GRAVITATION AND GAUGE SYMMETRIES

Series in High Energy Physics, Cosmology and Gravitation

Other books in the series

Electron–Positron Physics at the Z
M G Green, S L Lloyd, P N Ratoff and D R Ward

Non-accelerator Particle Physics
Paperback edition
H V Klapdor-Kleingrothaus and A Staudt

Ideas and Methods of Supersymmetry and Supergravity
***or* A Walk Through Superspace**
Revised edition
I L Buchbinder and S M Kuzenko

Pulsars as Astrophysical Laboratories for Nuclear and Particle Physics
F Weber

Classical and Quantum Black Holes
Edited by P Fré, V Gorini, G Magli and U Moschella

Particle Astrophysics
Revised paperback edition
H V Klapdor-Kleingrothaus and K Zuber

The World in Eleven Dimensions
Supergravity, Supermembranes and M-Theory
Edited by M J Duff

Gravitational Waves
Edited by I Ciufolini, V Gorini, U Moschella and P Fré

SERIES IN HIGH ENERGY PHYSICS,
COSMOLOGY AND GRAVITATION

Series Editors
B Foster, L Grishchuk, E W Kolb, M A H MacCallum,
D H Perkins and B F Schutz

GRAVITATION AND GAUGE SYMMETRIES

Milutin Blagojević

*Institute of Physics,
Belgrade, Yugoslavia*

IoP

INSTITUTE OF PHYSICS PUBLISHING
BRISTOL AND PHILADELPHIA

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

ISBN 0 7503 0767 6

Library of Congress Cataloging-in-Publication Data are available

Cover image: A burst of outgoing gravitational radiation originating from a grazing merger of two black holes. (Courtesy Konrad-Zuse-Zentrum fuer Informationstechnik Berlin (ZIB) and Max-Planck-Institut fuer Gravitationsphysik (Albert-Einstein-Institut, AEI). Simulation: AEI Numerical Relativity Group in cooperation with WashU and NCSA (Ed Seidel, Wai-Mo Suen *et al*). Visualization: Werner Benger.)

Commissioning Editor: James Revill
Production Editor: Simon Laurenson
Production Control: Sarah Plenty
Cover Design: Victoria Le Billon
Marketing Executive: Laura Serratrice

Published by Institute of Physics Publishing, wholly owned by The Institute of Physics, London

Institute of Physics Publishing, Dirac House, Temple Back, Bristol BS1 6BE, UK

US Office: Institute of Physics Publishing, The Public Ledger Building, Suite 1035, 150 South Independence Mall West, Philadelphia, PA 19106, USA

Typeset in LaTeX 2_ε by Text 2 Text, Torquay, Devon
Printed in the UK by MPG Books Ltd, Bodmin, Cornwall

Contents

Preface

The concept of a unified description of the basic physical interactions has evolved in parallel with the development of our understanding of their dynamical structure. It has its origins in Maxwell's unification of electricity and magnetism in the second half of the nineteenth century, matured in Weyl's and Kaluza's attempts to unify gravity and electromagnetism at the beginning of the last century, and achieved its full potential in the 1970s, in the process of unifying the weak and electromagnetic and also, to some extent, the strong interactions. The biggest barrier to this attractive idea comes from the continual resistance of gravity to join the other basic interactions in the framework of a unified, consistent quantum theory.

As the theory of electromagnetic, weak and strong interactions developed, the concept of (internal) gauge invariance came of age and established itself as an unavoidable dynamical principle in particle physics. It is less well known that the principle of equivalence, one of the prominent characteristics of the gravitational interaction, can also be expressed as a (spacetime) gauge symmetry. This book is intended to shed light upon the connection between the intrinsic structure of *gravity* and the principle of *gauge invariance*, which may lead to a consistent *unified* theory.

The first part of this book, chapters 1–6, gives a systematic account of the structure of gravity as a theory based on spacetime gauge symmetries. Some basic properties of space, time and gravity are reviewed in the first, introductory chapter. Chapter 2 deals with elements of the global Poincaré and conformal symmetries, which are necessary for the exposition of their localization; the structure of the corresponding gauge theories is explored in chapters 3 and 4. Then, in chapters 5 and 6, we present the basic features of the Hamiltonian dynamics of Poincaré gauge theory, discuss the relation between gauge symmetries and conservation laws and introduce the concept of gravitational energy and other conserved charges. The second part of the book treats the most promising attempts to build a unified field theory containing gravity, on the basis of the gauge principle. In chapters 7 and 8 we discuss the possibility of constructing gravity as a field theory in flat spacetime. Chapters 9–11 yield an exposition of the ideas of supersymmetry and supergravity, Kaluza–Klein theory and string theory—these ideas can hardly be avoided in any attempt to build a unified theory of basic physical interactions.

This book is intended to provide a pedagogical survey of the subject of gravity from the point of view of particle physics and gauge theories at the graduate level. The book is written as a self-contained treatise, which means that I assume no prior knowledge of gravity and gauge theories on the part of the reader. Of course, some familiarity with these subjects will certainly facilitate the reader to follow the exposition. Although the gauge approach differs from the more standard geometric approach, it leads to the same mathematical and physical structures.

The first part of the book (chapters 2–6) has evolved from the material covered in the one-semester graduate course *Gravitation II*, taught for about 20 years at the University of Belgrade. Chapters 9–11 have been used as the basis for a one-semester graduate course on the unification of fundamental interactions.

Special features

The following remarks are intended to help the reader in an efficient use of the book.

- Examples in the text are used to illustrate and clarify the main exposition.
- The exercises given at the end of each chapter are an integral part of the book. They are aimed at illustrating, completing, applying and extending the results discussed in the text.
- Short comments on some specific topics are given at the end of each chapter, in order to illustrate the relevant research problems and methods of investigation.
- The appendix consists of 13 separate sections (A–M), which have different relationships with the main text.
  - Technical appendices J and M (Dirac spinors, Fourier expansion) are *indispensable* for the exposition in chapters 9 and 11.
  - Appendices A, H and I (internal local symmetries, Lorentz and Poincaré group) are *very useful* for the exposition in chapters 3 (A) and 9 (H, I).
  - Appendices C, D, E, F, G and L (de Sitter gauge theory, scalar–tensor theory, Ashtekar's formulation of general relativity, constraint algebra and gauge symmetries, covariance, spin and interaction of massless particles, and Chern–Simons gravity) are *supplements* to the main exposition, and may be studied according to the reader's choice.
  - The material in appendices B and K (differentiable manifolds, symmetry groups and manifolds) is *not necessary* for the main exposition. It gives a deeper mathematical foundation for the geometric considerations in chapters 3, 4 and 10.
- The bibliography contains references that document the material covered in the text. Several references for each chapter, which I consider as the most suitable for further reading, are denoted by the symbol •.

- Chapters 4, 7 and 8 can be omitted in the first reading, without influencing the internal coherence of the exposition. Chapters 9–11 are largely independent of each other, and can be read in any order.

Acknowledgments

The material presented in this book has been strongly influenced by the research activities of the Belgrade group on *Particles and Fields*, during the last 20 years. I would like to mention here some of my teachers and colleagues who, in one or another way, have significantly influenced my own understanding of gravity. They are: Rastko Stojanović and Marko Leko, my first teachers of tensor calculus and general relativity; Djordje Živanović, an inspiring mind for many of us who were fascinated by the hidden secrets of gravity; Paul Senjanović, who brought the Dirac canonical approach all the way from City College in New York to Belgrade; Ignjat Nikolić and Milovan Vasilić, first my students and afterwards respected collaborators; Branislav Sazdović, who silently introduced supersymmetry and superstrings into our everyday life; then, Dragan Popović, the collaborator from an early stage on our studies of gravity, and Djordje Šijački, who never stopped insisting on the importance of symmetries in gravity.

For illuminating comments and criticisms on a preliminary version of the manuscript, I wish to thank my colleagues Tatjana Vukašinac, Milovan Vasilić, Djordje Šijački, Branislav Sazdović, Nenad Manojlović, Aleksandar Bogojević, Friedrich Hehl and Eckehard Mielke. I am especially grateful for their valuable help to my students Olivera Mišković and Dejan Stojković, who carefully studied the first version of the manuscript and solved all the exercises with patience. I owe special thanks to Ignjat Nikolić, coauthor of the manuscript *The Theory of Gravity II*, written in 1986 for the graduate school of physics in Belgrade, for allowing me to use a large part of my contribution to that text, and include some of his formulations (in sections 5.2 and 8.4) and exercises. Finally, it is my pleasure to thank Friedrich Hehl and Yuri Obukhov for their sincere support during this project.

The exposition of Poincaré gauge theory and its Hamiltonian structure is essentially based on the research done in collaboration with Ignjat Nikolić and Milovan Vasilić. The present book incorporates a part of the spirit of that collaboration, which was born from our longlasting joint work on the problems of gravity.

The sections on teleparallel theory, Chern–Simons gauge theory and three-dimensional gravity have been written as an update to the first version of the manuscript (published in 1997, in Serbian). I would like to thank Antun Balaš and Olivera Mišković for their assistance at this stage, and to express my gratitude for the hospitality at the Primorska Institute for Natural Sciences and Technology, Koper, Slovenia, where part of the update was completed.

Milutin Blagojević
Belgrade, April 2001

Chapter 1

Space, time and gravitation

Theories of special and general relativity represent a great revolution in our understanding of the structure of space and time, as well as of their role in the formulation of physical laws. While special relativity (SR) describes the influence of physical reality on the general properties of and the relation between *space* and *time*, the geometry of spacetime in general relativity (GR) is connected to the nature of *gravitational* interaction. Perhaps the biggest barrier to a full understanding of these remarkable ideas lies in the fact that we are not always ready to suspect the properties of space and time that are built into our consciousness by everyday experience.

In this chapter we present an overview of some aspects of the structure of space, time and gravitation, which are important for our understanding of gravitation as a gauge theory. These aspects include:

- the development of the principle of relativity from classical mechanics and electrodynamics, and its influence on the structure of space and time; and
- the formulation of the principle of equivalence, and the introduction of gravitation and the corresponding geometry of curved spacetime.

The purpose of this exposition is to illuminate those properties of space, time and gravitation that have had an important role in the development of GR, and still have an influence on various attempts to build an alternative approach to gravity (Sciama 1969, Weinberg 1972, Rindler 1977, Hoffmann 1983).

1.1 Relativity of space and time

Historical introduction

In order to get a more complete picture of the influence of relativity theories on the development of the concepts of space and time, we recall here some of the earlier ideas on this subject.

In Ancient Greece, the movement of bodies was studied philosophically. Many of the relevant ideas can be found in the works of Aristotle (fourth century BC) and other Greek philosophers. As an illustration of their conception of the nature of movement, we display here the following two statements.

- The speed of a body in free fall depends on its weight; heavy bodies fall faster than lighter ones.
- The earth is placed at the fixed centre of the universe.

While the first statement was so obvious that practically everyone believed in it, different opinions existed about the second one. One of the earliest recorded proposals that the earth might move belongs to the Pythagorean Philolaus (fifth century BC). Two centuries later, the idea appeared again in a proposal of the Greek astronomer Aristarchus (third century BC). However, it was not persuasive for most of the ancient astronomers. From a number of arguments against the idea of a moving earth, we mention Aristotle's. He argued that if the earth were moving, then a stone thrown straight up from the point A would fall at another point B, since the original point A would 'run away' in the direction of earth's movement. However, since the stone falls back at the same point from which it is thrown, he concluded that the earth does not move.

For a long time, the developments of physics and astronomy have been closely connected. Despite Aristarchus, the ancient Greek astronomers continued to believe that the earth is placed at the fixed centre of the universe. This geocentric conception of the universe culminated in Ptolemy's work (second century AD). The Ptolemaic system endured for centuries without major advances. The birth of modern astronomy started in the 16th century with the work of Copernicus (1473–1543), who dared to propose that the universe is heliocentric, thus reviving Aristarchus' old idea. According to this proposal, it is not the earth but the sun that is fixed at the centre of everything, while the earth and other planets move around the sun. The Danish astronomer Brahe (1546–1601) had his own ideas concerning the motion of planets. With the belief that the clue for all answers lies in *measurements*, he dedicated his life to precise astronomical observations of the positions of celestial bodies. On the basis of these data Kepler (1571–1630) was able to deduce his well known laws of planetary motion. With these laws, it became clear *how* the planets move; a search for an answer to the question *why* the planets move led Newton (1643–1727) to the discovery of the law of gravitation.

A fundamental change in the approach to physical phenomena was made by the Italian scientist Galileo Galilei (1564–1642). Since he did not believe in Aristotle's 'proofs', he began a systematic analysis and *experimental verification* of the laws of motion. By careful measurement of spatial distances and time intervals during the motion of a body along an inclined plane, he found new relations between distances, time intervals and velocities, that were unknown in Aristotle's time. What were Galileo's answers to the questions of free fall and the motion of the earth?

- Studying the problem of free fall, Galileo discovered that all bodies fall with the same acceleration, no matter what their masses are nor what they are made of. This is the essence of the *principle of equivalence*, which was used later by Einstein (1879–1955) to develop GR.
- Trying to understand the motion of the earth, Galileo concluded that the uniform motion of the earth cannot be detected by means of any internal mechanical experiment (thereby overturning Aristotle's arguments about the immobile earth). The conclusion about the equivalence of different (inertial) reference frames moving with constant velocities relative to each other, known as the *principle of relativity*, has been of basic importance for the development of Newton's mechanics and Einstein's SR.
- Let us mention one more discovery by Galileo. The velocity of a body moving along an inclined plane changes in time. The cause of the change is the gravitational attraction of that body and the earth. When the attraction is absent and there is no force acting on the body, its velocity remains constant. This is the well known *law of inertia* of classical mechanics.

The experiments performed by Galileo may be considered to be the origin of modern physics. His methods of research and the results obtained show, by their simplicity and their influence on future developments in physics, all the beauty and power of the scientific truth. He studied the motion of bodies by asking *where* and *when* something happens. Since then, measurements of space and time have been an intrinsic part of physics.

The concepts of time and space are used in physics only with reference to physical objects. What are these entities by themselves? 'What is time—if nobody asks me, I know, but if I want to explain it to someone, then I do not know' (St Augustine; a citation from J R Lucas (1973)). Time and space are connected with change and things that change. The challenge of physics is not to define space and time precisely, but to *measure* them precisely.

Relativity of motion and the speed of light

In Galileo's experiments we find embryos of the important ideas concerning space and time, ideas which were fully developed later in the works of Newton and Einstein.

Newton's classical mechanics, in which Galileo's results have found a natural place, is based on the following three laws:

1. A particle moves with constant velocity if no force acts on it.
2. The acceleration of a particle is proportional to the force acting on it.
3. The forces of action and reaction are equal and opposite.

Two remarks will clarify the content of these laws. First, the force appearing in the second law originates from interactions with other bodies, and should be known from independent considerations (e.g. Newton's law of gravitation). Only then can the second law be used to determine the acceleration stemming from a given force. Second, physical quantities like velocity, acceleration, etc, are defined always and only *relative* to some reference frame.

Galilean principle of relativity. The laws of Newtonian mechanics do not always hold in their simplest form, as stated earlier. If, for instance, an observer is placed on a disc rotating relative to the earth, he/she will sense a 'force' pushing him/her toward the periphery of the disc, which is not caused by any interaction with other bodies. Here, the acceleration is not a consequence of the usual force, but of the so-called inertial force. Newton's laws hold in their simplest form only in a family of reference frames, called *inertial* frames. This fact represents the essence of the Galilean principle of relativity (PR):

PR: The laws of mechanics have the same form in all inertial frames.

