \times T^*Q \otimes V_G P, \qquad \text{where} \\ V_G P_C &= V_G(C \times P) = C \times V_G P. \end{aligned} \tag{2.346}$$

Then the strength $F_\mathcal{A}$ of the connection (2.345) is the $V_G P-$valued horizontal 2–form on C,

$$\begin{aligned} F_\mathcal{A} &= \frac{1}{2} d_\mathcal{A}\mathcal{A} = \frac{1}{2}[\mathcal{A},\mathcal{A}]_{FN} \in \wedge^2(C) \otimes V_G P(Q), \\ F_\mathcal{A} &= (da^r_\mu \wedge dq^\mu + \frac{1}{2} c^r_{pq} a^p_\alpha a^q_\mu dq^\alpha \wedge dq^\mu) \otimes e_r. \end{aligned} \tag{2.347}$$

Note that, given a global section connection A of the connection bundle $C \to Q$, the pull–back $A^*F_\mathcal{A} = F_A$ is the strength (2.332) of the principal connection A.

Let us take the pull–back of the form $F_\mathcal{A}$ onto $J^1(Q,C)$ with respect to the fibration (2.343), and consider the $V_G P-$valued horizontal 2–form

$$\begin{aligned} \mathcal{F} &= h_0(F_\mathcal{A}) = \frac{1}{2}\mathcal{F}^r_{\lambda\mu} dq^\alpha \wedge dq^\mu \otimes e_r, \\ \mathcal{F}^r_{\lambda\mu} &= a^r_{\lambda\mu} - a^r_{\mu\lambda} + c^r_{pq} a^p_\alpha a^q_\mu, \end{aligned} \tag{2.348}$$

where h_0 is the horizontal projection (1.214). Note that

$$\mathcal{F}/2 : J^1(Q,C) \to C \times \wedge^2 T^*Q \otimes V_G P \tag{2.349}$$

is an affine map over C of constant rank. Hence, its kernel $C_+ = \operatorname{Ker}\mathcal{F}$ is the affine subbundle of $J^1(Q,C) \to C$, and we have a desired canonical splitting

$$J^1(Q,C) = C_+ \oplus C_- = C_+ \oplus (C \times \wedge^2 T^*Q \otimes V_G P), \tag{2.350}$$

$$a^r_{\lambda\mu} = \frac{1}{2}(a^r_{\lambda\mu} + a^r_{\mu\lambda} - c^r_{pq} a^p_\alpha a^q_\mu) + \frac{1}{2}(a^r_{\lambda\mu} - a^r_{\mu\lambda} + c^r_{pq} a^p_\alpha a^q_\mu), \tag{2.351}$$

over C of the jet space $J^1(Q,C)$. The corresponding canonical projections are

$$\pi_1 = \mathcal{S} : J^1(Q,C) \to C_+, \qquad \mathcal{S}^r_{\lambda\mu} = \frac{1}{2}(a^r_{\lambda\mu} + a^r_{\mu\lambda} - c^r_{pq} a^p_\alpha a^q_\mu), \tag{2.352}$$

and $\pi_2 = \mathcal{F}/2$ given by (2.349).

Gauge Transformations

In classical gauge theory, several classes of gauge transformations are examined [GMS97, MM92, Soc91]. A most general gauge transformation is defined as an automorphism Φ_P of a principal $G-$bundle P which is equivariant under the canonical action (1.128) of the group G on G, i.e.,

$$R_g \circ \Phi_P = \Phi_P \circ R_g, \qquad (g \in G).$$

Such an automorphism of P induces the corresponding automorphism

$$\Phi_Y : (pG, G^{-1}v) \to (\Phi_P(p)G, G^{-1}v) \tag{2.353}$$

of the $P-$associated bundle Y (2.337) and the corresponding automorphism

$$\Phi_C : J^1(Q,P)/G \to J^1\Phi_P(J^1(Q,P))/G \tag{2.354}$$

of the connection bundle C.

Every vertical automorphism of a principal bundle P is represented as

$$\Phi_P(p) = pf(p), \qquad (p \in P), \tag{2.355}$$

where f is a $G-$valued equivariant function on P, i.e.,

$$f(pg) = g^{-1}f(p)g, \qquad (g \in G). \tag{2.356}$$

There is a 1–1 correspondence between these functions and the global sections s of the group bundle

$$P^G = (P \times G)/G, \tag{2.357}$$

whose typical fibre is the group G, subject to the adjoint representation of the structure group G. Therefore, P^G (2.357) is also called the *adjoint bundle*. There is the canonical fibre–to–fibre action of the group bundle P^G on any $P-$associated bundle Y by the law

$$P^G \times Y \to Y, \qquad ((pG, G^{-1}gG), (pG, G^{-1}v)) \mapsto (pG, G^{-1}gv).$$

Then the above–mentioned correspondence is given by the relation

$$P^G \times P \to P, \qquad (s(\pi_P(p)), p) \mapsto pf(p),$$

where P is a $G-$bundle associated to itself. Hence, the gauge group $\mathfrak{G}(P)$ of vertical automorphisms of a principal $G-$bundle $P \to Q$ is isomorphic to the group of global sections of the $P-$associated group bundle (2.357).

In order to study the gauge invariance of one or another object in gauge theory, it suffices to examine its invariance under an arbitrary 1–parameter subgroup $[\Phi_P]$ of the gauge group. Its infinitesimal generator is a $G-$invariant vertical vector–field ξ on a principal bundle P or, equivalently, a section

$$\xi = \xi^p(x)e_p \tag{2.358}$$

of the gauge algebra bundle $V_G P \to Q$ (1.132). We will call it a *gauge vector–field*. One can think of its components $\xi^p(q)$ as being gauge parameters. Gauge vector–fields (2.358) are transformed under the infinitesimal generators of gauge transformations (i.e., other gauge vector–fields) ξ' by the adjoint representation

$$\mathfrak{L}_{\xi'}\xi = [\xi', \xi] = c^p_{rq}\xi'^r\xi^q e_p, \qquad (\xi, \xi' \in V_G P(Q)).$$

Therefore, gauge parameters are subject to the coadjoint representation

$$\xi' : \xi^p \mapsto - c^p_{rq}\xi'^r\xi^q. \tag{2.359}$$

Given a gauge vector–field ξ (2.358) seen as the infinitesimal generator of a 1–parameter gauge group $[\Phi_P]$, let us get the gauge vector–fields on a $P-$associated bundle Y and the connection bundle C.

The corresponding gauge vector–field on the $P-$associated vector bundle $Y \to Q$ issues from the relation (2.353), and reads

$$\xi_Y = \xi^p I^i_p \partial_i,$$

where I_p are generators of the group G, acting on the typical fibre V of Y.

The gauge vector–field ξ (2.358) acts on elements a (1.202) of the connection bundle by the law

$$\mathfrak{L}_\xi a = [\xi, a]_{FN} = (-\partial_\alpha \xi^r + c^r_{pq}\xi^p a^q_\alpha) dq^\alpha \otimes e_r.$$

In view of the vertical splitting (2.346), this quantity can be regarded as the vertical vector–field

$$\xi_C = (-\partial_\alpha \xi^r + c^r_{pq}\xi^p a^q_\alpha)\partial^\alpha_r \tag{2.360}$$

on the connection bundle C, and is the infinitesimal generator of the 1–parameter group $[\Phi_C]$ of vertical automorphisms (2.354) of C, i.e., a desired gauge vector–field on C.

Lagrangian Gauge Theory

Classical gauge theory of unbroken symmetries on a principal $G-$bundle $P \to Q$ deals with two types of fields. These are gauge potentials identified to global sections of the connection bundle $C \to Q$ (1.201) and matter fields represented by global sections of a $P-$associated vector bundle Y (2.337), called a matter bundle. Therefore, the total configuration space of classical gauge theory is the product of jet bundles

$$J^1(X, Y)_{tot} = J^1(X, Y) \times J^1(Q, C). \tag{2.361}$$

Let us study a gauge–invariant Lagrangian on this configuration space.

A total gauge vector–field on the product $C \times Y$ reads

$$\xi_{YC} = (-\partial_\alpha \xi^r + c^r_{pq}\xi^p a^q_\alpha)\partial^\alpha_r + \xi^p I^i_p \partial_i = (u^{A\lambda}_p \partial_\alpha \xi^p + u^A_p \xi^p)\partial_A, \qquad (2.362)$$

where we use the collective index A, and put the notation

$$u^{A\lambda}_p \partial_A = -\delta^r_p \partial^\alpha_r, \qquad u^A_p \partial_A = c^r_{qp} a^q_\alpha \partial^\alpha_r + I^i_p \partial_i.$$

A Lagrangian L on the configuration space (2.361) is said to be gauge–invariant if its Lie derivative $\mathfrak{L}_{J^1 \xi_{YC}} L$ along any gauge vector–field ξ (2.358) vanishes. Then the *first variational formula* (2.70) leads to the strong equality

$$0 = (u^A_p \xi^p + u^{A\mu}_p \partial_\mu \xi^p)\delta_A \mathcal{L} + d_\alpha[(u^A_p \xi^p + u^{A\mu}_p \partial_\mu \xi^p)\pi^\alpha_A], \qquad (2.363)$$

where $\delta_A \mathcal{L}$ are the variational derivatives (2.223) of L and the total derivative reads

$$d_\alpha = \partial_\alpha + a^p_{\lambda\mu}\partial^\mu_p + y^i_\alpha \partial_i.$$

Due to the arbitrariness of gauge parameters ξ^p, this equality falls into the system of strong equalities

$$u^A_p \delta_A \mathcal{L} + d_\mu(u^A_p \pi^\mu_A) = 0, \qquad (2.364)$$

$$u^{A\mu}_p \delta_A \mathcal{L} + d_\alpha(u^{A\mu}_p \pi^\alpha_A) + u^A_p \pi^\mu_A = 0, \qquad (2.365)$$

$$u^{A\lambda}_p \pi^\mu_A + u^{A\mu}_p \pi^\alpha_A = 0. \qquad (2.366)$$

Substituting (2.365) and (2.366) in (2.364), we get the well–known constraints

$$u^A_p \delta_A \mathcal{L} - d_\mu(u^{A\mu}_p \delta_A \mathcal{L}) = 0$$

for the variational derivatives of a gauge–invariant Lagrangian L.

Treating the equalities (2.364) – (2.366) as the equations for a gauge–invariant Lagrangian, let us solve these equations for a Lagrangian

$$L = \mathcal{L}(t, q^i, a^r_\mu, a^r_{\lambda\mu})\omega : J^1(Q, C) \to \wedge^n T^*Q \qquad (2.367)$$

without matter fields. In this case, the equations (2.364) – (2.366) read

$$c^r_{pq}(a^p_\mu \partial^\mu_r \mathcal{L} + a^p_{\lambda\mu}\partial^{\lambda\mu}_r \mathcal{L}) = 0, \qquad (2.368)$$

$$\partial^\mu_q \mathcal{L} + c^r_{pq} a^p_\alpha \partial^{\mu\lambda}_r \mathcal{L} = 0, \qquad (2.369)$$

$$\partial^{\mu\lambda}_p \mathcal{L} + \partial^{\lambda\mu}_p \mathcal{L} = 0. \qquad (2.370)$$

Let rewrite them relative to the coordinates $(a^q_\mu, \mathcal{S}^r_{\mu\lambda}, \mathcal{F}^r_{\mu\lambda})$ (2.348) and (2.352), associated to the canonical splitting (2.350) of the jet space $J^1(Q, C)$. The equation (2.370) reads

$$\frac{\partial \mathcal{L}}{\partial \mathcal{S}^r_{\mu\lambda}} = 0. \qquad (2.371)$$

Then a simple computation brings the equation (2.369) into the form

$$\partial^\mu_q \mathcal{L} = 0. \tag{2.372}$$

The equations (2.371) and (2.372) shows that the gauge–invariant Lagrangian (2.367) factorizes through the strength $\mathcal{F}$ (2.348) of gauge potentials. As a consequence, the equation (2.368) takes the form

$$c^r_{pq}\mathcal{F}^p_{\lambda\mu}\frac{\partial \mathcal{L}}{\partial \mathcal{F}^r_{\lambda\mu}} = 0.$$

It admits a unique solution in the class of quadratic Lagrangians which is the conventional *Yang–Mills Lagrangian* L_{YM} of gauge potentials on the configuration space $J^1(Q,C)$. In the presence of a background world metric g on the base Q, it reads

$$L_{YM} = \frac{1}{4\varepsilon^2}a^G_{pq}g^{\lambda\mu}g^{\beta\nu}\mathcal{F}^p_{\lambda\beta}\mathcal{F}^q_{\mu\nu}\sqrt{|g|}\omega, \qquad (\text{where } \ g=\det(g_{\mu\nu})), \tag{2.373}$$

where a^G is a $G-$invariant bilinear form on the Lie algebra of $\mathfrak{g}_r$ and ε is a coupling constant.

Hamiltonian Gauge Theory

Let us consider gauge theory of principal connections on a principal bundle $P \longrightarrow X$ with a structure Lie group G. Principal connections on $P \longrightarrow X$ are represented by sections of the affine bundle

$$C = J^1(Q,P)/G \longrightarrow X, \tag{2.374}$$

modelled over the vector bundle $T^*X \otimes V_GP$ [GMS97]. Here, $V_GP = VP/G$ is the fibre bundle in Lie algebras $\mathfrak{g}$ of the group G. Given the basis $\{\varepsilon_r\}$ for $\mathfrak{g}$, we get the local fibre bases $\{e_r\}$ for V_GP. The connection bundle C (2.374) is coordinated by (x^μ, a^r_μ) such that, written relative to these coordinates, sections $A = A^r_\mu dx^\mu \otimes e_r$ of $C \longrightarrow X$ are the familiar local connection 1–forms, regarded as *gauge potentials*.

There is 1–1 correspondence between the sections $\xi = \xi^r e_r$ of $V_GP \longrightarrow X$ and the vector–fields on P which are infinitesimal generators of 1–parameter groups of vertical automorphisms (gauge transformations) of P. Any section ξ of $V_GP \longrightarrow X$ induces the vector–field on C, given by

$$u(\xi) = u^k_\mu \frac{\partial}{\partial a^r_\mu} = (c^r_{pq}a^p_\mu\xi^q + \partial_\mu\xi^r)\frac{\partial}{\partial a^r_\mu}, \tag{2.375}$$

where c^r_{pq} are the structure constants of the Lie algebra $\mathfrak{g}$.

The configuration space of gauge theory is the jet space $J^1(Q,C)$ equipped with the coordinates $(x^\alpha, a^m_\alpha, a^m_{\mu\lambda})$. It admits the canonical splitting (2.299) given by the coordinate expression

$$a^r_{\mu\lambda} = \mathcal{S}^r_{\mu\lambda} + \mathcal{F}^r_{\mu\lambda} = \frac{1}{2}(a^r_{\mu\lambda} + a^r_{\lambda\mu} - c^r_{pq}a^p_\mu a^q_\alpha) + \frac{1}{2}(a^r_{\mu\lambda} - a^r_{\lambda\mu} + c^r_{pq}a^p_\mu a^q_\alpha),$$

where $\mathcal{F}$ is the strength of gauge fields up to the factor $1/2$. The *Yang–Mills Lagrangian* L_{YM} on the configuration space $J^1(Q,C)$ is given by

$$L_{YM} = a^G_{pq} g^{\lambda\mu} g^{\beta\nu} \mathcal{F}^p_{\lambda\beta} \mathcal{F}^q_{\mu\nu} \sqrt{|g|}\omega, \qquad (g = \det(g_{\mu\nu})), \tag{2.376}$$

where a^G is a non–degenerate $G-$invariant metric in the dual of the Lie algebra of $\mathfrak{g}$ and g is a pseudo–Riemannian metric on X.

The phase–space Π (2.261) of the gauge theory is with the canonical coordinates $(x^\alpha, a^p_\alpha, p^{\mu\lambda}_q)$. It admits the canonical splitting (2.300) given by the coordinate expression

$$p^{\mu\lambda}_m = \mathcal{R}^{\mu\lambda}_m + \mathcal{P}^{\mu\lambda}_m = p^{(\mu\lambda)}_m + p^{[\mu\lambda]}_m = \frac{1}{2}(p^{\mu\lambda}_m + p^{\lambda\mu}_m) + \frac{1}{2}(p^{\mu\lambda}_m - p^{\lambda\mu}_m). \tag{2.377}$$

With respect to this splitting, the Legendre map induced by the Lagrangian (2.376) takes the form

$$p^{(\mu\lambda)}_m \circ \widehat{L}_{YM} = 0, \tag{2.378}$$

$$p^{[\mu\lambda]}_m \circ \widehat{L}_{YM} = 4a^G_{mn} g^{\mu\alpha} g^{\lambda\beta} \mathcal{F}^n_{\alpha\beta} \sqrt{|g|}. \tag{2.379}$$

The equalities (2.378) define the Lagrangian constraint space N_L of Hamiltonian gauge theory. Obviously, it is an imbedded submanifold of Π, and the Lagrangian L_{YM} is almost–regular.

In order to construct an associated Hamiltonian, let us consider a connection Γ (2.295) on the fibre bundle $C \to X$ which take their values into $\mathrm{Ker}\,\widehat{L}$, i.e.,

$$\Gamma^r_{\lambda\mu} - \Gamma^r_{\mu\lambda} + c^r_{pq}a^p_\lambda a^q_\mu = 0.$$

Given a symmetric linear connection K on X and a principal connection B on $P \to X$, this connection reads

$$\Gamma^r_{\lambda\mu} = \frac{1}{2}[\partial_\mu B^r_\alpha + \partial_\alpha B^r_\mu - c^r_{pq}a^p_\alpha a^q_\mu + c^r_{pq}(a^p_\alpha B^q_\mu + a^p_\mu B^q_\alpha)] - K_\alpha{}^\beta{}_\mu(a^r_\beta - B^r_\beta).$$

The corresponding Hamiltonian (2.303) associated to L_{YM} is

$$\mathcal{H}_\Gamma = p^{\lambda\mu}_r \Gamma^r_{\lambda\mu} + a^{mn}_G g_{\mu\nu} g_{\lambda\beta} p^{[\mu\lambda]}_m p^{[\nu\beta]}_n \sqrt{|g|}.$$

Then we get the Lagrangian

$$\mathcal{L}_N = p^{[\lambda\mu]}_r \mathcal{F}^r_{\lambda\mu} - a^{mn}_G g_{\mu\nu} g_{\lambda\beta} p^{[\mu\lambda]}_m p^{[\nu\beta]}_n \sqrt{|g|}$$

(2.306) on the Lagrangian constraint manifold (2.378) and its pull–back

$$L_\Pi = \mathcal{L}_\Pi \omega, \qquad \mathcal{L}_\Pi = p^{\lambda\mu}_r \mathcal{F}^r_{\lambda\mu} - a^{mn}_G g_{\mu\nu} g_{\lambda\beta} p^{[\mu\lambda]}_m p^{[\nu\beta]}_n \sqrt{|g|}, \tag{2.380}$$

(2.309) onto Π.

Both the Lagrangian L_{YM} (2.376) on C and the Lagrangian L_Π (2.380) on Π are invariant under gauge transformations whose infinitesimal generators are the lifts

$$j^1u(\xi) = (c^r_{pq}a^p_\mu\xi^q + \partial_\mu\xi^r)\frac{\partial}{\partial a^r_\mu} + (c^r_{pq}(a^p_{\lambda\mu}\xi^q + a^p_\mu\partial_\alpha\xi^q) + \partial_\alpha\partial_\mu\xi^r)\frac{\partial}{\partial a^r_{\lambda\mu}},$$
$$\overline{u}(\xi) = j^1u(\xi) - c^r_{pq}p^{\lambda\mu}_r\xi^q\frac{\partial}{\partial p^{\lambda\mu}_p},$$

of the vector–fields (2.375) onto $J^1(Q,C)$ and $\Pi\times J^1(Q,C)$, respectively. We have the transformation laws

$$j^1u(\xi)(\mathcal{F}^r_{\lambda\mu}) = c^r_{pq}\mathcal{F}^p_{\lambda\mu}\xi^q, \qquad j^1u(\xi)(\mathcal{S}^r_{\lambda\mu}) = c^r_{pq}\mathcal{S}^p_{\lambda\mu}\xi^q + c^r_{pq}a^p_\mu\partial_\alpha\xi^q + \partial_\alpha\partial_\mu\xi^r.$$

Therefore, one can choose the gauge conditions

$$g^{\lambda\mu}S^r_{\lambda\mu}(x) - \alpha^r(x) = \frac{1}{2}g^{\lambda\mu}(\partial_\alpha a^r_\mu(x) + \partial_\mu a^r_\alpha(x)) - \alpha^r(x) = 0,$$

which are the familiar generalized Lorentz gauge. The corresponding second-order differential operator (2.325) reads

$$M^r_s\xi^s = g^{\lambda\mu}(\frac{1}{2}c^r_{pq}(\partial_\alpha a^r_\mu + \partial_\mu a^r_\alpha)\xi^q + c^r_{pq}a^p_\mu\partial_\alpha\xi^q + \partial_\alpha\partial_\mu\xi^r).$$

Passing to the Euclidean space and repeating the above quantization procedure, we come to the generating functional

$$\begin{aligned} Z = \mathcal{N}^{-1}\int\exp\{\int(&p^{\alpha\mu}_r\mathcal{F}^r_{\alpha\mu} - a^{mn}_G g_{\mu\nu}g_{\alpha\beta}p^{\mu\alpha}_m p^{\nu\beta}_n\sqrt{|g|} \\ &- \frac{1}{8}a^G_{rs}g^{\alpha\nu}g^{\alpha\mu}(\partial_\alpha a^r_\nu + \partial_\nu a^r_\alpha)(\partial_\alpha a^s_\mu + \partial_\mu a^s_\alpha) \\ &- g^{\alpha\mu}\overline{c}_r(\frac{1}{2}c^r_{pq}(\partial_\alpha a^r_\mu + \partial_\mu a^r_\alpha)c^q + c^r_{pq}a^p_\mu c^q_\alpha + c^r_{\alpha\mu}) \\ &+ \mathrm{i}J^\mu_r a^r_\mu + \mathrm{i}J^r_{\mu\alpha}p^{\mu\alpha}_r)\omega\}\prod_x[d\overline{c}][dc][dp(x)][da(x)]. \end{aligned}$$

Its integration with respect to momenta restarts the familiar generating functional of gauge theory.

Gauge Conservation Laws

On–shell, the strong equality (2.363) becomes the *weak Noether conservation law*

$$d_\alpha[(u^A_p\xi^p + u^{A\mu}_p\partial_\mu\xi^p)\pi^\alpha_A] \approx 0 \tag{2.381}$$

of the *Noether current*

$$\mathfrak{J}^\alpha = -(u_p^A \xi^p + u_p^{A\mu} \partial_\mu \xi^p)\pi_A^\alpha. \tag{2.382}$$

Therefore, the equalities (2.364) – (2.366) on–shell lead to the familiar *Noether identities*

$$d_\mu(u_p^A \pi_A^\mu) \approx 0, \tag{2.383}$$

$$d_\alpha(u_p^{A\mu} \pi_A^\alpha) + u_p^A \pi_A^\mu \approx 0, \tag{2.384}$$

$$u_p^{A\alpha} \pi_A^\mu + u_p^{A\mu} \pi_A^\alpha = 0 \tag{2.385}$$

for a gauge–invariant Lagrangian L. They are equivalent to the weak equality (2.381) due to the arbitrariness of the gauge parameters $\xi^p(q)$.

The expressions (2.381) and (2.382) shows that both the Noether conservation law and the Noether current depend on gauge parameters. The weak identities (2.383) – (2.385) play the role of the necessary and sufficient conditions in order that the Noether conservation law (2.381) is maintained under changes of gauge parameters. This means that, if the equality (2.381) holds for gauge parameters ξ, it does so for arbitrary deviations $\xi + \delta\xi$ of ξ. In particular, the Noether conservation law (2.381) is maintained under gauge transformations, when gauge parameters are transformed by the coadjoint representation (2.359).

It can be seen that the equalities (2.383) – (2.385) are not mutually independent, but (2.383) is a corollary of (2.384) and (2.385). This property reflects the fact that, in accordance with the strong equalities (2.365) and (2.366), the Noether current (2.382) is brought into the superpotential form

$$\mathfrak{J}^\alpha = \xi^p u_p^{A\alpha} \delta_A \mathcal{L} - d_\mu(\xi^p u_p^{A\mu} \pi_A^\alpha), \qquad U^{\mu\alpha} = -\xi^p u_p^{A\mu} \pi_A^\alpha,$$

(2.222). Since a matter field Lagrangian is independent of the jet coordinates $a_{\alpha\mu}^p$, the *Noether superpotential*, $U^{\mu\alpha} = \xi^p \pi_p^{\mu\alpha}$, depends only on gauge potentials. The corresponding integral relation (2.224) reads

$$\int_{N^{n-1}} s^* \mathfrak{J}^\alpha \omega_\alpha = \int_{\partial N^{n-1}} s^*(\xi^p \pi_p^{\mu\alpha}) \omega_{\mu\alpha}, \tag{2.386}$$

where N^{n-1} is a compact oriented $(n-1)$D submanifold of Q with the boundary ∂N^{n-1}. One can think of (2.386) as being the integral relation between the Noether current (2.382) and the gauge field, generated by this current. In electromagnetic theory seen as a $U(1)$ gauge theory, the similar relation between an electric current and the electromagnetic field generated by this current is well known. However, it is free from gauge parameters due to the peculiarity of Abelian gauge models.

Note that the Noether current (2.382) in gauge theory takes the superpotential form (2.222) because the infinitesimal generators of gauge transformations (2.362) depend on derivatives of gauge parameters.

Topological Gauge Theories

The field models that we have investigated so far show that when a background world metric is present, the stress–energy–momentum (SEM) transformation law becomes the covariant conservation law of the *metric SEM–tensor* (see subsection on SEM–tensors 2.2.6 below). Topological gauge theories exemplify the field models in the absence of a world metric [GMS05].

Let us consider the *Chern–Simons gauge theory* on a 3D base manifold X^3 [BBR91, Wit89]. Physical interpretation of this fundamental model will be given in subsection 2.2.8 below.

Let $P \longrightarrow X^3$ be a principal bundle with a structure semisimple Lie group G and C the corresponding bundle of principal connections which is coordinated by (x^α, k^m_α). The *Chern–Simons Lagrangian density* is given by the coordinate expression

$$L_{CS} = \frac{1}{2k} a^G_{mn} \varepsilon^{\alpha\lambda\mu} k^m_\alpha (\mathcal{F}^n_{\lambda\mu} + \frac{1}{3} c^n_{pq} k^p_\alpha k^q_\mu) d^3x \tag{2.387}$$

where $\varepsilon^{\alpha\lambda\mu}$ is the skew–symmetric Levi–Civita tensor.

Note that the Lagrangian density (2.387) is not gauge–invariant and globally defined. At the same time, it gives the globally defined Euler–Lagrangian operator

$$\mathcal{E}_{L_{CS}} = \frac{1}{k} a^G_{mn} \varepsilon^{\alpha\lambda\mu} \mathcal{F}^n_{\lambda\mu} dk^m_\alpha \omega.$$

Thus, the gauge transformations in the Chern–Simons model appear to be the generalized invariant transformations which keep invariant the Euler–Lagrangian equations, but not the Lagrangian density. Solutions of these equations are the curvature–free principal connections A on the principal bundle $P \longrightarrow X^3$.

Though the Chern–Simons Lagrangian density is not invariant under gauge transformations, we still have the Noether–type conservation law (2.381) in which the total conserved current is the standard Noether current (2.382) plus the additional term as follows.

Let $u_{\mathfrak{g}}$ be the principal vector–field (2.360) on the bundle of principal connections C. We calculate the Lie derivative

$$\mathfrak{L}_{\overline{u}_{\mathfrak{g}}} L_{CS} = \frac{1}{k} a^G_{mn} \varepsilon^{\alpha\lambda\mu} \partial_\alpha (\alpha^m \partial_\alpha A^n_\mu) d^3x.$$

Hence, the Noether transformation law (2.381) becomes the conservation law

$$d_\alpha {T_{CS}}^\alpha = d_\alpha (T^\alpha + \frac{1}{k} a^G_{mn} \varepsilon^{\alpha\lambda\mu} \alpha^m \partial_\alpha A^n_\mu) \approx 0, \qquad \text{where} \tag{2.388}$$

$$T^\alpha = \pi^{\mu\lambda}_n {u_{\mathfrak{g}}}_\mu^n = \frac{1}{k} a^G_{mn} \varepsilon^{\alpha\lambda\mu} A^m_\alpha {u_{\mathfrak{g}}}_\mu^n$$

is the standard *Noether current*. After simplification, the conservation law (2.388) takes the form

$$d_\alpha(\frac{1}{k}a^G_{mn}\varepsilon^{\alpha\lambda\mu}\alpha^m F^n_{\alpha\mu}) \approx 0.$$

In the Chern–Simons model, the total conserved current T_{CS} is equal to zero. At the same time, if we add the *Chern–Simons Lagrangian density* to the Yang–Mills one, T_{CS} plays the role of the massive term and makes the contribution into the standard Noether current of the Yang–Mills gauge theory.

Let τ be a vector–field on the base X and τ_B its lift onto the bundle C by means of a section B of C. Remind that the vector–fields τ_B are the general principal vector–fields associated with local 1–parameter groups of general gauge isomorphisms of C. We calculate

$$\mathfrak{L}_{\overline{\tau}_B} L_{CS} = \frac{1}{k}a^G_{mn}\varepsilon^{\alpha\lambda\mu}\partial_\alpha(\tau^\nu B^m_\nu \partial_\alpha A^n_\mu)\, d^3x.$$

The corresponding SEM transformation law takes the form

$$d_\alpha \mathcal{J}_{CS}{}^\alpha = d_\alpha(\mathcal{J}^\alpha - \frac{1}{2k}a^G_{mn}\varepsilon^{\alpha\lambda\mu}\tau^\nu B^m_\nu \partial_\alpha A^n_\mu) \approx 0, \qquad \text{where} \tag{2.389}$$
$$\mathcal{J}^\alpha = \pi^{\mu\lambda}_n[\tau^\nu \partial_\nu A^m_\mu - \tau^\nu(\partial_\mu B^n_\nu - c^n_{pq}A^p_\mu B^q_\nu) - \partial_\mu \tau^\nu (B^n_\nu - A^n_\nu)] - \delta^\alpha_\nu \tau^\nu \mathcal{L}_{CS}$$

is the standard SEM–tensor relative to the lift τ_B of the vector–field τ.

Let A be a critical section. We consider the lift of the vector–field τ on X onto C by means of the principal connection $B = A$. Then, the energy–momentum conservation law (2.389) becomes the conservation law

$$d_\alpha[\frac{1}{k}a^G_{mn}\tau^\nu \varepsilon^{\alpha\lambda\mu}A^m_\alpha \mathcal{F}^n_{\nu\mu} - \tau^\alpha \mathcal{L}_{CS}] = d_\alpha(-\frac{1}{6k}\tau^\alpha \varepsilon^{\alpha\nu\mu}c_{npq}A^n_\alpha A^p_\nu A^q_\mu) \approx 0. \tag{2.390}$$

Note that, since the gauge symmetry of the Chern–Simons Lagrangian density is broken, the energy–momentum conservation law (2.390) fails to be invariant under gauge transformations.

Let us consider Lagrangian densities of topological gauge models which are invariant under the general gauge isomorphisms of the bundle C. Though they imply the zero Euler–Lagrangian operators, the corresponding strong identities may be used as the superpotential terms when such a topological Lagrangian density is added to the Yang–Mills one.

Let $P \longrightarrow X$ be a principal bundle with the structure Lie group G. Let us consider the bundle $J^1(Q,P) \longrightarrow C$. This also is a G principal bundle. Due to the canonical vertical splitting $VP = P \times \mathfrak{g}_l$, where $\mathfrak{g}_l$ is the left Lie algebra of the group G, the complementary map (1.162) of $J^1(Q,P)$ defines the canonical $\mathcal{G}$-valued one–form θ on $J^1(Q,P)$. This form is the connection form of the canonical principal connection on the principal bundle $J^1(Q,P) \longrightarrow C$ [Gar72]. Moreover, if $\Gamma_A : P \longrightarrow J^1(Q,P)$ is a principal connection on P and A the corresponding connection form, we have $\Gamma^*_A\theta = A$. If Ω and R_A are the curvature 2–forms of the connections θ and A respectively, then $\Gamma^*_A\Omega = R_A$.

Local connection 1–forms on C associated with the canonical connection θ are given by the coordinate expressions $k^m_\mu dx^\mu \otimes I_m$. The corresponding curvature two–form on C reads

$$\Omega_C = (dk^m_\mu \wedge dx^\mu - \frac{1}{2} c^m_{nl} k^n_\mu k^l_\nu dx^\nu \wedge dx^\mu) \otimes I_m.$$

Let $I(\mathfrak{g})$ be the algebra of real $G-$invariant polynomial on the Lie algebra $\mathfrak{g}$ of the group G. Then, there is the well–known *Weyl homomorphism* of $I(\mathfrak{g})$ into the *de Rham cohomology algebra* $H^*(C, \mathbb{R})$. Using this isomorphism, every k-linear element $r \in I(\mathfrak{g})$ is represented by the cohomology class of the closed characteristic $2k-$form $r(\Omega_C)$ on C. If A is a section of C, we have $A^* r(\Omega_C) = r(F)$, where F is the strength of A and $r(F)$ is the corresponding characteristic form on X.

Let $\dim X$ be even and a characteristic $n-$form $r(\Omega_C)$ on C exist. This is a Lepagian form which defines a gauge–invariant Lagrangian density

$$L_r = h_0(r(\Omega_C))$$

on the jet space $J^1(Q, C)$. The Euler–Lagrangian operator associated with L_r is equal to zero. Then, for any projectable vector–field u on C, we have the strong relation (2.241):

$$\mathfrak{L}_{\overline{u}} h_0(r(\Omega_C)) = h_0(d\overline{u} \rfloor r(\Omega_C)).$$

If u is a general principal vector–field on C, this relation takes the form

$$d_H(\overline{u} \rfloor r(\Omega_C)) = 0.$$

For example, let $\dim X = 4$ and the group G be semisimple. Then, the characteristic Chern–Pontryagin 4–form

$$r(\Omega_C) = a^G_{mn} \Omega^n_C \wedge \Omega^m_C.$$

It is the Lepagian equivalent of the *Chern–Pontryagin Lagrangian density*

$$L = \frac{1}{k} a^G_{mn} \varepsilon^{\alpha\beta\mu\nu} \mathcal{F}^n_{\alpha\beta} \mathcal{F}^m_{\mu\nu} d^4x$$

of the topological Yang–Mills theory.

2.2.6 Modern Geometrodynamics

In this subsection we present some modern developments of the classical Einstein–Wheeler geometrodynamics that we briefly reviewed as a motivation to our geometrical machinery.

Stress–Energy–Momentum Tensors

While in analytical mechanics there exists the conventional differential energy conservation law, in field theory it does not exist (see [Sar98]). Let F be a smooth manifold. In time–dependent mechanics on the phase–space $\mathbb{R}\times T^*F$ coordinated by $(t, y^i, \dot{y}_i)$ and on the configuration space $\mathbb{R}\times TF$ coordinated by $(t, y^i, \dot{y}^i)$, the Lagrangian energy and the construction of the Hamiltonian formalism require the prior choice of a connection on the bundle $\mathbb{R}\times F\to\mathbb{R}$. However, such a connection is usually hidden by using the natural trivial connection on this bundle. Therefore, given a Hamiltonian function $\mathcal{H}$ on the phase–space manifold $\mathbb{R}\times T^*F$, we have the usual *energy conservation law*

$$\frac{d\mathcal{H}}{dt}\approx\frac{\partial\mathcal{H}}{\partial t} \tag{2.391}$$

where by '$\approx$' is meant the weak identity modulo the Hamiltonian equations. Given a Lagrangian function $\mathcal{L}$ on the configuration manifold $\mathbb{R}\times TF$, there exists the fundamental identity

$$\frac{\partial\mathcal{L}}{\partial t}+\frac{d}{dt}(\dot{y}^i(t)\frac{\partial\mathcal{L}}{\partial\dot{y}^i}-\mathcal{L})\approx 0 \tag{2.392}$$

modulo the equations of motion. It is the *energy conservation law* in the following sense. Let $\widehat{L}$ be the Legendre morphism given by $\dot{y}_i\circ\widehat{L}=\partial_{\dot{y}^i}\mathcal{L}$, and $Q=\operatorname{Im}\widehat{L}$ the Lagrangian constraint manifold. Let $\mathcal{H}$ be a Hamiltonian function associated with $\mathcal{L}$ and $\widehat{H}$ the momentum morphism, $\dot{y}^i\circ\widehat{H}=\partial_{\dot{y}_i}\mathcal{H}$. Every solution r of the Hamiltonian equations of $\mathcal{H}$ which lives on Q yields the solution $\widehat{H}\circ r$ of the Euler–Lagrangian equations of $\mathcal{L}$. Then, the identity (2.392) on $\widehat{H}\circ r$ recovers the energy transformation law (2.391) on r.[5]

Recall that in field theory, classical fields are described by sections of a fibre bundle $Y\to X$, while their dynamics is phrased in terms of jet spaces. We restrict ourselves to the *first–order Lagrangian formalism* when the configuration space is $J^1(X,Y)$. Given fibred coordinates (x^μ, y^i) of Y, the jet space $J^1(X,Y)$ is equipped with the adapted coordinates (x^μ, y^i, y^i_μ). Recall that the first–order Lagrangian density on $J^1(X,Y)$ is defined to be the morphism

$$\begin{aligned}&L: J^1(X,Y)\to\wedge^n T^*X, \qquad (n=\dim X),\\ &L=\mathcal{L}(x^\mu,y^i,y^i_\mu)\omega, \qquad \text{with} \qquad \omega=dx^1\wedge\ldots\wedge dx^n,\end{aligned}$$

while the corresponding first–order Euler–Lagrangian equations for sections $\overline{s}$ of the jet bundle $J^1(X,Y)\to X$ read

[5] There are different Hamiltonian functions associated with the same singular Lagrangian function as a rule. Given such a Hamiltonian function, the Lagrangian constraint space Q plays the role of the primary constraint space, and the Dirac procedure can be used in order to get the final constraint space where a solution of the Hamiltonian equations exists [CLM94, GNH78].

$$\partial_\alpha \overline{s}^i = \overline{s}^i_\alpha, \qquad \partial_i \mathcal{L} - (\partial_\alpha + \overline{s}^j_\alpha \partial_j + \partial_\alpha \overline{s}^j_\mu \partial^\mu_j)\partial^\alpha_i \mathcal{L} = 0. \tag{2.393}$$

As before, we consider the *Lie derivatives of Lagrangian densities* in order to get *differential conservation laws*. Let $u = u^\mu(x)\partial_\mu + u^i(y)\partial_i$ be a projectable vector–field on $Y \to X$ and $\overline{u}$ its jet lift (1.165) onto $J^1(X,Y) \to X$. Given L, let us calculate the Lie derivative $\mathfrak{L}_{\overline{u}} L$. We get the identity

$$\overline{s}^* \mathfrak{L}_{\overline{u}} L \approx -\frac{d}{dx^\alpha}[\pi^\alpha_i (u^\mu \overline{s}^i_\mu - u^i) - u^\alpha \mathcal{L}]\omega, \qquad \text{with} \qquad \pi^\mu_i = \partial^\mu_i \mathcal{L}, \tag{2.394}$$

modulo the Euler–Lagrangian equations (2.393). In particular, if u is a vertical vector–field this identity becomes the current conservation law exemplified by the Noether identities in gauge theory [Sar94].

Let now $\tau = \tau^\alpha \partial_\alpha$ be a vector–field on X and

$$\tau_\Gamma = \tau^\mu(\partial_\mu + \Gamma^i_\mu \partial_i) \tag{2.395}$$

its horizontal lift onto Y by a connection Γ on $Y \to X$. In this case, the identity (2.394) takes the form

$$\overline{s}^* \mathfrak{L}_{\overline{\tau}_\Gamma} L \approx -\frac{d}{dx^\alpha}[\tau^\mu \mathcal{J}_\Gamma{}^\alpha{}_\mu(\overline{s})]\omega, \tag{2.396}$$

$$\text{where} \qquad \mathcal{J}_\Gamma{}^\alpha{}_\mu(\overline{s}) = [\pi^\alpha_i (y^i_\mu - \Gamma^i_\mu) - \delta^\alpha_\mu \mathcal{L}] \circ \overline{s}^i_\mu$$

is the stress–energy–momentum (SEM) tensor on a field $\overline{s}$ relative to the connection Γ. We here restrict ourselves to this particular case of SEM–tensors [KT79].

For example, let us choose the trivial local connection $\Gamma^i_\mu = 0$. In this case, the identity (2.396) recovers the well–known conservation law

$$\frac{\partial \mathcal{L}}{\partial x^\alpha} + \frac{d}{dx^\alpha}\mathcal{J}^\alpha{}_\mu(\overline{s}) \approx 0$$

of the canonical energy–momentum tensor

$$\mathcal{J}^\alpha{}_\mu(\overline{s}) = \pi^\alpha_i \overline{s}^i_\mu - \delta^\alpha_\mu \mathcal{L}. \tag{2.397}$$

Physicists often lose sight of the fact that (2.397) fails to be a mathematical well–behaved object. The crucial point lies in the fact that the Lie derivative

$$\mathfrak{L}_{\overline{\tau}_\Gamma} L = \{\partial_\mu \tau^\mu \mathcal{L} + [\tau^\mu \partial_\mu + \tau^\mu \Gamma^i_\mu \partial_i + (\partial_\alpha(\tau^\mu \Gamma^i_\mu) + \tau^\mu y^j_\alpha \partial_j \Gamma^i_\mu - y^i_\mu \partial_\alpha \tau^\mu)\partial^\alpha_i]\mathcal{L}\}\omega$$

is almost never equal to zero. Therefore, it is not obvious how to choose the true energy–momentum tensor.

The canonical energy–momentum tensor (2.397) in gauge theory is symmetrized by hand in order to get the gauge–invariant one. In gauge theory in the presence of a background world metric g, the identity (2.396) is brought into the covariant conservation law for the *metric SEM–tensor*,

$$\nabla_\alpha t^\alpha{}_\mu \approx 0. \tag{2.398}$$

In Einstein's general relativity, the covariant conservation law (2.398) issues directly from the gravitation equations. It is concerned with the zero-spin matter in the presence of the gravitational field generated by this matter, though the matter is not required to satisfy the motion equations. The total energy–momentum conservation law for matter and gravity is introduced by hand. It is usually written as

$$\frac{d}{dx^\mu}[(-g)^N(t^{\lambda\mu} + t_g{}^{\lambda\mu})] \approx 0, \tag{2.399}$$

where the energy–momentum pseudotensor $t_g{}^{\lambda\mu}$ of a metric gravitational field is defined to satisfy the relation

$$(-g)^N(t^{\lambda\mu} + t_g{}^{\lambda\mu}) \approx \frac{1}{2\kappa}\partial_\sigma\partial_\alpha[(-g)^N(g^{\lambda\mu}g^{\sigma\alpha} - g^{\sigma\mu}g^{\lambda\alpha}]$$

modulo the Einstein equations. However, the conservation law (2.399) appears satisfactory only in cases of asymptotic–flat gravitational fields and of a background metric.

Moreover, the covariant conservation law (2.398) fails to take place in the affine–metric gravitation theory and in the gauge gravitation theory, e.g., in the presence of fermion fields.

Thus, we have not any conventional energy–momentum conservation law in Lagrangian field theory. In particular, one may take different SEM–tensors for different field models and, moreover, different SEM–tensors for different solutions of the same field equations. Just the latter in fact is the above-mentioned symmetrization of the canonical energy–momentum tensor in gauge theory.

Gauge theory exemplifies constraint field theories. Contemporary field models are almost always the constraint ones. To describe them, let us turn to the Hamiltonian formalism.

When applied to field theory, the conventional Hamiltonian formalism takes the form of the instantaneous Hamiltonian formalism where canonical variables are field functions at a given instant of time. The corresponding phase–space is infinite–dimensional, so that the Hamiltonian equations in the bracket form fail to be differential equations.

The true partners of the Lagrangian formalism in classical field theory are polysymplectic and multisymplectic Hamiltonian machineries where canonical momenta correspond to derivatives of fields with respect to all world coordinates, not only the temporal one [CCI91, Sar93]. We here follow the multimomentum Hamiltonian formulation of field theory when the phase–space of fields is the Legendre bundle over Y

$$\Pi = \wedge^n T^*X \otimes TX \otimes V^*Y, \tag{2.400}$$

which is coordinated by $(x^\alpha, y^i, p_i^\alpha)$ [Sar93, Sar94]. Every Lagrangian density L on $J^1(X,Y)$ implies the Legendre morphism

$$\widehat{L} : J^1(X,Y) \to \Pi, \qquad p_i^\mu \circ \widehat{L} = \pi_i^\mu.$$

The Legendre bundle (2.400) carries the polysymplectic form

$$\Omega = dp_i^\alpha \wedge dy^i \wedge \omega \otimes \partial_\alpha. \tag{2.401}$$

Recall that one says that a connection γ on the fibred Legendre manifold $\Pi \to X$ is a *Hamiltonian connection* if the form $\gamma \rfloor \Omega$ is closed. Then, a Hamiltonian form H on Π is defined to be an exterior form such that

$$dH = \gamma \rfloor \Omega \tag{2.402}$$

for some Hamiltonian connection γ. The key point lies in the fact that every Hamiltonian form admits the following splitting

$$H = p_i^\alpha dy^i \wedge \omega_\alpha - p_i^\alpha \Gamma_\alpha^i \omega - \widetilde{\mathcal{H}}_\Gamma \omega = p_i^\alpha dy^i \wedge \omega_\alpha - \mathcal{H}\omega, \qquad \omega_\alpha = \partial_\alpha \rfloor \omega, \tag{2.403}$$

where Γ is a connection on $Y \to X$.

Given the splitting (2.403), the equality (2.402) becomes the Hamiltonian equations

$$\partial_\alpha r^i = \partial_\alpha^i \mathcal{H}, \qquad \partial_\alpha r_i^\alpha = -\partial_i \mathcal{H} \tag{2.404}$$

for sections r of $\Pi \to X$.

The Hamiltonian equations (2.404) are the multimomentum generalization of the standard Hamiltonian equations in mechanics. The corresponding multimomentum generalization of the conventional energy conservation law (2.391) is the weak identity

$$\tau^\mu[(\partial_\mu + \Gamma_\mu^i \partial_i - \partial_i \Gamma_\mu^j r_j^\alpha \partial_\lambda^i)\widetilde{\mathcal{H}}_\Gamma - \frac{d}{dx^\alpha} T_\Gamma{}^\alpha{}_\mu(r)] \approx \tau^\mu r_i^\alpha R_{\lambda\mu}^i, \tag{2.405}$$

$$T_\Gamma{}^\alpha{}_\mu(r) = [r_i^\alpha \partial_\mu^i \widetilde{\mathcal{H}}_\Gamma - \delta_\mu^\alpha (r_i^\alpha \partial_\alpha^i \widetilde{\mathcal{H}}_\Gamma - \widetilde{\mathcal{H}}_\Gamma)], \tag{2.406}$$

$$\text{where} \qquad R_{\lambda\mu}^i = \partial_\alpha \Gamma_\mu^i - \partial_\mu \Gamma_\alpha^i + \Gamma_\alpha^j \partial_j \Gamma_\mu^i - \Gamma_\mu^j \partial_j \Gamma_\alpha^i$$

is the curvature of the connection Γ. One can think of the tensor (2.406) as being the Hamiltonian SEM–tensor.

If a Lagrangian density is regular, the multimomentum Hamiltonian formalism is equivalent to the Lagrangian formalism, otherwise in case of degenerate Lagrangian densities. In field theory, if a Lagrangian density is not regular, the Euler–Lagrangian equations become underdetermined and require supplementary gauge–type conditions. In gauge theory, they are the familiar gauge conditions. However, in general case, the gauge–type conditions remain elusive. In the framework of the multimomentum Hamiltonian formalism, they appear automatically as a part of the Hamiltonian equations. The key point consists in the fact that, given a degenerate Lagrangian density, one must consider a family of different associated Hamiltonian forms in order to exhaust all solutions of the Euler–Lagrangian equations [CCI91, Sar93].

Lagrangian densities of all realistic field models are quadratic or affine in the velocity coordinates y^i_μ. Complete family of Hamiltonian forms associated with such a Lagrangian density always exists [Sar93, Sar94]. Moreover, these Hamiltonian forms differ from each other only in connections Γ in the splitting (2.403). Different connections are responsible for different gauge–type conditions mentioned above. They are also the connections which one should use in construction of the Hamiltonian SEM–tensors (2.406).

The identity (2.405) remains true in the first–order Lagrangian theories of gravity. In this work, we examine the metric-affine gravity where independent dynamical variables are world metrics and general linear connections. The energy–momentum conservation law in the affine–metric gravitation theory is not widely discussed. We construct the Hamiltonian SEM–tensor for gravity. In case of the affine Hilbert–Einstein Lagrangian density, it is equal to

$$T^\alpha_\mu = \frac{1}{2\kappa}\delta^\alpha_\mu R\sqrt{-g}$$

and the total conservation law (2.405) for matter and gravity is reduced to the conservation law for matter in the presence of a background world metric, otherwise in case of quadratic Lagrangian densities.

Lagrangian SEM–Tensors

Given a Lagrangian density L, the jet space $J^1(X,Y)$ carries the associated Poincaré–Cartan form [Sar98]

$$\Xi_L = \pi^\alpha_i dy^i \wedge \omega_\alpha - \pi^\alpha_i y^i_\alpha \omega + \mathcal{L}\omega \tag{2.407}$$

and the Lagrangian polysymplectic form

$$\Omega_L = (\partial_j \pi^\alpha_i dy^j + \partial^\mu_j \pi^\alpha_i dy^j_\mu) \wedge dy^i \wedge \omega \otimes \partial_\alpha.$$

Using the pull–back of these forms onto the repeated jet space $J^1J^1(X,Y)$, one can construct the exterior generating form on $J^1J^1(X,Y)$,

$$\begin{aligned} \Lambda_L &= d\Xi_L - \lambda\rfloor\Omega_L = [y^i_{(\lambda)} - y^i_\alpha)d\pi^\alpha_i + (\partial_i - \widehat{\partial}_\alpha\partial^\alpha_i)\mathcal{L}dy^i] \wedge \omega, \quad (2.408)\\ \lambda &= dx^\alpha \otimes \widehat{\partial}_\alpha, \qquad \widehat{\partial}_\alpha = \partial_\alpha + y^i_{(\lambda)}\partial_i + y^i_{\mu\lambda}\partial^\mu_i. \end{aligned}$$

Its restriction to the sesquiholonomic jet space $\widehat{J}^2(X,Y)$ defines the first–order Euler–Lagrangian operator

$$\begin{aligned} &\mathcal{E}'_L : \widehat{J}^2(X,Y) \longrightarrow \wedge^{n+1}T^*Y, \qquad \text{given by}\\ &\mathcal{E}'_L = \delta_i\mathcal{L}dy^i \wedge \omega = [\partial_i - (\partial_\alpha + y^i_\alpha\partial_i + y^i_{\mu\lambda}\partial^\mu_i)\partial^\alpha_i]\mathcal{L}dy^i \wedge \omega, \quad (2.409) \end{aligned}$$

corresponding to L. The restriction of the form (2.408) to the second–order jet space $J^2(X,Y)$ of Y recovers the familiar variational Euler–Lagrangian operator

$$\mathcal{E}_L : J^2(X,Y) \longrightarrow \wedge^{n+1}T^*Y,$$

given by the expression (2.409), but with symmetric coordinates $y^i_{\mu\lambda} = y^i_{\lambda\mu}$.

Let $\overline{s}$ be a section of the jet bundle $J^1(X,Y) \longrightarrow X$ such that its jet prolongation $j^1\overline{s}$ takes its values into $\mathrm{Ker}\,\mathcal{E}'_L$ given by the coordinate relations

$$\partial_i\mathcal{L} - (\partial_\alpha + y^j_\alpha\partial_j + y^j_{\mu\lambda}\partial^\mu_j)\partial^\alpha_i\mathcal{L} = 0.$$

Then, $\overline{s}$ satisfies the first–order Euler–Lagrangian equations (2.393). These equations are equivalent to the second–order Euler–Lagrangian equations

$$\partial_i\mathcal{L} - (\partial_\alpha + \partial_\alpha s^j\partial_j + \partial_\alpha\partial_\mu s^j\partial^\mu_j)\partial^\alpha_i\mathcal{L} = 0. \tag{2.410}$$

for sections s of $Y \longrightarrow X$ where $\overline{s} = j^1 s$.

We have the following differential conservation laws on solutions of the first–order Euler–Lagrangian equations.

Given a Lagrangian density L on $J^1(X,Y)$, let us consider its pull–back onto $\widehat{J}^2(X,Y)$. Let u be a projectable vector–field on $Y \longrightarrow X$ and $\overline{u}$ its jet lift (1.165) onto $J^1(X,Y) \longrightarrow X$. Its pull–back onto $J^1J^1(X,Y)$ has the canonical horizontal splitting (1.170) given by the expression

$$\overline{u} = \overline{u}_H + \overline{u}_V = u^\alpha(\partial_\alpha + y^i_{(\lambda)}\partial_i + y^i_{\mu\lambda}\partial^\mu_i) + [(u^i - y^i_{(\lambda)}u^\alpha)\partial_i + (u^i_\mu - y^i_{\mu\lambda}u^\alpha)\partial^\mu_i].$$

Let us calculate the Lie derivative $\mathfrak{L}_{\overline{u}}L$. We have

$$\begin{aligned} \mathfrak{L}_{\overline{u}}L &= \widehat{\partial}_\alpha[\pi^\lambda_i(u^i - u^\mu y^i_\mu) + u^\alpha\mathcal{L}]\omega + \overline{u}_V\rfloor\mathcal{E}'_L, \\ \widehat{\partial}_\alpha &= \partial_\alpha + y^i_\alpha\partial_i + y^i_{\mu\lambda}\partial^\mu_i. \end{aligned} \tag{2.411}$$

Being restricted to $\mathrm{Ker}\,\mathcal{E}'_L$, the equality (2.411) is written

$$\begin{aligned} &\partial_\alpha u^\alpha\mathcal{L} + [u^\alpha\partial_\alpha + u^i\partial_i + (\partial_\alpha u^i + y^j_\alpha\partial_j u^i - y^i_\mu\partial_\alpha u^\mu)\partial^\alpha_i]\mathcal{L} \\ &\approx \widehat{\partial}_\alpha[\pi^\alpha_i(u^i - u^\mu y^i_\mu) + u^\alpha\mathcal{L}]. \end{aligned} \tag{2.412}$$

On solutions $\overline{s}$ of the first–order Euler–Lagrangian equations, the weak identity (2.412) becomes the differential conservation law

$$\overline{s}^*\mathfrak{L}_{\overline{u}}L \approx d(\overline{u}\rfloor\Xi_L \circ \overline{s}),$$

which takes the coordinate form (2.394).

In particular, let τ_Γ be the horizontal lift (2.395) of a vector–field τ on X onto $Y \to X$ by a connection Γ on Y. In this case, the identity (2.412) is written

$$\tau^\mu\{[\partial_\mu + \Gamma^i_\mu\partial_i + (\partial_\alpha\Gamma^i_\mu + y^j_\alpha\partial_j\Gamma^i_\mu)\partial^\alpha_i]\mathcal{L} + \widehat{\partial}_\alpha[\pi^\alpha_i(y^i_\mu - \Gamma^i_\mu) - \delta^\alpha_\mu\mathcal{L}] \approx 0. \tag{2.413}$$

On solutions $\overline{s}$ of the first–order Euler–Lagrangian equations, the identity (2.413) becomes the differential conservation law (2.396) where $\mathcal{J}_\Gamma{}^\alpha{}_\mu(\overline{s})$ are coefficients of the T^*X-valued form on X,

$$\mathcal{J}_\Gamma(\overline{s}) = -(\Gamma \rfloor \Xi_L) \circ \overline{s} = [\pi_i^\alpha(\overline{s}_\mu^i - \Gamma_\mu^i) - \delta_\mu^\alpha \mathcal{L}]\, dx^\mu \otimes \omega_\alpha. \tag{2.414}$$

This conservation law takes the coordinate form

$$\tau^\mu\{[\partial_\mu + \Gamma_\mu^i\partial_i + (\partial_\alpha\Gamma_\mu^i + \overline{s}_\alpha^j\partial_j\Gamma_\mu^i)\partial_i^\alpha]\mathcal{L} + \frac{d}{dx^\alpha}[\pi_i^\alpha(\overline{s}_\mu^i - \Gamma_\mu^i) - \delta_\mu^\alpha\mathcal{L}] \approx 0.$$

SEM Conservation Laws

Every projectable vector–field u on the bundle $Y \longrightarrow X$ which covers a vector–field τ on the base X is represented as the sum of a vertical vector–field on $Y \longrightarrow X$ and some lift of τ onto Y. Hence, any differential transformation law (2.239) can be represented as a superposition of some transformation law associated with a vertical vector–field on the bundle $Y \longrightarrow X$ and the one induced by the lift of a vector–field on the base X onto Y. Therefore, we can reduce our consideration to transformation laws associated with these two types of vector–fields on Y.

Vertical vector–fields result in transformation and conservation laws of Noether currents. In general case, a vector–field τ on a base X induces a vector–field on Y only by means of some connection on the bundle $Y \longrightarrow X$. Such lifts result in the transformation laws of the SEM–tensors.

Given a bundle $Y \to X$, let τ be a vector–field on X and

$$\tau_\Gamma = \tau \rfloor \Gamma = \tau^\mu(\partial_\mu + \Gamma_\mu^i\partial_i)$$

its horizontal lift onto $Y \to X$ by means of a connection on Y, given by

$$\Gamma = dx^\mu \otimes (\partial_\mu + \Gamma_\mu^i\partial_i).$$

In this case, the weak identity (2.412) is written

$$\begin{aligned}&\partial_\mu\tau^\mu\mathcal{L} + [\tau^\mu\partial_\mu + \tau^\mu\Gamma_\mu^i\partial_i + (\partial_\alpha(\tau^\mu\Gamma_\mu^i) + \tau^\mu y_\alpha^j\partial_j\Gamma_\mu^i - y_\mu^i\partial_\alpha\tau^\mu)\partial_i^\alpha]\mathcal{L}\\ &- \widehat{\partial}_\alpha[\pi_i^\alpha(\tau^\mu\Gamma_\mu^i - \tau^\mu y_\mu^i) + \delta_\mu^\alpha\tau^\mu\mathcal{L}] \approx 0.\end{aligned} \tag{2.415}$$

One can simplify it as follows:

$$\tau^\mu\{[\partial_\mu + \Gamma_\mu^i\partial_i + (\partial_\alpha\Gamma_\mu^i + y_\alpha^j\partial_j\Gamma_\mu^i)\partial_i^\alpha]\mathcal{L} - \widehat{\partial}_\alpha[\pi_i^\alpha(\Gamma_\mu^i - y_\mu^i) + \delta_\mu^\alpha\mathcal{L}] \approx 0.$$

Let us emphasize that this relation takes place for arbitrary vector–field τ on X. Therefore, it is equivalent to the system of the weak identities

$$[\partial_\mu + \Gamma_\mu^i\partial_i + (\partial_\alpha\Gamma_\mu^i + y_\alpha^j\partial_j\Gamma_\mu^i)\partial_i^\alpha]\mathcal{L} - \widehat{\partial}_\alpha[\pi_i^\alpha(\Gamma_\mu^i - y_\mu^i) + \delta_\mu^\alpha\mathcal{L}] \approx 0. \tag{2.416}$$

On solutions s of the Euler–Lagrangian equations, the weak identity (2.415) becomes the weak transformation law

$$s^*\mathfrak{L}_{\overline{\tau}_\Gamma}L + \frac{d}{dx^\alpha}[\tau^\mu\mathcal{J}_\Gamma{}^\alpha{}_\mu(s)]\omega \approx 0$$

and to the equivalent system of the weak transformation laws

$$[\partial_\mu + \Gamma^i_\mu \partial_i + (\partial_\alpha \Gamma^i_\mu + \partial_\alpha s^j \partial_j \Gamma^i_\mu)\partial^\alpha_i]\mathcal{L} + \frac{d}{dx^\alpha}[\pi^\alpha_i(\partial_\mu s^i - \Gamma^i_\mu) - \delta^\alpha_\mu \mathcal{L}] \approx 0 \quad (2.417)$$

where $\mathcal{J}_\Gamma{}^\alpha{}_\mu(s)$ is the SEM–tensor given by the components of the T^*X-valued $(n-1)-$form on X,

$$\mathcal{J}_\Gamma(s) = -(\Gamma \rfloor \Xi_L) \circ s = [\pi^\alpha_i(\partial_\mu s^i - \Gamma^i_\mu) - \delta^\alpha_\mu \mathcal{L}] dx^\mu \otimes \omega_\alpha .$$

It is clear that the first and the second terms in (2.417) taken separately fail to be well–behaved objects. Therefore, only their combination may result in the satisfactory transformation or conservation law.

For example, let a Lagrangian density L depend on a background metric g on the base X. In this case, we have

$$\partial_\mu \mathcal{L} = -t^\alpha_\beta \sqrt{|g|}\, \Gamma^\beta_{\mu\alpha}, \qquad \text{where} \qquad t^\alpha_\beta = g^{\alpha\gamma} t_{\gamma\beta}$$

is the *metric SEM–tensor* (by definition), while $\Gamma^\beta_{\mu\alpha}$ are the *Christoffel symbols* of the metric g. Then, the weak transformation law (2.417) takes the form

$$\begin{aligned} -t^\alpha_\beta \sqrt{|g|}\Gamma^\beta_{\mu\alpha} &+ [\Gamma^i_\mu \partial_i + (\partial_\alpha \Gamma^i_\mu + \partial_\alpha s^j \partial_j \Gamma^i_\mu)\partial^\alpha_i]\mathcal{L} \\ &+ \frac{d}{dx^\alpha}[\pi^\alpha_i(\partial_\mu s^i - \Gamma^i_\mu) - \delta^\alpha_\mu \mathcal{L}] \;\approx\; 0, \end{aligned}$$

and, under suitable conditions of symmetries of the Lagrangian density L, it may become the covariant conservation law $\nabla_\alpha t^\alpha_\beta = 0$ where ∇_α denotes the covariant derivative relative to the connection $\Gamma^\beta_{\mu\alpha}$.

Note that, if we consider another Lepagian equivalent of the Lagrangian density L, the SEM transformation law takes the form

$$\begin{aligned} & s^* \mathfrak{L}_{\overline{\tau}_\Gamma} L + \frac{d}{dx^\alpha}[\tau^\mu \mathcal{J}'_\Gamma{}^\alpha{}_\mu(s)]\, \omega \approx 0, \\ & \text{where} \qquad \mathcal{J}'_\Gamma{}^\alpha{}_\mu = \mathcal{J}_\Gamma{}^\alpha{}_\mu - \frac{d}{dx^\nu}[(\partial_\mu s^i - \Gamma^i_\mu) c^{\lambda\nu}_i], \end{aligned}$$

that is, the SEM–tensors $\mathcal{J}'_\Gamma{}^\alpha{}_\mu$ and $\mathcal{J}_\Gamma{}^\alpha{}_\mu$ differ from each other in the superpotential–type term: $-\frac{d}{dx^\nu}[(\partial_\mu s^i - \Gamma^i_\mu) c^{\lambda\nu}_i]$.

In particular, if the bundle Y has a fibre metric a^Y_{ij}, one can choose

$$c^{\mu\nu}_i = a^Y_{ij} g^{\mu\alpha} g^{\nu\beta} R^j_{\alpha\beta}\,,$$

where R is the curvature of the connection Γ on the bundle Y and g is a metric on X. In this case, the superpotential contribution into the SEM–tensor is equal to $-\frac{d}{dx^\nu}[a^Y_{ij} g^{\lambda\alpha} g^{\nu\beta}(\partial_\mu s^i - \Gamma^i_\mu) R^j_{\alpha\beta}]$.

Let us now consider the weak identity (2.415) when a vector–field τ on the base X induces a vector–field on Y by means of different connections Γ and Γ' on $Y \longrightarrow X$. Their difference result in the weak identity

$$[\tau^\mu\sigma^i_\mu\partial_i + (\partial_\alpha(\tau^\mu\sigma^i_\mu) + y^j_\alpha\partial_j(\tau^\mu\sigma^i_\mu))\partial^\alpha_i]\mathcal{L} - \widehat{\partial}_\alpha[\pi^\alpha_i\tau^\mu\sigma^i_\mu] \approx 0 \qquad (2.418)$$

where $\sigma = \Gamma' - \Gamma$ is a soldering form on the bundle $Y \longrightarrow X$ and

$$\tau\rfloor\sigma = \tau^\mu\sigma^i_\mu\partial_i \qquad (2.419)$$

is a vertical vector–field. It is clear that the identity (2.418) is exactly the weak identity (2.412) in case of the vertival vector–field (2.419).

It follows that every SEM transformation law contains a Noether transformation law. Conversely, every Noether transformation law associated with a vertical vector–field u_V on $Y \longrightarrow X$ can be get as the difference of two SEM transformation laws if the vector–field u_V takes the form $u_V = \tau\rfloor\sigma$, where σ is some soldering form on Y and τ is a vector–field on X. In field theory, this representation fails to be unique. On the contrary, in Newtonian mechanics there is the 1–1 correspondence between the vertical vector–fields and the soldering forms on the bundle $\mathbb{R} \times F \longrightarrow F$.

Note that one can consider the pull–back of the first–order Lagrangian density L and their Lepagian equivalents onto the infinite order jet space $J^\infty Y$. In this case, there exists the canonical lift τ^∞_H (1.215) of a vector–field τ on X onto $J^\infty Y$. One can treat this lift as the horizontal lift of τ by means of the canonical connection on the bundle $J^\infty Y \longrightarrow X$, given by

$$\Gamma_\infty = dx^\mu\otimes(\partial_\mu + y^i\partial_i + y^i_\alpha\partial^\alpha_i + \cdots).$$

Multimomentum Hamiltonian Formalism

Let Π be the Legendre bundle (2.400) coordinated by $(x^\alpha, y^i, p^\alpha_i)$. By $J^1(X,\Pi)$ is meant the first–order jet space of $\Pi \longrightarrow X$. It is coordinated by $(x^\alpha, y^i, p^\alpha_i, y^i_{(\mu)}, p^\alpha_{i\mu})$. The Legendre manifold Π carries the generalized Liouville form

$$\theta = -p^\alpha_i dy^i \wedge \omega \otimes \partial_\alpha$$

and the polysymplectic form Ω (2.401).

The Hamiltonian formalism in fibred manifolds is formulated intrinsically in terms of Hamiltonian connections which play the role similar to that of Hamiltonian vector–fields in the symplectic geometry [Sar93].

We say that a jet field (resp. a connection)

$$\gamma = dx^\alpha \otimes (\partial_\alpha + \gamma^i_{(\lambda)}\partial_i + \gamma^\mu_{i\lambda}\partial^i_\mu)$$

on the Legendre manifold $\Pi \longrightarrow X$ is a Hamiltonian jet field (resp. a Hamiltonian connection) if the following exterior form is closed:

$$\gamma\rfloor\Omega = dp^\alpha_i \wedge dy^i \wedge \omega_\alpha + \gamma^\alpha_{i\lambda}dy^i \wedge \omega - \gamma^i_{(\lambda)}dp^\alpha_i \wedge \omega.$$

An exterior $n-$form H on the Legendre manifold Π is called a *Hamiltonian form* if, on an open neighborhood of each point of Π, there exists a *Hamiltonian jet–field* satisfying the equation $\gamma\rfloor\Omega = dH$, i.e., if there exists a

Hamiltonian connection satisfying the equation (2.402). Hamiltonian connections constitute an affine subspace of connections on $\Pi \to X$. The following construction shows that this subspace is not empty.

Every connection Γ on $Y \to X$ is lifted to the connection

$$\gamma = \widetilde{\Gamma} = dx^\alpha \otimes [\partial_\alpha + \Gamma^i_\alpha(y)\partial_i + (-\partial_j \Gamma^i_\alpha(y) p^\mu_i - K^\mu{}_{\nu\lambda}(x) p^\nu_j + K^\alpha{}_{\alpha\lambda}(x) p^\mu_j)\partial^j_\mu]$$

on $\Pi \to X$, where

$$K = dx^\alpha \otimes (\partial_\alpha + K^\mu{}_{\nu\lambda}\dot{x}_\mu \frac{\partial}{\partial \dot{x}_\nu})$$

is a linear symmetric connection on T^*X. We have the equality

$$\widetilde{\Gamma} \rfloor \Omega = d(\Gamma \rfloor \theta).$$

This equality shows that $\widetilde{\Gamma}$ is a Hamiltonian connection and

$$H_\Gamma = \Gamma \rfloor \theta = p^\alpha_i dy^i \wedge \omega_\alpha - p^\alpha_i \Gamma^i_\alpha \omega$$

is a Hamiltonian form.

Let H be a Hamiltonian form. For any exterior horizontal density $\widetilde{H} = \widetilde{\mathcal{H}}\omega$ on $\Pi \longrightarrow X$, the form $H + \widetilde{H}$ is a Hamiltonian form. Conversely, if H and H' are Hamiltonian forms, their difference $H - H'$ is an exterior horizontal density on $\Pi \longrightarrow X$.

Thus, Hamiltonian forms constitute an affine space modelled on a linear space of the exterior horizontal densities on $\Pi \longrightarrow X$. It follows that every Hamiltonian form on Π can be given by the expression (2.403) where Γ is some connection on $Y \longrightarrow X$. Moreover, a Hamiltonian form has the canonical splitting (2.403) as follows.

Every Hamiltonian form H implies the momentum map

$$\widehat{H} : \Pi \longrightarrow J^1(X, Y), \qquad y^i_\alpha \circ \widehat{H} = \partial^i_\alpha \mathcal{H},$$

and the associated connection $\Gamma_H = \widehat{H} \circ \widehat{0}$ on Y where $\widehat{0}$ is the global zero section of $\Pi \to Y$. As a consequence, we have the canonical splitting

$$H = H_{\Gamma_H} - \widetilde{H}.$$

The Hamiltonian operator $\mathcal{E}_H$ of a Hamiltonian form H is defined to be the first–order differential operator on $\Pi \to X$,

$$\begin{aligned} &\mathcal{E}_H \ : \ j^1\Pi \to \wedge^{n+1} T^*\Pi, \\ &\mathcal{E}_H = dH - \widehat{\Omega} = [(y^i_{(\lambda)} - \partial^i_\alpha \mathcal{H}) dp^\alpha_i - (p^\alpha_{i\lambda} + \partial_i \mathcal{H}) dy^i] \wedge \omega \quad (2.420) \\ &\text{where} \quad \widehat{\Omega} = dp^\alpha_i \wedge dy^i \wedge \omega_\alpha + p^\alpha_{i\lambda} dy^i \wedge \omega - y^i_{(\lambda)} dp^\alpha_i \wedge \omega \end{aligned}$$

is the pull–back of the multisymplectic form (2.401) onto $j^1\Pi$.

For any connection γ on $\Pi \to X$, we have

$$\mathcal{E}_H \circ \gamma = dH - \gamma \rfloor \Omega.$$

It follows that γ is a Hamiltonian connection for a Hamiltonian form H iff it takes its values into $\operatorname{Ker} \mathcal{E}_H$ given by the coordinate relations

$$y^i_{(\lambda)} = \partial^i_\alpha \mathcal{H}, \qquad p^\alpha_{i\lambda} = -\partial_i \mathcal{H}. \tag{2.421}$$

Let a Hamiltonian connection γ has an integral section r of $\Pi \longrightarrow X$, that is, $\gamma \circ r = j^1 r$. Then, the algebraic equations (2.421) are brought into the first–order differential Hamiltonian equations (2.404).

Now we consider relations between Lagrangian and Hamiltonian formalisms. A Hamiltonian form H is defined to be associated with a Lagrangian density L if it satisfies the relations

$$\widehat{L} \circ \widehat{H}|_Q = Id_Q, \qquad Q = \widehat{L}(J^1(X,Y)),$$
$$H = H_{\widehat{H}} + L \circ \widehat{H},$$

which take the coordinate form

$$\partial^\mu_i \mathcal{L}(x^\alpha, y^j, \partial^j_\alpha \mathcal{H}) = p^\mu_i, \qquad \mathcal{L}(x^\alpha, y^j, \partial^j_\alpha \mathcal{H}) = p^\mu_i \partial^i_\mu \mathcal{H} - \mathcal{H}.$$

Note that there are different Hamiltonian forms associated with the same singular Lagrangian density.

Bearing in mind physical application, we restrict our consideration to so–called *semiregular Lagrangian densities* L when the preimage $\widehat{L}^{-1}(q)$ of each point of $q \in Q$ is the connected submanifold of $J^1(X,Y)$. In this case, all Hamiltonian forms associated with a semiregular Lagrangian density L coincide on the Lagrangian constraint space Q, and the Poincaré–Cartan form Ξ_L is the pull–back

$$\Xi_L = H \circ \widehat{L}, \qquad \pi^\alpha_i y^i_\alpha - \mathcal{L} = \mathcal{H}(x^\mu, y^i, \pi^\alpha_i),$$

of any associated multimomentum Hamiltonian form H by the Legendre morphism $\widehat{L}$ [Zak92]. Also the generating form (2.408) is the pull–back of

$$\Lambda_L = \mathcal{E}_H \circ J^1 \widehat{L}$$

of the Hamiltonian operator (2.420) of any Hamiltonian form H associated with a semiregular Lagrangian density L. As a consequence, we get the following correspondence between solutions of the Euler–Lagrangian equations and the Hamiltonian equations [Sar94, Zak92].

Let a section r of $\Pi \longrightarrow X$ be a solution of the Hamiltonian equations (2.404) for a Hamiltonian form H associated with a semiregular Lagrangian density L. If r lives on the Lagrangian constraint space Q, the section $\overline{s} = \widehat{H} \circ r$ of $J^1(X,Y) \longrightarrow X$ satisfies the first–order Euler–Lagrangian equations (2.393). Conversely, given a semiregular Lagrangian density L, let $\overline{s}$ be a solution of

the first–order Euler–Lagrangian equations (2.393). Let H be a Hamiltonian form associated with L so that

$$\widehat{H} \circ \widehat{L} \circ \overline{s} = \overline{s}. \tag{2.422}$$

Then, the section $r = \widehat{L} \circ \overline{s}$ of $\Pi \longrightarrow X$ is a solution of the Hamiltonian equations (2.404) for H. For sections $\overline{s}$ and r, we have the relations

$$\overline{s} = j^1 s, \qquad \text{and} \qquad s = \pi_{\Pi Y} \circ r,$$

where s is a solution of the second–order Euler–Lagrangian equations (2.410).

We shall say that a family of Hamiltonian forms H associated with a semiregular Lagrangian density L is complete if, for each solution $\overline{s}$ of the first–order Euler–Lagrangian equations (2.393), there exists a solution r of the Hamiltonian equations (2.404) for some Hamiltonian form H from this family so that

$$r = \widehat{L} \circ \overline{s}, \qquad \overline{s} = \widehat{H} \circ r, \qquad \overline{s} = J^1(\pi_{\Pi Y} \circ r). \tag{2.423}$$

Such a complete family exists iff, for each solution $\overline{s}$ of the Euler–Lagrangian equations for L, there exists a Hamiltonian form H from this family so that the condition (2.422) holds.

We do not discuss here existence of solutions of Euler–Lagrangian and Hamiltonian equations. Note that, in contrast with mechanics, there are different Hamiltonian connections associated with the same multimomentum Hamiltonian form in general. Moreover, in field theory when the primary constraint space is the Lagrangian constraint space Q, there is a family of Hamiltonian forms associated with the same Lagrangian density as a rule. In practice, one can choose either the Hamiltonian equations or solutions of the Hamiltonian equations such that these solutions live on the constraint space.

Hamiltonian SEM–Tensors

Let H be a Hamiltonian form on the Legendre bundle Π over a fibre bundle $Y \longrightarrow X$. We have the following differential conservation law on solutions of the Hamiltonian equations [Sar98].

Let r be a section of the fibred Legendre manifold $\Pi \longrightarrow X$. Given a connection Γ on $Y \longrightarrow X$, we consider the T^*X-valued $(n-1)-$form

$$\begin{aligned} &T_\Gamma(r) = -(\Gamma \rfloor H) \circ r, \\ &T_\Gamma(r) = [r_i^\alpha(\partial_\mu r^i - \Gamma_\mu^i) - \delta_\mu^\alpha(r_i^\alpha(\partial_\alpha r^i - \Gamma_\alpha^i) - \widetilde{\mathcal{H}}_\Gamma)]dx^\mu \otimes \omega_\alpha, \end{aligned} \tag{2.424}$$

on X where $\widetilde{\mathcal{H}}_\Gamma$ is the Hamiltonian density in the splitting (2.403) of H with respect to the connection Γ.

Let $\tau = \tau^\alpha \partial_\alpha$ be a vector–field on X. Given a connection Γ on $Y \to X$, it induces the projectable vector–field

$$\widetilde{\tau}_\Gamma = \tau^\alpha\partial_\alpha + \tau^\alpha\Gamma^i_\alpha\partial_i + (-\tau^\mu p^\alpha_j\partial_i\Gamma^j_\mu - p^\alpha_i\partial_\mu\tau^\mu + p^\mu_i\partial_\mu\tau^\alpha)\partial^i_\alpha$$

on the Legendre bundle Π. Let us calculate the Lie derivative $\mathfrak{L}_{\widetilde{\tau}_\Gamma}\widetilde{H}_\Gamma$ on a section r. We have

$$\begin{aligned}(\mathfrak{L}_{\widetilde{\tau}_\Gamma}\widetilde{H}_\Gamma)\circ r &= \{\partial_\alpha\tau^\alpha\widetilde{\mathcal{H}}_\Gamma + [\tau^\alpha\partial_\alpha \\ &+ \tau^\alpha\Gamma^i_\alpha\partial_i + (-\tau^\mu r^\alpha_j\partial_i\Gamma^j_\mu - r^\alpha_i\partial_\mu\tau^\mu + r^\mu_i\partial_\mu\tau^\alpha)\partial^i_\alpha]\widetilde{\mathcal{H}}_\Gamma\}\omega \\ &= \tau^\mu r^\alpha_i R^i_{\lambda\mu}\omega + d(\tau^\mu T_\Gamma{}^\alpha{}_\mu(r)\omega_\alpha) - (\widetilde{\tau}_{\Gamma V}\rfloor\mathcal{E}_H)\circ r, \qquad (2.425)\end{aligned}$$

where $\widetilde{\tau}_{\Gamma V}$ is the vertical part of the canonical horizontal splitting (1.170) of the vector–field $\widetilde{\tau}_V$ on Π over $j^1\Pi$. If r is a solution of the Hamiltonian equations, the equality (2.425) becomes the conservation law (2.405). The form (2.424) modulo the Hamiltonian equations reads

$$T_\Gamma(r) \approx [r^\alpha_i(\partial^i_\mu\mathcal{H} - \Gamma^i_\mu) - \delta^\alpha_\mu(r^\alpha_i\partial^i_\alpha\mathcal{H} - \mathcal{H})]dx^\mu\otimes\omega_\alpha. \qquad (2.426)$$

For example, if $X = \mathbb{R}$ and Γ is the trivial connection, we have $T_\Gamma(r) = \mathcal{H}dt$, where $\mathcal{H}$ is a Hamiltonian function. Then, the identity (2.405) becomes the conventional energy conservation law (2.391) in mechanics.

Unless $n = 1$, the identity (2.405) cannot be regarded directly as the energy–momentum conservation law. To clarify its physical meaning, we turn to the Lagrangian formalism.

Let a Hamiltonian form H be associated with a semiregular Lagrangian density L. Let r be a solution of the Hamiltonian equations of H which lives on the Lagrangian constraint space Q and $\overline{s}$ the associated solution of the first–order Euler–Lagrangian equations of L so that they satisfy the conditions (2.423). Then, we have

$$T_\Gamma(r) = \mathcal{J}_\Gamma(\widetilde{H}\circ r), \qquad T_\Gamma(\widetilde{L}\circ\overline{s}) = \mathcal{J}_\Gamma(\overline{s}),$$

where $\mathcal{J}_\Gamma$ is the SEM–tensor (2.414).

It follows that, on the Lagrangian constraint space Q, the form (2.426) can be treated the Hamiltonian SEM–tensor relative to the connection Γ on $Y\to X$.

At the same time, the examples below show that, in several field models, the equality (2.405) is brought into the covariant conservation law (2.398) for the metric SEM–tensor.

In the Lagrangian formalism, the metric SEM–tensor is defined to be

$$\sqrt{-g}t_{\alpha\beta} = 2\frac{\partial\mathcal{L}}{\partial g^{\alpha\beta}}.$$

In case of a background world metric g, this object is well–behaved. In the framework of the multimomentum Hamiltonian formalism, one can introduce the similar tensor

$$\sqrt{-g}t_H{}^{\alpha\beta} = 2\frac{\partial\mathcal{H}}{\partial g_{\alpha\beta}}. \qquad (2.427)$$

If a Hamiltonian form H is associated with a semiregular Lagrangian density L, there are the equalities

$$t_H{}^{\alpha\beta}(q) = -g^{\alpha\mu}g^{\beta\nu}t_{\mu\nu}(x^\alpha, y^i, \partial^i_\alpha \mathcal{H}(q)), \qquad (q \in Q),$$
$$t_H{}^{\alpha\beta}(x^\alpha, y^i, \pi^\alpha_i(z)) = -g^{\alpha\mu}g^{\beta\nu}t_{\mu\nu}(z), \qquad \widehat{H} \circ \widehat{L}(z) = z.$$

In view of these equalities, we can think of the tensor (2.427) restricted to the Lagrangian constraint space Q as being the Hamiltonian metric SEM–tensor. On Q, the tensor (2.427) does not depend upon choice of a Hamiltonian form H associated with L. Therefore, we shall denote it by the common symbol t. Set

$$t^\lambda{}_\alpha = g_{\alpha\nu}t^{\lambda\nu}.$$

In the presence of a background world metric g, the identity (2.405) takes the form

$$t^\lambda{}_\alpha\{^\alpha{}_{\lambda\mu}\}\sqrt{-g} + (\Gamma^i_\mu\partial_i - \partial_i\Gamma^j_\mu r^\alpha_j\partial^i_\alpha)\widetilde{\mathcal{H}}_\Gamma \approx \frac{d}{dx^\alpha}T_\Gamma{}^\alpha{}_\mu + r^\alpha_i R^i_{\alpha\mu}\,, \qquad (2.428)$$

where by $\{^\alpha{}_{\lambda\mu}\}$ are meant the Christoffel symbols of the world metric g.

SEM Tensors in Gauge Theory

In this subsection, following [Sar98] we consider the gauge theory of principal connections treated as gauge potentials. Here, the manifold X is assumed to be oriented and provided with a nondegenerate fibre metric $g_{\mu\nu}$ in the tangent bundle of X. We denote $g = \det(g_{\mu\nu})$.

Let $P \to X$ be a principal bundle with a structure Lie group G which acts freely and transitively on P on the right: $r_g : p \mapsto pg, \quad (p \in P,\, g \in G)$. A principal connection A on $P \to X$ is defined to be a G-equivariant connection on P such that $j^1 r_g \circ A = A \circ r_g$ for each canonical morphism r_g. Recall that there is the 1–1 correspondence between the principal connections on a principal bundle $P \to X$ and the global sections of the quotient bundle

$$C = J^1(X, P)/G \to X. \qquad (2.429)$$

The bundle (2.429) is the affine bundle modelled on the vector bundle $\overline{C} = T^*X \otimes (VP/G)$. Given a bundle atlas Ψ^P of P, the bundle C has the fibred coordinates (x^μ, k^m_μ) so that $(k^m_\mu \circ A)(x) = A^m_\mu(x)$ are coefficients of the local connection 1–form of a principal connection A with respect to the atlas Ψ^P. The 1–jet space $J^1(X, C)$ of the fibre bundle $C \longrightarrow X$ is coordinated by $(x^\mu, k^m_\mu, k^m_{\mu\lambda})$.

There exists the canonical splitting over C, given by

$$J^1(X, C) = C_+ \oplus C_- = (J^2P/G) \oplus (\wedge^2 T^*X \otimes V^G P), \qquad (2.430)$$
$$k^m_{\mu\lambda} = \frac{1}{2}(k^m_{\mu\lambda} + k^m_{\lambda\mu} + c^m_{nl}k^n_\alpha k^l_\mu) + \frac{1}{2}(k^m_{\mu\lambda} - k^m_{\lambda\mu} - c^m_{nl}k^n_\alpha k^l_\mu).$$

The corresponding surjections read:

$$\mathcal{S} : J^1(X,C) \to C_+, \qquad \mathcal{S}^m_{\lambda\mu} = k^m_{\mu\lambda} + k^m_{\lambda\mu} + c^m_{nl}k^n_\alpha k^l_\mu,$$
$$\mathcal{F} : J^1(X,C) \to C_-, \qquad \mathcal{F}^m_{\lambda\mu} = k^m_{\mu\lambda} - k^m_{\lambda\mu} - c^m_{nl}k^n_\alpha k^l_\mu.$$

On the configuration space (2.430), the conventional Yang–Mills Lagrangian density L_{YM} of gauge potentials in the presence of a background world metric is given by the expression

$$L_{YM} = \frac{1}{4\varepsilon^2} a^G_{mn} g^{\lambda\mu} g^{\beta\nu} \mathcal{F}^m_{\lambda\beta} \mathcal{F}^n_{\mu\nu} \sqrt{|g|}\omega, \tag{2.431}$$

where a^G is a nondegenerate $G-$invariant metric in the Lie algebra $\mathfrak{g}$ of G. The Legendre morphism associated with the Lagrangian density (2.431) takes the form

$$p^{(\mu\lambda)}_m \circ \widehat{L}_{YM} = 0, \tag{2.432}$$
$$p^{[\mu\lambda]}_m \circ \widehat{L}_{YM} = \varepsilon^{-2} a^G_{mn} g^{\lambda\alpha} g^{\mu\beta} \mathcal{F}^n_{\alpha\beta} \sqrt{|g|}. \tag{2.433}$$

The equation (2.432) defines the constraint space of gauge theory.

Given a symmetric connection K on the tangent bundle TX, every principal connection B on P induces the connection

$$\Gamma^m_{\mu\lambda} = \partial_\mu B^m_\alpha - c^m_{nl}k^n_\mu B^l_\alpha - K^\beta{}_{\mu\lambda}(B^m_\beta - k^m_\beta) \tag{2.434}$$

on the bundle of principal connections C.

Let τ be a vector–field on the base X and

$$\tau_{BK} = \tau^\alpha\{\partial_\alpha + [\partial_\mu B^m_\alpha - c^m_{nl}k^n_\mu B^l_\alpha - K^\beta{}_{\mu\lambda}(B^m_\beta - k^m_\beta)]\partial^\mu_m\} \tag{2.435}$$

its horizontal lift onto C by means of the connection (2.434). For every vector–field τ, one can choose the connection K on the tangent bundle TX which has τ as the geodesic field. In this case, the horizontal lift (2.435) of the vector–field τ becomes its canonical lift

$$\tau_B = \tau^\alpha\partial_\alpha + [\tau^\alpha(\partial_\mu B^m_\alpha - c^m_{nl}k^n_\mu B^l_\alpha) + \partial_\mu\tau^\alpha(B^m_\alpha - k^m_\alpha)]\partial^\mu_m\,, \tag{2.436}$$

by means of the principal connection B on the principal bundle P [GM90]. The vector–field (2.436) is just the general principal vector–field on C that has been mentioned in the previous Section. Hence, the Lie derivative of the Lagrangian density (2.431) by the jet lift $\overline{\tau}_B$ of the field τ_B becomes

$$\mathfrak{L}_{\overline{\tau}_B} L_{YM} = (\partial_\alpha\tau^\alpha\mathcal{L}_{YM} + \tau^\alpha\partial_\alpha\mathcal{L}_{YM} - \mathcal{F}^m_{\mu\nu}\partial_\alpha\tau^\mu\pi^{\nu\lambda}_m)\omega.$$

The corresponding SEM transformation law takes the form

$$\partial_\alpha\tau^\alpha\mathcal{L}_{YM} - \tau^\mu t^\alpha_\beta\sqrt{|g|}\Gamma^\beta_{\mu\alpha} - \mathcal{F}^m_{\mu\nu}\partial_\alpha\tau^\mu\pi^{\nu\lambda}_m \approx \tag{2.437}$$
$$\widehat{\partial}_\alpha[\pi^{\nu\lambda}_m(\tau^\mu(\partial_\nu B^m_\mu - c^m_{nl}k^n_\nu B^l_\mu) + \partial_\nu\tau^\mu(B^m_\mu - k^m_\mu) - \tau^\mu k^m_{\nu\mu}) + \delta^\alpha_\mu\tau^\mu\mathcal{L}_{YM}],$$

where

$$t^{\alpha}_{\beta} = \frac{1}{\sqrt{|g|}}(\pi^{\nu\alpha}_{m}\mathcal{F}^{m}_{\beta\nu} - \delta^{\alpha}_{\beta}\mathcal{L}_{YM})$$

is the metric SEM–tensor of gauge potentials.

Note that, in general case of the principal connection B, the corresponding SEM transformation law (2.437) differs from the covariant conservation law in the Noether conservation law

$$\widehat{\partial}_{\alpha}(\pi^{\nu\lambda}_{m}u_{\mathfrak{g}}{}^{m}_{\nu}) \approx 0, \qquad \text{where}$$
$$u_{\mathfrak{g}} = (\partial_{\nu}\alpha^{m} + c^{m}_{nl}k^{l}_{\nu}\alpha^{n})\partial^{\nu}_{m}, \qquad \alpha^{m} = \tau^{\mu}(B^{m}_{\mu} - A^{m}_{\mu})$$

is the principal vector–field (2.360) on C.

Following the general procedure [Sar93, Sar94], let us consider connections on the fibre bundle $C \to X$ which take their values into $\operatorname{Ker}\widehat{L}_{YM}$:

$$\Gamma : C \to C_{+}, \qquad \Gamma^{m}_{\mu\lambda} - \Gamma^{m}_{\lambda\mu} - c^{m}_{nl}k^{n}_{\alpha}k^{l}_{\mu} = 0. \tag{2.438}$$

Moreover, we can restrict ourselves to connections of the following type. Every principal connection B on P induces the connection Γ_B (2.438) on C such that

$$\Gamma_B \circ B = \mathcal{S} \circ j^1 B,$$
$$\Gamma_B{}^{m}_{\mu\lambda} = \frac{1}{2}[c^{m}_{nl}k^{n}_{\alpha}k^{l}_{\mu} + \partial_{\mu}B^{m}_{\alpha} + \partial_{\alpha}B^{m}_{\mu} - c^{m}_{nl}(k^{n}_{\mu}B^{l}_{\alpha} + k^{n}_{\alpha}B^{l}_{\mu})] - \Gamma^{\beta}_{\mu\lambda}(B^{m}_{\beta} - k^{m}_{\beta}).$$

For all these connections, the following *Hamiltonian forms*

$$H_B = p^{\mu\lambda}_{m}dk^{m}_{\mu} \wedge \omega_{\alpha} - p^{\mu\lambda}_{m}\Gamma_B{}^{m}_{\mu\lambda}\omega - \widetilde{\mathcal{H}}_{YM}\omega, \tag{2.439}$$
$$\widetilde{\mathcal{H}}_{YM} = \frac{\varepsilon^2}{4}a_G^{mn}g_{\mu\nu}g_{\lambda\beta}p^{[\mu\lambda]}_{m}p^{[\nu\beta]}_{n}|g|^{-1/2},$$

are associated with the Lagrangian density L_{YM} and constitute a complete family. The corresponding Hamiltonian equations for sections r of $\Pi \longrightarrow X$ read

$$\partial_{\alpha}p^{\mu\lambda}_{m} = -c^{n}_{lm}k^{l}_{\nu}p^{[\mu\nu]}_{n} + c^{n}_{ml}B^{l}_{\nu}p^{(\mu\nu)}_{n} - \Gamma^{\mu}_{\lambda\nu}p^{(\lambda\nu)}_{m}, \tag{2.440}$$
$$\partial_{\alpha}k^{m}_{\mu} + \partial_{\mu}k^{m}_{\alpha} = 2\Gamma_B{}^{m}_{(\mu\lambda)}, \tag{2.441}$$

plus the equation (2.433). The equations (2.433) and (2.440) restricted to the constraint space (2.432) are the familiar Yang–Mills equations. Different Hamiltonian forms (2.439) lead to the different equations (2.441). The equation (2.441) is independent of canonical momenta and plays the role of the gauge–type condition. Its solution is $k(x) = B$.

Let A be a solution of the Yang–Mills equations. There exists the Hamiltonian form $H_{B=A}$ (2.439) such that $r_A = \widehat{L}_{YM} \circ A$ is a solution of the corresponding Hamiltonian equations (2.440), (2.441) and (2.433) on the constraint space (2.432).

On the solution r_A, the curvature of the connection Γ_A is reduced to

$$R^m_{\lambda\alpha\mu} = \frac{1}{2}(\partial_\alpha F^m_{\alpha\mu} - c^m_{qn}k^q_\alpha F^n_{\alpha\mu} - \Gamma^\beta_{\alpha\lambda}F^m_{\beta\mu} - \Gamma^\beta_{\mu\lambda}F^m_{\alpha\beta}) =$$
$$\frac{1}{2}[(\partial_\alpha F^m_{\lambda\mu} - c^m_{qn}k^q_\alpha F^n_{\lambda\mu} - \Gamma^\beta_{\lambda\alpha}F^m_{\mu\beta}) - (\partial_\mu F^m_{\lambda\alpha} - c^m_{qn}k^q_\mu F^n_{\lambda\alpha} - \Gamma^\beta_{\lambda\mu}F^m_{\alpha\beta})]$$

where $F = \mathcal{F} \circ A$ is the strength of A. If we set

$$S^\alpha{}_\mu = p^{[\alpha\lambda]}_m \partial^m_{\alpha\mu}\widetilde{\mathcal{H}}_{YM} = \frac{\varepsilon^2}{2\sqrt{|g|}} a^{mn}_G g_{\mu\nu}g_{\alpha\beta}p^{[\alpha\lambda]}_m p^{[\beta\nu]}_n,$$

then we have

$$S^\alpha{}_\mu = \frac{1}{2}p^{[\alpha\lambda]}\mathcal{F}^m_{\mu\alpha}, \qquad \widetilde{\mathcal{H}}_{YM} = \frac{1}{2}S^\alpha{}_\alpha.$$

Using (2.432), (2.433) and (2.440), we get the relations

$$\partial^\beta_n \Gamma_A{}^m_{\alpha\mu} p^{\alpha\lambda}_m \partial^n_{\beta\lambda}\widetilde{\mathcal{H}}_{YM} = \Gamma^\beta_{\alpha\mu}S^\alpha{}_\beta, \qquad r_A{}^{[\lambda\alpha]}_m R^m_{\lambda\alpha\mu} = \partial_\alpha S^\alpha{}_\mu(r_A) - \Gamma^\beta_{\mu\lambda}S^\alpha{}_\beta(r_A)$$

and we find that

$$t^\alpha{}_\mu\sqrt{|g|} = 2S^\alpha{}_\mu - \frac{1}{2}\delta^\alpha_\mu S^\alpha{}_\alpha, \qquad T_{\Gamma_A}{}^\alpha{}_\mu(r_A) = S^\alpha{}_\mu(r_A) - \frac{1}{2}\delta^\alpha_\mu S^\alpha{}_\alpha(r_A),$$
$$t^\alpha{}_\mu(r_A)\sqrt{|g|} = T^\alpha_{\Gamma_A\mu}(r_A) + S^\lambda{}_\mu(r_A).$$

Hence, the identity (2.428) in gauge theory is brought into the covariant energy–momentum conservation law

$$\nabla_\alpha t^\alpha{}_\mu(r_A) \approx 0.$$

The Lagrangian partner of the Hamiltonian SEM–tensor $T_{\Gamma_A}(r_A)$ is the SEM–tensor $\mathcal{J}_{\Gamma_A}(A)$ (2.414) on the solution A relative to the connection Γ_A on the bundle C. This is exactly the familiar symmetrized canonical energy–momentum tensor of gauge potentials.

SEM Tensors of Matter Fields

In gauge theory, matter fields possessing only internal symmetries are described by sections of a vector bundle $Y = (P \times V)/G$, associated with a principal bundle P [Sar98]. It has a $G-$invariant fibre metric a^Y. Because of the canonical vertical splitting $VY = Y \times Y$, the metric a^Y is a fibre metric in the vertical tangent bundle $VY \to X$. Every principal connection A on a principal bundle P yields the associated connection

$$\Gamma = dx^\alpha \otimes [\partial_\alpha + A^m_\mu(x)I_m{}^i{}_j y^j \partial_i], \tag{2.442}$$

where $A^m_\mu(x)$ are coefficients of the local connection 1–form and I_m are generators of the structure group G on the standard fibre V of the bundle Y.

On the configuration space $J^1(X,Y)$, the regular Lagrangian density of matter fields in the presence of a background connection Γ on Y reads

$$L_{(m)} = \frac{1}{2}a^Y_{ij}[g^{\mu\nu}(y^i_\mu - \Gamma^i_\mu)(y^j_\nu - \Gamma^j_\nu) - m^2 y^i y^j]\sqrt{|g|}\omega. \tag{2.443}$$

The Legendre bundle of the vector bundle $Y \longrightarrow X$ is $\Pi = \wedge^n T^*X \otimes TX \otimes Y^*$. The unique Hamiltonian form on Π associated with the Lagrangian density $L_{(m)}$ (2.443) is written

$$H_{(m)} = p^\alpha_i dy^i \wedge \omega_\alpha - p^\alpha_i \Gamma^i_\alpha \omega - \frac{1}{2}(a^{ij}_Y g_{\mu\nu} p^\mu_i p^\nu_j |g|^{-1} + m^2 a^Y_{ij} y^i y^j)\sqrt{|g|}\omega, \tag{2.444}$$

where a_Y is the fibre metric in V^*Y dual to a^Y. There is the 1–1 correspondence between the solutions of the first–order Euler–Lagrangian equations of the regular Lagrangian density (2.443) and the solutions of the Hamiltonian equations of the Hamiltonian form (2.444).

To examine the conservation law (2.428), let us take the same Hamiltonian SEM–tensor relative to the connection Γ (2.442) for all solutions r of the Hamiltonian equations. The following equality motivates the option above. We have

$$T^\alpha_{\Gamma\mu}(r) = t^\alpha{}_\mu(r)\sqrt{|g|} = [a^{ij}_Y g_{\mu\nu} r^\alpha_i p^\nu_j |g|^{-1}$$
$$-\frac{1}{2}\delta^\alpha_\mu (a^{ij}_Y g_{\alpha\nu} r^\alpha_i r^\nu_j |g|^{-1} + m^2 a^Y_{ij} r^i r^j)]\sqrt{|g|}.$$

The gauge invariance condition $I_m{}^j{}_i r^\alpha_j \partial^i_\alpha \widetilde{\mathcal{H}} = 0$ also takes place. Then, it can be observed that the identity (2.428) reduces to the familiar covariant energy–momentum conservation law

$$\sqrt{|g|}\nabla_\alpha t^\alpha{}_\mu(r) \approx -r^\alpha_i F^m_{\lambda\mu} I_m{}^i{}_j y^j.$$

SEM Tensors in Affine–Metric Gravitation Theory

Now we can apply the Hamiltonian SEM–tensor machinery to gravitation theory [Sar98, GS96]. Here, X^4 is a 4D world manifold which obeys the well–known topological conditions in order that a gravitational field exists on X^4.

Recall that the contemporary concept of gravitational interaction is based on the gauge gravitation theory with two types of gravitational fields: tetrad gravitational fields and Lorentz gauge potentials. In absence of fermion matter, one can choose by gravitational variables a pseudo–Riemannian metric g on a world space–time manifold X^4 and a general linear connections K on the tangent bundle of X^4. We call them a world metric and a world connection respectively. Here we are not concerned with the matter interacting with a general linear connection, for it is non–Lagrangian and hypothetical as a rule.

Let $LX \longrightarrow X^4$ be the principal bundle of linear frames in the tangent spaces to X^4. Its structure group is $GL^+(4,\mathbb{R})$. The world connections are

associated with the principal connections on the principal bundle $LX \longrightarrow X^4$. Hence, there is the 1–1 correspondence between the world connections and the global sections of the quotient bundle

$$C = J^1(X^4, LX)/GL^+(4, \mathbb{R}). \tag{2.445}$$

We therefore can apply the standard procedure of the Hamiltonian gauge theory in order to describe the configuration and phase–spaces of world connections [Sar93, Sar94].

Also, there is the 1–1 correspondence between the world metrics g on X^4 and the global sections of the bundle Σ of pseudo–Riemannian bilinear forms in tangent spaces to X^4. This bundle is associated with the GL_4-principal bundle LX. The 2–fold covering of the bundle Σ is the quotient bundle $LX/SO(3,1)$.

The total configuration space of the affine–metric gravitational variables is the product

$$J^1(X^4, C) \times J^1(X^4, \Sigma). \tag{2.446}$$

coordinated by $(x^\mu, g^{\alpha\beta}, k^\alpha{}_{\beta\mu}, g^{\alpha\beta}{}_\alpha, k^\alpha{}_{\beta\mu\lambda})$. Also, the total phase–space Π of the affine–metric gravity is the product of the Legendre bundles over the above–mentioned bundles C and Σ. It has the corresponding canonical coordinates $(x^\mu, g^{\alpha\beta}, k^\alpha{}_{\beta\mu}, p_{\alpha\beta}{}^\alpha, p_\alpha{}^{\beta\mu\lambda})$.

On the configuration space (2.446), the Hilbert–Einstein Lagrangian density of general relativity reads

$$\begin{aligned} L_{HE} &= -\frac{1}{2\kappa} g^{\beta\lambda} \mathcal{F}^\alpha{}_{\beta\alpha\lambda} \sqrt{-g}\omega, \qquad \text{with} \\ \mathcal{F}^\alpha{}_{\beta\nu\lambda} &= k^\alpha{}_{\beta\lambda\nu} - k^\alpha{}_{\beta\nu\lambda} + k^\alpha{}_{\varepsilon\nu} k^\varepsilon{}_{\beta\lambda} - k^\alpha{}_{\varepsilon\lambda} k^\varepsilon{}_{\beta\nu}. \end{aligned} \tag{2.447}$$

The corresponding Legendre morphism is given by the expressions

$$\begin{aligned} & p_{\alpha\beta}{}^\alpha \circ \widehat{L}_{HE} = 0, \\ & p_\alpha{}^{\beta\nu\lambda} \circ \widehat{L}_{HE} = \pi_\alpha{}^{\beta\nu\lambda} = \frac{1}{2\kappa}(\delta^\nu_\alpha g^{\beta\lambda} - \delta^\alpha_\alpha g^{\beta\nu})\sqrt{-g}, \end{aligned} \tag{2.448}$$

which define the constraint space of general relativity in the affine–metric variables.

Now, let us consider the following connections on the bundle $C \times \Sigma$ in order to construct a complete family of Hamiltonian forms associated with the Lagrangian density (2.447).

Let K be a world space–time connection and

$$\begin{aligned} & \Gamma_K{}^\alpha{}_{\beta\nu\lambda} = \frac{1}{2}[k^\alpha{}_{\varepsilon\nu} k^\varepsilon{}_{\beta\lambda} - k^\alpha{}_{\varepsilon\lambda} k^\varepsilon{}_{\beta\nu} + \partial_\alpha K^\alpha{}_{\beta\nu} + \partial_\nu K^\alpha{}_{\beta\lambda} \\ & -2K^\varepsilon{}_{(\nu\lambda)}(K^\alpha{}_{\beta\varepsilon} - k^\alpha{}_{\beta\varepsilon}) + K^\varepsilon{}_{\beta\lambda} k^\alpha{}_{\varepsilon\nu} + K^\varepsilon{}_{\beta\nu} k^\alpha{}_{\varepsilon\lambda} - K^\alpha{}_{\varepsilon\lambda} k^\varepsilon{}_{\beta\nu} - K^\alpha{}_{\varepsilon\nu} k^\varepsilon{}_{\beta\lambda}] \end{aligned}$$

be the corresponding connection on the bundle C (2.445). Let K' be another symmetric world connection. Building on these connections, we set up the following connection on the bundle $C \times \Sigma$,

$$\Gamma^{\alpha\beta}{}_{\alpha} = -K'^{\alpha}{}_{\varepsilon\lambda}g^{\varepsilon\beta} - K'^{\beta}{}_{\varepsilon\lambda}g^{\alpha\varepsilon},$$
$$\Gamma^{\alpha}{}_{\beta\nu\lambda} = \Gamma_K{}^{\alpha}{}_{\beta\nu\lambda} - \frac{1}{2}R^{\alpha}{}_{\beta\nu\lambda}\,, \tag{2.449}$$

where $R^{\alpha}{}_{\beta\nu\lambda}$ is the *Riemann curvature tensor* of K.

For all connections (2.449), the following Hamiltonian forms are associated with the Lagrangian density L_{HE} and constitute a complete family:

$$\begin{aligned} H_{HE} &= (p_{\alpha\beta}{}^{\alpha}dg^{\alpha\beta} + p_{\alpha}{}^{\beta\nu\lambda}dk^{\alpha}{}_{\beta\nu}) \wedge \omega_{\alpha} - \mathcal{H}_{HE}\omega, \\ \mathcal{H}_{HE} &= -p_{\alpha\beta}{}^{\alpha}(K'^{\alpha}{}_{\varepsilon\lambda}g^{\varepsilon\beta} + K'^{\beta}{}_{\varepsilon\lambda}g^{\alpha\varepsilon}) \\ &\quad + p_{\alpha}{}^{\beta\nu\lambda}\Gamma_K{}^{\alpha}{}_{\beta\nu\lambda} - \frac{1}{2}R^{\alpha}{}_{\beta\nu\lambda}(p_{\alpha}{}^{\beta\nu\lambda} - \pi_{\alpha}{}^{\beta\nu\lambda}) \\ &= -p_{\alpha\beta}{}^{\alpha}(K'^{\alpha}{}_{\varepsilon\lambda}g^{\varepsilon\beta} + K'^{\beta}{}_{\varepsilon\lambda}g^{\alpha\varepsilon}) + p_{\alpha}{}^{\beta\nu\lambda}\Gamma^{\alpha}{}_{\beta\nu\lambda} + \widetilde{\mathcal{H}}_{HE}, \\ \widetilde{\mathcal{H}}_{HE} &= \frac{1}{2\kappa}R\sqrt{-g}. \end{aligned} \tag{2.450}$$

Given the Hamiltonian form H_{HE} (2.450) plus a Hamiltonian form H_M for matter, we have the corresponding Hamiltonian equations

$$\partial_{\alpha}g^{\alpha\beta} + K'^{\alpha}{}_{\varepsilon\lambda}g^{\varepsilon\beta} + K'^{\beta}{}_{\varepsilon\lambda}g^{\alpha\varepsilon} = 0, \tag{2.451}$$

$$\partial_{\alpha}k^{\alpha}{}_{\beta\nu} = \Gamma_K{}^{\alpha}{}_{\beta\nu\lambda} - \frac{1}{2}R^{\alpha}{}_{\beta\nu\lambda}, \tag{2.452}$$

$$\begin{aligned} &\partial_{\alpha}p_{\alpha\beta}{}^{\alpha} = p_{\varepsilon\beta}{}^{\sigma}K'^{\varepsilon}{}_{\alpha\sigma} + p_{\varepsilon\alpha}{}^{\sigma}K'^{\varepsilon}{}_{\beta\sigma} \\ &-\frac{1}{2\kappa}(R_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}R)\sqrt{-g} - \frac{\partial\mathcal{H}_M}{\partial g^{\alpha\beta}}, \end{aligned} \tag{2.453}$$

$$\begin{aligned} &\partial_{\alpha}p_{\alpha}{}^{\beta\nu\lambda} = -p_{\alpha}{}^{\varepsilon[\nu\gamma]}k^{\beta}{}_{\varepsilon\gamma} + p_{\varepsilon}{}^{\beta[\nu\gamma]}k^{\varepsilon}{}_{\alpha\gamma} - p_{\alpha}{}^{\beta\varepsilon\gamma}K^{\nu}{}_{(\varepsilon\gamma)} \\ &-p_{\alpha}{}^{\varepsilon(\nu\gamma)}K^{\beta}{}_{\varepsilon\gamma} + p_{\varepsilon}{}^{\beta(\nu\gamma)}K^{\varepsilon}{}_{\alpha\gamma}\,, \end{aligned} \tag{2.454}$$

plus the motion equations of matter. The Hamiltonian equations (2.451) and (2.452) are independent of canonical momenta and so, reduce to the gauge–type conditions. The equation (2.452) breaks into the following two parts,

$$F^{\alpha}{}_{\beta\lambda\nu} = R^{\alpha}{}_{\beta\nu\lambda}, \qquad \text{and} \tag{2.455}$$

$$\begin{aligned} &\partial_{\nu}(K^{\alpha}{}_{\beta\lambda} - k^{\alpha}{}_{\beta\lambda}) + \partial_{\alpha}(K^{\alpha}{}_{\beta\nu} - k^{\alpha}{}_{\beta\nu}) - 2K^{\varepsilon}{}_{(\nu\lambda)}(K^{\alpha}{}_{\beta\varepsilon} - k^{\alpha}{}_{\beta\varepsilon}) \\ &+ K^{\varepsilon}{}_{\beta\lambda}k^{\alpha}{}_{\varepsilon\nu} + K^{\varepsilon}{}_{\beta\nu}k^{\alpha}{}_{\varepsilon\lambda} - K^{\alpha}{}_{\varepsilon\lambda}k^{\varepsilon}{}_{\beta\nu} - K^{\alpha}{}_{\varepsilon\nu}k^{\varepsilon}{}_{\beta\lambda} = 0, \end{aligned} \tag{2.456}$$

where F is the curvature of the connection $k(x)$. It is clear that the gauge–type conditions (2.451) and (2.452) are satisfied by

$$k(x) = K, \qquad K'^{\alpha}{}_{\beta\lambda} = \Gamma^{\alpha}_{\beta\lambda}. \tag{2.457}$$

When restricted to the constraint space (2.448), the Hamiltonian equations (2.453) and (2.454) become

$$\frac{1}{\kappa}(R_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}R)\sqrt{-g} = -\frac{\partial\mathcal{H}_M}{\partial g^{\alpha\beta}}, \tag{2.458}$$

$$D_\alpha(\sqrt{-g}g^{\nu\beta}) - \delta^\nu_\alpha D_\alpha(\sqrt{-g}g^{\lambda\beta}) + \sqrt{-g}[g^{\nu\beta}(k^\alpha{}_{\alpha\lambda} - k^\alpha{}_{\lambda\alpha}) + g^{\lambda\beta}(k^\nu{}_{\lambda\alpha} - k^\nu{}_{\alpha\lambda}) + \delta^\nu_\alpha g^{\lambda\beta}(k^\mu{}_{\mu\lambda} - k^\mu{}_{\lambda\mu})] = 0, \tag{2.459}$$

where $D_\alpha g^{\alpha\beta} = \partial_\alpha g^{\alpha\beta} + k^\alpha{}_{\mu\lambda}g^{\mu\beta} + k^\beta{}_{\mu\lambda}g^{\alpha\mu}$.

Substituting the equation (2.455) into the equation (2.458), we get the *Einstein equations*

$$\frac{1}{\kappa}(F_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}F) = -t_{\alpha\beta}, \tag{2.460}$$

where $t_{\alpha\beta}$ is the metric SEM–tensor of matter. The equations(2.459) and (2.460) are the familiar equations of affine–metric gravity. In particular, the former is the equation for torsion and nonmetricity terms of the general linear connection $k(x)$. In the absence of matter sources of a general linear connection, it admits the well–known solution

$$k^\alpha{}_{\beta\nu} = \Gamma^\alpha_{\beta\nu} - \frac{1}{2}\delta^\alpha_\nu V_\beta, \qquad D_\alpha g^{\beta\gamma} = V_\alpha g^{\beta\gamma},$$

where V_α is an arbitrary covector–field corresponding to the well–known projective freedom.

Let $s = (k(x), g(x))$ be a solution of the Euler–Lagrangian equations of the first–order Hilbert–Einstein Lagrangian density (2.447) and r the corresponding solution of the Hamiltonian equations of the Hamiltonian form (2.450) where K and K' are given by the expressions (2.457). For this solution r, let us take the SEM–tensor T_s (2.406) relative to the connection (2.449) where K and K' are given by the expressions (2.457). It reads

$$T_s{}^\alpha{}_\mu = \delta^\alpha_\mu\widetilde{\mathcal{H}}_{HE} = \frac{1}{2\kappa}\delta^\alpha_\mu R\sqrt{-g}$$

and the identity (2.405) takes the form

$$(\partial_\mu + \Gamma^{\alpha\beta}{}_\mu\partial_{\alpha\beta} + \Gamma^i_\mu\partial_i - \partial_i\Gamma^j_\mu p^\alpha_j\partial^i_\lambda)(\widetilde{\mathcal{H}}_{HE} + \widetilde{\mathcal{H}}_M)$$
$$\approx \frac{d}{dx^\alpha}(T_s{}^\alpha{}_\mu + T_M{}^\alpha{}_\mu) + p_\alpha{}^{\beta\nu\lambda}R^\alpha{}_{\beta\nu\lambda\mu} + p^\alpha_i R^i_{\lambda\mu} \tag{2.461}$$

where T_M is the SEM–tensor for matter.

One can verify that the SEM–tensor T_s meets the condition

$$(\partial_\mu + \Gamma^{\alpha\beta}{}_\mu\partial_{\alpha\beta})\widetilde{\mathcal{H}}_{HE} = \frac{d}{dx^\alpha}T_s{}^\alpha{}_\mu, \tag{2.462}$$

so that on solutions (2.457), the curvature of the connection (2.449) vanishes. Hence, the identity (2.461) is reduced to the conservation law (2.428) of matter in the presence of a background metric. The gravitation SEM–tensor is eliminated from the conservation law because the Hamiltonian form $\mathcal{H}_{HE}$ is

affine in all canonical momenta. Note that only gauge–type conditions (2.451), (2.452) and the motion equations of matter have been used.

At the same time, since the canonical momenta $p_{\alpha\beta}{}^{\alpha}$ of the world metric are equal to zero, the Hamiltonian equation (2.453) on the Lagrangian constraint space becomes

$$\partial_{\alpha\beta}(\widetilde{\mathcal{H}}_{HE} + \widetilde{\mathcal{H}}_M) = 0.$$

Hence, the equality (2.461) takes the form

$$\pi_{\alpha}{}^{\beta\nu\lambda}\partial_\mu R^{\alpha}{}_{\beta\nu\lambda} + (\partial_\mu + \Gamma^i_\mu\partial_i - \partial_i\Gamma^j_\mu p^\alpha_j\partial^i_\lambda)\widetilde{\mathcal{H}}_M \approx \frac{d}{dx^\alpha}(T_s{}^{\alpha}{}_{\mu} + T_M{}^{\lambda}{}_{\mu}) + p^\alpha_i R^i_{\lambda\mu}. \tag{2.463}$$

This is the form of the energy–momentum conservation law which we observe also in case of quadratic Lagrangian densities of affine–metric gravity. Substituting the equality (2.462) into (2.463), we get the above result.

As a test case of quadratic Lagrangian densities of affine–metric gravity, let us examine the sum

$$L = (-\frac{1}{2\kappa}g^{\beta\lambda}\mathcal{F}^{\alpha}{}_{\beta\alpha\lambda} + \frac{1}{4\varepsilon}g_{\alpha\gamma}g^{\beta\sigma}g^{\nu\mu}g^{\lambda\varepsilon}\mathcal{F}^{\alpha}{}_{\beta\nu\lambda}\mathcal{F}^{\gamma}{}_{\sigma\mu\varepsilon})\sqrt{-g}\omega \tag{2.464}$$

of the Hilbert–Einstein Lagrangian density and the Yang–Mills one. The corresponding Legendre map reads

$$p_{\alpha\beta}{}^{\alpha} \circ \widehat{L} = 0, \tag{2.465}$$

$$p_{\alpha}{}^{\beta(\nu\lambda)} \circ \widehat{L} = 0, \tag{2.466}$$

$$p_{\alpha}{}^{\beta[\nu\lambda]} \circ \widehat{L} = \pi_{\alpha}{}^{\beta\nu\lambda} + \frac{1}{\varepsilon}g_{\alpha\gamma}g^{\beta\sigma}g^{\nu\mu}g^{\lambda\varepsilon}\mathcal{F}^{\gamma}{}_{\sigma\varepsilon\mu}\sqrt{-g}. \tag{2.467}$$

The relations (2.465) and (2.466) defines the Lagrangian constraint space.

Let us consider two connections on the bundle $C \times \Sigma$,

$$\Gamma^{\alpha\beta}{}_{\alpha} = -K'^{\alpha}{}_{\varepsilon\lambda}g^{\varepsilon\beta} - K'^{\beta}{}_{\varepsilon\lambda}g^{\alpha\varepsilon}, \qquad \text{and} \qquad \Gamma^{\alpha}{}_{\beta\nu\lambda} = \Gamma_K{}^{\alpha}{}_{\beta\nu\lambda}, \tag{2.468}$$

where the notations of the expression (2.449) are used. The corresponding Hamiltonian forms

$$\begin{aligned} H &= (p_{\alpha\beta}{}^{\alpha}dg^{\alpha\beta} + p_{\alpha}{}^{\beta\nu\lambda}dk^{\alpha}{}_{\beta\nu}) \wedge \omega_\alpha - \mathcal{H}\omega, \\ \mathcal{H} &= -p_{\alpha\beta}{}^{\alpha}(K'^{\alpha}{}_{\varepsilon\lambda}g^{\varepsilon\beta} + K'^{\beta}{}_{\varepsilon\lambda}g^{\alpha\varepsilon}) + p_{\alpha}{}^{\beta\nu\lambda}\Gamma_K{}^{\alpha}{}_{\beta\nu\lambda} + \widetilde{\mathcal{H}}, \\ \widetilde{\mathcal{H}} &= \frac{\varepsilon}{4}g^{\alpha\gamma}g_{\beta\sigma}g_{\nu\mu}g_{\lambda\varepsilon}(p_{\alpha}{}^{\beta[\nu\lambda]} - \pi_{\alpha}{}^{\beta\nu\lambda})(p_{\gamma}{}^{\sigma[\mu\varepsilon]} - \pi_{\gamma}{}^{\beta\mu\varepsilon}), \end{aligned} \tag{2.469}$$

are associated with the Lagrangian density (2.464) and constitute a complete family.

Given the Hamiltonian form (2.469) plus the Hamiltonian form H_M for matter, we have the corresponding Hamiltonian equations

$$\partial_\alpha g^{\alpha\beta} + K'^{\alpha}{}_{\varepsilon\lambda} g^{\varepsilon\beta} + K'^{\beta}{}_{\varepsilon\lambda} g^{\alpha\varepsilon} = 0, \tag{2.470}$$

$$\partial_\alpha k^{\alpha}{}_{\beta\nu} = \Gamma^{\alpha}_{K\beta\nu\lambda} + \varepsilon g^{\alpha\gamma} g_{\beta\sigma} g_{\nu\mu} g_{\lambda\varepsilon} (p_\gamma{}^{\sigma[\mu\varepsilon]} - \pi_\gamma{}^{\beta\mu\varepsilon}), \tag{2.471}$$

$$\partial_\alpha p_{\alpha\beta}{}^{\alpha} = -\frac{\partial \mathcal{H}}{\partial g^{\alpha\beta}} - \frac{\partial \mathcal{H}_M}{\partial g^{\alpha\beta}}, \tag{2.472}$$

$$\begin{aligned}\partial_\alpha p_\alpha{}^{\beta\nu\lambda} &= -p_\alpha{}^{\varepsilon[\nu\gamma]} k^{\beta}{}_{\varepsilon\gamma} + p_\varepsilon{}^{\beta[\nu\gamma]} k^{\varepsilon}{}_{\alpha\gamma} \\ &\quad - p_\alpha{}^{\beta\varepsilon\gamma} K^{\nu}{}_{(\varepsilon\gamma)} - p_\alpha{}^{\varepsilon(\nu\gamma)} K^{\beta}{}_{\varepsilon\gamma} + p_\varepsilon{}^{\beta(\nu\gamma)} K^{\varepsilon}{}_{\alpha\gamma}\end{aligned} \tag{2.473}$$

plus the motion equations for matter. The equation (2.471) breaks into the equation (2.467) and the gauge–type condition (2.456). The gauge–type conditions (2.470) and (2.456) have the solution (2.457). Substituting the equation (2.471) into the equation (2.472) on the constraint space (2.465), we get the quadratic Einstein equations. Substitution of the equations (2.466) and (2.467) into the equation (2.473) results into the Yang–Mills generalization of the equation (2.459),

$$\partial_\alpha p_\alpha{}^{\beta\nu\lambda} + p_\alpha{}^{\varepsilon[\nu\gamma]} k^{\beta}{}_{\varepsilon\gamma} - p_\varepsilon{}^{\beta[\nu\gamma]} k^{\varepsilon}{}_{\alpha\gamma} = 0.$$

Consider now the splitting of the Hamiltonian form (2.469) with respect to the connection (2.449) and the Hamiltonian density

$$\widetilde{\mathcal{H}}_\Gamma = \widetilde{\mathcal{H}} + \frac{1}{2} p_\alpha{}^{\beta\nu\lambda} R^{\alpha}{}_{\beta\nu\lambda}.$$

Let $s = (k(x), g(x))$ be a solution of the Euler–Lagrangian equations of the Lagrangian density (2.464) and r the corresponding solution of the Hamiltonian equations of the Hamiltonian form (2.469) where K and K' are given by the expressions (2.457). For this solution r, let us take the SEM–tensor T_s (2.406) relative to the connection (2.449) where K and K' are given by the expressions (2.457). It reads

$$\begin{aligned}T_s{}^{\alpha}{}_{\mu} &= \frac{1}{2} p_\alpha{}^{\beta[\nu\lambda]} R^{\alpha}{}_{\beta\nu\mu} + \frac{\varepsilon}{2} g^{\alpha\gamma} g_{\beta\sigma} g_{\nu\delta} g_{\mu\varepsilon} p_\alpha{}^{\beta[\nu\lambda]} (p_\gamma{}^{\sigma[\delta\varepsilon]} - \pi_\gamma{}^{\sigma\delta\varepsilon}) \\ &\quad - \delta^{\alpha}_{\mu} (\widetilde{\mathcal{H}} + \frac{\varepsilon}{2} g^{\alpha\gamma} g_{\beta\sigma} g_{\nu\delta} g_{\tau\varepsilon} \pi_\alpha{}^{\beta\nu\tau} (p_\gamma{}^{\sigma[\delta\varepsilon]} - \pi_\gamma{}^{\sigma\delta\varepsilon}))\end{aligned}$$

and is equal to

$$\frac{1}{\varepsilon} R_\alpha{}^{\beta\nu\lambda} R^{\alpha}{}_{\beta\nu\mu} + \pi_\alpha{}^{\beta\nu\lambda} R^{\alpha}{}_{\beta\nu\mu} - \delta^{\alpha}_{\mu} (\frac{1}{4\varepsilon} R_\alpha{}^{\beta\nu\lambda} R^{\alpha}{}_{\beta\nu\lambda} + \frac{1}{2\kappa} R).$$

The weak identity (2.405) now becomes

$$\begin{aligned}&(\partial_\mu + \Gamma^{\alpha\beta}{}_{\mu} \partial_{\alpha\beta} + \Gamma^{i}_{\mu} \partial_i - \partial_i \Gamma^{j}_{\mu} p^{\alpha}_{j} \partial^{i}_{\lambda} - p_\alpha{}^{\beta\nu\lambda} \frac{\partial}{\partial k^\sigma{}_{\gamma\delta}} \Gamma_K{}^{\alpha}{}_{\beta\nu\mu} \frac{\partial}{\partial P_\sigma{}^{\gamma\delta\lambda}}) (\widetilde{\mathcal{H}}_\Gamma + \widetilde{\mathcal{H}}_M) \\ &\qquad \approx \frac{d}{dx^\alpha} (T_s{}^{\alpha}{}_{\mu} + T_M{}^{\alpha}{}_{\mu}) + p_\alpha{}^{\beta\nu\lambda} R^{\alpha}{}_{\beta\nu\lambda\mu} + p^{\alpha}_{i} R^{i}_{\lambda\mu}\end{aligned}$$

and can be simplified to

$$(\partial_\mu + \Gamma^i_\mu\partial_i - \partial_i\Gamma^j_\mu p^\alpha_j\partial^i_\lambda)\widetilde{\mathcal{H}}_M - p_\alpha{}^{\beta\nu\lambda}\frac{\partial}{\partial k^\sigma{}_{\gamma\delta}}\Gamma_K{}^\alpha{}_{\beta\nu\mu}\frac{\partial}{\partial p_\sigma{}^{\gamma\delta\lambda}}\widetilde{\mathcal{H}}_\Gamma$$

$$\approx \frac{d}{dx^\alpha}(T_s{}^\alpha{}_\mu + T_M{}^\alpha{}_\mu) + p^\alpha_i R^i_{\lambda\mu}\,, \qquad \text{where} \tag{2.474}$$

$$p_\alpha{}^{\beta\nu\lambda}\frac{\partial}{\partial k^\sigma{}_{\gamma\delta}}\Gamma_K{}^\alpha{}_{\beta\nu\mu}\frac{\partial}{\partial p_\sigma{}^{\gamma\delta\lambda}}\widetilde{\mathcal{H}}_\Gamma \tag{2.475}$$

$$= \frac{1}{\kappa}k^\gamma{}_{\beta\mu}(g^{\beta\nu}R^\alpha{}_{\gamma\alpha\nu} - g^{\alpha\nu}R^\beta{}_{\alpha\nu\gamma})\sqrt{-g} - k^\gamma{}_{(\beta\mu)}p_\alpha{}^{\nu\beta\lambda}R^\alpha{}_{\nu\gamma\lambda}.$$

Let us choose the local geodetic coordinate system at a point $x \in X$. Relative to this coordinate system, the equality (2.474) at x becomes the conservation law

$$(\partial_\mu + \Gamma^i_\mu\partial_i - \partial_i\Gamma^j_\mu p^\alpha_j\partial^i_\lambda)\widetilde{\mathcal{H}}_M \approx \frac{d}{dx^\alpha}(T_s{}^\alpha{}_\mu + T_M{}^\alpha{}_\mu) + p^\alpha_i R^i_{\lambda\mu}.$$

For example, in gauge theory, we have

$$\frac{d}{dx^\alpha}(T_\Gamma{}^\alpha{}_\mu + t_M{}^\alpha{}_\mu) = 0,$$

where t_M is the metric SEM–tensor of matter.

Gauge Systems of Gravity and Fermion Fields

In physical reality, one observes three types of field systems: gravitational fields, fermion fields, and gauge fields associated with internal symmetries (see [GS97]). If the gauge invariance under internal symmetries is kept in the presence of a gravitational field, Lagrangian densities of gauge fields must depend on a metric gravitational field only.

In the gauge gravitation theory, gravity is represented by pairs (h, A_h) of gravitational fields h and associated Lorentz connections A_h [HMM95, Sar92]. The connection A_h is usually identified with both a connection on a world manifold X and a *spinor connection* on the *spinor bundle* $S_h \to X$ whose sections describe Dirac fermion fields ψ_h in the presence of the gravitational field h. The problem arises when Dirac fermion fields are described in the framework of the affine–metric gravitation theory. In this case, the fact that a world connection is some Lorentz connection may result from the field equations, but it cannot be assumed in advance. There are models where the world connection is not a Lorentz connection [HMM95]. Moreover, it may happen that a world connection is the Lorentz connection with respect to different gravitational fields [Tho93]. At the same time, a Dirac fermion field can be regarded only in a pair (h, ψ_h) with a certain gravitational field h.

One has to define the representation of cotangent vectors to X by the Dirac's $\gamma-$matrices in order to construct the Dirac operator. Given a *tetrad gravitational field* $h(x)$, we have the representation

$$\gamma_h : dx^\mu \mapsto \widehat{dx}^\mu = h^\mu_a \gamma^a.$$

However, different gravitational fields h and h' yield the nonequivalent representations γ_h and $\gamma_{h'}$.

It follows that fermion–gravitation pairs (h, ψ_h) are described by sections of the *composite spinor bundle*

$$S \to \Sigma \to X, \tag{2.476}$$

where $\Sigma \to X$ is the bundle of gravitational fields h where values of h play the role of parameter coordinates, besides the familiar world coordinates [Sar92]. In particular, every spinor bundle $S_h \to X$ is isomorphic to the restriction of $S \to \Sigma$ to $h(X) \subset \Sigma$. Performing this restriction, we come to the familiar case of a field model in the presence of a gravitational field $h(x)$. The feature of the dynamics of field systems on the composite bundle (2.476) lies in the fact that we have the modified covariant differential of fermion fields which depend on derivatives of gravitational fields h.

As a consequence, we get the following *covariant derivative of Dirac fermion fields* in the presence of a gravitational field $h(x)$:

$$\widetilde{D}_\alpha = \partial_\alpha - \frac{1}{2} A^{abc}{}_\mu (\partial_\alpha h^\mu_c + K^\mu{}_{\nu\lambda} h^\nu_c) I_{ab}, \tag{2.477}$$
$$A^{abc}{}_\mu = \frac{1}{2}(\eta^{ca} h^b_\mu - \eta^{cb} h^a_\mu),$$

where K is a general linear connection on a world manifold X,[6] η is the Minkowski metric, and $I_{ab} = \frac{1}{4}[\gamma_a, \gamma_b]$ are generators of the *spinor Lie group* $L_s = SL(2, \mathbb{C})$.

The covariant derivative (2.477) has been considered by [AM91, PO82, TW95]. The relation (2.479) correspond to the canonical decomposition of the Lie algebra of the general linear group. By the well–known theorem [KN63/9], every general linear connection being projected onto the Lie algebra of the Lorentz group induces a Lorentz connection.

In our opinion, the advantage of the covariant derivative (2.477), consists in the fact that, being derived in the framework of the gauge gravitation theory, it may be also applied to the affine–metric gravitation theory and the

[6] The connection

$$\widetilde{K}^{ab}{}_\alpha = A^{abc}{}_\mu (\partial_\alpha h^\mu_c + K^\mu{}_{\nu\lambda} h^\nu_c) \tag{2.478}$$

is not the connection

$$K^k{}_{m\lambda} = h^k_\mu (\partial_\alpha h^\mu_m + K^\mu{}_{\nu\lambda} h^\nu_m) = K^{ab}{}_\alpha (\eta_{am}\delta^k_b - \eta_{bm}\delta^k_a)$$

written with respect to the reference frame $h^a = h^a_\alpha dx^\alpha$, but there is the relation

$$\widetilde{K}^{ab}{}_\alpha = \frac{1}{2}(K^{ab}{}_\alpha - K^{ba}{}_\alpha). \tag{2.479}$$

If K is a Lorentz connection A_h, then the connection $\widetilde{K}$ given by (2.478) is consistent with K itself.

conventional Einstein's gravitation theory. We are not concerned here with the general problem of equivalence of metric, affine and affine–metric theories of gravity [FK82]. At the same time, when K is the *Levi–Civita connection* of h, the Lagrangian density of fermion fields which uses the covariant derivative (2.477) becomes that in the Einstein's gravitation theory. It follows that the configuration space of metric (or tetrad) gravitational fields and general linear connections may play the role of the universal configuration space of realistic gravitational models. In particular, one then can think of the generalized Komar superpotential as being the universal superpotential of energy–momentum of gravity [GS95].

We follow [GS97] in the geometrical approach to field theory when classical fields are described by global sections of a fibre bundle $Y \longrightarrow X$ over a smooth world space–time manifold X. Their dynamics is phrased in terms of jet spaces [Sar93, Sau89]. Recall that a kth–order differential operator on sections of a fibre bundle $Y \longrightarrow X$ is defined to be a bundle morphism of the jet bundle $J^k(X,Y) \longrightarrow X$ to a vector bundle over X.

In particular, given bundle coordinates (x^μ, y^i) of a fibre bundle $Y \longrightarrow X$, the 1–jet space $J^1(X,Y)$ of Y has the adapted coordinates (x^μ, y^i, y^i_μ), where $y^i_\mu(j^1_x s) = \partial_\mu s^i(x)$.

There is the 1–1 correspondence between the connections on the fibre bundle $Y \to X$ and the global sections $\Gamma = dx^\alpha \otimes (\partial_\alpha + \Gamma^i_\alpha \partial_i)$ of the affine jet bundle $J^1(X,Y) \to Y$. Every connection Γ on $Y \to X$ induces the first–order differential operator on Y,

$$D_\Gamma : J^1(X,Y) \longrightarrow T^*X \otimes VY, \qquad D_\Gamma = (y^i_\alpha - \Gamma^i_\alpha) dx^\alpha \otimes \partial_i,$$

which is called the *covariant differential* relative to the connection Γ.

Recall that in the *first–order Lagrangian formalism*, the 1–jet space $J^1(X,Y)$ of Y plays the role of the finite–dimensional configuration space of fields represented by sections s of a bundle $Y \to X$. A first–order Lagrangian density $L : J^1(X,Y) \longrightarrow \wedge^n T^*X$ is defined to be a *horizontal density* $L = \mathcal{L}(x^\mu, y^i, y^i_\mu)\omega$ on the jet bundle $J^1(X,Y) \to X$, where $\omega = dx^1 \wedge ... \wedge dx^n$, $(n = \dim X)$. Since the jet bundle $J^1(X,Y) \to Y$ is affine, every polynomial Lagrangian density of field theory factors through $L : J^1(X,Y) \stackrel{D}{\longrightarrow} T^*X \otimes VY \to \wedge^n T^*X$, where D is the covariant differential on Y, and VY is the vertical tangent bundle of Y.

Let us consider the gauge theory of gravity and fermion fields. By X is further meant an oriented 4D world manifold which satisfies the well–known topological conditions in order that gravitational fields and spinor structure can exist on X. To summarize these conditions, we assume that X is not compact and that the tangent bundle of X is trivial [GS97].

Let LX be the principal bundle of oriented linear frames in tangent spaces to X. In gravitation theory, its structure group $GL^+(4,\mathbb{R})$ is reduced to the connected Lorentz group $L = SO(1,3)$. It means that there exists a reduced subbundle L^hX of LX whose structure group is L. In accordance with the

well–known theorem, there is the 1–1 correspondence between the reduced L subbundles L^hX of LX and the global sections h of the quotient bundle

$$\Sigma = LX/L \to X. \tag{2.480}$$

These sections h describe gravitational fields on X, for the bundle (2.480) is the 2–folder covering of the bundle of pseudo–Riemannian metrics on X.

Given a section h of Σ, let Ψ^h be an atlas of LX such that the corresponding local sections z^h_ξ of LX take their values into L^hX. With respect to Ψ^h and a holonomic atlas $\Psi^T = \{\psi^T_\xi\}$ of LX, a gravitational field h can be represented by a family of GL_4-valued tetrad functions

$$h_\xi = \psi^T_\xi \circ z^h_\xi, \qquad dx^\alpha = h^\alpha_a(x)h^a. \tag{2.481}$$

By the Lorentz connections A_h associated with a gravitational field h are meant the principal connections on the reduced subbundle L^hX of LX. They give rise to principal connections on LX and to spinor connections on the L_s-lift P_h of L^hX.

Given a Minkowski space M, let $Cl_{1,3}$ be the *complex Clifford algebra*[7] generated by elements of M. A spinor space V is defined to be a minimal left ideal of $Cl_{1,3}$ on which this algebra acts on the left. We have the representation $\gamma : M \otimes V \to V$ of elements of the Minkowski space $M \subset Cl_{1,3}$ by Dirac's matrices γ on V.

Let us consider a bundle of complex Clifford algebras $Cl_{1,3}$ over X whose structure group is the Clifford group of invertible elements of $Cl_{1,3}$. Its subbundles are both a spinor bundle $S_M \to X$ and the bundle $Y_M \to X$ of Minkowski spaces of generating elements of $Cl_{1,3}$. To describe Dirac fermion fields on a world manifold X, one must require Y_M to be isomorphic to the cotangent bundle T^*X of X. It takes place if there exists a reduced L subbundle L^hX such that

$$Y_M = (L^hX \times M)/L.$$

Then, the spinor bundle

[7] Recall that *Clifford algebras* are a type of associative algebra, named after English geometer W. Clifford. They can be thought of as one of the possible generalizations of the complex numbers and quaternions. The theory of Clifford algebras is intimately connected with the theory of quadratic forms and orthogonal transformations. The most important Clifford algebras are those over $\mathbb{R}$ and $\mathbb{C}$ equipped with *nondegenerate quadratic forms*. Recall that every nondegenerate quadratic form on a finite–dimensional real vector space is equivalent to the standard diagonal form $Q(x) = x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_{p+q}^2$, where $n = p + q$ is the dimension of the vector space. The pair of integers (p, q) is called the *signature* of the quadratic form. Similarly, one can define Clifford algebras on complex vector spaces. Every nondegenerate quadratic form on a complex vector space is equivalent to the standard diagonal form $Q(z) = z_1^2 + z_2^2 + \cdots + z_n^2$, so there is essentially only one Clifford algebra in each dimension. One can show that the complex Clifford algebra may be obtained as the complexification of the real one.

$$S_M = S_h = (P_h \times V)/L_s \tag{2.482}$$

is associated with the L_s-lift P_h of L^hX. In this case, there exists the representation

$$\gamma_h : T^*X \otimes S_h = (P_h \times (M \otimes V))/L_s \longrightarrow (P_h \times \gamma(M \times V))/L_s = S_h \tag{2.483}$$

of cotangent vectors to a world manifold X by Dirac's γ-matrices on elements of the spinor bundle S_h. As a shorthand, one can write

$$\widehat{dx}^\alpha = \gamma_h(dx^\alpha) = h^\alpha_a(x)\gamma^a.$$

Given the representation (2.483), we shall say that sections of the spinor bundle S_h describe Dirac fermion fields in the presence of the gravitational field h. Let a principal connection on S_h be given by

$$A_h = dx^\alpha \otimes (\partial_\alpha + \frac{1}{2}A^{ab}{}_\alpha I_{ab}{}^A{}_B \psi^B \partial_A).$$

Given the corresponding covariant differential D and the representation γ_h (2.483), one can construct the Dirac operator on the spinor bundle S_h, as

$$\mathcal{D}_h = \gamma_h \circ D : J^1 S_h \to T^*X \otimes VS_h \to VS_h, \tag{2.484}$$
$$\dot{y}^A \circ \mathcal{D}_h = h^\alpha_a \gamma^{aA}{}_B (y^B_\alpha - \frac{1}{2}A^{ab}{}_\alpha I_{ab}{}^A{}_B y^B).$$

Different gravitational fields h and h' define nonequivalent representations γ_h and $\gamma_{h'}$. It follows that a Dirac fermion field must be regarded only in a pair with a certain gravitational field. There is the 1–1 correspondence between these pairs and sections of the composite spinor bundle (2.476).

Recall that we have a *composite bundle*

$$Y \to \Sigma \to X \tag{2.485}$$

of a bundle $Y \to X$ denoted by Y_Σ and a bundle $\Sigma \to X$. It is coordinated by $(x^\alpha, \sigma^m, y^i)$ where (x^μ, σ^m) are coordinates of Σ and y^i are the fibre coordinates of Y_Σ. We further assume that Σ has a global section.

The application of composite bundles to field theory is founded on the following [Sar92]. Given a global section h of Σ, the restriction Y_h of Y_Σ to $h(X)$ is a subbundle of $Y \to X$. There is the 1–1 correspondence between the global sections s_h of Y_h and the global sections of the composite bundle (2.485) which cover h. Therefore, one can think of sections s_h of Y_h as describing fermion fields in the presence of a background parameter field h, whereas sections of the composite bundle Y describe all the pairs (s_h, h). The configuration space of these pairs is the 1–jet space $J^1(X, Y)$ of the composite bundle Y.

Every connection

$$A_\Sigma = dx^\alpha \otimes (\partial_\alpha + \widetilde{A}^i_\alpha \partial_i) + d\sigma^m \otimes (\partial_m + A^i_m \partial_i)$$

on the bundle Y_Σ induces the *horizontal splitting*

$$VY = VY_\Sigma \oplus (Y \times V\Sigma), \qquad \text{locally given by}$$
$$\dot{y}^i\partial_i + \dot{\sigma}^m\partial_m = (\dot{y}^i - A^i_m\dot{\sigma}^m)\partial_i + \dot{\sigma}^m(\partial_m + A^i_m\partial_i).$$

Using this splitting, one can construct the first–order differential operator (1.197) on the composite bundle Y, namely

$$\widetilde{D} : J^1(X,Y) \to T^*X \otimes VY_\Sigma, \qquad \widetilde{D} = dx^\alpha \otimes (y^i_\alpha - \widetilde{A}^i_\alpha - A^i_m\sigma^m_\alpha)\partial_i. \quad (2.486)$$

This operator possess the following property. Given a global section h of Σ, let Γ be a connection on Σ whose integral section is h, that is, $\Gamma \circ h = j^1h$. Note that the differential (2.486) restricted to $J^1(X,Y)_h \subset J^1(X,Y)$ becomes the familiar covariant differential relative to the connection on Y_h,

$$A_h = dx^\alpha \otimes [\partial_\alpha + (A^i_m\partial_\alpha h^m + \widetilde{A}^i_\alpha)\partial_i].$$

Thus, it is $\widetilde{D}$ that we may use in order to construct a Lagrangian density

$$L : J^1(X,Y) \xrightarrow{\widetilde{D}} T^*X \otimes VY_\Sigma \to \wedge^n T^*X$$

for sections of the composite bundle Y.

In particular, in gravitation theory, we have the composite bundle $LX \to \Sigma \to X$, where Σ is the quotient bundle (2.480) and $LX_\Sigma = LX \to \Sigma$ is the $L-$principal bundle. Let P_Σ be the L_s-principal lift of LX_Σ such that $P_\Sigma/L_s = \Sigma$ and $LX_\Sigma = r(P_\Sigma)$. In particular, there is the imbedding of the L_s-lift P_h of L^hX onto the restriction of P_Σ to $h(X)$ [GS97].

Let us consider the composite spinor bundle (2.476) where $S_\Sigma = (P_\Sigma \times V)/L_s$ is associated with the L_s-principal bundle P_Σ. Note that, given a global section h of Σ, the restriction S_Σ to $h(X)$ is the spinor bundle S_h (2.482) whose sections describe Dirac fermion fields in the presence of the gravitational field h.

Let us give the principal bundle LX with a holonomic atlas $\{\psi^T_\xi, U_\xi\}$ and the principal bundles P_Σ and LX_Σ with associated atlases $\{z^s_\epsilon, U_\epsilon\}$ and $\{z_\epsilon = r \circ z^s_\epsilon\}$. With respect to these atlases, the composite spinor bundle is equipped with the bundle coordinates $(x^\alpha, \sigma^\mu_a, \psi^A)$ where (x^α, σ^μ_a) are coordinates of the bundle Σ such that σ^μ_a are the matrix components of the group element $(\psi^T_\xi \circ z_\epsilon)(\sigma)$, $\sigma \in U_\epsilon$, $\pi_{\Sigma X}(\sigma) \in U_\xi$. Given a section h of Σ, we have $(\sigma^\alpha_a \circ h)(x) = h^\alpha_a(x)$, where $h^\alpha_a(x)$ are the tetrad functions (2.481).

Let us consider the bundle of Minkowski spaces $(LX \times M)/L \to \Sigma$ associated with the $L-$principal bundle LX_Σ. Since LX_Σ is trivial, it is isomorphic to the pull–back $\Sigma \times T^*X$ which we denote by the same symbol T^*X. Then, one can define the bundle morphism γ_Σ over Σ, given by

$$\gamma_\Sigma : T^*X \otimes S_\Sigma \to S_\Sigma, \qquad \widehat{dx}^\alpha = \gamma_\Sigma(dx^\alpha) = \sigma^\alpha_a\gamma^a. \quad (2.487)$$

When restricted to $h(X) \subset \Sigma$, the map (2.487) becomes the morphism γ_h (2.483). We use this morphism in order to construct the total Dirac operator on the composite spinor bundle S (2.476).

Let

$$\widetilde{A} = dx^\alpha \otimes (\partial_\alpha + \widetilde{A}^B_\alpha \partial_B) + d\sigma^\mu_a \otimes (\partial^a_\mu + A^{Ba}{}_\mu \partial_B)$$

be a principal connection on the bundle S_Σ and $\widetilde{D}$ the corresponding differential (2.486). We have the first–order differential operator on S, given by

$$\mathcal{D} = \gamma_\Sigma \circ \widetilde{D} : J^1 S \to T^*X \otimes VS_\Sigma \to VS_\Sigma,$$
$$\dot{\psi}^A \circ \mathcal{D} = \sigma^\alpha_a \gamma^{aA}{}_B (\psi^B_\alpha - \widetilde{A}^B_\alpha - A^{Ba}{}_\mu \sigma^\mu_{a\lambda}).$$

One can think of it as being the total Dirac operator since, for every section h, the restriction of $\mathcal{D}$ to $J^1 S_h \subset J^1 S$ becomes the Dirac operator $\mathcal{D}_h$ (2.484) relative to the connection on the bundle S_h, given by

$$A_h = dx^\alpha \otimes [\partial_\alpha + (\widetilde{A}^B_\alpha + A^{Ba}{}_\mu \partial_\alpha h^\mu_a)\partial_B].$$

In order to construct the differential $\widetilde{D}$ (2.486) on $J^1(X,S)$ in explicit form, let us consider the principal connection on the bundle LX_Σ which is given by the local connection form

$$\widetilde{A} = (\widetilde{A}^{ab}{}_\mu dx^\mu + A^{abc}{}_\mu d\sigma^\mu_c) \otimes I_{ab}, \tag{2.488}$$
$$\widetilde{A}^{ab}{}_\mu = \frac{1}{2} K^\nu{}_{\lambda\mu} \sigma^\alpha_c (\eta^{ca}\sigma^b_\nu - \eta^{cb}\sigma^a_\nu),$$
$$A^{abc}{}_\mu = \frac{1}{2}(\eta^{ca}\sigma^b_\mu - \eta^{cb}\sigma^a_\mu), \tag{2.489}$$

where K is a general linear connection on TX and (2.489) corresponds to the canonical left–invariant connection on the bundle $GL^+(4,\mathbb{R}) \to GL^+(4,\mathbb{R})/L$.

Therefore, the differential $\widetilde{D}$ relative to the connection (2.488) reads

$$\widetilde{D} = dx^\alpha \otimes [\partial_\alpha - \frac{1}{2} A^{abc}{}_\mu (\sigma^\mu_{c\lambda} + K^\mu{}_{\nu\lambda}\sigma^\nu_c) I_{ab}{}^A{}_B \psi^B \partial_A]. \tag{2.490}$$

Given a section h, the connection $\widetilde{A}$ (2.488) is reduced to the Lorentz connection $\widetilde{K}$ (2.478) on $L^h X$, and the differential (2.490) leads to the covariant derivatives of fermion fields (2.477). We will use the differential (2.490) in order to construct a Lagrangian density of Dirac fermion fields. Their Lagrangian density is defined on the configuration space $J^1(X, S \oplus S^+)$ coordinated by $(x^\mu, \sigma^\mu_a, \psi^A, \psi^+_A, \sigma^\mu_{a\lambda}, \psi^A_\alpha, \psi^+_{A\lambda})$. It reads

$$\begin{aligned} L_\psi = \{ & \frac{\mathrm{i}}{2}[\psi^+_A (\gamma^0\gamma^\alpha)^A{}_B (\psi^B_\alpha - \frac{1}{2} A^{abc}{}_\mu (\sigma^\mu_{c\lambda} + K^\mu{}_{\nu\lambda}\sigma^\nu_c) I_{ab}{}^B{}_C \psi^C) \\ & - (\psi^+_{A\lambda} - \frac{1}{2} A^{abc}{}_\mu (\sigma^\mu_{c\lambda} + K^\mu{}_{\nu\lambda}\sigma^\nu_c)\psi^+_C I^{+C}_{ab}{}_A)(\gamma^0\gamma^\alpha)^A{}_B \psi^B] \\ & - m\psi^+_A (\gamma^0)^A{}_B \psi^B \} \sigma^{-1}\omega, \end{aligned} \tag{2.491}$$

where $\gamma^\mu = \sigma^\mu_a \gamma^a$, and $\sigma = \det(\sigma^\mu_a)$, while $\psi^+_A(\gamma^0)^A{}_B\psi^B$ is the Lorentz–invariant fibre metric in the bundle $S \oplus S^*$ [Cra91].

One can show that

$$\frac{\partial \mathcal{L}_\psi}{\partial K^\mu{}_{\nu\lambda}} + \frac{\partial \mathcal{L}_\psi}{\partial K^\mu{}_{\lambda\nu}} = 0.$$

Hence, the Lagrangian density (2.491) depends on the torsion of the general linear connection K only. In particular, it follows that, if K is the Levi-Civita connection of a gravitational field $h(x)$, after the substitution $\sigma^\nu_c = h^\nu_c(x)$, the Lagrangian density (2.491) becomes the familiar Lagrangian density of fermion fields in the Einstein's gravitation theory.

Hawking–Penrose Quantum Gravity and Black Holes

In their search for quantum gravity, S. Hawking and R. Penrose use the *straightforward* application of quantum theory to general relativity [HE79, Pen89, HP96], rather than following the more fashioned string theory approach (described below).

According to Hawking, "Einstein's general relativity is a beautiful theory that agrees with every observation that has been made so far. It might require modifications on the Planck scale, and it might be only a *low energy approximation* to some more fundamental theory, like e.g., superstring theory, but it will not affect many of the predictions that can be get from gravity..." [HE79].

Space–Time Manifold, Gravity, Black Holes and Big Bang

The crucial technique for investigating *Hawking–Penrose singularities and black holes*, has been the study of the global causal structure of space–time [HE79]. Define $I^+(p)$ to be the set of all points of the space–time manifold M that can be reached from the point p by future directed time like curves. One can think of $I^+(p)$ as the set of all events that can be influenced by what happens at p. One now considers the boundary $\dot{I}^+(S)$ of the future of a set S. It is easy to see that this boundary cannot be *time–like*. For in that case, a point q just outside the boundary would be to the future of a point p just inside. Nor can the boundary of the future be *space–like*, except at the set S itself. For in that case every past directed curve from a point q, just to the future of the boundary, would cross the boundary and leave the future of S. That would be a contradiction with the fact that q is in the future of S. Therefore, the boundary of the future is null apart from at S itself.

To show that each generator of the boundary of the future has a past end point on the set, one has to impose some global condition on the causal structure. The strongest and physically most important condition is that of *global space–time hyperbolicity*.[8] The significance of global hyperbolicity for singularity theorems stems from the following [HE79, HP96]. Let U be globally hyperbolic and let p and q be points of U that can be joined by a time like or

[8] Recall that an open set U is said to be globally hyperbolic if:

null curve. Then there is a time–like or null–geodesic between p and q which maximizes the length of time like or null curves from p to q. The method of proof is to show the space of all time like or null curves from p to q is compact in a certain topology. One then shows that the length of the curve is an upper semi–continuous function on this space. It must therefore attain its maximum and the curve of maximum length will be a geodesic because otherwise a small variation will give a longer curve.

One can now consider the second variation of the length of a geodesic γ. One can show that γ can be varied to a longer curve if there is an infinitesimally neighboring geodesic from p which intersects γ again at a point r between p and q. The point r is said to be conjugate to p. One can illustrate this by considering two points p and q on the surface of the Earth. Without loss of generality one can take p to be at the north pole. Because the Earth has a positive definite metric rather than a Lorentzian one, there is a geodesic of minimal length, rather than a geodesic of maximum length. This minimal geodesic will be a line of longitude running from the north pole to the point q. But there will be another geodesic from p to q which runs down the back from the north pole to the south pole and then up to q. This geodesic contains a point conjugate to p at the south pole where all the geodesics from p intersect. Both geodesics from p to q are stationary points of the length under a small variation. But now in a positive definite metric the second variation of a geodesic containing a conjugate point can give a shorter curve from p to q. Thus, on the Earth, the geodesic that goes down to the south pole and then comes up is not the shortest curve from p to q.

The reason one gets conjugate points in space–time is that gravity is an attractive force. It therefore curves space–time in such a way that neighboring geodesics are bent towards each other rather than away. One can see this from the *Newman–Penrose equation*

$$\frac{d\rho}{dv} = \rho^2 + \sigma^{ij}\sigma_{ij} + \frac{1}{n}R_{\alpha\beta}l^\alpha l^\beta, \qquad (\alpha, \beta = 0, 1, 2, 3)$$

where $n = 2$ for null geodesics and $n = 3$ for time–like geodesics. Here v is an affine parameter along a congruence of geodesics, with tangent vector l^α which are hypersurface orthogonal. The quantity ρ is the average rate of convergence of the geodesics, while σ measures the shear. The term $R_{\alpha\beta}l^\alpha l^\beta$ gives the direct gravitational effect of the matter on the convergence of the geodesics. By the *Einstein equation* (1.3), it will be non–negative for any null vector l^α if the matter obeys the so–called *weak energy condition*, which says that the energy density T_{00} is non–negative in any frame, i.e.,

1. For every pair of points p and q in U the intersection of the future of p and the past of q has compact closure. In other words, it is a bounded diamond shaped region.
2. Strong causality holds on U. That is, there are no closed or almost closed time–like curves contained in U.

$$T_{\alpha\beta}v^{\alpha}v^{\beta} \geq 0, \tag{2.492}$$

for any time–like vector v^{α}, is obeyed by the classical SEM–tensor of any reasonable matter [HE79, HP96].

Suppose the weak energy condition holds, and that the null geodesics from a point p begin to converge again and that ρ has the positive value ρ_0. Then the Newman–Penrose equation would imply that the convergence ρ would become infinite at a point q within an affine parameter distance $\frac{1}{\rho_0}$ if the null geodesic can be extended that far. If $\rho = \rho_0$ at $v = v_0$ then $\rho \geq \frac{1}{\rho^{-1}+v_0-v}$. Thus there is a conjugate point before $v = v_0 + \rho^{-1}$.

Infinitesimally neighboring null geodesics from p will intersect at q. This means the point q will be conjugate to p along the null geodesic γ joining them. For points on γ beyond the conjugate point q there will be a variation of γ that gives a time like curve from p. Thus γ cannot lie in the boundary of the future of p beyond the conjugate point q. So γ will have a future end point as a generator of the boundary of the future of p.

The situation with time–like geodesics is similar, except that the *strong energy condition* [HE79, HP96],

$$T_{\alpha\beta}v^{\alpha}v^{\beta} \geq \frac{1}{2}v^{\alpha}v_{\alpha}T, \tag{2.493}$$

that is required to make $R_{\alpha\beta}l^{\alpha}l^{\beta}$ non–negative for every time like vector l^{α}, is rather stronger than the weak energy condition (2.492). However, it is still physically reasonable, at least in an averaged sense, in classical theory. If the strong energy condition holds, and the time like geodesics from p begin converging again, then there will be a point q conjugate to p.

Finally there is the *generic energy condition*, which says:

1. The strong energy condition holds.
2. Every time–like or null geodesic has a point where $l_{[a}R_{b]cd[e}l_{f]}l^{c}l^{d} \neq 0$.

One normally thinks of a space–time singularity as a region in which the curvature becomes unboundedly large. However, the trouble with this definition is that one could simply leave out the singular points and say that the remaining manifold was the whole of space–time. It is therefore better to define space–time as the maximal manifold on which the metric is suitably smooth. One can then recognize the occurrence of singularities by the existence of incomplete geodesics that cannot be extended to infinite values of the affine parameter.

Hawking–Penrose Singularity Theorems

Hawking–Penrose Singularity is defined as follows [HE79, Pen89, HP96]:

A space–time manifold is singular if it is time–like or null geodesically incomplete but cannot be embedded in a larger space–time manifold.

This definition reflects the most objectionable feature of singularities, that there can be particles whose history has a beginning or end at a finite time. There are examples in which geodesic incompleteness can occur with the curvature remaining bounded, but it is thought that generically the curvature will diverge along incomplete geodesics. This is important if one is to appeal to quantum effects to solve the problems raised by singularities in classical general relativity.

Singularity Theorems include:

1. Energy condition (i.e., weak (2.492), strong (2.493), or generic (2.2.6)).
2. Condition on global structure (e.g., there should not be any closed time–like curves).
3. Gravity strong enough to trap a region (so that nothing could escape).

The various singularity theorems show that space–time must be time like or null geodesically incomplete if different combinations of the three kinds of conditions hold. One can weaken one condition if one assumes stronger versions of the other two. The Hawking–Penrose Singularity theorems have the generic energy condition, the strongest of the three energy conditions. The global condition is fairly weak, that there should be no closed time like curves. And the no escape condition is the most general, that there should be either a trapped surface or a closed space like three surface.

The theorems predict singularities in two situations. One is in the future in the gravitational collapse of stars and other massive bodies. Such singularities would be an end of time, at least for particles moving on the incomplete geodesics. The other situation in which singularities are predicted is in the past at the beginning of the present expansion of the universe.

The prediction of singularities means that classical general relativity is not a complete theory. Because the singular points have to be cut out of the space–time manifold one cannot define the field equations there and cannot predict what will come out of a singularity. With the singularity in the past the only way to deal with this problem seems to be to appeal to quantum gravity. But the singularities that are predicted in the future seem to have a property that Penrose has called, Cosmic Censorship. That is they conveniently occur in places like black holes that are hidden from external observers. So any break down of predictability that may occur at these singularities will not affect what happens in the outside world, at least not according to classical theory.

Hawking Cosmic Censorship Hypothesis says: "Nature abhors a naked singularity" [HE79, HP96]. However, there is unpredictability in the quantum theory. This is related to the fact that gravitational fields can have intrinsic entropy which is not just the result of coarse graining. *Gravitational entropy*, and the fact that time has a beginning and may have an end, are the two main themes of Hawking's research, because they are the ways in which gravity is distinctly different from other physical fields.

The fact that gravity has a quantity that behaves like entropy was first noticed in the purely classical theory. It depends on Penrose's Cosmic Cen-

sorship Conjecture. This is unproved but is believed to be true for suitably general initial data and state equations.

One makes the approximation of treating the region around a collapsing star as asymptotically flat. Then, as Penrose showed, one can conformally embed the space–time manifold M in a manifold with boundary $\bar{M}$. The boundary ∂M will be a null surface and will consist of two components, future and past null infinity, called $\mathcal{I}^+$ and $\mathcal{I}^-$. One says that *weak* Cosmic Censorship holds if two conditions are satisfied. First, it is assumed that the null geodesic generators of $\mathcal{I}^+$ are complete in a certain conformal metric. This implies that observers far from the collapse live to an old age and are not wiped out by a thunderbolt singularity sent out from the collapsing star. Second, it is assumed that the past of $\mathcal{I}^+$ is globally hyperbolic. This means there are no naked singularities that can be seen from large distances. Penrose has also a stronger form of Cosmic Censorship which assumes that the whole space–time is globally hyperbolic.

Weak Cosmic Censorship Hypothesis reads:

1. $\mathcal{I}^+$ and $\mathcal{I}^-$ are complete.
2. $I^-(\mathcal{I}^+)$ is globally hyperbolic.

If weak Cosmic Censorship holds, the singularities that are predicted to occur in gravitational collapse cannot be visible from $\mathcal{I}^+$. This means that there must be a region of space–time that is not in the past of $\mathcal{I}^+$. This region is said to be a **black hole** because no light or anything else can escape from it to infinity. The boundary of the black hole region is called the *event horizon*. Because it is also the boundary of the past of $\mathcal{I}^+$ the event horizon will be generated by null–geodesic segments that may have past end points but don't have any future end points. It then follows that if the weak energy condition holds the generators of the horizon cannot be converging. For if they were they would intersect each other within a finite distance [HE79, Pen89, HP96].

This implies that the area of a cross section of the event horizon can never decrease with time and in general will increase. Moreover if two black holes collide and merge together the area of the final black hole will be greater than the sum of the areas of the original black holes. This is very similar to the behavior of entropy according to the *Second Law of Thermodynamics:*[9].

Second Law of Black Hole Mechanics: $\delta\mathcal{A} \geq 0$.

Second Law of Thermodynamics: $\delta S \geq 0$.

The similarity with thermodynamics is increased by what is called the First Law of Black Hole Mechanics, which relates the change in mass of a black hole to the change in the area of the event horizon and the change in its angular momentum and electric charge. One can compare this to the First Law of Thermodynamics which gives the change in internal energy in terms of the change in entropy and the external work done on the system [HE79, HP96]:

[9] Recall that Second Law of Thermodynamics states: Entropy can never decrease and the entropy of a total system is greater than the sum of its constituent parts

First Law of Black Hole Mechanics: $\delta E = \frac{\kappa}{8\pi}\delta\mathcal{A} + \Omega\delta J + \Phi\delta Q$.
First Law of Thermodynamics: $\delta E = T\delta S + P\delta V$.

One sees that if the *area* $\mathcal{A}$ of the *event horizon* is *analogous to entropy* S then the quantity analogous to *temperature* is what is called the *surface gravity of the black hole* κ. This is a measure of the strength of the gravitational field on the event horizon. The similarity with thermodynamics is further increased by the so–called Zeroth Law of Black Hole Mechanics: the surface gravity is the same everywhere on the event horizon of a time independent black hole [HE79].

Zeroth Law of Black Hole Mechanics:
κ is the same everywhere on the horizon of a time independent black hole.

Zeroth Law of Thermodynamics:
T is the same everywhere for a system in thermal equilibrium.

Encouraged by these similarities Bekenstein proposed that some multiple of the area of the event horizon actually was the entropy of a black hole. He suggested a generalized Second Law: the sum of this black hole entropy and the entropy of matter outside black holes would never decrease (see [SV96]).

Generalized Second Law: $\delta(S + c\mathcal{A}) \geq 0$.

However, this proposal was not consistent. If black holes have an entropy proportional to horizon area $\mathcal{A}$ they should also have a non zero temperature proportional to surface gravity.

Path–Integral Model for Black Holes

Recall that the fact that gravity is attractive means that it will tend to draw the matter in the universe together to form objects like stars and galaxies. These can support themselves for a time against further contraction by thermal pressure, in the case of stars, or by rotation and internal motions, in the case of galaxies. However, eventually the heat or the angular momentum will be carried away and the object will begin to shrink. If the mass is less than about one and a half times that of the Sun the contraction can be stopped by the degeneracy pressure of electrons or neutrons. The object will settle down to be a white dwarf or a neutron star respectively. However, if the mass is greater than this limit there is nothing that can hold it up and stop it continuing to contract. Once it has shrunk to a certain critical size the gravitational field at its surface will be so strong that the light cones will be bent inward [HE79, HP96].