The concepts of force and acceleration in Newton's laws are defined relative to an inertial frame. Both of them have the same value in two inertial frames, moving relative to each other with a constant velocity. This can be seen by observing that the space and time coordinates† in two such frames S and S' (we assume that S' moves in the x-direction of S with constant velocity v) are related in the following way:

$$x' = x - vt \qquad y' = y \qquad z' = z \qquad t' = t. \tag{1.1}$$

These relations, called *Galilean transformations*, represent the mathematical realization of the Galilean PR. If a particle moves along the x-axis of the frame S, its velocities, measured in S and S', respectively, are connected by the relations

$$u'_1 = u_1 - v \qquad u'_2 = u_2 \qquad u'_3 = u_3 \tag{1.2}$$

representing the *classical velocity addition* law ($u_1 = \mathrm{d}x/\mathrm{d}t$, etc). This law implies that the acceleration is the same in both frames. Also, the gravitational force $m_1 m_2/r^2$, for instance, has the same value in both frames.

Similar considerations lead us to conclude that there is an infinite set of inertial frames, all moving uniformly relative to each other. What property singles out the class of inertial frames from all the others in formulating the laws of classical mechanics?

Absolute space. In order to answer this and other similar questions, Newton introduced the concept of absolute space, that is given *a priori* and independently of the distribution and motion of matter. Each inertial frame moves with a constant velocity relative to absolute space and inertial forces appear as a consequence of the acceleration relative to this space.

† Standard coordinates in inertial frames are orthonormal Cartesian spatial coordinates (x, y, z) and a time coordinate t.

Trying to prove the physical relevance of acceleration relative to absolute space, Newton performed the following experiment. He filled a vessel with water and set it to rotate relative to the frame of distant, fixed stars (absolute space). The surface of the water was at first flat, although the vessel rotated. Then, due to the friction between the water and the vessel, the water also began to rotate, its surface started to take a concave form, and the concavity increased until the water was rotating at the same rate as the vessel. From this behaviour Newton drew the conclusion that the appearance of inertial forces (measured by the concavity of the surface of water) does not depend on the acceleration relative to other objects (the vessel), but only on the acceleration relative to absolute space.

Absolute space did not explain the selected role of inertial frames; it only clarified the problem. Introduction of absolute space is not consistent within classical mechanics itself. The physical properties of absolute space are very strange. Why can we only observe accelerated and not uniform motion relative to absolute space? Absolute space is usually identified with the frame of fixed stars. Well-founded objections against absolute space can be formulated in the form of the following statements:

- The existence of absolute space contradicts the internal logic of classical mechanics since, according to Galilean PR, none of the inertial frames can be singled out.
- Absolute space does not explain inertial forces since they are related to acceleration with respect to any one of the inertial frames.
- Absolute space acts on physical objects by inducing their resistance to acceleration but it cannot be acted upon.

Thus, absolute space did not find its natural place within classical mechanics, and the selected role of inertial frames remained essentially unexplained.

The speed of light. Galilean PR holds for all phenomena in mechanics. In the last century, investigation of electricity, magnetism and light aroused new interest in understanding the PR. Maxwell (1831–79) was able to derive equations which describe electricity and magnetism in a unified way. The light was identified with electromagnetic waves, and physicists thought that it propagated through a medium called the *ether*. For an observer at rest relative to the ether, the speed of light is $c = 3 \times 10^{10}$ cm s^{-1}, while for an observer moving towards the light source with velocity v, the speed of light would be $c' = c + v$, on the basis of the classical velocity addition law. The ether was for light the same as air is for sound. It was one kind of realization of Newton's absolute space. Since it was only in the reference frame of ether that the speed of light was c, the speed of light could be measured in various reference frames and the one at rest relative to the ether could be found. If there was one such frame, PR would not hold for electromagnetism. The fate of absolute space was hidden in the nature of electromagnetic phenomena.

Although the classical velocity addition law has many confirmations in classical mechanics, it *does not hold* for the propagation of light. Many experiments have shown that

$$c' = c.$$

The speed of light is the same in all inertial frames, at all times and in all directions, independently of the motion of the source and/or the observer. This fact represents a cornerstone of SR. It contradicts classical kinematics but must be accepted on the basis of the experimental evidence. The constancy of the speed of light made the ether unobservable and eliminated it from physics forever.

A convincing experimental resolution of the question of the relativity of light phenomena was given by Michelson and Morley in 1887. They measured the motion of the light signal from a source on the moving earth and showed that its velocity is independent of the direction of motion. From this, we conclude that

- since the motion of an observer relative to the ether is unobservable, the PR also holds for light phenomena; and
- the speed of light does not obey the classical velocity addition law, but has the same value in all inertial frames.

Note that in the first statement the PR puts all inertial frames on an equal footing without implying Galilean transformations between them, since these contain the classical velocity addition law, which contradicts the second statement. Thus, it becomes clear that the PR must take a new mathematical form, one that differs from (1.1).

From space and time to spacetime

The results of previous considerations can be expressed in the form of the two postulates on which SR is based.

The first postulate is a generalization of Galilean PR not only to light phenomena, but to the whole of physics and is often called Einstein's principle of relativity.

P1. Physical laws have the same form in all inertial frames.

Although (P1) has a form similar to Galilean PR, the contents of the two are essentially different. Indeed, Galilean PR is realized in classical mechanics in terms of Galilean transformations and the classical velocity addition law, which does not hold for light signals. The realization of (P1) is given in terms of Lorentz transformations, as we shall soon see.

The second postulate is related to the experimental fact concerning the speed of light.

P2. The speed of light is finite and equal in all inertial frames.

The fact that the two postulates are not in agreement with the classical velocity addition law cannot be explained within Newtonian mechanics. Einstein found a simple explanation of this puzzle by a careful analysis of the space and time characteristics of physical events. He came to the conclusion that the concepts of time and space are *relative*, i.e. dependent on the reference frame of an observer.

The moment at which an event happens (e.g. the flash of a bulb) may be determined by using clocks. Let $\mathrm{T_A}$ be a clock at point A; the time of an event at A is determined by the position of the clock hands of $\mathrm{T_A}$ at the moment of the occurrence of that event. If the clock $\mathrm{T_1}$ is at $\mathrm{A_1}$, and the bulb is placed at some distant point $\mathrm{A_2}$, then $\mathrm{T_1}$ does not register the moment of the bulb flash at $\mathrm{A_2}$, but the moment the signal arrives at $\mathrm{A_1}$. We can place another clock at $\mathrm{A_2}$ which will measure the moment of the flash, but that is not enough. The clocks $\mathrm{T_1}$ and $\mathrm{T_2}$ have to be synchronized: if the time of the flash at $\mathrm{A_2}$ is t_2, then the time of the arrival of the signal at $\mathrm{A_1}$, according to T_1, must be $t_1 = t_2 +$ (the travelling time of the signal). By this procedure we have defined the *simultaneity* of distant events: taking into account the travelling time of the signal we know which position of the clock hands at $\mathrm{A_1}$ is synchronized with the bulb flash at $\mathrm{A_2}$.

A set of synchronized clocks $\mathrm{T_1}, \mathrm{T_2}, \ldots$, disposed at all points of reference frame S, enables the measurement of time t of an arbitrary event in S. According to this definition, the concept of simultaneity of two events is related to a given inertial reference frame S. This notion of simultaneity is relative, as it depends on the inertial frame of an observer.

Using similar arguments, we can conclude that space lengths and time intervals are also relative quantities.

Lorentz transformations. Classical ideas about space and time, which are expressed by Galilean transformations, have to be changed in accordance with postulates (P1) and (P2). These postulates imply a new connection between two inertial frames S and S', which can be expressed by the *Lorentz transformation* of coordinates:

$$\begin{aligned} x' &= \frac{x - vt}{\sqrt{1 - v^2/c^2}} \qquad & t' &= \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}} \\ y' &= y & z' &= z. \end{aligned} \tag{1.3}$$

In the limit of small velocities v, the Lorentz transformation reduces to the Galilean one.

From (1.3) we obtain a new law for the *addition of velocities*:

$$u'_1 = \frac{u_1 - v}{1 - u_1 v/c^2} \qquad u'_2 = u_2, \qquad u'_3 = u_3. \tag{1.4}$$

The qualitative considerations concerning the relativity of space and time can now be put into a precise mathematical form.

We begin by the relativity of lengths. Consider a rigid rod fixed in an inertial frame S', whose (proper) length is $\Delta x' = x_2' - x_1'$. The length of the rod in another inertial frame S is determined by the positions of its ends at the same moment of the S-time: $\Delta x = x_2(t) - x_1(t)$. From expressions (1.3) it follows that $\Delta x = \Delta x'\sqrt{1 - v^2/c^2}$. The length of the moving rod, measured from S, is less than its length in the rest frame S', $\Delta x < \Delta x'$. This effect is called the *length contraction*.

In order to clarify the relativity of time intervals, we consider a clock fixed in S'. Its 'tick' and 'tack' can be described by coordinates (x', t_1') and (x', t_2'). Using Lorentz transformation (1.3), we obtain the relation $\Delta t' = \Delta t\sqrt{1 - v^2/c^2}$, which shows that the time interval between two strikes of the clock is shortest in its rest frame, $\Delta t' < \Delta t$. Since the rate assigned to a moving clock is always longer than its proper rate, we talk about *time dilatation*. An interesting phenomenon related to this effect is the so-called twin paradox.

Both length contraction and time dilatation are real physical effects.

Four-dimensional geometry. The connection between the space and time coordinates of two inertial frames, moving with respect to one other with some velocity v, is given by the Lorentz transformation (1.3). It is easily seen that the general transformation between the two inertial frames includes spatial rotations and translations as well as time translations. The resultant set of transformations is known as the set of Poincaré transformations. Since these transformations 'mix' space and time coordinates, it turns out that it is more natural to talk about four-dimensional *spacetime* than about space and time separately. Of course, although space and time have equally important roles in spacetime, there is a clear physical distinction between them. This is seen in the form of the Lorentz transformation, and this has an influence on the geometric properties of spacetime.

The invariance of the expression $s^2 = c^2t^2 - x^2 - y^2 - z^2$ with respect to the Poincaré transformations represents a basic characteristic of the spacetime continuum—the Minkowski space M_4. The points in M_4 may be labelled by coordinates (t, x, y, z), and are called events. The expression s^2 has the role of the squared 'distance' between the events $(0, 0, 0, 0)$ and (t, x, y, z). In the same way as the squared distance in the Euclidean space E_3 is invariant under Galilean transformations, the expression s^2 in M_4 is invariant under Poincaré transformations. It is convenient to introduce the squared differential distance between neighbouring events:

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2. \tag{1.5a}$$

This equation can be written in a more compact, tensor form:

$$\mathrm{d}s^2 = \eta_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu \tag{1.5b}$$

where $\mathrm{d}x^\mu = (\mathrm{d}t, \mathrm{d}x, \mathrm{d}y, \mathrm{d}z)$, $\eta_{\mu\nu} = \mathrm{diag}(+, -, -, -)$ is the *metric* of M_4, and a summation over repeated indices is understood.

Lorentz transformations have the form $x'^{\mu} = \Lambda^{\mu}{}_{\nu} x^{\nu}$, where the coefficients $\Lambda^{\mu}{}_{\nu}$ are determined by equation (1.3). The set of four quantities which transform according to this rule is called a vector of M_4. The geometric formalism can be further developed by introducing general tensors; Lorentz transformations can be understood as 'rotations' in M_4 (since they do not change the 'lengths' of the vectors), etc. The analogy with the related concepts of Euclidean geometry is substantial, but not complete. While Euclidean metric is positive definite, i.e. $\mathrm{d}s^2$ is positive, the Minkowskian metric is indefinite, i.e. $\mathrm{d}s^2$ may be positive, negative or zero. The distance between two points in M_4 may be zero even when these points are not identical. However, this does not lead to any essential difference in the mathematical treatment of M_4 compared to the Euclidean case. The indefinite metric is a mathematical expression of the distinction between space and time.

The geometric formulation is particularly useful for the generalization of this theory and construction of GR.

1.2 Gravitation and geometry

The principle of equivalence

Clarification of the role of inertial frames in the formulation of physical laws is not the end of the story of relativity. Attempts to understand the physical meaning of the accelerated frames led Einstein to the general theory of space, time and gravitation.

Let us observe possible differences between the inertial and gravitational properties of a Newtonian particle. Newton's second law of mechanics can be written in the form $\boldsymbol{F} = m_{\mathrm{i}}\boldsymbol{a}$, where m_{i} is the so-called inertial mass, which measures inertial properties (resistance to acceleration) of a given particle. The force acting on a particle in a homogeneous gravitational field $\boldsymbol{g}$ has the form $\boldsymbol{F}_g = m_g \boldsymbol{g}$, where m_{g} is the gravitational mass of the particle, which may be regarded as the gravitational analogue of the electric charge. Experiments have shown that the ratio $m_{\mathrm{g}}/m_{\mathrm{i}}$ is the same for all particles or, equivalently, that all particles experience the same acceleration in a given *gravitational field*. This property has been known for a long time as a consequence of Galileo's experiments with particles moving along an inclined plane. It is also true in an inhomogeneous gravitational field provided we restrict ourselves to small regions of spacetime. The uniqueness of the motion of particles is a specific property of the gravitational interaction, which does not hold for any other force in nature.

On the other hand, all free particles in an *accelerated frame* have the same acceleration. Thus, for instance, if a train accelerates its motion relative to the earth, all the bodies on the train experience the same acceleration relative to the train, independently of their (inertial) masses. According to this property, as noticed by Einstein, the dynamical effects of a gravitational field and an accelerated frame cannot be distinguished. This is the essence of the principle of equivalence (PE).

PE. Every non-inertial frame is locally equivalent to some gravitational field.

The equivalence holds only *locally*, in small regions of space and time, where 'real' fields can be regarded as homogeneous.

Expressed in a different way, the PE states that a given gravitational field can be locally compensated for by choosing a suitable reference frame—a freely falling (non-rotating) laboratory. In each such frame, all the laws of mechanics or, more generally, the laws of physics have the same form as in an inertial frame. For this reason, each freely falling reference frame is called a locally inertial frame.

PE′. At every point in an arbitrary gravitational field we can choose a locally inertial frame in which the laws of physics take the same form as in SR.

We usually make a distinction between the weak and strong PE. If we restrict this formulation to the laws of mechanics, we have the weak PE. On the other hand, if 'the laws of physics' means *all* the laws of physics, we have Einstein's PE in its strongest form (sometimes, this 'very strong' version of the PE is distinguished from its 'medium-strong' form, which refers to all non-gravitational laws of physics) (for more details see, for instance, Weinberg (1972) and Rindler (1977)).

In previous considerations we used Newtonian mechanics and gravitation to illustrate the meaning of the (weak) PE. As previously mentioned, the first experimental confirmation of the equality of m_{i} and m_{g} (in suitable units) was given by Galileo. Newton tested this equality by experiments with pendulums of equal length but different composition. The same result was verified later, with a better precision, by Eötvös (1889; with an accuracy of 1 part in 10^9), Dicke (1964; 1 part in 10^{11}) and Braginsky and Panov (1971; 1 part in 10^{12}). Besides, all experimental evidence in favour of GR can be taken as an indirect verification of the PE.