If the Cosmic Censorship Conjecture is correct the trapped surface and the singularity it predicts cannot be visible from far away. Thus there must be a region of space–time from which it is not possible to escape to infinity. This region is said to be a black hole. Its boundary is called the event horizon and it is a null surface formed by the light rays that just fail to get away to infinity. As we saw in the last subsection, the area $\mathcal{A}$ of a cross section of the event horizon can never decrease, at least in the classical theory. This, and perturbation calculations of spherical collapse, suggest that black holes will settle down to a stationary state.

Recall that the *Schwarzschild metric form*, given by

$$ds^2 = -(1 - \frac{2M}{r})dt^2 + (1 - \frac{2M}{r})^{-1}dr^2 \ + \ r^2(d\theta^2 + \sin^2\theta d\phi^2),$$

represents the gravitational field that a black hole would settle down to if it were non rotating. In the usual r and t coordinates there is an apparent singularity at the Schwarzschild radius $r = 2M$. However, this is just caused by a bad choice of coordinates. One can choose other coordinates in which the metric is regular there.

Now, if one performs the *Wick rotation*, $t = \mathrm{i}\tau$, one gets a positive definite metric, usually called Euclidean even though they may be curved. In the *Euclidean–Schwarzschild metric*

$$ds^2 = x^2\left(\frac{d\tau}{4M}\right)^2 + \left(\frac{r^2}{4M^2}\right)^2 dx^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)$$

there is again an apparent singularity at $r = 2M$. However, one can define a new radial coordinate x to be $4M(1 - 2Mr^{-1})^{\frac{1}{2}}$.

The metric in the $x - \tau$ plane then becomes like the origin of polar coordinates if one identifies the coordinate τ with period $8\pi M$. Similarly, other Euclidean black hole metrics will have apparent singularities on their horizons which can be removed by identifying the imaginary time coordinate with period $\frac{2\pi}{\kappa}$.

To see the significance of having imaginary time identified with some period β, let us consider the *amplitude* to go from some field configuration ϕ_1 on the surface t_1 to a configuration ϕ_2 on the surface t_2. This will be given by the matrix element of $e^{\mathrm{i}H(t_2 - t_1)}$. However, one can also represent this amplitude as a *path integral* over all fields ϕ between t_1 and t_2 which agree with the given fields ϕ_1 and ϕ_2 on the two surfaces,

$$< \phi_2, t_2|\phi_1, t_1 >=< \phi_2|\exp(-\mathrm{i}H(t_2 - t_1))|\phi_1 >= \int D[\phi]\exp(\mathrm{i}A[\phi]).$$

One now chooses the time separation $(t_2 - t_1)$ to be pure imaginary and equal to β. One also puts the initial field ϕ_1 equal to the final field ϕ_2 and sums over a complete basis of states ϕ_n. On the left one has the expectation value of $e^{-\beta H}$ summed over all states. This is just the *thermodynamic partition function* Z at the temperature $T = \beta^{-1}$,

$$Z = \sum < \phi_n|\exp(-\beta H)|\phi_n >= \int D[\phi]\exp(-A[\phi]). \tag{2.494}$$

On the r.h.s. of this equation one has a path integral. One puts $\phi_1 = \phi_2$ and sums over all field configurations ϕ_n. This means that effectively one is doing the path integral over all fields ϕ on a space–time that is identified periodically in the imaginary time direction with period β. Thus the partition function for the field ϕ at temperature T is given by a path integral over all

fields on a Euclidean space–time. This space–time is periodic in the imaginary time direction with period $\beta = T^{-1}$ [HE79, HP96].

If one calculates the path integral in flat space–time identified with period β in the imaginary time direction one gets the usual result for the partition function of black body radiation. However, as we have just seen, the Euclidean–Schwarzschild solution is also periodic in imaginary time with period $\frac{2\pi}{\kappa}$. This means that fields on the Schwarzschild background will behave as if they were in a thermal state with temperature $\frac{\kappa}{2\pi}$.

The periodicity in imaginary time explained why the messy calculation of frequency mixing led to radiation that was exactly thermal. However, this derivation avoided the problem of the very high frequencies that take part in the frequency mixing approach. It can also be applied when there are interactions between the quantum fields on the background. The fact that the path integral is on a periodic background implies that all physical quantities like expectation values will be thermal. This would have been very difficult to establish in the frequency mixing approach [HE79, HP96].

One can extend these interactions to include interactions with the gravitational field itself. One starts with a background metric g_0 such as the Euclidean–Schwarzschild metric that is a solution of the classical field equations. One can then expand the action A in a power series in the perturbations δg about g_0, as

$$A[g] = A[g_0] + A_2(\delta g)^2 + A_3(\delta g)^3 + ...$$

Here, the linear term vanishes because the background is a solution of the field equations. The quadratic term can be regarded as describing gravitons on the background while the cubic and higher terms describe interactions between the gravitons. The path integral over the quadratic terms are finite. There are non renormalizable divergences at two loops in pure gravity but these cancel with the fermions in super–gravity theories. It is not known whether super–gravity theories have divergences at three loops or higher because no one has been brave or foolhardy enough to try the calculation. Some recent work indicates that they may be finite to all orders. But even if there are higher loop divergences they will make very little difference except when the background is curved on the scale of the Planck length (10^{-33} cm).

More interesting than the higher order terms is the zeroth order term, the action of the background metric g_0 [HE79, HP96],

$$A = -\frac{1}{16\pi}\int R(-g)^{\frac{1}{2}}\ d^4x + \frac{1}{8\pi}\int K(\pm h)^{\frac{1}{2}}\ d^3x.$$

Recall that the usual *Einstein–Hilbert action* for general relativity is the volume integral of the scalar curvature R. This is zero for vacuum solutions so one might think that the action of the Euclidean-Schwarzschild solution was zero. However, there is also a surface term in the action proportional to the integral of K, the trace of the second fundamental form of the boundary

surface. When one includes this and subtracts off the surface term for flat space one finds the action of the Euclidean–Schwarzschild metric is $\frac{\beta^2}{16\pi}$ where β is the period in imaginary time at infinity. Thus the dominant contribution to the path integral for the partition function Z given by (2.494), is $e^{\frac{-\beta^2}{16\pi}}$,

$$Z = \sum \exp(-\beta E_n) = \exp\left(-\frac{\beta^2}{16\pi}\right).$$

If one differentiates $\log Z$ with respect to the period β one gets the expectation value of the energy, or in other words, the mass,

$$< E >= -\frac{d}{d\beta}(\log Z) = \frac{\beta}{8\pi}.$$

So this gives the mass $M = \frac{\beta}{8\pi}$. This confirms the relation between the mass and the period, or inverse temperature, that we already knew. However, one can go further. By standard thermodynamic arguments, the log of the partition function is equal to minus the free energy F divided by the temperature T, i.e., $\log Z = -\frac{F}{T}$. And the free energy is the mass or energy plus the temperature times the entropy S, i.e., $F =< E > +TS$. Putting all this together one sees that the action of the black hole gives an entropy of $4\pi M^2$,

$$S = \frac{\beta^2}{16\pi} = 4\pi M^2 = \frac{1}{4}\mathcal{A}.$$

This is exactly what is required to make the laws of black holes the same as the laws of thermodynamics [HE79, HP96]. The reason why does one get this intrinsic gravitational entropy which has no parallel in other quantum field theories, is that gravity allows different topologies for the space–time manifold.

In the case we are considering the Euclidean–Schwarzschild solution has a boundary at infinity that has topology $S^2 \times S^1$. The S^2 is a large space like two sphere at infinity and the S^1 corresponds to the imaginary time direction which is identified periodical. One can fill in this boundary with metrics of at least two different topologies. One is the Euclidean–Schwarzschild metric. This has topology $R^2 \times S^2$, that is the Euclidean two plane times a two sphere. The other is $R^3 \times S^1$, the topology of Euclidean flat space periodically identified in the imaginary time direction. These two topologies have different Euler numbers. The *Euler number* of periodically identified flat space is zero, while that of the Euclidean–Schwarzschild solution is two,

$$\text{Total action} = M(\tau_2 - \tau_1).$$

The significance of this is as follows: on the topology of periodically identified flat space one can find a periodic time function τ whose gradient is no where zero and which agrees with the imaginary time coordinate on the boundary at infinity. One can then work out the action of the region between two surfaces

τ_1 and τ_2. There will be two contributions to the action, a volume integral over the matter Lagrangian, plus the Einstein–Hilbert Lagrangian and a surface term. If the solution is time independent the surface term over $\tau = \tau_1$ will cancel with the surface term over $\tau = \tau_2$. Thus the only net contribution to the surface term comes from the boundary at infinity. This gives half the mass times the imaginary time interval $(\tau_2 - \tau_1)$. If the mass is non–zero there must be non–zero matter fields to create the mass. One can show that the volume integral over the matter Lagrangian plus the Einstein–Hilbert Lagrangian also gives $\frac{1}{2}M(\tau_2 - \tau_1)$. Thus the total action is $M(\tau_2 - \tau_1)$. If one puts this contribution to the log of the partition function into the thermodynamic formulae one finds the expectation value of the energy to be the mass, as one would expect. However, the entropy contributed by the background field will be zero.

However, the situation is different with the Euclidean–Schwarzschild solution, which says:

$$\text{Total action including corner contribution} = M(\tau_2 - \tau_1)$$

$$\text{Total action without corner contribution} = \frac{1}{2}M(\tau_2 - \tau_1)$$

Because the Euler number is two rather than zero one cannot find a time function τ whose gradient is everywhere non–zero. The best one can do is choose the imaginary time coordinate of the Schwarzschild solution. This has a fixed two sphere at the horizon where τ behaves like an angular coordinate. If one now works out the action between two surfaces of constant τ the volume integral vanishes because there are no matter fields and the scalar curvature is zero. The trace K surface term at infinity again gives $\frac{1}{2}M(\tau_2 - \tau_1)$. However there is now another surface term at the horizon where the τ_1 and τ_2 surfaces meet in a corner. One can evaluate this surface term and find that it also is equal to $\frac{1}{2}M(\tau_2 - \tau_1)$. Thus the total action for the region between τ_1 and τ_2 is $M(\tau_2 - \tau_1)$. If one used this action with $\tau_2 - \tau_1 = \beta$ one would find that the entropy was zero. However, when one looks at the action of the Euclidean Schwarzschild solution from a $4-$dimensional point of view rather than a $3+1$, there is no reason to include a surface term on the horizon because the metric is regular there. Leaving out the surface term on the horizon reduces the action by one quarter the area of the horizon, which is just the intrinsic gravitational entropy of the black hole [HE79, HP96].

Quantum Cosmology

According to Hawking, *cosmology* used to be considered a *pseudo–science* and the preserve of physicists who may have done useful work in their earlier years but who had gone mystic in their dotage. There is a serious objection that cosmology cannot predict anything about the universe unless it makes some assumption about the *initial conditions*. Without such an assumption, all one can say is that things are as they are now because they were as they were

at an earlier stage. Yet many people believe that science should be concerned only with the local laws which govern how the universe evolves in time. They would feel that the boundary conditions for the universe that determine how the universe began were a question for metaphysics or religion rather than science [HE79, HP96].

Hawking–Penrose theorems showed that according to general relativity there should be a singularity in our past. At this singularity the field equations could not be defined. Thus classical general relativity brings about its own downfall: it predicts that it cannot predict the universe. For Hawking this sounds rally disturbing: If the laws of physics could break down at the beginning of the universe, why couldn't they break down any where. In quantum theory it is a principle that anything can happen if it is not absolutely forbidden. Once one allows that singular histories could take part in the path integral they could occur any where and predictability would disappear completely. If the laws of physics break down at singularities, they could break down any where.

The only way to have a scientific theory is if the laws of physics hold everywhere including at the beginning of the universe. One can regard this as a triumph for the *Principle of Democracy*: Why should the beginning of the universe be exempt from the laws that apply to other points. If all points are equal one cannot allow some to be more equal than others.

To implement the idea that the laws of physics hold everywhere, one should take the path integral only over non–singular metrics. One knows in the ordinary path integral case that the measure is concentrated on non–differentiable paths. But these are the completion in some suitable topology of the set of smooth paths with well defined action. Similarly, one would expect that the path integral for quantum gravity should be taken over the completion of the space of smooth metrics. What the path integral cannot include is metrics with singularities whose action is not defined.

In the case of black holes we saw that the path integral should be taken over Euclidean, that is, positive definite metrics. This meant that the singularities of black holes, like the Schwarzschild solution, did not appear on the Euclidean metrics which did not go inside the horizon. Instead the horizon was like the origin of polar coordinates. The action of the Euclidean metric was therefore well defined. One could regard this as a quantum version of Cosmic Censorship: the break down of the structure at a singularity should not affect any physical measurement.

It seems, therefore, that the path integral for quantum gravity should be taken over non–singular Euclidean metrics. But what should the boundary conditions be on these metrics. There are two, and only two, natural choices. The first is metrics that approach the flat Euclidean metric outside a compact set. The second possibility is metrics on manifolds that are compact and without boundary. Therefore, the natural choices for path integral for quantum gravity are [HE79, HP96]: (i) asymptotically Euclidean metrics, and (ii) compact metrics without boundary. The first class of asymptotically Euclidean

metrics is appropriate for *scattering* calculations. In these one sends particles in from infinity and observes what comes out again to infinity. All measurements are made at infinity where one has a flat background metric and one can interpret small fluctuations in the fields as particles in the usual way. One doesn't ask what happens in the interaction region in the middle. That is why one does a path integral over all possible histories for the interaction region, that is, over all asymptotically Euclidean metrics. However, in cosmology one is interested in measurements that are made in a finite region rather than at infinity. We are on the inside of the universe not looking in from the outside. To see what difference this makes let us first suppose that the path integral for cosmology is to be taken over all asymptotically Euclidean metrics.

The so–called No Boundary Proposal of Hartle and Hawking reads [HE79, HP96]: The path integral for quantum gravity should be taken over all compact Euclidean metrics. One can paraphrase this as: *the boundary condition of the universe is that it has no boundary.* According to Hawking, this no boundary proposal seems to account for the universe we live in. That is an isotropic and homogeneous expanding universe with small perturbations. We can observe the spectrum and statistics of these perturbations in the fluctuations in the microwave background. The results so far agree with the predictions of the no boundary proposal. It will be a real test of the proposal and the whole Euclidean quantum gravity program when the observations of the microwave background are extended to smaller angular scales.

In order to use the no boundary proposal to make predictions, it is useful to introduce a concept that can describe the state of the universe at one time:

$$\text{Probability of induced metric } h_{ij} \text{ on } \Sigma = \int_{\substack{\text{metrics on } M \text{ that} \\ \text{induce } h_{ij} \text{ on } \Sigma}} d[g]\ \exp(-A[g]).$$

Consider the probability that the space–time manifold M contains an embedded three dimensional manifold Σ with induced metric h_{ij}. This is given by a path integral over all metrics g_{ab} on M that induce h_{ij} on Σ. If M is simply–connected, which we will assume, the surface Σ will divide M into two parts M^+ and M^- [HE79, HP96],

$$\text{Probability of } h_{ij} = \Psi^+(h_{ij}) \times \Psi^-(h_{ij}), \qquad \text{where}$$
$$\Psi^+(h_{ij}) = \int_{\substack{\text{metrics on } M^+ \text{ that} \\ \text{induce } h_{ij} \text{ on } \Sigma}} d[g]\ \exp(-A[g]).$$

In this case, the probability for Σ to have the metric h_{ij} can be factorized. It is the product of two wave functions Ψ^+ and Ψ^-. These are given by path integrals over all metrics on M^+ and M^- respectively, that induce the given three metric h_{ij} on Σ. In most cases, the two wave functions will be equal and we will drop the superscripts $+$ and $-$. Ψ is called the wave function of the universe. If there are matter fields ϕ, the wave function will also depend on their values ϕ_0 on Σ. But it will not depend explicitly on time because there is

no preferred time coordinate in a closed universe. The no boundary proposal implies that the wave function of the universe is given by a path integral over fields on a compact manifold M^+ whose only boundary is the surface Σ. The path integral is taken over all metrics and matter fields on M^+ that agree with the metric h_{ij} and matter fields ϕ_0 on Σ.

One can describe the position of the surface Σ by a function τ of three coordinates x_i on Σ. But the wave function defined by the path integral cannot depend on τ or on the choice of the coordinates x_i. This implies that the wave function Ψ has to obey four functional differential equations. Three of these equations are called the momentum constraint One can describe the position of the surface Σ by a function τ of three coordinates x_i on Σ. But the wave function defined by the path integral cannot depend on τ or on the choice of the coordinates x_i. This implies that the wave function Ψ has to obey four functional differential equations. Three of these equations are called the *momentum constraint equation*: $\left(\frac{\partial \Psi}{\partial h_{ij}}\right)_{;j} = 0$. They express the fact that the wave function should be the same for different 3 metrics h_{ij} that can be get from each other by transformations of the coordinates x_i. The fourth equation is called the *Wheeler–DeWitt equation*

$$\left(G_{ijkl}\frac{\partial^2}{\partial h_{ij}\partial h_{kl}} - h^{\frac{1}{2}\,3}R\right)\Psi = 0.$$

It corresponds to the independence of the wave function on τ. One can think of it as the *Schrödinger equation for the universe*. But there is no time derivative term because the wave function does not depend on time explicitly.

In order to estimate the wave function of the universe, one can use the *saddle point approximation* to the path integral as in the case of black holes. One finds a Euclidean metric g_0 on the manifold M^+ that satisfies the field equations and induces the metric h_{ij} on the boundary Σ. One can then expand the action A in a power series around the background metric g_0,

$$A[g] = A[g_0] + \frac{1}{2}\delta g A_2 \delta g + ...$$

As before, the term linear in the perturbations vanishes. The quadratic term can be regarded as giving the contribution of gravitons on the background and the higher order terms as interactions between the gravitons. These can be ignored when the radius of curvature of the background is large compared to the Planck scale. Therefore, according to [HE79, HP96] we have

$$\Psi \approx \frac{1}{(\det\, A_2)^{\frac{1}{2}}}\exp(-A[g_o]).$$

Consider now a situation in which there are no matter fields but there is a positive cosmological constant Λ. Let us take the surface Σ to be a three sphere and the metric h_{ij} to be the round three sphere metric of radius a.

Then the manifold M^+ bounded by Σ can be taken to be the four ball. The metric that satisfies the field equations is part of a four sphere of radius $\frac{1}{H}$ where $H^2 = \frac{\Lambda}{3}$,

$$A = \frac{1}{16\pi}\int (R - 2\Lambda)(-g)^{\frac{1}{2}} d^4x + \frac{1}{8\pi}\int K(\pm h)^{\frac{1}{2}} d^3x.$$

For a 3–sphere Σ of radius less than $\frac{1}{H}$ there are two possible Euclidean solutions: either M^+ can be less than a hemisphere or it can be more. However there are arguments that show that one should pick the solution corresponding to less than a hemisphere.

One can interpret the wave function Ψ as follows. The real time solution of the Einstein equations with a Λ term and maximal symmetry is de Sitter space (see, e.g., [Wit98]). This can be embedded as a hyperboloid in five dimensional Minkowski space. Here, we have two choices:

1. *Lorentzian–de Sitter metric*,

$$ds^2 = -dt^2 + \frac{1}{H^2}\cosh Ht(dr^2 + \sin^2 r(d\theta^2 + \sin^2\theta d\phi^2)).$$

   One can think of it as a closed universe that shrinks down from infinite size to a minimum radius and then expands again exponentially. The metric can be written in the form of a Friedmann universe with scale factor $\cosh Ht$. Putting $\tau = it$ converts the cosh into cos giving the Euclidean metric on a four sphere of radius $\frac{1}{H}$.
2. *Euclidean metric*,

$$ds^2 = d\tau^2 + \frac{1}{H^2}\cos H\tau(dr^2 + \sin^2 r(d\theta^2 + \sin^2\theta d\phi^2)).$$

   Thus one gets the idea that a wave function which varies exponentially with the three metric h_{ij} corresponds to an imaginary time Euclidean metric. On the other hand, a wave function which oscillates rapidly corresponds to a real time Lorentzian metric.

Like in the case of the pair creation of black holes, one can describe the spontaneous creation of an exponentially expanding universe. One joins the lower half of the Euclidean four sphere to the upper half of the Lorentzian hyperboloid.

Unlike the black hole pair creation, one couldn't say that the de Sitter universe was created out of field energy in a pre-existing space. Instead, it would quite literally be created out of nothing: not just out of the vacuum but out of absolutely nothing at all because there is nothing outside the universe. In the Euclidean regime, the de Sitter universe is just a closed space like the surface of the Earth but with two more dimensions ([Wit98]). If the cosmological constant is small compared to the Planck value, the curvature of the Euclidean four sphere should be small. This will mean that the saddle point

approximation to the path integral should be good, and that the calculation of the wave function of the universe will not be affected by our ignorance of what happens in very high curvatures.

One can also solve the field equations for boundary metrics that aren't exactly the round three sphere metric. If the radius of the three sphere is less than $\frac{1}{H}$, the solution is a real Euclidean metric. The action will be real and the wave function will be exponentially damped compared to the round three sphere of the same volume. If the radius of the three sphere is greater than this critical radius there will be two complex conjugate solutions and the wave function will oscillate rapidly with small changes in h_{ij}.

Any measurement made in cosmology can be formulated in terms of the wave function. Thus the no boundary proposal makes cosmology into a science because one can predict the result of any observation. The case we have just been considering of no matter fields and just a cosmological constant does not correspond to the universe we live in. Nevertheless, it is a useful example, both because it is a simple model that can be solved fairly explicitly and because, as we shall see, it seems to correspond to the early stages of the universe.

Although it is not obvious from the wave function, a de Sitter universe has thermal properties rather like a black hole. One can see this by writing the de Sitter metric in a *static form* (rather like the Schwarzschild solution)

$$ds^2 = -(1 - H^2 r^2)dt^2 + (1 - H^2 r^2)^{-1} dr^2 + \ r^2(d\theta^2 + \sin^2\theta d\phi^2).$$

There is an apparent singularity at $r = \frac{1}{H}$. However, as in the Schwarzschild solution, one can remove it by a coordinate transformation and it corresponds to an event horizon.

If one returns to the static form of the de Sitter metric and put $\tau = it$ one gets a Euclidean metric. There is an apparent singularity on the horizon. However, by defining a new radial coordinate and identifying τ with period $\frac{2\pi}{H}$, one gets a regular Euclidean metric which is just the four sphere. Because the imaginary time coordinate is periodic, de Sitter space and all quantum fields in it will behave as if they were at a temperature $\frac{H}{2\pi}$. As we shall see, we can observe the consequences of this temperature in the fluctuations in the microwave background. One can also apply arguments similar to the black hole case to the action of the Euclidean–de Sitter solution [Wit98]. One finds that it has an intrinsic entropy of $\frac{\pi}{H^2}$, which is a quarter of the area of the event horizon. Again this entropy arises for a topological reason: the Euler number of the four sphere is two. This means that there cannot be a global time coordinate on Euclidean–de Sitter space. One can interpret this *cosmological entropy* as reflecting an observers lack of knowledge of the universe beyond his event horizon [HE79, HP96]:

$$\text{Euclidean metric periodic with period } \frac{2\pi}{H} \Rightarrow \left\{ \begin{array}{c} \text{Temperature } T = \frac{H}{2\pi}, \\ \text{Area } \mathcal{A} \text{ of event horizon} = \frac{4\pi}{H^2}, \\ \text{Entropy } S \ = \frac{\pi}{H^2}. \end{array} \right.$$

2.2.7 Topological Phase Transitions and Hamiltonian Chaos

Phase Transitions in Hamiltonian Systems

Recall that *phase transitions* (PTs) are phenomena which bring about *qualitative* physical changes at the macroscopic level in presence of the same microscopic forces acting among the constituents of a system. Their mathematical description requires to translate into *quantitative* terms the mentioned qualitative changes. The standard way of doing this is to consider how the values of thermodynamic observables, get in laboratory experiments, vary with temperature, or volume, or an external field, and then to associate the experimentally observed discontinuities at a PT to the appearance of some kind of singularity entailing a loss of analyticity. Despite the smoothness of the statistical measures, after the *Yang–Lee theorem* [YL52] we know that in the $N \to \infty$ limit non–analytic behaviors of thermodynamic functions are possible whenever the analyticity radius in the complex fugacity plane shrinks to zero, because this entails the loss of *uniform convergence* in N (number of degrees of freedom) of any sequence of real–valued thermodynamic functions, and all this depends on the distribution of the zeros of the grand canonical partition function. Also the other developments of the rigorous theory of PTs [Geo88, Rue78], identify PTs with the loss of analyticity.

In this subsection we will address a recently proposed geometric approach to thermodynamic phase transitions (see [CCC97, FCS99, FPS00, FP04]). Given any Hamiltonian system, the configuration space can be equipped with a metric, in order to get a Riemannian geometrization of the dynamics. At the beginning, several numerical and analytical studies of a variety of models showed that the fluctuation of the curvature becomes singular at the transition point. Then the following conjecture was proposed in [CCC97]: The phase transition is determined by a change in the topology of the configuration space, and the loss of analyticity in the thermodynamic observables is nothing but a consequence of such topological change. The latter conjecture is also known as the *topological hypothesis*.

The topological hypothesis states that suitable topology changes of equipotential submanifolds of the Hamiltonian system's configuration manifold can entail thermodynamic phase transitions [FPS00]. The authors of the topological hypothesis gave both a theoretical argument and numerical demonstration in case of $2d$ lattice φ^4 model. They considered classical many–particle (or many–subsystem) systems described by standard *mechanical Hamiltonians*

$$H(p,q) = \sum_{i=1}^{N} \frac{p_i^2}{2m} + V(q), \tag{2.495}$$

where the coordinates $q^i = q^i(t)$ and momenta $p_i = p_i(t)$, $(i = 1, ..., N)$, have continuous values and the system's potential energy $V(q)$ is bounded below.

Now, assuming a large number of subsystems N, the statistical behavior of physical systems described by Hamiltonians of the type (2.495) is usually

encompassed, in the system's canonical ensemble, by the *partition function* in the system's phase–space

$$Z_N(\beta) = \int \prod_{i=1}^{N} dp_i dq^i e^{-\beta H(p,q)} = \left(\frac{\pi}{\beta}\right)^{\frac{N}{2}} \int \prod_{i=1}^{N} dq^i e^{-\beta V(q)}$$
$$= \left(\frac{\pi}{\beta}\right)^{\frac{N}{2}} \int_0^\infty dv\, e^{-\beta v} \int_{M_v} \frac{d\sigma}{\|\nabla V\|}, \tag{2.496}$$

where the last term is written using a *co–area formula* [Fed69], and v labels the *equipotential hypersurfaces* M_v of the system's configuration manifold M,

$$M_v = \{(q^1, \ldots, q^N) \in \mathbb{R}^N | V(q^1, \ldots, q^N) = v\}. \tag{2.497}$$

Equation (2.496) shows that for Hamiltonians (1.17) the relevant statistical information is contained in the *canonical configurational partition function*

$$Z_N^C = \int \prod_{i=1}^{N} dq^i \exp[-\beta V(q)].$$

Therefore, partition function Z_N^C is decomposed – in the last term of equation (2.496) – into an infinite summation of geometric integrals, $\int_{M_v} d\sigma / \|\nabla V\|$, defined on the $\{M_v\}_{v\in\mathbb{R}}$. Once the microscopic interaction potential $V(q)$ is given, the configuration space of the system is automatically foliated into the family $\{M_v\}_{v\in\mathbb{R}}$ of these equipotential hypersurfaces. Now, from standard statistical mechanical arguments we know that, at any given value of the inverse temperature β, the larger the number N of particles the closer to $M_v \equiv M_{u_\beta}$ are the microstates that significantly contribute to the averages – computed through $Z_N(\beta)$ – of thermodynamic observables. The hypersurface M_{u_β} is the one associated with the average potential energy computed at a given β,

$$u_\beta = (Z_N^C)^{-1} \int \prod_{i=1}^{N} dq^i V(q) \exp[-\beta V(q)].$$

Thus, at any β, if N is very large the effective support of the canonical measure shrinks very close to a single $M_v = M_{u_\beta}$.

Explicitly, the topological hypothesis reads: the basic origin of a phase transition lies in a suitable topology change of the $\{M_v\}$, occurring at some v_c. This topology change induces the singular behavior of the thermodynamic observables at a phase transition. By change of topology we mean that $\{M_v\}_{v<v_c}$ are not diffeomorphic to the $\{M_v\}_{v>v_c}$. In other words, canonical measure should 'feel' a big and sudden change of the topology of the equipotential hypersurfaces of its underlying support, the consequence being the appearance of the typical signals of a phase transition.

This point of view has the interesting consequence that – also at finite N – in principle *different* mathematical objects, i.e., manifolds of different

cohomology type, could be associated to *different* thermodynamical phases, whereas from the point of view of measure theory [YL52] the only mathematical property available to signal the appearance of a phase transition is the loss of analyticity of the grand–canonical and canonical averages, a fact which is compatible with analytic statistical measures only in the mathematical $N \to \infty$ limit.

As it is conjectured that the counterpart of a phase transition is a breaking of diffeomorphicity among the surfaces M_v, it is appropriate to choose a *diffeomorphism invariant* to probe if and how the topology of the M_v changes as a function of v. This is a very challenging task because we have to deal with high dimensional manifolds. Fortunately a topological invariant exists whose computation is feasible, yet demands a big effort. Recall (from subsection 1.2.9 above) that this is the *Euler characteristic*, a diffeomorphism invariant of the system's configuration manifold, expressing its fundamental topological information.

Geometry of the Largest Lyapunov Exponent

Now, the topological hypothesis has recently been promoted into a *topological theorem* [FP04]. The new theorem says that non–analyticity is the 'shadow' of a more fundamental phenomenon occurring in the system's configuration manifold: a *topology change* within the family of equipotential hypersurfaces (2.497). This topological approach to PTs stems from the numerical study of the Hamiltonian dynamical counterpart of phase transitions, and precisely from the observation of *discontinuous* or *cuspy patterns*, displayed by the *largest Lyapunov exponent* at the transition energy (or temperature).

Recall that the *Lyapunov exponents* measure the strength of *dynamical chaos* and cannot be measured in laboratory experiments, at variance with thermodynamic observables, thus, being genuine dynamical observables they are only measurable in numerical simulations of the microscopic dynamics. To get a hold of the reason why the largest Lyapunov exponent λ_1 should probe configuration space topology, let us first remember that for standard Hamiltonian systems, λ_1 is computed by solving the *tangent dynamics equation* for Hamiltonian systems (see Jacobi equation of geodesic deviation (1.90)),

$$\ddot{\xi}_i + \left(\frac{\partial^2 V}{\partial q^i \partial q^j} \right)_{q(t)} \xi^j = 0, \tag{2.498}$$

which, for the nonlinear Hamiltonian system

$$\begin{aligned} \dot{q}^1 &= p_1, & \dot{p}_1 &= -\partial_{q^1} V, \\ &\dots & &\dots \\ \dot{q}^N &= p_N, & \dot{p}_N &= -\partial_{q^N} V, \end{aligned}$$

expands into *linearized Hamiltonian dynamics*

$$\begin{aligned}
\dot{\xi}_1 &= \xi_{N+1}, & \dot{\xi}_{N+1} &= -\sum_{j=1}^{N} \left(\frac{\partial^2 V}{\partial q_1 \partial q_j} \right)_{q(t)} \xi_j, \\
&\dots & &\dots \\
\dot{\xi}_n &= \xi_{2N}, & \dot{\xi}_{2N} &= -\sum_{j=1}^{N} \left(\frac{\partial^2 V}{\partial q_N \partial q_j} \right)_{q(t)} \xi_j.
\end{aligned} \tag{2.499}$$

Using (2.498) we can get the analytical expression for the largest Lyapunov exponent

$$\lambda_1 = \lim_{t\to\infty} \frac{1}{t} \log \frac{\left[\xi_1^2(t) + \dots + \xi_N^2(t) + \dot{\xi}_1^2(t) + \dots + \dot{\xi}_N^2(t)\right]^{1/2}}{\left[\xi_1^2(0) + \dots + \xi_N^2(0) + \dot{\xi}_1^2(0) + \dots + \dot{\xi}_N^2(0)\right]^{1/2}}. \tag{2.500}$$

If there are critical points of V in configuration space, that is points $q_c = [\overline{q}^1, \dots, \overline{q}^N]$ such that $\nabla V(q)|_{q=q_c} = 0$, according to the *Morse lemma* (see e.g., [Hir76]), in the neighborhood of any critical point q_c there always exists a coordinate system $\tilde{q}(t) = [\overline{q}^1(t), \dots, \overline{q}^N(t)]$ for which

$$V(\tilde{q}) = V(q_c) - \left(\overline{q}^1\right)^2 - \dots - \left(\overline{q}^k\right)^2 + \left(\overline{q}^{k+1}\right)^2 + \dots + \left(\overline{q}^N\right)^2, \tag{2.501}$$

where k is the index of the critical point, i.e., the number of negative eigenvalues of the Hessian of V. In the neighborhood of a critical point, equation (2.501) yields

$$\partial^2 V / \partial q^i \partial q^j = \pm \delta_{ij},$$

which, substituted into equation (2.498), gives k *unstable directions* which contribute to the exponential growth of the norm of the tangent vector $\xi = \xi(t)$. This means that the strength of dynamical chaos, measured by the largest Lyapunov exponent λ_1, is affected by the existence of critical points of V. In particular, let us consider the possibility of a sudden variation, with the potential energy v, of the number of critical points (or of their indexes) in configuration space at some value v_c, it is then reasonable to expect that the pattern of $\lambda_1(v)$ – as well as that of $\lambda_1(E)$ since $v = v(E)$ – will be consequently affected, thus displaying *jumps* or *cusps* or other *singular patterns* at v_c.

On the other hand, recall that *Morse theory* teaches us that the existence of critical points of V is associated with topology changes of the hypersurfaces $\{M_v\}_{v\in\mathbb{R}}$, provided that V is a good *Morse function* (that is: bounded below, with no vanishing eigenvalues of its Hessian matrix). Thus the existence of critical points of the potential V makes possible a conceptual link between dynamics and configuration space topology, which, on the basis of both direct and indirect evidence for a few particular models, has been formulated as a topological hypothesis about the relevance of topology for PTs phenomena (see [FPS00, FP04, GM04]).

Here we give two simple examples of standard Hamiltonian systems of the form (2.495), namely Peyrard–Bishop system and mean–field XY model.

Peyrard–Bishop Hamiltonian System

The *Peyrard–Bishop system* [PB89][10] exhibits a *second–order phase transition*. It is defined by the following potential energy

$$V(q) = \sum_{i=1}^{N} \left[\frac{K}{2}(q^{i+1} - q^i)^2 + D(e^{-aq^i} - 1)^2 + Dhaq^i \right], \tag{2.502}$$

which represents the energy of a string of N base pairs of reduced mass m. Each hydrogen bond is characterized by the stretching q^i and its conjugate momentum $p_i = m\dot{q}^i$. The elastic transverse force between neighboring pairs is tuned by the constant K, while the energy D and the inverse length a determine, respectively, the plateau and the narrowness of the on–site potential well that mimics the interaction between bases in each pair. It is understood that K, D, and a are all positive parameters. The transverse, external stress $h \geq 0$ is a computational tool useful in the evaluation of the susceptibility. Our interest in it lies in the fact that a phase transition can occur only when $h = 0$. We assume periodic boundary conditions.

The transfer operator technique [DTP02] maps the problem of computing the classical partition function into the easier task of evaluating the lowest energy eigenvalues of a 'quantum' mechanical Morse oscillator (no real quantum mechanics is involved, since the temperature plays the role of $\hbar$). One can then observe that, as the temperature increases, the number of levels belonging to the discrete spectrum decreases, until for some critical temperature $T_c = 2\sqrt{2KD}/(ak_B)$ only the continuous spectrum survives. This passage from a localized ground state to an unnormalizable one corresponds to the second–order phase transition of the statistical model. Various critical exponents can be analytically computed and all applicable scaling laws can be checked. The simplicity of this model permits an analytical computation of the largest Lyapunov exponent by exploiting the geometric method proposed in [CCC97].

Mean–Field XY Hamiltonian System

The mean–field XY model describes a system of N equally coupled planar classical rotators (see [AR95, CCP99]). It is defined by a Hamiltonian of the class (2.495) where the potential energy is

$$V(\varphi) = \frac{J}{2N} \sum_{i,j=1}^{N} \left[1 - \cos(\varphi_i - \varphi_j)\right] - h \sum_{i=1}^{N} \cos\varphi_i. \tag{2.503}$$

Here $\varphi_i \in [0, 2\pi]$ is the rotation angle of the $i-$th rotator and h is an external field. Defining at each site i a classical spin vector $\mathbf{s}_i = (\cos\varphi_i, \sin\varphi_i)$ the

[10] The Peyrard–Bishop system has been proposed as a simple model for describing the DNA thermally induced denaturation [GM04].

model describes a planar (XY) Heisenberg system with interactions of equal strength among all the spins. We consider only the ferromagnetic case $J > 0$; for the sake of simplicity, we set $J = 1$. The equilibrium statistical mechanics of this system is exactly described, in the thermodynamic limit, by the *mean–field theory* [AR95]. In the limit $h \to 0$, the system has a continuous phase transition, with classical critical exponents, at $T_c = 1/2$, or $\varepsilon_c = 3/4$, where $\varepsilon = E/N$ is the energy per particle.

The Lyapunov exponent λ_1 of this system is extremely sensitive to the phase transition. According to reported numerical simulations (see [CCP99]), $\lambda_1(\varepsilon)$ is positive for $0 < \varepsilon < \varepsilon_c$, shows a sharp maximum immediately below the critical energy, and drops to zero at ε_c in the thermodynamic limit, where it remains zero in the whole region $\varepsilon > \varepsilon_c$, which corresponds to the thermodynamic disordered phase. In fact in this phase the system is integrable, reducing to an assembly of uncoupled rotators.

Euler Characteristics of Hamiltonian Systems

Recall that *Euler characteristic* χ is a number that is a characterization of the various classes of geometric figures based only on the topological relationship between the numbers of vertices V, edges E, and faces F, of a geometric Figure. This number, $\chi = F - E + V$, is the same for all figures the boundaries of which are composed of the same number of connected pieces. Therefore, the Euler characteristic is a *topological invariant*, i.e., any two geometric figures that are homeomorphic to each other have the same Euler characteristic.

More specifically, a standard way to analyze a geometric Figure is to fragment it into other more familiar objects and then to examine how these pieces fit together. Take for example a surface M in the Euclidean 3D space. Slice M into pieces that are curved triangles (this is called a triangulation of the surface). Then count the number F of faces of the triangles, the number E of edges, and the number V of vertices on the tesselated surface. Now, no matter how we triangulate a compact surface Σ, its Euler characteristic, $\chi(\Sigma) = F - E + V$, will always equal a constant which is characteristic of the surface and which is invariant under diffeomorphisms $\phi : \Sigma \to \Sigma'$.

At higher dimensions this can be again defined by using higher dimensional generalizations of triangles (simplexes) and by defining the Euler characteristic $\chi(M)$ of the nD manifold M to be the alternating sum:

$$\begin{aligned} &\{\text{number of points}\} - \{\text{number of 2-simplices}\} \quad + \\ &\{\text{number of 3-simplices}\} - \{\text{number of 4-simplices}\} \quad + \quad ... \end{aligned}$$

i.e.,

$$\chi(M) = \sum_{k=0}^{n} (-1)^k (\text{number of faces of dimension } k). \tag{2.504}$$

and then define the Euler characteristic of a manifold as the Euler characteristic of any simplicial complex homeomorphic to it. With this definition,

circles and squares have Euler characteristic 0 and solid balls have Euler characteristic 1.

The Euler characteristic χ of a manifold is closely related to its *genus* g as $\chi = 2 - 2g$.[11]

Recall that in differential topology a more standard definition of $\chi(M)$ is

$$\chi(M) = \sum_{k=0}^{n} (-1)^k \, b_k(M), \tag{2.505}$$

where b_k are the kth *Betti numbers* of M.

In general, it would be hopeless to try to practically calculate $\chi(M)$ from (2.505) in the case of non–trivial physical models at large dimension. Fortunately, there is a possibility given by the *Gauss–Bonnet formula*, that relates $\chi(M)$ with the total *Gauss–Kronecker curvature* of the manifold, (compare with (1.83) and (1.92))

$$\chi(M) = \gamma \int_M K_G \, d\sigma, \tag{2.506}$$

which is valid for even dimensional hypersurfaces of Euclidean spaces $\mathbb{R}^N$ [here $\dim(M) = n \equiv N - 1$], and where:

$$\gamma = 2/\operatorname{vol}(S_1^n)$$

is twice the inverse of the volume of an $n-$dimensional sphere of unit radius S_1^n; K_G is the Gauss–Kronecker curvature of the manifold;

$$d\sigma = \sqrt{\det(g)} \, dx^1 dx^2 \cdots dx^n$$

is the invariant volume measure of M and g is its Riemannian metric (induced from $\mathbb{R}^N$). Let us briefly sketch the meaning and definition of the Gauss–Kronecker curvature. The study of the way in which an $n-$surface M curves around in $\mathbb{R}^N$ is measured by the way the normal direction changes as we move from point to point on the surface. The rate of change of the normal direction ξ at a point $x \in M$ in direction v is described by the *shape operator*

$$L_x(v) = -\mathcal{L}_v \xi = [v, \xi],$$

where v is a tangent vector at x and $\mathcal{L}_v$ is the Lie derivative, hence

[11] Recall that the *genus* of a topological space such as a surface is a topologically invariant property defined as the largest number of nonintersecting simple closed curves that can be drawn on the surface without separating it, i.e., an integer representing the maximum number of cuts that can be made through it without rendering it disconnected. This is roughly equivalent to the number of holes in it, or handles on it. For instance: a point, line, and a sphere all have genus 0; a torus has genus 1, as does a coffee cup as a solid object (solid torus), a Möbius strip, and the symbol 0; the symbols 8 and B have genus 2; etc.

$$L_x(v) = -(\nabla\xi_1 \cdot v, \ldots, \nabla\xi_{n+1} \cdot v);$$

gradients and vectors are represented in $\mathbb{R}^N$. As L_x is an operator of the tangent space at x into itself, there are n independent eigenvalues $\kappa_1(x), \ldots, \kappa_n(x)$ which are called the principal curvatures of M at x [Tho79]. Their product is the Gauss–Kronecker curvature:

$$K_G(x) = \prod_{i=1}^{n} \kappa_i(x) = \det(L_x).$$

Alternatively, recall that according to the *Morse theory*, it is possible to understand the topology of a given manifold by studying the regular critical points of a smooth *Morse function* defined on it. In our case, the manifold M is the configuration space $\mathbb{R}^N$ and the natural choice for the Morse function is the potential $V(q)$. Hence, one is lead to define the family M_v (2.497) of submanifolds of M.

A full characterization of the topological properties of M_v generally requires the critical points of $V(q)$, which means solving the equations

$$\partial_{q^i} V = 0, \qquad (i = 1, \ldots, N). \tag{2.507}$$

Moreover, one has to calculate the indexes of all the critical points, that is the number of negative eigenvalues of the Hessian $\partial^2 V/(\partial q^i \partial q_j)$. Then the Euler characteristic $\chi(M_v)$ can be computed by means of the formula

$$\chi(M_v) = \sum_{k=0}^{N} (-1)^k \mu_k(M_v), \tag{2.508}$$

where $\mu_k(M_v)$ is the total number of critical points of $V(q)$ on M_v which have index k, i.e., the so–called *Morse numbers* of a manifold M, which happen to be upper bounds of the Betti numbers,

$$b_k(M) \leq \mu_k(M) \qquad (k = 0, \ldots, n). \tag{2.509}$$

Among all the Morse functions on a manifold M, there is a special class, called *perfect Morse functions*, for which the Morse inequalities (2.509) hold as equalities. Perfect Morse functions characterize completely the topology of a manifold.

Now, we continue with our two examples started before.

Peyrard–Bishop System. If applied to any generic model, calculation of (2.508) turns out to be quite formidable, but the exceptional simplicity of the Peyrard–Bishop model (2.502) makes it possible to carry on completely the topological analysis without invoking equation (2.508).

For the potential in exam, equation (2.507) results in the nonlinear system

$$\frac{a}{R}(q^{i+1} - 2q^i + q^{i-1}) = h - 2(e^{-2aq^i} - e^{-aq^i}),$$

where $R = Da^2/K$ is a dimensionless ratio. It is easy to verify that a particular solution is given by

$$q^i = -\frac{1}{a} \ln \frac{1 + \sqrt{1+2h}}{2}, \qquad (i = 1, \ldots, N).$$

The corresponding minimum of potential energy is

$$V_{\min} = ND \left(\frac{1 + h - \sqrt{1+2h}}{2} - h \ln \frac{1 + \sqrt{1+2h}}{2} \right).$$

Mean–Field XY Model. In the case of the mean–field XY model (2.503) it is possible to show analytically that a topological change in the configuration space exists and that it can be related to the thermodynamic phase transition. Consider again the family M_v of submanifolds of the configuration space defined in (2.497); now the potential energy per degree of freedom is that of the mean–field XY model, i.e.,

$$\mathcal{V}(\varphi) = \frac{V(\varphi)}{N} = \frac{J}{2N^2} \sum_{i,j=1}^{N} \left[1 - \cos(\varphi_i - \varphi_j)\right] - h \sum_{i=1}^{N} \cos \varphi_i,$$

where $\varphi_i \in [0, 2\pi]$. Such a function can be considered a Morse function on M, so that, according to Morse theory, all these manifolds have the same topology until a critical level $\mathcal{V}^{-1}(v_c)$ is crossed, where the topology of M_v changes.

A change in the topology of M_v can only occur when v passes through a critical value of $\mathcal{V}$. Thus in order to detect topological changes in M_v we have to find the critical values of $\mathcal{V}$, which means solving the equations

$$\partial_{\varphi_i} \mathcal{V}(\varphi) = 0, \qquad (i = 1, \ldots, N). \tag{2.510}$$

For a general potential energy function $\mathcal{V}$, the solution of (2.510) would be a formidable task, but in the case of the mean–field XY model, the mean–field character of the interaction greatly simplifies the analysis, allowing an analytical treatment of (2.510); moreover, a projection of the configuration space onto a 2D plane is possible [CCP99, CPC03].

2.2.8 Topological String Theory

Three Pillars of Modern Physics

Arguably, the three most influential geniuses that shaped the world of the 20th Century physics, and at the same time showed the pathway to the current superstring theory ('of everything'), are:

1. In the first third of the Century, it had been Albert Einstein.
2. In the second third of the Century, it was Richard Feynman.
3. At the end of the Century – and still today, it has been Edward Witten.

It is well–known that *Einstein* had three periods of his scientific career:

1. Before 1905, when he formulated Special Relativity in a quick series of papers published in Annalen der Physik (the most prestigious physics journal of the time). This early period was dominated by his 'thought experiments', i.e., 'concrete physical images', described in the language of non–professional mathematics. You can say, it was almost pure visualization. This quick and powerful series of ground–braking papers (with just enough maths to be accepted by scientific community) gave him a reputation of the leading physicist and scientist.[12]
2. Although an original and brilliant theory, Special Relativity was not complete, which was obvious to Einstein. So, he embarked onto the general relativity voyage, incorporating gravitation. Now, for this goal, his maths was not strong enough. He spent 10 years fighting with gravity, using the 'hard' Riemannian geometry, and talking to the leading mathematician of the time, David Hilbert. At the end, they both submitted the same gravitational equations of general relativity (only derived in different ways) to Annalen der Physik in November of 1915.
3. Although even today considered as the most elegant physical theory, General Relativity is still not complete: it cannot live together in the same world with quantum mechanics. So, Einstein embarked onto the last journey of his life, the search for *unified field theory* – and he 'failed'[13] after 30 years of unsuccessful struggle with a task to big for one man.

Feynman's story is very different. All his life he was a profoundly original scientist, similar to the young Einstein. He refused to take anybody's word for anything, which meant that he had to reinvent for himself almost the whole of physics. It took him five years of concentrated work to reinvent quantum mechanics. At the end, he got a new version of quantum mechanics that he (and only he) could understand. In orthodox physics it was said: Suppose an electron is in this state at a certain time, then you calculate its future behavior by solving Schrodinger equation. Instead of this, Feynman said simply: "The electron does whatever it likes." A history of the electron is any possible path in space and time. The behavior of the electron is just the result of adding together all histories according to some simple rules that Feynman worked out. His path–integral and related Feynman diagrams, for long defied rigorous mathematical foundation. However, it is still the most powerful calculation tool in quantum (and statistical) mechanics. Later, Feynman generalized it to encompass physical fields – which led to his version of quantum

[12] Recall that the Nobel Prize was 'in the air' for Einstein for more than 15 years; at the end he got it in 1921, for his discovery of the Photo–Electric Effect.

[13] Einstein 'failed' in the same way as Hilbert 'failed' with his Program of axiomatic formalization of all mathematical sciences. Their apparent 'failure' still influences development of physics and mathematics, apparently converging into superstring theory. Their joined work on gravity is called the *Einstein–Hilbert action*.

electrodynamics (the first prototype of a quantum field theory) – and his Nobel Prize. All his career he consistently distrusted official mathematics and invented his own maths underpinned with a direct physical intuition.

If the story ended here, we might say that *visual physical intuition* is leading the way of science. However, the story does not end here. The leading authority in contemporary physics is *Ed Witten*, a physicist who did not get the Nobel Prize, but rather the Fields Medal – together with his string theory.[14] Witten works at the same place where Einstein spent the last 30 years of his life – at the Princeton Institute of Advanced Study. He is dreaming Einstein's dream: a *unified theory of everything*, using the most powerful maths possible. His prophecy, delivered at a turn of the Century, has been: "In the 21 fist Century, mathematics will be dominated by string theory."

When superstring theory arrived in physics in 1984 as a potential theory of the universe, it was considered by mainstream physicists as little better than religion in terms of constituting a viable, testable theory. In string theory, the fundamental particles were string–like, rather than point particles; the universe had 10 or 11 dimensions, rather than four; and the theory itself existed at an energy so far from earthly energies that it took a leap of enormous faith to imagine the day when an experiment could ever test it. Quite simply, string theory seemed an excessively esoteric pursuit, which it still is.

Ed Witten on String Theory

Witten's contributions to string theory have been many, including the critical time in 1995 when he gave the field a much–needed boost by showing how the five different variations of the theory then competing with one another actually all belonged under one umbrella (see [Wit95]). In his words, "String theory is an attempt at a deeper description of nature by thinking of an elementary particle not as a little point but as a little loop of vibrating string. One of the basic things about a string is that it can vibrate in many different shapes or forms, which gives music its beauty. If we listen to a tuning fork, it sounds harsh to the human ear. And that's because you hear a pure tone rather than the higher overtones that you get from a piano or violin that give music its richness and beauty. So in the case of one of these strings it can oscillate in many different forms—analogously to the overtones of a piano string. And those different forms of vibration are interpreted as different elementary particles: quarks, electrons, photons. All are different forms of vibration of the same basic string. Unity of the different forces and particles is achieved because they all come from different kinds of vibrations of the same basic string. In the case of string theory, with our present understanding, there would be nothing more basic than the string. It is surprising that replacing the elementary particle with a string leads to such a big change in things. I'm tempted to say that it has to do with the fuzziness it introduces. So the

[14] Witten joined the 'old Green–Schwarz bosonic string community' after he won his Fields Medal for topological quantum field theory (TQFT)

particle is spread out. But it turns out that everything about space–time is a little bit spread out; it's blurred. You have to start doing some calculations to really see it. It's hard to explain it just in words or by drawing pictures. Spreading out the particle into a string is a step in the direction of making everything we're familiar with fuzzy. You enter a completely new world where things aren't at all what you're used to. It's as surprising in its own way as the fuzziness that much of physics acquired in light of quantum mechanics and the Heisenberg uncertainty principle.

In Einstein's general relativity the geometrical structure of space can change but not its topology. Topology is the property of something that doesn't change when you bend it or stretch it as long as you don't break anything. You can imagine a bowling ball and you can imagine a coffee cup that has a handle—the coffee cup is different topologically because there's a handle. Even if you could bend it or stretch it, as long as you don't break it, it's still got that handle, which makes it topologically different. Quantum mechanics brought an unexpected fuzziness into physics because of quantum uncertainty, the Heisenberg uncertainty principle. String theory does so again because a point particle is replaced by a string, which is more spread out. And even though it's a naive statement, it leads in the right direction: in string theory, space–time becomes fuzzy.

String theory requires 10 dimensions of space. Thus, you need the extra dimensions. At first people didn't like them too much, but they've got a big benefit, which is that the ability of string theory to describe all the elementary particles and their forces along with gravity depends on using the extra dimensions. You have that one basic string, but it can vibrate in many ways. But we're trying to get a lot of particles because experimental physicists have discovered a lot of particles. The electron and its heavy cousins the neutrinos, the quarks, photons, gravitons, and so on. There is really a big zoo of elementary particles that you're trying to explain. Having those extra dimensions and therefore many ways the string can vibrate in many different directions turns out to be the key to being able to describe all the particles that we see. The theory has to be interpreted that extra dimensions beyond the ordinary four dimensions the three spatial dimensions plus time are sufficiently small that they haven't been observed yet. So we would hope to test the theory, conceivably directly at accelerators. I suspect that's a long shot. More likely we'll do it indirectly by making more precise calculations about elementary particles based on the existence of extra dimensions. If I take the theory as we have it now, literally, I would conclude that extra dimensions really exist. They're part of nature. We don't really know how big they are yet, but we hope to explore that in various ways. They're beyond our ordinary experience just like atomic nuclei are. On the other hand, we don't understand the theory too completely, and because of this fuzziness of space–time, the very concept of space–time and space–time dimensions isn't precisely defined. I suspect that the fuzziness of space–time will play more of a role in the eventual answer than we understand now... I guess it's possible that string theory could

be wrong. But if it is in fact wrong, it's amazing that it's been so rich and has survived so many brushes with catastrophe and has linked up with the established physical theories in so many ways, providing so many new insights about them. I wouldn't have thought that a wrong theory should lead us to understand better the ordinary quantum field theories or to have new insights about the quantum states of black holes."

Quantum Geometry Framework

To start our review on topological string theory, here we depict a general quantum geometry framework (see e.g., [Wit98]).

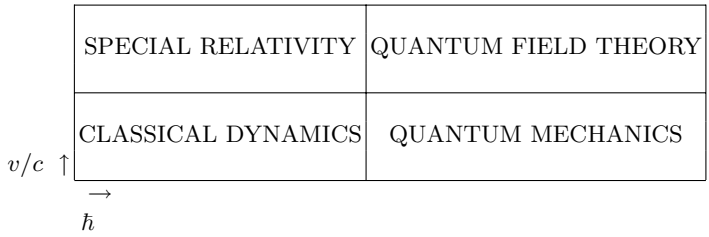


Fig. 2.1. The deformation from classical dynamics to quantum field theory (see text for explanation).

The relationship between non–relativistic classical mechanics and quantum field theory (see [Col88]) can be summarized as in Figure 2.1. We see that the horizontal axis corresponds to the *Planck constant* $\hbar$ (divided by the typical action of the system being studied), while the vertical axis corresponds to v/c, the ration of motion velocity and light velocity.

Similarly, in the superstring theory there are also two relevant expansion parameters, as shown in Figure 2.2. Here we see that the horizontal axis corresponds to the value of the *string coupling constant*, g_s, while the vertical axis corresponds to the value of the dimensionless *sigma model coupling* α'/R^2 with R being a typical radius of a compactified portion of space). In the extreme $\alpha' = g_s = 0$ limit, for instance, we recover relativistic particle dynamics. For nonzero g_s we recover point particle quantum field theory. For $g_s = 0$ and nonzero α' we are studying classical string theory. In general though, we need to understand the theory for arbitrary values of these parameters (see [Gre96]).

Quantum stringy geometry postulates the existence of 6D *Calabi–Yau manifolds* at every point of the space–time (see, e.g., [CHS85]). These curled–up local manifolds transform according to the general *orbifolding* procedure, as will be described below.

	CLASSICAL STRINGY GEOMETRY	QUANTUM STRINGY GEOMETRY
$\alpha'/R^2 \uparrow$	CLASSICAL RIEMANNIAN GEOMETRY	QUANTUM RIEMANNIAN GEOMETRY

$\rightarrow$
g_s

Fig. 2.2. The deformation from classical Riemannian geometry to *quantum stringy geometry* (see text for explanation).

Green–Schwarz Bosonic Strings and Branes

Here, we briefly describe the *world–sheet dynamics* of the *Green–Schwarz bosonic string theory*, and (more generally), bosonic $p-$brane theory, the predecessor of the current superstring theory (see [Sch93, GSW87] for details).

World–Line Description of a Point Particle

Recall that a point particle sweeps out a trajectory called world–line in space–time. This can be described by functions $x^\mu(\tau)$, that describe how the world–line, parameterized by τ, is embedded in the space–time, whose coordinates are denoted $x^\mu\,(\mu = 0, 1, 2, 3)$. For simplicity, let us assume that the space–time is flat Minkowski space with a *Lorentz metric tensor*

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Then, the Lorentz–invariant line element (metric form) is given by

$$ds^2 = -\eta_{\mu\nu} dx^\mu dx^\nu.$$

In normal units ($\hbar = c = 1$), the *action* for a particle of mass m is given by

$$S = -m \int ds.$$

This could be generalized to a curved space–time by replacing $\eta_{\mu\nu}$ by a Riemannian metric tensor $g_{\mu\nu}(x)$, but (for simplicity) we will not do so here. In terms of the embedding functions, $x^\mu(t)$, the action can be rewritten as

$$S[x] = -m \int d\tau \sqrt{-\eta_{\mu\nu} \dot{x}^\mu \dot{x}^\nu},$$

where overdot represents the derivative with respect to τ. An important property of this action is invariance under local reparametrizations. This is a kind of gauge invariance, whose meaning is that the form of S is unchanged under an arbitrary reparametrization of the world–line $\tau \to \tau(\tilde{\tau})$. Actually, one should require that the function $\tau(\tilde{\tau})$ is smooth and monotonic $\left(\frac{d\tau}{d\tilde{\tau}} > 0\right)$. The reparametrization invariance is a 1D analog of the 4D general coordinate invariance of general relativity. Mathematicians refer to this kind of symmetry as diffeomorphism invariance.

The reparametrization invariance of S allows us to choose a gauge. A nice choice is the *static gauge*, $x^0 = \tau$. In this gauge (renaming the parameter to t) the action becomes

$$S = -m \int \sqrt{1 - v_i^2}\, dt, \qquad \text{where} \qquad v_i = \frac{dx_i}{dt}.$$

Requiring this action to be stationary under an arbitrary variation of $x_i(t)$ gives the Euler–Lagrangian equations

$$\frac{dp_i}{dt} = 0, \qquad \text{where} \qquad p_i = \frac{\delta S}{\delta v_i} = \frac{m v_i}{\sqrt{1 - v_i^2}},$$

which is the usual result. So we see that usual relativistic kinematics follows from the action $S = -m \int ds$.

p−Branes and World–Volume Actions

We can now generalize the analysis of the massive point particle to a generic *p−brane*, which is characterized by its *tension* T_p. The action in this case involves the invariant $(p+1)$D volume and is given by

$$S_p = -T_p \int d\mu_{p+1},$$

where the invariant volume element is

$$d\mu_{p+1} = \sqrt{-\det(-\eta_{\mu\nu} \partial_\alpha x^\mu \partial_\beta x^\nu)}\, d^{p+1}\sigma.$$

Here the embedding of the $p-$brane into dD space–time is given by functions $x^\mu(\sigma^\alpha)$. The index $\alpha = 0, \ldots, p$ labels the $p+1$ coordinates σ^α of the $p-$brane world–volume and the index $\mu = 0, \ldots, d-1$ labels the d coordinates x^μ of the dD space–time. We have defined $\partial_\alpha x^\mu = \frac{\partial x^\mu}{\partial \sigma^\alpha}$.The determinant operation acts on the $(p+1) \times (p+1)$ matrix whose rows and columns are labeled by α and β. The tension T_p is interpreted as the mass per unit volume of the $p-$brane. For a $0-$brane, it is just the mass.

Let us now specialize to the string, $p = 1$. Evaluating the determinant gives the *Nambu–Goto action* (see subsection 1.5.5 above)

$$S[x] = -T \int d\sigma d\tau \sqrt{\dot{x}^2 x'^2 - (\dot{x} \cdot x')^2},$$

where we have defined $\sigma^0 = \tau$, $\sigma^1 = \sigma$, and $\dot{x}^\mu = \frac{\partial x^\mu}{\partial \tau}$, $x'^\mu = \frac{\partial x^\mu}{\partial \sigma}$. The above action is equivalent to the action

$$S[x,h] = -\frac{T}{2}\int d^2\sigma\sqrt{-h}h^{\alpha\beta}\eta_{\mu\nu}\partial_\alpha x^\mu \partial_\beta x^\nu, \tag{2.511}$$

where $h_{\alpha\beta}(\sigma,\tau)$ is the world–sheet metric, $h = \det h_{\alpha\beta}$, and $h^{\alpha\beta} = (h_{\alpha\beta})^{-1}$ is the inverse of $h_{\alpha\beta}$. The *Euler–Lagrangian equations* obtained by varying $h^{\alpha\beta}$ are

$$T_{\alpha\beta} = \partial_\alpha x \cdot \partial_\beta x - \frac{1}{2}h_{\alpha\beta}h^{\gamma\delta}\partial_\gamma x \cdot \partial_\delta x = 0.$$

In addition to reparametrization invariance, the action $S[x,h]$ has another local symmetry, called *conformal invariance*, or, *Weyl invariance*. Specifically, it is invariant under the replacement

$$h_{\alpha\beta} \to \Lambda(\sigma,\tau)h_{\alpha\beta}, \qquad x^\mu \to x^\mu.$$

This local symmetry is special to the $p = 1$ case (strings).

The two reparametrization invariance symmetries of $S[x,h]$ allow us to choose a gauge in which the three functions $h_{\alpha\beta}$ (this is a symmetric 2×2 matrix) are expressed in terms of just one function. A convenient choice is the *conformally flat gauge*

$$h_{\alpha\beta} = \eta_{\alpha\beta}\mathrm{e}^{\phi(\sigma,\tau)}.$$

Here, $\eta_{\alpha\beta}$ denoted the 2D Minkowski metric of a flat world–sheet. However, $h_{\alpha\beta}$ is only 'conformally flat', because of the factor e^ϕ. Classically, substitution of this gauge choice into $S[x,h]$ leaves the gauge–fixed action

$$S = \frac{T}{2}\int d^2\sigma\eta^{\alpha\beta}\partial_\alpha x \cdot \partial_\beta x. \tag{2.512}$$

Quantum mechanically, the story is more subtle. Instead of eliminating h via its classical field equations, one should perform a Feynman path integral, using standard machinery to deal with the local symmetries and gauge fixing. When this is done correctly, one finds that in general ϕ does not decouple from the answer. Only for the special case $d = 26$ does the quantum analysis reproduce the formula we have given based on classical reasoning. Otherwise, there are correction terms whose presence can be traced to a conformal anomaly (i.e., a quantum–mechanical breakdown of the conformal invariance).

The gauge–fixed action is quadratic in the x's. Mathematically, it is the same as a theory of d free scalar fields in two dimensions. The equations of motion obtained by varying x^μ are free 2D wave equations:

$$\ddot{x}^\mu - x''^\mu = 0.$$

However, this is not the whole story, because we must also take account of the constraints $T_{\alpha\beta} = 0$, which evaluated in the conformally flat gauge, read

$$T_{01} = T_{10} = \dot{x} \cdot x' = 0, \qquad T_{00} = T_{11} = \frac{1}{2}(\dot{x}^2 + x'^2) = 0.$$

Adding and subtracting gives

$$(\dot{x} \pm x')^2 = 0. \tag{2.513}$$

Boundary Conditions

To go further, one needs to choose boundary conditions. There are three important types. For a closed string one should impose periodicity in the spatial parameter σ. Choosing its range to be π (as is conventional)

$$x^\mu(\sigma, \tau) = x^\mu(\sigma + \pi, \tau).$$

For an open string (which has two ends), each end can be required to satisfy either *Neumann or Dirichlet boundary conditions* for each value of μ,

$$\begin{aligned} \text{Neumann}: \quad & \frac{\partial x^\mu}{\partial \sigma} = 0 \qquad \text{at} \quad \sigma = 0 \quad \text{or} \quad \pi, \\ \text{Dirichlet}: \quad & \frac{\partial x^\mu}{\partial \tau} = 0 \qquad \text{at} \quad \sigma = 0 \quad \text{or} \quad \pi. \end{aligned}$$

The Dirichlet condition can be integrated, and then it specifies a space–time location on which the string ends. The only way this makes sense is if the open string ends on a physical object – it ends on a D–brane.[15] If all the open–string boundary conditions are Neumann, then the ends of the string can be anywhere in the space–time. The modern interpretation is that this means that there are space–time–filling D–branes present.

Let us now consider the closed–string case in more detail. The general solution of the 2d wave equation is given by a sum of 'right–movers' and 'left–movers': $x^\mu(\sigma, \tau) = x^\mu_R(\tau - \sigma) + x^\mu_L(\tau + \sigma)$. These should be subject to the following additional conditions:

- $x^\mu(\sigma, \tau)$ is real,
- $x^\mu(\sigma + \pi, \tau) = x^\mu(\sigma, \tau)$, and
- $(x'_L)^2 = (x'_R)^2 = 0$ (these are the $T_{\alpha\beta} = 0$ constraints in (2.513)).

The first two of these conditions can be solved explicitly in terms of Fourier series:

$$\begin{aligned} x^\mu_R &= \frac{1}{2}x^\mu + \ell_s^2 p^\mu(\tau - \sigma) + \frac{\mathrm{i}}{\sqrt{2}}\ell_s \sum_{n \neq 0} \frac{1}{n} \alpha^\mu_n \mathrm{e}^{-2in(\tau - \sigma)} \\ x^\mu_L &= \frac{1}{2}x^\mu + \ell_s^2 p^\mu(\tau + \sigma) + \frac{\mathrm{i}}{\sqrt{2}}\ell_s \sum_{n \neq 0} \frac{1}{n} \tilde{\alpha}^\mu_n \mathrm{e}^{-2in(\tau - \sigma)}, \end{aligned}$$

[15] D here stands for *Dirichlet*.

where the expansion parameters α_n^μ, $\tilde{\alpha}_n^\mu$ satisfy $\alpha_{-n}^\mu = (\alpha_n^\mu)^\dagger$, $\tilde{\alpha}_{-n}^\mu = (\tilde{\alpha}_n^\mu)^\dagger$.The center–of–mass coordinate x^μ and momentum p^μ are also real. The fundamental string length scale ℓ_s is related to the tension T by

$$T = \frac{1}{2\pi\alpha'}, \quad \alpha' = \ell_s^2.$$

The parameter α' is called the universal Regge slope, since the string modes lie on linear parallel Regge trajectories with this slope.

Canonical Quantization

The analysis of closed–string left–moving modes, closed–string right–moving modes, and open–string modes are all very similar. Therefore, to avoid repetition, we focus on the closed–string right–movers. Starting with the gauge–fixed action in (2.512), the canonical momentum of the string is

$$p^\mu(\sigma,\tau) = \frac{\delta S}{\delta \dot{x}^\mu} = T\dot{x}^\mu.$$

Canonical quantization (this is just free 2d field theory for scalar fields) gives

$$[p^\mu(\sigma,\tau), x^\nu(\sigma',\tau)] = -\mathrm{i}\hbar\eta^{\mu\nu}\delta(\sigma-\sigma').$$

In terms of the Fourier modes (setting $\hbar = 1$) these become

$$[p^\mu, x^\nu] = -\mathrm{i}\eta^{\mu\nu}, \qquad [\alpha_m^\mu, \alpha_n^\nu] = m\delta_{m+n,0}\eta^{\mu\nu}, \qquad [\tilde{\alpha}_m^\mu, \tilde{\alpha}_n^\nu] = m\delta_{m+n,0}\eta^{\mu\nu},$$

and all other commutators vanish.

Recall that a quantum–mechanical harmonic oscillator can be described in terms of raising and lowering operators, usually called $a^\dagger$ and a, which satisfy $[a, a^\dagger] = 1$. We see that, aside from a normalization factor, the expansion coefficients α_{-m}^μ and α_m^μ are raising and lowering operators. There is just one problem. As $\eta^{00} = -1$, the time components are proportional to oscillators with the wrong sign ($[a, a^\dagger] = -1$). This is potentially very bad, because such oscillators create states of negative norm, which could lead to an inconsistent quantum theory (with negative probabilities, etc.). Fortunately, as we will explain, the $T_{\alpha\beta} = 0$ constraints eliminate the negative–norm states from the physical spectrum.

The classical constraint for the right–moving closed–string modes, $(x_R')^2 = 0$, has Fourier components

$$L_m = \frac{T}{2}\int_0^\pi \mathrm{e}^{-2im\sigma}(x_R')^2 d\sigma = \frac{1}{2}\sum_{n=-\infty}^{\infty} \alpha_{m-n}\cdot\alpha_n,$$

which are called *Virasoro operators*. Since α_m^μ does not commute with α_{-m}^μ, L_0 needs to be normal–ordered:

$$L_0 = \frac{1}{2}\alpha_0^2 + \sum_{n=1}^{\infty}\alpha_{-n}\cdot\alpha_n.$$

Here $\alpha_0^\mu = \ell_s p^\mu/\sqrt{2}$, where p^μ is the momentum.

Calabi–Yau Manifolds, Orbifolds and Mirror Symmetry

Calabi–Yau Manifolds

Fundamental geometrical objects in string theory are the so–called *Calabi–Yau manifolds* [Cal57, Yau78]. A Calabi–Yau manifold is a compact *Ricci–flat Kähler manifold* (see subsection 1.2.12 above) with a vanishing first Chern class. A Calabi–Yau manifold of complex dimension n is also called a Calabi–Yau $n-$fold, which is a manifold with an $SU(n)$ holonomyi.e., it admits a global nowhere vanishing holomorphic $(n,0)-$form.

For example, in one complex dimension, the only examples are family of *tori*. Note that the Ricci–flat metric on the torus is actually a flat metric, so that the holonomy is the trivial group SU(1). In particular, 1D *Calabi–Yau manifolds* are also called elliptic curves.

In two complex dimensions, the torus T^4 and the $K3$ surfaces[16] are the only examples. T^4 is sometimes excluded from the classification of being a Calabi–Yau, as its holonomy (again the trivial group) is a proper subgroup of $SU(2)$, instead of being isomorphic to $SU(2)$. On the other hand, the holonomy group of a $K3$ surface is the full $SU(2)$ group, so it may properly be called a Calabi–Yau in 2D.

In three complex dimensions, classification of the possible Calabi–Yau manifolds is an open problem. One example of a 3D Calabi–Yau is the quintic threefold in CP^4.

In string theory, the term *compactification* refers to 'curling up' the extra dimensions (6 in the superstring theory), usually on Calabi–Yau spaces or on orbifolds. The mechanism behind this type of compactification is described by the Kaluza–Klein theory.

In the most conventional superstring models, 10 conjectural dimensions in string theory are supposed to come as 4 of which we are aware, carrying some kind of fibration with fiber dimension 6. Compactification on Calabi–Yau $n-$folds are important because they leave some of the original supersymmetry unbroken. More precisely, compactification on a Calabi–Yau 3−fold (with real dimension 6) leaves one quarter of the original supersymmetry unbroken.

Orbifolds

Recall that in topology, an *orbifold* is a generalization of a manifold, a topological space (called the underlying space) with an orbifold structure. The underlying space locally looks like a quotient of a Euclidean space under the action of a finite group of isometries.

The formal orbifold definition goes along the same lines as a definition of manifold, but instead of taking domains in $\mathbb{R}^n$ as the target spaces of charts

[16] Recall that $K3$ surfaces are compact, complex, simply–connected surfaces, with trivial canonical line bundle, named after three algebraic geometers, Kummer, Kähler and Kodaira. Otherwise, they are *hyperkähler manifolds* of real dimension 4 with $SU(2)$ holonomy.

one should take domains of finite quotients of $\mathbb{R}^n$. A (topological) orbifold O, is a Hausdorff topological space X with a countable base, called the *underlying space*, with an *orbifold structure*, which is defined by orbifold atlas, given as follows.

An *orbifold chart* is an open subset $U \subset X$ together with open set $V \subset \mathbb{R}^n$ and a continuous map $\varphi : U \to V$ which satisfy the following property: there is a finite group Γ acting by linear transformations on V and a homeomorphism $\theta : U \to V/\Gamma$ such that $\varphi = \theta \circ \pi$, where π denotes the projection $V \to V/\Gamma$. A collection of orbifold charts, $\{\varphi_i = U_i \to V_i\}$, is called the *orbifold atlas* if it satisfies the following properties:

(i) $\cup_i U_i = X$;

(ii) if $\varphi_i(x) = \varphi_j(y)$ then there is a neighborhood $x \in V_x \subset V_i$ and $y \in V_y \subset V_j$ as well as a homeomorphism $\psi : V_x \to V_y$ such that $\varphi_i = \varphi_j \circ \psi$.

The orbifold atlas defines the orbifold structure completely and we regard two orbifold atlases of X to give the same orbifold structure if they can be combined to give a larger orbifold atlas. One can add differentiability conditions on the gluing map in the above definition and get a definition of *smooth* (C^∞) *orbifolds* in the same way as it was done for manifolds.

The main example of underlying space is a quotient space of a manifold under the action of a finite group of diffeomorphisms, in particular manifold with boundary carries natural orbifold structure, since it is $\mathbb{Z}_2-$factor of its double. A factor space of a manifold along a smooth S^1-action without fixed points cares an orbifold structure. The orbifold structure gives a natural stratification by open manifolds on its underlying space, where one strata corresponds to a set of singular points of the same type.

Note that one topological space can carry many different orbifold structures. For example, consider the orbifold O associated with a factor space of a $2-$sphere S^2 along a rotation by π. It is homeomorphic to S^2, but the natural orbifold structure is different. It is possible to adopt most of the characteristics of manifolds to orbifolds and these characteristics are usually different from the correspondent characteristics of the underlying space. In the above example, its orbifold fundamental group of O is $\mathbb{Z}_2$ and its *orbifold Euler characteristic* is 1.

Manifold orbifolding denotes an operation of wrapping, or folding in the case of mirrors, to superimpose all equivalent points on the original manifold – to get a new one.

In string theory, the word *orbifold* has a new flavor. In physics, the notion of an orbifold usually describes an object that can be globally written as a coset M/G where M is a manifold (or a theory) and G is a group of its isometries (or symmetries). In string theory, these symmetries do not need to have a geometric interpretation. The so–called *orbifolding* is a general procedure of string theory to derive a new string theory from an old string theory in which the elements of the group G have been identified with the identity. Such a procedure reduces the number of string states because the states must be invariant under G, but it also increases the number of states because of

the extra twisted sectors. The result is usually a new, perfectly smooth string theory.

Mirror Symmetry

The so–called *mirror symmetry* is a surprising relation that can exist between two Calabi–Yau manifolds. It happens, usually for two such 6D manifolds, that the shapes may look very different geometrically, but nevertheless they are equivalent if they are employed as hidden dimensions of a (super)string theory. More specifically, mirror symmetry relates two manifolds M and W whose *Hodge numbers* $h^{1,1} = \dim H^{1,1}$ and $h^{1,2} = \dim H^{1,2}$ are swapped; string theory compactified on these two manifolds leads to identical physical phenomena (see [Gre00]).

[Str90] showed that mirror symmetry is a special example of the so–called $T-$duality: the Calabi–Yau manifold may be written as a fiber bundle whose fiber is a 3D torus $T^3 = S^1 \times S^1 \times S^1$. The simultaneous action of $T-$duality on all three dimensions of this torus is equivalent to mirror symmetry.

Mirror symmetry allowed the physicists to calculate many quantities that seemed virtually incalculable before, by invoking the 'mirror' description of a given physical situation, which can be often much easier. Mirror symmetry has also become a very powerful tool in mathematics, and although mathematicians have proved many rigorous theorems based on the physicists' intuition, a full mathematical understanding of the phenomenon of mirror symmetry is still lacking.

Topological Field Theories

Unfortunately, there is no such thing as a crash course in string theory, but the necessary background can be found in the classic two–volume monographs [GSW87] and [Pol98]. A good introduction to conformal field theory is given in the lecture notes [Sch96]. The basics of topological string theory were laid out in a series of beautiful papers by E. Witten in 1990s [Wit88a, Wit89, Wit88c, Wit88b, Wit90, Wit91, Wit92, Wit95, Wit91], and more or less completed in a seminal paper [BCO94]. Reviews about topological *string* theory, usually also contain a discussion of topological field theory. The first one of these is the review [DVV91] from the same period. However, if we want to dig even deeper, there is the 900–page book [HKK03], which discusses topological string theory from the point of view of mirror symmetry.

In particular, there exists a mathematically rigorous, axiomatic definition of topological field theories due to [Ati88]. Instead of giving this definition, we will define topological field theory in a more physically–intuitive way, but as a result somewhat less rigorous way.

Recall (from subsection 1.5.5 above) that the output of a QFT is given by its observables: correlation functions of products of operators,

$$\langle \mathcal{O}_1(x_1) \cdots \mathcal{O}_n(x_n) \rangle_b. \tag{2.514}$$

Here, the $\mathcal{O}_i(x)$ are *physical* operators of the theory. What one calls 'physical' is part of the definition of the theory, but it is important to realize that in general not all combinations of fields are viewed as physical operators. For example, in a gauge theory, we usually require the observables to arise from gauge–invariant operators. That is, $\mathrm{Tr}\, F$ would be one of the $\mathcal{O}_i$, but $\mathrm{Tr}\, A$ or A itself would not.

The subscript b in the above formula serves as a reminder that the correlation function is usually calculated in a certain background. That is, the definition of the theory may involve a choice of a Riemannian manifold M on which the theory lives, it may involve choosing a metric on M, it may involve choosing certain coupling constants, and so on.

The definition of a topological field theory is now as follows. Suppose that we have a quantum field theory where the background choices involve a choice of manifold M and a choice of metric h on M. Then the theory is called a *topological field theory* if the observables (2.514) do *not* depend on the choice of metric h. Let us stress that it is part of the definition that h is a background field – in particular, we do *not* integrate over h in the path integral. One may wonder what happens if, once we have a topological field theory, we do make the metric h dynamical and integrate over it. This is exactly what we will do once we start considering topological *string* theories.

Note that the word 'topological' in the definition may be somewhat of a misnomer [Von05]. The reason is that the above definition does not strictly imply that the observables depend only on the topology of M – there may be other background choices hidden in b on which they depend as well. For example, in the case of a complex manifold M, correlation functions will in general not only depend on the topology of M and its metric, but also on our specific way of combining the $2d$ real coordinates on M into d complex ones. This choice, a *complex structure*, is part of the background of the quantum field theory, and correlation functions in a topological field theory will in general still depend on it.

If our quantum field theory has general coordinate invariance, as we will usually assume to be the case, then the above definition has an interesting consequence. The reason is that in such a case we can do an arbitrary general coordinate transformation, changing both the coordinates on M and its metric, under which the correlation functions should be invariant. Then, using the topological invariance, we can transform back the metric to its old value. The combined effect is that we have only changed the x_i in (2.514). That is, in a generally coordinate invariant topological field theory, the observables do not depend on the insertion points of the operators.

Chern Classes

Inspired by the identification of a connection with a gauge field, let us consider the analogue of the *non–Abelian field strength*,

$$F = dA - A \wedge A,$$

where $A \wedge A$ is a shorthand for $A^I_{Ji} A^J_{Kj}\, dx^i \wedge dx^j$ (note that $A \wedge A \neq 0$). A short calculation shows that on the overlap of two patches of M (or equivalently, under a gauge transformation), this quantity transforms as

$$F_{(b)} = \Lambda_{(ba)} F_{(a)} \Lambda^{-1}_{(ba)},$$

from which we see that F can be viewed as a section of a true vector bundle of Lie algebra valued 2–forms. In particular, we can take its trace and obtain a genuine 2–form:

$$c_1 = \frac{\mathrm{i}}{2\pi} \operatorname{Tr}(F),$$

where the prefactor is convention. It can be seen that this 2–form is closed:

$$d(\operatorname{Tr}(F)) = \operatorname{Tr}(dF) = -\operatorname{Tr}(d(A \wedge A)) = -\operatorname{Tr}(dA \wedge A - A \wedge dA) = 0.$$

Therefore, we can take its cohomology class, for which we would like to argue that it is a topological invariant. Note that this construction is independent of the choice of coordinates on M. Moreover, it is independent of gauge transformations. However, on a general vector bundle there may be connections which cannot be reached in this way from a given connection. Changing to such a connection is called a 'large gauge transformation', and from what we have said it is not clear a priori that the Chern classes do not depend on this choice of equivalence class of connections. However, with some work we can also prove this fact. The invariant $[c_1]$ is called the *first Chern class*. In fact, it might be better to call it a 'relative topological invariant': given a base manifold M of fixed topology, we can topologically distinguish vector bundles over it by calculating the above cohomology class.

By taking the trace of F, we loose a lot of information. There turns out to be a lot more topological information in F, and it can be extracted by considering the expression

$$c(F) = \det\left(1 + \frac{\mathrm{i}F}{2\pi}\right),$$

where 1 is the identity matrix of the same size as the elements of the Lie algebra of G. Again, it can be checked that this expression is invariant under a change of coordinates for M and under a change of connection. Since the matrix components inside the determinant consist of the 0–form 1 and the 2–form F, expanding the determinant will lead to an expression consisting of forms of all even degrees. One writes this as

$$c(F) = c_0(F) + c_1(F) + c_2(F) + \ldots$$

The sum terminates either at the highest degree encountered in expanding the determinant, or at the highest allowed even form on M. Note that $c_0(F) = 1$, and $[c_1(F)]$ is exactly the first Chern class we defined above. The cohomology class of c_n is called the *nth Chern class*.

As an almost trivial example, let us consider the case of a product bundle $M \times W$. In this case there is a global section $g(x)$ of the principal bundle P, and we can use this to construct a connection $A = -gdg^{-1}$, so that [Von05]

$$F = -dgdg^{-1} - gdg^{-1}gdg^{-1} = -dgdg^{-1} + dgdg^{-1} = 0.$$

Thus, for a trivial bundle with this connection, $c_0 = 1$ and $c_n = 0$ for all $n > 0$.

Chern–Simons theory

The easiest way to construct a topological field theory is to construct a theory where both the *action* S (or, *quantum measure* $\mathrm{e}^{\mathrm{i}S}$) and the fields do not include the metric at all. Such topological field theories are called 'Schwarz–type' topological field theories. This may sound like a trivial solution to the problem, but nevertheless it can lead to quite interesting results. To see this, let us consider familiar example: *Chern–Simons gauge theory* on a 3D manifold M – now from a physical point of view.

Recall from subsection 2.2.5 that Chern–Simons theory is a gauge theory – that is, it is constructed from a vector bundle E over the base space M, with a structure group (gauge group) G and a connection (gauge field) A. The Lagrangian of Chern–Simons theory is then given by

$$L = \mathrm{Tr}(A \wedge dA - \frac{2}{3} A \wedge A \wedge A).$$

It is a straightforward exercise to check how this Lagrangian changes under the gauge transformation

$$\tilde{A} = gAg^{-1} - gdg^{-1},$$

and one finds

$$\begin{aligned}\tilde{L} &\equiv \mathrm{Tr}(\tilde{A} \wedge d\tilde{A} - \frac{2}{3}\tilde{A} \wedge \tilde{A} \wedge \tilde{A}) \\ &= \mathrm{Tr}(A \wedge dA - \frac{2}{3} A \wedge A \wedge A) - d\,\mathrm{Tr}(gA \wedge dg^{-1}) + \frac{1}{3}\,\mathrm{Tr}(gdg^{-1} \wedge dg \wedge dg^{-1}).\end{aligned}$$

The second term is a total derivative, so if M does not have a boundary, the action, being the integral of L over M, does not get a contribution from this term. The last term is not a total derivative, but its integral turns out to be a topological invariant of the map $g(x)$, which is quantized as

$$\frac{1}{24\pi^2} \int_M \mathrm{Tr}(gdg^{-1} \wedge dg \wedge dg^{-1}) = m \in \mathbb{Z}.$$

From this, we see that if we define the *action* as

$$S = \frac{k}{4\pi} \int_M L,$$

with k an integer, the action changes by $2\pi km$ under gauge transformations, and the quantum measure $\mathrm{e}^{\mathrm{i}S}$ is invariant.

So from this discussion we seem to arrive at the conclusion that the *partition function*

$$Z = \int \mathcal{D}[A]\, \mathrm{e}^{\mathrm{i}S[A]}$$

for a line bundle of a fixed topology E is a topological invariant of M, as are the correlation functions of gauge–invariant operators such as $\mathrm{Tr}\, F$. However, there is one more detail we have to worry about: there may be an anomaly in the quantum theory. That is, it may not be possible to define the path integral measure $\mathcal{D}[A]$ in a gauge–invariant way.

One way to see what problems can arise is to note that to actually calculate the path integral, one has to pick a gauge condition on A. That is, we have to pick one representative of A in each equivalence class under gauge transformations. To make such a choice will in general require a choice of metric. For example, from electromagnetism (where E is a 1D complex line bundle and $G = U(1)$) we know that a useful gauge is the Feynman gauge, in which the equation of motion for A becomes

$$\Delta A = 0.$$

As we have seen before, the Laplacian Δ is an operator which, through the Hodge star, depends on the metric, and hence the results we find will a priori be metric dependent. To show that the results are truly metric independent, one needs to show that the quantum results do not depend on our arbitrary choice of gauge.

We will not go into the details of this, but state that one can show that Chern–Simons theory on a compact 3–manifold is anomaly–free, so our naive argument above was correct, and one can indeed calculate topological invariants of M in this way.

Let us briefly discuss the kind of topological invariants that Chern–Simons theory can lead to. Recall that one can construct a Lie group element g from a Lie algebra element $\mathcal{A}$ as $g = \mathrm{e}^{\mathcal{A}}$. Now suppose we have a path $\gamma(t)$ inside M. Suppose that we chop up γ into very small line elements given by tangent vectors $\dot{\gamma}\delta t$. Then we can insert this tangent vector into the connection 1–form A, and get a Lie algebra element. As we have seen, it is precisely this Lie algebra element which transports vectors in E along this small distance: we have to multiply these vectors by $1 + \mathcal{A}$. This is a linear approximation to the finite transformation $\mathrm{e}^{\mathcal{A}}$. So if we transport a vector along the entire closed curve γ, it will return multiplied by a group element

$$g = \lim_{\delta t \to 0} \left[\exp\left(A\left(\dot{\gamma}(0)\right)\delta t\right) \exp\left(A\left(\dot{\gamma}(\delta t)\right)\delta t\right) \exp\left(A\left(\dot{\gamma}(2\delta t)\right)\delta t\right) \cdots \right].$$

Now it is tempting to add all the exponents and write their sum in the limit as an integral, but this is not quite allowed since the different group elements may not commute, so $\mathrm{e}^X \mathrm{e}^Y \neq \mathrm{e}^{X+Y}$. Therefore, one uses the following notation,

$$g = P \exp \int_\gamma A,$$

where P stands for path ordering, while the element g is called the *holonomy* of A around the closed curve γ. An interesting gauge and metric independent object turns out to be the trace of this group element. This trace is called the *Wilson loop* $W_\gamma(A)$, given by

$$W_\gamma(A) = \mathrm{Tr}(P) \exp \int_\gamma A.$$

The topological invariants we are interested in are now the correlation functions of such Wilson loops in Chern–Simons theory. Since these correlation functions are independent of the parametrization of M, we can equivalently say that they will be independent of the precise location of the loop γ; we have in fact constructed a topological invariant of the embedding of γ inside M. This embedding takes the shape of a knot, so the invariants we have constructed are knot invariants. One can show that the invariants are actually polynomials in the variable $y = \exp 2\pi i/(k+2)$, where k is the integer 'coupling constant' of the Chern–Simons theory.

The above construction is due to E. Witten, and was carried out in [Wit89]. Before Witten's work, several polynomial invariants of knots were known, one of the simplest ones being the so–called *Jones polynomial*. It can be shown that many of these polynomials arise as special cases of the above construction, where one takes a certain structure group G, $SU(2)$ for the Jones polynomial, and a certain vector bundle (representation) E, the fundamental representation for the Jones polynomial. That is, using this 'trivial' topological field theory, Witten was able to reproduce a large number of the known knot invariants in a unified framework, and construct a great number of new invariants as well [Von05].

Cohomological Field Theories

Even though the above example leads to quite interesting topological invariants, the construction itself is somewhat trivial: given the absence of anomalies that we mentioned, the independence of the metric is completely manifest throughout the procedure. There exists a different way of constructing topological field theories in which the definition of the theory *does* use a metric, but one can still show that the partition function and the physical correlation functions of the theory are metric–independent. The theories constructed in this way are called topological theories of 'Witten–type', or cohomological field theories.

Cohomological field theories are field theories that possess a very special type of symmetry. Recall that from Noether's theorem a global symmetry of a theory leads to a conserved charge S, and that after quantizing the theory, the symmetry is generated by the corresponding operator:

$$\delta_\epsilon \mathcal{O}_i = \mathrm{i}\epsilon[S, \mathcal{O}_i], \qquad \text{or} \qquad \delta_\epsilon \mathcal{O}_i = \mathrm{i}\epsilon\{S, \mathcal{O}_i\},$$

depending on whether S and $\mathcal{O}_i$ are fermionic or bosonic. Furthermore, the symmetry–invariant states $|j\rangle$ satisfy

$$S\,|j\rangle = 0. \tag{2.515}$$

In particular, if the symmetry is not spontaneously broken, the vacuum of such a theory will be symmetric, i.e., $S\,|0\rangle = 0$, and expectation values of operators will be unchanged after a symmetry transformation:

$$\langle 0|\mathcal{O}_i + \delta\mathcal{O}_i|0\rangle = \langle 0|\mathcal{O}_i|0\rangle + \mathrm{i}\epsilon\langle 0|S\mathcal{O}_i \pm \mathcal{O}_i S|0\rangle = \langle 0|\mathcal{O}_i|0\rangle,$$

since S annihilates the vacuum. Note that, to linear order in a 'small parameter' ϵ, the second term in the first line (with $|0\rangle$ replaced by an arbitrary state $|\psi\rangle$) can also be obtained if instead of on the operators, we let the symmetry operator S act on the *state* as

$$|\psi\rangle \to |\psi\rangle + \mathrm{i}\epsilon S|\psi\rangle. \tag{2.516}$$

In case S is the Hamiltonian, this is the infinitesimal version of the well–known transition between the Schrödinger and Heisenberg pictures. Note that the equations (2.515) and (2.516) already contain some flavor of cohomology. Cohomological field theories are theories where this analogy can be made exact.

With this in mind, the first property of a cohomological field theory will not come as a surprise: it should contain a fermionic symmetry operator Q which squares to zero, $Q^2 = 0$. This may seem like a strange requirement for a field theory, but symmetries of this type occur for example when we have a gauge symmetry and fix it by using the Faddeev–Popov procedure; the resulting theory will then have a global *BRST–symmetry*, which satisfies precisely this constraint. Another example is found in supersymmetry, where one also encounters symmetry operators that square to zero, as we will see in detail later on.

The second property a cohomological field theory should have is really a definition: we define the physical operators in this theory to be the operators that are closed under the action of this $Q-$operator[17]:

$$\{Q, \mathcal{O}_i\} = 0. \tag{2.517}$$

Again, this may seem to be a strange requirement for a physical theory, but again it naturally appears in BRST quantization, and for example in conformal field theories, where we have a 1–1 correspondence between operators and states. In such theories, the symmetry requirement (2.515) on the states translates into the requirement (2.517) on the operators.

[17] From now on, we denote both the commutator and the anti-commutator by curly brackets, unless it is clear that one of the operators inside the brackets is bosonic.

Thirdly, we want to have a theory in which the $Q-$symmetry is not spontaneously broken, so the vacuum is symmetric. Note that this implies the equivalence

$$\mathcal{O}_i \sim \mathcal{O}_i + \{Q, \Lambda\}. \tag{2.518}$$

The reason for this is that the expectation value of an operator product involving a $Q-$exact operator $\{Q, \Lambda\}$ takes the form

$$\langle 0|\mathcal{O}_{i_1} \cdots \mathcal{O}_{i_j}\{Q, \Lambda\}\mathcal{O}_{i_{j+1}} \cdots \mathcal{O}_{i_n}|0\rangle = \langle 0|\mathcal{O}_{i_1} \cdots \mathcal{O}_{i_j}(Q\Lambda - \Lambda Q)\mathcal{O}_{i_{j+1}} \cdots \mathcal{O}_{i_n}|0\rangle,$$

and each term vanishes separately, e.g.,

$$\begin{aligned}\langle 0|\mathcal{O}_{i_1} \cdots \mathcal{O}_{i_j} Q\Lambda\mathcal{O}_{i_{j+1}} \cdots \mathcal{O}_{i_n}|0\rangle &= \pm\langle 0|\mathcal{O}_{i_1} \cdots Q\mathcal{O}_{i_j}\Lambda\mathcal{O}_{i_{j+1}} \cdots \mathcal{O}_{i_n}|0\rangle \\ &= \pm\langle 0|Q\mathcal{O}_{i_1} \cdots \mathcal{O}_{i_j}\Lambda\mathcal{O}_{i_{j+1}} \cdots \mathcal{O}_{i_n}|0\rangle = 0\,,\end{aligned} \tag{2.519}$$

where we made repeated use of (2.517). Together, (2.517) and (2.518) mean that our physical operators are $Q-$cohomology classes.

The fourth and final requirement for a cohomological field theory is that the *metric SEM–tensor* is $Q-$exact,

$$T_{\alpha\beta} \equiv \frac{\delta S}{\delta h^{\alpha\beta}} = \{Q, G_{\alpha\beta}\} \tag{2.520}$$

for some operator $G_{\alpha\beta}$. The physical interpretation of this is the following. The integrals of the components $T_{0\alpha}$ over a space–like hypersurface are conserved quantities. For example, the integral of T_{00} gives the Hamiltonian:

$$H = \int_{space} T_{00}.$$

Similarly, T_{0a} for $a \neq 0$ give the momentum charges. Certainly, the Hamiltonian H should commute with all symmetry operators of the theory, and one usually takes the other space–time symmetries to commute with the internal symmetries as well. The choice of the first lower index 0 here is related to a choice of Lorentz frame, so in general the integrals of all $T_{\alpha\beta}$ will commute with the internal symmetries. In a local theory, it is then natural to assume that also the *densities* commute with Q. However, (2.520) is an even stronger requirement: the SEM densities should not only commute with Q (that is, be $Q-$closed), but they have to do so in a trivial way (they should be $Q-$commutators, that is, $Q-$exact) for the theory to be cohomological.

This fourth requirement is the crucial one in showing the topological invariance of the theory. Let us consider the functional $h^{\alpha\beta}-$derivative of an observable:

$$\begin{aligned}\frac{\delta}{\delta h^{\alpha\beta}}\langle \mathcal{O}_{i_1} \cdots \mathcal{O}_{i_n}\rangle &= \frac{\delta}{\delta h^{\alpha\beta}}\left(\int D\phi\, \mathcal{O}_{i_1} \cdots \mathcal{O}_{i_n} \mathrm{e}^{\mathrm{i}S[\phi]}\right) \\ = \mathrm{i}\int D\phi\, \mathcal{O}_{i_1} \cdots \mathcal{O}_{i_n} \frac{\delta S}{\delta h^{\alpha\beta}} \mathrm{e}^{\mathrm{i}S[\phi]} &= \mathrm{i}\langle \mathcal{O}_{i_1} \cdots \mathcal{O}_{i_n}\{Q, G_{\alpha\beta}\}\rangle = 0,\end{aligned}$$

where in the last line we used the same argument as in (2.519). One might be worried about the operator ordering in going from the operator formalism to the path integral formalism and back. We should really have inserted a time ordering operator on the r.h.s. in the first step above. However, the result then shows us that we can arbitrarily change the metric – and hence the time ordering of the operators, so with hindsight we may actually think of the operators as being arbitrarily ordered. Finally, we have assumed that our operators do not depend explicitly on the metric.

A very practical way to ensure (2.520) is to use a Lagrangian which itself is $Q-$exact, $L = \{Q, V\}$, for some operator V. This choice has an extra virtue, which we can see if we explicitly include Planck's constant in our description: the quantum measure then reads

$$\exp \frac{\mathrm{i}}{\hbar} \left\{ Q, \int_M V \right\}.$$

Then, we can use exactly the same argument as before to show that

$$\frac{d}{d\hbar} \langle \mathcal{O}_{i_1} \cdots \mathcal{O}_{i_n} \rangle = 0.$$

That is, the correlators we are interested in are independent of $\hbar$, and we can therefore calculate them *exactly* in the classical limit.

Descent equations

An important property of cohomological field theories is that, given a scalar physical operator on M – where by 'scalar' we mean an operator that does not transform under coordinate transformations of M, so in particular it has no $\alpha-$indices – we can construct further operators which behave like $p-$forms on M. The basic observation is that we can integrate (2.520) over a spatial hypersurface to get a similar relation for the momentum operators:

$$P_\alpha = \{Q, G_\alpha\},$$

where G_α is a fermionic operator. Now consider the operator

$$\mathcal{O}^{(1)}_\alpha = \mathrm{i}\{G_\alpha, \mathcal{O}^{(0)}\},$$

where $\mathcal{O}^{(0)}(x)$ is a scalar physical operator: $\{Q, \mathcal{O}^{(0)}(x)\} = 0$. Let us calculate

$$\begin{aligned} \frac{d}{dx^\alpha} \mathcal{O}^{(0)} &= \mathrm{i}[P_\alpha, \mathcal{O}^{(0)}] = [\{Q, G_\alpha\}, \mathcal{O}^{(0)}] \\ &= \pm \mathrm{i}\{\{G_\alpha, \mathcal{O}^{(0)}\}, Q\} - \mathrm{i}\{\{\mathcal{O}^{(0)}, Q\}, G_\alpha\} = \{Q, \mathcal{O}^{(1)}_\alpha\}. \end{aligned}$$

In going from the second to the third line, we have used the Jacobi identity. The first sign in the third line depends on whether $\mathcal{O}^{(0)}$ is bosonic or fermionic, but there is no sign ambiguity in the last line. By defining the 1–form operator

$$\mathcal{O}^{(1)} = \mathcal{O}^{(1)}_\alpha dx^\alpha$$

we can write this result as

$$d\mathcal{O}^{(0)} = \{Q, \mathcal{O}^{(1)}\}.$$

Then, we can integrate this equation over a closed curve $\gamma \subset M$ to find

$$\{Q, \int_\gamma \mathcal{O}^{(1)}\} = 0.$$

That is, by constructing a $\int_\gamma \mathcal{O}^{(1)}$ for each $\mathcal{O}^{(0)}$, we have found a whole class of new, non–local, physical operators.

The above procedure can now be repeated in exactly the same way starting from $\mathcal{O}^{(1)}$, and doing this we find a whole tower of $p-$form operators:

$$\begin{aligned} \{Q, \mathcal{O}^{(0)}\} = 0, \qquad & \{Q, \mathcal{O}^{(1)}\} = d\mathcal{O}^{(0)}, \qquad \{Q, \mathcal{O}^{(2)}\} = d\mathcal{O}^{(1)}, \\ \cdots, \qquad & \{Q, \mathcal{O}^{(n)}\} = d\mathcal{O}^{(n-1)}, \qquad 0 = d\mathcal{O}^{(n)}. \end{aligned}$$

The last equation is trivial, since there are no $(n+1)-$forms on an nD smooth manifold.

Following the same reasoning, the integral of $\mathcal{O}^{(p)}$ over a pD submanifold of M is now a physical operator. This gives us a large class of new physical operators, starting from the scalar ones. Note that these operators, being integrated over a submanifold, are inherently nonlocal. Nevertheless, they can have a very physical interpretation. Particularly important examples of this are the 'top–form' operators $\mathcal{O}^{(n)}$ that can be integrated over the whole manifold, leading to

$$\left\{Q, \int_M \mathcal{O}^{(n)}\right\} = 0.$$

This implies that we can add terms $t^a \mathcal{O}_a^{(n)}$, with t^a arbitrary coupling constants, to our Lagrangian without spoiling the fact that the theory is cohomological. These deformations of the theory will be important to us later.

2D Cohomological Field Theories

Since string theories are 2D field theories, we will in particular be interested in cohomological field theories in two dimensions. These theories have some extra properties which will be important in our discussion. Let us begin by reminding the reader of the relation between states in the operator formalism of quantum field theories, and boundary conditions in the path–integral formalism. In its simplest form, this relation looks like

$$\int_{BC1}^{BC2} D\phi \cdots \mathrm{e}^{\mathrm{i}S[\phi]} = \langle BC1 | \, T(\cdots) \, | BC2 \rangle. \tag{2.521}$$

Here, we included a time–ordering operator for completeness, but as we have stated before, for the theories we are interested in this ordering is irrelevant. The notation $|BCi\rangle$ indicates the state corresponding to the incoming

or outgoing boundary condition. For example, if on the path integral side we prescribe all fields at a certain initial time, $\phi(t = t_1) = f(t_1)$, on the operator side this corresponds to the incoming state satisfying

$$\phi(t_1)|BC1\rangle = f(t_1)|BC1\rangle. \tag{2.522}$$

where on the l.h.s. we have an operator acting, but on the r.h.s. there is a simple scalar multiplication. More generally, in the operator formalism we can have linear combinations of states of the type (2.522). Therefore, we should allow for linear combinations on the l.h.s. of (2.521) as well. In other words, in the path–integral formalism, *a state is an operator which adds a number (a weight) to each possible boundary condition on the fields.* From this point of view, the states in (2.522) are like 'delta–functionals': they assign weight 1 to the boundary condition $\phi(t_1) = f(t_1)$, and weight 0 to all other boundary conditions.

Now, let us specialize to 2D field theories. Here, the boundary of a compact manifold is a set of circles. Let us for simplicity assume that the 'incoming' boundary consists of a single circle. We can now define a state in the above sense by doing a path integral over a second surface with the topology of a hemisphere. This path integral gives a number for every boundary condition of the fields on the circle, and this is exactly what a state in the path–integral formalism should do. In particular, one can use this procedure to define a state corresponding to every operator $\mathcal{O}_a$ by inserting $\mathcal{O}_a$ on the hemisphere and then stretching this hemisphere to infinite size. An expectation value in the operator formalism, such as

$$\langle \mathcal{O}_a | \mathcal{O}_b(x_2) \mathcal{O}_c(x_3) | \mathcal{O}_d \rangle_{cyl}, \tag{2.523}$$

on a cylinder of finite length, can then schematically be drawn as in Figure 2.3. Here, instead of first doing the path integrals over the semi–infinite hemispheres and inserting the result in the path integral over the cylinder, one can just as well integrate over the whole surface at once. However, note that in topological field theories, there is no need to do the stretching, since the path integral only depends on the topology of the surface, and hence not on the size of the hemisphere [Von05].

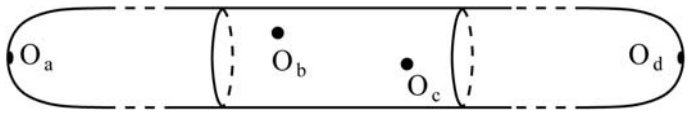


Fig. 2.3. A graphical representation of the correlation function (2.523).

We assume that all states that we are interested in are of the above form, which in particular means that we will integrate only over Riemann surfaces

without boundary. Moreover, we assume these surfaces to be orientable. An important property of topological field theories in two dimensions is now that its correlation functions factorize in the following way:

$$\langle \mathcal{O}_1 \cdots \mathcal{O}_n \rangle_\Sigma = \sum_{a,b} \langle \mathcal{O}_1 \cdots \mathcal{O}_i \mathcal{O}_a \rangle_{\Sigma_1} \eta^{ab} \langle \mathcal{O}_b \mathcal{O}_{i+1} \cdots \mathcal{O}_n \rangle_{\Sigma_2}, \tag{2.524}$$

where the genus of Σ is the sum of the genera of Σ_1 and Σ_2. This statement is explained in Figure 2.4. By using the topological invariance, we can deform a Riemann surface Σ with a set of operator insertions in such a way that it develops a long tube. From general quantum field theory, we know that if we stretch this tube long enough, only the asymptotic states – that is, the states in the physical part of the Hilbert space – will propagate. But as we have just argued, instead of inserting these asymptotic states, we may just as well insert the corresponding operators at a finite distance. However, to conclude that this leads to (2.524), we have to show that this definition of 'physical states' – being the ones that need to be inserted as asymptotic states – agrees with our previous definition in terms of $Q-$cohomology. Let us argue that it does by first writing the factorization as

$$\langle \mathcal{O}_1 \cdots \mathcal{O}_n \rangle_\Sigma = \sum_{A,B} \langle \mathcal{O}_1 \cdots \mathcal{O}_i | \mathcal{O}_A \rangle_{\Sigma_1} \eta^{AB} \langle \mathcal{O}_B | \mathcal{O}_{i+1} \cdots \mathcal{O}_n \rangle_{\Sigma_2}. \tag{2.525}$$

where the $\mathcal{O}_A$ with capital index now correspond to a *complete* basis of asymptotic states in the Hilbert space. The reader may be more used to this type of expressions in the case where $\eta_{AB} = \delta_{AB}$, but since we have not shown that with our definitions $\langle \mathcal{O}_A | \mathcal{O}_B \rangle = \delta_{AB}$, we have to work with this more general form of the identity operator, where η is a metric that we will determine in a moment.

Now, we can write the Hilbert space as a direct sum, $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, where $\mathcal{H}_0$ consists of states $|\psi\rangle$ for which $Q|\psi\rangle = 0$ and $\mathcal{H}_1$ is its orthogonal complement. Since

$$Q\left(\mathcal{O}_1 \ldots \mathcal{O}_i |0\rangle\right) = 0, \tag{2.526}$$

the states in $\mathcal{H}_1$ are in particular orthogonal to states of the form $\mathcal{O}_1 \ldots \mathcal{O}_i |0\rangle$, and hence the states in $\mathcal{H}_1$ do not contribute to the sum in (2.525). Moreover, changing $\mathcal{O}_A$ to $\mathcal{O}_A + \{Q, \Lambda\}$ does not change the result in (2.525), so we only need to sum over a basis of $\mathcal{H}_0 / \Im(\{Q, \cdot\})$, which is exactly the space $\mathcal{H}_{phys}$ of 'topologically physical' states.

Finally, let us determine the metric η^{ab}. We can deduce its form by factorizing the 2–point function

$$C_{ab} = \langle \mathcal{O}_a \mathcal{O}_b \rangle \tag{2.527}$$

in the above way, resulting in

$$C_{ab} = C_{ac} \eta^{cd} C_{db}. \tag{2.528}$$

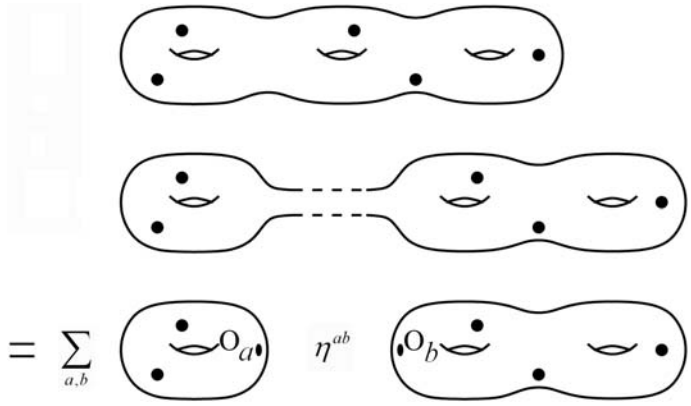


Fig. 2.4. A correlation function on a Riemann surface factorizes into correlation functions on two Riemann surfaces of lower genus (adapted from [Von05]).

In other words, we find that the metric η^{ab} is the matrix inverse of the 2–point function C_{ab}, which for this reason we will write as η_{ab} from now on.

One can apply a similar procedure to 'cut open' internal loops in a Riemann surface Σ, so that we get

$$\langle \mathcal{O}_1 \cdots \mathcal{O}_n \rangle_\Sigma = (-1)^{F_a} \eta^{ab} \sum_{a,b} \langle \mathcal{O}_a \mathcal{O}_b \mathcal{O}_1 \cdots \mathcal{O}_n \rangle_{\Sigma'}, \tag{2.529}$$

where the genus of Σ' is one less than the genus of Σ. The factor $(-1)^{F_a}$ multiplies the expression on the r.h.s. by -1 if the inserted operator $\mathcal{O}_a$ (and hence also $\mathcal{O}_b$) is fermionic. Proving that it needs to be included is not straightforward, but one can think of it as the 'stringy version' of the well–known statement from quantum field theory that fermion loops add an extra minus sign to a Feynman diagram.

The reader should convince himself that together, the equations (2.524) and (2.529) imply that we can reduce all $n-$point correlation functions to products of 3–point functions on the sphere. We denote these important quantities by

$$c_{abc} \equiv \langle \mathcal{O}_a \mathcal{O}_b \mathcal{O}_c \rangle_0, \tag{2.530}$$

where the label 0 denotes the genus of the sphere. By using (2.524) to separate two insertion points on a sphere, we see that

$$\langle \cdots \mathcal{O}_a \mathcal{O}_b \cdots \rangle_\Sigma = \sum_{c,d} \langle \cdots \mathcal{O}_c \cdots \rangle_\Sigma \, \eta^{cd} c_{abd} = \sum_c \langle \cdots \mathcal{O}_c \cdots \rangle_\Sigma \, c_{ab}{}^c,$$

where we raised an index of c_{abc} with the metric η. We can view the above result as the definition of an operator algebra with structure constants $c_{ab}{}^c$:

$$\mathcal{O}_a \mathcal{O}_b = \sum_c c_{ab}{}^c \mathcal{O}_c. \tag{2.531}$$

From the metric η_{ab} and the structure constants $c_{ab}{}^{c}$, we can now calculate any correlation function in the cohomological field theory.

Topological Strings

The 2D field theories we have constructed are already very similar to string theories. However, one ingredient from string theory is missing: in string theory, the world–sheet theory does not only involve a path integral over the maps ϕ^i to the target space and their fermionic partners, but also a path integral over the world–sheet metric $h_{\alpha\beta}$. So far, we have set this metric to a fixed background value.

We have also encountered a drawback of our construction. Even though the theories we have found can give us some interesting 'semi–topological' information about the target spaces, one would like to be able to define general nonzero $n-$point functions at genus g instead of just the partition function at genus one and the particular correlation functions we calculated at genus zero.

It turns out that these two remarks are intimately related. In this section we will go from topological field theory to topological string theory by introducing integrals over all metrics, and in doing so we will find interesting nonzero correlation functions at any genus (see [Von05]).

Coupling to Topological Gravity

In coupling an ordinary field theory to gravity, one has to perform three steps.

- First of all, one rewrites the Lagrangian of the theory in a covariant way by replacing all the flat metrics by the dynamical ones, introducing covariant derivatives and multiplying the measure by a factor of $\sqrt{\det h}$.
- Secondly, one introduces an Einstein–Hilbert term as the 'kinetic' term for the metric field, plus possibly extra terms and fields to preserve the symmetries of the original Lagrangian.
- Finally, one has to integrate the resulting theory over the space of all metrics.

Here we will not discuss the first two steps in this procedure. As we have seen in our discussion of topological field theories, the precise form of the Lagrangian only plays a comparatively minor role in determining the properties of the theory, and we can derive many results without actually considering a Lagrangian. Therefore, let us just state that it is possible to carry out the analog of the first two steps mentioned above, and construct a Lagrangian with a 'dynamical' metric which still possesses the topological $Q-$symmetry we have constructed. The reader who is interested in the details of this construction is referred to the paper [Wit90] and to the lecture notes [DVV91].

The third step, integrating over the space of all metrics, is the one we will be most interested in here. Naively, by the metric independence of our

theories, integrating their partition functions over the space of all metrics, and then dividing the results by the volume of the topological 'gauge group', would be equivalent to multiplication by a factor of 1,

$$Z[h_0] \stackrel{?}{=} \frac{1}{G_{top}} \int \mathcal{D}[h]\, Z[h], \tag{2.532}$$

for any arbitrary background metric h_0. There are several reasons why this naive reasoning might go wrong:

- There may be metric configurations which cannot be reached from a given metric by continuous changes.
- There may be anomalies in the topological symmetry at the quantum level preventing the conclusion that all gauge fixed configurations are equivalent.
- The volume of G_{top} is infinite, so even if we could rigorously define a path integral the above multiplication and division would not be mathematically well–defined.

For these reasons, we should really be more careful and precisely define what we mean by the *'integral over the space of all metrics'*. Let us note the important fact that just like in ordinary string theory (and even before twisting), the 2D sigma models become conformal field theories when we include the metric in the Lagrangian. This means that we can borrow the technology from string theory to integrate over all conformally equivalent metrics. As is well known, and as we will discuss in more detail later, the conformal symmetry group is a huge group, and integrating over conformally equivalent metrics leaves only a nD integral over a set of world–sheet moduli. Therefore, our strategy will be to use the analogy to ordinary string theory to first do this integral over all conformally equivalent metrics, and then perform the integral over the remaining nD moduli space.

In integrating over conformally equivalent metrics, one usually has to worry about conformal anomalies. However, here a very important fact becomes our help. To understand this fact, it is useful to rewrite our twisting procedure in a somewhat different language (see [Von05]).

Let us consider the SEM–tensor $T_{\alpha\beta}$, which is the conserved Noether current with respect to global translations on $\mathbb{C}$. From conformal field theory, it is known that $T_{z\bar{z}} = T_{\bar{z}z} = 0$, and the fact that T is a conserved current, $\partial_\alpha T^\alpha{}_\beta = 0$, means that $T_{zz} \equiv T(z)$ and $T_{\bar{z}\bar{z}} \equiv \bar{T}(\bar{z})$ are (anti–)holomorphic in z. One can now expand $T(z)$ in Laurent modes,

$$T(z) = \sum L_m z^{-m-2}. \tag{2.533}$$

The L_m are called the *Virasoro generators*, and it is a well–known result from conformal field theory that in the quantum theory their commutation relations are

$$[L_m, L_n] = (m-n)L_{m+n} + \frac{c}{12} m(m^2-1)\delta_{m+n}.$$

The number c depends on the details of the theory under consideration, and it is called the central charge. When this central charge is nonzero, one runs into a technical problem. The reason for this is that the equation of motion for the metric field reads

$$\frac{\delta S}{\delta h^{\alpha\beta}} = T_{\alpha\beta} = 0.$$

In conformal field theory, one imposes this equation as a constraint in the quantum theory. That is, one requires that for physical states $|\psi\rangle$,

$$L_m|\psi\rangle = 0 \qquad (\text{for all } m \in \mathbb{Z}).$$

However, this is clearly incompatible with the above commutation relation unless $c = 0$. In string theory, this value for c can be achieved by taking the target space of the theory to be 10D. If $c \neq 0$ the quantum theory is problematic to define, and we speak of a 'conformal anomaly' [Von05].

The whole above story repeats itself for $\bar{T}(\bar{z})$ and its modes $\bar{L}_m$. At this point there is a crucial difference between open and closed strings. On an open string, left–moving and right–moving vibrations are related in such a way that they combine into standing waves. In our complex notation, 'left–moving' translates into '$z-$dependent' (i.e. holomorphic), and 'right–moving' into '$\bar{z}-$dependent' (i.e. anti–holomorphic). Thus, on an open string all holomorphic quantities are related to their anti–holomorphic counterparts. In particular, $T(z)$ and $\bar{T}(\bar{z})$, and their modes L_m and $\bar{L}_m$, turn out to be complex conjugates. There is therefore only one independent algebra of Virasoro generators L_m.

On a *closed string* on the other hand, which is the situation we have been studying so far, left– and right–moving waves are completely independent. This means that all holomorphic and anti–holomorphic quantities, and in particular $T(z)$ and $\bar{T}(\bar{z})$, are independent. One therefore has two sets of Virasoro generators, L_m and $\bar{L}_m$.

Let us now analyze the problem of central charge in the twisted theories. To twist the theory, we have used the $U(1)-$symmetries. Any global $U(1)-$symmetry of our theory has a conserved current J_α. The fact that it is conserved again means that $J_z \equiv J(z)$ is holomorphic and $J_{\bar{z}} \equiv \bar{J}(\bar{z})$ is anti–holomorphic. Once again, on an open string J and $\bar{J}$ will be related, but in the closed string theory we are studying they will be independent functions. In particular, this means that we can view a global $U(1)-$symmetry as really consisting of two independent, left– and right–moving, $U(1)-$symmetries, with generators F_L and F_R.

Note that the sum of $U(1)-$symmetries $F_V + F_A$ only acts on objects with a + index. That is, it acts purely on left–moving quantities. Similarly, $F_V - F_A$ acts purely on right–moving quantities. From our discussion above, it is therefore natural to identify these two symmetries with the two components of a single global $U(1)$ symmetry:

$$F_V = \frac{1}{2}(F_L + F_R) \qquad F_A = \frac{1}{2}(F_L - F_R).$$

A more detailed construction shows that this can indeed be done.

Let us expand the left–moving conserved $U(1)-$current into Laurent modes,

$$J(z) = \sum J_m z^{-m-1}. \tag{2.534}$$

The commutation relations of these modes with one another and with the Virasoro modes can be calculated, either by writing down all of the modes in terms of the fields of the theory, or by using more abstract knowledge from the theory of superconformal symmetry algebras. In either case, one finds

$$\begin{aligned}[L_m, L_n] &= (m-n)L_{m+n} + \frac{c}{12}m(m^2-1)\delta_{m+n}[L_m, J_n] \\ &= -nJ_{m+n}[J_m, J_n] = \frac{c}{3}m\delta_{m+n}.\end{aligned}$$

Note that the same central charge c appears in the $J-$ and in the $L-$commutators. This turns out to be crucial.

Following the standard Noether procedure, we can now construct a conserved charge by integrating the conserved current $J(z)$ over a space–like slice of the $z-$plane. In string theory, the physical time direction is the radial direction in the $z-$plane, so a space–like slice is just a curve around the origin. The integral is therefore calculated using the *Cauchy theorem*,

$$F_L = \oint_{z=0} J(z)dz = 2\pi\mathrm{i}J_0.$$

In the quantum theory, it will be this operator that generates the $U(1)_L-$symmetry. Now recall that to twist the theory we want to introduce new Lorentz rotation generators,

$$M_A = M - F_V = M - \frac{1}{2}(F_L + F_R) M_B = M - F_A = M - \frac{1}{2}(F_L - F_R).$$

A well–known result from string theory (see [Von05]) is that the generator of Lorentz rotations is $M = 2\pi\mathrm{i}(L_0 - \bar{L}_0)$. Therefore, we find that the twisting procedure in this new language amounts to

$$\begin{aligned} A : L_{0,A} &= L_0 - \frac{1}{2}J_0, \qquad \bar{L}_{0,A} = \bar{L}_0 + \frac{1}{2}\bar{J}_0, \\ B : L_{0,B} &= L_0 - \frac{1}{2}J_0, \qquad \bar{L}_{0,B} = \bar{L}_0 - \frac{1}{2}\bar{J}_0. \end{aligned}$$

Let us now focus on the left–moving sector; we see that for both twistings the new Lorentz rotation generator is the difference of L_0 and $\frac{1}{2}J_0$. The new Lorentz generator should also correspond to a conserved 2–tensor, and from (2.533) and (2.534) there is a very natural way to obtain such a current:

$$\tilde{T}(z) = T(z) + \frac{1}{2}\partial J(z), \tag{2.535}$$

which clearly satisfies $\bar{\partial}\tilde{T} = 0$ and

$$\tilde{L}_m = L_m - \frac{1}{2}(m+1)J_m, \tag{2.536}$$

so in particular we find that $\tilde{L}_0$ can serve as $L_{0,A}$ or $L_{0,B}$. We should apply the same procedure (with a minus sign in the $A-$model case) in the right–moving sector. Equations (2.535) and (2.536) tell us how to implement the twisting procedure not only on the conserved charges, but on the whole $N = 2$ superconformal algebra – or at least on the part consisting of the $J-$ and $L-$modes, but a further investigation shows that this is the only part that changes. We have motivated, but not rigorously derived (2.535); for a complete justification the reader is referred to the original papers [LVW89] and [CV91].

Now, we come to the crucial point. The algebra that the new modes $\tilde{L}_m$ satisfy can be directly calculated from (2.535), and we find

$$[\tilde{L}_m, \tilde{L}_n] = (m-n)\tilde{L}_{m+n}.$$

That is, there is no central charge left. This means that we do not have any restriction on the dimension of the theory, and topological strings will actually be well–defined in target spaces of any dimension.

From this result, we see that we can integrate our partition function over conformally equivalent metrics without having to worry about the conformal anomaly represented by the nonzero central charge. After having integrated over this large part of the space of all metrics, it turns out that there is a nD integral left to do. In particular, it is known that one can always find a conformal transformation which in the neighborhood of a chosen point puts the metric in the form $h_{\alpha\beta} = \eta_{\alpha\beta}$, with η the usual flat metric with diagonal entries ± 1. (Or, $+1$ in the Euclidean setting.) On the other hand, when one considers the global situation, it turns out that one cannot always enforce this gauge condition everywhere. For example, if the world–sheet is a torus, there is a left–over complex parameter τ that cannot be gauged away. The easiest way to visualize this parameter (see [Von05]) is by drawing the resulting torus in the complex plane and rescaling it in such a way that one of its edges runs from 0 to 1; the other edge then runs from 0 to τ, see Figure 2.5. It seems intuitively clear that a conformal transformation – which should leave all angles fixed – will never deform τ, and even though intuition often fails when considering conformal mappings, in this case this can indeed be proven. Thus, τ is really a modular parameter which we need to integrate over. Another fairly intuitive result is that any locally flat torus can, after a rescaling, be drawn in this form, so τ indeed is the only modulus of the torus.

More generally, one can show that a Riemann surface of genus g has $m_g = 3(g-1)$ complex modular parameters. As usual, this is the *virtual* dimension of the moduli space. If $g > 1$, one can show that this virtual dimension equals the

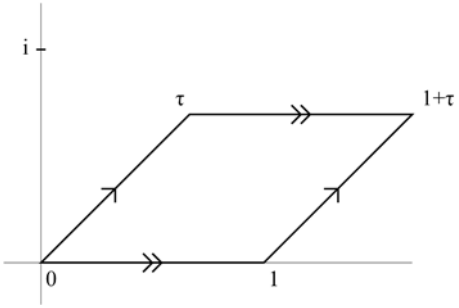


Fig. 2.5. The only modulus τ of a torus T^2.

actual dimension. For $g = 0$, the sphere, we have a negative virtual dimension $m_g = -3$, but the actual dimension is 0: there is always a flat metric on a surface which is topologically a sphere (just consider the sphere as a plane with a point added at infinity), and after having chosen this metric there are no remaining parameters such as τ in the torus case. For $g = 1$, the virtual dimension is $m_g = 0$, but as we have seen the actual dimension is 1.

We can explain these discrepancies using the fact that, after we have used the conformal invariance to fix the metric to be flat, the sphere and the torus have leftover symmetries. In the case of the sphere, it is well known in string theory that one can use these extra symmetries to fix the positions of three labelled points. In the case of the torus, after fixing the metric to be flat we still have rigid translations of the torus left, which we can use to fix the position of a single labelled point. To see how this leads to a difference between the virtual and the actual dimensions, let us for example consider tori with n labelled points on them. Since the virtual dimension of the moduli space of tori without labelled points is 0, the virtual dimension of the moduli space of tori with n labelled points is n. One may expect that at some point (and in fact, this happens already when $n = 1$), one reaches a sufficiently generic situation where the virtual dimension really is the actual dimension. However, even in this case we can fix one of the positions using the remaining conformal (translational) symmetry, so the positions of the points only represent $n - 1$ moduli. Hence, there must be an nth modulus of a different kind, which is exactly the shape parameter τ that we have encountered above. In the limiting case where $n = 0$, this parameter survives, thus causing the difference between the virtual and the real dimension of the moduli space.

For the sphere, the reasoning is somewhat more formal: we analogously expect to have three 'extra' moduli when $n = 0$. In fact, three extra parameters are present, but they do not show up as moduli. They must be viewed as the three parameters which need to be added to the problem to find a 0D moduli space.

Since the cases $g = 0, 1$ are thus somewhat special, let us begin by studying the theory on a Riemann surface with $g > 1$. To arrive at the topological string correlation functions, after gauge fixing we have to integrate over the remaining moduli space of complex dimension $3(g-1)$. To do this, we need to fix a measure on this moduli space. That is, given a set of $6(g-1)$ tangent vectors to the moduli space, we want to produce a number which represents the size of the volume element spanned by these vectors, see Figure 2.6. We should do this in a way which is invariant under coordinate redefinitions of both the moduli space and the world–sheet. Is there a 'natural' way to do this?

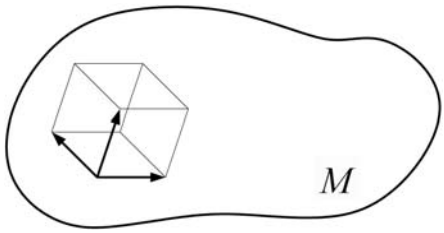


Fig. 2.6. A measure on the moduli space M assigns a number to every set of three tangent vectors. This number is interpreted as the volume of the element spanned by these vectors.

To answer this question, let us first ask how we can describe the tangent vectors to the moduli space (see [Von05]). In two dimensions, conformal transformations are equivalent to holomorphic transformations: $z \mapsto f(z)$. It thus seems natural to assume that the moduli space we have left labels different complex structures on Σ, and indeed this can be shown to be the case. Therefore, a tangent vector to the moduli space is an infinitesimal change of complex structure, and these changes can be parameterized by holomorphic 1–forms with anti–holomorphic vector indices,

$$dz \mapsto dz + \epsilon \mu_{\bar{z}}^{z}(z) d\bar{z}.$$

The dimension counting above tells us that there are $3(g-1)$ independent $(\mu_i)_{\bar{z}}^{z}$, plus their $3(g-1)$ complex conjugates which change $d\bar{z}$. So the tangent space is spanned by these $\mu_i(z,\bar{z}), \bar{\mu}_i(z,\bar{z})$. How do we get a number out of a set of these objects? Since μ_i has a z and a $\bar{z}$ index, it seems natural to integrate it over Σ. However, the $z-$index is an upper index, so we need to lower it first with some tensor with two $z-$indices. It turns out that a good choice is to use the $Q-$partner G_{zz} of the SEM–tensor component T_{zz}, and thus to define the integration over moduli space as

$$\int_{M_g} \prod_{i=1}^{3g=3} \left(dm^i d\bar{m}^i \int_{\Sigma} G_{zz} (\mu_i)_{\bar{z}}^{z} \int_{\Sigma} G_{\bar{z}\bar{z}} (\bar{\mu}_i)_{z}^{\bar{z}} \right). \tag{2.537}$$

Note that by construction, this integral is also invariant under a change of basis of the moduli space. There are several reasons why using G_{zz} is a natural choice. First of all, this choice is analogous to what one does in bosonic string theory. There, one integrates over the moduli space using exactly the same formula, but with G replaced by the conformal ghost b. This ghost is the BRST–partner of the SEM–tensor in exactly the same way as G is the $Q-$partner of T. Secondly, one can make the not unrelated observation that since $\{Q, G\} = T$, we can still use the standard arguments to show independence of the theory of the parameters in a Lagrangian of the form $L = \{Q, V\}$. The only difference is that now we also have to commute Q through G to make it act on the vacuum, but since $T_{\alpha\beta}$ itself is the derivative of the action with respect to the metric $h_{\alpha\beta}$, the terms we obtain in this way amount to integrating a total derivative over the moduli space. Therefore, apart from possible boundary terms these contributions vanish. Note that this reasoning also gives us an argument for using G_{zz} instead of T_{zz} (which is more or less the only other reasonable option) in (2.537): if we had chosen T_{zz} then all path integrals would have been over total derivatives on the moduli space, and apart from boundary contributions the whole theory would have become trivial.

If we consider the vector and axial charges of the full path integral measure, including the new path integral over the world–sheet metric h, we find a surprising result. Since the world–sheet metric does not transform under $R-$symmetry, naively one might expect that its measure does not either. However, this is clearly not correct since one should also take into account the explicit $G-$insertions in (2.537) that do transform under $R-$symmetry. From the $N = 2$ superconformal algebra (or, more down–to–earth, from expressing the operators in terms of the fields), it follows that the product of G and $\bar{G}$ has vector charge zero and axial charge 2. Therefore, the total vector charge of the measure remains zero, and the axial charge gets an extra contribution of $6(g-1)$, so we find a total axial $R-$charge of $6(g-1) - 2m(g-1)$. From this, we see that the case of complex target space dimension 3 is very special: here, the axial charge of the measure vanishes for any g, and hence the partition function is nonzero at every genus. If $m > 3$ and $g > 1$, the total axial charge of the measure is negative, and we have seen that we cannot cancel such a charge with local operators. Therefore, for these theories only the partition function at $g = 1$ and a specific set of correlation functions at genus zero give nonzero results. Moreover, for $m = 2$ and $m = 1$, the results can be shown to be trivial by other arguments. Therefore, a Calabi–Yau threefold is by far the most interesting target space for a topological string theory. It is a 'happy coincidence' (see [Von05]) that this is exactly the dimension we are most interested in from the string theory perspective.

Finally, let us come back to the special cases of genus 0 and 1. At genus zero, the Riemann surface has a single point as its moduli space, so there are no extra integrals or $G-$insertions to worry about. Therefore, we can copy the topological field theory result saying that we have to introduce local operators with total degree (m, m) in the theory. The only remnant of the fact that we

are integrating over metrics is that we should also somehow fix the remaining three symmetries of the sphere. The most straightforward way to do this is to consider 3–point functions with insertions on three labelled points. As a gauge choice, we can then for example require these points to be at the points 0, 1 and ∞ in the compactified complex plane. For example, in the $A-$model on a Calabi–Yau threefold, the 3–point function of three operators corresponding to $(1,1)-$forms would thus give a nonzero result.

In the case of the torus, we have seen that there is one 'unexpected' modular parameter over which we have to integrate. This means we have to insert one $G-$ and one $\bar{G}-$operator in the measure, which spoils the absence of the axial anomaly we had for $g = 1$ in the topological field theory case. However, we also must fix the one remaining translational symmetry, which we can do by inserting a local operator at a labelled point. Thus, we can restore the axial $R-$charge to its zero value by choosing this to be an operator of degree $(1,1)$.

Summarizing, we have found that in topological string theory on a target Calabi–Yau 3–fold, we have a nonvanishing 3–point function of total degree (3,3) at genus zero; a nonvanishing 1–point function of degree (1,1) at genus one, and a nonvanishing partition ('zero–point') function at all genera $g > 1$.

Nonlocal Operators

In one respect, what we have achieved is great progress: we can now for any genus define a nonzero partition function (or for low genus a correlation function) of the topological string theory. On the other hand, we would also like to define correlation functions of an arbitrary number of operators at these genera. As we have seen, the insertion of extra local operators in the correlation functions is not possible, since any such insertion will spoil our carefully constructed absence of $R-$symmetry anomalies. Therefore, we have to introduce nonlocal operators.

There is one class of nonlocal operators which immediately becomes mind. Before we saw, using the descent equations, that for every local operator we can define a corresponding 1–form and a 2–form operator. If we check the axial and vector charges of these operators, we find that if we start with an operator of degree $(1,1)$, the 2–form operator we end up with actually has vanishing axial and vector charges. This has two important consequences. First of all, we can add the integral of this operator to our action [Von05],

$$S[t] = S_0 + t^a \int \mathcal{O}_a^{(2)},$$

without spoiling the axial and vector symmetry of the theory. Secondly, we can insert the integrated operator into correlation functions,

$$\langle \int \mathcal{O}_1^{(2)} \dots \int \mathcal{O}_n^{(2)} \rangle$$

and still get a nonzero result by the vanishing of the axial and vector charges. These two statements are related: one obtains such correlators by differentiating $S[t]$ with respect to the appropriate t's, and then setting all $t^a = 0$.

A few remarks are in place here. First of all, recall that the integration over the insertion points of the operators can be viewed as part of the integration over the moduli space of Riemann surfaces, where now we label a certain number of points on the Riemann surface. From this point of view, the $g = 0, 1$ cases fit naturally into the same framework. We could unite the descendant fields into a world–sheet *super–field*,

$$\Phi_a = \mathcal{O}_a^{(0)} + \mathcal{O}_{a\alpha}^{(1)}\theta^\alpha + \mathcal{O}_{a\alpha\beta}^{(2)}\theta^\alpha\theta^\beta$$

where we formally replaced each dz and $d\bar{z}$ by corresponding fermionic coordinates θ^z and $\theta^{\bar{z}}$. Now, one can write the above correlators as integrals over n copies of this *super–space*,

$$\int \prod_{s=1}^{n} d^2 z_s d^2\theta_s \, \langle \Phi_{a_1}(z_1, \theta_1) \cdots \Phi_{a_n}(z_n, \theta_n) \rangle$$

The integration prescription at genus 0 and 1 tells us to fix 3 and 1 points respectively, so we need to remove this number of super–space integrals. Then, integrating over the other super–space coordinates, the genus 0 correlators indeed become

$$\langle \mathcal{O}_{a_1}^{(0)} \mathcal{O}_{a_2}^{(0)} \mathcal{O}_{a_3}^{(0)} \int \mathcal{O}_{a_4}^{(2)} \cdots \int \mathcal{O}_{a_n}^{(2)} \rangle$$

From this prescription we note that these expressions are symmetric in the exchange of all a_i and a_j. In particular, this means that the genus zero 3–point functions at arbitrary t,

$$c_{abc}[t] = \langle \mathcal{O}_a^{(0)} \mathcal{O}_b^{(0)} \mathcal{O}_c^{(0)} \rangle [t]$$

have symmetric derivatives:

$$\frac{\partial c_{abc}}{\partial t^d} = \frac{\partial c_{abd}}{\partial t^c},$$

and similarly with permuted indices. These equations can be viewed as integrability conditions, and using the Poincaré lemma we see that they imply that

$$c_{ijk}[t] = \frac{\partial Z_0[t]}{\partial t^i \partial t^j \partial t^k}.$$

for some function $Z_0[t]$. Following the general philosophy that $n-$point functions are nth derivatives of the $t-$dependent partition function, we see that $Z_0[t]$ can be naturally thought of as the partition function at genus zero. Similarly, the partition function at genus 1 can be defined by integrating up the one-point functions once.

The quantities we have calculated above should be semi–topological invariants, meaning that they only depend on 'half' of the moduli (either the Kähler ones or the complex structure ones) of the target space. For example, in the $A-$model we find the Gromov–Witten invariants. In the $B-$model, it turns out that $F_0[t] = \ln Z_0[t]$ is actually a quantity we already knew: it is the prepotential of the Calabi–Yau manifold. A discussion of why this is the case can be found in the paper [BCO94]. The higher genus partition functions can be thought of as 'quantum corrections' to the prepotential.

Finally, there is a type of operator we have not discussed at all so far. Recall that in the topological string theory, the metric itself is now a dynamical field. We could not include the metric in our physical operators, since this would spoil the topological invariance. However, the metric is part of a $Q-$multiplet, and the highest field in this multiplet is a scalar field which is usually labelled φ. (It should not be confused with the fields ϕ^i.) We can get more correlation functions by inserting operators φ^k and the operators related to them by the descent equations into the correlation functions. These operators are called 'gravitational descendants'. Even the case where the power is $k = 0$ is nontrivial; it does not insert any operator, but it does label a certain point, and hence changes the moduli space one integrates over. This operator is called the 'puncture operator'.

All of this seems to lead to an enormous amount of semi-topological target space invariants that can be calculated, but there are many recursion relations between the several correlators. This is similar to how we showed before that all correlators for the cohomological field theories follow from the 2–and 3–point functions on the sphere. Here, it turns out that the set of all correlators has a structure which is related to the theory of integrable hierarchies. Unfortunately, a discussion of this is outside the scope of both these lectures and the author's current knowledge.

The Holomorphic Anomaly

We have now defined the partition function and correlation functions of topological string theory, but even though the expressions we obtained are much simpler than the path integrals for ordinary quantum field or string theories, it would still be very hard to explicitly calculate them. Fortunately, it turns out that the $t-$dependent partition and correlation functions are actually 'nearly holomorphic' in t, and this is a great aid in exactly calculating these quantities.

Let us make this 'near holomorphy' more precise. As we have seen, calculating correlation functions of primary operators in topological string theories amounts to taking $t-$derivatives of the corresponding perturbed partition function $Z[t]$ and consequently setting $t = 0$. Recall that $Z[t]$ is defined through adding terms to the action of the form

$$t^a \int_\Sigma \mathcal{O}_a^{(2)}, \tag{2.538}$$

Let us for definiteness consider the $B-$twisted model. We want to show that the above term is Q_B-exact. For simplicity, we assume that $\mathcal{O}_a^{(2)}$ is a bosonic operator, but what we are about to say can by inserting a few signs straightforwardly be generalized to the fermionic case. From the descent equations we studied in the subsection 2.2.8 above, we know that

$$(\mathcal{O}_a^{(2)})_{+-} = -\{G_+, [G_-, \mathcal{O}_a^{(0)}]\}, \tag{2.539}$$

where G_+ is the charge corresponding to the current G_{zz}, and G_- the one corresponding to $G_{\bar{z}\bar{z}}$. We can in fact express $G_\pm$ in terms of the $N = (2,2)$ supercharges $\mathcal{Q}$. So, according to [Von05], we have

$$\begin{aligned} H = 2\pi\mathrm{i}(L_0 + \bar{L}_0) &= \frac{1}{2}\{\mathcal{Q}_+, \bar{\mathcal{Q}}_+\} - \frac{1}{2}\{\mathcal{Q}_-, \bar{\mathcal{Q}}_-\}P \\ = 2\pi\mathrm{i}(L_0 - \bar{L}_0) &= \frac{1}{2}\{\mathcal{Q}_+, \bar{\mathcal{Q}}_+\} + \frac{1}{2}\{\mathcal{Q}_-, \bar{\mathcal{Q}}_-\}. \end{aligned}$$

Thus, we find that the left– and right–moving SEM charges satisfy

$$T_+ = 2\pi\mathrm{i}L_0 = \frac{1}{2}\{\mathcal{Q}_+, \bar{\mathcal{Q}}_+\}T_- = 2\pi\mathrm{i}\bar{L}_0 = -\frac{1}{2}\{\mathcal{Q}_-, \bar{\mathcal{Q}}_-\}.$$

To find G in the $B-$model, we should write these charges as commutators with respect to $Q_B = \bar{\mathcal{Q}}_+ + \bar{\mathcal{Q}}_-$, which gives

$$T_+ = \frac{1}{2}\{Q_B, \mathcal{Q}_+\}T_- = -\frac{1}{2}\{Q_B, \mathcal{Q}_-\},$$

so we arrive at the conclusion that for the $B-$model,

$$G_+ = \frac{1}{2}\mathcal{Q}_+ G_- = -\frac{1}{2}\mathcal{Q}_-.$$

Now, we can rewrite (2.539) as

$$\begin{aligned} (\mathcal{O}_a^{(2)})_{+-} &= -\{G_+, [G_-, \mathcal{O}_a^{(0)}]\} = \frac{1}{4}\{\mathcal{Q}_+, [\mathcal{Q}_-, \mathcal{O}_a^{(0)}]\} \\ &= \frac{1}{8}\{\bar{Q}_B, [(\mathcal{Q}_- - \mathcal{Q}_+), \mathcal{O}_a^{(0)}]\}, \end{aligned} \tag{2.540}$$

which proves our claim that $\mathcal{O}_a^{(2)}$ is Q_B-exact.

An $N = (2,2)$ sigma model with a real action does, apart from the term (2.538), also contain a term

$$t^{\bar{a}} \int_\Sigma \bar{\mathcal{O}}_a^{(2)}, \tag{2.541}$$

where $t^{\bar{a}}$ is the complex conjugate of t^a. It is not immediately clear that $\bar{\mathcal{O}}_a^{(2)}$ is a physical operator: we have seen that physical operators in the $B-$model correspond to forms that are $\bar{\partial}-$closed, but the complex conjugate of such a

form is $\partial-$closed. However, by taking the complex conjugate of (2.540), we see that

$$(\bar{\mathcal{O}}_a^{(2)})_{+-} = \frac{1}{8}\{Q_B, [(\bar{\mathcal{Q}}_- - \bar{\mathcal{Q}}_+), \bar{\mathcal{O}}_a^{(0)}]\},$$

so not only is the operator Q_B-closed, it is even Q_B-exact. This means that we can add terms of the form (2.541) to the action, and taking $t^{\bar{a}}-$derivatives inserts Q_B-exact terms in the correlation functions. Naively, we would expect this to give a zero result, so all the physical quantities seem to be $t-$independent, and thus holomorphic in t. We will see in a moment that this naive expectation turns out to be almost right, but not quite.

However, before doing so, let us comment briefly on the generalization of the above argument in the case of the $A-$model. It seems that a straightforward generalization of the argument fails, since Q_A is its own complex conjugate, and the complex conjugate of the de Rham operator is also the same operator. However, note that the $N = (2,2)-$theory has a different kind of 'conjugation symmetry': we can exchange the two supersymmetries, or in other words, exchange θ^+ with $\bar{\theta}^+$ and θ^- with $\bar{\theta}^-$. This exchanges Q_A with an operator which we might denote as $Q_{\bar{A}} \equiv \mathcal{Q}_+ + \bar{\mathcal{Q}}_-$. Using the above argument, we then find that the physical operators $\mathcal{O}_a^{(2)}$ are $Q_{\bar{A}}-$exact, and that their conjugates in the new sense are Q_A-exact. We can now add these conjugates to the action with parameters $t^{\bar{a}}$, and we again naively find independence of these parameters. In this case it is less natural to choose t^a and $t^{\bar{a}}$ to be complex conjugates, but we are free to choose this particular 'background point' and study how the theory behaves if we then vary t^a and $t^{\bar{a}}$ independently.

Now, let us see how the naive argument showing independence of the theory of $t^{\bar{a}}$ fails. In fact, the argument above would certainly hold for topological *field* theories. However, in topological string theories (see [Von05]), we have to worry about the insertions in the path integral of

$$G \cdot \mu_i \equiv \int d^2z\, G_{zz}\, (\mu_i)^z_{\bar{z}},$$

and their complex conjugates, when commuting the Q_B towards the vacuum and making sure it gives a zero answer. Indeed, the Q_B-commutator of the above factor is not zero, but it gives

$$\{Q_B, G \cdot \mu_i\} = T \cdot \mu_i.$$

Now recall that $T_{\alpha\beta} = \partial_{h^{\alpha\beta}} S$. We did not give a very precise definition of μ_i above, but we know that it parameterizes the change in the metric under an infinitesimal change of the coordinates m_i on the moduli space. One can make this intuition precise, and then finds the following 'chain rule': $T \cdot \mu_i = \partial_{m^i} S$. Inserting this into the partition function, we find that

$$\frac{\partial F_g}{\partial t^{\bar{a}}}$$
$$= \int_{M_g} \prod_{i=1}^{3g-3} dm^i d\bar{m}^i \sum_{j,k} \frac{\partial^2}{\partial m^j \partial \bar{m}^k} \left\langle (\prod_{l\neq j} \int \mu_l \cdot G)(\prod_{l\neq k} \int \bar{\mu}_l \cdot \bar{G}) \int \bar{\mathcal{O}}_a^{(2)} \right\rangle ,$$

where $F_g = \ln Z_g$ is the free energy at genus g, and the reason F_g appears in the above equation instead of Z_g is, as usual in quantum field theory, that the expectation values in the r.h.s. are normalized such that $\langle 1 \rangle = 1$, and so the l.h.s. should be normalized accordingly and equal $Z_g^{-1} \partial_{\bar{a}} Z_g = \partial_{\bar{a}} F_g$ [Von05].

Thus, as we have claimed before, we are integrating a total derivative over the moduli space of genus g surfaces. If the moduli space did not have a boundary, this would indeed give zero, but in fact the moduli space does have a boundary. It consists of the moduli which make the genus g surface degenerate. This can happen in two ways: an internal cycle of the genus g surface can be pinched, leaving a single surface of genus $g-1$, as in Figure 2.7 (a), or the surface can split up into two surfaces of genus g_1 and $g_2 = g - g_1$, as depicted in Figure 2.7 (b). By carefully considering the boundary contributions to the integral for these two types of boundaries, it was shown in [BCO94] that

$$\frac{\partial F_g}{\partial t^{\bar{a}}} = \frac{1}{2} c_{\bar{a}\bar{b}\bar{c}} e^{2K} G^{\bar{b}d} G^{\bar{c}e} \left(D_d D_e F_{g-1} + \sum_{r=1}^{g-1} D_d F_r D_e F_{g-r} \right) ,$$

where G is the so–called *Zamolodchikov metric* on the space parameterized by the coupling constants $t^a, t^{\bar{a}}$; K is its Kähler potential, and the D_a are covariant derivatives on this space. The coefficients $c_{\bar{a}\bar{b}\bar{c}}$ are the 3–point functions on the sphere of the operators $\bar{\mathcal{O}}_a^{(0)}$. We will not derive the above formula in detail, but the reader should notice that the contributions from the two types of boundary are quite clear.

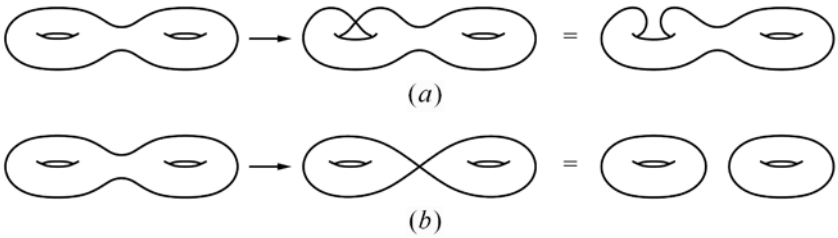


Fig. 2.7. At the boundary of the moduli space of genus g surfaces, the surfaces degenerate because certain cycles are pinched. This either lowers the genus of the surface (a) or breaks the surface into two lower genus ones (b) (adapted from [Von05]).

Using this formula, one can inductively determine the $t^{\bar{a}}$ dependence on the partition functions if the holomorphic t^a-dependence is known. Holomor-

phic functions on complex spaces (or more generally holomorphic sections of complex vector bundles) are quite rare: usually, there is only a nD space of such functions. The same turns out to hold for our topological string partition functions: even though they are not quite holomorphic, their anti–holomorphic behavior is determined by the holomorphic dependence on the coordinates, and as a result there is a finite number of coefficients which determines them.

Thus, just from the above structure and without doing any path integrals, one can already determine the topological string partition functions up to a finite number of constants. This leads to a feasible program for completely determining the topological string partition function for a given target space and at given genus. From the holomorphic anomaly equation, one first has to find the general form of the partition function. Then, all one has left to do is to fix the unknown constants. Here, the fact that in the $A-$model the partition function counts a number of points becomes our help: by requiring that the $A-$model partition functions are integral, one can often fix the unknown constants and completely determine the $t-$dependent partition function. In practice, the procedure is still quite elaborate, so we will not describe any examples here, but several have been worked out in detail in the literature. Once again, the pioneering work for this can be found in the paper [BCO94].

Geometrical Transitions

Conifolds

Recall that a *conifold* is a generalization of the notion of a manifold. Unlike manifolds, a conifold can (or, should) contain conical singularities i.e., points whose neighborhood looks like a cone with a certain base. The base is usually a 5D manifold.

In string theory, a *conifold transition* represents such an evolution of the Calabi–Yau manifold in which its fabric rips and repairs itself, yet with mild and acceptable physical consequences in the context of string theory. However, the tears involved are more severe than those in an 'weaker' *flop transition* (see [Gre00]). The geometrically singular conifolds were shown to lead to completely smooth physics of strings. The divergences are 'smeared out' by D3–branes wrapped on the shrinking 3–sphere S^3, as originally pointed out by A. Strominger, who, together with D. Morrison and B. Greene have also found that the topology near the conifold singularity can undergo a topological phase–transition (see subsection 2.2.7). It is believed that nearly all Calabi–Yau manifolds can be connected via these 'critical transitions'.

More precisely, the conifold is the simplest example of a non–compact Calabi–Yau 3–fold: it is the set of solutions to the equation

$$x_1 x_2 - x_3 x_4 = 0$$

in $\mathbb{C}^4$. The resulting manifold is a cone, meaning in this case that any real multiple of a solution to this equation is again a solution. The point $(0, 0, 0, 0)$

is the 'tip' of this cone, and it is a singular point of the solution space. Note that by writing

$$x_1 = z_1 + iz_2, \qquad x_2 = z_1 - iz_2, \qquad x_3 = z_3 + iz_4, \qquad x_4 = -z_3 + iz_4,$$

where the z_i are still *complex* numbers, one can also write the equation as

$$z_1^2 + z_2^2 + z_3^2 + z_4^2 = 0.$$

Writing each z_i as $a_i + ib_i$, with a_i and b_i real, we get the two equations

$$|a|^2 - |b|^2 = 0, \qquad a \cdot b = 0. \tag{2.542}$$

Here $a \cdot b = \sum_i a_i b_i$ and $|a|^2 = a \cdot a$. Since the geometry is a cone, let us focus on a 'slice' of this cone given by

$$|a|^2 + |b|^2 = 2r^2,$$

for some $r \in \mathbb{R}$. On this slice, the first equation in (2.542) becomes

$$|a|^2 = r^2, \tag{2.543}$$

which is the equation defining a 3–sphere S^3 of radius r. The same holds for b, so both a and b lie on 3–spheres. However, we also have to take the second equation in (2.542) into account. Let us suppose that we fix an a satisfying (2.543). Then b has to lie on a 3–sphere, but also on the plane through the origin defined by $a \cdot b = 0$. That is, b lies on a 2–sphere. This holds for every a, so the slice we are considering is a fibration of 2–spheres over the 3–sphere. With a little more work, one can show that this fibration is trivial, so the conifold is a cone over $S^2 \times S^3$.

Since the conifold is a singular geometry, we would like to find geometries which approximate it, but which are non–singular. There are two interesting ways in which this can be done. The simplest way is to replace the defining equation by

$$x_1 x_2 - x_3 x_4 = \mu^2. \tag{2.544}$$

From the two equations constraining a and b, we now see that $|a|^2 \geq \mu^2$. In other words, the parameter r should be at least μ. At $r = \mu$, the $a-$sphere still has finite radius μ, but the $b-$sphere shrinks to zero size. This geometry is called the *deformed conifold*. Even though this is not clear from the picture, from the equation (2.544) one can straightforwardly show that it is nonsingular. One can also show that it is topologically equivalent to the cotangent bundle on the 3–sphere, T^*S^3. Here, the S^3 on which the cotangent bundle is defined is exactly the S^3 at the 'tip' of the deformed conifold.

The second way to change the conifold geometry arises from studying the two equations

$$x_1 A + x_3 B = 0, \qquad x_4 A + x_2 B = 0. \tag{2.545}$$

Here, we require A and B to be homogeneous complex coordinates on a $\mathbb{C}P^1$, i.e.,

$$(A, B) \neq (0, 0), \qquad (A, B) \sim (\lambda A, \lambda B)$$

where λ is any nonzero complex number. If one of the x_i is nonzero, say x_1, one can solve for A or B, e.g., $A = -\frac{x_3 B}{x_1}$, and insert this in the other equation to get

$$x_1 x_2 - x_3 x_4 = 0$$

which is the conifold equation. However, if all x_i are zero, any A and B solve the system of equations (2.545). In other words, we have constructed a geometry which away from the former singularity is completely the same as the conifold, but the singularity itself is replaced by a $\mathbb{C}P^1$, which topologically is the same as an S^2. From the defining equations one can again show that the resulting geometry is nonsingular, so we have now replaced our conifold geometry by the so–called *resolved conifold*.

Topological D–branes

Since topological string theories are in many ways similar to an ordinary (bosonic) string theories, one natural question which arises is: are there also open topological strings which can end on D–branes? To answer the above question rigorously, we would have to study boundary conditions on world–sheets with boundaries which preserve the $Q-$symmetry.

In the $A-$model, one can only construct 3D–branes wrapping so–called 'Lagrangian' submanifolds of M. Here, 'Lagrangian' means that the Kähler form ω vanishes on this submanifold. In the $B-$model, one can construct D–branes of any even dimension, as long as these branes wrap *holomorphic* submanifolds of M.

Just like in ordinary string theory, when we consider open topological strings ending on a D–brane, there should be a field theory on the brane world–volume describing the low–energy physics of the open strings. Moreover, since we are studying topological theories, one may expect such a theory to inherit the property that it only depends on a restricted amount of data of the manifolds involved. A key example is the case of the $A-$model on the deformed conifold, $M = T^* S^3$, where we wrap ND–branes on the S^3 in the base. (One can show that this is indeed a Lagrangian submanifold.) In ordinary string theory, the world–volume theory on ND–branes has a $U(N)$ gauge symmetry, so putting the ingredients together we can make the guess that the world–volume theory is a 3D topological field theory with $U(N)$ gauge symmetry. There is really only one candidate for such a theory: the *Chern–Simons gauge theory*. Recall that it consists of a single $U(N)$ gauge field, and has the action

$$S = \frac{k}{4\pi} \int_{S^3} \mathrm{Tr} \left(A \wedge dA + \frac{2}{3} A \wedge A \wedge A \right). \tag{2.546}$$

Before the invention of D–branes, E. Witten showed that this is indeed the theory one gets. In fact, he showed even more: this theory actually describes

the *full* topological string–field theory on the D–branes, even without going to a low–energy limit [Wit95].

Let us briefly outline the argument that gives this result. In his paper, Witten derived the open string–field theory action for the open $A-$model topological string; it reads

$$S = \int \mathrm{Tr}\left(\mathcal{A} * Q_A \mathcal{A} + \frac{2}{3}\mathcal{A} * \mathcal{A} * \mathcal{A}\right).$$

The form of this action is very similar to Chern–Simons theory, but its interpretation is completely different: $\mathcal{A}$ is a *string–field* (a wave function on the space of all maps from an open string to the space–time manifold), Q_A is the topological symmetry generator, which has a natural action on the string–field, and $*$ is a certain noncommutative product. Witten shows that the topological properties of the theory imply that only the constant maps contribute, so $\mathcal{A}$ becomes a field on M – and since open strings can only end on D–branes, it actually becomes a field on S^3. Moreover, recall that Q_A can be interpreted as a de Rham differential. Using these observations and the precise definition of the star product one can indeed show that the string–field theory action reduces to Chern–Simons theory on S^3.

Topological Strings and Black Hole Attractors

Topological string theory is naturally related to *black hole dynamics* (see subsection 2.2.6 above). Namely, critical string theory compactified on Calabi–Yau manifolds has played a central role in both the mathematical and physical development of modern string theory. The physical relevance of the data provided by the topological string $\hat{c} = 6$ (of A and B types) has been that it computes $F-$type terms in the corresponding four dimensional theory [BCO94, AGN94]. These higher–derivative $F-$type terms for Type II superstring on a Calabi–Yau manifold are of the general form

$$\int d^4x d^4\theta (W_{ab}W^{ab})^g F_g(X^\Lambda), \tag{2.547}$$

where W_{ab} is the *graviphoton super–field* of the $N = 2$ *super–gravity* and X^Λ are the *vector multiplet fields*. The lowest component of W is F the graviphoton field strength and the highest one is the Riemann tensor. The lowest components of X^Λ are the complex scalars parameterizing Calabi–Yau moduli and their highest components are the associated $U(1)$ vector fields. These terms contribute to multiple graviphoton–graviton scattering. (2.547) includes (after θ integrations) an R^2F^{2g-2} term. The topological string partition function Z_{top} represents the canonical ensemble for multi–particle spinning five dimensional black holes [BMP97, KKV99].

Recently, [OSV04] proposed a simple and direct relationship between the second–quantized topological string partition function Z_{top} and black hole partition function Z_{BH} in four dimensions of the form

$$Z_{BH}(p^\Lambda, \phi^\Lambda) = |Z_{\rm top}(X^\Lambda)|^2, \qquad \text{where} \qquad X^\Lambda = p^\Lambda + \frac{\rm i}{\pi}\phi^\Lambda$$

in a certain *Kähler gauge*. The l.h.s. here is evaluated as a function of integer magnetic charges p^Λ and continuous electric potentials ϕ^Λ, which are conjugate to integer electric charges q_Λ. The r.h.s. is the holomorphic square of the partition function for a gas of topological strings on a Calabi–Yau whose moduli are those associated to the charges/potentials $(p^\Lambda, \phi^\Lambda)$ via the attractor equations [OSV04]. Both sides of (2.548) are defined in a perturbation expansion in $1/Q$, where Q is the graviphoton charge carried by the black hole.[18] The nonperturbative completion of either side of (2.548) might in principle be defined as the partition function of the holographic CFT dual to the black hole, as in [SV96]. Then we have the triple equality,

$$Z_{CFT} = Z_{BH} = |Z_{\rm top}|^2.$$

The existence of fundamental connection between 4D black holes and the topological string might have been anticipated from the following observation. Calabi–Yau spaces have two types of moduli: Kähler and complex structure. The world–sheet twisting which produces the A (B) model topological string from the critical superstring eliminates all dependence on the complex structure (Kähler) moduli at the perturbative level. Hence the perturbative topological string depends on only half the moduli. Black hole entropy on the other hand, insofar as it is an intrinsic property of the black hole, cannot depend on any externally specified moduli. What happens at leading order is that the moduli in vector multiplets are driven to attractor values at the horizon which depend only on the black hole charges and not on their asymptotically specified values. Hypermultiplet vevs on the other hand are not fixed by an attractor mechanism but simply drop out of the entropy formula. It is natural to assume this is valid to all orders in a $1/Q$ expansion. Hence the perturbative topological string and the large black hole partition functions depend on only half the Calabi–Yau moduli. It would be surprising if string theory produced two functions on the same space that were not simply related. Indeed [OSV04] argued that they were simply related as in (2.548).

Supergravity Area–Entropy Formula

Recall that a well–known hypothesis by J. Bekenstein and S. Hawking states that *the entropy of a black hole is proportional to the area of its horizon* (see [HE79]). This *area* is a function of the black hole mass, or in the extremal case, of its charges. Here we review the leading *semiclassical area–entropy formula* for a general $N = 2,\ d = 4$ extremal black hole characterized by magnetic and electric charges (p^Λ, q_Λ), recently reviewed in [OSV04]. The asymptotic values of the moduli in vector multiplets, parameterized by complex projective coordinates $X^\Lambda,\ (\Lambda = 0, 1, \ldots, n_V)$ in the black hole solution, are arbitrary.

[18] The string coupling g_s is in a hypermultiplet and decouples from the computation.

These moduli couple to the electromagnetic fields and accordingly vary as a function of the radius. At the horizon they approach an attractor point whose location in the moduli space depends only on the charges. The locations of these attractor points can be found by looking for supersymmetric solutions with constant moduli. They are determined by the attractor equations,

$$p^\Lambda = \mathrm{Re}[CX^\Lambda], \qquad q_\Lambda = \mathrm{Re}[CF_{0\Lambda}], \tag{2.548}$$

where $F_{0\Lambda} = \partial F_0/\partial X^\Lambda$ are the *holomorphic periods*, and the subscript 0 distinguishes these from the string loop corrected periods to appear in the next subsection. Both (p^Λ, q_Λ) and $(X^\Lambda, F_{0\Lambda})$ transform as vectors under the $Sp(2n+2;Z)$ duality group.

The $(2n_v+2)$ real equations (2.548) determine the (n_v+2) complex quantities (C, X^Λ) up to Kähler transformations, which act as

$$K \to K - f(X) - \bar{f}(\bar{X}), \quad X^\Lambda \to \mathrm{e}^f X^\Lambda, \qquad F_0 \to \mathrm{e}^{2f} F_0, \quad C \to \mathrm{e}^{-f} C,$$

where the *Kähler potential* K is given by

$$\mathrm{e}^{-K} = \mathrm{i}(\bar{X}^\Lambda F_{0\Lambda} - X^\Lambda \bar{F}_{0\Lambda}).$$

We could at this point set $C = 1$ and fix the Kähler gauge but later we shall find other gauges useful. It is easy to see that (as required) the charges (p^Λ, q_Λ) determined by the attractor equations (2.548) are invariant under Kähler transformations. Given the horizon attractor values of the moduli determined by (2.548) the *Bekenstein–Hawking entropy* S_{BH} may be written as

$$S_{BH} = \frac{1}{4} Area = \pi |Q|^2,$$

where $Q = Q_m + \mathrm{i}Q_\mathrm{e}$ is a complex combination of the magnetic and electric graviphoton charges and

$$|Q|^2 = \frac{\mathrm{i}}{2}\left(q_\Lambda \bar{C}\bar{X}^\Lambda - p^\Lambda \bar{C}\bar{F}_{0\Lambda}\right) = \frac{C\bar{C}}{4}\mathrm{e}^{-K}.$$

The normalization of Q here is chosen so that $|Q|$ equals the radius of the two sphere at the horizon.

It is useful to rephrase the above results in the context of type IIB superstrings in terms of geometry of Calabi–Yau. In this case the attractor equations fix the complex geometry of the Calabi–Yau. The electric/magentic charges correlate with three cycles of Calabi–Yau. Choosing a symplectic basis for the three cycles gives a choice of the splitting to electric and magnetic charges. Let A_Λ denote a basis for the electric three cycles, B^Σ the dual basis for the magnetic charges and Ω the holomorphic 3–form at the attractor point. Ω is fixed up to an overall multiplication by a complex number $\Omega \to \lambda\Omega$. There is a unique choice of λ such that the resulting Ω has the property that

$$p^\Lambda = \int_{A_\Lambda} \mathrm{Re}\,\Omega = \mathrm{Re}[CX^\Lambda], \qquad q_\Lambda = \int_{B^\Lambda} \mathrm{Re}\,\Omega = \mathrm{Re}[CF_{0\Lambda}],$$

where $\mathrm{Re}\,\Omega = \frac{1}{2}(\Omega + \overline{\Omega})$.