The PE and local Poincaré symmetry. It is very interesting, but not widely known, that the PE can be expressed, using the language of modern physics, as the principle of *local symmetry*. To see this, we recall that at each point in a given gravitational field we can choose a locally inertial reference frame $S(x)$ (on the basis of the PE). The frame $S(x)$ can be obtained from an arbitrarily fixed frame $S_0 \equiv S(x_0)$ by

- translating S_0, so as to bring its origin to coincide with that of $S(x)$ and
- performing Lorentz 'rotations' on S_0, until its axes are brought to coincide with those of $S(x)$.

Four translations and six Lorentz 'rotations' are the elements of the Poincaré group of transformations, the parameters of which depend on the point x at which the locally inertial frame $S(x)$ is defined. Since the laws of physics have the same form in all locally inertial frames (on the basis of the PR), these Poincaré transformations are symmetry transformations. Thus, an arbitrary gravitational field is characterized by the group of local, x-dependent Poincaré transformations, acting on the set of all locally inertial frames. When the gravitational field is absent, we return to SR and the group of global, x-independent Poincaré transformations.

Physics and geometry

The physical content of geometry. The properties of space and time cannot be deduced by pure mathematical reasoning, omitting all reference to physics. There are many possible geometries that are equally good from the mathematical point of view, but not so many if we make use of the physical properties of nature. We believed in Euclidean geometry for more than 2000 years, as it was very convincing with regard to the description of physical reality. However, its logical structure was not completely clear. Attempts to purify Euclid's system of axioms led finally, in the 19th century, to the serious acceptance of non-Euclidean geometry as a logical possibility. Soon after that, new developments in physics, which resulted in the discovery of SR and GR, showed that non-Euclidean geometry is not only a mathematical discipline, but also part of physics. We shall now try to clarify how physical measurements can be related to the geometric properties of space and time.

For mathematicians, geometry is based on some elementary concepts (such as a point, straight line, etc), which are intuitively more or less clear, and certain statements (axioms), which express the most fundamental relations between these concepts. All other statements in geometry can be proved on the basis of some definite mathematical methods, which are considered to be true within a given mathematical structure. Thus, the question of the truthfulness of a given geometric statement is equivalent to the question of the 'truthfulness' of the related set of axioms. It is clear, however, that such a question has no meaning within the geometry itself.

For physicists, the space is such as is seen in experiments; that space is, at least, the space relevant for physics. Therefore, if we assign a definite physical meaning to the basic geometric concepts (e.g. straight line $\equiv$ path of the light ray, etc), then questions of the truthfulness of geometric statements become questions of physics, i.e. questions concerning the relations between the relevant physical objects. This is how physical measurements become related to geometric properties.

Starting from physically defined measurements of space and time in SR, we are naturally led to introduce the Minkowskian geometry of spacetime. To illustrate what happens in GR, we shall consider the geometry on a flat disc, rotating uniformly (relative to an inertial frame) around the axis normal to its plane, passing through the centre of the disc. There is an observer on the disc,

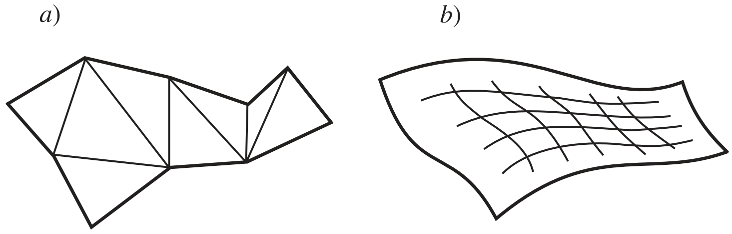

Figure 1.1. Approximate realization (a) of a curved surface (b).

trying to test spacetime geometry by physical measurements. His/her conclusions will also be valid locally for true gravitational fields, on the basis of the PE. Assume that the observer sits at the centre of the disc, and has two identical clocks: one of them is placed at the centre, the other at some point on the periphery. The observer will see that the clock on the periphery is running slower (time dilatation from SR). Consequently, clocks at various positions in the gravitational field run faster or slower, depending on the local strength of the field. There is no definition of time that is pertinent to the whole spacetime in general. The observer will also conclude that the length of a piece of line orthogonal to the radius of the disc will be shortened (length contraction from SR). Therefore, the ratio of the circumference of the circle to its radius will be smaller than 2π, so that the Euclidean geometry of space does not hold in GR.

Geometry of curved surfaces. We have seen that in spacetime, within a limited region, we can always choose a suitable reference frame, called the local inertial frame. Taking spacetime apart into locally inertial components, we can apply the laws of SR in each such component and derive various dynamical conclusions. Reconstruction of the related global dynamical picture, based on the PE (and some additional, simple geometric assumptions), gives rise to GR.

This procedure for dissecting spacetime into locally inertial components, out of which we can reconstruct, using the PE, the global structure of spacetime containing the gravitational field, can be compared geometrically with an attempt to build a curved surface, approximately, from a lot of small, plane elements 'continuously' bound to each other (figure 1.1). A locally flat surface is a geometric analogue of locally inertial spacetime.

GR makes essential use of curved spaces (or, more accurately curved spacetimes). Since the curved, four-dimensional spacetime cannot be visualized, let us try to understand the basic features of curved spaces by considering a two-dimensional surface X_2. We shall focus our attention on those geometric properties of X_2, which could be determined by an intelligent, two-dimensional being entirely confined to live and measure in the surface, a being that does not know anything about the embedding Euclidean space E_3. Such properties determine the *intrinsic geometry* on X_2.

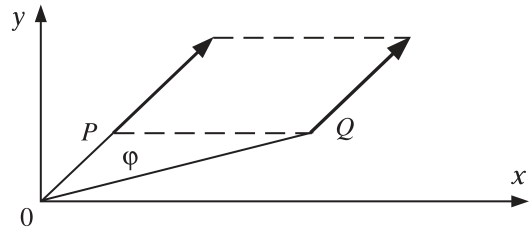

Figure 1.2. Vector components change under the parallel transport.

On a general surface we cannot set up Cartesian coordinates; the best we can do is to introduce two arbitrary families of coordinate lines and label them by $u^\alpha = (u^1, u^2)$. These curvilinear (or Gaussian) coordinates do not always have a direct geometric interpretation.

Let us now imagine a surface X_2 embedded in E_3, and consider a point P in X_2 with Cartesian coordinates (x, y, z) in E_3; since P is in X_2, these coordinates can be expressed as functions of u^α. The squared distance between two neighbouring points in X_2, $P(x, y, z)$ and $Q(x+\mathrm{d}x, y+\mathrm{d}y, z+\mathrm{d}z)$, is given by $\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$. Going over to u^α we obtain

$$\mathrm{d}s^2 = g_{\alpha\beta}\,\mathrm{d}u^\alpha\,\mathrm{d}u^\beta \tag{1.6}$$

where the set of functions $g_{\alpha\beta}(u^1, u^2)$ defines the *metric* on X_2. The metric is an essential intrinsic property of X_2, independent of its embedding in E_3. A two-dimensional continuum X_2, equipped with a squared differential distance according to (1.6), becomes the *metric space*, $G_2 = (X_2, \boldsymbol{g})$ (it is now clear that the symbol X_2 is used to denote an abstract two-dimensional continuum, disregarding all metric properties of the real surface).

Studying the motion of particles in X_2 our observer could have arrived at the idea of the tangent vector, $\boldsymbol{a} = (a^1, a^2)$. Consider, further, two tangent vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, defined at neighbouring points P and Q. If we wish to compare these two vectors we have to know how one of them can be transported to the position of the other. This transport is called *parallel transport*. The continuum X_2 together with the parallel transport law, denoted by Γ, is called a *linearly connected* (or *affine*) space, $L_2 = (X_2, \Gamma)$.

The components of a vector in curvilinear coordinates are expected to change under parallel transport. This can be easily seen by considering the parallel transport of a vector in a plane, using polar coordinates (r, θ). As shown in figure 1.2, if the unit vector $\boldsymbol{a}$ at point P has components $(a^r = 1, a^\theta = 0)$, after parallel transport from P to Q its components become $(a^r = \cos\varphi, a^\theta = \sin\varphi)$.

A two-dimensional observer should determine the rule of parallel transport of vectors in accordance with his/her own experience. This rule is generally independent of the concept of a metric. If X_2 is equipped with a rule of parallel transport and the independent concept of a metric, we obtain the *linearly connected metric* space $(X_2, \Gamma, \boldsymbol{g}) = (L_2, \boldsymbol{g}) = (G_2, \Gamma)$.

It is very probable that our two-dimensional observer will demand that the lengths of vectors should remain unchanged under parallel transport. After adopting this very natural property, which relates Γ and $\boldsymbol{g}$, the linearly connected metric space becomes the *Riemann–Cartan* space U_2. The usual surfaces have some additional geometric structure, and belong to the class of *Riemann* spaces V_2.

This discussion was slightly more general than strictly necessary for studying surfaces embedded in E_3, in order to emphasize that the intrinsic geometry on X_2 is defined in terms of

- the metric $\boldsymbol{g}$ and
- the rule of parallel transport Γ.

These basic concepts of the intrinsic geometry of surfaces can be extended directly to spaces with higher dimensions, although in this case we lose the intuitive geometric picture.

Any Riemann space is locally Euclidean, i.e. its metric is positive definite ($\mathrm{d}s^2 > 0$). A slight generalization of this case, that consists in admitting metrics that are pseudo-Euclidean, leads to a pseudo-Riemannian geometry or Riemannian geometry with an indefinite metric.

Relativity, covariance and Mach's ideas

Relativity and covariance. We have seen previously that inertial frames play a particularly important role in SR. Einstein considered this to be an expression of the incompleteness of SR. He wanted to generalize the relativity of inertial motions to the relativity of all motions, including accelerated ones. This idea can be formulated as the general principle of relativity:

> *General PR: The form of physical laws is the same in all reference frames.*

Equations expressing physical laws are often formulated using specific coordinates. Thus, for instance, spatial positions in an inertial laboratory L can be determined using Cartesian coordinates (x, y, z) or the coordinates obtained from these by a rotation around the z-axis for a fixed angle φ. If we allow the angle φ to change with time as $\varphi = \omega t$, new coordinates can be naturally attached to a new physical reference frame, rotating relative to L with the angular velocity ω around the z-axis. We see that coordinate transformations can be interpreted as describing transitions from one to another laboratory (the so-called active interpretation). Thus, it becomes clear that the general PR can be realized with the help of the principle of general covariance:

> *General covariance: The form of physical laws does not depend on the choice of coordinates.*

General covariance is a technical way to express the general PR.

When mathematicians study the geometric properties of space, they use certain geometric objects with properties that do not depend on the choice of coordinates. The covariance means that physical laws can be expressed in terms of these geometric objects. In classical physics such objects are tensors, so that covariant equations must be tensorial.

General covariance seems to be a powerful principle which determines acceptable forms of physical equations. Is it really so? The next example shows a certain triviality of this principle. Consider Newton's second law in an inertial frame S, $m\boldsymbol{a} = \boldsymbol{F}$, which is invariant under the group of Galilean transformations G. Going over to a non-inertial frame $\widetilde{S}$ that has linear acceleration $\boldsymbol{a}_0$ and angular velocity $\boldsymbol{\omega}$ relative to S, Newton's equation transforms into

$$m\boldsymbol{a} = \boldsymbol{F} - m\boldsymbol{a}_0 - 2m\boldsymbol{\omega} \times \boldsymbol{v} - m\dot{\boldsymbol{\omega}} \times \boldsymbol{r} - m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{r}). \tag{1.7}$$

The transition from S to $\widetilde{S}$ can be described as a Galilean transformation (a spatial translation combined with a rotation) with time-dependent parameters. The set of these transformations forms a group, denoted as $\widetilde{G}$, which is larger than the Galilean group G. In this way we obtained the $\widetilde{G}$-covariant form of Newton's equation. Other physical equations can also be covariantized in a similar manner, which means that covariance implies that there are no stringent conditions on the form of the physical equations.

It is important to note that equation (1.7) is completely determined by its original form in S and the process of covariantization. The original equation has Galilean symmetry, and the covariantization 'knows' of this property, since it is based on $\widetilde{G}$ transformations. Using modern language, we can describe this procedure as an incomplete *localization* of Galilean symmetry, based on time-dependent parameters. Since the procedure does not depend on the form of the original equation but only on its global symmetry, it is clear that covariantization does not restrict the form of physically acceptable equations.

In spite of this conclusion, we know that Einstein's covariant equations have experimental consequences that are essentially different from those of Newton's theory of gravity, so that general covariance still seems to be an important principle for physics. This dilemma can be resolved in the following way. If we restrict ourselves to weak gravitational fields for simplicity, we obtain an approximate, non-covariant form of Einstein's theory, known as the Pauli–Fierz field theory in M_4. Comparing this field theory with Newton's action-at-a-distance theory we easily find significant differences; in particular, these theories have different global symmetries (Poincaré and Galilean, respectively). It is therefore quite natural that their physical contents are essentially different. This difference has nothing to do with the covariance of Einstein's equations.

After clarifying this point, we are now ready to conclude that general covariance by itself does not have any physical content. However, we should note that Einstein related covariance to the PE, so that general covariance (transition to accelerated frames) represents an important technical procedure for *introducing the gravitational field* into any physical theory. Thus, Einstein's theory of gravity is obtained by unifying the principle of general covariance with the PE:

General covariance + PE $\Rightarrow$ *theory of gravity.*

Here, general covariance stands essentially for the general PR, while the PE tells us not only how to introduce gravity, but also the kind of physics we have in the absence of gravity.

General covariance is only a statement about the effects of gravity, it does not imply the validity of SR in locally inertial frames. Although the local validity of SR is usually assumed, theoretically it is not necessary; for instance, we could assume Galilean relativity in locally inertial frames. This is particularly clear if we think of gravity as a locally invariant theory.

Covariance and symmetry. Although the concepts of general covariance and symmetry have certain formal similarities, they are, in fact, essentially different. In order to illustrate this difference, consider the interval in $G_4 = (X_4, \boldsymbol{g})$,

$$\mathrm{d}s^2 = g_{\mu\nu}(x)\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu \tag{1.8a}$$

which is expressed in some coordinate system K. This equation is covariant with respect to general coordinate transformations $x \to x' = x'(x)$, since it has the same form in the new coordinate system K' as it had in K:

$$\mathrm{d}s^2 = g'_{\mu\nu}(x')\,\mathrm{d}x^{\mu'}\,\mathrm{d}x^{\nu'}. \tag{1.8b}$$

Here, the coordinate transformation induces a related functional change in the metric, $g_{\mu\nu} \to g'_{\mu\nu}$, where the prime denotes a new function. Now, if we restrict ourselves to those transformations that do not change the form of the metric, $g'_{\mu\nu}(x) = g_{\mu\nu}(x)$, we obtain symmetry transformations (or isometries) of the space G_4.

To see the meaning of this definition in Minkowski space, let us consider equation (1.8a) in an inertial frame S, where $g_{\mu\nu} = \eta_{\mu\nu}$ is the metric of M_4. If we look for coordinate transformations that do not change the form of the metric, we obtain Poincaré transformations. They express the symmetry of physical laws under transition from one inertial frame to another in M_4.

Mach's principle. In his attempts to build GR, Einstein was led by the idea of abolishing the privileged role of inertial frames from physics. Even if it were possible to develop a satisfactory theory of gravity within SR, he would not have stopped there—he had to go beyond SR. On this route he was greatly influenced by the ideas of the philosopher E Mach (1836–1916), who gave the first constructive critique of Newton's understanding of inertia and absolute space, and proposed a new approach to these questions.

The conclusion drawn by Newton on the basis of his experiment with a rotating vessel (that the acceleration of the water with respect to the vessel does not produce inertial forces, they are produced only by the acceleration of the water with respect to absolute space) cannot be considered as well founded. Newton's experiment only shows that the rotation of the water relative to a *small* vessel produces no noticeable inertial forces. It is not clear how the experiment would have turned out if the vessel had been more massive—these were the suspicions raised by Mach. He believed that water rotating relative to a very massive vessel would remain flat, since it would be at rest with respect to the very close and massive vessel, while the water that does not rotate would be curved, since its motion with respect to the vessel would be accelerated. In other words, inertial forces are caused by the acceleration of the water relative to the average mass distribution in the universe, not with respect to absolute space.

We shall now illustrate the difference between Newton and Mach by another example. Consider an elastic sphere that is rotating relative to some inertial frame, getting larger at the equator. How does the sphere 'know' that it is rotating and hence must deform? Newton might have said that it was accelerated relative to absolute space, which thereby causes the inertial forces that deformed it. Mach would have said that the sphere was accelerated relative to distant masses in the universe, which was the cause of deformation.

It is interesting to observe that both Newton and Mach explained the appearance of inertial forces by the *acceleration* of a body, not by its velocity. Why is that so? Mach's considerations were general in nature, he was not able to give any quantitative description of these ideas. The inertial influence of cosmic matter was clearly defined in Einstein's GR (inertia is a manifestation of the geometry of spacetime, geometry is affected by the presence of matter) which, to some extent, realizes Mach's abstract ideas in a concrete physical theory.

Mach's principle can be roughly expressed in the form of the following (not completely independent) statements:

- Space as such plays no role in physics; it is merely an abstraction from the totality of spatial relations between material objects.
- The inertial properties of every particle are determined by its interaction with all the other masses in the universe.
- Any motion is relative and can be determined only with respect to all the other masses in the universe.
- Inertial forces are caused by an acceleration of a body relative to the average mass distribution in the universe.

It is often believed that Mach's principle is so general that it cannot be checked experimentally so that, essentially, it has no physical content. Yet, there are examples showing clearly that this is not so. In spite of this, there is still no direct experimental verification (Rindler 1977).

Bearing in mind the strong influence of Mach's principle on Einstein, it is interesting to clarify to what extent this principle has been justified by GR. We can say that GR has realized only some of Mach's ideas; indeed, solutions of Einstein's field equations are determined not only by matter distribution, but also by boundary conditions. Hence, the same mass distribution may give rise to different gravitational fields. Mach wanted to abolish space in its own right and replace it with the relative configuration of matter; Einstein retained spacetime but made it non-absolute. Although Mach's principle helped stimulate the development of GR, there is no reason today to consider it as a basic principle of physics.

Perspectives of further developments

Einstein's theory of gravity predicted a number of physical effects which can be completely verified experimentally. Nevertheless, we should stress that certain properties of this theory deserve critical analysis.

GR is characterized by the existence of *singularities*, which are generic features of solutions describing both localized physical systems (black holes) and cosmology (big bang). A spacetime singularity can be intuitively characterized by an infinite growth of some physical quantities in some regions of spacetime and the analogous pathological behaviour of the related geometric objects. Both cosmology and black holes represent very interesting phenomena in gravitational physics—their features cast light on the intrinsic limitations of GR and serve as a signpost towards a new, consistent approach to gravitation.

The standard cosmological model predicts that at some finite time in the past, the universe was in a singular state (big bang). For many years it was generally believed that this prediction was due merely to the simplifying assumptions of the model. However, the singularity theorems of GR have shown that singularities are generic features of cosmological solutions. We expect that the singular behaviour can be avoided either by modifying classical GR and/or by taking quantum effects into account.

General singularity theorems also apply to black holes—gravitational objects that inevitably arise at the endpoint of the thermonuclear evolution of some stars. After a finite proper time all the matter of the star meets at one point, where the density of matter and the curvature of spacetime become infinite.

A cosmological singularity means that there existed a moment of time in the past, such that nothing before that moment, even time itself, did not have any sense, did not exist. The singularity of the black hole is related to the future, and means that for an observer falling into the black hole, there will come a moment at which not only the observer's life will end, but also time itself will go out of existence. In these extreme conditions, where the density of matter becomes very large and the predictions of GR contradict the fundamental concepts of classical physics, quantum effects must be taken into account. Could these effects alone remove the singularity or do we need to modify classical GR—that is the question for the future.

Gravity is the only basic physical interaction for which no consistent quantum formulation exists. Some people think that in order to construct quantum gravity we need to radically change our understanding of the structure of space and time. If this were correct, it would be the second time in the last century that gravity has played a key role in the development of our concepts of space and time.

Many attempts to quantize gravity have been unsuccessful. We do not now whether the remarkable ideas of *gauge symmetry, supergravity, Kaluza–Klein theory* or *string theory* will lead us to a unified, quantum theory of the fundamental interactions. The former successes of these ideas motivate us to study them carefully in the hope that they will bring us closer to a consistent formulation of quantum gravity and its unification with the other fundamental interactions.

Chapter 2

Spacetime symmetries

The physics of elementary particles and gravitation is successfully described by Lagrangian field theory. The dynamical variables in this theory are fields $\phi(x)$ and the dynamics is determined by a function of the fields and their derivatives, $\mathcal{L}(\phi, \partial_\mu\phi)$, called the Lagrangian. Equations of motion are given as the Euler–Lagrange equations of the variational problem $\delta_\phi I = 0$ for the action integral $I = \int d^4x\, \mathcal{L}$.

In physical processes at low energies the gravitational field does not play a significant role, since the gravitational interaction is extremely weak. The structure of spacetime without gravity is determined by the relativity principle and the existence of a finite, maximal velocity of propagation of physical signals. The unity of these two principles, sometimes called Einstein's relativity principle, represents the basis for special relativity theory. Spacetimes based on Einstein's relativity have the structure of the Minkowski space M_4. The equivalence of inertial reference frames is expressed by the *Poincaré symmetry* in M_4.

Due to the fact that some physical constants are dimensional, the physical world is not invariant under a change of scale. In physical processes at high energies, where dimensional constants become practically negligible, we expect scale invariance to appear, so that the relevant symmetry becomes the *Weyl symmetry* or, the even higher, *conformal symmetry*. Conformal symmetry is a broken symmetry in basic physical interactions. Of particular interest for string theory is conformal symmetry in two dimensions.

Bearing in mind the importance of these spacetime symmetries in particle physics, in this chapter we shall give a review of those properties of Poincaré and conformal symmetries that are of interest for their *localization* and the construction of related gravitational theories. Similar ideas can be, and have been, applied to other symmetry groups. Understanding gravity as a theory based on local spacetime symmetries represents an important step towards the unification of all fundamental interactions.

2.1 Poincaré symmetry

Poincaré transformations

The Minkowski spacetime M_4 is a four-dimensional arena, successfully used to describe all physical phenomena except gravitation. The physical observer in spacetime uses some *reference frame*, usually associated with an imagined extension of a physical object, and endowed with *coordinates* x^μ ($\mu = 0, 1, 2, 3$), serving to identify physical events. In M_4 preferred reference frames, called *inertial* frames, exist. An inertial observer can always choose coordinates, global inertial coordinates, such that the infinitesimal interval between points $P(x)$ and $Q(x + \mathrm{d}x)$ has the form

$$\mathrm{d}s^2 = \eta_{\mu\nu}\mathrm{d}x^\mu\,\mathrm{d}x^\nu \tag{2.1}$$

where $\eta_{\mu\nu} = (1, -1, -1, -1)$ is the metric tensor in the inertial frame $S(x)$. Since the interval between P and Q does not depend on the choice of reference frame, the transition to another reference frame $S'(x')$ implies

$$\eta_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = g'_{\mu\nu}(x')\,\mathrm{d}x'^\mu\,\mathrm{d}x'^\nu$$

where $g'_{\mu\nu}$ is the metric in S'. The new reference frame is inertial if there exist coordinates x'^μ such that $g'_{\mu\nu} = \eta_{\mu\nu}$, i.e. if the *form* of the metric is not changed by the transition $S \to S'$.

The form variation of a field $F(x)$, $\delta_0 F(x) \equiv F'(x) - F(x)$, should be clearly distinguished from its total variation, $\delta F(x) \equiv F'(x') - F(x)$. When $x' - x = \xi$ is infinitesimally small, we get

$$\delta F(x) \approx \delta_0 F(x) + \xi^\mu \partial_\mu F(x).$$

Form variation and differentiation are commuting operations. Coordinate transformations $x \to x'$ which do not change the form of the metric define the *isometry group* of a given space. The isometry group of M_4 is the group of global *Poincaré transformations* $P(1, 3)$.

In order to find out the form of infinitesimal Poincaré transformations, let us consider the coordinate transformation

$$x^\mu \to x'^\mu = x^\mu + \xi^\mu(x) \tag{2.2}$$

relating the metrics $\eta_{\mu\nu}$ and $g'_{\mu\nu}$ by the equation

$$g'_{\mu\nu}(x') = \frac{\partial x^\lambda}{\partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\nu}\eta_{\lambda\rho} \approx \eta_{\mu\nu} - (\xi_{\mu,\nu} + \xi_{\nu,\mu}).$$

Form invariance of the metric is now expressed by the following *Killing equation*:

$$\delta_0\eta_{\mu\nu} \equiv g'_{\mu\nu}(x) - \eta_{\mu\nu} \approx -(\xi_{\mu,\nu} + \xi_{\nu,\mu}) = 0. \tag{2.3}$$

Expanding $\xi^\mu(x)$ in a power series in x,

$$\xi^\mu(x) = \varepsilon^\mu + \omega^\mu{}_\nu x^\nu + \omega^\mu{}_{\nu\rho} x^\nu x^\rho + \cdots$$

with $\varepsilon^\mu, \omega^\mu{}_\nu, \ldots$ constant parameters, condition (2.3) yields

$$\omega^{\mu\nu} + \omega^{\nu\mu} = 0, \qquad \varepsilon^\mu = \text{arbitrary}$$

while the remaining parameters vanish. Thus, the infinitesimal, *global* Poincaré transformations,

$$\xi^\mu(x) = \varepsilon^\mu + \omega^\mu{}_\nu x^\nu \tag{2.4}$$

are defined in terms of ten *constant* parameters $\omega^{\mu\nu} = -\omega^{\nu\mu}$ and ε^μ (Lorentz rotations and translations).

Finite Poincaré transformations are inhomogeneous, linear transformations

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu + a^\mu$$

where the matrix $\Lambda = (\Lambda^\mu{}_\nu)$ is determined by the form invariance of the Minkowski metric: $\eta = \Lambda^T \eta \Lambda$.

Lie algebra and its representations

In order to define the action of the Poincaré group on fields and introduce the related generators and their Lie algebra, let us first consider a field $\varphi(x)$, which is scalar with respect to the transformations (2.4): $\varphi'(x') = \varphi(x)$. As a consequence, the change of form of φ is given by

$$\delta_0 \varphi(x) = -(\omega^\mu{}_\nu x^\nu + \varepsilon^\mu)\partial_\mu \varphi(x).$$

If we define the generators of the transformation of a general field $\phi(x)$ by

$$\delta_0 \phi(x) = (\tfrac{1}{2}\omega^{\mu\nu} M_{\mu\nu} + \varepsilon^\mu P_\mu)\phi(x) \tag{2.5}$$

their coordinate representation in the case of a scalar field has the form

$$M_{\mu\nu} = x_\mu \partial_\nu - x_\nu \partial_\mu \equiv L_{\mu\nu} \qquad P_\mu = -\partial_\mu .$$

We can easily verify that they satisfy the Lie algebra

$$\begin{aligned}
[M_{\mu\nu}, M_{\lambda\rho}] &= \eta_{\nu\lambda} M_{\mu\rho} - \eta_{\mu\lambda} M_{\nu\rho} - (\lambda \leftrightarrow \rho) \equiv \tfrac{1}{2} f_{\mu\nu,\lambda\rho}{}^{\tau\sigma} M_{\tau\sigma} \\
[M_{\mu\nu}, P_\lambda] &= \eta_{\nu\lambda} P_\mu - \eta_{\mu\lambda} P_\nu \\
[P_\mu, P_\nu] &= 0.
\end{aligned} \tag{2.6}$$

In the general case of an arbitrary field ϕ the generators have the form

$$M_{\mu\nu} = L_{\mu\nu} + \Sigma_{\mu\nu} \qquad P_\mu = -\partial_\mu \tag{2.7}$$

where $\Sigma_{\mu\nu}$ is the spin part of $M_{\mu\nu}$.

Example 1. A change in the Dirac spinor field $\psi_\alpha(x)$ under the action of the Poincaré group is given by $\psi'(x') = S(\omega)\psi(x)$, where $S(\omega)$ is a matrix satisfying $S^{-1}\gamma^\mu S = \Lambda^\mu{}_\nu\gamma^\nu$, and the γ^μ are the Dirac matrices. For infinitesimal transformations $\Lambda^\mu{}_\nu = \delta^\mu_\nu + \omega^\mu{}_\nu$, we find that $S = 1 + \frac{1}{8}\omega^{\mu\nu}[\gamma_\mu, \gamma_\nu]$, i.e.

$$\Sigma^D_{\mu\nu} = \tfrac{1}{4}[\gamma_\mu, \gamma_\nu] \equiv \sigma_{\mu\nu}.$$

The vector field $V^\mu(x)$ transforms as $V'^\mu(x') = \Lambda^\mu{}_\nu V^\nu(x)$. For infinitesimal transformations we have $\Lambda^\mu{}_\nu = \delta^\mu_\nu + \omega^\mu{}_\nu = \delta^\mu_\nu + \frac{1}{2}\omega^{\lambda\rho}(\Sigma^1_{\lambda\rho})^\mu{}_\nu$, therefore

$$(\Sigma^1_{\lambda\rho})^\mu{}_\nu = \delta^\mu_\lambda\eta_{\rho\nu} - \delta^\mu_\rho\eta_{\lambda\nu}.$$

The procedure described in this example can be directly generalized and used to find the form of generators for a general field $\phi(x)$, transforming as $\phi'(x') = S(\omega)\phi(x)$ under Poincaré transformations.

The same result can be obtained by using another approach, known in group theory as the method of induced representations. This method gives a prescription of how to extend any representation of the Lorentz subgroup on $\phi(0)$ to a representation of the Poincaré group on $\phi(x)$, using the definition of $\phi(x)$ and the algebra of generators. To demonstrate the procedure, let us first introduce finite Poincaré transformations on fields:

$$\phi'(x) = (G\phi)(x) \qquad G(\omega, a) = \exp(\tfrac{1}{2}\omega \cdot M + a \cdot P) \tag{2.8}$$

where ω and a are the parameters of transformation and $\omega \cdot M \equiv \omega^{\mu\nu}M_{\mu\nu}$. In particular, translations and Lorentz transformations are represented as

$$T(a) = \exp(a \cdot P) \qquad \Lambda(\omega) = \exp(\tfrac{1}{2}\omega \cdot M).$$

The generators P and M act *on fields* and change their form in accordance with the algebra (2.6).