In terms of this choice, the black hole entropy can be written as

$$S_{BH} = \frac{\pi}{4}\int_{CY} \Omega \wedge \overline{\Omega}.$$

Higher–Order Corrections

$F-$term corrections to the action are encoded in a string loop corrected holomorphic prepotential

$$F(X^\Lambda, W^2) = \sum_{h=0}^{\infty} F_h(X^\Lambda) W^{2h}, \tag{2.549}$$

where F_h can be computed by topological string amplitudes (as we review in the next section) and W^2 involves the square of the anti–self dual graviphoton field strength. This obeys the homogeneity equation

$$X^\Lambda \partial_\Lambda F(X^\Lambda, W^2) + W\partial_W F(X^\Lambda, W^2) = 2F(X^\Lambda, W^2). \tag{2.550}$$

Near the black hole horizon, the attractor value of W^2 obeys $C^2W^2 = 256$, and therefore the exact attractor equations read

$$p^\Lambda = \mathrm{Re}[CX^\Lambda], \qquad q_\Lambda = \mathrm{Re}\left[CF_\Lambda\left(X^\Lambda, \frac{256}{C^2}\right)\right]. \tag{2.551}$$

This is essentially the only possibility consistent with *symplectic invariance*. It has been then argued that the entropy as a function of the charges is

$$S_{BH} = \frac{\pi \mathrm{i}}{2}(q_\Lambda \bar{C}\bar{X}^\Lambda - p^\Lambda \bar{C}\bar{F}_\Lambda) + \frac{\pi}{2}\,\mathrm{Im}[C^3\partial_C F], \tag{2.552}$$

where F_Λ, X^Λ and C are expressed in terms of the charges using (2.551).

Topological Strings

Partition Functions for Black Hole and Topological Strings. The notion of *topological string* was introduced in [Wit90]. Subsequently a connection between them and superstring was discovered: It was shown in [BCO94, AGN94], that the superstring loop corrected $F-$terms (2.549) can be computed as topological string amplitudes. The purpose of this subsection is to translate the super–gravity notation of the previous section to the topological string notation.

The second quantized partition function for the topological string may be written

$$Z_{\rm top}(t^A, g_{\rm top}) = \exp\left[F_{\rm top}(t^A, g_{\rm top})\right], \qquad \text{where}$$
$$F_{\rm top}(t^A, g_{\rm top}) = \sum_h g_{\rm top}^{2h-2}\, F_{{\rm top},h}(t^A),$$

and $F_{{\rm top},h}$ is the $h-$loop *topological string amplitude*. The Kähler moduli are expressed in the flat coordinates

$$t^A = \frac{X^A}{X^0} = \theta^A + {\rm i}r^A,$$

where r^A are the Kähler classes of the Calabi–Yau M and θ^A are periodic $\theta^A \sim \theta^A + 1$.

We would like to determine relations between super–gravity quantities and topological string quantities. Using the homogeneity property (2.550) and the expansion (2.549), the holomorphic prepotential in super–gravity can be expressed as

$$\begin{aligned} F(CX^\Lambda, 256) &= (CX^0)^2 F\left(\frac{X^\Lambda}{X^0}, \frac{256}{(CX^0)^2}\right) \\ &= \sum_{h=0}^{\infty} (CX^0)^{2-2h} f_h(t^A), \end{aligned} \tag{2.553}$$

where $f_h(t^A)$ is related to $F_h(X^\Lambda)$ in (2.549) as

$$f_h(t^A) = 16^{2h} F_h\left(\frac{X^\Lambda}{X^0}\right).$$

This suggests an identification of the form $f_h(t^A) \sim F_{{\rm top},h}(t^A)$ and $g_{\rm top} \sim (CX^0)^{-1}$. For later purposes, we need precise relations between super–gravity and topological string quantities, including numerical coefficients. These can be determined by studying the limit of a large Calabi–Yau space.

In the *super–gravity notation*, the genus 0 and 1 terms in the large volume are given by

$$\begin{aligned} F\left(CX^\Lambda, 256\right) &= C^2 D_{ABC} \frac{X^A X^B X^C}{X^0} - \frac{1}{6} c_{2A} \frac{X^A}{X^0} + \cdots \\ &= (CX^0)^2 D_{ABC} t^A t^B t^C - \frac{1}{6} c_{2A} t^A + \cdots, \\ \text{where} \quad c_{2A} &= \int_M c_2 \wedge \alpha_A, \end{aligned}$$

with c_2 being the *second Chern class* of M, and $C_{ABC} = -6D_{ABC}$ are the 4–cycle intersection numbers. These terms are normalized so that the mixed entropy S_{BH} is given by (2.552). On the other hand, the topological string amplitude in this limit is given by

$$F_{\rm top} = -\frac{(2\pi)^3 {\rm i}}{g_{\rm top}^2} D_{ABC} t^A t^B t^C - \frac{\pi {\rm i}}{12} c_{2A} t^A + \cdots \tag{2.554}$$

The normalization here is fixed by the holomorphic anomaly equations in [BCO94], which are nonlinear equations for $F_{{\rm top},h}$.

Comparing the one–loop terms in (2.553) and (2.554), which are independent of $g_{\rm top}$, we find

$$F(CX^\Lambda, 256) = -\frac{2{\rm i}}{\pi} F_{\rm top}(t^A, g_{\rm top}).$$

Given this, we can compare the genus 0 terms to find

$$g_{\rm top} = \pm \frac{4\pi {\rm i}}{CX^0}.$$

This implies

$$\begin{aligned} \ln Z_{BH} &= -\pi \ {\rm Im}\left[F(CX^\Lambda, 256)\right] = F_{\rm top} + \bar{F}_{\rm top} \qquad \text{and} \\ Z_{BH}(\phi^\Lambda, p^\Lambda) &= |Z_{\rm top}(t^A, g_{\rm top})|^2, \qquad \text{with} \\ t^A &= \frac{p^A + {\rm i}\phi^A/\pi}{p^0 + {\rm i}\phi^0/\pi}, \qquad g_{\rm top} = \pm \frac{4\pi {\rm i}}{p^0 + {\rm i}\phi^0/\pi}. \end{aligned}$$

Supergravity Approach to Z_{BH}. The above relation

$$Z_{BH} = |Z_{\rm top}|^2 \tag{2.555}$$

can have a simpler super–gravity derivation [OSV04].

A main ingredient in this derivation is the observation that the $N = 2$ super–gravity coupled to vector multiplets can be written as the action

$$S = \int d^4x d^4\theta \ ({\rm super}--{\rm volume\ form}) + h.c. = \int d^4x \sqrt{-g} R + ..., \tag{2.556}$$

where the super–volume form in the above depends non–trivially on curvature of the fields. This reproduces the ordinary action after integrating over $d^4\theta$ and picking up the θ^4 term in the super–volume. In the context of black holes the boundary terms accompanying (2.556) give the classical black hole entropy.

We now become the derivation of (2.555). As was observed in [BCO94, AGN94], topological string computes the terms

$$F = \sum_{h=0}^{\infty} \int d^4x d^4\theta F_h(X) (W^2)^g + c.c. \tag{2.557}$$

There are various terms one can get from the above action after integrating over $d^4\theta$. Let us concentrate on one of the terms which turns out to be the relevant one for us: Take the top components of X^Λ and W^2, and absorb the $d^4\theta$ integral from the super–volume measure as in (2.556). We will work in

the gauge $X^0 \sim 1$ and thus $C \sim 1/g_{\rm top}$. As noted before in the near–horizon black hole geometry in this gauge the top component $W^2 \sim 1/C^2 \sim g_{\rm top}^2$ and the X^Λ are fixed by the attractor mechanism. Thus, we have the black hole free energy

$$\begin{aligned}\ln Z_{BH} &= \sum_{h=0}^{\infty} g_{\rm top}^{2h} F_{{\rm top},h}(X^\Lambda/X^0) \int d^4x d^4\theta + c.c. \\ &= \sum_{g=0}^{\infty} (g_{\rm top})^{2h-2} F_{{\rm top},h}(X^\Lambda/X^0) + c.c. \\ &= 2\ {\rm Re}\ F_{\rm top}, \qquad ({\rm using} \int d^4x d^4\theta \sim 1/g_{\rm top}^2).\end{aligned}$$

Upon exponentiation this leads to (2.555).

Here we have shown that if we consider one absorption of θ^4 term in (2.557) upon $d^4\theta$ integral we get the desired result. That there be no other terms is not obvious. For example another way to absorb the θ's would have given the familiar term R^2F^{2g-2} where F is the graviphoton field. However, such terms do not contribute in the black hole background. It would be nice to find a simple way to argue why these terms do not contribute and that we are left with this simple absorption of the θ integrals.

2.2.9 Turbulence and Chaos Field Theory

Recall that *chaos theory*, of which *turbulence* is the most extreme form, started in 1963, when Ed Lorenz from MIT took the *Navier–Stokes equations* from viscous fluid dynamics and reduced them into three first–order coupled nonlinear ODEs (see subsection 2.1.3 above), to demonstrate the idea of sensitive dependence upon initial conditions and associated *chaotic behavior*.

It is well–known that the viscous fluid evolves according to the nonlinear Navier–Stokes PDEs[19]

$$\dot{\mathbf{u}} + \mathbf{u} \cdot \boldsymbol{\nabla} \mathbf{u} + \nabla p/\rho = \nu \Delta \mathbf{u} + \mathbf{f}, \tag{2.558}$$

[19] Recall that the Navier–Stokes equations, named after C.L. Navier and G.G. Stokes, are a set of PDEs that describe the motion of liquids and gases, based on the fact that changes in momentum of the particles of a fluid are the product of changes in pressure and dissipative viscous forces acting inside the fluid. These viscous forces originate in molecular interactions and dictate how viscous a fluid is, so the Navier–Stokes PDEs represent a dynamical statement of the balance of forces acting at any given region of the fluid. They describe the physics of a large number of phenomena of academic and economic interest (they are useful to model weather, ocean currents, water flow in a pipe, motion of stars inside a galaxy, flow around an airfoil (wing); they are also used in the design of aircraft and cars, the study of blood flow, the design of power stations, the analysis of the effects of pollution, etc).

where $\mathbf{u} = \mathbf{u}(x^i, t)$, $(i = 1, 2, 3)$ is the fluid 3D velocity, $p = p(x^i, t)$ is the pressure field, ρ, ν are the fluid density and viscosity coefficient, while $\mathbf{f} = \mathbf{f}(x^i, t)$ is the nonlinear external energy source. To simplify the problem, we can impose to $\mathbf{f}$ the so–called *Reynolds condition*, $\langle \mathbf{f} \cdot \mathbf{u} \rangle = \varepsilon$, where ε is the average rate of energy injection.

Fluid dynamicists believe that *Navier–Stokes equations* (2.558) *accurately describe turbulence*. A mathematical proof of the global regularity of the solutions to the Navier–Stokes equations is a very challenging problem and yet such a proof or disproof does not solve the problem of turbulence. However, it may help understanding turbulence. Turbulence is more of a dynamical system problem. We will see below that studies on chaos in PDEs indicate that turbulence can have *Bernoulli shift dynamics* which results in the wandering of a turbulent solution in a fat domain in the phase space; thus, turbulence can not be averaged. The hope is that turbulence can be controlled [Li04].

The first demonstration of existence of an unstable recurrent pattern in a 3D turbulent hydrodynamic flow was performed in [KK01], using the full numerical simulation, a 15,422-dimensional discretization of the 3D Plane Couette turbulence at the *Reynolds number* $Re = 400$.[20] The authors found an important unstable spatio–temporally periodic solution, a single unstable recurrent pattern.

Classical Chaos in Lorenz and Laser ODEs

Before we focus on the turbulent geometry of the Navier–Stokes PDEs (2.558), let us briefly review the Lorenz reduced system of nonlinear ODEs

$$\dot{x} = a(y - x), \qquad \dot{y} = bx - y - xz, \qquad \dot{z} = xy - cz, \tag{2.559}$$

where x, y and z are dynamical variables, constituting the 3D *phase–space* of the *Lorenz flow*; and a, b and c are the parameters of the system. Originally, Lorenz used this model to describe the unpredictable behavior of the weather,

[20] Recall that the Reynolds number Re is the most important dimensionless number in fluid dynamics and provides a criterion for determining *dynamical similarity*. Where two similar objects in perhaps different fluids with possibly different flowrates have similar fluid flow around them, they are said to be dynamically similar. Re is the ratio of inertial forces to viscous forces and is used for determining whether a flow will be *laminar* or *turbulent*. Laminar flow occurs at low Reynolds numbers, where viscous forces are dominant, and is characterized by smooth, constant fluid motion, while turbulent flow, on the other hand, occurs at high Res and is dominated by inertial forces, producing random eddies, vortices and other flow fluctuations. The transition between laminar and turbulent flow is often indicated by a critical Reynolds number (Re_{crit}), which depends on the exact flow configuration and must be determined experimentally. Within a certain range around this point there is a region of gradual transition where the flow is neither fully laminar nor fully turbulent, and predictions of fluid behavior can be difficult.

where x is the rate of convective overturning (convection is the process by which heat is transferred by a moving fluid), y is the horizontal temperature overturning, and z is the vertical temperature overturning; the parameters are: $a \equiv P-$proportional to the Prandtl number (ratio of the fluid viscosity of a substance to its thermal conductivity, usually set at 10), $b \equiv R-$proportional to the Rayleigh number (difference in temperature between the top and bottom of the system, usually set at 28), and $c \equiv K-$a number proportional to the physical proportions of the region under consideration (width to height ratio of the box which holds the system, usually set at 8/3). The Lorenz system (2.559) has the properties:

1. *symmetry*: $(x, y, z) \rightarrow (-x, -y, z)$ for all values of the parameters, and
2. the $z-$axis ($x = y = 0$) is *invariant* (i.e., all trajectories that start on it also end on it).

Nowadays, it is well–known that the Lorenz model is a paradigm for low–dimensional chaos in dynamical systems in synergetics and this model or its modifications are widely investigated in connection with modelling purposes in meteorology, hydrodynamics, laser physics, superconductivity, electronics, oil industry, chemical and biological kinetics, etc.

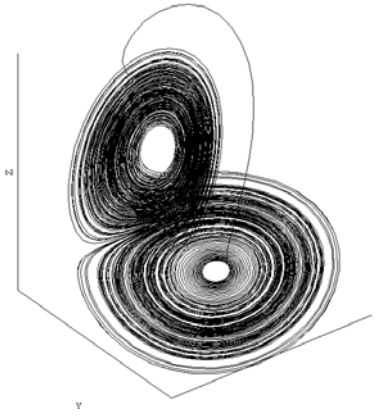


Fig. 2.8. The celebrated 'Lorenz–mask' strange attractor, obtained by simulating the equations (2.559) in $Matlab^{TM}$.

The 3D *phase–portrait* of the Lorenz system (2.8) shows the celebrated '*Lorenz mask*', a special type of (*fractal*, or '*strange*') chaotic *Lorenz attractor* (see Figure 2.8). It depicts the famous '*butterfly effect*', (i.e., sensitive dependence on initial conditions) – the idea in meteorology that the flapping of a butterfly's wing will create a disturbance that in the chaotic motion of the atmosphere will become amplified eventually to change the large scale

atmospheric motion, so that the long term behavior becomes impossible to forecast. The Lorenz mask has the following characteristics:

1. Trajectory does not intersect itself in three dimensions,
2. Trajectory is not periodic or transient,
3. General form of the shape does not depend on initial conditions, and
4. Exact sequence of loops is very sensitive to the initial conditions.

In 1975, H. Haken showed in [Hak83, Hak93] that the Lorenz equations (2.8) were isomorphic to the *Maxwell–Haken laser equations*, that were the starting point for *Haken's synergetics*,

$$\dot{E} = \sigma(P - E), \qquad \dot{P} = \beta(ED - P), \qquad \dot{D} = \gamma(\sigma - 1 - D - \sigma EP),$$

Here, the variables in the Lorenz equations, namely x,y and z correspond to the slowly varying amplitudes of the electric field E and polarization P and the inversion D respectively in the Maxwell–Haken equations. The parameters are related via $c = \frac{\gamma}{\beta}$, $a = \frac{\sigma}{\beta}$ and $b = \sigma + 1$, where γ is the relaxation rate of the inversion, β is the relaxation rate of the polarization, σ is the field relaxation rate, and σ represents the normalized pump power.

Turbulent Flow

Recall that in fluid dynamics, *turbulent flow* is a flow regime characterized by low momentum diffusion, high momentum convection, and rapid variation of pressure and velocity in space and time. Flow that is not turbulent is called *laminar flow*. Also, recall that the *Reynolds number Re* characterizes whether flow conditions lead to laminar or turbulent flow. The structure of turbulent flow was first described by A. Kolmogorov. Consider the flow of water over a simple smooth object, such as a sphere. At very low speeds the flow is laminar, i.e., the flow is locally smooth (though it may involve vortices on a large scale). As the speed increases, at some point the transition is made to turbulent (or, chaotic) flow. In turbulent flow, unsteady vortices[21] appear on many scales and interact with each other. Drag due to boundary layer skin friction increases. The structure and location of boundary layer separation often changes, sometimes resulting in a reduction of overall drag. Because laminar–turbulent transition is governed by Reynolds number, the same transition occurs if the size of the object is gradually increased, or the viscosity of the fluid is decreased, or if the density of the fluid is increased.

[21] Recall that a *vortex* can be any circular or rotary flow that possesses vorticity. Vortex represents a spiral whirling motion (i.e., a spinning turbulent flow) with closed streamlines. The shape of media or mass rotating rapidly around a center forms a vortex. It is a flow involving rotation about an arbitrary axis.

Vorticity Dynamics

Vorticity $\omega = \omega(x^i, t)$, $(i = 1, 2, 3)$ is a geometrical concept used in fluid dynamics, which is related to the amount of 'circulation' or 'rotation' in a fluid. More precisely, *vorticity* is the circulation per unit area at a point in the flow field, or formally, $\omega = \nabla \times \mathbf{u}$, where $\mathbf{u} = \mathbf{u}(x^i, t)$ is the fluid velocity. It is a vector quantity, whose direction is (roughly speaking) along the axis of the swirl. The movement of a fluid can be said to be vortical if the fluid moves around in a circle, or in a helix, or if it tends to spin around some axis. Such motion can also be called *solenoidal*. In the atmospheric sciences, vorticity is a property that characterizes large–scale rotation of air masses. Since the atmospheric circulation is nearly horizontal, the 3D vorticity is nearly vertical, and it is common to use the vertical component as a scalar vorticity.

A vortex can be seen in the spiraling motion of air or liquid around a center of rotation. Circular current of water of conflicting tides form vortex shapes. Turbulent flow makes many vortices. A good example of a vortex is the atmospheric phenomenon of a whirlwind or a *tornado*. This whirling air mass mostly takes the form of a helix, column, or spiral. Tornadoes develop from severe thunderstorms, usually spawned from squall lines and *supercell thunderstorms*, though they sometimes happen as a result of a *hurricane*.[22] Another example is a mesovortex on the scale of a few miles (smaller than a hurricane but larger than a tornado). On a much smaller scale, a vortex is usually formed as water goes down a drain, as in a sink or a toilet. This occurs in water as the revolving mass forms a whirlpool.[23] This whirlpool is caused by water flowing out of a small opening in the bottom of a basin or reservoir. This swirling flow structure within a region of fluid flow opens downward from the water surface. In the hydrodynamic interpretation of the behavior of electromagnetic fields, the acceleration of electric fluid in a particular direction creates a positive vortex of magnetic fluid. This in turn creates around itself a corresponding negative vortex of electric fluid.

Dynamical Similarity and Eddies

In order for two flows to be similar they must have the same geometry and equal Reynolds numbers. When comparing fluid behavior at homologous points in a model and a full–scale flow, we have $Re^* = Re$, where quantities marked with * concern the flow around the model and the other the real flow.

[22] Recall that a hurricane is a much larger, swirling body of clouds produced by evaporating warm ocean water and influenced by the Earth's rotation. In particular, polar vortex is a persistent, large–scale cyclone centered near the Earth's poles, in the middle and upper troposphere and the stratosphere. Similar, but far greater, vortices are also seen on other planets, such as the permanent Great Red Spot on Jupiter and the intermittent Great Dark Spot on Neptune.

[23] Recall that a whirlpool is a swirling body of water produced by ocean tides or by a hole underneath the vortex, where water drains out, as in a bathtub.

This allows us to perform experiments with reduced models in water channels or wind tunnels, and correlate the data to the real flows. Note that true dynamic similarity may require matching other dimensionless numbers as well, such as the Mach number used in compressible flows, or the Froude number that governs free-surface flows.

In a turbulent flow, there is a range of scales of the fluid motions, sometimes called *eddies*. A single packet of fluid moving with a bulk velocity is called an *eddy*. The size of the largest scales (eddies) are set by the overall geometry of the flow. For instance, in an industrial smoke–stack, the largest scales of fluid motion are as big as the diameter of the stack itself. The size of the smallest scales is set by Re. As Re increases, smaller and smaller scales of the flow are visible. In the smoke–stack, the smoke may appear to have many very small bumps or eddies, in addition to large bulky eddies. In this sense, Re is an indicator of the range of scales in the flow. The higher the Reynolds number, the greater the range of scales.

Spatio–Temporal Chaos and Turbulence in PDEs

In their first edition of Fluid Mechanics [LL59], Landau and Lifschitz proposed a *route to turbulence* in spatio–temporal fluid systems. Since then, much work, in dynamical systems, experimental fluid dynamics, and many other fields has been done concerning the routes to turbulence. Ever since the discovery of chaos in low–dimensional systems, researchers have been trying to use the concept of chaos to understand turbulence [RT71]. recall that there are two types of fluid motions: laminar flows and turbulent flows. Laminar flows look regular, and turbulent flows are non–laminar and look irregular. Chaos is more precise, for example, in terms of the so–called *Bernoulli shift dynamics*. On the other hand, even in low–dimensional systems, there are solutions which look irregular for a while, and then look regular again. Such a dynamics is often called a *transient chaos*.

Low–dimensional chaos is the starting point of a long journey toward understanding turbulence. To have a better connection between chaos and turbulence, one has to study chaos in PDEs [Li04].

Sine–Gordon Equation

Consider the simple perturbed *sine–Gordon equation* [Li04c]

$$u_{tt} = c^2 u_{xx} + \sin u + \epsilon[-a u_t + \cos t \ \sin^3 u], \tag{2.560}$$

subject to periodic boundary condition

$$u(t, x + 2\pi) = u(t, x) \ ,$$

as well as even or odd constraint,

$$u(t, -x) = u(t, x), \qquad \text{or} \qquad u(t, -x) = -u(t, x) \ ,$$

where u is a real–valued function of two real variables (t, x), c is a real constant, $\epsilon \geq 0$ is a small perturbation parameter, and $a > 0$ is an external parameter. One can view (2.560) as a *flow* (u, u_t) defined in the phase–space manifold $M \equiv H^1 \times L^2$, where H^1 and L^2 are the Sobolev spaces on $[0, 2\pi]$. A point in the phase–space manifold M corresponds to two profiles, $(u(x), u_t(x))$. [Li04c] has proved that there exists a homoclinic orbit $(u, u_t) = h(t, x)$ asymptotic to $(u, u_t) = (0, 0)$. Let us define two orbits segments

$$\eta_0 : \ (u, u_t) = (0, 0) \ , \qquad \text{and} \qquad \eta_1 : \ (u, u_t) = h(t, x) \ , \qquad (t \in [-T, T] \ ).$$

When T is large enough, η_1 is almost the entire homoclinic orbit (chopped off in a small neighborhood of $(u, u_t) = (0, 0)$). To any binary sequence

$$a = \{\cdots a_{-2} a_{-1} a_0, a_1 a_2 \cdots\} \ , \qquad (a_k \in \{0, 1\}), \tag{2.561}$$

one can associate a pseudo–orbit

$$\eta_a = \{\cdots \eta_{a_{-2}} \eta_{a_{-1}} \eta_{a_0}, \eta_{a_1} \eta_{a_2} \cdots\} \ .$$

The pseudo–orbit η_a is not a true orbit but rather 'almost an orbit'. One can prove that for any such pseudo–orbit η_a, there is a unique true orbit in its neighborhood [Li04c]. Therefore, each binary sequence labels a true orbit. All these true orbits together form a chaos. In order to talk about sensitive dependence on initial data, one can introduce the *product topology* by defining the neighborhood basis of a binary sequence

$$a^* = \{\cdots a^*_{-2} a^*_{-1} a^*_0, a^*_1 a^*_2 \cdots\} \qquad \text{as} \qquad \Omega_N = \{a \ : \quad a_n = a^*_n \ , \quad |n| \leq N\} \ .$$

The Bernoulli shift on the binary sequence (2.561) moves the comma one step to the right. Two binary sequences in the neighborhood Ω_N will be of order Ω_1 away after N iterations of the Bernoulli shift. Since the binary sequences label the orbits, the orbits will exhibit the same feature. In fact, the Bernoulli shift is topologically conjugate to the perturbed sine–Gordon flow.

Replacing a homoclinic orbit by its fattened version – a homoclinic tube, or by a heteroclinic cycle, or by a heteroclinically tubular cycle; one can still obtain the same Bernoulli shift dynamics. Also, adding diffusive perturbation $\epsilon b u_{txx}$ to (2.560), one can still prove the existence of homoclinics or heteroclinics, but the Bernoulli shift result has not been established [Li04c].

Complex Ginzburg–Landau Equation

Consider the complex–valued *Ginzburg–Landau equation* [Li04a, Li04b],

$$\mathrm{i} q_t = q_{xx} + 2[|q|^2 - \omega^2] + \mathrm{i}\epsilon[q_{xx} - \alpha q + \beta] \ , \tag{2.562}$$

which is subject to periodic boundary condition and even constraint

$$q(t, x + 2\pi) = q(t, x) \ , \quad q(t, -x) = q(t, x) \ ,$$

where q is a complex-valued function of two real variables (t, x), (ω, α, β) are positive constants, and $\epsilon \geq 0$ is a small perturbation parameter. In this case, one can prove the existence of homoclinic orbits [Li04a]. But the Bernoulli shift dynamics was established under generic assumptions [Li04b].

A real fluid example is the amplitude equation of Faraday water wave, which is also a complex Ginzburg–Landau equation [Li04d],

$$\mathrm{i}q_t = q_{xx} + 2[|q|^2 - \omega^2] + \mathrm{i}\epsilon[q_{xx} - \alpha q + \beta\bar{q}] \ , \tag{2.563}$$

subject to the same boundary condition as (2.562). For the first time, one can prove the existence of homoclinic orbits for a water wave equation (2.563). The Bernoulli shift dynamics was also established under generic assumptions [Li04d]. That is, one can prove the existence of chaos in water waves under generic assumptions.

The nature of the complex Ginzburg–Landau equation is a parabolic equation which is near a hyperbolic equation. The same is true for the perturbed sine–Gordon equation with the diffusive term $\epsilon b u_{txx}$ added. They contain effects of diffusion, dispersion, and nonlinearity. The Navier–Stokes equations are diffusion–advection equations. The advective term is missing from the perturbed sine–Gordon equation and the complex Ginzburg–Landau equation. However, the modified KdV equation (1.52) does contain an advective term. In principle, perturbed modified KdV equation should have the same feature as the perturbed sine–Gordon equation. Turbulence happens when the diffusion is weak, i.e., in the near hyperbolic regime. One should hope that turbulence should share some of the features of chaos in the perturbed sine–Gordon equation. There is a popular myth that turbulence is fundamentally different from chaos because turbulence contains many unstable modes. In both the perturbed sine–Gordon equation and the complex Ginzburg–Landau equation, one can incorporate as many unstable modes as one likes, the resulting Bernoulli shift dynamics is still the same. On a computer, the solution with more unstable modes may look rougher, but it is still chaos [Li04].

In a word, dynamics of strongly nonlinear classical fields is 'turbulent', not 'laminar'.

On the other hand, field theories such as 4-dimensional QCD or gravity have many dimensions, symmetries, tensorial indices. They are far too complicated for exploratory forays into this forbidding terrain. Instead, we consider a simple spatio–temporally chaotic nonlinear system of physical interest [CCP96].

Kuramoto–Sivashinsky System

One of the simplest and extensively studied spatially extended dynamical systems is the Kuramoto–Sivashinsky (KS) system [Kur76, Siv77]

$$u_t = (u^2)_x - u_{xx} - \nu u_{xxxx}, \tag{2.564}$$

which arises as an amplitude equation for interfacial instabilities in a variety of contexts. The so–called *flame front* $u(x,t)$ has compact support, with $x \in [0, 2\pi]$ a periodic space coordinate. The u^2 term makes this a nonlinear system, $t \geq 0$ is the time, and ν is a 4–order 'viscosity' damping parameter that irons out any sharp features. Numerical simulations demonstrate that as the viscosity decreases (or the size of the system increases), the *flame front* becomes increasingly unstable and turbulent. The task of the theory is to describe this spatio-temporal turbulence and yield quantitative predictions for its measurable consequences.

For any finite spatial resolution, the KS system (2.564) follows approximately for a finite time a pattern belonging to a finite alphabet of admissible patterns, and the long term dynamics can be thought of as a walk through the space of such patterns, just as chaotic dynamics with a low dimensional attractor can be thought of as a succession of nearly periodic (but unstable) motions. The periodic orbit gives the machinery that converts this intuitive picture into precise calculation scheme that extracts asymptotic time predictions from the short time dynamics. For extended systems the theory gives a description of the asymptotics of partial differential equations in terms of recurrent spatio–temporal patterns.

The KS periodic orbit calculations of Lyapunov exponents and escape rates [CCP96] demonstrate that the *periodic orbit theory* predicts observable averages for deterministic but classically chaotic spatio–temporal systems. The main problem today is not how to compute such averages – periodic orbit theory as well as direct numerical simulations can handle that – but rather that there is no consensus on what the sensible experimental observables worth are predicting [Cvi00].

Burgers Dynamical System

Consider the following *Burgers dynamical system* on a *functional manifold* $M \subset C^k(\mathbb{R};\mathbb{R})$:

$$u_t = uu_x + u_{xx}, \tag{2.565}$$

where $u \in M$, $t \in \mathbb{R}$ is an evolution parameter. The flow of (2.565) on M can be recast into a set of 2–forms $\{\alpha\} \subset \Lambda^2(J(\mathbb{R}^2;\mathbb{R}))$ upon the adjoint jet–manifold $J(\mathbb{R}^2;\mathbb{R})$ as follows [BPS98]:

$$\begin{aligned}\{\alpha\} = \big\{ & du^{(0)} \wedge dt - u^{(1)} dx \wedge dt = \alpha^1,\ du^{(0)} \wedge dx + u^{(0)} du^{(0)} \wedge dt \\ & + du^{(1)} \wedge dt = \alpha^2 :\ \left(x, t; u^{(0)}, u^{(1)}\right)^\tau \in M^4 \subset J^1(\mathbb{R}^2;\mathbb{R}) \big\},\end{aligned} \tag{2.566}$$

where M^4 is some finite–dimensional submanifold in $J^1(\mathbb{R}^2;\mathbb{R}))$ with coordinates $(x, t,\ u^{(0)} = u, u^{(1)} = u_x)$. The set of 2–forms (2.566) generates the closed ideal $\mathfrak{I}(\alpha)$, since

$$d\alpha^1 = dx \wedge \alpha^2 - u^{(0)} dx \wedge \alpha^1, \qquad d\alpha^2 = 0, \tag{2.567}$$

the integral submanifold $\bar{M} = \{x, t \in \mathbb{R}\} \subset M^4$ being defined by the condition $\mathfrak{I}(\alpha) = 0$. We now look for a reduced 'curvature' 1–form $\Gamma \in \Lambda^1(M^4) \otimes \mathcal{G}$, belonging to some not yet determined Lie algebra $\mathcal{G}$. This 1–form can be represented using (2.566), as follows:

$$\Gamma = b^{(x)}(u^{(0)}, u^{(1)})dx + b^{(t)}(u^{(0)}, u^{(1)})dt, \tag{2.568}$$

where elements $b^{(x)}, b^{(t)} \in \mathcal{G}$ satisfy such determining equations [BPS98]

$$\begin{array}{lll} \frac{\partial b^{(x)}}{\partial u^{(0)}} = g_2, & \frac{\partial b^{(x)}}{\partial u^{(1)}} = 0, & \frac{\partial b^{(t)}}{\partial u^{(0)}} = g_1 + g_2 u^{(0)}, \\ \frac{\partial b^{(t)}}{\partial u^{(1)}} = g_2, & [b^{(x)}, b^{(t)}] = -u^{(1)} g_1. & \end{array} \tag{2.569}$$

The set (2.569) has the following unique solution

$$\begin{array}{l} b^{(x)} = A_0 + A_1 u^{(0)}, \\ b^{(t)} = u^{(1)} A_1 + \frac{{u^{(0)}}^2}{2} A_1 + [A_1, A_0] u^{(0)} + A_2, \end{array} \tag{2.570}$$

where $A_j \in \mathcal{G}$, $j = \overline{0,2}$, are some constant elements on M of a Lie algebra $\mathcal{G}$ under search, enjoying the next Lie structure equations (see subsection 1.2.8 above):

$$\begin{array}{l} [A_0, A_2] = 0, \\ [A_0, [A_1, A_0]] + [A_1, A_2] = 0, \\ [A_1, [A_1, A_0]] + \frac{1}{2}[A_0, A_1] = 0. \end{array} \tag{2.571}$$

From (2.569) one can see that the curvature 2–form $\Omega \in span_{\mathbb{R}}\{A_1, [A_0, A_1] : A_j \in \mathcal{G},\ j = 0, 1\}$. Therefore, reducing via the *Ambrose–Singer theorem* the associated principal fibred frame space $P(M; G = GL(n))$ to the principal fibre bundle $P(M; G(h))$, where $G(h) \subset G$ is the corresponding holonomy Lie group of the connection Γ on P, we need to satisfy the following conditions for the set $\mathcal{G}(h) \subset \mathcal{G}$ to be a Lie subalgebra in $\mathcal{G}$: $\nabla_x^m \nabla_t^n \Omega \in \mathcal{G}(h)$ for all $m, n \in \mathbb{Z}_+$.

Let us try now to close the above transfinitive procedure requiring that [BPS98]

$$\mathcal{G}(h) = \mathcal{G}(h)_0 = span_{\mathbb{R}}\{\nabla_x^m \nabla_x^n \Omega \in \mathcal{G} :\ m + n = 0\} \tag{2.572}$$

This means that

$$\mathcal{G}(h)_0 = span_{\mathbb{R}}\{A_1, A_3 = [A_0, A_1]\}. \tag{2.573}$$

To enjoy the set of relations (2.571) we need to use expansions over the basis (2.573) of the external elements $A_0, A_2 \in \mathcal{G}(h)$:

$$A_0 = q_{01} A_1 + q_{13} A_3, \qquad A_2 = q_{21} A_1 + q_{23} A_3. \tag{2.574}$$

Substituting expansions (2.574) into (2.571), we get that $q_{01} = q_{23} = \lambda,\ q_{21} = -\lambda^2/2$ and $q_{03} = -2$ for some arbitrary real parameter $\lambda \in \mathbb{R}$, that is $\mathcal{G}(h) = span_{\mathbb{R}}\{A_1, A_3\}$, where

$$[A_1, A_3] = A_3/2; \qquad A_0 = \lambda A_1 - 2A_3, \qquad A_2 = -\lambda^2 A_1/2 + \lambda A_3. \quad (2.575)$$

As a result of (2.575) we can state that the holonomy Lie algebra $\mathcal{G}(h)$ is a real 2D one, assuming the following $(2 \times 2)-$matrix representation [BPS98]:

$$\begin{aligned} A_1 &= \begin{pmatrix} 1/4 & 0 \\ 0 & -1/4 \end{pmatrix}, \qquad A_3 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \\ A_0 &= \begin{pmatrix} \lambda/4 & -2 \\ 0 & -\lambda/4 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} -\lambda^2/8 & \lambda \\ 0 & \lambda^2/8 \end{pmatrix}. \end{aligned} \quad (2.576)$$

Thereby from (2.568), (2.570) and (2.576) we obtain the *reduced curvature 1–form* $\Gamma \in \Lambda^1(M) \otimes \mathcal{G}$,

$$\Gamma = (A_0 + uA_1)dx + ((u_x + u^2/2)A_1 - uA_3 + A_2)dt, \quad (2.577)$$

generating *parallel transport* of vectors from the representation space Y of the holonomy Lie algebra $\mathcal{G}(h)$:

$$dy + \Gamma y = 0 \quad (2.578)$$

upon the integral submanifold $\bar{M} \subset M^4$ of the ideal $\mathcal{I}(\alpha)$, generated by the set of 2–forms (2.566). The result (2.578) means also that the Burgers dynamical system (2.565) is endowed with the standard Lax type representation, having the spectral parameter $\lambda \in \mathbb{R}$ necessary for its integrability in quadratures.

Nonlinear Fluid Dynamics

In this subsection we will derive the general form of the Navier–Stokes equations (2.558) in nonlinear fluid dynamics.

Continuity Equation

Recall that the most important equation in fluid dynamics, as well as in general continuum mechanics, is the celebrated *equation of continuity*, (we explain the symbols in the following text)

$$\partial_t \rho + \mathrm{div}(\rho \mathbf{u}) = 0. \quad (2.579)$$

As a warm–up for turbulence, we will derive the continuity equation (2.579), starting from the *mass conservation principle*. Let dm denote an infinitesimal mass of a fluid particle. Then, using the absolute time derivative operator $\overline{\dot{(\ )}} \equiv \frac{D}{dt}$ (see Appendix), the mass conservation principle reads

$$\overline{\dot{dm}} = 0. \quad (2.580)$$

If we further introduce the fluid density $\rho = dm/dv$, where dv is an infinitesimal volume of a fluid particle, then the mass conservation principle (2.580) can be rewritten as

$$\dot{\overline{\rho dv}} = 0,$$

which is the absolute derivative of a product, and therefore expands into

$$\dot{\rho} dv + \rho \dot{\overline{dv}} = 0. \tag{2.581}$$

Now, as the fluid density $\rho = \rho(x^k, t)$ is a function of both time t and spatial coordinates x^k, for $k = 1, 2, 3$, that is, a *scalar–field*, its total time derivative $\dot{\rho}$, figuring in (2.581), is defined by

$$\begin{aligned} \dot{\rho} &= \partial_t \rho + \partial_{x^k} \rho \, \partial_t x^k \equiv \partial_t \rho + \rho_{;k} u^k, \\ \text{or, in vector form} \qquad \dot{\rho} &= \partial_t \rho + \operatorname{grad}(\rho) \cdot \mathbf{u}, \end{aligned} \tag{2.582}$$

where $u^k = u^k(x^k, t) \equiv \mathbf{u}$ is the velocity vector–field of the fluid.

Regarding $\dot{\overline{dv}}$, the other term figuring in (2.581), we start by expanding an elementary volume dv along the sides $\{dx^i_{(p)}, dx^j_{(q)}, dx^k_{(r)}\}$ of an elementary parallelepiped, as

$$dv = \frac{1}{3!} \delta^{pqr}_{ijk} dx^i_{(p)} dx^j_{(q)} dx^k_{(r)}, \qquad (i, j, k, p, q, r = 1, 2, 3)$$

so that its absolute derivative becomes

$$\begin{aligned} \dot{\overline{dv}} &= \frac{1}{2!} \delta^{pqr}_{ijk} \dot{\overline{dx^i}}_{(p)} dx^j_{(q)} dx^k_{(r)} \\ &= \frac{1}{2!} u^i_{;l} \delta^{pqr}_{ijk} dx^l_{(p)} dx^j_{(q)} dx^k_{(r)} \qquad (\text{using } \dot{\overline{dx^i}}_{(p)} = u^i_{;l} dx^l_{(p)}), \end{aligned}$$

which finally simplifies into

$$\dot{\overline{dv}} = u^k_{;k} dv \equiv \operatorname{div}(\mathbf{u})\, dv. \tag{2.583}$$

Substituting (2.582) and (2.583) into (2.581) gives

$$\dot{\overline{\rho dv}} \equiv \left(\partial_t \rho + \rho_{;k} u^k\right) dv + \rho u^k_{;k} dv = 0. \tag{2.584}$$

As we are dealing with arbitrary fluid particles, $dv \neq 0$, so from (2.584) follows

$$\partial_t \rho + \rho_{;k} u^k + \rho u^k_{;k} \equiv \partial_t \rho + (\rho u^k)_{;k} = 0. \tag{2.585}$$

Equation (2.585) is the covariant form of the continuity equation, which in standard vector notation becomes (2.579), i.e., $\partial_t \rho + \operatorname{div}(\rho \mathbf{u}) = 0$.

Forces Acting on a Fluid

A fluid contained in a finite volume is subject to the action of both *volume forces* F^i and *surface forces* S^i, which are respectively defined by

$$F^i = \int_v \rho f^i dv, \qquad \text{and} \qquad S^i = \oint_a \sigma^{ij} da_j. \tag{2.586}$$

Here, f^i is a force vector acting on an elementary mass dm, so that the elementary volume force is given by

$$dF^i = f^i dm = \rho f^i dv,$$

which is the integrand in the volume integral on l.h.s of (2.586). $\sigma^{ij} = \sigma^{ij}(x^k, t)$ is the stress tensor–field of the fluid, so that the elementary force acting on the closed oriented surface a is given by

$$dS^i = \sigma^{ij} da_j,$$

where da_j is an oriented element of the surface a; this is the integrand in the surface integral on the r.h.s of (2.586).

On the other hand, the elementary momentum dK^i of a fluid particle (with elementary volume dv and elementary mass $dm = \rho dv$) equals the product of dm with the particle's velocity u^i, i.e.,

$$dK^i = u^i dm = \rho u^i dv,$$

so that the total momentum of the finite fluid volume v is given by the volume integral

$$K^i = \int_v \rho u^i dv. \tag{2.587}$$

Now, the *Newtonian–like force law for the fluid* states that the time derivative of the fluid momentum equals the resulting force acting on it, $\dot{K}^i = \mathcal{F}^i$, where the resulting force $\mathcal{F}^i$ is given by the sum of surface and volume forces,

$$\mathcal{F}^i = S^i + F^i = \oint_a \sigma^{ij} da_j + \int_v \rho f^i dv. \tag{2.588}$$

From (2.587), taking the time derivative and using $\dot{\overline{\rho dv}} = 0$, we get

$$\dot{K}^i = \int_v \rho \dot{u}^i dv,$$

where $\dot{u}^i = \dot{u}^i(x^k, t) \equiv \dot{\mathbf{u}}$ is the acceleration vector–field of the fluid, so that (2.588) gives

$$\oint_a \sigma^{ij} da_j + \int_v \rho(f^i - \dot{u}^i) dv = 0. \tag{2.589}$$

Now, assuming that the stress tensor $\sigma^{ij} = \sigma^{ij}(x^k, t)$ does not have any singular points in the volume v bounded by the closed surface a, we can transform the surface integral in (2.589) in the volume one, i.e.,

$$\oint_a \sigma^{ij} da_j = \int_v \sigma^{ij}_{;j} dv, \tag{2.590}$$

where $\sigma^{ij}_{;j}$ denotes the *divergence of the stress tensor*. The expression (2.590) shows us that the resulting surface force acting on the closed surface a equals the flux of the stress tensor through the surface a. Using this expression, we can rewrite (2.589) in the form

$$\int_v \left(\sigma^{ij}_{;j} + \rho f^i - \rho \dot{u}^i\right) dv = 0.$$

As this equation needs to hold for an arbitrary fluid element $dv \neq 0$, it implies the dynamical equation of motion for the fluid particles, also called the *first Cauchy law of motion*,

$$\sigma^{ij}_{;j} + \rho f^i = \rho \dot{u}^i. \tag{2.591}$$

Constitutive and Dynamical Equations

Recall that, in case of a homogenous isotropic viscous fluid, the stress tensor σ^{ij} depends on the strain–rate tensor–field $e^{ij} = e^{ij}(x^k, t)$ of the fluid in such a way that

$$\sigma^{ij} = -pg^{ij}, \qquad \text{when} \qquad e^{ij} = 0,$$

where the scalar function $p = p(x^k, t)$ represents the pressure field. Therefore, pressure is independent on the strain–rate tensor e^{ij}. Next, we introduce the viscosity tensor–field $\beta^{ij} = \beta^{ij}(x^k, t)$, as

$$\beta^{ij} = \sigma^{ij} + pg^{ij}, \tag{2.592}$$

which depends exclusively on the strain–rate tensor (i.e., $\beta^{ij} = 0$ whenever $e^{ij} = 0$). A viscous fluid in which the viscosity tensor β^{ij} can be expressed as a function of the strain–rate tensor e^{ij} in the form

$$\beta^{ij} = \alpha_1(e_I, e_{II}, e_{III})g^{ij} + \alpha_2(e_I, e_{II}, e_{III})e^{ij} + \alpha_3(e_I, e_{II}, e_{III})e^i_k e^{kj}, \tag{2.593}$$

where $\alpha_l = \alpha_l(e_I, e_{II}, e_{III})$, $(l = 1, 2, 3)$ are scalar functions of the basic invariants (e_I, e_{II}, e_{III}) of the strain–rate tensor e^{ij}, is called the *Stokes fluid*.

If we take only the linear terms in (2.593), we get the constitutive equation for the *Newtonian fluid*,

$$\beta^{ij} = \alpha_1 e_I g^{ij} + \alpha_2 e^{ij}, \tag{2.594}$$

which is, therefore, a linear approximation of the constitutive equation (2.593) for the Stokes fluid.

If we now put (2.594) into (2.592) we get the dynamical equation for the Newtonian fluid,

$$\sigma^{ij} = -pg^{ij} + \mu e_I g^{ij} + 2\eta e^{ij}, \tag{2.595}$$

If we put $\mu = \eta_V - \frac{2}{3}\eta$, where η_V is called the *volume viscosity* coefficient, while η is called the *shear viscosity* coefficient, we can rewrite (2.595) as

$$\sigma^{ij} = -pg^{ij} + \left(\eta_V - \frac{2}{3}\eta\right) e_I g^{ij} + 2\eta e^{ij}. \tag{2.596}$$

Navier–Stokes Equations

From the constitutive equation of the Newtonian viscous fluid (2.596), by taking the divergence, we get

$$\sigma^{ij}_{;j} = -p_{;j} g^{ij} + \left(\eta_V - \frac{2}{3}\eta\right) e_{I;j} g^{ij} + 2\eta e^{ij}_{;j}.$$

However, as $e_{I;j} = u^k_{;kj}$ as well as

$$e^{ij}_{;j} = \frac{1}{2}(u^{i;j} + u^{j;i})_{;j} = \frac{1}{2}(u^{i;j}_j + u^{j;i}_j) = \frac{1}{2}\Delta u^i + \frac{1}{2}u^{ki}_{;k}$$

we get

$$\sigma^{ij}_{;j} = -p_{;j} g^{ij} + \left(\eta_V - \frac{2}{3}\eta\right) u^k_{;kj} g^{ij} + \eta \Delta u^i + \eta u^k_{;kj} g^{ij},$$

$$\text{or} \qquad \sigma^{ij}_{;j} = -p_{;j} g^{ij} + \left(\eta_V - \frac{1}{3}\eta\right) u^k_{;kj} g^{ij} + \eta \Delta u^i.$$

If we now substitute this expression into (2.591) we get

$$\rho \dot{u}^i = \rho f^i - p_{;j} g^{ij} + \left(\eta_V - \frac{1}{3}\eta\right) u^k_{;kj} g^{ij} + \eta \Delta u^i, \tag{2.597}$$

that is a system of 3 scalar PDEs called the *Navier–Stokes equations*, which in vector form read

$$\rho \dot{\mathbf{u}} = \rho \mathbf{f} - \operatorname{grad} p + \left(\eta_V - \frac{1}{3}\eta\right) \operatorname{grad}(\operatorname{div} \mathbf{u}) + \eta \Delta \mathbf{u}. \tag{2.598}$$

In particular, for incompressible fluids, $\operatorname{div} \mathbf{u} = 0$, we have

$$\dot{\mathbf{u}} = \mathbf{f} - \frac{1}{\rho} \operatorname{grad} p + \nu \Delta \mathbf{u}, \qquad \text{where} \qquad \nu = \frac{\eta}{\rho} \tag{2.599}$$

is the coefficient of kinematic viscosity.

Computational Fluid Dynamics

It is possible to numerically solve the Navier–Stokes equations (2.598–2.599) for laminar flow cases and for turbulent flows when all of the relevant length scales can be contained on the grid (a direct numerical simulation).[24] In general however, the range of length scales appropriate to the problem is larger than even today's massively parallel computers can model. In these cases, turbulent flow simulations require the introduction of a turbulence model. The so–called *Reynolds–averaged Navier–Stokes equations*,[25] and *large eddy simulations*,[26] are two techniques for dealing with these scales.

In many instances, other equations (mostly convective-diffusion equations) are solved simultaneously with the Navier–Stokes equations. These other equations can include those describing species concentration, chemical reactions, heat transfer, etc. More advanced codes allow the simulation of more complex cases involving multi-phase flows (e.g., liquid/gas, solid/gas, liquid/solid) or non–Newtonian fluids, such as blood.

In all of these approaches the same basic procedure is followed: (i) the geometry of the problem is defined, (ii) the volume occupied by the fluid is divided into the mesh of discrete cells, (iii) the physical modelling is defined (e.g., the equations of motions + enthalpy + species conservation), (iv) boundary conditions are defined (this involves specifying the fluid behavior

[24] Direct numerical simulation (DNS) captures all of the relevant scales of turbulent motion, so no model is needed for the smallest scales. This approach is extremely expensive, if not intractable, for complex problems on modern computing machines, hence the need for models to represent the smallest scales of fluid motion.

[25] Reynolds–averaged Navier–Stokes equations (RANS) is the oldest approach to turbulence modelling. In this method, an ensemble version of the governing equations is solved, which introduces new apparent stresses known as Reynolds stress. This adds a second–order tensor of unknown variables for which various models can give different levels of closure. It is a common misconception that the RANS equations do not apply to flows with a time-varying mean flow because these equations are 'time–averaged'. In fact, non–stationary flows can equally be treated. This is sometimes referred to as URANS. There is nothing inherent in Reynolds averaging to preclude this, but the turbulence models used to close the equations are valid only as long as the time over which these changes in the mean occur is large compared to the time scales of the turbulent motion containing most of the energy.

[26] Large eddy simulations (LES) is a technique in which the smaller eddies are filtered and modelled using a sub–grid scale model, while the larger energy carrying eddies are simulated. This method generally requires a more refined mesh than a RANS model, but a far coarser mesh than a DNS solution. The so–called detached eddy simulations (DES) is a modification of RANS, in which the model switches to a subgrid scale formulation in regions fine enough for LES calculations. Regions near solid boundaries and where the turbulent length scale is less than the maximum grid dimension are assigned the RANS mode of solution. As the turbulent length scale exceeds the grid dimension, the regions are solved using the LES mode, so the grid resolution is not as demanding as pure LES.

and properties at the boundaries of the problem; for transient problems, the initial conditions are also defined), (v) the equations are solved iteratively as either steady–state or transient, (vi) the resulting numerical solution is visualized and further analyzed using 3D computer graphics methods.

Common discretization methods currently in use are:
(i) *Finite volume method* is the standard approach used most often in commercial software and research codes, in which the governing equations are solved on discrete control volumes. This integral approach yields a method that is inherently conservative (i.e., quantities such as density remain physically meaningful),

$$\frac{\partial}{\partial t}\iiint QdV + \iint Fd\mathbf{A} = 0,$$

where Q is the vector of conserved variables, and F is the vector of fluxes.
(ii) *Finite element method* (FEM) is popular for structural analysis of solids, but is also applicable to fluids. The FEM formulation requires, however, special care to ensure a conservative solution.
(iii) *Finite difference method* has historical importance and is simple to program.
(iv) *Boundary element method*, in which the boundary occupied by the fluid is divided into surface mesh.

Chaos Field Theory

In [Cvi00], Cvitanovic re–examined the path–integral formulation and the role that the classical solutions play in quantization of strongly nonlinear fields. In the path integral formulation of a field theory the dominant contributions come from saddle–points, the classical solutions of equations of motion. Usually one imagines *one dominant saddle point*, the 'vacuum' (see Figure 2.9, (a)).

The *Feynman diagrams* of quantum electrodynamics (QED) and quantum chromodynamics (QCD), associated to their path integrals, give us a visual and intuitive scheme to calculate the correction terms to this starting semiclassical, *Gaussian saddlepoint approximation*. But there might be other saddles (Figure 2.9, (b)). That field theories might have a rich repertoire of classical solutions became apparent with the discovery of *instantons* [BPS75], analytic solutions of the classical $SU(2)$ *Yang–Mills relation*, and the realization that the associated *instanton vacua* receive contributions from countable ∞'s of saddles. What is not clear is whether these are the important classical saddles. Cvitanovic asks the question: could it be that the strongly nonlinear theories are dominated by altogether different classical solutions?

The search for the classical solutions of nonlinear field theories such as the *Yang–Mills* and *gravity* has so far been neither very successful nor very systematic. In modern field theories the main emphasis has been on symmetries (compactly collected in action functionals that define the theories) as guiding

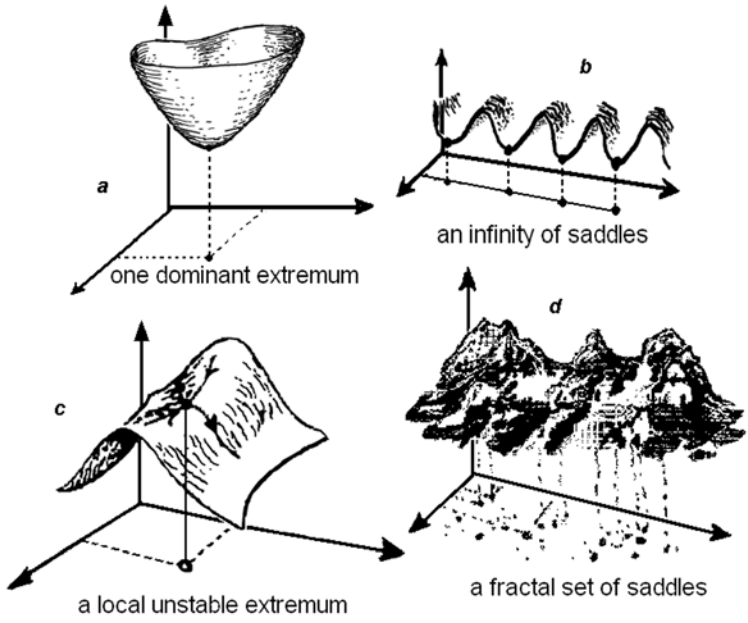


Fig. 2.9. Path integrals and chaos field theory (see text for explanation).

principles in writing down the actions. But writing down a differential equation is only the start of the story; even for systems as simple as 3 coupled ordinary differential equations one in general has no clue what the nature of the long time solutions might be.

These are hard problems, and in explorations of modern field theories the dynamics tends to be is neglected, and understandably so, because the wealth of the classical solutions of nonlinear systems can be truly bewildering. If the classical behavior of these theories is anything like that of the field theories that describe the classical world – the hydrodynamics, the magneto–hydrodynamics, the *Burgers dynamical system* (2.565), *Ginzburg–Landau equation* (2.562), or *Kuramoto–Sivashinsky equation* (2.564), there should be very many solutions, with very few of the important ones analytical in form; the strongly nonlinear classical field theories are turbulent, after all. Furthermore, there is not a dimmest hope that such solutions are either beautiful or analytic, and there is not much enthusiasm for grinding out numerical solutions as long as one lacks ideas as what to do with them.

By late 1970's it was generally understood that even the simplest nonlinear systems exhibit chaos. Chaos is the norm also for generic Hamiltonian flows, and for path integrals that implies that instead of a few, or countably few saddles (Figure 2.9, (c)), classical solutions populate fractal sets of saddles (Figure 2.9, (d)). For the path–integral formulation of quantum mechanics such solutions were discovered and accounted for by [Gut90] in late 1960's.

In this framework the spectrum of the theory is computed from a set of its unstable classical periodic solutions and quantum corrections. The new aspect is that the individual saddles for classically chaotic systems are nothing like the harmonic oscillator degrees of freedom, the quarks and gluons of QCD – they are all unstable and highly nontrivial, accessible only by numerical techniques.

So, if one is to develop a semiclassical field theory of systems that are *classically chaotic* or *turbulent*, the problem one faces is twofold [Cvi00]

1. Determine, classify, and order by relative importance the classical solutions of nonlinear field theories.
2. Develop methods for calculating perturbative corrections to the corresponding classical saddles.

2.3 Nonlinear Control Systems

2.3.1 The Basis of Modern Geometrical Control

Introduction to Geometrical Nonlinear Control

In this section we give a brief introduction to *geometrical nonlinear control systems*. Majority of techniques developed under this name consider the so–called *affine nonlinear MIMO–systems* of the form (see [Isi89, NS90, Lew95, LM97, Lew98])

$$\dot{x}(t) = f_0(x(t)) + u^i(t) f_i(x(t)), \qquad (i = 1, ..., m) \tag{2.600}$$

where $t \mapsto x(t)$ is a curve in a system's *state manifold* M. The vector–field f_0 is called the *drift vector–field*, describing the dynamics of the system in the absence of controls, and the vector–fields $f_1, ..., f_m$ are the *input vector–fields* or *control vector–fields*, indicating how we are able to actuate the system. The vector–fields $f_0, f_1, ..., f_m$ are assumed to be real analytic. We do not ask for any sort of linear independence of the control vector–fields $f_1, ..., f_m$. We shall suppose that the controls $u : [0, T] \to U$ are locally integrable with U some subset of $\mathbb{R}^m$. We allow the length T of the interval on which the control is defined to be arbitrary. It is convenient to denote by $\tau(u)$ the right endpoint of the interval for a given control u. For a fixed U we denote by $\mathcal{U}$ the collection of all measurable controls taking their values in U. To be concise about this, a *control affine system* is a triple $\Sigma = (M, \mathcal{F} = \{f_0, f_1, ..., f_m\}, U)$, with all objects as defined above. A *controlled trajectory* for Σ is a pair (c, u), where $u \in \mathcal{U}$ and where $c : [0, \tau(u)] \to M$ is defined so that

$$\dot{c}(t) = f_0(c(t)) + u^i(t) f_i(c(t)).$$

One can show that for admissible controls, the curve c will exist at least for sufficiently small times, and that the initial condition $c(0) = x_0$ uniquely defines c on its domain of definition.

For $x \in M$ and $T > 0$ we define several types of *reachable sets* as:

$$\mathcal{R}_\Sigma(x,T) = \{c(T) : \\ (c,u) \text{ is a controlled trajectory for } \Sigma \text{ with } \tau(u) = T \text{ and } c(0) = x\},$$
$$\mathcal{R}_\Sigma(x, \leq T) = \cup_{t\in[0,T]}\mathcal{R}_\Sigma(x,t), \qquad \mathcal{R}_\Sigma(x) = \cup_{t\geq 0}\mathcal{R}_\Sigma(x,t),$$

that allow us to give several definitions of controllability as follows. Let $\Sigma = (M, \mathcal{F}, U)$ be a control affine system and let $x \in M$. We say that:

1. Σ is *accessible* from x if $\text{int}(\mathcal{R}_\Sigma(x)) \neq 0$.
2. Σ is *strongly accessible* from x if $\text{int}(\mathcal{R}_\Sigma(x,T)) \neq 0$ for each $T > 0$.
3. Σ is *locally controllable* from x if $x \in \text{int}(\mathcal{R}_\Sigma(x))$.
4. Σ is *small–time locally controllable* (STLC) from x if there exists $T > 0$ so that $x \in \text{int}(\mathcal{R}_\Sigma(x, \leq T))$ for each $t \in [0,T]$.
5. Σ is *globally controllable* from x if $(\mathcal{R}_\Sigma(x)) = M$.

For example, a typical simple system that is accessible but not controllable is given by the following data:

$$M = \mathbb{R}^2,\ m = 1,\ U = [-1,1],$$
$$\dot{x} = u, \qquad \dot{y} = x^2.$$

This system is (not obviously) accessible from $(0,0)$, but is (obviously) not locally controllable from that same point. Note that although $\mathcal{R}_\Sigma((0,0), \leq T)$ has nonempty interior, the initial point $(0,0)$ is not in that interior. Thus this is a system that is not controllable in any sense. Note that the system is also strongly accessible.

Feedback Linearization

Recall that the core of control theory is the idea of the *feedback*. In case of nonlinear control, this implies *feedback linearization*.

Exact Feedback Linearization

The idea of feedback linearization is to algebraically transform the nonlinear system dynamics into a fully or partly linear one so that the linear control techniques can be applied. Note that this is not the same as a conventional linearization using Jacobians. In this subsection we will present the modern, geometrical, Lie–derivative based techniques for exact feedback linearization of nonlinear control systems.

The Lie Derivative and Lie Bracket in Control Theory. Recall (see (1.2.6) above) that given a scalar function $h(x)$ and a vector–field $f(x)$, we define a new scalar function, $\mathcal{L}_f h = \nabla h f$, which is the Lie derivative of h w.r.t. f, i.e., the directional derivative of h along the direction of the vector f. Repeated Lie derivatives can be defined recursively:

$$\mathcal{L}_f^0 h = h, \qquad \mathcal{L}_f^i h = \mathcal{L}_f \left(\mathcal{L}_f^{i-1} h \right) = \nabla \left(\mathcal{L}_f^{i-1} h \right) f, \qquad (\text{for } \; i = 1, 2, ...)$$

Or given another vector–field, g, then $\mathcal{L}_g \mathcal{L}_f h(x)$ is defined as

$$\mathcal{L}_g \mathcal{L}_f h = \nabla \left(\mathcal{L}_f h \right) g.$$

For example, if we have a control system

$$\dot{x} = f(x), \qquad y = h(x),$$

with the state $x = x(t)$ and the output y, then the derivatives of the output are:

$$\dot{y} = \frac{\partial h}{\partial x} \dot{x} = \mathcal{L}_f h, \qquad \text{and} \qquad \ddot{y} = \frac{\partial L_f h}{\partial x} \dot{x} = \mathcal{L}_f^2 h.$$

Also, recall that the curvature of two vector–fields, g_1, g_2, gives a non–zero Lie bracket, $[g_1, g_2]$ ((1.2.6) see Figure 2.10). Lie bracket motions can generate new directions in which the system can move.

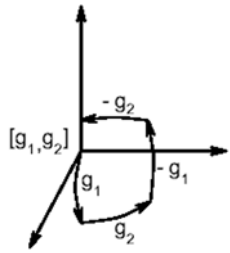


Fig. 2.10. The so–called 'Lie bracket motion' is possible by appropriately modulating the control inputs (see text for explanation).

In general, the Lie bracket of two vector–fields, $f(x)$ and $g(x)$, is defined by

$$[f, g] = Ad_f g = \nabla g f - \nabla f g = \frac{\partial g}{\partial x} f - \frac{\partial f}{\partial x} g,$$

where $\nabla f = \partial f / \partial x$ is the Jacobian matrix. We can define Lie brackets recursively,

$$Ad_f^0 g = g, \qquad Ad_f^i g = [f, Ad_f^{i-1} g], \qquad (\text{for } \; i = 1, 2, ...)$$

Lie brackets have the properties of bilinearity, skew–commutativity and Jacobi identity.

For example, if

$$f = \begin{pmatrix} \cos x_2 \\ x_1 \end{pmatrix}, \qquad g = \begin{pmatrix} x_1 \\ 1 \end{pmatrix},$$

then we have

$$[f,g] = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \cos x_2 \\ x_1 \end{pmatrix} - \begin{pmatrix} 0 & -\sin x_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ 1 \end{pmatrix} = \begin{pmatrix} \cos x_2 + \sin x_2 \\ -x_1 \end{pmatrix}.$$

Input/Output Linearization. Given a single–input single–output (SISO) system

$$\dot{x} = f(x) + g(x)\, u, \qquad y = h(x), \tag{2.601}$$

we want to formulate a linear–ODE relation between output y and a new input v. We will investigate (see [Isi89, SI89, Wil00]):

- How to generate a linear input/output relation.
- What are the internal dynamics and zero–dynamics associated with the input/output linearization?
- How to design stable controllers based on the I/O linearization.

This linearization method will be exact in a finite domain, rather than tangent as in the local linearization methods, which use Taylor series approximation. Nonlinear controller design using the technique is called exact feedback linearization.

Algorithm for Exact Feedback Linearization. We want to find a nonlinear compensator such that the closed–loop system is linear (see Figure 2.11). We will consider only affine SISO systems of the type (2.601), i.e, $\dot{x} = f(x) + g(x)\, u$, $y = h(x)$, and we will try to construct a *control law* of the form

$$u = p(x) + q(x)\, v, \tag{2.602}$$

where v is the setpoint, such that the *closed–loop nonlinear system*

$$\dot{x} = f(x) + g(x)\, p(x) + g(x)\, q(x)\, v, \qquad y = h(x),$$

is linear from command v to y.

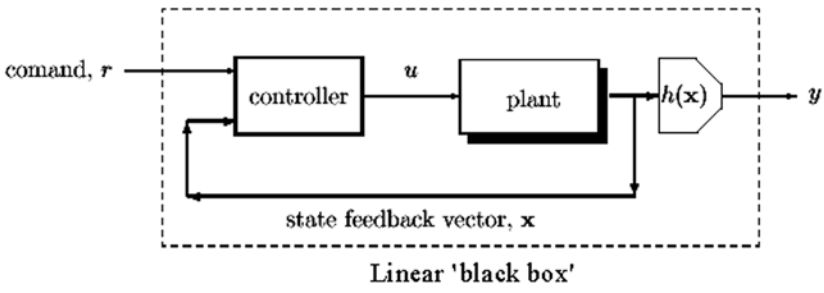


Fig. 2.11. Feedback linearization (see text for explanation).

The main idea behind the feedback linearization construction is to find a nonlinear change of coordinates which transforms the original system into one which is linear and controllable, in particular, a chain of integrators. The difficulty is finding the output function $h(x)$ which makes this construction possible.

We want to design an exact nonlinear feedback controller. Given the nonlinear affine system, $\dot{x} = f(x) + g(x)$, $y = h(x)$,.we want to find the controller functions $p(x)$ and $q(x)$. The unknown functions inside our controller (2.602) are given by:

$$p(x) = \frac{-\left(\mathcal{L}_f^r h(x) + \beta_1 \mathcal{L}_f^{r-1} h(x) + ... + \beta_{r-1} \mathcal{L}_f h(x) + \beta_r h(x)\right)}{\mathcal{L}_g \mathcal{L}_f^{r-1} h(x)},$$

$$q(x) = \frac{1}{\mathcal{L}_g \mathcal{L}_f^{r-1} h(x)}, \tag{2.603}$$

which are comprised of Lie derivatives, $\mathcal{L}_f h(x)$. Here, the *relative order*, r, is the smallest integer r such that $\mathcal{L}_g \mathcal{L}_f^{r-1} h(x) \neq 0$. For linear systems r is the difference between the number of poles and zeros.

To get the *desired response*, we choose the r parameters in the β polynomial to describe how the output will respond to the setpoint, v (pole–placement).

$$\frac{d^r y}{dt^r} + \beta_1 \frac{d^{r-1} y}{dt^{r-1}} + ... + \beta_{r-1} \frac{dy}{dt} + \beta_r y = v.$$

Here is the proposed algorithm [Isi89, SI89, Wil00]):

1. Given nonlinear SISO process, $\dot{x} = f(x, u)$, and output equation $y = h(x)$, then:
2. Calculate the relative order, r.
3. Choose an rth order desired linear response using pole–placement technique (i.e., select β). For this could be used a simple rth order low–pass filter such as a Butterworth filter.
4. Construct the exact linearized nonlinear controller (2.603), using Lie derivatives and perhaps a symbolic manipulator (Mathematica or Maple).
5. Close the loop and get a linear input–output black–box (see Figure 2.11).
6. Verify that the result is actually linear by comparing with the desired response.

Relative Degree

A nonlinear SISO system

$$\dot{x} = f(x) + g(x)\, u, \qquad y = h(x),$$

is said to have *relative degree* r at a point x_o if (see [Isi89, NS90])

1. $L_g L_f^k h(x) = 0$ for all x in a neighborhood of x_o and all $k < r - 1$; and

2. $L_g L_f^{r-1} h(x_o) \neq 0$.

For example, *controlled Van der Pol oscillator* has the state–space form

$$\dot{x} = f(x) + g(x)\, u = \begin{bmatrix} x_2 \\ 2\omega\zeta\,(1 - \mu x_1^2)\, x_2 - \omega^2 x_1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u.$$

Suppose the output function is chosen as $y = h(x) = x_1$. In this case we have

$$L_g h(x) = \frac{\partial h}{\partial x} g(x) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0, \qquad \text{and}$$
$$L_f h(x) = \frac{\partial h}{\partial x} f(x) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_2 \\ 2\omega\zeta\,(1 - \mu x_1^2)\, x_2 - \omega^2 x_1 \end{bmatrix} = x_2.$$

Moreover

$$L_g L_f h(x) = \frac{\partial (L_f h)}{\partial x} g(x) = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 1,$$

and thus we see that the Vand der Pol oscillator system has relative degree 2 at any point x_o.

However, if the output function is, for instance $y = h(x) = \sin x_2$, then $L_g h(x) = \cos x_2$. The system has relative degree 1 at any point x_o, provided that $(x_o)_2 \neq (2k+1)\,\pi/2$. If the point x_o is such that this condition is violated, no relative degree can be defined.

As another example, consider a *linear system* in the state–space form

$$\dot{x} = A\, x + B\, u, \qquad y = C\, x.$$

In this case, since $f(x) = A\, x$, $g(x) = B$, $h(x) = C\, x$, it can be seen that

$$L_f^k h(x) = C\, A^k\, x, \qquad \text{and therefore,}$$
$$L_g L_f^k h(x) = C\, A^k\, B.$$

Thus, the integer r is characterized by the conditions

$$C\, A^k\, B = 0, \qquad \text{for all } k < r - 1$$
$$C\, A^{r-1}\, B \neq 0, \qquad \text{otherwise.}$$

It is well–known that the integer satisfying these conditions is exactly equal to the *difference* between the degree of the denominator polynomial and the degree of the numerator polynomial of the transfer function

$$H(s) = C\,(sI - A)^{-1} B$$

of the system.

Approximative Feedback Linearization

Consider a SISO system

$$\dot{x} = f(x) + g(x)\, u, \tag{2.604}$$

where f and g are smooth vector–fields defined on a compact contractible region M of $\mathbb{R}^n$ containing the origin. (Typically, M is a closed ball in $\mathbb{R}^n$.) We assume that $f(0) = 0$, i.e., that the origin is an equilibrium for $\dot{x} = f(x)$. The classical problem of feedback linearization can be stated as follows: find in a neighborhood of the origin a smooth change of coordinates $z = \Phi(x)$ (a local diffeomorphism) and a smooth feedback law $u = k(x) + l(x)\, u_{new}$ such that the closed–loop system in the new coordinates with new control is linear,

$$\dot{z} = Az + B\, u_{new},$$

and controllable (see [BH96]). We usually require that $\Phi(0) = 0$. We assume that the system (2.604) has the *linear controllability* property

$$\dim(\text{span}\{g, Ad_f g, ..., Ad_f^{n-1} g\}) = n, \qquad \text{for all } \; x \in M \tag{2.605}$$

(where Ad_f^i are iterated Lie brackets of f and g). We define the *characteristic distribution* for (2.604)

$$\mathcal{D} = \text{span}\{g, Ad_f g, ..., Ad_f^{n-2} g\},$$

which is an $(n-1)$D smooth distribution by assumption of linear controllability (2.605). We call any nowhere vanishing $1-$form ω annihilating $\mathcal{D}$ a characteristic $1-$form for (2.604). All the characteristic $1-$forms for (2.604) can be represented as multiples of some fixed characteristic $1-$form ω_0 by a smooth nowhere vanishing function (zero–form) β. Suppose that there is a nonvanishing β so that $\beta\omega_0$ is exact, i.e., $\beta\omega_0 = d\alpha$ for some smooth function α, where d denotes the exterior derivative. Then ω_0 is called *integrable* and is called an integrating factor for ω_0. The following result is standard in nonlinear control: Suppose that the system (2.604) has the linear controllability property (2.605) on M. Let $\mathcal{D}$ be the characteristic distribution and ω_0 be a characteristic $1-$form for (2.604). The following statements are equivalent:

1. Equation (2.604) is feedback linearizable in a neighborhood of the origin in M;
2. D is involutive in a neighborhood of the origin in M; and
3. ω_0 is integrable in a neighborhood of the origin in M.

As is well known, a generic nonlinear system is not feedback linearizable for $n > 2$. However, in some cases, it may make sense to consider *approximate feedback linearization*.

Namely, if one can find a feedback linearizable system close to (2.604), there is hope that a control designed for the feedback linearizable system and

applied to (2.604) will give satisfactory performance if the feedback linearizable system is close enough to (2.604). The first attempt in this direction goes back to [Kre84], where it was proposed to apply to (2.604) a change of variables and feedback that yield a system of the form

$$\dot{z} = Az + B\,u_{new} + O(z, u_{new}),$$

where the term $O(z, u_{new})$ contains higher–order terms. The aim was to make $O(z, u_{new})$ of as high order as possible. Then we can say that the system (2.604) is approximately feedback linearized in a small neighborhood of the origin. Later [HT93] introduced a new algorithm to achieve the same goal with fewer steps.

Another idea has been investigated in [HSK92]. Roughly speaking, the idea was to neglect nonlinearities in (2.604) responsible for the failure of the involutivity condition in above theorem. This approach happened to be successful in the ball–and–beam system, when neglect of centrifugal force acting on ball yielded a feedback linearizable system. Application of a control scheme designed for the system with centrifugal force neglected to the original system gave much better results than applying a control scheme based on classical Jacobian linearization. This approach has been further investigated in [XH94, XH95] for the purpose of approximate feedback linearization about the manifold of constant operating points. However, a general approach to deciding which nonlinearities should be neglected to get the best approximation has not been set forth.

All of the above–mentioned work dealt with applying a change of coordinates and a preliminary feedback so that the resulting system looks like linearizable part plus nonlinear terms of highest possible order around an equilibrium point or an equilibrium manifold. However, in many applications one requires a large region of operation for the nonlinearizable system. In such a case, demanding the nonlinear terms to be neglected to be of highest possible order may, in fact, be quite undesirable. One might prefer that the nonlinear terms to be neglected be small in a uniform sense over the region of operation. In tis section we propose an approach to approximate feedback linearization that uses a change of coordinates and a preliminary feedback to put a system (2.604) in a perturbed Brunovsky form,

$$\dot{z} = Az + B\,u_{new} + P(z) + Q(z)\,u_{new}), \tag{2.606}$$

where $P(z)$ and $Q(z)$ vanish at $z = 0$ and are 'small' on M. We get upper bounds on uniform norms of P and Q (depending on some measures of noninvolutivity of $\mathcal{D}$) on any compact, contractible M.

A different, indirect approach was presented in [BH96]. In this section, the authors present an approach for finding feedback linearizable systems that approximate a given SISO nonlinear system on a given compact region of the state–space. First, they it is shown that if the system is close to being involutive, then it is also close to being linearizable. Rather than working directly with the characteristic distribution of the system, the authors work with

characteristic 1−forms, i.e., with the 1−forms annihilating the characteristic distribution. It is shown that homotopy operators can be used to decompose a given characteristic 1−form into an exact and an antiexact part. The exact part is used to define a change of coordinates to a normal form that looks like a linearizable part plus nonlinear perturbation terms. The nonlinear terms in this normal form depend continuously on the antiexact part, and they vanish whenever the antiexact part does. Thus, the antiexact part of a given characteristic 1−form is a measure of nonlinearizability of the system. If the nonlinear terms are small, by neglecting them we get a linearizable system approximating the original system. One can design control for the original system by designing it for the approximating linearizable system and applying it to the original one. We apply this approach for design of locally stabilizing feedback laws for nonlinear systems that are close to being linearizable.

Let us start with approximating characteristic 1−forms by exact forms using *homotopy operators* (compare with equation (1.27) above). Namely, on any contractible region M one can define a linear operator H that satisfies

$$\omega = d(H\omega) + Hd\omega \tag{2.607}$$

for any form ω. The homotopy identity (2.607) allows to decompose any given 1−form into the *exact part* $d(H\omega)$ and an 'error part' $\epsilon = Hd\omega$, which we call the *antiexact part of* ω. For given ω_0 annihilating $\mathcal{D}$ and a scaling factor β we define $\alpha_\beta = H\beta w_0$ and $\epsilon_\beta = Hd\beta w_0$. The 1−form ϵ_β measures how exact $\omega_\beta = \beta w_0$ is. If it is zero, then ω_β is exact and the system (2.604) is linearizable, and the zero–form α_β and its first $n-1$ Lie derivatives along f are the new coordinates. In the case that ω_0 is not exactly integrable, i.e., when no exact integrating factor β exists, we choose β so that $d\beta w_0$ is *smallest* in some sense (because this also makes ϵ_β small). We call this β an *approximate integrating factor for* ω_0. We use the zero–form α_β and its first $n-1$ Lie derivatives along f as the new coordinates as in the linearizable case. In those new coordinates the system (2.604) is in the form

$$\dot{z} = Az + Bru + Bp + Eu,$$

where r and p are smooth functions, $r \neq 0$ around the origin, and the term E (the obstruction to linearizablity) depends linearly on ϵ_β and some of its derivatives. We choose $u = r^{-1}(u_{new} - p)$, where u_{new} is a new control variable. After this change of coordinates and control variable the system is of the form (2.606) with $Q = r^{-1}E$, $P = -r^{-1}pE$. We get estimates on the uniform norm of Q and P (via estimates on r, p, and E) in terms of the error 1−form ϵ_β, for any fixed β, on any compact, contractible manifold M. Most important is that Q and P depend in a continuous way on ϵ_β and some of its derivatives, and they vanish whenever ϵ does (see [BH96]).

Controllability

Linear Controllability

Recall that a system is said to be *controllable* if the set of all states it can reach from initial state $x_0 = x(0)$ at the fixed time $t = T$ contains a ball $\mathcal{B}$ around x_0. Again, a system is called *small time locally controllable* (STLC) iff the ball $\mathcal{B}$ for $t \leq T$ contains a neighborhood of x_0.[27]

In the case of a linear system in the standard state–space form (see subsection (2.4.3) above)

$$\dot{x} = Ax + Bu, \tag{2.608}$$

where A is the $n \times n$ *state matrix* and B is the $m \times n$ *input matrix*, all controllability definitions coincide, i.e.,

$$0 \to x(T), \qquad x(0) \to 0, \qquad x(0) \to x(T),$$

where T is either fixed or free.

Rank condition states: System (2.608) is controllable iff the matrix

$$W_n = \left(B\, AB\, ...\, A^{n-1}B\right) \qquad \text{has full rank.}$$

In the case of nonlinear systems the corresponding result is get using the formalism of Lie brackets, as Lie algebra is to nonlinear systems as matrix algebra is to linear systems.

Nonlinear Controllability

Nonlinear MIMO–systems are generally described by differential equations of the form (see [Isi89, NS90, Goo98]):

$$\dot{x} = f(x) + g_i(x)\, u^i, \qquad (i = 1, ..., n), \tag{2.609}$$

defined on a smooth $n-$manifold M, where $x \in M$ represents the state of the control system, $f(x)$ and $g_i(x)$ are vector–fields on M and the u^i are control inputs, which belong to a set of *admissible controls*, $u^i \in U$. The system (2.609) is called *driftless*, or *kinematic*, or *control linear* if $f(x)$ is identically zero; otherwise, it is called a *system with drift*, and the vector–field $f(x)$ is called the *drift term*. The flow $\phi_t^g(x_0)$ represents the solution of the differential equation $\dot{x} = g(x)$ at time t starting from x_0. Geometrical way to understand the *controllability* of the system (2.609) is to understand the geometry of the vector–fields $f(x)$ and $g_i(x)$.

Example: Car–Parking Using Lie Brackets In this popular example, the driver has two different transformations at his disposal. He/she can turn

[27] The above definition of controllability tells us only whether or not something can reach an open neighborhood of its starting point, but does not tell us how to do it. That is the point of the *trajectory generation*.

the steering wheel, or he/she can drive the car forward or back. Here, we specify the state of a car by four coordinates: the (x, y) coordinates of the center of the rear axle, the direction θ of the car, and the angle ϕ between the front wheels and the direction of the car. L is the constant length of the car. Therefore, the configuration manifold of the car is 4D, $M = (x, y, \theta, \phi)$.

Using (2.609), the driftless car kinematics can be defined as:

$$\dot{x} = g_1(x)\, u_1 + g_2(x)\, u_2, \tag{2.610}$$

with two vector–fields $g_1, g_2 \in \mathcal{X}^k(M)$.

The infinitesimal transformations will be the vector–fields

$$g_1(x) \equiv \text{DRIVE} = \cos\theta \frac{\partial}{\partial x} + \sin\theta \frac{\partial}{\partial y} + \frac{\tan\phi}{L} \frac{\partial}{\partial \theta} \equiv \begin{pmatrix} \cos\theta \\ \sin\theta \\ \frac{1}{L}\tan\phi \\ 0 \end{pmatrix},$$

and $$g_2(x) \equiv \text{STEER} = \frac{\partial}{\partial \phi} \equiv \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Now, STEER and DRIVE do not commute; otherwise we could do all your steering at home before driving of on a trip. Therefore, we have a Lie bracket

$$[g_2, g_1] \equiv [\text{STEER}, \text{DRIVE}] = \frac{1}{L\cos^2\phi} \frac{\partial}{\partial \theta} \equiv \text{ROTATE}.$$

The operation $[g_2, g_1] \equiv \text{ROTATE} \equiv [\text{STEER,DRIVE}]$ is the infinitesimal version of the sequence of transformations: steer, drive, steer back, and drive back, i.e.,

$$\{\text{STEER}, \text{DRIVE}, \text{STEER}^{-1}, \text{DRIVE}^{-1}\}.$$

Now, ROTATE can get us out of some parking spaces, but not tight ones: we may not have enough room to ROTATE out. The usual tight parking space restricts the DRIVE transformation, but not STEER. A truly tight parking space restricts STEER as well by putting your front wheels against the curb.

Fortunately, there is still another commutator available:

$$[g_1, [g_2, g_1]] \equiv [\text{DRIVE}, [\text{STEER}, \text{DRIVE}]] = [[g_1, g_2], g_1] \equiv$$
$$[\text{DRIVE}, \text{ROTATE}] = \frac{1}{L\cos^2\phi} \left(\sin\theta \frac{\partial}{\partial x} - \cos\theta \frac{\partial}{\partial y} \right) \equiv \text{SLIDE}.$$

The operation $[[g_1, g_2], g_1] \equiv \text{SLIDE} \equiv [\text{DRIVE,ROTATE}]$ is a displacement at right angles to the car, and can get us out of any parking place. We just need to remember to steer, drive, steer back, drive some more, steer, drive back, steer back, and drive back:

$$\{\text{STEER}, \text{DRIVE}, \text{STEER}^{-1}, \text{DRIVE}, \text{STEER}, \text{DRIVE}^{-1}, \text{STEER}^{-1}, \text{DRIVE}^{-1}\}.$$

We have to reverse steer in the middle of the parking place. This is not intuitive, and no doubt is part of the problem with parallel parking.

Thus from only two controls u_1 and u_2 we can form the vector–fields DRIVE $\equiv g_1$, STEER $\equiv g_2$, ROTATE $\equiv [g_2, g_1]$, and SLIDE $\equiv [[g_1, g_2], g_1]$, allowing us to move anywhere in the configuration manifold M. The car kinematics $\dot{x} = g_1 u_1 + g_2 u_2$ is thus expanded as:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \\ \dot{\phi} \end{pmatrix} = \text{DRIVE} \cdot u_1 \ + \text{STEER} \cdot u_2 \ \equiv \begin{pmatrix} \cos\theta \\ \sin\theta \\ \frac{1}{L}\tan\phi \\ 0 \end{pmatrix} \cdot u_1 \ + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \cdot u_2 \ .$$

The *parking theorem* says: One can get out of any parking lot that is larger than the car.

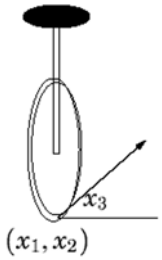


Fig. 2.12. Classical unicycle problem (see text for explanation).

The Unicycle Example. Now, consider the unicycle example (see Figure 2.12). Here we have

$$g_1 = \begin{pmatrix} \cos x_3 \\ \sin x_3 \\ 0 \end{pmatrix}, \qquad g_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \qquad [g_1, g_2] = \begin{pmatrix} \sin x_3 \\ -\cos x_3 \\ 0 \end{pmatrix}.$$

The unicycle system is full rank and therefore controllable.

Controllability Condition

Nonlinear controllability is an extension of linear controllability. The nonlinear MIMO system

$$\dot{x} = f(x) + g(x)\, u \qquad \text{is controllable}$$

if the set of vector–fields $\{g, [f, g], ..., [f^{n-1}, g]\}$ is independent.

For example, for the kinematic car system of the form (2.610), the *nonlinear controllability criterion* reads: If the Lie bracket tree:

$g_1,\ g_2,\ [g_1, g_2],\ [[g_1, g_2], g_1],\ [[g_1, g_2], g_2],\ [[[g_1, g_2], g_1], g_1],\ \ [[[g_1, g_2], g_1], g_2],$ $[[[g_1, g_2], g_2], g_1],\ \ [[[g_1, g_2], g_2], g_2], ...$

– has *full rank* then the system is *controllable* [Isi89, NS90, Goo98]. In this case the combined input

$$(u_1, u_2) = \begin{cases} (1,0), & t \in [0, \varepsilon] \\ (0,1), & t \in [\varepsilon, 2\varepsilon] \\ (-1,0), & t \in [2\varepsilon, 3\varepsilon] \\ (0,-1), & t \in [3\varepsilon, 4\varepsilon] \end{cases}$$

gives the motion $x(4\varepsilon) = x(0) + \varepsilon^2 [g_1, g_2] + O(\varepsilon^3)$, with the flow given by (see (1.32) below)

$$F_t^{[g_1,g_2]} = \lim_{n\to\infty} \left(F^{-g_2}_{\sqrt{t/n}} F^{-g_1}_{\sqrt{t/n}} F^{g_2}_{\sqrt{t/n}} F^{g_1}_{\sqrt{t/n}} \right)^n .$$

Distributions

In control theory, the set of all possible directions in which the system can move, or the set of all points the system can reach, is of obvious fundamental importance. Geometrically, this is related to *distributions*.

Recall from subsection 1.3.3 above that a *distribution* $\Delta \subset \mathcal{X}^k(M)$ on a smooth nD manifold M is a subbundle of its tangent bundle TM, which assigns a subspace of the tangent space T_xM to each point $x \in M$ in a smooth way. The dimension of $\Delta(x)$ over $\mathbb{R}$ at a point $x \in M$ is called the rank of Δ at x.

A distribution Δ is involutive if, for any two vector–fields $X, Y \in \Delta$, their Lie bracket $[X, Y] \in \Delta$.

A function $f \in C^\infty(M)$ is called an *integral* of Δ if $df(x) \in \Delta^0(x)$ for each $x \in M$. An *integral manifold* of Δ is a submanifold N of M such that $T_xN \subset \Delta(x)$ for each $x \in N$. A distribution Δ is *integrable* if, for any $x \in M$, there is a submanifold $N \subset M$, whose dimension is the same as the rank of Δ at x,.containing x such that the tangent bundle, TN, is exactly Δ restricted to N, i.e., $TN = \Delta|_N$. Such a submanifold is called the *maximal integral manifold* through x.

It is natural to consider distributions generated by the vector–fields appearing in the sequence of flows (1.31). In this case, consider the distribution defined by

$$\Delta = \text{span}\{f; g_1 ... g_m\},$$

where the span is taken over the set of smooth real–valued functions. Denote by $\bar{\Delta}$ the *involutive closure* of the distribution Δ, which is the closure of Δ under bracketing. Then, $\bar{\Delta}$ is the smallest subalgebra of $\mathcal{X}^k(M)$ which contains $\{f; g_1 ... g_m\}$. We will often need to 'add' distributions. Since distributions are, pointwise, vector spaces, define the sum of two distributions,

$$(\Delta_1 + \Delta_2)(x) = \Delta_1(x) + \Delta_2(x).$$

Similarly, define the intersection

$$(\Delta_1 \cap \Delta_2)(x) = \Delta_1(x) \cap \Delta_2(x).$$

More generally, we can arrive at a distribution via a *family of vector–fields*, which is a subset $\mathcal{V} \subset \mathcal{X}^k(M)$. Given a family of vector–fields $\mathcal{V}$, we may define a distribution on M by

$$\Delta_{\mathcal{V}}(x) = \langle X(x) | X \in \mathcal{V} \rangle_{\mathbb{R}}.$$

Since $\mathcal{X}^k(M)$ is a Lie algebra, we may ask for the smallest Lie subalgebra of $\mathcal{X}^k(M)$ which contains a family of vector–fields $\mathcal{V}$. It will be denoted as $\overline{Lie}(\mathcal{V})$, and will be represented by the set of vector–fields on M generated by repeated Lie brackets of elements in $\mathcal{V}$. Let $\mathcal{V}^{(0)} = \mathcal{V}$ and then iteratively define a sequence of families of vector–fields by

$$\mathcal{V}^{(i+1)} = \mathcal{V}^{(i)} \cup \{[X, Y] | X \in \mathcal{V}^{(0)} = \mathcal{V} \text{ and } Y \in \mathcal{V}^{(i)}\}.$$

Now, every element of $\overline{Lie}(\mathcal{V})$ is a linear combination of repeated Lie brackets of the form

$$[Z_k, [Z_{k-1}, [\cdot\cdot, [Z_2, Z_1] \cdot\cdot]]]$$

where $Z_i \in \mathcal{V}$ for $i = 1, ..., k$.

Foliations

Recall that related to integrable distributions are *foliations*.

The *Frobenius theorem* asserts that integrability and involutivity are equivalent, at least locally. Thus, associated with an involutive distribution is a partition Φ of M into disjoint connected immersed submanifolds called *leaves*. This partition Φ is called a *foliation*. More precisely, a foliation $\mathcal{F}$ of a smooth manifold M is a collection of disjoint immersed submanifolds of M whose disjoint union equals M. Each connected submanifold of $\mathcal{F}$ is called a *leaf* of the foliation. Given an integrable distribution Δ, the collection of maximal integral manifolds for Δ defines a foliation on M, denoted by $\mathcal{F}_D$.

A foliation $\mathcal{F}$ of M defines an equivalence relation on M whereby two points in M are equivalent if they lie in the same leaf of $\mathcal{F}$. The set of equivalence classes is denoted $M/\mathcal{F}$ and is called the *leaf space* of $\mathcal{F}$. A foliation $\mathcal{F}$ is said to be simple if $M/\mathcal{F}$ inherits a manifold structure so that the projection from M to $M/\mathcal{F}$ is a surjective submersion.

In control theory, foliation leaves are related to the set of points that a control system can reach starting from a given initial condition. A foliation Φ of M defines an equivalence relation on M whereby two points in M are equivalent if they lie in the same leaf of Φ. The set of equivalence classes is denoted M/Φ and is called the *leaf space* of Φ.

Philip Hall Basis

Given a set of vector–fields $\{g_1 ... g_m\}$, define the *length* of a *Lie product* as

$$l(g_i) = 1, \qquad l([A, B]) = l(A) + l(B), \qquad (\text{for } \; i = 1, ..., m),$$

where A and B may be Lie products. A *Philip Hall basis* is an ordered set of Lie products $H = \{B_i\}$ satisfying:

1. $g_i \in H, \qquad (i = 1, ..., m)$;
2. If $l(B_i) < l(B_j)$, then $B_i < B_j$; and
3. $[B_i, B_j] \in H$ iff
   (a) $B_i, B_j \in H$ and $B_i < B_j$, and
   (b) either $B_j = g_k$ for some k or $B_j = [B_l, B_r]$ with $B_l, B_r \in H$ and $B_l \leq B_i$.

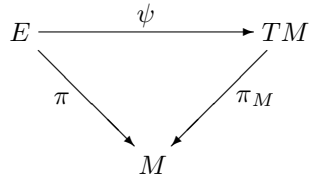
Essentially, the ordering aspect of the Philip Hall basis vectors accounts for skew symmetry and Jacobi identity to determine a basis.

2.3.2 Geometrical Control of Mechanical Systems

Much of the existing work on control of mechanical systems has relied on the presence of specific structure. The most common examples of the types of structure assumed are symmetry (conservation laws) and constraints. While it may seem counter–intuitive that constraints may help in control theory, this is sometimes in fact the case. The reason is that the constraints give extra forces (forces of constraint) which can be used to advantage. probably, the most interesting work is done from the Lagrangian (respectively Hamiltonian) perspective where we study systems whose Lagrangians are 'kinetic energy minus potential energy' (resp. 'kinetic energy plus potential energy'). For these *simple mechanical control systems*, the controllability questions are different than those typically asked in nonlinear control theory. In particular, one is often more interested in what happens to configurations rather than states, which are configurations and velocities (resp. momenta) for these systems (see [Lew95, LM97]).

Abstract Control System

In general, a nonlinear control system Σ can be represented as a triple (Σ, M, f), where M is the system's *state–space* manifold with the tangent bundle TM and the general fibre bundle E, and f is a smooth map, such that the following bundle diagram commutes [Man98]



where $\psi : (x, u) \mapsto (x, f(x, u))$, π_M is the natural projection of TM on M, the projection $\pi : E \to M$ is a smooth fibre bundle, and the fibers of E represent the *input spaces*. If one chooses fibre–respecting coordinates (x, u) for E, then locally this definition reduces to $\psi : (x, u) \mapsto (x, \psi(x, u))$, i.e.,

$$\dot{x} = \psi(x, u).$$

The specific form of the map ψ, usually used in nonlinear control, is $\psi : (x, u) \mapsto (x, f(x) + g(x, u))$, with $g(x, 0) = 0$, producing standard nonlinear system equation

$$\dot{x} = f(x) + g(x, u).$$

Controllability of a Linear Control System

Consider a linear biomechanical control system:

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{2.611}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $A \in L(\mathbb{R}^n, \mathbb{R}^n)$, and $B \in L(\mathbb{R}^m, \mathbb{R}^n)$. One should think of $t \mapsto u(t)$ as being a specified input signal, i.e., a function on the certain time interval, $[0, T]$. Now, control theory wants to design the signal to make the *state* $t \mapsto x(t)$ do what we want. What this is may vary, depending on the situation at hand. For example, one may want to steer from an initial state x^i to a final state x_f, perhaps in an optimal way. Or, one may wish to design $u : \mathbb{R}^n \to \mathbb{R}^m$ so that some state, perhaps $x = 0$, is stable for the dynamical system $\dot{x}(t) = Ax + Bu(x)$, which is called *state feedback* (often one asks that u be linear). One could also design u to be a function of both x and t, etc.

One of the basic control questions is *controllability*, which comes in many guises. Basically we are asking for 'reachable' points. In particular,

$$\mathcal{R}(0) = \text{span}_{\mathbb{R}}\{[B|AB|...|A^{n-1}B]\},$$

which is the smallest $A-$invariant subspace containing $\text{Im}(B)$, denotes the set of points reachable from $0 \in \mathbb{R}^n$. For the linear system (2.611), the basic controllability questions have definite answers. We want to do something similar for a class of simple mechanical systems [Lew95, LM97].

Affine Control System and Local Controllability

The nonlinear control system that we most often consider in humanoid robotics (see next section) has state–space M, a smooth $n-$manifold, and is *affine* in the controls. Thus it has the form (see [Lew95, LM97])

$$\dot{x} = f(x) + u^a g_a(x), \qquad (x \in M), \tag{2.612}$$

where $f, g_1, ..., g_m$ are vector–fields on M. The *drift* vector–field $f = f(x)$ describes how the system would evolve in the absence of any inputs. Each of the *control* vector–fields $g_1, ..., g_m$ specifies a direction in which one can supply actuation. To fully specify the control system properly, one should also specify the type of control action to be considered. Here we consider our controls to be taken from the set: $U = \{u : \mathbb{R} \to \mathbb{R}^m \,|\, u \text{ is piecewise constant}\}$. This class of controls is sufficient to deal with all analytic control systems. More generally, one may wish to consider measurable functions which take their values in a subset of $\mathbb{R}^m$.

Given an *affine control system* (2.612), it is possible to define a family of vector–fields on M by: $V_\Sigma = \{f + u^a g_a \,|\, u \in \mathbb{R}^m\}$.

A solution of the system (2.612) is a pair (γ, u), where $\gamma : [0, T] \to M$ is a piecewise smooth curve on M and $u \in U$ such that

$$\dot{\gamma}(t) = f(\gamma(t)) + u^a(t)\, g_a(\gamma(t)), \qquad \text{for each} \quad t \in [0, T].$$

The *reachable set* from x_0 in time T is

$$\begin{aligned} \mathcal{R}(x_0, T) = \{x | \exists \gamma : [0, T] \to M \quad & \text{and} \\ u \;:\; [0, T] \to \mathbb{R}^m \quad & \text{satisfying} \quad (2.612) \\ \text{with} \;\; \gamma(0) = x_0 \quad & \text{and} \quad \gamma(T) = x\}. \end{aligned}$$

Note that since the system has drift f, when we reach the point $\gamma(T)$ we will not remain there if this is not an equilibrium point for f. Also, we have, $\mathcal{R}(x_0, \leq T) = \cup_{0<t\leq T} \mathcal{R}(x_0, T)$.

Let $x_0 \in M$, let V be a neighborhood of x_0, and let $T > 0$. We say that equation (2.612) represents a *locally accessible system* at x_0 if $\mathcal{R}(x_0, \leq T)$ contains an open subset of M for each V and for each T sufficiently small. Furthermore, we say that the system (2.612) is *small–time local controllability* (STLC, see [Sus83, Sus87]), if it is locally accessible and if x_0 is in the interior of $\mathcal{R}(x_0, \leq T)$ for each V and for each T sufficiently small.

Lagrangian Control Systems

Simple Mechanical Control Systems

As a motivation/prototype of a simple mechanical control system, consider a simple *robotic leg* (see Figure 2.13), in which inputs are: (1) an internal torque F^1 moving the leg relative to the body and (2) a force F^2 extending the leg. This system is 'controllable' in the sense that, starting from rest, one can reach any configuration from a given initial configuration. However, as a traditional control system, it is not controllable because of conservation of angular momentum. If one asks for the *states* (i.e., configurations and velocities) reachable from configurations with zero initial velocity, one finds that not all states are reachable. This is a consequence of the fact that angular momentum

is conserved, even with inputs. Thus if one starts with zero momentum, the momentum will remain zero (this is what enables one to treat the system as *nonholonomic*). Nevertheless, all configurations are accessible. This suggests that the question of controllability is different depending on whether one is interested in configurations or states. We will be mainly interested in reachable configurations. Considering the system with just one of the two possible input forces is also interesting. In the case where we are just allowed to use F^2, the possible motions are quite simple; one can only move the ball on the leg back and forth. With just the force F^1 available, things are a bit more complicated. But, for example, one can still say that no matter how you apply the force, the ball with never move 'inwards' [Lew95, LM97].

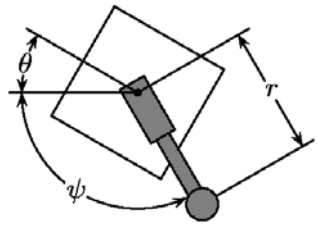


Fig. 2.13. A simple robotic leg (see text for explanation).

In general, simple mechanical control systems are characterized by:

- An nD configuration manifold M;
- A Riemannian metric g on M;
- A potential energy function V on M; and
- m linearly independent 1−forms, $F^1, ..., F^m$ on M (input forces; e.g., in the case of the simple robotic leg, $F^1 = d\theta - d\psi$ and $F^2 = dr$).

When we say these systems are not amenable to liberalization–based methods, we mean that their liberalizations at zero velocity are not controllable, and that they are not feedback linearizable. This makes simple mechanical control systems a non–trivial class of nonlinear control systems, especially from the point of view of control design.

As a basic example to start with, consider a *planar rigid body* (see Figure 2.14), with coordinates (x, y, θ). Inputs are (1) force pointing towards center of mass, $F^1 = \cos\theta dx + \sin\theta dy$, (2) force orthogonal to line to center of mass, $F^2 = -\sin\theta dx + \cos\theta dy - hd\theta$, and (3) torque at center of mass $F^3 = d\theta$. The planar rigid body, although seemingly quite simple, can be actually interesting. Clearly, if one uses all three inputs, the system is *fully actuated*, and so boring for investigating reachable configurations. But if one takes various combinations of one or two inputs, one gets a pretty nice sampling of what can happen for these systems. For example, all possible combinations of two inputs allow one to reach all configurations. Using F^1 or F^3 alone give simple,

1D reachable sets, similar to using F^2 for the robotic leg (as we are always starting with zero initial velocity). However, if one is allowed to only use F^2, then it is not quite clear what to expect, at least just on the basis of intuition.

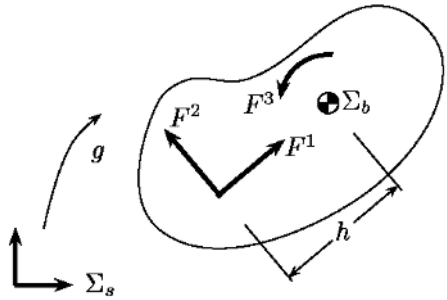


Fig. 2.14. Coordinate systems of a planar rigid body.

It turns out that our simplifying assumptions, i.e., zero initial velocity and restriction of our interest to configurations (i.e., as all problem data is on M, we expect answers to be describable using data on M), makes our task much simpler. In fact, the computations without these assumptions have been attempted, but have yet to yield coherent answers.

Now, we are interested in how do the input 1−forms $F^1, ..., F^m$ interact with the unforced mechanics of the system as described by the kinetic energy Riemannian metric. That is, what is the analogue of linear system's 'the smallest $A-$invariant subspace containing $\mathrm{Im}(B)$' – for simple mechanical control systems?

Motion and Controllability in Affine Connections

If we start with the local Riemannian metric form $g \longmapsto g_{ij}(q)\, dq^i dq^j$, then we have a kinetic energy Lagrangian $L(q, v) = g_{ij}(q)\, \dot{q}^i \dot{q}^j$, and consequently the Euler–Lagrangian equations read

$$\frac{d}{dt}\partial_{\dot{q}^i} L - \partial_{q^i} L \equiv g_{ij}\ddot{q}^j + \left(\partial_{q^k} g_{ij} - \frac{1}{2}\partial_{q^i} g_{jk}\right) \dot{q}^j \dot{q}^k = u_a F_i^a, \qquad (i = 1, ..., n).$$

Now multiply this by g^{li} and take the symmetric part of the coefficient of $\dot{q}^j \dot{q}^k$ to get $\ddot{q}^l + \Gamma^l_{jk}\, \dot{q}^j \dot{q}^k = u^a Y^l_a$, $(l = 1, ..., n,)$, where Γ^i_{jk} are the *Christoffel symbols* (2.5) for the *Levi–Civita connection* ∇ (see (1.2.9) above). So, the equations of motion an be rewritten

$$\nabla_{\dot{\gamma}(t)} \dot{\gamma}(t) = u^a(t)\, Y_a\left(\gamma(t)\right), \qquad (a = 1, ..., m),$$

where $Y_a = (F^a)^\sharp$, while $\sharp : T^*M \to TM$ is the 'sharp'–isomorphism associated with the Riemannian metric g.

Now, there is nothing to be gained by using a Levi–Civita connection, or by assuming that the vector–fields come from 1−forms. At this point, perhaps the generalization to an arbitrary affine connection seems like a senseless abstraction. However, as we shall see, this abstraction allows us to include another large class of mechanical control systems. So we will study the control system

$$\nabla_{\dot\gamma(t)}\dot\gamma(t) = u^a(t)\, Y_a\left(\gamma(t)\right) \left[+Y_0\left(\gamma(t)\right)\right], \tag{2.613}$$

with ∇ a general affine connection on M, and $Y_1 ..., Y_m$ linearly independent vector–fields on M. The 'optional' term $Y_0 = Y_0\left(\gamma(t)\right)$ in (2.613) indicates how potential energy may be added. In this case $Y_0 = -\operatorname{grad} V$ (however, one looses nothing by considering a general vector–field instead of a gradient) [Lew98].

A *solution* to (2.613) is a pair (γ, u) satisfying (2.613) where $\gamma : [0, T] \to M$ is a curve and $u : [0; T] \to \mathbb{R}^m$ is bounded and measurable.

Let U be a neighborhood of $q_0 \in M$ and denote by $\mathcal{R}^U_M(q_0, T)$ those points in M for which there exists a solution (γ, u) with the following properties:

1. $\gamma(t) \in U$ for $t \in [0, T]$;
2. $\dot\gamma(0) = 0_q$; and
3. $\gamma(T) \in T_qM$.

Also $\mathcal{R}^U_M(q_0, \leq T) = \cup_{0\leq t\leq T}\mathcal{R}^U_M(q_0, t)$. Now, regarding the local controllability, we are only interested in points which can be reached without taking 'large excursions'. Control problems which are local in this way have the advantage that they can be characterized by Lie brackets. So, we want to describe our *reachable set* $\mathcal{R}^U_M(q, \leq T)$ for the simple mechanical control system (2.613). The system (2.613) is *locally configuration accessible* (LCA) at q if there exists $T > 0$ so that $\mathcal{R}^U_M(q, \leq t)$ contains a non–empty open subset of M for each neighborhood U of q and each $t \in]0, T]$. Also, (2.613) is *locally configuration controllable* (LCC) at q if there exists $T > 0$ so that $\mathcal{R}^U_M(q, \leq t)$ contains a neighborhood of q for each neighborhood U of q and each $t \in]0, T]$. Although sound very similar, the notions of local configuration accessibility and local configuration controllability are genuinely different (see Figure 2.15). Indeed, one need only look at the example of the robotic leg with the F^1 input. In this example one may show that the system is LCA, but is not LCC [Lew98].

Local Configuration Accessibility

The accessibility problem is solved by looking at Lie brackets. For this we need to recall the definition of the *vertical lift* [Lew98]:

$$\operatorname{verlift}(Y(v_q)) = \left.\frac{d}{dt}\right|_{t=0} (v_q + tY(q)),$$

in local coordinates, if $Y = Y^i\partial_{q^i}$, then $\operatorname{verlift}(Y) = Y^i\partial_{v^i}$. Now we can rewrite (2.613) in the first–order form:

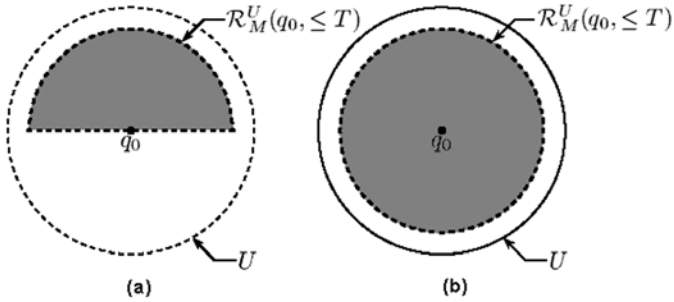


Fig. 2.15. Difference between the notions of local configuration accessibility (a), and local configuration controllability (b).

$$\dot{v} = Z(v) + u^a \,\mathrm{verlift}(Y_a(v)),$$

where Z is the geodesic spray for ∇.

We evaluate all brackets at 0_q (recall that $T_{0_q}TM \simeq T_qM \oplus T_qM$). Here, the first component we think of as being the 'horizontal' bit which is tangent to the zero section in TM, and we think of the second component as being the 'vertical' bit which is the tangent space to the fibre of $\tau_M : TM \to M$.

To get an answer to the local configuration accessibility problem, we employ standard nonlinear control techniques involving Lie brackets. Doing so gives us our first look at the *symmetric product*, $\langle X : Y \rangle = \nabla_X Y + \nabla_Y X$. Our sample brackets suggest that perhaps the only things which appear in the bracket computations are symmetric products and Lie brackets of the input vector–fields $Y_1, ..., Y_m$.

Here are some sample brackets:

(i) $[Z, \mathrm{verlift}(Y_a)](0_q) = (-Y_a(q), 0)$;
(ii) $[\mathrm{verlift}(Y_a), [Z, \mathrm{verlift}(Y_b)]](0_q) = (0, \langle Y_a : Y_b \rangle (q))$;
(iii) $[[Z, \mathrm{verlift}(Y_a)], [Z, \mathrm{verlift}(Y_b)]](0_q) = ([Y_a, Y_b](q), 0)$.

Now, let C_{ver} be the closure of $\mathrm{span}\{Y_1, ..., Y_m\}$ under symmetric product. Also, let C_{hor} be the closure of C_{ver} under Lie bracket. So, we assume C_{ver} and C_{hor} to be distributions (i.e., of constant rank) on M. The closure of $\mathrm{span}\{Z, \mathrm{verlift}(Y_1), ..., \mathrm{verlift}(Y_m)\}$ under Lie bracket, when evaluated at 0_q, is then the distribution

$$q \mapsto C_{hor}(q) \oplus C_{ver}(q) \subset T_qM \oplus T_qM.$$

Proving that the involutive closure of $\mathrm{span}\{Z, \mathrm{verlift}(Y_1), ..., \mathrm{verlift}(Y_m)\}$ is equal at 0_q to $C_{hor}(q) \oplus C_{ver}(q)$ is a matter of computing brackets, samples of which are given above, and seeing the patterns to suggest an inductive proof. The brackets for these systems are very structured. For example, the brackets of input vector–fields are identically zero. Many other brackets vanish identically, and many more vanish when evaluated at 0_q.

C_{hor} is integrable: let Λ_q be the maximal integral manifold through $q \in M$. Then, $\mathcal{R}^U_M(q, \leq T)$ is contained in Λ_q, and $\mathcal{R}^U_M(q, \leq T)$ contains a non–empty open subset of Λ_q. In particular, if $rank(C_{hor}) = n$ then (2.613) is LCA [Lew95, LM97]. This theorem gives a 'computable' description of the reachable sets (in the sense that we can calculate Λ_q by solving some over–determined nonlinear PDE's). But it does not give the kind of insight that we had with the 'smallest $A-$invariant subspace containing $\mathrm{Im}(B)$'.

Recall that a submanifold N of M is *totally geodesic* if every geodesic with initial velocity tangent to N remains on N. This can be weakened to distributions: a distribution D on M is *geodesically invariant* if for every geodesic $\gamma : [0, T] \to M$, $\dot{\gamma}(0) \in D_{\gamma(0)}$ implies $\dot{\gamma}(t) \in D_{\gamma(t)}$ for $t \in]0, T]$.

D is geodesically invariant i it is closed under symmetric product [Lew98]. This theorem says that the symmetric product plays for geodesically invariant distributions the same role the Lie bracket plays for integrable distributions. This result was key in providing the geometrical description of the reachable configurations.

An integrable distribution is *geodesically generated distribution* if it is the involutive closure of a geodesically invariant distribution. This basically means that one may reach all points on a leaf with geodesics lying in some subdistribution. The picture one should have in mind with the geometry of the reachable sets is a foliation of M by geodesically generated (immersed) submanifolds onto which the control system restricts if the initial velocity is zero. The idea is that when we start with zero velocity we remain on leaves of the foliation defined by C_{hor} [LM97, Lew00a]. Note that for cases when the affine connection possesses no geodesically invariant distributions, the system (2.613) is automatically LCA. This is true, for example, of S^2 with the affine connection associated with its round metric.

Clearly C_{ver} is the smallest geodesically invariant distribution containing $\mathrm{span}\{Y_1, ..., Y_m\}$. Also, C_{hor} is geodesically generated by span $\mathrm{span}\{Y_1, ..., Y_m\}$. Thus $\mathcal{R}^U_M$ is contained in, and contains a non–empty open subset of, the distribution geodesically generated by $\mathrm{span}\{Y_1, ..., Y_m\}$. Note that the pretty decomposition we have for systems with no potential energy does not exist at this point for systems with potential energy.

Local Configuration Controllability

The problem of configuration controllability is harder than the one of configuration accessibility. Following [LM99, Lew00a], we will call a symmetric product in $\{Y_1, ..., Y_m\}$ *bad* if it contains an even number of each of the input vector–fields. Otherwise we will call it *good*. The *degree* is the total number of vector–fields. For example, $\langle\langle Y_a : Y_b\rangle : \langle Y_a : Y_b\rangle\rangle$ is bad and of degree 4, and $\langle Y_a : \langle Y_b : Y_b\rangle\rangle$ is good and of degree 3. If each bad symmetric product at q is a linear combination of good symmetric products of lower degree, then (2.613) is LCC at q.

Now, the single–input case can be solved completely: The system (2.613) with $m = 1$ is LCC iff $\dim(M) = 1$ [LM99].

Systems With Nonholonomic Constraints

Let us now add to the data a distribution D defining nonholonomic constraints. One of the interesting things about this affine connection approach is that we can easily integrate into our framework systems with nonholonomic constraints. As a simple example, consider a *rolling disk* (see Figure 2.16), with two inputs: (1) a 'rolling' torque, $F^1 = d\theta$ and (2) a 'spinning' torque, $F^2 = d\phi$. It can be analyzed as a nonholonomic system (see [Lew99, Lew00a]).

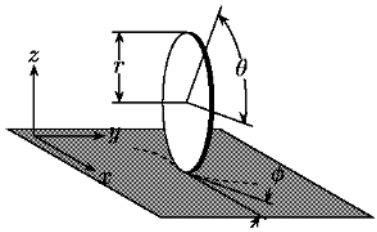


Fig. 2.16. Rolling disk problem (see text for explanation).

The control equations for a simple mechanical control system with constraints are:

$$\nabla_{\dot\gamma(t)}\dot\gamma(t) = \lambda(t) + u^a(t)\, Y_a\left(\gamma(t)\right)\left[-\operatorname{grad} V\left(\gamma(t)\right)\right], \qquad \dot\gamma(t) \in D_{\gamma(t)},$$

where $\lambda(t) \in D^{\perp}_{\gamma(t)}$ are *Lagrangian multipliers*.

Examples

1. Recall that for the simple robotic leg (Figure 2.13) above, Y_1 was internal torque and Y_2 was extension force. Now, in the following three cases:
(i) both inputs active – this system is LCA and LCC (satisfies sufficient condition);
(ii) Y_1 only, it is LCA but not LCC; and
(iii) Y_2 only, it is not LCA.
In theses three cases, C_{hor} is generated by the following linearly independent vector–fields:
(i) both inputs: $\{Y_1, Y_2, [Y_1, Y_2]\}$;
(ii) Y_1 only: $\{Y_1, \langle Y_1 : Y_1\rangle, \langle Y_1 : \langle Y_1 : Y_1\rangle\rangle\}$; and
(iii) Y_2 only: $\langle Y_2\rangle$.
Recall that with both inputs the system was not accessible in TM as a consequence of conservation of angular momentum. With the input Y_2 only, the control system behaves very simply when given zero initial velocity. The ball on the end of the leg just gets moved back and forth. This reflects the foliation of M by the maximal integral manifolds of C_{hor}, which are evidently 1D in

this case. With the Y_1 input, recall that the ball will always go 'outwards' no matter what one does with the input. Thus the system is not LCC. But apparently (since $rank(C_{hor}) = \dim(M)$) one can reach a non–empty open subset of M. The behavior exhibited in this case is typical of what one can expect for single–input systems with no potential energy.

2. For the planar rigid body (Figure 2.14) above, we have the following five cases:
(i) Y_1 and Y_2 active, this system is LCA and LCC (satisfies sufficient condition);
(ii) Y_1 and Y_3, it is LCA and LCC (satisfies sufficient condition);
(iii) Y_1 only or Y_3 only, not LCA;
(iv) Y_2 only, LCA but not LCC; and
(v) Y_2 and Y_3: LCA and LCC (fails sufficient condition).

Now, with the inputs Y_1 or Y_3 alone, the motion of the system is simple. In the first case the body moves along the line connecting the point of application of the force and the center of mass, and in the other case the body simply rotates. The equations in (x, y, θ) coordinates are

$$\ddot{x} = \frac{\cos\theta}{m}u^1 - \frac{\sin\theta}{m}u^2, \qquad \ddot{y} = \frac{\sin\theta}{m}u^1 + \frac{\cos\theta}{m}u^2, \qquad \ddot{\theta} = \frac{1}{J}\left(u^3 - hu^2\right),$$

which illustrates that the $\theta-$equation decouples when only Y_3 is applied. We make a change of coordinates for the case where we have only Y_1: $(\xi, \eta, \psi) = (x\cos\theta + y\sin\theta, -x\sin\theta + y\cos\theta, \theta)$. In these coordinates we have

$$\ddot{\xi} - 2\dot{\eta}\dot{\psi} - \xi\dot{\psi}^2 = \frac{1}{m}u^1, \qquad \ddot{\eta} + 2\dot{\xi}\dot{\psi} - \eta\dot{\psi}^2 = 0, \qquad \dot{\psi} = 0,$$

which illustrates the decoupling of the $\xi-$equation in this case.

C_{hor} has the following generators:
(i) Y_1 and Y_2: $\{Y_1, Y_2, [Y_1, Y_2]\}$;
(ii) Y_1 and Y_3: $\{Y_1, Y_3, [Y_1, Y_3]\}$;
(iii) Y_1 only or Y_3 only: $\{Y_1\}$ or $\{Y_3\}$;
(iv) Y_2 only: $\{Y_2, \langle Y_2 : Y_2\rangle, \langle Y_2 : \langle Y_2 : Y_2\rangle\rangle\}$;
(v) Y_2 and Y_3 $\{Y_2, Y_3, [Y_2, Y_3]\}$.

3. Recall that for the rolling disk (Figure 2.16) above, Y_1 was 'rolling' input and Y_2 was 'spinning' input. Now, in the following three cases:
(i) Y_1 and Y_2 active, this system is LCA and LCC (satisfies sufficient condition);
(ii) Y_1 only: not LCA; and
(iii) Y_2 only: not LCA.
In theses three cases, C_{hor} has generators:
(i) Y_1 and Y_2: $\{Y_1, Y_2, [Y_1, Y_2], [Y_2, [Y_1, Y_2]]\}$;
(ii) Y_1 only: $\{Y_1\}$; and
(iii) Y_2 only: $\{Y_2\}$.
The rolling disk passes the good/bad symmetric product test. Another way to show that it is LCC is to show that the inputs allow one to follow any curve

which is admitted by the constraints. Local configuration controllability then follows as the constraint distribution for the rolling disk has an involutive closure of maximal rank [Lew99].

Categorical Structure of Control Affine Systems

Control affine systems make a category $\mathcal{CAS}$ (see [Elk99]). The category $\mathcal{CAS}$ has the following data:

- An *object* in $\mathcal{CAS}$ is a pair $\sum = (M, \mathfrak{F} = \{f_0, f_1, ..., f_m\})$ where $\mathfrak{F}$ is a family of vector–fields

$$\dot{x}(t) = f_0(x(t)) + u^a(t) f_a(x(t))$$

  on the manifold M.
- A *morphism* sending $\sum = (M, \mathfrak{F} = \{f_0, f_1, ..., f_m\})$ to $\sum' = (M', \mathfrak{F}' = \{f'_0, f'_1, ..., f'_{m'}\})$ is a triple $(\psi, \lambda_0, \Lambda)$ where $\psi : M \to M'$, $\lambda_0 : M \to \mathbb{R}^{m'}$, and $\Lambda : M \to L(\mathbb{R}^m, \mathbb{R}^{m'})$ are smooth maps satisfying:
  1. $T_x\psi(f_a(x)) = \Lambda^\alpha_a(x) f'_\alpha(\psi(x)),\ \ a \in \{1, ..., m\},$ and
  2. $T_x\psi(f_0(x)) = f'_0(\psi(x)) + \lambda^\alpha_0 f'_\alpha(\psi(x))$.

  This corresponds to a change of state–input by

$$(x, u) \longmapsto (\psi(x), \lambda_0(x) + \Lambda(x)u).$$

Elkin [Elk99] discusses equivalence, inclusion, and factorization in the category $\mathcal{CAS}$. Using categorical language, he considers local equivalence for various classes of nonlinear control systems, including single–input systems, systems with involutive input distributions, and systems with three states and two inputs.

Lie–Adaptive Control

In this subsection we develop the concept of *machine learning* in the framework of Lie derivative control formalism (see (2.3.1) above). Consider an nD, SISO system in the standard affine form (2.601), rewritten here for convenience:

$$\dot{x}(t) = f(x) + g(x)\, u(t), \qquad y(t) = h(x), \tag{2.614}$$

As already stated, the feedback control law for the system (2.614) can be defined using Lie derivatives $\mathcal{L}_f h$ and $\mathcal{L}_g h$ of the system's output h along the vector–fields f and g.

If the SISO system (2.614) is a relatively simple (quasilinear) system with *relative degree* $r = 1$ it can be rewritten in a quasilinear form

$$\dot{x}(t) = \gamma_i(t)\, f_i(x) + d_j(t)\, g_j(x)\, u(t), \tag{2.615}$$

where γ_i $(i = 1, ..., n)$ and d_j $(j = 1, ..., m)$ are system's parameters, while f_i and g_j are smooth vector–fields.

In this case the feedback control law for *tracking* the *reference signal* $y_R = y_R(t)$ is defined as (see [Isi89, NS90])

$$u = \frac{-\mathcal{L}_f h + \dot{y}_R + \alpha\,(y_R - y)}{\mathcal{L}_g h}, \tag{2.616}$$

where α denotes the feedback gain.

Obviously, the problem of reference signal tracking is relatively simple and straightforward if we know all the system's parameters $\gamma_i(t)$ and $d_j(t)$ of (2.615). The question is can we apply a similar control law if the system parameters are unknown?

Now we have much harder problem of *adaptive signal tracking*. However, it appears that the feedback control law can be actually cast in a similar form (see [SI89, Gom94]):

$$\widehat{u} = \frac{-\widehat{\mathcal{L}_f h} + \dot{y}_R + \alpha\,(y_R - y)}{\widehat{\mathcal{L}_g h}}, \tag{2.617}$$

where Lie derivatives $\mathcal{L}_f h$ and $\mathcal{L}_g h$ of (2.616) have been replaced by their estimates $\widehat{\mathcal{L}_f h}$ and $\widehat{\mathcal{L}_g h}$, defined respectively as

$$\widehat{\mathcal{L}_f h} = \widehat{\gamma}_i(t)\,\mathcal{L}_{f_i} h, \qquad \widehat{\mathcal{L}_g h} = \widehat{d}_j(t)\,\mathcal{L}_{g_i} h,$$

in which $\widehat{\gamma}_i(t)$ and $\widehat{d}_j(t)$ are the estimates for $\gamma_i(t)$ and $d_j(t)$.

Therefore, we have the straightforward control law even in the uncertain case, provided that we are able to estimate the unknown system parameters. Probably the best known *parameter update law* is based on the so–called *Lyapunov criterion* (see [SI89]) and given by

$$\dot{\psi} = -\gamma\,\epsilon\,W, \tag{2.618}$$

where $\psi = \{\gamma_i - \widehat{\gamma}_i, d_j - \widehat{d}_j\}$ is the parameter estimation error, $\epsilon = y - y_R$ is the output error, and γ is a positive constant, while the matrix W is defined as:

$$W = \left[W_1^T\, W_2^T\right]^T, \qquad \text{with}$$

$$W_1 = \begin{bmatrix} \mathcal{L}_{f_1} h \\ \vdots \\ \mathcal{L}_{f_n} h \end{bmatrix}, \qquad W_2 = \begin{bmatrix} \mathcal{L}_{g_1} h \\ \vdots \\ \mathcal{L}_{g_m} h \end{bmatrix} \cdot \frac{-\widehat{\mathcal{L}_f h} + \dot{y}_R + \alpha\,(y_R - y)}{\widehat{\mathcal{L}_g h}}.$$

The proposed adaptive control formalism (2.617–2.618) can be efficiently applied wherever we have a problem of tracking a given signal with an output of a SISO–system (2.614–2.615) with unknown parameters.

2.3.3 Hamiltonian Optimal Control and Maximum Principle

Hamiltonian Control Systems

Hamiltonian control system on a symplectic manifold (P, ω) is defined as an affine control system whose drift and control vector–fields are Hamiltonian. It can be written as

$$\dot{p} = X_H(p) + u^a X_a(p),$$

where the vector–fields X_a are assumed to be Hamiltonian with Hamiltonian H_a for $a = 1, ..., m$. Examples of systems which are (at least locally) Hamiltonian control systems are those which evolve on the symplectic manifold T^*M and where the control Hamiltonians are simply coordinate functions on M.

Alternatively, Hamiltonian control systems can be defined on Poisson manifolds. However, for the purposes of this subsection, it will be more natural to work within the Poisson context. Recall that given a smooth Hamiltonian function $h : M \to \mathbb{R}$, on the Poisson manifold M, the Poisson bracket $\{,\} : C^\infty(M) \times C^\infty(M) \to C^\infty(M)$ (such that $\{f, g\} = -\{g, f\}$, $\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} = 0$, and $\{fg, h\} = \{f, h\}g + f\{g, h\}$) allows us to obtain a Hamiltonian vector–field X_h with Hamiltonian h through the equality

$$\mathcal{L}_{X_h} f = \{f, h\}, \qquad \text{for all} \quad f \in C^\infty(M),$$

where $\mathcal{L}_{X_h} f$ is the Lie derivative of f along X_h. Note that the vector–field X_h is well defined since the Poisson bracket verifies the Leibniz rule and therefore defines a derivation on $C^\infty(M)$ (see [MR99]). Furthermore $C^\infty(M)$ equipped with a Poisson bracket is a Lie algebra, called a Poisson algebra. Also, we say that the Poisson structure on M is nondegenerate if the $\{,\}-$associated map $B^\# : T^*M \to TM$ defined by

$$dg(B^\#(x)(df)) = B(x)(df, dg),$$

(where df denotes the exterior derivative of f) is an isomorphism for every $x \in M$.

An *affine Hamiltonian control* system $\Sigma = (U, M, h)$ consists of a smooth manifold U (the *input space*), a Poisson manifold M with nondegenerate Poisson bracket (the *state–space*), and a smooth function $H : M \times U \to \mathbb{R}$ (the *controlled Hamiltonian*). Furthermore, H is locally of the form $H = h_0 + h_i u^i$ $(i = 1, ..., n)$, with h_i locally defined smooth real valued maps and u^i local coordinates for U [TP01].

Using the controlled Hamiltonian and the Poisson structure on M we can recover the familiar system map $F : M \times U \to TM$, locally given by

$$F = X_{h_0} + X_{h_i} u^i,$$

and defines an affine distribution on M given by

$$\mathcal{D}_M(x) = X_{h_0}(x) + \text{span}\{X_{h_1}(x), X_{h_2}(x), ..., X_{h_n}(x)\}.$$

This distribution captures all the possible directions of motion available at a certain point x, and therefore describes a control system, up to a parametrization by control inputs. This affine distribution will is our main object of interest here, and we will assume that the rank of $\mathcal{D}_M$ does not change with x. Furthermore, we denote an affine distribution $\mathcal{D}_M$ by $X + \Delta$, where X is a vector–field and Δ a distribution. When this affine distribution is defined by a Hamiltonian control system we have $X = X_{h_0}$ and $\Delta = \text{span}\{X_{h_1}(x), X_{h_2}(x), ..., X_{h_n}(x)\}$. A similar reasoning is possible at the level of Hamiltonians. Locally, we can define the following affine space of smooth maps

$$\mathcal{H}_M = h_0 + \text{span}_{\mathbb{R}}\{h_1, h_1, ..., h_n\},$$

which defines $\mathcal{D}_M$ by the equality

$$\mathcal{D}_M = B^{\#}(d\mathcal{H}_M),$$

where we used the notation $d\mathcal{H}_M$ to denote the set $\cup_{h\in\mathcal{H}_M} dh$. We also use the notation $\mathcal{H}_M = h_0 + H_\Delta$ for an affine space of smooth maps where h_0 is a smooth map and H_Δ a linear space of smooth maps.

Having defined Hamiltonian control systems we turn to their trajectories or solutions: A smooth curve $\gamma : I \to M$, $I \subseteq \mathbb{R}_0^+$ is called a *trajectory of control system* $\Sigma = (U, M, H)$, iff there exists a curve $\gamma^U : I \to U$ satisfying [TP01]

$$\dot{y}(t) = F(\gamma(t), \gamma^U(t)), \qquad \text{for every } \ t \in I.$$

Now, given a Hamiltonian control system and a desired property, an abstracted Hamiltonian system is a reduced system that preserves the property of interest while ignoring modelling detail (see [TP01]). Property preserving abstractions of control systems are important for reducing the complexity of their analysis or design. From an analysis perspective, given a large scale control system and a property to be verified, one extracts a smaller abstracted system with equivalent properties. Checking the property on the abstraction is then equivalent to checking the property on the original system. From a design perspective, rather than designing a controller for the original large scale system, one designs a controller for the smaller abstracted system, and then refines the design to the original system while incorporating modelling detail.

This approach critically depends on whether we are able to construct hierarchies of abstractions as well as characterize conditions under which various properties of interest propagate from the original to the abstracted system and vice versa. In [PLS00], hierarchical abstractions of linear control systems were extracted using computationally efficient constructions, and conditions under which controllability of the abstracted system implied controllability of the original system were obtained. This led to extremely efficient hierarchical controllability algorithms. In the same spirit, abstractions of nonlinear control affine systems were considered in [PS02], and the canonical construction for linear systems was generalized to nonlinear control affine systems.

In [TP01], abstractions of Hamiltonian control systems are considered, which are control systems completely specified by controlled Hamiltonians. This additional structure allows to simplify the abstraction process by working with functions instead of vector–fields or distributions as is the case for general nonlinear systems [PS02]. This is possible since the controlled Hamiltonian contains all the relevant information that must be captured by the abstracted system. On the other hand, to be able to relate the dynamics induced by the controlled Hamiltonians, we need to restrict the class of abstracting maps to those that preserve the Hamiltonian structure. More precisely, given a Hamiltonian control system on a Poisson manifold M, and a (quotient) Poisson map $\phi : M \to N$, one presents a canonical construction that extracts an abstracted Hamiltonian control system on N. One then characterizes abstracting maps for which the original and abstracted system are equivalent from a local accessibility point of view [TP01].

Pontryagin's Maximum Principle

Recall that the *Pontryagin Maximum Principle* (PMP, see [PBK62, IK80]) applies to a general optimization problem called a *Bolza problem*. To apply PMP to optimal control, we need to define Hamiltonian function:

$$H(\psi, x, u) = (\psi, f(x, u)) = \psi_i f^i(x, u), \qquad (i = 1, ..., n). \tag{2.619}$$

Then in order for a control $u(t)$ and a trajectory $x(t)$ to be *optimal*, it is necessary that there exist a nonzero absolutely continuous vector function $\psi(t) = (\psi_0(t), \psi_1(t), ..., \psi_n(t))$ corresponding to the functions $u(t)$ and $x(t)$ such that:

1. The function $H(\psi(t), x(t), u(t))$ attains its maximum at the point $u = u(t)$ almost everywhere in the interval $t_0 \leq t \leq T$,

$$H(\psi(t), x(t), u(t)) = \max_{u \in U} H(\psi(t), x(t), u(t)).$$

2. At the *terminal time* T, the following relations are satisfied: $\psi_0(T) \leq 0$ and $H(\psi(T), x(T), u(T)) = 0$.

PMP states the following algorithm: To maximize the set of *steering functions* $\gamma_i x^i(t)$ (with n constants γ_i) for controlling the changes in the *state variables*

$$\dot{x}^i(t) = f^i(x^i, u_k), \qquad (i = 0, 1, ..., n, \quad k = 1, ..., m),$$

we maximize at each instant the Hamiltonian function (2.619), where

$$\dot{\psi}_i = -\psi_j \frac{\partial f^j}{\partial x^i} \qquad \text{and} \qquad \psi_i(T) = \gamma_i.$$

Affine Control Systems

Now, let us look at PMP as applied to the *affine control system* (see [Lew00b])

$$\dot{y}(t) = f_0(\gamma(t)) + u^a(t)\, f_a(\gamma(t)),$$

with $\gamma(t) \in M$, u taking values in $U \subset \mathbb{R}^m$, and objective function $L(x,u)$.

We need to have the control Hamiltonian on $U \times T^*M$:

$$H(\alpha_x, u) = \underbrace{\alpha_x(f_0(x))}_{H_1} + \underbrace{\alpha_x(u^a f_a(x))}_{H_2} - \underbrace{L(x,u)}_{H_3}.$$

One of several consequences of the PMP is that if (u,γ) is a *minimizer* then there exists a $1-$form field λ along γ with the property that $t \mapsto \lambda(t)$ is an integral curve for the time–dependent Hamiltonian $(\alpha_x, u) \mapsto H(\alpha_x, u)$. The Hamiltonian $H(\alpha_x, u)$ is a sum of three terms, and so too will be the Hamiltonian vector–field.

Let us look at the first term, that with (old) Hamiltonian $H_1 = \alpha_x(f_0(x))$. In local coordinates X_{H_1} is written as

$$\dot{x}^i = f_0^i(x), \qquad \dot{p}_i = -\frac{\partial f_0^j(x)}{\partial x^i} p_j. \tag{2.620}$$

X_{H_1} is the *cotangent lift* of f_0 and, following [Lew00b], we denote it $f_0^{T^*}$. So we want to understand $f_0^{T^*}$ on TM with $f_0 = Z$.

Let f_0 be a vector–field on a general manifold N with f_0^T its *tangent lift* defined by

$$f_0^T(v_x) = \left.\frac{d}{dt}\right|_{t=0} T_x F_t(v_x),$$

where F_t denotes the flow of f_0. Therefore, f_0^T is the 'linearization' of f_0 and in local coordinates it is given by (compare with (2.620))

$$\dot{x}^i = f_0^i(x), \qquad \dot{v}^i = -\frac{\partial f_0^i(x)}{\partial x^j} v^j.$$

The flow of f_0^T measures how the integral curves of f_0 change as we change the initial condition in the direction of v_x.

Now, perhaps we can understand Z^T on TM with $f_0 = Z$ in the discussion of tangent lift. Let $\gamma(t)$ be a geodesic. By varying the initial condition for the geodesic we generate an 'infinitesimal variation' which satisfies the *extended Jacobi equation*,

$$\nabla^2_{\dot{y}(t)} \xi(t) + R(\xi(t), \dot{y}(t))\, \dot{y}(t) + \nabla_{\dot{y}(t)} \left(T(\xi(t), \dot{y}(t))\right) = 0. \tag{2.621}$$

To make the 'connection' between Z^T and the Jacobi equation, we perform constructions on the tangent bundle using the spray Z. ∇ comes from a linear

connection on M which induces an *Ehresmann connection* on $\tau_M : TM \to M$. Thus we may write $T_{v_q}TM \simeq T_qM \oplus T_qM$. Now, if $I_M : TTM \to TTM$ is the canonical involution then $I_M^* Z^T$ is a spray. We use $I_M^* Z^T$ to induce an Ehresmann connection on $\tau_{TM} : TTM \to TM$. Thus,

$$T_{X_{v_q}}TTM \simeq T_{v_q}TM \oplus T_{v_q}TM \simeq \underbrace{T_qM \oplus T_qM}_{geodesic\,equations} \oplus \underbrace{T_qM \oplus T_qM}_{variation\,equations} .$$

One represents Z^T in this splitting and determines that the Jacobi equation sits 'inside' one of the four components. Now one applies similar constructions to T^*TM and Z^{T^*} to derive a $1-$form version of the Jacobi equation (2.621), the so–called *adjoint Jacobi equation* [Lew00b]:

$$\nabla^2_{\dot{y}(t)}\lambda(t) + R^*(\lambda(t), \dot{y}(t))\,\dot{y}(t) - T^*\left(\nabla_{\dot{y}(t)}\lambda(t), \dot{y}(t)\right) = 0, \tag{2.622}$$

where we have used $\langle R^*(\alpha, u)v; \omega\rangle = \langle \alpha; R(\omega, u)v\rangle$, and $\langle T^*(\alpha, u); \omega\rangle = \langle \alpha; T(\omega, u)\rangle$.

The adjoint Jacobi equation forms the backbone of a general statement of the PMP for affine connection control systems. When objective function is the Lagrangian $L(u, v_q) = \frac{1}{2}g(v_q, v_q)$, when ∇ is the *Levi–Civita connection* for the Riemannian metric g, and when the system is fully actuated, then we recover the equation of [NHP89]

$$\nabla^3_{\dot{y}(t)}\,\dot{y}(t) + R\left(\nabla_{\dot{y}(t)}\dot{y}(t), \dot{y}(t)\right) = 0.$$

Therefore, the adjoint Jacobi equation (2.622) captures the interesting part of the Hamiltonian vector–field Z^{T^*}, which comes from the PMP, in terms of affine geometry, i.e., from Z^{T^*} follows

$$\nabla_{\dot{y}(t)}\dot{y}(t) = 0, \qquad \nabla^2_{\dot{y}(t)}\lambda(t) + R^*(\lambda(t), \dot{y}(t))\,\dot{y}(t) - T^*\left(\nabla_{\dot{y}(t)}\lambda(t), \dot{y}(t)\right) = 0.$$

The geometry of Z on TM gives a way of globally pulling out the adjoint Jacobi equation from the PMP in an intrinsic manner, which is not generally possible in the PMP [Lew00b].

2.3.4 Path–Integral Optimal Control of Stochastic Systems

A path–integral based optimal control model for nonlinear stochastic systems has recently been developed in [Kap05]. The author addressed the role of noise and the issue of efficient computation in stochastic optimal control problems. He considered a class of nonlinear control problems that can be formulated as a *path integral* and where the noise plays the role of temperature. The path integral displays symmetry breaking and there exist a critical noise value that separates regimes where optimal control yields qualitatively different solutions. The path integral can be computed efficiently by Monte Carlo integration or by Laplace approximation, and can therefore be used to solve high dimensional stochastic control problems.

Recall that *optimal control of nonlinear systems in the presence of noise* is a very general problem that occurs in many areas of science and engineering. It underlies *autonomous system behavior*, such as the control of movement and planning of actions of animals and robots, but also optimization of financial investment policies and control of chemical plants. The problem is stated as: *given that the system is in this configuration at this time, what is the optimal course of action to reach a goal state at some future time.* The *cost* of each time course of actions consists typically of a path contribution, that specifies the amount of work or other cost of the trajectory, and an end cost, that specifies to what extend the trajectory reaches the goal state.

Also recall that in the absence of noise, the optimal control problem can be solved in two ways: using (i) the *Pontryagin Maximum Principle* (PMP, see previous subsection), which represents a pair of ordinary differential equations that are similar to the Hamiltonian equations; or (ii) the *Hamilton–Jacobi–Bellman* (HJB) equation, which is a partial differential equation (PDE) [BK64].

In the presence of *Wiener noise*, the PMP formalism is replaced by a set of stochastic differential equations (SDEs), which become difficult to solve (compare with [YZ99]). The inclusion of noise in the HJB framework is mathematically quite straightforward, yielding the so–called *stochastic HJB equation* [Ste93]. However, its solution requires a *discretization of space and time* and the computation becomes intractable in both memory requirement and CPU time in high dimensions. As a result, deterministic control can be computed efficiently using the PMP approach, but stochastic control is intractable due to the curse of dimensionality.

For small noise, one expects that optimal stochastic control resembles optimal deterministic control, but for larger noise, the optimal stochastic control can be entirely different from the deterministic control [RN03]. However, there is currently no good understanding how *noise* affects optimal control.

In this subsection, we address both the issue of efficient computation and the role of noise in stochastic optimal control. We consider a class of nonlinear stochastic control problems, that can be formulated as a statistical mechanics problem. This class of control problems includes arbitrary dynamical systems, but with a limited control mechanism. It contains linear–quadratic [Ste93] control as a special case. We show that under certain conditions on the noise, the HJB equation can be written as a *linear* PDE

$$-\partial_t \psi = H\psi, \tag{2.623}$$

with H a (non–Hermitian) operator. Equation (2.623) must be solved subject to a boundary condition at the end time. As a result of the linearity of (2.623), the solution can be obtained in terms of a diffusion process evolving forward in time, and can be written as a path integral. The path–integral has a direct interpretation as a free energy, where noise plays the role of temperature.

This link between stochastic optimal control and a free energy has an immediate consequence that phenomena that allow for a free energy description,

typically display phase transitions. [Kap05] has argued that for stochastic optimal control one can identify a critical noise value that separates regimes where the optimal control has been qualitatively different. He showed how the Laplace approximation can be combined with Monte Carlo sampling to efficiently calculate the optimal control.

Path–Integral Approach to Nonlinear Stochastic Optimal Control

Let x^i be an nD stochastic variable that is subject to the SDE

$$dx^i = (b^i(x^i,t) + u^i)dt + d\xi^i \tag{2.624}$$

with $d\xi^i$ being an nD Wiener process with $\left\langle d\xi_i d\xi_j \right\rangle = \nu_{ij}dt$, and functions ν_{ij} independent of x^i, u^i and time t. The term $b^i(x^i,t)$ is an arbitrary nD function of x^i and t, and u^i represents an nD vector of control variables. Given the value of x^i at an initial time t, the *stochastic optimal control problem* is to find the control path $u^i(\cdot)$ that minimizes

$$C(x^i,t,u^i(\cdot)) = \left\langle \phi(x^i(t_f)) + \int_t^{t_f} d\tau (\frac{1}{2}u_i(\tau)Ru^i(\tau) + V(x^i(\tau),\tau)) \right\rangle_{x^i}, \tag{2.625}$$

with R a matrix, $V(x^i,t)$ a time–dependent potential, and $\phi(x^i)$ the *end cost*. The brackets $\langle\rangle_{x^i}$ denote expectation value with respect to the stochastic trajectories (2.624) that start at x^i.

One defines the *optimal cost–to–go function* from any time t and state x^i as

$$J(x^i,t) = \min_{u^i(\cdot)} C(x^i,t,u^i(\cdot)).$$

J satisfies the following stochastic HJB equation [Kap05]

$$\begin{aligned} -\partial_t J(x^i,t) &= \min_{u^i}\left(\frac{1}{2}u_i R u^i + V + (b_i+u_i)\partial_{x^i}J(x^i,t) + \frac{1}{2}\nu_{ij}\partial_{x^i x^j}J(x^i,t)\right) \\ &= -\frac{1}{2}R^{-1}\partial_{x^i}J(x^i,t)\partial_{x^i}J + V + b_i\partial_{x^i}J(x^i,t) + \frac{1}{2}\nu_{ij}\partial_{x^i x^j}J(x^i,t), \end{aligned} \tag{2.626}$$

where $b_i = (b^i)^T$, and $u_i = (u^i)^T$, and

$$u^i = -R^{-1}\partial_{x^i}J(x^i,t) \tag{2.627}$$

is the optimal control at the point (x^i,t). The HJB equation is nonlinear in J and must be solved with end boundary condition $J(x^i,t_f) = \phi(x^i)$.

Let us define $\psi(x^i,t)$ through the Log Transform

$$J(x^i,t) = -\lambda \log \psi(x^i,t), \tag{2.628}$$

and assume that there exists a scalar λ such that

$$\lambda\delta_{ij} = (R\nu)_{ij}, \tag{2.629}$$

with δ_{ij} the Kronecker delta. In the one dimensional case, such a λ can always be found. In the higher dimensional case, this restricts the matrices $R \propto (\nu_{ij})^{-1}$. Equation (2.629) reduces the dependence of optimal control on the nD noise matrix to a scalar value λ that will play the role of temperature, while (2.626) reduces to the linear equation (2.623) with

$$H = -\frac{V}{\lambda} + b_i\partial_{x^i} + \frac{1}{2}\nu_{ij}\partial_{x^ix^j}J(x^i,t).$$

Let $\rho(y^i,\tau|x^i,t)$ with $\rho(y^i,t|x^i,t) = \delta(y^i - x^i)$ describe a *diffusion process* for $\tau > t$ defined by the *Fokker–Planck equation*

$$\partial_\tau\rho = H^\dagger\rho = -\frac{V}{\lambda}\rho - \partial_{x^i}(b_i\rho) + \frac{1}{2}\nu_{ij}\partial_{x^ix^j}J(x^i,t)\rho \tag{2.630}$$

with $H^\dagger$ the Hermitian–conjugate of H. Then $A(\tau) = \int dy^i\rho(y^i,\tau|x^i,t)\psi(y^i,\tau)$ is independent of τ and in particular $A(t) = A(t_f)$. It immediately follows that

$$\psi(x^i,t) = \int dy^i\rho(y^i,t_f|x^i,t)\exp(-\phi(y^i)/\lambda) \tag{2.631}$$

We arrive at the important conclusion that $\psi(x^i,t)$ can be computed either by backward integration using (2.623) or by forward integration of a diffusion process given by (2.630).

We can write the integral in (2.631) as a path integral. Following [Kap05] we can divide the time interval $t \to t_f$ in n_1 intervals and write $\rho(y^i,t_f|x^i,t) = \prod_{i=1}^{n_1}\rho(x_i^i,t_i|x_{i-1}^i,t_{i-1})$ and let $n_1 \to \infty$. The result is

$$\psi(x^i,t) = \int [dx^i]_{x^i}\exp\left(-\frac{1}{\lambda}S(x^i(t\to t_f))\right) \tag{2.632}$$

with $\int[dx^i]_{x^i}$ an integral over all paths $x^i(t\to t_f)$ that start at x^i and with

$$S(x^i(t\to t_f)) = \phi(x^i(t_f) + \int_t^{t_f} d\tau(\frac{1}{2}(\dot{x}^i - b_i(x^i,\tau))R(\dot{x}^i - b^i(x^i,\tau)) + V(x^i,\tau)) \tag{2.633}$$

the Action associated with a path. From (2.628) and (2.632), the cost–to–go $J(x,t)$ becomes a log partition sum (i.e., a free energy) with temperature λ.

Monte Carlo Sampling

The path integral (2.632) can be estimated by stochastic integration from t to t_f of the diffusion process (2.630) in which particles get annihilated at a rate $V(x^i,t)/\lambda$ [Kap05]:

$$x^i = x^i + b^i(x^i,t)dt + d\xi^i, \qquad \text{with probability} \quad 1 - Vdt/\lambda$$
$$x^i = \dagger, \qquad \text{with probability} \quad Vdt/\lambda \tag{2.634}$$

where † denotes that the particle is taken out of the simulation. Denote the trajectories by $x^i_\alpha(t \to t_f)$, $(\alpha = 1, \ldots, N)$. Then, $\psi(x^i,t)$ and u^i are estimated as

$$\hat{\psi}(x^i,t) = \sum_{\alpha \in \text{alive}} w_\alpha, \qquad u^i dt = \frac{1}{\hat{\psi}(x^i,t)} \sum_{\alpha \in \text{alive}}^{N} w_\alpha d\xi^i_\alpha(t), \tag{2.635}$$
$$\text{with} \qquad w_\alpha = \frac{1}{N} \exp(-\phi(x^i_\alpha(t_f))/\lambda),$$

where 'alive' denotes the subset of trajectories that do not get killed along the way by the † operation. The normalization $1/N$ ensures that the annihilation process is properly taken into account. Equation (2.635) states that optimal control at time t is obtained by averaging the initial directions of the noise component of the trajectories $d\xi^i_\alpha(t)$, weighted by their success at t_f.

The above sampling procedure can be quite inefficient, when many trajectories get annihilated. One of the simplest procedures to improve it is by importance sampling. We replace the diffusion process that yields $\rho(y^i, t_f|x^i,t)$ by another diffusion process, that will yield $\rho'(y^i,t_f|x^i,t) = \exp(-S'/\lambda)$. Then (2.632) becomes,

$$\psi(x^i,t) = \int [dx^i]_{x^i} \exp\left(-S'/\lambda\right) \exp\left(-(S-S')/\lambda\right).$$

The idea is to chose ρ' such as to make the sampling of the path integral as efficient as possible. Following [Kap05], here we use the Laplace approximation, which is given by the k deterministic trajectories $x_\beta(t \to t_f)$ that minimize the Action

$$J(x^i,t) \approx -\lambda \log \sum_{\beta=1}^{k} \exp(-S(x^i_\beta(t \to t_f)/\lambda).$$

The Laplace approximation ignores all fluctuations around the modes and becomes exact in the limit $\lambda \to 0$. The Laplace approximation can be computed efficiently, requiring $\mathcal{O}(n^2 m^2)$ operations, where m is the number of time discretization.

For each Laplace trajectory, we can define a diffusion processes ρ'_β according to (2.634) with $b^i(x^i,t) = x^i_\beta(t)$. The estimators for ψ and u^i are given again by (2.635), but with weights

$$w_\alpha = \frac{1}{N} \exp\left(-\left(S(x^i_\alpha(t \to t_f)) - S'_\beta(x^i_\alpha(t \to t_f))\right)/\lambda\right).$$

S is the original Action (2.633) and S'_β is the new Action for the Laplace guided diffusion. When there are multiple Laplace trajectories one should include all of these in the sample.

2.3.5 Life: Complex Dynamics of Gene Regulatory Networks

A *living organism* represents a complex interconnection of many control units that form a *gene regulatory network*.[28] In developmental biology, clusters of DNA sequence elements, the so–called *cis–regulatory module*, are target sites for transcription factors. One cis–regulatory module controls a set of spatio–temporal gene expressions [LD05]. One transcription factor can interact with many modules, and one module is controlled by many transcription factors. Thus, the spatio–temporal variation of a gene expression is a consequence of an interconnected network of interactions.

In the last five years we have witnessed a growing interest in experiments within the field of systems biology that require mathematical models to describe the experimental results [OTK02, BMT04, MZA03, GBL03, HDR02, HYW03, BS00]. A model for gene regulatory networks is also closely related with *synthetic biology*, the engineering counterpart of systems biology [Bre04, Ark01, EB01]. Similar to the development of the field of electronics, where complex equipment is built on interconnected simple devices, the field of synthetic biology aims to build simple molecular devices for later use in more complex molecular machines [Alo03, GEH02]. To build a robust, reliable, and simple device, the molecular engineer needs to have a mathematical description of the system in order to evaluate the number, range and meaning of a group of parameters that are critical for the device functionality.

Briefly, the so–called *messenger RNA* (mRNA) is RNA that encodes and carries information from DNA[29] to sites of protein synthesis. The brief 'life cycle' of an mRNAs begins with transcription and ultimately ends in degradation. During their life, mRNAs may also be processed, edited,[30] and trans-

[28] Recall that an *irreducible complex system* is a single system which is composed of several interacting parts that contribute to the basic function, and where the removal of any one of the parts causes the system to effectively cease functioning.

[29] During transcription, RNA polymerase makes a copy of a gene from the DNA to mRNA as needed. This process is similar in eukaryotes and prokaryotes. One notable difference, however, is that eukaryotic RNA polymerase associates with mRNA processing enzymes during transcription so that processing can proceed quickly after the start of transcription. The short–lived, unprocessed or partially processed, product is termed pre–mRNA; once completely processed, it is termed *mature mRNA*.

[30] In some instances, an mRNA will be edited, changing the nucleotide composition of that mRNA. An example in humans is the apolipoprotein B mRNA, which is edited in some tissues, but not others. The editing creates an early stop codon, which upon translation, produces a shorter protein.

ported prior to translation.[31] Eukaryotic mRNAs often require extensive processing and transport, while prokaryotic mRNAs do not.[32]

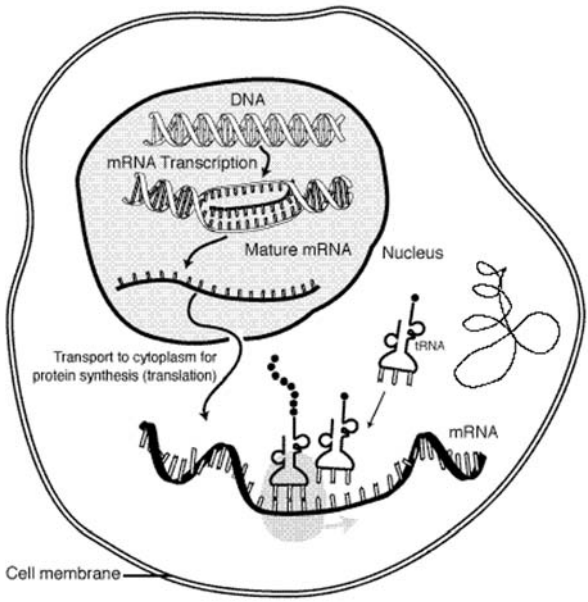


Fig. 2.17. The 'life cycle' of an mRNA in a eukaryotic cell: RNA is transcribed in the nucleus; once completely processed, it is transported to the cytoplasm and translated by the ribosome; at the end of its life, the mRNA is degraded (adapted from [Wik05]).

[31] *Translation* is the second process of protein biosynthesis (part of the overall process of gene expression). In translation, *messenger RNA* is decoded to produce a specific polypeptide according to the rules specified by the genetic code. Translation is necessarily preceded by transcription. Similarly to transcription, translation proceeds in three phases: initiation, elongation and termination (all describing the growth of the amino acid chain, or polypeptide that is the product of translation). The capacity of disabling or inhibiting translation in protein biosynthesis is used by antibiotics such as: anisomycin, cycloheximide, chloramphenicol and tetracycline.

[32] Processing of mRNA differs greatly between eukaryotes and prokaryotes. Prokaryotic mRNA is essentially mature upon transcription and requires no processing (except in rare cases). Eukaryotic pre-mRNA, however, requires extensive processing. A fully processed mRNA includes the 5' cap, 5' UTR, coding region, 3' UTR, and poly(A) tail. Coding regions are composed of codons, which are decoded and translated into protein by the ribosome. Coding regions begin with the start codon and end with the one of three possible stop codons. In addition to protein–coding, portions of coding regions may also serve as regulatory sequences.

Recently, it has been proposed that signal generators controlled by light, can be incorporated into the gene regulatory network [LW05]. With the help of these *light–controlled signal generators* [SHT02], different types of signal perturbations could be imposed on the gene regulatory network. The mRNA is controlled by an input signal generator and thus the gene state will evolve in time. Such a generator can be practically constructed using an yeast two–hybrid system. The light–switch is based on phytochrome that is synthesized in darkness in the Q1 form. A red light photon of wavelength 664 nm shined on the Q1 form of the protein transforms it in the form Q2. Figure 2.18 presents the state of the switch after the effect of the corresponding wavelength took place. When Q2 absorbs a far red light of wavelength 748 nm, the molecule Q goes back to its original form, Q1. These transitions take milliseconds. The protein P interacts only with the Q2 form, recruiting thus the activation domain to the target promoter. In this position, the promoter is open and the gene is transcribed. After the desired time elapsed, the gene can be turned off by a photon from a far red light source. Using a sequence of red and far red light pulses the molecular switch can be opened and closed.

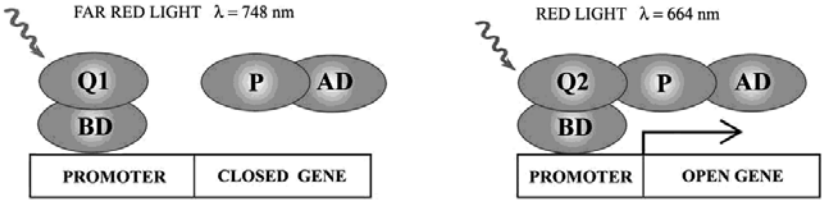


Fig. 2.18. Input signal generator for mRNA1 control. Here, AD is activation domain, BD is binding domain, Q is a protein that changes its form upon light exposure, from Q1 to Q2 and back, P is a protein that interacts only with Q2 (adapted from [SHT02]).

In [LW05], the Master Equation was solved for genetic systems that were linear in the transition probabilities. The network's response to signal generators was expressed in terms of a transfer matrix for the first and second order moments of the stochastic process.

More recently, the structure of a *stochastic nonlinear gene regulatory network* has been uncovered by studying its response to *input signal generators* [AL05]. Each genetic regulatory network has been built on a set of interacting molecular species, described by a set of n *gene state coordinates* $q = (q^1, \ldots, q^n)$ (see Figure 2.19, for a nonlinear connection of two linear systems: System 1, with mRNA1 and protein1 as molecular species, nonlinearly coupled with System 2 with mRNA2 and protein2 as its molecular species). Namely, each state coordinate q^i represents the number of molecules for the component i of the state q, so that the components $i = 1 \ldots n$ can represent different proteins, mRNAs, or the same protein but in different configurations

or localizations (in nucleus, on the membrane, in Golgi apparatus, etc.). The state coordinates q^i change in time due to a set of possible *transitions* ϵ. If at time t the state coordinate is q, then at time $t+dt$ the state coordinate will be $q+\epsilon$, with ϵ being one of the possible transitions. Each transition is governed by its transition probability $T_\epsilon(q,t)$ that depends on the state coordinates q^i and time t.

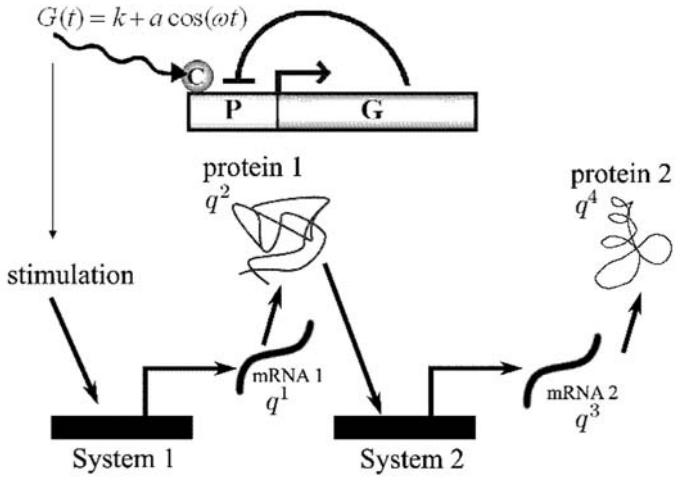


Fig. 2.19. Up: A schematic of an autoregulatory genetic network; here, the gene G is under the influence of a cofactor C that rhythmically modulates the activity of the promoter P. Down: A simple input–output network composed of two interconnected (i.e., nonlinearly–coupled) *gene regulatory systems*, described by the set of gene coordinates $q = (q^1, q^2, q^3, q^4)$ (adapted from [LW05] and [AL05]).

To excite a gene regulatory network, an experimental biologist has to act on it through a set of *signal generators* $M_\epsilon^m(t)$, to get the *transition probability*

$$T_\epsilon(q,t) = \sum_m M_\epsilon^m e_{\overline{m}}(q) + G(t),$$

where $G(t)$ denotes the part of the signal generator that modulates the mRNA1 transcription, $\mathbf{e}_{\overline{m}}(q)$ are polynomials given by decreasing factorials in coordinates q^i, while $\overline{m}$ stands for a set $\overline{m} = (m_1, m_2, ..., m_n)$ of integers.

For example, one of the basic elements of a gene regulatory network is a gene that controls its own transcription, the so–called *Hill feedback control* [LRR02]. Here, the protein acts on mRNA production through a simple transfer function of the form $\frac{a}{b+p^2}$, where p is the Laplace complex variable. When the number of protein molecules increases, the rate of mRNA production will decrease, stabilizing the system's transcription and translation. This kind of feedback control is employed in the description of many biological systems. In

[OTL04] it was used to explain the appearance of multistability in the lactose utilization network of Escherichia coli, while in [EL00] it was used to describe a stable oscillator constructed from three genes that repress themselves in a closed loop.

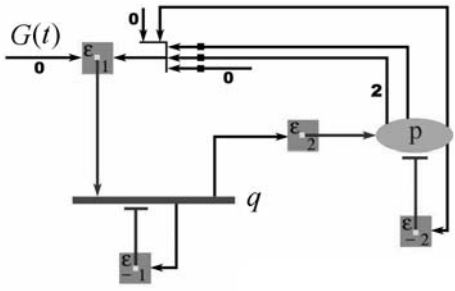


Fig. 2.20. Autoregulatory gene. The feedback has a Hill coefficient of 2.

Using the basic building blocks (the state q, the transitions ϵ and the transition probabilities T_ϵ), the structure of the Hill feedback system (Figure 2.20) is given by [LRR02]:

$\epsilon_1 = (1,0)$	$\epsilon_{-1} = (-1,0)$	$\epsilon_2 = (0,1)$	$\epsilon_{-2} = (0,-1)$
$T_{\epsilon_1} = G(t) + \frac{a_1 + a_2 p}{b_1 + b_2 p + b_3 p(p-1)}$	$T_{\epsilon_{-1}} = \gamma_q q$	$T_{\epsilon_2} = Kq$	$T_{\epsilon_{-2}} = \gamma_p p$

The *Master equation* for the probability evolution $P(q-\epsilon,t)$ of the gene regulatory network $P(q^1,...q^n,t)$ to be in the state q at time t is given by

$$\partial_t P(q,t) = \sum_\epsilon T_\epsilon(q-\epsilon,t)P(q-\epsilon,t) - \sum_\epsilon T_\epsilon(q,t)P(q,t). \qquad (2.636)$$

We are interested in the mean values for differen tmolecules, $\langle q^{\overline{m}}\rangle = \sum_q q^{\overline{m}} P(q,t)$, as well as their correlations. The generating function for such variables is $F(z,t)$, the $\mathcal{Z}$–transform of the network probability $P(q,t)$ given by (2.636),

$$F(z,t) \equiv \mathcal{Z}(P(q,t)) = \sum_q z^q P(q,t), \qquad \text{with} \qquad \begin{cases} z = (z_1,\ldots,z_n), \\ z^q = z_1^{q^1} \ldots z_n^{q^n}. \end{cases}$$

Using the boundary condition, $\overline{m} \geq -\epsilon$, for every m in $T_\epsilon(q,t)$ with $M_\epsilon^m(t) \neq 0$, we get the evolution equation for the variable $F(z,t)$,

$$\partial_t F = \sum_{\epsilon,m} (z^{\epsilon+\overline{m}} - z^{\overline{m}}) M_\epsilon^m(t) \partial_m F, \qquad (2.637)$$

where we used the property that $\mathcal{Z}(e_{\overline{m}}(q)P(q,t))) = z^{\overline{m}}\partial_m F(z,t)$.

Now, using

$$\begin{aligned}
\partial_\alpha(z^\epsilon) &= Q_\alpha(\epsilon) z^{\epsilon-\overline{\alpha}}, \qquad \text{with} \qquad Q_\alpha(\epsilon) = \epsilon_\alpha, \qquad \text{we get} \\
\dot{F}_\alpha &= R^m_\alpha(t) F_m\,, \qquad \text{where} \\
R^m_\alpha(t) &= \sum_\epsilon (Q_\alpha(\epsilon+\overline{m}) - Q_\alpha(\overline{m})) M^m_\epsilon(t),
\end{aligned}$$

(here, Greek letters refer to a 1D index running from 1 to n, while Latin letters refer to a tensor index $m = \alpha_1\alpha_2\alpha_3...$). Similarly, we have

$$\begin{aligned}
\dot{F}_{\alpha\beta} &= R^m_{\alpha\beta}(t) F_m + R^m_\alpha(t) F_{m\beta} + R^m_\beta(t) F_{m\alpha}\,, \qquad \text{with} \\
R^m_{\alpha\beta}(t) &= \sum_\epsilon (Q_{\alpha\beta}(\epsilon+\overline{m}) - Q_{\alpha\beta}(\overline{m})) M^m_\epsilon(t)\,.
\end{aligned} \tag{2.638}$$

The general equations for the F_m variables are obtained by applying the operator $\partial_{\alpha_1\ldots\alpha_n}$ to (2.637). We write the action of this operator on a product of two functions as

$$\partial_{\alpha_1\ldots\alpha_n}(fg) = \left\{\partial_{\alpha_1\ldots\alpha_k} f\, \partial_{\alpha_{k+1}\ldots\alpha_n} g\right\}_\alpha\,,$$

where the braces indicate the summation for all pairs of disjoint sets $(\alpha_1\ldots\alpha_k)$, $(\alpha_{k+1}\ldots\alpha_n)$ that form a partition of the tensor index $\alpha_1\ldots\alpha_n$. When listing all possible partitions, we take care that a permutation of the elements of a set does not change said set. Also R^m with an empty index set is zero, because $Q(\epsilon) = 1$, which comes from $z^\epsilon = Q(\epsilon) z^\epsilon$.

Then the evolution equation for the time variation of $F_m(t)$ is [AL05]

$$\dot{F}_{\alpha_1\ldots\alpha_n} = \left\{R^m_{\alpha_1\ldots\alpha_k}(t) F_{m\alpha_{k+1}\ldots\alpha_n}\right\}_\alpha\,, \qquad \text{(summing over } m\text{)} \tag{2.639}$$

The tensor $R^m_{\alpha_1\ldots\alpha_k}(t)$ is given by (2.638) with $\alpha_1\ldots\alpha_k$ instead of the index $\alpha\beta$. To see the structure of the equation (2.639) we specialize it for $n = 3$,

$$\begin{aligned}
\dot{F}_{\alpha_1\alpha_2\alpha_3} &= R^m_{\alpha_1\alpha_2\alpha_3}(t) F_m \\
&+ R^m_{\alpha_1\alpha_2}(t) F_{m\alpha_3} + R^m_{\alpha_1\alpha_3}(t) F_{m\alpha_2} + R^m_{\alpha_2\alpha_3}(t) F_{m\alpha_1} \\
&+ R^m_{\alpha_1}(t) F_{m\alpha_2\alpha_3} + R^m_{\alpha_2}(t) F_{m\alpha_1\alpha_3} + R^m_{\alpha_3}(t) F_{m\alpha_1\alpha_2}.
\end{aligned} \tag{2.640}$$

The equation (2.639) has a similar structure to a *bilinear MIMO–system* [Isi89]. Recall that a bilinear system is represented by a time evolution equation that is linear in state, linear in control, but not jointly linear in both,

$$\dot{x}(t) = A(t)x(t) + \sum_{k=1}^{n} N_k(t) u_k(t) x(t) + B(t)u(t), \tag{2.641}$$

where $x \in \mathbb{R}^n$ and A, N_k, B are appropriate matrices, while the controls are u_k, $(k = 1\ldots n)$ and the state is described by x [Moh91]. Obviously, system (2.641) is a special case of the general nonlinear MIMO–system (2.600).

Using the equation (2.639), several gene regulatory applications have been studied in [AL05], including (i) design of a logic pulse; (ii) a molecular *Michaelis–Menten amplifier*; and (iii) the interference of three signal generators in E2F1 regulatory element. Also, the *Young tableaux* formalism from the Lie group theory is used to classify the symmetries of (2.639) and to keep its bilinear structure, as follows.

The change of variable $F(z,t) = e^{X(z,t)}$ is similar with the change from moments to cumulants [McCul87]. As $F(z,t)$ generates the factorial moments, $X(z,t)$ generates the factorial cumulants. In this way, the time–dependent variables, F_m, are replaced by $X_m = \partial_m X(z,t) \mid_{z=1}$. The transformation relations between F_m and X_m follow from the *Faà Di Bruno's formula* for the derivative of the composition of functions [CS96]. To keep the bilinear structure, we need to construct an appropriate index notation for the terms of the Faà Di Bruno's formula. For example, the fourth–order derivative of F at $z = 1$ is

$$\begin{aligned} F_{\alpha\beta\gamma\delta} = {} & X_{\alpha\beta\gamma\delta} + X_{\alpha\beta\gamma}X_{\delta} + X_{\alpha\beta\delta}X_{\gamma} + X_{\alpha\gamma\delta}X_{\beta} + X_{\beta\gamma\delta}X_{\alpha} \\ & + X_{\alpha\beta}X_{\gamma\delta} + X_{\alpha\gamma}X_{\beta\delta} + X_{\alpha\delta}X_{\beta\gamma} + X_{\alpha\beta}X_{\gamma}X_{\delta} + X_{\alpha\gamma}X_{\beta}X_{\delta} \\ & + X_{\alpha\delta}X_{\beta}X_{\gamma} + X_{\beta\gamma}X_{\alpha}X_{\delta} + X_{\beta\delta}X_{\alpha}X_{\gamma} + X_{\gamma\delta}X_{\alpha}X_{\beta} + X_{\alpha}X_{\beta}X_{\gamma}X_{\delta} \end{aligned} \tag{2.642}$$

Given the index $\alpha\beta\gamma\delta$ from the l.h.s of (2.642), we need to generate all the indices that appear in its the r.h.s. If the term $X_{\alpha\gamma}X_{\beta}X_{\delta}$ is present in the sum, then any symmetric version of it, like $X_{\beta}X_{\gamma\alpha}X_{\delta}$, cannot be present. If we classify all possible symmetries of a term, then we find an index notation that will eliminate all the equivalent terms. The symmetries come from the commutativities of the product and the partial derivatives.

If we use the corresponding Young tableaux filled with indices $\alpha, \beta, \gamma, ...$, to represent products of X_m variables,

$$\mathrm{X}\begin{array}{|c|c|c|}\hline \alpha & \beta & \gamma \\ \hline \delta \\ \cline{1-1}\end{array} = X_{\alpha\beta\gamma}X_{\delta},$$

then (2.642) becomes (see [AL05]):

$$\begin{aligned} \mathrm{F}_{\begin{array}{|c|c|c|c|}\hline \alpha & \beta & \gamma & \delta \\ \hline \end{array}} = {} & \mathrm{X}_{\begin{array}{|c|c|c|c|}\hline \alpha & \beta & \gamma & \delta \\ \hline \end{array}} \\ & + \mathrm{X}_{\begin{array}{|c|c|c|}\hline \alpha & \beta & \gamma \\ \hline \delta \\ \cline{1-1}\end{array}} + \mathrm{X}_{\begin{array}{|c|c|c|}\hline \alpha & \beta & \delta \\ \hline \gamma \\ \cline{1-1}\end{array}} + \mathrm{X}_{\begin{array}{|c|c|c|}\hline \alpha & \gamma & \delta \\ \hline \beta \\ \cline{1-1}\end{array}} + \mathrm{X}_{\begin{array}{|c|c|c|}\hline \beta & \gamma & \delta \\ \hline \alpha \\ \cline{1-1}\end{array}} \\ & + \mathrm{X}_{\begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \gamma & \delta \\ \hline\end{array}} + \mathrm{X}_{\begin{array}{|c|c|}\hline \alpha & \gamma \\ \hline \beta & \delta \\ \hline\end{array}} + X_{\begin{array}{|c|c|}\hline \alpha & \delta \\ \hline \beta & \gamma \\ \hline\end{array}} \\ & + \mathrm{X}_{\begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \gamma \\ \cline{1-1} \delta \\ \cline{1-1}\end{array}} + \mathrm{X}_{\begin{array}{|c|c|}\hline \alpha & \gamma \\ \hline \beta \\ \cline{1-1} \delta \\ \cline{1-1}\end{array}} + \mathrm{X}_{\begin{array}{|c|c|}\hline \alpha & \delta \\ \hline \beta \\ \cline{1-1} \gamma \\ \cline{1-1}\end{array}} + \mathrm{X}_{\begin{array}{|c|c|}\hline \beta & \gamma \\ \hline \alpha \\ \cline{1-1} \delta \\ \cline{1-1}\end{array}} + \mathrm{X}_{\begin{array}{|c|c|}\hline \beta & \delta \\ \hline \alpha \\ \cline{1-1} \gamma \\ \cline{1-1}\end{array}} + \mathrm{X}_{\begin{array}{|c|c|}\hline \gamma & \delta \\ \hline \alpha \\ \cline{1-1} \beta \\ \cline{1-1}\end{array}} \\ & + \mathrm{X}_{\begin{array}{|c|}\hline \alpha \\ \hline \beta \\ \hline \gamma \\ \hline \delta \\ \hline\end{array}} \end{aligned}$$

The rows of a Young tableau are listed in decreasing order of their length, which will enforce an order in the product of the variables X_m. Thus the symmetry due to the commutativity of the product is lifted.

For example, if $Y=$  , $m = \alpha\beta\gamma\delta$, $\sigma(1) = 1, \sigma(2) = 2, \sigma(3) = 4$ and $\sigma(4) = 3$ we have:

Y[m]=  $Y[m^\sigma]=$ 

and $X_{Y[m]} = X_{\alpha\beta\gamma}X_\delta \qquad X_{Y[m^\sigma]} = X_{\alpha\beta\delta}X_\gamma.$ (2.643)

The terms in the Faà Di Bruno formula are generated using filled Young tableaux and a set of representative permutations σ_i, $(i = 1 \dots J)$, where J is $|Y|!/\mathrm{card}(H^Y)$, chosen form each set of the coset space,

$$S_{|Y|}/H^Y = \{\sigma_1 H^Y, \dots, \sigma_J H^Y\},$$

so that the Faà Di Bruno formula in Young tableaux notation reads

$$F_m = \sum_{|Y|=|m|} \sum_{\sigma \in S_{|Y|}/H^Y} X_{Y[m^\sigma]}.$$

For partial derivatives with z not fixed to 1 the formula is similar

$$\partial_m F(z,t) = \sum_{|Y|=|m|} \sum_{\sigma \in S_{|Y|}/H^Y} \partial_{Y[m^\sigma]} X(z,t) e^{X(z,t)},$$

with the convention that the derivation, with respect to a filled Young tableau, is the product of the derivatives along each line of the tableau. Thus, using the example (2.643) we have

$$\partial_{Y[m]} X(z,t) = \partial_\alpha \partial_\beta \partial_\gamma X(z,t) \partial_\delta X(z,t)\,.$$

Now, the equation (2.637) in the variable $X(z,t)$ becomes

$$\partial_t X(z,t) = \sum_{m,\epsilon} (z^{\epsilon+\overline{m}} - z^{\overline{m}}) M_\epsilon^m(t) [\sum_{|Y|=|m|} \sum_{\sigma \in S_{|Y|}/H^Y} \partial_{Y[m^\sigma]} X(z,t)]. \tag{2.644}$$

The time–dependent variables will now be $\partial_m X(z,t) \mid_{z=1}$, so that we need to take partial derivatives with respect to z of (2.644). The concatenation notation $\partial_{\alpha\beta} = \partial_\alpha \partial_\beta$ has to be generalized for the filled Young tableaux

$$\partial_{\alpha|Y[m]} X(z,t) = \partial_\alpha (\partial_{Y[m]} X(z,t)).$$

Therefore, the concatenation $\alpha|Y[m]$ means that a box containing α must be glued to each row of $Y[m]$ and the object thus obtained must be rearranged into a lexicographical order filled Young tableau. Inductively we can define

$$\partial_{\alpha|\beta|\dots\gamma|Y[m]} = \partial_\alpha (\partial_{\beta|\dots|\gamma|Y[m]}).$$

The concatenation notation is also applied to the X_m variable,

$$X_{\alpha|Y[m]} = \partial_{\alpha|Y[m]} X(z,t) \mid_{z=1} .$$

The equations for the factorial cumulants are now, as a consequence of (2.644), given by

$$\dot{X}_\alpha = R^m_\alpha(t) \sum_{Y,\sigma} X_{Y[m^\sigma]},$$

$$\dot{X}_{\alpha\beta} = R^m_{\alpha\beta}(t) \sum_{Y,\sigma} X_{Y[m^\sigma]} + R^m_\alpha(t) \sum_{Y,\sigma} X_{\beta|Y[m^\sigma]} + R^m_\beta(t) \sum_{Y,\sigma} X_{\alpha|Y[m^\sigma]},$$

with $\sum_{Y,\sigma}$ being a short notation for the the sums over Y and σ in (2.644). In general, for polynomial transition probabilities and factorial cumulants, we get the time evolution equation [AL05]:

$$\dot{X}_{\alpha_1\alpha_2\ldots\alpha_n} = \left\{ R^m_{\alpha_1\alpha_2\ldots\alpha_k}(t) \sum_{Y,\sigma} X_{\alpha_{k+1}|\ldots|\alpha_n|Y[m^\sigma]} \right\}_\alpha .$$

2.4 Human–Like Biomechanics

Recall from [II05] that modern unified geometrical basis for both human biomechanics and humanoid robotics represents the *constrained* $SE(3)-$*group*, i.e., the so–called *special Euclidean group of rigid–body motions in 3D space* (see, e.g., [MLS94, PC05]). In other words, during human movement, in each movable human joint there is an action of a constrained $SE(3)-$group. Therefore, constrained $SE(3)-$group represents *general kinematics* of human–like joints. The corresponding nonlinear dynamics problem (resolved mainly for aircraft and spacecraft dynamics) is called the *dynamics on* $SE(3)-$*group*, while the associated nonlinear control problem (resolved mainly for general helicopter control) is called the *control on* $SE(3)-$*group*.

Recall that the Euclidean $SE(3)-$group is defined as a semidirect (noncommutative) product of 3D rotations and 3D translations, $SE(3) := SO(3) \triangleright \mathbb{R}^3$ [MLS94, PC05, II05]). Its most important subgroups are the following:

Subgroup	*Definition*
$SO(3)$, group of rotations in 3D (a spherical joint)	Set of all proper orthogonal $3 \times 3-$ rotational matrices
$SE(2)$, special Euclidean group in 2D (all planar motions)	Set of all $3 \times 3-$ matrices: $\begin{bmatrix} \cos\theta & \sin\theta & r_x \\ -\sin\theta & \cos\theta & r_y \\ 0 & 0 & 1 \end{bmatrix}$
$SO(2)$, group of rotations in 2D subgroup of $SE(2)-$group (a revolute joint)	Set of all proper orthogonal $2 \times 2-$ rotational matrices included in $SE(2)-$group
$\mathbb{R}^3$, group of translations in 3D (all spatial displacements)	Euclidean 3D vector space

In the next subsection we give detailed analysis of these subgroups, as well as the total $SE(3)-$group.

2.4.1 Lie Groups and Symmetries in Biomechanics

Lie Groups of Joint Rotations

Local kinematics at each rotational robot or (synovial) human joint, is defined as a *group action* of an nD constrained rotational Lie group $SO(n)$ on the Euclidean space $\mathbb{R}^n$. In particular, there is an action of $SO(2)$–group in uniaxial human joints (cylindrical, or *hinge joints*, like knee and elbow) and an action of $SO(3)$–group in three–axial human joints (spherical, or *ball–and–socket joints*, like hip, shoulder, neck, wrist and ankle). In both cases, $SO(n)$ acts, with its operators of rotation, on the vector $x = \{x^\mu\}$, $(i = 1, 2, 3)$ of external, Cartesian coordinates of the parent body–segment, depending, at the same time, on the vector $q = \{q^s\}$, $(s = 1, \cdots, n)$ on n group–parameters, i.e., joint angles.

Each joint rotation $R \in SO(n)$ defines a map

$$R : x^\mu \mapsto \dot{x}^\mu, \qquad R(x^\mu, q^s) = R_{q^s} x^\mu,$$

where $R_{q^s} \in SO(n)$ are joint group operators. The vector $v = \{v_s\}$, $(s = 1, \cdots, n)$ of n infinitesimal generators of these rotations, i.e., joint angular velocities, given by

$$v_s = -[\frac{\partial R(x^\mu, q^s)}{\partial q^s}]_{q=0} \frac{\partial}{\partial x^\mu},$$

constitute an nD Lie algebra $\mathfrak{so}(n)$ corresponding to the joint rotation group $SO(n)$. Conversely, each joint group operator R_{q^s}, representing a one–parameter subgroup of $SO(n)$, is defined as the exponential map of the corresponding joint group generator v_s

$$R_{q^s} = \exp(q^s v_s). \tag{2.645}$$

The exponential map (2.645) represents a solution of the joint operator differential equation in the joint group–parameter space $\{q^s\}$

$$\frac{dR_{q^s}}{dq^s} = v_s R_{q^s}.$$

Uniaxial Group of Joint Rotations

The uniaxial joint rotation in a single Cartesian plane around a perpendicular axis, e.g., $xy-$plane about the z axis, by an internal joint angle θ, leads to the following transformation of the joint coordinates

$$\dot{x} = x\cos\theta - y\sin\theta, \qquad \dot{y} = x\sin\theta + y\cos\theta.$$

In this way, the joint $SO(2)-$group, given by

$$SO(2) = \left\{ R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} | \theta \in [0, 2\pi] \right\},$$

acts in a canonical way on the Euclidean plane $\mathbb{R}^2$ by

$$SO(2) = \left(\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \begin{pmatrix} x \\ w \end{pmatrix} \right) \longmapsto \begin{pmatrix} x\cos\theta & -y\sin\theta \\ x\sin\theta & y\cos\theta \end{pmatrix}.$$

Its associated Lie algebra $\mathfrak{so}(2)$ is given by

$$\mathfrak{so}(2) = \left\{ \begin{pmatrix} 0 & -t \\ t & 0 \end{pmatrix} | t \in \mathbb{R} \right\},$$

since the curve $\gamma_\theta \in SO(2)$ given by

$$\gamma_\theta : t \in \mathbb{R} \longmapsto \gamma_\theta(t) = \begin{pmatrix} \cos t\theta & -\sin t\theta \\ \sin t\theta & \cos t\theta \end{pmatrix} \in SO(2),$$

passes through the identity $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and then

$$\left.\frac{d}{dt}\right|_{t=0} \gamma_\theta(t) = \begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix},$$

so that I_2 is a basis of $\mathfrak{so}(2)$, since $\dim(SO(2)) = 1$.

The *exponential map* $\exp : \mathfrak{so}(2) \to SO(2)$ is given by

$$\exp \begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix} = \gamma_\theta(1) = \begin{pmatrix} \cos t\theta & -\sin t\theta \\ \sin t\theta & \cos t\theta \end{pmatrix}.$$

The *infinitesimal generator* of the action of $SO(2)$ on $\mathbb{R}^2$, i.e., joint angular velocity v, is given by

$$v = -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y},$$

since

$$v_{\mathbb{R}^2}(x,y) = \left.\frac{d}{dt}\right|_{t=0} \exp(tv)(x,y) = \left.\frac{d}{dt}\right|_{t=0} \begin{pmatrix} \cos tv & -\sin tv \\ \sin tv & \cos tv \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

The *momentum map* (see subsection 1.2.11 below) $J : T^*\mathbb{R}^2 \to \mathbb{R}$ associated to the lifted action of $SO(2)$ on $T^*\mathbb{R}^2 \simeq \mathbb{R}^4$ is given by

$$\begin{aligned} J(x,y,p_1,p_2) &= xp_y - yp_x, \qquad \text{since} \\ J(x,y,p_x,p_y)(\xi) &= (p_x dx + p_y dy)(v_{\mathbb{R}^2}) = -vp_x y + -vp_y x. \end{aligned}$$

The Lie group $SO(2)$ acts on the symplectic manifold $(\mathbb{R}^4, \omega = dp_x \wedge dx + dp_y \wedge dx)$ by

$$\begin{aligned} &\phi\left(\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, (x,y,p_x,p_y)\right) \\ &= (x\cos\theta - y\sin\theta,\ x\sin\theta + y\cos\theta,\ p_x\cos\theta - p_y\sin\theta,\ p_x\sin\theta + p_y\cos\theta). \end{aligned}$$

Three–Axial Group of Joint Rotations

The three–axial $SO(3)-$group of human–like joint rotations depends on three parameters, Euler joint angles $q^i = (\varphi, \psi, \theta)$, defining the rotations about the Cartesian coordinate triedar (x, y, z) placed at the joint pivot point. Each of the Euler angles are defined in the constrained range $(-\pi, \pi)$, so the joint group space is a constrained sphere of radius π.

Let $G = SO(3) = \{A \in \mathcal{M}_{3\times 3}(\mathbb{R}) : A^t A = I_3, \det(A) = 1\}$ be the group of rotations in $\mathbb{R}^3$. It is a Lie group and $\dim(G) = 3$. Let us isolate its one–parameter joint subgroups, i.e., consider the three operators of the finite joint rotations $R_\varphi, R_\psi, R_\theta \in SO(3)$, given by

$$R_\varphi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{bmatrix}, \quad R_\psi = \begin{bmatrix} \cos\psi & 0 & \sin\psi \\ 0 & 1 & 0 \\ -\sin\psi & 0 & \cos\psi \end{bmatrix}, \quad R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

corresponding respectively to rotations about $x-$axis by an angle φ, about $y-$axis by an angle ψ, and about $z-$axis by an angle θ.

The total three–axial joint rotation A is defined as the product of above one–parameter rotations $R_\varphi, R_\psi, R_\theta$, i.e., $A = R_\varphi \cdot R_\psi \cdot R_\theta$ is equal

$$A = \begin{bmatrix} \cos\psi\cos\varphi - \cos\theta\sin\varphi\sin\psi & \cos\psi\cos\varphi + \cos\theta\cos\varphi\sin\psi & \sin\theta\sin\psi \\ -\sin\psi\cos\varphi - \cos\theta\sin\varphi\sin\psi & -\sin\psi\sin\varphi + \cos\theta\cos\varphi\cos\psi & \sin\theta\cos\psi \\ \sin\theta\sin\varphi & -\sin\theta\cos\varphi & \cos\theta \end{bmatrix}.$$

However, the order of these matrix products matters: different order products give different results, as the matrix product is *noncommutative product*. This

is the reason why Hamilton's *quaternions*[33] are today commonly used to parameterize the $SO(3)-$group, especially in the field of 3D computer graphics.

The one–parameter rotations $R_\varphi, R_\psi, R_\theta$ define curves in $SO(3)$ starting from $I_3 = \begin{pmatrix} 1\,0\,0 \\ 0\,1\,0 \\ 0\,0\,1 \end{pmatrix}$. Their derivatives in $\varphi = 0, \psi = 0$ and $\theta = 0$ belong to the associated *tangent Lie algebra* $\mathfrak{so}(3)$. That is the corresponding infinitesimal generators of joint rotations – joint angular velocities $v_\varphi, v_\psi, v_\theta \in \mathfrak{so}(3)$ – are respectively given by

$$v_\varphi = \begin{bmatrix} 0\,0\,0 \\ 0\,0\,-1 \\ 0\,1\,0 \end{bmatrix} = -y\frac{\partial}{\partial z} + z\frac{\partial}{\partial y}, \qquad v_\psi = \begin{bmatrix} 0\,0\,1 \\ 0\,0\,0 \\ -1\,0\,0 \end{bmatrix} = -z\frac{\partial}{\partial x} + x\frac{\partial}{\partial z},$$

$$v_\theta = \begin{bmatrix} 0\,-1\,0 \\ 1\,1\,0 \\ 0\,0\,0 \end{bmatrix} = -x\frac{\partial}{\partial y} + y\frac{\partial}{\partial x}.$$

Moreover, the elements are linearly independent and so

$$\mathfrak{so}(3) = \left\{ \begin{bmatrix} 0 & -a & b \\ a & 0 & -\gamma \\ -b & \gamma & 0 \end{bmatrix} | a, b, \gamma \in \mathbb{R} \right\}.$$

The Lie algebra $\mathfrak{so}(3)$ is identified with $\mathbb{R}^3$ by associating to each $v = (v_\varphi, v_\psi, v_\theta) \in \mathbb{R}^3$ the matrix $v \in \mathfrak{so}(3)$ given by $v = \begin{bmatrix} 0 & -a & b \\ a & 0 & -\gamma \\ -b & \gamma & 0 \end{bmatrix}$. Then we have the following identities:

1. $\widehat{u \times v} = [\hat{u}, v]$; and
2. $u \cdot v = -\frac{1}{2}\,\mathrm{Tr}(\hat{u} \cdot v)$.

The exponential map $\exp : \mathfrak{so}(3) \to SO(3)$ is given by *Rodrigues relation*

$$\exp(v) = I + \frac{\sin\|v\|}{\|v\|}v + \frac{1}{2}\left(\frac{\sin\frac{\|v\|}{2}}{\frac{\|v\|}{2}}\right)^2 v^2,$$

where the norm $\|v\|$ is given by

$$\|v\| = \sqrt{(v^1)^2 + (v^2)^2 + (v^3)^2}.$$

[33] Recall that the set of Hamilton's *quaternions* $\mathbb{H}$ represents an extension of the set of complex numbers $\mathbb{C}$. We can compute a rotation about the unit vector, $\mathbf{u}$ by an angle θ. The quaternion q that computes this rotation is

$$q = \left(\cos\frac{\theta}{2}\ ,\ u\sin\frac{\theta}{2}\right).$$

The the dual, *cotangent Lie algebra* $\mathfrak{so}(3)^*$, includes the three joint angular momenta $p_\varphi, p_\psi, p_\theta \in \mathfrak{so}(3)^*$, derived from the joint velocities v by multiplying them with corresponding moments of inertia.

Note that the parameterization of $SO(3)-$rotations is the subject of continuous research and development in many theoretical and applied fields of mechanics, such as rigid body, structural, and multibody dynamics, robotics, spacecraft attitude dynamics, navigation, image processing, etc. For a complete discussion on the classical attitude representations see [Fri88, Mla91, Shu93, STP95]. In addition, a modern vectorial parameterization of finite rotations, encompassing the mentioned earlier developments as well as Gibbs, Wiener, and Milenkovic parameterizations [Mla99, BT03].

Euclidean Groups of Total Joint Motions

Biomechanically realistic joint movement is predominantly rotational, plus restricted translational (translational motion in human joints is observed after reaching the limit of rotational amplitude). Gross translation in any human joint means joint dislocation, which is a severe injury. Obvious models for uniaxial and triaxial joint motions are *special Euclidean groups of rigid body motions*, $SE(2)$ and $SE(3)$, respectively.

Special Euclidean Group in the Plane

The motion in uniaxial human joints is naturally modelled by the *special Euclidean group in the plane*, $SE(2)$. It consists of all transformations of $\mathbb{R}^2$ of the form $Az + a$, where $z, a \in \mathbb{R}^2$, and

$$A \in SO(2) = \left\{ \text{matrices of the form } \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \right\}.$$

In other words [MR99], group $SE(2)$ consists of matrices of the form: $(R_\theta, a) = \begin{pmatrix} R_\theta & a \\ 0 & I \end{pmatrix}$, where $a \in \mathbb{R}^2$ and R_θ is the rotation matrix: $R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$, while I is the 3×3 identity matrix. The inverse $(R_\theta, a)^{-1}$ is given by

$$(R_\theta, a)^{-1} = \begin{pmatrix} R_\theta & a \\ 0 & I \end{pmatrix}^{-1} = \begin{pmatrix} R_{-\theta} & -R_{-\theta}a \\ 0 & I \end{pmatrix}.$$

The Lie algebra $\mathfrak{se}(2)$ of $SE(2)$ consists of 3×3 block matrices of the form

$$\begin{pmatrix} -\xi J & v \\ 0 & 0 \end{pmatrix}, \qquad \text{where} \qquad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad (J^T = J^{-1} = -J),$$

with the usual commutator bracket. If we identify $\mathfrak{se}(2)$ with $\mathbb{R}^3$ by the isomorphism

$$\begin{pmatrix} -\xi J & v \\ 0 & 0 \end{pmatrix} \in \mathfrak{se}(2) \longmapsto (\xi, v) \in \mathbb{R}^3,$$

then the expression for the Lie algebra bracket becomes

$$[(\xi, v_1, v_2), (\zeta, w_1, w_2)] = (0, \zeta v_2 - \xi w_2, \xi w_1 - \zeta v1) = (0, \xi J^T w - \zeta J^T v),$$

where $v = (v_1, v_2)$ and $w = (w_1, w_2)$.

The *adjoint group action* of

$$(R_\theta, a) \begin{pmatrix} R_\theta & a \\ 0 & I \end{pmatrix} \qquad \text{on} \qquad (\xi, v) = \begin{pmatrix} -\xi J & v \\ 0 & 0 \end{pmatrix}$$

is given by the *group conjugation*,

$$\begin{pmatrix} R_\theta & a \\ 0 & I \end{pmatrix} \begin{pmatrix} -\xi J & v \\ 0 & 0 \end{pmatrix} \begin{pmatrix} R_{-\theta} & -R_{-\theta} a \\ 0 & I \end{pmatrix} = \begin{pmatrix} -\xi J & \xi J a + R_\theta v \\ 0 & 0 \end{pmatrix},$$

or, in coordinates [MR99],

$$Ad_{(R_\theta, a)}(\xi, v) = (\xi, \xi J a + R_\theta v). \tag{2.646}$$

In proving (2.646) we used the identity $R_\theta J = J R_\theta$. Identify the dual algebra, $\mathfrak{se}(2)^*$, with matrices of the form $\begin{pmatrix} \frac{\mu}{2} J & 0 \\ \alpha & 0 \end{pmatrix}$, via the nondegenerate pairing given by the trace of the product. Thus, $\mathfrak{se}(2)^*$ is isomorphic to $\mathbb{R}^3$ via

$$\begin{pmatrix} \frac{\mu}{2} J & 0 \\ \alpha & 0 \end{pmatrix} \in \mathfrak{se}(2)^* \longmapsto (\mu, \alpha) \in \mathbb{R}^3,$$

so that in these coordinates, the pairing between $\mathfrak{se}(2)^*$ and $\mathfrak{se}(2)$ becomes

$$\langle (\mu, \alpha), (\xi, v) \rangle = \mu \xi + \alpha \cdot v,$$

that is, the usual dot product in $\mathbb{R}^3$. The *coadjoint group action* is thus given by

$$Ad^*_{(R_\theta, a)^{-1}}(\mu, \alpha) = (\mu - R_\theta \alpha \cdot J a + R_\theta \alpha). \tag{2.647}$$

Formula (2.647) shows that the coadjoint orbits are the cylinders $T^* S^1_\alpha = \{(\mu, \alpha) | \, \|\alpha\| = \text{const}\}$ if $\alpha \neq 0$ together with the points are on the $\mu-$axis. The canonical cotangent bundle projection $\pi : T^* S^1_\alpha \to S^1_\alpha$ is defined as $\pi(\mu, \alpha) = \alpha$.

Special Euclidean Group in the 3D Space

The most common group structure in human–like biomechanics is the *special Euclidean group in 3D space*, $SE(3)$. It is defined as a semidirect (noncommutative) product of 3D rotations and 3D translations, $SO(3) \triangleright \mathbb{R}^3$.

The Heavy Top

As a starting point consider a rigid body (see (2.1.1) below) moving with a fixed point but under the influence of gravity. This problem still has a configuration space $SO(3)$, but the symmetry group is only the circle group S^1, consisting of rotations about the direction of gravity. One says that gravity has broken the symmetry from $SO(3)$ to S^1. This time, eliminating the S^1 symmetry mysteriously leads one to the larger Euclidean group $SE(3)$ of rigid motion of $\mathbb{R}^3$. Conversely, we can start with $SE(3)$ as the configuration space for the rigid–body and 'reduce out' translations to arrive at $SO(3)$ as the configuration space (see [MR99]).

The equations of motion for a rigid body with a fixed point in a gravitational field give an interesting example of a system that is Hamiltonian (see (2.1.1)) relative to a *Lie–Poisson bracket* (see (2.4.2)). The underlying Lie algebra consists of the algebra of infinitesimal Euclidean motions in $\mathbb{R}^3$.

The basic phase–space we start with is again $T^*SO(3)$, parameterized by Euler angles and their conjugate momenta. In these variables, the equations are in canonical Hamiltonian form. However, the presence of gravity breaks the symmetry, and the system is no longer $SO(3)$ invariant, so it cannot be written entirely in terms of the body angular momentum p. One also needs to keep track of Γ, the 'direction of gravity' as seen from the body. This is defined by $\Gamma = A^{-1}k$, where k points upward and A is the element of $SO(3)$ describing the current configuration of the body. The equations of motion are

$$\begin{aligned}
\dot{p}_1 &= \frac{I_2 - I_3}{I_2 I_3} p_2 p_3 + Mgl(\Gamma^2 \chi^3 - \Gamma^3 \chi^2),\\
\dot{p}_2 &= \frac{I_3 - I_1}{I_3 I_1} p_3 p_1 + Mgl(\Gamma^3 \chi^1 - \Gamma^1 \chi^3),\\
\dot{p}_3 &= \frac{I_1 - I_2}{I_1 I_2} p_1 p_2 + Mgl(\Gamma^1 \chi^2 - \Gamma^2 \chi^1),\\
&\text{and} \qquad \dot{\Gamma} = \Gamma \times \Omega,
\end{aligned}$$

where Ω is the body's angular velocity vector, I_1, I_2, I_3 are the body's principal moments of inertia, M is the body's mass, g is the acceleration of gravity, χ is the body fixed unit vector on the line segment connecting the fixed point with the body's center of mass, and l is the length of this segment.

The Euclidean Group and Its Lie Algebra

An element of $SE(3)$ is a pair (A, a) where $A \in SO(3)$ and $a \in \mathbb{R}^3$. The action of $SE(3)$ on $\mathbb{R}^3$ is the rotation A followed by translation by the vector a and has the expression

$$(A, a) \cdot x = Ax + a.$$

Using this formula, one sees that multiplication and inversion in $SE(3)$ are given by

$$(A,a)(B,b) = (AB, Ab + a) \qquad \text{and} \qquad (A,a)^{-1} = (A^{-1}, -A^{-1}a),$$

for $A, B \in SO(3)$ and $a, b \in \mathbb{R}^3$. The identity element is $(l, 0)$.

The Lie algebra of the Euclidean group $SE(3)$ is $\mathfrak{se}(3) = \mathbb{R}^3 \times \mathbb{R}^3$ with the Lie bracket

$$[(\xi, u), (\eta, v)] = (\xi \times \eta, \xi \times v - \eta \times u). \tag{2.648}$$

The Lie algebra of the Euclidean group has a structure that is a special case of what is called a *semidirect product*. Here it is the *product of the group of rotations with the corresponding group of translations*. It turns out that semidirect products occur under rather general circumstances when the symmetry in T^*G is broken.

The dual Lie algebra of the Euclidean group $SE(3)$ is $\mathfrak{se}(3)^* = \mathbb{R}^3 \times \mathbb{R}^3$ with the same Lie bracket (2.648). For the further details on adjoint orbits in $\mathfrak{se}(3)$ as well as coadjoint orbits in $\mathfrak{se}(3)^*$ see [MR99].

Symplectic Group in Hamiltonian Mechanics

Let $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$, with I the $n \times n$ identity matrix. Now, $A \in L(\mathbb{R}^{2n}, \mathbb{R}^{2n})$ is called a *symplectic matrix* if $A^T J\, A = J$. Let $Sp(2n, \mathbb{R})$ be the set of $2n \times 2n$ symplectic matrices. Taking determinants of the condition $A^T J\, A = J$ gives $\det A = \pm 1$, and so $A \in GL(2n, \mathbb{R})$. Furthermore, if $A, B \in Sp(2n, R)$, then $(AB)^T J(AB) = B^T A^T JAB = J$. Hence, $AB \in Sp(2n, \mathbb{R})$, and if $A^T J\, A = J$, then $JA = (A^T)^{-1} J = (A^{-1})^T J$, so $J = (A-1)^T\, JA^{-1}$, or $A^{-1} \in Sp(2n, \mathbb{R})$. Thus, $Sp(2n, \mathbb{R})$ is a group [MR99].

The *symplectic Lie group*

$$Sp(2n, \mathbb{R}) = \left\{ A \in GL(2n, \mathbb{R}) : A^T J\, A = J \right\}$$

is a noncompact, connected Lie group of dimension $2n^2 + n$. Its Lie algebra

$$\mathfrak{sp}(2n, \mathbb{R}) = \left\{ A \in L(\mathbb{R}^{2n}, \mathbb{R}^{2n}) : A^T J\, A = J = 0 \right\},$$

called the *symplectic Lie algebra*, consists of the $2n \times 2n$ matrices A satisfying $A^T J\, A = 0$ [MR99].

Consider a particle of mass m moving in a potential $V(q)$, where $q^i = (q^1, q^2, q^3) \in \mathbb{R}^3$. Newtonian second law states that the particle moves along a curve $q(t)$ in $\mathbb{R}^3$ in such a way that $m\ddot{q}^i = -\operatorname{grad} V(q^i)$. Introduce the momentum $p_i = m\dot{q}^i$, and the energy

$$H(q,p) = \frac{1}{2m} \sum_{i=1}^{3} p_i^2 + V(q).$$

Then

$$\frac{\partial H}{\partial q^i} = \frac{\partial V}{\partial q^i} = -m\ddot{q}^i = -\dot{p}_i, \quad \text{and}$$
$$\frac{\partial H}{\partial p_i} = \frac{1}{m} p_i = \dot{q}^i, \qquad (i = 1, 2, 3),$$

and hence Newtonian law $F = m\ddot{q}^i$ is equivalent to Hamiltonian equations

$$\dot{q}^i = \frac{\partial H}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial H}{\partial q^i}.$$

Now, writing $z = (q^i, p_i)$ [MR99],

$$J \operatorname{grad} H(z) = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} \frac{\partial H}{\partial q^i} \\ \frac{\partial H}{\partial p_i} \end{pmatrix} = \left(\dot{q}^i, \dot{p}_i\right) = \dot{z},$$

so Hamiltonian equations read

$$\dot{z} = J \operatorname{grad} H(z). \tag{2.649}$$

Now let $f : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3 \times \mathbb{R}^3$ and write $w = f(z)$. If $z(t)$ satisfies Hamiltonian equations (2.649) then $w(t) = f(z(t))$ satisfies $\dot{w} = A^T \dot{z}$, where $A^T = [\partial w^i / \partial z^j]$ is the Jacobian matrix of f. By the chain rule,

$$\dot{w} = A^T J \operatorname*{grad}_z H(z) = A^T J\, A \operatorname*{grad}_w H(z(w)).$$

Thus, the equations for $w(t)$ have the form of Hamiltonian equations with energy $K(w) = H(z(w))$ iff $A^T J\, A = J$, that is, iff A is symplectic. A nonlinear transformation f is canonical iff its Jacobian matrix is symplectic. $Sp(2n, \mathbb{R})$ is the linear invariance group of classical mechanics [MR99].

Group Structure of the Biomechanical Manifold M

Purely Rotational Biomechanical Manifold

Kinematics of an n–segment human–body chain (like arm, leg or spine) is usually defined as a map between *external coordinates* (usually, end–effector coordinates) x^r $(r = 1, \ldots, n)$ and *internal joint coordinates* q^i $(i = 1, \ldots, N)$ (see [IS01, Iva02, IP01b, IP01b, Iva05a]). The *forward kinematics* are defined as a nonlinear map $x^r = x^r(q^i)$ with a corresponding linear vector functions $dx^r = \partial x^r / \partial q^i \; dq^i$ of differentials: and $\dot{x}^r = \partial x^r / \partial q^i \; \dot{q}^i$ of velocities. When the rank of the configuration–dependent Jacobian matrix $J \equiv \partial x^r / \partial q^i$ is less than n the *kinematic singularities* occur; the onset of this condition could be detected by the *manipulability measure*. The *inverse kinematics* are defined conversely by a nonlinear map $q^i = q^i(x^r)$ with a corresponding linear vector functions $dq^i = \partial q^i / \partial x^r \; dx^r$ of differentials and $\dot{q}^i = \partial q^i / \partial x^r \; \dot{x}^r$ of velocities. Again, in the case of *redundancy* $(n < N)$, the inverse kinematic problem

admits infinite solutions; often the *pseudo–inverse* configuration–control is used instead: $\dot{q}^i = J^* \dot{x}^r$, where $J^* = J^T(J J^T)^{-1}$ denotes the Moore–Penrose pseudo–inverse of the Jacobian matrix J.

Humanoid joints, that is, internal coordinates q^i $(i = 1, \dots, N)$, constitute a smooth configuration manifold M, described as follows. Uniaxial, 'hinge' joints represent constrained, rotational Lie groups $SO(2)^i_{cnstr}$, parameterized by constrained angles $q^i_{cnstr} \equiv q^i \in [q^i_{\min}, q^i_{\max}]$. Three–axial, 'ball–and–socket' joints represent constrained rotational Lie groups $SO(3)^i_{cnstr}$, parameterized by constrained Euler angles $q^i = q^{\phi_i}_{cnstr}$ (in the following text, the subscript 'cnstr' will be omitted, for the sake of simplicity, and always assumed in relation to internal coordinates q^i).

All $SO(n)$–joints are Hausdorff C^∞–manifolds with atlases (U_α, u_α); in other words, they are paracompact and metrizable smooth manifolds, admitting Riemannian metric.

Let A and B be two smooth manifolds described by smooth atlases (U_α, u_α) and (V_β, v_β), respectively. Then the family $(U_\alpha \times V_\beta, u_\alpha \times v_\beta : U_\alpha \times V_\beta \to \mathbb{R}^m \times \mathbb{R}^n)(\alpha, \beta) \in A \times B$ is a smooth atlas for the direct product $A \times B$. Now, if A and B are two Lie groups (say, $SO(n)$), then their *direct product* $G = A \times B$ is at the same time their direct product as smooth manifolds and their direct product as algebraic groups, with the product law

$$(a_1, b_1)(a_2, b_2) = (a_1 a_2, b_1 b_2), \qquad (a_{1,2} \in A,\ b_{1,2} \in B).$$

Generalizing the direct product to N rotational joint groups, we can draw an *anthropomorphic product–tree* (see Figure 2.21) using a line segment '–' to represent direct products of human $SO(n)$–joints. This is our basic model of the biomechanical configuration manifold M.

Let T_qM be a tangent space to M at the point q. The *tangent bundle* TM represents a union $\cup_{q \in M} T_qM$, together with the standard topology on TM and a natural smooth manifold structure, the dimension of which is twice the dimension of M. A vector–field X on M represents a section $X : M \to TM$ of the tangent bundle TM.

Analogously let T^*_qM be a cotangent space to M at q, the dual to its tangent space T_qM. The *cotangent bundle* T^*M represents a union $\cup_{q \in M} T^*_qM$, together with the standard topology on T^*M and a natural smooth manifold structure, the dimension of which is twice the dimension of M. A $1-$form θ on M represents a section $\theta : M \to T^*M$ of the cotangent bundle T^*M.

We refer to the tangent bundle TM of biomechanical configuration manifold M as the *velocity phase–space* manifold, and to its cotangent bundle T^*M as the *momentum phase–space* manifold.

Reduction of the Rotational Biomechanical Manifold

The biomechanical configuration manifold M (Figure 2.21) can be (for the sake of the brain–like motor control) reduced to N–torus T^N, in three steps, as follows.

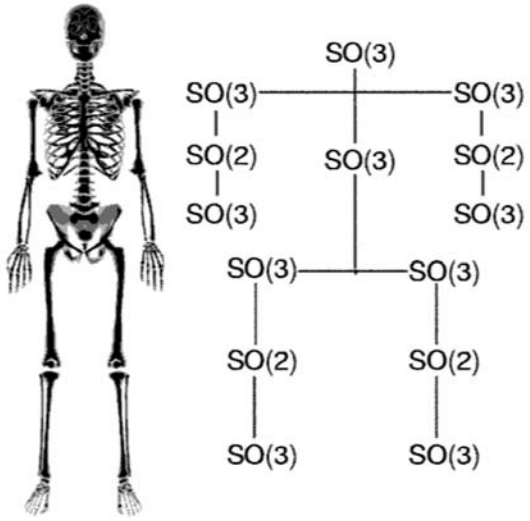


Fig. 2.21. Purely rotational, whole–body biomechanical manifold, with a single $SO(3)-$joint representing the whole spinal movability.

First, a single three–axial $SO(3)-$joint can be reduced to the direct product of three uniaxial $SO(2)-$joints, in the sense that three hinge joints can produce any orientation in space, just as a ball–joint can. Algebraically, this means reduction (using symbol '$\gtrsim$') of each of the three $SO(3)$ rotation matrices to the corresponding $SO(2)$ rotation matrices

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{pmatrix} \gtrsim \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$$

$$\begin{pmatrix} \cos\psi & 0 & \sin\psi \\ 0 & 1 & 0 \\ -\sin\psi & 0 & \cos\psi \end{pmatrix} \gtrsim \begin{pmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{pmatrix}$$

$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \gtrsim \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

In this way we can set the *reduction equivalence relation* $SO(3) \gtrsim SO(2) \triangleright SO(2) \triangleright SO(2)$, where '$\triangleright$' denotes the noncommutative *semidirect product* (see (2.4.1) above).

Second, we have a homeomorphism: $SO(2) \sim S^1$, where S^1 denotes the constrained unit circle in the complex plane, which is an Abelian Lie group.

Third, let I^N be the unit cube $[0,1]^N$ in $\mathbb{R}^N$ and '$\sim$' an equivalence relation on $\mathbb{R}^N$ get by 'gluing' together the opposite sides of I^N, preserving their

orientation. The manifold of human–body configurations (Figure 2.21) can be represented as the *quotient space* of $\mathbb{R}^N$ by the space of the integral lattice points in $\mathbb{R}^N$, that is a constrained ND torus T^N (2.684),

$$\mathbb{R}^N/Z^N = I^N/\sim \cong \prod_{i=1}^{N} S_i^1 \equiv \{(q^i, i = 1, \ldots, N) : \mathrm{mod}\ 2\pi\} = T^N. \quad (2.650)$$

Since S^1 is an Abelian Lie group, its N–fold tensor product T^N is also an Abelian Lie group, the toral group, of all nondegenerate diagonal $N \times N$ matrices. As a Lie group, the biomechanical configuration space $M \equiv T^N$ has a natural Banach manifold structure with local internal coordinates $q^i \in U$, U being an open set (chart) in T^N.

Conversely by 'ungluing' the configuration space we get the primary unit cube. Let '$\sim^*$' denote an equivalent decomposition or 'ungluing' relation. By the Tychonoff product–topology theorem, for every such quotient space there exists a 'selector' such that their quotient models are homeomorphic, that is, $T^N/\sim^* \approx A^N/\sim^*$. Therefore I^N represents a 'selector' for the configuration torus T^N and can be used as an N–directional 'command–space' for the topological control of human motion. Any subset of DOF on the configuration torus T^N representing the joints included in human motion has its simple, rectangular image in the command space – selector I^N. Operationally, this resembles what the *brain–motor–controller*, the *cerebellum*, actually performs on the highest level of human motor control.

The Complete Biomechanical Manifold

The full kinematics of a whole human–like body can be split down into *five kinematic chains*: one for each leg and arm, plus one for spine with the head. In all five chains internal joint coordinates, namely n_1 constrained rotations x^k_{rt} together with n_2 of even more constrained translations x^j_{tr} (see Figure 2.22), constitute a smooth nD anthropomorphic configuration manifold M, with local coordinates x^i, $(i = 1, \ldots, n)$. That is, the motion space in each joint is defined as a semidirect (noncommutative) product of the Lie group $SO(n)$ of constrained rotations and a corresponding Lie group $\mathbb{R}^n$ of even more restricted translations. More precisely, in each movable human–like joint we have an action of the constrained special Euclidean $SE(3)$ group (see (2.4.1) above). The joints themselves are linked by direct (commutative) products.

Realistic Human Spine Manifold

The high–resolution human spine manifold is a dynamical chain consisting of 25 constrained $SE(3)-$ joints. Each movable spinal joint has 6 DOF: 3 dominant rotations, (performed first in any free spinal movement), restricted to about 7 angular degrees and 3 secondary translations (performed after reaching the limit of rotational amplitude), restricted to about 5 mm (see Figure 2.23).

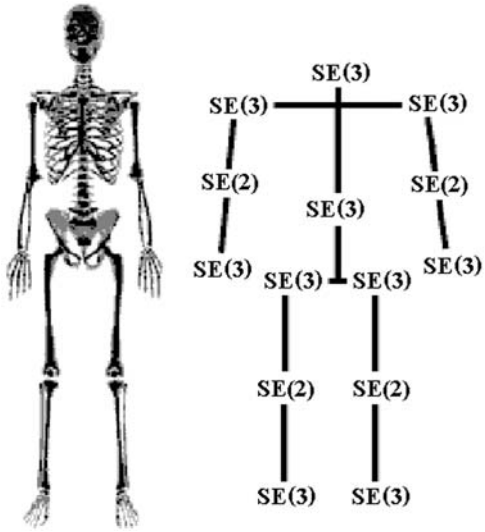


Fig. 2.22. A medium–resolution, whole–body biomechanical manifold, with just a single $SE(3)-$joint representing the spinal movability.

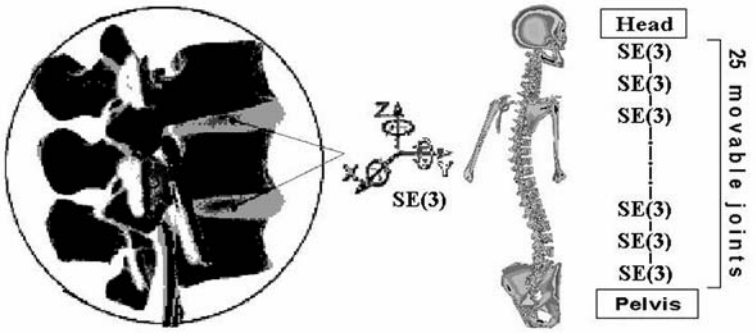


Fig. 2.23. The high–resolution human spine manifold is a dynamical chain consisting of 25 constrained $SE(3)-$joints.

Now, $SE(3) = SO(3) \triangleright \mathbb{R}^3$ is a non–compact group, so there is no any natural metric given by the kinetic energy on $SE(3)$, and consequently, no natural controls in the sense of geodesics on $SE(3)$. However, both of its subgroups, $SO(3)$ and $\mathbb{R}^3$, are compact with quadratic metric forms defined by standard line element $g_{ij}dq^i dq^j$, and therefore admit optimal muscular–like controls in the sense of geodesics (see section 1.2.9 below).

2.4.2 Muscle–Driven Hamiltonian Biomechanics

We will develop our Hamiltonian geometry on the configuration biomechanical manifold M in three steps, following the standard symplectic geometry prescription (see subsection 1.2.11 above):

Step A Find a symplectic *momentum phase–space* (P, ω).

Recall that a symplectic structure on a smooth manifold M is a nondegenerate closed 2−form ω on M, i.e., for each $x \in M$, $\omega(x)$ is nondegenerate, and $d\omega = 0$.

Let $T^*_x M$ be a cotangent space to M at m. The cotangent bundle T^*M represents a union $\cup_{m\in M} T^*_x M$, together with the standard topology on T^*M and a natural smooth manifold structure, the dimension of which is twice the dimension of M. A 1−form θ on M represents a section $\theta : M \to T^*M$ of the cotangent bundle T^*M.

$P = T^*M$ is our momentum phase–space. On P there is a nondegenerate symplectic 2−form ω is defined in local joint coordinates $q^i, p_i \in U$, U open in P, as $\omega = dq^i \wedge dp_i$ ('$\wedge$' denotes the wedge or exterior product). In that case the coordinates $q^i, p_i \in U$ are called canonical. In a usual procedure the canonical 1−form θ is first defined as $\theta = p_i dq^i$, and then the canonical 2–form ω is defined as $\omega = -d\theta$.

A *symplectic phase–space manifold* is a pair (P, ω).

Step B Find a *Hamiltonian vector–field* X_H on (P, ω).

Let (P, ω) be a symplectic manifold. A vector–field $X : P \to TP$ is called *Hamiltonian* if there is a smooth function $F : P \longrightarrow \mathbb{R}$ such that $i_X\omega = dF$ ($i_X\omega$ denotes the *interior product* or *contraction* of the vector–field X and the 2–form ω). X is *locally Hamiltonian* if $i_X\omega$ is closed.

Let the smooth real–valued *Hamiltonian function* $H : P \to \mathbb{R}$, representing the total biomechanical energy $H(q, p) = T(p) + V(q)$ (T and V denote kinetic and potential energy of the system, respectively), be given in local canonical coordinates $q^i, p_i \in U$, U open in P. The *Hamiltonian vector–field* X_H, condition by $i_{X_H}\omega = dH$, is actually defined via symplectic matrix J, in a local chart U, as

$$X_H = J\nabla H = \left(\partial_{p_i} H, -\partial_{q^i} H\right), \qquad J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix},$$

where I denotes the $n \times n$ identity matrix and ∇ is the gradient operator.

Step C Find a *Hamiltonian phase–flow* ϕ_t of X_H.

Let (P, ω) be a symplectic phase–space manifold and $X_H = J\nabla H$ a Hamiltonian vector–field corresponding to a smooth real–valued Hamiltonian function $H : P \to \mathbb{R}$, on it. If a unique one–parameter group of diffeomorphisms $\phi_t : P \to P$ exists so that $\frac{d}{dt}|_{t=0}\, \phi_t x = J\nabla H(x)$, it is called the *Hamiltonian phase–flow*.

A smooth curve $t \mapsto \left(q^i(t),\, p_i(t)\right)$ on (P, ω) represents an *integral curve* of the Hamiltonian vector–field $X_H = J\nabla H$, if in the local canonical coordinates $q^i, p_i \in U$, U open in P, *Hamiltonian canonical equations* (2.6) hold.

An integral curve is said to be *maximal* if it is not a restriction of an integral curve defined on a larger interval of $\mathbb{R}$. It follows from the standard theorem on the *existence* and *uniqueness* of the solution of a system of ODEs with smooth r.h.s, that if the manifold (P, ω) is Hausdorff, then for any point $x = (q^i, p_i) \in U$, U open in P, there exists a maximal integral curve of $X_H = J\nabla H$, passing for $t = 0$, through point x. In case X_H is complete, i.e., X_H is C^p and (P, ω) is compact, the maximal integral curve of X_H is the Hamiltonian phase–flow $\phi_t : U \to U$.

The phase–flow ϕ_t is *symplectic* if ω is constant along ϕ_t, i.e., $\phi_t^* \omega = \omega$
($\phi_t^* \omega$ denotes the *pull–back* of ω by ϕ_t),
iff $\mathcal{L}_{X_H} \omega = 0$
($\mathcal{L}_{X_H} \omega$ denotes the *Lie derivative* of ω upon X_H).

Symplectic phase–flow ϕ_t consists of canonical transformations on (P, ω), i.e., diffeomorphisms in canonical coordinates $q^i, p_i \in U$, U open on all (P, ω) which leave ω invariant. In this case the *Liouville theorem* is valid: ϕ_t *preserves* the *phase volume* on (P, ω). Also, the system's total energy H is conserved along ϕ_t, i.e., $H \circ \phi_t = \phi_t$.

Recall that the Riemannian metrics $g = <, >$ on the configuration manifold M is a positive–definite quadratic form $g : TM \to \mathbb{R}$, in local coordinates $q^i \in U$, U open in M, given by (2.1–2.2) above. Given the metrics g_{ij}, the system's Hamiltonian function represents a momentum p–dependent quadratic form $H : T^*M \to \mathbb{R}$ – the system's kinetic energy $H(p) = T(p) = \frac{1}{2} < p, p >$, in local canonical coordinates $q^i, p_i \in U_p$, U_p open in T^*M, given by

$$H(p) = \frac{1}{2} g^{ij}(q, m)\, p_i p_j, \tag{2.651}$$

where $g^{ij}(q, m) = g_{ij}^{-1}(q, m)$ denotes the *inverse* (contravariant) material *metric tensor*

$$g^{ij}(q, m) = \sum_{\chi=1}^{n} m_\chi \delta_{rs} \frac{\partial q^i}{\partial x^r} \frac{\partial q^j}{\partial x^s}.$$

T^*M is an *orientable* manifold, admitting the standard *volume form*

$$\Omega_{\omega_H} = \frac{(-1)^{\frac{N(N+1)}{2}}}{N!} \omega_H^N.$$

For Hamiltonian vector–field, X_H on M, there is a base integral curve $\gamma_0(t) = \big(q^i(t), p_i(t)\big)$ iff $\gamma_0(t)$ is a *geodesic*, given by the one–form *force equation*

$$\dot{\tilde{p}}_i \equiv \dot{p}_i + \Gamma^i_{jk}\, g^{jl} g^{km}\, p_l p_m = 0, \qquad \text{with} \qquad \dot{q}^k = g^{ki} p_i, \tag{2.652}$$

where Γ^i_{jk} denote *Christoffel symbols* of an affine *Levi–Civita connection* on M, defined upon the *Riemannian metric* $g = <, >$ by (2.5).

The l.h.s $\dot{\tilde{p}}_i$ of the covariant momentum equation (2.652) represents the *intrinsic* or *Bianchi covariant derivative* of the momentum with respect to

time t. Basic relation $\dot{\tilde{p}}_i = 0$ defines the *parallel transport* on T^N, the simplest form of human–motion dynamics. In that case Hamiltonian vector–field X_H is called the *geodesic spray* and its phase–flow is called the *geodesic flow*.

For Earthly dynamics in the gravitational *potential* field $V : M \to \mathbb{R}$, the Hamiltonian $H : T^*M \to \mathbb{R}$ (2.651) extends into potential form

$$H(p,q) = \frac{1}{2} g^{ij} p_i p_j + V(q),$$

with Hamiltonian vector–field $X_H = J\nabla H$ still defined by canonical equations (2.6).

A general form of a *driven*, non–conservative Hamiltonian equations reads:

$$\dot{q}^i = \partial_{p_i} H, \qquad \dot{p}_i = F_i - \partial_{q^i} H, \tag{2.653}$$

where $F_i = F_i(t,q,p)$ represent any kind of joint–driving *covariant torques*, including active neuro–muscular–like controls, as functions of time, angles and momenta, as well as passive dissipative and elastic joint torques. In the covariant momentum formulation (2.652), the non–conservative Hamiltonian equations (2.653) become

$$\dot{\tilde{p}}_i \equiv \dot{p}_i + \Gamma^i_{jk}\, g^{jl} g^{km}\, p_l p_m = F_i, \qquad \text{with} \qquad \dot{q}^k = g^{ki} p_i.$$

Hamiltonian–Poisson Biomechanical Systems

Recall from subsection 2.1.1 above that Hamiltonian–Poisson mechanics is a generalized form of classical Hamiltonian mechanics. Let $(P, \{\})$ be a Poisson manifold and $H \in C^\infty(P, \mathbb{R})$ a smooth real valued function on P. The vector–field X_H defined by

$$X_H(F) = \{F, H\},$$

is the Hamiltonian vector–field with energy function H. The triple $(P, \{\}, H)$ we call the *Hamiltonian–Poisson biomechanical system* (HPBS) [MR99, Put93, IP01a]. The map $F \mapsto \{F, H\}$ is a derivation on the space $C^\infty(P, \mathbb{R})$, hence it defines a vector–field on P. The map $F \in C^\infty(P, \mathbb{R}) \mapsto X_F \in \mathcal{X}(P)$ is a Lie algebra anti–homomorphism, i.e., $[X_F, X_g] = -X_{\{F,g\}}$.

Let $(P, \{\}, H)$ be a HPBS and ϕ_t the flow of X_H. Then for all $F \in C^\infty(P, \mathbb{R})$ we have the *conservation of energy*:

$$H \circ \phi_t = H,$$

and the *equations of motion in Poisson bracket form*,

$$\frac{d}{dt}\,(F \circ \phi_t) = \{F, H\} \circ \phi_t = \{F \circ \phi_t, H\},$$

that is, the above Poisson evolution equation (2.21) holds. Now, the function F is constant along the integral curves of the Hamiltonian vector–field X_H iff

$$\{F, H\} = 0.$$

ϕ_t preserves the Poisson structure.

Next we present two main examples of HPBS.

'Ball–and–Socket' Joint Dynamics in Euler Vector Form

The dynamics of human body–segments, classically modelled *via* Lagrangian formalism (see [Hat77b, Iva91, ILI95, II05]), may be also prescribed by Euler's equations of rigid body dynamics. The equations of motion for a free rigid body, described by an observer fixed on the moving body, are usually given by *Euler's vector equation*

$$\dot{p} = p \times w. \tag{2.654}$$

Here $p, w \in \mathbb{R}^3$, $p_i = I_i w_i$ and I_i $(i = 1, 2, 3)$ are the principal moments of inertia, the coordinate system in the segment is chosen so that the axes are principal axes, w is the angular velocity of the body and p is the corresponding angular momentum.

The kinetic energy of the segment is the Hamiltonian function $H : \mathbb{R}^3 \to \mathbb{R}$ given by [IP01a]

$$H(p) = \frac{1}{2} p \cdot w$$

and is a conserved quantity for (2.654).

The vector space $\mathbb{R}^3$ is a Lie algebra with respect to the bracket operation given by the usual cross product. The space $\mathbb{R}^3$ is paired with itself *via* the usual dot product. So if $F : \mathbb{R}^3 \to \mathbb{R}$, then $\delta F/\delta p = \nabla F(p)$ and the (–) Lie–Poisson bracket $\{F, G\}_-(p)$ is given *via* (2.20) by the triple product

$$\{F, G\}_-(p) = -p \cdot (\nabla F(p) \times \nabla G(p)).$$

Euler's vector equation (2.654) represents a generalized Hamiltonian system in $\mathbb{R}^3$ relative to the Hamiltonian function $H(p)$ and the (–) Lie–Poisson bracket $\{F, G\}_-(p)$. Thus the Poisson manifold $(\mathbb{R}^3, \{F, G\}_-(p))$ is defined and the abstract Poisson equation is equivalent to Euler's equation (2.654) for a body segment and associated joint.

Solitary Model of Muscular Contraction

Recall that the so–called *sliding filament theory of muscular contraction* was developed in 1950s by Nobel Laureate A. Huxley [HN54, Hux57]. At a deeper level, the basis of the molecular model of muscular contraction is represented by oscillations of *Amid I peptide groups* with associated dipole electric momentum inside a spiral structure of myosin filament molecules (see [Dav81, Dav91]).

There is a simultaneous resonant interaction and strain interaction generating a collective interaction directed along the axis of the spiral. The resonance excitation jumping from one peptide group to another can be represented as an exciton, the local molecule strain caused by the static effect of excitation as a phonon and the resultant collective interaction as a *soliton*.

The simplest model of Davydov's solitary particle–waves is given by the *nonlinear Schrödinger equation* [IP01a]

$$i\partial_t\psi = -\partial_{x^2}\psi + 2\chi|\psi|^2\psi, \tag{2.655}$$

for $-\infty < x < +\infty$. Here $\psi(x,t)$ is a smooth complex–valued wave function with initial condition $\psi(x,t)|_{t=0} = \psi(x)$ and χ is a nonlinear parameter. In the linear limit ($\chi = 0$) (2.655) becomes the ordinary Schrödinger equation for the wave function of the free 1D particle with mass $m = 1/2$.

We may define the infinite–dimensional phase–space manifold $\mathcal{P} = \{(\psi, \bar{\psi}) \in S(\mathbb{R}, \mathbf{C})\}$, where $S(\mathbb{R}, \mathbf{C})$ is the Schwartz space of rapidly–decreasing complex–valued functions defined on $\mathbb{R}$). We define also the algebra $\boldsymbol{\chi}(\mathcal{P})$ of observables on $\mathcal{P}$ consisting of real–analytic functional derivatives $\delta F/\delta\psi$, $\delta F/\delta\bar{\psi} \in S(\mathbb{R}, \mathbf{C})$.

The Hamiltonian function $H : \mathcal{P} \longrightarrow \mathbb{R}$ is given by

$$H(\psi) = \int_{-\infty}^{+\infty} \left(\left| \frac{\partial\psi}{\partial x} \right|^2 + \chi|\psi|^4 \right) dx$$

and is equal to the total energy of the soliton. It is a conserved quantity for (4.3) (see [Sei95]).

The Poisson bracket on $\boldsymbol{\chi}(\mathcal{P})$ represents a direct generalization of the classical nD Poisson bracket

$$\{F, G\}_+(\psi) = \mathrm{i} \int_{-\infty}^{+\infty} \left(\frac{\delta F}{\delta\psi} \frac{\delta G}{\delta\bar{\psi}} - \frac{\delta F}{\delta\bar{\psi}} \frac{\delta G}{\delta\psi} \right) dx. \tag{2.656}$$

It manifestly exhibits skew–symmetry and satisfies Jacobi identity. The functionals are given by $\delta F/\delta\psi = -i\{F, \bar{\psi}\}$ and $\delta F/\delta\bar{\psi} = i\{F, \psi\}$. Therefore the algebra of observables $\boldsymbol{\chi}(\mathcal{P})$ represents the Lie algebra and the Poisson bracket is the (+) Lie–Poisson bracket $\{F, G\}_+(\psi)$.

The nonlinear Schrödinger equation (2.655) for the solitary particle–wave is a Hamiltonian system on the Lie algebra $\boldsymbol{\chi}(\mathcal{P})$ relative to the (+) Lie–Poisson bracket $\{F, G\}_+(\psi)$ and Hamiltonian function $H(\psi)$. Therefore the Poisson manifold $(\boldsymbol{\chi}(\mathcal{P}), \{F, G\}_+(\psi))$ is defined and the abstract Poisson evolution equation (2.21), which holds for any smooth function $F : \boldsymbol{\chi}(\mathcal{P}) \to \mathbb{R}$, is equivalent to equation (2.655).

A more subtle model of soliton dynamics is provided by the *Korteveg–de Vries equation* [IP01a]

$$f_t - 6ff_x + f_{xxx} = 0, \qquad (f_x = \partial_x f), \tag{2.657}$$

where $x \in \mathbb{R}$ and f is a real–valued smooth function defined on $\mathbb{R}$ (compare with (1.52) above). This equation is related to the ordinary Schrödinger equation by the inverse scattering method [Sei95, IP01a].

We may define the infinite–dimensional phase–space manifold $\mathcal{V} = \{f \in S(\mathbb{R})\}$, where $S(\mathbb{R})$ is the Schwartz space of rapidly–decreasing real–valued functions $\mathbb{R}$). We define further $\boldsymbol{\chi}(\mathcal{V})$ to be the algebra of observables consisting of functional derivatives $\delta F/\delta f \in S(\mathbb{R})$.

The Hamiltonian $H : \mathcal{V} \to \mathbb{R}$ is given by

$$H(f) = \int_{-\infty}^{+\infty} (f^3 + \frac{1}{2} f_x^2)\, dx$$

and gives the total energy of the soliton. It is a conserved quantity for (2.657) (see [Sei95]).

As a real–valued analogue to (2.656), the (+) Lie–Poisson bracket on $\boldsymbol{\chi}(\mathcal{V})$ is given *via* (2.19) by

$$\{F, G\}_+(f) = \int_{-\infty}^{+\infty} \frac{\delta F}{\delta f} \frac{d}{dx} \frac{\delta G}{\delta f}\, dx.$$

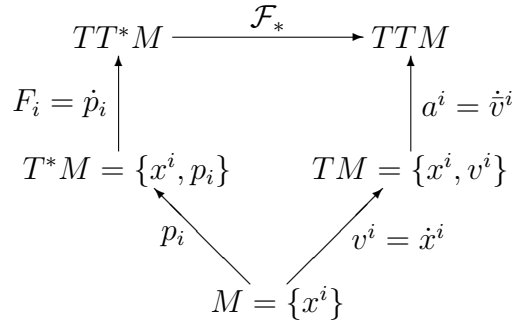
Again it possesses skew–symmetry and satisfies Jacobi identity. The functionals are given by $\delta F/\delta f = \{F, f\}$.

The Korteveg–de Vries equation (KdV1), describing the behavior of the molecular solitary particle–wave, is a Hamiltonian system on the Lie algebra $\boldsymbol{\chi}(\mathcal{V})$ relative to the (+) Lie–Poisson bracket $\{F, G\}_+(f)$ and the Hamiltonian function $H(f)$. Therefore, the Poisson manifold $(\boldsymbol{\chi}(\mathcal{V}), \{F, G\}_+(f))$ is defined and the abstract Poisson evolution equation (2.21), which holds for any smooth function $F : \boldsymbol{\chi}(\mathcal{V}) \to \mathbb{R}$, is equivalent to (2.657).

2.4.3 Biomechanical Functors

The Covariant Force Functor

Recall (see subsection 3.1.5 in Appendix) that int the realm of biomechanics the central concept is the *covariant force law*, $F_i = m g_{ij} a^j$ [II05]. In categorical language, it represents the *covariant force functor* $\mathcal{F}_*$ defined by commutative diagram:



which states that the force 1–form $F_i = \dot{p}_i$, defined on the mixed tangent–cotangent bundle TT^*M, causes the acceleration vector–field $a^i = \dot{\tilde{v}}^i$, defined on the second tangent bundle TTM of the configuration manifold M.

The corresponding *contravariant acceleration functor* is defined as its inverse map $\mathcal{F}^* : TTM \longrightarrow TT^*M$.

In the following subsections we present several Lie functors, as they are used in modern biomechanical research, all being different formulations of the covariant force law, $F_i = mg_{ij}a^j$, and giving different Lie representations of the fundamental covariant force functor $\mathcal{F}_* : TT^*M \longrightarrow TTM$.

Lie–Lagrangian Biomechanical Functor

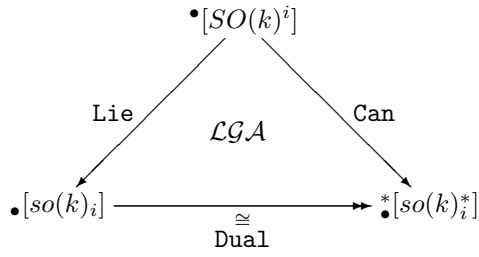
Now we develop the *Lie–Lagrangian biomechanical functor* using a modern, nonlinear formulation of the classical robotics structure (see [Iva05b, Iva05c]):

Kinematics	$\rightarrow$	*Dynamics*	$\rightarrow$	*Control*
Lie groups	$\rightarrow$	*Exterior Lagrangian*	$\rightarrow$	*Lie derivative*

The conservative part of generalized Lagrangian formalism, as used in biomechanics, is derived from Lagrangian conservative energy function. It describes the motion of the conservative skeleton, which is free of control and dissipation. According to the Liouville theorem, this conservative dynamics is structurally unstable due to the phase–space spreading effect, caused by the growth of entropy (see [Iva91, ILI95, IS01, II05]. The dissipative part is derived from nonlinear dissipative function, and describes quadratic joint dampings, which prevent entropy growth. Its driving part represents equivalent muscular torques F_i acting in all DOF (or just in active joints, as used in the affine input control), in the form of force–time and force–velocity signals.

Joint Kinematics

Recall that human joints represented by internal coordinates x^i $(i = 1, \ldots, n)$, constitute an nD smooth biomechanical configuration manifold M (see Figure 2.21). Now we are going to perform some categorical transformations on the biomechanical configuration manifold M. If we apply the *functor* `Lie` to the *category* ${}^{\bullet}[SO(k)^i]$ of rotational Lie groups $SO(k)^i$ and their homomorphisms we get the category ${}_{\bullet}[so(k)_i]$ of corresponding *tangent* Lie algebras $so(k)_i$ and their homomorphisms. If we further apply the *isomorphic* functor `Dual` to the category ${}_{\bullet}[so(k)_i]$ we get the dual category ${}^{*}_{\bullet}[so(k)^*_i]$ of *cotangent*, or, *canonical* Lie algebras $so(k)^*_i$ and their homomorphisms. To go directly from ${}^{\bullet}[SO(k)^i]$ to ${}^{*}_{\bullet}[so(k)^*_i]$ we use the canonical functor `Can` [IS01, Iva02, IB05, Iva05a]. Therefore we have a commutative triangle



Both the tangent algebras $so(k)_i$ and the cotangent algebras $so(k)_i^*$ contain infinitesimal group generators, angular velocities $\dot{x}^i = \dot{x}^{\phi_i}$ in the first case and canonical angular momenta $p_i = p_{\phi_i}$ in the second. As Lie group generators, angular velocities and angular momenta satisfy the respective commutation relations $[\dot{x}^{\phi_i}, \dot{x}^{\psi_i}] = \epsilon_\theta^{\phi\psi} \dot{x}^{\theta_i}$ and $[p_{\phi_i}, p_{\psi_i}] = \epsilon^\theta_{\phi\psi} p_{\theta_i}$, where the structure constants $\epsilon_\theta^{\phi\psi}$ and $\epsilon^\theta_{\phi\psi}$ constitute totally antisymmetric third–order tensors.

In this way, the functor $\texttt{Dual}_G : \texttt{Lie} \cong \texttt{Can}$ establishes a geometrical duality between kinematics of angular velocities $\dot{x}^i$ (involved in Lagrangian formalism on the tangent bundle of M) and that of angular momenta p_i (involved in Hamiltonian formalism on the cotangent bundle of M). This is analyzed below. In other words, we have two functors `Lie` and `Can` from a category of Lie groups (of which ${}^\bullet[SO(k)^i]$ is a subcategory) into a category of their Lie algebras (of which ${}_\bullet[so(k)_i]$ and ${}^*_\bullet[so(k)_i^*]$ are subcategories), and a natural equivalence (functor isomorphism) between them defined by the functor $\texttt{Dual}_G$. (As angular momenta p_i are in a bijective correspondence with angular velocities $\dot{x}^i$, every component of the functor $\texttt{Dual}_G$ is invertible.)

Applying the functor `Lie` to the biomechanical configuration manifold M (Figure 2.21), we get the product–tree of the same anthropomorphic structure, but having tangent Lie algebras $so(k)_i$ as vertices, instead of the groups $SO(k)^i$. Again, applying the functor `Can` to M, we get the product–tree of the same anthropomorphic structure, but this time having cotangent Lie algebras $so(k)_i^*$ as vertices.

The functor `Lie` defines the *second–order Lagrangian formalism on the tangent bundle* TM (i.e., the *velocity phase–space manifold*) while the functor `Can` defines *the first–order canonical Hamiltonian formalism on the cotangent bundle* T^*M (i.e., the *momentum phase–space manifold*). As these two formalisms are related by the isomorphic functor `Dual`, they are equivalent. In this section we shall follow the Lagrangian functor `Lie`, using the powerful formalism of exterior differential systems and integral variational principles [Gri83a, BM82]. For the parallel, Hamiltonian treatment along the functor `Can`, more suitable for chaos theory and stochastic generalizations, see [IS01, Iva02].

Exterior Lagrangian Dynamics

Let $\Omega^p(M) = \sum \omega_I dx^I$ denote the space of differential $p-$forms on M. That is, if multi–index $I \subset \{1, \dots, n\}$ is a subset of p elements then we have a $p-$form $dx^I = dx^i_1 \wedge dx^i_2 \wedge \cdots \wedge dx^i_p$ on M. We define the exterior derivative on M as $d\omega = \sum \frac{\partial \omega^I}{\partial x_p} dx_p \wedge dx^I$ (compare with (1.4.7) above).

Now, from exterior differential systems point of view (see subsection 1.2.5 above as well as [Gri83a]), human–like motion represents an n DOF neuro–musculo–skeletal system Ξ, evolving in time on its nD configuration manifold M, (with local coordinates x^i, $i = 1, ..., n$) as well as on its tangent bundle TM (with local coordinates $(x^i; \dot{x}^i)$).

For the system Ξ we will consider a *well–posed variational problem* $(I, \omega; \varphi)$, on an associated $(2n+1)$–D jet space $X = J^1(\mathbb{R}, M) \cong \mathbb{R} \times TM$, with local canonical variables $(t; x^i; \dot{x}^i)$.

Here, (I, ω) is called a *Pfaffian exterior differential system* on X (see [Gri83a]), given locally as

$$\begin{cases} \theta^i = dx^i - \dot{x}^i \omega = 0 \\ \omega \equiv dt \neq 0 \end{cases}, \tag{2.658}$$

with the structure equations

$$d\theta^i = -d\dot{x}^i \wedge \omega.$$

Integral manifolds $N \in J^1(\mathbb{R}, M)$ of the Pfaffian system (I, ω) are locally one–jets $t \to (t, x(t), \dot{x}(t))$ of curves $x = x(t) : \mathbb{R} \to M$.

φ is a $1-$form

$$\varphi = L\, \omega, \tag{2.659}$$

where $L = L(t, x, \dot{x})$ is the system's *Lagrangian* function defined on X, having both coordinate and velocity partial derivatives, respectively denoted by $L_{x^i} \equiv \partial_{x^i} L$, and $L_{\dot{x}^i} \equiv \partial_{\dot{x}^i} L$.

A variational problem $(I, \omega; \varphi)$ is said to be *strongly non–degenerate*, or *well–posed* [Gri83a], if the determinant of the matrix of mixed velocity partials of the Lagrangian is positive definite, i.e.,

$$\det \| L_{\dot{x}^i \dot{x}^j} \| > 0.$$

The *extended Pfaffian system*

$$\begin{cases} \theta^i = 0 \\ dL_{\dot{x}^i} - L_{x^i}\, \omega = 0 \\ \omega \neq 0 \end{cases}.$$

generates classical *Euler–Lagrangian equations*

$$\frac{d}{dt} L_{\dot{x}^i} = L_{x^i}, \tag{2.660}$$

describing the *control–free, dissipation–free, conservative skeleton dynamics.*

If an *integral manifold* N satisfies the Euler–Lagrangian equations (2.660) of a well–posed variational problem on X then

$$\frac{d}{dt} \left(\int_{N_t} \varphi \right)_{t=0} = 0$$

for any *admissible variation* $N_t \in N$ that satisfies the *endpoint conditions* $\omega = \theta^i = 0$.

Theorem: *Under the above conditions, both the Lagrangian dynamics with initial conditions*

$$\begin{cases} \frac{d}{dt} L_{\dot{x}^i} = L_{x^i} \\ x(t_0) = x_0, \qquad \dot{x}(t_0) = \dot{x}_0 \end{cases}$$

and the Lagrangian dynamics with endpoint conditions

$$\begin{cases} \frac{d}{dt} L_{\dot{x}^i} = L_{x^i} \\ x(t_0) = x_0, \qquad x(t_1) = x_1 \end{cases}$$

have unique solutions. For the proof, see [Gri83a].

Now, if M is a smooth Riemannian manifold, its metric $g =< . >$ is locally given by a positive definite quadratic form

$$ds^2 = g_{ij}(x)\, dx^i dx^j, \tag{2.661}$$

where the metric tensor g_{ij} is a C^∞ symmetric matrix $g(x) = \|g_{ij}(x)\|$.

Kinetic energy of the system Ξ is a function $T = T(x, \dot{x})$ on the tangent bundle TM, which induces a positive definite quadratic form in each fibre $T_xM \subset TM$. In local coordinates, it is related to the Riemannian metric (2.661) by: $T\,\omega^2 = \frac{1}{2}\, ds^2$.

If potential energy of the system Ξ is a function $U = U(x)$ on M, then the autonomous Lagrangian is defined as $L(x, \dot{x}) = T(x, \dot{x}) - U(x)$, i.e., kinetic minus potential energy.

The condition of well–posedness is satisfied, as

$$\det \|L_{\dot{x}^i \dot{x}^j}\| = \det \|g_{ij}(x)\| > 0.$$

Now, the *covariant Euler–Lagrangian equations* (2.660) expand as

$$\frac{d}{dt}\left(g_{ij}(x(t))\, \dot{x}^j(t)\right) = \frac{1}{2}\, \left(\partial_{x^i} g_{jk}(x(t))\, \dot{x}^j(t)\, \dot{x}^k(t)\right) - F_i(x(t)), \tag{2.662}$$

where $F_i(x(t)) = \frac{\partial U(x(t))}{\partial \dot{x}^i}$ denote the *gradient force 1–forms.*

Letting $\left\|g^{ij}(x)\right\|$ be the inverse matrix to $\|g_{ij}(x)\|$ and introducing the *Christoffel symbols*

$$\Gamma^i_{jk} = g^{il}\Gamma_{jkl}, \qquad \Gamma_{jkl} = \frac{1}{2}\left(\partial_{x^j} g_{kl} + \partial_{x^k} g_{jl} - \partial_{x^l} g_{jk}\right)$$

the equations (2.662) lead to the classical *contravariant form* (see [Iva91, IP01b, II05])

$$\ddot{x}^i(t) + \Gamma^i_{jk}(x(t))\, \dot{x}^j(t)\, \dot{x}^k(t) = -F^i(x(t)), \tag{2.663}$$

where $F^i(x(t)) = g^{ij}(x) \frac{\partial U(x(t))}{\partial \dot{x}^j}$ denote the *gradient force vector–fields.*

The above theorem implies that both the Lagrangian dynamics with initial conditions

$$\begin{cases} \ddot{x}^i(t) + \Gamma^i_{jk}(x(t))\, \dot{x}^j(t)\, \dot{x}^k(t) = -F^i(x(t)) \\ x(t_0) = x_0, \qquad \dot{x}(t_0) = \dot{x}_0 \end{cases} \tag{2.664}$$

and the Lagrangian dynamics with endpoint conditions

$$\begin{cases} \ddot{x}^i(t) + \Gamma^i_{jk}(x(t))\,\dot{x}^j(t)\,\dot{x}^k(t) = -F^i(x(t)) \\ x(t_0) = x_0, \qquad x(t_1) = x_1 \end{cases} \tag{2.665}$$

have unique solutions. We consider the system (2.664) to be the valid basis of human–like dynamics, and the system (2.665) to be the valid basis of the finite biomechanics control.

Now, recall that any smooth $n-$manifold M induces an $n-$*category* $\Pi_n(M)$, its fundamental $n-$*groupoid*. In $\Pi_n(M)$, 0–cells are *points* in M; 1–cells are *paths* in M (i.e., parameterized smooth maps $f : [0,1] \to M$); 2–cells are *smooth homotopies* (denoted by $\simeq$) *of paths* relative to endpoints (i.e., parameterized smooth maps $h : [0,1] \times [0,1] \to M$); 3–cells are *smooth homotopies of homotopies* of paths in M (i.e., parameterized smooth maps $j : [0,1] \times [0,1] \times [0,1] \to M$). Categorical *composition* is defined by *pasting* paths and homotopies, which gives the *recursive homotopy dynamics* (see below).

On the other hand, to describe the biomechanical realism, we have to generalize (2.663), so to include any other type of *external* contravariant forces (including excitation and contraction dynamics of muscular–like actuators, as well as nonlinear dissipative joint forces) to the r.h.s of (2.663). In this way, we get the *general form* of *contravariant Lagrangian dynamics*

$$\ddot{x}^i(t) + \Gamma^i_{jk}(x(t))\,\dot{x}^j(t)\,\dot{x}^k(t) = \mathcal{F}^i\left(t, x(t), \dot{x}(t)\right), \tag{2.666}$$

or, in exterior, *covariant* form

$$\frac{d}{dt} L_{\dot{x}^i} - L_{x^i} = \mathcal{F}_i\left(t, x(t), \dot{x}(t)\right). \tag{2.667}$$

Recursive homotopy dynamics:

$\mathtt{0-cell}$: $x_0 \bullet$ $\qquad x_0 \in M$; $\qquad$ in the higher cells below: $t, s \in [0,1]$;

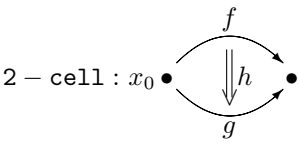
$\mathtt{1-cell}$: $x_0 \bullet \xrightarrow{\quad f \quad} \bullet\, x_1$ $\qquad f : x_0 \simeq x_1 \in M$,

$f : [0,1] \to M$, $f : x_0 \mapsto x_1$, $x_1 = f(x_0)$, $f(0) = x_0$, $f(1) = x_1$;

e.g., linear path: $f(t) = (1-t)\,x_0 + t\,x_1$; $\qquad$ or

e.g., Euler–Lagrangian $f-$dynamics with endpoint conditions (x_0, x_1) :

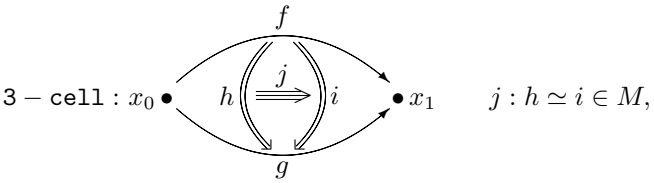
$\frac{d}{dt} f_{\dot{x}^i} = f_{x^i}$, $\quad$ with $\quad x(0) = x_0, \quad x(1) = x_1, \quad (i = 1, ..., n)$;



$h : f \simeq g \in M,$

$h : [0,1] \times [0,1] \to M$, $h : f \mapsto g$, $g = h(f(x_0))$,
$h(x_0, 0) = f(x_0)$, $h(x_0, 1) = g(x_0)$, $h(0, t) = x_0$, $h(1, t) = x_1$
e.g., linear homotopy: $h(x_0, t) = (1 - t)\, f(x_0) + t\, g(x_0)$; or
e.g., homotopy between two Euler–Lagrangian $(f, g)-$dynamics
with the same endpoint conditions (x_0, x_1) :

$$\frac{d}{dt} f_{\dot{x}^i} = f_{x^i}, \quad \text{and} \quad \frac{d}{dt} g_{\dot{x}^i} = g_{x^i} \quad \text{with} \quad x(0) = x_0, \quad x(1) = x_1;$$



$j : h \simeq i \in M,$

$j : [0,1] \times [0,1] \times [0,1] \to M$, $j : h \mapsto i$, $i = j(h(f(x_0)))$
$j(x_0, t, 0) = h(f(x_0))$, $j(x_0, t, 1) = i(f(x_0))$,
$j(x_0, 0, s) = f(x_0)$, $j(x_0, 1, s) = g(x_0)$,
$j(0, t, s) = x_0$, $j(1, t, s) = x_1$
e.g., linear composite homotopy: $j(x_0, t, s) = (1 - t)\, h(f(x_0)) + t\, i(f(x_0))$;
or, homotopy between two homotopies between above two Euler-Lagrangian $(f, g)-$dynamics with the same endpoint conditions (x_0, x_1).

Lie–Hamiltonian Biomechanical Functor

The three fundamental and interrelated obstacles facing any researcher in the field of human–like musculo–skeletal dynamics, could be identified as [IS01]:

1. *Deterministic chaos*,
2. *Stochastic forces*, and
3. *Imprecision of measurement* (or estimation) of the system numbers (SN): inputs, parameters and initial conditions.

Recall that the deterministic chaos is manifested as an irregular and unpredictable time evolution of purely deterministic nonlinear systems. If a nonlinear system is started twice, from slightly different initial conditions, its time evolution differs exponentially, while in case of a linear system, the difference in time evolution is linear.

Again, recall that the stochastic dynamics is based on the concept of *Markov process*[34], which represents the probabilistic analogue to the deterministic dynamics. The property of a *Markov chain* of prime importance for human–motion dynamics is the existence of an *invariant distribution of states*: we start with an initial state x_0 whose absolute probability is 1. Ultimately the states should be distributed according to a specified distribution.

Recall that *Brownian dynamics* represents the phase–space trajectories of a collection of particles that individually obey *Langevin rate equations* (see [Gar85]) in the field of force (i.e., the particles interact with each other via some deterministic force). For one free particle the Langevin equation of motion is given by

$$m\dot{v} = R(t) - \beta v,$$

where m denotes the mass of the particle and v its velocity. The r.h.s represents the coupling to a *heat bath*; the effect of the random force $R(t)$ is to heat the particle. To balance overheating (on the average), the particle is subjected to *friction* β.

Noe, between pure deterministic (in which all DOF of the system in consideration are explicitly taken into account, leading to classical dynamical equations like Hamiltonian) and pure stochastic dynamics (Markov process), there is so–called *hybrid dynamics*, particularly the Brownian dynamics, in which some of DOF are represented only through their *stochastic influence* on others.

System theory and artificial intelligence have long investigated the topic of *uncertainty* in measurement, modelling and simulation. Research in artificial intelligence has enriched the spectrum of available techniques to deal with uncertainty by proposing a theory of possibility, based on the theory of fuzzy sets (see [Yag87, DP80, Cox92, Cox94]). The field of qualitative reasoning and simulation [BK92] is also interested in modelling incompletely known systems where qualitative values are expressed by intervals. However, qualitative simulation techniques reveal a low predictive power in presence of complex models. In this section we have combined qualitative and quantitative methods, in spirit of [Bon95, IS01].

In this section we will deal with the general biomechanics from the point of view that mathematically and logically approaches a *general theory of systems*, i.e., that *makes the unique framework* for both linear and nonlinear, discrete and continuous, deterministic and stochastic, crisp and fuzzy, SISO and MIMO–systems, and *generalizes the robot dynamics* elaborated in the literature (see [VJ69, VJF70, VFJ70, VS72, VS73, IN92, Hur93, SGL93, SK93]), including all necessary DOF to match the physiologically realistic human–like motion. Yet, we wish *to avoid* all the mentioned fundamental system *obsta-*

[34] Recall that the Markov process is characterized by a *lack of memory*, i.e., the statistical properties of the immediate future are uniquely determined by the present, regardless of the past (see [Gar85]).

cles. To achieve this goal we have formulated the general *biomechanics functor machine*, covering a union of the three intersected frameworks:

1. Muscle–driven, dissipative, Hamiltonian (nonlinear, both discrete and continuous) MIMO–system;
2. Stochastic forces (including dissipative fluctuations and 'Master' jumps); and
3. Fuzzy system numbers.

The Abstract Functor Machine

In this subsection we define the *abstract functor machine* [IS01] (compare with [AAM76]) by a two–step generalization of the Kalman's modular theory of linear MIMO–systems. The first generalization puts the Kalman's theory into the category `Vect` of vector spaces and linear operators (see [MacL71] for technical details about categorical language), thus formulating the unique, categorical formalism valid both for the discrete– and continuous–time MIMO–systems.

We start with the unique, *continual–sequential state equation*

$$\dot{x}(t+1) = Ax(t) + Bu(t), \qquad y(t) = Cx(t), \tag{2.668}$$

where the nD vector spaces of *state* $X \ni x$, *input* $U \ni u$, and *output* $Y \ni y$ have the corresponding linear operators, respectively $A : X \to X$, $B : U \to X$, and $C : X \to Y$. The modular system theory comprises the *system dynamics*, given by a pair (X, A), together with a *reachability map* $e : U \to X$ of the pair (B, A), and an *observability map* $m : X \to Y$ of the pair (A, C). If the reachability map e is surjection the system dynamics (X, A) is called *reachable*; if the observability map m is injection the system dynamics (X, A) is called *observable*. If the system dynamics (X, A) is both reachable and observable, a *composition* $r = m \circ e : U \to Y$ defines the *total system's response*, which is given by solution of equation (2.668). If the unique solution to the continual–sequential state equation exists, it gives the answer to the (minimal) *realization problem*: find the system S that realizes the given response $r = m \circ e : U \to Y$ (in the smallest number of discrete states and in the shortest time).

In categorical language, the system dynamics in the category `Vect` is a pair (X, A), where $X \in \mathtt{Ob}(\mathtt{Vect})$ is an object in `Vect` and $A : X \to X \in \mathtt{Mor}(\mathtt{Vect})$ is a `Vect`–morphism. A *decomposable system* in `Vect` is such a sextuple $S \equiv (X, A, U, B, Y, C)$ that (X, A) is the system dynamics in `Vect`, a `Vect`–morphism $B : U \to X$ is an *input map*, and a `Vect`–morphism $C : X \to Y$ is an *output map*. Any object in `Vect` is characterized by mutually *dual*[35] notions of its *degree* (a number of its input morphisms) and its *codegree* (a number of its output morphisms). Similarly, any decomposable system S in

[35] Recall that in categorical language *duality* means reversing the (arrows of) morphisms; the knowledge of one of the two mutually dual terms automatically implies the knowledge of the other.

`Vect` has a *reachability map* given by an epimorphism $e = A \circ B : U \to X$ and its dual *observability map* given by a monomorphism $m = C \circ A : X \to Y$; their composition $r = m \circ e : U \to Y$ in `Mor(Vect)` defines the total system's response in `Vect` given by the unique solution of the continual–sequential state equation (2.668).

The second generalization gives an extension of the continual–sequential MIMO–system theory: from the linear category `Vect` – to an arbitrary nonlinear category $\mathcal{K}$. We do this extension (see [IS01]) by formally applying the *action* of the *nonlinear process–functor* $\mathcal{F} : \mathcal{K} \Rightarrow \mathcal{K}$ *on* the decomposable system $S \equiv (X, A, U, B, Y, C)$ in `Vect`. Under the action of the process functor $\mathcal{F}$ the linear system dynamics (X, A) in `Vect` transforms into a nonlinear $\mathcal{F}$–dynamics $(\mathcal{F}[X], \mathcal{F}[A])$ in $\mathcal{K}$, creating the *functor machine* in $\mathcal{K}$ represented by a *nonlinear decomposable system* $\mathcal{F}[S] \equiv (\mathcal{F}[X], \mathcal{F}[A], \mathcal{F}[U], \mathcal{F}[B], \mathcal{F}[Y], \mathcal{F}[C])$. The reachability map transforms into the *input process* $\mathcal{F}[e] = \mathcal{F}[A] \circ \mathcal{F}[B] : \mathcal{F}[U] \to \mathcal{F}[X]$, while its dual, observability map transforms into the *output process* $\mathcal{F}[m] = \mathcal{F}[C] \circ \mathcal{F}[A] : \mathcal{F}[X] \to \mathcal{F}[Y]$. In this way the total response of the linear system $r = m \circ e : U \to Y$ in `Mor(Vect)` transforms into the *nonlinear system behavior* $\mathcal{F}[r] = \mathcal{F}[m] \circ \mathcal{F}[e] : \mathcal{F}[U] \to \mathcal{F}[Y]$ in `Mor`$(\mathcal{K})$. Obviously, $\mathcal{F}[r]$, if exists, is given by a nonlinear $\mathcal{F}$–transform of the linear state equation (2.668).

The purpose of this section is to formulate a nonlinear $\mathcal{F}$–transform for the linear state equation (2.668) for biomechanics, i.e., the biomechanics functor machine. In subsequent sections we give a three–step development of a fuzzy–stochastic–Hamiltonian formulation for the biomechanics functor machine $\mathcal{F}[S]$, with a corresponding nonlinear system behavior $\mathcal{F}[r]$.

Muscle–Driven, Dissipative, Hamiltonian Biomechanics

In this subsection we choose the functor `Can`, as the first–order Hamiltonian formalism is more suitable for both stochastic and fuzzy generalizations to follow. Recall that the general deterministic Hamiltonian biomechanics, representing the canonical functor `Can` : $\mathcal{S}^{\bullet}[SO(n)^i] \Rightarrow \mathcal{S}^{*}_{\bullet}[so(n)^{*}_{i}]$, is given by dissipative, driven $\delta-$Hamiltonian equations,

$$\dot{q}^i = \frac{\partial H}{\partial p_i} + \frac{\partial R}{\partial p_i}, \tag{2.669}$$

$$\dot{p}_i = F_i - \frac{\partial H}{\partial q^i} + \frac{\partial R}{\partial q^i}, \tag{2.670}$$

$$q^i(0) = q^i_0, \qquad p_i(0) = p^0_i, \tag{2.671}$$

including *contravariant* equation (2.669) – the *velocity vector–field*, and *covariant* equation (2.670) – the *force* $1-$*form*, together with initial joint angles and momenta (2.671). Here $(i = 1, \ldots, N)$, and $R = R(q, p)$ denotes the Raileigh nonlinear (biquadratic) dissipation function, and $F_i = F_i(t, q, p)$ are covariant driving torques of *equivalent muscular actuators*, resembling muscular excitation and contraction dynamics in rotational form.

The velocity vector–field (2.669) and the force 1−form (2.670) together define the generalized Hamiltonian vector–field X_H, which geometrically represents the *section* of the momentum phase–space manifold T^*M, which is itself the cotangent bundle of the biomechanical configuration manifold M; the Hamiltonian (total energy) function $H = H(q, p)$ is its generating function.

As a Lie group, the configuration manifold M is Hausdorff [AMR88, MR99, Pos86]. Therefore, for $x = (q^i, p_i) \in U_p$, U_p open in T^*M, there exists a unique one–parameter group of diffeomorphisms $\phi_{\delta_t} : T^*M \to T^*M$, the generalized deterministic $\delta-$Hamiltonian phase–flow

$$\begin{aligned} &\phi_{\delta_t} : G_1 \times T^*M \to T^*M : (p(0), q(0)) \mapsto (p(t), q(t)), \qquad (2.672)\\ &(\phi_{\delta_t} \circ \phi_{\delta_s} = \phi_{\delta_{t+s}}, \quad \phi_{\delta_0} = \text{identity}), \end{aligned}$$

given by (2.669–2.671) such that

$$\frac{d}{dt} \mid_{t=0} \phi_{\delta_t} x = J \nabla H(x).$$

The $\delta-$Hamiltonian system (2.669–2.671), with its $\delta-$Hamiltonian phase–flow ϕ_{δ_t} (2.672), i.e., the canonical functor `Can`, represents our first, continual–deterministic model for the biomechanics functor machine $\mathcal{F}[S]$ with the nonlinear system behavior $\mathcal{F}[r]$. In the two subsequent sections we generalize this model to include discrete stochastic forces and fuzzy SN.

Stochastic–Lie–Hamiltonian Biomechanical Functor

In terms of the Markov stochastic process, we can interpret the deterministic $\delta-$Hamiltonian biomechanical system (2.669–2.671) as deterministic drift corresponding to the *Liouville equation*. Thus, we can naturally (in the sense of Langevin) add the covariant vector $\sigma_i(t)$ of stochastic forces (diffusion fluctuations and discontinuous–Master jumps) $\sigma_i(t) = B_{ij}[q^i(t), t]\, dW^j(t)$ to the canonical force equation. In this way we get *stochastic* $\sigma-$Hamiltonian biomechanical system, a *stochastic transformation* `Stoch`[`Can`] of the canonical functor `Can`,

$$dq^i = \left(\frac{\partial H}{\partial p_i} + \frac{\partial R}{\partial p_i} \right) dt, \qquad (2.673)$$

$$dp_i = \left(F_i - \frac{\partial H}{\partial q^i} + \frac{\partial R}{\partial q^i} \right) dt + \sigma_i(t), \qquad (2.674)$$

$$\sigma_i(t) = B_{ij}[q^i(t), t]\, dW^j(t), \qquad q^i(0) = q^i_0, \qquad p_i(0) = p^0_i.$$

In our low–dimensional example–case of symmetrical 3D load–lifting, the velocity and force $\sigma-$Hamiltonian biomechanics equations (2.673–2.674) become

$$dq^i = \left(p_i \left\{ [J_i]^{-1} + \left[m_i \left(\sum_{j=1}^{i} L_j \cos q^j \right)^2 \right]^{-1} \right\} + \frac{\partial R}{\partial p_i} \right) dt,$$

$$dp_i = B_{ij}[q^i(t), t]\, dW^j(t) + \left(F_i - g \sum_{j=i}^{10-i} L_j m_j \sin q^j \right.$$

$$\left. - \sum_{j=i}^{10-i} L_j \sin q^j p_i p_j \left[m_i \left(\sum_{k=1}^{i} L_k \cos q^k \right)^3 \right]^{-1} + \frac{\partial R}{\partial q^i} \right) dt.$$

Recall that Ito quadratic cotangent bundle I^*Q^N is defined as a Whitney sum

$$I^*Q^N = T^*Q^N \oplus SQ^N,$$

where SQ^N corresponds to *stochastic tensor bundle*, whose elements are 2nd–order tensor–fields composed of continual diffusion fluctuations and discontinuous jumps at every point of the manifold Q^N. On I^*Q^N is defined a non–degenerate, stochastic 2−form α which is closed, i.e., $d\alpha = 0$, and exact, i.e., $\alpha = d\beta$, where 1−form β represents a section $\beta : Q^N \to I^*Q^N$ of the Ito bundle I^*Q^N.

Now, the stochastic Hamiltonian vector–field Ξ_H represents a section $\Xi_H : Q^N \to IQ^N$ of the *Ito quadratic tangent bundle* IQ^N, also defined as a Whitney sum

$$IQ^N = TM \oplus SQ^N.$$

The quadratic character of Ito stochastic fibre–bundles corresponds to the second term (trace of the 2nd–order tensor–field) of associate *stochastic Taylor expansion* (see [Elw82, May81]).

Through stochastic σ−Hamiltonian biomechanical system (2.673–2.674), the deterministic δ−Hamiltonian phase–flow ϕ_{δ_t} (2.672), extends into stochastic σ−Hamiltonian phase–flow ϕ_{σ_t}

$$\phi_{\sigma_t} : G_1 \times I^*M \to I^*M : (p(0), q(0)) \mapsto (p(t), q(t)), \qquad (2.675)$$
$$(\phi_{\sigma_t} \circ \phi_{\sigma_s} = \phi_{\sigma_{t+s}}, \quad \phi_{\sigma_0} = \text{identity}),$$

where I^*M denotes *Ito quadratic cotangent bundle* (see [Elw82, May81]) of biomechanical configuration manifold M.

Besides the σ−Hamiltonian phase–flow ϕ_{σ_t} (2.675), including N individual random–phase trajectories, we can also define (see [Elw82]) an *average* or *mean* $\langle\sigma\rangle$ − Hamiltonian flow $\langle\phi\rangle_{\sigma_t}$

$$\langle\phi\rangle_{\sigma_t} : G_1 \times I^*M \to I^*M : (\langle p(0)\rangle, \langle q(0)\rangle) \mapsto (\langle p(t)\rangle, \langle q(t)\rangle),$$
$$(\langle\phi\rangle_{\sigma_t} \circ \langle\phi\rangle_{\sigma_s} = \langle\phi\rangle_{\sigma_{t+s}}, \quad \langle\phi\rangle_{\sigma_0} = \text{identity}),$$

which stochastically corresponds to the trajectory of the center of mass in the human–like dynamics, approximatively lumbo–sacral spinal $SO(3)$−joint.

The necessary conditions for existence of a unique *non–anticipating solution* of the $\sigma-$Hamiltonian biomechanical system in a fixed time interval are *Lipschitz condition* and *growth condition* (see [Elw82, May81]). For constructing an approximate solution a simple iterative Cauchy–Euler procedure could be used to calculate (q^i_{k+1}, p_i^{k+1}) from the knowledge of (q^i_k, p^k_i) on the mesh of time points t^k, $k = 1, \ldots, s$, by adding discrete δ–Hamiltonian drift–terms $A^i(q^i_k)\Delta t^k$ and $A_i(p^k_i)\Delta t^k$,as well as a stochastic term $B_{ij}(q^k_i, t^k)\Delta W^j_k$.

$\sigma-$Hamiltonian biomechanical system (2.673–2.674), with its $\sigma-$Hamiltonian phase–flow ϕ_{σ_t} (2.675), i.e., the functor `Stoch`[`Can`], represents our second, continual–discrete stochastic model for the biomechanics functor machine $\mathcal{F}[S]$ with the nonlinear system behavior $\mathcal{F}[r]$. In the next section we generalize this model once more to include fuzzy SN.

Fuzzy–Stochastic–Lie–Hamiltonian Functor

Generally, a *fuzzy differential equation* model (FDE–model, for short) is a symbolic description expressing a state of incomplete knowledge of the continuous world, and is thus an abstraction of an infinite set of ODEs models. Qualitative simulation (see [BK92]) predicts the set of possible behaviors consistent with a FDE model and an initial state. Specifically, as a FDE we consider an ordinary deterministic (i.e., *crisp*) differential equation (CDE) in which some of the parameters (i.e., coefficients) or initial conditions are *fuzzy numbers*, i.e., uncertain and represented in a *possibilistic* form. As a *solution* of a FDE we consider a time evolution of a *fuzzy region of uncertainty* in the system's phase–space, which corresponds to its the *possibility distribution*.

Recall that a fuzzy number is formally defined as a convex, normalized *fuzzy set* [DP80, Cox92, Cox94]. The concept of fuzzy numbers is an extension of the notion of real numbers: it encodes approximate quantitative knowledge. It is not probabilistic, but rather a possibilistic distribution. The mathematics of fuzzy numbers is founded on the *extension principle*, introduced by Zadeh [Yag87]. This principle gives a general method for extending standard mathematical concepts in order to deal with fuzzy quantities [DP80].

Let $\Phi : Y^1 \times Y^2 \times \cdots \times Y^n \to Z$ be a deterministic map such that $z = \Phi(y^1, y^2, \ldots, y^n)$ for all $z \in Z$, $y^i \in Y^i$. The extension principle allows us to induce from n input fuzzy sets $\bar{y}^i$ on Y^i an output fuzzy set $\bar{z}$ on Z through Φ given by

$$\mu_{\bar{z}}(t) = \sup_{t=\Phi(s^1,\ldots,s^n)} \min(\mu_{\bar{y}^1}(s^1), \ldots, \mu_{\bar{y}^n}(s^n)),$$

$$\text{or} \qquad \mu_{\bar{z}}(t) = 0 \qquad \text{if} \qquad \Phi^{-1}(t) = \emptyset,$$

where $\Phi^{-1}(t)$ denotes the inverse image of t and $\mu_{\bar{y}^i}$ is the membership function of $\bar{y}^i$, $(i = 1, \ldots, n)$.

The extension principle gives a method to calculate the fuzzy value of a fuzzy map but, in practice, its application is not feasible because of the infinite number of computations it would require. The simplest way of efficiently

applying the extension principle is in the form of iterative repetition of several crisp Hamiltonian simulations (see [Bon95, IS01, PI03, PI04]), within the range of included fuzzy SN.

Fuzzification of the crisp deterministic $\delta-$Hamiltonian biomechanical system (2.669–2.671) gives the fuzzified μ–Hamiltonian biomechanical system, namely δ–Hamiltonian biomechanical system with fuzzy SN, i.e., the *fuzzy transformation* `Fuzzy`[`Can`] of the canonical functor `Can`

$$\dot{q}^i = \frac{\partial H(q,p,\sigma)}{\partial p_i} + \frac{\partial R}{\partial p_i}, \tag{2.676}$$

$$\dot{p}_i = \bar{F}_i(q,p,\sigma) - \frac{\partial H(q,p,\sigma)}{\partial q^i} + \frac{\partial R}{\partial q^i}, \tag{2.677}$$

$$q^i(0) = \bar{q}_0^i, \qquad p_i(0) = \bar{p}_i^0, \qquad (i = 1, \ldots, N). \tag{2.678}$$

Here $\sigma = \sigma_\mu$ (with $\mu \geq 1$) denote fuzzy sets of conservative parameters (segment lengths, masses and moments of inertia), dissipative joint dampings and actuator parameters (amplitudes and frequencies), while the bar $\bar{(.)}$ over a variable (.) denotes the corresponding fuzzified variable.

In our example–case of symmetrical 3D load–lifting, the fuzzified $\mu-$Hamiltonian biomechanical system (2.676–2.678) becomes

$$\dot{q}^i = p_i \left\{ [\bar{J}_i]^{-1} + \left[\bar{m}_i \left(\sum_{j=1}^{i} \bar{L}_j \cos q^j \right)^2 \right]^{-1} \right\} + \frac{\partial R}{\partial p_i},$$

$$\dot{p}_i = \bar{F}_i(t, q^i, p_i, \{\sigma\}_\mu) - g \sum_{j=i}^{10-i} \bar{L}_j \bar{m}_j \sin q^j$$

$$- \sum_{j=i}^{10-i} \bar{L}_j \sin q^j p_i p_j \left[\bar{m}_i \left(\sum_{k=1}^{i} \bar{L}_k \cos q^k \right)^3 \right]^{-1} + \frac{\partial R}{\partial q^i}$$

$$q^i(0) = \bar{q}_0^i, \qquad p_i(0) = \bar{p}_i^0, \qquad (i = 1, \ldots, 9).$$

In this way, the crisp δ–Hamiltonian phase–flow ϕ_{δ_t} (2.672) extends into fuzzy–deterministic μ–Hamiltonian phase–flow ϕ_{μ_t}

$$\phi_{\mu_t} : G_1 \times T^*M \to T^*M : (\bar{p}_i^0, \bar{q}_0^i) \mapsto (p(t), q(t)),$$
$$(\phi_{\mu_t} \circ \phi_{\mu_s} = \phi_{\mu_{t+s}}, \quad \phi_{\mu_0} = \text{identity}).$$

Similarly, fuzzification of crisp stochastic $\sigma-$Hamiltonian biomechanical system (2.673–2.674) gives fuzzy–stochastic $[\mu\sigma]-$Hamiltonian biomechanical system, namely stochastic $\sigma-$Hamiltonian biomechanical system with fuzzy SN, i.e., the *fuzzy–stochastic transformation* `Fuzzy`[`Stoch`[`Can`]] of the canonical functor `Can`

$$dq^i = \left(\frac{\partial H(q,p,\sigma)}{\partial p_i} + \frac{\partial R}{\partial p_i} \right) dt, \tag{2.679}$$

$$dp_i = B_{ij}[q^i(t),t]\, dW^j(t) + \left(\bar{F}_i(q,p,\sigma) - \frac{\partial H(q,p,\sigma)}{\partial q^i} + \frac{\partial R}{\partial q^i} \right) dt, \tag{2.680}$$

$$q^i(0) = \bar{q}_0^i, \qquad p_i(0) = \bar{p}_i^0. \tag{2.681}$$

In our example–case of symmetrical 3D load–lifting, the velocity and force $[\mu\sigma]-$Hamiltonian biomechanics equations (2.679–2.680) become

$$dq^i = \left(p_i \left\{ [\bar{J}_i]^{-1} + \left[\bar{m}_i \left(\sum_{j=1}^{i} \bar{L}_j \cos q^j \right)^2 \right]^{-1} \right\} + \frac{\partial R}{\partial p_i} \right) dt,$$

$$dp_i = B_{ij}[q^i(t),t]\, dW^j(t) + \left(\bar{F}_i(t,q^i,p_i,\{\sigma\}_\mu) - g \sum_{j=i}^{10-i} \bar{L}_j \bar{m}_j \sin q^j \right.$$

$$\left. - \sum_{j=i}^{10-i} \bar{L}_j \sin q^j p_i p_j \left[\bar{m}_i \left(\sum_{k=1}^{i} \bar{L}_k \cos q^k \right)^3 \right]^{-1} + \frac{\partial R}{\partial q^i} \right) dt.$$

In this way, the crisp stochastic σ–Hamiltonian phase–flow ϕ_{σ_t} (2.675) extends into fuzzy–stochastic $[\mu\sigma]$–Hamiltonian phase–flow $\phi_{[\mu\sigma]_t}$

$$\phi_{[\mu\sigma]_t} : G_1 \times I^*M \to I^*M : (\bar{p}_i^0, \bar{q}_0^i) \mapsto (p(t), q(t)), \tag{2.682}$$
$$(\phi_{[\mu\sigma]_t} \circ \phi_{[\mu\sigma]_s} = \phi_{[\mu\sigma]_{t+s}}, \quad \phi_{[\mu\sigma]_0} = \text{identity}).$$

$[\mu\sigma]-$Hamiltonian biomechanical system (2.679–2.681), with its phase–flow $\phi_{[\mu\sigma]_t}$ (2.682), i.e., the functor `Fuzzy[Stoch[Can]]`, represents our final, continual–discrete and fuzzy–stochastic model for the biomechanics functor machine $\mathcal{F}[S]$ with the nonlinear system behavior $\mathcal{F}[r]$.

2.4.4 Biomechanical Topology

(Co)Chain Complexes in Biomechanics

In this section we present the category of (co)chain complexes, as used in modern biomechanics. The central concept in cohomology theory is the *category* $\mathbf{S}^\bullet(\mathbb{C})$ of *generalized cochain complexes* in an Abelian category $\mathbb{C}$ [Die88]. The *objects* of the category $\mathbf{S}^\bullet(\mathbb{C})$ are infinite sequences

$$A^\bullet : \cdots \longrightarrow A^{n-1} \xrightarrow{d^{n-1}} A^n \xrightarrow{d^n} A^{n+1} \longrightarrow \cdots$$

where, for each $n \in \mathbb{Z}$, A^n is an object of $\mathbb{C}$ and d^n a morphism of $\mathbb{C}$, with the conditions

$$d^{n-1} \circ d^n = 0$$

for every $n \in \mathbb{Z}$. When $A^n = 0$ for $n < 0$, one speaks of *cochain complexes*. The d^n are called *coboundary operators*.

The *morphisms* of the category $\mathbf{S}^\bullet(\mathbb{C})$ are sequences $f^\bullet = (f^n) : A^\bullet \to B^\bullet$ where, for each $n \in \mathbb{Z}$, $f^n : A^n \to B^n$ is a morphism of $\mathbb{C}$, and in the diagram

$$\begin{array}{ccccccccc} \cdots & \longrightarrow & A^{n-1} & \xrightarrow{d^{n-1}} & A^n & \xrightarrow{d^n} & A^{n+1} & \longrightarrow & \cdots \\ & & \downarrow{\scriptstyle f^{n-1}} & & \downarrow{\scriptstyle f^n} & & \downarrow{\scriptstyle f^{n+1}} & & \\ \cdots & \longrightarrow & B^{n-1} & \xrightarrow{d^{n-1}} & B^n & \xrightarrow{d^n} & B^{n+1} & \longrightarrow & \cdots \end{array} \tag{2.683}$$

all squares are commutative; one says the f^n commute with the coboundary operators. One has $\operatorname{Im} d^{n+1} \subset \operatorname{Ker} d^n \subset A^n$ for every $n \in \mathbb{Z}$; the quotient $H^n(A^\bullet) = \operatorname{Ker} d^n / \operatorname{Im} d^{n+1}$ is called the nth cohomology object of $A^\bullet$. From (2.683) it follows that there is a morphism

$$H^n(f^\bullet) : H^n(A^\bullet) \to H^n(B^\bullet)$$

deduced canonically from $f^\bullet$, and

$$(A^\bullet, f^\bullet) \Rightarrow (H^n(A^\bullet), H^n(f^\bullet))$$

is a *covariant functor* from $\mathbf{S}^\bullet(\mathbb{C})$ to $\mathbb{C}$.

The *cohomology exact sequence*: if three cochain complexes $A^\bullet, B^\bullet, C^\bullet$ are elements of a short exact sequence of morphisms

$$0 \longrightarrow A^\bullet \longrightarrow B^\bullet \longrightarrow C^\bullet \longrightarrow 0$$

then there exists an infinite sequence of canonically defined morphisms $d^n : H^n(C^\bullet) \to H^{n-1}(A^\bullet)$ such that the sequence

$$\cdots \longrightarrow H^n(A^\bullet) \longrightarrow H^n(B^\bullet) \longrightarrow H^n(C^\bullet) \longrightarrow H^{n-1}(A^\bullet) \longrightarrow \cdots$$

is *exact*, that is the *image* of each homomorphism in the sequence is exactly the *kernel* of the next one.

The *dual* to the category $\mathbf{S}^\bullet(\mathbb{C})$ is the *category* of $\mathbf{S}_\bullet(\mathbb{C})$ of *generalized chain complexes*. Its objects and morphisms are get by formal inversion of all arrows and lowering all indices.

Biomechanical (Co)Homologies

Let $\mathcal{M}^\bullet$ denote the Abelian category of cochains, (i.e., p–forms) on the biomechanical configuration manifold M (see Figure 2.22). When $\mathcal{C} = \mathcal{M}^\bullet$, we have the category $\mathcal{S}^\bullet(\mathcal{M}^\bullet)$ of generalized cochain complexes $A^\bullet$ in $\mathcal{M}^\bullet$, and if $A' = 0$ for $n < 0$ we have a subcategory $\mathcal{S}^\bullet_{\mathcal{DR}}(\mathcal{M}^\bullet)$ of the *de Rham differential complexes* in $\mathcal{M}^\bullet$

$$A^{\bullet}_{DR}: 0 \to \Omega^0(M) \xrightarrow{d} \Omega^1(M) \xrightarrow{d} \Omega^2(M) \cdots \xrightarrow{d} \Omega^n(M) \xrightarrow{d} \cdots .$$

Here $A' = \Omega^n(M)$ is the vector space over $\mathbb{R}$ of all p–forms ω on M (for $p = 0$ the smooth functions on M) and $d_n = d : \Omega^{n-1}(M) \to \Omega^n(M)$ is the exterior differential. A form $\omega \in \Omega^n(M)$ such that $d\omega = 0$ is a closed form or n–cocycle. A form $\omega \in \Omega^n(M)$ such that $\omega = d\theta$, where $\theta \in \Omega^{n-1}(M)$, is an exact form or n–coboundary. Let $Z^n(M) = \operatorname{Ker} d$ (resp. $B^n(M) = \operatorname{Im} d$ denote a real vector space of cocycles (resp. coboundaries) of degree n. Since $d_{n+1} \circ d_n = d^2 = 0$, we have $B^n(M) \subset Z^n(M)$. The quotient vector space

$$H^n_{DR}(M) = \operatorname{Ker} d / \operatorname{Im} d = Z^n(M)/B^n(M)$$

is the *de Rham cohomology group*. The elements of $H^n_{DR}(M)$ represent equivalence sets of cocycles. Two cocycles ω_1, ω_2 belong to the same equivalence set, or are cohomologous (written $\omega_1 \sim \omega_2$) iff they differ by a coboundary $\omega_1 - \omega_2 = d\theta$. The de Rham's cohomology class of any form $\omega \in \Omega^n(M)$ is $[\omega] \in H^n_{DR}(M)$. The de Rham differential complex (1) can be considered as a system of second–order DEs $d^2\theta = 0$, $\theta \in \Omega^{n-1}(M)$ having a solution represented by $Z^n(M) = \operatorname{Ker} d$.

Analogously let $\mathcal{M}_{\bullet}$ denote the Abelian category of chains on the configuration manifold M. When $\mathcal{C} = \mathcal{M}_{\bullet}$, we have the category $\mathcal{S}_{\bullet}(\mathcal{M}_{\bullet})$ of generalized chain complexes $A_{\bullet}$ in $\mathcal{M}_{\bullet}$, and if $A_n = 0$ for $n < 0$ we have a subcategory $\mathcal{S}^{\mathcal{C}}_{\bullet}(\mathcal{M}_{\bullet})$ of chain complexes in $\mathcal{M}_{\bullet}$

$$A_{\bullet}: 0 \leftarrow C^0(M) \xleftarrow{\partial} C^1(M) \xleftarrow{\partial} C^2(M) \cdots \xleftarrow{\partial} C^n(M) \xleftarrow{\partial} \cdots .$$

Here $A_n = C^n(M)$ is the vector space over $\mathbb{R}$ of all finite chains C on the manifold M and $\partial_n = \partial : C^{n+1}(M) \to C^n(M)$. A finite chain C such that $\partial C = 0$ is an $n-$cycle. A finite chain C such that $C = \partial B$ is an $n-$boundary. Let $Z_n(M) = \operatorname{Ker} \partial$ (resp. $B_n(M) = \operatorname{Im} \partial$) denote a real vector space of cycles (resp. boundaries) of degree n. Since $\partial_{n+1} \circ \partial_n = \partial^2 = 0$, we have $B_n(M) \subset Z_n(M)$. The quotient vector space

$$H^C_n(M) = \operatorname{Ker} \partial / \operatorname{Im} \partial = Z_n(M)/B_n(M)$$

is the $n-$homology group. The elements of $H^C_n(M)$ are equivalence sets of cycles. Two cycles C_1, C_2 belong to the same equivalence set, or are homologous (written $C_1 \sim C_2$), iff they differ by a boundary $C_1 - C_2 = \partial B$). The homology class of a finite chain $C \in C^n(M)$ is $[C] \in H^C_n(M)$.

The dimension of the $n-$cohomology (resp. $n-$homology) group equals the nth Betti number b^n (resp. b_n) of the manifold M. *Poincaré lemma* says that on an open set $U \in M$ diffeomorphic to $\mathbb{R}^N$, all closed forms (cycles) of degree $p \geq 1$ are exact (boundaries). That is, the Betti numbers satisfy $b^p = 0$ (resp. $b = 0$), for $p = 1, \ldots, n$.

The *de Rham theorem* states the following. The map Φ: $H_n \times H^n \to \mathbb{R}$ given by $([C], [\omega]) \to \langle C, \omega \rangle$ for $C \in Z_n, \omega \in Z^n$ is a bilinear nondegenerate map which establishes the duality of the groups (vector spaces) H_n and H^n and the equality $b_n = b^n$.

Configuration Manifold Reduction and its Euler Characteristic

Recall (see subsection (2.4.1) above), that for the purpose of high–level control, the rotational biomechanical configuration manifold M (Figure 2.21), could be first, reduced to an $n-torus$, and second, transformed into an $n-cube$ 'hyper–joystick', using the following topological techniques (see [IP01b, Iva02, Iva05a]).

Let S^1 denote the constrained unit circle in the complex plane, which is an Abelian Lie group. Firstly, we propose two reduction homeomorphisms, using the noncommutative semidirect product '$\triangleright$' of the constrained $SO(2)-$groups:

$$SO(3) \gtrsim SO(2) \triangleright SO(2) \triangleright SO(2), \qquad \text{and} \qquad SO(2) \approx S^1.$$

Next, let I^n be the unit cube $[0,1]^n$ in $\mathbb{R}^n$ and '$\sim$' an equivalence relation on $\mathbb{R}^n$ get by 'gluing' together the opposite sides of I^n, preserving their orientation. Therefore, the manifold M can be represented as the quotient space of $\mathbb{R}^n$ by the space of the integral lattice points in $\mathbb{R}^n$, that is an oriented and constrained nD torus T^n:

$$\mathbb{R}^n/Z^n = I^n/\sim \approx \prod_{i=1}^{n} S_i^1 \equiv \{(q^i,\, i=1,\dots,N) : \text{mod } 2\pi\} = T^n. \qquad (2.684)$$

Now, using the *de Rham theorem* and the *homotopy axiom* for the de Rham cohomologies, we can calculate the *Euler–Poincaré characteristics* for T^n as well as for its two bundles, TT^n and T^*T^n, as (see [Iva02, Iva05a])

$$\chi(T^n, TT^n) = \sum_{p=1}^{n} (-1)^p b_p\,, \qquad \text{where } b_p \text{ are the } \textit{Betti numbers} \text{ defined as}$$

$$b^0 = 1,\;\; b^1 = n, \dots\; b^p = \binom{n}{p}, \dots b^{n-1} = n,\;\; b^n = 1,\;\; (p = 0, ..., n).$$

Morse Theory in Biomechanics

Morse Geometry of a Biomechanical Manifold

Recall that on any smooth manifold M there exist many Riemannian metrics g. Each of these metrics is *locally defined* in a particular point $q \in M$ as a symmetric $(0,2)$ tensor–field such that $g|_q : T_qM \times T_qM \to \mathbb{R}$ is a positively defined inner product for each point $q \in M$. In an open local chart $U \in M$ containing the point q, this metric is given as $g|_q \mapsto g_{ij}(q)\, dq^i dq^j$. With each metric $g|_q$ there is associated a *local geodesic* on M.

Now, two main *global geodesics problems* on the biomechanical configuration manifold M with the Riemannian metrics g, can be formulated as follows (compare with subsection 1.2.9 above):

1. *Is there a minimal geodesic* $\gamma_0(t)$ *between two points* A and B on M? In other words, does an arc of geodesic $\gamma_0(t)$ with extremities A, B actually have minimum length among all rectifiable curves $\gamma(t) = (q^i(t), p_i(t))$ joining A and B?
2. *How many geodesic arcs are there joining two points* A and B on M?

Locally these problems have a *complete answer*: each point of the biomechanics manifold M has an open neighborhood V such that for any two distinct points A, B of V *there is* exactly *one* arc of a geodesic *contained* in V and joining A and B, and it is the *unique minimal geodesic* between A and B.

Recall (see subsection (1.2.9) above), that seven decades ago, Morse considered the set $\Omega = \Omega(M; A, B)$ of *piecewise smooth paths* on a Riemannian manifold M having fixed extremities A, B, defined as continuous maps $\gamma : [0, 1] \to M$ such that $\gamma(0) = A$, $\gamma(1) = B$, and there were a finite number of points

$$t_0 = 0 < t_1 < t_2 < \cdots < t_{m-1} < t_m = 1, \tag{2.685}$$

such that in every *closed interval* $[t_i, t_{i+1}]$, γ was a $C^\infty-$function. The parametrization was always chosen such that for $t_j \leq t \leq t_{j+1}$,

$$t-t_j = \frac{t_{j+1} - t_j}{l_j} \int_{t_j}^{t} \parallel \frac{d\gamma}{du} \parallel \, du, \qquad \text{with} \qquad l_j = \int_{t_j}^{t_{j+1}} \parallel \frac{d\gamma}{du} \parallel \, du. \tag{2.686}$$

In other words, $t - t_j$ was proportional to the length of the image of $[t_j, t]$ by γ. Then

$$L(\gamma) = \sum_{j=0}^{m} l_j,$$

the length of γ, was a function of γ in Ω. A minimal arc from A to B should be a path γ for which $L(\gamma)$ is *minimum* in Ω, and a geodesic arc from A to B should be a path that is a 'critical point' for the function L. This at first has no meaning, since Ω is not a differential manifold; the whole of Morse's theory consists in showing that it is possible to substitute for Ω genuine differential manifolds to which his results on critical points can be applied ([Mor34]).

To study the geodesics joining two points A, B it is convenient, instead of working with the length $L(\gamma)$, to work with the *energy of a path* $\gamma : [A, B] \to M$, defined by ([Die88])

$$E_A^B(\gamma) = \int_A^B \parallel \frac{d\gamma}{du} \parallel^2 \, du. \tag{2.687}$$

With the chosen parametrization (2.686), $E(\gamma) = (B - A)L(\gamma)^2$, and the extremals of E are again the geodesics, but the computations are easier with E.

Morse theory can be divided into several steps (see [Mil63]).

Step 1 is essentially a presentation of the classical Lagrangian method that brings to light the analogy with the critical points of a $C^\infty-$ function on M. No topology is put on Ω; a *variation* of a path $\gamma \in \Omega$ is a continuous map α into M, defined in a product $]-\varepsilon, \varepsilon[\times [0,1]$ with the following properties:

1. $\alpha(0,t) = \gamma(t)$;
2. $\alpha(u,0) = A$, $\alpha(u,1) = B$ for $-\varepsilon < u < \varepsilon$; and
3. There is a decomposition (2.685) such that α is C^∞ in each set

$$]-\varepsilon, \varepsilon[\times [t_i, t_{i+1}].$$

A *variation vector–field* $t \mapsto W(t)$ is associated to each variation α, where $W(t)$ is a tangent vector in the tangent space $T_{\gamma(t)}M$ to M, defined by

$$W(t) = \partial_u \alpha(0,t). \tag{2.688}$$

It is a continuous map of $[0,1]$ into the tangent bundle TM, *smooth* in each interval $[t_i, t_{i+1}]$. These maps are the substitute for the *tangent vectors* at the *point* γ; they form an infinite–dimensional vector space written $T\Omega(\gamma)$.

More generally the interval $]-\varepsilon, \varepsilon[$ can be replaced in the definition of a variation by a neighborhood of 0 in some $\mathbb{R}^n$, defining an $n-$*parameter variation.*

A *critical path* $\gamma_0 \in \Omega$ for a function $F : \Omega \longrightarrow \mathbb{R}$ is defined by the condition that for every variation α of γ_0 the function

$$u \mapsto F(\alpha(u, \cdot))$$

is derivable for $u = 0$ and its derivative is 0.

Step 2 is a modern presentation of the formulas of Riemannian geometry, giving the *first variation* and *second variation* of the energy (2.687) of a path $\gamma_0 \in \Omega$, which form the basis of Jacobi results.

First consider an arbitrary path $\omega_0 \in \Omega$, its *velocity* $\dot{\omega}(t) = d\omega/dt$, and its *acceleration* in the Riemannian sense

$$\overline{\ddot{\omega}}(t) = \nabla_t \dot{\omega}(t),$$

where ∇_t denotes the Bianchi covariant derivative. They belong to $T_{\omega(t)}M$ for each $t \in [0,1]$, are defined and continuous in each interval $[t_i, t_{i+1}]$ in which ω is smooth, and have limits at both extremities. Now let α be a variation of ω and $t \mapsto W(t)$ be the corresponding variation vector–field (2.688). The *first variation formula* gives the first derivative

$$\frac{1}{2}\frac{d}{du}E(\alpha(u,\cdot))|_{u=0} = -\sum_i (W(t_i)|\dot{\omega}(t_i+) - \dot{\omega}(t_i-)) - \int_0^1 (W(t)|\overline{\ddot{\omega}}(t))\, dt,$$

where $(x|y)$ denotes the scalar product of two vectors in a tangent space. It follows from this formula that $\gamma_0 \in \Omega$ is a critical path for E iff γ is a *geodesic.*

Next, fix such a geodesic γ and consider a two–parameter variation:

$$\alpha : U \times [0,1] \to M,$$

where U is a neighborhood of 0 in $\mathbb{R}^2$, so that

$$\alpha(0,0,t) = \gamma(t), \qquad \partial_{u_1}\alpha(0,0,t) = W_1(t), \qquad \partial_{u_2}\alpha(0,0,t) = W_2(t),$$

in which W_1 and W_2 are in $T\Omega(\gamma)$. The *second variation formula* gives the mixed second derivative

$$\frac{1}{2}\frac{\partial^2}{\partial u_1 \partial u_2} E(\alpha(u_1,u_2,\cdot)))|_{(0,0)} = -\sum_i (W_2(t_i)|\nabla_t W_1(t_i+) - \nabla_t W_1(t_i-)) - \int_0^1 (W_2(t)|\nabla_t^2 W_1(t) + R(V(t) \wedge W_1(t)) \cdot V(t))\, dt, \qquad (2.689)$$

where $Z \mapsto R(X \wedge Y) \cdot Z$ is the curvature of the Levi–Civita connection. The l.h.s of (2.689) is thus a *bilinear symmetric form*

$$(W_1, W_2) \mapsto E_{**}(W_1, W_2)$$

on the product $T\Omega(\gamma) \times T\Omega(\gamma)$. For a one–parameter variation α

$$E_{**}(W,W) = \frac{1}{2}\frac{d^2}{du^2} E(\alpha(u,\cdot))|_{u=0},$$

from which it follows that if γ is a *minimal* geodesic in Ω, $E_{**}(W,W) \geq 0$ in $T\Omega(\gamma)$. As usual, we shall speak of E_{**} indifferently as a symmetric bilinear form or as a quadratic form $W \mapsto E_{**}(W,W)$.

Formula (2.689) naturally leads to the junction with Jacobi work (see [Die88]): consider the smooth vector–fields $t \mapsto J(t)$ along $\gamma \in M$, satisfying the equation

$$\nabla_t^2 J(t) + R(V(t) \wedge J(t)) \cdot V(t) = 0 \qquad \text{for } 0 \leq t \leq 1. \qquad (2.690)$$

With respect to a frame along γ moving by parallel translation on M this relation is equivalent to a system of n linear homogeneous ODEs of order 2 with $C^\infty-$coefficients; the solutions J of (2.690) are called the *Jacobi fields* along γ and form a vector space of dimension $2n$. If for a value $a \in]0,1]$ of the parameter t there exists a Jacobi field along γ that is not identically 0 but *vanishes for* $t = 0$ *and* $t = a$, then the points $A = \gamma(0)$ and $r = \gamma(a)$ are conjugate along γ with a *multiplicity* equal to the dimension of the vector space of Jacobi fields vanishing for $t = 0$ and $t = a$.

Jacobi fields on the biomechanical configuration manifold M may also be defined as variation vector–fields for *geodesic variations* of the path $\gamma \in M$: they are $C^\infty-$maps

$$\alpha :]-\varepsilon, \varepsilon[\times [0,1] \to M,$$

such that for any $u \in]-\varepsilon, \varepsilon[\,, t \mapsto \alpha(u,t)$ is a geodesic and $\alpha(0,t) = \gamma(t)$.

It can be proved that the Jacobi fields along $\gamma \in M$ that vanish at A and B (hence belong to $T\Omega(\gamma)$) are exactly the vector–fields $J \in T\Omega(\gamma)$ such that

$$E_{**}(J, W) = 0$$

for every $W \in T\Omega(\gamma)$. Although $T\Omega(\gamma)$ is infinite–dimensional, the form E_{**} is again called *degenerate* if the vector space of the Jacobi fields vanishing at A and B is note reduced to 0 and the dimension of that vector space is called the *nullity* of E_{**}. Therefore, E_{**} is thus degenerate iff A and B are conjugate along γ and the nullity of E_{**} is the multiplicity of B.

Step 3 is the beginning of Morse's contributions (see [Mil63]). He first considered a *fixed* geodesic $\gamma : [0,1] \to M$ with extremities $A = \gamma(0)$, $B = \gamma(1)$ and the bilinear symmetric form $E_{**} : T\Omega(\gamma) \times T\Omega(\gamma) \to \mathbb{R}$. By analogy with the finite–dimensional quadratic form, the *index* of E_{**} is defined as the maximum dimension of a vector subspace of $T\Omega(\gamma)$ in which E_{**} is *strictly negative* (i.e., nondegenerate and taking values $E_{**}(W,W) < 0$ except for $W = 0$). Morse's central result gives the value of the index of E_{**} and is known as the *index theorem*.

Suppose a subdivision (2.685) is chosen such that each arc $\gamma([t_{i-1}, t_i])$ is contained in an open set $U_i \subset M$ such that any two points of U_i are joined by a unique geodesic arc contained in U_i that is *minimal*; $\gamma([t_{i-1}, t_i])$ is such an arc. In the infinite–dimensional vector space $T\Omega(\gamma)$, consider the two vector subspaces:

1. $T\Omega(\gamma; t_0, t_1, \cdots, t_m)$ consisting of all continuous vector–fields $t \mapsto W(t)$ along γ, vanishing for $t = 0$ and $t = 1$, such that each restriction $W|[t_{i-1}, t_i]$ is a *Jacobi field* (hence smooth) along $\gamma([t_{i-1}, t_i])$; that subspace is finite–dimensional;
2. T' consisting of the vector–fields $t \mapsto W(t)$ along γ, such that $W(t_0) = 0$, $W(t_1) = 0, \cdots, W(t_m) = 0$.

$T\Omega(\gamma)$ is then the *direct sum* $T\Omega(\gamma; t_0, t_1, \cdots, t_m) \oplus T'$; these two subspaces are orthogonal for the bilinear form E_{**}, and E_{**} is *strictly positive* in T', so that the index of E_{**} is equal to the index of its restriction to the subspace $T\Omega(\gamma; t_0, t_1, \cdots, t_m)$.

To calculate the nullity and index of E_{**}, due to this decomposition, apply their definitions either to vector subspaces of $T\Omega(\gamma)$ or to vector subspaces of $T\Omega(\gamma;\ t_0, t_1, \cdots, t_m)$. The computation of the index of E_{**} is done by considering the geodesic arc $\gamma_\tau : [0, \tau] \to M$, the restriction of γ to $[0, \tau]$, and its energy

$$E(\gamma_\tau) = \tau \int_0^\tau \parallel \frac{d\gamma}{du} \parallel^2 \, du.$$

E^τ_{**} is the corresponding quadratic form on $T\Omega(\gamma_\tau)$, and $\lambda(\tau)$ is its index; one studies the variation of $\lambda(\tau)$ when *tau* varies from 0 to 1, and $\lambda(1)$ is the index of E_{**}.

The index theorem says: *the index of E_{**} is the sum of the multiplicities of the points conjugate to A along B and distinct from B.*

We have seen that the dimension of $T\Omega(\gamma; t_0, t_1, \cdots, t_m)$ is finite; it follows that the index of E_{**} is always *finite*, and therefore the number of points conjugate to A along γ is also *finite*.

Step 4 of Morse theory introduces a *topology* on the set $\Omega = \Omega(M; A, B)$. On the biomechanical configuration manifold M the usual topology can be defined by a *distance* $\rho(A, B)$, the g.l.b. of the lengths of all piecewise smooth paths joining A and B. For any pair of paths ω_1, ω_2 in $\Omega(M; A, B)$, consider the function $d(\omega_1, \omega_2) \in M$

$$d(\omega_1, \omega_2) = \sup_{0 \leqslant t \leqslant 1} \rho(\omega_1(t), \omega_2(t)) + \sqrt{\int_0^1 (\dot{s}_1 - \dot{s}_2)^2 \, dt},$$

where $s_1(t)$ (resp. $s_2(t)$) is the length of the path $\tau \mapsto \omega_1(\tau)$ (resp. $\tau \mapsto \omega_2(\tau)$) defined in $[0, t]$. This distance on Ω such that the function $\omega \mapsto E_A^B(\omega)$ is *continuous* for that distance.

Morse Homology of a Biomechanical Manifold

Morse Functions and Boundary Operators. Let $f : M \to \mathbb{R}$ represents a $C^\infty-$function on the biomechanical configuration manifold M. Recall that $z = (q, p) \in M$ is the *critical point* of f if $df(z) \equiv df[(q, p)] = 0$. In local coordinates $(x^1, ..., x^n) = (q^1, ..., q^n, p_1, ..., p_n)$ in a neighborhood of z, this means $\frac{\partial f}{\partial x^i}(z) = 0$ for $i = 1, ..., n$. The Hessian of f at a critical point z defines a symmetric bilinear form $\nabla df(z) = d^2 f(z)$ on $T_z M$, in local coordinates $(x^1, ..., x^n)$ represented by the matrix $\left(\frac{\partial^2 f}{\partial x^i \partial x^j}\right)$. Index and nullity of this matrix are called index and nullity of the critical point z of f.

Now, we assume that all critical points $z_1, ..., z_n$ of $f \in M$ are nondegenerate in the sense that the Hessians $d^2 f(z_i)$, $i = 1, ..., m$, have maximal rank. Let z be such a critical point of f of *Morse index* s (= number of negative eigenvalues of $d^2 f(z_i)$, counted with multiplicity). The eigenvectors corresponding to these negative eigenvalues then span a subspace $V_z \subset T_z M$ of dimension s. We choose an orthonormal basis $e_1, ..., e_s$ of V_z w.r.t. the Riemannian metric g on M (induced by the system's kinetic energy), with dual basis $dx^1, ..., dx^s$. This basis then defines an orientation of V_z which we may also represent by the s–form $dx^1 \wedge ... \wedge dx^s$. We now let z' be another critical point of f, of Morse index $s - 1$. We consider paths $\gamma(t)$ of the steepest descent of f from z to z', i.e., integral curves of the vector–field $-\nabla f(\gamma)$. Thus $\gamma(t)$ defines the *gradient flow* of f

$$\dot{\gamma}(t) = -\nabla f\left(\gamma(t)\right), \qquad \text{with} \qquad \begin{cases} \lim_{t \to -\infty} \gamma(t) = z, \\ \lim_{t \to \infty} \gamma(t) = z' \end{cases}. \tag{2.691}$$

A path $\gamma(t)$ obviously depends on the Riemannian metric g on M as

$$\nabla f = g^{ij}\,\partial_{x^i} f\,\partial_{x^j} f.$$

From [Sma60, Sma67] it follows that for a generic metric g, the Hessian $\nabla df(y)$ has only nondegenerate eigenvalues. Having a metric g induced by the system's kinetic energy, we let $\tilde{V}_y \subset T_yM$ be the space spanned by the eigenvectors corresponding to the $s-1$ lowest eigenvalues. Since z' has Morse index $s-1$, $\nabla df(z') = d^2 f(z')$ has precisely $s-1$ negative eigenvalues. Therefore, $\tilde{V}_{z'} \equiv \lim_{t\to\infty} \tilde{V}_{\gamma(t)} = V_{z'}$, while the unit tangent vector of γ at z', i.e., $\lim_{t\to\infty} \frac{\dot\gamma(t)}{\|\dot\gamma(t)\|}$, lies in the space of directions corresponding to positive eigenvalues and is thus orthogonal to $V_{z'}$. Likewise, the unit tangent vector v_z of γ at z, while contained in V_z, is orthogonal to $\tilde{V}_z$, because it corresponds to the largest one among the s negative eigenvalues of $d^2 f(z)$. Taking the interior product $i(v_z)\,dx^1 \wedge \ldots \wedge dx^s$ defines an orientation of $\tilde{V}_z$. Since $\tilde{V}_y$ depends smoothly on y, we may transport the orientation of $\tilde{V}_z$ to $\tilde{V}_{z'}$ along γ. We then define $n_\gamma = +1$ or -1, depending on whether this orientation of $\tilde{V}_{z'}$ coincides with the chosen orientation of $V_{z'}$ or not, and further define $n(z,z') = \sum_\gamma n_\gamma$, where the sum is taken over all such paths γ of the steepest descent from p to p'.

Now, let M^s be the *set of critical points* of f of *Morse index* s, and let H_f^s be the vector space over $\mathbb{R}$ spanned by the elements of M^s. We define a *boundary operator*

$$\begin{aligned} \delta \;:\; H_f^{s-1} &\to H_f^s, \qquad \text{by putting, for } z' \in M^{s-1},\\ \delta(z') &= \sum_{n\in M^s} n(z',z)\,z, \quad \text{and extending } \delta \text{ by linearity.} \end{aligned}$$

This operator satisfies $\delta^2 = 0$ and therefore defines a cohomology theory. Using *Conley's continuation principle*, Floer [Flo88] showed that the resulting cohomology theories resulting from different choices of f are canonically isomorphic.

In his QFT–based rewriting the Morse topology, Ed Witten [Wit82] considered also the operators:

$$\begin{aligned} d_t \;=\; e^{-tf} d e^{tf}, \qquad &\text{their adjoints}: \quad d_t^* = e^{tf} d e^{-tf},\\ &\text{as well as their Laplacian:} \quad \Delta_t = d_t d_t^* + d_t^* d_t. \end{aligned}$$

For $t=0$, Δ_0 is the standard *Hodge–de Rham Laplacian*, whereas for $t\to\infty$, one has the following expansion

$$\Delta_t = dd^* + d^*d + t^2\,\|df\|^2 + t\sum_{k,j} \frac{\partial^2 h}{\partial x^k \partial x^j}[i\,\partial_{x^k}, dx^j],$$

where $(\partial_{x^k})_{k=1,\ldots,n}$ is an orthonormal frame at the point under consideration. This becomes very large for $t\to\infty$, except at the critical points of f, i.e., where $df = 0$. Therefore, the eigenvalues of Δ_t will concentrate near the

critical points of f for $t \to \infty$, and we get an *interpolation* between de Rham cohomology and Morse cohomology.

Morse Homology on M. Now, following [Mil99, IP05b], for any Morse function f on the configuration manifold M we denote by $\mathrm{Crit}_p(f)$ the set of its critical points of index p and define $C_p(f)$ as a free Abelian group generated by $\mathrm{Crit}_p(f)$. Consider the gradient flow generated by (2.691). Denote by $\mathcal{M}_{f,g}(M)$ the set of all $\gamma : \mathbb{R} \to M$ satisfying (2.691) such that

$$\int_{-\infty}^{+\infty} \left| \frac{d\gamma}{dt} \right|^2 dt < \infty.$$

The spaces

$$\mathcal{M}_{f,g}(x^-, x^+) = \{\gamma \in \mathcal{M}_{f,g}(M) \mid \gamma(t) \to x^{\pm} \text{ as } t \to \pm\infty\}$$

are smooth manifolds of dimension $m(x^+) - m(x^-)$, where $m(x)$ denotes the Morse index of a critical point x. Note that

$$\mathcal{M}_{f,g}(x, y) \cong W_g^u(x) \cap W_g^s(y),$$

where $W_g^s(y)$ and $W_g^u(x)$ are the stable and unstable manifolds of the gradient flow (2.691). For generic g the intersection above is transverse (Morse–Smale condition). The group $\mathbb{R}$ acts on $\mathcal{M}_{f,g}(x, y)$ by $\gamma \mapsto \gamma(\cdot + t)$. We denote

$$\widehat{\mathcal{M}}_{f,g}(x, y) = \mathcal{M}_{f,g}(x, y)/\mathbb{R}.$$

The manifolds $\widehat{\mathcal{M}}_{f,g}(x, y)$ can be given a coherent orientation σ (see [Sch93]).

Now, we can define the boundary operator, as

$$\partial : C_p(f) \to C_{p-1}(f), \qquad \partial x = \sum_{y \in \mathrm{Crit}_{p-1}(f)} n(x, y) y,$$

where $n(x, y)$ is the number of points in 0D manifold $\widehat{\mathcal{M}}_{f,g}(x, y)$ counted with the sign with respect to the orientation σ. The proof of $\partial \circ \partial = 0$ is based on gluing and cobordism arguments [Sch93]. Now Morse homology groups are defined by

$$H_p^{\mathrm{Morse}}(f) = \mathrm{Ker}(\partial)/\mathrm{Im}(\partial).$$

For generic choices of Morse functions f_1 and f_2 the groups $H_p(f_1)$ and $H_p(f_2)$ are isomorphic. Furthermore, they are isomorphic to the singular homology group of M, i.e.,

$$H_p^{\mathrm{Morse}}(f) \cong H_p^{\mathrm{sing}}(M),$$

for generic f [Mil65].

The construction of isomorphism is given (see [Mil99, IP05b]) as

$$h_{\alpha\beta} : H_p(f^\alpha) \to H_p(f^\beta), \tag{2.692}$$

for generic Morse functions f^α, f^β. Consider the 'connecting trajectories', i.e., the solutions of non–autonomous equation

$$\dot\gamma = -\nabla f_t^{\alpha\beta}, \tag{2.693}$$

where $f_t^{\alpha\beta}$ is a homotopy connecting f^α and f^β such that for some $R>0$

$$f_t^{\alpha\beta} \equiv \begin{cases} f^\alpha \text{ for } t \le -R \\ f^\beta \text{ for } t \ge R \end{cases}.$$

For $x^\alpha \in \mathrm{Crit}_p(f^\alpha)$ and $x^\beta \in \mathrm{Crit}_p(f^\beta)$ denote

$$\mathcal{M}_{f^{\alpha\beta},g}(x^\alpha, x^\beta) = \{\gamma \ : \ \gamma \text{ satisfies (2.693) and } \lim_{t\to-\infty}\gamma = x^\alpha,\ \lim_{t\to\infty}\gamma = x^\beta\}.$$

As before, $\mathcal{M}_{f^{\alpha\beta},g}$ is a smooth finite–dimensional manifold. Now, define

$$\begin{aligned} (h_{\alpha\beta})_\sharp \ &: \ C_p(f^\alpha) \to C_p(f^\beta), \qquad \text{by} \\ (h_{\alpha\beta})_\sharp x^\alpha &= \sum_{x^\beta \in \mathrm{Crit}_p(f^\beta)} n(x^\alpha, x^\beta) x^\beta, \text{ for } x^\alpha \in \mathrm{Crit}_p(f^\alpha), \end{aligned}$$

where $n(x^\alpha, x^\beta)$ is the algebraic number of points in 0D manifold $\mathcal{M}_{f^{\alpha\beta},g}(x^\alpha, x^\beta)$ counted with the signs defined by the orientation of $\mathcal{M}_{f^{\alpha\beta},g}$. Homomorphisms $(h_{\alpha\beta})_\sharp$ commute with ∂ and thus define the homomorphisms $h_{\alpha\beta}$ in homology which, in addition, satisfy $h_{\alpha\beta}\circ h_{\beta\gamma} = h_{\alpha\gamma}$.

Now, if we fix a Morse function $f : M \to \mathbb{R}$ instead of a metric g, we establish the isomorphism (see [Mil99, IP05b])

$$h_{\alpha\beta} : H_p(g^\alpha, f) \to H_p(g^\beta, f)$$

between the two Morse homology groups defined by means of two generic metrics g^α and g^β in a similar way, by considering the 'connecting trajectories',

$$\dot\gamma = -\nabla^{g_t^{\alpha\beta}} f. \tag{2.694}$$

Here $g_t^{\alpha\beta}$ is a homotopy connecting g^α and g^β such that for some $R>0$

$$g_t^{\alpha\beta} \equiv \begin{cases} g^\alpha \text{ for } t \le -R, \\ g^\beta \text{ for } t \ge R, \end{cases}$$

and ∇^g is a gradient defined by metric g.

Note that f is decreasing along the trajectories solving autonomous gradient equation (2.691). Therefore, the boundary operator ∂ preserves the downward filtration given by level sets of f. In other words, if we denote

$$\begin{aligned} \mathrm{Crit}_p^\lambda(f) &= \mathrm{Crit}_p(f) \cap f^{-1}((-\infty, \lambda]), \qquad \text{and} \\ C_p^\lambda(f) &= \text{ free Abelian group generated by } \mathrm{Crit}_p^\lambda(f), \end{aligned}$$

then the boundary operator ∂ restricts to $\partial^\lambda : C_p^\lambda(f) \to C_{p-1}^\lambda(f)$. Obviously, $\partial^\lambda \circ \partial^\lambda = 0$, thus we can define the relative Morse homology groups

$$H_p^\lambda(f) = \mathrm{Ker}(\partial^\lambda)/\mathrm{Im}(\partial^\lambda).$$

Following the standard algebraic construction, we define (relative) Morse cohomology. We set

$$\begin{aligned} C_\lambda^p(f) &= \mathrm{Hom}(C_p^\lambda(f), \mathbb{Z}), \qquad \text{and} \\ \delta^\lambda &: C_\lambda^p(f) \to C_\lambda^{p+1}(f), \quad \langle \delta^\lambda a, x\rangle = \langle a, \partial^\lambda x\rangle \end{aligned}$$

and define

$$H_\lambda^p(f) = \mathrm{Ker}(\delta^\lambda)/\mathrm{Im}(\delta^\lambda).$$

Since $\mathrm{Crit}_p(f)$ is finite, we have $H_p^\lambda(f) = H_p(f)$ and $H_\lambda^p(f) = H^p(f)$.

Hodge–De Rham Theory in Biomechanics

Hodge Laplacian

A single biomechanical configuration manifold M can be equipped with many different Riemannian metrics g in local coordinates (apart from the one generated by its kinetic energy)

$$g = g_{ij}(u^1, u^2, ..., u^n)\, du^i\, du^j.$$

Beltrami had shown that it is always possible for such a metric to define an operator (depending on the metric) that generalizes the usual Laplacian on $\mathbb{R}^n$ and therefore induces the notion of harmonic functions on the Riemannian manifold [BM82].

Hodge theory was described by H. Weyl as 'one of the landmarks in the history of mathematics in the 20th Century'. Hodge showed that it was possible to define a notion of *harmonic exterior differential form*: the metric g on M canonically defines a metric on the tangent bundle TM, hence also, by standard multilinear algebra, a metric on any bundle of tensors on M. In particular, let $(\alpha, \beta) \mapsto g_p(\alpha, \beta)$ be the positive nondegenerate symmetric bilinear form defined on the vector space of $p-$forms on M. As M is orientable, this defines a *duality* between $p-$forms and $(n-p)-$forms: to each $p-$form α is associated a $(n-p)-$form $*\alpha$, defined by the linear Hodge star operator $*$ (see subsection 1.2.5), characterized by the relations

$$\beta \wedge (*\alpha) = g_p(\alpha, \beta)\, v, \qquad ** \alpha = (-1)^{p(n-p)}\alpha,$$

for all $p-$forms α, β, where v is the volume form on the Riemannian manifold M. If d is the exterior derivative, it has a *transposed* (adjoint) operator for that duality, the codifferential δ, defined as

$$\delta = -(*) \circ d \circ (*),$$

which maps $p-$forms onto $(p-1)-$forms, such that

$$\delta\alpha = (-1)^{np+n+1} * d * \alpha.$$

Recall (from subsection 1.2.5 above) that the *Hodge Laplacian*, defined as

$$\Delta = d \circ \delta + \delta \circ d,$$

transforms $p-$forms into $p-$forms and generalizes Beltrami's Laplacian (1.30), which is the special case for $p = 0$ (up to a sign). This defines *harmonic* (real or complex valued) *$p-$forms* as those for which $\Delta\alpha = 0$, or equivalently, $d\alpha = \delta\alpha = 0$.

In other words, let dv be the volume element of the chosen metric g. Then for every $p-$form α we can define a *norm functional*

$$\|\alpha\| = \int_X (\alpha, *\alpha)_g dv,$$

for which the *Euler–Lagrangian equation* becomes $\Delta\alpha = 0$.

Now, the pth Betti number of M can be defined as

$$b_p = \dim \mathrm{Ker}\, \Delta_p,$$

so that the Euler–Poincaré characteristics of M is given by

$$\chi(M) = \sum_{p=0}^{n} (-1)^p b_p = \sum_{p=0}^{n} (-1)^p \dim \mathrm{Ker}\, \Delta_p. \tag{2.695}$$

Finally, for any $(p-1)-$form α, $(p+1)-$form β, and harmonic $p-$form γ ($\Delta\gamma = 0$) on the biomechanical configuration manifold M, the celebrated *Hodge–de Rham decomposition* of a $p-$form ω [Gri83b, Voi02] gives

$$\omega = d\alpha + \delta\beta + \gamma.$$

Now, recall from section 1.2.12, that a large class of symplectic manifolds is given by the *Kähler manifolds*. Let M be a smooth manifold and g a Riemannian metric on M. Let J be a complex structure on M, that is, $J : TM \to TM$, $J^2 = -\mathrm{Id}$, and J is g–orthogonal. M is called a *Kähler manifold* if $\nabla j = 0$, where ∇ is the *Levi–Civita connection* of g and J is regarded as a $(1,1)$ tensor–field. Define a $2-$form ω on M by $\omega(X,Y) = g(JX,Y)$, for each vector–field X, Y on M. Then (M,ω) is a symplectic manifold.

Hodge theory takes place on the cohomology of the compact orientable configuration manifold M and reflects the subtle interplay of the following basic additional linear structures one can impose on M:

- Symplectic structure $\omega \in \Gamma_{C^\infty}(M, \Lambda^2 T^\vee_M)$, where ω is nondegenerate, $d\omega = 0$.

- Riemannian structure $g \in \Gamma_{C^\infty}(M, S^2T^\vee_M)$, where g is positive definite.
- Complex structure $J \in \Gamma_{C^\infty}(M, \mathrm{End}(T_M))$, where $J^2 = -\mathrm{id}$, and J is integrable.

The data (M, ω, g, J) satisfy the *Kähler condition* if ω, g and J are *compatible* in the sense that

$$\omega(\bullet, J(\bullet)) = g(\bullet, \bullet),$$

where $\bullet$ is the strong compatibility condition allowing the comparison of different cohomology theories.

Recall that the de Rham cohomology of (M, J) is defined as

$$H^k_{DR}(M) = \frac{\mathrm{Ker}\left(\Omega^k(M) \xrightarrow{d} \Omega^{k+1}(M)\right)}{\mathrm{Im}\left(\Omega^{k-1}(M) \xrightarrow{d} \Omega^k(M)\right)}.$$

de Rham cohomology classes are represented by harmonic (natural) differential forms.

Let (M, g) be a compact oriented (real or complex) Riemannian manifold. Let dv be the volume element of g. Then for every $k-$form α we can define

$$\|\alpha\| = \int_M (\alpha, \bar{\alpha})_g dv.$$

The *Euler–Lagrangian equation* for the norm functional turns out to be $d\alpha = \delta\alpha = 0$. A $k-$form $\alpha \in \Omega^k(M)$ is called *harmonic* if it satisfies one of the following equivalent conditions:

- α is *closed* and $\|\alpha\| \leq \|\alpha + d\beta\|$ for all $\beta \in \Omega^{k-1}(M)$.
- $d\alpha = \delta\alpha = 0$.
- $\Delta\alpha = 0$, where $\Delta = d\delta + \delta d$ is the *Hodge Laplacian*.

Hodge–Weyl theorem [Gri83b, Voi02] states that *every de Rham cohomology class has a unique harmonic representative*.

Heat Kernel and Thermodynamics on M

Besides pure mechanical consideration of biomechanical system, there is another biophysical point of view – thermodynamical, compatible with the human motion [Hil38]. Namely, the *heat equation* on the biomechanical configuration manifold M,

$$\partial_t a(t) = \Delta a(t), \qquad \text{with initial condition} \quad a(0) = \alpha,$$

has a unique solution for every $t \in [0, \infty)$ and every $p-$form α on M. If we think of α as an *initial temperature distribution* on M then as the configuration manifold cools down, according to the classical heat equation, the temperature should approach a *steady state* which should be *harmonic* [Dav89].

To prove this, we define a *stationary* and hence *harmonic* operator $H(\alpha) = \lim_{t\to\infty} a(t)$. Also, a map $\alpha \to G(\alpha)$ with

$$G(\alpha) = \int_0^\infty a(t)\, dt$$

is *orthogonal to the space of harmonic forms* and satisfies

$$\Delta G(\alpha) = \int_0^\infty \Delta a(t)\, dt = -\int_0^\infty \partial_t a(t)\, dt = \alpha - H(\alpha).$$

Here, the map $\alpha \to H(\alpha)$ is called *harmonic projection* and the map $\alpha \to G(\alpha)$ is called *Green's operator*.

In particular, for each $p-$form α we get a *unique decomposition*

$$\alpha = H(\alpha) + \Delta G(\alpha).$$

This proves the existence of a *harmonic representative* in every de Rham cohomology class, as follows.

Let $\alpha \in \Omega^p(M)$ be a closed form. Then

$$\alpha = H(\alpha) + dd^* G(\alpha) + d^* dG(\alpha).$$

But the three terms in this sum are orthogonal and so

$$\|d^* dG(\alpha)\| = \langle d^* dG(\alpha), \alpha\rangle = \langle dG(\alpha), d\alpha\rangle = 0,$$

since α is closed. Thus $H(\alpha)$ is *cohomologous* to α.

This thermal reflection on the biomechanics topology complies with the basic biophysics of human muscles (see [Hil38]).

Lagrangian–Hamiltonian Duality in Biomechanics

The present section uncovers the underlying dual geometro–topological structure beneath the general biomechanics. It presents a parallel development of Hamiltonian and Lagrangian formulations of biomechanics (see [IS01, Iva02, IP01b, IP01b, Iva05a]), proves both differential–geometrical and algebraic–topo-logical dualities between these two formulations, and finally establishes a *unique functorial relation* between biomechanics geometry and biomechanics topology.

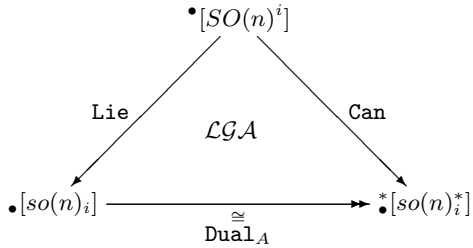
Lagrangian formulation of biomechanics is performed on the *tangent bundle* TM, while Hamiltonian formulation is performed on the *cotangent bundle* T^*M. Both *Riemannian* and *symplectic* geometry are used. The geometrical duality (see [KMS93, BM82]) of Lie groups and algebras between these two biomechanics formulations is proved as an existence of natural equivalence between Lie and canonical functors. The topological duality (see [DP97]) between these two biomechanics formulations is proved as an existence of natural equivalence between Lagrangian and Hamiltonian functors in both *homology* and *cohomology* categories. In the case of reduced configuration manifold, the Betti numbers and Euler–Poincaré characteristic are given.

Geometrical Duality Theorem for M

Theorem. There is a geometrical duality between rotational Lagrangian and Hamiltonian biomechanical formulations on M (as given by Figure 2.21). In categorical terms, there is a unique natural geometrical equivalence

$$\texttt{Dual}_G : \texttt{Lie} \cong \texttt{Can}$$

in biomechanics (symbols are described in the next subsection).
Proof. The proof has two parts: Lie–functorial and geometrical.
Lie–Functorial Proof. If we apply the functor Lie on the category ${}^{\bullet}[SO(n)^i]$ (for $n = 2, 3$ and $i = 1, \ldots, N$) of rotational Lie groups $SO(n)^i$ (and their homomorphisms) we get the category ${}_{\bullet}[so(n)_i]$ of corresponding *tangent* Lie algebras $so(n)_i$ (and their homomorphisms). If we further apply the isomorphic functor Dual to the category ${}_{\bullet}[so(n)_i]$ we get the dual category ${}^{*}_{\bullet}[so(n)^*_i]$ of *cotangent*, or, *canonical* Lie algebras $so(n)^*_i$ (and their homomorphisms). To go directly from ${}^{\bullet}[SO(n)^i]$ to ${}^{*}_{\bullet}[so(n)^*_i]$ we use the canonical functor Can. Therefore, we have a commutative triangle:



Applying the functor Lie on the biomechanical configuration manifold M, we get the product–tree of the same anthropomorphic structure, but having tangent Lie algebras $so(n)_i$ as vertices, instead of the groups $SO(n)^i$. Again, applying the functor Can on M, we get the product–tree of the same anthropomorphic structure, but this time having cotangent Lie algebras $so(n)^*_i$ as vertices. Both the tangent algebras $so(n)_i$ and the cotangent algebras $so(n)^*_i$ contain infinitesimal group generators: angular velocities $\dot{q}^i = \dot{q}^{\phi_i}$ – in the first case, and canonical angular momenta $p_i = p_{\phi_i}$ – in the second case [IS01]. As Lie group generators, both the angular velocities and the angular momenta satisfy the commutation relations: $[\dot{q}^{\phi_i}, \dot{q}^{\psi_i}] = \epsilon^{\phi\psi}_{\theta}\, \dot{q}^{\theta_i}$ and $[p_{\phi_i}, p_{\psi_i}] = \epsilon^{\theta}_{\phi\psi}\, p_{\theta_i}$, respectively, where the structure constants $\epsilon^{\phi\psi}_{\theta}$ and $\epsilon^{\theta}_{\phi\psi}$ constitute the totally antisymmetric third–order tensors.