Next, we introduce the field ϕ_u obtained by translation from $\phi(0)$:

$$\phi_u = T(u)\phi(0).$$

Using the relation $T(x)\phi_u = \phi_{x+u}$ we can conclude that the translation generator acts on ϕ_u as $(P_\mu\phi)_u = (-\partial_\mu\phi)_u$, which implies $\phi_u = \phi(-u)$.

The action of the Lorentz rotation on ϕ_u yields

$$\phi'_u \equiv \Lambda(\omega)\phi_u = T(\bar{u})\Lambda(\omega)\phi(0) \tag{2.9}$$

where $\bar{u}$ is implicitly defined by $\Lambda(\omega)T(u) = T(\bar{u})\Lambda(\omega)$ (figure 2.1). For infinitesimal ω we find that

$$\begin{aligned}(1 + \tfrac{1}{2}\omega \cdot M)\mathrm{e}^{u\cdot P} &= \mathrm{e}^{u\cdot P}\mathrm{e}^{-u\cdot P}(1 + \tfrac{1}{2}\omega \cdot M)\mathrm{e}^{u\cdot P} \\ &= \mathrm{e}^{u\cdot P}(1 + \tfrac{1}{2}\omega \cdot M + [\tfrac{1}{2}\omega \cdot M, u \cdot P]) \\ &= \exp[(u^\mu + \omega^\mu{}_\nu u^\nu)P_\mu](1 + \tfrac{1}{2}\omega \cdot M)\end{aligned}$$

$$\begin{tikzcd} \phi(0) \arrow[r, "T(u)"] \arrow[d, "\Lambda(\omega)"'] & \phi_u \arrow[d, "\Lambda(\omega)"] \\ \phi'(0) \arrow[r, "T(\bar{u})"'] & \phi'_u \end{tikzcd}$$

Figure 2.1. Construction of the Lorentz generators.

so that $\bar{u}^\mu = u^\mu + \omega^\mu{}_\nu u^\nu$. Here we used the formula

$$\mathrm{e}^{-B} A \mathrm{e}^{B} = A + [A, B] + \tfrac{1}{2!}[[A, B], B] + \cdots$$

and the algebra of the Poincaré group. A similar result can be found for finite transformations. Finally, the action of Lorentz transformation on $\phi(0)$ in (2.9) is realized as a linear transformation acting on the spinorial indices of $\phi(0)$ (which are not explicitly written),

$$\phi'(0) = \Lambda(\omega)\phi(0) = \mathrm{e}^{\omega\cdot\Sigma/2}\phi(0)$$

where $\Sigma_{\mu\nu}$ is a matrix representation of $M_{\mu\nu}$. In this way we find that

$$\phi'(-u) = T(\bar{u})\mathrm{e}^{\omega\cdot\Sigma/2}\phi(0) = \mathrm{e}^{\omega\cdot\Sigma/2}\phi(-\bar{u})$$

since $T(\bar{u})$ commutes with the matrix part and acts directly on $\phi(0)$. Now, if we perform an additional translation $T(-2u)$, and introduce the natural notation $x = u$, the final result takes the form:

$$\phi'(x) = \mathrm{e}^{\omega\cdot\Sigma/2}\phi(\tilde{x}) \qquad \tilde{x}^\mu \equiv x^\mu - \omega^\mu{}_\nu x^\nu. \tag{2.10a}$$

The differential version of this relation reads as

$$\delta_0\phi(x) = \tfrac{1}{2}\omega^{\mu\nu}(L_{\mu\nu} + \Sigma_{\mu\nu})\phi(x) \tag{2.10b}$$

and proves the general form of generator (2.7).

The basic result of this consideration is equation (2.10a), which can be rewritten as $\phi'(x') = S(\omega)\phi(x)$, where $S(\omega) = \exp(\omega\cdot\Sigma/2)$. Such a transformation law is typical for relativistic field theories. The same method can be used to study other groups with a similar basic structure (Bergshoeff 1983, Sohnius 1985).

Invariance of the action and conservation laws

Invariance of a theory under spacetime transformations can be expressed in terms of some restrictions on the Lagrangian that are different from the case of internal symmetries. To show this, consider an action integral defined over a spacetime region Ω,

$$I(\Omega) = \int_\Omega \mathrm{d}^4x\, \mathcal{L}(\phi, \partial_k\phi; x)$$

where we allow for the possibility that $\mathcal{L}$ may depend explicitly on x. The change of $I(\Omega)$ under spacetime transformations $x' = x + \xi(x)$ has the form

$$\delta I = \int_{\Omega'} \mathrm{d}^4x'\, \mathcal{L}'(\phi'(x'), \partial'_k\phi'(x'); x') - \int_\Omega \mathrm{d}^4x\, \mathcal{L}(\phi(x), \partial_k\phi(x); x).$$

Introducing the Jacobian $\partial(x')/\partial(x) \approx 1 + \partial_\mu\xi^\mu$, we see that the action integral is invariant if (Kibble 1961)

$$\Delta\mathcal{L} \equiv \delta_0\mathcal{L} + \xi^\mu\partial_\mu\mathcal{L} + (\partial_\mu\xi^\mu)\mathcal{L} = 0 \tag{2.11}$$

where

$$\begin{aligned}\delta_0\mathcal{L} &\equiv \mathcal{L}(\phi + \delta_0\phi, \partial_k\phi + \delta_0\partial_k\phi; x) - \mathcal{L}(\phi, \partial_k\phi; x)\\ &= \frac{\partial\mathcal{L}}{\partial\phi}\delta_0\phi + \frac{\partial\mathcal{L}}{\partial\phi_{,k}}\delta_0\phi_{,k}.\end{aligned}$$

The Lagrangian $\mathcal{L}$ satisfying the invariance condition (2.11) is usually called an invariant density. The following two remarks will be useful for further applications:

(i) This derivation is based on the assumption that $\delta_0\eta_{\mu\nu} = 0$.
(ii) Condition (2.11) can be relaxed by demanding a weaker condition $\Delta\mathcal{L} = \partial_\mu\Lambda^\mu$; in this case the action changes by a surface term, but the equations of motion remain invariant.

If we now replace the Poincaré expressions for ξ^μ and $\delta_0\phi$ in (2.11), the coefficients multiplying $\omega^{\mu\nu}/2$ and ξ^μ vanish yielding

$$\begin{gathered}\frac{\partial\mathcal{L}}{\partial\phi}\Sigma_{\mu\nu}\phi + \frac{\partial\mathcal{L}}{\partial\phi_{,\rho}}[\Sigma_{\mu\nu}\partial_\rho - (\eta_{\nu\rho}\partial_\mu - \eta_{\mu\rho}\partial_\nu)]\phi = 0\\ \partial_\mu\mathcal{L} - \frac{\partial\mathcal{L}}{\partial\phi}\partial_\mu\phi - \frac{\partial\mathcal{L}}{\partial\phi_{,\nu}}\partial_\mu\phi_{,\nu} = 0.\end{gathered} \tag{2.12}$$

The first identity is the condition of Lorentz invariance, while the second one, related to translational invariance, is equivalent to the absence of any explicit x dependence in $\mathcal{L}$, as we could have expected.

Conserved currents. The expression for $\Delta\mathcal{L}$ can be rewritten as

$$\Delta\mathcal{L} = \frac{\delta\mathcal{L}}{\delta\phi}\delta_0\phi + \partial_\mu J^\mu \qquad J^\mu \equiv \frac{\partial\mathcal{L}}{\partial\phi_{,\mu}}\delta_0\phi + \mathcal{L}\xi^\mu$$

where $\delta\mathcal{L}/\delta\phi \equiv \partial\mathcal{L}/\partial\phi - \partial_\mu(\partial\mathcal{L}/\partial\phi_{,\mu})$. Then, assuming the equations of motion to hold, $\delta\mathcal{L}/\delta\phi = 0$, the invariance condition leads to the differential *conservation law*:

$$\partial_\mu J^\mu = 0 \qquad J^\mu = \tfrac{1}{2}\omega^{\nu\lambda} M^\mu{}_{\nu\lambda} - \varepsilon^\nu T^\mu{}_\nu \tag{2.13}$$

where $T^\mu{}_\nu$ and $M^\mu{}_{\nu\lambda}$ are the *canonical* currents—the energy–momentum and angular momentum tensors, respectively,

$$\begin{aligned} T^\mu{}_\nu &= \frac{\partial\mathcal{L}}{\partial\phi_{,\mu}}\partial_\nu\phi - \delta^\mu_\nu\mathcal{L} \\ M^\mu{}_{\nu\lambda} &= (x_\nu T^\mu{}_\lambda - x_\lambda T^\mu{}_\nu) - S^\mu{}_{\nu\lambda} \end{aligned} \tag{2.14}$$

and $S^\mu{}_{\nu\lambda}$ is the canonical spin tensor,

$$S^\mu{}_{\nu\lambda} = -\frac{\partial\mathcal{L}}{\partial\phi_{,\mu}}\Sigma_{\nu\lambda}\phi.$$

We note that the canonical energy–momentum tensor is, in general, not symmetric.

Since the parameters $\omega^{\nu\lambda}$ and ε^ν are constant, equation (2.13) implies the conservation of energy–momentum and angular momentum currents:

$$\begin{aligned} &\partial_\mu T^\mu{}_\nu = 0 \\ \partial_\mu M^\mu{}_{\nu\lambda} = 0 \qquad &\text{or} \qquad \partial_\mu S^\mu{}_{\nu\lambda} = T_{\nu\lambda} - T_{\lambda\nu}. \end{aligned} \tag{2.15}$$

This is a typical case of Noether's theorem, stating that to each parameter of a continuous symmetry in the Lagrangian, there corresponds a differentially conserved current. The integral over the three-space of the null component of the current defines the related 'charge':

$$\begin{aligned} P^\nu &= \int \mathrm{d}^3x\, T^{0\nu} \\ M^{\nu\lambda} &= \int \mathrm{d}^3x\, M^{0\nu\lambda} = \int \mathrm{d}^3x\, (x^\nu T^{0\lambda} - x^\lambda T^{0\nu} - S^{0\nu\lambda}). \end{aligned} \tag{2.16}$$

It should be stressed that the usual conservation in time of these charges does not hold automatically, but only if the related flux integrals through the boundary of the three-space vanish (this is particularly important in electrodynamics and gravitation).

The canonical currents (2.14) can be defined even when the Lagrangian is not Poincaré invariant, but then, of course, they are not conserved.

Example 2. The canonical energy–momentum tensor for the scalar field, described by the Lagrangian $\mathcal{L}_\mathrm{S} = \frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi + \lambda\varphi^4$, has the form

$$T_{\mu\nu} = \partial_\mu\varphi\partial_\nu\varphi - \eta_{\mu\nu}\mathcal{L}_\mathrm{S}$$

while the spin tensor vanishes.

For the antisymmetrized Dirac Lagrangian, $\mathcal{L}_D = \frac{1}{2}(i\bar{\psi}\gamma^k\overleftrightarrow{\partial}_k\psi - 2m\bar{\psi}\psi)$, with $\overleftrightarrow{\partial}_k \equiv \overrightarrow{\partial}_k - \overleftarrow{\partial}_k$, we find that

$$T^\mu{}_\nu = \tfrac{1}{2}i\bar{\psi}\gamma^\mu\overleftrightarrow{\partial}_\nu\psi - \delta^\mu_\nu\mathcal{L}_D \qquad S^\mu{}_{\nu\lambda} = \tfrac{1}{2}i\varepsilon^\mu{}_{\nu\lambda\sigma}\bar{\psi}\gamma_5\gamma^\sigma\psi.$$

In both cases, we can easily check the conservation laws by making use of the equations of motion.

The Belinfante tensor. The angular momentum $M^{\nu\lambda}$ is composed from two parts: the first one is the integral of moments of $T^{0\lambda}$, the second is the integral of the spin tensor. However, it is possible to modify the energy–momentum tensor in such a way that the value of the four-momentum P^ν remains unchanged, while the angular momentum reduces effectively to the first part only (Treiman *et al* 1972). To show this, we note that the replacement

$$T^{\mu\nu} \to T^{\mu\nu} + \partial_\lambda W^{\lambda\mu\nu}$$

where $W^{\lambda\mu\nu} = -W^{\mu\lambda\nu}$, does not change the value of the integral defining P^μ. Using this ambiguity we can define the Belinfante tensor,

$$T_B^{\mu\nu} = T^{\mu\nu} - \tfrac{1}{2}\partial_\lambda(S^{\mu\nu\lambda} + S^{\nu\mu\lambda} - S^{\lambda\nu\mu}) \tag{2.17}$$

which is *symmetric*: $T_B^{\mu\nu} - T_B^{\nu\mu} = T^{\mu\nu} - T^{\nu\mu} + \partial_\lambda S^{\lambda\nu\mu} = 0$. The integral defining the angular momentum can be now simplified:

$$M^{\nu\lambda} = \int d^3x\,(x^\nu T_B^{0\lambda} - x^\lambda T_B^{0\nu}).$$

The physical significance of the Belinfante tensor in Einstein's GR will be clarified in chapter 3.

Thus, the conservation laws of the energy–momentum and angular momentum currents, i.e. translational and Lorentz invariance, can be expressed by two properties of a *single* object—the Belinfante tensor:

$$\partial_\mu T_B^{\mu\nu} = 0 \qquad T_B^{\mu\nu} = T_B^{\nu\mu}. \tag{2.18}$$

These two relations are equivalent to conditions (2.15).

2.2 Conformal symmetry

Conformal transformations and Weyl rescaling

Conformal coordinate transformations. We have seen that Poincaré transformations can be defined as the coordinate transformations in M_4 that do not change the form of the metric $\eta_{\mu\nu}$. Conformal transformations of coordinates in M_4 are defined by demanding that the form of the metric is changed according to the following simple rule:

$$g'_{\mu\nu}(x) = s(x)\eta_{\mu\nu} \qquad s(x) > 0. \tag{2.19}$$

These transformations leave the light-cone structure ($ds^2 = 0$) and the 'angles' between vectors invariant, and define the group of conformal coordinate transformations $C(1, 3)$.

When $s(x)$ is close to 1, an infinitesimal change in the form of the metric tensor is given by the relation $\delta_0\eta_{\mu\nu} = g'_{\mu\nu}(x) - \eta_{\mu\nu} = [s(x) - 1]\eta_{\mu\nu}$, which, together with equation (2.3), leads to

$$-(\xi_{\mu,\nu} + \xi_{\nu,\mu}) = (s - 1)\eta_{\mu\nu}.$$

This result remains unchanged if we replace M_4 with a D-dimensional Minkowski space M_D. Transition to M_D is useful for discussing the specific properties of the conformal group $C(1, D - 1)$ in $D = 2$. By observing that contraction of the last equation yields $-2(\partial \cdot \xi) = (s - 1)D$, we obtain

$$\xi_{\mu,\nu} + \xi_{\nu,\mu} = \tfrac{2}{D}(\partial \cdot \xi)\eta_{\mu\nu} \tag{2.20a}$$

We shall refer to this equation as the *conformal Killing equation* in M_D. It implies that

$$[\eta_{\mu\nu}\Box + (D - 2)\partial_\mu\partial_\nu]\partial \cdot \xi = 0 \tag{2.20b}$$

where we can clearly see the specific nature of the case $D = 2$, which will be analysed later.