In this way, the functor $\texttt{Dual}_G : \texttt{Lie} \cong \texttt{Can}$ establishes the unique geometrical duality between kinematics of angular velocities $\dot{q}^i$ (involved in *Lagrangian* formalism on the tangent bundle of M) and kinematics of angular momenta p_i (involved in *Hamiltonian* formalism on the cotangent bundle of M), which is analyzed below. In other words, we have two functors, Lie and Can, from the *category of Lie groups* (of which ${}^{\bullet}[SO(n)^i]$ is a subcategory) into the *category of* (their) *Lie algebras* (of which ${}_{\bullet}[so(n)_i]$ and ${}^{*}_{\bullet}[so(n)^*_i]$ are

subcategories), and a unique natural equivalence between them defined by the functor $\texttt{Dual}_G$. (As angular momenta p_i are in a bijective correspondence with angular velocities $\dot{q}^i$, every component of the functor $\texttt{Dual}_G$ is invertible.) ■

Geometrical Proof. Geometrical proof is given along the lines of Riemannian and symplectic geometry of mechanical systems, as follows (see 2.4.2 above, as well as [MR99, IS01, Iva02, IP01b, Iva05a]). Recall that the Riemannian metric $g = <,>$ on the configuration manifold M is a positive–definite quadratic form $g : TM \to \mathbb{R}$, given in local coordinates $q^i \in U$ (U open in M) as

$$g_{ij} \mapsto g_{ij}(q,m)\, dq^i dq^j, \qquad \text{where}$$
$$g_{ij}(q,m) \;=\; m_\mu \delta_{rs} \frac{\partial x^r}{\partial q^i} \frac{\partial x^s}{\partial q^j}$$

is the covariant material metric tensor g, defining a relation between internal and external coordinates and including n segmental masses m_μ. The quantities x^r are external coordinates ($r, s = 1, \ldots, 6n$) and $i, j = 1, \ldots, N \equiv 6n - h$, where h denotes the number of holonomic constraints.

The *Lagrangian* of the system is a quadratic form $L : TM \to \mathbb{R}$ dependent on velocity v and such that $L(v) = \frac{1}{2} < v, v >$. It is given by

$$L(v) \;=\; \frac{1}{2} g_{ij}(q,m)\, v^i v^j$$

in local coordinates q^i, $v^i = \dot{q}^i \in U_v$ (U_v open in TM). The *Hamiltonian* of the system is a quadratic form $H : T^*M \to \mathbb{R}$ dependent on momentum p and such that $H(p) = \frac{1}{2} < p, p >$. It is given by

$$H(p) \;=\; \frac{1}{2} g^{ij}(q,m)\, p_i p_j$$

in local canonical coordinates $q^i, p_i \in U_p$ (U_p open in T^*M). The inverse (contravariant) metric tensor g^{-1}, is defined as

$$g^{ij}(q,m) \;=\; m_\mu \delta_{rs} \frac{\partial q^i}{\partial x^r} \frac{\partial q^j}{\partial x^s}.$$

For any smooth function L on TM, the *fibre derivative*, or *Legendre transformation*, is a diffeomorphism $\mathbb{F}L : TM \to T^*M$, $\mathbb{F}(w) \cdot v = < w, v >$, from the momentum phase–space manifold to the velocity phase–space manifold associated with the metric $g = <,>$. In local coordinates q^i, $v^i = \dot{q}^i \in U_v$ (U_v open in TM), $\mathbb{F}L$ is given by $(q^i, v^i) \mapsto (q^i, p_i)$.

Recall that on the momentum phase–space manifold T^*M exists:
(i) A unique canonical 1−form θ_H with the property that, for any 1−form β on the configuration manifold M, we have $\beta^* \theta_H = \beta$. In local canonical coordinates $q^i, p_i \in U_p$ (U_p open in T^*M) it is given by $\theta_H = p_i dq^i$.
(ii) A unique nondegenerate Hamiltonian symplectic 2−form ω_H, which is closed ($d\omega_H = 0$) and exact ($\omega_H = d\theta_H = dp_i \wedge dq^i$). Each body segment has, in the general $SO(3)$ case, a sub–phase–space manifold $T^*SO(3)$ with

$$\omega_H^{(sub)} \;=\; dp_\phi \wedge d\phi \;+\; dp_\psi \wedge d\psi \;+\; dp_\theta \wedge d\theta.$$

Analogously, on the velocity phase–space manifold TM exists:
(i) A unique 1−form θ_L, defined by the *pull–back* $\theta_L = (\mathbb{F}L) * \theta_H$ of θ_H by $\mathbb{F}L$. In local coordinates q^i, $v^i = \dot{q}^i \in U_v$ (U_v open in TM) it is given by $\theta_L = L_{v^i} dq^i$, where $L_{v^i} \equiv \partial L/\partial v^i$.
(ii) A unique nondegenerate Lagrangian symplectic 2−form ω_L, defined by the pull–back $\omega_L = (\mathbb{F}L) * \omega_H$ of ω_H by $\mathbb{F}L$, which is closed ($d\omega_L = 0$) and exact ($\omega_L = d\theta_L = dL_{v^i} \wedge dq^i$).

Both T^*M and TM are orientable manifolds, admitting the standard volumes given respectively by

$$\Omega_{\omega_H}, = \frac{(-1)^{\frac{N(N+1)}{2}}}{N!}\omega_H^N, \qquad \text{and} \qquad \Omega_{\omega_L} = \frac{(-1)^{\frac{N(N+1)}{2}}}{N!}\omega_L^N,$$

in local coordinates $q^i, p_i \in U_p$ (U_p open in T^*M), resp. q^i, $v^i = \dot{q}^i \in U_v$ (U_v open in TM). They are given by

$$\begin{aligned} \Omega_H \;&=\; dq^1 \wedge \cdots \wedge dq^N \wedge dp_1 \wedge \cdots \wedge dp_N, \qquad \text{and} \\ \Omega_L \;&=\; dq^1 \wedge \cdots \wedge dq^N \wedge dv^1 \wedge \cdots \wedge dv^N. \end{aligned}$$

On the velocity phase–space manifold TM we can also define the *action* $A : TM \to \mathbb{R}$ by $A(v) = \mathbb{F}L(v) \cdot v$ and the energy $E = A - L$. In local coordinates q^i, $v^i = \dot{q}^i \in U_v$ (U_v open in TM) we have $A = v^i L_{v^i}$, so $E = v^i L_{v^i} - L$. The Lagrangian vector–field X_L on TM is determined by the condition $i_{X_L}\omega_L = dE$. Classically, it is given by the second–order Lagrangian equations

$$\frac{d}{dt}L_{v^i} \;=\; L_{q^i}. \tag{2.696}$$

The Hamiltonian vector–field X_H is defined on the momentum phase–space manifold T^*M by the condition $i_{X_H}\omega = dH$. The condition may be expressed equivalently as $X_H = J\nabla H$, where $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$.

In local canonical coordinates $q^i, p_i \in U_p$ (U_p open in T^*M) the vector–field X_H is classically given by the first–order Hamiltonian canonical equations

$$\dot{q}^i \;=\; \partial_{p_i} H, \qquad \dot{p_i} \;=\; -\,\partial_{q^i} H. \tag{2.697}$$

As a Lie group, the configuration manifold M is Hausdorff. Therefore for $x = (q^i, p_i) \in U_p$ (U_p open in T^*M) there exists a unique one–parameter group of diffeomorphisms $\phi_t : T^*M \to T^*M$ such that $\frac{d}{dt}|_{t=0}\, \phi_t x = J\nabla H(x)$. This is termed *Hamiltonian phase–flow* and represents the maximal integral curve $t \mapsto (q^i(t), p_i(t))$ of the Hamiltonian vector–field X_H passing through the point x for $t = 0$.

The flow ϕ_t is *symplectic* if ω_H is constant along it (that is, $\phi_t^*\omega_H = \omega_H$) iff its Lie derivative vanishes (that is, $\mathcal{L}_{X_H}\omega_H = 0$). A symplectic flow consists

of canonical transformations on T^*M, that is, local diffeomorphisms that leave ω_H invariant. By Liouville theorem, a symplectic flow ϕ_t preserves the phase volume on T^*M. Also, the total energy $H = E$ of the system is conserved along ϕ_t, that is, $H \circ \phi_t = \phi_t$.

Lagrangian flow can be defined analogously (see [AM78, MR99]).

For a Lagrangian (resp. a Hamiltonian) vector–field X_L (resp. X_H) on M, there is a base integral curve $\gamma_0(t) = (q^i(t), v^i(t))$ (resp. $\gamma_0(t) = (q^i(t), p_i(t))$) iff$\gamma_0(t)$ is a geodesic. This is given by the contravariant velocity equation

$$\dot{q}^i = v^i, \qquad \dot{v}^i + \Gamma^i_{jk} v^j v^k = 0, \tag{2.698}$$

in the former case, and by the covariant momentum equation

$$\dot{q}^k = g^{ki} p_i, \qquad \dot{p}_i + \Gamma^i_{jk} g^{jl} g^{km} p_l p_m = 0, \tag{2.699}$$

in the latter. As before, Γ^i_{jk} denote the Christoffel symbols of an affine connection ∇ in an open chart U on M, defined by the Riemannian metric $g = <,>$ as: $\Gamma^i_{jk} = g^{il}\Gamma_{jkl}$, $\Gamma_{jkl} = \frac{1}{2}\left(\partial_{q^j} g_{kl} + \partial_{q^k} g_{jl} - \partial_{q^l} g_{jk}\right)$.

The l.h.s $\dot{\dot{v}}^i = \dot{v}^i + \Gamma^i_{jk} v^j v^k$ (resp. $\dot{\dot{p}}_i = \dot{p}_i + \Gamma^i_{jk} g^{jl} g^{km} p_l p_m$) in the second parts of (2.698) and (2.699) represent the *Bianchi covariant derivative* of the velocity (resp. momentum) with respect to t. Parallel transport on M is defined by $\dot{\dot{v}}^i = 0$, (resp. $\dot{\dot{p}}_i = 0$). When this applies, X_L (resp. X_H) is called the *geodesic spray* and its flow the *geodesic flow*.

For the dynamics in the gravitational potential field $V : M \to \mathbb{R}$, the Lagrangian $L : TM \to \mathbb{R}$ (resp. the Hamiltonian $H : T^*M \to \mathbb{R}$) has an extended form

$$L(v, q) = \frac{1}{2} g_{ij} v^i v^j - V(q),$$
$$\text{(resp.} \qquad H(p, q) = \frac{1}{2} g^{ij} p_i p_j + V(q)).$$

A Lagrangian vector–field X_L (resp. Hamiltonian vector–field X_H) is still defined by the second–order Lagrangian equations (2.696, 2.698) (resp. first–order Hamiltonian equations (2.697, 2.699)).

The fibre derivative $\mathbb{F}L : TM \to T^*M$ thus maps Lagrangian equations (2.696, 2.698) into Hamiltonian equations (2.697, 2.699). Clearly there exists a diffeomorphism $\mathbb{F}H : T^*M \to TM$, such that $\mathbb{F}L = (\mathbb{F}H)^{-1}$. In local canonical coordinates $q^i, p_i \in U_p$ (U_p, open in T^*M) this is given by $(q^i, p_i) \mapsto (q^i, v^i)$ and thus maps Hamiltonian equations (2.697, 2.699) into Lagrangian equations (2.696, 2.698).

A general form of the forced, non–conservative Hamiltonian equations (resp. Lagrangian equations) is given as

$$\dot{q}^i = \frac{\partial H}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial H}{\partial q^i} + F_i(t, q^i, p_i),$$
$$\text{(resp.} \qquad \frac{d}{dt}\frac{\partial L}{\partial v^i} - \frac{\partial L}{\partial q^i} = F_i(t, q^i, v^i)).$$

Here the $F_i(t,\, q^i,\, p_i)$ (resp. $F_i(t,\, q^i,\, v^i)$) represent any kind of *covariant forces*, including dissipative and elastic joint forces, as well as actuator drives and control forces, as a function of time, coordinates and momenta. In covariant form we have

$$\begin{aligned} & \dot{q^k} \;=\; g^{ki} p_i, \qquad \dot{p}_i \;+\; \Gamma^i_{jk}\, g^{jl} g^{km}\, p_l p_m \;=\; F_i(t,\, q^i,\, p_i), \\ (\text{resp.} \quad & \dot{q^i} \;=\; v^i, \qquad \dot{v}^i \;+\; \Gamma^i_{jk}\, v^j v^k \;=\; g^{ij}\, F_j(t,\, q^i,\, v^i)). \end{aligned} \qquad \blacksquare$$

This proves the existence of the unique natural geometrical equivalence

$$\mathtt{Dual}_G : \mathtt{Lie} \cong \mathtt{Can}$$

in the rotational biomechanics.

Topological Duality Theorem for M

In this section we want to prove that the general biomechanics can be *equivalently* described in terms of two *topologically dual functors* `Lag` and `Ham`, from `Diff`, the *category of smooth manifolds* (and their smooth maps) of class C^p, into `Bund`, the *category of vector bundles* (and vector–bundle maps) of class C^{p-1}, with $p \geq 1$. `Lag` is physically represented by the second–order Lagrangian formalism on $TM \in \mathtt{Bund}$, while `Ham` is physically represented by the first–order Hamiltonian formalism on $T^*M \in \mathtt{Bund}$.

Theorem. There is a topological duality between Lagrangian and Hamiltonian formalisms on M (as given by Figure 2.21). In categorical terms, there is a unique natural topological equivalence

$$\mathtt{Dual}_T : \mathtt{Lag} \cong \mathtt{Ham}$$

in the general biomechanics.

Proof. The proof has two parts: cohomological and homological.

Cohomological Proof. If $\mathcal{C} = \mathcal{H}^\bullet\mathcal{M}$ (resp. $\mathcal{C} = \mathcal{L}^\bullet\mathcal{M}$) represents the Abelian category of cochains on the momentum phase–space manifold T^*M (resp. the velocity phase–space manifold TM), we have the category $\mathcal{S}^\bullet(\mathcal{H}^\bullet\mathcal{M})$ (resp. $\mathcal{S}^\bullet(\mathcal{L}^\bullet\mathcal{M})$) of generalized cochain complexes $A^\bullet$ in $\mathcal{H}^\bullet\mathcal{M}$ (resp. $\mathcal{L}^\bullet\mathcal{M}$) and if $A' = 0$ for $n < 0$ we have a subcategory $\mathcal{S}^\bullet_{\mathcal{DR}}(\mathcal{H}^\bullet\mathcal{M})$ (resp. $\mathcal{S}^\bullet_{\mathcal{DR}}(\mathcal{L}^\bullet\mathcal{M})$) of de Rham differential complexes in $\mathcal{S}^\bullet(\mathcal{H}^\bullet\mathcal{M})$ (resp. $\mathcal{S}^\bullet(\mathcal{L}^\bullet\mathcal{M})$)

$$\begin{aligned} & A^\bullet_{DR} : 0 \to \Omega^0(T^*M) \xrightarrow{d} \Omega^1(T^*M) \xrightarrow{d} \\ & \qquad \xrightarrow{d} \Omega^2(T^*M) \xrightarrow{d} \cdots \xrightarrow{d} \Omega^N(T^*M) \xrightarrow{d} \cdots \\ (\text{resp.} \quad & A^\bullet_{DR} : 0 \to \Omega^0(TM) \xrightarrow{d} \Omega^1(TM) \xrightarrow{d} \Omega^2(TM) \xrightarrow{d} \\ & \qquad \cdots \xrightarrow{d} \Omega^N(TM) \xrightarrow{d} \cdots), \end{aligned}$$

where $A' = \Omega^N(T^*M)$ (resp. $A' = \Omega^N(TM)$) is the vector space of all N–forms on T^*M (resp. TM) over $\mathbb{R}$.

Let $Z^N(T^*M) = \mathrm{Ker}(d)$ (resp. $Z^N(T) = \mathrm{Ker}(d)$) and $B^N(T^*M) = \mathrm{Im}(d)$ (resp. $B^N(TM) = \mathrm{Im}(d)$) denote respectively the real vector spaces of cocycles and coboundaries of degree N. Since $d_{N+1}d_N = d^2 = 0$, it follows that $B^N(T^*M) \subset Z^N(T^*M)$ (resp. $B^N(TM) \subset Z^N(TM)$). The quotient vector space

$$H^N_{DR}(T^*M) = \mathrm{Ker}(d)/\mathrm{Im}(d) = Z^N(T^*M)/B^N(T^*M)$$
$$(\text{resp.} \quad H^N_{DR}(TM) = \mathrm{Ker}(d)/\mathrm{Im}(d) = Z^N(TM)/B^N(TM)),$$

we refer to as the de Rham cohomology group (vector space) of T^*M (resp. TM). The elements of $H^N_{DR}(T^*M)$ (resp. $H^N_{DR}(TM)$) are equivalence sets of cocycles. Two cocycles ω_1 and ω_2 are cohomologous, or belong to the same equivalence set (written $\omega_1 \sim \omega_2$) iff they differ by a coboundary $\omega_1 - \omega_2 = d\theta$. Any form $\omega_H \in \Omega^N(T^*M)$ (resp. $\omega_L \in \Omega^N(TM)$ has a de Rham cohomology class $[\omega_H] \in H^N_{DR}(T^*M)$ (resp. $[\omega_L] \in H^N_{DR}(TM)$).

Hamiltonian symplectic form $\omega_H = dp_i \wedge dq_i$ on T^*M (resp. Lagrangian symplectic form $\omega_L = dL_{v^i} \wedge dq^i$ on TM) is by definition both a closed $2-$form or two–cocycle and an exact $2-$form or two–coboundary. Therefore the 2D–de Rham cohomology group of human motion is defined as a quotient vector space

$$H^2_{DR}(T^*M) = Z^2(T^*M)/B^2(T^*M)$$
$$(\text{resp.} \quad H^2_{DR}(TM) = Z^2(TM)/B^2(TM)).$$

As T^*M (resp. TM) is a compact Hamiltonian symplectic (resp. Lagrangian symplectic) manifold of dimension $2N$, it follows that ω_H^N (resp. ω_L^N) is a volume element on T^*M (resp. TM), and the $2N$D de Rham's cohomology class $\left[\omega_H^N\right] \in H^{2N}_{DR}(T^*M)$ (resp. $\left[\omega_L^N\right] \in H^{2N}_{DR}(TM)$) is nonzero. Since $\left[\omega_H^N\right] = [\omega_H]^N$ (resp. $\left[\omega_L^N\right] = [\omega_L]^N$), then $[\omega_H] \in H^2_{DR}(T^*M)$ (resp. $[\omega_L] \in H^2_{DR}(TM)$) and all of its powers up to the N–th must be zero as well. The existence of such an element is a necessary condition for T^*M (resp. TM) to admit a Hamiltonian symplectic structure ω_H (resp. Lagrangian symplectic structure ω_L).

The de Rham complex $A^\bullet_{DR}$ on T^*M (resp. TM) can be considered as a system of second–order ODEs $d^2\theta_H = 0$, $\theta_H \in \Omega^N(T^*M)$ (resp. $d^2\theta_L = 0$, $\theta_L \in \Omega^N(TM)$) having a solution represented by $Z^N(T^*M)$ (resp. $Z^N(TM)$). In local coordinates q^i, $p_i \in U_p$ (U_p open in T^*M) (resp. q^i, $v^i \in U_v$ (U_v open in TM)) we have $d^2\theta_H = d^2(p_i dq^i) = d(dp_i \wedge dq^i) = 0$, (resp. $d^2\theta_L = d^2(L_{v^i} dq^i) = d(dL_{v^i} \wedge dq^i) = 0$). ■

Homological Proof. If $\mathcal{C} = \mathcal{H}_\bullet\mathcal{M}$, (resp. $\mathcal{C} = \mathcal{L}_\bullet\mathcal{M}$) represents an Abelian category of chains on T^*M (resp. TM), we have a category $\mathcal{S}_\bullet(\mathcal{H}_\bullet\mathcal{M})$ (resp. $\mathcal{S}_\bullet(\mathcal{L}_\bullet\mathcal{M})$) of generalized chain complexes $\mathcal{A}_\bullet$ in $\mathcal{H}_\bullet\mathcal{M}$ (resp. $\mathcal{L}_\bullet\mathcal{M}$), and if $A = 0$ for $n < 0$ we have a subcategory $\mathcal{S}^C_\bullet(H_\bullet M)$ (resp. $\mathcal{S}^C_\bullet(L_\bullet M)$) of chain complexes in $\mathcal{H}_\bullet\mathcal{M}$ (resp. $\mathcal{L}_\bullet\mathcal{M}$)

$$A_\bullet : 0 \leftarrow C^0(T^*M) \xleftarrow{\partial} C^1(T^*M) \xleftarrow{\partial} C^2(T^*M) \xleftarrow{\partial} \cdots$$
$$\cdots \xleftarrow{\partial} C^n(T^*M) \xleftarrow{\partial} \cdots$$
$$(\text{resp.} \quad A_\bullet : 0 \leftarrow C^0(TM) \xleftarrow{\partial} C^1(TM) \xleftarrow{\partial} C^2(TM) \xleftarrow{\partial} \cdots$$
$$\cdots \xleftarrow{\partial} C^n(TM) \xleftarrow{\partial} \cdots).$$

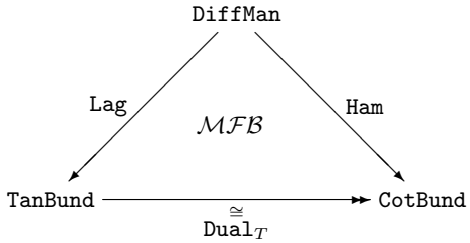
Here $A_N = C^N(T^*M)$ (resp. $A_N = C^N(TM)$) is the vector space of all finite chains C on T^*M (resp. TM) over $\mathbb{R}$, and $\partial_N = \partial : C^{N+1}(T^*M) \to C^N(T^*M)$ (resp. $\partial_N = \partial : C^{N+1}(TM) \to C^N(TM)$). A finite chain C such that $\partial C = 0$ is an $N-$cycle. A finite chain C such that $C = \partial B$ is an $N-$boundary. Let $Z_N(T^*M) = \mathrm{Ker}(\partial)$ (resp. $Z_N(TM) = \mathrm{Ker}(\partial)$) and $B_N(T^*M) = \mathrm{Im}(\partial)$ (resp. $B_N(TM) = \mathrm{Im}(\partial)$) denote respectively real vector spaces of cycles and boundaries of degree N. Since $\partial_{N-1}\partial_N = \partial^2 = 0$, then $B_N(T^*M) \subset Z_N(T^*M)$ (resp. $B_N(TM) \subset Z_N(TM)$). The quotient vector space

$$H_N^C(T^*M) = Z_N(T^*M)/B_N(T^*M)$$
$$(\text{resp.} \quad H_N^C(TM) = Z_N(TM)/B_N(TM))$$

represents an ND *biomechanics homology group* (vector space). The elements of $H_N^C(T^*M)$ (resp. $H_N^C(TM)$) are equivalence sets of cycles. Two cycles C_1 and C_2 are homologous, or belong to the same equivalence set (written $C_1 \sim C_2$) iff they differ by a boundary $C_1 - C_2 = \partial B$. The homology class of a finite chain $C \in C^N(T^*M)$ (resp. $C \in C^N(TM)$) is $[C] \in H_N^C(T^*M)$ (resp. $[C] \in H_N^C(TM)$). ■

Lagrangian Versus Hamiltonian Duality

In this way, we have proved a commutativity of a triangle:



which implies the existence of the unique natural topological equivalence

$$\mathtt{Dual}_T : \mathtt{Lag} \cong \mathtt{Ham}$$

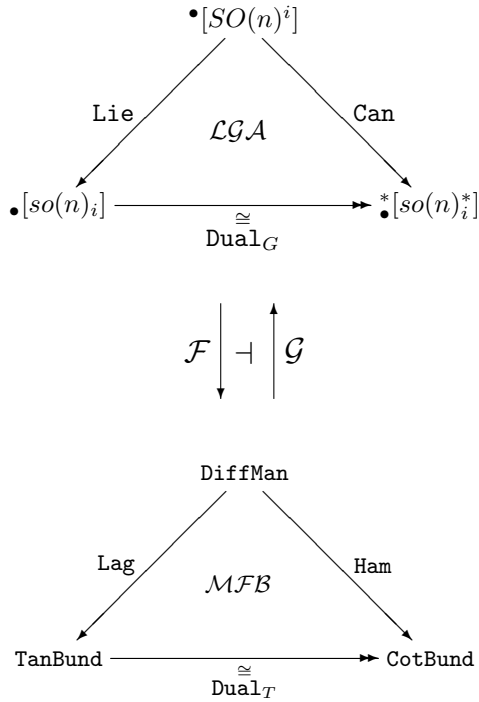
in the rotational biomechanics.

Globally Dual Structure of Rotational Biomechanics

Theorem. Global dual structure of the rotational biomechanics is defined by the unique natural equivalence

$$\texttt{Dyn} : \texttt{Dual}_G \cong \texttt{Dual}_T.$$

Proof. This unique functorial relation, uncovering the natural equivalence between *geometrical* and *topological* structures of biomechanics:



– has been established by parallel development of Lagrangian and Hamiltonian biomechanics formulations, i.e., functors `Lag`(`Lie`) and `Ham`(`Can`). ■

2.5 Neurodynamics

2.5.1 Microscopic Neurodynamics and Quantum Brain

Biochemistry of Microtubules

Recent developments/efforts to understand aspects of the brain function at the *sub–neural* level are discussed in [Nan95]. Microtubules (MTs), protein polymers constructing the cytoskeleton of a neuron, participate in a wide variety of dynamical processes in the cell. Of special interest for this subsection is the

MTs participation in bioinformation processes such as *learning* and *memory*, by possessing a well–known binary error–correcting code $[K_1(13, 2^6, 5)]$ with 64 words. In fact, MTs and DNA/RNA are *unique* cell structures that possess a code system. It seems that the MTs' code system is strongly related to a kind of *mental code* in the following sense. The MTs' periodic paracrystalline structure make them able to support a *superposition* of coherent quantum states, as it has been recently conjectured by Hameroff and Penrose [HP96], representing an *external* or *mental order*, for sufficient time needed for *efficient quantum computing*.

Living organisms are collective assemblies of cells which contain collective assemblies of organized material, including membranes, organelles, nuclei, and the *cytoplasm*, the bulk interior medium of living cells. Dynamic rearrangements of the cytoplasm within *eucaryotic cells*, the cells of all animals and almost all plants on Earth, account for their changing shape, movement, etc. This extremely important cytoplasmic structural and dynamical organization is due to the presence of networks of inteconnected protein polymers, which are referred to as the *cytosceleton* due to their bone–like structure [HP96, Dus84]. The cytoskeleton consists of MT's, actin microfilaments, intermediate filaments and an *organizing complex*, the *centrosome* with its chief component the *centriole*, built from two bundles of microtubules in a separated **T** shape. Parallel–arrayed MTs are interconnected by cross–bridging proteins (*MT–Associated Proteins*: MAPs) to other MTs, organelle filaments and membranes to form *dynamic networks* [HP96, Dus84]. MAPs may be contractile, structural, or enzymatic. A very important role is played by contractile MAPs, like dynein and kinesin, through their participation in cell movements as well as in intra-neural, or axoplasmic transport which moves material and thus is of fundamental importance for the *maintenance* and *regulation* of *synapses* (see, e.g., [Ecc64]). The structural bridges formed by MAPs stabilize MTs and prevent their disassembly. The MT–MAP 'complexes' or *cytosceletal networks* determine the cell architecture and dynamic functions, such a *mitosis*, or *cell division*, *growth*, *differentiation*, *movement*, and for us here the very crucial, *synapse formation and function*, all essential to the living state. It is usually said that *microtubules* are ubiquitous through the entire biology [HP96, Dus84].

MTs are hollow cylinders comprised of an exterior surface of cross–section diameter 25 nm ($1\ nm = 10^{-9}$ meters) with 13 arrays (protofilaments) of protein dimers called tubulines [Dus84]. The interior of the cylinder, of cross–section diameter 14 nm, contains *ordered water* molecules, which implies the existence of an electric dipole moment and an electric field. The arrangement of the dimers is such that, if one ignores their size, they resemble triangular lattices on the MT surface. Each dimer consists of two hydrophobic protein pockets, and has an unpaired electron. There are two possible positions of the electron, called α and β *conformations*. When the electron is in the $\beta-$conformation there is a 29^o distortion of the electric dipole moment as compared to the α conformation.

In standard models for the simulation of the MT dynamics [STZ93, SZT98], the 'physical' DOF – relevant for the description of the energy transfer – is the projection of the electric dipole moment on the longitudinal symmetry axis ($x-$axis) of the MT cylinder. The 29^o distortion of the $\beta-$conformation leads to a displacement u_n along the $x-$axis, which is thus the relevant physical DOF.

There has been speculation for quite some time that MTs are involved in information processing: it has been shown that the particular geometrical arrangement (packing) of the tubulin protofilaments obeys an error–correcting mathematical code known as the $K_2(13, 2^6, 5)-$code [KHS93]. Error correcting codes are also used in classical computers to protect against errors while in quantum computers special error correcting algorithms are used to protect against errors by preserving quantum coherence among qubits.

Information processing occurs via interactions among the MT proto–filament chains. The system may be considered as similar to a model of *interacting Ising chains* on a triangular lattice, the latter being defined on the plane stemming from filleting open and flattening the cylindrical surface of MT. Classically, the various dimers can occur in either α or β conformations. Each dimer is influenced by the neighboring dimers resulting in the possibility of a transition. This is the basis for classical information processing, which constitutes the picture of a (classical) cellular automaton.

Kink Soliton Model of MT–Dynamics

The *quantum nature* of an MT network results from the *assumption* that each dimer finds itself in a *superposition* of α and β conformations. Viewed as a *two–state quantum mechanical system*, the MT tubulin dimers couple to conformational changes with $10^{-9} - 10^{-11}$sec transitions, corresponding to an angular frequency $\omega \sim \mathcal{O}(10^{10}) - \mathcal{O}(10^{12})$ Hz [Nan95].

The *quantum computer* character of the MT network [Pen89] results from the assumption that each dimer finds itself in a superposition of α and β conformations [Ham87]. There is a macroscopic coherent state among the various chains, which lasts for $\mathcal{O}(1\,\mathrm{sec})$ and constitutes the 'preconscious' state [Nan95]. The interaction of the chains with (non–critical stringy) quantum gravity, then, induces self–collapse of the wave function of the coherent MT network, resulting in quantum computation.

In [EMN92, EMN99, MN95a, MN95b, Nan95] the authors assumed that the collapse occurs mainly due to the interaction of each chain with quantum gravity, the interaction from neighboring chains being taken into account by including mean–field interaction terms in the dynamics of the displacement field of each chain. This amounts to a modification of the effective potential by anharmonic oscillator terms. Thus, the effective system under study is 2D, possessing one space and one time coordinate.

Let u_n be the displacement field of the nth dimer in a MT chain. The continuous approximation proves sufficient for the study of phenomena asso-

ciated with energy transfer in biological cells, and this implies that one can make the replacement

$$u_n \to u(x,t), \tag{2.700}$$

with x a spatial coordinate along the longitudinal symmetry axis of the MT. There is a time variable t due to fluctuations of the displacements $u(x)$ as a result of the dipole oscillations in the dimers.

The effects of the neighboring dimers (including neighboring chains) can be phenomenologically accounted for by an effective potential $V(u)$. In the kink–soliton model[36] of ref. [STZ93, SZT98] a double–well potential was used, leading to a classical kink solution for the $u(x,t)$ field. More complicated interactions are allowed in the picture of Ellis *et al.*, where more generic polynomial potentials have been considered.

The effects of the surrounding water molecules can be summarized by a *viscous force* term that damps out the dimer oscillations,

$$F = -\gamma \partial_t u, \tag{2.701}$$

with γ determined phenomenologically at this stage. This friction should be viewed as an environmental effect, which however does not lead to energy dissipation, as a result of the non–trivial solitonic structure of the ground–state and the non–zero constant force due to the electric field. This is a well known result, directly relevant to energy transfer in biological systems.

In mathematical terms the effective equation of motion for the relevant field DOF $u(x,t)$ reads:

$$u''(\xi) + \rho u'(\xi) = P(u), \tag{2.702}$$

where $\xi = x - vt$, $u'(\xi) = du/d\xi$, v is the velocity of the soliton, $\rho \propto \gamma$ [STZ93, SZT98], and $P(u)$ is a polynomial in u, of a certain degree, stemming from the variations of the potential $V(u)$ describing interactions among the MT chains. In the mathematical literature there has been a classification of solutions of equations of this form. For certain forms of the potential the solutions include *kink solitons* that may be responsible for dissipation–free energy transfer in biological cells:

$$u(x,t) \sim c_1 \left(\tanh[c_2(x - vt)] + c_3\right), \tag{2.703}$$

where c_1, c_2, c_3 are constants depending on the parameters of the dimer lattice model. For the form of the potential assumed in the model of [STZ93, SZT98] there are solitons of the form $u(x,t) = c'_1 + \frac{c'_2 - c'_1}{1+e^{c'_3(c'_2-c'_1)(x-vt)}}$, where again $c'_i,\ i = 1, \ldots 3$ are appropriate constants.

A *semiclassical quantization* of such solitonic states has been considered by Ellis *et al.*. The result of such a quantization yields a *modified soliton equation* for the (quantum corrected) field $u_q(x,t)$ [TF91]

[36] Recall that kinks are solitary (non–dispersive) waves arising in various 1D (bio)physical systems.

$$\partial_t^2 u_q(x,t) - \partial_x^2 u_q(x,t) + \mathcal{M}^{(1)}[u_q(x,t)] = 0, \tag{2.704}$$

with the notation

$$M^{(n)} = e^{\frac{1}{2}(G(x,y,t)-G_0(x,y))\frac{\partial^2}{\partial z^2}} U^{(n)}(z)|_{z=u_q(x,t)}, \qquad U^{(n)} \equiv d^n U/dz^n.$$

The quantity U denotes the potential of the original soliton Hamiltonian, and $G(x,y,t)$ is a bilocal field that describes quantum corrections due to the modified boson field around the soliton. The quantities $M^{(n)}$ carry information about the quantum corrections. For the kink soliton (2.703) the quantum corrections (2.704) have been calculated explicitly in [TF91], thereby providing us with a concrete example of a large–scale quantum coherent state.

A typical propagation velocity of the kink solitons (e.g., in the model of [STZ93, SZT98]) is $v \sim 2$ m/sec, although, models with $v \sim 20$ m/sec have also been considered. This implies that, for moderately long microtubules of length $L \sim 10^{-6}$ m, such kinks transport energy without dissipation in

$$t_F \sim 5 \times 10^{-7} \text{ s}. \tag{2.705}$$

Such time scales are comparable to, or smaller in magnitude than, the decoherence time scale of the above–described coherent (solitonic) states $u_q(x,t)$. This implies the possibility that fundamental quantum mechanical phenomena may then be responsible for frictionless energy (and signal) transfer across microtubular arrangements in the cell [Nan95].

Open Liouville Neurodynamics and Self–Similarity

Recall that neurodynamics has its physical behavior both on the *macroscopic*, classical, *inter–neuronal level*, and on the *microscopic*, quantum, *intra–neuronal level*. On the *macroscopic* level, various models of neural networks (NNs, for short) have been proposed as goal–oriented models of the specific neural functions, like for instance, function–approximation, pattern–recognition, classification, or control (see, e.g., [Hay94]). In the physically–based, Hopfield–type models of NNs [Hop82), Hop84] the information is stored as a content–addressable memory in which synaptic strengths are modified after the Hebbian rule (see [Heb49]. Its retrieval is made when the network with the symmetric couplings works as the point–attractor with the fixed points. Analysis of both *activation* and *learning dynamics* of Hopfield–Hebbian NNs using the techniques of statistical mechanics [DHS91], gives us with the most important information of storage capacity, role of noise and recall performance.

On the other hand, on the general *microscopic* intra–cellular level, energy transfer across the cells, without dissipation, had been first conjectured to occur in biological matter by [FK83]. The phenomenon conjectured by them was based on their 1D superconductivity model: in 1D electron systems with holes, the formation of *solitonic structures* due to electron–hole pairing results

in the transfer of electric current without dissipation. In a similar manner, Frölich and Kremer conjectured that energy in biological matter could be transferred without dissipation, if appropriate solitonic structures are formed inside the cells. This idea has lead theorists to construct various models for the energy transfer across the cell, based on the formation of *kink* classical solutions (see [STZ93, SZT98].

The interior of living cells is structurally and dynamically organized by *cytoskeletons*, i.e., networks of protein polymers. Of these structures, *microtubules* (MTs, for short) appear to be the most fundamental (see [Dus84]). Their dynamics has been studied by a number of authors in connection with the mechanism responsible for dissipation–free energy transfer. Hameroff and Penrose [Ham87] have conjectured another fundamental role for the MTs, namely being responsible for *quantum computations* in the human neurons. [Pen89, Pen94, Pen97] further argued that the latter is associated with certain aspects of quantum theory that are believed to occur in the cytoskeleton MTs, in particular quantum superposition and subsequent collapse of the wave function of coherent MT networks. These ideas have been elaborated by [MN95a, MN95b] and [Nan95], based on the quantum–gravity EMN–language of [EMN92, EMN99] where MTs have been physically modelled as non-critical (SUSY) bosonic strings. It has been suggested that the neural MTs are the microsites for the emergence of stable, macroscopic quantum coherent states, identifiable with the *preconscious states*; stringy–quantum space–time effects trigger an organized collapse of the coherent states down to a specific or *conscious state*. More recently, [TVP99] have presented the evidence for biological self–organization and pattern formation during embryogenesis.

Now, we have two space–time biophysical scales of neurodynamics. Naturally the question arises: are these two scales somehow inter-related, is there a space–time self–similarity between them?

The purpose of this subsection is to prove the formal positive answer to the self–similarity question. We try to describe neurodynamics on both physical levels by the *unique form* of a single equation, namely *open Liouville equation*: NN–dynamics using its classical form, and MT–dynamics using its quantum form in the Heisenberg picture. If this formulation is consistent, that would prove the *existence* of the *formal neurobiological space–time self–similarity*.

Hamiltonian Framework

Suppose that on the macroscopic NN–level we have a conservative Hamiltonian system acting in a $2N$D symplectic phase–space $T^*Q = \{q^i(t), p_i(t)\}$, $(i = 1 \dots N)$ (which is the cotangent bundle of the NN–configuration manifold $Q = \{q^i\}$), with a Hamiltonian function $H = H(q^i, p_i, t) : T^*Q \times \mathbb{R} \to \mathbb{R}$. The conservative dynamics is defined by classical Hamiltonian canonical equations (1.17). Recall that within the conservative Hamiltonian framework, we can apply the formalism of classical Poisson brackets: for any two functions $A = A(q^i, p_i, t)$ and $B = B(q^i, p_i, t)$ their Poisson bracket is defined as

$$[A,B]=\left(\frac{\partial A}{\partial q^i}\frac{\partial B}{\partial p_i}-\frac{\partial A}{\partial p_i}\frac{\partial B}{\partial q^i}\right).$$

Conservative Classical System

Any function $A(q^i,p_i,t)$ is called a *constant* (or integral) of motion of the conservative system (1.17) if

$$\dot{A}\equiv\partial_t A+[A,H]=0,\qquad\text{which implies}\qquad\partial_t A=-[A,H]\,.\tag{2.706}$$

For example, if $A=\rho(q^i,p_i,t)$ is a *density function* of ensemble phase–points (or, a probability density to see a state $x(t)=(q^i(t),p_i(t))$ of *ensemble* at a moment t), then equation

$$\partial_t\rho=-[\rho,H]=-\mathrm{i}L\,\rho\tag{2.707}$$

represents the *Liouville theorem*, where L denotes the (Hermitian) *Liouville operator*

$$\mathrm{i}L=[...,H]\equiv\left(\frac{\partial H}{\partial p_i}\frac{\partial}{\partial q^i}-\frac{\partial H}{\partial q^i}\frac{\partial}{\partial p_i}\right)=\mathrm{div}(\rho\dot{\mathbf{x}}),$$

which shows that the conservative Liouville equation (2.707) is actually equivalent to the mechanical *continuity equation*

$$\partial_t\rho+\mathrm{div}(\rho\dot{\mathbf{x}})=0.\tag{2.708}$$

Conservative Quantum System

We perform the formal quantization of the conservative equation (2.707) in the Heisenberg picture: all variables become Hermitian operators (denoted by '$\wedge$'), the symplectic phase–space $T^*Q=\{q^i,p_i\}$ becomes the Hilbert state–space $\mathcal{H}=\mathcal{H}_{\hat{q}^i}\otimes\mathcal{H}_{\hat{p}_i}$ (where $\mathcal{H}_{\hat{q}^i}=\mathcal{H}_{\hat{q}^1}\otimes...\otimes\mathcal{H}_{\hat{q}^N}$ and $\mathcal{H}_{\hat{p}_i}=\mathcal{H}_{\hat{p}_1}\otimes...\otimes\mathcal{H}_{\hat{p}_N}$), the classical Poisson bracket $[\,,\,]$ becomes the quantum commutator $\{\,,\,\}$ multiplied by $-\mathrm{i}/\hbar$

$$[\,,\,]\longrightarrow-\mathrm{i}\{\,,\,\}\qquad(\hbar=1\text{ in normal units})\,.\tag{2.709}$$

In this way the classical Liouville equation (2.707) becomes the *quantum Liouville equation*

$$\partial_t\hat{\rho}=\mathrm{i}\{\hat{\rho},\hat{H}\}\,,\tag{2.710}$$

where $\hat{H}=\hat{H}(\hat{q}^i,\hat{p}_i,t)$ is the Hamiltonian evolution operator, while

$$\hat{\rho}=P(a)|\Psi_a><\Psi_a|,\qquad\text{with}\qquad\mathrm{Tr}(\hat{\rho})=1,$$

denotes the von Neumann *density matrix operator*, where each quantum state $|\Psi_a>$ occurs with probability $P(a)$; $\hat{\rho}=\hat{\rho}(\hat{q}^i,\hat{p}_i,t)$ is closely related to another von Neumann concept: *entropy* $S=-\mathrm{Tr}(\hat{\rho}[\ln\hat{\rho}])$.

Open Classical System

We now move to the *open (nonconservative) system*: on the macroscopic NN–level the *opening operation* equals to the *adding* of a *covariant* vector of external (dissipative and/or motor) forces $F_i = F_i(q^i, p_i, t)$ to (the r.h.s of) the covariant Hamiltonian *force equation*, so that Hamiltonian equations get the *open (dissipative and/or forced) form*

$$\dot{q}^i = \frac{\partial H}{\partial p_i}, \qquad \dot{p}_i = F_i - \frac{\partial H}{\partial q^i}\,. \tag{2.711}$$

In the framework of the open Hamiltonian system (2.711), dynamics of any function $A(q^i, p_i, t)$ is defined by the *open evolution equation*:

$$\partial_t A = -[A, H] + \Phi,$$

where $\Phi = \Phi(F_i)$ represents the general form of the scalar force term.

In particular, if $A = \rho(q^i, p_i, t)$ represents the density function of ensemble phase–points, then its dynamics is given by the (dissipative/forced) *open Liouville equation*:

$$\partial_t \rho = -[\rho, H] + \Phi\,. \tag{2.712}$$

In particular, the scalar force term can be cast as a linear Poisson–bracket form

$$\Phi = F_i[A, q^i]\,, \qquad \text{with} \qquad [A, q^i] = -\frac{\partial A}{\partial p_i}\,. \tag{2.713}$$

Now, in a similar way as the conservative Liouville equation (2.707) resembles the continuity equation (2.708) from continuum dynamics, also the open Liouville equation (2.712) resembles the probabilistic *Fokker–Planck equation* from statistical mechanics. If we have a ND stochastic process $x(t) = (q^i(t), p_i(t))$ defined by the vector *Itô SDE*

$$dx(t) = f(x,t)\,dt + G(x,t)\,dW,$$

where f is a ND vector function, W is a KD Wiener process, and G is a $N \times K$D matrix valued function, then the corresponding probability density function $\rho = \rho(x, t|\dot{x}, t')$ is defined by the ND Fokker–Planck equation (see, e.g., [Gar85])

$$\partial_t \rho = -\operatorname{div}[\rho\, f(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x_i \partial x_j}(Q_{ij}\,\rho)\,, \tag{2.714}$$

where $Q_{ij} = \big(G(x,t)\,G^T(x,t)\big)_{ij}$. It is obvious that the Fokker–Planck equation (2.714) represents the particular, stochastic form of our general open Liouville equation (2.712), in which the scalar force term is given by the (second–derivative) noise term

$$\Phi = \frac{1}{2}\frac{\partial^2}{\partial x_i \partial x_j}(Q_{ij}\,\rho)\ .$$

Equation (2.712) will represent the open classical model of our macroscopic NN–dynamics.

Continuous Neural Network Dynamics

The generalized NN–dynamics, including two special cases of *graded response neurons* (GRN) and *coupled neural oscillators* (CNO), can be presented in the form of a stochastic *Langevin rate equation*

$$\dot{\sigma}_i = f_i + \eta_i(t), \tag{2.715}$$

where $\sigma_i = \sigma_i(t)$ are the continual neuronal variables of ith neurons (representing either membrane action potentials in case of GRN, or oscillator phases in case of CNO); J_{ij} are individual synaptic weights; $f_i = f_i(\sigma_i, J_{ij})$ are the deterministic forces (given, in GRN–case, by $f_i = \sum_j J_{ij} \tanh[\gamma\sigma_j] - \sigma_i + \theta_i$, with $\gamma > 0$ and with the θ_i representing injected currents, and in CNO–case, by $f_i = \sum_j J_{ij} \sin(\sigma_j - \sigma_i) + \omega_i$, with ω_i representing the natural frequencies of the individual oscillators); the noise variables are given as $\eta_i(t) = \lim_{\Delta\to 0} \zeta_i(t)\sqrt{2T/\Delta}$ where $\zeta_i(t)$ denote uncorrelated Gaussian distributed random forces and the parameter T controls the amount of noise in the system, ranging from $T = 0$ (deterministic dynamics) to $T = \infty$ (completely random dynamics).

More convenient description of the neural random process (2.715) is provided by the Fokker–Planck equation describing the time evolution of the probability density $P(\sigma_i)$

$$\partial_t P(\sigma_i) = -\frac{\partial}{\partial\sigma_i}\left(f_i P(\sigma_i)\right) + T\frac{\partial^2}{\partial\sigma_i^2} P(\sigma_i). \tag{2.716}$$

Now, in the case of deterministic dynamics $T = 0$, equation (2.716) can be put into the form of the conservative Liouville equation (2.707), by making the substitutions: $P(\sigma_i) \to \rho$, $f_i = \dot{\sigma}_i$, and $[\rho, H] = \mathrm{div}(\rho\,\dot{\sigma}_i) \equiv \sum_i \frac{\partial}{\partial\sigma_i}(\rho\,\dot{\sigma}_i)$, where $H = H(\sigma_i, J_{ij})$. Further, we can formally identify the stochastic forces, i.e., the second–order noise–term $T\sum_i \frac{\partial^2}{\partial\sigma_i^2}\rho$ with $F^i[\rho, \sigma_i]$, to get the open Liouville equation (2.712).

Therefore, on the NN–level deterministic dynamics corresponds to the conservative system (2.707). Inclusion of stochastic forces corresponds to the system opening (2.712), implying the *macroscopic arrow of time*.

Open Quantum System

By formal quantization of equation (2.712) with the scalar force term defined by (2.713), in the same way as in the case of the conservative dynamics, we get the *quantum open Liouville equation*

$$\partial_t\hat{\rho} = \mathrm{i}\{\hat{\rho}, \hat{H}\} + \hat{\Phi}, \qquad \text{with} \qquad \hat{\Phi} = -\mathrm{i}\hat{F}_i\{\hat{\rho}, \hat{q}^i\}, \tag{2.717}$$

where $\hat{F}_i = \hat{F}_i(\hat{q}^i, \hat{p}_i, t)$ represents the covariant quantum operator of external friction forces in the Hilbert state–space $\mathcal{H} = \mathcal{H}_{\hat{q}^i} \otimes \mathcal{H}_{\hat{p}_i}$.

Equation (2.717) will represent the open quantum–friction model of our microscopic MT–dynamics. Its system–independent properties are [EMN92, EMN99, MN95a, MN95b, Nan95]:

1. Conservation of probability P

$$\partial_t P = \partial_t[\mathrm{Tr}(\hat{\rho})] = 0.$$

2. Conservation of energy E, on the average

$$\partial_t \langle\langle E \rangle\rangle \equiv \partial_t[\mathrm{Tr}(\hat{\rho}\, E)] = 0.$$

3. Monotonic increase in entropy

$$\partial_t S = \partial_t[-\mathrm{Tr}(\hat{\rho} \ln \hat{\rho})] \geq 0,$$

and thus automatically and naturally implies a *microscopic arrow of time*, so essential in realistic biophysics of neural processes.

Non–Critical Stringy MT–Dynamics

In EMN–language of non–critical (SUSY) bosonic strings, our MT–dynamics equation (2.717) reads

$$\partial_t \hat{\rho} = \mathrm{i}\{\hat{\rho}, \hat{H}\} - \mathrm{i}\hat{g}_{ij}\{\hat{\rho}, \hat{q}^i\}\hat{\dot{q}}^j\,, \tag{2.718}$$

where the target–space density matrix $\hat{\rho}(\hat{q}^i, \hat{p}_i)$ is viewed as a function of coordinates $\hat{q}^i$ that parameterize the couplings of the generalized $\sigma-$models on the bosonic string world–sheet, and their conjugate momenta $\hat{p}_i$, while $\hat{g}_{ij} = \hat{g}_{ij}(\hat{q}^i)$ is the quantum operator of the *positive definite metric* in the space of couplings. Therefore, the covariant quantum operator of external friction forces is in EMN–formulation given as $\hat{F}_i(\hat{q}^i, \hat{\dot{q}}^i) = \hat{g}_{ij}\,\hat{\dot{q}}^j$.

Equation (2.718) establishes the conditions under which a large–scale coherent state appearing in the MT–network, which can be considered responsible for loss–free energy transfer along the tubulins.

Equivalence of Neurodynamic Forms

It is obvious that both the macroscopic NN–equation (2.712) and the microscopic MT–equation (2.717) have the same open Liouville form, which implies the arrow of time. These proves the existence of the formal neuro–biological space–time self–similarity.

In this way, we have described neurodynamics of both NN and MT ensembles, belonging to completely different biophysical space–time scales, by the unique form of open Liouville equation, which implies the arrow of time. The existence of the formal *neuro–biological self–similarity* has been proved.

Dissipative Quantum Brain Model

The *conservative brain* model was originally formulated within the framework of the quantum field theory (QFT) by [RU67] and subsequently developed

in [STU78, STU79, JY95, JPY96]. The conservative brain model has been recently extended to the *dissipative quantum dynamics* in the work of G. Vitiello and collaborators [Vit95, AV00, PV99, Vit01, PV03, PV04].

The canonical quantization procedure of a dissipative system requires to include in the formalism also the system representing the environment (usually the heat bath) in which the system is embedded. One possible way to do that is to depict the environment as the time–reversal image of the system [CRV92]: the environment is thus described as the *double* of the system in the time–reversed dynamics (the system image in the mirror of time).

Within the framework of dissipative QFT, the brain system is described in terms of an *infinite collection of damped harmonic oscillators* A_κ (the simplest prototype of a dissipative system) representing the DWQ [Vit95]. Now, the collection of damped harmonic oscillators is ruled by the Hamiltonian [Vit95, CRV92]

$$\begin{aligned} H &= H_0 + H_I, \qquad \text{with} \\ H_0 &= \hbar\Omega_\kappa(A_\kappa^\dagger A_\kappa - \tilde{A}_\kappa^\dagger \tilde{A}_\kappa), \qquad H_I = \mathrm{i}\hbar\Gamma_\kappa(A_\kappa^\dagger \tilde{A}_\kappa^\dagger - A_\kappa \tilde{A}_\kappa), \end{aligned}$$

where Ω_κ is the frequency and Γ_κ is the damping constant. The $\tilde{A}_\kappa$ modes are the 'time–reversed mirror image' (i.e., the 'mirror modes') of the A_κ modes. They are the doubled modes, representing the environment modes, in such a way that κ generically labels their degrees of freedom. In particular, we consider the damped harmonic oscillator (DHO)

$$m\ddot{x} + \gamma\dot{x} + \kappa x = 0, \tag{2.719}$$

as a simple prototype for dissipative systems (with intention that thus get results also apply to more general systems). The damped oscillator (2.719) is a non–Hamiltonian system and therefore the customary canonical quantization procedure cannot be followed. However, one can face the problem by resorting to well known tools such as the *density matrix* ρ and the *Wigner function* W.

Let us start with the special case of a *conservative particle* in the absence of friction γ, with the standard Hamiltonian, $H = -(\hbar\partial_x)^2/2m + V(x)$.

Recall (from the previous subsection) that the *density matrix equation of motion*, i.e., *quantum Liouville equation*, is given by

$$\mathrm{i}\hbar\dot{\rho} = [H, \rho]. \tag{2.720}$$

The density matrix function ρ is defined by

$$\langle x + \frac{1}{2}y|\rho(t)|x - \frac{1}{2}y\rangle = \psi^*(x + \frac{1}{2}y, t)\psi(x - \frac{1}{2}y, t) \equiv W(x, y, t),$$

with the associated standard expression for the *Wigner function* (see [FH65]),

$$W(p, x, t) = \frac{1}{2\pi\hbar}\int W(x, y, t)\, \mathrm{e}^{\left(-\mathrm{i}\frac{py}{\hbar}\right)} dy.$$

Now, in the coordinate $x-$representation, by introducing the notation

$$x_{\pm} = x \pm \frac{1}{2}y, \tag{2.721}$$

the Liouville equation (2.720) can be expanded as

$$\begin{aligned} &\mathrm{i}\hbar\,\partial_t\langle x_+|\rho(t)|x_-\rangle = \\ &\left\{-\frac{\hbar^2}{2m}\left[\partial^2_{x_+} - \partial^2_{x_-}\right] + [V(x_+) - V(x_-)]\right\}\langle x_+|\rho(t)|x_-\rangle, \end{aligned} \tag{2.722}$$

while the Wigner function $W(p,x,t)$ is now given by

$$\begin{aligned} \mathrm{i}\hbar\,\partial_t W(x,y,t) &= H_o W(x,y,t), \qquad \text{with} \\ H_o &= \frac{1}{m}p_x p_y + V(x+\frac{1}{2}y) - V(x-\frac{1}{2}y), \\ \text{and} \qquad p_x &= -\mathrm{i}\hbar\partial_x, \qquad p_y = -\mathrm{i}\hbar\partial_y. \end{aligned} \tag{2.723}$$

The new Hamiltonian H_o (2.723) may be get from the corresponding Lagrangian

$$L_o = m\dot{x}\dot{y} - V(x+\frac{1}{2}y) + V(x-\frac{1}{2}y). \tag{2.724}$$

In this way, Vitiello concluded that the density matrix and the Wigner function formalism *required*, even in the conservative case (with zero mechanical resistance γ), the introduction of a 'doubled' set of coordinates, $x_{\pm}$, or, alternatively, x and y. One may understand this as related to the introduction of the 'couple' of indices *necessary* to label the density matrix elements (2.722).

Let us now consider the case of the *particle interacting* with a *thermal bath* at temperature T. Let f denote the *random force* on the particle at the position x due to the bath. The interaction Hamiltonian between the bath and the particle is written as

$$H_{int} = -fx. \tag{2.725}$$

Now, in the *Feynman–Vernon formalism* (see [Fey72]), the *effective action* $A[x,y]$ for the particle is given by

$$A[x,y] = \int_{t_i}^{t_f} L_o(\dot{x},\dot{y},x,y)\,dt + I[x,y],$$

with L_o defined by (2.724) and

$$\mathrm{e}^{\frac{\mathrm{i}}{\hbar}I[x,y]} = \langle(\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}\int_{t_i}^{t_f} f(t)x_-(t)dt})_-(\mathrm{e}^{\frac{\mathrm{i}}{\hbar}\int_{t_i}^{t_f} f(t)x_+(t)dt})_+\rangle, \tag{2.726}$$

where the symbol $\langle . \rangle$ denotes the average with respect to the thermal bath; '$(.)_+$' and '$(.)_-$' denote time ordering and anti–time ordering, respectively; the coordinates $x_{\pm}$ are defined as in (2.721). If the interaction between the

bath and the coordinate x (2.725) were turned off, then the operator f of the bath would develop in time according to $f(t) = \mathrm{e}^{\mathrm{i}H_\gamma t/\hbar} f \mathrm{e}^{-\mathrm{i}H_\gamma t/\hbar}$, where H_γ is the Hamiltonian of the isolated bath (decoupled from the coordinate x). $f(t)$ is then the force operator of the bath to be used in (2.726).

The interaction $I[x, y]$ between the bath and the particle has been evaluated in [SVW95] for a linear passive damping due to thermal bath by following Feynman–Vernon and Schwinger [FH65]. The final result from [SVW95] is:

$$I[x,y] = \frac{1}{2}\int_{t_i}^{t_f} dt\,[x(t)F_y^{ret}(t) + y(t)F_x^{adv}(t)] + \frac{\mathrm{i}}{2\hbar}\int_{t_i}^{t_f}\int_{t_i}^{t_f} dtds\,N(t-s)y(t)y(s),$$

where the retarded force on y, F_y^{ret}, and the advanced force on x, F_x^{adv}, are given in terms of the retarded and advanced Green functions $G_{ret}(t-s)$ and $G_{adv}(t-s)$ by

$$F_y^{ret}(t) = \int_{t_i}^{t_f} ds\,G_{ret}(t-s)y(s), \qquad F_x^{adv}(t) = \int_{t_i}^{t_f} ds\,G_{adv}(t-s)x(s),$$

respectively. In (2.727), $N(t-s)$ is the *quantum noise* in the fluctuating random force given by: $N(t-s) = \frac{1}{2}\langle f(t)f(s) + f(s)f(t)\rangle$.

The real and the imaginary part of the action are given respectively by

$$\mathrm{Re}\,(A[x,y]) = \int_{t_i}^{t_f} L\,dt, \tag{2.727}$$

$$L = m\dot{x}\dot{y} - \left[V(x+\frac{1}{2}y) - V(x-\frac{1}{2}y)\right] + \frac{1}{2}\left[xF_y^{ret} + yF_x^{adv}\right], \tag{2.728}$$

$$\text{and} \qquad \mathrm{Im}\,(A[x,y]) = \frac{1}{2\hbar}\int_{t_i}^{t_f}\int_{t_i}^{t_f} N(t-s)y(t)y(s)\,dtds. \tag{2.729}$$

Equations (2.727–2.729), are *exact* results for linear passive damping due to the bath. They show that in the classical limit '$\hbar \to 0$' nonzero y yields an 'unlikely process' in view of the large imaginary part of the action implicit in (2.729). Nonzero y, indeed, may lead to a negative real exponent in the evolution operator, which in the limit $\hbar \to 0$ may produce a negligible contribution to the probability amplitude. On the contrary, at quantum level nonzero y accounts for quantum noise effects in the fluctuating random force in the system–environment coupling arising from the imaginary part of the action (see [SVW95]).

When in (2.728) we use

$$F_y^{ret} = \gamma\dot{y} \qquad \text{and} \qquad F_x^{adv} = -\gamma\dot{x} \qquad \text{we get,}$$

$$L(\dot{x},\dot{y},x,y) \;=\; m\dot{x}\dot{y} - V\left(x+\frac{1}{2}y\right) + V\left(x-\frac{1}{2}y\right) + \frac{\gamma}{2}(x\dot{y}-y\dot{x}). \quad (2.730)$$

By using

$$V\left(x \pm \frac{1}{2}y\right) = \frac{1}{2}\kappa(x \pm \frac{1}{2}y)^2$$

in (2.730), the DHO equation (2.719) and its complementary equation for the y coordinate

$$m\ddot{y} - \gamma\dot{y} + \kappa y = 0. \quad (2.731)$$

are derived. The $y-$oscillator is the time–reversed image of the $x-$oscillator (2.719). From the manifolds of solutions to equations (2.719) and (2.731), we could choose those for which the y coordinate is constrained to be zero, they simplify to

$$m\ddot{x} + \gamma\dot{x} + \kappa x = 0, \qquad y = 0.$$

Thus we get the classical damped oscillator equation from a Lagrangian theory at the expense of introducing an 'extra' coordinate y, later constrained to vanish. Note that the constraint $y(t) = 0$ is *not* in violation of the equations of motion since it is a true solution to (2.719) and (2.731).

2.5.2 Macroscopic Neurodynamics

Hopfield's Neural Nets

Ising–Spin Hopfield Neurons

Recall that Hopfield [Hop82)] gives a collection of simple threshold automata, called formal neurons by McCulloch & Pitts: two–state, 'all–or–none', firing or nonfiring units that can be modeled by Ising spins (uniaxial magnets) $\{S_i\}$ such that $S_i = \pm 1$ (where $1 = |\uparrow\rangle$–'spin up' and $-1 = |\downarrow\rangle$–'spin down'; the label of the neuron is i and ranges between 1 and the size of the network N). The neurons are connected by synapses J_{ij}.

A (firing) patterns $\{\xi_i^\mu\}$ represent specific S_i-spin configurations (where the label of the pattern is μ and ranges between 1 and q).

Using random patterns $\xi_i^\mu = \pm 1$ with equal probability 1/2, we have the synaptic efficacy J_{ij} of jth neuron operating on ith neuron given by

$$J_{ij} = N^{-1}\xi_i^\mu\xi_j^\mu \equiv N^{-1}\boldsymbol{\xi}_\mathbf{i}\cdot\boldsymbol{\xi}_\mathbf{j}. \quad (2.732)$$

Postsynaptic potential (PSP) represents an internal local field

$$h_i(t) = J_{ij}S_j(t). \quad (2.733)$$

Now, the sequential (threshold) dynamics is defined in the form of discrete equation

$$S_i(t+\Delta t) = \mathrm{sgn}[h_i(t)]. \quad (2.734)$$

Dynamics (2.734) is equivalent to the rule that the state of a neuron is changed, or a spin is flipped if and only if the total network energy, given by Ising Hamiltonian

$$H_N = -\frac{1}{2} J_{ij} S_i S_j \tag{2.735}$$

is lowered [Hop82), HT85]. Therefore, the Ising Hamiltonian H_N represents the monotonically decreasing Lyapunov function for the sequential dynamics (2.734), which converges to a local minimum or ground state of H_N. This holds for any symmetric coupling $J_{ij} = J_{ji}$ with $J_{ii} = 0$ and if spin–updating in (2.734) is asynchronous. In this case the patterns $\{\xi_i^\mu\}$ after convergence become identical, or very near to, ground states of H_N, each of them at the bottom of the valley.

Data are stored in the neural net if, by a suitable choice of the J_{ij}, several specific patterns $\{\xi_i^\mu\}$ are made local minima of H_N. If this can be achieved, the neural net will function as content–addressable or (auto)associative memory. A network state which 'somehow resembles' one of the stored prototypes corresponds to a location in the energy landscape which is close enough to the minimum representing that prototype to lie in its basin of attraction. By spontaneously moving downhill, or relaxing to the energy minimum, the network recalls the data or reconstructs the prototype.

Suppose that we have somehow stored several (stationary) patterns $\{\xi_i^\mu\}$ in the J_{ij} and that the system is offered a noisy version of one of them. If the noise was not too strong, the system remains in the valley associated with that pattern and under its natural dynamics it will relax to the energy minimum were the stored patterns live. That is, the system has recalled the pattern.

In statistical mechanics, one is usually given the synapses J_{ij} and one of the first tasks consists in finding the minima of the Ising Hamiltonian H_N. However, in the theory of neural networks, one is given the patterns $\{\xi_i^\mu\}$ and one is asked to solve the inverse problem: finding synapses J_{ij} such that the patterns $\{\xi_i^\mu\}$ are minima of the Hamiltonian H_N.

To see why the Hopfield model with synapses given by (2.732) has patterns $\{\xi_i^\mu\}$ as attractors of the dynamics (2.734), note that the sequential dynamical law embodies a two–step process, the evolution of the local field (PSP) (2.733), which is a linear operation, and a nonlinear decision process (2.734).

Graded–Response Hopfield Networks

Recall that graded–response neurons have continuous input–output relation (like nonlinear amplifiers) of the form $V_i = g_i(\lambda u_i)$, where u_i denotes the input at i, a constant λ is called the gain parameter, and V_i is the output [Hop84]. Usually, g_i are taken to be sigmoid functions, odd, and monotonically increasing (e.g., $g(\cdot) = \frac{1}{2}(1 + \tanh(\cdot))$, while discrete Ising spins have $g_i(u_i) = \mathrm{sgn}_i(u_i)$. The behavior of the continuous Hopfield neural network is usually described by a set of coupled RC–transient equations

$$C_i \dot{u}_i = I_i + J_{ij} V_j - \frac{u_i}{R_i} \tag{2.736}$$

where $u_i = g^{-1}(V_i)$, R_i and C_i denote input capacitance and resistance, and I_i represents an external source.

The Hamiltonian of the continuous system (2.736) is given by

$$H = -\frac{1}{2} J_{ij} V_i V_j + \sum_{i=1}^{N} R_i^{-1} \int_0^{V_i} dV g^{-1}(V) - I_i V_i. \tag{2.737}$$

However, according to [Hop84], the synapses J_{ij} retain the form (2.732) with random patterns $\xi_i^\mu = \pm 1$ with equal probability 1/2, and the synaptic symmetry $J_{ij} = J_{ji}$ implies that the continuous Hamiltonian (2.737) represents a Lyapunov function of the system (2.736), i.e., H decreases under the continual neurodynamics governed by equation (2.736) as time proceeds.

More general form of synapses is

$$J_{ij} = N^{-1} Q(\xi_i; \xi_j),$$

for some synaptic kernel Q on $R^q \times R^q$. The vector ξ_i varies as i travels from 1 to N, but remains on a corner of the Hamming hypercube $[-1, 1]^q$.

Hopfield overlaps

Assuming that the number q of stored patterns is small compared to the number of neurons, i.e., $q/N \to 0$, we find that the synapses (2.732) give rise to a local field of the form

$$h_i = \xi_i^\mu m_\mu, \qquad \text{where} \qquad m_\mu = N^{-1} \xi_i^\mu S_i \tag{2.738}$$

is the *auto–overlap* (or simply *overlap*)[37] of the network state $\{S_i\}$ with the pattern $\{\xi_i^\mu\}$, measuring the proximity between them. We can see that $m_\mu = 1$ (like peak–up in auto–correlation) if $\{S_i\}$ and $\{\xi_i^\mu\}$ are identical patterns, $m_\mu = -1$ (like peak–down in autocorrelation) if they are each other's complement, and $m_\mu = \mathrm{O}(1/\sqrt{N})$ if they are uncorrelated (like no–peak in auto–correlation) with each other. Overlaps m_μ are related to the Hamming distance d_μ between the patterns (the fraction of spins which differ) by $d_\mu = \frac{1}{2}(1 - m_\mu)$.

As a pattern ξ_i^μ represents (in the simplest case) a specific Ising–spin S_i–configuration, then $(\xi_i^\mu)^2 = 1$. If $S_i = \xi_i^\mu$ for all i, then $m_\mu = 1$. Conversely, if $m_\mu = 1$, then $S_i = \xi_i^\mu$. In all other cases $m_\mu < 1$, by the Cauchy–Schwartz inequality. If ξ_i^μ and S_i are uncorrelated, we may expect m_μ to be of the order of $N^{-1/2}$, since the sum consists of N terms, each containing a ξ_i^μ. On the other hand, if the S_i are positively correlated with ξ_i^μ, then m_μ is of the order of unity. So the overlaps give the global information about the network

[37] The auto–overlap resembles the auto–correlation function of a time–series, where distinct peaks indicate that the series at the certain time t is similar to the series at time $t + \Delta t$.

and hence are good order parameters, in terms of Haken's synergetics (see [Hak83]). Also, according to [Hop84], the extension to the continual network is straightforward.

Using overlaps, the Ising Hamiltonian becomes

$$H_N = -\frac{1}{2} N \sum_{\mu=1}^{q} m_\mu^2.$$

The similarity between two different patterns ξ_i^μ and ξ_i^ν is measured by their mutual overlap or *cross–overlap* $m_{\mu\nu}$ (in other parlance it is called Karhunen–Loeve matrix (see [FS92]), which extracts the principal components from a data set)[38], equal

$$m_{\mu\nu} = N^{-1} \xi_i^\mu \xi_i^\nu. \tag{2.739}$$

For similar patterns the cross–overlap is close to unity whereas for uncorrelated patterns it is random variable with zero mean and small $(1/\sqrt{N})$ variance.

The symmetric Hopfield synaptic matrix J_{ij} can be expressed in terms of the cross–overlaps $m_{\mu\nu}$ as

$$J_{ij} = N^{-1} \xi_i^\mu (m_{\mu\nu})^{-1} \xi_j^\nu = J_{ji},$$

where $(m_{\mu\nu})^{-1}$ denotes the *Moore–Penrose pseudoinverse* of the cross–overlap matrix $m_{\mu\nu}$.

Besides the Hopfield model, the proposed pattern–overlap picture can be extended to cover some more sophisticated kinds of associative memory, to two of them:

(i) The so–called *forgetful memories*, characterized by iterative synaptic prescription

$$J_{ij}^{(\mu)} = \phi(\epsilon \xi_i^\mu \xi_j^\mu + J_{ij}^{(\mu-1)}),$$

for some small parameter ϵ and some odd function ϕ. If $\phi(\cdot)$ saturates as $|\cdot| \to \infty$, the memory creates storage capacity for new patterns by forgetting the old ones.

(ii) The so–called *temporal associative memories*, which can store and retrieve a sequence of patterns, through synapses

$$N J_{ij} = \xi_i^\mu \xi_j^\mu + \epsilon \xi_i^{(\mu+1)} \xi_j^\mu,$$

where the second term on the right is associated with a temporal delay, so that one can imagine that the second term 'pushes' the neural system through an energy landscape created by the first term.

[38] The cross–overlap resembling the cross–correlation function of two time–series, with several distinct peaks, indicating that the two series are very similar at each point in time where the peaks occur

Spiking Neural Nets

Recently, Izhikevich [Izh04] discussed biological plausibility and computational efficiency of some of the most useful models of *spiking and bursting neurons* (see Figure 2.24). He compared their applicability to large–scale simulations of cortical neural networks.

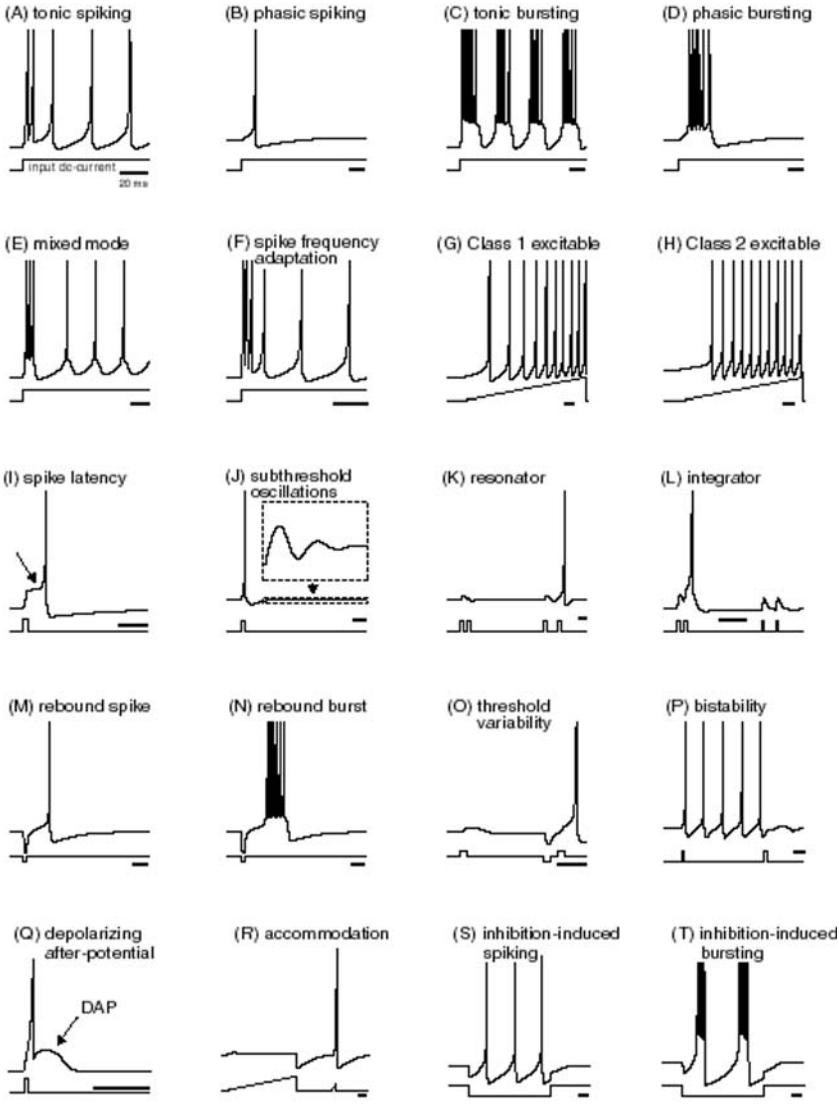


Fig. 2.24. Neuro–computational features of biological neurons (with permission from E. Izhikevich).

Following [Izh04], we present some widely used models of spiking and bursting neurons that can be expressed in the form of ODEs. Throughout this subsection, v denotes the membrane potential. All the parameters in the models are chosen so that v has mV scale and the time has ms scale. To compare computational cost, we assume that each model, written as a dynamical system $\dot{x} = f(x)$, is implemented using the simplest, fixed–step first–order Euler method, with the integration time step chosen to achieve a reasonable numerical accuracy.

Integrate–and–Fire Neuron

One of the most widely used models in computational neuroscience is the *leaky integrate–and–fire neuron*, (I&F neuron, for short) given by

$$\dot{v} = I + a - bv, \qquad \text{If} \quad v \geq v_{trsh} \quad \text{Then} \quad v \leftarrow c,$$

where v is the membrane potential, I is the input current, and a, b, c, and v_{trsh} are the parameters. When the membrane potential v reaches the threshold value v_{trsh}, the neuron is said to fire a *spike*, and v is reset to c. The I&F neuron can fire tonic spikes with constant frequency, and it is an integrator. The I&F neuron is *Class 1 excitable system* [Izh99]; it can fire tonic spikes with constant frequency, and it is an integrator. It is the simplest model to implement when the integration time step τ is 1 ms. Because I&F has only one variable, it cannot have phasic spiking, bursting of any kind, rebound responses, threshold variability, bistability of attractors, or autonomous chaotic dynamics. Because of the fixed threshold, the spikes do not have latencies. In summary, despite its simplicity, I&F is one of the worst models to use in simulations, unless one wants to prove analytical results [Izh04].

Integrate–and–Fire Neuron with Adaptation

The I&F model is $1D$, hence it cannot burst or have other properties of cortical neurons. One may think that having a second linear equation

$$\dot{v} = I + a - bv + g(d - v), \qquad \dot{g} = (e\delta(t) - g)/\tau,$$

describing activation dynamics of a high–threshold $K-$current, can make an improvement, e.g., endow the model with spike–frequency adaptation. Indeed, each firing increases the $K-$activation gate via Dirac $\delta-$function and produces an outward current that slows down the frequency of tonic spiking. This model is fast, yet still lacks many important properties of cortical spiking neurons.

Integrate–and–Fire–or–Burst Neuron

The *integrate–and–fire–or–burst neuron* model is given by

$$\dot{v} = I + a - bv + gH(v - v_h)h(v_T - v),$$

$$\text{If} \quad v \geq v_{trsh} \quad \text{Then} \quad v \leftarrow c, \qquad \dot{h} = \begin{cases} \frac{-h}{\tau^-}, & \text{if} \quad v > v_h, \\ \frac{1-h}{\tau^+}, & \text{if} \quad v < v_h \end{cases}$$

to model thalamo–cortical neurons. Here h describes the inactivation of the calcium $T-$current, g, v_h, v_T, τ^+ and τ^- are parameters describing dynamics of the $T-$current, and H is the Heaviside step function. Having this kind of a second variable creates the possibility for bursting and other interesting regimes [Izh04], but is already a much slower (depending on the value of v).

Complex–Valued Resonate–and–Fire Neuron

The *resonate–and–fire neuron* is a complex–valued (i.e., $2D$) analogue of the I&F neuron [Izh01], given by

$$\dot{z} = I + (b + iw)z, \qquad \text{if} \quad \text{Im}\, z = a_{trsh} \quad \text{then} \quad z \longleftarrow z_0(z), \tag{2.740}$$

where $z = x + iy \in \mathbb{C}$ is a complex–valued variable that describes oscillatory activity of the neuron. Here b, w, and a_{trsh} are parameters, $i = \sqrt{-1}$, and $z_0(z)$ is an arbitrary function describing activity–dependent after–spike reset. (2.740) is equivalent to the linear system

$$\dot{x} = bx - wy, \qquad \dot{y} = wx + by,$$

where the real part x is the current–like variable, while the imaginary part y is the voltage–like variable. The resonate–and–fire model is simple and efficient. When the frequency of oscillation $w = 0$, it becomes an integrator.

Quadratic Integrate–and–Fire Neuron

An alternative to the leaky I&F neuron is the *quadratic I&F neuron*, also known as the *theta–neuron*, or the Ermentrout–Kopell canonical model [Erm96, Gut98]. It can be presented as

$$\dot{v} = I + a(v - v_{rest})(v - v_{trsh}), \qquad \text{If} \quad v = v_{trsh} \quad \text{Then} \quad v \leftarrow v_{rest},$$

where v_{rest} and v_{trsh} are the resting and threshold values of the membrane potential. This model is canonical in the sense that any *Class 1 excitable system* [Izh99] described by smooth ODEs can be transformed into this form by a continuous change of variables. It takes only seven operations to simulate 1 ms of the model, and this should be the model of choice when one simulates large–scale networks of integrators. Unlike its linear analogue, the quadratic I&F neuron has spike latencies, activity dependent threshold (which is v_{trsh} only when $I = 0$), and bistability of resting and tonic spiking modes.

FitzHugh–Nagumo Neuron

The parameters in the *FitzHugh–Nagumo neuron* model

$$\dot{v} = a + bv + cv^2 + dv^3 - u, \qquad \dot{u} = \varepsilon(ev - u),$$

can be tuned so that the model describes spiking dynamics of many resonator neurons. Since one needs to simulate the shape of each spike, the time step

in the model must be relatively small, e.g., $\tau = 0.25\,ms$. Since the model is a $2D$ system of ODEs, without a reset, it cannot exhibit autonomous chaotic dynamics or bursting. Adding noise to this, or some other $2D$ models, allows for stochastic bursting.

Hindmarsh–Rose Neuron

The *Hindmarsh–Rose thalamic neuron* model [RH89] can be written as a $3D$ ODE system

$$\dot{v} = I + u - F(v) - w, \qquad \dot{u} = G(v) - u, \qquad \dot{w} = (H(v) - w)/\tau,$$

where F, G, and H are some functions. This model is quite expensive to implement as a large–scale spike simulator [Izh04].

Morris–Lecar Neuron

Morris and Lecar [ML81] suggested a simple $2D$ model to describe oscillations in barnacle giant muscle fiber. Because it has biophysically meaningful and measurable parameters, the *Morris–Lecar neuron* model became quite popular in computational neuroscience community. It consists of a membrane potential equation with instantaneous activation of Ca current and an additional equation describing slower activation of K current,

$$\begin{aligned}
&C\dot{V} = I - g_L(V - V_L) - g_{Ca} m_\infty(V)(V - V_{Ca}) - g_K n(V - V_K),\\
&\dot{n} = \lambda(V)(n_\infty(V) - n), \qquad \text{where}\\
&m_\infty(V) = \frac{1}{2}\left(1 + \tanh\left[\frac{V - V_1}{V_2}\right]\right), \qquad \text{and}\\
&n_\infty(V) = \frac{1}{2}\left(1 + \tanh\left[\frac{V - V_3}{V_4}\right]\right), \qquad \lambda(V) = \bar{\lambda}\cosh\left[\frac{V - V_3}{2V_4}\right],
\end{aligned}$$

with parameters: $C = 20\,\mu F/cm^2$, $g_L = 2\,mmho/cm^2$, $V_L = -50\,mV$, $g_{Ca} = 4\,mmho/cm^2$, $V_{Ca} = 10\,mV$, $g_K = 8\,mmho/cm^2$, $V_K = -70\,mV$, $V_1 = 0\,mV$, $V_2 = 15\,mV$, $V_3 = 10\,mV$, $V_4 = 10\,mV$, $\bar{\lambda} = 0.1\,s^{-1}$, and applied current $I(\mu A/cm^2)$. The model can exhibit various types of spiking, but could exhibit tonic bursting only when an additional equation is added, e.g., slow inactivation of Ca current. In this case, the model becomes equivalent to the *Hodgkin–Huxley neuron model* [HH52, Hod64], which is extremely expensive to implement.

Activation Dynamics of Graded–Response Networks

To design neurodynamical classifier activation dynamics, we start from the fully recurrent, ND, RC transient circuit, i.e., a nonlinear vector differential equation [Hay94, Kos92, Iva95]:

$$C_j\dot{v}_j = I_j - \frac{v_j}{R_j} + w_{ij}f_i(v_i), \qquad (i,j = 1,...,N), \tag{2.741}$$

where $v_j = v_j(t)$ represent the activation potentials in the jth neuron, C_j and R_j denote input capacitances and leakage resistances, synaptic weights w_{ij} represent conductances, I_j represent the total currents flowing toward the input nodes, and the functions f_i are sigmoidal.

Geometrically, equation (2.741) defines a smooth autonomous vector–field $X(t)$ in ND neurodynamical phase–space manifold M, and its (numerical) solution for the given initial potentials $v_j(0)$ defines the autonomous neurodynamical phase–flow $\Phi(t) : v_j(0) \to v_j(t)$ on M.

In AI parlance, equation (2.741) represents a generalization of three well–known recurrent NN models [Hay94, Kos92]:
(i) continuous Hopfield model,
(ii) Grossberg ART–family cognitive system, and
(iii) Hecht–Nielsen counter–propagation network.

Physiologically, equation (2.741) is based on the Nobel–awarded *Hodgkin–Huxley equation* of the neural action potential (for the single squid giant axon membrane) as a function of the conductances g of sodium, potassium and leakage [HH52, Hod64]:

$$C\dot{v} = I(t) - g_{Na}(v - v_{Na}) - g_K(v - v_K) - g_L(v - v_L),$$

where bracket terms represent the electromotive forces acting on the ions.

The continuous Hopfield circuit model:

$$C_j\dot{v}_j = I_j - \frac{v_j}{R_j} + T_{ij}u_i, \qquad (i,j = 1,...,N), \tag{2.742}$$

where u_i are output functions from processing elements, and T_{ij} is the inverse of the resistors connection–matrix becomes equation (2.741) if we put $T_{ij} = w_{ij}$ and $u_i = f_i[v_j(t)]$.

The Grossberg analogous ART2 system is governed by activation equation:

$$\varepsilon\dot{v}_j = -Av_j + (1 - Bv_j)I_j^+ - (C + Dv_j)I_j^-, \qquad (j = 1,...,N),$$

where A, B, C, D are positive constants (A is dimensionally conductance), $0 \leq \varepsilon << 1$ is the fast–variable factor (dimensionally capacitance), and I_j^+, I_j^- are excitatory and inhibitory inputs to the jth processing unit, respectively.

General Cohen–Grossberg activation equations [CG83] have the form:

$$\dot{v}_j = -a_j(v_j)[b_j(v_j) - f_k(v_k)m_{jk}], \qquad (j = 1,...,N), \tag{2.743}$$

and the Cohen–Grossberg theorem ensures the global stability of the system (2.743). If

$$a_j = 1/C_j, \qquad b_j = v_j/R_j - I_j, \qquad f_j(v_j) = u_j,$$

and constant $m_{ij} = m_{ji} = T_{ij}$, the system (2.743) reduces to the Hopfield circuit model (2.742).

The Hecht–Nielsen counter–propagation network is governed by the activation equation:

$$\dot{v}_j = -Av_j + (B - v_j)I_j - v_j \sum_{k \neq j} I_k, \qquad (j = 1, ..., N),$$

where A, B are positive constants and I_j are input values for each processing unit.

Provided some simple conditions are satisfied, namely, say symmetry of weights $w_{ij} = w_{ij}$, non–negativity of activations v_j and monotonicity of transfer functions f_j, the system (1) is globally asymptotically stable (in the sense of Liapunov energy functions). The fixed points (stable states) of the system correspond to the fundamental memories to be stored, so it works as content–addressable memory (CAM). The initial state of the system (2.741) lies inside the basin of attraction of its fixed points, so that its initial state is related to appropriate memory vector. Various variations on this basic model are reported in the literature [Hay94, Kos92], and more general form of the vector–field can be given, preserving the above stability conditions. Here is a two–step generalization.

First, all of N neurons of equation (2.741) belong to the same layer, so the M–layer generalized network is naturally governed by tensorial 'field'–equation:

$$C_j^L \, \dot{v}_j^L = I_j^L - \frac{v_j^L}{R_j^L} + w_{ij}^L \, f_i^L(v_i^L), \qquad (i, j = 1, ..., N, \quad L = 1, ..., M), \quad (2.744)$$

where, in the pattern–recognition formulation, js denote the pixels and Ls correspond to the clusters.

The basic two–layer example of the tensorial system (2.744) is Kosko's adaptive bidirectional associative memory (ABAM) system, which represents a two–layer generalization of the Cohen–Grossberg system (2.743):

$$\dot{v}_j = -a_j(v_j) \, [b_j(v_j) - f_k(v_k) \, m_{jk}], \qquad (j = 1, ..., N), \qquad (2.745)$$
$$\dot{u}_k = -a_k(u_k) \, [b_k(u_k) - f_j(u_j) \, m_{jk}], \qquad (k = 1, ..., N). \qquad (2.746)$$

Kosko has proved the global stability for the ABAM system (2.745) in case of signal Hebbian and competitive learning [Kos92]. He has also proved the global stability for its stochastic generalization random adaptive bidirectional associative memory (RABAM) system:

$$\dot{v}_j = -a_j(v_j) \, [b_j(v_j) - f_k(v_k) \, m_{jk}] + n_j, \qquad (j = 1, ..., N),$$
$$\dot{u}_k = -a_k(u_k) \, [b_k(u_k) - f_j(u_j) \, m_{jk}] + n_k, \qquad (k = 1, ..., N).$$

where $n_{j,k}$ represents an additive zero–mean Gaussian white noise.

Second, the fuzzyfied model of the system (2.744) can be get by replacing the inputs I_j^L with the fuzzy membership functions Ψ_j^L for each node (pixel) j in each layer (cluster) L:

$$C_j^L \dot{v}_j^L = \Psi_j^L - \frac{v_j^L}{R_j^L} + w_{ij}^L f_i^L(v_i^L), \qquad (i,j = 1,...,N, \quad L = 1,...,M),$$

Other possibilities for the fuzzification include fuzzy parameters (C_j^L and R_j^L) and fuzzy initial conditions ($v_j^L(0)$). Even activation functions $f_j^L(\cdot)$ could be subject to the fuzzification process.

Learning Dynamics of Common Graded–Response Networks

A continuous unsupervised learning paradigm, i.e., continuous update law for the third–order, mixed, synaptic tensor w_{ij}^L, can be proposed in two forms (generalized from [Kos92]):

Hebbian Learning Scheme

The generalized *Hebbian learning* scheme [Heb49] is defined by

$$\dot{w}_{ij}^L = -w_{ij}^L + \Phi_{ij}^L(v_j^L, u_j^L, w_{ij}^L), \qquad (i,j = 1,..,N, \quad L = 1,...,M),$$

where three terms from the left to the right denote respectively a new–update value, an old value and innovation of the synaptic tensor w_{ij}^L. In this case the nonlinear functions Φ_{ij}^L are usually defined by one of the following learning models:
(i) signal Hebbian learning, with innovation:

$$\Phi_{ij}^L = f_j^L(v_j^L) f_i^L(u_j^L);$$

(ii) differential Hebbian learning, with innovation:

$$\Phi_{ij}^L = f_j^L(v_j^L) f_i^L(u_j^L) + \dot{f}_j^L(v_j^L) \dot{f}_i^L(u_j^L),$$

where overdots denote 'signal velocities';
(iii) random signal Hebbian learning, with innovation:

$$\Phi_{ij}^L = f_j^L(v_j^L) f_i^L(u_j^L) + n_{ij}^L,$$

where n_{ij}^L denotes an additive, tensorial, zero–mean, Gaussian white–noise process independent of the main 'innovation signal'; and
(iv) random differential signal Hebbian learning, with innovation:

$$\Phi_{ij}^L = f_j^L(v_j^L) f_i^L(u_j^L) + \dot{f}_j^L(v_j^L) \dot{f}_i^L(u_j^L) + n_{ij}^L.$$

Tensorial competitive learning scheme

Tensorial *competitive learning* scheme has four particular update laws:
(i) deterministic competitive learning:

$$\dot{w}_{ij}^L = f_i^L(u_j^L)[f_j^L(v_j^L) - w_{ij}^L];$$

(ii) differential competitive learning:

$$\dot{w}_{ij}^L = \dot{f}_i^L(u_j^L)[f_j^L(v_j^L) - w_{ij}^L];$$

(iii) random competitive learning, and

$$\dot{w}_{ij}^L = f_i^L(u_j^L)[f_j^L(v_j^L) - w_{ij}^L] + n_{ij}^L;$$

(iv) random differential competitive learning:

$$\dot{w}_{ij}^L = \dot{f}_i^L(u_j^L)[f_j^L(v_j^L) - w_{ij}^L] + n_{ij}^L;$$

General Neurodynamical Stability

Lyapunov stability

A dynamical system $\dot{x}_j = F_j(x_j), j = 1, ..., N$ is stable if some Lyapunov function L decreases along the system's trajectories: $\dot{L} \leq 0$. A system is asymptotically stable if it strictly decreases along the trajectories: $\dot{L} < 0$. In a stable equilibrium the trajectory may hover arbitrarily close to the equilibrium point without reaching it. In an asymptotically stable system the state trajectory reaches the equilibrium, and in general reaches it exponentially fast. Lyapunov functions are usually chosen in a quadratic form, representing (negative) potential $L = -\sum_{j=1}^{N} x_j^2$ or kinetic energy $L = -\sum_{j=1}^{N} \dot{x}_j^2$ of the system. By the chain rule, $\dot{L} = \frac{\partial L}{\partial x_j}\dot{x}_j$. In particular, gradient systems $\dot{x}_j = -\frac{\partial L}{\partial x_j}$ are stable.

For neural–networks purposes Lyapunov functions need only decrease and be bounded [Kos92].

Convergence versus stability

In a single neural network both neurons and synapses change in time. Therefore, we have three dynamical systems [Kos92]:
(i) the neural dynamical system $\dot{x}_j$,
(ii) the synaptic dynamical system $\dot{w}_{jk}$, and
(iii) the joint neuronal–synaptic dynamical system $(\dot{x}_j, \dot{w}_{jk})$.

Equilibrium is steady state (for fixed–point attractors).

Stability is neuronal equilibrium: $\dot{x}_j = 0$.

Convergence is synaptic equilibrium: $\dot{w}_{jk} = 0$.

Then global stability is joint neuronal–synaptic steady state:
both $\dot{x}_j = 0$ and $\dot{w}_{jk} = 0$ hold.

Neurons fluctuate faster than synapses. In feedback neural networks this dynamical asymmetry creates the stability–convergence dilemma: learning tends to destroy the neural patterns being learned.

GBAM Neurodynamical Classifier

System Architecture

Mathematically, the GBAM is a tensor field system (q, p, W) defined on a manifold M called the GBAM manifold. The system (q, p, W) includes two nonlinearly coupled (yet non–chaotic and stable) subsystems (see Figure 2.25): (i) activation (q, p)–dynamics, where q and p represent neuronal 1D tensor–fields, and (ii) self–organized learning W–dynamics, where W is a symmetric synaptic 2D tensor–field.

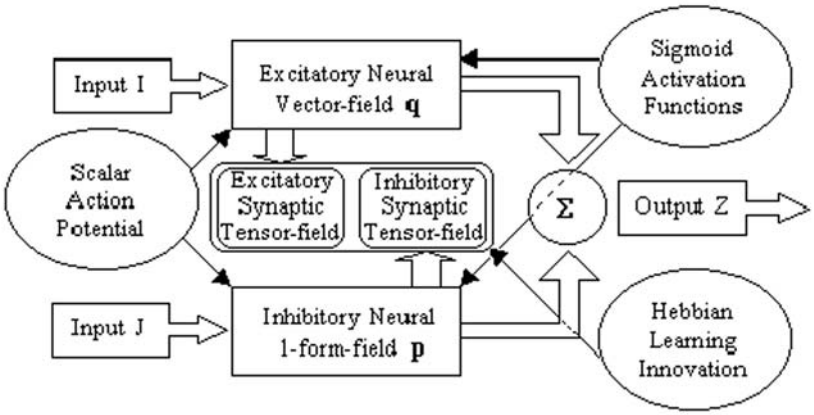


Fig. 2.25. Diagram of the GBAM neurodynamical system.

GBAM Activation Dynamics

The GBAM–manifold M can be viewed as a Banach space with a $C^\infty-$smooth structure on it, so that in each local chart U open in M, an nD smooth coordinate system U_α exists.

GBAM–activation (q, p)–dynamics, is defined as a system of two coupled, first–order oscillator tensor–fields, dual to each other, in a local Banach chart U_α, $\alpha = 1, ..., n$ on M:

1. An excitatory neural vector–field $q^\alpha = q^\alpha(t) : M \to TM$, being a cross–section of the tangent bundle TM; and

2. An inhibitory neural 1–form $p_\alpha = p_\alpha(t) : M \to T^*M$, being a cross–section of the cotangent bundle T^*M.

To start with conservative linear (q, p)–system, we postulate the GBAM scalar activation–potential V to be a negative bilinear form:

$$V = -\tfrac{1}{2}\omega_{\alpha\beta}q^\alpha q^\beta - \tfrac{1}{2}\omega^{\alpha\beta}p_\alpha p_\beta + q^\alpha p_\alpha, \qquad (\alpha, \beta = 1, ..., n), \qquad (2.747)$$

where n is the number of neurons in each neural field, while $\omega_{\alpha\beta}$ and $\omega^{\alpha\beta}$ represent respectively inhibitory–covariant and excitatory–contravariant components of the symmetric (with zero–trace) coupling GBAM synaptic tensor W.

The Lyapunov–stable, conservative, linear $(q, \mathbf{p})$–dynamics is given as a bidirectional (excitatory–inhibitory) gradient system:

$$\dot{q}^\alpha = -\frac{\partial V}{\partial p_\alpha} = \omega^{\alpha\beta} p_\beta - q^\alpha, \qquad \dot{p}_\alpha = -\frac{\partial V}{\partial q^\alpha} = \omega_{\alpha\beta} q^\beta - p_\alpha. \tag{2.748}$$

As W is a symmetric and zero–trace synaptic coupling tensor, the conservative linear dynamics (2.748) is equivalent to the rule that the state of each neuron (in both excitatory and inhibitory neural fields) is changed in time if and only if the scalar action potential V, defined by relation (2.747), is lowered. Therefore, the scalar action potential V is a monotonically non–increasing Lyapunov function $\dot{V} \leq 0$ for the conservative linear (q, p)–dynamics (2.748), which converges to a local minimum or ground state of V.

Applying the inputs I^α and J_α, we get the non–conservative linear (q, p)–system equations:

$$\dot{q}^\alpha = I^\alpha + \omega^{\alpha\beta} p_\beta - q^\alpha, \qquad \dot{p}_\alpha = J_\alpha + \omega_{\alpha\beta} q^\beta - p_\alpha. \tag{2.749}$$

Further, applying the sigmoid GBAM activation functions $S_\alpha(\cdot)$ and $S^\alpha(\cdot)$ to the synaptic product–terms, we get the non-conservative nonlinear (q, p)–system equations, which generalize the transient RC–circuit neurodynamical model:

$$\dot{q}^\alpha = I^\alpha + \omega^{\alpha\beta} S_\beta(p_\beta) - q^\alpha, \qquad \dot{p}_\alpha = J_\alpha + \omega_{\alpha\beta} S^\beta(q^\beta) - p_\alpha. \tag{2.750}$$

The equations in (2.750) represent a 2–input system that can be applied e.g., to classification of two–feature data. The generalization to an N–input system working in a ND *feature–space* is given by

$$\dot{q}^\alpha_e = I^\alpha_e + \omega^{\alpha\beta}_e S_\beta(p_\beta) - q^\alpha_e, \qquad \dot{p}^o_\alpha = I^o_\alpha + \omega^o_{\alpha\beta} S^\beta(q^\beta) - p^o_\alpha, \tag{2.751}$$

where $e(= 2, 4, ..., N)$ and $o(= 1, 3, ..., N-1)$ denote respectively even and odd partitions of the total sample of N features.

The GBAM model (2.751) gives a generalization of four well–known recurrent NN models:

1. Continuous Hopfield amplifier–circuit model [Hop84]

$$C_j \dot{v}_j = I_j - \frac{v_j}{R_j} + T_{ij} u_i, \qquad (i, j = 1, ..., N),$$

where $v_j = v_j(t)$ represent the activation potentials in the jth processing unit, C_j and R_j denote input capacitances and leakage resistances, $u_i = f_i[v_j(t)]$ are output functions from processing elements, and $T_{ij} = w_{ij}$ is the inverse of the resistors connection–matrix; and the functions f_i are sigmoidal.

2. Cohen–Grossberg general ART–system [CG83],

$$\dot{v}_j = -a_j(v_j)[b_j(v_j) - f_k(v_k)m_{jk}], \qquad (j = 1, ..., N),$$

with proved asymptotical stability.

3. Hecht–Nielsen counter–propagation network [Hec87],

$$\dot{v}_j = -Av_j + (B - v_j)I_j - v_j I_k,$$

where A, B are positive constants and I_j are input values for each processing unit.

4. Kosko's BAM (ABAM and RABAM) bidirectional models [Kos92]

$$\dot{v}_j = -a_j(v_j)[b_j(v_j) - f_k(v_k)m_{jk}],$$
$$\dot{u}_k = -a_k(u_k)[b_k(u_k) - f_j(u_j)m_{jk}],$$

which is globally stable for the cases of signal and random–signal Hebbian learning.

GBAM Self–Organized Learning Dynamics

The continuous (and at least C^1–differentiable) unsupervised update law for the coupling synaptic GBAM tensor–field W can be viewed both as an inhibitory–covariant Hebbian learning scheme, generalized from [Kos92]:

$$\dot{\omega}_{\alpha\beta} = -\omega_{\alpha\beta} + \Phi_{\alpha\beta}(q^\alpha, p_\alpha), \qquad (\alpha, \beta = 1, ..., n), \tag{2.752}$$

and, as an excitatory–contravariant Hebbian learning scheme:

$$\dot{\omega}^{\alpha\beta} = -\omega^{\alpha\beta} + \Phi^{\alpha\beta}(q^\alpha, p_\alpha), \tag{2.753}$$

where the three terms from the left to the right denote respectively the new–update value, the old value and the innovation of the synaptic tensor W. In this case the nonlinear (usually sigmoid) innovation functions $\Phi_{\alpha\beta}$ and $\Phi^{\alpha\beta}$ are defined by one of following four Hebbian models:

Signal Hebbian learning, with innovation in both variance–forms:

$$\begin{aligned} \Phi_{\alpha\beta} &= S_\alpha(q^\alpha)\, S_\beta(p_\beta), \\ \Phi^{\alpha\beta} &= S^\alpha(q^\alpha)\, S^\beta(p_\beta); \end{aligned} \tag{2.754}$$

Differential Hebbian learning, with innovation in both variance–forms:

$$\begin{aligned} \Phi_{\alpha\beta} &= S_\alpha(q^\alpha)S_\beta(p_\beta) + \dot{S}_\alpha(q^\alpha)\dot{S}_\beta(p_\beta), \\ \Phi^{\alpha\beta} &= S^\alpha(q^\alpha)S^\beta(p_\beta) + \dot{S}^\alpha(q^\alpha)\dot{S}^\beta(p_\beta), \end{aligned} \tag{2.755}$$

where $\dot{S}-$terms denote the so–called 'signal velocities' (for details see [Kos92]).

Random signal Hebbian learning, with innovation in both variance–forms:

$$\begin{aligned} \Phi_{\alpha\beta} &= S_\alpha(q^\alpha)\, S_\beta(p_\beta) + n_{\alpha\beta}, \\ \Phi^{\alpha\beta} &= S^\alpha(q^\alpha)\, S^\beta(p_\beta) + n^{\alpha\beta}, \end{aligned} \tag{2.756}$$

where $n_{\alpha\beta} = \{n_{\alpha\beta}(t)\}$, $n^{\alpha\beta} = \{n^{\alpha\beta}(t)\}$ respectively denote covariant and contravariant additive, zero–mean, Gaussian white–noise processes independent of the main innovation signal; and

Random differential signal Hebbian learning, with innovation in both variance–forms:

$$\begin{aligned} \Phi_{\alpha\beta} &= S_\alpha(q^\alpha) S_\beta(p_\beta) + \dot{S}_\alpha(q^\alpha)\dot{S}_\beta(p_\beta) + n_{\alpha\beta}, \\ \Phi^{\alpha\beta} &= S^\alpha(q^\alpha) S^\beta(p_\beta) + \dot{S}^\alpha(q^\alpha)\dot{S}^\beta(p_\beta) + n^{\alpha\beta}. \end{aligned} \tag{2.757}$$

Total GBAM $(q, p, W)-$neurodynamics and biological interpretation

Total GBAM tensorial neurodynamics is defined as a union of the neural oscillatory activation (q, p)–dynamics (2.751) and the synaptic learning W–dynamics (2.757), namely

$$\begin{aligned} \dot{q}_e^\alpha &= I_e^\alpha + \omega_e^{\alpha\beta} S_\beta(p_\beta) - q_e^\alpha, \\ \dot{p}_\alpha^o &= I_\alpha^o + \omega_{\alpha\beta}^o S^\beta(q^\beta) - p_\alpha^o, \\ \dot{\omega}_e^{\alpha\beta} &= -\omega_e^{\alpha\beta} + \Phi_e^{\alpha\beta}(q^\alpha, p_\alpha), \\ \dot{\omega}_{\alpha\beta}^o &= -\omega_{\alpha\beta}^o + \Phi_{\alpha\beta}^o(q^\alpha, p_\alpha), \qquad (\alpha, \beta = 1, ..., n), \end{aligned} \tag{2.758}$$

where the tensorial innovation Φ–functions are given by one of Hebbian models (2.754–2.757), $\alpha(= 1, ..., n)$ is the number of continuous–time (or, graded–response) neurons in each neural–activation field, $e(= 2, 4, ..., N)$ and $o(= 1, 3, ..., N-1)$ denote respectively even and odd partitions of the total sample of N features.

Artificial neural networks are generally inspired by biological neural systems, but in fact, some important features of biological systems are not present in most artificial neural networks. In particular, unidirectional neural networks, which include all associative neural networks except the BAM model introduced by [Kos92], do not resemble oscillatory biological neural systems. GBAM is a generalization of Kosko's ABAM and RABAM neural systems and inherits their oscillatory (excitatory/inhibitory) neuro-synaptic behavior. Such oscillatory behavior is a basic characteristic of a number of biological systems. Examples of similar oscillatory neural ensembles in the human nervous system are:

– Motoneurons and Renshaw interneurons in the spinal cord;
– Pyramidal and basket cells in the hippocampus;
– Mitral and granule cells in the olfactory bulb;
– Pyramidal cells and thalamic inter–neurons in cortico–thalamic system;
– Interacting excitatory and inhibitory populations of neurons found in the cerebellum, olfactory cortex, and neocortex, all representing the basic

mechanisms for the generation of oscillating (EEG–monitored) activity in the brain.

Therefore, GBAM can be considered as a model for any of above–mentioned oscillatory biological neural systems.

Self–Organizing Lie–Derivative Neuro–Classifier

A *Lie–derivative neuro–classifier* is a self–organized, associative–memory machine, represented by oscillatory (excitatory/inhibitory) tensor–field–system (x, v, ω) on the Banach manifold M. It consists of continual neural activation (x, y)–dynamics and self–organizing synaptic learning ω–dynamics.

The continual activation $(x, y)-$dynamics, is defined as a system of two coupled, first–order oscillator tensor–fields, dual to each other, in a local Banach chart U_α, $(\alpha = 1, ..., n)$ on M:

1) an excitatory neural vector–field $x^i = x^i(t) : M \to TM$, representing a cross–section of the tangent bundle TM; and

2) an inhibitory neural one–form $y_i = y_i(t) : M \to T^*M$, representing a cross–section of the cotangent bundle T^*M.

The self–organized learning $\omega-$dynamics is performed on a second–order symmetrical synaptic tensor–field $\omega = \omega(t)$, given by its covariant components $\omega_{ij} = \omega_{ij}(t)$ and its contravariant components $\omega^{ij} = \omega^{ij}(t)$, where $i, j = 1, ..., n$.

Starting with the Lyapunov–stable, negative scalar neural action potential:

$$U = -\tfrac{1}{2}(\omega_{ij}x^i x^j + \omega^{ij}y_i y_j), \qquad (i, j = 1, ..., n),$$

the (x, y)–dynamics is given in two versions, which are compared and contrasted:

(1) the **Lie–linear** neurodynamics with first–order Lie derivatives

$$\dot{x}^i = J^i + L_X U, \qquad \dot{y}_i = J_i + L_Y U,$$

and

(2) the **Lie–quadratic** neurodynamics with both first and second–order Lie derivatives

$$\dot{x}^i = J^i + L_X U + L_X L_X U, \qquad \dot{y}_i = J_i + L_Y U + L_Y L_Y U,$$

where $X = S_i(x^i)$, $Y = S_i(y_i)$, S_i represent sigmoid activation functions, while $L_X L_X, L_Y L_Y : F(M) \to F(M)$ denote the *second–order* (*iterated*) *Lie derivatives*.

Self–organized learning ω–dynamics is presented in the form of differential Hebbian learning scheme in both covariant and contravariant forms

$$\dot{\omega}_{ij} = -\omega_{ij} + S_i(x^i)S_j(y_j) + \dot{S}_i(x^i)\dot{S}_j(y_j), \qquad \text{and}$$
$$\dot{\omega}^{ij} = -\omega^{ij} + S_i(x^i)S_j(y_j) + \dot{S}_i(x^i)\dot{S}_j(y_j), \qquad (i, j = 1, ..., n),$$

respectively.

Self–Organizing Lie–Poisson Neuro–Classifier

A *Lie–Poisson neuro–classifier* is a tensor–field–system $\{\mu\} = (q, p, \omega)$ on a Poisson manifold $(\mathbf{g}^*, \{F, H(\mu)\}_{\pm}(\mu))$. Like a GBAM neuro–classifier, it consists of continual activation $(q, p)-$dynamics and self–organized learning ω–dynamics. In this case, both dynamics are defined by "neural activation form" of the abstract Lie–Poisson evolution equation

$$\dot{F} = \{S(F), H(\mu)\}, \tag{2.759}$$

where $S(\cdot) = \tanh(\cdot)$ denotes the sigmoid activation function. A Hamiltonian function $H(\mu)$, representing the total network energy, is given in the form

$$H(\mu) = \frac{1}{2}\omega_{ij}\delta^{ij} + \frac{1}{2}\omega^{ij}\delta_{ij}, \qquad (i, j = 1, ..., n),$$

where δ^{ij} and δ_{ij} are Kronecker tensors, while $\omega^{ij} = \omega^{ij}(q^i)$ and $\omega_{ij} = \omega_{ij}(p_i)$ correspond to the contravariant and covariant components of the functional–coupling synaptic tensor $\omega = \omega(q, p)$, defined respectively by

$$\omega^{ij} = \varepsilon\, q^i q^j, \qquad \omega_{ij} = \tau\, p_i p_j,$$

with random coefficients ε and τ.

Activation (q, p)–dynamics are given by

$$\dot{q}^i = I^i + \{S(q^i), H(\mu)\}, \qquad \dot{p}_i = J_i + \{S(p_i), H(\mu)\},$$

where I^i and J_i represent the two input features.

Two types of self–organized learning ω–dynamics are presented and compared:

Lie–Poisson learning dynamics, in which synaptic update law is given by inhibitory–covariant and excitatory–contravariant form of equation (2.759):

$$\dot{\omega}_{ij} = \{S(\omega_{ij}), H(\mu)\}, \qquad \dot{\omega}^{ij} = \{S(\omega^{ij}), H(\mu)\},$$

respectively.

Differential Hebbian learning (see [Kos92] for details), in both inhibitory–covariant and excitatory–contravariant learning form:

$$\dot{\omega}_{ij} = -\omega_{ij} + \Phi_{ij}(q^i, p_i), \qquad \dot{\omega}^{ij} = -\omega^{ij} + \Phi^{ij}(q^i, p_i),$$

with innovations defined in both variance–forms as:

$$\Phi_{ij} = S_i(q^i)S_j(p_j) + \dot{S}_i(q^i)\dot{S}_j(p_j), \qquad \Phi^{ij} = S^i(q^i)S^j(p_j) + \dot{S}^i(q^i)\dot{S}^j(p_j).$$

2.5.3 Oscillatory Phase Neurodynamics

In coupled oscillatory neuronal systems, under suitable conditions, the original dynamics can be reduced theoretically to a simpler phase dynamics. The state of the ith neuronal oscillatory system can be then characterized by a single phase variable φ_i representing the timing of the neuronal firings. The typical dynamics of *oscillator neural networks* are described by the *Kuramoto model* [Kur84, HI97, Str00], consisting of N equally weighted, all–to–all, phase–coupled limit–cycle oscillators, where each oscillator has its own natural frequency ω_i drawn from a prescribed distribution function:

$$\dot{\varphi}_i \;=\; \omega_i + \frac{K}{N}\sum_{i=1}^{N} J_{ij}\sin(\varphi_j - \varphi_i + \beta_{ij}). \tag{2.760}$$

Here, J_{ij} and β_{ij} are parameters representing the effect of the interaction, while $K \geq 0$ is the coupling strength. For simplicity, we assume that all natural frequencies ω_i are equal to some fixed value ω_0. We can then eliminate ω_0 by applying the transformation $\varphi_i \to \varphi_i + \omega_0 t$. Using the complex representation $W_i = \exp(i\varphi_i)$ and $C_{ij} = J_{ij}\exp(i\beta_{ij})$ in (2.760), it is easily found that all neurons relax toward their stable equilibrium states, in which the relation $W_i = h_i/|h_i|$ ($h_i = \sum_{j=1}^{N} C_{ij}W_j$) is satisfied. Following this line of reasoning, as a synchronous update version of the oscillator neural network we can consider the alternative discrete form [AN99],

$$W_i(t+1) = \frac{h_i(t)}{|h_i(t)|}, \qquad h_i(t) = C_{ij}W_j(t). \tag{2.761}$$

Now we will attempt to construct an extended model of the oscillator neural networks to retrieve sparsely coded phase patterns. In equation (2.761), the complex quantity h_i can be regarded as the local field produced by all other neurons. We should remark that the phase of this field, h_i, determines the timing of the ith neuron at the next time step, while the amplitude $|h_i|$ has no effect on the retrieval dynamics (2.761). It seems that the amplitude can be thought of as the strength of the local field with regard to emitting spikes. Pursuing this idea, as a natural extension of the original model we stipulate that the system does not fire and stays in the resting state if the amplitude is smaller than a certain value. Therefore, we consider a network of N oscillators whose dynamics are governed by

$$W_i(t+1) = f(|h_i(t)|)\frac{h_i(t)}{|h_i(t)|}, \qquad h_i(t) = C_{ij}W_j(t). \tag{2.762}$$

We assume that $f(x) = \Theta(x - H)$, where the real variable H is a threshold parameter and $\Theta(x)$ is the unit step function; $\Theta(x) = 1$ for $x \geq 0$ and 0 otherwise. Therefore, the amplitude $|W_i^t|$ assumes a value of either 1 or 0, representing the state of the ith neuron as firing or non–firing. Consequently,

the neuron can emit spikes when the amplitude of the local field $h_i(t)$ is greater than the threshold parameter H.

Now, let us define a set of P patterns to be memorized as $\xi_i^\mu = A_i^\mu \exp(i\theta_i^\mu)$ $(\mu = 1, 2, \ldots, P)$, where θ_i^μ and A_i^μ represent the phase and the amplitude of the ith neuron in the μth pattern, respectively. For simplicity, we assume that the θ_i^μ are chosen at random from a uniform distribution between 0 and 2π. The amplitudes A_i^μ are chosen independently with the probability distribution

$$P(A_i^\mu) = a\delta(A_i^\mu - 1) + (1 - a)\delta(A_i^\mu),$$

where a is the mean activity level in the patterns. Note that, if $H = 0$ and $a = 1$, this model reduces to (2.761).

For the synaptic efficacies, to realize the function of the associative memory, we adopt the *generalized Hebbian rule* in the form

$$C_{ij} = \frac{1}{aN}\xi_i^\mu \tilde{\xi}_j^\mu, \tag{2.763}$$

where $\tilde{\xi}_j^\mu$ denotes the complex conjugate of ξ_j^μ. The overlap $M_\mu(t)$ between the state of the system and the pattern μ at time t is given by

$$M_\mu(t) = m_\mu(t)\, e^{i\varphi_\mu(t)} = \frac{1}{aN}\tilde{\xi}_j^\mu W_j(t), \tag{2.764}$$

In practice, the rotational symmetry forces us to measure the correlation of the system with the pattern μ in terms of the amplitude component $m_\mu(t) = |M_\mu(t)|$.

Let us consider the situation in which the network is recalling the pattern ξ_i^1; that is, $m_1(t) = m(t) \sim O(1)$ and $m_\mu(t) \sim O(1/\sqrt{N})(\mu \neq 1)$. The local field $h_i(t)$ in (2.762) can then be separated as

$$h_i(t) = C_{ij}W_j(t) = m_t e^{i\varphi_1(t)}\xi_i^1 + z_i(t), \tag{2.765}$$

where $z_i(t)$ is defined by

$$z_i(t) = \frac{1}{aN}\xi_i^\mu \tilde{\xi}_j^\mu W_j(t). \tag{2.766}$$

The first term in (2.765) acts to recall the pattern, while the second term can be regarded as the noise arising from the other learned patterns. The essential point in this analysis is the treatment of the second term as *complex Gaussian noise* characterized by

$$< z_i(t) >= 0, \qquad < |z_i(t)|^2 >= 2\sigma(t)^2. \tag{2.767}$$

We also assume that $\varphi_1(t)$ remains a constant, that is, $\varphi_1(t) = \varphi_0$. By applying the method of statistical neurodynamics to this model under the above assumptions [AN99], we can study the retrieval properties analytically. As a

result of such analysis we have found that the retrieval process can be characterized by some macroscopic order parameters, such as $m(t)$ and $\sigma(t)$.

From (2.764), we find that the overlap at time $t+1$ is given by

$$m(t+1) = \left\langle\!\left\langle f(|m(t)+z(t)|)\frac{m(t)+z(t)}{|m(t)+z(t)|}\right\rangle\!\right\rangle, \tag{2.768}$$

where $\langle\!\langle \cdots \rangle\!\rangle$ represents an average over the complex Gaussian $z(t)$ with mean 0 and variance $2\sigma(t)^2$. For the noise $z(t+1)$, in the limit $N \to \infty$ we get [AN99]

$$\begin{aligned} z_i(t+1) \sim \frac{1}{aN}\sum_{j=1}^{N}\sum_{\mu=2}^{P}\xi_i^\mu\tilde{\xi}_j^\mu f(|h_{j,\mu}(t)|)\frac{h_{j,\mu}(t)}{|h_{j,\mu}(t)|} \\ + z_i(t)\left(\frac{f'(|h_{j,\mu}(t)|)}{2} + \frac{f(|h_{j,\mu}(t)|)}{2|h_{j,\mu}(t)|}\right), \end{aligned} \tag{2.769}$$

where $h_{j,\mu}(t) = 1/aN\sum_{k=1}^{N}\sum_{\nu\neq mu,1}^{P}\xi_j^\nu\tilde{\xi}_k^\nu W_k(t)$.

Kuramoto Synchronization Model

The microscopic individual level dynamics of the Kuramoto model (2.760) is easily visualized by imagining oscillators as points running around on the unit circle. Due to rotational symmetry, the average frequency $\Omega = \sum_{i=1}^{N}\omega_i/N$ can be set to 0 without loss of generality; this corresponds to observing dynamics in the co–rotating frame at frequency Ω.

The governing equation (2.760) for the ith oscillator phase angle φ_i can be simplified to

$$\dot{\varphi}_i = \omega_i + \frac{K}{N}\sum_{i=1}^{N}\sin(\varphi_j - \varphi_i), \quad 1 \le i \le N. \tag{2.770}$$

It is known that as K is increased from 0 above some critical value K_c, more and more oscillators start to get synchronized (or phase–locked) until all the oscillators get fully synchronized at another critical value of K_{tp}. In the choice of $\Omega = 0$, the fully synchronized state corresponds to an exact steady state of the 'detailed', fine–scale problem in the co–rotating frame.

Such synchronization dynamics can be conveniently summarized by considering the fraction of the synchronized (phase–locked) oscillators, and conventionally described by a *complex–valued order parameter* [Kur84, Str00], $re^{i\psi} = \frac{1}{N}e^{i\varphi_j}$, where the radius r measures the phase coherence, and ψ is the average phase angle.

Transition from Full to Partial Synchronization

Following [MK05], here we restate certain facts about the nature of the second transition mentioned above, a transition between the full and the partial synchronization regime at $K = K_{tp}$, in the direction of decreasing K.

A fully synchronized state in the continuum limit corresponds to the solution to the mean–field type alternate form of equation (2.770),

$$\dot{\varphi}_i = \Omega = \omega_i + rK \sum_{i=1}^{N} \sin(\psi - \varphi_i), \tag{2.771}$$

where Ω is the common angular velocity of the fully synchronized oscillators (which is set to 0 in our case). Equation (2.771) can be further rewritten as

$$\frac{\Omega - \omega_i}{rK} = \sum_{i=1}^{N} \sin(\psi - \varphi_i), \tag{2.772}$$

where the absolute value of the r.h.s is bounded by unity.

As K approaches K_{tp} from above, the l.h.s for the 'extreme' oscillator (the oscillator in a particular family that has the maximum value of $|\Omega - \omega_i|$) first exceeds unity, and a real–valued solution to (2.772) ceases to exist. Different random draws of ω_i's from $g(\omega)$ for a finite number of oscillators result in slightly different values of K_{tp}. K_{tp} appears to follow the Gumbel type extreme distribution function [KN00], just as the maximum values of $|\Omega - \omega_i|$ do:

$$p(K_{tp}) = \sigma^{-1} e^{-(K_{tp}-\mu)/\sigma} \exp[-e^{-(K_{tp}-\mu)/\sigma}],$$

where σ and μ are parameters.

Lyapunov Chaotic Synchronization

The notion of *conditional Lyapunov exponents* was introduced by Pecora and Carroll in their study of synchronization of chaotic systems. First, in [PC91], they generalized the idea of driving a stable system to the situation when the drive signal is chaotic. This leaded to the concept of conditional Lyapunov exponents and also generalized the usual criteria of the linear stability theorem. They showed that driving with chaotic signals can be done in a robust fashion, rather insensitive to changes in system parameters. The calculation of the stability criteria leaded naturally to an estimate for the convergence of the driven system to its stable state. The authors focussed on a homogeneous driving situation that leaded to the construction of synchronized chaotic subsystems. They applied these ideas to the Lorenz and Rössler systems, as well as to an electronic circuit and its numerical model. Later, in [PC98], they showed that many coupled oscillator array configurations considered in the literature could be put into a simple form so that determining the stability

of the synchronous state could be done by a master stability function, which could be tailored to one's choice of stability requirement. This solved, once and for all, the problem of synchronous stability for any linear coupling of that oscillator.

It turns out, that, like the full Lyapunov exponent, the conditional exponents are well defined ergodic invariants, which are reliable quantities to quantify the relation of a global dynamical system to its constituent parts and to characterize dynamical self–organization [Men98].

Given a dynamical system defined by a map $f : M \to M$, with $M \subset R^m$ the conditional exponents associated to the splitting $R^k \times R^{m-k}$ are the eigenvalues of the limit

$$\lim_{n\to\infty} \left(D_k f^{n*}(x) D_k f^n(x)\right)^{\frac{1}{2n}} ,$$

where $D_k f^n$ is the $k \times k$ diagonal block of the full Jacobian.

Mendes [Men98] proved that existence of the conditional Lyapunov exponents as well–defined ergodic invariants was guaranteed under the same conditions that established the existence of the Lyapunov exponents.

Recall that for measures μ that are absolutely continuous with respect to the Lebesgue measure of M or, more generally, for measures that are smooth along unstable directions (SBR measures) Pesin's [Pes77] identity holds

$$h(\mu) = \sum_{\lambda_i > 0} \lambda_i,$$

relating *Kolmogorov–Sinai entropy* $h(\mu)$ to the sum of the Lyapunov exponents. By analogy we may define the *conditional exponent entropies* [Men98] associated to the splitting $R^k \times R^{m-k}$ as the sum of the positive conditional exponents counted with their multiplicity

$$h_k(\mu) = \sum_{\xi_i^{(k)} > 0} \xi_i^{(k)}, \qquad h_{m-k}(\mu) = \sum_{\xi_i^{(m-k)} > 0} \xi_i^{(m-k)}.$$

The Kolmogorov–Sinai entropy of a dynamical system measures the rate of information production per unit time. That is, it gives the amount of randomness in the system that is not explained by the defining equations (or the minimal model [CY89]). Hence, the conditional exponent entropies may be interpreted as a measure of the randomness that would be present if the two parts $S^{(k)}$ and $S^{(m-k)}$ were uncoupled. The difference $h_k(\mu) + h_{m-k}(\mu) - h(\mu)$ represents the effect of the coupling.

Given a dynamical system S composed of N parts $\{S_k\}$ with a total of m degrees of freedom and invariant measure μ, one defines a *measure of dynamical self–organization* $I(S, \Sigma, \mu)$ as

$$I(S, \Sigma, \mu) = \sum_{k=1}^{N} \left\{h_k(\mu) + h_{m-k}(\mu) - h(\mu)\right\} .$$

For each system S, this quantity will depend on the partition Σ into N parts that one considers. $h_{m-k}(\mu)$ always denotes the conditional exponent entropy of the complement of the subsystem S_k. Being constructed out of ergodic invariants, $I(S, \Sigma, \mu)$ is also a well–defined ergodic invariant for the measure μ. $I(S, \Sigma, \mu)$ is formally similar to a mutual information. However, not being strictly a mutual information, in the information theory sense, $I(S, \Sigma, \mu)$ may take negative values.

2.5.4 Neural Path–Integral Model for the Cerebellum

Recall that human motion is naturally driven by synergistic action of more than 600 skeletal muscles. While the muscles generate driving torques in the moving joints, subcortical neural system performs both local and global (loco)motion control: first reflexly controlling contractions of individual muscles, and then orchestrating all the muscles into synergetic actions in order to produce efficient movements. While the local reflex control of individual muscles is performed on the *spinal control level*, the global integration of all the muscles into coordinated movements is performed within the *cerebellum*.

All hierarchical subcortical neuro–muscular physiology, from the bottom level of a single muscle fiber, to the top level of cerebellar muscular synergy, acts as a *temporal* $< out|in >$ *reaction*, in such a way that the higher level acts as a command/control space for the lower level, itself representing an abstract image of the lower one:

1. At the *muscular level*, we have *excitation–contraction dynamics* [Hat77a, Hat78, Hat77b], in which $< out|in >$ is given by the following sequence of nonlinear diffusion processes: *neural-action-potential* $\rightsquigarrow$ *synaptic-potential*
   $\rightsquigarrow$*muscular-action-potential*$\rightsquigarrow$*excitation-contraction-coupling* $\rightsquigarrow$ *muscle-tension-generating* [Iva91, II05]. Its purpose is the generation of muscular forces, to be transferred into driving torques within the joint anatomical geometry.
2. At the *spinal level*, $< out|in >$ is given by *autogenetic–reflex stimulus–response control* [Hou79]. Here we have a neural image of all individual muscles. The main purpose of the spinal control level is to give both positive and negative feedbacks to stabilize generated muscular forces within the 'homeostatic' (or, more appropriately, 'homeokinetic') limits. The individual muscular actions are combined into flexor–extensor (or agonist–antagonist) pairs, mutually controlling each other. This is the mechanism of *reciprocal innervation of agonists and inhibition of antagonists*. It has a purely mechanical purpose to form the so–called *equivalent muscular actuators* (EMAs), which would generate driving torques $T_i(t)$ for all movable joints.
3. At the *cerebellar level*, $< out|in >$ is given by *sensory–motor integration* [HBB96]. Here we have an abstracted image of all autogenetic reflexes.

The main purpose of the cerebellar control level is integration and fine tuning of the action of all active EMAs into a synchronized movement, by *supervising* the individual autogenetic reflex circuits. At the same time, to be able to perform in new and unknown conditions, the cerebellum is continuously adapting its own neural circuitry by unsupervised (self–organizing) learning. Its action is subconscious and automatic, both in humans and in animals.

Naturally, we can ask the question: Can we assign a single $< out|in >$ measure to all these neuro–muscular stimulus–response reactions? We think that we can do it; so in this Letter, we propose the concept of *adaptive sensory–motor transition amplitude* as a unique measure for this temporal $< out|in >$ relation. Conceptually, this $< out|in > -amplitude$ can be formulated as the '*neural path integral*':

$$< out|in > \equiv \underset{amplitude}{\langle motor|sensory\rangle} = \int \mathcal{D}[w,x]\, \mathrm{e}^{\mathrm{i}\, S[x]}. \tag{2.773}$$

Here, the integral is taken over all *activated* (or, 'fired') *neural pathways* $x^i = x^i(t)$ of the cerebellum, connecting its input $sensory-$state with its output $motor-$state, symbolically described by *adaptive neural measure* $\mathcal{D}[w,x]$, defined by the weighted product (of discrete time steps)

$$\mathcal{D}[w,x] = \lim_{n\longrightarrow\infty} \prod_{t=1}^{n} w^i(t)\, dx^i(t),$$

in which the *synaptic weights* $w^i = w^i(t)$, included in all active neural pathways $x^i = x^i(t)$, are updated by the unsupervised *Hebbian–like learning* rule 1.245, namely

$$w^i(t+1) = w^i(t) + \frac{\sigma}{\eta}(w^i_d(t) - w^i_a(t)), \tag{2.774}$$

where $\sigma = \sigma(t)$, $\eta = \eta(t)$ represent local neural *signal* and *noise* amplitudes, respectively, while superscripts d and a denote *desired* and *achieved* neural states, respectively. Theoretically, equations (2.773–2.774) define an $\infty-$*dimensional neural network*. Practically, in a computer simulation we can use $10^7 \leq n \leq 10^8$, roughly corresponding to the number of neurons in the cerebellum.

The exponent term $S[x]$ in equation (2.773) represents the *autogenetic–reflex action*, describing reflexly–induced motion of all active EMAs, from their initial $stimulus-$state to their final $response-$state, along the family of extremal (i.e., Euler–Lagrangian) paths $x^i_{\min}(t)$. ($S[x]$ is properly derived in (2.777–2.778) below.)

Spinal Autogenetic Reflex Control

Recall (from Introduction) that at the spinal control level we have the autogenetic reflex *motor servo* [Hou79], providing the local, reflex feedback loops for

individual muscular contractions. A voluntary contraction force F of human skeletal muscle is reflexly excited (positive feedback $+F^{-1}$) by the responses of its *spindle receptors* to stretch and is reflexly inhibited (negative feedback $-F^{-1}$) by the responses of its *Golgi tendon organs* to contraction. Stretch and unloading reflexes are mediated by combined actions of several autogenetic neural pathways, forming the *motor servo*.

In other words, branches of the afferent fibers also synapse with with interneurons that inhibit motor neurons controlling the antagonistic muscles – *reciprocal inhibition*. Consequently, the stretch stimulus causes the antagonists to relax so that they cannot resists the shortening of the stretched muscle caused by the main reflex arc. Similarly, firing of the Golgi tendon receptors causes inhibition of the muscle contracting too strong and simultaneous *reciprocal activation* of its antagonist. Both mechanisms of reciprocal inhibition and activation performed by the autogenetic circuits $+F^{-1}$ and $-F^{-1}$, serve to generate the well–tuned EMA–driving torques T_i.

Now, once we have properly defined the symplectic musculo–skeletal dynamics [Iva04] on the biomechanical (momentum) phase–space manifold T^*M^N, we can proceed in formalizing its hierarchical subcortical neural control. By introducing the *coupling Hamiltonians* $H^m = H^m(q,p)$, selectively corresponding *only* to the $M \leq N$ *active joints*, we define the *affine Hamiltonian control function* $H_{aff} : T^*M^N \to \mathbb{R}$, in local canonical coordinates on T^*M^N given by (adapted from [NS90] for the biomechanical purpose)

$$H_{aff}(q,p) = H_0(q,p) - H^m(q,p)\, T_m, \qquad (m = 1, \ldots, M \leq N), \tag{2.775}$$

where $T_m = T_m(t,q,p)$ are affine feedback torque one–forms, different from the initial driving torques T_i acting in all the joints. Using the affine Hamiltonian function (2.775), we get the *affine Hamiltonian servo–system* [Iva04],

$$\begin{aligned} \dot{q}^i &= \frac{\partial H_0(q,p)}{\partial p_i} - \frac{\partial H^m(q,p)}{\partial p_i} T_m, \\ \dot{p}_i &= -\frac{\partial H_0(q,p)}{\partial q^i} + \frac{\partial H^m(q,p)}{\partial q^i} T_m, \\ q^i(0) = q^i_0, \quad p_i(0) &= p^0_i, \qquad (i = 1, \ldots, N; \quad m = 1, \ldots, M \leq N). \end{aligned} \tag{2.776}$$

The affine Hamiltonian control system (2.776) gives our formal description for the autogenetic spinal motor–servo for all $M \leq N$ activated (i.e., working) EMAs.

Cerebellum – the Comparator

Having, thus, defined the spinal reflex control level, we proceed to model the top subcortical commander/controller, the *cerebellum*. It is a brain region anatomically located at the bottom rear of the head (the hindbrain), directly above the brainstem, which is important for a number of subconscious and

automatic motor functions, including motor learning. It processes information received from the motor cortex, as well as from proprioceptors and visual and equilibrium pathways, and gives 'instructions' to the motor cortex and other subcortical motor centers (like the basal nuclei), which result in proper balance and posture, as well as smooth, coordinated skeletal movements, like walking, running, jumping, driving, typing, playing the piano, etc. Patients with cerebellar dysfunction have problems with precise movements, such as walking and balance, and hand and arm movements. The cerebellum looks *similar in all animals*, from fish to mice to humans. This has been taken as evidence that it performs a common function, such as regulating motor learning and the timing of movements, in all animals. Studies of simple forms of motor learning in the vestibulo–ocular reflex and eye–blink conditioning are demonstrating that timing and amplitude of learned movements are encoded by the cerebellum.

The cerebellum is responsible for coordinating precisely timed $< out|in >$ activity by integrating motor output with ongoing sensory feedback. It receives extensive projections from sensory–motor areas of the cortex and the periphery and directs it back to premotor and motor cortex [Ghe90, Ghe91]. This suggests a role in sensory–motor integration and the timing and execution of human movements. The cerebellum stores patterns of motor control for frequently performed movements, and therefore, its circuits are changed by experience and training. It was termed the *adjustable pattern generator* in the work of J. Houk and collaborators [HBB96]. Also, it has become the inspiring 'brain–model' in the recent robotic research [SA98, Sch98].

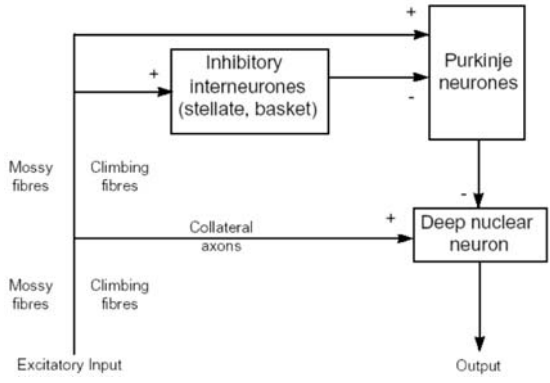


Fig. 2.26. Schematic $< out|in >$ organization of the primary cerebellar circuit. In essence, excitatory inputs, conveyed by collateral axons of Mossy and Climbing fibers activate directly neurones in the Deep cerebellar nuclei. The activity of these latter is also modulated by the inhibitory action of the cerebellar cortex, mediated by the Purkinje cells.

Comparing the number of its neurons ($10^7 - 10^8$), to the size of conventional neural networks, suggests that artificial neural nets *cannot* satisfactorily model the function of this sophisticated 'super–bio–computer', as its dimensionality is virtually infinite. Despite a lot of research dedicated to its structure and function (see [HBB96] and references there cited), the real nature of the cerebellum still remains a 'mystery'.

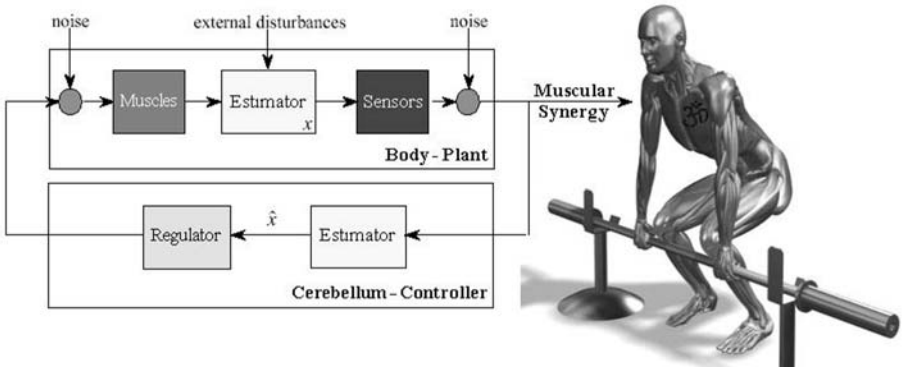


Fig. 2.27. The cerebellum as a motor controller.

The main function of the cerebellum as a motor controller is depicted in Figure 2.27. A coordinated movement is easy to recognize, but we know little about how it is achieved. In search of the neural basis of coordination, a model of spinocerebellar interactions was recently presented in [AG05], in which the structural and functional organizing principle is a division of the cerebellum into discrete micro–complexes. Each micro–complex is the recipient of a specific motor error signal - that is, a signal that conveys information about an inappropriate movement. These signals are encoded by spinal reflex circuits and conveyed to the cerebellar cortex through climbing fibre afferents. This organization reveals salient features of cerebellar information processing, but also highlights the importance of systems level analysis for a fuller understanding of the neural mechanisms that underlie behavior.

Hamiltonian Action and Neural Path Integral

Here, we propose a *quantum–like adaptive control* approach to modelling the 'cerebellar mystery'. Corresponding to the affine Hamiltonian control function (2.775) we define the *affine Hamiltonian control action*,

$$S_{aff}[q,p] = \int_{t_{in}}^{t_{out}} d\tau \left[p_i \dot{q}^i - H_{aff}(q,p) \right] . \tag{2.777}$$

From the affine Hamiltonian action (2.777) we further derive the associated expression for the *neural phase–space path integral* (in normal units),

representing the *cerebellar sensory–motor amplitude* $< out|in >$,

$$\begin{aligned} \left\langle q^i_{out}, p_i^{out} | q^i_{in}, p_i^{in} \right\rangle &= \int \mathcal{D}[w,q,p]\, \mathrm{e}^{\mathrm{i}\, S_{aff}[q,p]} \qquad (2.778) \\ &= \int \mathcal{D}[w,q,p] \exp\left\{ \mathrm{i} \int_{t_{in}}^{t_{out}} d\tau \left[p_i \dot{q}^i - H_{aff}(q,p) \right] \right\}, \\ \text{with} \qquad \int \mathcal{D}[w,q,p] &= \int \prod_{\tau=1}^{n} \frac{w^i(\tau) dp_i(\tau) dq^i(\tau)}{2\pi}, \end{aligned}$$

where $w_i = w_i(t)$ denote the cerebellar synaptic weights positioned along its neural pathways, being continuously updated using the Hebbian–like self–organizing learning rule (2.774). Given the transition amplitude $< out|in >$ (2.778), the *cerebellar sensory–motor transition probability* is defined as its absolute square, $| < out|in > |^2$.

In (2.778), $q^i_{in} = q^i_{in}(t)$, $q^i_{out} = q^i_{out}(t)$; $p_i^{in} = p_i^{in}(t)$, $p_i^{out} = p_i^{out}(t)$; $t_{in} \leq t \leq t_{out}$, for all discrete time steps, $t = 1, ..., n \longrightarrow \infty$, and we are allowing for the affine Hamiltonian $H_{aff}(q,p)$ to depend upon all the $(M \leq N)$ EMA–angles and angular momenta collectively. Here, we actually systematically took a discretized differential time limit of the form $t_\sigma - t_{\sigma-1} \equiv d\tau$ (both σ and τ denote discrete time steps) and wrote $\frac{(q^i_\sigma - q^i_{\sigma-1})}{(t_\sigma - t_{\sigma-1})} \equiv \dot{q}^i$. For technical details regarding the path integral calculations on Riemannian and symplectic manifolds (including the standard regularization procedures), see [Kla97, Kla00].

Now, motor learning occurring in the cerebellum can be observed using functional MR imaging, showing changes in the cerebellar action potential, related to the motor tasks (see, e.g., [MA02]). To account for these electro–physiological currents, we need to add the *source* term $J_i(t)q^i(t)$ to the affine Hamiltonian action (2.777), (the current $J_i = J_i(t)$ acts as a source $J_i A^i$ of the *cerebellar electrical potential* $A^i = A^i(t)$),

$$S_{aff}[q,p,J] = \int_{t_{in}}^{t_{out}} d\tau \left[p_i \dot{q}^i - H_{aff}(q,p) + J_i q^i \right],$$

which, subsequently gives the cerebellar path integral with the action potential source, coming either from the motor cortex or from other subcortical areas.

Note that the standard *Wick rotation*: $t \mapsto it$ (see [Kla97, Kla00]), makes all our path integrals real, i.e.,

$$\int \mathcal{D}[w,q,p]\, \mathrm{e}^{\mathrm{i}\, S_{aff}[q,p]} \quad \underrightarrow{Wick} \quad \int \mathcal{D}[w,q,p]\, \mathrm{e}^{-\, S_{aff}[q,p]},$$

while their subsequent discretization gives the standard thermodynamic *partition functions*,

$$Z = \sum_j \mathrm{e}^{-w_j E^j / T}, \qquad (2.779)$$

where E^j is the energy eigenvalue corresponding to the affine Hamiltonian $H_{aff}(q,p)$, T is the temperature–like environmental control parameter, and the sum runs over all energy eigenstates (labelled by the index j). From (2.779), we can further calculate all statistical and thermodynamic system properties (see [Fey72]), as for example, *transition entropy* $S = k_B \ln Z$, etc.

2.5.5 Intelligent Robot Control

In this subsection we show a dynamic model of an intelligent robot interacting with its environment. Recall that a general robot dynamics is described by a nonlinear vector/matrix differential equation (see [Vuk75, VP82, VS82, VK85a, VK85b, VSK85, VP85, VBS89])

$$H(q)\,\ddot{q} + h(q,\dot{q}) + J^T(q)\,F = \tau,$$

where, $q = q(t)$ is an nD vector of robot generalized coordinates; $H(q)$ is an $n \times n$ positive definite matrix of inertia moments of the manipulation mechanics; $h(q,\dot{q})$ is an nD nonlinear function of centrifugal, Coriolis, and gravitational moments; $\tau = \tau(t)$ is an nD vector of input control; $J^T(q)$ is an $n \times n$ Jacobian matrix connecting the velocities of robot end–effector and the velocities of robot generalized coordinates; and $F = F(t)$ is an mD vector of generalized forces, or, of generalized forces and moments from the environment acting on the end–effector.

In the frame of robot joint coordinates, the model of environment dynamics can be presented in the form [KV98, KV03a, KV03b]

$$M(q)\,\ddot{q} + L(q,\dot{q}) = S^T(q)\,F,$$

where $M(q) \in \mathbb{R}^{n\times n}$ is a nonsingular matrix; $L(q,\dot{q}) \in \mathbb{R}^n$ is a nonlinear vector function; and $S^T(q) \in \mathbb{R}^{n\times n}$ is the matrix with $rank(S) = n$.

The end–effector of the manipulator is constrained on static geometric surfaces, $\Phi(q) = 0$, where $\Phi(q) \in \mathbb{R}^m$ is the holonomic constraint function.

In practice, it is convenient to adopt a simplified model of the environment, taking into account the dominant effects, such as stiffness, $F = K'(x - x_0)$, or an environment damping during the tool motion, $F = B'x$, where $K' \in \mathbb{R}^{n\times n}$, $B' \in \mathbb{R}^{n\times n}$ are semidefinite matrices describing the environment stiffness and damping, respectively, and $x_0 \in \mathbb{R}^n$ denotes the coordinate vector in Cartesian coordinates of the point of contact between the end–effector (tool) and a constraint surface. However, it is more appropriate to adopt the relationship defined by specification of the target impedance

$$F = M'\Delta\ddot{x} + B'\Delta\dot{x} + K'\Delta x, \qquad \text{where} \qquad \Delta x = x - x_0,$$

and M' is a positive definite inertia matrix. The matrices M', B', K' define the target impedance which can be selected to correspond to various objectives of the given manipulation task.

In the case of contact with the environment, the robot control task can be described as robot motion along a programmed trajectory $q_p(t)$ representing a twice continuously differentiable function, when a desired force of interaction $F_p(t)$ acts between the robot and the environment. Thus, the programmed motion $q_p(t)$ and the desired interaction force $F_p(t)$ must satisfy the following relation

$$F_p(t) \equiv f\left(q_p(t), \dot{q}_p(t), \ddot{q}_p(t)\right).$$

The control problem for robot interacting with dynamic environment is to define the control $\tau(t)$ for $t \geq t_0$, that satisfies the target conditions

$$\lim_{t\to\infty} q(t) \to q_p(t), \qquad \lim_{t\to\infty} F(t) \to F_p(t).$$

As a first example, the control algorithm based on stabilization of the robot motion with a preset quality of transient responses is considered, which has the form

$$\tau = H(q)[\ddot{q}_p - KP\eta - KD\dot{\eta}] + h(q, \dot{q}) + J^T(q)F.$$

The familly of desired transient responses is specified by the vector differential equation

$$\ddot{\eta} = -KP\eta - KD\dot{\eta}, \qquad \eta(t) = q(t) - q_p(t), \tag{2.780}$$

where $KP \in \mathbb{R}^{n\times n}$ is the diagonal matrix of position feedback gains, and $KD \in R^{n\times n}$ is the diagonal matrix of velocity feedback gains. The right side of (2.780), i.e., PD–regulator is chosen such that the system defined by (2.780) is asymptotically stable in the whole. The values of matrices KP and KD can be chosen according to algebraic stability conditions.

The proposed control law represents a version of the well–known computed torque method including force term which uses dynamic robot model and the available on–line information from the position, velocity and force sensors. Here the model of robot environment does not have any influence on the performance of the control algorithm.

As the second example, control algorithm based on stabilization of the interaction force with a preset quality of transient responses is considered, which has the form

$$\begin{aligned}\tau = {} & H(q)\, M^{-1}(q)\, [-L(q, \dot{q}) + S^T(q)\, F] + h(q, \dot{q}) \\ & + J^T(q) \left\{ F_p - \int_{t_0}^{t} \left[KFP\, \mu(\omega) + KFI \int_{t_0}^{t} \mu(\omega)\, dt \right] d\omega \right\},\end{aligned}$$

where $\mu(t) = F(t) - F_p(t)$; $KFP \in \mathbb{R}^{n\times n}$ is the matrix of proportional force feedback gains; and $KFI \in \mathbb{R}^{n\times n}$ is the matrix of integral force feedback gains. Here, it has been assumed that the interaction force in transient process should behave according to the following differential equation

$$\dot{\mu}(t) = Q(\mu), \qquad Q(\mu) = -KFP\,\mu - KFI \int_{t_0}^{t} \mu\, dt. \tag{2.781}$$

PI force regulator (continuous vector function of Q) is chosen such that the system defined by (2.781) is asymptotically stable in the whole. In this case, environment dynamics model has explicit influence on the performance of contact control algorithm, also having influence on PI force local gains. Note that without knowing a sufficiently accurate environment model (parameters of matrices $M(q), L(q,\dot{q}), S(q)$) it is not possible to determine the nominal contact force $F_p(t)$.

2.5.6 Brain–Like Control Functor in Biomechanics

In this final section we propose our most recent model [IB05] of the complete biomechanical *brain–like control functor*. This is a neurodynamical reflection on our *covariant force law*, $F_i = mg_{ij}a^j$, and its associated *covariant force functor* $\mathcal{F}_* : TT^*M \to TTM$ (see section 2.4.3 above).

Recall that traditional hierarchical robot control (see, e.g., [VS82, VBS89]) consists of three levels: the *executive* control–level (at the bottom) performs tracking of nominal trajectories in internal–joint coordinates, the *strategic* control–level (at the top) performs 'planning' of trajectories of an end–effector in external–Cartesian coordinates, and the *tactical* control–level (in the middle) connects other two levels by means of inverse kinematics.

The modern version of the hierarchical robot control includes decision–making done by the neural (or, neuro–fuzzy) classifier to adapt the (manipulator) control to dynamically changing environment.

On the other hand, the so–called 'intelligent' approach to robot control typically represents a form of function approximation, which is itself based on some combination of neuro–fuzzy–genetic computations. Many special issues and workshops focusing on physiological models for robot control reflect the increased attention for the development of *cerebellar models* [Sma99, SA98, Sch99, Sch98, Arb98] for learning robot control with functional decomposition, where the main result could be formulated as: *the cerebellum is more then just the function approximator.*

In this section we try to fit between these three approaches for humanoid control, emphasizing the role of muscle–like robot actuators. We propose a new, physiologically based, tensor–invariant, hierarchical force control (FC, for short) for the physiologically realistic biomechanics. We consider the muscular torque one–forms F_i as the most important component of human–like motion; therefore we propose the sophisticated hierarchical system for the subtle F_i–control: corresponding to the spinal, the cerebellar and cortical levels of human motor control. F_i are first set–up as testing input–signals to biomechanics, and then covariantly updated as feedback $1-$forms u_i on each FC–level. On the spinal FC–level the nominal joint–trajectory tracking is proposed in the form of affine Hamiltonian control; here the driving torques are given

corrections by spinal–reflex controls. On the cerebellar FC–level, the relation is established between canonical joint coordinates q^i, p_i and gradient *neural–image coordinates* x^i, y_i, representing bidirectional, self–organized, associative memory machine; here the driving torques are given the cerebellar corrections. On the cortical FC–level the topological 'hyper–joystick' is proposed as the central FC command–space, selector, with the fuzzy–logic feedback–control map defined on it, giving the cortical corrections to the driving torques.

The model of the spinal FC–level formulated here resembles *autogenetic motor servo*, acting on the spinal–reflex level of the human locomotor control. The model of the cerebellar FC–level formulated here mimics the self–organizing, associative function of the excitatory granule cells and the inhibitory Purkinje cells of the cerebellum [HBB96]. The model of the cortical FC–level presented in this section mimics the synergistic *regulation of locomotor conditioned reflexes* by the cerebellum [HBB96].

We believe that (already mentioned) extremely high order of the driving force redundancy in biomechanics justifies the formulation of the three–level force control system. Also, both brain–like control systems can be easily extended to give $SE(3)-$based force control for moving inverse kinematics (IK) chains of legs and arms.

Functor Control Machine

In this subsection we define the functor control–machine (compare with section (2.4.3) above), for the learning control with functional decomposition, by a two–step generalization of the Kalman's theory of linear MIMO–feedback systems. The first generalization puts the Kalman's theory into the pair of mutually dual linear categories `Vect` and `Vect`* of vector spaces and linear operators, with a 'loop–functor' representing the closed–loop control, thus formulating the unique, categorical formalism valid both for the discrete and continual MIMO–systems.

We start with the unique, *feedforward continual–sequential state equation*

$$\dot{x}(t+1) = Ax(t) + Bu(t), \qquad y(t) = Cx(t), \tag{2.782}$$

where the nD vector spaces of *state* $X \ni x$, *input* $U \ni u$, and *output* $Y \ni y$ have the corresponding linear operators, respectively $A : X \to X$, $B : U \to X$, and $C : X \to Y$. The modular system theory comprises the *system dynamics*, given by a pair (X, A), together with a *reachability map* $e : U \to X$ of the pair (B, A), and an *observability map* $m : X \to Y$ of the pair (A, C). If the reachability map e is surjection the system dynamics (X, A) is called *reachable*; if the observability map m is injection the system dynamics (X, A) is called *observable*. If the system dynamics (X, A) is both reachable and observable, a *composition* $r = m \circ e : U \to Y$ defines the *total system's response*, which is given by solution of equation (2.782). If the unique solution to the continual–sequential state equation exists, it gives the answer to the (minimal) *realization problem*: find the system S that realizes the given response

$r = m \circ e : U \to Y$ (in the smallest number of discrete states and in the shortest time).

The inverse map $r^{-1} = e^{-1} \circ m^{-1} : Y \to U$ of the total system's response $r : U \to Y$ defines the linear *feedback operator* $K : Y \to U$, given by standard feedback equation

$$u(t) = Ky(t). \tag{2.783}$$

In categorical language, the feedforward system dynamics in the category $\mathtt{Vect}$ is a pair (X, A), where $X \in \mathtt{Ob}(\mathtt{Vect})$ is an object in $\mathtt{Vect}$ and $A : X \to X \in \mathtt{Mor}(\mathtt{Vect})$ is a $\mathtt{Vect}$–morphism. A *feedforward decomposable system* in $\mathtt{Vect}$ is such a sixtuple $S \equiv (X, A, U, B, Y, C)$ that (X, A) is the system dynamics in $\mathtt{Vect}$, a $\mathtt{Vect}$–morphism $B : U \to X$ is an *input map*, and a $\mathtt{Vect}$–morphism $C : X \to Y$ is an *output map*. Any object in $\mathtt{Vect}$ is characterized by mutually dual notions of its *degree* (a number of its input morphisms) and its *codegree* (a number of its output morphisms). Similarly, any decomposable system S in $\mathtt{Vect}$ has a *reachability map* given by an epimorphism $e = A \circ B : U \to X$ and its dual *observability map* given by a monomorphism $m = C \circ A : X \to Y$; their composition $r = m \circ e : U \to Y$ in $\mathtt{Mor}(\mathtt{Vect})$ defines the total system's response in $\mathtt{Vect}$ given by the unique solution of the continual–sequential state equation (2.782) [IS01].

The dual of the total system's response, defined by the feedback equation (2.783), is the *feedback morphism* $K = e^{-1} \circ m^{-1} : Y \to U$ belonging to the dual category $\mathtt{Vect}^*$.

In this way, the linear, closed–loop, continual–sequential MIMO–system (2.782–2.783) represents the *linear iterative loop functor* $\mathcal{L} : \mathtt{Vect} \Rightarrow \mathtt{Vect}^*$.

Our second generalization represents a *natural system process* $\Xi[\mathcal{L}]$, that transforms the linear loop functor $\mathcal{L} : \mathtt{Vect} \Rightarrow \mathtt{Vect}^*$ – into the *nonlinear loop functor* $\mathcal{NL} : \mathcal{CAT} \Rightarrow \mathcal{CAT}^*$ between two mutually dual nonlinear categories $\mathcal{CAT}$ and $\mathcal{CAT}^*$. We apply the natural process Ξ, separately

1. To the feedforward decomposable system
   $S \equiv (X, A, U, B, Y, C)$ in $\mathtt{Vect}$, and
2. To the feedback morphism $K = e^{-1} \circ m^{-1} : Y \to U$ in $\mathtt{Vect}^*$.

Under the action of the natural process Ξ, the linear feedforward system dynamics (X, A) in $\mathtt{Vect}$ transforms into a nonlinear feedforward Ξ–dynamics $(\Xi[X], \Xi[A])$ in $\mathcal{CAT}$, represented by a *nonlinear feedforward decomposable system*, $\Xi[S] \equiv (\Xi[X], \Xi[A], \Xi[U], \Xi[B], \Xi[Y], \Xi[C])$.

The reachability map transforms into the *input process* $\Xi[e] = \Xi[A] \circ \Xi[B] : \Xi[U] \longrightarrow \Xi[X]$, while its dual, observability map transforms into the *output process* $\Xi[m] = \Xi[C] \circ \Xi[A] : \Xi[X] \longrightarrow \Xi[Y]$. In this way the total response of the linear system $r = m \circ e : U \to Y$ in $\mathtt{Mor}(\mathtt{Vect})$ transforms into the *nonlinear system behavior*, $\Xi[r] = \Xi[m] \circ \Xi[e] : \Xi[U] \longrightarrow \Xi[Y]$ in $\mathtt{Mor}(\mathcal{CAT})$. Obviously, $\Xi[r]$, if exists, is given by a nonlinear Ξ–transform of the linear state equations (2.782–2.783).

Analogously, the linear feedback morphism $K = e^{-1} \circ m^{-1} : Y \to U$ in $\mathtt{Mor}(\mathtt{Vect}^*)$ transforms into the nonlinear feedback morphism $\Xi[K] = \Xi[e^{-1}] \circ \Xi[m^{-1}] : \Xi[Y] \to \Xi[U]$ in $\mathtt{Mor}(\mathcal{CAT}^*)$.

In this way, the natural system process $\Xi : \mathcal{L} \Rrightarrow \mathcal{NL}$ is established. That means that the nonlinear loop functor $L = \Xi[\mathcal{L}] : \mathcal{CAT} \Rightarrow \mathcal{CAT}^*$ is defined out of the linear, closed–loop, continual–sequential MIMO–system (2.782).

In this section we formulate the nonlinear loop functor $L = \Xi[\mathcal{L}] : \mathcal{CAT} \Rightarrow \mathcal{CAT}^*$ for various hierarchical levels of muscular–like FC.

Spinal Control Level

Our first task is to establish the nonlinear loop functor $L = \Xi[\mathcal{L}] : \mathcal{EX} \Rightarrow \mathcal{EX}^*$ on the category $\mathcal{EX}$ of spinal FC–level.

Recall that our dissipative, driven $\delta-$Hamiltonian biomechanical system on the configuration manifold M is, in local canonical–symplectic coordinates $q^i, p_i \in U_p$ on the momentum phase–space manifold T^*M, given by autonomous equations

$$\dot{q}^i = \frac{\partial H_0}{\partial p_i} + \frac{\partial R}{\partial p_i}, \qquad (i = 1, \dots, N) \tag{2.784}$$

$$\dot{p}_i = F_i - \frac{\partial H_0}{\partial q^i} + \frac{\partial R}{\partial q^i}, \tag{2.785}$$

$$q^i(0) = q^i_0, \qquad p_i(0) = p^0_i, \tag{2.786}$$

including contravariant equation (2.784) – the velocity vector–field, and covariant equation (2.785) – the force 1−form, together with initial joint angles q^i_0 and momenta p^0_i. Here the physical Hamiltonian function $H_0 : T^*M \to \mathbb{R}$ represents the total biomechanical energy function, in local canonical coordinates $q^i, p_i \in U_p$ on T^*M given by

$$H_0(q, p) = \frac{1}{2} g^{ij}\, p_i\, p_j + V(q),$$

where $g^{ij} = g^{ij}(q, m)$ denotes the contravariant material metric tensor.

Now, the *control Hamiltonian function* $H_\gamma : T^*M \to \mathbb{R}$ of FC is in local canonical coordinates on T^*M defined by [NS90]

$$H_\gamma(q, p, u) = H_0(q, p) - q^i\, u_i, \qquad (i = 1, \dots, N) \tag{2.787}$$

where $u_i = u_i(t, q, p)$ are feedback–control 1−forms, representing the spinal FC–level $u-$corrections to the covariant torques $F_i = F_i(t, q, p)$.

Using $\delta-$Hamiltonian biomechanical system (2.784–2.786) and the control Hamiltonian function (2.787), control $\gamma_\delta-$Hamiltonian FC–system can be defined as

$$\dot{q}^i = \frac{\partial H_\gamma(q,p,u)}{\partial p_i} + \frac{\partial R(q,p)}{\partial p_i},$$
$$\dot{p}_i = F_i - \frac{\partial H_\gamma(q,p,u)}{\partial q^i} + \frac{\partial R(q,p)}{\partial q^i},$$
$$o^i = -\frac{\partial H_\gamma(q,p,u)}{\partial u_i}, \qquad (i = 1,\dots,N)$$
$$q^i(0) = q_0^i, \qquad p_i(0) = p_i^0,$$

where $o^i = o^i(t)$ represent FC natural outputs which can be different from commonly used joint angles.

If nominal reference outputs $o_R^i = o_R^i(t)$ are known, the simple PD *stiffness–servo* [Whi87] could be formulated, via *error function* $e(t) = o^j - o_R^j$, in covariant form

$$u_i = K_o \delta_{ij}(o^j - o_R^j) + K_{\dot{o}} \delta_{ij}(\dot{o}^j - \dot{o}_R^j), \tag{2.788}$$

where Ks are the control–gains and δ_{ij} is the Kronecker tensor.

If natural outputs o^i *actually are* the joint angles and nominal canonical trajectories $\left(q_R^i = q_R^i(t),\, p_i^R = p_i^R(t)\right)$ are known, then the stiffness–servo (2.788) could be formulated in canonical form as

$$u_i = K_q \delta_{ij}(q^i - q_R^i) + K_p(p_i - p_i^R).$$

Now, using the fuzzified $\mu-$Hamiltonian biomechanical system with fuzzy system numbers (i.e, imprecise segment lengths, masses and moments of inertia, joint dampings and muscular actuator parameters)

$$\dot{q}^i = \frac{\partial H_0(q,p,\sigma_\mu)}{\partial p_i} + \frac{\partial R}{\partial p_i}, \tag{2.789}$$
$$\dot{p}_i = \bar{F}_i - \frac{\partial H_0(q,p,\sigma_\mu)}{\partial q^i} + \frac{\partial R}{\partial q^i}, \tag{2.790}$$
$$q^i(0) = \bar{q}_0^i, \qquad p_i(0) = \bar{p}_i^0, \qquad (i = 1,\dots,N), \tag{2.791}$$

(see 2.4.3 above) and the control Hamiltonian function (2.787), $\gamma_\mu-$Hamiltonian FC–system can be defined as

$$\dot{q}^i = \frac{\partial H_\gamma(q,p,u,\sigma_\mu)}{\partial p_i} + \frac{\partial R(q,p)}{\partial p_i},$$
$$\dot{p}_i = \bar{F}_i - \frac{\partial H_\gamma(q,p,u,\sigma_\mu)}{\partial q^i} + \frac{\partial R(q,p)}{\partial q^i},$$
$$\bar{o}^i = -\frac{\partial H_\gamma(q,p,u,\sigma_\mu)}{\partial u_i}, \qquad q^i(0) = \bar{q}_0^i, \qquad p_i(0) = \bar{p}_i^0,$$

where $\bar{o}^i = \bar{o}^i(t)$ represent the fuzzified natural outputs.

Finally, applying stochastic forces (diffusion fluctuations $B_{ij}[q^i(t),t]$ and discontinuous jumps in the form of ND Wiener process $W^j(t)$), i.e., using the fuzzy–stochastic $[\mu\sigma]-$Hamiltonian biomechanical system

$$dq^i = \left(\frac{\partial H_0(q,p,\sigma_\mu)}{\partial p_i} + \frac{\partial R}{\partial p_i}\right) dt, \tag{2.792}$$

$$dp_i = B_{ij}[q^i(t),t]\, dW^j(t) + \left(\bar{F}_i - \frac{\partial H_0(q,p,\sigma_\mu)}{\partial q^i} + \frac{\partial R}{\partial q^i}\right) dt, \tag{2.793}$$

$$q^i(0) = \bar{q}_0^i, \qquad p_i(0) = \bar{p}_i^0. \tag{2.794}$$

(see 2.4.3 above), and the control Hamiltonian function (2.787), $\gamma_{\mu\sigma}$–Hamiltonian FC–system can be defined as

$$\begin{aligned}
dq^i &= \left(\frac{\partial H_\gamma(q,p,u,\sigma_\mu)}{\partial p_i} + \frac{\partial R(q,p)}{\partial p_i}\right) dt,\\
dp_i &= B_{ij}[q^i(t),t]\, dW^j(t) + \left(\bar{F}_i - \frac{\partial H_\gamma(q,p,u,\sigma_\mu)}{\partial q^i} + \frac{\partial R(q,p)}{\partial q^i}\right) dt,\\
d\bar{o}^i &= -\frac{\partial H_\gamma(q,p,u,\sigma_\mu)}{\partial u_i} dt, \qquad (i = 1,\dots,N)\\
q^i(0) &= \bar{q}_0^i, \qquad p_i(0) = \bar{p}_i^0.
\end{aligned}$$

If we have the case that not all of the configuration joints on the configuration manifold M are active in the specified robot task, we can introduce the coupling Hamiltonians $H^j = H^j(q,p)$, $j = 1,\dots,M \le N$, corresponding to the system's active joints, and we come to *affine Hamiltonian function* $H_a : T^*M \to \mathbb{R}$, in local canonical coordinates on T^*M given as [NS90]

$$H_a(q,p,u) = H_0(q,p) - H^j(q,p)\, u_j. \tag{2.795}$$

Using $\delta-$Hamiltonian biomechanical system (2.784–2.786) and the affine Hamiltonian function (2.795), affine $a_\delta-$Hamiltonian FC–system can be defined as

$$\dot{q}^i = \frac{\partial H_0(q,p)}{\partial p_i} - \frac{\partial H^j(q,p)}{\partial p_i} u_j + \frac{\partial R}{\partial p_i}, \tag{2.796}$$

$$\dot{p}_i = F_i - \frac{\partial H_0(q,p)}{\partial q^i} + \frac{\partial H^j(q,p)}{\partial q^i} u_j + \frac{\partial R}{\partial q^i}, \tag{2.797}$$

$$o^i = -\frac{\partial H_a(q,p,u)}{\partial u_i} = H^j(q,p), \tag{2.798}$$

$$q^i(0) = q_0^i, \qquad p_i(0) = p_i^0, \tag{2.799}$$

$$(i = 1,\dots,N; \qquad j = 1,\dots,M \le N).$$

Using the Lie–derivative exact feedback linearization (see (2.3.1) above), and applying the *constant relative degree* r (see [Isi89, SI89]) to all N joints of the affine $a_\delta-$Hamiltonian FC–system (2.796–2.799), the control law for asymptotic tracking the reference outputs o_R^j could be formulated as

$$u_j = \frac{\dot{o}_R^{(r)j} - \mathcal{L}_f^{(r)} H^j + \sum_{s=1}^{r} \gamma_{s-1}(o_R^{(s-1)j} - \mathcal{L}_f^{(s-1)} H^j)}{\mathcal{L}_g \mathcal{L}_f^{(r-1)} H^j},$$

where standard MIMO–vector–fields f and g are given by

$$f = \left(\frac{\partial H_0}{\partial p_i}, -\frac{\partial H_0}{\partial q^i} \right), \qquad g = \left(-\frac{\partial H^j}{\partial p_i}, \frac{\partial H^j}{\partial q^i} \right)$$

and γ_{s-1} are the coefficients of linear differential equation of order r for the error function $e(t) = o^j - o_R^j$

$$e^{(r)} + \gamma_{r-1} e^{(r-1)} + \cdots + \gamma_1 e^{(1)} + \gamma_0 e = 0.$$

Using the fuzzified $\mu-$Hamiltonian biomechanical system (2.789–2.791) and the affine Hamiltonian function (2.795), affine $a_\mu-$Hamiltonian FC–system can be defined as

$$\begin{aligned}
\dot{q}^i &= \frac{\partial H_0(q,p,\sigma_\mu)}{\partial p_i} - \frac{\partial H^j(q,p,\sigma_\mu)}{\partial p_i} u_j + \frac{\partial R(q,p)}{\partial p_i}, \\
\dot{p}_i &= \bar{F}_i - \frac{\partial H_0(q,p,\sigma_\mu)}{\partial q^i} + \frac{\partial H^j(q,p,\sigma_\mu)}{\partial q^i} u_j + \frac{\partial R(q,p)}{\partial q^i}, \\
\bar{o}^i &= -\frac{\partial H_a(q,p,u,\sigma_\mu)}{\partial u_i} = H^j(q,p,\sigma_\mu), \\
q^i(0) &= \bar{q}_0^i, \qquad p_i(0) = \bar{p}_i^0, \qquad (i = 1,\ldots,N;\ j = 1,\ldots,M \leq N).
\end{aligned}$$

Using the fuzzy–stochastic $[\mu\sigma]-$Hamiltonian biomechanical system (2.792–2.794) and the affine Hamiltonian function (2.795), affine $a_{\mu\sigma}-$Hamiltonian FC–system can be defined as

$$\begin{aligned}
dq^i &= \left(\frac{\partial H_0(q,p,\sigma_\mu)}{\partial p_i} - \frac{\partial H^j(q,p,\sigma_\mu)}{\partial p_i} u_j + \frac{\partial R(q,p)}{\partial p_i} \right) dt, \\
dp_i &= B_{ij}[q^i(t),t]\, dW^j(t) \qquad + \\
&\quad \left(\bar{F}_i - \frac{\partial H_0(q,p,\sigma_\mu)}{\partial q^i} + \frac{\partial H^j(q,p,\sigma_\mu)}{\partial q^i} u_j + \frac{\partial R(q,p)}{\partial q^i} \right) dt, \\
d\bar{o}^i &= -\frac{\partial H_a(q,p,u,\sigma_\mu)}{\partial u_i} dt = H^j(q,p,\sigma_\mu)\, dt, \\
q^i(0) &= \bar{q}_0^i, \qquad p_i(0) = \bar{p}_i^0, \qquad (i = 1,\ldots,N;\ j = 1,\ldots,M \leq N).
\end{aligned}$$

Being high–degree and highly nonlinear, all of these affine control systems are extremely sensitive upon the variation of parameters, inputs, and initial conditions. The sensitivity function S of the affine Hamiltonian $H_a(q,p,u)$ upon the parameters β_i (representing segment lengths L_i, masses m_i, moments of inertia J_i and joint dampings b_i, see [IS01, Iva91]), is in the case of $a_\delta-$Hamiltonian FC–system defined as

$$S(H,\beta) = \frac{\beta_i}{H_a(q,p,u)} \frac{\partial H_a(q,p,u)}{\partial \beta_i},$$

and similarly in other two $a_\mu-$ and $a_{\mu\sigma}-$ cases.