Now going back to $D > 2$, let us look for a solution of (2.20) using a power series expansion for $\xi^\mu(x)$. It follows that the third derivatives of ξ must vanish, so that the general solution is, at most, quadratic in x:

$$\xi^\mu(x) = \varepsilon^\mu + \omega^\mu{}_\nu x^\nu + \rho x^\mu + (c^\mu x^2 - 2c \cdot x x^\mu). \tag{2.21}$$

In $D = 4$ the solution is determined with 15 constant parameters: 10 parameters $(\varepsilon^\mu, \omega^{\mu\nu})$ correspond to *Poincaré transformations*, one parameter ρ defines *dilatations* (or scale transformations), and four parameters c^μ define *special conformal transformations* (SCT). Note that conformal coordinate transformations in M_4 are nonlinear.

Weyl rescaling. There is a transformation group in physics which is similar to $C(1, 3)$ in appearance, but has an essentially different structure. These transformations change the metric according to the rule

$$\eta_{\mu\nu} \to g^r_{\mu\nu}(x) = e^{2\lambda(x)}\eta_{\mu\nu} \tag{2.22a}$$

and are called Weyl or conformal *rescalings*. If the set of dynamical variables also involves some other fields ϕ, Weyl rescaling is extended by

$$\phi(x) \to \phi^r(x) = e^{w\lambda(x)}\phi(x) \tag{2.22b}$$

where w is a real number, called the *weight*, or Weyl dimension, of the field ϕ. The weight of the metric η is, by convention, taken to be $w_\eta = 2$.

Weyl rescaling looks like a $C(1,3)$ transformation in which the coordinate part is neglected. The set of Weyl rescalings defines the Abelian group W_{r}, while the group of conformal transformations $C(1,3)$ is non-Abelian. The collection of all $C(1,3)$ and W_{r} transformations is known as the extended conformal group (Fulton *et al* 1962).

Conformal algebra and finite transformations

In order to find the form of the Lie algebra of $C(1,3)$, we shall start with a field $\varphi(x)$, which is scalar under the transformations (2.21). Then,

$$\delta_0\varphi(x) = -[\omega^\mu{}_\nu x^\nu + \varepsilon^\mu + \rho x^\mu + (c^\mu x^2 - 2c\cdot x x^\mu)]\partial_\mu\varphi(x).$$

After introducing the generators of $C(1,3)$ for a general field $\phi(x)$,

$$\delta_0\phi(x) = (\tfrac{1}{2}\omega^{\mu\nu}M_{\mu\nu} + \varepsilon^\mu P_\mu + \rho D + c^\mu K_\mu)\phi(x) \tag{2.23}$$

we easily obtain their form in the space of scalar fields:

$$\begin{gathered} M_{\mu\nu} = L_{\mu\nu} \qquad P_\mu = -\partial_\mu \\ D = -x\cdot\partial \\ K_\mu = 2x_\mu x\cdot\partial - x^2\partial_\mu. \end{gathered}$$

These generators define the Lie algebra of the conformal group $C(1,3)$:

$$\begin{gathered} [M_{\mu\nu}, M_{\lambda\rho}] = \tfrac{1}{2} f_{\mu\nu,\lambda\rho}{}^{\sigma\tau} M_{\sigma\tau} \\ [M_{\mu\nu}, P_\lambda] = \eta_{\nu\lambda}P_\mu - \eta_{\mu\lambda}P_\nu \qquad [P_\mu, P_\nu] = 0 \\ [M_{\mu\nu}, D] = 0 \qquad [P_\mu, D] = -P_\mu \qquad [D, D] = 0 \\ [M_{\mu\nu}, K_\lambda] = \eta_{\nu\lambda}K_\mu - \eta_{\mu\lambda}K_\nu \\ [P_\mu, K_\nu] = 2(M_{\mu\nu} + \eta_{\mu\nu}D) \\ [D, K_\mu] = -K_\mu \qquad [K_\mu, K_\nu] = 0. \end{gathered} \tag{2.24}$$

The first three commutators define the Poincaré algebra. The generators (M, P, D) define a subalgebra corresponding to the *Weyl group* $W(1,3)$.

In the general case of an arbitrary field ϕ the generators have the form

$$\begin{gathered} M_{\mu\nu} = L_{\mu\nu} + \Sigma_{\mu\nu} \qquad P_\mu = -\partial_\mu \\ D = -x\cdot\partial + \Delta \\ K_\mu = (2x_\mu x\cdot\partial - x^2\partial_\mu) + 2(x^\nu\Sigma_{\mu\nu} - x_\mu\Delta) + \kappa_\mu \end{gathered} \tag{2.25}$$

where $\Sigma_{\mu\nu}$, Δ and κ_μ are matrix representations of $M_{\mu\nu}$, D and K_μ, acting on the components of ϕ. The result can be obtained by the method of induced representations, as in the case of the Poincaré algebra, using

$$\phi'(0) = \exp(\tfrac{1}{2}\omega \cdot \Sigma + \rho\Delta + c \cdot \kappa)\phi(0).$$

If ϕ belongs to an irreducible representation of the Lorentz group, the relation $[\Sigma_{\mu\nu}, \Delta] = 0$ implies, on the basis of Schur's lemma, that Δ is proportional to the unit matrix,

$$\Delta = dI \tag{2.26a}$$

where d is a real number, known as the *scale dimension* of ϕ. Then, from $[\Delta, \kappa_\mu] = -\kappa_\mu$ we find that

$$\kappa_\mu = 0. \tag{2.26b}$$

Scale transformations. Conformal transformations defined by the generators (2.25) refer to a general representation of the conformal group, and may be different from the usual general coordinate transformations (GCT) based on (2.21). Thus, for instance, a scale transformation and the related GCT of $\eta_{\mu\nu}$ are given by

$$\xi^\mu = \rho x^\mu : \quad \delta_0\eta_{\mu\nu} = \rho D\eta_{\mu\nu} = \rho d_\eta \eta_{\mu\nu} \qquad \tilde{\delta}_o\eta_{\mu\nu} = -2\rho\eta_{\mu\nu}$$

where d_η is the scale dimension of η. The usual choice $d_\eta = 0$ yields the natural geometric dimension of η and leads to a very simple scale transformation rule, $\delta_0\eta_{\mu\nu} = 0$, but it differs from $\tilde{\delta}_0\eta$. To 'explain' the difference, consider an extended conformal transformation of η, consisting of a GCT with $\xi^\mu = \rho x^\mu$ and a Weyl rescaling with $\mathrm{e}^{w\rho} \approx 1 + w\rho$:

$$(\tilde{\delta}_0 + \delta_w)\eta_{\mu\nu} = (-2\rho + w_\eta\rho)\eta_{\mu\nu}.$$

In tensor analysis, quantities of this type are called tensor densities of weight w. Hence, a general scale transformation of η, with $d_\eta = 0$, can be interpreted as a GCT of the tensor density, provided $w_\eta = 2$.

The same interpretation may be given to SCTs. Conformal transformations in Riemann spaces will be studied in chapter 4.

Example 3. The dynamics of the free scalar field is determined by the Lagrangian $\mathcal{L}_\mathrm{S} = \frac{1}{2}\eta^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi - \frac{1}{2}m^2\varphi^2$. Under scale transformations,

$$\delta x^\mu = \rho x^\mu \qquad \delta_0\varphi = -\rho(x \cdot \partial - d)\varphi \qquad \delta_0\eta^{\mu\nu} = 0$$

the Lagrangian changes as $\delta_0\mathcal{L}_\mathrm{S} = -\rho[x \cdot \partial\mathcal{L}_\mathrm{S} + (1-d)(\partial\varphi)^2 + dm^2\varphi^2]$. If we adopt $d(\varphi) = -1$, it follows

$$\delta_0\mathcal{L}_\mathrm{S} = -\rho[(x \cdot \partial + 4)\mathcal{L}_\mathrm{S} + m^2\varphi^2].$$

Since the invariance condition (2.11) has the form $\delta_0\mathcal{L} + \rho(x \cdot \partial + 4)\mathcal{L} = 0$, it is fulfilled only for $m = 0$.

Example 4. The free Dirac field is described by the antisymmetrized Lagrangian $\mathcal{L}_\mathrm{D}$ (example 2). If we apply scale transformations,

$$\delta x^\mu = \rho x^\mu \qquad \delta_0\psi = -\rho(x \cdot \partial - d)\psi \qquad \delta_0\eta_{\mu\nu} = 0$$

the change in $\mathcal{L}_\mathrm{D}$ is $\delta_0\mathcal{L}_\mathrm{D} = -\rho[x \cdot \mathcal{L}_\mathrm{D} + (1-2d)\frac{1}{2}\mathrm{i}\bar\psi\gamma \cdot \overleftrightarrow{\partial}\psi + 2dm\bar\psi\psi]$. The choice $d(\psi) = -3/2$ yields

$$\delta_0\mathcal{L}_\mathrm{D} = -\rho[(x \cdot \partial + 4)\mathcal{L}_\mathrm{D} + m\bar\psi\psi]$$

and, again, the action is invariant only for $m = 0$.

Similar arguments show that the theory of the free electromagnetic field is scale invariant if $d(A) = -1$.

The scale dimension d determines the transformation law of dynamical variables (fields) under dilatations, whereas dimensional parameters are left unchanged. Dilatations are different from the transformations of dimensional analysis, which scale not only dynamical variables, but also dimensional parameters. The field $\phi_1 = l^n\phi$ has the same scale dimension as ϕ, although its *natural* dimension is higher for n units.

If scale invariance holds in nature, the relation $\mathrm{e}^{\rho D}P^2\mathrm{e}^{-\rho D} = \mathrm{e}^{2\rho}P^2$ defines the transformation law of masses. This implies that (i) the mass spectrum is either continuous (if $m^2 \neq 0$) or (ii) all the masses vanish. Both possibilities are unacceptable, as they contradict the properties of the physical mass spectrum. Scale invariance is a broken symmetry in a world with non-vanishing, discrete masses.

Finite transformations. Finite elements of $C(1,3)$ in the space of fields, which are connected to identity, have the form

$$G(\omega, a, \rho, c) = \exp(\tfrac{1}{2}\omega \cdot M + a \cdot P + \rho D + c \cdot K) \tag{2.27}$$

where ω, a, ρ and c are finite parameters. We shall now find the form of the finite conformal transformations of coordinates in M_4. Finite Poincaré transformations and dilatations (Weyl group) are easily found from the corresponding infinitesimal expressions:

$$\begin{gathered} T(a)x^\mu = x^\mu + a^\mu \qquad \Lambda(\omega)x^\mu = \Lambda^\mu{}_\nu(\omega)x^\nu \\ D(\rho)x^\mu = \mathrm{e}^\rho x^\mu. \end{gathered} \tag{2.28}$$

To find finite SCT, we first introduce the *inversion*:

$$x'^\mu = Ix^\mu = -x^\mu/x^2 \qquad x^2 \equiv \eta_{\mu\nu}x^\mu x^\nu \neq 0. \tag{2.29}$$

This discrete transformation is also conformal, since $g'_{\mu\nu}(x) = \eta_{\mu\nu}/(x^2)^2$. The importance of inversion is expressed by the following theorem (Dubrovin *et al* 1979):

> *Every smooth conformal transformation of a pseudo-Euclidean (Euclidean) space of dimension $D \geq 3$ can be given as a composition of isometry, dilatation and inversion.*

Now, consider the following composite transformation:

$$K(c)x^\mu = I \cdot T(-c) \cdot Ix^\mu = \frac{x^\mu + c^\mu x^2}{1 + 2c \cdot x + c^2 x^2}. \tag{2.30}$$

$K(c)$ is conformal, as a composition of inversion and translation, and for small c^μ it reduces to the infinitesimal SCT; therefore, $K(c)$ represents the finite SCT.

Nonlinearity. The conformal algebra is isomorphic to that of the group $SO(2,4)$. The latter may be considered as a set of pseudo-orthogonal transformations in a six-dimensional flat space M_6 with the metric $\eta_{ab} = (\eta_{\mu\nu}, -1, 1)$. Since the generators of these transformations, m_{ab}, satisfy the $SO(2,4)$ Lie algebra, the isomorphism is proved by establishing the following correspondence:

$$\begin{aligned} M_{\mu\nu} &\to m_{\mu\nu} \qquad & P_\mu &\to m_{\mu 4} + m_{\mu 5} \\ K_\mu &\to m_{\mu 4} - m_{\mu 5} \qquad & D &\to m_{45}. \end{aligned}$$

The geometric meaning of the nonlinearity of conformal transformations can be clarified by considering the connection between $C(1,3)$ and $SO(2,4)$. The coordinate $SO(2,4)$ transformations in M_6 are *linear*. By projecting these transformations on M_4 we obtain a *nonlinear* realization of the group $SO(2,4)$, that coincides with the action of $C(1,3)$ on M_4. The fact that the orbital part of K_μ can be expressed in terms of P_μ and $L_{\mu\nu}$ (with x dependent coefficients) is a direct consequence of such a specific realization.

Conformal symmetry and conserved currents

Since conformal transformations are consistent with $\delta_0\eta_{\mu\nu} = 0$, the conformal invariance of a theory can be expressed by condition (2.11), where ξ and $\delta_0\phi$ are given by equations (2.21) and (2.23), respectively. For translations and Lorentz transformations $\delta_0\phi$ has the form (2.5), while for dilatations and SCTs we have

$$\begin{aligned} \delta_0^{\mathrm{D}}\phi &= -\rho(x \cdot \partial - d)\phi \\ \delta_0^{\mathrm{K}}\phi &= c^\mu[(2x_\mu x \cdot \partial - x^2\partial_\mu) + 2(x^\nu\Sigma_{\mu\nu} - x_\mu d)]\phi. \end{aligned} \tag{2.31}$$

Vanishing of the coefficients multiplying the Poincaré parameters in (2.11) yields the conditions (2.12); vanishing of the coefficients multiplying ρ and c^μ leads to new conditions (Treiman *et al* 1972):

$$-\frac{\partial\mathcal{L}}{\partial\phi}d\phi+\frac{\partial\mathcal{L}}{\partial\phi_{,\nu}}(-d+1)\phi_{,\nu}-4\mathcal{L}=0 \tag{2.32a}$$

$$2x_\mu\left[-\frac{\partial\mathcal{L}}{\partial\phi}d\phi+\frac{\partial\mathcal{L}}{\partial\phi_{,\nu}}(-d+1)\phi_{,\nu}-4\mathcal{L}\right]+2V_\mu=0 \tag{2.32b}$$

where

$$V_\mu\equiv\frac{\partial\mathcal{L}}{\partial\phi_{,\nu}}(\Sigma_{\mu\nu}-\eta_{\mu\nu}d)\phi. \tag{2.33}$$

In this derivation, conditions (2.12) of Poincaré invariance are used.

If the scale dimension is chosen so that

$$d(\text{fermion field})=-\tfrac{3}{2}\qquad d(\text{boson field})=-1$$

the kinetic energy term of a typical Lagrangian satisfies condition (2.32*a*) for scale invariance. These values for d correspond to the *canonical* dimensions of fields, in units of length, as they ensure that the kinetic part of the action has dimension zero. Scale invariance requires that $d(\mathcal{L})=-4$, i.e. that $\mathcal{L}$ does not contain any dimensional parameters.

Equation (2.32*b*) shows that for conformal invariance *two conditions* must be fulfilled: (i) the theory should be scale invariant; and (ii) the quantity V_μ should be a total divergence,

$$V^\mu=\partial_\lambda\sigma^{\lambda\mu}. \tag{2.34}$$

Indeed, in this case $\delta I=2\int c_\mu V^\mu\,\mathrm{d}^4x$ is a surface term that does not influence the equations of motion. Remarkably, the second condition turns out to be true for all renormalizable field theories (involving spins 0, $\frac{1}{2}$ and 1), although scale invariance is, in general, broken. Consequently, for these theories conformal invariance is equivalent to scale invariance.

The invariance of the theory, in conjunction with the equations of motion, leads to the differential conservation law for the current J^μ, as in (2.13). Using expressions (2.21) and (2.31) for ξ and $\delta_0\phi$, we obtain

$$J^\mu=\tfrac{1}{2}\omega^{\nu\lambda}M^\mu{}_{\nu\lambda}-\varepsilon^\nu T^\mu{}_\nu-\rho D^\mu+c^\nu K^\mu{}_\nu \tag{2.35}$$

where D^μ and $K^\mu{}_\nu$ are the *canonical* dilatation and (special) conformal currents, respectively:

$$D^\mu=x_\nu T^{\mu\nu}-\frac{\partial\mathcal{L}}{\partial\phi_{,\mu}}d\phi \tag{2.36a}$$

$$K^\mu{}_\nu=(2x_\nu x^\lambda-\delta^\lambda_\nu x^2)T^\mu{}_\lambda+2\frac{\partial\mathcal{L}}{\partial\phi_{,\mu}}x^\lambda(\Sigma_{\nu\lambda}-\eta_{\nu\lambda}d)\phi-2\sigma^\mu{}_\nu. \tag{2.36b}$$

With constant parameters, the condition $\partial\cdot J=0$ yields (2.15) and

$$\partial_\mu D^\mu=0\qquad\partial_\mu K^\mu{}_\nu=0. \tag{2.37}$$

The improved energy–momentum tensor. In the previous section we introduced the Belinfante tensor, the properties of which express translation and Lorentz invariance in a very simple way. The conformal current D^μ is similar to expression (2.14) for $M^\mu{}_{\nu\lambda}$, in the sense that it involves a term containing $T_{\lambda\rho}$, and an extra term. The extra term in $M^\mu{}_{\nu\lambda}$ is removed by introducing a new energy–momentum tensor, the Belinfante tensor. Can we use the same idea once again?

The answer is affirmative for a large class of theories. To simplify the exposition, let us consider the case of a scalar field φ,

$$\mathcal{L}_\mathrm{S} = \tfrac{1}{2}\partial_\mu\varphi\partial^\mu\varphi + \lambda\varphi^4. \tag{2.38}$$

Since the canonical and Belinfante energy–momentum tensors are the same, it follows that

$$D^\mu = x_\nu T_\mathrm{B}^{\mu\nu} + \tfrac{1}{2}\partial^\mu\varphi^2.$$

Now, we define

$$\theta^{\mu\nu} = T_\mathrm{B}^{\mu\nu} - \tfrac{1}{6}(\partial^\mu\partial^\nu - \eta^{\mu\nu}\partial^2)\varphi^2. \tag{2.39}$$

The added term is symmetrical and divergenceless, so it does not change any essential property of the Belinfante tensor and makes no contribution to P^μ and $M^{\mu\nu}$. Using $\theta^{\mu\nu}$ instead of the Belinfante tensor in D^μ yields

$$D^\mu = x_\nu\theta^{\mu\nu} + \tfrac{1}{6}\partial_\nu(x^\nu\partial^\mu\varphi^2 - x^\mu\partial^\nu\varphi^2).$$

The last term can be dropped as a divergence of an antisymmetric tensor, leading to

$$D^\mu = x_\nu\theta^{\mu\nu}. \tag{2.40}$$

From this result we easily obtain

$$\partial_\mu D^\mu = \theta^\mu{}_\mu. \tag{2.41}$$

Thus, in analogy with the construction of the Belinfante tensor, we can introduce a new, improved energy–momentum tensor $\theta^{\mu\nu}$, such that the Poincaré and scale symmetry can be expressed in terms of the properties of a *single* tensor (Callan *et al* 1970, Coleman 1973a, b):

$$\partial_\mu\theta^{\mu\nu} = 0 \qquad \theta^{\mu\nu} = \theta^{\nu\mu} \qquad \theta^\mu{}_\mu = 0. \tag{2.42}$$

Next, we shall try to express the conformal current $K^\mu \equiv c_\nu K^{\mu\nu}$ in terms of $\theta^{\mu\nu}$. For the scalar field theory (2.38) we find that

$$V^\mu = \tfrac{1}{2}\partial^\mu\varphi^2 \qquad \sigma^{\mu\nu} = \tfrac{1}{2}\eta^{\mu\nu}\varphi^2.$$

It follows from (2.36) that

$$K^\mu = \xi^K_\lambda T_\mathrm{B}^{\mu\lambda} + (c\cdot x\partial^\mu - c^\mu)\varphi^2.$$

Introducing $\theta^{\mu\nu}$ we find that

$$K^\mu = -\xi^K_\lambda \theta^{\mu\lambda} + \tfrac{1}{6}\partial_\lambda (X^{\lambda\mu} - X^{\mu\lambda})$$

where $X^{\lambda\mu} \equiv -\xi^\lambda_K \partial^\mu \varphi^2 + 2c^\lambda x^\mu \varphi^2$. The last term can be dropped without loss, so that, finally,

$$K^\mu = -\xi^K_\lambda \theta^{\mu\lambda}. \tag{2.43}$$

From this it follows that

$$\partial_\mu K^\mu = 2(\partial \cdot \xi^K)\theta^\mu{}_\mu \tag{2.44}$$

by using the conformal Killing equation for ξ^K. Thus, for scalar field theory conformal invariance reduces to scale invariance.

Several comments are now in order.

(a) In contrast to the Belinfante tensor, $\theta^{\mu\nu}$ *cannot* be constructed for an arbitrary field theory. Indeed, if such a construction were possible, scale invariance would imply conformal invariance, which is not always true. There are examples of field theories which are scale invariant, but not conformally invariant.
(b) The improved energy–momentum tensor can be constructed whenever V^μ is a total divergence. This is true for many field theories, but *not for all*. Explicit computation shows that $V^\mu = 0$ for spin-$\frac{1}{2}$ and spin-1 renormalizable field theories. In these theories $\theta^{\mu\nu}$ has the same form as in (2.39), and expressions (2.40) and (2.43) for D^μ and $K^{\mu\nu}$, respectively, are also unchanged, so that both scale and conformal invariance are measured by $\theta^\mu{}_\mu$.
(c) Although we can modify the energy–momentum tensor without changing the related conservation laws, various definitions may have different dynamical significance. The role of $\theta^{\mu\nu}$ in gravitational theories will be examined in chapter 4.

Conformal transformations in $D = 2$

In two dimensions the conformal Killing equation still has the solutions (2.21), but since now the third derivatives of $\xi(x)$ are not required to vanish, there are infinitely many other solutions, too. As a consequence, the algebra of conformal transformations is *infinite-dimensional* (Ginsparg 1990).

The conformal Killing equation in $D = 2$ has the form

$$\partial_0 \xi^0 = \partial_1 \xi^1 \qquad \partial_0 \xi^1 = \partial_1 \xi^0. \tag{2.45a}$$

Going to the light-cone coordinates, $x^\pm = (x^0 \pm x^1)/\sqrt{2}$, these equations imply

$$\begin{aligned} \partial_+ \xi^- = 0 \quad &\Rightarrow \quad \xi^- = \xi^-(x^-) \\ \partial_- \xi^+ = 0 \quad &\Rightarrow \quad \xi^+ = \xi^+(x^+). \end{aligned}$$

It is now convenient to go over to Euclidean space, $x = x^1$, $y = \mathrm{i}x^0$, and introduce complex coordinates:

$$\begin{aligned} x^+ = \bar{z} = (x - \mathrm{i}y)/\sqrt{2} \qquad & -x^- = z = (x + \mathrm{i}y)/\sqrt{2} \\ \xi^+(x^+) = \bar{\xi}(\bar{z}) \qquad & -\xi^-(x^-) = \xi(z). \end{aligned}$$

Then, the conformal Killing equations become the Cauchy–Riemann equations of complex analysis,

$$\partial_x \xi^x = \partial_y \xi^y \qquad \partial_x \xi^y = -\partial_y \xi^x \tag{2.45b}$$

and the conformal transformations correspond to analytic or anti-analytic mappings:

$$z' = z + \xi(z) \qquad \bar{z}' = \bar{z} + \bar{\xi}(\bar{z}).$$

These mappings conserve the intersection angles between curves in the complex plane.

Every solution $\xi(z)$ that is regular at $z = 0$ can be expanded in a Taylor series,

$$\xi(z) = a_{-1} + a_0 z + a_1 z^2 + \cdots \tag{2.46}$$

and depends on an infinite number of parameters $(a_{-1}, a_0, a_1, \ldots)$. To find the algebra of these transformations we shall consider a scalar analytic function F, $F'(z') = F(z)$. The change of F under the transformations $z' = z + a_n z^{n+1}$, $n \geq -1$, is given by

$$\delta_0 F(z) = a_n L_n F(z) \qquad L_n \equiv -z^{n+1}\partial_z \qquad n \geq -1. \tag{2.47}$$

The generators of anti-analytic transformations are denoted by $\bar{L}_n$. The L_ns and $\bar{L}_n$s satisfy the (classical) *Virasoro algebra*:

$$[L_n, L_m] = (n - m)L_{n+m} \qquad [\bar{L}_n, \bar{L}_m] = (n - m)\bar{L}_{n+m} \tag{2.48}$$

while the L_ns and $\bar{L}_m$s commute†.

The quantity

$$L(z) = \sum_{n\geq -1} a_n L_n = -\sum_{n\geq -1} a_n z^{n+1}\partial_z$$

that generates conformal transformations with parameter (2.48), is regular as $z \to 0$. To investigate the behaviour of $L(z)$ as $z \to \infty$, we perform the change of variables $z = 1/w$:

$$L(z) = -\sum_{n\geq -1} a_n w^{-n-1} \frac{\partial w}{\partial z}\partial_w = \sum_{n\geq -1} a_n w^{1-n}\partial_w.$$

† The Virasoro algebra can be extended by the generators $\{L_n, \bar{L}_n \mid n < -1\}$, which are not regular at $z = 0$. The form of the algebra remains unchanged.

The regularity as $w \to 0$ allows only values $n \leq 1$. Thus, only the conformal transformations generated by $\{L_{\pm 1}, L_0\}$ are well-defined on the whole Riemann sphere $C^2 \cup \{\infty\}$. The generators $\{L_{\pm 1}, L_0\}$ define a subalgebra of the Virasoro algebra:

$$[L_{\pm 1}, L_0] = \pm L_{\pm 1} \qquad [L_{+1}, L_{-1}] = 2L_0.$$

The same is true for $\{\bar{L}_{\pm 1}, \bar{L}_0\}$.

The generators $\{L_{\pm 1}, L_0\} \cup \{\bar{L}_{\pm 1}, \bar{L}_0\}$ define the global conformal group in $D = 2$—the group of conformal transformations that are well-defined and invertible on the whole Riemann sphere. The precise correspondence with the earlier results reads:

$$\begin{aligned} P_0 &= \bar{L}_{-1} - L_{-1} \qquad & M_{01} &= L_0 - \bar{L}_0 \qquad & K_0 &= L_1 - \bar{L}_1 \\ P_1 &= L_{-1} + \bar{L}_{-1} \qquad & D &= L_0 + \bar{L}_0 \qquad & K_1 &= L_1 + \bar{L}_1. \end{aligned} \tag{2.49}$$

The finite form of the global conformal transformations is given by the linear fractional (or Möbius) transformations:

$$L_{\mathrm{F}}: \quad z \to \frac{az+b}{cz+d} \qquad ad - bc = 1.$$

Introducing the $SL(2, C)$ matrices

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad \det M = 1$$

wee can show that L_{F} and $SL(2, C)/Z_2$ are isomorphic mappings (matrices M and $-M$ correspond to the same L_{F} transformation). Conformal transformations (2.28) and (2.30) are expressed by the L_{F} mappings in the following way:

$$\begin{aligned} T:& \quad \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \qquad & \Lambda:& \quad \begin{pmatrix} \mathrm{e}^{\mathrm{i}\omega/2} & 0 \\ 0 & \mathrm{e}^{-\mathrm{i}\omega/2} \end{pmatrix} \\ D:& \quad \begin{pmatrix} \mathrm{e}^{\rho/2} & 0 \\ 0 & \mathrm{e}^{-\rho/2} \end{pmatrix} \qquad & K:& \quad \begin{pmatrix} 1 & 0 \\ -2\bar{c} & 1 \end{pmatrix} \end{aligned}$$

where a and c are complex, and ω and ρ real parameters.

The existence of local conformal transformations, in addition to the global ones, is unique to two dimensions. We should note that only the global conformal transformations define a true group, since the remaining ones do not have inverses on the whole Riemann sphere.

Spontaneously broken scale invariance

The physical properties of classical field theory are determined with respect to the *ground state* (or vacuum), which is defined as the state of lowest energy. In the language of quantum theory, field excitations around the ground state are called particle excitations and their energy, momentum, etc, are defined with respect to the ground state.

It is instructive to investigate a theory involving a set of n real, scalar fields $\varphi = (\varphi^a)$,

$$\mathcal{L} = \sum \tfrac{1}{2}\partial_\mu\varphi^a\partial^\mu\varphi^a - U(\varphi). \tag{2.50}$$

From the form of the Hamiltonian,

$$\mathcal{H} = \sum \tfrac{1}{2}(\partial_\alpha\varphi^a)^2 + U(\varphi)$$

we conclude that the state of lowest energy is one for which φ is a constant, $(\varphi^a)_0 = v^a$, provided the values v^a are the minima of the potential $U(\varphi)$.

There is no reason why an invariance of the Lagrangian, described by a group G, should also be an invariance of the ground state. Physical symmetries depend essentially on the structure of the ground state and can be realized in the following ways (Coleman 1975).

(i) The symmetry may be *manifest*, in which case the ground state has the same symmetry as the Lagrangian and the field components within each irreducible multiplet of G have the same mass. Manifest symmetry may be broken explicitly by adding a non-invariant term to $\mathcal{L}$.

(ii) The symmetry may be *spontaneously broken* (or hidden). This is the case when the ground state is not invariant under G. The masses within G-multiplets are now not equal. If G is a continuous internal symmetry, then, assuming some additional conditions, the mass spectrum is characterized by *Goldstone's theorem*, which tells us that the theory must contain massless bosons, one for each broken infinitesimal symmetry of the ground state. (Gauge theories do not obey those additional conditions.)

Example 5. If the number of scalar fields in (2.50) is $n = 3$, and

$$U(\varphi) = \frac{\lambda}{4!}(\varphi^2 - v^2)^2 \qquad \varphi^2 \equiv \sum(\varphi^a)^2$$

the Lagrangian is invariant under $SO(3)$ rotations, $\varphi^a \to R^a{}_b\varphi^b$. The minima of U lie on the sphere $\varphi^2 = v^2$; it is irrelevant which point on this sphere is chosen as the ground state. Let us choose $v^1 = 0$, $v^2 = 0$, $v^3 = v$. Then, rotations in the planes 1–2 and 1–3 are not symmetries of the ground state, while rotations around the third direction leave the ground state invariant. In order to examine physical consequences, let us define new fields by

$$\varphi^1 = \eta^1 \qquad \varphi^2 = \eta^2 \qquad \varphi^3 = v + \eta^3.$$

When the potential is expressed in terms of the new fields, we easily find that η^1 and η^2 are massless. The number of massless fields is equal to the number of generators that break the symmetry of the ground state.