The three affine FC–level systems a_δ, a_μ and $a_{\mu\sigma}$, resemble (in a fuzzy–stochastic–Hamiltonian form), Houk's autogenetic motor servo of muscle spindle and Golgi tendon proprioceptors [Hou79], correcting the covariant driving torques $F_i = F_i(t,q,p)$ by local 'reflex controls' $u_i(t,q,p)$. They form the nonlinear loop functor $L = \Xi[\mathcal{L}] : \mathcal{EX} \Rightarrow \mathcal{EX}^*$.

Cerebellar Control Level

Our second task is to establish the nonlinear loop functor $L = \Xi[\mathcal{L}] : \mathcal{TA} \Rightarrow \mathcal{TA}^*$ on the category $\mathcal{TA}$ of the cerebellar FC–level. Here we propose an oscillatory neurodynamical $(x,y,\boldsymbol{\omega})$–system (adapted from [IJB99a]), a bidirectional, self–organized, associative–memory machine, resembling the function of a set of excitatory granule cells and inhibitory Purkinje cells in the middle layer of the cerebellum (see [EIS67, HBB96]). The neurodynamical $(x,y,\boldsymbol{\omega})$–system acts on *neural–image manifold* M^N_{im} of the configuration manifold M^N as a pair of smooth, '$1-1$' and 'onto' maps (Ψ,Ψ^{-1}), where $\Psi : M^N \to M^N_{im}$ represents the feedforward map, and $\Psi^{-1} : M^N_{im} \to M^N$ represents the feedback map. Locally, it is defined in Riemannian neural coordinates $x^i, y_i \in V_y$ on M^N_{im}, which are in bijective correspondence with symplectic joint coordinates $q^i, p_i \in U_p$ on T^*M.

The $(x,y,\boldsymbol{\omega})$–system is formed out of two distinct, yet nonlinearly–coupled neural subsystems, with $A^i(q)$ (2.802) and $B_i(p)$ (2.803) as system inputs, and the feedback–control $1-$forms u_i (2.808) as system outputs:

1. Granule cells excitatory (contravariant) and Purkinje cells inhibitory (covariant) activation (x,y)–dynamics (2.800–2.803), defined respectively by a vector–field $x^i = x^i(t) : M \to TM$, representing a cross–section of the tangent bundle TM, and a $1-$form $y_i = y_i(t) : M \to T^*M$, representing a cross–section of the cotangent bundle T^*M; and
2. Excitatory and inhibitory unsupervised learning $(\boldsymbol{\omega})$–dynamics (2.803–2.805) generated by random differential Hebbian learning process (2.806–2.808), defined respectively by contravariant synaptic tensor–field $\omega^{ij} = \omega^{ij}(t) : M \to TTM^N_{im}$ and covariant synaptic tensor–field $\omega_{ij} = \omega_{ij}(t) : M \to T^*T^*M$, representing cross–sections of contravariant and covariant tensor bundles, respectively.

The system equations are defined as

$$\dot{x}^i = A^i(q) + \omega^{ij}\, f_j(y) - x^i, \tag{2.800}$$

$$\dot{y}_i = B_i(p) + \omega_{ij}\, f^j(x) - y_i, \tag{2.801}$$

$$A^i(q) = K_q(q^i - q^i_R), \tag{2.802}$$

$$B_i(p) = K_p(p^R_i - p_i), \tag{2.803}$$

$$\dot{\omega}^{ij} = -\omega^{ij} + I^{ij}(x,y), \tag{2.804}$$

$$\dot{\omega}_{ij} = -\omega_{ij} + I_{ij}(x,y), \tag{2.805}$$

$$I^{ij} = f^i(x)\, f^j(y) + \dot{f}^i(x)\, \dot{f}^j(y) + \sigma^{ij}, \tag{2.806}$$

$$I_{ij} = f_i(x)\, f_j(y) + \dot{f}_i(x)\, \dot{f}_j(y) + \sigma_{ij}, \tag{2.807}$$

$$u_i = \frac{1}{2}(\delta_{ij}\, x^i + y_i), \qquad (i,j = 1,\dots,N). \tag{2.808}$$

Here $\boldsymbol{\omega}$ is a symmetric 2nd order synaptic tensor–field; $I^{ij} = I^{ij}(x,y,\sigma)$ and $I_{ij} = I_{ij}(x,y,\sigma)$ respectively denote contravariant–excitatory and covariant–inhibitory random differential Hebbian innovation–functions with tensorial Gaussian noise $\boldsymbol{\sigma}$ (in both variances); fs and $\dot{f}$s denote sigmoid activation functions ($f = \tanh(.)$) and corresponding signal velocities ($\dot{f} = 1 - f^2$), respectively in both variances;

$A^i(q)$ and $B_i(p)$ are contravariant–excitatory and covariant–inhibitory neural inputs to granule and Purkinje cells, respectively; u_i are the corrections to the feedback–control 1−forms on the cerebellar FC–level.

Nonlinear activation (x,y)–dynamics (2.800–2.803), describes a two–phase biological neural oscillator field, in which excitatory neural field excites inhibitory neural field, which itself reciprocally inhibits the excitatory one. (x,y)–dynamics represents a nonlinear extension of a linear, Lyapunov–stable, conservative, gradient system, defined in local neural coordinates $x^i, y_i \in V_y$ on T^*M as

$$\dot{x}^i = -\frac{\partial \Phi}{\partial y_i} = \omega^{ij} y_j - x^i, \qquad \dot{y}_i = -\frac{\partial \Phi}{\partial x^i} = \omega_{ij} x^j - y_i. \tag{2.809}$$

The gradient system (2.809) is derived from scalar, neuro-synaptic action potential $\Phi : T^*M \to \mathbb{R}$, given by a negative, smooth bilinear form in $x^i, y_i \in V_y$ on T^*M as

$$-2\Phi = \omega_{ij} x^i x^j + \omega^{ij} y_i y_j - 2x^i y_i, \qquad (i,j = 1,\dots,N), \tag{2.810}$$

which itself represents a Ψ–image of the Riemannian metrics $g : TM \to \mathbb{R}$ on the configuration manifold M.

The nonlinear oscillatory activation (x,y)–dynamics (2.800–2.803) is get from the linear conservative dynamics (2.809) by adding configura-tion–dependent inputs A^i and B_i, as well as sigmoid activation functions f_j and f^j, respectively. It represents an interconnected pair of excitatory and inhibitory neural fields.

Both variant–forms of learning $(\boldsymbol{\omega})$–dynamics (2.804–2.805) are given by generalized unsupervised (self–organizing) Hebbian learning scheme (see

[Kos92]) in which $\dot{\omega}_{ij}$ (resp. $\dot{\omega}^{ij}$) denotes the new–update value, $-\omega_{ij}$ (resp. $-\omega^{ij}$) corresponds to the old value and $I_{ij}(x^i, y_j)$ (resp. $I^{ij}(x^i, y_j)$) is the innovation function of the symmetric 2nd order synaptic tensor-field $\boldsymbol{\omega}$. The nonlinear innovation functions I_{ij} and I^{ij} are defined by random differential Hebbian learning process (2.806–2.807). As $\boldsymbol{\omega}$ is symmetric and zero-trace coupling synaptic tensor, the conservative linear activation dynamics (2.809) is equivalent to the rule that the state of each neuron (in both neural fields) is changed in time iff the scalar action potential Φ (2.810), is lowered. Therefore, the scalar action potential Φ represents the monotonically decreasing Lyapunov function (such that $\dot{\Phi} \leq 0$) for the conservative linear dynamics (2.809), which converges to a local minimum or ground state of Φ. That is to say, the system (2.809) moves in the direction of decreasing the scalar action potential Φ, and when both $\dot{x}^i = 0$ and $\dot{y}_i = 0$ for all $i = 1, \ldots, N$, the steady state is reached.

In this way, the neurodynamical $(x, y, \boldsymbol{\omega})-$system acts as tensor–invariant self–organizing (excitatory / inhibitory) associative memory machine, resembling the set of granule and Purkinje cells of cerebellum [HBB96].

The feedforward map $\Psi : M \to M$ is realized by the inputs $A^i(q)$ and $B_i(p)$ to the $(x, y, \boldsymbol{\omega})$–system, while the feedback map $\Psi^{-1} : M \to M$ is realized by the system output, i.e., the feedback–control $1-$forms $u_i(x, y)$. These represent the cerebellar FC–level corrections to the covariant torques $F_i = F_i(t, q, p)$.

The tensor–invariant form of the oscillatory neurodynamical $(x, y, \boldsymbol{\omega})$–system (2.800–2.808) resembles the associative action of the granule and Purkinje cells in the tunning of the limb cortico–rubro–cerebellar recurrent network [HBB96], giving the cerebellar correction $u_i(x, y)$ to the covariant driving torques $F_i = F_i(t, q, p)$. In this way $(x, y, \boldsymbol{\omega})$–system forms the nonlinear loop functor $L = \Xi[\mathcal{L}] : \mathcal{TA} \Rightarrow \mathcal{TA}^*$.

Cortical Control Level

Our third task is to establish the nonlinear loop functor $L = \Xi[\mathcal{L}] : \mathcal{ST} \Rightarrow \mathcal{ST}^*$ on the category $\mathcal{ST}$ of the cortical FC–level.

Recall that for the purpose of cortical control, the purely rotational biomechanical manifold M could be firstly reduced to N–*torus* and subsequently transformed to N–*cube* ('hyper–joystick'), using the following geometrical techniques (see (2.4.1) above).

Denote by S^1 the constrained unit circle in the complex plane. This is an Abelian Lie group. We have two reduction homeomorphisms

$$SO(3) \gtrsim SO(2) \triangleright SO(2) \triangleright SO(2), \qquad \text{and} \qquad SO(2) \approx S^1,$$

where '$\triangleright$' denotes the noncommutative semidirect product.

Next, let I^N be the unit cube $[0, 1]^N$ in $\mathbb{R}^N$ and '$\sim$' an equivalence relation on $\mathbb{R}^N$ get by 'gluing' together the opposite sides of I^N, preserving their

orientation. Therefore, M can be represented as the quotient space of $\mathbb{R}^N$ by the space of the integral lattice points in $\mathbb{R}^N$, that is a constrained torus T^N:

$$\mathbb{R}^N/Z^N = I^N/\sim \cong \prod_{i=1}^{N} S_i^1 \equiv \{(q^i,\, i=1,\ldots,N) : \text{mod } 2\pi\} = T^N.$$

In the same way, the momentum phase–space manifold T^*M can be represented by T^*T^N.

Conversely by 'ungluing' the configuration space we get the primary unit cube. Let '$\sim^*$' denote an equivalent decomposition or 'ungluing' relation. By the *Tychonoff product–topology theorem*, for every such quotient space there exists a 'selector' such that their quotient models are homeomorphic, that is, $T^N/\sim^* \approx A^N/\sim^*$. Therefore I_q^N represents a 'selector' for the configuration torus T^N and can be used as an N–directional '$\hat{q}$–command–space' for FC. Any subset of DOF on the configuration torus T^N representing the joints included in the general biomechanics has its simple, rectangular image in the rectified $\hat{q}$–command space – selector I_q^N, and any joint angle q^i has its rectified image $\hat{q}^i$.

In the case of an end–effector, $\hat{q}^i$ reduces to the position vector in external–Cartesian coordinates z^r $(r = 1,\ldots,3)$. If orientation of the end–effector can be neglected, this gives a topological solution to the standard inverse kinematics problem.

Analogously, all momenta $\hat{p}_i$ have their images as rectified momenta $\hat{p}_i$ in the $\hat{p}$–command space – selector I_p^N. Therefore, the total momentum phase–space manifold T^*T^N gets its 'cortical image' as the $\widehat{(q,p)}$–command space, a trivial $2N$D bundle $I_q^N \times I_p^N$.

Now, the simplest way to perform the feedback FC on the cortical $\widehat{(q,p)}$–command space $I_q^N \times I_p^N$, and also to mimic the cortical–like behavior [1,2], is to use the $2N$D fuzzy–logic controller, in pretty much the same way as in popular 'inverted pendulum' examples [Kos92, Kos96].

We propose the fuzzy feedback–control map Ξ that maps all the rectified joint angles and momenta into the feedback–control 1−forms

$$\Xi : (\hat{q}^i(t),\, \hat{p}_i(t)) \mapsto u_i(t,q,p), \tag{2.811}$$

so that their corresponding universes of discourse, $\hat{M}^i = (\hat{q}^i_{max} - \hat{q}^i_{min})$, $\hat{P}_i = (\hat{p}_i^{max} - \hat{p}_i^{min})$ and $U_i = (u_i^{max} - u_i^{min})$, respectively, are mapped as

$$\Xi : \prod_{i=1}^{N} \hat{M}M^i \times \prod_{i=1}^{N} \hat{P}_i \to \prod_{i=1}^{N} U_i. \tag{2.812}$$

The $2N$–D map Ξ (2.811–2.812) represents a *fuzzy inference system*, defined by (adapted from [IJB99b]):

1. *Fuzzification* of the crisp *rectified* and *discretized* angles, momenta and controls using Gaussian–bell membership functions

$$\mu_k(\chi) = \exp[-\frac{(\chi - m_k)^2}{2\sigma_k}], \qquad (k = 1, 2, \ldots, 9),$$

where $\chi \in D$ is the common symbol for $\hat{q}^i$, $\hat{p}_i$ and $u_i(q,p)$ and D is the common symbol for $M^i, \hat{P}_i$ and $_i$; the mean values m_k of the seven partitions of each universe of discourse D are defined as $m_k = \lambda_k D + \chi_{min}$, with partition coefficients λ_k uniformly spanning the range of D, corresponding to the set of nine linguistic variables $L = \{NL, NB, NM, NS, ZE, PS, PM, PB, PL\}$; standard deviations are kept constant $\sigma_k = D/9$. Using the linguistic vector L, the 9×9 FAM (fuzzy associative memory) matrix (a 'linguistic phase–plane'), is heuristically defined for each human joint, in a symmetrical weighted form

$$\mu_{kl} = \varpi_{kl} \exp\{-50[\lambda_k + u(q,p)]^2\}, \qquad (k, l = 1, 2, \ldots, 9)$$

with weights $\varpi_{kl} \in \{0.6, 0.6, 0.7, 0.7, 0.8, 0.8, 0.9, 0.9, 1.0\}$.

2. *Mamdani inference* is used on each FAM–matrix μ_{kl} for all human joints: (i) $\mu(\hat{q}^i)$ and $\mu(\hat{p}_i)$ are combined inside the fuzzy IF–THEN rules using AND (Intersection, or Minimum) operator,

$$\mu_k[\bar{u}_i(q,p)] = \min_l\{\mu_{kl}(\hat{q}^i),\, \mu_{kl}(\hat{p}_i)\}.$$

(ii) the output sets from different IF–THEN rules are then combined using OR (Union, or Maximum) operator, to get the final output, fuzzy–covariant torques,

$$\mu[u_i(q,p)] = \max_k\{\mu_k[\bar{u}_i(q,p)]\}.$$

3. *Defuzzification* of the fuzzy controls $\mu[u_i(q,p)]$ with the 'center of gravity' method

$$u_i(q,p) = \frac{\int \mu[u_i(q,p)]\, du_i}{\int du_i},$$

to update the crisp feedback–control 1−forms $u_i = u_i(t,q,p)$. These represent the cortical FC–level corrections to the covariant torques $F_i = F_i(t,q,p)$.

Operationally, the construction of the cortical $\widehat{(q,p)}$–command space $I_q^N \times I_p^N$ and the $2N$D feedback map Ξ (2.811–2.812), mimic the regulation of locomotor conditioned reflexes by the motor cortex [HBB96], giving the cortical correction to the covariant driving torques F_i. Together they form the nonlinear loop functor $\mathcal{NL} = \Xi[\mathcal{L}] : \mathcal{ST} \Rightarrow \mathcal{ST}^*$.

A sample output from the leading human–motion simulator, *Human Biodynamics Engine* (developed by the authors in Defence Science & Technology Organisation, Australia), is given in Figure 2.28, giving the sophisticated 264 DOF analysis of adult male running with the speed of 5 m/s.

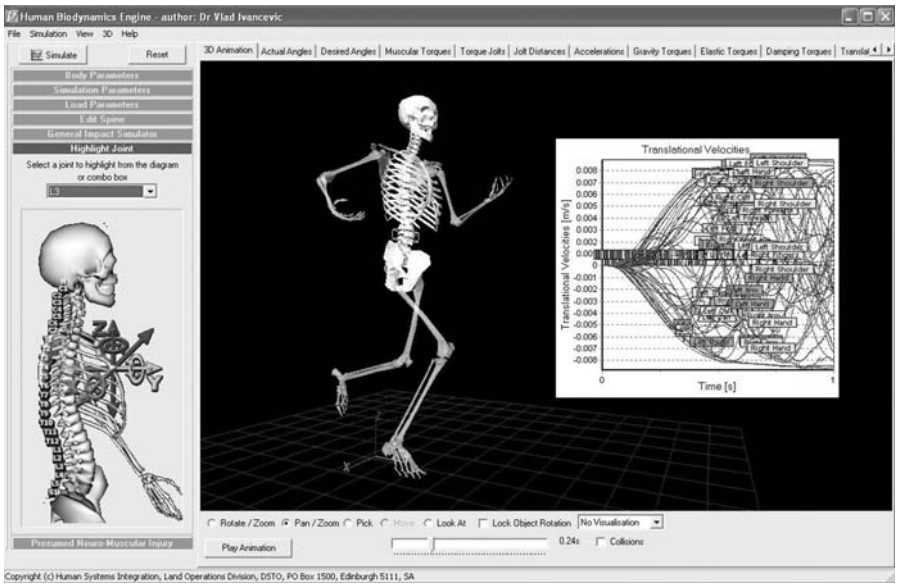


Fig. 2.28. Sample output from the Human Biodynamics Engine: running with the speed of 5 m/s.

2.5.7 Concurrent and Weak Functorial Machines

In this section we first present a concept of machine concurrency, as a modern development of parallel and distributed information processing realized in ANNs, and after that an abstract concept of functorial machine.

Concurrent Machines and Higher–Dimensional Automata

A *sequential machine* (i.e., an ordinary state–machine without concurrency) is a set of states, also called 0−transitions, and a set of 1−transitions from a given state to another one.

On the other hand, a *concurrent machine* consist of a set of states and a set of 1−transitions, but has also the capability of carrying out several 1−transitions at the same time. A concurrent machine is a computer in which several tasks can be performed at the same time. This can be a true parallelism like in the case of several processors running concurrently, or a mono–processor machine in which a unique CPU shares its time between several different tasks.

The *homotopical analysis of concurrency* consists of dealing with concurrent machines using the tools and methods of algebraic topology by working with execution paths up to homotopy [Gou95, FGR98]. The geometric model of concurrent processes allows to formalize very precisely the notion of dihomotopy. Two concurrent processes (i.e., the corresponding geometric

models) are said dihomotopic when they can be geometrically continuously deformed to each other without changing their computer–scientific properties (see [Gau02a, Gau02b, Gau02c, GG03, Gau03]).

Consider a system with finitely many processes running altogether. We assume that each process starts at (local time) 0 and finishes at (local time) 1; the P and V actions correspond to sequences of real numbers between 0 and 1, which reflect the order of the P's and V's. The initial state is $(0, \ldots, 0)$ and the final state is $(1, \ldots, 1)$.

The basic idea is to give a description of what can happen when several processes are modifying shared resources. Given a shared resource a, we see it as its associated semaphore that rules its behavior with respect to processes. For example, if a is an ordinary shared variable, it is customary to use its semaphore to ensure that only one process at a time can write on it (this is mutual exclusion). A semaphore is nothing but a register which counts the number of times a shared object can still be accessed by processes. In the case of usual shared variables, this register is initialized with value 1, processes trying to access (read or write) on the corresponding variable compete in order to get it first, then the semaphore value is decreased: we say that the semaphore has been locked by the process. When it is equal to zero, all processes trying to access this semaphore are blocked, waiting for the process which holds the lock to relinquish it, typically when it has finished reading or writing on the corresponding variable: the value of the semaphore is then increased.

For example, when the semaphores are initialized with value one, meaning that they are associated with shared variables accessed in a mutually exclusive manner, they are called binary semaphores. When a shared data (identified with its semaphore) can be accessed by one or more processes, meaning that the corresponding semaphore has been initialized with a value greater than one, it is called a counting semaphore.

Given n deterministic sequential processes $Q_1, \ldots, Q_n$, abstracted as a sequence of locks and unlocks on (semaphores associated with) shared objects, $Q_i = R^1 a_i^1 . R^2 a_i^2 \cdots R^{n_i} a_i^{n_i}$ (R^k being P or V, there is a natural way to understand the possible behaviors of their concurrent execution, by associating to each process a coordinate line in $\mathbb{R}^n$. The state of the system corresponds to a point in $\mathbb{R}^n$, whose ith coordinate describes the state (or 'local time') of the ith processor.

Consider a system with finitely many processes running altogether. We assume that each process starts at (local time) 0 and finishes at (local time) 1; the P and V actions correspond to sequences of real numbers between 0 and 1, which reflect the order of the P's and V's. The initial state is $(0, \ldots, 0)$ and the final state is $(1, \ldots, 1)$. An example consisting of the two processes $T_1 = Pa.Pb.Vb.Va$ and $T_2 = Pb.Pa.Va.Vb$ induces the 2D *progress graph* of Figure 2.29.

The shaded area represents states which are not allowed in any execution path, since they correspond to mutual exclusion. Such states constitute the *forbidden area*. An *execution path* is a path from the initial state $(0, \ldots, 0)$ to

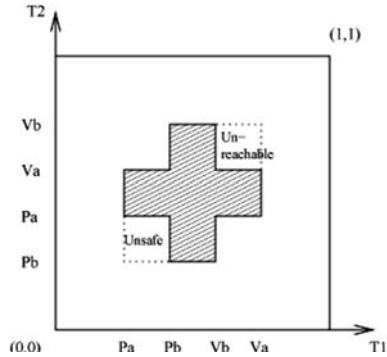


Fig. 2.29. Example of a *progress graph* (adapted from [Gau02a]).

the final state $(1, \ldots, 1)$ avoiding the forbidden area and increasing in each coordinate – time cannot run backwards. We call these paths *directed paths* or *dipaths*. This entails that paths reaching the states in the dashed square underneath the forbidden region, marked 'unsafe' are deemed to deadlock, i.e., they cannot possibly reach the allowed terminal state which is $(1, 1)$ here. Similarly, by reversing the direction of time, the states in the square above the forbidden region, marked 'unreachable', cannot be reached from the initial state, which is $(0, 0)$ here. Also notice that all terminating paths above the forbidden region are 'equivalent' in some sense, given that they are all characterized by the fact that T_2 gets a and b before T_1 (as far as resources are concerned, we call this a *schedule*). Similarly, all paths below the forbidden region are characterized by the fact that T_1 gets a and b before T_2 does.

On this picture, one can already recognize many ingredients that are at the center of the main problem of algebraic topology, namely the classification of shapes modulo 'elastic deformation'. As a matter of fact, the actual coordinates that are chosen for representing the times at which Ps and Vs occur are unimportant, and these can be 'stretched' in any manner, so the properties (deadlocks, schedules etc.) are invariant under some notion of deformation, or *homotopy*. This is only a particular kind of homotopy though, and this explains why a new theory has to be designed. We call it (in subsequent work) *directed homotopy* or *dihomotopy* in the sense that it should preserve the direction of time. For example, the two homotopic shapes, all of which have two holes, of Figure 2.30 and Figure 2.31 have a different number of dihomotopy classes of dipaths. In Figure 2.30 there are essentially four dipaths up to dihomotopy (i.e., four schedules corresponding to all possibilities of accesses of resources a and b) whereas in Figure 2.31, there are essentially three dipaths up to dihomotopy [Gau02a, Gau02b, Gau02c, GG03, Gau03].

The natural combinatorial notion which discretizes this topological framework is that of a *precubical set*, which is a collection of points (states), edges (transitions), squares, cubes and hypercubes (higher–dimensional transitions

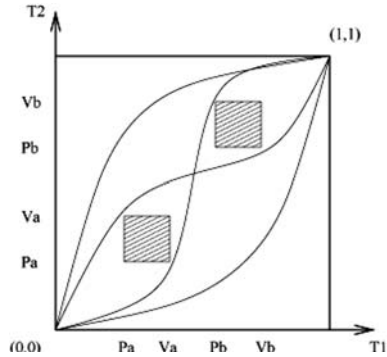


Fig. 2.30. The progress graph corresponding to $Pa.Va.Pb.Vb \mid Pa.Va.Pb.Vb$ (adapted from [Gau02a]).

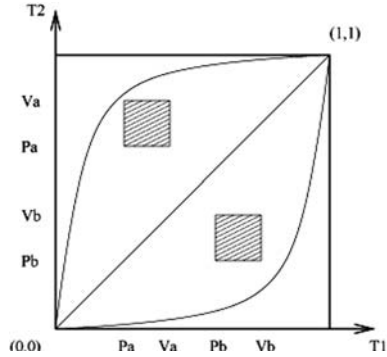


Fig. 2.31. The progress graph corresponding to $Pb.Vb.Pa.Va \mid Pa.Va.Pb.Vb$ (adapted from [Gau02a]).

representing the truly–concurrent execution of some number of actions). This was introduced in [Pra91] as well as possible formalizations using $n-$categories (see the next subsection), and a notion of homotopy [Lei03]. These precubical sets are called *higher–dimensional automata* (HDA) because it really makes sense to consider a hypercube as some form of transition, as in transition systems, used in semantics of programming languages.

In general, an HDA labelled over an alphabet A, is a 7–tuple (Q, d, s, t, I, F, l) where Q is a set (of states), $d : Q \to Nat$, associating to each state q a dimension $d(q)$; states of dimension 0 are called *nodes*, states of dimension 1 are *edges*, states of dimension n represent lists of n *transitions firing concurrently*, $s, t : Nat^*Q \to Q$ are partial functions ($s(k, q)$ indicates the start and $t(q, k)$ the termination of the k-th transition in the list of transitions q) satisfying [Pra91]:

$$s(k, q) \text{ defined iff } t(k, q) \text{ defined iff } k < d(q),$$

$$k < d(q) \geq d(s(k,q)) = d(t(k,q)) = d(q) - 1,$$
$$i \leq j \geq s(j,(s(i,q)) = s(i,s(j+1,q)),$$
$$i \leq j \geq s(j,(t(i,q)) = t(i,s(j+1,q)),$$
$$i \leq j \geq t(j,(s(i,q)) = s(i,t(j+1,q)),$$
$$i \leq j \geq t(j,(t(i,q)) = t(i,t(j+1,q)).$$

These requirements say that $s(k,\ )$ and $t(k,\ )$ applied to a state q representing a list of $d(q) > k$ transitions yield states representing lists of $d(q)-1$ transitions obtained by leaving out transition number k and renumbering the transitions with number $> k$. $s(0,)$ and $t(0,)$ applied on edges just yield the beginning and ending nodes of those edges. Also, I is an element of Q with $d(I) = 0$ (the initial state), F is a subset of Q whose members q satisfy $d(q) = 0$ (the set of final states), $l : Q \to A$ is a partial function (the labelling function) defined on states of dimension 1, satisfying [Pra91]:

$$d(q) = 2 \ \text{ and } (k = 0 \quad \text{or} \quad k = 1) \ \text{ imply } \ l(s(k,q)) = l(t(k,q)).$$

This requirement says that opposite edges in a square have the same label. This because they represent the same transition, scheduled either before or after the firing of another transition. Transitions or events can be defined as equivalence classes of edges (with respect to the finest equivalence identifying $s(k,q)$ and $t(k,q)$ for any q with $d(q) = 2$ and $(k = 0 \text{ or } k = 1)$), and represent occurrences of actions (elements of A) indicated by their label.

Let (Q, d, s, t, I, F, l) be an HDA. The *projection function* $p : Nat^*Q \to Q$ is a partial function defined by $p(k,q) = s(0, s(1, ...s(k-1, s(k+1, ...s(d(q) - 1, q)...))...))$ if $k < d(q)$, and $p(k,q)$ undefined if $k \geq d(q)$. Clearly, $d(p(k,q)) = 1$ and $s(0, d(p(i,q))) = s(0, d(p(j,q)))$ for $i, j, k < d(q)$.

Now the labelling function can be extended to a total function $l : Q \to A^*$ by defining:

$$l(q) = l(p(0,q))l(s(0,q)) \ \text{ if } d(q) > 1,$$
$$\text{and } l(q) = \text{the empty string } \ \text{if } d(q) = 0.$$

A *path* in a higher–dimensional automaton (Q, d, s, t, I, F, l) is a sequence $p(0)p(1)...p(n)$ in $Q^*(n \geq 0)$ such that for any $0 < i < n + 1$ there is a k in Nat such that $p(i-1) = s(k, p(i))$ or $p(i) = t(k, p(i-1))$.

Two paths P and P' are adjacent, denoted $P \leftrightarrow P'$ if there are paths Q, R and states p, p' such that $d(p)$ differs from $d(p')$ and $P = QpR$, while $P' = Qp'R$ [Pra91].

A path P' is an extension of a path P, denoted $P < P'$ if there is a path Q with $P' = PQ$.

A *bisimulation* between two higher–dimensional automata (Q, d, s, t, I, F, l) and $(Q', d', s', t', I', F', l')$ labelled over A is a binary relation R between their paths, such that their initial states (viewed as paths) are related, as well as:

– if $(p(0)p(1)...p(n))\ R\ (q(0)q(1)...q(m))$ then $n = m$, $p(n) \in F$ iff $q(n) \in F'$, and for $0 \leq k \leq n : l(p(k)) = l'(q(k))$ (and thus $d(p(k)) = d'(p(k))$),
– if PRQ and $P \leftrightarrow P'$ then there is a Q' with $Q \leftrightarrow Q'$ and $P'RQ'$,

– if PRQ and $Q \leftrightarrow Q'$ then there is a P' with $P \leftrightarrow P'$ and $P'RQ'$,
– if PRQ and $P < P'$ then there is a Q' with $Q < Q'$ and $P'RQ'$,
– if PRQ and $Q < Q'$ then there is a P' with $P < P'$ and $P'RQ'$.

Two higher–dimensional automata are called *bisimulation equivalent* if there exists a bisimulation between them. Note that bisimulation equivalence is an equivalence indeed. For 1D automata (with $d(q) < 2$ for all $q \in Q$) a path is just an alternating sequence of nodes and edges – each edge going from the node before it to the node after it, no two paths are adjacent and the notion of bisimulation equivalence defined above coincides with the classical one [Pra91].

Let $A = (Q, d, s, t, I, F, l)$ be an HDA. *Homotopy* is the smallest equivalence relation on the paths of A containing adjacency. The unfolding $U(A)$ of A is defined as the automaton $(Q', d', s', t', I', F', l')$, where: Q' is the set of all paths in A starting at I, modulo homotopy, $d'[p(1)...p(n)] = d(p(n))$; $s'([p(1)...p(n)], k) = \{q(1)...q(n-1) | q(1)...q(n-1)p(n)$ is homotopic with $p(1)...p(n)$ and $s(p(n), k) = q(n-1)\}$; $t'([p(1)...p(n)], k) = [p(1)...p(n)t(p(n), k)]$; $I' = [I]$; $F' = \{[p(1)...p(n)] | p(n)$ in $F\}$; and $l'[p(1)...p(n)] = l(p(n))$. It is straightforward to check that $U(A)$ is well–defined and is an automata indeed [Pra91].

Two HDA (Q, d, s, t, I, F, l) and $(Q', d', s', t', I', F', l')$ labelled over A are *isomorphic* if there exists a bijection $f : Q \to Q'$ (an isomorphism) satisfying [Pra91]:

$$d'(f(q)) = d(q),\ f(I) = I',\ f(q) \text{ in } F' \text{ iff } q \in F,\ l'(f(q)) = l(q);$$
$$\{s'(k, f(q)) | k < d(q)\} = \{f(s(k, q)) | k < d(q)\}; \text{ and}$$
$$\{t'(k, f(q)) | k < d(q)\} = \{f(t(k, q)) | k < d(q)\}.$$

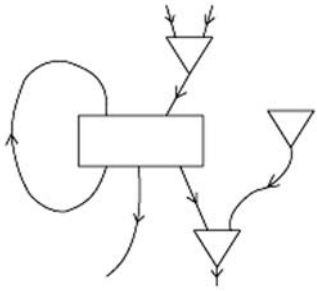
The weak requirements for s and t allow to change the order in the list of transitions represented by a state.

Working with dihomotopy classes of HDA allows to solve the *state–space explosion problem*: the number of spaces to be traversed might be exponential in the number of processes involved (see [Gau02a, Gau02b, Gau02c, GG03, Gau03]).

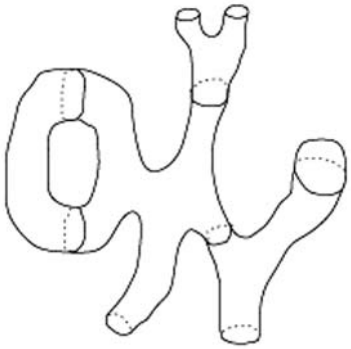
Natural Geometrical Structures

Closely related to the higher–dimensional automata are various *natural geometrical structures*, most of which are commonly called tangles.

For example, a 2D flow–chart–like complex $1D-$structure could be a diagram of the form [Lei02, Lei03]:



Its 3D–generalization is a surface diagram with the same information–flow:



Moreover, if we allow crossings, as in a *braid*:

then we start getting pictures that look like *knots* which are again related to higher categorical structures [Lei02, Bae97].

A category $\mathcal{C}$ with only one object is a *monoid* (= semigroup with unit) M. A 2−category $\mathcal{C}$ with only one 0−cell is a *monoidal category* $\mathcal{M}$. A *braided monoidal category* is a monoidal category equipped with a map called *braiding*

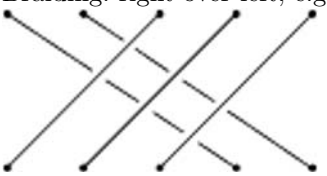
$$A \otimes B \xrightarrow{\beta_{A,B}} B \otimes A\,,$$

for each pair A, B of objects.

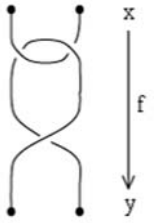
The canonical example of a braided monoidal category is $\mathcal{BR}$ [Lei03]. This has:

1. Objects: natural numbers $0, 1, \ldots$;
2. Morphisms: braids, e.g.,

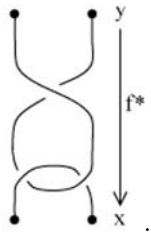
(taken up to deformation);
there are no morphisms $m \longrightarrow n$ when $m \neq n$;
3. Tensor product: placing side–by–side (which on objects means addition); and
4. Braiding: right over left, e.g.,



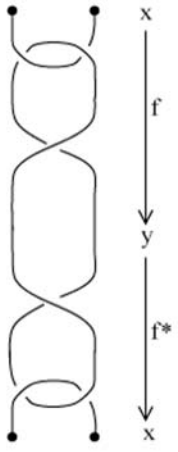
Knots, links and braids are all special cases of *tangles* (see [RT90]). The mysterious relationships between topology, algebra and physics amount in large part to the existence of interesting functors from various topologically defined categories to **Hilbert**, the category of Hilbert spaces. These topologically defined categories are always $*-$*categories*, and the really interesting functors from them to **Hilbert** are always $*-$*functors*, which preserve the $*-$structure. Physically, the $*$ operation corresponds to *reversing the direction of time*. For example, there is a $*-$category whose objects are collections of points and whose morphisms are tangles (see [Bae97, BD98]):



We can think of this morphism $f : x \to y$ as representing the trajectories of a collection of particles and antiparticles, where particles and antiparticles can be created or annihilated in pairs. Reversing the direction of time, we obtain the 'dual' morphism $f^* : y \to x$:



This morphism is not the inverse of f, since the composite $f \circ f^*$ is a nontrivial tangle:



Indeed, any groupoid becomes a $*-$category if we set $f^* = f^{-1}$ for every morphism f.

The above example involves 1D curves in 3D space. More generally, topological quantum field theory studies nD manifolds embedded in $(n + k)$D space–time, which in the $k \to \infty$ limit appear as 'abstract' nD manifolds. It appears that these are best described using certain '$n-$categories with duals', meaning $n-$categories in which every $j-$morphism f has a dual f^*.

Therefore, a tangle is a box in 3D space with knotted and linked string embedded within it and a certain number of strands of that string emanating from the surface of the box. There are no open ends of string inside the box. We usually think of some subset of the strands as *inputs* to the tangle and the remaining strands as the *outputs* from the tangle. Usually the inputs are arranged to be drawn vertically and so that they enter tangle from below, while the outputs leave the tangle from above. The tangle itself (within the box) is arranged as nicely as possible with respect to a vertical direction. This means that a definite vertical direction is chosen, and that the tangle intersects planes perpendicular to this direction transversely except for a finite collection of critical points. These basic critical points are local maxima and local minima for the space curves inside the tangle. Two tangles configured with respect to the same box are *ambient isotopic* if there is an isotopy in

three space carrying one to the other that fixes the input and output strands of each tangle. We can *compose* two tangles A and B where the number of output strand of A is equal to the number of input strands of B. Composition is accomplished by joining each output strand of A to a corresponding input strand of B [KR95, KR99, Kau94].

A *tangle diagram* is a box in the plane, arranged parallel to a chosen vertical direction with a left–right ordered sequence of input strands entering the bottom of the box, and a left–right ordered sequence of output strands emanating from the top of the box. Inside the box is a diagram of the tangle represented with crossings (broken arc indicating the undercrossing line) in the usual way for knot and links. We assume, as above, that the tangle is represented so that it is transverse to lines perpendicular to the vertical except for a finite number of points in the vertical direction along the tangle. It is said that the *tangle is well arranged, or Morse* with respect to the vertical direction when these transversality conditions are met. At the critical points we will see a local maximum, a local minimum or a crossing in the diagram. Tangle composition is well–defined (for matching input/output counts) since the input and output strands have an ordering (from left to right for the reader facing the plane on which the tangle diagram is drawn). Note that the cardinality of the set of input strands or output strands can be equal to zero. If they are both zero, then the tangle is simply a knot or link diagram arranged well with respect to the vertical direction [KR95, KR99, Kau94].

The *Reidemeister moves* are a set of moves on diagrams that combinatorially generate isotopy for knots, links and and tangles [Rei48]. If two tangles are equivalent in 3D space, then corresponding diagrams of these tangles can be obtained one from another, by a sequence of Reidemeister moves. Each move is confined to the tangle box and keeps the input and output strands of the tangle diagram fixed.

Two (tangle) diagrams are said to be *regularly isotopic* if one can be obtained from the other by a sequence of Reidemeister moves of type 0,2,3 (move number 1 is not used in regular isotopy).

If A and B are given tangles, we denote the composition of A and B by AB where the diagram of A is placed below the diagram of B and the output strands of A are connected to the input strands of B. If the cardinalities of the sets of input and output strands are zero, then we simple place one tangle below the other to form the product [KR95, KR99, Kau94].

Along with tangle composition, as defined in the previous paragraph, we also have an operation of *product* or *juxtaposition* of tangles. To juxtapose two tangles A and B simply place their diagrams side by side with A to the left of B and regard this new diagram as a new tangle whose inputs are the inputs of A followed by the inputs of B, and whose outputs are the outputs of A followed by the outputs of B. We denote the tangle product of A and B by $A \otimes B$.

It remains to describe the equivalence relation on tangles that makes them represent regular isotopy classes of embedded string. Every tangle is a compo-

sition of *elementary tangles* where an elementary tangle is one of the following list: a *cup* (a single minimum – zero inputs, two outputs), a *cap* (a single maximum – two inputs, zero outputs), a *crossing* (a single local crossing diagram – two inputs and two outputs).

Ultimate Conceptual Machines: Weak $n-$Categories

As traditionally conceived, an $n-$category is an algebraic structure having objects or $0-$morphisms, $1-$morphisms between $0-$morphisms, $2-$morphisms between $1-$morphisms, and so on up to $n-$morphisms. There should be various ways of composing $j-$morphisms, and these composition operations should satisfy various laws, such as associativity laws. In the so–called *strict* $n-$*categories*, these laws are equations. While well–understood and tractable, strict $n-$categories are insufficiently general for many applications: what one usually encounters in nature are *weak* $n-$*categories*, in which composition operations satisfy the appropriate laws *only up to equivalence*. Here the idea is that $n-$morphisms are equivalent precisely when they are equal, while for $j < n$ an equivalence between $j-$morphisms is recursively defined as a $(j+1)-$morphism from one to the other that is invertible up to equivalence [BD98].

Now, what makes it difficult to define weak $n-$categories is that laws formulated as equivalences should satisfy laws of their own – the so–called *coherence laws* – so that one can manipulate them with some of the same facility as equations. Moreover, these coherence laws should also be equivalences satisfying their own coherence laws, again up to equivalence, and so on [BD98].

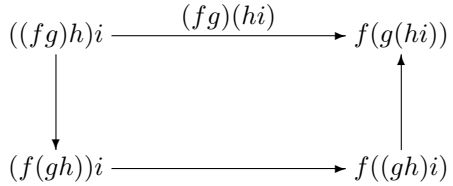
For example, a weak $1-$category is just an ordinary category. In a category, composition of $1-$morphisms is associative:

$$(fg)h = f(gh).$$

Weak $2-$categories first appeared in the work of Bénabou [Ben67], under the name of *bicategories*. In a bicategory, composition of $1-$morphisms is associative only up to an invertible $2-$morphism, the '*associator*':

$$A_{f,g,h} \colon (fg)h \to f(gh).$$

The associator allows one to re–bracket parenthesized composites of arbitrarily many $1-$morphisms, but there may be many ways to use it to go from one parenthesization to another. For all these to be equal, the associator must satisfy a coherence law, the pentagon identity, which says that the following diagram commutes:



where all the arrows are 2−morphisms built using the associator. Weak 3−categories or *tricategories* were defined by [GPS95]. In a tricategory, the pentagon identity holds only up to an invertible 3−morphism, which satisfies a further coherence law of its own.

When one explicitly lists the coherence laws this way, the definition of weak n−category tends to grow ever more complicated with increasing n. To get around this, one must carefully study the origin of these coherence laws. So far, most of our insight into coherence laws has been won through homotopy theory, where it is common to impose equations *only up to homotopy*, with these homotopies satisfying coherence laws, again up to homotopy, and so on. For example, the pentagon identity and higher coherence laws for associativity first appeared in Stasheff's work on the structure inherited by a space equipped with a homotopy equivalence to a space with an associative product [Sta63]. Subsequent work led to a systematic treatment of coherence laws in homotopy theory through the formalism of *topological operads* [Ada78].

Underlying the connection between homotopy theory and n−category theory is a hypothesis made quite explicit by Grothendieck [Gro83]: to any topological space one should be able to associate an n−category having points as objects, paths between points as 1−morphisms, certain paths of paths as 2−morphisms, and so on, with certain homotopy classes of n−fold paths as n−morphisms. This should be a special sort of weak n−category called a *weak n−groupoid*, in which all j−morphisms ($0 < j \leq n$) are equivalences. Moreover, the process of assigning to each space its *fundamental n−groupoid*, as Grothendieck called it, should set up a complete correspondence between the theory of homotopy n−types (spaces whose homotopy groups vanish above the nth) and the theory of weak n−groupoids. This hypothesis explains why all the coherence laws for weak n−groupoids should be deducible from homotopy theory. It also suggests that weak n−categories will have features not found in homotopy theory, owing to the presence of j−morphisms that are not equivalences [BD98].

Homotopy theory also makes it clear that when setting up a theory of n−categories, there is some choice involved in the shapes of ones j−morphisms – or in the language of topology, *j−cells*. The traditional approach to n−categories is *globular*. This means that for $j > 0$, each j−cell $f\colon x \to y$ has two $(j-1)$−cells called its *source*, $sf = x$, and *target*, $tf = y$, which for $j > 1$ satisfy

$$s(sf) = s(tf), \qquad t(sf) = t(tf)).$$

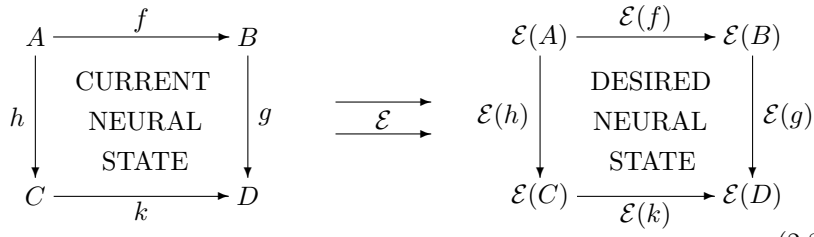
Thus a j−cell can be visualized as a *globe*, a jD ball whose boundary is divided into two $(j-1)$D hemispheres corresponding to its source and target. However, in homotopy theory, the simplicial approach is much more popular. In a *simplicial set*, each j−cell f is shaped like a jD simplex, and has $j+1$ faces, certain $(j-1)$−cells $d_0 f, \ldots, d_n f$. In addition to these there are $(j+1)$−cells $i_0 f, \ldots, i_{n+1} f$ called *degeneracies*, and the face and degeneracy maps satisfy certain well–known relations [BD98].

2.5.8 Brain–Mind Functorial Machines

In this section we propose two models of the *brain–mind functorial machines*: the first one is a psychologically–motivated top–down machine, while the second one is physically–motivated bottom–up solitary machine.

Neurodynamical 2−Functor

Here we define the goal–directed cognitive neurodynamics as an evolution 2−functor $\mathcal{E}$ given by

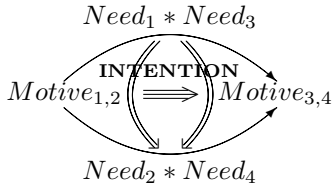


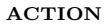
(2.813)

In (2.813), $\mathcal{E}$ represents a projection map from the *source* 2−*category* of the current neural state, defined as a commutative square of small categories $A, B, C, D, \ldots$ of *current neural ensembles* and their causal interrelations $f, g, h, k, \ldots$, onto the *target* 2−*category* of the desired neural state, defined as a commutative square of small categories $\mathcal{E}(A), \mathcal{E}(B)$, $\mathcal{E}(C), \mathcal{E}(D), \ldots$ of *evolved neural ensembles* and their causal interrelations $\mathcal{E}(f), \mathcal{E}(g), \mathcal{E}(h), \mathcal{E}(k)$.

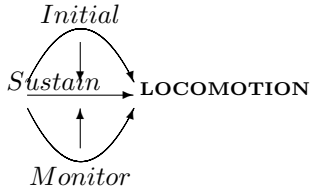
The evolution 2−functor $\mathcal{E}$ can be horizontally decomposed in the following three neurodynamic components (see [Lew97, AL91]):

1. *Intention*, defined as a 3−cell:



2. *Action*, defined as a 1−cell: 

3. *Locomotion*, defined as a 2−cell:



Now, each causal arrow in (2.813), say $f : A \to B$, stands for a generic 'neuro–morphism', representing a self–organized, oscillatory neurodynamic system.

We define a generic neuro–morphism f to be a nonlinear tensor–field $(x, y, \boldsymbol{\omega})$–system (2.814–2.819), acting as a bidirectional associative memory machine on a ND Riemannian manifold M^N of the human cortex. It is formed out of two distinct, yet nonlinearly–coupled neural subsystems:

1. Activation (x, y)–dynamics (2.814–2.815), defined as an interplay of an excitatory vector–field $x^i = x^i(t) : M^N \to TM$, representing a cross–section of the tangent bundle TM, and and an inhibitory 1−form $y_i = y_i(t) : M^N \to T^*M$, representing a cross–section of the cotangent bundle T^*M.
2. Excitatory and inhibitory unsupervised learning $(\boldsymbol{\omega})$–dynamics (2.816–2.819) generated by random differential Hebbian learning process (2.818–2.819), defined respectively by contravariant synaptic tensor–field $\omega^{ij} = \omega^{ij}(t) : M^N \to TTM^N_{im}$ and covariant synaptic tensor–field $\omega_{ij} = \omega_{ij}(t) : M^N \to T^*T^*M$, representing cross–sections of contravariant and covariant tensor bundles, respectively.

$(x, y, \boldsymbol{\omega})$–system is analytically defined as a set of N coupled neurodynamic equations:

$$\dot{x}^i = A^i + \omega^{ij}\, f_j(y) - x^i, \tag{2.814}$$

$$\dot{y}_i = B_i + \omega_{ij}\, f^j(x) - y_i, \tag{2.815}$$

$$\dot{\omega}^{ij} = -\omega^{ij} + I^{ij}(x, y), \tag{2.816}$$

$$\dot{\omega}_{ij} = -\omega_{ij} + I_{ij}(x, y), \tag{2.817}$$

$$I^{ij} = f^i(x)\, f^j(y) + \dot{f}^i(x)\, \dot{f}^j(y) + \sigma^{ij}, \tag{2.818}$$

$$I_{ij} = f_i(x)\, f_j(y) + \dot{f}_i(x)\, \dot{f}_j(y) + \sigma_{ij}, \tag{2.819}$$

$$(i, j = 1, \ldots, N).$$

Here $\boldsymbol{\omega}$ is a symmetric, second–order synaptic tensor–field; $I^{ij} = I^{ij}(x, y, \sigma)$ and $I_{ij} = I_{ij}(x, y, \sigma)$ respectively denote contravariant–excitatory and covariant–inhibitory random differential Hebbian innovation–functions with tensorial Gaussian noise $\boldsymbol{\sigma}$ (in both variances); fs and $\dot{f}$s denote sigmoid activation functions ($f = \tanh(.)$) and corresponding signal velocities ($\dot{f} = 1 - f^2$), respectively in both variances; $A^i = A^i(t)$ and $B_i = B_i(t)$ are contravariant–excitatory and covariant–inhibitory neural inputs to the corresponding cortical cells, respectively;

Nonlinear activation (x, y)–dynamics, describes a two–phase biological neural oscillator field, in which the excitatory neural field excites the inhibitory neural field, which itself reciprocally inhibits the excitatory one. (x, y)–dynamics represents a nonlinear extension of a linear, Lyapunov–stable, conservative, gradient system, defined in local neural coordinates $x^i, y_i \in V_y$ on T^*M as

$$\dot{x}^i = -\frac{\partial \Phi}{\partial y_i} = \omega^{ij} y_j - x^i, \qquad \dot{y}_i = -\frac{\partial \Phi}{\partial x^i} = \omega_{ij} x^j - y_i. \tag{2.820}$$

The gradient system (2.820) is derived from scalar, neuro–synaptic action potential $\Phi : T^*M \to \mathbb{R}$, given by a negative, smooth bilinear form in $x^i, y_i \in V_y$ on T^*M as

$$-2\Phi = \omega_{ij} x^i x^j + \omega^{ij} y_i y_j - 2x^i y_i, \qquad (i, j = 1, \ldots, N),$$

which itself represents a Ψ–image of the Riemannian metrics $g : TM \to \mathbb{R}$ on the configuration manifold M^N.

The nonlinear oscillatory activation (x, y)–dynamics (2.814–2.817) is get from the linear conservative dynamics (2.820), by adding configuration dependent inputs A^i and B_i, as well as sigmoid activation functions f_j and f^j, respectively. It represents an interconnected pair of excitatory and inhibitory neural fields.

Both variant–forms of learning $(\boldsymbol{\omega})$–dynamics (2.816–2.817) are given by a generalized unsupervised (self–organizing) Hebbian learning scheme (see [Kos92]) in which $\dot{\omega}_{ij}$ (resp. $\dot{\omega}^{ij}$) denotes the new–update value, $-\omega_{ij}$ (resp. ω^{ij}) corresponds to the old value and $I_{ij}(x^i, y_j)$ (resp. $I^{ij}(x^i, y_j)$) is the innovation function of the symmetric 2nd order synaptic tensor–field $\boldsymbol{\omega}$. The nonlinear innovation functions I_{ij} and I^{ij} are defined by random differential Hebbian learning process (2.818–2.819). As $\boldsymbol{\omega}$ is a symmetric and zero–trace coupling synaptic tensor, the conservative linear activation dynamics (2.820) is equivalent to the rule that 'the state of each neuron (in both neural fields) is changed in time if, and only if, the scalar action potential Φ (52), is lowered'. Therefore, the scalar action potential Φ represents the monotonically decreasing Lyapunov function (such that $\dot{\Phi} \leq 0$) for the conservative linear dynamics (2.820), which converges to a local minimum or ground state of Φ. That is to say, the system (2.820) moves in the direction of decreasing the scalar action potential Φ, and when both $\dot{x}^i = 0$ and $\dot{y}_i = 0$ for all $i = 1, \ldots, N$, the steady state is reached.

Solitary 'Thought Nets' and 'Emerging Mind'

Synergetic 'Thought Solitons'

Recall that synergetics teaches us that *order parameters* (and their spatio–temporal evolution) are *patterns, emerging from chaos*. In our opinion, the most important of these order parameters, both natural and man made, are *solitons*, because of their self–organizing quality to create order out of chaos. From this perspective, *nonlinearity* – the essential characteristic of nature – is the *cause* of both *chaos* and *order*. Recall that the solitary particle–waves, also called the 'light bullets', are localized space–time excitations $\Psi(x, t)$, propagating through a certain medium Ω with constant velocities v_j. They describe a variety of nonlinear wave phenomena in one dimension and playing important roles in optical fibers, many branches of physics, chemistry and biology.

To derive our solitary network we start with modelling the conservative '*thought solitons*', using the following three classical nonlinear equations, defining the time evolution of the spatio–temporal wave function $\Psi(x,t)$ (which is smooth, and either complex–, or real–valued) (see [NMP84, For90, AC91, IP01a]; also compare with (2.4.2) above):

1. *Nonlinear Schrödinger (NS) equation*

$$\mathrm{i}\Psi_t \;=\; 2\mu|\Psi|^2\Psi - \Psi_{xx}\,, \tag{2.821}$$

where $\Psi = \Psi(x,t)$ is a complex-valued wave function with initial condition $\Psi(x,t)|_{t=0} \;=\; \Psi(x)$ and μ is a nonlinear parameter representing field strength. In the linear limit ($a \;=\; 0$) NS becomes the ordinary Schrödinger equation for the wave function of the free 1D particle with mass $m \;=\; 1/2$. Its Hamiltonian function

$$H_{NS} \;=\; \int_{-\infty}^{+\infty} \Big(\,\mu|\Psi|^4 + |\Psi_x|^2\Big)\,dx,$$

is equal to the total and conserved energy of the soliton. NS describes, for example, nonlinear Faraday resonance in a vertically oscillating water, an easy–plane ferromagnet with a combination of a stationary and a high–frequency magnetic fields, and the effect of phase–sensitive amplifiers on solitons propagating in optical fibers.
2. *Korteveg–de Vries (KdV) equation*

$$\Psi_t \;=\; 6\Psi\Psi_x \;-\Psi_{xxx}\,,$$

with Hamiltonian (total conserved energy) given by

$$H_{KdV} \;=\; \int_{-\infty}^{+\infty} \left(\Psi^3 \;+\; \frac{1}{2}\Psi_x^2\right) dx.$$

KdV is related to the ordinary Schrödinger equation by the inverse scattering method. KdV is a well–known model of 1D turbulence that was derived in various physical contexts, including chemical–reaction waves, propagation of combustion fronts in gases, surface waves in a film of a viscous liquid flowing along an inclined plane, patterns in thermal convection, rapid solidification, and others. Its discretization gives the Lotka–Voltera equation

$$\dot{x}^j(t) = x^j(t)\left(x^{j+1}(t) - x^{j-1}(t)\right),$$

which appears in a model of struggle for existence of biological species.
3. *Sine–Gordon (SG) equation*

$$\Psi_{tt} = \Psi_{xx} - \sin\Psi,$$

with Hamiltonian (total conserved enegy) given by

$$H_{SG} = \int_{-\infty}^{+\infty} \left(\Psi_t^2 + \Psi_x^2 + \cos\Psi\right) dx.$$

SG gives one of the simplest models of the unified field theory, can be found in the theory of dislocations in metals, in the theory of Josephson junctions and so on. It can be used also in interpreting certain biological processes like DNA dynamics. Its discretization gives a *system of coupled pendulums*.

Discrete solitons exist also in the form of the *soliton celular automata* (SCA) [PST86]. SCA is a 1(space)+1(time)–dimensional 'box and ball system' made of infinite number of zeros (or, boxes) and finite number of ones (or, balls). The value of the jth SCA cell a_t^j at a discrete time time t, is given as

$$a_{t+1}^j = \begin{cases} 1, & \text{if} \quad a_t^j = 0 \quad \text{and} \quad \sum_{i=-\infty}^{j-1} u_t^i > \sum_{i=-\infty}^{j-1} a_{t+1}^i \,, \\ \\ 0, & \text{otherwise,} \end{cases}$$

where $a_t^j = 0$ is assumed for $|j| \gg 1$. Any state of the SCA consists purely of solitons (particularly, KdV–solitons), possessing conserved quantities of the form of H_{KdV}. All of these properties have motivated a number of suggestive applications for a new kind of computational architecture that will use these evolution patterns of SCA in order to give a 'gateless' implementation of logical operations.

In practice, both SCA and KdV are usually approximated by the *Toda lattice equation*,

$$\ddot{q}^i = e^{q^{i+1}-q^i} - e^{q^i-q^{i-1}}, \qquad (i = 1, ..., N) \tag{2.822}$$

with quasiperiodic $\quad q^{N+i}(t) = q^i(t) + c, \qquad$ or,

fast–dacaying boundary conditions $\quad \lim_{i\to-\infty} q^i(t) = 0, \qquad \lim_{i\to+\infty} q^i(t) = c.$

The Toda equation (2.822) is a *gradient Newtonian equation* of motion

$$\ddot{q}^i = -\partial_q^i V, \qquad V(q) = \sum_{i=1}^{N} e^{q^{i+1}-q^i}.$$

Otherwise, the Toda equation represents a Hamiltonian system

$$\dot{q}^i = p_i, \qquad \dot{p}_i = e^{q^{i+1}-q^i} - e^{q^i-q^{i-1}},$$

with the phase–space $P = \mathbb{R}^{2N}$ with coordinates (p_i, q^i), standard Poisson structure

$$\{p_i, p_j\} = \{q^i, q^j\} = 0, \qquad \{p_i, q^j\} = \delta_i^j,$$

and Hamiltonian function $\quad H = \sum_{i=1}^{N} (\frac{1}{2} p_i^2 + e^{q^{i+1}-q^i}), \qquad (i, j = 1, ..., N).$

Next, to make our conservative thought solitons *open to the environment*, we have to modify them by adding:

1. Input from the senses, in the form of the Weber–Fechner's law,

$$S(t) = a_r \log s^r(t), \qquad (r = 1, ..., 5), \tag{2.823}$$

where $S = S(t)$ is the *sensation*, $s^r = s^r(t)$ the vector of stimuli from the five senses, and a_r a constant vector; and

2. Disturbances, in the form of additive, zero–mean Gaussian white noise $\eta = \eta(t)$, independent from the main soliton–signal.

In this way, we get the *modified solitary equations*:

$$\begin{aligned}
\text{MNS}: \quad & \mathrm{i}\Psi_t = 2\mu|\Psi|^2\Psi - \Psi_{xx} + a_r \log s^r \Psi + \eta, \\
\text{MKdV}: \quad & \Psi_t = 6\Psi\Psi_x - \Psi_{xxx} + a_r \log s^r \Psi + \eta, \\
\text{MSG}: \quad & \Psi_{tt} = \Psi_{xx} - \sin\Psi + a_r \log s^r \Psi + \eta,
\end{aligned}$$

representing the *three different models of the thought units*.

Now we will form a single emerging order–parameter, the general factor, that we call the *Mind*. It behaves like an orchestrated ensemble of thought solitons, defined as systems of trainable, coupled soliton equations. Their tensor couplings perform *self–organizing associative learning by trial and error*, similar to that of the neural ensemble.

The dynamics of the *soliton ensemble*, representing our model of the 'mind' can be described as one of the following three soliton systems; each of them performs *learning, growing* and *competing* between each other, and *communicates with environment* through the *sensory inputs* and the *heating noise*:

1. Coupled modified nonlinear Schrödinger equations

$$\begin{aligned}
\mathrm{i}\Psi_t^k = & -\Psi_{xx}^k + 2\mu^k \sum_{j\neq k} |\Psi^k|^2 \mathbf{W}_k^j S^j(\Psi^j) \\
& + \nu^k \Psi^k (1 - \epsilon^k \Psi^k) + a_r \log s^r \Psi^k + \eta^k,
\end{aligned}$$

2. Coupled modified Korteveg–de Vries equations

$$\begin{aligned}
\Psi_t^k = & \; 6\Psi_x^k \Psi^k - \Psi_{xxx}^k + \sum_{j\neq k} \mathbf{W}_k^j S^j(\Psi^j) \\
& + \nu^k \Psi^k (1 - \epsilon^k \Psi^k) + a_r \log s^r \Psi^k + \eta^k,
\end{aligned}$$

3. Coupled modified Sine–Gordon equations

$$\Psi_{tt}^k = \Psi_{xx}^k - \sin\Psi^k + \sum_{j\neq k} + \nu^k \Psi^k (1 - \epsilon^k \Psi^k) + a_r \log s^r \Psi^k + \eta^k,$$

where $\Psi^k = \Psi^k(x,t)$, $(k = 1, ..., n)$ is the set of wave functions of the solitary thoughts, $S(\cdot)$ represents the sigmoidal threshold functions, ν^k and ϵ^k are *growing* and *competition* parameters.

$\mathbf{W}_k^j = \mathbf{W}_k^j(\Psi)$ are *tensorial learning couplings*, evolving according to the Hebbian learning scheme (see [Kos92]):

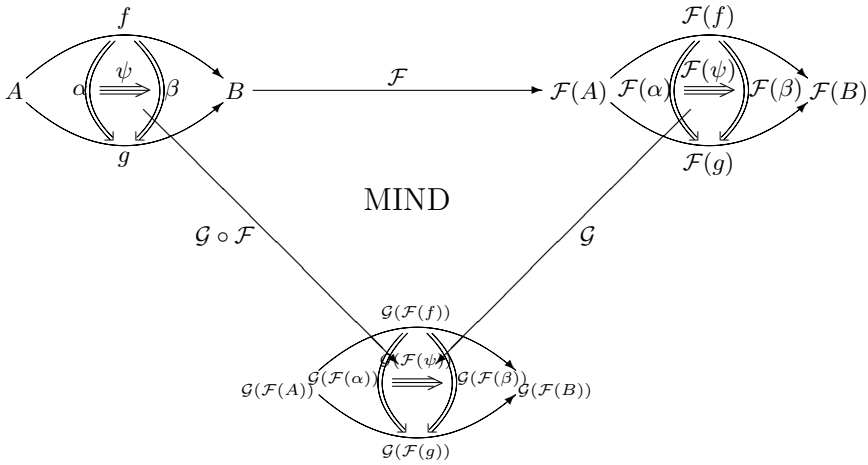
$$\dot{\mathbf{W}}_k^j = -\mathbf{W}_k^j + \Phi_k^j(\Psi^k, \Psi^j),$$

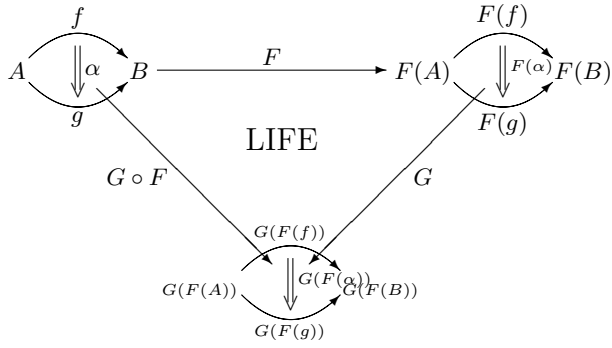
with *innovation* defined in tensor signal form (here $\dot{S}(\cdot) = 1 - \tanh(\cdot)$)

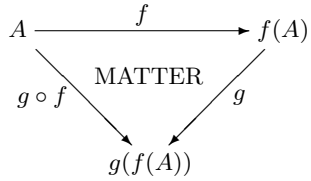
$$\Phi_k^j = S^j(\Psi^j)\, S^k(\Psi^k) + \dot{S}^j(\Psi^j)\, \dot{S}^k(\Psi^k).$$

Emerging Categorical Structure: $MATTER \Rightarrow LIFE \Rightarrow MIND$

The solitary thought nets effectively simulate the following 3−categorical structure of MIND, emerging from the 2−categorical structure of LIFE, which is itself emerging from the 1−categorical structure of MATTER:







2.6 Psycho–Socio–Economic Dynamics

2.6.1 Force–Field Psychodynamics

In this section, which is written in the fashion of the *quantum brain*, we present the top level of natural biodynamics, using geometrical generalization of the *Feynman path integral*. To formulate the basics of *force–field psychodynamics*, we use the *action–amplitude picture* of the $BODY \rightleftarrows MIND$ adjunction:

$$\downarrow \textbf{ Deterministic (causal) world of } \textbf{\textit{Human BODY}} \downarrow$$

$$Action : S[q^n] = \int_{t_{in}}^{t_{out}} (E_k - E_p + Wrk + Src^{\pm})\, dt$$

$$- -$$

$$Amplitude : \langle out|in\rangle \;=\; \oint\!\!\!\!\!\!\sum \mathcal{D}[w_n q^n]\, \mathrm{e}^{\mathrm{i}S[q^n]}$$

$$\uparrow \textbf{ Probabilistic (fuzzy) world of } \textbf{\textit{Human MIND}} \uparrow$$

In the action integral, E_k, E_p, Wrk and $Src^{\pm}$ denote the kinetic end potential energies, work done by dissipative/driving forces and other energy sources/sinks, respectively. In the amplitude integral, the peculiar sign $\oint\!\!\!\!\!\!\sum$ denotes integration along smooth paths and summation along discrete Markov chains; i is the imaginary unit, w_n are synaptic–like weights, while $\mathcal{D}$ is the Feynman path differential (defined below) calculated along the configuration trajectories q^n. The action $S[q^n]$, through the *least action principle* $\delta S = 0$, leads to all biodynamic equations considered so far (in generalized Lagrangian and Hamiltonian form). At the same time, the action $S[q^n]$ figures in the exponent of the path integral $\oint\!\!\!\!\!\!\sum$, defining the probability transition amplitude $\langle out|in\rangle$. In this way, the whole body dynamics is incorporated in the mind dynamics. This *adaptive path integral* represents an *infinite–dimensional neural network*, suggesting an infinite capacity of human brain/mind.

For a long time the cortical systems for *language and actions* were believed to be independent modules. However, according to the recent research of [Pul05], as these systems are reciprocally connected with each other, information about language and actions might interact in distributed neuronal assemblies. A critical case is that of action words that are semantically related to different parts of the body (e.g. 'pick', 'kick', 'lick',...). The author suggests that the comprehension of these words might specifically, rapidly and automatically activate the motor system in a somatotopic manner, and that their comprehension rely on activity in the action system.

Motivational Cognition in the Life Space Foam

Applications of nonlinear dynamical systems (NDS) theory in psychology have been encouraging, if not universally productive/effective [Met97]. Its historical antecedents can be traced back to Piaget's [PHE92] and Vygotsky's [Vyg82] interpretations of the dynamic relations between action and thought, Lewinian theory of social dynamics and cognitive–affective development [Lew51, Gol99], and Bernstein's [Ber47] theory of self–adjusting, goal–driven motor action.

Now, both the original *Lewinian force–field theory* in psychology (see [Lew51, Gol99]) and modern decision–field dynamics (see [BT93, RBT01, BD02]) are based on the classical Lewinian concept of an individual's *life space*.[39] As a topological construct, Lewinian life space represents a person's psychological environment that contains *regions* separated by dynamical permeable *boundaries*. As a field construct, on the other hand, the life space is not empty: each of its regions is characterized by *valence* (ranging from positive or negative and resulting from an interaction between the person's *needs* and the dynamics of their *environment*). Need is an energy construct, according to Lewin. It creates *tension* in the person, which, in combination with other tensions, initiates and sustains behavior. Needs vary from the most primitive urges to the most idiosyncratic intentions and can be both internally generated (e.g., thirst or hunger) and stimulus–induced (e.g., an urge to buy something in response to a TV advertisement). Valences are, in essence, personal values dynamically derived from the person's needs and attached to various regions in their life space. As a field, the life space generates forces pulling the person towards positively–valenced regions and pushing them away from regions with negative valence. Lewin's term for these forces is *vectors*. Combinations of multiple vectors in the life space cause the person to move from one region towards another. This movement is termed *locomotion* and it may range from overt behavior to cognitive shifts (e.g., between alternatives in a decision–making process). Locomotion normally results in crossing the boundaries between regions. When their permeability is degraded, these boundaries become *barriers* that restrain locomotion. Life space model, thus, offers a meta–theoretical language to describe a wide range of behaviors, from goal–directed action to intrapersonal conflicts and multi–alternative decision–making.

In order to formalize the Lewinian life–space concept, a set of *action principles* need to be associated to Lewinian force–fields, (loco)motion paths (representing mental abstractions of biomechanical paths [Iva04]) and life space geometry. As an extension of the Lewinian concept, in this paper we introduce a new concept of *life–space foam* (LSF, see Figure 2.32). According to this new concept, Lewin's life space can be represented as a *geometrical functor* with globally smooth macro–dynamics, which is at the same time underpinned by

[39] The work presented in this subsection has been developed in collaboration with Dr. Eugene Aidman, Senior Research Scientist, Human Systems Integration, Land Operations Division, Defence Science & Technology Organisation, Australia.

wildly fluctuating, non–smooth, local micro–dynamics, describable by *Feynman's*: (i) *sum–over–histories* $\oint_{paths}$, (ii) *sum–over–fields* $\oint_{fields}$, and (iii) *sum–over–geometries* $\oint_{geom}$.

LSF is thus a two–level *geometrodynamical functor*, representing these two distinct types of dynamics within the Lewinian life space. At its *macroscopic spatio–temporal level*, LSF appears as a 'nice & smooth' geometrical functor with globally predictable dynamics – formally, a smooth $n-$dimensional manifold M with local Riemannian metrics $g_{ij}(x)$, smooth force–fields and smooth (loco)motion paths, as conceptualized in the Lewinian theory. To model the global and smooth macro–level LSF–paths, fields and geometry, we use the general physics–like *principle of the least action*.

Now, the apparent smoothness of the macro–level LSF is achieved by the existence of another level underneath it. This *micro–level* LSF is actually a collection of wildly fluctuating force–fields, (loco)motion paths, curved regional geometries and topologies with holes. The micro–level LSF is proposed as an extension of the Lewinian concept: it is characterized by uncertainties and fluctuations, enabled by microscopic time–level, microscopic transition paths, microscopic force–fields, local geometries and varying topologies with holes. To model these fluctuating microscopic LSF–structures, we use three instances of *adaptive path integral*, defining a multi–phase and multi–path (also multi–field and multi–geometry) *transition* process from *intention* to the goal–driven *action*.

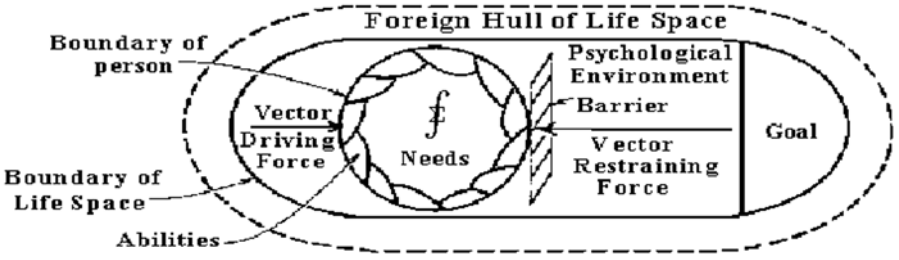


Fig. 2.32. Diagram of the *life space foam*: Lewinian life space with an adaptive path integral acting inside it and generating microscopic fluctuation dynamics.

We use the new LSF concept to develop modelling framework for motivational dynamics (MD) and induced cognitive dynamics (CD).

According to Heckhausen (see [Hec77]), *motivation* can be thought of as a process of *energizing* and *directing the action*. The process of energizing can be represented by Lewin's *force–field analysis* and Vygotsky's *motive formation* (see [Vyg82, AL91]), while the process of directing can be represented by *hierarchical action control* (see [Ber47, Ber35, Kuh85]).

Motivation processes both precede and coincide with every goal–directed action. Usually these motivation processes include the sequence of the following four feedforward *phases* [Vyg82, AL91]: (*)

1. *Intention Formation* $\mathcal{F}$, including: decision making, commitment building, etc.
2. *Action Initiation* $\mathcal{I}$, including: handling conflict of motives, resistance to alternatives, etc.
3. *Maintaining the Action* $\mathcal{M}$, including: resistance to fatigue, distractions, etc.
4. *Termination* $\mathcal{T}$, including parking and avoiding addiction, i.e., staying in control.

With each of the phases $\{\mathcal{F}, \mathcal{I}, \mathcal{M}, \mathcal{T}\}$ in (*), we can associate a *transition propagator* – an ensemble of (possibly crossing) feedforward paths propagating through the 'wood of obstacles' (including topological holes in the LSF, see Figure 2.33), so that the complete *transition functor* $\mathcal{T}A$ is a product of propagators (as well as sum over paths). All the phases–propagators are controlled by a unique *Monitor* feedback process.

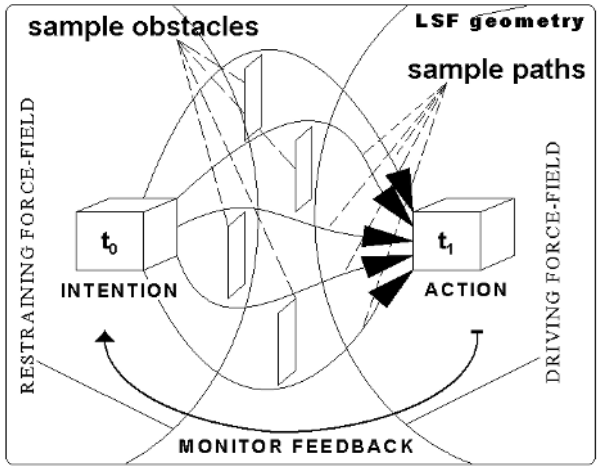


Fig. 2.33. *Transition–propagator* corresponding to each of the motivational phases $\{\mathcal{F}, \mathcal{I}, \mathcal{M}, \mathcal{T}\}$, consisting of an ensemble of feedforward paths propagating through the 'wood of obstacles'. The paths affected by driving and restraining force–fields, as well as by the local LSF–geometry. Transition goes from *Intention*, occurring at a sample time instant t_0, to *Action*, occurring at some later time t_1. Each propagator is controlled by its own *Monitor* feedback. All together they form the transition functor $\mathcal{T}A$.

In this subsection we propose an *adaptive path integral* formulation for the motivational–transition functor $\mathcal{T}A$. In essence, we sum/integrate over differ-

ent paths and make a product (composition) of different phases–propagators. Recall that this is the most general description of the general *Markov stochastic process*.

We will also attempt to demonstrate the utility of the same LSF–formalisms in representing cognitive functions, such as memory, learning and decision making. For example, in the classical *Stimulus encoding* $\longrightarrow$ *Search* $\longrightarrow$ *Decision* $\longrightarrow$ *Response* sequence [Ste69, Ash94], the environmental input–triggered *sensory memory* and *working memory* (WM) can be interpreted as operating at the micro–level force–field under the executive control of the *Monitor* feedback, whereas *search* can be formalized as a *control* mechanism guiding retrieval from the long–term memory (LTM, itself shaped by learning) and filtering material relevant to decision making into the WM. The essential measure of these mental processes, the *processing speed* (essentially determined by Sternberg's reaction–time) can be represented by our (loco)motion speed $\dot{x}$.

Six Faces of the Life Space Foam

The LSF has three forms of appearance: $paths + field + geometries$, acting on both macro–level and micro–level, which is six modes in total. In this section, we develop three least action principles for the macro–LSF–level and three adaptive path integrals for the micro–LSF–level. While developing our psycho–physical formalism, we will address the behavioral issues of motivational fatigue, learning, memory and decision making.

General Formalism

At both macro– and micro–levels, the total LSF represents a union of transition paths, force–fields and geometries, formally written as

$$\begin{aligned} LSF_{total} &:= LSF_{paths} \bigcup LSF_{fields} \bigcup LSF_{geom} \\ &\equiv \oint\!\!\!\!\!\!\cdot_{paths} + \oint\!\!\!\!\!\!\cdot_{fields} + \oint\!\!\!\!\!\!\cdot_{geom} . \end{aligned} \tag{2.824}$$

Corresponding to each of the three LSF–subspaces in (2.824) we formulate:

1. The *least action principle*, to model deterministic and predictive, macro–level MD & CD, giving a unique, global, causal and smooth path–field–geometry on the macroscopic spatio–temporal level; and
2. Associated *adaptive path integral* to model uncertain, fluctuating and probabilistic, micro–level MD & CD, as an ensemble of local paths–fields–geometries on the microscopic spatio–temporal level, to which the global macro–level MD & CD represents both time and ensemble *average* (which are equal according to the *ergodic hypothesis*).

In the proposed formalism, transition paths $x^i(t)$ are affected by the force–fields $\varphi^k(t)$, which are themselves affected by geometry with metric g_{ij}.

Global Macro–Level of LSF_{total}. In general, at the *macroscopic* LSF–level we first formulate the *total action* $S[\Phi]$, the central quantity in our formalism that has psycho–physical dimensions of $Energy \times Time = Effort$, with immediate cognitive and motivational applications: *the greater the action – the higher the speed of cognitive processes and the lower the macroscopic fatigue* (which includes all sources of physical, cognitive and emotional fatigue that influence motivational dynamics). The action $S[\Phi]$ depends on macroscopic paths, fields and geometries, commonly denoted by an abstract field symbol Φ^i. The action $S[\Phi]$ is formally defined as a temporal integral from the *initial* time instant t_{ini} to the *final* time instant t_{fin},

$$S[\Phi] = \int_{t_{ini}}^{t_{fin}} \mathfrak{L}[\Phi]\, dt, \tag{2.825}$$

with *Lagrangian density* given by

$$\mathfrak{L}[\Phi] = \int d^n x\, \mathcal{L}(\Phi_i, \partial_{x^j}\Phi^i),$$

where the integral is taken over all n coordinates $x^j = x^j(t)$ of the LSF, and $\partial_{x^j}\Phi^i$ are time and space partial derivatives of the Φ^i-variables over coordinates.

Second, we formulate the *least action principle* as a minimal variation δ of the action $S[\Phi]$

$$\delta S[\Phi] = 0, \tag{2.826}$$

which, using techniques from the calculus of variations gives, in the form of the so–called Euler–Lagrangian equations, a shortest (loco)motion path, an extreme force–field, and a life–space geometry of minimal curvature (and without holes). In this way, we effectively derive a *unique globally smooth transition functor*

$$\mathcal{TA} : INTENTION_{t_{ini}} \Rrightarrow ACTION_{t_{fin}}, \tag{2.827}$$

performed at a macroscopic (global) time–level from some initial time t_{ini} to the final time t_{fin}.

In this way, we get macro–objects in the global LSF: a single path described Newtonian–like equation of motion, a single force–field described by Maxwellian–like field equations, and a single obstacle–free Riemannian geometry (with global topology without holes).

For example, recall that in the period 1945–1949 J. Wheeler and R. Feynman developed their *action-at-a-distance electrodynamics* [WF49], in complete experimental agreement with the classical Maxwell's electromagnetic theory, but at the same time avoiding the complications of divergent self–interaction of the Maxwell's theory as well as eliminating its infinite number of field degrees of freedom. In Wheeler–Feynman view, "Matter consists of electrically charged particles," so they found a form for the action directly involving the

motions of the charges only, which upon variation would give the Newtonian–like equations of motion of these charges. Here is the expression for this action in the flat space–time, which is in the core of quantum electrodynamics:

$$S[x;t_i,t_j] = \frac{1}{2}m_i \int (\dot{x}^i_\mu)^2\, dt_i + \frac{1}{2}e_i e_j \int\int \delta(I^2_{ij})\, \dot{x}^i_\mu(t_i)\dot{x}^j_\mu(t_j)\, dt_i dt_j$$
$$\text{with} \tag{2.828}$$
$$I^2_{ij} = \left[x^i_\mu(t_i) - x^j_\mu(t_j)\right]\left[x^i_\mu(t_i) - x^j_\mu(t_j)\right],$$

where $x^i_\mu = x^i_\mu(t_i)$ is the four–vector position of the ith particle as a function of the proper time t_i, while $\dot{x}^i_\mu(t_i) = dx^i_\mu/dt_i$ is the velocity four–vector. The first term in the action (2.828) is the ordinary mechanical action in Euclidean space, while the second term defines the electrical interaction of the charges, representing the Maxwell–like field (it is summed over each pair of charges; the factor $\frac{1}{2}$ is to count each pair once, while the term $i = j$ is omitted to avoid self–action; the interaction is a double integral over a delta function of the square of space–time interval I^2 between two points on the paths; thus, interaction occurs only when this interval vanishes, that is, along light cones [WF49]).

Now, from the point of view of Lewinian geometrical force–fields and (loco)motion paths, we can give the following life–space interpretation to the Wheeler–Feynman action (2.828). The mechanical–like locomotion term occurring at the single time t, needs a covariant generalization from the flat 4D Euclidean space to the nD smooth Riemannian manifold, so it becomes (see e.g., [Iva04])

$$S[x] = \frac{1}{2}\int_{t_{ini}}^{t_{fin}} g_{ij}\, \dot{x}^i \dot{x}^j\, dt,$$

where g_{ij} is the Riemannian metric tensor that generates the total 'kinetic energy' of (loco)motions in the life space.

The second term in (2.828) gives the sophisticated definition of Lewinian force–fields that drive the psychological (loco)motions, if we interpret electrical charges e_i occurring at different times t_i as motivational charges – needs.

Local Micro–Level of LSF_{total}. After having properly defined macro–level MD & CD, with a unique transition map F (including a unique motion path, driving field and smooth geometry), we move down to the *microscopic* LSF–level of rapidly fluctuating MD & CD, where we cannot define a unique and smooth path–field–geometry. The most we can do at this level of *fluctuating uncertainty*, is to formulate an adaptive path integral and calculate overall probability amplitudes for ensembles of local transitions from one LSF–point to the neighboring one. This *probabilistic transition micro–dynamics* functor is defined by a multi–path (field and geometry, respectively) and multi–phase *transition amplitude* $\langle Action|Intention\rangle$ of corresponding to the globally–smooth transition map (2.827). This absolute square of this probability amplitude gives the *transition probability* of occurring the final state of *Action* given the initial state of $Intention$,

$$P(Action|Intention) = |\langle Action|Intention\rangle|^2.$$

The total transition amplitude from the state of *Intention* to the state of *Action* is defined on LSF_{total}

$$\mathcal{TA} \equiv \langle Action|Intention\rangle_{total} : INTENTION_{t_0} \Rrightarrow ACTION_{t_1}\,, \quad (2.829)$$

given by adaptive generalization of the Feynman's path integral [FH65, Fey72, Fey98]. The transition map (2.829) calculates the *overall probability amplitude* along a multitude of wildly fluctuating paths, fields and geometries, performing the *microscopic* transition from the micro–state $INTENTION_{t_0}$ occurring at initial micro–time instant t_0 to the micro–state $ACTION_{t_1}$ at some later micro–time instant t_1, such that all micro–time instants fit inside the global transition interval $t_0, t_1, ..., t_s \in [t_{ini}, t_{fin}]$. It is symbolically written as

$$\langle Action|Intention\rangle_{total} := \oint \mathcal{D}[w\Phi]\, \mathrm{e}^{\mathrm{i}S[\Phi]}, \quad (2.830)$$

where the Lebesgue integration is performed over all continuous $\Phi^i_{con} = paths + field + geometries$, while summation is performed over all discrete processes and regional topologies Φ^j_{dis}). The symbolic differential $\mathcal{D}[w\Phi]$ in the general path integral (2.830), represents an *adaptive path measure*, defined as a weighted product

$$\mathcal{D}[w\Phi] = \lim_{N\longrightarrow\infty} \prod_{s=1}^{N} w_s d\Phi^i_s, \qquad (i = 1, ..., n = con + dis), \quad (2.831)$$

which is in practice satisfied with a large N corresponding to infinitesimal temporal division of the four motivational phases (*). Technically, the path integral (2.830) calculates the *amplitude* for the transition functor $\mathcal{TA} : Intention \Rrightarrow Action$.

In the exponent of the path integral (2.830) we have the action $S[\Phi]$ and the imaginary unit $i = \sqrt{-1}$ (i can be converted into the real number -1 using the so–called *Wick rotation*, see next subsection).

In this way, we get a range of micro–objects in the local LSF at the short time–level: ensembles of rapidly fluctuating, noisy and crossing paths, force–fields, local geometries with obstacles and topologies with holes. However, by averaging process, both in time and along ensembles of paths, fields and geometries, we recover the corresponding global MD & CD variables.

Infinite–Dimensional Neural Network. The adaptive path integral (2.830) incorporates the *local learning process* according to the standard formula: $New\ Value = Old\ Value + Innovation$. The general *weights* $w_s = w_s(t)$ in (2.831) are updated by the $MONITOR$ feedback during the transition process, according to one of the two standard neural learning schemes, in which the micro–time level is traversed in discrete steps, i.e., if $t = t_0, t_1, ..., t_s$ then $t + 1 = t_1, t_2, ..., t_{s+1}$:

1. A *self–organized, unsupervised* (e.g., Hebbian–like [Heb49]) learning rule:

$$w_s(t+1) = w_s(t) + \frac{\sigma}{\eta}(w_s^d(t) - w_s^a(t)), \tag{2.832}$$

where $\sigma = \sigma(t)$, $\eta = \eta(t)$ denote *signal* and *noise*, respectively, while superscripts d and a denote *desired* and *achieved* micro–states, respectively; or

2. A certain form of a *supervised gradient descent learning*:

$$w_s(t+1) \;=\; w_s(t) - \eta \nabla J(t), \tag{2.833}$$

where η is a small constant, called the *step size*, or the *learning rate* and $\nabla J(n)$ denotes the gradient of the 'performance hyper–surface' at the $t-$th iteration.

Both Hebbian and supervised learning are used for the local decision making process (see below) occurring at the intention formation faze $\mathcal{F}$.

In this way, local micro–level of LSF_{total} represents an infinite–dimensional neural network. In the cognitive psychology framework, our adaptive path integral (2.830) can be interpreted as *semantic integration* (see [BF71, Ash94]).

Motion and Decision Making in LSF_{paths}

On the macro–level in the subspace LSF_{paths} we have the (loco)*motion action principle*

$$\delta S[x] = 0,$$

with the *Newtonian–like action* $S[x]$ given by

$$S[x] = \int_{t_{ini}}^{t_{fin}} dt\, [\frac{1}{2} g_{ij}\, \dot{x}^i \dot{x}^j + \varphi^i(x^i)], \tag{2.834}$$

where overdot denotes time derivative, so that $\dot{x}^i$ represents *processing speed*, or (loco)motion velocity vector. The first bracket term in (2.834) represents the kinetic energy T,

$$T = \frac{1}{2} g_{ij}\, \dot{x}^i \dot{x}^j,$$

generated by the *Riemannian metric tensor* g_{ij}, while the second bracket term, $\varphi^i(x^i)$, denotes the family of potential force–fields, driving the (loco)motions $x^i = x^i(t)$ (the *strengths* of the fields $\varphi^i(x^i)$ depend on their positions x^i in LSF, see LSF_{fields} below). The corresponding Euler–Lagrangian equation gives the Newtonian–like equation of motion

$$\frac{d}{dt} T_{\dot{x}^i} - T_{x^i} = -\varphi^i_{x^i}, \tag{2.835}$$

(subscripts denote the partial derivatives), which can be put into the standard Lagrangian form

$$\frac{d}{dt}L_{\dot{x}^i} = L_{x^i}, \qquad \text{with} \qquad L = T - \varphi^i(x^i).$$

In the next subsection we use the micro–level implications of the action S[x] as given by (2.834), for dynamical descriptions of the local decision–making process.

On the micro–level in the subspace LSF_{paths}, instead of a single path defined by the Newtonian–like equation of motion (2.835), we have an ensemble of fluctuating and crossing paths with weighted probabilities (of the unit total sum). This ensemble of micro–paths is defined by the simplest instance of our adaptive path integral (2.830), similar to the Feynman's original *sum over histories*,

$$\langle Action|Intention\rangle_{paths} = \oint\!\!\!\!\!\!\int \mathcal{D}[wx]\, \mathrm{e}^{\mathrm{i}S[x]}, \tag{2.836}$$

where $\mathcal{D}[wx]$ is a functional measure on the *space of all weighted paths*, and the exponential depends on the action $S[x]$ given by (2.834). This procedure can be redefined in a mathematically cleaner way if we Wick–rotate the time variable t to imaginary values $t \mapsto \tau = it$, thereby making all integrals real:

$$\oint\!\!\!\!\!\!\int \mathcal{D}[wx]\, \mathrm{e}^{\mathrm{i}S[x]} \xrightarrow{Wick} \oint\!\!\!\!\!\!\int \mathcal{D}[wx]\, \mathrm{e}^{-S[x]}. \tag{2.837}$$

Discretization of (2.837) gives the *thermodynamic–like partition function*

$$Z = \sum_j \mathrm{e}^{-w_j E^j/T}, \tag{2.838}$$

where E^j is the motion energy eigenvalue (reflecting each possible motivational energetic state), T is the temperature–like environmental control parameter, and the sum runs over all motion energy eigenstates (labelled by the index j). From (2.838), we can further calculate all thermodynamic–like and statistical properties of MD & CD (see e.g., [Fey72]), as for example, *transition entropy* $S = k_B \ln Z$, etc.

From cognitive perspective, our adaptive path integral (2.836) calculates all (alternative) pathways of information flow during the transition $Intention \longrightarrow Action$.

In the language of transition–propagators, the integral over histories (2.836) can be decomposed into the product of propagators (i.e., Fredholm kernels or Green functions) corresponding to the cascade of the four motivational phases (*)

$$\langle Action|Intention\rangle_{paths} = \oint\!\!\!\!\!\!\int dx^{\mathcal{F}} dx^{\mathcal{I}} dx^{\mathcal{M}} dx^{\mathcal{T}} K(\mathcal{F},\mathcal{I}) K(\mathcal{I},\mathcal{M}) K(\mathcal{M},\mathcal{T}), \tag{2.839}$$

satisfying the Schrödinger–like equation (see e.g., [Dir82])

$$i\,\partial_t \langle Action|Intention\rangle_{paths} = H_{Action}\, \langle Action|Intention\rangle_{paths}, \tag{2.840}$$

where H_{Action} represents the Hamiltonian (total energy) function available at the state of *Action*. Here our 'golden rule' is: the higher the H_{Action}, the lower the microscopic fatigue.

In the connectionist language, our propagator expressions (2.839–2.840) represent *activation dynamics*, to which our $Monitor$ process gives a kind of *backpropagation* feedback, a version of the basic supervised learning (2.833).

Mechanisms of Decision–Making under Uncertainty. The basic question about our local decision making process, occurring under uncertainty at the intention formation faze $\mathcal{F}$, is: Which alternative to choose? (see [RBT01, Gro82, Gro99, Gro88, Ash94]). In our path–integral language this reads: Which path (alternative) should be given the highest probability weight w? Naturally, this problem is iteratively solved by the learning process (2.832–2.833), controlled by the $MONITOR$ feedback, which we term *algorithmic approach*.

In addition, here we analyze qualitative mechanics of the local decision making process under uncertainty, as a *heuristic approach*. This qualitative analysis is based on the micro–level interpretation of the Newtonian–like action $S[x]$, given by (2.834) and figuring both processing speed $\dot{x}$ and LTM (i.e., the force–field $\varphi(x)$, see next subsection). Here we consider three different cases:

1. If the potential $\varphi(x)$ is not very dependent upon position $x(t)$, then the more direct paths contribute the most, as longer paths, with higher mean square velocities $[\dot{x}(t)]^2$ make the exponent more negative (after Wick rotation (2.837)).
2. On the other hand, suppose that $\varphi(x)$ does indeed depend on position x. For simplicity, let the potential increase for the larger values of x. Then a direct path does not necessarily give the largest contribution to the overall transition probability, because the integrated value of the potential is higher than over another paths.
3. Finally, consider a path that deviates widely from the direct path. Then $\varphi(x)$ decreases over that path, but at the same time the velocity $\dot{x}$ increases. In this case, we expect that the increased velocity $\dot{x}$ would more than compensate for the decreased potential over the path.

Therefore, the most important path (i.e., the path with the highest weight w) would be one for which any smaller integrated value of the surrounding field potential $\varphi(x)$ is more than compensated for by an increase in kinetic–like energy $\frac{m}{2}\dot{x}^2$. In principle, this is neither the most direct path, nor the longest path, but rather a middle way between the two. Formally, it is the path along which the average Lagrangian is minimal,

$$< \frac{m}{2}\dot{x}^2 + \varphi(x) > \longrightarrow \min, \tag{2.841}$$

i.e., the *path that requires minimal memory* (both LTM and WM, see LSF_{fields} below) and *processing speed*. This mechanical result is consistent with the 'filter theory' of *selective attention* [Bro58], proposed in an attempt to explain a

range of the existing experimental results. This theory postulates a low level filter that allows only a limited number of percepts to reach the brain at any time. In this theory, the importance of conscious, directed attention is minimized. The type of attention involving low level filtering corresponds to the concept of *early selection* [Bro58].

Although we termed this 'heuristic approach' in the sense that we can instantly feel both the processing speed $\dot{x}$ and the LTM field $\varphi(x)$ involved, there is clearly a psycho–physical rule in the background, namely the averaging minimum relation (2.841).

From the decision making point of view, all possible paths (alternatives) represent the *consequences* of decision making. They are, by default, *short–term consequences*, as they are modelled in the micro–time–level. However, the path integral formalism allows calculation of the *long–term consequences*, just by extending the integration time, $t_{fin} \longrightarrow \infty$. Besides, this *averaging decision mechanics* – choosing the optimal path – actually performs the 'averaging lift' in the LSF: from micro– to the macro–level.

Force–Fields and Memory in LSF_{fields}

At the macro–level in the subspace LSF_{fields} we formulate the *force–field action principle*

$$\delta S[\varphi] = 0, \tag{2.842}$$

with the action $S[\varphi]$ dependent on Lewinian force–fields $\varphi^i = \varphi^i(x)$ $(i = 1, ..., N)$, defined as a temporal integral

$$S[\varphi] = \int_{t_{ini}}^{t_{fin}} \mathfrak{L}[\varphi]\, dt, \tag{2.843}$$

with Lagrangian density given by

$$\mathfrak{L}[\varphi] = \int d^n x\, \mathcal{L}(\varphi_i, \partial_{x^j}\varphi^i),$$

where the integral is taken over all n coordinates $x^j = x^j(t)$ of the LSF, and $\partial_{x^j}\varphi^i$ are partial derivatives of the field variables over coordinates.

On the micro–level in the subspace LSF_{fields} we have the Feynman–type *sum over fields* φ^i $(i = 1, ..., N)$ given by the adaptive path integral

$$\langle Action|Intention\rangle_{fields} = \oint \mathcal{D}[w\varphi]\, \mathrm{e}^{\mathrm{i}S[\varphi]} \xrightarrow{Wick} \oint \mathcal{D}[w\varphi]\, \mathrm{e}^{-S[\varphi]}, \tag{2.844}$$

with action $S[\varphi]$ given by temporal integral (2.843). (Choosing special forms of the force–field action $S[\varphi]$ in (2.844) defines micro–level MD & CD, in the LSF_{fields} space, that is similar to standard quantum–field equations, see e.g., [Ram90].) The corresponding partition function has the form similar to (2.838), but with field energy levels.

Regarding topology of the force fields, we have in place *n−categorical Lagrangian–field structure* on the Riemannian LSF manifold M,

$$\Phi^i : [0,1] \to M,\ \Phi^i : \Phi^i_0 \mapsto \Phi^i_1,$$

generalized from the recursive homotopy dynamics (2.668) above, using

$$\frac{d}{dt} f_{\dot{x}^i} \ = \ f_{x^i} \longrightarrow \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial_\mu \Phi^i} \right) = \frac{\partial \mathcal{L}}{\partial \Phi^i},$$
$$\text{with} \quad [x_0, x_1] \longrightarrow [\Phi^i_0, \Phi^i_1].$$

Relationship between Memory and Force–Fields. As already mentioned, the subspace LSF_{fields} is related to our *memory storage* [Ash94]. Its global macro–level represents the *long–term memory* (LTM), defined by the least action principle (2.842), related to *cognitive economy* in the model of *semantic memory* [Rat78, CQ69]. Its local micro–level represents *working memory* (WM), a limited–capacity 'bottleneck' defined by the adaptive path integral (2.844). According to our formalism, each of Miller's 7 ± 2 units [Mil56] of the local WM are adaptively stored and averaged to give the global LTM capacity (similar to the physical notion of potential). This averaging memory lift, from WM to LTM represents *retroactive interference*, while the opposite direction, given by the path integral (2.844) itself, represents *proactive interference*. Both retroactive and proactive interferences are examples of the impact of cognitive contexts on memory. Motivational contexts can exert their influence, too. For example, a reduction in task–related recall following the completion of the task is one of the clearest examples of force–field influences on memory: the amount of details remembered of a task declines as the force–field tension to complete the task is reduced by actually completing it.

Once defined, the global LTM potential $\varphi = \varphi(x)$ is then affecting the locomotion transition paths through the path action principle (2.834), as well as general learning (2.832–2.833) and decision making process (2.841).

On the other hand, the two levels of LSF_{fields} fit nicely into the two levels of processing framework, as presented by [CL72], as an alternative to theories of separate stages for sensory, working and long–term memory. According to the *levels of processing framework*, stimulus information is processed at multiple levels simultaneously depending upon its characteristics. In this framework, our macro–level memory field, defined by the fields action principle (2.842), corresponds to the *shallow memory*, while our micro–level memory field, defined by the adaptive path integral (2.844), corresponds to the *deep memory*.

Geometries, Topologies and Noise in LSF_{geom}

On the macro–level in the subspace LSF_{geom} representing an $n-$dimensional smooth manifold M with the global Riemannian metric tensor g_{ij}, we formulate the *geometrical action principle*

$$\delta S[g_{ij}] = 0,$$

where $S = S[g_{ij}]$ is the $n-$dimensional *geodesic action* on M,

$$S[g_{ij}] = \int d^n x \sqrt{g_{ij}\, dx^i dx^j}. \tag{2.845}$$

The corresponding Euler–Lagrangian equation gives the *geodesic equation* of the *shortest path* in the manifold M,

$$\ddot{x}^i + \Gamma^i_{jk}\, \dot{x}^j\, \dot{x}^k = 0,$$

where the symbol Γ^i_{jk} denotes the so–called *affine connection* which is the source of *curvature*, which is geometrical description for *noise* (see [Ing97, Ing98]). The higher the local curvatures of the LSF–manifold M, the greater the noise in the life space. This noise is the source of our micro–level fluctuations. It can be internal or external; in both cases it curves our micro–LSF.

Otherwise, if instead we choose an $n-$dimensional Hilbert–like action (see [MTW73]),

$$S[g_{ij}] = \int d^n x \sqrt{\det |g_{ij}|}R, \tag{2.846}$$

where R is the scalar curvature (derived from Γ^i_{jk}), we get the $n-$dimensional Einstein–like equation: $G_{ij} = 8\pi T_{ij}$, where G_{ij} is the Einstein–like tensor representing geometry of the LSF manifold M (G_{ij} is the trace–reversed Ricci tensor R_{ij}, which is itself the trace of the *Riemann curvature tensor* of the manifold M), while T_{ij} is the $n-$dimensional *stress–energy–momentum* tensor. This equation explicitly states that *psycho–physics of the LSF is proportional to its geometry*. T_{ij} is important quantity, representing motivational *energy*, geometry–imposed *stress* and *momentum* of (loco)motion. As before, we have our 'golden rule': *the greater the $T_{ij}-$components, the higher the speed of cognitive processes and the lower the macroscopic fatigue.*

The choice between the geodesic action (2.845) and the Hilbert action (2.846) depends on our interpretation of time. If time is not included in the LSF manifold M (non–relativistic approach) then we choose the geodesic action. If time is included in the LSF manifold M (making it a relativistic–like $n-$dimensional space–time) then the Hilbert action is preferred. The first approach is more related to the information processing and the working memory. The later, space–time approach can be related to the long–term memory: we usually recall events closely associated with the times of their happening.

On the micro–level in the subspace LSF_{geom} we have the adaptive *sum over geometries*, represented by the path integral over all local (regional) Riemannian metrics $g_{ij} = g_{ij}(x)$ varying from point to point on M (modulo diffeomorphisms),

$$\langle Action|Intention\rangle_{geom} = \oint \mathcal{D}[wg_{ij}]\, \mathrm{e}^{\mathrm{i}S[g_{ij}]} \xrightarrow{Wick} \oint \mathcal{D}[wg_{ij}]\, \mathrm{e}^{-S[g_{ij}]}, \tag{2.847}$$

where $\mathcal{D}[g_{ij}]$ is diffeomorphism equivalence class of $g_{ij}(x) \in M$.

To include the topological structure (e.g., a number of holes) in M, we can extend (2.847) as

$$\langle Action|Intention\rangle_{geom/top} = \sum_{\text{topol.}} \oint\!\!\!\!\!\!\int \mathcal{D}[wg_{ij}]\, \mathrm{e}^{\mathrm{i}S[g_{ij}]}, \tag{2.848}$$

where the topological sum is taken over all connectedness–components of M determined by the *Euler characteristic* χ of M. This type of integral defines the *theory of fluctuating geometries*, a propagator between $(n-1)-$dimensional boundaries of the $n-$dimensional manifold M. One has to contribute a meaning to the integration over geometries. A key ingredient in doing so is to approximate (using simplicial approximation and Regge calculus [MTW73]) in a natural way the smooth structures of the manifold M by piecewise linear structures (mostly using topological simplices Δ). In this way, after the Wick–rotation (2.837), the integral (2.847–2.848) becomes a *simple statistical system*, given by partition function $Z = \sum_{\Delta} \frac{1}{C_{\Delta}} e^{-S_{\Delta}}$, where the summation is over all triangulations Δ of the manifold M, while C_T is the order of the automorphism group of the performed triangulation.

Micro–Level Geometry: the source of noise and stress in LSF. The subspace LSF_{geom} is the source of noise, fluctuations and obstacles, as well as psycho–physical stress. Its micro–level is adaptive, reflecting the human ability to efficiently act within the noisy environment and under the stress conditions. By averaging it produces smooth geometry of certain curvature, which is at the same time the smooth psycho–physics. This macro–level geometry directly affects the memory fields and indirectly affects the (loco)motion transition paths.

The Mental Force Law. As an effective summary of this section, we state that the psychodynamic transition functor $\mathcal{TA} : INTENTION_{t_{ini}} \Rrightarrow ACTION_{t_{fin}}$, defined by the generic path integral (2.830), can be interpreted as a *mental force law*, analogous to our musculo–skeletal *covariant force law*, $F_i = mg_{ij}a^j$, and its associated *covariant force functor* $\mathcal{F}_* : TT^*M \to TTM$ [II05].

2.6.2 Geometrical Dynamics of Human Crowd

In this subsection we formulate crowd representation model as an emotion–field induced collective behavior of individual autonomous agents [IS00].

It is well–known that *crowd behavior* is more influenced by *collective emotion* than by cognition. Recall from previous subsection that according to *Lewinian psychodynamics*, human behavior is largely determined by underlying *forces* (needs). For him, a force–field is defined as 'the totality of coexisting motivational facts which are conceived of as mutually interdependent' [Lew97]. He also stresses psychological *direction* and *velocity* of behavior.

On the other hand, a number of factor–analysis based studies show that human emotion is not a single quantity, but rather a *multidimensional space*.

For example, in [SFM98], authors assessed emotions with single adjective descriptors using standard linear factor analysis, by examining semantic as well as cognitive, motivational, and intensity features of emotions. The focus was on seven negative emotions common to several emotion typologies: anger, fear, sadness, shame, pity, jealousy, and contempt. For each of these emotions, seven items were generated corresponding to cognitive appraisal about the self, cognitive appraisal about the environment, action tendency, action fantasy, synonym, antonym, and intensity range of the emotion, respectively. These findings set the groundwork for the construction of an instrument to assess emotions multicomponentially.

Crowd Hypothesis

We consider a *human crowd* $\mathcal{C}$ as a group of m *autonomous agents* $\mathcal{A}_i$ $(i = 1, ..., m)$, each of which carries its own nD *motivational factor–structure* (compare with the standard BDI–agents, as described in subsection 2.6.8 below). This nonlinear factor structure, which can be get using modern *nonlinear factor analysis* techniques (see [YA01, Ame93, WA98, WA00]; also compare with subsection 2.6.9 below), is defined by n hypothetical motivational factor–coordinates $\mathbf{q}_i = \{q_i^\mu\}$, $(\mu = 1, \ldots, n)$, spanning the smooth nD *motivational factor manifold* M_i for each autonomous agent $\mathcal{A}_i$.

We understand crowd representation as an environmental field–induced collective behavior of individual autonomous agents. To model it in a general geometrodynamical framework, we firstly define the *behavior* of each agent $\mathcal{A}_i$ as a *motion* $\boldsymbol{\pi}_i$ along his motivational manifold M_i, caused by his own *emotion field* Φ_i, which is an active (motor) subset of M_i.

Secondly, we formulate a collective geometrodynamical model for the crowd, considered as a union $\mathcal{C} = \cup_i \mathcal{A}_i$,in the form of a divergence equation for the total crowd's SEM–tensor $\mathbf{C}$.[40]

Geometrodynamics of Individual Agents

To formulate individual agents' geometrodynamics, we firstly derive two higher geometrical structures from a motivational factor manifold M_i corresponding to an agent $\mathcal{A}_i$: (i) the *agent's velocity phase–space*, defined as a tangent bundle TM_i, and (ii) the *agent's momentum phase–space*, defined as a cotangent bundle T^*M_i.

Now, the sections of TM_i we call the agent's vector–fields $\mathbf{v}_i$, which can be expanded in terms of the basis vector–fields $\{e_\mu^i \equiv \partial_{q_i^\mu}\}$, as $\mathbf{v}_i = v_i^\mu e_\mu^i$. Similarly, the sections of T^*M_i we call the agent's one–forms $\boldsymbol{\alpha}_i$, which can be expanded in terms of the basis one–forms $\{\omega_i^\mu \equiv dq_i^\mu\}$, as $\boldsymbol{\alpha}_i = \alpha_\mu^i \, \omega_i^\mu$.

[40] Throughout the text we use the following index convention: we label individual agents using Latin indices, and individual motivational factors using Greek indices; summation convention is applied only to Greek factor indices.

Here d denotes the *exterior derivative* (such that $dd = 0$). In particular, the *velocity* vector–fields $\dot{\mathbf{q}}_i$ are defined on TM_i as $\dot{\mathbf{q}}_i = \dot{q}_i^\mu e_\mu^i$,while the *momentum* one–forms $\boldsymbol{\pi}_i$ are defined on T^*M_i as $\boldsymbol{\pi}_i = \pi_\mu^i\, \omega_i^\mu$.

Also, all factor–configuration manifolds M_i are assumed to be Riemannian, admitting the metrics $\mathbf{g}_i = \langle \omega_i^\mu, \omega_i^\eta \rangle$, determined by *kinetic energies* $T_i = \langle \dot{q}_i^\mu, \dot{q}_i^\eta \rangle$ of individual agents $\mathcal{A}_i$, each with the metric tensor $\mathbf{g}^i = g^i_{\mu\eta}\, \omega_i^\mu \otimes \omega_i^\eta$. This implies that all vector–fields $\mathbf{v}_i$ and one–forms $\boldsymbol{\alpha}_i$ are related by $\mathbf{g}-$induced scalar products $\left\langle \omega_i^\mu, e_\eta^i \right\rangle = \delta_\eta^{i\mu}$.

Emotional/Environmental Fields and Induced Agents' Motions

Now, for each autonomous agent $\mathcal{A}_i$ three additional geometrodynamical objects are defined as:

1. *Emotional/environmental potential* one–form $\boldsymbol{\alpha}_i = \alpha_\mu^i\, \omega_i^\mu$, which is the gradient of some scalar function $\mathbf{f}_i = f_i^\mu(q_i^\mu)$ on M_i,

$$\boldsymbol{\alpha}_i = d\mathbf{f}_i, \qquad \text{in components,} \qquad \alpha_\mu^i = \partial_{q_i^\mu} f_i^\mu \omega_i^\mu.$$

2. *Psycho–physical current* vector–field $\mathbf{J}_i = J_i^\mu\, e_\mu^i$ on M_i, defined through its *motivational charge* e_i as

$$J_i^\mu = e_i \int_{t_i} \dot{q}_i^\mu\, \delta^n \left[q_i^\mu(t_i)\right]\, dt_i,$$

   where $\delta^n = \delta^n \left[q_i^\mu(t_i)\right]$ denotes the nD impulse delta–function defined on M_i.
3. *Emotional/environmental psycho–physical field* is a two–form

$$\boldsymbol{\Phi}_i = \Phi^i_{\mu\eta}\, \omega_i^\mu \otimes \omega_i^\eta = \frac{1}{2} \Phi^i_{\mu\eta}\, \omega_i^\mu \wedge \omega_i^\eta$$

   on M_i, defined as the exterior derivative (i.e., curl) of the emotional/environmental potential $\boldsymbol{\alpha}_i$,

$$\boldsymbol{\Phi}_i = d\boldsymbol{\alpha}_i, \qquad \text{in components,} \qquad \Phi^i_{\mu\eta} = \alpha^i_{\eta;\mu} - \alpha^i_{\mu;\eta}.$$

The emotional/environmental psycho–physical field $\boldsymbol{\Phi}_i$ is governed by two standard field equations:

$$d\boldsymbol{\Phi}_i = dd\boldsymbol{\alpha}_i = 0, \qquad \text{in components,} \qquad \Phi^i_{[\mu\eta;\nu]} = 0, \tag{2.849}$$

and

$$\operatorname{div} \boldsymbol{\Phi}_i = \mathbf{g}^i \mathbf{J}_i, \qquad \text{in components,} \qquad \Phi^{i;\eta}_{\mu\eta} = g^i_{\mu\eta}\, J_i^\eta, \tag{2.850}$$

where $[\mu\eta;\nu] \equiv \mu\eta;\nu + \eta\nu;\mu + \upsilon\mu;\eta$. The first field equation (2.849) states that for each agent $\mathcal{A}_i$ the motivational tension–field $\boldsymbol{\Phi}_i$ is *curl–free*, while the second field equation (2.850) states that the environmental psycho–physical field $\boldsymbol{\Phi}_i$ has its *source* in the psycho–physical current $\mathbf{J}_i$.

Now, the *behavioral equation* for each agent $\mathcal{A}_i$, induced by his/her environmental psycho–physical field $\mathbf{\Phi}_i$, reads

$$\dot{\boldsymbol{\pi}}_i = e_i \mathbf{\Phi}_i \dot{\mathbf{q}}_i, \qquad \text{in components,} \qquad \dot{\pi}^i_\mu = e_i \Phi^i_{\mu\eta}\, \dot{q}^\eta_i.$$

This equation states that the force of an individual agent's motion, or his/her behavior, equals the product of his/her psycho–physical charge, environmental psycho–physical field, and velocity (speed) of behavior.

Collective Crowd Geometrodynamics

Now we define the total crowd's geometrodynamics as a union $\mathcal{C} = \cup_i \mathcal{A}_i$ of all individual agents' geometrodynamics, to model their emotion–induced behavior. For this we use the total *crowd's energy–momentum tensor* (CEM) and its divergence equation of motion.

As a union of individual Riemannian manifolds, the total crowd's manifold $\mathcal{M} = \cup_i M_i$ is also Riemannian, with the metrics $\mathfrak{g} = \sum_i \mathbf{g}^i$ equal to the sum of individual metrics $\mathbf{g}^i = g^i_{\mu\eta}\, \omega^\mu_i \otimes \omega^\eta_i$. Now, the crowd's CEM tensor $\mathbf{C} = C_{\mu\eta}\, \omega^\mu_{\mathcal{M}} \otimes \omega^\eta_{\mathcal{M}} = \frac{1}{2} C_{\mu\eta}\, \omega^\mu_{\mathcal{M}} \wedge \omega^\eta_{\mathcal{M}}$ (where $\omega^\mu_{\mathcal{M}}$ denote the basis one–forms on $\mathcal{M}$) has two parts, $\mathbf{C} = \mathbf{C}^{(E)} + \mathbf{C}^{(B)}$, in components, $C_{\mu\eta} = C^{(E)}_{\mu\eta} + C^{(B)}_{\mu\eta}$, corresponding to the total crowd's emotion and behavior, which we define respectively as follows:

1. The CEM's *emotional part* $\mathbf{C}^{(E)}$ is in components defined as a sum of individual agents' motivational–tension fields,

$$C^{(E)}_{\mu\eta} = \sum_{i=1}^{M} \left(\Phi^i_{\mu\nu} \Phi^\nu_{i\eta} - g^i_{\mu\eta} \Phi^i_{\nu\lambda} \Phi^{\nu\lambda}_i \right),$$

   so the equation of the crowd's emotion is defined in the form of the divergence equation

$$C^{(E);\eta}_{\mu\eta} = \sum_{i=1}^{M} \Phi^i_{\mu\eta} \Phi^{\eta;\nu}_{i\nu} = -\sum_{i=1}^{M} \Phi^i_{\mu\eta}\, J^\eta_i.$$

   This equation says that CEM's *emotional part* $\mathbf{C}^{(E)}$ is a collective motivational field with a *sink*.
2. The CEM's *behavioral part* $\mathbf{C}^{(B)}$ as a sum of individual agents' motivational–currents times momenta:

$$C^{(B)}_{\mu\eta} = \sum_{i=1}^{M} g^i_{\mu\eta} \int_{t_i} \pi^i_\eta(t_i)\, \dot{q}^\eta_i(t_i)\, \delta^n \left[q^\mu_i(t_i) \right] dt_i,$$

   so the equation of the crowd's behavior is defined in the form of the divergence equation

$$C^{(B);\eta}_{\mu\eta} = \sum_{i=1}^{M} g^i_{\mu\eta} \int_{t_i} \pi^i_\eta(t_i)\, \dot{q}^\eta_i(t_i)\, \partial_{q^\eta_i}\, \delta^n\left[q^\mu_i(t_i)\right]\, dt_i$$
$$= \sum_{i=1}^{M} e_i \int_{t_i} \Phi^i_{\mu\eta}\, \dot{q}^\eta_i\, \delta^n\left[q^\mu_i(t_i)\right]\, dt_i = \sum_{i=1}^{M} \Phi^i_{\mu\eta}\, J^\eta_i.$$

This equation says that CEM's *behavioral part* $\mathbf{C}^{(B)}$ is a collective behavioral field with a *source*.

Therefore, the divergence equation for the total crowd's CEM tensor, represents the crowd's *motivation–behavior conservation law*

$$\operatorname{div} \mathbf{C} = \operatorname{div}\left(\mathbf{C}^{(E)} + \mathbf{C}^{(B)}\right) = 0, \qquad \text{in components,} \qquad C^{(E);\eta}_{\mu\eta} + C^{(B);\eta}_{\mu\eta} = 0.$$

This gives our basic representation of an isolated crowd as a conservative spatio–temporal dynamical system. Naturally, additional crowd's energy–momentum sources and sinks can violate this basic motivational–behavior conservation law.

2.6.3 Dynamical Games on Lie Groups

In this section we propose a general approach to modelling *conflict resolution manoeuvres* for land, sea and airborne *vehicles*, using dynamical games on Lie groups. We use the generic name 'vehicle' to represent all planar vehicles, namely land and sea vehicles, as well as fixed altitude motion of aircrafts (see, e.g., [LGS98, TPS98]). First, we elaborate on the two–vehicle conflict resolution manoeuvres, and after that discuss the multi–vehicle manoeuvres.

We explore special features of the dynamical games solution when the underlying dynamics correspond to left–invariant control systems on Lie groups. We show that the 2D (i.e., planar) motion of a vehicle may be modelled as a control system on the Lie group $SE(2)$. The proposed algorithm surrounds each vehicle with a circular protected zone, while the simplification in the derivation of *saddle* and *Nash strategies* follows from the use of symplectic reduction techniques [MR99]. To model the two–vehicle conflict resolution, we construct the safe subset of the state–space for one of the vehicles using zero–sum non–cooperative dynamic game theory [BO95] which we specialize to the $SE(2)$ group. If the underlying continuous dynamics are left–invariant control systems, reduction techniques can be used in the computation of safe sets.

Configuration Models for Planar Vehicles

The configuration of each individual vehicle is described by an element of the *Lie group* $SE(2)$ of rigid–body motions in $\mathbb{R}^2$. Let $g_i \in SE(2)$ denote the

configurations of vehicles labelled i, with $i = 1, 2$. The motion of each vehicle may be modelled as a left–invariant vector–field on $SE(2)$:

$$\dot{g}_i = g_i X_i, \tag{2.851}$$

where the vector–fields X_i belong to the vector space $\mathfrak{se}(2)$, the *Lie algebra* associated with the group $SE(2)$.

Each $g_i \in SE(2)$ can be represented in standard local coordinates (x_i, y_i, θ_i) as

$$g_i = \begin{bmatrix} \cos\theta_i & -\sin\theta_i & x_i \\ \sin\theta_i & \cos\theta_i & y_i \\ 0 & 0 & 1 \end{bmatrix},$$

where x_i, y_i is the position of vehicle i and θ_i is its orientation, or heading. The associated Lie algebra is $\mathfrak{se}(2)$, with $X_i \in \mathfrak{se}(2)$ represented as

$$X_i = \begin{bmatrix} 0 & -\omega_i & v_i \\ \omega_i & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

where v_i and ω_i represent the *translational* (linear) and *rotational* (angular) *velocities*, respectively.

Now, to determine *dynamics of the relative configuration* of two vehicles, we perform a change (transformation) of coordinates, to place the identity element of the group $SE(2)$ on vehicle 1. If $g^{rel} \in SE(2)$ denotes the relative configuration of vehicle 2 with respect to vehicle 1, then

$$g_2 = g_1 g^{rel} \Longrightarrow g^{rel} = g_1^{-1} g_2.$$

Differentiation with respect to time yields the dynamics of the relative configuration:

$$\dot{g}^{rel} = g^{rel} X_2 - X_1 g^{rel},$$

which expands into:

$$\begin{aligned} \dot{x}^{rel} &= -v_1 + v_2 \cos\theta^{rel} + \omega_1 y^{rel}, \\ \dot{y}^{rel} &= v_2 \sin\theta^{rel} - \omega_1 x^{rel}, \\ \dot{\theta}^{rel} &= \omega_2 - \omega_1. \end{aligned}$$

Two–Vehicles Conflict Resolution Manoeuvres

Next, we seek *control strategies* for each vehicle, which are *safe* under (possible) *uncertainty* in the actions of neighbouring vehicle. For this, we expand the dynamics of two vehicles (2.851),

$$\dot{g}_1 = g_1 X_1, \qquad \dot{g}_2 = g_2 X_2,$$

and write it in the matrix form as

$$\dot{g} = gX, \tag{2.852}$$

with

$$g = \begin{bmatrix} g_1 & 0 \\ 0 & g_2 \end{bmatrix}, \qquad X = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix},$$

in which g is an element in the configuration manifold $M = SE(2) \times SE(2)$, while the vector–fields $X_i \in \mathfrak{se}(2) \times \mathfrak{se}(2)$ are linearly parameterised by velocity inputs $(\omega_1, v_1) \in \mathbb{R}^2$ and $(\omega_2, v_2) \in \mathbb{R}^2$.

The goal of each vehicle is to maintain *safe operation*, meaning that

(i) the vehicles remain outside of a specified *target set* T with boundary ∂T, defined by

$$T = \{g \in M | l(g) < 0\},$$

where $l(g)$ is a differentiable circular function,

$$l(g) = (x_2 - x_1)^2 + (y_2 - y_1)^2 - \rho^2$$

(with ρ denoting the *radius of a circular protected zone*) defines the minimum allowable lateral separation between vehicles; and

(ii)

$$dl(g) \neq 0 \quad \text{on} \quad \partial T = \{g \in M | l(g) = 0\},$$

where d represents the *exterior derivative* (a unique generalization of the gradient, divergence and curl).

Now, due to possible uncertainty in the actions of vehicle 2, the safest possible strategy of vehicle 1 is to drive along a *trajectory* which guarantees that the minimum allowable separation with vehicle 2 is maintained regardless of the actions of vehicle 2. We therefore formulate this problem as a *zero–sum dynamical game* with two players: control vs. disturbance. The *control* is the action of vehicle 1,

$$u = (\omega_1, v_1) \in U,$$

and the *disturbance* is the action of vehicle 2,

$$d = (\omega_2, v_2) \in D.$$

Here the control and disturbance sets, U and D, are defined as

$$U = ([\omega_1^{\min}, \omega_1^{\max}], [v_1^{\min}, v_1^{\max}]),$$
$$D = ([\omega_2^{\min}, \omega_2^{\max}], [v_2^{\min}, v_2^{\max}])$$

and the corresponding control and disturbance functional spaces, $\mathcal{U}$ and $\mathcal{D}$ are defined as:

$$\mathcal{U} = \{u(\cdot) \in PC^0(\mathbb{R}^2) | u(t) \in U, t \in \mathbb{R}\},$$
$$\mathcal{D} = \{d(\cdot) \in PC^0(\mathbb{R}^2) | d(t) \in U, t \in \mathbb{R}\},$$

where $PC^0(\mathbb{R}^2)$ is the space of piecewise continuous functions over $\mathbb{R}^2$.

We define the *cost of a trajectory* $g(t)$ which starts at state g at initial time $t \leq 0$, evolves according to (2.852) with input $(u(\cdot), d(\cdot))$, and ends at the final state $g(0)$ as:

$$\begin{aligned} & J(g, u(\cdot), d(\cdot), t) \ : \ SE(2) \times SE(2) \times \mathcal{U} \times \mathcal{D} \times \mathbb{R}_- \to \mathbb{R}, \\ \text{such that} \ \ & J(g, u(\cdot), d(\cdot), t) = l(g(0)), \end{aligned} \tag{2.853}$$

where 0 is the final time (without loss of generality). Thus the cost depends only on the final state $g(0)$ (the Lagrangian, or running cost, is identically zero). The game is won by vehicle 1 if the terminal state $g(0)$ is either outside T or on ∂T (i.e., $J(g, 0) \geq 0$), and is won by vehicle 2 otherwise.

This two–player zero–sum dynamical game on $SE(2)$ is defined as follows. Consider the matrix system (2.852), $\dot{g} = gX$, over the time interval $[t, 0]$ where $t < 0$ with the cost function $J(g, u(\cdot), d(\cdot), t)$ defined by (2.853) As vehicle 1 attempts to maximize this cost assuming that vehicle 2 is acting blindly, the optimal control action and worst disturbance actions are calculated as

$$u^* = \arg \max_{u \in \mathcal{U}} \min_{d \in \mathcal{D}} J(g, u(\cdot), d(\cdot), t), \qquad d^* = \arg \min_{d \in \mathcal{D}} \max_{u \in \mathcal{U}} J(g, u(\cdot), d(\cdot), t).$$

The game is said to have a *saddle solution* (u^*,d^*) if the resulting optimal cost $J^*(g, t)$ does not depend on the order of play, i.e., on the order in which the maximization and minimization is performed:

$$J^*(g, t) = \max_{u \in \mathcal{U}} \min_{d \in \mathcal{D}} J(g, u(\cdot), d(\cdot), t) = \min_{d \in \mathcal{D}} \max_{u \in \mathcal{U}} J(g, u(\cdot), d(\cdot), t).$$

Using this saddle solution we calculate the 'losing states' for vehicle 1, called the *predecessor* $Pre_t(T)$ of the target set T,

$$Pre_t(T) = \{g \in M | J(g, u^*(\cdot), d(\cdot), t) < 0\}.$$

Symplectic Reduction and Dynamical Games on $SE(2)$

Since vehicles 1 and 2 have dynamics given by left–invariant control systems on the Lie group $SE(2)$, we have

$$X_1 = \xi^1 \omega_1 + \xi^2 v_1, \qquad X_2 = \xi^1 \omega_2 + \xi^2 v_2,$$

with ξ^1, ξ^2 being two of the three basis elements for the tangent Lie algebra $\mathfrak{se}(2)$ given by

$$\xi^1 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \xi^2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \xi^3 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

If p_1 (resp. p_2) is a cotangent vector–field to $SE(2)$ at g_1 (resp. g_2), belonging to the cotangent (dual) Lie algebra $\mathfrak{se}(2)^*$, we can define the *momentum functions* for both vehicles:

$$P_1^1 = < p_1, g_1\xi^1 >, P_1^2 =< p_1, g_1\xi^2 >, P_1^3 =< p_1, g_1\xi^3 >,$$
$$P_2^1 = < p_2, g_2\xi^1 >, P_2^2 =< p_2, g_2\xi^2 >, P_2^3 =< p_2, g_2\xi^3 >,$$

which can be compactly written as

$$P_i^j =< p_i, g_i\xi^j > .$$

Defining $p = (p_1, p_2) \in \mathfrak{se}(2)^* \times \mathfrak{se}(2)^*$, the *optimal cost* for the two-player, zero-sum dynamical game is given by

$$J^*(g,t) = \max_{u\in\mathcal{U}} \min_{d\in\mathcal{D}} J(g, u(\cdot), d(\cdot), t) = \max_{u\in\mathcal{U}} \min_{d\in\mathcal{D}} l(g(0)).$$

The *Hamiltonian* $H(g, p, u, d)$ is given by

$$H(g,p,u,d) = P_1^1\omega_1 + P_1^2 v_1 + P_2^1\omega_1 + P_2^2 v_1$$

for control and disturbance inputs $(\omega_1, v_1) \in U$ and $(\omega_2, v_2) \in D$ as defined above. It follows that the *optimal Hamiltonian* $H^*(g,p)$, defined on the cotangent bundle $T^*SE(2)$, is given by

$$\begin{aligned} H^*(g,p) = {} & P_1^1 \frac{\omega_1^{\max} + \omega_1^{\min}}{2} + P_2^1 \frac{\omega_2^{\max} + \omega_2^{\min}}{2} + |P_1^1| \frac{\omega_1^{\max} - \omega_1^{\min}}{2} \\ & - |P_1^1| \frac{\omega_2^{\max} - \omega_2^{\min}}{2} + P_1^2 \frac{v_1^{\max} + v_1^{\min}}{2} + P_2^2 \frac{v_2^{\max} + v_2^{\min}}{2} \\ & + |P_1^2| \frac{v_1^{\max} - v_1^{\min}}{2} - |P_1^2| \frac{v_2^{\max} - v_2^{\min}}{2} \end{aligned}$$

and the *saddle solution* (u^*, d^*) is given by

$$u^* = \arg\max_{u\in\mathcal{U}} \min_{d\in\mathcal{D}} H(g,p,u,d), \qquad d^* = \arg\min_{d\in\mathcal{D}} \max_{u\in\mathcal{U}} H(g,p,u,d). \quad (2.854)$$

Note that $H(g,p,u,d)$ and $H^*(g,p)$ do not depend on the state g and costate p directly, rather through the momentum functions P_1^j, P_2^j. This is because the dynamics are determined by left–invariant vector fields on the Lie group and the Lagrangian is state independent [MR99].

The optimal Hamiltonian $H^*(g,p)$ determines a 12D Hamiltonian vector–field X_{H^*} on the symplectic manifold $T^*M = SE(2) \times SE(2) \times \mathfrak{se}(2)^* \times \mathfrak{se}(2)^*$ (which is the cotangent bundle of the configuration manifold M), defined by Hamiltonian equations

$$X_{H^*} : \dot{g} = \frac{\partial H^*(g,p)}{\partial p}, \qquad \dot{p} = -\frac{\partial H^*(g,p)}{\partial g},$$

with initial condition at time t being $g(t) = g$ and final condition at time 0 being $p(0) = dl(g(0))$. In general, to solve for the saddle solution (2.854), one needs to solve the ODE system for all states. However since the original system on $M = SE(2) \times SE(2)$ is left–invariant, it induces generic symmetries

in the Hamiltonian dynamics on T^*M , referred to as Marsden–Weinstein reduction of Hamiltonian systems on symplectic manifolds, see [MR99]. In general for such systems one only needs to solve an ODE system with half of the dimensions of the underlying symplectic manifold.

For the two-vehicle case we only need to solve an ODE system with 6 states. That is exactly given by the dynamics of the 6 momentum functions

$$\dot{P}_i^j = L_{X_{H^*}} P_i^j = \{P_i^j, H^*(g,p)\}, \tag{2.855}$$

for $i,j = 1,2$, which is the *Lie derivative* of P_i^j with respect to the Hamiltonian vector–field X_{H^*}. In the equation (2.855), the bracket $\{\cdot,\cdot\}$ is the *Poisson bracket* [IP01a], giving the commutation relations:

$$\begin{array}{lll} \{P_1^1, P_1^2\} = P_1^3, & \{P_1^2, P_1^3\} = 0, & \{P_1^3, P_1^1\} = P_1^2, \\ \{P_2^1, P_2^2\} = P_2^3, & \{P_2^2, P_2^3\} = 0, & \{P_2^3, P_2^1\} = P_2^2. \end{array}$$

Using these commutation relations, equation (2.855) can be written explicitly:

$$\begin{aligned}
\dot{P}_1^1 &= P_1^3 \left(\frac{v_1^{\max} + v_1^{\min}}{2} + sign(P_1^2) \frac{v_1^{\max} + v_1^{\min}}{2} \right), \\
\dot{P}_1^2 &= P_1^3 \left(-\frac{\omega_1^{\max} + \omega_1^{\min}}{2} - sign(P_1^1) \frac{\omega_1^{\max} - \omega_1^{\min}}{2} \right), \\
\dot{P}_1^3 &= P_1^2 \left(\frac{\omega_1^{\max} + \omega_1^{\min}}{2} + sign(P_1^1) \frac{\omega_1^{\max} - \omega_1^{\min}}{2} \right), \\
\dot{P}_2^1 &= P_2^3 \left(\frac{v_2^{\max} + v_2^{\min}}{2} + sign(P_2^2) \frac{v_2^{\max} + v_2^{\min}}{2} \right), \\
\dot{P}_2^2 &= P_2^3 \left(-\frac{\omega_2^{\max} + \omega_2^{\min}}{2} - sign(P_2^1) \frac{\omega_2^{\max} - \omega_2^{\min}}{2} \right), \\
\dot{P}_2^3 &= P_2^2 \left(\frac{\omega_2^{\max} + \omega_2^{\min}}{2} + sign(P_2^1) \frac{\omega_2^{\max} - \omega_2^{\min}}{2} \right).
\end{aligned}$$

The final conditions for the variables $P_1^j(t)$ and $P_2^j(t)$ are get from the boundary of the safe set as

$$P_1^j(0) =< d_1 l(g), g_1 \xi^j >, \qquad P_2^j(0) =< d_2 l(g), g_2 \xi^j >,$$

where d_1 is the derivative of l taken with respect to its first argument g_1 only (and similarly for d_2). In this way, $P_1^j(t)$ and $P_2^j(t)$ are get for $t \leq 0$. Once this has been calculated, the optimal input $u^*(t)$ and the worst disturbance $d^*(t)$ are given respectively as

$$u^*(t) = \begin{cases} \omega_1^*(t) = \begin{cases} \omega_1^{\max} & \text{if} \quad P_1^1(t) > 0 \\ \omega_1^{\min} & \text{if} \quad P_1^1(t) < 0 \end{cases} \\ v_1^*(t) = \begin{cases} v_1^{\max} & \text{if} \quad P_1^2(t) > 0 \\ v_1^{\min} & \text{if} \quad P_1^2(t) < 0 \end{cases} \end{cases}$$

$$d^*(t) = \begin{cases} \omega_2^*(t) = \begin{cases} \omega_2^{\max} & \text{if} \quad P_2^1(t) > 0 \\ \omega_2^{\min} & \text{if} \quad P_2^1(t) < 0 \end{cases} \\ v_2^*(t) = \begin{cases} v_2^{\max} & \text{if} \quad P_2^2(t) > 0 \\ v_2^{\min} & \text{if} \quad P_2^2(t) < 0 \end{cases} \end{cases}.$$

Nash Solutions for Multi–Vehicle Manoeuvres

The methodology introduced in the previous sections can be generalized to find conflict–resolutions for multi–vehicle manoeuvres. Consider the three–vehicle dynamics:

$$\dot{g} = gX, \tag{2.856}$$

with

$$g = \begin{bmatrix} g_1 & 0 & 0 \\ 0 & g_2 & 0 \\ 0 & 0 & g_3 \end{bmatrix}, \qquad X = \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{bmatrix},$$

where g is an element in the configuration space $M = SE(2) \times SE(2) \times SE(2)$ and $X \in \mathfrak{se}(2) \times \mathfrak{se}(2) \times \mathfrak{se}(2)$ is linearly parameterised by inputs (ω_1, v_1), (ω_2, v_2) and (ω_3, v_3).

Now, the target set T is defined as

$$T = \{g \in M | l_1(g) < 0 \vee l_2(g) < 0 \vee l_3(g) < 0\},$$

where

$$\begin{aligned} l_1(g) &= \min\{(x_2 - x_1)^2 + (y_2 - y_1)^2 - \rho^2, \quad (x_3 - x_1)^2 + (y_3 - y_1)^2 - \rho^2\}, \\ l_2(g) &= \min\{(x_3 - x_2)^2 + (y_3 - y_2)^2 - \rho^2, \quad (x_1 - x_2)^2 + (y_1 - y_2)^2 - \rho^2\}, \\ l_3(g) &= \min\{(x_2 - x_3)^2 + (y_2 - y_3)^2 - \rho^2, \quad (x_1 - x_3)^2 + (y_1 - y_3)^2 - \rho^2\}. \end{aligned}$$

The control inputs $u = (u_1, u_2, u_3)$ are the actions of vehicle 1, 2 and 3:

$$u_i = (\omega_i, v_i) \in U_i,$$

where U_i are defined as

$$U_i = ([\omega_i^{\min}, \omega_i^{\max}], [v_i^{\min}, v_i^{\max}]).$$

Clearly, this can be generalized to N vehicles.

The cost functions $J_i(g, \{u_i(\cdot)\}, t)$ are defined as

$$J_i(g, \{u_i(\cdot)\}, t) : \prod_{i=1}^{N} SE_i(2) \times \prod_{i=1}^{N} \mathcal{U}_i \times \mathbb{R}_- \to \mathbb{R},$$

such that $J_i(g, \{u_i(\cdot)\}, t) = l_i(g(0))$.

The simplest non–cooperative solution strategy is a so–called *non–cooperative Nash equilibrium* (see e.g., [BO95]). A set of controls u_i^*, $(i = 1, ..., N)$ is said to be a *Nash strategy*, if for each player modification of that strategy under the assumption that the others play their Nash strategies results in a decrease in his payoff, that is for $i = 1, ..., N$, and $\forall u_i(\cdot)$,

$$J_i(u_1, ..., u_i, ..., u_N) \leq J_i(u_1^*, ..., u_i^*, ..., u_N^*), \qquad (u \neq u^*).$$

(Note that Nash equilibria may not be unique. It is also easy to see that for the two–player zero–sum game, a Nash equilibrium is a saddle solution with $J = J_1 = -J_2$.)

For N vehicles, the momentum functions are defined as in the two–vehicle case:

$$P_i^j =< p_i, g_i \xi^j >,$$

with $p_i \in \mathfrak{se}(2)^*$ for $i = 1, ..., N$ and ξ^j defined as above.

Then the Hamiltonian $H(g, p, u_1, ...u_N)$ is given by

$$H(g, p, u_1, ...u_N) = P_i^1 \omega_i + P_i^2 v_i.$$

The first case we consider is one in which all the vehicles are *cooperating*, meaning that each tries to avoid conflict assuming the others are doing the same. In this case, the optimal Hamiltonian $H^*(g, p)$ is

$$H^*(g, p) = \max_{u_i \in U_i} H(g, p, u_1, ...u_N).$$

For example, if $N = 3$, one may solve for (u_1^*, u_2^*, u_3^*), on the 9D quotient space T^*M/M, so that the optimal control inputs are given as

$$u_i^*(t) = \begin{cases} \omega_i^*(t) = \begin{cases} \omega_i^{\max} & \text{if} \quad P_i^1(t) > 0 \\ \omega_i^{\min} & \text{if} \quad P_i^1(t) < 0 \end{cases} \\ v_i^*(t) = \begin{cases} v_i^{\max} & \text{if} \quad P_i^2(t) > 0 \\ v_i^{\min} & \text{if} \quad P_i^2(t) < 0 \end{cases} \end{cases}.$$

One possibility for the optimal Hamiltonian corresponding to the *non–cooperative case* is

$$H^*(g, p) = \max_{u_1 \in U_1} \max_{u_2 \in U_2} \max_{u_3 \in U_3} H(g, p, u_1, u_2, u_3).$$

2.6.4 Nonlinear Dynamics of Option Pricing

Classical theory of option pricing is based on the results found in 1973 by Black and Scholes [BS73] and, independently, Merton [Mer73]. Their pioneering work starts from the basic assumption that the asset prices follow the dynamics of a particular stochastic process (geometrical Brownian motion), so that they have a lognormal distribution [Hul00, PB99]. In the case

of an efficient market with no arbitrage possibilities, no dividends and constant volatilities, they found that the price of each financial derivative is ruled by an ordinary partial differential equation, known as the (Nobel–Prize winning) *Black–Scholes–Merton* (BSM) formula. In the most simple case of a so–called European option, the BSM equation can be explicitly solved to get an analytical formula for the price of the option [Hul00, PB99]. When we consider other financial derivatives, which are commonly traded in real markets and allow anticipated exercise and/or depend on the history of the underlying asset, the BSM formula fails to give an analytical result. Appropriate numerical procedures have been developed in the literature to price exotic financial derivatives with path–dependent features, as discussed in detail in [Hul00, WDH93, PBS01]. The aim of this work is to give a contribution to the problem of efficient option pricing in financial analysis, showing how it is possible to use path integral methods to develop a fast and precise algorithm for the evaluation of option prices.

Following recent studies on the application of the path integral approach to the financial market as appeared in the econophysics literature (see [Mat02] for a comprehensive list of references), in [MNM02] the authors proposed an original, efficient path integral algorithm to price financial derivatives, including those with path-dependent and early exercise features, and to compare the results with those get with the standard procedures known in the literature.

Theory and Simulations of Option Pricing

Classical Theory and Path–Dependent Options

The basic ingredient for the development of a theory of option pricing is a suitable model for the time evolution of the asset prices. The assumption of the BSM model is that the price S of an asset is driven by a *Brownian motion* and verifies the stochastic differential equation (SDE) [Hul00, PB99]

$$dS = \mu S dt + \sigma S dw, \tag{2.857}$$

which, by means of the *Itô lemma*, can be cast in the form of an arithmetic Brownian motion for the logarithm of S

$$d(\ln S) = A dt + \sigma dw, \tag{2.858}$$

where σ is the *volatility*, $A = \left(\mu - \sigma^2/2\right)$, μ is the drift parameter and w is the realization of a Wiener process. Due to the properties of a Wiener process, (2.858) may be written as

$$d(\ln S) = A dt + \sigma \epsilon \sqrt{dt}, \tag{2.859}$$

where ϵ follows from a standardized normal distribution with mean 0 and variance 1. Thus, in terms of the logarithms of the asset prices $z' = \ln S', z =$

$\ln S$, the conditional transition probability $p(z'|z)$ to have at the time t' a price S' under the hypothesis that the price was S at the time $t < t'$ is given by [PB99, BRT99]

$$p(z'|z) = \frac{1}{\sqrt{2\pi(t'-t)\sigma^2}} \exp\left\{-\frac{[z'-(z+A(t'-t))]^2}{2\sigma^2(t'-t)}\right\}, \tag{2.860}$$

which is a gaussian distribution with mean $z+A(t'-t)$ and variance $\sigma^2(t'-t)$. If we require the options to be exercised only at specific times $t_i, i = 1, \cdots, n$, the asset price, between two consequent times t_{i-1} and t_i, will follow (2.859) and the related transition probability will be

$$p(z_i|z_{i-1}) = \frac{1}{\sqrt{2\pi\Delta t\sigma^2}} \exp\left\{-\frac{[z_i-(z_{i-1}+A\Delta t)]^2}{2\sigma^2\Delta t}\right\}, \tag{2.861}$$

with $\Delta t = t_i - t_{i-1}$.

A time–evolution model for the asset price is strictly necessary in a theory of option pricing because the fair price at time $t = 0$ of an option $\mathcal{O}$, without possibility of anticipated exercise before the expiration date or maturity T (a so–called *European option*), is given by the scaled expectation value [Hul00]

$$\mathcal{O}(0) = e^{-rT} E[\mathcal{O}(T)], \tag{2.862}$$

where r is the risk–free interest and $E[\cdot]$ indicates the mean value, which can be computed only if a model for the asset underlying the option is understood. For example, the value $\mathcal{O}$ of an European call option at the maturity T will be $\max\{S_T - X, 0\}$, where X is the strike price, while for an European put option the value $\mathcal{O}$ at the maturity will be $\max\{X - S_T, 0\}$. It is worth emphasizing, for what follows, that the case of an European option is particularly simple, since in such a situation the price of the option can be evaluated by means of analytical formulae, which are get by solving the BSM partial differential equation with the appropriate boundary conditions [Hul00, PB99]. On the other hand, many further kinds of options are present in the financial markets, such as American options (options which can be exercised at any time up to the expiration date) and exotic options [Hul00], i.e., derivatives with complicated payoffs or whose value depend on the whole time evolution of the underlying asset and not just on its value at the end. For such options with path-dependent and early exercise features no exact solutions are available and pricing them correctly is a great challenge.

In the case of options with possibility of anticipated exercise before the expiration date, the above discussion needs to be generalized, by introducing a slicing of the time interval T. Let us consider, for definiteness, the case of an option which can be exercised within the maturity but only at the times $t_1 = \Delta t, t_2 = 2\Delta t, \ldots, t_n = n\Delta t = T$. At each time slice t_{i-1} the value $\mathcal{O}_{i-1}$ of the option will be the maximum between its expectation value at the time

t_i scaled with $e^{-r\Delta t}$ and its value in the case of anticipated exercise $\mathcal{O}_{i-1}^{Y}$. If S_{i-1} denotes the price of the underlying asset at the time t_{i-1}, we can thus write for each $i = 1, \ldots, n$

$$\mathcal{O}_{i-1}(S_{i-1}) = \max\left\{\mathcal{O}_{i-1}^{Y}(S_{i-1}), e^{-r\Delta t}E[\mathcal{O}_i|S_{i-1}]\right\}, \tag{2.863}$$

where $E[\mathcal{O}_i|S_{i-1}]$ is the conditional expectation value of $\mathcal{O}_i$, i.e., its expectation value under the hypothesis of having the price S_{i-1} at the time t_{i-1}. In this way, to get the actual price $\mathcal{O}_0$, it is necessary to proceed backward in time and calculate $\mathcal{O}_{n-1}, \ldots, \mathcal{O}_1$, where the value $\mathcal{O}_n$ of the option at maturity is nothing but $\mathcal{O}_n^{Y}(S_n)$. It is therefore clear that evaluating the price of an option with early exercise features means to simulate the evolution of the underlying asset price (to get the $\mathcal{O}_i^{Y}$) and to calculate a (usually large) number of expectation conditional probabilities.

Standard Numerical Procedures

To value derivatives when analytical formulae are not available, appropriate numerical techniques have to be advocated. They involve the use of Monte Carlo (MC) simulation, binomial trees (and their improvements) and finite–difference methods [Hul00, WDH93].

A natural way to simulate price paths is to discretize (2.859) as

$$\ln S(t + \Delta t) - \ln S(t) = A\Delta t + \sigma\epsilon\sqrt{\Delta t},$$

or, equivalently,

$$S(t + \Delta t) = S(t)\exp\left[A\Delta t + \sigma\epsilon\sqrt{\Delta t}\right], \tag{2.864}$$

which is correct for any $\Delta t > 0$, even if finite. Given the spot price S_0, i.e., the price of the asset at time $t = 0$, one can extract from a standardized normal distribution a value ϵ_k, $(k = 1, \ldots, n)$ for the random variable ϵ to simulate one possible path followed by the price by means of (2.864):

$$S(k\Delta t) = S((k-1)\Delta t)\exp\left[A\Delta t + \sigma\epsilon_k\sqrt{\Delta t}\right].$$

Iterating the procedure m times, one can simulate m price paths $\{(S_0, S_1^{(j)}, S_2^{(j)}, \ldots, S_n^{(j)} \equiv S_T^{(j)}) : j = 1, \ldots, m\}$ and evaluate the price of the option. In such a MC simulation of the stochastic dynamics of asset price (Monte Carlo random walk) the mean values $E[\mathcal{O}_i|S_{i-1}], i = 1, \ldots, n$ are given by

$$E[\mathcal{O}_i|S_{i-1}] = \frac{\mathcal{O}_i^{(1)} + \mathcal{O}_i^{(2)} + \cdots + \mathcal{O}_i^{(m)}}{m},$$

with no need to calculate transition probabilities because, through the extraction of the possible ϵ values, the paths are automatically weighted according

to the probability distribution function of (2.861). Unfortunately, this method leads to an estimated value whose numerical error is proportional to $m^{-1/2}$. Thus, even if it is powerful because of the possibility to control the paths and to impose additional constrains (as it is usually required by exotic and path-dependent options), the MC random walk is extremely time consuming when precise predictions are required and appropriate variance reduction procedures have to be used to save CPU time [Hul00]. This difficulty can be overcome by means of the method of the binomial trees and its extensions (see [Hul00] and references therein), whose main idea stands in a deterministic choice of the possible paths to limit the number of intermediate points. At each time step the price S_i is assumed to have only two choices: increase to the value $uS_i, u > 1$ or decrease to $dS_i, 0 < d < 1$, where the parameters u and d are given in terms of σ and Δt in such a way to give the correct values for the mean and variance of stock price changes over the time interval Δt. Also finite difference methods are known in the literature [Hul00] as an alternative to time-consuming MC simulations. They give the value of the derivative by solving the differential equation satisfied by the derivative, by converting it into a difference equation. Although tree approaches and finite difference methods are known to be faster than the MC random walk, they are difficult to apply when a detailed control of the history of the derivative is required and are also computationally time consuming when a number of stochastic variables is involved [Hul00]. It follows that the development of efficient and fast computational algorithms to price financial derivatives is still a key issue in financial analysis.

Option Pricing via Path Integrals

Recall that the path integral method is an integral formulation of the dynamics of a stochastic process. It is a suitable framework for the calculation of the transition probabilities associated to a given stochastic process, which is seen as the convolution of an infinite sequence of infinitesimal short-time steps [BRT99, Sch81]. For the problem of option pricing, the path–integral method can be employed for the explicit calculation of the expectation values of the quantities of financial interest, given by integrals of the form [BRT99]

$$E[\mathcal{O}_i|S_{i-1}] = \int dz_i p(z_i|z_{i-1})\mathcal{O}_i(e^{z_i}), \tag{2.865}$$

where $z = \ln S$ and $p(z_i|z_{i-1})$ is the transition probability. $E[\mathcal{O}_i|S_{i-1}]$ is the conditional expectation value of some functional $\mathcal{O}_i$ of the stochastic process. For example, for an European call option at the maturity T the quantity of interest will be $\max\{S_T - X, 0\}$, X being the strike price. As already emphasized, and discussed in the literature [Hul00, WDH93, PBS01, RT02, Mat02], the computational complexity associated to this calculation is generally great: in the case of exotic options, with path-dependent and early exercise features,

integrals of the type (2.865) cannot be analytically solved. As a consequence, we demand two things from a path integral framework: a very quick way to estimate the transition probability associated to a stochastic process (2.859) and a clever choice of the integration points with which evaluate the integrals (2.865). In particular, our aim is to develop an efficient calculation of the probability distribution without losing information on the path followed by the asset price during its time evolution.

Transition Probability

The probability distribution function related to a SDE verifies the *Chapman–Kolmogorov equation* [PB99]

$$p(z''|z') = \int dz p(z''|z)p(z|z'), \tag{2.866}$$

which states that the probability (density) of a transition from the value z' (at time t') to the value z'' (at time t'') is the 'summation' over all the possible intermediate values z of the probability of separate and consequent transitions $z' \to z$, $z \to z''$. As a consequence, if we consider a finite time interval $[t', t'']$ and we apply a time slicing, by considering $n+1$ subintervals of length $\Delta t = (t'' - t')/n + 1$, we can write, by iteration of (2.866)

$$p(z''|z') = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} dz_1 \cdots dz_n p(z''|z_n)p(z_n|z_{n-1}) \cdots p(z_1|z'),$$

which, thanks to (2.860), can be written as [MNM02]

$$\int_{-\infty}^{+\infty} \cdots \tag{2.867}$$
$$\cdots \int_{-\infty}^{+\infty} dz_1 \cdots dz_n \frac{1}{\sqrt{(2\pi\sigma^2\Delta t)^{n+1}}} \exp\left\{ -\frac{1}{2\sigma^2\Delta t} \sum_{k=1}^{n+1} [z_k - (z_{k-1} + A\Delta t)]^2 \right\}.$$

In the limit $n \to \infty$, $\Delta t \to 0$ such that $(n+1)\Delta t = (t'' - t')$ (infinite sequence of infinitesimal time steps), the expression (2.867), as explicitly shown in [BRT99], exhibits a Lagrangian structure and it is possible to express the transition probability in the path integral formalism as a convolution of the form [BRT99]

$$p(z'', t''|z', t') = \int_{\mathcal{C}} \mathcal{D}[\sigma^{-1}\tilde{z}] \exp\left\{ -\int_{t'}^{t''} L(\tilde{z}(\tau), \dot{\tilde{z}}(\tau); \tau) d\tau \right\},$$

where L is the Lagrangian, given by

$$L(\tilde{z}(\tau), \dot{\tilde{z}}(\tau); \tau) = \frac{1}{2\sigma^2} \left[\dot{\tilde{z}}(\tau) - A \right]^2,$$

and the integral is performed (with functional measure $\mathcal{D}[\cdot]$) over the paths $\tilde{z}(\cdot)$ belonging to $\mathcal{C}$, i.e., all the continuous functions with constrains $\tilde{z}(t') \equiv z'$, $\tilde{z}(t'') \equiv z''$. As carefully discussed in [BRT99], a path integral is well defined only if both a continuous formal expression and a discretization rule are given. As done in many applications, the Itô prescription is adopted here (see subsection 1.5.2 above).

A first, naïve evaluation of the transition probability (2.867) can be performed via Monte Carlo simulation, by writing (2.867) as

$$p(z'', t''|z', t') = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \prod_{i}^{n} dg_i \frac{1}{\sqrt{2\pi\sigma^2 \Delta t}} \exp\left\{-\frac{1}{2\sigma^2 \Delta t}\left[z'' - (z_n + A\Delta t)\right]^2\right\}, \quad (2.868)$$

in terms of the variables g_i defined by the relation

$$dg_k = \frac{dz_k}{\sqrt{2\pi\sigma^2 \Delta t}} \exp\left\{-\frac{1}{2\sigma^2 \Delta t}\left[z_k - (z_{k-1} + A\Delta t)\right]^2\right\}, \quad (2.869)$$

and extracting each g_i from a gaussian distribution of mean $z_{k-1} + A\Delta t$ and variance $\sigma^2 \Delta t$. However, as we will see, this method requires a large number of calls to get a good precision. This is due to the fact that each g_i is related to the previous g_{i-1}, so that this implementation of the path integral approach can be seen to be equivalent to a naïve MC simulation of random walks, with no variance reduction.

By means of appropriate manipulations [Sch81] of the integrand entering (2.867), it is possible, as shown in the following, to get a path integral expression which will contain a factorized integral with a constant kernel and a consequent variance reduction. If we define $z'' = z_{n+1}$ and $y_k = z_k - kA\Delta t$, $k = 1, \ldots, n$, we can express the *transition probability distribution* as

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} dy_1 \cdots dy_n \frac{1}{\sqrt{(2\pi\sigma^2 \Delta t)^{n+1}}} \cdot \exp\left\{-\frac{1}{2\sigma^2 \Delta t} \sum_{k=1}^{n+1} [y_k - y_{k-1}]^2\right\}, \quad (2.870)$$

in order to get rid of the contribution of the drift parameter. Now let us extract from the argument of the exponential function a quadratic form

$$\begin{aligned} \sum_{k=1}^{n+1} [y_k - y_{k-1}]^2 &= y_0^2 - 2y_1 y_0 + y_1^2 + y_1^2 - 2y_1 y_2 + \ldots + y_{n+1}^2 \\ &= y^t M y + [y_0^2 - 2y_1 y_0 + y_{n+1}^2 - 2y_n y_{n+1}], \end{aligned} \quad (2.871)$$

by introducing the nD array y and the nxn matrix M defined as [MNM02]

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{pmatrix}, \qquad M = \begin{pmatrix} 2 & -1 & 0 & \cdots & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 \\ 0 & \cdots & -1 & 2 & -1 & 0 \\ 0 & \cdots & \cdots & -1 & 2 & -1 \\ 0 & \cdots & \cdots & \cdots & -1 & 2 \end{pmatrix}, \tag{2.872}$$

where M is a real, symmetric, non singular and tridiagonal matrix. In terms of the eigenvalues m_i of the matrix M, the contribution in (2.871) can be written as

$$y^t M y = w^t O^t M O w = w^t M_d w = \sum_{i=1}^{n} m_i w_i^2, \tag{2.873}$$

by introducing the orthogonal matrix O which diagonalizes M, with $w_i = O_{ij} y_j$. Because of the orthogonality of O, the Jacobian

$$J = \det \left| \frac{dw_i}{dy_k} \right| = \det |O_{ki}|,$$

of the transformation $y_k \to w_k$ equals 1, so that $\prod_{i=1}^{n} dw_i = \prod_{i=1}^{n} dy_i$. After some algebra, (2.871) can be written as

$$\sum_{k=1}^{n+1} [y_k - y_{k-1}]^2 = \sum_{i=1}^{n} m_i w_i^2 + y_0^2 - 2y_1 y_0 + y_{n+1}^2 - 2y_n y_{n+1} =$$

$$\sum_{i=1}^{n} m_i \left[w_i - \frac{(y_0 O_{1i} + y_{n+1} O_{ni})}{m_i} \right]^2 + y_0^2 + y_{n+1}^2 - \sum_{i=1}^{n} \frac{(y_0 O_{1i} + y_{n+1} O_{ni})^2}{m_i}. \tag{2.874}$$

Now, if we introduce new variables h_i obeying the relation

$$dh_i = \sqrt{\frac{m_i}{2\pi\sigma^2 \Delta t}} \exp \left\{ -\frac{m_i}{2\sigma^2 \Delta t} \left[w_i - \frac{(y_0 O_{1i} + y_{n+1} O_{ni})}{m_i} \right]^2 \right\} dw_i, \tag{2.875}$$

it is possible to express the *finite–time probability distribution* $p(z''|z')$ as [MNM02]

$$\begin{aligned} &\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \prod_{i=1}^{n} dy_i \frac{1}{\sqrt{(2\pi\sigma^2 \Delta t)^{n+1}}} \exp \left\{ -\frac{1}{2\sigma^2 \Delta t} \sum_{k=1}^{n+1} [y_k - y_{k-1}]^2 \right\} \\ &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \prod_{i=1}^{n} dw_i \frac{1}{\sqrt{(2\pi\sigma^2 \Delta t)^{n+1}}} e^{-(y_0^2 + y_{n+1}^2)/2\sigma^2 \Delta t} \\ &\quad \times \exp \left\{ -\frac{1}{2\sigma^2 \Delta t} \sum_{i=1}^{n} \left[m_i \left(w_i - \frac{(y_0 O_{1i} + y_{n+1} O_{ni})}{m_i} \right)^2 - \frac{(y_0 O_{1i} + y_{n+1} O_{ni})^2}{m_i} \right] \right\} \\ &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \prod_{i=1}^{n} dh_i \frac{1}{\sqrt{2\pi\sigma^2 \Delta t \det(M)}} \\ &\quad \times \exp \left\{ -\frac{1}{2\sigma^2 \Delta t} \left[y_0^2 + y_{n+1}^2 + \sum_{i=1}^{n} \frac{(y_0 O_{1i} + y_{n+1} O_{ni})^2}{m_i} \right] \right\}. \end{aligned} \tag{2.876}$$

The probability distribution function, as given by (2.876), is an integral whose kernel is a constant function (with respect to the integration variables) and which can be factorized into the n integrals

$$\int_{-\infty}^{+\infty} dh_i \exp\left\{-\frac{1}{2\sigma^2 \Delta t}\frac{(y_0 O_{1i} + y_{n+1} O_{ni})^2}{m_i}\right\}, \tag{2.877}$$

given in terms of the h_i, which are gaussian variables that can be extracted from a normal distribution with mean $(y_0 O_{1i} + y_{n+1} O_{ni})^2/m_i$ and variance $\sigma^2 \Delta t/m_i$. Differently to the first, naïve implementation of the path integral, now each h_i is no longer dependent on the previous h_{i-1}, and importance sampling over the paths is automatically accounted for.

It is worth noticing that, by means of the extraction of the random variables h_i, we are creating price paths, since at each intermediate time t_i the asset price is given by

$$S_i = \exp\{\sum_{k=1}^{n} O_{ik} h_k + iA\Delta t\}. \tag{2.878}$$

Therefore, this path integral algorithm can be easily adapted to the cases in which the derivative to be valued has, in the time interval $[0, T]$, additional constraints, as in the case of interesting path–dependent options, such as Asian and barrier options [Hul00].

Integration Points

The above illustrated method represents a powerful and fast tool to calculate the transition probability in the path integral framework and it can be employed if we need to value a generic option with maturity T and with possibility of anticipated exercise at times $t_i = i\Delta t$ $(n\Delta t = T)$ [MNM02]. As a consequence of this time slicing, one must numerically evaluate $n - 1$ mean values of the type (9), in order to check at any time t_i, and for any value of the stock price, whether early exercise is more convenient with respect to holding the option for a future time. To keep under control the computational complexity and the time of execution, it is mandatory to limit as far as possible the number of points for the integral evaluation. This means that we would like to have a linear growth of the number of integration points with the time. Let us suppose to evaluate each mean value

$$E[\mathcal{O}_i|S_{i-1}] = \int dz_i \, p(z_i|z_{i-1}) \mathcal{O}_i(e^{z_i}),$$

with p integration points, i.e., considering only p fixed values for z_i. To this end, we can create a grid of possible prices, according to the dynamics of the stochastic process as given by (2.859)

$$z(t + \Delta t) - z(t) = \ln S(t + \Delta t) - \ln S(t) = A\Delta t + \epsilon\sigma\sqrt{\Delta t}. \tag{2.879}$$

Starting from z_0, we thus evaluate the expectation value $E[\mathcal{O}_1|S_0]$ with $p = 2m+1, m \in \mathbb{N}$ values of z_1 centered on the mean value $E[z_1] = z_0 + A\Delta t$ and which differ from each other of a quantity of the order of $\sigma\sqrt{\Delta t}$

$$z_1^j = z_0 + A\Delta t + j\sigma\sqrt{\Delta t}, \qquad (j = -m, \ldots, +m).$$

Going on like this, we can evaluate each expectation value $E[\mathcal{O}_2|z_1^j]$ get from each one of the z_1's created above with p values for z_2 centered around the mean value

$$E[z_2|z_1^j] = z_1^j + A\Delta t = z_0 + 2A\Delta t + j\sigma\sqrt{\Delta t}.$$

Iterating the procedure until the maturity, we create a deterministic grid of points such that, at a given time t_i, there are $(p-1)i+1$ values of z_i, in agreement with the request of linear growth. This procedure of selection of integration points, together with the calculation of the transition probability previously described, is the basis of the path integral simulation of the price of a generic option.

By applying the results derived above, we have at disposal an efficient path integral algorithm both for the calculation of transition probabilities and the evaluation of option prices. In [MNM02] the application of the above path–integral method to European and American options in the BSM model was illustrated and comparisons with the results were get with the standard procedures known in the literature were shown. First, the path integral simulation of the probability distribution of the logarithm of the stock prices, $p(lnS)$, as a function of the logarithm of the stock price, for a BSM–like stochastic model, was given by (2.858). Once the transition probability has been computed, the price of an option could be computed in a path integral approach as the conditional expectation value of a given functional of the stochastic process. For example, the price of an European call option was given by

$$\mathcal{C} = e^{-r(T-t)} \int_{-\infty}^{+\infty} dz_f \, p(z_f, T|z_i, t) \max[e^{z_f} - X, 0], \tag{2.880}$$

while for an European put it will be

$$\mathcal{P} = e^{-r(T-t)} \int_{-\infty}^{+\infty} dz_f \, p(z_f, T|z_i, t) \max[X - e^{z_f}, 0], \tag{2.881}$$

where r is the risk–free interest rate. Therefore just 1D integrals need to be evaluated and they can be precisely computed with standard quadrature rules.

Continuum Limit and American Options

In the specific case of an *American option*, the possibility of exercise at any time up to the expiration date allows to develop, within the path integral formalism, a specific algorithm, which, as shown in the following, is precise and very quick [MNM02].

Given the time slicing considered above, the case of American options requires the limit $\Delta t \longrightarrow 0$ which, putting $\sigma \longrightarrow 0$, leads to a delta–like transition probability

$$p(z, t + \Delta t | z_t, t) \approx \delta(z - z_t - A\Delta t).$$

This means that, apart from volatility effects, the price z_i at time t_i will have a value remarkably close to the expected value $\bar{z} = z_{i-1} + A\Delta t$, given by the drift growth. In order to take care of the volatility effects, a possible solution is to estimate the integral of interest, i.e.,

$$E[\mathcal{O}_i | S_{i-1}] = \int_{-\infty}^{+\infty} dz\, p(z|z_{i-1}) \mathcal{O}_i(e^z), \tag{2.882}$$

by inserting in (2.882) the analytical expression for the $p(z|z_{i-1})$ transition probability

$$\begin{aligned} p(z|z_{i-1}) &= \frac{1}{\sqrt{2\pi\Delta t\sigma^2}} \exp\left\{ -\frac{(z - z_{i-1} - A\Delta t)^2}{2\sigma^2 \Delta t} \right\} \\ &= \frac{1}{\sqrt{2\pi\Delta t\sigma^2}} \exp\left\{ -\frac{(z - \bar{z})^2}{2\sigma^2 \Delta t} \right\}, \end{aligned}$$

together with a Taylor expansion of the kernel function $\mathcal{O}_i(e^z) = f(z)$ around the expected value $\bar{z}$. Hence, up to the second–order in $z - \bar{z}$, the kernel function becomes

$$f(z) = f(\bar{z}) + (z - \bar{z}) f'(\bar{z}) + \frac{1}{2} f''(\bar{z})(z - \bar{z})^2 + O((z - \bar{z})^3),$$

which induces

$$E[\mathcal{O}_i | S_{i-1}] = f(\bar{z}) + \frac{\sigma^2}{2} f''(\bar{z}), + \ldots,$$

since the first derivative does not give contribution to (2.882), being the integral of an odd function over the whole z range. The second derivative can be numerically estimated as

$$f''(\bar{z}) = \frac{1}{\delta_\sigma^2} [f(\bar{z} + \delta_\sigma) - 2f(\bar{z}) + f(\bar{z} - \delta_\sigma)],$$

with $\delta_\sigma = O(\sigma\sqrt{\Delta t})$, as dictated by the dynamics of the stochastic process.

2.6.5 Command/Control in Human–Robot Interactions

Suppose that we have a *human–robot team*, consisting of m robots and n humans. To be able to put the modelling of the fully controlled human–robot team performance into the rigorous geometrical settings, we suppose that all possible behaviors of m robots can be described by a set of continuous and smooth, time–dependent *robot configuration coordinates* $x^r = x^r(t)$, while all

robot–related behaviors of n humans can be described by a set of continuous and smooth, time–dependent *human configuration coordinates* $q^h = q^h(t)$. In other words, all robot coordinates, $x^r = x^r(t)$, constitute the smooth Riemannian manifold M^r_g (such that $r = 1, ..., \dim(M^r_g)$), with the positive–definite metric form

$$g \mapsto ds^2 = g_{rs}(x)dx^r dx^s \tag{2.883}$$

similarly, all human coordinates $q^h = q^h(t)$, constitute a smooth Riemannian manifold N^h_a (such that $h = 1, ..., \dim(N^h_a)$) , with the positive–definite metric form

$$a \mapsto d\sigma^2 = a_{hk}(q)dq^h dq^k. \tag{2.884}$$

In this Riemannian geometry settings, the feedforward command/control action of humans upon robots is defined by a smooth map,

$$\mathcal{C} : N^h_a \to M^r_g,$$

which is in local coordinates given by a general (nonlinear) functional transformation

$$x^r = x^r(q^h), \qquad (r = 1, ..., \dim(M^r_g);\ h = 1, ..., \dim(N^h_a)), \tag{2.885}$$

while its inverse, the feedback map from robots to humans is defined by a smooth map,

$$\mathcal{F} = \mathcal{C}^{-1} : M^r_g \to N^h_a,$$

which is in local coordinates given by an inverse functional transformation

$$q^h = q^h(x^r), \qquad (h = 1, ..., \dim(N^h_a);\ r = 1, ..., \dim(M^r_g)). \tag{2.886}$$

Now, although the coordinate transformations (2.885) and (2.886) are completely general, nonlinear and even unknown at this stage, there is something known and simple about them: the corresponding transformations of differentials are *linear and homogenous*, namely

$$dx^r = \frac{\partial x^r}{\partial q^h} dq^h, \qquad \text{and} \qquad dq^h = \frac{\partial q^h}{\partial x^r} dx^r,$$

which imply linear and homogenous transformations of robot and human velocities,

$$\dot{x}^r = \frac{\partial x^r}{\partial q^h} \dot{q}^h, \qquad \text{and} \qquad \dot{q}^h = \frac{\partial q^h}{\partial x^r} \dot{x}^r. \tag{2.887}$$

Relation (2.887), representing two autonomous dynamical systems, given by two sets of ordinary differential equations (ODEs), geometrically defines two *velocity vector–fields*: (i) robot velocity vector–field, $v^r \equiv v^r(x^r, t) := \dot{x}^r(x^r, t)$; and human velocity vector–field, $u^h \equiv u^h(q^h, t) := \dot{q}^h(q^h, t)$. Recall that a vector–field defines a single vector at each point x^r (in some domain U) of a manifold in case. Its solution gives the *flow*, consisting of *integral curves*

of the vector–field, such that all the vectors from the vector–field are tangent to integral curves at different points $x^i \in U$. Geometrically, a velocity vector–field is defined as a cross–section of the tangent bundle of the manifold. In our case, the robot velocity vector–field $v^r = \dot{x}^r(x^r, t)$ represents a cross–section of the robot tangent bundle TM^r_g, while the human velocity vector–field $u^h = \dot{q}^h(q^h, t)$ represents a cross–section of the human tangent bundle TN^h_a. In this way, two local velocity vector–fields, v^r and u^h, give local representations for the following two global *tangent maps*,

$$T\mathcal{C} : TN^h_a \to TM^r_g, \qquad \text{and} \qquad T\mathcal{F} : TM^r_g \to TN^h_a.$$

To be able to proceed along the geometrodynamical line, we need next to formulate the two corresponding acceleration vector–fields, $a^r \equiv a^r(x^r, \dot{x}^r, t)$ and $w^h \equiv w^h(q^h, \dot{q}^h, t)$, as time rates of change of the two velocity vector–fields v^r and u^h. Now, recall that the acceleration vector–field is defined as the *absolute time derivative*, $\dot{v}^r = \frac{D}{dt}v^r$, of the velocity vector–field. In our case, we have the robotic acceleration vector–field $a^r := \dot{v}^r$ defined on M^r_g by

$$a^r := \dot{v}^r = \dot{v}^r + \Gamma^r_{st} v^s v^t = \ddot{x}^r + \Gamma^r_{st} \dot{x}^s \dot{x}^t, \tag{2.888}$$

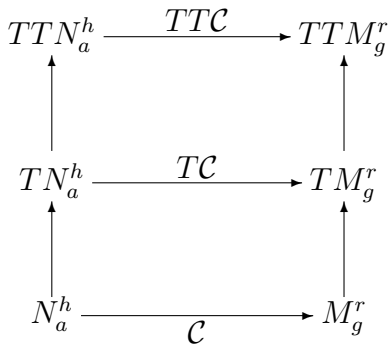
and the human acceleration vector–field $w^h := \dot{u}^h$ defined on N^h_a by

$$w^h := \dot{u}^h = \dot{u}^h + \Gamma^h_{jk} u^j u^k = \ddot{x}^r + \Gamma^h_{jk} \dot{q}^j \dot{q}^k, \tag{2.889}$$

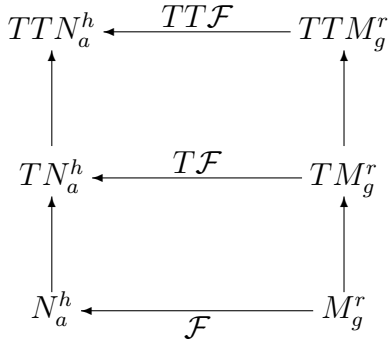
Geometrically, an acceleration vector–field is defined as a cross–section of the second tangent bundle of the manifold. In our case, the robot acceleration vector–field $a^r = \dot{v}^r(x^r, \dot{x}^r, t)$, given by the ODEs (2.888), represents a cross–section of the second robot tangent bundle TTM^r_g, while the human acceleration vector–field $w^h = \dot{u}^h(q^h, \dot{q}^h, t)$, given by the ODEs (2.889), represents a cross–section of the second human tangent bundle TTN^h_a. In this way, two local acceleration vector–fields, a^r and w^h, give local representations for the following two *second tangent maps*,

$$TT\mathcal{C} : TTN^h_a \to TTM^r_g, \qquad \text{and} \qquad TT\mathcal{F} : TTM^r_g \to TTN^h_a.$$

In other words, we have the feedforward command/control commutative diagram:



as well as the feedback commutative diagram:



These two commutative diagrams formally define the global feedforward and feedback human–robot interactions at the positional, velocity, and acceleration levels of command and control.

2.6.6 Nonlinear Dynamics of Complex Nets

Recall that many systems in nature, such as neural nets, food webs, metabolic systems, co–authorship of papers, the worldwide web, etc. can be represented as *complex networks*, or *small–world networks* (see, e.g., [WS98, DM03]). In particular, it has been recognized that many networks have scale–free topology; the distribution of the degree obeys the power law, $P(k) \sim k^{-\gamma}$. The study of the scale–free network now attracts the interests of many researchers in mathematics, physics, engineering and biology [Ich04].

Another important aspect of complex networks is their dynamics, describing e.g., the spreading of viruses in the Internet, change of populations in a food web, and synchronization of neurons in a brain. In particular, [Ich04] studied the synchronization of the random network of oscillators. His work follows the previous studies (see [Str00]) that showed that mean–field type synchronization, that Kuramoto observed in globally–coupled oscillators [Kur84], appeared also in the small–world networks.

Continuum Limit of the Kuramoto Net

Ichinomiya started with the standard network with N nodes, described by a variant of the *Kuramoto model* (2.760). At each node, there exists an oscillator and the phase of each oscillator θ_i is evolving according to

$$\dot{\theta}_i = \omega_i + K \sum_j a_{ij} \sin(\theta_j - \theta_i), \tag{2.890}$$

where K is the coupling constant, a_{ij} is 1 if the nodes i and j are connected, and 0 otherwise; ω_i is a random number, whose distribution is given by the function $N(\omega)$.

For the analytic study, it is convenient to use the *continuum limit equation*. We define $P(k)$ as the distribution of nodes with degree k, and $\rho(k,\omega;t,\theta)$ the density of oscillators with phase θ at time t, for given ω and k. We assume that $\rho(k,\omega;t,\theta)$ is normalized as

$$\int_0^{2\pi} \rho(k,\omega;t,\theta)d\theta = 1.$$

For simplicity, we also assume $N(\omega) = N(-\omega)$. Thus, we suppose that the collective oscillation corresponds to the stable solution, $\dot{\rho} = 0$.

Now we construct the continuum limit equation for the network of oscillators. The evolution of ρ is determined by the *continuity equation* $\partial_t \rho = -\partial_\theta(\rho v)$, where v is defined by the continuum limit of the r.h.s of (2.890). Because one randomly selected edge connects to the node of degree k, frequency ω, phase θ with the probability $kP(k)N(\omega)\rho(k,\omega;t,\theta)/\int dkkP(k)$, $\rho(k,\omega;t,\theta)$ obeys the equation

$$\partial_t \rho(k,\omega;t,\theta) = -\partial_\theta[\rho(k,\omega;t,\theta)\,(\omega\ + \\ + \frac{Kk\int d\omega' \int dk' \int d\theta' N(\omega')P(k')k'\rho(k',\omega';t,\theta')\sin(\theta-\theta')}{\int dk' P(k')k'})].$$

The mean–field solution of this equation was studied by [Ich04].

Path–Integral Approach to Complex Nets

Recently, [Ich05] introduced the *path–integral* (see subsection 4.4.6 above) *approach* in studying the dynamics of complex networks. He considered the stochastic generalization of the Kuramoto network (2.890), given by

$$\dot{x}_i = f_i(x_i) + \sum_{j=1}^{N} a_{ij} g(x_i, x_j) + \xi_i(t), \tag{2.891}$$

where $f_i = f_i(x_i)$ and $g_{ij} = g(x_i, x_j)$ are functions of network activations x_i, $\xi_i(t)$ is a *random force* that satisfies $\langle \xi_i(t) = 0 \rangle$, $\langle \xi_i(t)\xi_j(t') \rangle = \delta_{ij}\delta(t-t')\sigma^2$. He assumed $x_i = x_{i,0}$ at $t = 0$. In order to discuss the dynamics of this system, he introduced the so–called *Matrin–Siggia–Rose* (MSR) *generating functional* Z given by [Dom78]

$$Z[\{l_{ik}\},\{\bar{l}_{ik}\}] = \left(\frac{1}{\pi}\right)^{NN_t} \left\langle \int \prod_{i=1}^{N}\prod_{k=0}^{N_t} dx_{ik} d\bar{x}_{ik} e^{-S} \exp(l_{ik}x_{ik} + \bar{l}_{ik}\bar{x}_{ik}) J \right\rangle,$$

where the *action* S is given by

$$S = \sum_{ik}[\frac{\sigma^2 \Delta t}{2}\bar{x}_{ik}^2 + i\bar{x}_{ik}\{x_{ik} - x_{i,k-1} - \Delta t(f_i(x_{i,k-1}) + \sum_j a_{ij} g(x_{i,k-1}, x_{j,k-1}))\}],$$

and $\langle \cdots \rangle$ represents the average over the ensemble of networks. J is the functional Jacobian term,

$$J = \exp\left(-\frac{\Delta t}{2} \sum_{ijk} \frac{\partial(f_i(x_{ik}) + a_{ij} g(x_{ik}, x_{jk}))}{\partial x_{ik}} \right).$$

Ichinomiya considered such a form of the network model (2.891) in which

$$a_{ij} = \begin{cases} 1 & \text{with probability } p_{ij}, \\ 0 & \text{with probability } 1 - p_{ij}. \end{cases}$$

Note that p_{ij} can be a function of variables such as i or j. For example, in the 1D chain model, p_{ij} is 1 if $|i - j| = 1$, else it is 0. The average over all networks can be expressed as

$$\left\langle \exp\left[\sum_{ik} i\Delta t \bar{x}_{ik} \sum_j a_{ij} g(x_{i,k-1}, x_{j,k-1}) \right] \right\rangle$$
$$= \prod_{ij} \left[p_{ij} \exp\left\{ \sum_k i\Delta t \bar{x}_{ik} g(x_{i,k-1}, x_{j,k-1}) \right\} + 1 - p_{ij} \right],$$

so we get

$$\langle e^{-S} \rangle = \exp(-S_0) \prod_{ij} \left[p_{ij} \exp\left\{ \sum_k i\Delta t \bar{x}_{ik} g(x_{i,k-1}, x_{j,k-1}) \right\} + 1 - p_{ij} \right],$$

where $S_0 = \sum_{ik} \frac{\sigma^2 \Delta t}{2} \bar{x}_{ik}^2 + i\bar{x}_{ik}\{x_{ik} - x_{i,k-1} - \Delta t f_i(x_{i,k-1})\}.$

This expression can be applied to the dynamics of any complex network model. [Ich05] applied this model to analysis of the *Kuramoto transition in random sparse networks*.

2.6.7 Complex Adaptive Systems: Common Characteristics

According to [AEH05], a *complex adaptive system* (CAS) consists of inhomogeneous, interacting adaptive agents, where the word *adaptive* means *capable of learning*. An *emergent property* of a CAS is a property of the system as a whole which does not exist at the individual elements (agents) level.

Typical CAS examples are the brain, the immune system, the economy, social systems, ecology, etc... Most of living systems are CAS.

Therefore to understand a complex system one has to study the system as a whole and not to decompose it into its constituents. This totalistic approach is against the standard reductionist one, which tries to decompose any system to its constituents and hopes that by understanding the elements one can understand the whole system.

The standard approaches to CAS modelling are: (i) ODEs, difference equations and PDEs; (ii) Cellular automata (CA) [Ila01]; (iii) Evolutionary game theory [HS98]; (iv) Various agent based models; (v) Complex networks (see previous subsection); and (vi) Fractional calculus [Sta00]. Most of these approaches are included in [Boc04].

Both the ODE and PDE approaches have some difficulties as follows [LSA03]: (i) ODE and PDE assumes that local fluctuations have been smoothed out, (ii) typically they neglect correlations between movements of different species, and (iii) they assume instantaneous results of interactions.

Most biological systems show delay and do not satisfy the above assumptions. They concluded that a cellular automata (CA) [Ila01] type system called microscopic simulation is more suitable to model complex biological systems. We agree that CA type systems are more suitable to model complex biological systems but such systems suffer from a main drawback namely the difficulty of getting analytical results. The known analytical results about CA type systems are very few compared to the known results about ODE and PDE.

A compromise was presented in [AEH05] in the form of a PDE, which avoids the delay and the correlations drawbacks. It is called *telegraph reaction diffusion equation* (TRD). To overcome the non–delay weakness in the *Fick law*,[41] it is replaced by

$$J(x,t) + \tau\, \partial_t J(x,t) = -D\, \partial_x c, \tag{2.892}$$

where the flux $J(x,t)$ relaxes, with some given characteristic time constant τ and c is the concentration of the diffusing substance. Combining (2.892) with the equation of continuity, one gets the *modified diffusion equation*, or the *telegraph equation*, $\dot{c} = (D - \tau)\, \partial_{x^2} c$. The corresponding TRD equaton is given by

$$\tau \ddot{c} + \left(1 - \frac{df(c)}{dc}\right) \dot{c} = D\, \partial_{x^2} c + f(c), \tag{2.893}$$

where $f(c)$ is a polynomial in c. Moreover it is known that TRD results from correlated random walk. This supports the conclusion that TRD equation (2.893) is more suitable for modeling complex systems than the usual diffusion one.

[41] Recall that *diffusion* through a fluid is can be described by the *Fick equation*

$$\partial_t T = D\, \nabla^2 c, \qquad D > 0,$$

where T is the temperature, c is the concentration of a certain substance dissolved in the fluid and D is the mass diffusivity.

For example, a *human immune system* as a CAS was elaborated in [SC01]. The emergent properties of the immune system (IS) included:

* The ability to distinguish any substance (typically called antigen Ag) and determine whether it is damaging or not. If Ag is non–damaging (damaging) then, typically, IS tolerates it (responds to it).
* If it decides to respond to it then IS determines whether to eradicate it or to contain it.
* The ability to memorize most previously encountered Ag, which enables it to mount a more effective reaction in any future encounters. This is the basis of vaccination processes.
* IS is complex thus it has a network structure.
* The immune network is not homogeneous since there are effectors with many connections and others with low number of connections.
* The Ag, which enters our bodies, has extremely wide diversity. Thus mechanisms have to exist to produce immune effectors with constantly changing random specificity to be able to recognize these Ag. Consequently IS is an adaptive complex system.
* Having said that, one should notice that the wide diversity of IS contains the danger of autoimmunity (attacking the body). Thus mechanisms that limit autoimmunity should exist.
* In addition to the primary clonal deletion mechanism, two further brilliant mechanisms exist: The first is that the IS network is a threshold or 'window' one i.e., no activation exists if the Ag quantity is too low or too high (This is called low and high zone tolerance).
* Thus an auto reactive immune effector (i.e., an immune effector that attacks the body to which it belongs) will face so many self–antigens that it has to be suppressed due to the high zone tolerance mechanism.
* Another mechanism against autoimmunity is the second signal given by antigen presenting cells (APC). If the immune effector is self reactive then, in most cases, it does not receive the second signal thus it becomes anergic.
* Also long term memory can be explained by the phenomena of high and low zone tolerance where IS tolerates Ag if its quantity is too high or too low. So persisting Ag is possible and continuous activation of immune effectors may occur.
* There is another possible explanation for long term memory using the immune system (extremal dynamics).
* Thus design principles of IS can explain important phenomena of IS.

The following Summary on CAS was given in [AEH05] (see also the literature cited there):

(i) CAS should be studied as a whole hence reductionist point of view may not be reliable in some cases.

(ii) CAS are open with nonlinear local interactions hence: (1) Long range prediction is highly unlikely. (2) When studying a CAS take into consideration the effects of its perturbation on related systems e.g. perturbation of lake Victoria has affected mosquitoes' numbers hence the locals quality of life. This is also relevant to the case of natural disasters where an earthquake at a city can cause a widespread power failure at other cities. (3) Expect side effects to any 'seemingly wise' decision. (4) Mathematical and computer models may be helpful in reducing such side effects.
(iii) Optimization in CAS should be multi–objective (not single objective).
(iv) CAS are very difficult to control. Interference at highly connected sites may be a useful approach. The interlinked nature of CAS elements complicates both the unpredictability and controllability problems. It also plays an important role in innovations spread.
(v) Memory effects should not be neglected in CAS. This lends more support for the proposed TRD equation (2.893). Also, memory games have been studied. Also, delay and fractional calculus are relevant to CAS.
(vi) Mathematical topics motivated by CAS include ODE and PDE (non–autonomous, delayed, periodic coefficients, stability and persistence), multi–objective optimization (including biologically motivated methods e.g., ant colony optimization, extremal optimization, genetic algorithms, etc.), difference equations, cellular automata, networks, fractional calculus, control (e.g., bounded delayed control of distributed systems), game theory, nonlinear dynamics and fuzzy mathematics.

2.6.8 FAM Functors and Real–Life Games

Recall that the *agent theory* concerns the definition of the so–called *belief–desire–intention agents* (BDI–agents, for short), as well as multi–agent systems, properties, architectures, communication, cooperation and coordination capabilities (see [RG98]). Its practical side concerns the agent languages and platforms for programming and experimenting with agents. According to [Fer99], a BDI agent is a physical or virtual entity which: (i) is capable of acting in an environment, (ii) can communicate directly with other agents, (iii) is driven by a set of tendencies (in the form of individual objectives or of a satisfaction/survival function which it tries to optimize), (iv) possesses resources of its own, which is capable of perceiving its environment (but to a limited extent), (v) has only a partial representation of its environment (and perhaps none at all), (vi) possesses skills and can offer services, (vii) may be able to reproduce itself, (viii) whose behavior tends towards satisfying its objectives, taking account of the resources and skills available to it and depending on its perception, its representation and the communications it receives. Agents' actions affect the environment which, in turn, affects future decisions of agents.

Multi–agent systems have already been successfully applied in numerous fields (see [Fer99] for the review). For example, the emerging field of *cellular*

robotics relates to building robots on modular basis: a robot (say a manipulator arm), is considered as a multi–agent system, and each of its constituent parts (arm segments) is regarded as an agent (with the joints describing constraints on the set of acceptable movements). The coordination of movement is thus a result of coordination involving an assembly of agents. It has been claimed that techniques similar to those used for modelling the multi–agent based eco–systems make it possible to accomplish complex movements with a minimum of computation. The head agent attempts to achieve the goal which has been set for it: if it can do it itself, it makes the movement and the system comes to a halt; if not, it brings in the neighboring agent, giving it goals which will help the head agent to get nearer to its objective. The process is repeated recursively, each agent trying to achieve the goals set for it by transmitting its desiderata to the neighboring agent.

This concept looks fine at the *low resolution level*. However, if we want a more sophisticated, high–resolution behavior of the arm manipulator, then we need to calculate a few derivatives of the nice and simple command 'Move'. This would give us proper robot kinematics: velocities and accelerations. In addition, if we want some more realism, we need to include inertial, gravity and friction forces, as well as to compensate them by driving forces. In this way we come to the proper dynamical system. Even further, to make it a proper control system, we need to add some feedbacks.

In this subsection we propose a *high–resolution agent model*, suitable for human performance modelling and real–life games.

Adaptive Fuzzy Associative Maps

Our high–resolution agent model is a flexible and adaptive functorial structure. Its basic unit is a generic nonlinear MIMO–System

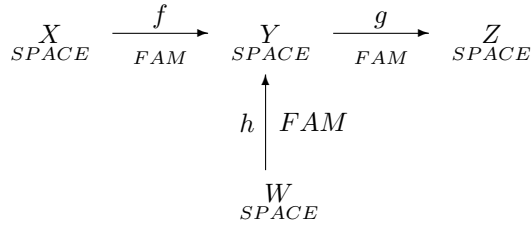
$$\underset{SPACE}{INPUT} \xrightarrow[FAM]{f} \underset{SPACE}{OUTPUT} \tag{2.894}$$

where $\underset{SPACE}{INPUT}$ and $\underset{SPACE}{OUTPUT}$ are nonlinear functional spaces consisting of a certain number of *adaptive fuzzy processes* (i.e., temporal fuzzy variables), partitioned by overlapping membership functions, which are either standard triangular/trapesoidal functions, or Gaussian functions of the form $\mu(z) = \exp\left[\frac{-(z-m)^2}{2\sigma^2}\right]$, where $z = z(t)$ denotes any fuzzy variable, while m and σ are its adjustable parameters, mean and standard deviation, respectively. The corresponding adaptive $n \times m-$dimensional map $\mathbb{R}^n \xrightarrow[FAM]{f} \mathbb{R}^m$ denotes the generalized *fuzzy associative memory* (FAM) (see [Kos92, Kos96]), that is, a (possibly sparse) matrix R^{ij} of fuzzy IF–THEN rules of the form:

IF x_1 is $\mu(x)_1^i$ AND ... x_n is $\mu(x)_n^i$ THEN y_1 is $\mu(y)_1^j$ AND ... y_m is $\mu(y)_m^j$

each with its associated *rule–importance weight* $w_{ij} \in [0, 1]$.

Given several FAM–maps of the form (2.894), we can perform their *composition* as well as *fusion* of the type



Using composition and fusion of FAM–maps, we can form hierarchical networks of nonlinear MIMO systems, for modelling human behavior real–life games.

Both the rule–weights and Gaussian parameters (means and standard deviations) are adjusted using two forms of unsupervised trial–and–error learning: *Hebbian learning*, minimizing the error [Heb49] and *reinforcement learning*, maximizing the reward [SB98]. We use both types of learning only as a minor parameter–adjusting tool, secondary to the primary fuzzy–logic *instruction*, setting–up the system's structure (i.e., FAM–map configuration). The learning is performed in discrete time steps, according to the general self–organized rule:

$$new\ value\ =\ old\ value\ +\ innovation$$

The local learning dynamics is performed as a combination of the two different ways:

1. Immediate, feedback–based, *Hebbian learning* rule [Heb49], occurring after each input–output system pass, comparing achieved output execution with the desired one and not considering the long–term rewards. Here the system learns its *optimal performance*:

   $$innovation = |desired\ output - achieved\ output|$$

   e.g., in case of the tennis game, this means the proper execution (i.e., precision, speed and spin) of the shot, without regard to the winning–points. It follows the simple tactic: 'Let me do my best shot, regardless of the opponent's actions'.
2. Long–term, reward–based, *reinforcement learning* rule [SB98], occurring only after winning points and remembering the rewards. Here the system learns its *optimal policy*:

   $$innovation = |reward - penalty|$$

   In the tennis case, this means only the winning–point, no matter what was the actual execution of the shot. It follows the competitive tactics: 'The goal is to beat the opponent, no matter how (within the rules)'.

Tennis Game Application

Now we will formulate a continuous–time *attack* (AT) and *counter–attack* (CA) model for the *tennis game*, with discrete–time learning. Tennis represents a competitive real–life situation within fully controlled conditions, depending on the individual's technical[42], tactical[43] and strategic[44] skills as well as operational psycho–physical strengths. Here, we will formulate an adaptive, fuzzy–dynamics AT & CA, tennis performance model. Using compositions and fusions of FAM–maps (as described above), we will design a generic simulator for counter–attack and attack performance dynamics, based on the tennis game. We have chosen the tennis game as it allows: (i) complete analysis in every detail, (ii) full experimental validation, and (iii) training both in real–life situations and in a machine–learning environment.

Attack Model: Tennis Serve

A. Simple Attack: Serve Only. The simple AT–dynamics is represented by a single FAM map

$$\underset{SPACE}{TARGET} \xrightarrow[FAM]{f^{AT}} \underset{SPACE}{ATTACK}$$

In the case of simple tennis serve, this AT–scenario reads

$$\underset{OPPONENT-IN}{O \ni o_m} \xrightarrow{f^{AT}} \underset{SERVE-OUT}{SR \ni sr_n}$$

where the two functional spaces, $O_{\dim=2} \ni o_m$ and $SR_{\dim=3} \ni sr_n$, contain the fuzzy variables $\{o_m = o_m(t)\}$ and $\{sr_n = sr_n(t)\}$, respectively opponent–related (target information) and serve–related, partitioned by overlapping Gaussians, $\mu(z) = \exp\left[\frac{-(z-m)^2}{2\sigma^2}\right]$, and defined as:

$$\underset{OPPONENT-IN}{O}: \begin{array}{l} o_1 = Opp.Posit.Left.Right : (center, medium, wide), \\ o_2 = Opp.Antcp.Lft.Rght : (runCenter, stay, runWide), \end{array}$$

$$\underset{SERVE-OUT}{SR}: \begin{array}{l} sr_1 = 1.Serve.Speed : (low, medium, high) \\ sr_2 = 2.Serve.Spin : (low, medium, high) \\ sr_3 = 3.Serve.Placement : (center, medium, wide) \end{array}$$

[42] *Technical skills* in sports are judged by the *biomechanical correctness of movements.*

[43] *Tactical skills* in sports include 'winners', 'changes of tempo', and 'applying pressure'.

[44] *Strategic skills* in sports include 'planning ahead' and 'weighting the relative importance of points' along the match.

In the fuzzy–matrix form this simple serve reads

$$\begin{bmatrix} {\scriptstyle O:\,OPPONENT-IN} \\ o_1 = Opp.Posit.Left.Right \\ o_2 = Opp.Anticip.Left.Right \end{bmatrix} \xrightarrow{f^{AT}} \begin{bmatrix} {\scriptstyle SR:\,SERVE-OUT} \\ sr_1 = 1.Serve.Speed \\ sr_2 = 2.Serve.Spin \\ sr_3 = 3.Serve.Place \end{bmatrix}$$

B. Attack–Maneuver: Serve–Volley. The generic advanced AT–dynamics is given by a composition of FAM maps

$$\underset{SPACE}{TARGET} \xrightarrow[FAM]{f^{AT}} \underset{SPACE}{ATTACK} \xrightarrow[FAM]{g^{AT}} \underset{SPACE}{MANEUVER}$$

In the case of advanced tennis serve, this AT–scenario reads

$$\underset{OPPONENT-IN}{O \ni o_m} \xrightarrow{f^{AT}} \underset{SERVE-OUT}{SR \ni sr_n} \xrightarrow{g^{AT}} \underset{RUN-VOLEY}{RV \ni rv_p}$$

where the new functional space, $RV_{\dim=2} \ni rv_p$, contains the opponent–anticipation driven volley–maneuver, expressed by fuzzy variables $\{rv_p = rv_p(t)\}$, partitioned by overlapping Gaussians and given by:

$$\underset{RUN-VOLEY}{RV} : \begin{matrix} rv_1 = RV.For : (baseLine, center, netClose) \\ rv_2 = RV.L.R. : (left, center, right) \end{matrix}$$

In the fuzzy–matrix form this advanced serve reads

$$\begin{bmatrix} {\scriptstyle O:\,OPPONENT-IN} \\ o_1 = Opp.Posit.L.R. \\ o_2 = Opp.Anticip.L.R. \end{bmatrix} \xrightarrow{f^{AT}} \begin{bmatrix} {\scriptstyle SR:\,SERVE-OUT} \\ sr_1 = 1.Serve.Speed \\ sr_2 = 2.Serve.Spin \\ sr_3 = 3.Serve.Place \end{bmatrix} \xrightarrow{g^{AT}} \begin{bmatrix} {\scriptstyle RV:\,RUN-VOLEY} \\ rv_1 = RV.For \\ rv_2 = RV.L.R \end{bmatrix}$$

Counter–Attack Model: Tennis Return

A. Simple Return. The simple CA–dynamics reads:

$$\underset{SPACE}{ATTACK} \xrightarrow[FAM]{f^{CA}} \underset{SPACE}{MANEUVER} \xrightarrow[FAM]{g^{CA}} \underset{SPACE}{RESPONSE}$$

In the case of simple tennis return, this CA–scenario consists purely of conditioned–reflex reaction, no decision process is involved, so it reads:

$$\underset{BALL-IN}{B \ni b_i} \xrightarrow{f^{CA}} \underset{RUNNING}{R \ni r_j} \xrightarrow{g^{CA}} \underset{SHOT-OUT}{S \ni s_k}$$

where the functional spaces $B_{\dim=5} \ni b_i$, $R_{\dim=3} \ni r_j$, $S_{\dim=4} \ni s_k$, contain the fuzzy variables $\{b_i = b_i(t)\}$, $\{r_j = r_j(t)\}$ and $\{s_k = s_k(t)\}$, respectively

defining the ball inputs, our player's running maneuver and his shot–response, i.e.,

$$\overset{B:\,BALL-IN}{\begin{bmatrix} b_1 = Dist.L.R. \\ b_2 = Dist.F.B. \\ b_3 = Dist.Vert \\ b_4 = Speed \\ b_5 = Spin \end{bmatrix}} \xrightarrow{f^{CA}} \overset{R:\,RUNNING}{\begin{bmatrix} r_1 = Run.L.R. \\ r_2 = Run.F.B. \\ r_3 = Run.Vert \end{bmatrix}} \xrightarrow{g^{CA}} \overset{S:\,SHOT-OUT}{\begin{bmatrix} s_1 = Backhand \\ s_2 = Forehand \\ s_3 = Voley \\ s_4 = Smash \end{bmatrix}}$$

Here, the existence of efficient weapons within the $\underset{SHOT-OUT}{S}$ arsenal–space, namely $s_k(t) : s_1 = Backhand, s_2 = Forehand, s_3 = Voley$ and $s_4 = Smash$, is assumed.

The universes of discourse for the fuzzy variables $\{b_i(t)\}$, $\{r_j(t)\}$ and $\{s_k(t)\}$, partitioned by overlapping Gaussians, are defined respectively as:

$$\underset{BALL-IN}{B}: \begin{array}{l} b_1 = Dist.L.R. : (veryLeft, left, center, right, veryRight), \\ b_2 = Dist.F.B. : (baseLine, center, netClose), \\ b_3 = Dist.Vert : (low, medium, high), \\ b_4 = Speed : (low, medium, high), \\ b_5 = Spin : (highTopSpin, lowTopSpin, flat, \\ \qquad\qquad lowBackSpin, highBackSpin). \end{array}$$

$$\underset{RUNNING}{R}: \begin{array}{l} r_1 = Run.L.R. : (veryLeft, left, center, right, veryRight), \\ r_2 = Run.F.B. : (closeFront, front, center, back, farBack), \\ r_3 = Run.Vert : (squat, normal, jump). \end{array}$$

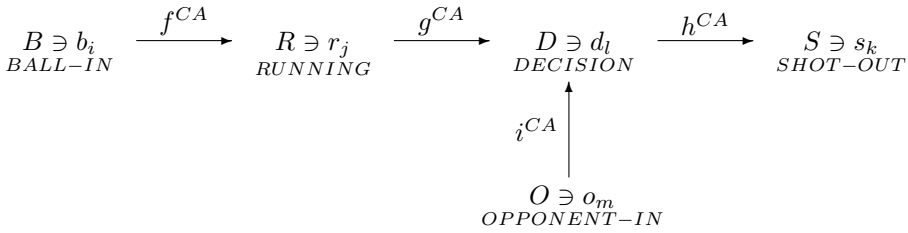
$$\underset{SHOT-OUT}{S}: \begin{array}{l} s_1 = Backhand : (low, medium, high), \\ s_2 = Forehand : (low, medium, high), \\ s_3 = Voley : (backhand, block, forehand), \\ s_4 = Smash : (low, medium, high). \end{array}$$

B. Advanced Return. The advanced CA–dynamics includes both the information about the opponent and (either conscious or subconscious) decision making. This generic CA–scenario is formulated as the following composition + fusion of FAM maps:

$$\underset{SPACE}{ATTACK} \xrightarrow[FAM]{f^{CA}} \underset{SPACE}{MANEUV} \xrightarrow[FAM]{g^{CA}} \underset{SPACE}{DECISION} \xrightarrow[FAM]{h^{CA}} \underset{SPACE}{RESPON}$$

$$\underset{SPACE}{DECISION} \uparrow {\scriptstyle i^{CA}\ FAM} \quad \underset{SPACE}{TARGET}$$

where we have added two new functional spaces, $\underset{SPACE}{TARGET}$ and $\underset{SPACE}{DECISION}$, respectively containing information about the opponent as a target, as well as our own aiming decision processes. In the case of advanced tennis return, this reads:



where the two additional functional spaces, $O_{\dim=4} \ni o_m$ and $D_{\dim=5} \ni d_l$, contain the fuzzy variables $\{o_m = o_m(t)\}$ and $\{d_l = d_l(t)\}$, respectively defining the opponent–related target information and the aim–related decision processes, both partitioned by overlapping Gaussians and defined as:

$$\underset{OPPONENT-IN}{O}: \begin{array}{l} o_1 = Opp.Posit.L.R.:(left, center, right), \\ o_2 = Opp.Posit.F.B.:(netClose, center, baseLine), \\ o_3 = Opp.Anticip.L.R.:(runLeft, stay, runRight), \\ o_4 = Opp.Anticip.F.B.:(runNet, stay, runBase). \end{array}$$

$$\underset{DECISION}{D}: \begin{array}{l} d_1 = Aim.L.R.:(left, center, right), \\ d_2 = Aim.F.B.:(netClose, center, baseLine), \\ d_3 = Aim.Vert:(low, medium, high), \\ d_4 = Aim.Speed:(low, medium, high), \\ d_5 = Aim.Spin:(highTopSpin, lowTopSpin, noSpin, \\ \qquad lowBackSpin, highBackSpin). \end{array}$$

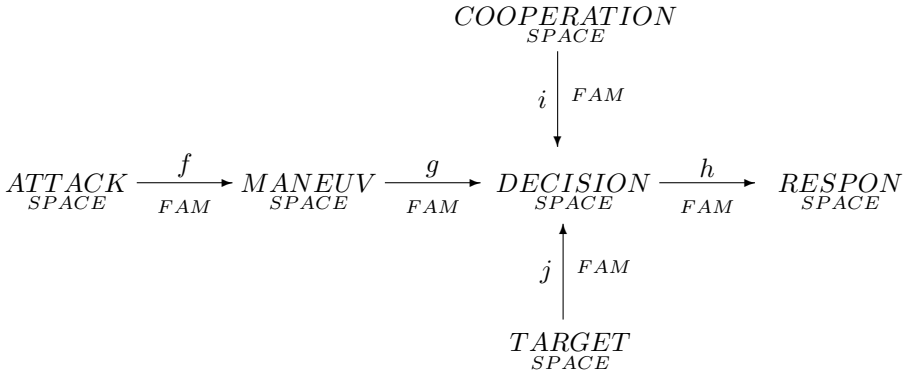
The corresponding fuzzy–matrices read:

$$\overset{B:\,BALL-IN}{\begin{bmatrix} b_1 = Dist.L.R. \\ b_2 = Dist.F.B. \\ b_3 = Dist.Vert \\ b_4 = Speed \\ b_5 = Spin \end{bmatrix}}, \quad \overset{R:\,RUNNING}{\begin{bmatrix} r_1 = Run.L.R. \\ r_2 = Run.F.B. \\ r_3 = Run.Vert \end{bmatrix}}, \quad \overset{D:\,DECISION}{\begin{bmatrix} d_1 = Aim.L.R. \\ d_2 = Aim.F.B. \\ d_3 = Aim.Vert \\ d_4 = Aim.Speed \\ d_5 = Aim.Spin \end{bmatrix}},$$

$$\overset{O:\,OPPONENT-IN}{\begin{bmatrix} o_1 = Opp.Posit.L.R. \\ o_2 = Opp.Posit.F.B. \\ o_3 = Opp.Anticip.L.R. \\ o_4 = Opp.Anticip.F.B. \end{bmatrix}}, \quad \overset{S:\,SHOT-OUT}{\begin{bmatrix} s_1 = Backhand \\ s_2 = Forehand \\ s_3 = Voley \\ s_4 = Smash \end{bmatrix}}.$$

Generic FAM–Agents

Finally, we briefly describe a generic *FAM–agent* as a following complex adaptive system, designed from five adaptive FAM–maps and six hybrid functional spaces:



Each generic FAM–agent is supposed to communicate with other FAM–agents and/or environment only by its inputs and outputs. It is both a feedback control system and an expert system, implementing: (i) feedbacks, (ii) instructions and (iii) common sense. It is a high–resolution human performance predictor–corrector system, which has to be designed in each specific case by consulting an expert from the field.

A *smart multi–agent organization* represents a brain–like sensory–motor system–of–systems, including hierarchical networks of generic FAM–agents, with the unique input and output.

2.6.9 Riemann–Finsler Approach to Information Geometry

Model Specification and Parameter Estimation

Model as a Parametric Family of Probability Distributions

From a statistical standpoint, observed data is a random sample from an unknown population. Ideally, the goal of modelling is to deduce the population that generated the observed data. Formally, a model is defined as a parametric family of probability distributions (see [MP03]).

Let us use $f(y|w)$ to denote the *probability distribution function* that gives the probability of observing *data*, $y = (y_1, \ldots, y_m)$, given the *model's parameter vector*, $w = (w_1, \ldots, w_k)$. Under the assumption that individual observations, y_i's, are independent of one another, $f(y|w)$ can be rewritten as a product of individual probability distribution functions,

$$f\left(y = (y_1, ..., y_m)\,|w\right) = f\left(y_1|w\right)\; f\left(y_2|w\right) ... \, f\left(y_m|w\right). \tag{2.895}$$

Parameter Estimation

Once a model is specified with its parameters and data have been collected, the model's ability to fit the data can be assessed. Model fit is measured by finding parameter values of the model that give the 'best' fit to the data in some defined sense – a procedure called *parameter estimation* in statistics.

There are two generally accepted methods of parameter estimation: *least–squares estimation* (LSE) and *maximum likelihood estimation* (MLE). In LSE, we seek the parameter values that *minimize the sum of squares error* (SSE) between observed data and a model's predictions:

$$SSE(w) = \sum_{i=1}^{m} \left(y_i - y_{i,prd}\,(w)\right)^2,$$

where $y_{i,prd}\,(w)$ denotes the model's prediction for observation y_i. In MLE, we seek the parameter values that are most likely to have produced the data. This is obtained by *maximizing* the *log–likelihood* of the observed data:

$$\mathbf{loglik}\,(w) = \sum_{i=1}^{m} \ln f(y_i|w).$$

By maximizing either the likelihood or the log–likelihood, the same solution is obtained because the two are monotonically related to each other. In practice, the log–likelihood is preferred for computational ease. The parameters that minimize the sum of squares error or the log–likelihood are called the LSE or MLE *estimates*, respectively.

For normally distributed data with constant variance, LSE and MLE are equivalent in the sense that both methods yield the same parameter estimates. For non–normal data such as proportions and response times, however, LSE estimates tend to differ from MLE estimates. Although LSE is often the 'de facto' method of estimation in cognitive psychology, MLE is a preferred method of estimation in statistics, especially for non–normal data. In particular, MLE is well–suited for statistical inference in hypothesis testing and model selection. Finding LSE or MLE estimates generally requires use of a numerical optimization procedure.

Model Evaluation and Testing

Qualitative Criteria

A model satisfies the *explanatory adequacy criterion* if its assumptions are plausible and consistent with established findings, and importantly, the theoretical account is reasonable for the cognitive process of interest. In other words, the model must be able to do more than redescribe observed data. The model must also be interpretable in the sense that the model makes sense and

is understandable. Importantly, the components of the model, especially, its parameters, must be linked to psychological processes and constructs. Finally, the model is said to be faithful to the extent that the model's ability to capture the underlying mental process originates from the theoretical principles embodied in the model, rather than from the choices made in its computational instantiation.

Quantitative Criteria

Falsifiability

This is a *necessary condition* for testing a model or theory, refers to *whether there exist potential observations that a model cannot describe* [Pop59]. If so, then the model is said to be falsifiable. An unfalsifiable model is one that can describe all possible data patterns in a given experimental situation. *There is no point in testing an unfalsifiable model.*

A heuristic rule for determining a model's falsifiability is already familiar to us: The model is falsifiable if the number of its free parameters is less than the number of data observations. This *counting rule*, however, turns out to be imperfect, in particular, for nonlinear models. To remedy limitations of the counting rule, [BS85] provided a *formal rule* for assessing a model's falsifiability, which yielded the counting rule as a special case. The rule states that a model is falsifiable if the *rank of its Jacobian matrix* is less than the number of data observations for all values of the parameters. Recall that the Jacobian matrix is defined in terms of partial derivatives as: $J_{ij}(w) = \partial E(y_j)/\partial w_i$ $(i = 1, ..., k; j = 1, ..., m)$ where $E(x)$ stands for the *expectation* of a random variable x.

Goodness of Fit

A model should also give a good description of the observed data. Goodness of fit refers to the *model's ability to fit* the particular set of observed data. Common examples of *goodness of fit measures* are the minimized sum of squares error (SSE), the mean squared error (MSE), the root mean squared error (RMSE), the percent variance accounted for (PVAF), and the maximum likelihood (ML).

The first four of these, defined below, are related to one another in a way that one can be written in terms of another:

$$\begin{aligned} MSE &= SSE\left(w^*_{LSE}\right)/m, \\ RMSE &= \sqrt{SSE\left(w^*_{LSE}\right)/m}, \\ PVAF &= 100\left(1 - SSE\left(w^*_{LSE}\right)/SST\right), \\ ML &= f\left(y|w^*_{MLE}\right). \end{aligned}$$

Here w^*_{LSE} is the parameter that minimizes $SSE(w)$, that is, an LSE estimate, and SST stands for the sum of squares total defined as $SST = \sum_i (y_i - y_{mean})^2$. ML is the probability distribution function maximized with respect to the model's parameters, evaluated at w^*_{MLE}, which is obtained through MLE.

Complexity

Not only should a model describe the data in hand well, but it should also do so in the least complex (i.e., simplest) way. Intuitively, complexity has to do with a model's inherent flexibility that enables it to fit a wide range of data patterns. There seem to be at least *two dimensions of model complexity*, the *number of parameters* and the model's *functional form*. The latter refers to the way the parameters are combined in the model equation. The more parameters a model has, the more complex it is. Importantly also, two models with the same number of parameters but different functional forms can differ significantly in their complexity. For example, it seems unlikely that two one–parameter models, $y = x + w$ and $y = e^{wx}$ are equally complex. The latter is probably much better at fitting data than the former.

It turns out that one can devise a quantitative measure of model complexity that takes into account both dimensions of complexity and at the same time is theoretically justified as well as intuitive. One example is the *geometric complexity* (GC) of a model [PMZ02] defined as:

$$GC = \frac{k}{2} \ln \frac{n}{2\pi} + \ln \int dw \sqrt{\det I(w)}, \tag{2.896}$$

or

$$GC = \text{parametric complexity} + \text{functional complexity},$$

where k is the number of parameters, n is the sample size, $I(w)$ is the *Fisher information matrix* (or, *covariance matrix*) defined as

$$I_{ij}(w) = -E\left[\partial^2 \ln f(y|w)/\partial w_i \partial w_j\right], i, j = 1, ..., k, \tag{2.897}$$

or

$$I_{ij} = -\text{Expect.Value}(\text{Hessian}(\mathbf{loglik}(w))).$$

Functional form effects of complexity are reflected in the second term of GC through $I(w)$. How do we interpret geometric complexity? The meaning of geometric complexity is related to the *number of 'different'* (i.e., distinguishable) *probability distributions* that a model can account for. The more distinguishable distributions that the model can describe by finely tuning its parameter values, the more complex it is ([MBP00]). For example, when geometric complexity is calculated for the following two-parameter psychophysical models, *Stevens' law* ($y = w_1 x^{w_2}$) and *Fechner's logarithmic law* ($y = w_1 \ln(x + w_2)$), the former turns out to be more complex than the latter [PMZ02].

Generalizability

The fourth quantitative criterion for model evaluation is generalizability. This criterion is defined as a *model's ability to fit not only the observed data at hand, but also new, as yet unseen data samples from the same probability distribution*. In other words, model evaluation should not be focused solely on how well a model fits observed data, but how well it fits data generated by the cognitive process underlying the data. This goal will be achieved best when generalizability is considered.

To summarize, these four quantitative criteria work together to assist in model evaluation and guide (even constrain) model development and selection. The model must be sufficiently complex, but not too complex, to capture the regularity in the data. Both a good fit to the data and good generalizability will ensure an appropriate degree of complexity, so that the model captures the regularity in the data. In addition, because of its broad focus, generalizability will constrain the power of the model, thus making it falsifiable. Although all four criteria are inter-related, generalizability may be the most important. By making it the guiding principle in model evaluation and selection, one cannot go wrong.

Selection Among Different Models

Since a model's generalizability is not directly observable, it must be estimated using observed data. The measure developed for this purpose trades off a model's fit to the data with its complexity, the aim being to select the model that is complex enough to capture the regularity in the data, but not overly complex to capture the ever-present random variation. Looked at in this way, generalizability embodies the principle of *Occam's razor* (or *principle of parsimony*, i.e., the requirement of maximal simplicity of cognitive models).

Model Selection Methods

Now, we describe specific measures of generalizability. Four representative *generalizability criteria* are introduced. They are the *Akaike Information Criterion* (AIC), the *Bayesian Information Criterion* (BIC), *crossvalidation* (CV), and *minimum description length* (MDL) (see a special Journal of Mathematical Psychology issue on *model selection*, in particular [MFB00]). In all four methods, the *maximized log–likelihood* is used as a *goodness–of–fit measure*, but they differ in how model complexity is conceptualized and measured.

AIC and BIC

AIC and BIC for a given model are defined as follows:

$$AIC = -2 \ln f(y|w^*) + 2k,$$
$$BIC = -2 \ln f(y|w^*) + k \ln n,$$

where w^* is a MLE estimate, k is the number of parameters and n is the sample size. For normally distributed errors with constant variance, the first term of both criteria, $-2\ln f(y|w^*)$, is reduced to $(n{\cdot}\ln(SSE(w^*)) + c_o)$ where c_o is a constant that does not depend upon the model. In each criterion, the first term represents a lack of fit measure, the second term represents a complexity measure, and together they represent a lack of generalizability measure. A lower value of the criterion means better generalizability. Therefore, the model that minimizes a given criterion should be chosen.

Complexity in AIC and BIC is a *function* of *only* the *number of parameters.* Functional form, another important dimension of model complexity, is not considered. For this reason, these methods are not recommended for comparing models with the same number of parameters but different functional forms. The other two selection methods, CV and MDL, described next, are sensitive to functional form as well as the number of parameters.

Cross–Validation

In CV, a model's generalizability is estimated without defining an explicit measure of complexity. Instead, models with more complexity than necessary to capture the regularity in the data are penalized through a *resampling procedure*, which is performed as follows: The observed data sample is divided into two sub–samples, *calibration* and *validation*. The calibration sample is then used to find the best–fitting values of a model's parameters by MLE or LSE. These values, denoted by w^*_{cal}, are then fixed and fitted, without any further tuning of the parameters, to the validation sample, denoted by y_{val}. The resulting fit to y_{val} by w^*_{cal} is called as the model's CV index and is taken as the model's generalizability estimate. If desired, this single–division–based CV index may be replaced by the average CV index calculated from multiple divisions of calibration and validation samples. The latter is a more accurate estimate of the model's generalizability, though it is also more computationally demanding.

The main attraction of cross–validation is its *ease of use*. All that is needed is a simple resampling routine that can easily be programmed on any desktop computer. The second attraction is that unlike AIC and BIC, CV is sensitive to the functional form dimension of model complexity, though how it works is unclear because of the implicit nature of the method. For these reasons, the method can be used in all modelling situations, including the case of comparing among models that differ in functional form but have the same number of parameters.

Minimum Description Length

MDL is a selection method that has its origin in *algorithmic coding theory* in computer science. According to MDL, both models and data are viewed as codes that can be compressed. The basic idea of this approach is that regularities in data necessarily imply the existence of *statistical redundancy*

and that the redundancy can be used to compress the data [Gru99, Gru00, GMP05]. Put another way, the amount of *regularity in data* is directly related to the *data description length*. The shorter the description of the data by the model, the better the approximation of the underlying regularity, and thus, the higher the model's generalizability is. Formally, MDL is defined as:

$$MDL = -\ln f(y|w^*) + \frac{k}{2}\ln\frac{n}{2\pi} + \ln\int dw\sqrt{\det I(w)}, \tag{2.898}$$

or

$$MDL = \text{lack–of–fit measure} + \text{param. complexity} + \text{functional complexity},$$

where the first term is the same lack of fit measure as in AIC and BIC; the second and third terms together represent the geometric complexity measure (2.896). In coding theory, MDL is interpreted as the length in bits of the shortest possible code that describes the data unambiguously with the help of a model. The *model with the minimum value of MDL* encodes the most regularity in the data, and therefore should be preferred.

The second term in (2.898), which captures the effects of model complexity due to the number of parameter (k), is a logarithmic function of sample size n. In contrast, the third term, which captures functional form effects, is not sensitive to sample size. This means that as sample size increases, the relative contribution of the effects due to functional form to those due to the number of parameters will be gradually reduced. Therefore, *functional form effects can be ignored for sufficiently large* n, in which case the MDL value becomes approximately equal to one half of the BIC value.

Probably the most desirable property of MDL over other selection methods is that its complexity measure takes into account the effects of both dimensions of model complexity, the number of parameters and functional form. The MDL complexity measure, unlike CV, shows explicitly how both factors contribute to model complexity. In short, *MDL is a sharper and more accurate method* than these three competitors. The price that is paid for MDL's superior performance is its *computational cost*. MDL can be laborious to calculate. First, the Fisher information matrix (2.897) must be obtained by calculating the second derivatives (i.e., Hessian matrix) of the log–likelihood function, $\ln f(y|w)$. This calculation can be non–trivial, though not impossible. Second, the square–root of the determinant of the Fisher information matrix must be integrated over the parameter space. This generally requires use of a numerical integration method such as *Markov Chain Monte Carlo* (see e.g., [GRS96]).

Riemannian Geometry of Minimum Description Length

From a geometric perspective, a *parametric model family of probability distributions* (2.895) forms a *Riemannian manifold embedded in the space of all probability distributions* (see [Ama85, AN00]). *Every distribution is a point in this space*, and the *collection of points created by varying the parameters*

of the model induces a manifold in which 'similar' distributions are mapped to 'nearby' points. The *infinitesimal distance between points* separated by the infinitesimal parameter differences dw^i is given by

$$ds^2 = g_{ij}(w)\, dw^i dw^j,$$

where $g_{ij}(w)$ is the *Riemannian metric tensor*. The Fisher information, $I_{ij}(w)$, defined by (2.897), is the natural metric on a manifold of distributions in the context of statistical inference [Ama85]. We argue that the MDL measure of model fitness has an attractive interpretation in such a geometric context.

The first term in MDL equation (2.898) estimates the accuracy of the model since the likelihood $f(y|w^*)$ measures the ability of the model to fit the observed data. The second and third terms are supposed to penalize model complexity; we will show that they have interesting geometric interpretations. Given the metric $I_{ij}(w) = g_{ij}(w)$ on the space of parameters, the *infinitesimal volume element* on the parameter manifold is

$$dV = dw\sqrt{\det I(w)} = \prod_{l=1}^{k} dw^l \sqrt{\det I(w)}.$$

The *Riemannian volume of the parameter manifold* is obtained by integrating dV over the space of parameters:

$$V_M = \int dV = \int dw\sqrt{\det I(w)}.$$

In other words, the *third term* (*functional complexity*) *in MDL penalizes models that occupy a large volume in the space of distributions.*

In fact, the volume measure V_M is related to the number of 'distinguishable' probability distributions indexed by the model. Because of the way the model family is embedded in the space of distributions, two different parameter values can index very similar distributions. If complexity is related to volumes occupied by model manifolds, the measure of volume should count only different, or distinguishable, distributions, and not the artificial coordinate volume. It is shown in [MBP00] that the volume V_M achieves this goal.

Selecting Among Qualitative Models

Application of any of the preceding selection methods requires that the models are *quantitative models*, each defined as a *parametric family of probability distributions.*

Pseudo–probabilistic MDL Approach

The 'pseudo–probabilistic' approach [Gru99] for selecting among qualitative models derives a selection criterion that is similar to the MDL criterion for

quantitative models, but it is a formulation that is closer to the original spirit of the MDL principle, which states:

'Given a data set D and a model M, the description length of the data, $DL_M(D)$, is given by the sum of (a) the description length of the data when encoded with help of the model, $DL(D|M)$, and (b) the description length of the model itself, $DL(M) : DL_M(D) = DL(D|M) + DL(M)$. Among a set of competing models, the best model is the one that minimizes $DLM(D)$.'

The above MDL principle is broad enough to include the MDL criterion for quantitative models as a specific instantiation. The first, lack–of–fit term of the quantitative criterion $(-\ln f(y|w^*))$ can be seen as $DL(D|M)$, whereas the second and third terms $(\frac{k}{2}\ln\frac{n}{2\pi} + \ln\int dw\sqrt{\det I(w)})$ represent geometric complexity as DL(M). Likewise, a computable criterion that implements the above principle can be obtained with the pseudo–probabilistic approach. It is derived from the *Kraft–Inequality theorem* in coding theory ([LV97]). The theorem proves that one can always associate arbitrary models with their 'equivalent' probability distributions in a procedure called *entropification* [Gru99].

MDL Criterion for Qualitative Models

Entropification proceeds as follows. We first 'construct' a parametric family of probability distributions for a given qualitative model in the following form:

$$p(y = (y_1, ..., y_m)|w) = \exp\left(-w\sum_{i=1}^{m} Err(y_{i,obs} - y_{i,prd}(w))\right)/Z(w).$$

In this equation, $Err(x)$ is an error function that measures the model's prediction performance such as $Err(x) = |x|$ or x^2, w is a scalar parameter, and $Z(w)$ is the normalizing factor defined as

$$Z(w) = \sum_{y_1} ... \sum_{y_m} p(y = (y_1, ..., y_m)|w).$$

The above formulation requires that each observation y_i be represented by a discrete variable that takes on a finite number of possible values representing the model's qualitative (e.g., ordinal) predictions.

Once a suitable error function, $Err(x)$, is chosen, the above probability distribution function is then used to fit observed data, and the best–fitting parameter w^* is sought by MLE. The description length of the data encoded with the help of the model is then obtained by taking the minus logarithm of the maximum likelihood (ML),

$$DL(D|M) = -\ln p(y|w^*).$$

The second term, $DL(M)$, the description length of the model itself, is obtained by counting the number of different data patterns the model can account for and then taking the logarithm of the resulting number. Putting these together, the desired MDL criterion for a qualitative model is given by

$$MDL_{qual} = w^* \sum_{i=1}^{m} Err\left(y_{i,obs} - y_{i,prd}\left(w^*\right)\right) + \ln Z\left(w^*\right) + \ln N,$$

where N is the number of all possible data patterns or data sets that the model predicts.

Finsler Approach to Information Geometry

Recall that information geometry has emerged from investigating the geometrical structure of a *family of probability distributions*, and has been applied successfully to various areas including statistical inference, control theory and multi–terminal information theory (see [Ama85, AN00]). In this subsection we give a brief review on a more general approach to information geometry, based on Finsler geometry (see subsection 1.2.10 above).

A *parameter–space of probability distributions*, defined by

$$M = \{x : p = p(r, x) \text{ is a probability distribution on } \mathbb{R}\}$$

represents a smooth manifold, called the *probability manifold* [Ama85, AN00]. On a probability manifold M we can define a *probability divergence* $D = D(x, y)$, as

$$D(x,y) = \int_M p(r,x) f(\frac{p(r,y)}{p(r,x)})\, dr,$$
$$\text{where } f(\cdot) \text{ is a convex function such that } f(1) = 0, \quad f''(1) = 1,$$

which satisfies the following conditions [She05]

$$\begin{aligned} &D(x,y) > 0 \qquad \text{if} \qquad x \neq y, \\ &D(x,y) = 0 \qquad \text{if} \qquad x = y, \\ &D(x,y) \neq D(y,x) \qquad \text{in general.} \end{aligned}$$

On the other hand, if $d = d(x, y)$ is a *probability distance* on a probability manifold M, satisfying the following standard conditions:

$$\begin{aligned} &d(x,x) = 0, \\ &d(x,y) > 0, \qquad \text{if} \qquad x \neq y, \\ &d(x,y) \leq d(x,r) + d(r,y) \qquad \text{(triangle inequality)}, \end{aligned}$$

then for any function $\psi = \psi(h)$ with $\psi(0) = 0, \psi(h) > 0$ for $h > 0$, the probability divergence on M is defined as

$$D(x,y) := \psi(d(x,y)). \tag{2.899}$$

Recall that a *Finsler metric* $L = L(x, y)$ is a function of tangent vectors y at a point $x \in M$, with the following properties:

$$\begin{aligned} L(x,ty) &= t^2 L(x,y) \qquad \text{for } \ t>0, \\ g_{ij}(x,y) &:= \frac{1}{2}\frac{\partial^2 L}{\partial y^i \partial y^j}(x,y) > 0, \\ F_x(y) &:= \sqrt{L(x,y)}, \qquad F_x(u+v) \ \le \ F_x(u) + F_x(v). \end{aligned} \tag{2.900}$$

This means that there is an inner product g_y at a pint $x \in M$, such that

$$g_y(u,v) = g_{ij}(x,y)u^i v^j,$$

so that our Finsler metric $L(x,y) \in M$, given by (2.900), becomes

$$L(x,y) = g_y(u,v) = g_{ij}(x)y^i y^j.$$

Therefore, in a special case when $g_{ij}(x,y) = g_{ij}(x)$ are independent of y, the Finsler metric $L(x,y)$ becomes a standard Riemannian metric $g_{ij}(x)y^i y^j$. In this way, all the material from the previous subsection can be generalized to Finsler geometry.

Now, $D(x,y) \in M$, given by (2.899), is called the *regular divergence*, if

$$2D(x,x+y) = L(x,y) + \frac{1}{2}L_{x^k}(x,y)y^k + \frac{1}{3}H(x,y) + o(|y|^3),$$

where $H = H(x,y) \in M$ is homogenous function of degree 3 in y, i.e.,

$$H(x,ty) = t^3 H(x,y) \qquad \text{for} \quad t>0.$$

A pair $\{L,H\} \in M$ is called a *Finsler information structure* [She05].

In a particular case when $L(x,y) = g_{ij}(x)y^i y^j$ is a Riemannian metric, and $H(x,y) = H_{ijk}(x)y^i y^j y^k$ is a polynomial, then we have *affine information structure* $\{L,H\} \in M$, which is described by a family of affine connections, called $\alpha-$connections by [Ama85, AN00]. However, in general, the induced information structure $\{L,H\} \in M$ is not affine, i.e., $L(x,y)$ is not Riemannian and $H(x,y)$ is not polynomial.

3

Appendix: Tensors and Functors

3.1 Elements of Classical Tensor Analysis

Physical and engineering laws must be independent of any particular coordinate systems used in describing them mathematically, if they are to be valid. In other words, all physical and engineering equations need to be tensorial or *covariant*. Therefore, for the reference purpose, in this subsection, we give the basic formulas from the standard tensor calculus, which is used throughout the text. The basic notational convention used in tensor calculus is *Einstein's summation convention* over repeated indices. More on this subject can be found in any standard textbook on mathematical methods for scientists and engineers, or mathematical physics (we recommend [MTW73]).

3.1.1 Transformation of Coordinates and Elementary Tensors

To introduce tensors, consider a standard linear nD matrix system, $\mathbf{Ax} = \mathbf{b}$. It can be rewritten in the so–called *covariant form* as

$$a_{ij}x^j = b_i\,, \qquad (i, j = 1, ..., n). \tag{3.1}$$

Here, i is a *free index* and j is a *dummy index* to be summed upon, so the expansion of (3.1) gives

$$\begin{aligned} a_{11}x^1 + a_{12}x^2 + ... + a_{1n}x^n &= b_1\,,\\ a_{21}x^1 + a_{22}x^2 + ... + a_{2n}x^n &= b_2\,,\\ &... \\ a_{n1}x^1 + a_{n2}x^2 + ... + a_{nn}x^n &= b_n\,, \end{aligned}$$

as expected from the original matrix form $\mathbf{Ax} = \mathbf{b}$. This indicial notation can be more useful than the matrix one, like e.g., in computer science, where indices would represent loop variables. However, the full potential of tensor analysis is to deal with nonlinear multivariate systems, which are untractable by linear matrix algebra and analysis. The core of this *nonlinear multivariate analysis* is *general functional transformation*.

Transformation of Coordinates

Suppose that we have two sets of curvilinear coordinates that are single–valued, continuous and smooth functions of time, $x^j = x^j(t),\ (j = 1, ..., m)$ and $\bar{x}^i = \bar{x}^i(t),\ (i = 1, ..., n)$, respectively, representing trajectories of motion of some physical or engineering system. Then a general $(m \times n)$D transformation (i.e., a nonlinear map) $x^j \mapsto \bar{x}^i$ is defined by the set of transformation equations

$$\bar{x}^i = \bar{x}^i(x^j), \qquad (i = 1, ..., n;\ j = 1, ..., m). \tag{3.2}$$

In case of the square transformation, $m = n$, we can freely exchange the indices, like e.g., in general relativity theory. On the other hand, in the general case of rectangular transformation, $m \neq n$, like e.g., in robotics, and we need to take care of these 'free' indices.

Now, if the *Jacobian determinant* of this *coordinate transformation* is different from zero,

$$\left| \frac{\partial \bar{x}^i}{\partial x^j} \right| \neq 0,$$

then the transformation (3.2) is reversible and the inverse transformation,

$$x^j = x^j(\bar{x}^i),$$

exists as well. Finding the inverse transformation is the problem of matrix inverse: in case of the square matrix it is well defined, although the inverse might not exist if the matrix is singular. However, in case of the square matrix, its proper inverse does not exist, and the only tool that we are left with is the so–called *Moore–Penrose pseudoinverse*, which gives an optimal solution (in the least–squares sense) of an overdetermined system of equations. Every (overdetermined) rectangular coordinate transformation induces a *redundant system*.

For example, in Euclidean 3D space $\mathbb{R}^3$, transformation from Cartesian coordinates $y^k = \{x, y, z\}$ into spherical coordinates $x^i = \{\rho, \theta, \varphi\}$ is given by

$$y^1 = x^1 \cos x^2 \cos x^3, \qquad y^2 = x^1 \sin x^2 \cos x^3, \qquad y^3 = x^1 \sin x^3, \tag{3.3}$$

with the Jacobian matrix given by

$$\left(\frac{\partial y^k}{\partial x^i} \right) = \begin{pmatrix} \cos x^2 \cos x^3 & -x^1 \sin x^2 \cos x^3 & -x^1 \cos x^2 \sin x^3 \\ \sin x^2 \cos x^3 & x^1 \cos x^2 \cos x^3 & -x^1 \sin x^2 \sin x^3 \\ \sin x^3 & 0 & x^1 \cos x^3 \end{pmatrix} \tag{3.4}$$

and the corresponding Jacobian determinant, $\left| \frac{\partial y^k}{\partial x^i} \right| = (x^1)^2 \cos x^3$.

An inverse transform is given by

$$x^1 = \sqrt{(y^1)^2 + (y^2)^2 + (y^3)^2}, \qquad x^2 = \arctan\left(\frac{y^2}{y^1} \right),$$

$$x^3 = \arctan\left(\frac{y^3}{\sqrt{(y^1)^2 + (y^2)^2}} \right), \qquad \text{with} \quad \left| \frac{\partial x^i}{\partial y^k} \right| = \frac{1}{(x^1)^2 \cos x^3}.$$

As an important engineering (robotic) example, we have a rectangular transformation from 6 DOF external, end–effector (e.g., hand) coordinates, into n DOF internal, joint–angle coordinates. In most cases this is a redundant manipulator system, with infinite number of possible joint trajectories.

Scalar Invariants

A *scalar invariant* (or, a zeroth order tensor) with respect to the transformation (3.2) is the quantity $\varphi = \varphi(t)$ defined as

$$\varphi(x^i) = \bar{\varphi}(\bar{x}^i),$$

which does not change at all under the coordinate transformation. In other words, φ is *invariant* under (3.2). Biodynamic examples of scalar invariants include various energies (kinetic, potential, biochemical, mental) with the corresponding kinds of work, as well as related thermodynamic quantities (free energy, temperature, entropy, etc.).

Vectors and Covectors

Any geometrical object $v^i = v^i(t)$ that under the coordinate transformation (3.2) transforms as

$$\bar{v}^i = v^j \frac{\partial \bar{x}^i}{\partial x^j}, \qquad \text{(remember, summing upon } j-\text{index)},$$

represents a *vector*, traditionally called a *contravariant vector*, or, a first–order contravariant tensor. Standard physical and engineering examples include both translational and rotational velocities and accelerations.

On the other hand, any geometrical object $v_i = v_i(t)$ that under the coordinate transformation (3.2) transforms as

$$\bar{v}_i = v_j \frac{\partial x^j}{\partial \bar{x}^i},$$

represents a *one–form* or *covector*, traditionally called a *covariant vector*, or, a first–order covariant tensor. Standard physical and engineering examples include both translational and rotational momenta, forces and torques.

Second–Order Tensors

Any geometrical object $t^{ik} = t^{ik}(t)$ that under the coordinate transformation (3.2) transforms as

$$\bar{t}^{ik} = t^{jl} \frac{\partial \bar{x}^i}{\partial x^j} \frac{\partial \bar{x}^k}{\partial x^l}, \qquad (i, k = 1, ..., n;\ j, l = 1, ..., m),$$

represents a *second–order contravariant tensor*. It can be get as an *outer product* of two contravariant vectors, $t^{ik} = u^i v^k$.

Any geometrical object $t_{ik} = t_{ik}(t)$ that under the coordinate transformation (3.2) transforms as

$$\bar{t}_{ik} = t_{jl} \frac{\partial x^j}{\partial \bar{x}^i} \frac{\partial x^l}{\partial \bar{x}^k},$$

represents a *second–order covariant tensor*. It can be get as an outer product of two covariant vectors, $t_{ik} = u_i v_k$.

Any geometrical object $t^i_k = t^i_k(t)$ that under the coordinate transformation (3.2) transforms as

$$\bar{t}^i_k = t^j_l \frac{\partial \bar{x}^i}{\partial x^j} \frac{\partial x^l}{\partial \bar{x}^k},$$

represents a *second–order mixed tensor*. It can be get as an outer product of a covariant vector and a contravariant vector, $t^i_k = u^i v_k$.

Standard physical and engineering examples examples include:

1. The fundamental (material) covariant metric tensor $\mathbf{g} \equiv g_{ik}$, i.e., inertia matrix, given usually by the transformation from Cartesian coordinates y^j to curvilinear coordinates x^i,

$$g_{ik} = \frac{\partial y^j}{\partial x^i} \frac{\partial y^j}{\partial x^k}, \qquad \text{(summing over } j\text{)}.$$

   It is used in the quadratic metric form ds^2 of the space in consideration (e.g., a certain physical or engineering configuration space)

$$ds^2 \equiv dy^j dy^j = g_{ik} dx^i dx^k,$$

   where the first term on the r.h.s denotes the *Euclidean metrics*, while the second term is the *Riemannian metric* of the space, respectively.
2. Its inverse $\mathbf{g}^{-1} \equiv g^{ik}$, given by

$$g^{ik} = (g_{ik})^{-1} = \frac{G_{ik}}{|g_{ik}|}, \qquad G_{ik} \text{ is the cofactor of the matrix } (g_{ik});$$

3. The *Kronecker–delta* symbol δ^i_k, given by

$$\delta^i_k = \begin{cases} 1 \text{ if } i = k \\ 0 \text{ if } i \neq k \end{cases},$$

   used to denote the metric tensor in Cartesian orthogonal coordinates. δ^i_k is a discrete version of the *Dirac δ−function*. The *generalized Kronecker–delta symbol* δ^{ijk}_{lmn} (in 3D) is the product of *Ricci antisymmetric tensors* ε^{ijk} and ε_{lmn},

$$\delta^{ijk}_{lmn} = \varepsilon^{ijk} \varepsilon_{lmn} = \begin{cases} 0 & \text{if at least two indices are equal} \\ +1 & \text{if both } ijk \text{ and } lmn \text{ are either even or odd} \\ -1 & \text{if one of } ijk, lmn \text{ is even and the other is odd} \end{cases}.$$

For example, to derive components of the metric tensor $\mathbf{g} \equiv g_{ij}$ in standard spherical coordinates, we use the relations (3.3–3.4) between the spherical coordinates $x^i = \{\rho, \theta, \varphi\}$ and the Cartesian coordinates $y^k = \{x, y, z\}$, and the definition, $g_{ij} = \frac{\partial y^k}{\partial x^i}\frac{\partial y^k}{\partial x^j}$, to get the metric tensor (in matrix form)

$$(g_{ij}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & (x^1)^2 \cos^2 x^3 & 0 \\ 0 & 0 & (x^1)^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \rho^2 \cos^2 \varphi & 0 \\ 0 & 0 & \rho^2 \end{pmatrix}, \tag{3.5}$$

and the inverse metric tensor

$$(g^{ij}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{(x^1)^2 \cos^2 x^3} & 0 \\ 0 & 0 & \frac{1}{(x^1)^2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\rho^2 \cos^2 \varphi} & 0 \\ 0 & 0 & \frac{1}{\rho^2} \end{pmatrix}. \tag{3.6}$$

Given a tensor, we can derive other tensors by raising and lowering its indices, by their multiplication with covariant and contravariant metric tensors. In this way, the so–called *associated tensors* to the given tensor are be formed. For example, v^i and v_i are associated tensors, related by

$$v_i = g_{ik} v^k \qquad \text{and} \qquad v^i = g^{ik} v_k.$$

Given two vectors, $\mathbf{u} \equiv u^i$ and $\mathbf{v} \equiv v^i$, their inner (dot, or scalar) product is given by

$$\mathbf{u} \cdot \mathbf{v} \equiv g_{ij} u^i v^j,$$

while their vector (cross) product (in 3D) is given by

$$\mathbf{u} \times \mathbf{v} \equiv \varepsilon_{ijk} u^j v^k.$$

Higher–Order Tensors

As a generalization of above tensors, consider a geometrical object $R^i_{kps} = R^i_{kps}(t)$ that under the coordinate transformation (3.2) transforms as

$$\bar{R}^i_{kps} = R^j_{lqt} \frac{\partial \bar{x}^i}{\partial x^j} \frac{\partial x^l}{\partial \bar{x}^k} \frac{\partial x^q}{\partial \bar{x}^p} \frac{\partial x^t}{\partial \bar{x}^s}, \qquad (\text{all indices} = 1, ..., n). \tag{3.7}$$

Clearly, $R^i_{kjl} = R^i_{kjl}(x, t)$ is a fourth order tensor, once contravariant and three times covariant, representing the central tensor in Riemannian geometry, called the *Riemann curvature tensor*. As all physical and engineering configuration spaces are Riemannian manifolds, they are all characterized by curvature tensors. In case $R^i_{kjl} = 0$, the corresponding Riemannian manifold reduces to the Euclidean space of the same dimension, in which $g_{ik} = \delta^i_k$.

If one contravariant and one covariant index of a tensor a set equal, the resulting sum is a tensor of rank two less than that of the original tensor. This process is called *tensor contraction*.

If to each point of a region in an nD space there corresponds a definite tensor, we say that a *tensor–field* has been defined. In particular, this is a *vector–field* or a *scalar–field* according as the tensor is of rank one or zero. It should be noted that a tensor or tensor field is not just the set of its components in one special coordinate system, but all the possible sets of components under any transformation of coordinates.

Tensor Symmetry

A tensor is called *symmetric* with respect to two indices of the same variance if its components remain unaltered upon interchange of the indices; e.g., $a_{ij} = a_{ji}$, or $a^{ij} = a^{ji}$. A tensor is called *skew–symmetric* (or, *antisymmetric*) with respect to two indices of the same variance if its components change sign upon interchange of the indices; e.g., $a_{ij} = -a_{ji}$, or $a^{ij} = -a^{ji}$. Regarding tensor symmetry, in the following we will prove several useful propositions.

(i) *Every second–order tensor can be expressed as the sum of two tensors, one of which is symmetric and the other is skew–symmetric.* For example, a second–order tensor a_{ij}, which is for $i, j = 1, ..., n$ given by the $n \times n-$matrix

$$a_{ij} = \begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{n2} \\ ... & ... & ... & ... \\ a_{n1} & a_{n2} & ... & a_{nn} \end{pmatrix},$$

can be rewritten as

$$\begin{aligned} a_{ij} &= \frac{1}{2}a_{ij} + \frac{1}{2}a_{ij} + \frac{1}{2}a_{ji} - \frac{1}{2}a_{ji}\,, \quad \text{that can be rearranged as} \\ &= \frac{1}{2}a_{ij} + \frac{1}{2}a_{ji} + \frac{1}{2}a_{ij} - \frac{1}{2}a_{ji}\,, \quad \text{which can be regrouped as} \\ &= \frac{1}{2}(a_{ij} + a_{ji}) + \frac{1}{2}(a_{ij} - a_{ji}), \quad \text{which can be written as} \\ &= a_{(ij)} + a_{[ij]}\,, \end{aligned}$$

where $a_{(ij)}$ denotes its symmetric part, while $a_{[ij]}$ denotes its skew–symmetric part, as required.

(ii) *Every quadratic form can be made symmetric.* For example, a quadratic form $a_{ij}x^i x^j$, that (for $i, j = 1, ..., n$) expands as

$$\begin{aligned} a_{ij}x^i x^j &= a_{11}x^1x^1 + a_{12}x^1x^2 + ... + a_{1n}x^1x^n \\ &+ a_{21}x^2x^1 + a_{22}x^2x^2 + ... + a_{2n}x^2x^n \\ &... \\ &+ a_{n1}x^nx^1 + a_{n2}x^nx^2 + ... + a_{nn}x^nx^n, \end{aligned}$$

with a non–symmetric second–order tensor a_{ij}, can be made symmetric in the following way.

$$
\begin{aligned}
a_{ij}x^i x^j &= \frac{1}{2}a_{ij}x^i x^j + \frac{1}{2}a_{ij}x^i x^j. \\
&\quad \text{If we swap indices in the second term, we get} \\
&= \frac{1}{2}a_{ij}x^i x^j + \frac{1}{2}a_{ji}x^j x^i, \quad \text{which is equal to} \\
&= \frac{1}{2}(a_{ij} + a_{ji})\, x^i x^j. \\
&\quad \text{If we now use a substitution,} \\
\frac{1}{2}(a_{ij} + a_{ji}) &\equiv b_{ij} = b_{ji}, \quad \text{we get} \\
a_{ij}x^i x^j &= b_{ij}x^i x^j,
\end{aligned}
$$

where a_{ij} is non–symmetric and b_{ij} is symmetric, as required.

(iii) *Every second–order tensor that is the sum* $a^{ij} = u^i v^j + u^j v^i$, or, $a_{ij} = u_i v_j + u_j v_i$ is *symmetric.* In both cases, if we swap the indices i and j, we get $a^{ji} = u^j v^i + u^i v^j$, (resp. $a_{ji} = u_j v_i + u_i v_j$), which implies that the tensor a^{ij} (resp. a_{ij}) is symmetric.

(iv) *Every second–order tensor that is the difference* $b^{ij} = u^i v^j - u^j v^i$, or, $b_{ij} = u_i v_j - u_j v_i$ is *skew–symmetric.* In both cases, if we swap the indices i and j, we get $b^{ji} = -(u^j v^i - u^i v^j)$, (resp. $b_{ji} = -(u_j v_i - u_i v_j)$), which implies that the tensor b^{ij} (resp. b_{ij}) is skew–symmetric.

3.1.2 Euclidean Tensors

Basis Vectors and the Metric Tensor in $\mathbb{R}^n$

The natural *Cartesian coordinate basis* in an nD Euclidean space $\mathbb{R}^n$ is defined as a set of nD unit vectors e^i given by

$$
e^1 = [\{1,0,0,...\}^t,\ e^2 = \{0,1,0,...\}^t,\ e^3 = \{0,0,1,...\}^t, ...,\ e^n = \{0,0,...,1\}^t],
$$

(where index t denotes transpose) while its dual basis e_i is given by:

$$
e_1 = [\{1,0,0,...\},\ e_2 = \{0,1,0,...\},\ e_3 = \{0,0,1,...\}, ...,\ e_n = \{0,0,...,1\}],
$$

(no transpose) where the definition of the dual basis is given by the Kronecker's $\delta-$symbol, i.e., the $n \times n$ identity matrix:

$$
e^i \cdot e_j = \delta^i_j = \begin{bmatrix} 1 & 0 & 0 & ... & 0 \\ 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 1 \end{bmatrix},
$$

that is the metric tensor in Cartesian coordinates equals $\mathbf{g} = \delta^i_j$. In general, (i.e., curvilinear) coordinate system, the metric tensor $\mathbf{g} = g_{ij}$ is defined as the scalar product of the dual basis vectors, i.e., the $n \times n$ matrix:

$$g_{ij} = e_i \cdot e_j = \begin{bmatrix} g_{11} & g_{12} & g_{13} & \cdots & g_{1n} \\ g_{21} & g_{22} & g_{23} & \cdots & g_{2n} \\ g_{31} & g_{32} & g_{33} & \cdots & g_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ g_{n1} & g_{n2} & g_{n3} & \cdots & g_{nn} \end{bmatrix}.$$

Tensor Products in $\mathbb{R}^n$

Let $\mathbf{u}$ and $\mathbf{v}$ denote two vectors in $\mathbb{R}^n$, with their components given by

$$u^i = u \cdot e^i, \quad \text{and} \quad v^j = v \cdot e^j,$$

where $u = |\mathbf{u}|$ and $v = |\mathbf{v}|$ are their respective norms (or, lengths). Then their inner product (i.e., scalar, or dot product) $\mathbf{u} \cdot \mathbf{v}$ is a scalar invariant S, defined as

$$S = u^i \cdot v^j = g_{ij} u^i v^j.$$

Besides the dot product of two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, there is also their tensor product (i.e., generalized vector, or cross product), which is a second–order tensor

$$\mathbf{T} = \mathbf{u} \otimes \mathbf{v}, \text{ in components, } T^{ij} = u^i \otimes v^j.$$

In the natural basis e_i this tensor is expanded as

$$\mathbf{T} = T^{ij} e_i \otimes e_j,$$

while its components in the dual basis read:

$$T^{ij} = T(e^i, e^j),$$

where $T = |\mathbf{T}|$ is its norm. To get its components in curvilinear coordinates, we need first to substitute it in Cartesian basis:

$$T^{ij} = T^{mn}(e_m \otimes e_n)(e^i, e^j),$$

then to evaluate it on the slots:

$$T^{ij} = T^{mn} e_m \cdot e^i \, e_n \cdot e^j,$$

and finally to calculate the other index configurations by lowering indices, by means of the metric tensor:

$$T^i_j = g_{jm} T^{im}, \qquad T_{ij} = g_{im} g_{jn} T^{mn}.$$

3.1.3 Tensor Derivatives on Riemannian Manifolds

Consider now some nD Riemannian manifold M with the metric form (i.e., line element) $ds^2 = g_{ik} dx^i dx^k$, as a configuration space for a certain physical or engineering system (e.g., human spine, or arm–shoulder complex).

Christoffel's Symbols

Partial derivatives of the metric tensor g_{ik} form themselves special symbols that do not transform as tensors (with respect to the coordinate transformation (3.2)), but nevertheless represent important quantities in tensor analysis. They are called *Christoffel symbols of the first kind*, defined by

$$\Gamma_{ijk} = \frac{1}{2}(\partial_{x^k} g_{ij} + \partial_{x^j} g_{ki} - \partial_{x^i} g_{jk}), \qquad \left(\text{remember, } \partial_{x^i} \equiv \frac{\partial}{\partial x^i}\right)$$

and *Christoffel symbols of the second kind*, defined by

$$\Gamma_{ij}^k = g^{kl}\Gamma_{ijl}.$$

The Riemann curvature tensor R_{ijk}^l (3.7) of the manifold M, can be expressed in terms of the later as

$$R_{ijk}^l = \partial_{x^j}\Gamma_{ik}^l - \partial_{x^k}\Gamma_{ij}^l + \Gamma_{rj}^l\Gamma_{ik}^r - \Gamma_{rk}^l\Gamma_{ij}^r.$$

For example, in 3D spherical coordinates, $x^i = \{\rho, \theta, \varphi\}$, with the metric tensor and its inverse given by (3.5, 3.6), it can be shown that the only nonzero Christoffel's symbols are:

$$\begin{aligned} &\Gamma_{12}^2 = \Gamma_{21}^2 = \Gamma_{13}^3 = \Gamma_{31}^3 = \frac{1}{\rho}, \qquad \Gamma_{23}^3 = \Gamma_{32}^2 = -\tan\theta, \\ &\Gamma_{22}^1 = -\rho, \qquad \Gamma_{33}^1 = -\rho\cos^2\theta, \qquad \Gamma_{33}^2 = \sin\theta\cos\theta. \end{aligned} \tag{3.8}$$

Geodesics

From the Riemannian metric form $ds^2 = g_{ik}dx^i dx^k$ it follows that the distance between two points t_1 and t_2 on a curve $x^i = x^i(t)$ in M is given by

$$s = \int_{t_1}^{t_2} \sqrt{g_{ik}\dot{x}^i\dot{x}^k}dt.$$

That curve $x^i = x^i(t)$ in M which makes the distance s a minimum is called a *geodesic* of the space M (e.g., in a sphere, the geodesics are arcs of great circles). Using the calculus of variations, the geodesics are found from the differential *geodesic equation*,

$$\ddot{x}^i + \Gamma_{jk}^i\dot{x}^j\dot{x}^k = 0, \tag{3.9}$$

where overdot means derivative upon the line parameter s.

For example, in 3D spherical coordinates $x^i = \{\rho, \theta, \varphi\}$, using (3.8), geodesic equation (3.9) becomes a system of three scalar ODEs,

$$\begin{aligned} &\ddot{\rho} - \rho\dot{\theta}^2 - \rho\cos^2\theta\dot{\varphi}^2 = 0, \qquad \ddot{\theta} + \frac{2}{\rho}\dot{\rho}\dot{\varphi} + \sin\theta\cos\theta\dot{\varphi}^2 = 0, \\ &\ddot{\varphi} + \frac{2}{\rho}\dot{\rho}\dot{\varphi} - 2\tan\theta\dot{\theta}\dot{\varphi} = 0. \end{aligned} \tag{3.10}$$

The Covariant Derivative

Ordinary total and partial derivatives of vectors (covectors) *do not transform as vectors* (covectors) with respect to the coordinate transformation (3.2). For example, let y^k be Cartesian coordinates and x^i be general curvilinear coordinates of a dynamical system (with $i,k = 1,...,n$). We have: $x^i(t) = x^i[y^k(t)]$, which implies that

$$\frac{dx^i}{dt} = \frac{\partial x^i}{\partial y^k}\frac{dy^k}{dt}, \qquad \text{or equivalently,} \qquad \dot{x}^i = \frac{\partial x^i}{\partial y^k}\dot{y}^k,$$

that is a transformation law for the contravariant vector, which means that the velocity $v^i \equiv \dot{x}^i \equiv \frac{dx^i}{dt}$ *is* a proper contravariant vector. However, if we perform another time differentiation, we get

$$\frac{d^2x^i}{dt^2} = \frac{\partial x^i}{\partial y^k}\frac{d^2y^k}{dt^2} + \frac{\partial^2 x^i}{\partial y^k \partial y^m}\frac{dy^k}{dt}\frac{dy^m}{dt},$$

which means that $\frac{d^2x^i}{dt^2}$ *is not* a proper vector.

$\frac{d^2x^i}{dt^2}$ is an acceleration vector only in a special case when x^i are another Cartesian coordinates; then $\frac{\partial^2 x^i}{\partial y^k \partial y^m} = 0$, and therefore the original coordinate transformation is linear, $x^i = a^i_k y^k + b^i$ (where a^i_k and b^i are constant).

Therefore, $\frac{d^2x^i}{dt^2}$ represents an acceleration vector only in terms of Newtonian mechanics in a Euclidean space $\mathbb{R}^n$, while it is not a proper acceleration vector in terms of Lagrangian or Hamiltonian mechanics in general curvilinear coordinates on a smooth manifold M^n. And we know that Newtonian mechanics in $\mathbb{R}^n$ is sufficient only for fairly simple mechanical systems.

The above is true for any tensors. So we need to find another derivative operator to be able to preserve their tensor character. The solution to this problem is called the *covariant derivative*.

The covariant derivative $v^i_{;k}$ of a contravariant vector v^i is defined as

$$v^i_{;k} = \partial_{x^k} v^i + \Gamma^i_{jk} v^j.$$

Similarly, the covariant derivative $v_{i;k}$ of a covariant vector v_i is defined as

$$v_{i;k} = \partial_{x^k} v_i - \Gamma^j_{ik} v_j.$$

Generalization for the higher order tensors is straightforward; e.g., the covariant derivative $t^j_{kl;q}$ of the third order tensor t^j_{kl} is given by

$$t^j_{kl;q} = \partial_{x^q} t^j_{kl} + \Gamma^j_{qs} t^s_{kl} - \Gamma^s_{kq} t^j_{sl} - \Gamma^s_{lq} t^j_{ks}.$$

The covariant derivative is the most important tensor operator in general relativity (its zero defines *parallel transport*) as well as the basis for defining other differential operators in mechanics and physics.

Covariant Form of Differential Operators

Here we give the covariant form of classical vector differential operators: gradient, divergence, curl and Laplacian.

Gradient. If $\varphi = \varphi(x^i, t)$ is a scalar field, the gradient one–form $\mathrm{grad}(\varphi)$ is defined by

$$\mathrm{grad}(\varphi) = \nabla\varphi = \varphi_{;i} = \partial_{x^i}\varphi.$$

Divergence. The divergence $\mathrm{div}(v^i)$ of a vector–field $v^i = v^i(x^i, t)$ is defined by contraction of its covariant derivative with respect to the coordinates $x^i = x^i(t)$, i.e., the contraction of $v^i_{;k}$, namely

$$\mathrm{div}(v^i) = v^i_{;i} = \frac{1}{\sqrt{g}}\partial_{x^i}(\sqrt{g}v^i).$$

Curl. The curl $\mathrm{curl}(\theta_i)$ of a one–form $\theta_i = \theta_i(x^i, t)$ is a second–order covariant tensor defined as

$$\mathrm{curl}(\theta_i) = \theta_{i;k} - \theta_{k;i} = \partial_{x^k}\theta_i - \partial_{x^i}\theta_k.$$

Laplacian. The Laplacian $\Delta\varphi$ of a scalar invariant $\varphi = \varphi(x^i, t)$ is the divergence of $\mathrm{grad}(\varphi)$, or

$$\Delta\varphi = \nabla^2\varphi = \mathrm{div}(\mathrm{grad}(\varphi)) = \mathrm{div}(\varphi_{;i}) = \frac{1}{\sqrt{g}}\partial_{x^i}(\sqrt{g}g^{ik}\partial_{x^k}\varphi).$$

The Absolute Derivative

The *absolute derivative* (or *intrinsic*, or *Bianchi's derivative*) of a contravariant vector v^i along a curve $x^k = x^k(t)$ is denoted by $\dot{\bar{v}}^i \equiv Dv^i/dt$ and defined as the inner product of the covariant derivative of v^i and $\dot{x}^k \equiv dx^k/dt$, i.e., $v^i_{;k}\dot{x}^k$, and is given by

$$\dot{\bar{v}}^i = \dot{v}^i + \Gamma^i_{jk}v^j\dot{x}^k.$$

Similarly, the absolute derivative $\dot{\bar{v}}_i$ of a covariant vector v_i is defined as

$$\dot{\bar{v}}_i = \dot{v}_i - \Gamma^j_{ik}v_j\dot{x}^k.$$

Generalization for the higher order tensors is straightforward; e.g., the absolute derivative $\dot{\bar{t}}^j_{kl}$ of the third order tensor t^j_{kl} is given by

$$\dot{\bar{t}}^j_{kl} = \dot{t}^j_{kl} + \Gamma^j_{qs}t^s_{kl}\dot{x}^q - \Gamma^s_{kq}t^j_{sl}\dot{x}^q - \Gamma^s_{lq}t^j_{ks}\dot{x}^q.$$

The absolute derivative is the most important differential operator in physics and engineering, as it is the basis for the *covariant form* of both Lagrangian and Hamiltonian equations of motion of many physical and engineering systems.

Application to Curve Geometry

Given three unit vectors: *tangent* τ^i, *principal normal* β^i, and *binormal* ν^i, as well as two scalar invariants: curvature K and torsion T, of a curve $\gamma(s) = \gamma[x^i(s)]$, the so–called *Frenet–Serret formulae* are valid[1]

$$\begin{aligned}
\dot{\tau}^i &\equiv \dot{\tau}^i + \Gamma^i_{jk}\tau^j\dot{x}^k = K\beta^i,\\
\dot{\beta}^i &\equiv \dot{\beta}^i + \Gamma^i_{jk}\beta^j\dot{x}^k = -(K\tau^i + T\nu^i),\\
\dot{\nu}^i &\equiv \dot{\nu}^i + \Gamma^i_{jk}\nu^j\dot{x}^k = T\beta^i.
\end{aligned}$$

Application to Mechanical Definitions of Acceleration and Force

In modern analytical mechanics, the two fundamental notions of *acceleration* and *force* in general curvilinear coordinates are substantially different from the corresponding terms in Cartesian coordinates as commonly used in engineering mechanics. Namely, the acceleration vector *is not* an ordinary time derivative of the velocity vector; 'even worse', the force, which is a paradigm of a vector in statics and engineering vector mechanics, *is not* a vector at all. Proper mathematical definition of the acceleration vector is the absolute time derivative of the velocity vector, while the force is a differential one–form.

To give a brief look at these 'weird mathematical beasts', consider a material dynamical system described by n curvilinear coordinates $x^i = x^i(t)$. First, recall from subsection 3.1.3 above, that an ordinary time derivative of the velocity vector $v^i(t) = \dot{x}^i(t)$ *does not transform as a vector* with respect to the general coordinate transformation (3.2). Therefore, $a^i \neq \dot{v}^i$. So, we need to use its absolute time derivative to define the acceleration vector (with $i, j, k = 1, ..., n$),

$$a^i = \dot{\bar{v}}^i \equiv \frac{Dv^i}{dt} = v^i_{;k}\dot{x}^k \equiv \dot{v}^i + \Gamma^i_{jk}v^jv^k \equiv \ddot{x}^i + \Gamma^i_{jk}\dot{x}^j\dot{x}^k, \tag{3.11}$$

which is equivalent to the l.h.s of the geodesic equation (3.9). Only in the particular case of Cartesian coordinates, the general acceleration vector (3.11) reduces to the familiar engineering form of the Euclidean acceleration vector[2], $\mathbf{a} = \dot{\mathbf{v}}$.

For example, in standard spherical coordinates $x^i = \{\rho, \theta, \varphi\}$, we have the components of the acceleration vector given by (3.10), if we now reinterpret overdot as the time derivative,

[1] In this paragraph, the overdot denotes the total derivative with respect to the line parameter s (instead of time t).

[2] Any Euclidean space can be defined as a set of Cartesian coordinates, while any Riemannian manifold can be defined as a set of curvilinear coordinates. Christoffel's symbols Γ^i_{jk} vanish in Euclidean spaces defined by Cartesian coordinates; however, they are nonzero in Riemannian manifolds defined by curvilinear coordinates.

$$a^{\rho} = \ddot{\rho} - \rho\dot{\theta}^2 - \rho\cos^2\theta\dot{\varphi}^2, \qquad a^{\theta} = \ddot{\theta} + \frac{2}{\rho}\dot{\rho}\dot{\varphi} + \sin\theta\cos\theta\dot{\varphi}^2,$$

$$a^{\varphi} = \ddot{\varphi} + \frac{2}{\rho}\dot{\rho}\dot{\varphi} - 2\tan\theta\dot{\theta}\dot{\varphi}.$$

Now, using (3.11), the *Newton's fundamental equation of motion, that is the basis of all science,* $\mathbf{F} = m\,\mathbf{a}$, gets the following tensorial form

$$F^i = ma^i = m\dot{v}^i = m(v^i_{;k}\dot{x}^k) \equiv m(\dot{v}^i + \Gamma^i_{jk}v^j v^k) = m(\ddot{x}^i + \Gamma^i_{jk}\dot{x}^j\dot{x}^k), \quad (3.12)$$

which defines Newtonian force as a contravariant vector.

However, modern Hamiltonian dynamics reminds us that: (i) Newton's own force definition was not really $\mathbf{F} = m\,\mathbf{a}$, but rather $\mathbf{F} = \dot{\mathbf{p}}$, where $\mathbf{p}$ is the system's momentum, and (ii) the momentum $\mathbf{p}$ is not really a vector, but rather a dual quantity, a differential one–form[3]. Consequently, the force, as its time derivative, is also a one–form (see Figure 3.1; also, compare with Figure Figure 1.8 above). This new force definition includes the precise definition of the mass distribution within the system, by means of its Riemannian metric tensor g_{ij}. Thus, (3.12) has to be modified as

$$F_i = mg_{ij}a^j \equiv mg_{ij}(\dot{v}^j + \Gamma^j_{ik}v^i v^k) = mg_{ij}(\ddot{x}^j + \Gamma^j_{ik}\dot{x}^i\dot{x}^k), \quad (3.13)$$

where the quantity mg_{ij} is called the *material metric tensor*, or *inertia matrix*. Equation (3.13) generalizes the notion of the Newtonian force $\mathbf{F}$, from Euclidean space $\mathbb{R}^n$ to the Riemannian manifold M.

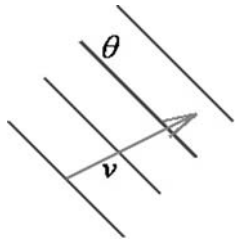


Fig. 3.1. A one–form θ (which is a family of parallel (hyper)surfaces, the so–called *Grassmann planes*) pierced by the vector v to give a scalar product $\theta(v) \equiv < \theta, v >= 2.6$ (see [MTW73] for technical details).

3.1.4 Tensor Mechanics in Brief

Recall that a material system is regarded from the dynamical standpoint as a collection of particles which are subject to interconnections and constraints

[3] For example, in Dirac's $< bra|ket >$ formalism, *ket*s are vectors, while *bra*s are one–forms; in matrix notation, columns are vectors, while rows are one–forms.

of various kinds (e.g., a rigid body is regarded as a number of particles rigidly connected together so as to remain at invariable distances from each other). The *number of independent coordinates which determine the configuration of a dynamical system completely* is called the *number of degrees of freedom* (DOF) of the system. In other words, this number, n, is the *dimension of the system's configuration manifold*. This viewpoint is the core of our geometrical dynamics.

For simplicity, let us suppose that we have a dynamical system with three DOF (e.g., a particle of mass M, or a rigid body of mass M with one point fixed); generalization to n DOF, with N included masses M_α, is straightforward. The configuration of our system at any time is then given by three coordinates $\{q^1, q^2, q^3\}$. As the coordinates change in value the dynamical system changes its configuration. Obviously, there is an infinite number of sets of independent coordinates which will determine the configuration of a dynamical system, but since the position of the system is completely given by any one set, these sets of coordinates must be functionally related. Hence, if $\bar{q}^i$ is any other set of coordinates, these quantities must be connected with q^i by formulae of the type

$$\bar{q}^i = \bar{q}^i(q^i), \qquad (i = 1, ..., n(= 3)). \tag{3.14}$$

Relations (3.14) are the equations of transformation from one set of dynamical coordinates to another and, in a standard tensorial way (see [MTW73]), we can define tensors relative to this coordinate transformation. The generalized coordinates q^i, $(i = 1, ..., n)$ constitute the system's *configuration manifold*.

In particular, in our ordinary Euclidean 3−dimensional (3D) space $\mathbb{R}^3$, the ordinary Cartesian axes are $x^i = \{x, y, z\}$, and the *system's center of mass* (COM) is given by

$$C^i = \frac{M_\alpha x^i_\alpha}{\sum_{\alpha=1}^N M_\alpha},$$

where Greek subscript α labels the masses included in the system. If we have a continuous distribution of matter $V = V(M)$ rather than the discrete system of masses M_α, all the α−sums should be replaced by volume integrals, the element of mass dM taking the place of M_α,

$$\sum_{\alpha=1}^{N} M_\alpha \Rightarrow \iiint_V dM.$$

An important quantity related to the system's COM is the double symmetric contravariant tensor

$$I^{ij} = M_\alpha x^i_\alpha x^j_\alpha, \tag{3.15}$$

called the *inertia tensor*, calculated relative to the origin O of the Cartesian axes $x^i_\alpha = \{x_\alpha, y_\alpha, z_\alpha\}$. If we are given a straight line through O, defined by

its unit vector λ^i, and perpendiculars p_α are drawn from the different particles on the line λ^i, the quantity

$$I(\lambda^i) = M_\alpha p_\alpha^2$$

is called the *moment of inertia* around λ^i. The moment of inertia $I(\lambda^i)$ can be expressed through inertia tensor (3.15) as

$$I(\lambda^i) = (Ig_{ij} - I_{ij})\lambda^i\lambda^j,$$

where g_{ij} is the system's *Euclidean 3D metric tensor* (as defined above), $I = g_{ij}I^{ij}$, and $I_{ij} = g_{rm}g_{sn}I^{mn}$ is the *covariant inertia tensor*. If we now consider the quadric Q whose equation is

$$(Ig_{ij} - I_{ij})x^i x^j = 1, \tag{3.16}$$

we find that the moment of inertia around λ^i is $1/R$, where R is the radius vector of Q in the direction of λ^i. The quadric Q defined by relation (3.16) is called the *ellipsoid of inertia* at the origin O. It has always three principal axes, which are called the *principal axes of inertia* at O, and the planes containing them in pairs are called the *principal planes of inertia* at O. The principal axes of inertia are given by the equations

$$(Ig_{ij} - I_{ij})\lambda^j = \theta\lambda^i,$$

where θ is a root of the determinant equation

$$|(I - \theta)g_{ij} - I_{ij}| = 0.$$

More generally, if we suppose that the points of our dynamical system are referred to rectilinear Cartesian axes x^i in a Euclidean $n-$dimensional (nD) space $\mathbb{R}^n$, then when we are given the time and a set of *generalized coordinates* q^i we are also given all the points x^i of the dynamical system, as the system is determined uniquely. Consequently, the x^i are functions of q^i and possibly also of the time, that is,

$$x^i = x^i(q^i, t).$$

If we restrict ourselves to the *autonomous dynamical systems* in which these equations do not involve t, i.e.,

$$x^i = x^i(q^i), \tag{3.17}$$

then differentiating (3.17) with respect to the time t gives

$$\dot{x}^i = \frac{\partial x^i}{\partial q^j}\dot{q}^j. \tag{3.18}$$

The quantities $\dot{q}^i$, which form a vector with reference to coordinate transformations (3.14), we shall call the *generalized velocity vector*. We see from

(3.18) that when the generalized velocity vector is given we know the velocity of each point of our system. Further, this gives us the system's *kinetic energy*,

$$E_{kin} = \frac{1}{2} M_\alpha g_{mn} \dot{x}^m_\alpha \dot{x}^n_\alpha = \frac{1}{2} M_\alpha g_{mn} \frac{\partial x^m_\alpha}{\partial q^i} \frac{\partial x^n_\alpha}{\partial q^j} \dot{q}^i \dot{q}^j. \tag{3.19}$$

Now, if we use the Euclidean metric tensor g_{ij} to define the *material metric tensor* G_{ij}, including the distribution of all the masses M_α of our system, as

$$G_{ij} = M_\alpha g_{mn} \frac{\partial x^m_\alpha}{\partial q^i} \frac{\partial x^n_\alpha}{\partial q^j}, \tag{3.20}$$

the kinetic energy (3.19) becomes a *homogenous quadratic form* in the generalized system's velocities $\dot{q}^i$,

$$E_{kin} = \frac{1}{2} G_{ij} \dot{q}^i \dot{q}^j. \tag{3.21}$$

From the transformation relation (3.20) we see that the material metric tensor G_{ij} is symmetric in i and j. Also, since E_{kin} is an invariant for all transformations of generalized coordinates, from (3.21) we conclude that G_{ij} is a double symmetric tensor. Clearly, this is the central quantity in classical tensor system dynamics. We will see later, that G_{ij} defines the Riemannian geometry of the system dynamics. For simplicity reasons, G_{ij} is often denoted by purely geometrical symbol g_{ij}, either assuming or neglecting the material properties of the system.

Now, let us find the equations of motion of our system. According to the *D'Alembert's Principle of virtual displacement*s, the equations of motion in Cartesian coordinates x^i in $\mathbb{R}^n$ are embodied in the single tensor equation

$$g_{mn}(M_\alpha \ddot{x}^m_\alpha - X^m_\alpha)\delta x^n_\alpha = 0, \tag{3.22}$$

where X^i_α is the total force vector acting on the particle M_α, while δx^i_α is the associated *virtual displacement* vector, so that the product $g_{ij} X^i_\alpha \delta x^j_\alpha$ is the *virtual work* of the system, and we can neglect in X^i_α all the internal or external forces which do not work in the displacement δx^i_α. If we give the system a small displacement compatible to with the constraints of the system, we see that this displacement may be effected by giving increments δq^i to the generalized coordinates q^i of the system, and these are related to the δx^i in accordance with the transformation formulae $\delta x^i_\alpha = \frac{\partial x^i_\alpha}{\partial q^j} \delta q^j$.

Furthermore, in this displacement the internal forces due to the constraints of the system will do no work, since these constraints are preserved, and consequently only the external forces will appear in (3.22), so it becomes

$$g_{mn} \left[M_\alpha \frac{d}{dt} \left(\frac{\partial x^m_\alpha}{\partial q^j} \dot{q}^j \right) \frac{\partial x^n_\alpha}{\partial q^i} - X^m_\alpha \frac{\partial x^n_\alpha}{\partial q^i} \right] \delta q^i = 0. \tag{3.23}$$

Now, using (3.19–3.21), we derive

$$M_\alpha g_{mn}\frac{d}{dt}\left(\frac{\partial x^m_\alpha}{\partial q^j}\dot{q}^j\right)\frac{\partial x^n_\alpha}{\partial q^i}=\frac{d}{dt}(G_{ij}\dot{q}^j)-\frac{1}{2}\frac{\partial G_{st}}{\partial q^i}\dot{q}^j\dot{q}^k=\frac{d}{dt}\left(\frac{\partial E_{kin}}{\partial \dot{q}^i}\right)-\frac{\partial E_{kin}}{\partial q^i}.$$

Also, if we put

$$F_i=g_{mn}X^m_\alpha\frac{\partial x^n_\alpha}{\partial q^i},\qquad \text{we get}$$
$$F_i\delta q^i=g_{mn}X^m_\alpha\delta x^n_\alpha=\delta W, \tag{3.24}$$

where δW is the virtual work done by the external forces in the small displacement δq^i, which shows that F_i is the covariant vector, called the *generalized force vector*. Now (3.23) takes the form

$$\left[\frac{d}{dt}\left(\frac{\partial E_{kin}}{\partial \dot{q}^i}\right)-\frac{\partial E_{kin}}{\partial q^i}-F_i\right]\delta q^i=0.$$

Since the coordinates q^i are independent this equation is true for all variations δq^i and we get as a final result the *covariant Lagrangian equations of motion*,

$$\frac{d}{dt}\left(\frac{\partial E_{kin}}{\partial \dot{q}^i}\right)-\frac{\partial E_{kin}}{\partial q^i}=F_i.$$

If the force system is conservative and E_{pot} is the system's *potential energy* given by

$$F_i=-\frac{\partial E_{pot}}{\partial q^i},$$

then, using (3.24) the Lagrangian equations take the standard form

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right)=\frac{\partial L}{\partial q^i}, \tag{3.25}$$

where the *Lagrangian function* $L=L(q,\dot{q})$ of the system is given by $L=E_{kin}-E_{pot}$ (since E_{pot} does not contain $\dot{q}^i$).

Now, the kinetic energy E_{kin} of the system, given by quadratic form (3.21), is always positive except when $\dot{q}^i$ is zero in which case E_{kin} vanishes. In other words, the quadratic form (3.21) is positive definite. Consequently, we can always find the *line* (or *arc*) *element*, defined by

$$ds^2=G_{ij}dq^idq^j. \tag{3.26}$$

A manifold in which ds^2 is given by relation of the type of (3.26), geometrically with g_{ij} instead of G_{ij}, is called a *Riemannian manifold*.

Riemannian Curvature Tensor

Every Riemannian manifold is characterized by the *Riemann curvature tensor*. In physical literature (see, e.g., [MTW73]) it is usually introduced through the *Jacobi equation of geodesic deviation*, showing the acceleration of the relative separation of nearby geodesics (the shortest, straight lines on the manifold). For simplicity, consider a sphere of radius a in $\mathbb{R}^3$. Here, Jacobi equation is pretty simple,

$$\frac{d^2\xi}{ds^2} + R\xi = 0,$$

where ξ is the *geodesic separation vector* (the so–called Jacobi vector–field), s denotes the geodesic arc parameter given by (3.26) and $R = 1/a^2$ is the *Gaussian curvature* of the surface.

In case of a higher–dimensional manifold M, the situation is naturally more complex, but the main structure of the Jacobi equation remains similar,

$$\frac{D^2\xi}{ds^2} + R(u, \xi, u) = 0,$$

where D denotes the covariant derivative and $R(u, \xi, u)$ is the curvature tensor, a three–slot linear machine. In components defined in a local coordinate chart (x^i) on M, this equation reads

$$\frac{D^2\xi^i}{ds^2} + R^i_{jkl}\frac{dx^j}{ds}\xi^k\frac{dx^l}{ds} = 0,$$

where R^i_{jkl} are the components of the *Riemann curvature tensor*.

Exterior Differential Forms

Recall that *exterior differential forms* are a special kind of antisymmetrical covariant tensors (see, e.g., [Rha84, Fla63]). Such tensor–fields arise in many applications in physics, engineering, and differential geometry. The reason for this is the fact that the classical vector operations of **grad**, **div**, and **curl** as well as the theorems of Green, Gauss, and Stokes can all be expressed concisely in terms of differential forms and the main operator acting on them, the exterior derivative d. Differential forms inherit all geometrical properties of the general tensor calculus and add to it their own powerful geometrical, algebraic and topological machinery (see Figures 3.2 and 3.3). Differential $p-$forms formally occur as *integrands* under ordinary integral signs in $\mathbb{R}^3$:

- a *line integral* $\int P\,dx + Q\,dy + R\,dz$ has as its integrand the *one–form* $\omega = P\,dx + Q\,dy + R\,dz$;
- a *surface integral* $\iint A\,dydz + B\,dzdx + C\,dxdy$ has as its integrand the *two–form* $\alpha = A\,dydz + B\,dzdx + C\,dxdy$;
- a *volume integral* $\iiint K\,dxdydz$ has as its integrand the *three–form* $\lambda = K\,dxdydz$.

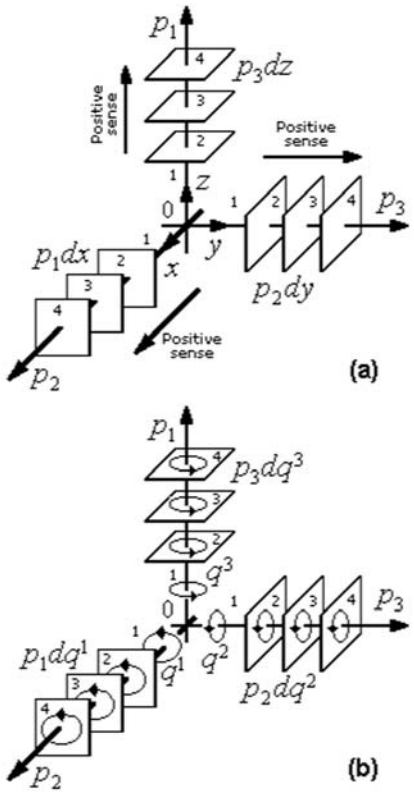


Fig. 3.2. Basis vectors and 1–forms in Euclidean $\mathbb{R}^3-$space: (a) Translational case; and (b) Rotational case.

By means of an *exterior derivative* d, a *derivation* that transforms $p-$forms into $(p+1)-$forms, these geometrical objects generalize ordinary vector differential operators in $\mathbb{R}^3$:

- a *scalar function* $f = f(x)$ is a zero–form;
- its *gradient* df, is a one–form[4]

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz;$$

- a *curl* $d\omega$, of a one–form ω above, is a two–form

[4] We use the same symbol, d, to denote both ordinary and exterior derivation, in order to avoid extensive use of the boldface symbols. It is clear from the context which derivative (differential) is in place: exterior derivative operates only on differential forms, while the ordinary differential operates mostly on coordinates.

$$d\omega = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) dydz + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) dzdx + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dxdy;$$

- a *divergence* $d\alpha$, of the two–form α above, is a three–form

$$d\alpha = \left(\frac{\partial A}{\partial x} + \frac{\partial B}{\partial y} + \frac{\partial C}{\partial z}\right) dxdydz.$$

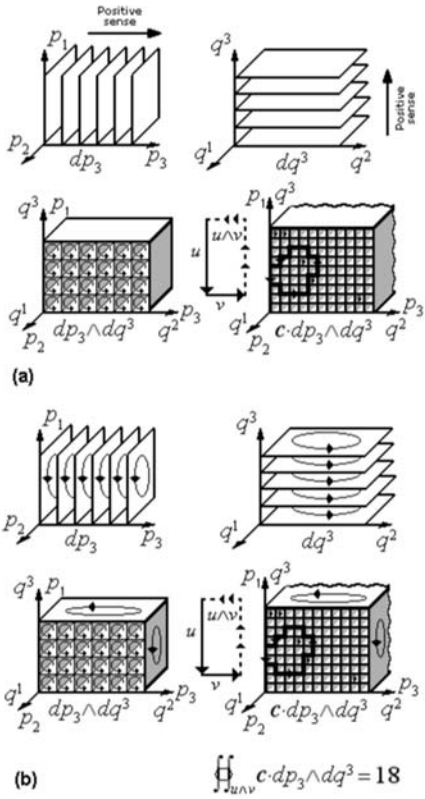


Fig. 3.3. Fundamental two–form and its flux in $\mathbb{R}^3$: (a) Translational case; (b) Rotational case. In both cases the flux through the plane $u\wedge v$ is defined as $\int\int_{u\wedge v} c\, dp_i dq^i$ and measured by the number of tubes crossed by the circulation oriented by $u \wedge v$.

Now, although visually intuitive, our Euclidean 3D space $\mathbb{R}^3$ is not sufficient for thorough physical or engineering analysis. The fundamental concept of a *smooth manifold*, locally topologically equivalent to the Euclidean nD space $\mathbb{R}^n$, is required (with or without Riemannian metric tensor defined on it). In general, a proper definition of exterior derivative d for a $p-$form β on a

smooth manifold M, includes the *Poincaré lemma*: $d(d\beta) = 0$, and validates the general integral *Stokes formula*

$$\int_{\partial M} \beta = \int_M d\beta,$$

where M is a $p-$dimensional *manifold with a boundary* and ∂M is its $(p-1)-$dimensional *boundary*, while the integrals have appropriate dimensions.

A $p-$form β is called *closed* if its exterior derivative is equal to zero,

$$d\beta = 0.$$

From this condition one can see that the closed form (the *kernel* of the exterior derivative operator d) is conserved quantity. Therefore, closed $p-$forms possess certain invariant properties, physically corresponding to the conservation laws.

A $p-$form β that is an exterior derivative of some $(p-1)-$form α,

$$\beta = d\alpha,$$

is called *exact* (the *image* of the exterior derivative operator d). By Poincaré Lemma, exact forms prove to be closed automatically,

$$d\beta = d(d\alpha) = 0.$$

Similarly to the components of a 3D vector v defined above, a one–form θ defined on an nD manifold M can also be expressed in components, using the coordinate basis $\{dx^i\}$ along the local nD coordinate chart $\{x^i\} \in M$, as

$$\theta \;=\; \theta_i \, dx^i.$$

Now, the components of the exterior derivative of θ are equal to the components of its *commutator* defined on M by

$$d\theta = \omega_{ij} \, dx^i \, dx^j,$$

where the components of the *form commutator* ω_{ij} are given by

$$\omega_{ij} \;=\; \left(\frac{\partial \theta_i}{\partial x^i} \;-\; \frac{\partial \theta_i}{\partial x^j} \right).$$

The space of all smooth $p-$forms on a smooth manifold M is denoted by $\Omega^p(M)$. The *wedge*, or *exterior product* of two differential forms, a $p-$form $\alpha \in \Omega^p(M)$ and a $q-$form $\beta \in \Omega^q(M)$ is a $(p+q)-$form $\alpha \wedge \beta$. For example, if $\theta = a_i dx^i$, and $\eta = b_j dx^j$, their wedge product $\theta \wedge \eta$ is given by

$$\theta \wedge \eta = a_i b_j dx^i dx^j,$$

so that the coefficients $a_i b_j$ of $\theta \wedge \eta$ are again smooth functions, being polynomials in the coefficients a_i of θ and b_j of η. The exterior product $\wedge$ is related to the exterior derivative $d : \Omega^p(M) \to \Omega^{p+1}(M)$, by

$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^p \alpha \wedge d\beta.$$

Another important linear operator is the *Hodge star* $* : \Omega^p(M) \to \Omega^{n-p}(M)$, where n is the dimension of the manifold M. This operator depends on the inner product (i.e., Riemannian metric) on M and also depends on the orientation (reversing orientation will change the sign). For any $p-$forms α and β,

$$** \alpha = (-1)^{p(n-p)}\alpha, \quad \text{and} \quad \alpha \wedge *\beta = \beta \wedge *\alpha.$$

Hodge star is generally used to define *dual* $(n-p)-$*forms* on nD smooth manifolds.

For example, in $\mathbb{R}^3$ with the ordinary Euclidean metric, if f and g are functions then (compare with the 3D forms of gradient, curl and divergence defined above)

$$\begin{aligned} df &= \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz, \\ *df &= \frac{\partial f}{\partial x}dydz + \frac{\partial f}{\partial y}dzdx + \frac{\partial f}{\partial z}dxdy, \\ df \wedge *dg &= \left(\frac{\partial f}{\partial x}\frac{\partial g}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial g}{\partial y} + \frac{\partial f}{\partial z}\frac{\partial g}{\partial z}\right) dxdydz = \Delta f \, dxdydz, \end{aligned}$$

where Δf is the *Laplacian* on $\mathbb{R}^3$. Therefore the three–form $df \wedge *dg$ is the Laplacian multiplied by the volume element, which is valid, more generally, in any local orthogonal coordinate system in any smooth domain $U \in \mathbb{R}^3$.

The subspace of all closed $p-$forms on M we will denote by $Z^p(M) \subset \Omega^p(M)$, and the sub-subspace of all exact $p-$forms on M we will denote by $B^p(M) \subset Z^p(M)$. Now, the quotient space

$$H^p(M) = \frac{Z^p(M)}{B^p M} = \frac{\text{Ker}\left(d : \Omega^p(M) \to \Omega^{p+1}(M)\right)}{\text{Im}\left(d : \Omega^{p-1}(M) \to \Omega^p(M)\right)}$$

is called the pth *de Rham cohomology group* (or vector space) of a manifold M. Two $p-$forms α and β on M are equivalent, or belong to the same *cohomology class* $[\alpha] \in H^p(M)$, if their difference equals $\alpha - \beta = d\theta$, where θ is a $(p-1)-$form on M.

3.1.5 The Covariant Force Law in Robotics and Biomechanics

Objective of this final tensor subsection is to generalize the fundamental Newtonian 3D equation, $\mathbf{F} = m\mathbf{a}$, for a generic robotic/biomechanical system, consisting of a number of flexibly–coupled rigid segments (see Figures 2.22–2.23 above), and thus to formulate the *covariant force law*.

To be able to apply the covariant formalism, we need to start with the suitable coordinate transformation (3.2), in this case as a relation between the 6 *external* $SE(3)$ rigid–body coordinates, $y^e = y^e(t)$ $(e = 1, ..., 6)$, and

$2n$ *internal joint coordinates*, $x^i = x^i(t)$ $(i = 1, ..., 2n)$ (n angles, forming the constrained $n-$torus T^n, plus n very restricted translational coordinates, forming the hypercube $I^n \subset \mathbb{R}^n$). Once we have these two sets of coordinates, external–y^e and internal–x^i, we can perform the general functional transformation (3.2) between them,

$$x^i = x^i(y^e). \tag{3.27}$$

Now, although the coordinate transformation (3.27) is nonlinear and even unknown at this stage, there is something known and simple about it: the corresponding transformation of differentials is *linear and homogenous*,

$$dx^i = \frac{\partial x^i}{\partial y^e} dy^e,$$

which implies the *linear and homogenous transformation of velocities,*

$$\dot{x}^i = \frac{\partial x^i}{\partial y^e} \dot{y}^e. \tag{3.28}$$

Our *internal velocity vector–field* is defined by the set of ODEs (3.28), at each *representative point* $x^i = x^i(t)$ of the system's configuration manifold $M = T^n \times I^n$, as $v^i \equiv v^i(x^i, t) := \dot{x}^i(x^i, t)$.

Note that in general, a *vector–field* represents a field of vectors defined at every point x^i within some region U (e.g., movable segments/joints only) of the total configuration manifold M (consisting of all the segments/joints). Analytically, vector–field is defined as a set of autonomous ODEs (in our case, the set (3.28)). Its solution gives the *flow*, consisting of *integral curves* of the vector–field, such that all the vectors from the vector–field are tangent to integral curves at different representative points $x^i \in U$. In this way, through every representative point $x^i \in U$ passes both a curve from the flow and its tangent vector from the vector–field. Geometrically, vector–field is defined as a cross–section of the tangent bundle TM, the so–called *velocity phase–space*. Its *geometrical dual* is the *1–form–field*, which represents a field of one–forms (see Figure 3.1), defined at the same representative points $x^i \in U$. Analytically, 1–form–field is defined as an *exterior differential system*, an algebraic dual to the autonomous set of ODEs. Geometrically, it is defined as a cross–section of the cotangent bundle T^*M, the so–called *momentum phase–space*. Together, the vector–field and its corresponding 1–form–field define the *scalar potential field* (e.g., kinetic and/or potential energy) at the same movable region $U \subset M$.

Next, we need to formulate the internal acceleration vector–field, $a^i \equiv a^i(x^i, \dot{x}^i, t)$, acting in all movable joints, and at the same time generalizing the Newtonian 3D acceleration vector $\mathbf{a}$.

According to Newton, acceleration is a rate–of–change of velocity. But, from the previous subsections, we know that $a^i \neq \dot{v}^i$. However,

$$a^i := \dot{\bar{v}}^i = \dot{v}^i + \Gamma^i_{jk} v^j v^k = \ddot{x}^i + \Gamma^i_{jk} \dot{x}^j \dot{x}^k. \tag{3.29}$$

Once we have the *internal acceleration vector–field* $a^i = a^i(x^i, \dot{x}^i, t)$, defined by the set of ODEs (3.29) (including *Levi–Civita connections* Γ^i_{jk} of the Riemannian configuration manifold M), we can finally define the *internal force 1–form field*, $F_i = F_i(x^i, \dot{x}^i, t)$, as a family of force one–forms, half of them rotational and half translational, acting in all movable joints,

$$F_i := mg_{ij}a^j = mg_{ij}(\dot{v}^j + \Gamma^j_{ik}v^iv^k) = mg_{ij}(\ddot{x}^j + \Gamma^j_{ik}\dot{x}^i\dot{x}^k), \tag{3.30}$$

where we have used the simplified *material metric tensor*, mg_{ij}, for the system (considering, for simplicity, all segments to have equal mass m), defined by its *Riemannian kinetic energy form*

$$T = \frac{1}{2}mg_{ij}v^iv^j.$$

Equation $F_i = mg_{ij}a^j$, defined properly by (3.30) at every representative point x^i of the system's configuration manifold M, formulates the sought for *covariant force law*, that generalizes the fundamental Newtonian equation, $\mathbf{F} = m\mathbf{a}$, for the generic physical or engineering system. Its meaning is:

Force 1–form–field = Mass distribution × Acceleration vector–field

In other words, *the field (or, family) of force one–forms F_i, acting in all movable joints (with constrained rotations on T^n and very restricted translations on I^n), causes both rotational and translational accelerations of all body segments, within the mass distribution mg_{ij}*[5]*, along the flow–lines of the vector–field a^j.*

3.2 Categories and Functors

In modern mathematical sciences whenever one defines a new class of mathematical objects, one proceeds almost in the next breath to say what kinds of maps between objects will be considered [Swi75]. A general framework for dealing with situations where we have some *objects* and *maps between objects*, like sets and functions, vector spaces and linear operators, points in a space and paths between points, etc. – gives the modern metalanguage of categories and functors. Categories are mathematical universes and functors are 'projectors' from one universe onto another. For this reason, in this book we extensively use this language, mainly following its founder, S. MacLane [MacL71].

[5] More realistically, instead of the simplified metric mg_{ij} we have the *material metric tensor* G_{ij} (3.20), including all k segmental masses m_χ, as well as the corresponding moments and products of inertia,

$$G_{ij}(x, m) = \sum_{\chi=1}^{k} m_\chi \delta_{rs} \frac{\partial y^r}{\partial x^i}\frac{\partial y^s}{\partial x^j}, \qquad (r, s = 1, ..., 6;\ i, j = 1, ..., 2n).$$

3.2.1 Maps

Notes from Set Theory

Given a map (or, a function) $f : A \to B$, the set A is called the *domain* of f, and denoted $\mathrm{Dom}\, f$. The set B is called the *codomain* of f, and denoted $\mathrm{Cod}\, f$. The codomain is not to be confused with the *range* of $f(A)$, which is in general only a subset of B.

A map $f : X \to Y$ is called *injective* or 1–1 or an *injection* if for every y in the codomain Y there is at most one x in the domain X with $f(x) = y$. Put another way, given x and x' in X, if $f(x) = f(x')$, then it follows that $x = x'$. A map $f : X \to Y$ is called *surjective* or *onto* or a *surjection* if for every y in the codomain $\mathrm{Cod}\, f$ there is at least one x in the *domain* X with $f(x) = y$. Put another way, the *range* $f(X)$ is equal to the codomain Y. A map is *bijective* iff it is both injective and surjective. Injective functions are called the *monomorphisms*, and surjective functions are called the *epimorphisms* in the *category of sets* (see below).

Two main classes of maps (or, functions) that we will use int this book are: (i) continuous maps (denoted as C^0-class), and (ii) smooth or differentiable maps (denoted as $C^\infty-$class). The former class is the core of topology, the letter of differential geometry. They are both used in the core concept of manifold.

A *relation* is any subset of a *Cartesian product* (see below). By definition, an *equivalence relation* α on a set X is a relation which is *reflexive, symmetrical* and *transitive*, i.e., relation that satisfies the following three conditions:

1. *Reflexivity*: each element $x \in X$ is equivalent to itself, i.e., $x \alpha x$,
2. *Symmetry*: for any two elements $x, x' \in X$, $x \alpha x'$ implies $x' \alpha x$, and
3. *Transitivity*: $a \leq b$ and $b \leq c$ implies $a \leq c$.

Similarly, a relation $\leq$ defines a *partial order* on a set S if it has the following properties:

1. *Reflexivity*: $a \leq a$ for all $a \in S$,
2. *Antisymmetry*: $a \leq b$ and $b \leq a$ implies $a = b$, and
3. *Transitivity*: $a \leq b$ and $b \leq c$ implies $a \leq c$.

A *partially ordered set* (or *poset*) is a set taken together with a partial order on it. Formally, a partially ordered set is defined as an ordered pair $P = (X, \leq)$, where X is called the *ground set* of P and $\leq$ is the partial order of P.

Notes From Calculus

Maps

Recall that a *map* (or, *function*) f is a *rule* that assigns to each element x in a set A exactly one element, called $f(x)$, in a set B. A map could be thought

of as a *machine* $[[f]]$ with $x-$input (the *domain* of f is the set of all possible inputs) and $f(x)-$output (the *range* of f is the set of all possible outputs) [Stu99]

$$x \to [[f]] \to f(x)$$

There are four possible ways to represent a function (or map): (i) verbally (by a description in words); (ii) numerically (by a table of values); (iii) visually (by a graph); and (iv) algebraically (by an explicit formula). The most common method for visualizing a function is its *graph*. If f is a function with domain A, then its graph is the set of ordered input–output pairs

$$\{(x, f(x)) : x \in A\}.$$

A generalization of the graph concept is a concept of a *cross–section of a fibre bundle*, which is one of the core geometrical objects for dynamics of complex systems.

Algebra of Maps

Let f and g be maps with domains A and B. Then the maps $f+g$, $f-g$, fg, and f/g are defined as follows [Stu99]

$$\begin{aligned}
(f+g)(x) &= f(x)+g(x) \qquad && \text{domain } = A \cap B,\\
(f-g)(x) &= f(x)-g(x) \qquad && \text{domain } = A \cap B,\\
(fg)(x) &= f(x)\,g(x) \qquad && \text{domain } = A \cap B,\\
\left(\frac{f}{g}\right)(x) &= \frac{f(x)}{g(x)} \qquad && \text{domain } = \{x \in A \cap B : g(x) \neq 0\}.
\end{aligned}$$

Compositions of Maps

Given two maps f and g, the composite map $f \circ g$ (also called the *composition* of f and g) is defined by

$$(f \circ g)(x) = f(g(x)).$$

The $(f \circ g)-$machine is composed of the $g-$machine (first) and then the $f-$machine [Stu99],

$$x \to [[g]] \to g(x) \to [[f]] \to f(g(x))$$

For example, suppose that $y = f(u) = \sqrt{u}$ and $u = g(x) = x^2 + 1$. Since y is a function of u and u is a function of x, it follows that y is ultimately a function of x. We calculate this by substitution

$$y = f(u) = f \circ g = f(g(x)) = f(x^2+1) = \sqrt{x^2+1}.$$

The Chain Rule

If f and g are both differentiable (or smooth, i.e., C^∞) maps and $h = f \circ g$ is the composite map defined by $h(x) = f(g(x))$, then h is differentiable and h' is given by the product [Stu99]

$$h'(x) = f'(g(x))\, g'(x).$$

In Leibniz notation, if $y = f(u)$ and $u = g(x)$ are both differentiable maps, then

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}.$$

The reason for the name *chain rule* becomes clear if we add another link to the chain. Suppose that we have one more differentiable map $x = h(t)$. Then, to calculate the derivative of y with respect to t, we use the chain rule twice,

$$\frac{dy}{dt} = \frac{dy}{du}\frac{du}{dx}\frac{dx}{dt}.$$

Integration and Change of Variables

1–1 continuous (i.e., C^0) map T with a nonzero *Jacobian* $\left|\frac{\partial(x,\ldots)}{\partial(u,\ldots)}\right|$ that maps a region S onto a region R, (see [Stu99]) we have the following substitution formulas:

1. for a single integral,

$$\int_R f(x)\, dx = \int_S f(x(u))\frac{\partial x}{\partial u} du,$$

2. for a double integral,

$$\iint_R f(x,y)\, dA = \iint_S f(x(u,v), y(u,v)) \left|\frac{\partial(x,y)}{\partial(u,v)}\right| dudv,$$

3. for a triple integral,

$$\iiint_R f(x,y,z)\, dV = \iint_S f(x(u,v,w), y(u,v,w), z(u,v,w)) \left|\frac{\partial(x,y,z)}{\partial(u,v,w)}\right| dudvdw$$

4. similarly for $n-$tuple integrals.

Notes from General Topology

Topology is a kind of *abstraction* of Euclidean geometry, and also a natural framework for the study of *continuity*.[6] Euclidean geometry is abstracted by

[6] Intuitively speaking, a function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous near a point x in its domain if its value does not jump there. That is, if we just take δx to be small

regarding triangles, circles, and squares as being the same basic object. Continuity enters because in saying this one has in mind a *continuous deformation* of a triangle into a square or a circle, or any arbitrary shape. On the other hand, a disk with a hole in the center is topologically different from a circle or a square because one cannot create or destroy holes by continuous deformations. Thus using topological methods one does not expect to be able to identify a geometrical figure as being a triangle or a square. However, one does expect to be able to detect the presence of gross features such as holes or the fact that the figure is made up of two disjoint pieces etc. In this way topology produces theorems that are usually qualitative in nature – they may assert, for example, the existence or non–existence of an object. They will not, in general, give the means for its construction [Nas83].

Topological Space

Study of topology starts with the fundamental notion of *topological space*. Let X be any *set* and $Y = \{X_\alpha\}$ denote a collection, finite or infinite of subsets of X. Then X and Y form a topological space provided the X_α and Y satisfy:

1. Any finite or infinite subcollection $\{Z_\alpha\} \subset X_\alpha$ has the property that $\cup Z_\alpha \in Y$, and
2. Any *finite subcollection* $\{Z_{\alpha_1}, ..., Z_{\alpha_n}\} \subset X_\alpha$ has the property that $\cap Z_{\alpha_i} \in Y$.

The set X is then called a topological space and the X_α are called *open sets*. The choice of Y satisfying (2) is said to give a topology to X.

Given two topological spaces X and Y, a *function* (or, a *map*) $f : X \to Y$ is *continuous* if the inverse image of an open set in Y is an open set in X.

The main general idea in topology is to study spaces which can be continuously deformed into one another, namely the idea of *homeomorphism*. If we have two topological spaces X and Y, then a map $f : X \to Y$ is called a homeomorphism iff

1. f is continuous (C^0), and
2. There exists an inverse of f, denoted f^{-1}, which is also continuous.

Definition (2) implies that if f is a homeomorphism then so is f^{-1}. Homeomorphism is the main topological example of *reflexive*, *symmetrical* and *transitive relation*, i.e., *equivalence relation*. Homeomorphism divides all topological spaces up into *equivalence classes*. In other words, a pair of topological spaces, X and Y, belong to the same equivalence class if they are homeomorphic.

---

enough, the two function values $f(x)$ and $f(x + \delta x)$ should approach each other arbitrarily closely. In more rigorous terms, this leads to the following definition: A function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous at $x \in \mathbb{R}$ if for all $\epsilon > 0$, there exists a $\delta > 0$ such that for all $y \in \mathbb{R}$ with $|y - x| < \delta$, we have that $|f(y) - f(x)| < \epsilon$. The whole function is called continuous if it is continuous at every point x.

The second example of topological equivalence relation is *homotopy*. While homeomorphism generates equivalence classes whose members are topological spaces, homotopy generates equivalence classes whose members are continuous (C^0) maps. Consider two continuous maps $f, g : X \to Y$ between topological spaces X and Y. Then the map f is said to be *homotopic* to the map g if f can be continuously deformed into g (see below for the precise definition of homotopy). Homotopy is an equivalence relation which divides the space of continuous maps between two topological spaces into equivalence classes [Nas83].

Another important notions in topology are *covering*, *compactness* and *connectedness*. Given a family of sets $\{X_\alpha\} = X$ say, then X is a *covering* of another set Y if $\cup X_\alpha$ contains Y. If all the X_α happen to be open sets the covering is called an *open covering*. Now consider the set Y and all its possible open coverings. The set Y is *compact* if for every open covering $\{X_\alpha\}$ with $\cup X_\alpha \supset Y$ there always exists a finite subcovering $\{X_1, ..., X_n\}$ of Y with $X_1 \cup ... \cup X_n \supset Y$. Again, we define a set Z to be *connected* if it cannot be written as $Z = Z_1 \cup Z_2$, where Z_1 and Z_2 are both open and $Z_1 \cap Z_2$ is an empty set.

Let $A_1, A_2, ..., A_n$ be closed subspaces of a topological space X such that $X = \cup_{i=1}^n A_i$. Suppose $f_i : A_i \to Y$ is a function, $1 \leq i \leq n$, iff

$$f_i | A_i \cap A_j = f_j | A_i \cap A_j, 1 \leq i, j \leq n. \tag{3.31}$$

In this case f is continuous iff each f_i is. Using this procedure we can define a C^0-function $f : X \to Y$ by cutting up the space X into closed subsets A_i and defining f on each A_i separately in such a way that $f|A_i$ is obviously continuous; we then have only to check that the different definitions agree on the *overlaps* $A_i \cap A_j$.

The *universal property of the Cartesian product*: let $p_X : X \times Y \to X$, and $p_Y : X \times Y \to Y$ be the *projections* onto the first and second factors, respectively. Given any pair of functions $f : Z \to X$ and $g : Z \to Y$ there is a unique function $h : Z \to X \times Y$ such that $p_X \circ h = f$, and $p_Y \circ h = g$. Function h is continuous iff both f and g are. This property characterizes X/α up to homeomorphism. In particular, to check that a given function $h : Z \to X$ is continuous it will suffice to check that $p_X \circ h$ and $p_Y \circ h$ are continuous.

The *universal property of the quotient*: let α be an equivalence relation on a topological space X, let X/α denote the *space of equivalence classes* and $p_\alpha : X \to X/\alpha$ the *natural projection*. Given a function $f : X \to Y$, there is a function $f' : X/\alpha \to Y$ with $f' \circ p_\alpha = f$ iff $x \alpha x'$ implies $f(x) = f(x')$, for all $x \in X$. In this case f' is continuous iff f is. This property characterizes X/α up to homeomorphism.

Homotopy

Now we return to the fundamental notion of homotopy. Let I be a compact unit interval $I = [0, 1]$. A *homotopy* from X to Y is a continuous function

$F : X \times I \to Y$. For each $t \in I$ one has $F_t : X \to Y$ defined by $F_t(x) = F(x,t)$ for all $x \in X$. The functions F_t are called the 'stages' of the homotopy. If $f, g : X \to Y$ are two continuous maps, we say f is homotopic to g, and write $f \simeq g$, if there is a homotopy $F : X \times I \to Y$ such that $F_0 = f$ and $F_1 = g$. In other words, f can be continuously deformed into g through the stages F_t. If $A \subset X$ is a subspace, then F is a homotopy relative to A if $F(a,t) = F(a,0)$, for all $a \in A, t \in I$.

The homotopy relation $\simeq$ is an equivalence relation. To prove that we have $f \simeq f$ is obvious; take $F(x, t = f(x)$, for all $x \in X$, $t \in I$. If $f \simeq g$ and F is a homotopy from f to g, then $G : X \times I \to Y$ defined by $G(x,t) = F(x, 1-t)$, is a homotopy from g to f, i.e., $g \simeq f$. If $f \simeq g$ with homotopy F and $g \simeq f$ with homotopy G, then $f \simeq h$ with homotopy H defined by

$$H(x,t) = \begin{cases} F(x,t), & 0 \leq t \leq 1/2 \\ G(x, 2t-1), & 1/2 \leq t \leq 1 \end{cases} .$$

To show that H is continuous we use the relation (3.31).

In this way, the set of all C^0-functions $f : X \to Y$ between two topological spaces X and Y, called the *function space* and denoted by Y^X, is partitioned into equivalence classes under the relation $\simeq$. The equivalence classes are called *homotopy classes*, the homotopy class of f is denoted by $[f]$, and the set of all homotopy classes is denoted by $[X;Y]$.

If α is an equivalence relation on a topological space X and $F : X \times I \to Y$ is a homotopy such that each stage F_t factors through X/α, i.e., $x \alpha x'$ implies $F_t(x) = F_t(x')$, then F induces a homotopy $F' : (X/\alpha) \times I \to Y$ such that $F' \circ (p_\alpha \times 1) = F$.

Homotopy theory has a range of applications of its own, outside topology and geometry, as for example in proving Cauchy theorem in complex variable theory, or in solving nonlinear equations of artificial neural networks.

A *pointed set* (S, s_0) is a set S together with a distinguished point $s_0 \in S$. Similarly, a *pointed topological space* (X, x_0) is a space X together with a distinguished point $x_0 \in X$. When we are concerned with pointed spaces $(X, x_0), (Y, y_0)$, etc., we always require that all functions $f : X \to Y$ shell preserve base points, i.e., $f(x_0) = y_0$, and that all homotopies $F : X \times I \to Y$ be relative to the base point, i.e., $F(x_0, t) = y_0$, for all $t \in I$. We denote the homotopy classes of base point–preserving functions by $[X, x_0; Y, y_0]$ (where homotopies are relative to x_0). $[X, x_0; Y, y_0]$ is a pointed set with base point f_0, the constant function: $f_0(x) = y_0$, for all $x \in X$.

A *path* $\gamma(t)$ from x_0 to x_1 in a topological space X is a continuous map $\gamma : I \to X$ with $\gamma(0) = x_0$ and $\gamma(1) = x_1$. Thus X^I is the space of all paths in X with the compact–open topology. We introduce a relation $\sim$ on X by saying $x_0 \sim x_1$ iff there is a path $\gamma : I \to X$ from x_0 to x_1. $\sim$ is clearly an equivalence relation, and the set of equivalence classes is denoted by $\pi_0(X)$. The elements of $\pi_0(X)$ are called the *path components*, or $0-$*components* of X. If $\pi_0(X)$ contains just one element, then X is called *path connected*, or $0-$*connected*. A *closed path*, or *loop* in X at the point x_0 is a path $\gamma(t)$ for

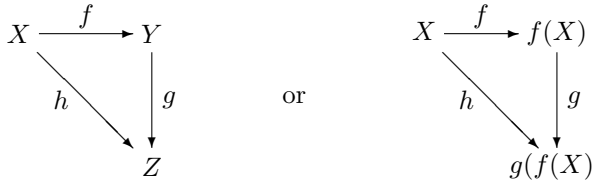
which $\gamma(0) = \gamma(1) = x_0$. The *inverse loop* $\gamma^{-1}(t)$ based at $x_0 \in X$ is defined by $\gamma^{-1}(t) = \gamma(1-t)$, for $0 \leq t \leq 1$. The *homotopy of loops* is the particular case of the above defined homotopy of continuous maps.

If (X, x_0) is a pointed space, then we may regard $\pi_0(X)$ as a pointed set with the 0−component of x_0 as a base point. We use the notation $\pi_0(X, x_0)$ to denote $p_0(X, x_0)$ thought of as a pointed set. If $f : X \to Y$ is a map then f sends 0−components of X into 0−components of Y and hence defines a function $\pi_0(f) : \pi_0(X) \to \pi_0(Y)$. Similarly, a base point–preserving map $f : (X, x_0) \to (Y, y_0)$ induces a map of pointed sets $\pi_0(f) : \pi_0(X, x_0) \to \pi_0(Y, y_0)$. In this way defined π_0 represents a 'functor' from the 'category' of topological (point) spaces to the underlying category of (point) sets (see the next subsection).

Combination of topology and calculus gives differential topology, or differential geometry.

Commutative Diagrams

The *category theory* (see below) was born with an observation that many properties of mathematical systems can be unified and simplified by a presentation with *commutative diagrams of arrows* [MacL71]. Each arrow $f : X \to Y$ represents a function (i.e., a map, transformation, operator); that is, a source (domain) set X, a target (codomain) set Y, and a rule $x \mapsto f(x)$ which assigns to each element $x \in X$ an element $f(x) \in Y$. A typical diagram of sets and functions is



This diagram is *commutative* iff $h = g \circ f$, where $g \circ f$ is the usual composite function $g \circ f : X \to Z$, defined by $x \mapsto g(f(x))$.

Similar commutative diagrams apply in other mathematical, physical and computing contexts; e.g., in the 'category' of all topological spaces, the letters X, Y, and Z represent topological spaces while f, g, and h stand for continuous maps. Again, in the category of all groups, X, Y, and Z stand for groups, f, g, and h for homomorphisms.

Less formally, composing maps is like following directed paths from one object to another (e.g., from set to set). In general, a diagram is commutative iff any two paths along arrows that start at the same point and finish at the same point yield the same 'homomorphism' via compositions along successive arrows. Commutativity of the whole diagram follows from commutativity of its triangular components (depicting a 'commutative flow', see Figure 3.4). Study of commutative diagrams is popularly called 'diagram chasing', and provides a powerful tool for mathematical thought.

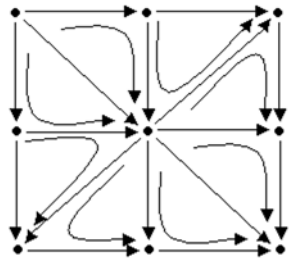
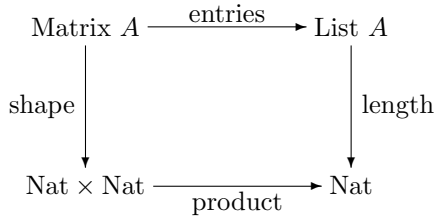


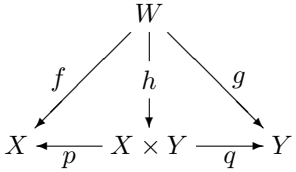
Fig. 3.4. A *commutative flow* (denoted by curved arrows) on a *triangulated digraph*. Commutativity of the whole diagram follows from commutativity of its triangular components.

As an example from linear algebra, consider an elementary diagrammatic description of matrices, using the following *pull–back diagram* [Bar93]:



asserts that a matrix is determined by its shape, given by a pair of natural numbers representing the number of rows and columns, and its data, given by the matrix entries listed in some specified order.

Many properties of mathematical constructions may be represented by *universal properties* of diagrams [MacL71]. Consider the *Cartesian product* $X \times Y$ of two sets, consisting as usual of all ordered pairs $\langle x, y\rangle$ of elements $x \in X$ and $y \in Y$. The projections $\langle x, y\rangle \mapsto x$, $\langle x, y\rangle \mapsto y$ of the product on its 'axes' X and Y are functions $p : X \times Y \to X$, $q : X \times Y \to Y$. Any function $h : W \to X \times Y$ from a third set W is uniquely determined by its composites $p \circ h$ and $q \circ h$. Conversely, given W and two functions f and g as in the diagram below, there is a unique function h which makes the following diagram commute:



This property describes the Cartesian product $X \times Y$ uniquely; the same diagram, read in the category of topological spaces or of groups, describes uniquely the Cartesian product of spaces or of the direct product of groups.

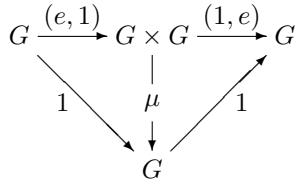
The construction 'Cartesian product' is technically called a 'functor' because it applies suitably both to the sets and to the functions between them; two functions $k : X \to X'$ and $l : Y \to Y'$ have a function $k \times l$ as their Cartesian product:

$$k \times l : X \times Y \to X' \times Y', \qquad \langle x, y \rangle \mapsto \langle kx, ly \rangle.$$

Groups and Related Algebraic Structures

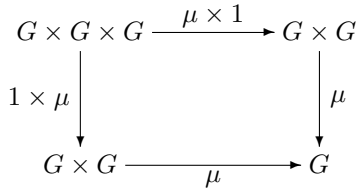
As already stated, the basic functional unit of lower biomechanics is the special Euclidean group $SE(3)$ of rigid body motions. In general, a *group* is a pointed set (G, e) with a *multiplication* $\mu : G \times G \to G$ and an *inverse* $\nu : G \to G$ such that the following diagrams commute [Swi75]:

1.



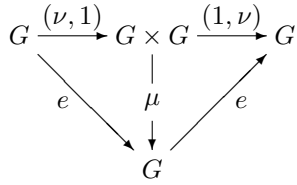
(e is a two–sided identity)

2.



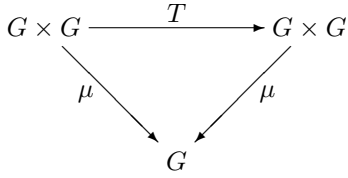
(associativity)

3.



(inverse).

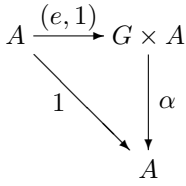
Here $e : G \to G$ is the constant map $e(g) = e$ for all $g \in G$. $(e, 1)$ means the map such that $(e, 1)(g) = (e, g)$, etc. A group G is called *commutative* or *Abelian group* if in addition the following diagram commutes



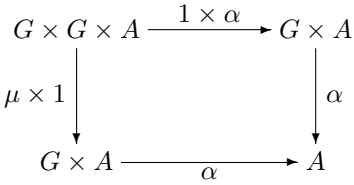
where $T : G \times G \to G \times G$ is the switch map $T(g_1, g_2) = (g_1, g_2)$, for all $(g_1, g_2) \in G \times G$.

A group G *acts* (on the left) on a set A if there is a function $\alpha : G \times A \to A$ such that the following diagrams commute [Swi75]:

1.



2.



where $(e, 1)(x) = (e, x)$ for all $x \in A$. The *orbits* of the action are the sets $Gx = \{gx : g \in G\}$ for all $x \in A$.

Given two groups $(G, *)$ and $(H, \cdot)$, a *group homomorphism* from $(G, *)$ to $(H, \cdot)$ is a function $h : G \to H$ such that for all x and y in G it holds that

$$h(x * y) = h(x) \cdot h(y).$$

From this property, one can deduce that h maps the identity element e_G of G to the identity element e_H of H, and it also maps inverses to inverses in the sense that $h(x^{-1}) = h(x)^{-1}$. Hence one can say that h is *compatible* with the *group structure*.

The *kernel* $\operatorname{Ker} h$ of a group homomorphism $h : G \to H$ consists of all those elements of G which are sent by h to the identity element e_H of H, i.e.,

$$\operatorname{Ker} h = \{x \in G : h(x) = e_H\}.$$

The *image* $\operatorname{Im} h$ of a group homomorphism $h : G \to H$ consists of all elements of G which are sent by h to H, i.e.,

$$\operatorname{Im} h = \{h(x) : x \in G\}.$$

The kernel is a *normal subgroup* of G and the image is a *subgroup* of H. The homomorphism h is *injective* (and called a *group monomorphism*) iff $\operatorname{Ker} h = e_G$, i.e., iff the kernel of h consists of the identity element of G only.

Similarly, a *ring* is a set S together with two binary operators $+$ and $*$ (commonly interpreted as addition and multiplication, respectively) satisfying the following conditions:

1. Additive associativity: For all $a, b, c \in S$, $(a + b) + c = a + (b + c)$,
2. Additive commutativity: For all $a, b \in S$, $a + b = b + a$,
3. Additive identity: There exists an element $0 \in S$ such that for all $a \in S$, $0 + a = a + 0 = a$,
4. Additive inverse: For every $a \in S$ there exists $-a \in S$ such that $a+(-a) = (-a) + a = 0$,
5. Multiplicative associativity: For all $a, b, c \in S$, $(a * b) * c = a * (b * c)$,
6. Left and right distributivity: For all $a, b, c \in S$, $a*(b+c) = (a*b)+(a*c)$ and $(b + c) * a = (b * a) + (c * a)$.

A ring (the term introduced by *David Hilbert*) is therefore an Abelian group under addition and a semigroup under multiplication. A ring that is commutative under multiplication, has a unit element, and has no divisors of zero is called an *integral domain*. A ring which is also a commutative multiplication group is called a *field*. The simplest rings are the integers $\mathbb{Z}$, polynomials $R[x]$ and $R[x, y]$ in one and two variables, and square $n \times n$ real matrices.

An *ideal* is a subset $\mathfrak{I}$ of elements in a ring R which forms an additive group and has the property that, whenever x belongs to R and y belongs to $\mathfrak{I}$, then xy and yx belong to $\mathfrak{I}$. For example, the set of even integers is an ideal in the ring of integers $\mathbb{Z}$. Given an ideal $\mathfrak{I}$, it is possible to define a factor ring $R/\mathfrak{I}$.

A ring is called *left* (respectively, *right*) *Noetherian* if it does not contain an infinite ascending chain of left (respectively, right) ideals. In this case, the ring in question is said to satisfy the ascending chain condition on left (respectively, right) ideals. A *ring* is said to be *Noetherian* if it is both left and right Noetherian. If a ring R is Noetherian, then the following are equivalent:

1. R satisfies the ascending chain condition on ideals.
2. Every ideal of R is finitely generated.
3. Every set of ideals contains a maximal element.

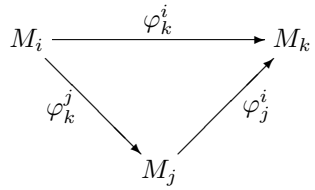
A *module* is a mathematical object in which things can be added together commutatively by multiplying coefficients and in which most of the rules of manipulating vectors hold. A module is abstractly very similar to a vector space, although in modules, coefficients are taken in rings which are much more general algebraic objects than the fields used in vector spaces. A module taking its coefficients in a ring R is called a module over R or $R-$module. Modules are the basic tool of homological algebra.

Examples of modules include the set of integers $\mathbb{Z}$, the cubic lattice in d dimensions $\mathbb{Z}^d$, and the group ring of a group. $\mathbb{Z}$ is a module over itself. It is closed under addition and subtraction. Numbers of the form $n\alpha$ for $n \in \mathbb{Z}$ and α a fixed integer form a submodule since, for $(n, m) \in \mathbb{Z}$, $n\alpha \pm m\alpha = (n \pm m)\alpha$ and $(n \pm m)$ is still in $\mathbb{Z}$. Also, given two integers a and b, the smallest module containing a and b is the module for their greatest common divisor, $\alpha = GCD(a, b)$.

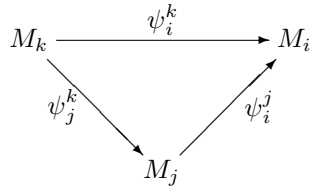
A module M is a *Noetherian module* if it obeys the ascending chain condition with respect to inclusion, i.e., if every set of increasing sequences of submodules eventually becomes constant. If a module M is Noetherian, then the following are equivalent:

1. M satisfies the ascending chain condition on submodules.
2. Every submodule of M is finitely generated.
3. Every set of submodules of M contains a maximal element.

Let I be a partially ordered set. A *direct system* of $R-$*modules* over I is an ordered pair $\{M_i, \varphi_j^i\}$ consisting of an indexed family of modules $\{M_i : i \in I\}$ together with a family of homomorphisms $\{\varphi_j^i : M_i \to M_j\}$ for $i \leq j$, such that $\varphi_i^i = 1_{M_i}$ for all i and such that the following diagram commutes whenever $i \leq j \leq k$



Similarly, an *inverse system* of $R-$*modules* over I is an ordered pair $\{M_i, \psi_i^j\}$ consisting of an indexed family of modules $\{M_i : i \in I\}$ together with a family of homomorphisms $\{\psi_i^j : M_j \to M_i\}$ for $i \leq j$, such that $\psi_i^i = 1_{M_i}$ for all i and such that the following diagram commutes whenever $i \leq j \leq k$



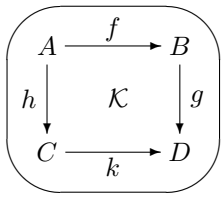
3.2.2 Categories

A category is a generic mathematical structure consisting of a collection of *objects* (sets with possibly additional structure), with a corresponding collection of *arrows*, or *morphisms*, between objects (agreeing with this additional structure). A category $\mathcal{K}$ is defined as a pair $(\mathtt{Ob}(\mathcal{K}), \mathtt{Mor}(\mathcal{K}))$ of generic objects $A, B, \ldots$ in $\mathtt{Ob}(\mathcal{K})$ and generic arrows $f : A \to B$, $g : B \to C, \ldots$ in $\mathtt{Mor}(\mathcal{K})$ between objects, with *associative composition*:

$$A \xrightarrow{f} B \xrightarrow{g} C = A \xrightarrow{f \circ g} C,$$

and *identity* (*loop*) arrow. (Note that in topological literature, $\mathtt{Hom}(\mathcal{K})$ or $\mathtt{hom}(\mathcal{K})$ is used instead of $\mathtt{Mor}(\mathcal{K})$; see [Swi75]).

A category $\mathcal{K}$ is usually depicted as a *commutative diagram* (i.e., a diagram with a common *initial object* A and *final object* D):



To make this more precise, we say that a *category* $\mathcal{K}$ is defined if we have:

1. A *class of objects* $\{A, B, C, ...\}$ of $\mathcal{K}$, denoted by $\mathtt{Ob}(\mathcal{K})$;
2. A *set of morphisms*, or *arrows* $\mathtt{Mor}_{\mathcal{K}}(A, B)$, with elements $f : A \to B$, defined for any *ordered pair* $(A, B) \in \mathcal{K}$, such that for two different pairs $(A, B) \neq (C, D)$ in $\mathcal{K}$, we have $\mathtt{Mor}_{\mathcal{K}}(A, B) \cap \mathtt{Mor}_{\mathcal{K}}(C, D) = \emptyset$;
3. For any *triplet* $(A, B, C) \in \mathcal{K}$ with $f : A \to B$ and $g : B \to C$, there is a *composition* of morphisms

$$\mathtt{Mor}_{\mathcal{K}}(B, C) \times \mathtt{Mor}_{\mathcal{K}}(A, B) \ni (g, f) \to g \circ f \in \mathtt{Mor}_{\mathcal{K}}(A, C),$$

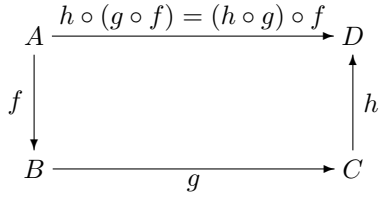
written schematically as

$$\frac{f : A \to B, \qquad g : B \to C}{g \circ f : A \to C}.$$

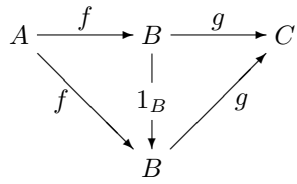
If we have a morphism $f \in \mathtt{Mor}_{\mathcal{K}}(A, B)$, (otherwise written $f : A \to B$, or $A \xrightarrow{f} B$), then $A = \mathrm{dom}(f)$ is a *domain* of f, and $B = \mathrm{cod}(f)$ is a *codomain* of f (of which *range* of f is a subset) and denoted $B = \mathrm{ran}(f)$.

To make $\mathcal{K}$ a category, it must also fulfill the following two properties:

1. *Associativity of morphisms*: for all $f \in \mathtt{Mor}_{\mathcal{K}}(A, B)$, $g \in \mathtt{Mor}_{\mathcal{K}}(B, C)$, and $h \in \mathtt{Mor}_{\mathcal{K}}(C, D)$, we have $h \circ (g \circ f) = (h \circ g) \circ f$; in other words, the following diagram is commutative



2. *Existence of identity morphism*: for every object $A \in \mathtt{Ob}(\mathcal{K})$ exists a unique identity morphism $1_A \in \mathtt{Mor}_{\mathcal{K}}(A, A)$; for any two morphisms $f \in \mathtt{Mor}_{\mathcal{K}}(A, B)$, and $g \in \mathtt{Mor}_{\mathcal{K}}(B, C)$, compositions with identity morphism $1_B \in \mathtt{Mor}_{\mathcal{K}}(B, B)$ give $1_B \circ f = f$ and $g \circ 1_B = g$, i.e., the following diagram is commutative:



The set of all morphisms of the category $\mathcal{K}$ is denoted

$$\mathtt{Mor}(\mathcal{K}) = \bigcup_{A,B \in Ob(\mathcal{K})} \mathtt{Mor}_{\mathcal{K}}(A,B).$$

If for two morphisms $f \in \mathtt{Mor}_{\mathcal{K}}(A,B)$ and $g \in \mathtt{Mor}_{\mathcal{K}}(B,A)$ the equality $g \circ f = 1_A$ is valid, then the morphism g is said to be *left inverse* (or *retraction*), of f, and f *right inverse* (or *section*) of g. A morphism which is both right and left inverse of f is said to be *two–sided inverse* of f.

A morphism $m : A \to B$ is called *monomorphism* in $\mathcal{K}$ (i.e., *1–1*, or *injection* map), if for any two parallel morphisms $f_1, f_2 : C \to A$ in $\mathcal{K}$ the equality $m \circ f_1 = m \circ f_2$ implies $f_1 = f_2$; in other words, m is monomorphism if it is *left cancellable*. Any morphism with a left inverse is monomorphism.

A morphism $e : A \to B$ is called *epimorphism* in $\mathcal{K}$ (i.e., *onto*, or *surjection* map), if for any two morphisms $g_1, g_2 : B \to C$ in $\mathcal{K}$ the equality $g_1 \circ e = g_2 \circ e$ implies $g_1 = g_2$; in other words, e is epimorphism if it is *right cancellable*. Any morphism with a right inverse is epimorphism.

A morphism $f : A \to B$ is called *isomorphism* in $\mathcal{K}$ (denoted as $f : A \cong B$) if there exists a morphism $f^{-1} : B \to A$ which is a two–sided inverse of f in $\mathcal{K}$. The relation of isomorphism is reflexive, symmetric, and transitive, i.e., equivalence relation.

For example, an isomorphism in the category of sets is called a set–isomorphism, or a *bijection*, in the category of topological spaces is called a topological isomorphism, or a *homeomorphism*, in the category of differentiable manifolds is called a differentiable isomorphism, or a *diffeomorphism*.

A morphism $f \in \mathtt{Mor}_{\mathcal{K}}(A,B)$ is *regular* if there exists a morphism $g : B \to A$ in $\mathcal{K}$ such that $f \circ g \circ f = f$. Any morphism with either a left or a right inverse is regular.

An object T is a *terminal object* in $\mathcal{K}$ if to each object $A \in \mathtt{Ob}(\mathcal{K})$ there is exactly one arrow $A \to T$. An object S is an *initial object* in $\mathcal{K}$ if to each object $A \in \mathtt{Ob}(\mathcal{K})$ there is exactly one arrow $S \to A$. A *null object* $Z \in \mathtt{Ob}(\mathcal{K})$ is an object which is both initial and terminal; it is unique up to isomorphism. For any two objects $A, B \in \mathtt{Ob}(\mathcal{K})$ there is a unique morphism $A \to Z \to B$ (the composite through Z), called the *zero morphism* from A to B.

A notion of subcategory is analogous to the notion of subset. A subcategory $\mathcal{L}$ of a category $\mathcal{K}$ is said to be a *complete subcategory* iff for any objects $A, B \in \mathcal{L}$, every morphism $A \to B$ of $\mathcal{L}$ is in $\mathcal{K}$.

A *groupoid* is a category in which every morphism is invertible. A typical groupoid is the *fundamental groupoid* $\Pi_1(X)$ of a topological space X. An

object of $\Pi_1(X)$ is a point $x \in X$, and a morphism $x \to x'$ of $\Pi_1(X)$ is a homotopy class of paths f from x to x'. The *composition* of paths $g : x' \to x''$ and $f : x \to x'$ is the path h which is 'f followed by g'. Composition applies also to homotopy classes, and makes $\Pi_1(X)$ a category and a groupoid (the inverse of any path is the same path traced in the opposite direction).

A *group* is a groupoid with one object, i.e., a *category with one object* in which *all morphisms are isomorphisms*. Therefore, if we try to generalize the concept of a group, keeping associativity as an essential property, we get the notion of a category.

A category is *discrete* if every morphism is an identity. A *monoid* is a category with one object. A *group* is a category with one object in which every morphism has a two–sided inverse under composition.

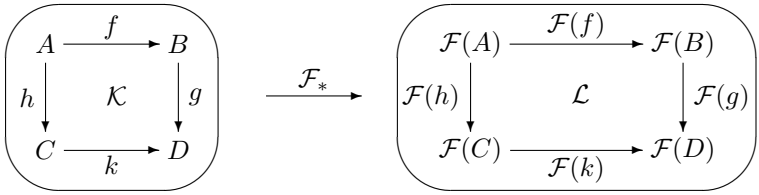
Homological algebra was the progenitor of category theory (see e.g., [Die88]). Generalizing L. Euler's formula $f + v = e + 2$ for the faces, vertices and edges of a convex polyhedron, E. Betti defined *numerical invariants of spaces* by formal addition and subtraction of faces of various dimensions; H. Poincaré formalized these and introduced homology. E. Noether stressed the fact that these calculations go on in Abelian groups, and that the operation ∂_n taking a face of dimension n to the alternating sum of faces of dimension $n-1$ which form its boundary is a homomorphism, and it also satisfies $\partial_n \circ \partial_{n+1} = 0$. There are many ways of approximating a given space by polyhedra, but the quotient $H_n = \operatorname{Ker} \partial_n / \operatorname{Im} \partial_{n+1}$ is an invariant, the *homology group*. Since Noether, the groups have been the object of study instead of their dimensions, which are the *Betti numbers*.

3.2.3 Functors

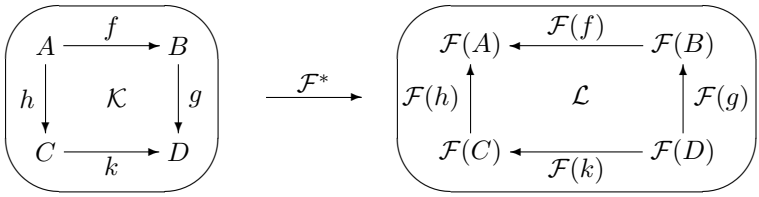
In algebraic topology, one attempts to assign to every topological space X some algebraic object $\mathcal{F}(X)$ in such a way that to every C^0-function $f : X \to Y$ there is assigned a homomorphism $\mathcal{F}(f) : \mathcal{F}(X) \longrightarrow \mathcal{F}(Y)$ (see [Swi75, DP97]). One advantage of this procedure is, e.g., that if one is trying to prove the non–existence of a C^0-function $f : X \to Y$ with certain properties, one may find it relatively easy to prove the non–existence of the corresponding algebraic function $\mathcal{F}(f)$ and hence deduce that f could not exist. In other words, $\mathcal{F}$ is to be a 'homomorphism' from one category (e.g., $\mathcal{T}$) to another (e.g., $\mathcal{G}$ or $\mathcal{A}$). Formalization of this notion is a *functor*.

A functor is a generic *picture* projecting one category into another. Let $\mathcal{K} = (\mathtt{Ob}(\mathcal{K}), \mathtt{Mor}(\mathcal{K}))$ be a *source* (or domain) *category* and $\mathcal{L} = (\mathtt{Ob}(\mathcal{L}), \mathtt{Mor}(\mathcal{L}))$ be a *target* (or codomain) category. A functor $\mathcal{F} = (\mathcal{F}_O, \mathcal{F}_M)$ is defined as a pair of maps, $\mathcal{F}_O : \mathtt{Ob}(\mathcal{K}) \to \mathtt{Ob}(\mathcal{L})$ and $\mathcal{F}_M : \mathtt{Mor}(\mathcal{K}) \to \mathtt{Mor}(\mathcal{L})$, preserving categorical symmetry (i.e., commutativity of all diagrams) of $\mathcal{K}$ in $\mathcal{L}$.

More precisely, a *covariant functor*, or simply a *functor*, $\mathcal{F}_* : \mathcal{K} \to \mathcal{L}$ is a *picture* in the target category $\mathcal{L}$ of (all objects and morphisms of) the source category $\mathcal{K}$:



Similarly, a *contravariant functor*, or a *cofunctor*, $\mathcal{F}^* : \mathcal{K} \to \mathcal{L}$ is a *dual picture* with reversed arrows:



In other words, a *functor* $\mathcal{F} : \mathcal{K} \to \mathcal{L}$ from a *source* category $\mathcal{K}$ to a *target* category $\mathcal{L}$, is a pair $\mathcal{F} = (\mathcal{F}_O, \mathcal{F}_M)$ of maps $\mathcal{F}_O : \mathtt{Ob}(\mathcal{K}) \to \mathtt{Ob}(\mathcal{L})$, $\mathcal{F}_M : \mathtt{Mor}(\mathcal{K}) \to \mathtt{Mor}(\mathcal{L})$, such that

1. If $f \in \mathtt{Mor}_{\mathcal{K}}(A, B)$ then $\mathcal{F}_M(f) \in \mathtt{Mor}_{\mathcal{L}}(\mathcal{F}_O(A), \mathcal{F}_O(B))$ in case of the *covariant* functor $\mathcal{F}_*$, and $\mathcal{F}_M(f) \in \mathtt{Mor}_{\mathcal{L}}(\mathcal{F}_O(B), \mathcal{F}_O(A))$ in case of the *contravariant* functor $\mathcal{F}^*$;
2. For all $A \in \mathtt{Ob}(\mathcal{K}) : \mathcal{F}_M(1_A) = 1_{\mathcal{F}_O(A)}$;
3. For all $f, g \in \mathtt{Mor}(\mathcal{K})$: if $\mathrm{cod}(f) = \mathrm{dom}(g)$, then $\mathcal{F}_M(g \circ f) = \mathcal{F}_M(g) \circ \mathcal{F}_M(f)$ in case of the *covariant* functor $\mathcal{F}_*$, and $\mathcal{F}_M(g \circ f) = \mathcal{F}_M(f) \circ \mathcal{F}_M(g)$ in case of the *contravariant* functor $\mathcal{F}^*$.

Category theory originated in algebraic topology, which tried to assign algebraic invariants to topological structures. The golden rule of such *invariants* is that they should be *functors*. For example, the *fundamental group* π_1 is a functor. Algebraic topology constructs a group called the *fundamental group* $\pi_1(X)$ from any topological space X, which keeps track of how many holes the space X has. But also, any map between topological spaces determines a homomorphism $\phi : \pi_1(X) \to \pi_1(Y)$ of the fundamental groups. So the fundamental group is really a functor $\pi_1 : \mathcal{T} \to \mathcal{G}$. This allows us to completely transpose any situation involving *spaces* and *continuous maps* between them to a parallel situation involving *groups* and *homomorphisms* between them, and thus reduce some topology problems to algebra problems.

Also, singular homology in a given dimension n assigns to each topological space X an Abelian group $H_n(X)$, its nth *homology group* of X, and also to each continuous map $f : X \to Y$ of spaces a corresponding homomorphism $H_n(f) : H_n(X) \to H_n(Y)$ of groups, and this in such a way that $H_n(X)$ becomes a functor $H_n : \mathcal{T} \to \mathcal{A}$.

The leading idea in the *use of functors in topology* is that H_n or π_n gives an algebraic picture or image not just of the topological spaces X, Y but also of all the continuous maps $f : X \to Y$ between them.

Similarly, there is a functor $\Pi_1 : \mathcal{T} \to \mathcal{G}$, called the 'fundamental groupoid functor', which plays a very basic role in algebraic topology. Here's how we get from any space X its 'fundamental groupoid' $\Pi_1(X)$. To say what the groupoid $\Pi_1(X)$ is, we need to say what its objects and morphisms are. The objects in $\Pi_1(X)$ are just the *points* of X and the morphisms are just certain equivalence classes of *paths* in X. More precisely, a morphism $f : x \to y$ in $\Pi_1(X)$ is just an equivalence class of continuous paths from x to y, where two paths from x to y are decreed equivalent if one can be continuously deformed to the other while not moving the endpoints. (If this equivalence relation holds we say the two paths are 'homotopic', and we call the equivalence classes 'homotopy classes of paths' (see [MacL71, Swi75]).

Another examples are covariant *forgetful* functors:

- From the category of topological spaces to the category of sets; it 'forgets' the topology–structure.
- From the category of metric spaces to the category of topological spaces with the topology induced by the metrics; it 'forgets' the metric.

For each category $\mathcal{K}$, the *identity functor* $I_{\mathcal{K}}$ takes every $\mathcal{K}-$object and every $\mathcal{K}-$morphism to itself.

Given a category $\mathcal{K}$ and its subcategory $\mathcal{L}$, we have an *inclusion functor* $\mathtt{In} : \mathcal{K} \longrightarrow \mathcal{K}$.

Given a category $\mathcal{K}$, a *diagonal functor* $\Delta : \mathcal{K} \longrightarrow \mathcal{K}$ takes each object $A \in \mathcal{K}$ to the object (A, A) in the product category $\mathcal{K} \times \mathcal{K}$.

Given a category $\mathcal{K}$ and a category of sets $\mathcal{S}$, each object $A \in \mathcal{K}$ determines a *covariant Hom–functor* $\mathcal{K}[A, \_] : \mathcal{K} \to \mathcal{S}$, a *contravariant Hom–functor* $\mathcal{K}[\_, A] : \mathcal{K} \longrightarrow \mathcal{S}$, and a *Hom–bifunctor* $\mathcal{K}[\_, \_] : \mathcal{K}^{op} \times \mathcal{K} \to \mathcal{S}$.

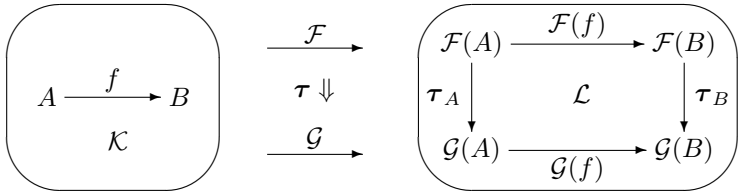
A functor $\mathcal{F} : \mathcal{K} \to \mathcal{L}$ is a *faithful functor* if for all $A, B \in \mathtt{Ob}(\mathcal{K})$ and for all $f, g \in \mathtt{Mor}_{\mathcal{K}}(A, B)$, $\mathcal{F}(f) = \mathcal{F}(g)$ implies $f = g$; it is a *full functor* if for every $h \in \mathtt{Mor}_{\mathcal{L}}(\mathcal{F}(A), \mathcal{F}(B))$, there is $g \in \mathtt{Mor}_{\mathcal{K}}(A, B)$ such that $h = \mathcal{F}(g)$; it is a *full embedding* if it is both full and faithful.

A *representation of a group* is a functor $\mathcal{F} : \mathcal{G} \to \mathcal{V}$.

Similarly, we can define a *representation of a category* to be a functor $\mathcal{F} : \mathcal{K} \to \mathcal{V}$ from the 2−category $\mathcal{K}$ (a 'big' category including all ordinary, or 'small' categories, see subsection (3.2.7) below) to the category of vector spaces $\mathcal{V}$. In this way, a category is a generalization of a group and group representations are a special case of category representations.

3.2.4 Natural Transformations

A *natural transformation* (i.e., a *functor morphism*) $\boldsymbol{\tau} : \mathcal{F} \stackrel{\cdot}{\to} \mathcal{G}$ is a *map between two functors of the same variance*, $(\mathcal{F}, \mathcal{G}) : \mathcal{K} \rightrightarrows \mathcal{L}$, preserving categorical symmetry:



More precisely, all functors of the same variance from a source category $\mathcal{K}$ to a target category $\mathcal{L}$ form themselves objects of the *functor category* $\mathcal{L}^{\mathcal{K}}$. Morphisms of $\mathcal{L}^{\mathcal{K}}$, called *natural transformations*, are defined as follows.

Let $\mathcal{F} : \mathcal{K} \to \mathcal{L}$ and $\mathcal{G} : \mathcal{K} \to \mathcal{L}$ be two functors of the same variance from a category $\mathcal{K}$ to a category $\mathcal{L}$. Natural transformation $\mathcal{F} \xrightarrow{\boldsymbol{\tau}} \mathcal{G}$ is a family of morphisms such that for all $f \in \mathtt{Mor}_{\mathcal{K}}(A, B)$ in the source category $\mathcal{K}$, we have $\mathcal{G}(f) \circ \boldsymbol{\tau}_A = \boldsymbol{\tau}_B \circ \mathcal{F}(f)$ in the target category $\mathcal{L}$. Then we say that the *component* $\boldsymbol{\tau}_A : \mathcal{F}(A) \to \mathcal{G}(A)$ *is natural in* A.

If we think of a functor $\mathcal{F}$ as giving a *picture* in the target category $\mathcal{L}$ of (all the objects and morphisms of) the source category $\mathcal{K}$, then a natural transformation $\boldsymbol{\tau}$ represents a set of morphisms mapping the picture $\mathcal{F}$ to another picture $\mathcal{G}$, preserving the commutativity of all diagrams.

An invertible natural transformation, such that all components $\boldsymbol{\tau}_A$ are isomorphisms) is called a *natural equivalence* (or, *natural isomorphism*). In this case, the inverses $(\boldsymbol{\tau}_A)^{-1}$ in $\mathcal{L}$ are the components of a natural isomorphism $(\boldsymbol{\tau})^{-1} : \mathcal{G} \xrightarrow{*} \mathcal{F}$. Natural equivalences are among the most important *metamathematical constructions* in algebraic topology (see [Swi75]).

For example, let $\mathcal{B}$ be the category of Banach spaces over $\mathbb{R}$ and bounded linear maps. Define $D : \mathcal{B} \to \mathcal{B}$ by taking $D(X) = X^* =$ Banach space of bounded linear functionals on a space X and $D(f) = f^*$ for $f : X \to Y$ a bounded linear map. Then D is a cofunctor. $D^2 = D \circ D$ is also a functor. We also have the identity functor $1 : \mathcal{B} \to \mathcal{B}$. Define $T : 1 \to D \circ D$ as follows: for every $X \in \mathcal{B}$ let $T(X) : X \to D^2X = X^{**}$ be the *natural inclusion* – that is, for $x \in X$ we have $[T(X)(x)](f) = f(x)$ for every $f \in X^*$. T is a natural transformation. On the subcategory of nD Banach spaces T is even a natural equivalence. The largest subcategory of $\mathcal{B}$ on which T is a natural equivalence is called the category of reflexive Banach spaces [Swi75].

As S. Eilenberg and S. MacLane first observed, 'category' has been defined in order to define 'functor' and 'functor' has been defined in order to define 'natural transformation' [MacL71]).

Compositions of Natural Transformations

Natural transformations can be *composed* in two different ways. First, we have an 'ordinary' composition: if $\mathcal{F}, \mathcal{G}$ and $\mathcal{H}$ are three functors from the source category $\mathcal{A}$ to the target category $\mathcal{B}$, and then $\boldsymbol{\alpha} : \mathcal{F} \stackrel{\cdot}{\to} \mathcal{G}$, $\boldsymbol{\beta} : \mathcal{G} \stackrel{\cdot}{\to} \mathcal{H}$ are two natural transformations, then the formula

$$(\beta \circ \alpha)_A = \beta_A \circ \alpha_A, \qquad \text{for all} \quad A \in \mathcal{A}, \tag{3.32}$$

defines a new natural transformation $\beta \circ \alpha : \mathcal{F} \dot{\to} \mathcal{H}$. This composition law is clearly associative and possesses a unit $1_{\mathcal{F}}$ at each functor $\mathcal{F}$, whose $\mathcal{A}$–component is $1_{\mathcal{F}\mathcal{A}}$.

Second, we have the *Godement product* of natural transformations, usually denoted by $*$. Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be three categories, $\mathcal{F}, \mathcal{G}$, $\mathcal{H}$ and $\mathcal{K}$ be four functors such that $(\mathcal{F}, \mathcal{G}) : \mathcal{A} \rightrightarrows \mathcal{B}$ and $(\mathcal{H}, \mathcal{K}) : \mathcal{B} \rightrightarrows \mathcal{C}$, and $\boldsymbol{\alpha} : \mathcal{F} \dot{\to} \mathcal{G}$, $\boldsymbol{\beta} : \mathcal{H} \dot{\to} \mathcal{K}$ be two natural transformations. Now, instead of (3.32), the Godement composition is given by

$$(\beta * \alpha)_A = \beta_{GA} \circ H(\alpha_A) = K(\alpha_A) \circ \beta_{FA}, \qquad \text{for all} \quad A \in \mathcal{A}, \tag{3.33}$$

which defines a new natural transformation $\beta * \alpha : \mathcal{H} \circ \mathcal{F} \dot{\to} \mathcal{K} \circ \mathcal{G}$.

Finally, the two compositions (3.32) and (3.32) of natural transformations can be combined as

$$(\delta * \gamma) \circ (\beta * \alpha) = (\delta \circ \beta) * (\gamma \circ \alpha),$$

where $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are three categories, $\mathcal{F}, \mathcal{G}$, $\mathcal{H}$, $\mathcal{K}$, $\mathcal{L}$, $\mathcal{M}$ are six functors, and $\boldsymbol{\alpha} : \mathcal{F} \dot{\to} \mathcal{H}$, $\boldsymbol{\beta} : \mathcal{G} \dot{\to} \mathcal{K}$, $\boldsymbol{\gamma} : \mathcal{H} \dot{\to} \mathcal{L}$, $\boldsymbol{\delta} : \mathcal{K} \dot{\to} \mathcal{M}$ are four natural transformations.

Dinatural Transformations

Double natural transformations are called *dinatural transformations*. An *end of a functor* $S : C^{op} \times C \to X$ is a universal dinatural transformation from a constant e to S. In other words, an end of S is a pair $\langle e, \omega \rangle$, where e is an object of X and $\omega : e \dddot{\to} S$ is a *wedge (dinatural) transformation* with the property that to every wedge $\beta : x \dddot{\to} S$ there is a unique arrow $h : x \to e$ of B with $\beta_c = \omega_c h$ for all $a \in C$. We call ω the *ending wedge* with *components* ω_c, while the object e itself, by abuse of language, is called the end of S and written with integral notation as $\int_c S(c, c)$; thus

$$S(c, c) \overset{\omega_c}{\to} \int_c S(c, c) = e.$$

Note that the 'variable of integration' c appears twice under the integral sign (once contravariant, once covariant) and is 'bound' by the integral sign, in that the result no longer depends on c and so is unchanged if 'c' is replaced by any other letter standing for an object of the category C. These properties are like those of the letter x under the usual integral symbol $\int f(x)\, dx$ of calculus.

Every end is manifestly a limit (see below) – specifically, a limit of a suitable diagram in X made up of pieces like $S(b, b) \to S(b, c) \to S(c, c)$.

For each functor $T : C \to X$ there is an isomorphism

$$\int_c S(c, c) = \int_c Tc \cong \operatorname{Lim} T,$$

valid when either the end of the limit exists, carrying the ending wedge to the limiting cone; the indicated notation thus allows us to write any limit as an integral (an end) without explicitly mentioning the dummy variable (the first variable c of S).

A functor $H : X \to Y$ is said to *preserve the end* of a functor $S : C^{op} \times C \to X$ when $\omega : e \dddot{\to} S$ an end of S in X implies that $H\omega : He \dddot{\to} HS$ is an and for HS; in symbols

$$H \int_c S(c,c) = \int_c HS(c,c).$$

Similarly, H *creates* the end of S when to each end $v : y \dddot{\to} HS$ in Y there is a unique wedge $\omega : e \dddot{\to} S$ with $H\omega = v$, and this wedge ω is an end of S.

The definition of the coend of a functor $S : C^{op} \times C \to X$ is dual to that of an end. A *coend* of S is a pair $\langle d, \zeta \rangle$, consisting of an object $d \in X$ and a wedge $\zeta : S \dddot{\to} d$. The object d (when it exists, unique up to isomorphism) will usually be written with an integral sign and with the bound variable c as superscript; thus

$$S(c,c) \overset{\zeta_c}{\to} \int^c S(c,c) = d.$$

The formal properties of coends are dual to those of ends. Both are much like those for integrals in calculus (see [MacL71], for technical details).

3.2.5 Limits and Colimits

In abstract algebra constructions are often defined by an abstract property which requires the existence of unique morphisms under certain conditions. These properties are called *universal properties*. The *limit* of a functor generalizes the notions of inverse limit and product used in various parts of mathematics. The dual notion, *colimit*, generalizes direct limits and direct sums. Limits and colimits are defined via universal properties and provide many examples of *adjoint functors*.

A *limit* of a covariant functor $\mathcal{F} : \mathcal{J} \to \mathcal{C}$ is an object L of $\mathcal{C}$, together with morphisms $\phi_X : L \to \mathcal{F}(X)$ for every object X of $\mathcal{J}$, such that for every morphism $f : X \to Y$ in $\mathcal{J}$, we have $\mathcal{F}(f)\phi_X = \phi_Y$, and such that the following *universal property* is satisfied: for any object N of $\mathcal{C}$ and any set of morphisms $\psi_X : N \to \mathcal{F}(X)$ such that for every morphism $f : X \to Y$ in $\mathcal{J}$, we have $\mathcal{F}(f)\psi_X = \psi_Y$, there exists precisely one morphism $u : N \to L$ such that $\phi_X u = \psi_X$ for all X. If $\mathcal{F}$ has a limit (which it need not), then the limit is defined up to a unique isomorphism, and is denoted by $\lim \mathcal{F}$.

Analogously, a *colimit* of the functor $\mathcal{F} : \mathcal{J} \to \mathcal{C}$ is an object L of $\mathcal{C}$, together with morphisms $\phi_X : \mathcal{F}(X) \to L$ for every object X of $\mathcal{J}$, such that for every morphism $f : X \to Y$ in $\mathcal{J}$, we have $\phi_Y \mathcal{F}(X) = \phi_X$, and such that the following universal property is satisfied: for any object N of $\mathcal{C}$ and any set of morphisms $\psi_X : \mathcal{F}(X) \to N$ such that for every morphism $f : X \to Y$ in $\mathcal{J}$,

we have $\psi_Y \mathcal{F}(X) = \psi_X$, there exists precisely one morphism $u : L \to N$ such that $u\phi_X = \psi_X$ for all X. The colimit of $\mathcal{F}$, unique up to unique isomorphism if it exists, is denoted by $\operatorname{colim} \mathcal{F}$.

Limits and colimits are related as follows: A functor $\mathcal{F} : \mathcal{J} \to \mathcal{C}$ has a colimit iff for every object N of $\mathcal{C}$, the functor $X \longmapsto Mor_{\mathcal{C}}(\mathcal{F}(X), N)$ (which is a covariant functor on the dual category $\mathcal{J}^{op}$) has a limit. If that is the case, then $Mor_{\mathcal{C}}(\operatorname{colim} \mathcal{F}, N) = \lim Mor_{\mathcal{C}}(\mathcal{F}(-), N)$ for every object N of $\mathcal{C}$.

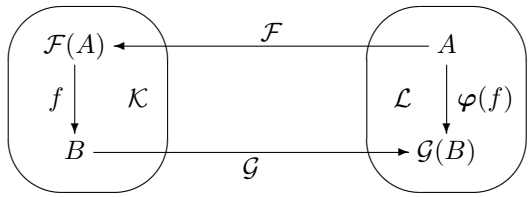
3.2.6 The Adjunction

The most important functorial operation is adjunction; as S. MacLane once said, "Adjoint functors arise everywhere" [MacL71].

The *adjunction* $\boldsymbol{\varphi} : \mathcal{F} \dashv \mathcal{G}$ between two functors $(\mathcal{F}, \mathcal{G}) : \mathcal{K} \leftrightarrows \mathcal{L}$ of *opposite variance* [Kan58], represents a *weak functorial inverse*

$$\frac{f : \mathcal{F}(A) \to B}{\boldsymbol{\varphi}(f) : A \to \mathcal{G}(B)}$$

forming a *natural equivalence* $\boldsymbol{\varphi} : \mathtt{Mor}_{\mathcal{K}}(\mathcal{F}(A), B) \xrightarrow{\boldsymbol{\varphi}} \mathtt{Mor}_{\mathcal{L}}(A, \mathcal{G}(B))$. The adjunction isomorphism is given by a *bijective correspondence* (a 1–1 and onto map on objects) $\boldsymbol{\varphi} : \mathtt{Mor}(\mathcal{K}) \ni f \to \boldsymbol{\varphi}(f) \in \mathtt{Mor}(\mathcal{L})$ of isomorphisms in the two categories, $\mathcal{K}$ (with a representative object A), and $\mathcal{L}$ (with a representative object B). It can be depicted as a (non–commutative) diagram



In this case $\mathcal{F}$ is called *left adjoint*, while $\mathcal{G}$ is called *right adjoint*.

In other words, an *adjunction* $F \dashv G$ *between two functors* $(\mathcal{F}, \mathcal{G})$ of opposite variance, from a source category $\mathcal{K}$ to a target category $\mathcal{L}$, is denoted by $(\mathcal{F}, \mathcal{G}, \boldsymbol{\eta}, \boldsymbol{\varepsilon}) : \mathcal{K} \leftrightarrows \mathcal{L}$. Here, $\mathcal{F} : \mathcal{L} \to \mathcal{K}$ is the *left (upper) adjoint functor*, $\mathcal{G} : \mathcal{L} \leftarrow \mathcal{K}$ is the *right (lower) adjoint functor*, $\boldsymbol{\eta} : 1_{\mathcal{L}} \to \mathcal{G} \circ \mathcal{F}$ is the *unit natural transformation* (or, *front adjunction*), and $\boldsymbol{\varepsilon} : \mathcal{F} \circ \mathcal{G} \to 1_{\mathcal{K}}$ is the *counit natural transformation* (or, *back adjunction*).

For example, $\mathcal{K} = \mathcal{S}$ is the category of sets and $\mathcal{L} = \mathcal{G}$ is the category of groups. Then $\mathcal{F}$ turns any set into the *free group* on that set, while the 'forgetful' functor $\mathcal{F}^*$ turns any group into the *underlying set* of that group. Similarly, all sorts of other 'free' and 'underlying' constructions are also left and right adjoints, respectively.

Right adjoints preserve *limits*, and left adjoints preserve *colimits*.

The category $\mathcal{C}$ is called a *cocomplete category* if every functor $\mathcal{F} : \mathcal{J} \to \mathcal{C}$ has a colimit. The following categories are cocomplete: $\mathcal{S}, \mathcal{G}, \mathcal{A}, \mathcal{T}$, and $\mathcal{PT}$.

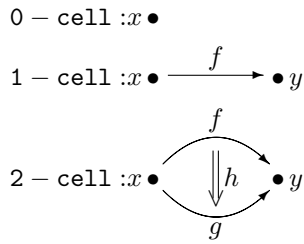
The importance of adjoint functors lies in the fact that every functor which has a left adjoint (and therefore is a right adjoint) is continuous. In the category $\mathcal{A}$ of Abelian groups, this e.g., shows that the kernel of a product of homomorphisms is naturally identified with the product of the kernels. Also, limit functors themselves are continuous. A covariant functor $\mathcal{F} : \mathcal{J} \to \mathcal{C}$ is *cocontinuous* if it transforms colimits into colimits. Every functor which has a right adjoint (and is a left adjoint) is cocontinuous.

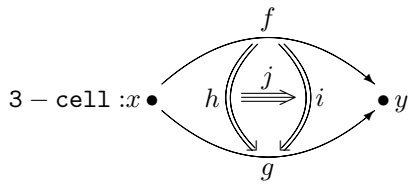
The *analogy* between *adjoint functors* and *adjoint linear operators* relies upon a deeper analogy: just as in quantum theory the inner product $\langle \phi, \psi \rangle$ represents the *amplitude* to pass from ϕ to ψ, in category theory $Mor(A, B)$ represents the *set of ways* to go from A to B. These are to Hilbert spaces as categories are to sets. The analogues of adjoint linear operators between Hilbert spaces are certain adjoint functors between 2−Hilbert spaces [Bae97, BD98]. Similarly, the *adjoint representation* of a Lie group G is the linearized version of the action of G on itself by conjugation, i.e., for each $g \in G$, the inner automorphism $x \mapsto gxg^{-1}$ gives a linear transformation $Ad(g) : \mathfrak{g} \to \mathfrak{g}$, from the Lie algebra $\mathfrak{g}$ of G to itself.

3.2.7 n−Categories

Generalization from 'Small' Categories to 'Big' n−Categories

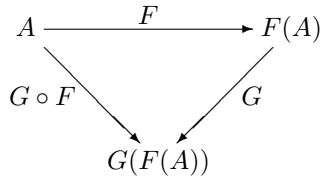
If we think of a point in geometrical space (either natural, or abstract) as an *object* (or, a 0−cell), and a path between two points as an *arrow* (or, a 1−*morphism*, or a 1−cell), we could think of a 'path of paths' as a 2−arrow (or, a 2−morphism, or a 2−cell), and a 'path of paths of paths' (or, a 3−morphism, or a 3−cell), etc. Here a 'path of paths' is just a continuous 1–parameter family of paths from between source and target points, which we can think of as tracing out a 2D surface, etc. In this way we get a 'skeleton' of an n−category, where a 1−category operates with 0−cells (objects) and 1−cells (arrows, causally connecting *source* objects with *target* ones), a 2−category operates with all the cells up to 2−cells [Ben67], a 3−category operates with all the cells up to 3−cells, etc. This skeleton clearly demonstrates the *hierarchical self–similarity* of *n–categories*:



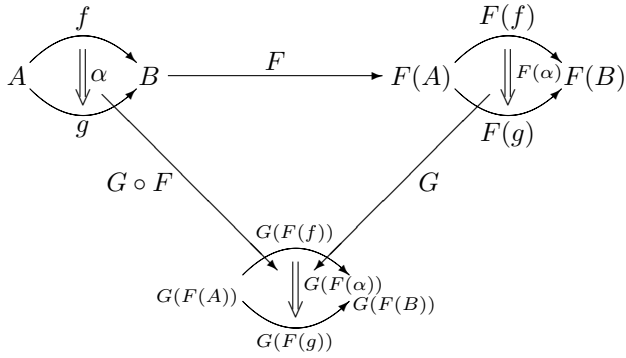


where triple arrow goes in the third direction, perpendicular to both single and double arrows. Categorical composition is defined by pasting arrows.

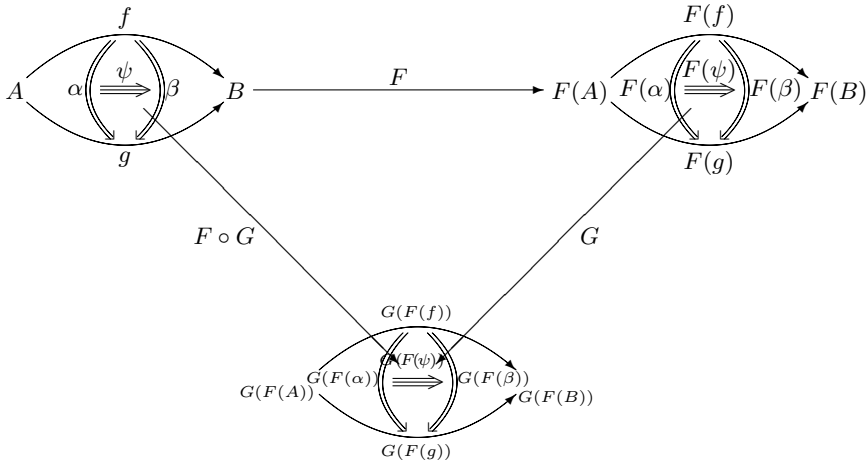
Thus, a 1−category can be depicted as a commutative triangle:



a 2−category is a commutative triangle:



a 3−category is a commutative triangle:



etc., up to $n-$categories.

Many deep–sounding results in mathematical sciences are get by the process of *categorification*[7] of the high school mathematics [CF94, BD98].

An $n-$category is a generic mathematical structure consisting of a collection of objects, a collection of arrows between objects, a collection of $2-$arrows between arrows [Ben67], a collection of $3-$arrows between $2-$arrows, and so on up to n [Bae97, BD98, Lei02, Lei03, Lei04].

More precisely, an $n-$category (for $n \geq 0$) consists of:

- $0-$cells, or objects, $A, B, \ldots$
- $1-$cells, or arrows, $A \xrightarrow{f} B$, with a composition

$$A \xrightarrow{f} B \xrightarrow{g} C = A \xrightarrow{g \circ f} C$$

- $2-$cells, 'arrows between arrows', $A \overset{f}{\underset{g}{\Downarrow \alpha}} B$, with vertical compositions (denoted by $\circ$) and horizontal compositions (denoted by $*$), respectively given by

$$A \overset{f}{\underset{g}{\Downarrow \alpha}} B \overset{h}{\underset{}{\Downarrow \beta}} = A \overset{f}{\underset{h}{\Downarrow \beta \circ \alpha}} B \qquad \text{and}$$

$$A \overset{f}{\underset{g}{\Downarrow \alpha}} A' \overset{f'}{\underset{g'}{\Downarrow \alpha'}} A'' = A \overset{f' \circ f}{\underset{g' \circ g}{\alpha' \Downarrow * \alpha}} A''$$

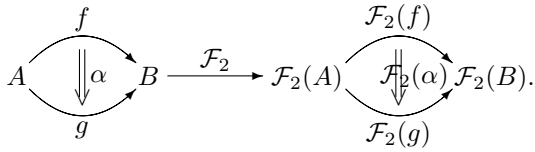
- $3-$cells, 'arrows between arrows between arrows', $A \overset{f}{\underset{g}{\alpha \Rrightarrow \Gamma \Rrightarrow \beta}} B$ (where the $\Gamma-$arrow goes in a direction perpendicular to f and α), with various kinds of vertical, horizontal and mixed compositions,

[7] Categorification means replacing sets with categories, functions with functors, and equations between functions by natural equivalences between functors. Iterating this process requires a theory of $n-$categories.

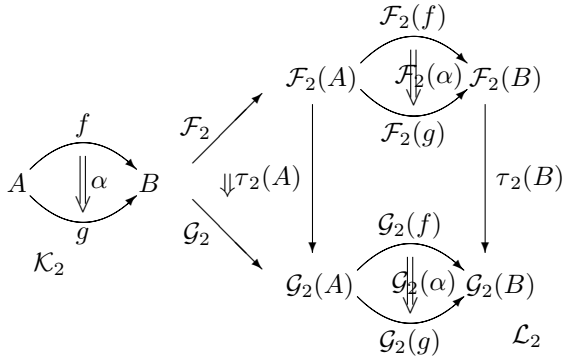
- etc., up to $n-$cells.

Calculus of $n-$categories has been developed as follows. First, there is $\mathcal{K}_2$, the 2–category of all ordinary (or small) categories. $\mathcal{K}_2$ has categories $\mathcal{K}, \mathcal{L}, ...$ as objects, functors $\mathcal{F}, \mathcal{G} : \mathcal{K} \rightrightarrows \mathcal{L}$ as arrows, and natural transformations, like $\tau : \mathcal{F} \stackrel{\cdot}{\longrightarrow} \mathcal{G}$ as 2–arrows.

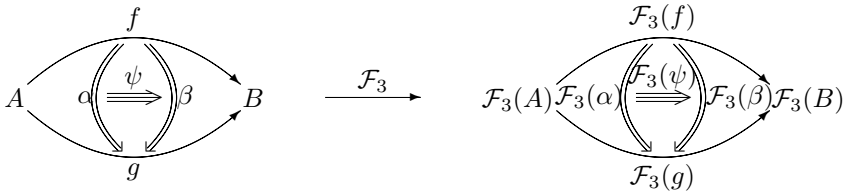
In a similar way, the arrows in a 3–category $\mathcal{K}_3$ are 2–functors $\mathcal{F}_2, \mathcal{G}_2, ...$ sending objects in $\mathcal{K}_2$ to objects in $\mathcal{L}_2$, arrows to arrows, and 2–arrows to 2–arrows, strictly preserving all the structure of $\mathcal{K}_2$



The 2–arrows in $\mathcal{K}_3$ are 2–natural transformations, like $\tau_2 : \mathcal{F}_2 \stackrel{2\cdot}{\Rightarrow} \mathcal{G}_2$ between 2–functors $\mathcal{F}_2, \mathcal{G}_2 : \mathcal{K}_2 \longrightarrow \mathcal{L}_2$ that sends each object in $\mathcal{K}_2$ to an arrow in $\mathcal{L}_2$ and each arrow in $\mathcal{K}_2$ to a 2–arrow in $\mathcal{L}_2$, and satisfies natural transformation–like conditions. We can visualize τ_2 as a prism going from one functorial picture of $\mathcal{K}_2$ in $\mathcal{L}_2$ to another, built using commutative squares:

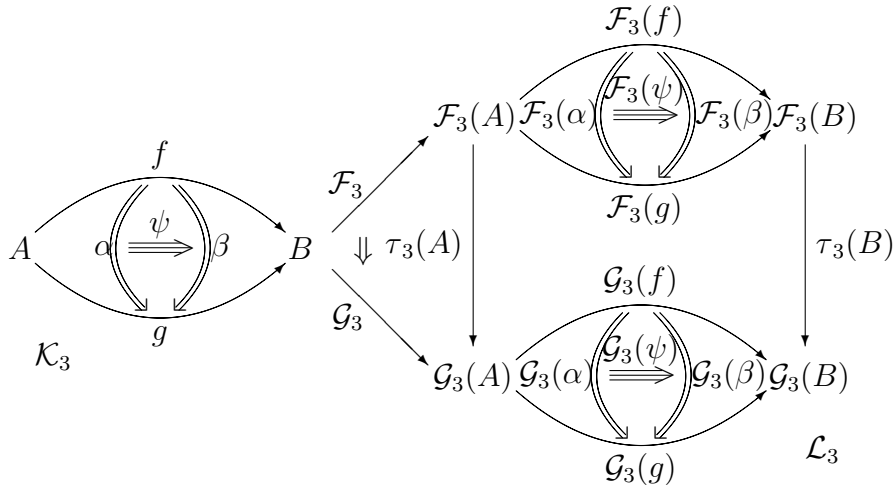


Similarly, the arrows in a 4–category $\mathcal{K}_4$ are 3–functors $\mathcal{F}_3, \mathcal{G}_3, ...$ sending objects in $\mathcal{K}_3$ to objects in $\mathcal{L}_3$, arrows to arrows, and 2–arrows to 2–arrows, strictly preserving all the structure of $\mathcal{K}_3$



The 2–arrows in $\mathcal{K}_4$ are 3–natural transformations, like $\tau_3 : \mathcal{F} \stackrel{3\cdot}{\Rightarrow} \mathcal{G}$ between 3–functors $\mathcal{F}_3, \mathcal{G}_3 : \mathcal{K}_3 \rightarrow \mathcal{L}_3$ that sends each object in $\mathcal{K}_3$ to a arrow in $\mathcal{L}_3$ and

each arrow in $\mathcal{K}_3$ to a 2–arrow in $\mathcal{L}_3$, and satisfies natural transformation–like conditions. We can visualize τ_3 as a prism going from one picture of $\mathcal{K}_3$ in $\mathcal{L}_3$ to another, built using commutative squares:



Topological Structure of $n-$Categories

We already emphasized the topological nature of ordinary category theory. This fact is even more obvious in the general case of $n-$categories (see [Lei02, Lei03, Lei04]).

Homotopy Theory

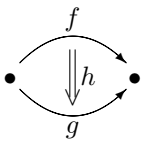
Any topological manifold M induces an $n-$*category* $\Pi_n(M)$ (its *fundamental* $n-$*groupoid*), in which 0–cells are *points* in M; 1–cells are *paths* in M (i.e., parameterized continuous maps $f : [0,1] \to M$); 2–cells are *homotopies* (denoted by $\simeq$) *of paths* relative to endpoints (i.e., parameterized continuous maps $h : [0,1] \times [0,1] \to M$); 3–cells are *homotopies of homotopies* of paths in M (i.e., parameterized continuous maps $j : [0,1] \times [0,1] \times [0,1] \to M$); categorical *composition* is defined by *pasting* paths and homotopies. In this way the following 'homotopy skeleton' emerges:

$$\texttt{0}-\texttt{cell} : x \bullet \qquad x \in M;$$

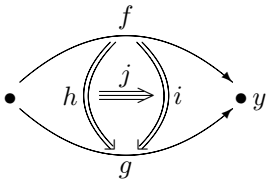
$$\texttt{1}-\texttt{cell} : x \bullet \xrightarrow{\ f\ } \bullet\, y \qquad f : x \simeq y \in M,$$

$f : [0,1] \to M$, $f : x \mapsto y$, $y = f(x)$, $f(0) = x$, $f(1) = y$;

e.g., linear path: $f(t) = (1-t)x + ty$;

$2-\texttt{cell}: x$  $y \qquad h : f \simeq g \in M,$

$h : [0,1] \times [0,1] \to M$, $h : f \mapsto g$, $g = h(f(x))$,
$h(x,0) = f(x)$, $h(x,1) = g(x)$, $h(0,t) = x$, $h(1,t) = y$
e.g., linear homotopy: $h(x,t) = (1-t)f(x) + tg(x)$;

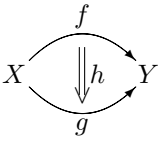
$3-\texttt{cell}: x$  $y \qquad j : h \simeq i \in M,$

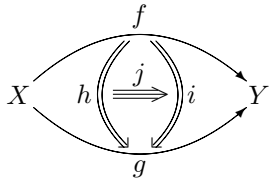
$j : [0,1] \times [0,1] \times [0,1] \to M$, $j : h \mapsto i$, $i = j(h(f(x)))$
$j(x,t,0) = h(f(x))$, $j(x,t,1) = i(f(x))$,
$j(x,0,s) = f(x)$, $j(x,1,s) = g(x)$,
$j(0,t,s) = x$, $j(1,t,s) = y$
e.g., linear composite homotopy: $j(x,t,s) = (1-t)\, h(f(x)) + t\, i(f(x))$.

If M is a *smooth* manifold, then all included paths and homotopies need to be *smooth*. Recall that a *groupoid* is a category in which every morphism is invertible; its special case with only one object is a *group*.

Category $\mathcal{TT}$

Topological $n-$category $\mathcal{TT}$ has:

- 0–cells: topological spaces X
- 1–cells: continuous maps $X \xrightarrow{f} Y$
- 2–cells: homotopies h between f and $g : X$  Y

  i.e., continuous maps $h : X \times [0,1] \to Y$, such that $\forall x \in X$, $h(x,0) = f(x)$ and $h(x,1) = g(x)$
- 3–cells: homotopies between homotopies : X  Y

  i.e., continuous maps $j : X \times [0,1] \times [0,1] \to Y$.

Category $\mathcal{CK}$

Consider an $n-$category $\mathcal{CK}$, which has:

- 0–cells: chain complexes A (of Abelian groups, say)
- 1–cells: chain maps $A \xrightarrow{f} B$
- 2–cells: chain homotopies $A \overset{f}{\underset{g}{\Downarrow \alpha}} B$,

  i.e., maps $\alpha : A \to B$ of degree 1
- 3–cells $A \overset{f}{\underset{g}{\alpha \Rightarrow \Gamma \Rightarrow \beta}} B$: homotopies between homotopies,

  i.e., maps $\Gamma : A \to B$ of degree 2 such that $d\Gamma - \Gamma d = \beta - \alpha$.

There ought to be some kind of map $\mathcal{CC} : \mathcal{TT} \Rightarrow \mathcal{CK}$ (see [Lei02, Lei03, Lei04]).

Categorification

Categorification is the process of finding category–theoretic analogs of set–theoretic concepts by replacing sets with categories, functions with functors, and equations between functions by natural isomorphisms between functors, which in turn should satisfy certain equations of their own, called 'coherence laws'. Iterating this process requires a theory of $n-$categories.

Categorification uses the following analogy between set theory and category theory [CF94, BD98]:

Set Theory	**Category Theory**
elements	objects
equations between elements	isomorphisms between objects
sets	categories
functions	functors
equations between functions	natural isomorphisms between functors

Just as sets have elements, categories have objects. Just as there are functions between sets, there are functors between categories. Now, the proper analog of an equation between elements is not an equation between objects, but an isomorphism. Similarly, the analog of an equation between functions is a natural isomorphism between functors.

3.2.8 Abelian Functorial Algebra

An *Abelian category* is a certain kind of category in which morphisms and objects can be added and in which kernels and cokernels exist and have the usual properties. The motivating prototype example of an Abelian category is the category of Abelian groups $\mathcal{A}$. Abelian categories are the framework for homological algebra (see [Die88]).

Given a homomorphism $f : A \to B$ between two objects $A \equiv \mathrm{Dom}\, f$ and $B \equiv \mathrm{Cod}\, f$ in an Abelian category $\mathcal{A}$, then its *kernel*, *image*, *cokernel* and *coimage* in $\mathcal{A}$ are defined respectively as:

$$\begin{aligned} \mathrm{Ker}\, f &= f^{-1}(e_B), & \mathrm{Coker}\, f &= \mathrm{Cod}\, f / \mathrm{Im}\, f, \\ \mathrm{Im}\, f &= f(A), & \mathrm{Coim}\, f &= \mathrm{Dom}\, f / \mathrm{Ker}\, f, \end{aligned}$$

where e_B is a unit of B [DP97].

In an Abelian category $\mathcal{A}$ a *composable* pair of arrows,

$$\bullet \xrightarrow{\ f\ } B \xrightarrow{\ g\ } \bullet$$

is *exact* at B iff $\mathrm{Im}\, f \equiv \mathrm{Ker}\, g$ (equivalence as subobjects of B) – or, equivalently, if $\mathrm{Coker}\, f \equiv \mathrm{Coim}\, g$ [MacL71].

For each arrow f in an Abelian category $\mathcal{A}$ the *triangular identities* read

$$\mathrm{Ker}(\mathrm{Coker}(\mathrm{Ker}\, f)) = \mathrm{Ker}\, f, \qquad \mathrm{Coker}(\mathrm{Ker}(\mathrm{Coker}\, f)) = \mathrm{Coker}\, f.$$

The diagram (with 0 the null object)

$$0 \longrightarrow A \xrightarrow{\ f\ } B \xrightarrow{\ g\ } C \longrightarrow 0 \tag{3.34}$$

is a *short exact sequence* when it is exact at A, at B, and at C.

Since $0 \to a$ is the zero arrow, exactness at A means just that f is *monic* (i.e., 1–1, or injective map); dually, exactness at C means that g is *epic* (i.e., onto, or surjective map). Therefore, (3.34) is equivalent to

$$f = \mathrm{Ker}\, g, \qquad g = \mathrm{Coker}\, f.$$

Similarly, the statement that $h = \mathrm{Coker}\, f$ becomes the statement that the sequence

$$A \xrightarrow{\ f\ } B \xrightarrow{\ g\ } C \longrightarrow 0$$

is exact at B and at C. Classically, such a sequence was called a short right exact sequence. Similarly, $k = \mathrm{Ker}\, f$ is expressed by a short left exact sequence

$$0 \longrightarrow A \xrightarrow{\ f\ } B \xrightarrow{\ g\ } C.$$

If $\mathcal{A}$ and $\mathcal{A}'$ are Abelian categories, an *additive functor* $\mathcal{F} : \mathcal{A} \to \mathcal{A}'$ is a functor from $\mathcal{A}$ to $\mathcal{A}'$ with

$$\mathcal{F}(f + f') = \mathcal{F}f + \mathcal{F}f',$$

for any parallel pair of arrows $f, f' : b \to c$ in $\mathcal{A}$. It follows that $\mathcal{F}0 = 0$.

A functor $\mathcal{F} : \mathcal{A} \to \mathcal{A}'$ between Abelian categories $\mathcal{A}$ and $\mathcal{A}'$ is, by definition, *exact* when it preserves all finite limits and all finite colimits. In particular, an exact functor preserves kernels and cokernels, which means that

$$\mathrm{Ker}(\mathcal{F}f) = \mathcal{F}(\mathrm{Ker}\, f) \qquad \text{and} \qquad \mathrm{Coker}(\mathcal{F}f) = \mathcal{F}(\mathrm{Coker}\, f);$$

then $\mathcal{F}$ also preserves images, coimages, and carries exact sequences to exact sequences. By construction of limits from products and equalizers and dual constructions, $\mathcal{F} : \mathcal{A} \to \mathcal{A}'$ is exact iff it is additive and preserves kernels and cokernels.

A functor $\mathcal{F}$ is *left exact* when it preserves all finite limits. In other words, $\mathcal{F}$ is left exact iff it is additive and $\mathrm{Ker}(\mathcal{F}f) = \mathcal{F}(\mathrm{Ker}\, f)$ for all f: the last condition is equivalent to the requirement that $\mathcal{F}$ preserves short left exact sequences.

Similarly, a functor $\mathcal{F}$ is *right exact* when it preserves all finite colimits. In other words, $\mathcal{F}$ is right exact iff it is additive and $\mathrm{Coker}(\mathcal{F}f) = \mathcal{F}(\mathrm{Coker}\, f)$ for all f: the last condition is equivalent to the requirement that $\mathcal{F}$ preserves short right exact sequences.

In an Abelian category $\mathcal{A}$, a *chain complex* is a sequence

$$...cone c_{n+1} \xrightarrow{\partial_{n+1}} c_n \xrightarrow{\partial_n} c_{n-1} \longrightarrow ...$$

of composable arrows, with $\partial_n \partial_{n+1} = 0$ for all n. The sequence need not be exact at c_n; the deviation from exactness is measured by the nth *homology object*

$$H_n c = \mathrm{Ker}(\partial_n : c_n \longrightarrow c_{n-1}) / \mathrm{Im}(\partial_{n+1} : c_{n+1} \longrightarrow c_n).$$

Similarly, a *cochain complex* in an Abelian category $\mathcal{A}$ is a sequence

$$...cone w_{n+1} \xrightarrow{d_{n+1}} w_n \xrightarrow{d_n} w_{n-1} \longrightarrow ...$$

of composable arrows, with $d_n d_{n+1} = 0$ for all n. The sequence need not be exact at w_n; the deviation from exactness is measured by the nth *cohomology object*

$$H^n w = \mathrm{Ker}(d_{n+1} : w_n \longrightarrow w_{n+1}) / \mathrm{Im}(d_n : w_{n-1} \longrightarrow w_n).$$

A *cycle* is a chain C such that $\partial C = 0$. A *boundary* is a chain C such that $C = \partial B$, for any other chain B.

A *cocycle* (a *closed form*) is a cochain ω such that $d\omega = 0$. A *coboundary* (an *exact form*) is a cochain ω such that $\omega = d\theta$, for any other cochain θ.

References

AAM76. Anderson, B.D., Arbib, M.A., Manes, E.G.: Foundations of System Theory: Finitary and Infinitary Conditions. Lecture Notes in Economics and Mathematical Systems Theory, Springer, New York, (1976)

AC91. Ablowitz, M.J., Clarkson, P.A.: Solitons, nonlinear evolution equations and inverse scattering. London Math. Soc., **149**, CUP, Cambridge, UK, (1991)

Ada78. Adams, J.F.: Infinite Loop Spaces. Princeton Univ. Press, Princeton, NJ, (1978)

ADJ97. Ambjørn, J., Durhuus, B., Jonsson, T.: Quantum geometry. Cambridge Monographs on Mathematical Physics, Cambridge Univ. Press, Cambridge, UK, (1997)

AEH05. Ahmed, E., Elgazzar, A.S., Hegazi, A.S.: An Overview of Complex Adaptive Systems. Mansoura J. Math. (to appear)

AFH86. Albeverio, S., Fenstat, J., Hoegh-Krohn, R., Lindstrom, T.: Nonstandard Methods in Stochastic Analysis and Mathematical Physics. Academic Press, New York, (1986)

AG05. Apps, R., Garwicz, M.: Anatomical and physiological foundations of cerebellar information processing. Nature Rev. Neurosci., **6**, 297–311, (2005)

AGM94. Alekseevsky, D.V., Grabowski, J., Marmo, G., Michor, P.W.: Poisson structures on the cotangent bundle of a Lie group or a principle bundle and their reductions. J. Math. Phys., **35**, 4909–4928, (1994)

AGM97. Alekseevsky, D., Grabowksi, J., Marmo, G., Michor, P.W.: Completely integrable systems: a generalization. Mod. Phys. Let. A, **12**(22), 1637–1648, (1997)

AGN94. Antoniadis, I., Gava, E., Narain, K.S., Taylor, T.R.: Topological amplitudes in string theory. Nucl. Phys. B **413**, 162, (1994)

AJL00a. Ambjørn, J., Jurkiewicz, J., Loll, R.: Lorentzian and Euclidean quantum gravity – analytical and numerical results. In M-Theory and Quantum Geometry, eds. L. Thorlacius and T. Jonsson, NATO Science Series, Kluwer, 382-449, (2000)

AJL00b. Ambjørn, J., Jurkiewicz, J., Loll, R.: A nonperturbative Lorentzian path integral for gravity. Phys. Rev. Lett. **85**, 924–927, (2000)

AJL01a. Ambjørn, J., Jurkiewicz, J., Loll, R.: Dynamically triangulating Lorentzian quantum gravity. Nucl. Phys. B **610**, 347–382, (2001)

AJL01b. Ambjørn, J., Jurkiewicz, J., Loll, R.: Nonperturbative 3d Lorentzian quantum gravity. Phys. Rev. D **64**, 044-011, (2001)

AJL01c. Ambjørn, J., Jurkiewicz, J., Loll, R.: Computer simulations of 3d Lorentzian quantum gravity. Nucl. Phys. B **94**, 689–692, (2001)

AK93. Ambjørn, J., Kristjansen, C.F.: Nonperturbative 2D quantum gravity and Hamiltonians unbounded from below. Int. J. Mod. Phys. A **8**, 1259–1282, (1993)

AL05. Achimescu, S., Lipan, O.: Signal Propagation in Nonlinear Stochastic Gene Regulatory Networks. 3rd Int. Conf. Path. Netw., Sys. Rhodes, Greece 2005.

AL91. Aidman, E.V., Leontiev, D.A.: From being motivated to motivating oneself: a Vygotskian perspective. Stud. Sov. Thought, **42**, 137–151, (1991)

AL98. Ambjørn, J., Loll, R.: Non-perturbative Lorentzian quantum gravity, causality and topology change. Nucl. Phys. B **536**, 407–434, (1998)

Alo03. Alon, U.: Biological Networks: The tinkerer as an engineer. Science **301**, 1866-1867, (2003)

AM78. Abraham, R., Marsden, J.: Foundations of Mechanics. Benjamin, Reading, (1978)

AM91. Aringazin, A., Mikhailov, A.: Matter fields in space–time with vector non–metricity. Clas. Quant. Grav. **8**, 1685, (1991)

Ama85. Amari, S.I.: Differential Geometrical Methods in Statistics. Springer, New York, (1985)

Ame93. Amemiya, Y.: On nonlinear factor analysis. Proc. Social Stat. Section. Ann. Meet. Ame. Stat. Assoc. 290–294, (1993)

AMR88. Abraham, R., Marsden, J., Ratiu, T.: Manifolds, Tensor Analysis and Applications. Springer, New York, (1988)

AN00. Amari, Nagaoka, H.: Methods of Information Geometry, Oxford Univ. Press and Amer. Math. Soc., (2000)

AN99. Aoyagi, T., Nomura, M.: Oscillator Neural Network Retrieving Sparsely Coded Phase Patterns. Phys. Rev. Lett. **83**, 1062–1065, (1999)

AR95. Antoni, M., Ruffo, S.: Clustering and relaxation in long-range Hamiltonian dynamics. Phys. Rev. E, **52**, 2361–2374, (1995)

Arb98. Arbib, M. (ed.): Handbook of Brain Theory and Neural Networks (2nd ed.). MIT Press, Cambridge, MA, (1998)

Ark01. Arkin, A.P.: Synthetic cell biology. Curr. Opin. Biotech., **12**, 638-644, (2001)

Arn78. Arnold, V.I.: Ordinary Differential Equations. MIT Press, Cambridge, MA, (1978)

Arn88. Arnold, V.I.: Geometrical Methods in the Theory of Ordinary differential equations. Springer, New York, (1988)

Arn89. Arnold, V.I.: Mathematical Methods of Classical Mechanics (2nd ed). Springer, New York, (1989)

Arn92. Arnold, V.I.: Catastrophe Theory. Springer, Berlin, (1992)

Arn93. Arnold, V.I.: Dynamical systems. Encyclopaedia of Mathematical Sciences, Springer, Berlin, (1993)

AS92. Abraham, R., Shaw, C.: Dynamics: the Geometry of Behavior. Addison–Wesley, Reading, MA, (1992)

Ash94. Ashcraft, M.H.: Human Memory and Cognition (2nd ed). HarperCollins, New York, (1994)

Ati88. Atiyah, M.F.: Topological quantum field theory. Publ. Math. IHES **68**, 175–186, (1988)

AV00. Alfinito, E., Vitiello, G.: Formation and life–time of memory domains in the dissipative quantum model of brain, Int. J. Mod. Phys. B, **14**, 853–868, (2000)

B-Y97. Bar-Yam, Y.: Dynamics of Complex Systems. Perseus Books, Reading, (1997)

Bae01. Baez, J.: The Meaning of Einstein's Equation. arXiv: gr-qc/0103044.

Bae02. Baez, J.: Categorified gauge theory. Lecture in the Joint Spring Meeting of the Pacific Northwest Geometry Seminar and Cascade Topology Seminar, (2002)

Bae97. Baez, J.: An introduction to $n-$categories. $7th$ Conference on Category Theory and Computer Science, E. Moggi and G. Rosolini (eds), Lecture Notes in Computer Science, Springer, Berlin, (1997)

Bar93. Barry Jay, C. : Matrices, Monads and the Fast Fourier Transform. Univ. Tech., Sidney, (1993)

Bau00. Baum, H.: Twistor and Killing spinors on Lorentzian manifolds and their relations to CR and Kaehler geometry. Int. Congr. Diff. Geom. in memory of Alfred Gray, Bilbao, Spain, (2000)

BBR91. Birmingham, D., Blau, M., Rakowski, M., Thompson, G.: Topological field theory. Phys. Rep. **209**, 129, (1991)

BCG91. Bryant, R., Chern, S., Gardner, R., Goldscmidt, H., Griffiths, P.: Exterior Differential Systems. Springer, Berlin, (1991)

BCO94. Bershadsky, M., Cecotti, S., Ooguri, H., Vafa, C.: Kodaira-Spencer theory of gravity and exact results for quantum string amplitudes. Commun. Math. Phys. **165**, 311, (1994)

BD02. Busemeyer, J.R., Diederich, A.: Survey of decision field theory. Math. Soc. Sci., **43**, 345–370, (2002)

BD95. Baez, J., Dolan, J.: Higher dimensional algebra and topological quantum field theory. J. Math. Phys. **36**, 6073–6105, (1995)

BD98. Baez, J., Dolan, J.: Higher–Dimensional Algebra III: $n-$categories and the Algebra of Opetopes. Adv. Math. **135**(2), 145–206, (1998)

Ben67. Bénabou, J.: Introduction to bicategories. In: Lecture Notes in Mathematics. Springer, New York, (1967)

Ber35. Bernstein, N.A.: Investigations in Biodynamics of Locomotion (in Russian). WIEM, Moscow, (1935)

Ber47. Bernstein, N.A.: On the structure of motion (in Russian). Medgiz, Moscow, (1947)

BF71. Bransford, J.D., Franks, J.J.: The Abstraction of Linguistic Ideas. Cogn. Psych., **2**, 331–350, (1971)

BFR04. Bagnoli, F., Franci, F., Rechtman, R.: Chaos in a simple cellular automata model of a uniform society. In Lec. Not. Comp. Sci., Vol. 3305, 513–522, Springer, London, (2004)

BGG03. Bryant, R., Griffiths, P., Grossman, D.: Exterior Differential Systems and Euler–Lagrange partial differential equations. Univ. Chicago Press, Chicago, (2003)

BGG89. Batlle, C., Gomis, J., Gràcia, X., Pons, J.M.: Noether's theorem and gauge transformations: application to the bosonic string and CP_2^{n-1} model. J. Math. Phys. **30**, 1345, (1989)

BH96. Banaszuk, A., Hauser, J.: Approximate feedback linearization: a homotopy operator approach. SIAM J. Cont. & Optim., **34**(5), 1533–1554, (1996)

BK64. Bellman, R., Kalaba, R.: Selected papers on mathematical trends in control theory. Dover, New York, (1964)

BK92. Berleant, D., Kuipers, B.: Qualitative–Numeric Simulation with Q3, in Recent Advances in Qualitative Physics. eds. Boi Faltings and Peter Struss, MIT Press, Cambridge, (1992)

BL92. Blackmore, D.L., Leu, M.C.: Analysis of swept volumes via Lie group and differential equations. Int. J. Rob. Res., **11**(6), 516–537, (1992)

Bla83. Blattner, R.: Nonlinear Partial Differential Operators and Quantization Procedure, in Proceedings, Clausthall 1981, Springer-Verlag, New York, 209-241, (1983)

BM00. Choquet-Bruhat, Y., DeWitt-Morete, C.: Analysis, Manifolds and Physics, Part II: 92 Applications (rev. ed). North-Holland, Amsterdam, (2000)

BM82. Choquet-Bruhat, Y., DeWitt-Morete, C.: Analysis, Manifolds and Physics (2nd ed). North-Holland, Amsterdam, (1982)

BMP97. Breckenridge, J.C., Myers, R.C., Peet, A.W., Vafa, C.: D–branes and spinning black holes. Phys. Lett. B **391**, 93, (1997)

BMT04. Basu, S., Mehreja, R., Thiberge, S., Chen, M.T., Weiss, R.: Spatiotemporal control of gene expression with pulse generating networks. Proc Natl. Acad. Sci. USA, **101**, 6355-6360, (2004)

BO95. Basar, T., Olsder, G.J.: Dynamic Noncooperative Game Theory (2nd ed.), Academic Press, New York, (1995)

Boc04. Boccara, N.: Modeling complex systems. Springer, Berlin, (2004)

Bon95. Bontempi, G.: Modelling with uncertainty in continuous dynamical systems: the probability and possibility approach. IRIDIA – ULB Technical Report, 95–16, (1995)

Boo86. Boothby, W.M.: An Introduction to Differentiable Manifolds and Riemannian Geometry, Academic Press, New York, (1986)

BP02. Benvenuto, N., Piazza, F.: On the complex backpropagation algorithm. IEEE Trans. Sig. Proc., **40**(4), 967–969, (1992)

BPS75. Belavin, A.A., Polyakov, A.M., Swartz, A.S., Tyupkin, Yu.S.: SU(2) instantpons discovered. Phys. Lett. B **59**, 85, (1975)

BPS98. Blackmore, D.L., Prykarpatsky, Y.A., Samulyak, R.V.: The Integrability of Lie-invariant Geometric Objects Generated by Ideals in the Grassmann Algebra. J. Nonlin. Math. Phys., **5**(1), 54–67, (1998)

Bre04. Brent, R.: A partnership between biology and engineering. Nature, **22**, 1211-1214, (2004)

Bro58. Broadbent, D.E.: Perception and communications. Pergamon Press, London, (1958)

BRT99. Bennati, E., Rosa-Clot, M., Taddei, S.: A Path Integral Approach to Derivative Security Pricing: I. Formalism and Analytical Results, Int. Journ. Theor. Appl. Finance **2**, 381, (1999)

BS00. Becskei, A., Serrano, L.: Engineering stability in gene networks by autoregulation. Nature, **405**, 590-593, (2000)

BS04. Bashkirov, D., Sardanashvily, G.: Covariant Hamiltonian Field Theory. Path Integral Quantization. arXiv:hep-th/0402057.

BS04. Bashkirov, D., Sardanashvily, G.: Covariant Hamiltonian Field Theory. Path Integral Quantization. Int. J. Theor. Phys. **43**, 1317-1333, (2004)

BS73. Black, F., Scholes, M.: The Pricing of Options and Corporate Liabilities. J. Polit. Econ. **81**, 637–659, (1973)

BS85. Bamber, D., van Santen, J.P.H.: How many parameters can a model have and still be testable? J. Math. Psych., **29**, 443–473, (1985)

BSM97. Breitenbach, G., Schiller, S., Mlynek, J.: Measurement of the quantum states of squeezed light. Nature, **387**, 471–475 (1997)

BT03. Bauchau, O.A., Trainelli, L.: The Vectorial Parameterization of Rotation. Non. Dyn., **31**(1), 71–92, (2003)

BT93. Busemeyer, J.R., Townsend, J.T.: Decision field theory: A dynamic–cognitive approach to decision making in an uncertain environment. Psych. Rev., **100**, 432–459, (1993)

Cal57. Calabi, E.: On Kähler manifolds with vanishing canonical class. In Algebraic geometry and topology: a symposium in honor of S. Lefschetz, eds. R.H. Fox, D.C. Spencer and A.W. Tucker, Princeton Univ. Press, (1957)

Can86. Canarutto, D.: Bundle splittings, connections and locally principle fibred manifolds. Bull. U.M.I. Algebra e Geometria, Seria VI **V-D**, 18, (1986)

CC99. Cao, H.D., Chow, B.: Recent Developments on the Ricci Flow. Bull. Amer. Math. Soc. **36**, 59–74, (1999)

CCC97. Caiani, L., Casetti, L., Clementi, C., Pettini, M.: Geometry of dynamics, Lyapunov exponents and phase transitions. Phys. Rev. Lett. **79**, 4361, (1997)

CCI91. Cariñena, J., Crampin, M., Ibort, L.: On the multisymplectic formalism for first order field theories. Diff. Geom. Appl. **1**, 345, (1991)

CCI91. Cariñena, J., Crampin, M., Ibort, L.: On the multisymplectic formalism. for first order field theories. *Diff. Geom. Appl.* **1**, 345, (1991)

CCP96. Christiansen, F., Cvitanovic, P., Putkaradze, V.: Hopf's last hope: spatio-temporal chaos in terms of unstable recurrent patterns. Nonlinearity, **10**, 1, (1997)

CCP99. Casetti, L., Cohen, E.G.D., Pettini, M.: Origin of the Phase Transition in a Mean–Field Model. Phys. Rev. Lett., **82**, 4160, (1999)

CD98. Chen, G., Dong, X.: From Chaos to Order. Methodologies, Perspectives and Application. World Scientific, Singapore, (1998)

CF93. Cariñena, J., Fernández-Núñez, J.: Geometric theory of time-dependent singular Lagrangians, Fortschr. Phys. **41**, 517, (1993)

CF94. Crane, L., Frenkel, I.: Four dimensional topological quantum field theory, Hopf categories, and the canonical bases. Jour. Math. Phys. **35**, 5136–5154, (1994)

CG83. Cohen, M.A., Grossberg, S.: Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. IEEE Trans. Syst., Man, Cybern., **13**(5), 815–826, (1983)

CGI95. Cariñena, J., Gomis, J., Ibort, L. and Román, N.: Canonical transformation theory for presymplectic systems. J. Math. Phys. **26**, 1961, (1985)

CH64. Conway, E.D., Hopf, E.: Hamilton's Theory and Generalized Solutions of the Hamilton–Jacobi Equation. J. Math. Mech. **13**, 939–986, (1964)

Che55. Chevalley, C.: Theorie differential equations groupes de Lie. Vol. 1–3. Hermann D.C., Paris, (1955)

Che96. Chern, S.S.: Riemannian Geometry as a Special Case of Finsler Geometry, Cont. Math., 51–58, Vol. 196, Amer. Math. Soc. Providence, RI, (1996)

CHS85. Candelas, P., Horowitz, G.T., Strominger, A., Witten, E.: Vacuum configurations for superstrings, Nucl. Phys. B **258**, 46, (1985)

CL72. Craik, F., Lockhart, R.: Levels of processing: A framework for memory research. J. Verb. Learn. & Verb. Behav., **11**, 671–684, (1972)

CL84. Cheng, T.-P., Li, L.-F.: Gauge Theory of Elementary Particle Physics. Clarendon Press, Oxford, (1984)

CLM94. Chinea, D., de León, M., Marrero, J.: The constraint algorithm for time-dependent Lagrangians, J. Math. Phys. **35**, 3410, (1994)

CLM94. Chinea, D., de León, M., Marrero, J.: The constraint algorithm for time-dependent Lagrangians. J. Math. Phys., **35**(7), 3410-3447, (1994)

Col88. Coleman, S.: Aspects of Symmetry : Selected Erice Lectures. Cambridge Univ. Press, (1988)

Cox92. Cox, E.: Fuzzy Fundamentals, IEEE Spectrum, 58–61, (1992)
Cox94. Cox, E.: The Fuzzy Systems Handbook. AP Professional, (1994)
CP02. Clementi, C. Pettini, M.: A geometric interpretation of integrable motions. Celest. Mech. & Dyn. Astr., **84**, 263–281, (2002)
CPC03. Casetti, L., Pettini, M., Cohen, E.G.D.: Phase transitions and topology changes in configuration space. J. Stat. Phys., **111**, 1091, (2003)
CQ69. Collins, A.M., Quillian, M.R.: Retrieval Time From Semantic Memory, J. Verb. Learn. & Verb. Behav., **8**, 240–248, (1969)
CR89. Cariñena, J., Rañada, M.: Poisson maps and canonical transformations for time-dependent Hamiltonian systems. J. Math. Phys. **30**, 2258, (1989)
CR93. Cariñena, J., Rañada, M.: Lagrangian systems with constraints. A geometric approach to the method of Lagrange multipliers. J. Phys. A. **26**, 1335, (1993)
Cra91. Crawford, J.: Clifford algebra: Notes on the spinor metric and Lorentz, Poincaré and conformal groups. J. Math. Phys. **32**, 576, (1991)
CRV92. Celeghini, E., Rasetti, M., Vitiello, G.: Quantum Dissipation. Annals Phys., **215**, 156, (1992)
CS96. Constantine, G.M., Savits, T.H.: A multivariate Faà di Bruno formula with applications. Trans. Amer. Math. Soc. **348**(2), 503–520, (1996)
CV91. Cecotti, S., Vafa, C.: Topological – antitopological fusion, Nucl. Phys. B **367**, 359, (1991)
Cvi00. Cvitanovic, P.: Chaotic field theory: a sketch. Physica A **288**, 61–80, (2000)
CY89. Crutchfield, J.P., Young, K.: Computation at the onset of chaos. In *Complexity, Entropy and the Physics of Information*, p. 223, SFI Studies in the Sciences Complexity, Vol. VIII, W.H. Zurek (Ed.), Addison-Wesley, Reading, MA, (1989)
Das02. Dasgupta, A.: The real Wick rotations in quantum gravity. JHEP, **0207**, (2002)
Dav81. Davydov, A.S.: Biology and Quantum Mechanics, Pergamon Press, New York, (1981)
Dav89. Davies, E.B.: Heat Kernels and Spectral Theory. Cambridge Univ. Press, (1989)
Dav91. Davydov, A.S.: Solitons in Molecular Systems. (2nd ed), Kluwer, Dordrecht, Ger, (1991)
DEF99. Deligne, P., Etingof, P., Freed, D.S., Jeffrey, L.C., Kazhdan, D., Morgan, J.W., Morrison, D.R., Witten, E.: Quantum Fields and Strings: A Course for Mathematicians, Am. Math. Soc., (1999)
DG03. Dragovic, V., Gajic, B.: The Wagner Curvature Tensor in Nonholonomic Mechanics. Reg. Chaot. Dyn., **8**(1), 105–124, (2003)
DHS91. Domany, E., van Hemmen, J.L., Schulten, K. (eds.): Models of Neural Networks. Springer, Berlin, (1991)

Die69. Dieudonne, J.A.: Foundations of Modern Analysis (in four volumes). Academic Press, New York, (1969)

Die88. Dieudonne, J.A.: A History of Algebraic and Differential Topology 1900–1960. Birkháuser, Basel, (1988)

Dir82. Dirac, P.A.M.: Principles of Quantum Mechanics. (4th ed.), Oxford Univ. Press, (1982)

DL01. Dasgupta, A., Loll, R.: A proper-time cure for the conformal sickness in quantum gravity. Nucl. Phys. B **606**, 357–379, (2001)

DM03. Dorogovtsev, S.N., Mendes, J.F.F.: Evolution of Networks. Oxford Univ. Press, (2003)

Dom78. de Dominicis, C.: Dynamics as a substitute for replicas in systems with quenched random impurities. Phys. Rev. B **18**, 4913–4919, (1978)

Dow02. Dowker, F.: Topology change in quantum gravity. In Proceedings of Stephen Hawking's 60th birthday conference, Cambridge, UK, (2002)

DP80. Dubois, D., Prade, H.: Fuzzy Sets and Systems. Academic Press, New York, (1980)

DP97. Dodson, C.T.J., Parker, P.E.: A User's Guide to Algebraic Topology. Kluwer, Dordrecht, (1997)

DR94. Dittrich, W., Reuter, M.: Classical and Quantum Dynamics. Springer Verlag, Berlin, (1994)

DTP02. Dauxois, T., Theodorakopoulos, N., Peyrard, M.: Thermodynamic instabilities in one dimension: correlations, scaling and solitons. J. Stat. Phys. **107**, 869, (2002)

Dus84. Dustin, P.: Microtubules. Springer, Berlin, (1984)

DVV91. Dijkgraaf, R., Verlinde, H., Verlinde, E.: Notes On Topological String Theory And 2D Quantum Gravity, in String Theory and Quantum Gravity, Proceedings of the Trieste Spring School 1990, (eds.) M. Green et al., World Scientific, 91–156, (1991)

Eas02. Eastwood, M.: Higher symmetries of the Laplacian. arXiv:hep-th/0206233, (2002)

EB01. Endy, D., Brent, R.: Modelling cellular behaviuor. Nature **409**, 391-395, (2001)

Ecc64. Eccles, J.C.: The Physiology of Synapses. Springer, Berlin, (1964)

EG91. Eastwood, M.G., Graham, C.R.: Invariants of conformal densities. Duke Math. Jour. **63**, 633–671, (1991)

EIS67. Eccles, J.C., Ito M., Szentagothai J.: The Cerebellum as a Neuronal Machine. Springer, Berlin, (1967)

EL00. Elowitz, M.B., Leibler, S.: A synthetic oscillatory network of transcriptional regulators. Nature, **403**, 335–338, (2000)

EL78. Eells, J., Lemaire, L.: A report on harmonic maps. Bull. London Math. Soc. **10**, 1-68, (1978)

Elk99. Elkin, V.I.: Reduction of Nonlinear Control Systems. A Differential Geometric Approach, Kluwer, Dordrecht, (1999)

Elw82. Elworthy, K.D.: Stochastic Differential Equations on Manifolds. Cambridge Univ. Press, Cambridge, (1982)

EMN92. Ellis, J., Mavromatos, N., Nanopoulos, D.V.: String theory modifies quantum mechanics. CERN-TH/6595, (1992)

EMN99. Ellis, J., Mavromatos, N., Nanopoulos, D.V.: A microscopic Liouville arrow of time. Chaos, Solit. Fract., **10**(2–3), 345–363, (1999)

EMR91. Echeverría-Enríquez, A., Muñoz-Lecanda, M., Román-Roy, N.: Geometrical setting of time-dependent regular systems. Alternative models, Rev. Math. Phys. **3**, 301, (1991)

EMR95. Echeverría-Enríquez, A., Muñoz-Lecanda, M., Román-Roy, N.: Non-standard connections in classical mechanics. J. Phys. A **28**, 5553, (1995)

EMR98. Echeverría-Enríquez, A., Muñoz-Lecanda, M., Román-Roy, N.: Multivector fields and connections. Setting Lagrangian equations in field theories. J. Math. Phys. **39**, 4578, (1998)

Erm96. Ermentrout, G.B.: Type I membranes, phase resetting curves, and synchrony. Neural Computation, **8**(5), 979–1001, (1996)

FCS99. Franzosi, R., Casetti, L., Spinelli, L., Pettini, M.: Topological aspects of geometrical signatures of phase transitions. Phys. Rev. E, **60**, 5009–5012, (1999)

Fed69. Federer, H.: Geometric Measure Theory. Springer, New York, (1969)

Fer99. Ferber, J.: Multi-Agent Systems. An Introduction to Distributed Artificial Intelligence. Addison-Wesley, Reading, MA, (1999)

Fey48. Feynman, R.P.: Space–time Approach to Nonrelativistic Quantum Mechanics. Rev. Mod. Phys. **20**, 367–387, (1948)

Fey51. Feynman, R.P.: An Operator Calculus having Applications in Quantum Electrodynamics. Phys. Rev., **84**, 108–128, (1951)

Fey51. Feynman, R.P.: An Operator Calculus Having Applications in Quantum Electrodynamics. Phys. Rev. **84**, 108–128, (1951)

Fey72. Feynman, R.P.: Statistical Mechanics, A Set of Lectures. WA Benjamin, Inc., Reading, Massachusetts, (1972)

Fey98. Feynman, R.P.: Quantum Electrodynamics. Advanced Book Classics, Perseus Publishing, (1998)

FFF94. Fatibene, L., Ferraris, M., Francaviglia, M.: Nöther formalism for conserved quantities in classical gauge field theories II. J. Math. Phys., **35**, 1644, (1994)

FGR98. Fajstrup, L., Goubault, E., Raussen, M.: Detecting Deadlocks in Concurrent Systems, In D. Sangiorgi R. de Simone, editor, CONCUR '98; Concurrency Theory, number 1466 in Lecture Notes in Computer Science, 332–347. Springer, Berlin, (1998)

FH65. Feynman, R.P., Hibbs, A.R.: Quantum Mechanics and Path Integrals, McGraw-Hill, New York, (1965)

FK82. Ferraris, M., Kijowski, J.: On the equivalence of the relativistic theories of gravitation. GRG **14**, 165, (1982)

FK83. Frölich, H., Kremer, F.: Coherent Excitations in Biological Systems. Springer, New York, (1983)

Fla63. Flanders, H.: Differential Forms with Applications to the Physical Sciences. Acad. Press, (1963)

Flo87. Floer, A.: Morse theory for fixed points of symplectic diffeomorphisms. Bull. AMS. **16**, 279, (1987)

Flo88. Floer, A.: Morse theory for Lagrangian intersections. J. Diff. Geom., **28**(9), 513–517, (1988)

For90. Fordy, A.P. (ed.): Soliton Theory: A Survey of Results. MUP, Manchester, UK, (1990)

FP04. Franzosi, R., Pettini, M.: Theorem on the origin of phase transitions. Phys. Rev. Lett., **92**(6), 60601, (2004)

FPR04. Forger, M., Paufler, C., Römer, H.: Hamiltonian Multivector Fields and Poisson Forms in Multisymplectic Field Theory. arXiv:math-ph/0407057.

FPS00. Franzosi, R., Pettini, M. Spinelli, L.: Topology and phase transitions: a paradigmatic evidence. Phys. Rev. Lett. 84(13), 2774–2777, (2000)

FR04. Forger, M., Römer, H.: Currents and the energy–momentum tensor in classical field theory: a fresh look at an old problem. Ann. Phys. (N.Y.) **309**, 306–389, (2004)

Fri88. Friberg, O.: A Set of Parameters for Finite Rotations and Translations. Comp. Met. App. Mech. Eng., **66**, 163–171, (1988)

FS92. Freeman, J.A., Skapura, D.M.: Neural Networks: Algorithms, Applications, and Programming Techniques. Addison-Wesley, Reading, MA, (1992)

Gal83. Gallavotti, G.: The Elements of Mechanics. Springer-Verlag, Berlin, (1983)

Gar72. García, P.: Connections and 1-jet fibre bundles. Rend. Sem. Univ. Padova, **47**, 227, (1972)

Gar77. García, P.: Gauge algebras, curvature and symplectic structure. J. Diff. Geom., **12**, 209, (1977)

Gar85. Gardiner, C.W.: Handbook of Stochastic Methods for Physics, Chemistry and Natural Sciences, (2nd ed). Springer-Verlag, New York, (1985)

Gau02a. Gaucher, P.: Homotopy invariants of higher–dimensional categories and concurrency in computer science. Mathematical Structure in Computer Science **10**(4), 481–524, (2000)

Gau02b. Gaucher, P.: From Concurrency to Algebraic Topology. Electronic Notes in Theoretical Computer Science **39**(2), 19, (2000)

Gau02c. Gaucher, P.: About the globular homology of higher–dimensional automata. Cahiers de Topologie et Géométrie Diff

Gau03. Gaucher, P.: The branching nerve of HDA and the Kan condition. Theo. Appl. Categ. **11**(3), 75–106, (2003)

GBL03. Gardner, T.S., di Bernardo, D., Lorenz, D., Collins, J.J.: Inferring genetic networks and identifying compound mode of action via expression profiling. Science, **301**, 102-105, (2003)

GEH02. Guet, C., Elowitz, M.B., Hsing, W., Leibler, S.: Combinatorial synthesis of genetic networks. Science, **296**(5572), 1466-1470, (2002)

Geo88. Georgii, H.O.: Gibbs Measures and Phase Transitions. Walter de Gruyter, Berlin, (1988)

GG03. Gaucher, P., Goubault, E.: Topological Deformation of Higher Dimensional Automata. Homology, Homotopy and Applications, **5**(2), 39–82, (2003)

Ghe90. Ghez, C.: Introduction to motor system. In: Kandel, E.K. and Schwarz, J.H. (eds.) Principles of neural science. 2nd ed. Elsevier, Amsterdam, 429–442, (1990)

Ghe91. Ghez, C.: Muscles: Effectors of the Motor Systems. In: Principles of Neural Science. 3rd Ed. (Eds. E.R. Kandel, J.H. Schwartz, T.M. Jessell), Appleton and Lange, Elsevier, 548–563, (1991)

Gia92. Giachetta G.: Jet manifolds in nonholonomic mechanics. J. Math. Phys. **33**, 1652, (1992)

GIM04. Gotay, M.J., Isenberg, J., Marsden, J.E., Montgomery, R.: Momentum Maps and Classical Relativistic Fields II: Canonical Analysis of Field Theories. arXiv:math-ph/0411032.

GIM98. Gotay, M.J., Isenberg, J., Marsden, J.E.: Momentum Maps and the Hamiltonian Structure of Classical Relativistic Fields. arXiv:hep/9801019.

GJ94. van Groesen E., De Jager E.M.: (Editors) Mathematical Structures in Continuous Dynamical Systems. Studies in Mathematical Physics, vol. 6 North-Holland, Amsterdam, (1994)

GK92. Georgiou, G.M., Koutsougeras, C.: Complex domain backpropagation, IEEE Trans. Circ. Sys., **39**(5), 330-334, (1992)

Gla63a. Glauber, R.J.: The Quantum Theory of Optical Coherence. Phys. Rev. **130**, 2529-2539, (1963)

Gla63b. Glauber, R.J.: Coherent and Incoherent States of the Radiation Field. Phys. Rev. **131**, 2766-2788, (1963)

GM04. Grinza, P., Mossa, A.: Topological origin of the phase transition in a model of DNA denaturation. Phys. Rev. Lett. **92**(15), 158102, (2004)

GM90. Giachetta, G., Mangiarotti, L.: Gauge-invariant and covariant operators in gauge theories. Int. J. Theor. Phys., **29**, 789, (1990)

GM92. Gotay, M.J., Marsden, J.E.: Stress-energy–momentum tensors and the Belinfante-Rosenfeld formula. Contemp. Math. **132**, AMS, Providence, 367–392, (1992)

GMP05. Grunwald, P., Myung, I.J., Pitt, M.A.(eds.): Advances in Minimum Description Length: Theory and Applications. MIT Press, Cambridge, MA (2005)

GMS02a. Giachetta, G., Mangiarotti, L., Sardanashvily, G.: Covariant geometric quantization of non-relativistic Hamiltonian mechanics. J. Math. Phys. **43**, 56, (2002)

GMS02b. Giachetta, G., Mangiarotti, L., Sardanashvily, G.: Geometric quantization of mechanical systems with time-dependent parameters. J. Math. Phys., **43**, 2882, (2002)

GMS05. Giachetta, G., Mangiarotti, L., Sardanashvily, G.: Lagrangian supersymmetries depending on derivatives. Global analysis and cohomology. Commun. Math. Phys. **259**, 103–128, (2005)

GMS97. Giachetta, G., Mangiarotti, L., Sardanashvily, G.: New Lagrangian and Hamiltonian Methods in Field Theory, World Scientific, Singapore, (1997)

GMS99. Giachetta, G., Mangiarotti, L., Sardanashvily, G.: Covariant Hamilton equations for field theory. J. Phys. A **32**, 6629, (1999)

GN80. Gotay, M.J., Nester, J.M.: Generalized constraint algorithm and special presymplectic manifolds. In G. E. Kaiser, J. E. Marsden, Geometric methods in mathematical physics, Proc. NSF-CBMS Conf., Lowell/Mass. 1979, Berlin: Springer-Verlag, Lect. Notes Math. **775**, 78–80, (1980)

GNH78. Gotay, M., Nester, J., Hinds, G.: Presymplectic manifolds and the Dirac-Bergmann theory of constraints. J. Math. Phys. **19**, 2388, (1978)

Gol99. Gold, M.: A Kurt Lewin Reader, the Complete Social Scientist. Am. Psych. Assoc., Washington, (1999)

Gom94. Gómez, J.C.: Using symbolic computation for the computer aided design of nonlinear (adaptive) control systems. Tech. Rep. EE9454, Dept. Electr. and Comput. Eng., Univ. Newcastle, Callaghan, NSW, AUS, (1994)

Goo98. Goodwine, J.W.: Control of Stratified Systems with Robotic Applications. PhD thesis, California Institute of Technology, Pasadena, Cal, (1998)

Got82. Gotay, M.: On coisotropic imbeddings of presymplectic manifolds. Proc. Amer. Math. Soc. **84**, 111, (1982)

Got91a. Gotay, M.J.: A multisymplectic framework for classical field theory and the calculus of variations. I: Covariant Hamiltonian formalism. In M. Francaviglia (ed.), Mechanics, analysis and geometry: 200 years after Lagrangian. North-Holland, Amsterdam, 203–235, (1991)

Got91b. Gotay, M.J.: A multisymplectic framework for classical field theory and the calculus of variations. II: Space + time decomposition. Differ. Geom. Appl. **1**(4), 375–390, (1991)

Got91c. Gotay, M.: A multisymplectic framework for classical field theory and the calculus of variations. I. Covariant Hamiltonian formalism. In: M.Francaviglia (ed.) Mechanics, Analysis and Geometry: 200 Years after Lagrange. 203-235. North-Holland, Amsterdam, (1991)

Gou95. Goubault, E.: The Geometry of Concurrency. PhD thesis, Ecole Normale Supérieure, (1995)

GPS95. Gordon, R., Power, A.J., Street, R.: Coherence for tricategories. Memoirs Amer. Math. Soc. **117**(558), (1995)

Gre00. Greene, B.R.: The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory. Random House, Parktown, SA, (2000)

Gre96. Greene, B.R.: String Theory on Calabi-Yau Manifolds. Lectures given at the TASI-96 summer school on Strings, Fields and Duality, (1996)

Gri83a. Griffiths, P.A.: Exterior Differential Systems and the Calculus of Variations, Birkhauser, Boston, (1983)

Gri83b. Griffiths, P.A.: Infinitesimal variations of Hodge structure. III. Determinantal varieties and the infinitesimal invariant of normal functions. Compositio Math., **50**(2-3), 267–324, (1983)

Gro69. Grossberg, S.: Embedding fields: A theory of learning with physiological implications. J. Math. Psych. **6**, 209–239, (1969)

Gro82. Grossberg, S.: Studies of Mind and Brain. Dordrecht, Holland, (1982)

Gro83. Grothendieck, A.: Pursuing stacks (Unpublished manuscript, distributed from UCNW). Bangor, UK, (1983)

Gro88. Grossberg, S.: Neural Networks and Natural Intelligence. MIT Press, Cambridge, MA, (1988)

Gro99. Grossberg, S.: How does the cerebral cortex work? Learning, attention and grouping by the laminar circuits of visual cortex. Spatial Vision **12**, 163–186, (1999)

GRS96. Gilks, W.R., Richardson, S., Spiegelhalter, D.J.: Markov Chain Monte Carlo in Practice. Chapman & Hall, (1996)

Gru00. Grunwald, P.: The minimum description length principle. J. Math. Psych., **44**, 133–152, (2000)

Gru99. Grunwald, P.: Viewing all models as 'probabilistic'. Proceedings of the Twelfth Annual Conference on Computational Learning Theory (COLT' 99), Santa Cruz, CA, (1999)

GS95. Giachetta, G., Sardanashvily, G.: Stress-energy–momentum of affine–metric gravity. Generalized Komar superpotential. arXiv: gr-qc/9511008.

GS96. Giachetta, G., Sardanashvily, G.: Stress-Energy-Momentum of Affine-Metric Gravity. Class.Quant.Grav. **13** L, 67–72, (1996)

GS97. Giachetta, G., Sardanashvily, G.: Dirac Equation in Gauge and Affine-Metric Gravitation Theories. Int. J. Theor. Phys. **36**, 125–142, (1997)

GS98. Grosche, C., Steiner, F.: Handbook of Feynman path integrals. Springer Tracts in Modern Physics 145, Springer, Berlin, (1998)

GSW87. Green, M.B., Schwarz, J.H., Witten, E.: Superstring Theory, 2 Vols., Cambridge Univ. Press, Cambridge, (1987)

Gun03. Gunion, J.F.: Class Notes on Path-Integral Methods. U.C. Davis, 230B, (2003)

Gut90. Gutzwiller, M.C.: Chaos in Classical and Quantum Mechanics, Springer, New York, (1990)

Gut90. Gutzwiller, M.: Chaos in Classical and Quantum Mechanics Springer, Berlin, (1990)

Gut98. Gutkin, B.S., Ermentrout, B.: Dynamics of membrane excitability determine interspike interval variability: A link between spike generation mechanisms and cortical spike train statistics. Neural Comput., **10**(5), 1047–1065, (1998)

Hak83. Haken, H.: Synergetics: An Introduction (3rd ed). Springer, Berlin, (1983)

Hak93. Haken, H.: Advanced Synergetics: Instability Hierarchies of Self-Organizing Systems and Devices (3nd ed.). Springer, Berlin. (1993)

Ham82. Hamilton, R.S.: Three-manifolds with positive Ricci curvature, J. Differential Geom. **17**, 255–306, (1982)

Ham87. Hameroff, S.R.: Ultimate Computing: Biomolecular Consciousness and Nanotechnology. North-Holland, Amsterdam, (1987)

Hat02. Hatcher, A.: Algebraic Topology. Cambridge Univ. Press, Cambridge, (2002)

Hat77a. Hatze, H.: A myocybernetic control model of skeletal muscle. Biol. Cyber. **25**, 103–119, (1977)

Hat77b. Hatze, H.: A complete set of control equations for the human musculoskeletal system. J. Biomech. **10**, 799–805, (1977)

Hat78. Hatze, H.: A general myocybernetic control model of skeletal muscle. Biol. Cyber., **28**, 143–157, (1978)

Hay94. Haykin, S. Neural Networks: A Comprehensive Foundation. Macmillan, (1994)

HBB96. Houk, J.C., Buckingham, J.T., Barto, A.G.: Models of the cerebellum and motor learning. Behavioral and Brain Sciences, **19**(3), 368–383, (1996)

HDR02. Hasty, J., Dolnik, M., Rottschafer, V., Collins, J.J.: A synthetic gene network for entraining and amplifying cellular oscillations. Phys. Rev. Lett. **88**(14), 1–4, (2002)

HE79. Hawking, S.W., Israel, W. (ed.): General relativity: an Einstein centenary survey. Cambridge Univ. Press, Cambridge, (1979)

Heb49. Hebb, D.O.: The Organization of Behavior, Wiley, New York, (1949)

Hec77. Heckhausen, H.: Achievement motivation and its constructs: a cognitive model. Motiv. Emot, **1**, 283–329, (1977)

Hec87. Hecht-Nielsen, R.: Counterpropagation networks. Applied Optics, **26**(23), 4979–4984, (1987)

Hee90. Heermann, D.W.: Computer Simulation Methods in Theoretical Physics. (2nd ed), Springer, Berlin, (1990)

HH52. Hodgkin, A.L., Huxley, A.F.: A quantitative description of membrane current and application to conduction and excitation in nerve. J. Physiol., **117**, 500–544, (1952)

HI97. Hoppensteadt, F.C., Izhikevich, E.M.: Weakly Connected Neural Networks. Springer, New York, (1997)

Hil38. Hill, A.V.: The heat of shortening and the dynamic constants of muscle, Proc. R. Soc. B, **76**, 136–195, (1938)

Hir66. Hirzebruch, F.: Topological Methods in Algebraic Geometry, Springer-Verlag, Berlin, (1966)

Hir76. Hirsch, M.W.: Differential Topology. Springer, New York, (1976)

HKK03. Hori, K., Katz, S., Klemm, A., Pandharipande, R., Thomas, R., Vafa, C., Vakil, R., Zaslow, E.: Mirror symmetry. Clay Mathematics Monographs 1, American Mathematical Society, Providence, Clay Mathematics Institute, Cambridge, MA, (2003)

HL84. Hamoui, A., Lichnerowicz, A.: Geometry of dynamical systems with time–dependent constraints and time-dependent Hamiltonians: An approach towards quantization. J. Math. Phys. **25**, 923, (1984)

HMM95. Hehl, F., McCrea, J., Mielke, E., Ne'eman, Y.: Metric–affine gauge theory of gravity. Phys. Rep. **258**, 1, (1995)

HN54. Huxley, A.F., Niedergerke, R.: Changes in the cross–striations of muscle during contraction and stretch and their structural interpretation. Nature, **173**, 973–976, (1954)

Hod64. Hodgkin, A.L.: The Conduction of the Nervous Impulse. Liverpool Univ. Press, Liverpool, (1964)

Hop82). Hopfield, J.J.: Neural networks and physical systems with emergent collective computational activity. Proc. Natl. Acad. Sci. USA., **79**, 2554–2558, (1982)

Hop84. Hopfield, J.J.: Neurons with graded response have collective computational properties like those of two–state neurons. Proc. Natl. Acad. Sci. USA, **81**, 3088–3092, (1984)

Hou79. Houk, J.C.: Regulation of stiffness by skeletomotor reflexes. Ann. Rev. Physiol., **41**, 99–114, (1979)

HP96. Hameroff, S.R., Penrose, R.: Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. In: Hameroff, S. R., Kaszniak, A.W. and Scott, A.C. Eds: Toward a Science of Consciousness: the First Tucson Discussion and Debates, 507–539. MIT Press, Cambridge, MA, (1996)

HP96. Hawking, S.W., Penrose, R.: The Nature of Space and Time. Princetone Univ. Press, Princetone, NJ, (1996)

HS74. Hirsch, M. Smale, S.: Differential Equations, Dynamical Systems, and Linear Algebra, Academic Press, New York, (1974)

HS98. Hofbauer, J., Sigmund, K.: Evolutionary games and population dynamics, Cambridge Univ. Press, U.K., (1998)

HSK92. Hauser, J., Sastry, S., Kokotovic, P.: Nonlinear control via approximate input–output linearization: The ball and beam example, IEEE Trans. Aut. Con., AC–37, 392–398, (1992)

HT85. Hopfield, J.J., Tank, D.W.: Neural computation of decisions in optimisation problems. Biol. Cybern., **52**, 114–152, (1985)

HT92. Henneaux, M., Teitelboim, C.: Quantization of Gauge systems. Princeton Univ. Press, (1992)

HT93. Hunt, L, Turi, J.: A new algorithm for constructing approximate transformations for nonlinear systems. IEEE Trans. Aut. Con., AC–38,1553–1556, (1993)

Hul00. Hull, J.C.: Options, Futures, and Other Derivatives. (4th ed.), Prentice Hall, New Jersey, (2000)

Hur93. Hurmuzlu, Y.: Dynamics of bipedal gait. J. Appl. Mech., **60**, 331–343, (1993)

Hux57. Huxley, A.F.: Muscle structure and theories of contraction. Progr. Biophys. Chem., **7**, 255–328, (1957)

HYW03. Hao, N., Yildirim, N., Wang, Y., Elston, T.C., Dohlman, H.G.: Regulators of G protein signaling and transient activation

of signaling: experimental and computational analysis reveals negative and positive feedback controls on G protein activity. J. Biol. Chem., 278, 46506-46515, (2003)

IB05. Ivancevic, V., Beagley, N.: Brain-like functor control machine for general humanoid biodynamics. Int. J. Math. Math. Sci. **11**, 1759–1779, (2005)

Ich04. Ichinomiya, T.: Frequency synchronization in random oscillator network. Phys. Rev. E **70**, 026116, (2004)

Ich05. Ichinomiya, T.: Path-integral approach to the dynamics in sparse random network. Phys. Rev. E **72**, 016109, (2005)

II05. Ivancevic, V., Ivancevic, T.: Human–Like Biomechanics. Springer, Mechanical Engineering Ser., (2005)

IJB99a. Ivancevic, T., Jain, L.C., Bottema, M.: New Two–feature GBAM–Neurodynamical Classifier for Breast Cancer Diagnosis. Proc. KES'99, IEEE Press, (1999)

IJB99b. Ivancevic, T., Jain, L.C., Bottema, M.: A New Two–Feature FAM–Matrix Classifier for Breast Cancer Diagnosis. Proc. KES'99, IEEE Press, (1999)

IK80. Iyanaga, S., Kawada, Y. (eds.): Pontryagin's Maximum Principle. In Encyclopedic Dictionary of Mathematics. MIT Press, Cambridge, MA, 295-296, (1980)

Ike90. Ikeda, S.: Some remarks on the Lagrangian Theory of Electromagnetism. Tensor, N.S. **49**, (1990)

Ila01. Ilachinski, A.: Cellular automata. World Scientific, Singapore, (2001)

ILI95. Ivancevic, V., Lukman, L., Ivancevic, T.: Selected Chapters in Human Biomechanics. Textbook (in Serbian). Univ. Novi Sad Press, Novi Sad, (1995)

IN92. Igarashi, E., Nogai, T.: Study of lower level adaptive walking in the saggital plane by a biped locomotion robot. Advanced Robotics, **6**, 441–459, (1992)

Ing97. Ingber, L.: Statistical mechanics of neocortical interactions: Applications of canonical momenta indicators to electroencephalography, Phys. Rev. E, **55**(4), 4578-4593, (1997)

Ing98. Ingber, L.: Statistical mechanics of neocortical interactions: Training and testing canonical momenta indicators of EEG, Mathl. Computer Modelling **27**(3), 33–64, (1998)

IP01a. Ivancevic, V., Pearce, C.E.M.: Poisson manifolds in generalised Hamiltonian biomechanics. Bull. Austral. Math. Soc. **64**, 515–526, (2001)

IP01b. Ivancevic, V., Pearce, C.E.M.: Topological duality in humanoid robot dynamics. ANZIAM J. **43**, 183–194, (2001)

IP05b. Ivancevic, V., Pearce, C.E.M.: Hamiltonian dynamics and Morse topology of humanoid robots. Gl. J. Mat. Math. Sci. (to appear)

IS00. Ivancevic, V., Snoswell, M.: Torson Sports Cybernetics. Torson Group. Inc., Adelaide, SA, (2000)

IS01. Ivancevic, V., Snoswell, M.: Fuzzy–stochastic functor machine for general humanoid–robot dynamics. IEEE Trans. on Sys, Man, Cyber. B, **31**(3), 319–330, (2001)

Isi89. Isidori, A.: Nonlinear Control Systems. An Introduction, (2nd ed) Springer, Berlin, (1989)

Ito60. Ito, K.: Wiener Integral and Feynman Integral. Proc. Fourth Berkeley Symp. Math., Stat., Prob., **2**, 227–238, (1960)

Iva02. Ivancevic, V.: Generalized Hamiltonian biodynamics and topology invariants of humanoid robots. Int. J. Mat. Mat. Sci., **31**(9), 555–565, (2002)

Iva04. Ivancevic, V.: Symplectic Rotational Geometry in Human Biomechanics. SIAM Rev., **46**(3), 455-474, (2004)

Iva05a. Ivancevic, V.: Dynamics of Humanoid Robots: Geometrical and Topological Duality. In Biomathematics: Modelling and simulation (ed. J.C. Misra), World Scientific, Singapore (to appear)

Iva05b. Ivancevic, V.: A Lagrangian model in human biomechanics and its Hodge–de Rham cohomology. Int. J. Appl. Math. Mech., **3**, 35–51, (2005)

Iva05c. Ivancevic, V.: Lie–Lagrangian model for realistic human biodynamics. Int. J. Humanoid Robotics (World Sci.) (to appear)

Iva91. Ivancevic, V.: Introduction to Biomechanical Systems: Modelling, Control and Learning (in Serbian). Scientific Book, Belgrade, (1991)

Iva95. Ivancevic, T.: Some Possibilities of Multilayered Neural Networks Application in Biomechanics of Muscular Contractions, Human Motion and Sport Training. Master Thesis (in Serbian), University of Novi Sad, YU, (1995)

Izh01. Izhikevich, E.M.: Resonate-and-fire neurons. Neu. Net., **14**, 883–894, (2001)

Izh04. Izhikevich, E.M.: Which model to use for cortical spiking neurons? IEEE Trans. Neu. Net., **15**, 1063–1070, (2004)

Izh99. Izhikevich, E.M.: Class 1 neural excitability, conventional synapses, weakly connected networks, and mathematical foundations of pulse-coupled models. IEEE Trans. Neu. Net., **10**, 499–507, (1999)

JPY96. Jibu, M., Pribram, K.H., Yasue, K.: From conscious experience to memory storage and retrieval: the role of quantum brain dynamics and boson condensation of evanescent photons, Int. J. Mod. Phys. B, **10**, 1735, (1996)

JY95. Jibu, M., Yasue, K.: Quantum brain dynamics and consciousness. John Benjamins, Amsterdam, (1995)

Kac51. Kac, M.: On Some Connection between Probability Theory and Differential and Integral Equations. Proc. 2nd Berkeley Sympos. Math. Stat. and Prob., 189-215, (1951)

Kal60. Kalman, R.E.: A new approach to linear filtering and prediction problems. Transactions of the ASME, Ser. D, J. Bas. Eng., **82**, 34–45, (1960)

Kan58. Kan, D.M.: Adjoint Functors. Trans. Am. Math. Soc. **89**, 294–329, (1958)

Kan98. Kanatchikov, I.: Canonical structure of classical field theory in the polymomentum phase space. Rep. Math. Physics, **41**(1), 49-90, (1998)

Kap05. Kappen, H.J.: A linear theory for control of nonlinear stochastic systems. Phys. Rev. Let. (to appear)

Kau94. Kauffman, L.H.: Knots and Physics, (2nd ed.) World Scientific, Singapore (1994)

KFA69. Kalman, R.E., Falb, P., Arbib, M.A.: Topics in Mathematical System Theory. McGraw Hill, New York, (1969)

KHS93. Koruga, D.L., Hameroff, S.I., Sundareshan, M.K., Withers, J., Loutfy, R.: Fullerence C60: History, Physics, Nanobiology and Nanotechnology. Elsevier Science Pub, (1993)

KK01. Kawahara, G., Kida, S.: Periodic motion embedded in plane Couette turbulence: regeneration cycle and burst. J. Fluid Mech. **449**, 291–300, (2001)

KKV99. Katz, S., Klemm, A., Vafa, C.: M–theory, topological strings and spinning black holes. Adv. Theor. Math. Phys. **3**, 1445, (1999)

Kla00. Klauder, J.R.: Beyond Conventional Quantization, Cambridge Univ. Press, Cambridge, (2000)

Kla00. Klauder, J.R.: Beyond Conventional Quantization. Cambridge Univ. Press, Cambridge, UK, (2000)

Kla97. Klauder, J.R.: Understanding Quantization. Found. Phys. **27**, 1467–1483, (1997)

KLV85. Krasil'shchik, I., Lychagin, V., Vinogradov, A.: Geometry of Jet Spaces and Nonlinear Partial Differential Equations, Gordon and Breach, Glasgow, (1985)

KMS93. Kolar, I., Michor, P.W., Slovak, J.: Natural Operations in Differential Geometry. Springer, Berlin, (1993)

KN00. Kotz, S., Nadarajah, S.: Extreme Value Distributions. Imperial College Press, London, (2000)

KN63/9. Kobayashi, S., Nomizu, K.: Foundations of Differential Geometry, Vols. 1,2., Interscience Publ., New York, (1963/1969).

KO89. Kamron, N., Olver, P.J.: Le probléme d'equivalence à une divergence prés dans le calcul des variations des intégrales multiples. C. R. Acad. Sci. Paris,t. **308**, Sèrie I, 249–252, (1989)

Koc81. Kock, A.: Synthetic Differential Geometry, London Math.Soc. Lecture Notes Series No. 51, Cambridge Univ. Press, Cambridge, (1981)

Kos04. Kosmann-Schwarzbach, Y.: Derived brackets. Lett. Math. Phys. **69**, 61–87, (2004)

Kos92. Kosko, B.: Neural Networks and Fuzzy Systems, A Dynamical Systems Approach to Machine Intelligence. Prentice–Hall, New York, (1992)

Kos96. Kosko, B.: Fuzzy Engineering. Prentice Hall, New York, (1996)

KR03. Kock, A., Reyes, G.E.: Some calculus with extensive quantities: wave equation. Theory and Applications of Categories, **11**(14), 321–336, (2003)

KR95. Kauffman, L.H., Radford, D.E.: Invariants of 3

KR99. Kauffman, L.H., Radford, D.E.: Quantum algebra structures on $n \times n$ matrices. J. Algebra **213**, 405–436, (1999)

Kre84. Krener, A.: Approximate linearization by state feedback and coordinate change, Systems Control Lett., 5, 181–185, (1984)

Kru97. Krupkova, O.: The Geometry of Ordinary Variational Equations. Springer, Berlin, (1997)

KS76. Kijowski, J., Szczyrba, W.: A Canonical Structure for Classical Field Theories. Commun. Math. Phys. **46**, 183, (1976)

KT75. Kamber, F., Tondeur, P.: Foliated Bundles and Characteristic Classes. Lecture Notes in Math. 493, Springer, Berlin, (1975)

KT79. Kijowski, J., Tulczyjew, W.: A Symplectic Framework for Field Theories, Springer-Verlag, Berlin, (1979)

Kuh85. Kuhl, J.: Volitional Mediator of cognition-Behaviour consistency: Self-regulatory Processes and action versus state orientation (pp. 101-122). In: J. Kuhl & S. Beckman (Eds.) Action control: From Cognition to Behaviour. Springer, Berlin, (1985)

Kur76. Kuramoto, Y., Tsuzuki, T.: Persistent propagation of concentration waves in dissipative media far from thermal equilibrium. Progr. Theor. Physics **55**, 365, (1976)

Kur84. Kuramoto, Y.: Chemical Oscillations. Waves and Turbulence. Springer, New York, (1984)

KV03a. Katic, D., Vukobratovic, M.: Advances in Intelligent Control of Robotic Systems, Book series: Microprocessor-Based and Intelligent Systems Engineering, Kluwer Acad. Pub., Dordrecht, (2003)

KV03b. Katic, D., Vukobratovic, M.: Survey of intelligent control techniques for humanoid robots. J. Int. Rob. Sys, **37**, 117–141, (2003)

KV98. Katic, D., Vukobratovic, M.: A neural network-based classification of environment dynamics models for compliant control of manipulation robots. IEEE Trans. Sys., Man, Cyb., B, **28**(1), 58–69, (1998)

Lam94. Lamport, L.: The temporal logic of actions. ACM Transactions on Programming Languages and Systems, **16**(3), 872–923, (1994)

Lan95. Landsman, N.P.: Against the Wheeler–DeWitt equation. Class. Quan. Grav. L **12**, 119–124, (1995)

LD05. Levine, M., Davidson, E.H.: Gene regulatory networks for development. Proc Natl. Acad. Sci. USA, **102**(14), 4936–4942, (2005)

Lei02. Leinster, T.: A survey of definitions of $n-$category. Theor. Appl. Categ. **10**, 1–70, (2002)

Lei03. Leinster, T.: Higher Operads, Higher Categories, London Mathematical Society Lecture Notes Series, Cambridge Univ. Press, (2003)

Lei04. Leinster, T.: Operads in higher-dimensional category theory. Theor. Appl. Categ. **12**, 73–194, (2004)

Lew00a. Lewis, A.D.: Simple mechanical control systems with constraints, IEEE Trans. Aut. Con., **45**(8), 1420–1436, (2000)

Lew00b. Lewis, A.D.: Affine connection control systems. Proceedings of the IFAC Workshop on Lagrangian and Hamiltonian Methods for Nonlinear Control 128–133, Princeton, (2000)

Lew51. Lewin, K.: Field Theory in Social Science. Univ. Chicago Press, Chicago, (1951)

Lew95. Lewis, A.D.: Aspects of Geometric Mechanics and Control of Mechanical Systems. Technical Report CIT-CDS 95-017 for the Control and Dynamical Systems Option, California Institute of Technology, Pasadena, CA, (1995)

Lew97. Lewin, K.: Resolving Social Conflicts: Field Theory in Social Science, American Psych. Assoc., New York, (1997)

Lew98. Lewis, A.D.: Affine connections and distributions with applications to nonholonomic mechanics, Reports on Mathematical Physics, **42**(1/2), 135–164, (1998)

Lew99. Lewis, A.D.: When is a mechanical control system kinematic?, in Proceedings of the 38th IEEE Conf. Decis. Con., 1162–1167, IEEE, Phoenix, AZ, (1999)

LGS98. Lygeros, J., Godbole, D.N., Sastry, S.: Verified hybrid controllers for automated vehicles, IIEEE Trans. Aut. Con., **43**, 522-539, (1998)

Li04a. Li, Y.: Persistent homoclinic orbits for nonlinear Schrödinger equation under singular perturbation. Dyn. PDE, **1**(1), 87–123, (2004)

Li04b. Li, Y.: Existence of chaos for nonlinear Schrödinger equation under singular perturbation. Dyn. PDE. **1**(2), 225-237, (2004)

Li04c. Li, Y.: Homoclinic tubes and chaos in perturbed sine-Gordon equation. Cha. Sol. Fra., **20**(4), 791–798, 2004)

Li04d. Li, Y.: Chaos in Miles' equations. Cha. Sol. Fra., **22**(4), 965–974, (2004)

Li04. Li, Y.: Chaos in Partial Differential Equations. Int. Press, Sommerville, MA, (2004)

LL59. Landau, L.D., Lifshitz, E.M.: Fluid Mechanics. Pergamon Press, (1959)

LL98. Labastida, J.M.F., Lozano, C.: Lectures on topological quantum field theory, in Proceedings of the CERN–Santiago de Compostela–La Plata Meeting on 'Trends in Theoretical Physics', eds. H. Falomir, R. Gamboa, F. Schaposnik, Amer. Inst. Physics, New York, (1998)

LM03. Lopez, M.C., Marsden, J.E.: Some remarks on Lagrangian and Poisson reduction for field theories. J. Geom. Phys., **48**, 52-83, (2003)

LM87. Libermann, P., Marle, C.M.: Symplectic Geometry and Analytical Mechanics, Reidel, Dordrecht, (1987)

LM93. de León, M., Marrero, J.: Constrained time-dependent Lagrangian systems and Lagrangian submanifolds, J. Math. Phys. **34**, 622, (1993)

LM96. de León, M., Martín de Diego, D.: On the geometry of nonholonomic Lagrangian systems. J. Math. Phys. **37**, 3389, (1996)

LM97. Lewis, A.D., Murray, R.M.: Controllability of simple mechanical control systems, SIAM J. Con. Opt., **35**(3), 766–790, (1997)

LM99. Lewis, A.D., Murray, R.M.: Configuration controllability of simple mechanical control systems, SIAM Review, **41**(3), 555–574, (1999)

LMD04. de León, M., Martin, D., de Diego, A.: Santamaria-Merino: Symmetries in Classical Field Theory. arXiv:math-ph/0404013.

LMM97. de León, M., Marrero, J., Martín de Diego D.: Nonholonomic Lagrangian systems in jet manifolds. J. Phys. A. **30**, 1167, (1997)

Lol01. Loll, R.: Discrete Lorentzian quantum gravity. Nucl. Phys. B **94**, 96–107, (2001)

Lol98. Loll, R.: Discrete approaches to quantum gravity in four dimensions. Living Reviews in Relativity, **13**, (1998)

Loo00. Loo, K.: A Rigorous Real Time Feynman Path Integral and Propagator. J. Phys. A: Math. Gen, **33**, 9215–9239, (2000)

Loo99. Loo, K.: A Rigorous Real Time Feynman Path Integral. J. Math. Phys., **40**(1), 64–70, (1999)

Lor63. Lorenz, E.N.: Deterministic Nonperiodic Flow. J. Atmos. Sci., **20**, 130–141, (1963)

LP94. Langer, J., Perline, R.: Local geometric invariants of integrable evolution equations, J. Math. Phys., **35**(4), 1732–1737, (1994)

LRR02. Lee, T.I., Rinaldi, N.J., Robert, F., Odom D.T., Bar-Joseph Z., et al.: Transcriptional regulatory networks in Saccharomyces cerevisiae. Science, **298**, 799–804, (2002)

LS97. Louko, J., Sorkin, R.D.: Complex actions in two-dimensional topology change. Class. Quant. Grav. **14**, 179–204, (1997)

LSA03. Louzoun, Y., Solomon, S., Atlan, H., Cohen, I.R.: Proliferation and competition in discrete biological systems. Bull. Math. Biol. **65**, 375, (2003)

Lu90. Lu, J.-H.: Multiplicative and affine Poisson structures on Lie groups. PhD Thesis, Berkeley Univ., Berkeley, (1990)

Lu91. Lu, J.-H.: Momentum mappings and reduction of Poisson actions. in Symplectic Geometry, Groupoids, and Integrable Systems, eds.: P. Dazord and A. Weinstein, 209–225, Springer, New York, (1991)

LV97. Li, M., Vitanyi, P.: An Introduction to Kolmogorov Complexity and its Applications. Springer, New York, (1997)

LVW89. Lerche, W., Vafa, C., Warner, N.P.: Chiral Rings In N=2 Superconformal Theories, Nucl. Phys. B **324**, 427, (1989)

LW05. Lipan, O., Wong, W.H.: The use of oscillatory signals in the study of genetic networks. Proc. Natl. Acad. Sci. USA, **102**, 7063-7068, (2005)

MA02. Mascalchi, M. *et al.*: Proton MR Spectroscopy of the Cerebellum and Pons in Patients with Degenerative Ataxia, Radiology, **223**, 371, (2002)

MA94. Miron, R., Anastasiei, M.: The Geometry of Lagrangian Spaces: Theory and Applications, Kluwer Academic Publishers, (1994)

MacL71. MacLane, S.: Categories for the Working Mathematician. Springer, New York, (1971)

Man98. Manikonda, V.: Control and Stabilization of a Class of Nonlinear Systems with Symmetry. PhD Thesis, Center for Dynamics and Control of Smart Structures, Harvard Univ., Cambridge, MA, (1998)

Mat02. Matacz, A.: Path Dependent Option Pricing: the Path Integral Partial Averaging Method. J. Comp. Finance, **6**(2), (2002)

Mat82. Matsumoto, M.: Foundations of Finsler Geometry and Special Finsler Spaces. Kaisheisha Press, Kyoto, (1982)

May81. Mayer, P.A.: A Differential Geometric Formalism for the Ito Calculus. Lecture Notes in Mathematics, Vol. 851, Springer, New York, (1981)

MBP00. Myung, I.J., Balasubramanian, V., Pitt, M.A.: Counting probability distributions: differential geometry and model selection. Proceedings of the National Academy of Science, USA, **97**, 11170–11175, (2000)

McCul87. McCullagh, P.: Tensor methods in statistics. Monographs in statistics and applied probability. Chapman & Hall, Cambridge, UK, (1987)

Men98. Mendes, R.V.: Conditional exponents, entropies and a measure of dynamical self-organization. Phys. Let. A **248**, 167–1973, (1998)

Mer73. Merton, R.: Theory of Rational Option Pricing. Bell J. Econom. Managem. Sci. **4**, 141–183, (1973)

Met97. Metzger, M.A.: Applications of nonlinear dynamical systems theory in developmental psychology: Motor and cognitive development. Nonlinear Dynamics, Psychology, and Life Sciences, **1**, 55–68, (1997)

MFB00. Myung, I.J., Forster, M., Browne, M.W.: Special issue on model selection. J. Math. Psych., **44**, 1–2, (2000)

MFV90. Morandi, G., Ferrario, C., Lo Vecchio, G., Marmo, G., Rubano, C.: The inverse problem in the calculus of variations and the geometry of the tangent bundle, Phys. Rep. **188**, 147, (1990)

MH92. Meyer, K.R., Hall, G.R.: Introduction to Hamiltonian Dynamical Systems and the N–body Problem. Springer, New York, (1992)

Mic01. Michor, P.W.: Topics in Differential Geometry. Lecture notes of a course in Vienna, (2001)

Mil56. Miller, G.A.: The magical number seven, plus or minus two: Some limits on our capacity for processing information, Psych. Rev., **63**, 81–97, (1956)

Mil63. Milnor, J.: Morse Theory. Princeton Univ. Press, Princeton, (1963)

Mil65. Milnor, J.: Lectures on the H–Cobordism Theorem. Math. Notes. Princeton Univ. Press, Princeton, (1965)

Mil99. Milinković, D.: Morse homology for generating functions of Lagrangian submanifolds. Trans. Amer. Math. Soc. **351**(10), 3953–3974, (1999)

MK05. Moon, S.J., Kevrekidis, I.G.: An equation-free approach to coupled oscillator dynamics: the Kuramoto model example. Submitted to Int. J. Bifur. Chaos, (2005)

MKA88. Miron, R., Kirkovits, M.S., Anastasiei, M.: A Geometrical Model for Variational Problems of Multiple Integrals. Proc. Conf. Diff. Geom Appl. Dubrovnik, Yugoslavia, (1988)

ML81. Morris, C., Lecar, H.: Voltage oscillations in the barnacle giant muscle fiber. Biophys. J., **35**, 193–213, (1981)

Mla91. Mladenova, C.: Mathematical Modelling and Control of Manipulator Systems. Int. J. Rob. Comp. Int. Man., **8**(4), 233–242, (1991)

Mla99. Mladenova, C.: Applications of Lie Group Theory to the Modelling and Control of Multibody Systems. Mult. Sys. Dyn., **3**(4), 367–380, (1999)

MLS94. Murray, R.M., Li, X., Sastry, S.: Robotic Manipulation, CRC Press, Boco Raton, Fl, (1994)

MM92. Marathe, K., Martucci, G.: The Mathematical Foundations of Gauge Theories. North-Holland, Amsterdam, (1992)

MN95a. Mavromatos, N.E., Nanopoulos, D.V.: A Non-critical String (Liouville) Approach to Brain Microtubules: State Vector reduction, Memory coding and Capacity. ACT-19/95, CTP-TAMU-55/95, OUTP-95-52P, (1995)

MN95b. Mavromatos, N.E., Nanopoulos, D.V.: Non-Critical String Theory Formulation of Microtubule Dynamics and Quantum Aspects of Brain Function. ENSLAPP-A-524/95, (1995)

MNM02. Montagna, G., Nicrosini, O., Moreni, N.: A path integral way to option pricing. Physica A **310**, 450 – 466, (2002)

Moh91. Mohler, R.R.: Nonlinear systems, Vol. 2. Applications to bilinear control. Prentice Hall, Inc. (1991)

Mor34. Morse, M.: The Calculus of Variations in the Large. Amer. Math. Soc. Coll. Publ. No. 18, Providence, RI, (1934)

MOS99. Mangiarotti, L., Obukhov, Yu., Sardanashvily, G.: Connections in Classical and Quantum Field Theory. World Scientific, Singapore, (1999)

MP03. Myung, I.J., Pitt, M.A.: Model Evaluation, Testing and Selection. To appear in K. Lambert and R. Goldstone (eds.), Handbook of Cognition, Sage Publ., (2003)

MP94. Massa, E., Pagani, E.: Jet bundle geometry, dynamical connections and the inverse problem of Lagrangian mechanics. Ann. Inst. Henri Poincaré **61**, 17, (1994)

MPS98. Marsden, J.E., Patrick, G.W., Shkoller, S.: Multisymplectic Geometry, Variational Integrators, and Nonlinear PDEs. Comm. Math. Phys., **199**, 351–395, (1998)

MR92. Muñoz-Lecanda, M. and Román-Roy, N.: Lagrangian theory for presymplectic systems. Ann. Inst. Henrí Poincaré **57**, 27, (1992)

MR99. Marsden, J.E., Ratiu, T.S.: Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems. (2nd ed), Springer, New York, (1999)

MRS04. Munteanu, F., Rey, A.M., Salgado, M.: The Günther's formalism in classical field theory: momentum map and reduction. J. Math. Phys. **45**(5), 1730–1750, (2004)

MS00a. Mangiarotti, L., Sardanashvily, G.: Connections in Classical and Quantum Field Theory, World Scientific, Singapore, (2000)

MS00b. Mangiarotti, L., Sardanashvily, G.: Constraints in Hamiltonian time-dependent mechanics. J. Math. Phys. **41**, 2858, (2000)

MS00. Meyer, K.R., Schmidt, D.S.: From the restricted to the full three-body problem. Trans. Amer. Math. Soc., **352**, 2283–2299, (2000)

MS03. Moroianu, A., Semmelmann, U.: Twistor forms on Kähler manifolds. Ann. Scuola Norm. Sup. Pisa Cl. Sci. **2**(4), 823–845, (2003)

MS78. Modugno, M., Stefani, G.: Some results on second tangent and cotangent spaces. Quadernidell' Instituto di Matematica dell' Universit a di Lecce Q., **16**, (1978)

MS95. Marmo, G., Simoni, A., Stern, A.: Poisson Lie group symmetries for the isotropic rotator. Int. J. Mod. Phys. A **10**, 99–114, (1995)

MS98. Mangiarotti, L., Sardanashvily, G.: Gauge Mechanics. World Scientific, Singapore, (1998)

MS99. Mangiarotti, L., Sardanashvily, G.: On the geodesic form of non-relativistic dynamic equations. arXiv: math-ph/9906001.

MTW73. Misner, C.W., Thorne, K.S., Wheeler, J.A.: Gravitation. Freeman, San Francisco, (1973)

Mus99. Mustafa, M.T.: Restrictions on harmonic morphisms. Confomal Geometry and Dynamics (AMS), **3**, 102–115, (1999)

MW74. Marsden, J.E., Weinstein, A.: Reduction of symplectic manifolds with symmetry. Rept. Math. Phys. **5**, 121–130, (1974)

MZA03. Mangan, S., Zaslaver, A., Alon, U.: The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. JMB, **334**(2), 197-204, (2003)

Nan95. Nanopoulos, D.V.: Theory of Brain Function, Quantum Mechanics and Superstrings. CERN-TH/95128, (1995)

Nas83. Nash, C., Sen, S.: Topology and Geometry for Physicists. Academic Press, London, (1983)

Nay73. Nayfeh, A.H.: Perturbation Methods. Wiley, New York, (1973)

Nea00. Neagu, M.: Upon $h-$normal $\Gamma-$linear connection on $J^1(T, M)$. arXiv:math.DG/0009070.

Nea02. Neagu, M.: The Geometry of Autonomous Metrical Multi-Time Lagrange Space of Electrodynamics. Int. J. Math. Math. Sci. **29**, 7-16, (2002)

NHP89. Noakes, L., Heinzinger, G., Paden, B.: Cubic splines on curved spaces, IMA J. Math. Con. Inf. , **6**(4), 465–473, (1989)

NMP84. Novikov, S.P., Manakov, S.V., Pitaevskii, L.P., Zakharov, V.E.: Theory of Solitons, Plenum/Kluwer, Dordrecht, (1984)

NS90. Nijmeijer, H., van der Schaft, A.J.: Nonlinear Dynamical Control Systems. Springer, New York, (1990)

NU00a. Neagu, M. Udrişte, C.: Multi-Time Dependent Sprays and Harmonic Maps on $J^1(T, M)$, Third Conference of Balkan Society of Geometers, Politehnica University of Bucarest, Romania, (2000)

NU00b. Neagu, M. Udrişte, C.: Torsions and Curvatures on Jet Fibre Bundle $J^1(T, M)$. arXiv:math.DG/0009069.

Olv86. Olver, P.J.: Applications of Lie Groups to Differential Equations (2nd ed.) Graduate Texts in Mathematics, Vol. 107, Springer, New York, (1986)

Omo86. Omohundro, S.M.: Geometric Perturbation Theory in Physics. World Scientific, Singapore, (1986)

Ori01. Oriti, D.: Spacetime geometry from algebra: Spin foam models for non-perturbative quantum gravity. Rept. Prog. Phys. **64**, 1489–1544, (2001)

OSV04. Ooguri, H., Strominger, A., Vafa, C.: Black hole attractors and the topological string. Phys. Rev. D **70**, 106007, (2004)

OTK02. Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D., van Oudenaarden, A.: Regulation of noise in the expression of a single gene. Nature Gen., **31**,69-73, (2002)

OTL04. Ozbudak, E.M., Thattai, M., Lim, H.N., Shraiman, B.I., van Oudenaarden, A.: Multistability in the lactose utilization network of Escherichia coli. Nature, **427**(6976), 737–40, (2004)

PB89. Peyrard, M., Bishop, A.R.: Statistical mechanics of a nonlinear model for DNA denaturation. Phys. Rev. Lett. **62**, 2755, (1989)

PB99. Paul, W., Baschnagel, J.: Stochastic Processes: from Physics to Finance. Springer, Berlin, (1999)

PBK62. Pontryagin, L., Boltyanskii, V., Gamkrelidze, R., Mishchenko, E.: The mathematical theory of optimal processes. Interscience, (1962)

PBS01. Potters, M., Bouchaud, J.P., Sestovic, D.: Hedged Monte-Carlo: low variance derivative pricing with objective probabilities. Physica A **289**, 517-25, (2001)

PC05. Park, J., Chung, W.-K.: Geometric Integration on Euclidean Group With Application to Articulated Multibody Systems. IEEE Trans. Rob. **21**(5), 850–863, (2005)

PC91. Pecora, L.M., Carroll, T.L.: Driving systems with chaotic signals. Phys. Rev. A, **44**, 2374-2383, (1991)

PC98. Pecora, L.M., Carroll, T.L.: Master stability functions for synchronized coupled systems. Phys. Rev. Lett. **80**, 2109-2112, (1998)

Pen67. Penrose, R.: Twistor algebra. J. Math. Phys., **8**, 345–366, (1967)

Pen89. Penrose, R.: The Emperor's New Mind, Oxford Univ. Press, Oxford, (1989)

Pen94. Penrose, R.: Shadows of the Mind. Oxford Univ. Press, Oxford, (1994)

Pen97. Penrose, R.: The Large, the Small and the Human Mind. Cambridge Univ. Press, (1997)

Pes77. Pesin, Ya.B.: Lyapunov Characteristic Exponents and Smooth Ergodic Theory. Russ. Math. Surveys, **32**(4), 55–114, (1977)

Pet98. Petersen, P.: Riemannian Geometry. Springer, New York, (1998)

Pet99. Petersen, P.: Aspects of Global Riemannian Geometry. Bull. Amer. Math. Soc., **36**(3), 297–344, (1999)

PHE92. Piaget, J., Henriques, G., Ascher, E.: Morphisms and categories. Erlbaum Associates, Hillsdale, NJ, (1992)

PI03. Pearce, C.E.M., Ivancevic, V.: A generalised Hamiltonian model for the dynamics of human motion. In Differential Equations and Applications, Vol. 2, Eds. Y.J. Cho, J.K. Kim and K.S. Ha, Nova Science, New York, (2003)

PI04. Pearce, C.E.M., Ivancevic, V.: A qualitative Hamiltonian model for the dynamics of human motion. In Differential Equations and Applications, Vol. 3, Eds. Y.J. Cho, J.K. Kim and K.S. Ha, Nova Science, New York, (2004)

PLS00. Pappas, G.J., Lafferriere, G., Sastry, S.: Hierarchically consistent control systems. IEEE Trans. Aut. Con., **45**(6), 1144–1160, (2000)

PMZ02. Pitt, M.A., Myung, I.J., Zhang, S.: Toward a method of selecting among computational models of cognition Psychological Review, **109**(3), 472–491, (2002)

PO82. Ponomarev, V., Obukhov, Yu.: Generalized Einstein-Maxwell theory. GRG **14**, 309, (1982)

Pol98. Polchinski, J.: String theory. 2 Vols. Cambridge Univ. Press, (1998)

Pom78. Pommaret, J.: Systems of Partial Differential Equations and Lie Pseudogroups, Gordon and Breach, Glasgow, (1978)

Pop59. Popper, K.R: The Logic of Scientific Discovery. New York, NY: Basic Books, (1959)

Pos86. Postnikov, M.M.: Lectures in Geometry V, Lie Groups and Lie Algebras, Mir Publ., Moscow, (1986)

PR02. Paufler, C., Römer, H.: The geometry of Hamiltonian n–vector–fields in multisymplectic field theory. J. Geom. Phys. **44**(1), 52–69, (2002)

PR86. Penrose, R., Rindler, W.: Spinors and Space-time II. Cambr. Univ. Press, Cambridge, (1986)

Pra91. Pratt, V.: Modelling concurrency with geometry. in: Proc. of the 18th ACM Symposium on Principles of Programming Languages, (1991)

Pry96. Prykarpatsky, A.K.: Geometric models of the Blackmore's swept volume dynamical systems and their integrability, In: Proc. of the IMACS-95, Hamburg 1995, ZAMP, **247**(5), 720–724, (1996)

PS02. Pappas, G.J., Simic, S.: Consistent hierarchies of affine nonlinear systems. IEEE Trans. Aut. Con., **47**(5), 745–756, (2002)

PSS96. Pons, J.M., Salisbury, D.C., Shepley, L.C.: Gauge transformations in the Lagrangian and Hamiltonian formalisms of generally covariant theories. arXiv gr-qc/9612037, (1996)

PST86. Park, J.K., Steiglitz, K. and Thurston, W.P.: Soliton–like Behavior in Automata. Physica D, **19**, 423–432, (1986)
Pul05. Pulvermüller, F.: Brain mechanicsms kinking language and action. Nature Rev. Neurosci. **6**, 576–582, (2005)
Put93. Puta, M.: Hamiltonian Mechanical Systems and Geometric Quantization, Kluwer, Dordrecht, (1993)
PV03. Pessa, E., Vitiello, G.: Quantum noise, entanglement and chaos in the quantum field theory of mind/brain states, Mind and Matter, **1**, 59–79, (2003)
PV04. Pessa, E., Vitiello, G.: Quantum noise induced entanglement and chaos in the dissipative quantum model of brain, Int. J. Mod. Phys. B, **18** 841–858, (2004)
PV05. Potkonjak, V., Vukobratovic, M.: A generalized approach to modelling dynamics of human and humanoid motion, Int. J. Adv. Robot. Sys., **2**(1), pp. 21-45, (2005)
PV99. Pessa, E., Vitiello, G.: Quantum dissipation and neural net dynamics, Biolectrochem. Bioener., **48**, 339–342, (1999)
Ram90. Ramond, P.: Field Theory: a Modern Primer. Addison–Wesley, Reading, MA, (1990)
Rat78. Ratcliff, R.: A theory of memory retrieval. Psych. Rev., **85**, 59–108, (1978)
RBT01. Roe, R.M., Busemeyer, J.R., Townsend, J.T.: Multi-alternative decision field theory: A dynamic connectionist model of decision making. Psych. Rev., **108**, 370–392, (2001)
REB03. Rani, R., Edgar, S.B., Barnes, A.: Killing Tensors and Conformal Killing Tensors from Conformal Killing Vectors. Class.Quant.Grav., **20**, 1929–1942, (2003)
Reg61. Regge, T.: General relativity without coordinates. Nuovo Cim. A **19**, 558–571, (1961)
Rei48. Reidemeister, K.: Knotentheorie, Chelsea Pub., New York, (1948)
RG98. Rao, A.S., Georgeff, M.P.: Decision Procedures for BDI Logics. Journal of Logic and Computation, **8**(3), 292-343, (1998)
RH89. Rose, R.M., Hindmarsh, J.L.: The assembly of ionic currents in a thalamic neuron. I The three-dimensional model. Proc. R. Soc. Lond. B, **237**, 267–288, (1989)
Rha84. de Rham, G.: Differentiable Manifolds. Springer, Berlin, (1984)
Ric93. Ricca, R.L.: Torus knots and polynomial invariants for a class of soliton equations, Chaos, **3**(1), 83–91, (1993)
RN03. Russell, Norvig: Artificial Intelligence. A modern Approach. Prentice Hall, (2003)
Rov98. Rovelli, C.: Loop quantum gravity. Living Rev. Rel. **1**, (1998)
RS75. Reed, M., Simon, B.: Methods of modern mathematical physics, Vol. 2: Fourier analysis, self-adjointness. Academic Press, San Diego, (1975)
RT02. Rosa-Clot, M., Taddei, S.: A Path Integral Approach to Derivative Security Pricing. Int. J. Theor. Appl. Finance, **5**(2), 123–146, (2002)

RT71. Ruelle, D., Takens, F.: On the nature of turbulence. Comm. Math. Phys., **20**, 167–192, (1971)

RT90. Reshetikhin, N., Turaev, V.: Ribbon graphs and their invariants derived from quantum groups. Comm. Math. Phys. **127**, 1–26, (1990)

RU67. Ricciardi, L.M., Umezawa, H.: Brain physics and many-body problems, Kibernetik, **4**, 44, (1967)

Rue78. Ruelle, D.: Thermodynamic formalism. Encyclopaedia of Mathematics and its Applications. Addison–Wesley, Reading, MA, (1978)

Ryd96. Ryder, L.: Quantum Field Theory. Cambridge Univ. Press, (1996)

SA98. Schaal, S., Atkeson, C.G.: Constructive incremental learning from only local information. Neural Comput., **10**, 2047–2084, (1998)

Sar02a. Sardanashvily, G.: Ten lectures on jet manifolds in classical and quantum field theory. arXiv: math-ph/0203040.

Sar02b. Sardanashvily, G.: The Lyapunov stability of first order dynamic equations with respect to time-dependent Riemannian metrics. arXiv nlin.CD/0201060.

Sar02c. Sardanashvily, G.: The bracket and the evolution operator in covariant Hamiltonian field theory, arXiv: math-ph/0209001.

Sar03. Sardanashvily, G.: Geometric quantization of relativistic Hamiltonian mechanics. Int. J. Theor. Phys. **42**, 697-704, (2003)

Sar92. Sardanashvily, G., Zakharov, O.: Gauge Gravitation Theory. World Scientific, Singapore, (1992)

Sar92. Sardanashvily, G.: On the geometry of spontaneous symmetry breaking. J. Math. Phys. **33**, 1546, (1992)

Sar93. Sardanashvily, G.: Gauge Theory in Jet Manifolds. Hadronic Press, Palm Harbor, FL, (1993)

Sar94. Sardanashvily, G.: Energy-momentum conservation law in Hamiltonian field theory, arXiv: gr-qc/9412041.

Sar94. Sardanashvily, G.: Constraint field systems in multimomentum canonical variables. J. Math. Phys. **35**, 6584, (1994)

Sar95. Sardanashvily, G.: Generalized Hamiltonian Formalism for Field Theory. Constraint Systems., World Scientific, Singapore, (1995)

Sar97. Sardanashvily, G.: Stress-energy–momentum tensors in constraint field theories. J. Math. Phys. **38**, 847, (1997)

Sar98. Sardanashvily, G.: Hamiltonian time-dependent mechanics. J. Math. Phys. **39**, 2714, (1998)

Sar98. Sardanashvily, G.: Stress-Energy-Momentum Tensors in Constraint Field Theories. J. Math. Phys. **39**, 2714, (1998)

Sau89. Saunders, D.J.: The Geometry of Jet Bundles. Lond. Math. Soc. Lect. Notes Ser. **142**, Cambr. Univ. Pr., (1989)

SB98. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA, (1998)

SC01. Segel, L.A., Cohen, I.R.(eds.): Design principles for the immune system and other distributed autonomous systems, Oxford Univ. Press, (2001)
Sch81. Schulman, L.S.: Techniques and Applications of Path Integration, John Wiley & Sons, New York, (1981)
Sch93. Schwarz, M.: Morse Homology, Birkháuser, Basel, (1993)
Sch96. Schellekens, A.N.: Introduction to conformal field theory. Fortsch. Phys., **44**, 605, (1996)
Sch98. Schaal, S.: Robot learning. In M. Arbib (ed). Handbook of Brain Theory and Neural Networks (2nd ed.), MIT Press, Cambridge, (1998)
Sch99. Schaal, S.: Is imitation learning the route to humanoid robots?. Trends Cogn. Sci., **3**, 233–242, (1999)
Sei95. Seiler, W.M.: Involution and constrained dynamics II: The Faddeev-Jackiw approach, J. Phys. A, **28**, 7315–7331, (1995)
Sem02. Semmelmann, U.: Conformal Killing forms on Riemannian manifolds. arXiv math.DG/0206117, (2002)
SFM98. Skiffington, S., Fernandez, E., McFarland, K.: Towards a validation of multiple features in the assessment of emotions. Eur. J. Psych. Asses. **14**(3), (1998)
SGL93. Shih, C.L., Gruver, W. Lee, T.: Inverse kinematics and inverse dynamics for control of a biped walking machine. J. Robot. Syst., **10**, 531–555, (1993)
She05. Shen, Z.: Riemann–Finsler geometry with applications to information geometry, unpublished material, (2005)
SHT02. Shimizu-Sato, S., Huq, E., Tepperman, J.M., Quail, P.H.: A light-switchable gene promoter system. Nature, Biotech., **20**, 1041-1044, (2002)
Shu93. Shuster, M.D.: A Survey of Attitude Representations, J. Astr. Sci., **41**(4), 316–321, (1993)
SI89. Sastry, S.S., Isidori, A.: Adaptive control of linearizable systems. IEEE Trans. Aut. Con., **34**(11), 1123–1131, (1989)
Siv77. Sivashinsky, G.I.: Nonlinear analysis of hydrodynamical instability in laminar flames – I. Derivation of basic equations. Acta Astr. **4**, 1177, (1977)
SK93. Shih, C.L., Klein, C.A.: An adaptive gait for legged walking machines over rough terrain. IEEE Trans. Syst. Man, Cyber. A, **23**, 1150–1154, (1993)
SK98a. Shabanov, S.V., Klauder, J.R.: Path Integral Quantization and Riemannian-Symplectic Manifolds. Phys. Lett. B4 **35**, 343–349, (1998)
SK98b. Shabanov, S.V., Klauder, J.R.: Path Integral Quantization and Riemannian-Symplectic Manifolds. Phys. Lett. B **435**, 343-349, (1998)
Sma60. Smale, S.: The generalized Poincaré conjecture in higher dimensions, Bull. Amer. Math. Soc., **66**, 373–375, (1960)
Sma67. Smale, S.: Differentiable dynamical systems, Bull. Amer. Math. Soc., **73**, 747–817, (1967)
Sma99. van der Smagt, P.: (ed.) Self–Learning Robots. Workshop: Brainstyle Robotics, IEE, London, (1999)

Sni80. Sniatycki, J.: Geometric Quantization and Quantum Mechanics. Springer-Verlag, Berlin, (1980)

Soc91. Socolovsky, M.: Gauge transformations in fibre bundle theory. J. Math. Phys., **32**, 2522, (1991)

Spa82. Sparrow, C.: The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors. Springer, New York, (1982)

Sta00. Stanislavsky, A.A.: Memory effects and macroscopic manifestation of randomness. Phys. Rev. E **61**, 4752, (2000)

Sta63. Stasheff, J.D.: Homotopy associativity of $H-$spaces I & II. Trans. Amer. Math. Soc., **108**, 275–292, 293–312, (1963)

Ste69. Sternberg, S.: Memory-scanning: Mental processes revealed by reaction-time experiments. Am. Sci., **57**(4), 421–457, (1969)

Ste72. Steenrod, N.: The Topology of Fibre Bundles, Princeton Univ. Press, Princeton, (1972)

Ste93. Stengel, R.: Optimal control and estimation. Dover, New York, (1993)

Sto68. Stong, R.E.: Notes on Cobordism Theory. Princeton Univ. Press, Princeton, (1968)

STP95. Schaub, H., Tsiotras, P., Junkins, J.: Principal Rotation Representations of Proper N x N Orthogonal Matrices. Int. J. Eng. Sci., **33**,(15), 2277–2295, (1995)

Str00. Strogatz, S.: From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. Physica D, **143**, 1–20, (2000)

Str90. Strominger, A.: Special Geometry. Commun. Math. Phys. **133**, 163, (1990)

STU78. Stuart, C.I.J., Takahashi, Y., Umezawa, H.: On the stability and non-local properties of memory, J. Theor. Biol. **71**, 605–618, (1978)

STU79. Stuart, C.I.J., Takahashi, Y., Umezawa, H.: Mixed system brain dynamics: neural memory as a macroscopic ordered state, Found. Phys. **9**, 301, (1979)

Stu99. Stuart, J.: Calculus (4th ed.). Brooks/Cole Publ., Pacific Grove, CA, (1999)

STZ93. Satarić, M.V., Tuszynski, J.A., Zakula, R.B.: Kinklike excitations as an energy-transfer mechanism in microtubules. Phys. Rev. E, **48**, 589–597, (1993)

Sus83. Sussmann, H.J.: Lie brackets and local controllability: a sufficient condition for scalar–input systems, SIAM J. Con. Opt., **21**(5), 686–713, (1983)

Sus87. Sussmann, H.J.: A general theorem on local controllability, SIAM J. Con. Opt., **25**(1), 158–194, (1987)

SV96. Strominger, A., Vafa, C.: On the microscopic origin of the Bekenstein–Hawking entropy. Phys. Lett. B3 **79**, 99, (1996)

SVW95. Srivastava, Y.N., Vitiello, G., Widom, A.: Quantum dissipation and quantum noise, Annals Phys., **238**, 200, (1995)

SW72. Sulanke R., Wintgen P.: Differential geometry und faserbundel; bound 75, Veb. Deutscher Verlag der Wissenschaften, Berlin, (1972)

Swi75. Switzer, R.K.: Algebraic Topology – Homology and Homotopy. (in Classics in Mathematics), Springer, New York, (1975)

SWV99. Srivastava, Y.N., Widom, A., Vitiello, G.: Quantum measurements, information and entropy production. Int. J. Mod. Phys. B13, 3369–3382, (1999)

SZT98. Satarić, M.V., Zeković, S., Tuszynski, J.A., Pokorny, J.: Mössbauer effect as a possible tool in detecting nonlinear excitations in microtubules. Phys. Rev. E **58**, 6333–6339, (1998)

Tei83. Teitelboim, C.: Causality versus gauge invariance in quantum gravity and supergravity. Phys. Rev. Lett. **50**, 705–708, (1983)

TF91. Tsue, Y., Fujiwara, Y.: Time-Dependent Variational Approach to (1+1)-Dimensional Scalar-Field Solitons, Progr. Theor. Phys. **86**(2), 469–489, (1991)

Thi79. Thirring, W.: A Course in Mathematical Physics (in four volumes). Springer, New York, (1979)

Tho79. Thorpe, J.A.: Elementary Topics in Differential Geometry. Springer, New York, (1979)

Tho93. Thompson, G.: Non-uniqueness of metrics compatible with a symmetric connection. Class. Quant. Grav. **10**, 2035, (1993)

TP01. Tabuada, P., Pappas, G.J.: Abstractions of Hamiltonian Control Systems. Proceedings of the 40th IEEE Conf. Decis. Con., Orlando, FL, (2001)

TPS98. Tomlin, C., Pappas, G.J., Sastry, S.: Conflict resolution for air traffic management: A case study in multi-agent hybrid systems, IEEE Trans. Aut. Con., **43**, 509-521, (1998)

TVP99. Tabony, J., Vuillard, L., Papaseit, C.: Biological self-organisation and pattern formation by way of microtubule reaction-diffusion processes. Adv. Complex Syst. **2**(3), 221–276, (1999)

TW95. Tucker, R., Wang, C.: Black holes with Weyl charge and non-Riemannian waves. Class. Quant. Grav. **12**, 2587, (1995)

Udr00. Udriste, C.: Geometric dynamics. Kluwer Academic Publishers, (2000)

UN99. Udriste, C., Neagu, M.: Extrema of p-energy functional on a Finsler manifold. Diff. Geom. Dyn. Sys. 1(1), 10–19, (1999)

VAB04. Vukobratovic, M., Andric, D., Borovac, B.: How to achieve various gait patterns from single nominal. Int. J. Adv. Rob. Sys., **1**(2), 99–108, (2004)

Vai94. Vaisman, I.: Lectures on the Geometry of Poisson Manifolds. Birkhäuser Verlag, Basel, (1994)

VB04. Vukobratovic, M., Borovac, B.: Zero-Moment Point – Thirty Five Years of its Life. Int. J. Hum. Rob., **1**(1), 157–173, (2004)

VBB05. Vukobratovic M., Borovac B., Babkovic K.: Contribution to the Study of Anthropomorphism of Humanoid Robots. Int. J. Adv. Rob. Sys. (to appear)

VBS89. Vukobratovic, M., Borovac, B., Surla, D., Stokic, D.: Scientific Fundamentals of Robotics, Vol. 7, Biped Locomotion: Dynamics, Stability, Control, and Applications. Springer-Verlag, Berlin, (1989)

VFJ70. Vukobratovic, M., Juricic, D., Frank, A.: On the stability of biped locomotion. IEEE Trans. Biom. Eng., **17**(1), 25–36, (1970)

VHI87. Vukobratovic, M., Hristic, D., Ivancevic, V.: DC–muscular hybrid joint control system. Proc. IX Int. Symp. External Control of Human Extremities, Dubrovnik, (1987)

Vit01. Vitiello, G.: My Double Unveiled. John Benjamins, Amsterdam, (2001)

Vit95. Vitiello, G.: Dissipation and memory capacity in the quantum brain model, Int. J. Mod. Phys. B, **9**, 973–989, (1995)

VJ69. Vukobratovic, M., Juricic, D.: Contribution to the synthesis of biped gait. IEEE Trans. Biom. Eng., **16**(1), (1969)

VJF70. Vukobratovic, M., Juricic, D., Frank, A.: On the control and stability of one class of biped locomotion systems. ASME J. Bas. Eng., **3**, 328–332, (1970)

VK85a. Vukobratovic, M., Kircanski, M.: Scientific Fundamentals of Robotics, Vol. 3, Kinematics and Trajectories Synthesis of Manipulation Robots (in English), Springer-Verlag, (1985); also published in Chinese, Beijing, (1992)

VK85b. Vukobratovic, M., Kircanski, N.: Scientific Fundamentals of Robotics, Vol. 4, Real-Time Dynamics of Manipulation Robots (in English), Springer-Verlag, (1985); also published in Chinese, Beijing, (1992)

Voi02. Voisin, C.: Hodge Theory and Complex Algebraic Geometry I. Cambridge Univ. Press, Cambridge, (2002)

Von05. Vonk, M.: A mini-course on topological strings. arXiv: hep-th/0504147.

VP82. Vukobratovic, M., Potkonjak, V.: Scientific Fundamentals of Robotics, Vol. 1, Dynamics of Manipulation Robots: Theory and Application (in English), Springer-Verlag, (1982); also published in Japanese, extended version, Japanese Springer-Verlag, (1986); and in Chinese, Beijing, (1991)

VP85. Vukobratovic, M., Potkonjak, V.: Scientific Fundamentals of Robotics, Vol. 6, Applied Dynamics and CAD of Manipulation Robots, Springer-Verlag, (1985)

VPM03. Vukobratovic, M., Potkonjak, V., Matijevic, V.: Dynamics of Robots in Contact Tasks, Book series: Microprocessor-Based and Intelligent Systems Engineering, Kluwer Acad. Pub., Dordrecht, (2003)

VPR04. Vukobratovic, M., Potkonjak, V., Rodic, A.: Contribution to the Dynamic Study of Humanoid Robots Interacting with Dynamic Environment. Int. J. Rob., **22**, 439–447, (2004)

VPT04. Vukobratovic, M., Potkonjak, V., Tzafestas, S.: Human and Humanoid Dynamics – From the Past to the Future, J. Int. Rob. Sys., **41**, 65–84, (2004)

VR05. Vukobratovic, M., Rodic, A.: Contribution to the Integrated Control of Biped Locomotion Mechanisms Interacting With Dynamic Environment, Int. J. Adv. Robot. Sys., (in press)

VS72. Vukobratovic, M., Stepanenko, Y.: On the stability of anthropomorphic systems. Math. Biosci. **15**, 1–37, (1972)

VS73. Vukobratovic, M., Stepanenko, Y.: Mathematical models of general anthropomorphic systems. Math. Biosci. **17**, 191–242, (1973)

VS82. Vukobratovic, M., Stokic, D.: Scientific Fundamentals of Robotics, Vol. 2, Control of Manipulation Robots: Theory and Application (in English), Springer-Verlag, (1982); also published in Russian, "Nauka", Moskow, (1985); and in Chinese, Beijing, (1991)

VSK85. Vukobratovic, M., Stokic, D., Kircanski, N.: Scientific Fundamentals of Robotics, Vol. 5, Non-Adaptive and Adaptive Control of Manipulation Robots (in English), Springer-Verlag, (1985); also published in Russian, "Mir", Moskow, (1988); and in Chinese, Beijing, (1993)

Vuk75. Vukobratovic, M.: Legged Locomotion Robots and Anthropomorphic Mechanisms (in English). Mihailo Pupin, Belgrade, (1975); also published in Japanese, Nikkan Shimbun Ltd. Tokyo, (1975); in Russian, "MIR", Moscow, (1976); in Chinese, Beijing, (1983)

Vuk78. Vukobratovic, M.: Dynamics of Robots (in Japanese). Nikkan Shimbun Ltd, Tokyo, (1978)

Vyg82. Vygotsky, L.S. Historical meaning of the Psychological crisis. Collected works. Vol. 1. Pedag. Publ., Moscow, (1982)

WA00. Wall, M.M., Amemiya, Y.: Estimation for polynomial structural equation models. Journal of American Statistical Association, **95**, 929–940, (2000)

WA98. Wall, M.M., Amemiya, Y.: Fitting nonlinear structural equation models. Proc. Social Stat. Section. Ann. Meet. Ame. Stat. Assoc. 180–185, (1998)

WDH93. Wilmott, P., Dewynne, J., Howinson, S.: Option Pricing: Mathematical Models and Computation. Oxford Financial Press, (1993)

Wei90. Weinstein, A.: Affine Poisson structures. Internat. J. Math., **1**, 343–360, (1990)

WF49. Wheeler, J.A., Feynman, R.P.: Classical Electrodynamics in Terms of Direct Interparticle Action. Rev. Mod. Phys. **21**, 425–433, (1949)

Whe61. Wheeler, J.A.: Geometrodynamics and the Problem of Motion. Rev. Mod. Phys. **33**, 63–78, (1961)

Whe62. Wheeler, J.A.: Geometrodynamics. Academic Press, New York, (1962)

Whi87. Whitney, D.E.: Historical perspective and state of the art in robot force control. Int. J. Robot. Res., **6**(1), 3–14, (1987)

Wie61. Wiener, N.: Cybernetics. Wiley, New York, (1961)

Wig90. Wiggins, S.: Introduction to Applied Dynamical Systems and Chaos. Springer, New York, (1990)

Wik05. Wikipedia, the free encyclopedia. (2005) http://wikipedia.org.

Wil00. Wilson, D.: Nonlinear Control, Advanced Control Course (Student Version), Karlstad Univ., (2000)

Wil93. Willmore, T.J.: Riemannian Geometry. Oxford Univ. Press, Oxford, (1993)

Wil97. Williams, R.M.: Recent progress in Regge calculus. Nucl. Phys. B **57**, 73–81, (1997)

Wit82. Witten, E.: Supersymmetry and Morse theory. J. Diff. Geom., **17**, 661–692, (1982)

Wit88a. Witten, E.: Topological Sigma Models. Commun. Math. Phys. **118**, 411, (1988)

Wit88b. Witten, E.: Topological Quantum Field Theory. Commun. Math. Phys. **117**, 353, (1988)

Wit88c. Witten, E.: Topological Gravity. Phys. Lett. B **206**, 601, (1988)

Wit88. Witten, E: Topological quantum field theory, Commun. Math. Phys., **117**, 353, (1988)

Wit89. Witten, E.: Quantum Field Theory And The Jones Polynomial. Commun. Math. Phys. **121**, 351, (1989)

Wit90. Witten, E.: On the structure of the topological phase of two–dimensional gravity. Nucl. Phys. B **340**, 281, (1990)

Wit91. Witten, E.: Introduction To Cohomological Field Theories. Int. J. Mod. Phys. A **6**, 2775, (1991)

Wit92. Witten, E.: Mirror manifolds and topological field theory. In Essays on mirror manifolds, ed. S.-T. Yau, International Press, 120–158, (1992)

Wit95. Witten, E.: Chern–Simons gauge theory as a string theory. Prog. Math. **133**, 637, (1995)

Wit95. Witten, E: String Theory Dynamics in Various Dimensions. Nucl. Phys. B **443**, 85, (1995)

Wit98. Witten, E.: Magic, mystery, and matrix. Notices AMS, **45**(9), 1124–1129, (1998)

Wit98. Witten, E: Anti-de Sitter space and holography. Adv. Theor. Math. Phys., **2**, 253, (1998)

Woo92. Woodhouse, N.: Geometric Quantization. Clarendon Press, Oxford, (1992)

WS98. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature, **393**, 440-442, (1998)

WW83a. Wehner, M.F., Wolfer, W.G.: Numerical evaluation of path–integral solutions to Fokker–Planck equations. I., Phys. Rev. A **27**, 2663–2670, (1983)

WW83b. Wehner, M.F., Wolfer, W.G.: Numerical evaluation of path–integral solutions to Fokker–Planck equations. II. Restricted stochastic processes, Phys. Rev. A, **28**, 3003–3011, (1983)

XH94. Xu, Z., Hauser, J.: Higher order approximate feedback linearization about a manifold. J. Math. Sys. Est. Con., **4**, 451–465, (1994)

XH95. Xu, Z., Hauser, J.: Higher order approximate feedback linearization about a manifold for multi-input systems. IEEE Trans. Aut. Con, AC-40, 833–840, (1995)

YA01. Yalcin, I., Amemiya, Y.: Nonlinear factor analysis as a statistical method. Statistical Science, **16**, 275–294, (2001)

Yag87. Yager, R.R.: Fuzzy Sets and Applications: Selected Papers by L.A. Zadeh, Wiley, New York, (1987)

Yan52. Yano, K.: Some remarks on tensor fields and curvature. Ann. of Math. **55**(2), 328–347, (1952)

Yau78. Yau, S.-T.: On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation, I. Comm. Pure Appl. Math. **31**, 339–411, (1978)

YL52. Yang, C.N., Lee, T.D.: Statistical theory of equation of state and phase transitions I: Theory of condensation. Phys. Rev. **87**, 404–409, (1952)

YZ99. Yong, J., Zhou, X.: Stochastic controls. Hamiltonian Systems and HJB Equations. Springer, New York, (1999)

Zak92. Zakharov, O.: Hamiltonian formalism for nonregular Lagrangian theories in fibred manifolds. J. Math. Phys. **33**, 607, (1992)

Index

International Series on

MICROPROCESSOR-BASED AND INTELLIGENT SYSTEMS ENGINEERING

Editor: Professor S. G. Tzafestas, *National Technical University, Athens, Greece*

1. S.G. Tzafestas (ed.): *Microprocessors in Signal Processing, Measurement and Control*. 1983 ISBN 90-277-1497-5
2. G. Conte and D. Del Corso (eds.): *Multi-Microprocessor Systems for Real-Time Applications*. 1985 ISBN 90-277-2054-1
3. C.J. Georgopoulos: *Interface Fundamentals in Microprocessor-Controlled Systems*. 1985 ISBN 90-277-2127-0
4. N.K. Sinha (ed.): *Microprocessor-Based Control Systems*. 1986 ISBN 90-277-2287-0
5. S.G. Tzafestas and J.K. Pal (eds.): *Real Time Microcomputer Control of Industrial Processes*. 1990 ISBN 0-7923-0779-8
6. S.G. Tzafestas (ed.): *Microprocessors in Robotic and Manufacturing Systems*. 1991 ISBN 0-7923-0780-1
7. N.K. Sinha and G.P. Rao (eds.): *Identification of Continuous-Time Systems*. Methodology and Computer Implementation. 1991 ISBN 0-7923-1336-4
8. G.A. Perdikaris: *Computer Controlled Systems*. Theory and Applications. 1991 ISBN 0-7923-1422-0
9. S.G. Tzafestas (ed.): *Engineering Systems with Intelligence*. Concepts, Tools and Applications. 1991 ISBN 0-7923-1500-6
10. S.G. Tzafestas (ed.): *Robotic Systems*. Advanced Techniques and Applications. 1992 ISBN 0-7923-1749-1
11. S.G. Tzafestas and A.N. Venetsanopoulos (eds.): *Fuzzy Reasoning in Information, Decision and Control Systems*. 1994 ISBN 0-7923-2643-1
12. A.D. Pouliezos and G.S. Stavrakakis: *Real Time Fault Monitoring of Industrial Processes*. 1994 ISBN 0-7923-2737-3
13. S.H. Kim: *Learning and Coordination*. Enhancing Agent Performance through Distributed Decision Making. 1994 ISBN 0-7923-3046-3
14. S.G. Tzafestas and H.B. Verbruggen (eds.): *Artificial Intelligence in Industrial Decision Making, Control and Automation*. 1995 ISBN 0-7923-3320-9
15. Y.-H. Song, A. Johns and R. Aggarwal: *Computational Intelligence Applications to Power Systems*. 1996 ISBN 0-7923-4075-2
16. S.G. Tzafestas (ed.): *Methods and Applications of Intelligent Control*. 1997 ISBN 0-7923-4624-6
17. L.I. Slutski: *Remote Manipulation Systems*. Quality Evaluation and Improvement. 1998 ISBN 0-7932-4822-2
18. S.G. Tzafestas (ed.): *Advances in Intelligent Autonomous Systems.* 1999 ISBN 0-7932-5580-6
19. M. Teshnehlab and K. Watanabe: *Intelligent Control Based on Flexible Neural Networks*. 1999 ISBN 0-7923-5683-7
20. Y.-H. Song (ed.): *Modern Optimisation Techniques in Power Systems*. 1999 ISBN 0-7923-5697-7
21. S.G. Tzafestas (ed.): *Advances in Intelligent Systems*. Concepts, Tools and Applications. 2000 ISBN 0-7923-5966-6

International Series on
MICROPROCESSOR-BASED AND INTELLIGENT SYSTEMS ENGINEERING

22. S.G. Tzafestas (ed.): *Computational Intelligence in Systems and Control Design and Applications*. 2000 ISBN 0-7923-5993-3
23. J. Harris: *An Introduction to Fuzzy Logic Applications*. 2000 ISBN 0-7923-6325-6
24. J.A. Fernández and J. González: *Multi-Hierarchical Representation of Large-Scale Space*. 2001 ISBN 1-4020-0105-3
25. D. Katic and M. Vukobratovic: *Intelligent Control of Robotic Systems.* 2003 ISBN 1-4020-1630-1
26. M. Vukobratovic, V. Potkonjak and V. Matijevic: *Dynamics of Robots with Contact Tasks.* 2003 ISBN 1-4020-1809-6
27. M. Ceccarelli: *Fundamentals of Mechanics of Robotic Manipulation.* 2004 ISBN 1-4020-1810-X
28. V.G. Ivancevic and T.T. Ivancevic: *Human-Like Biomechanics.* A Unified Mathematical Approach to Human Biomechanics and Humanoid Robotics. 2005 ISBN 1-4020-4116-0
29. J. Harris: *Fuzzy Logic Applications in Engineering Science.* 2005 ISBN 1-4020-4077-6
30. M.D. Zivanovic and M.K. Vukobratovic: *Multi-Arm Cooperating Robots*. Dynamics and Control. 2006 ISBN 1-4020-4268-X
31. V.G. Ivancevic and T. Ivancevic: *Geometrical Dynamics of Complex Systems*. A Unified Modelling Approach to Physics, Control, Biomechanics, Neurodynamics and Psycho-Socio-Economical Dynamics. 2006 ISBN 1-4020-4544-1