Now, we shall examine scale invariance from this point of view. Mass terms in the Lagrangian explicitly break scale invariance. Can we interpret these terms as a consequence of spontaneous symmetry breaking?

The essential features of the problem can be seen by considering the massive Dirac field. The mass term can easily be made scale invariant by introducing a massless, scalar field φ, with scale dimension $d(\varphi) = -1$ (transforming as $\varphi'(x') = e^{-\rho}\varphi(x)$ under $x' = e^{\rho}x$). Indeed, the replacement $m\bar{\psi}\psi \to \lambda\bar{\psi}\psi\varphi$, combined with an additional free term for φ, yields the scale-invariant result:

$$\mathcal{L}' = \bar{\psi}\mathrm{i}\gamma\cdot\partial\psi - \lambda\bar{\psi}\psi\varphi + \tfrac{1}{2}\partial_\mu\varphi\partial^\mu\varphi. \qquad (2.51a)$$

Here, both ψ and φ are massless fields. Let us now assume that there exists some mechanism for spontaneous symmetry breaking, such that in the ground state we have $(\varphi)_0 = v$ $(\lambda v = m)$. Then, to describe excitations around the ground state we shall introduce a new field by the exponential parametrization,

$$\varphi = v e^{\sigma/v} = v + \sigma + \cdots$$

which is, in its lowest order, equivalent to $\varphi = v + \sigma$. The new field transforms *inhomogeneously* under the broken symmetry: $\sigma'(x') = \sigma(x) - \rho v$. Expressed in terms of $\sigma(x)$, the Lagrangian $\mathcal{L}'$ takes the form

$$\mathcal{L}' = \bar{\psi}\mathrm{i}\gamma\cdot\partial\psi - \lambda v\bar{\psi}\psi e^{\sigma/v} + \tfrac{1}{2}v^2\partial_\mu e^{\sigma/v}\partial^\mu e^{\sigma/v}. \qquad (2.51b)$$

This Lagrangian has spontaneously broken scale symmetry: the field $\sigma(x)$ is the Goldstone boson, and $\psi(x)$ becomes massive.

It is interesting to note that $\mathcal{L}'$ has not only scale, but also special conformal symmetry. On the other hand, if the ground-state value of φ is $(\varphi)_0 = v$, both scale and special conformal symmetries are spontaneously broken and we have only one Goldstone boson. This unusual situation is a consequence of the nonlinear realization of the conformal group in M_4.

Hidden scale invariance can be explicitly broken by adding a suitable term to the Lagrangian. A choice

$$\mathcal{L}_m = -\tfrac{1}{16}m_\sigma^2 v^2[e^{4\sigma/v} - (1 + 4\sigma/v)] \qquad (2.52)$$

represents a mass term for σ, as is easily seen by Taylor expansion in σ.

Exercises

1. Check that the theory of the free Dirac field is Poincaré invariant.
2. Construct the canonical energy–momentum tensor for scalar field theory, defined by

$$\mathcal{L}_S = \tfrac{1}{2}(\partial\varphi)^2 + gx^2\varphi.$$

Show that it is not a conserved quantity while the angular momentum is conserved.

3. Find the Belinfante tensor for
   (a) the free Dirac field,
   (b) the scalar field with φ^4 interaction and
   (c) the free electromagnetic field.
4. Show that
   (a) if ξ^μ is a solution of the Killing equation in M_4, then $\partial_\nu \partial_\rho \xi^\mu = 0$;
   (b) if ξ^μ satisfies the conformal Killing equation in M_4, then $\partial_\nu \partial_\rho \partial_\lambda \xi^\mu = 0$.
5. Calculate the conformal factor $s(x)$ characterizing the coordinate transformation of the metric $\eta_{\mu\nu}$ under (a) dilatations, (b) SCTs and (c) inversions.
6. Show that by applying the inversion to the right and left of a Lorentz transformation (dilatation), we obtain a Lorentz transformation (dilatation) again.
7. (a) Calculate the action of $I \cdot T(-c) \cdot I$ on x^μ.
   (b) Show that infinitesimal SCTs satisfy the condition $x'^\mu/x'^2 = x^\mu/x^2 + c^\mu$, which is equivalent to $IK(c) = T(-c)I$. Check whether this is true for finite SCTs.
8. Using the conformal Lie algebra prove the relations

$$\begin{aligned}
(1+\rho D)\mathrm{e}^{u\cdot P} &= \mathrm{e}^{(1+\rho)u\cdot P}(1+\rho D) + \mathcal{O}(\rho^2) \\
(1+c\cdot K)\mathrm{e}^{u\cdot P} &= \mathrm{e}^{(u^\mu + c^\mu u^2 - 2c\cdot u u^\mu)P_\mu} \\
&\quad \times (1 + c\cdot K + 2c^\mu u^\nu M_{\mu\nu} - 2c\cdot u D) + \mathcal{O}(c^2).
\end{aligned}$$

   Then, find the general form of the generators D and K in the field space.
9. Consider an extended conformal transformation of η, consisting of the GCT with $\xi^\mu = c^\mu x^2 - 2c\cdot x x^\mu$ and Weyl rescaling with $\mathrm{e}^{-2wc\cdot x} \approx 1 - 2wc\cdot x$. Can this transformation be interpreted as an SCT of η, with $d_\eta = 0$?
10. Consider a six-dimensional space M_6 in which the metric, in coordinates y^a, has the form $g_{ab} = (\eta_{\mu\nu}, -1, 1)$. On the hypercone $y^2 \equiv g_{ab}y^a y^b = 0$ introduce the coordinates

$$x^\mu = y^\mu/k \qquad k \equiv y_4 + y_5.$$

   Show that the action of $SO(2,4)$ on M_6 induces the conformal transformation of coordinates x^μ.
11. Show that the Lagrangian of a scalar field $\mathcal{L}_S(\varphi, \partial\varphi)$, that satisfies the conditions of scale invariance in M_4, has the general form $\mathcal{L}_S = f(\partial_\mu\varphi\partial^\mu\varphi/\varphi^4)\varphi^4$. Find the form of $\mathcal{L}_S$ in the case of conformal invariance.
12. Show that every Lagrangian in M_4 consisting of
   (a) a free Lagrangian for massless fields φ, ψ or A_μ and
   (b) an interaction of the form

$$\begin{gathered}
\varphi^4 \qquad \varphi\bar\psi\psi \qquad \varphi\bar\psi\gamma_5\psi \\
A^\mu\bar\psi\gamma_\mu\psi \qquad A^\mu[\varphi^*(\partial_\mu\varphi) - (\partial_\mu\varphi^*)\varphi] \qquad A^\mu A_\mu\varphi^*\varphi
\end{gathered}$$

defines a scale-invariant theory. Check also the conformal invariance of these Lagrangians.

13. Investigate the scale and conformal invariance of the theory of massless Dirac and scalar fields, with the interaction $\lambda\bar{\psi}\gamma^{\mu}\psi\partial_{\mu}\varphi$.
14. Investigate the scale and conformal invariance of the theory of massless field of spin $\frac{3}{2}$, defined by the Lagrangian

$$\mathcal{L} = -\tfrac{1}{2}\varepsilon^{\mu\nu\lambda\rho}\bar{\psi}_{\mu}\gamma_5\gamma_{\nu}\partial_{\rho}\psi_{\lambda} \qquad \psi_{\mu} = (\psi_{\mu\alpha}).$$

15. Check that the generators P, M, D and K, defined in (2.49), satisfy the conformal Lie algebra.
16. Prove that L_{F} and $SL(2, C)/Z_2$ are isomorphic mappings.
17. Find mappings L_{F} corresponding to the conformal transformations (2.28) and (2.30).
18. Find the divergence of the dilatation current corresponding to the explicit breaking of the dilatation symmetry by the term (2.52).

Chapter 3

Poincaré gauge theory

It is well known that the existence and interaction of certain fields, such as the electromagnetic field, can be closely related to the invariance properties of the theory. Thus, if the Lagrangian of matter fields is invariant under phase transformations with *constant* parameters α, the electromagnetic field can be introduced by demanding invariance under extended, *local* transformations, obtained by replacing α with a function of spacetime points, $\alpha \to \alpha(x)$. This idea was generalized by Yang and Mills (1954) to the case of $SU(2)$ symmetry. Studying these *internal* local symmetries (appendix A) is particularly interesting from the point of view of the generalization to local *spacetime* symmetries.

On the other hand, it is much less known that Einstein's GR is invariant under local Poincaré transformations. This property is based on the principle of equivalence, and gives a rich physical content to the concept of local or gauge symmetry. Instead of thinking of local Poincaré symmetry as derived from the principle of equivalence, the whole idea can be reversed, in accordance with the usual philosophy of gauge theories. When the gravitational field is absent, it has become clear from a host of experiments that the underlying symmetry of fundamental interactions is given by the Poincaré group. If we now want to make a physical theory invariant under local Poincaré transformations, it is necessary to introduce new, *compensating* fields, which, in fact, represent gravitational interactions.

> *Compensating fields cancel all the unwanted effects of local transformations and enable the existence of local symmetries.*

Localization of Poincaré symmetry leads to the Poincaré gauge theory of gravity, which contains GR as a special case (Kibble 1961, Sciama 1962). Here, in contrast to GR, at each point of spacetime there exists a whole class of local inertial frames, mutually related by Lorentz transformations. Using this freedom, allowed by the principle of equivalence, we can naturally introduce not only energy–momentum, but also the *spin* of matter fields into gravitational dynamics.

The exposition of Poincaré gauge theory (PGT) will be followed by the corresponding geometric interpretation, leading to Riemann–Cartan spacetime U_4. Then, we shall study in more detail two simple but important cases: Einstein–Cartan theory, representing a direct generalization of GR, and teleparallel theory. We shall also give a short account of some general dynamical features of the new gravitational theory.

3.1 Poincaré gauge invariance

We shall now analyse the process of transition from global to local Poincaré symmetry and find its relation to the gravitational interaction (Kibble 1961). Other spacetime symmetries (de Sitter, conformal, etc) can be treated in an analogous manner.

Localization of Poincaré symmetry

We assume that the spacetime has the structure of Minkowski space M_4. At each point of M_4, labelled by coordinates x^μ, we can define a local Lorentz reference frame, represented by an orthonormal tetrad—a set of four orthonormal, tangent vectors $\boldsymbol{e}_i(x)$, $\boldsymbol{e}_i \cdot \boldsymbol{e}_j = \eta_{ij}$. In global inertial coordinates x^μ, we can always choose the tetrad such that it coincides with a coordinate frame $\boldsymbol{e}_\mu(x)$ (a set of four vectors, tangent to coordinate lines at x), i.e. $\boldsymbol{e}_i = \delta_i^\mu \boldsymbol{e}_\mu$. Here, the Latin indices $(i, j, \ldots)$ refer to local Lorentz frames, while the Greek indices $(\mu, \nu, \ldots)$ refer to coordinate frames. Later, when we come to more general spaces, this distinction will become geometrically more important.

A matter field $\phi(x)$ in spacetime is always referred to a local Lorentz frame and its transformation law under the Poincaré group is of the form

$$\delta_0\phi = (\tfrac{1}{2}\omega \cdot M + \varepsilon \cdot P)\phi = (\tfrac{1}{2}\omega \cdot \Sigma + \xi \cdot P) \equiv \mathcal{P}\phi. \tag{3.1}$$

The matter Lagrangian $\mathcal{L}_\mathrm{M} = \mathcal{L}_\mathrm{M}(\phi, \partial_k\phi)$ is assumed to be invariant under global Poincaré transformations, which yields the conservation of the energy–momentum and angular momentum tensors. If we now generalize Poincaré transformations by replacing ten *constant* group parameters with some *functions* of spacetime points, the invariance condition (2.11) is violated for two reasons. First, the old transformation rule of $\partial_k\phi$,

$$\delta_0\partial_k\phi = \mathcal{P}\partial_k\phi + \omega_k{}^i\partial_i\phi \equiv \mathcal{P}_k{}^i\partial_i\phi \tag{3.2a}$$

is changed into

$$\delta_0\partial_k\phi = \mathcal{P}\partial_k\phi - \xi^\nu{}_{,k}\partial_\nu\phi + \tfrac{1}{2}\omega^{ij}{}_{,k}\Sigma_{ij}\phi. \tag{3.2b}$$

The second reason is that $\partial_\mu\xi^\mu \neq 0$. Thus, after using the conservation laws (2.15), we obtain

$$\Delta\mathcal{L}_\mathrm{M} = \tfrac{1}{2}\omega^{ij}{}_{,\mu}S^\mu{}_{ij} - (\xi^i{}_{,\mu} - \omega^i{}_\mu)T^\mu{}_i \neq 0.$$

The violation of local invariance can be compensated for by certain modifications of the original theory.

Covariant derivative. Let us first eliminate non-invariance stemming from the change in the transformation rule of $\partial_k\phi$. This can be accomplished by introducing a new Lagrangian

$$\mathcal{L}'_{\mathrm{M}} = \mathcal{L}_{\mathrm{M}}(\phi, \nabla_k\phi) \tag{3.3}$$

where $\nabla_k\phi$ is the *covariant derivative* of ϕ, which transforms according to the 'old rule' (3.2a):

$$\delta_0\nabla_k\phi = \mathcal{P}\nabla_k\phi + \omega_k{}^i\nabla_i\phi. \tag{3.4}$$

The new Lagrangian satisfies $\delta\mathcal{L}'_{\mathrm{M}} \equiv \delta_0\mathcal{L}'_{\mathrm{M}} + \xi\cdot\partial\mathcal{L}'_{\mathrm{M}} = 0$.

To show this, we begin by introducing the ω-covariant derivative, which eliminates the $\omega^{ij}{}_{,\mu}$ term in (3.2b):

$$\nabla_\mu\phi = \partial_\mu\phi + \delta_0(\omega)\phi|_{\omega\to A_\mu} = (\partial_\mu + A_\mu)\phi \qquad A_\mu \equiv \tfrac{1}{2}A^{ij}{}_\mu\Sigma_{ij} \tag{3.5}$$

where $\delta_0(\omega) = \delta_0(\omega, \xi = 0)$ and $A^{ij}{}_\mu$ are compensating fields. The condition

$$\delta_0\nabla_\mu\phi = \mathcal{P}\nabla_\mu\phi - \xi^\nu{}_{,\mu}\nabla_\nu\phi \tag{3.6}$$

determines the transformation properties of A_μ:

$$\begin{aligned} \delta_0 A_\mu &= [\mathcal{P}, A_\mu] - \xi^\nu{}_{,\mu}A_\nu - \omega_{,\mu} \qquad \omega \equiv \tfrac{1}{2}\omega^{ij}\Sigma_{ij} \\ \delta_0 A^{ij}{}_\mu &= \omega^i{}_s A^{sj}{}_\mu + \omega^j{}_s A^{is}{}_\mu - \omega^{ij}{}_{,\mu} - \xi^\lambda{}_{,\mu}A^{ij}{}_\lambda - \xi^\lambda\partial_\lambda A^{ij}{}_\mu. \end{aligned} \tag{3.7}$$

If we now rewrite equation (3.6) as

$$\delta_0\nabla_\mu\phi = \mathcal{P}_\mu{}^\nu\nabla_\nu\phi - (\xi^\nu{}_{,\mu} - \omega^\nu{}_\mu)\nabla_\nu\phi$$

we see that the last term, which is proportional to $\nabla_\nu\phi$, can be eliminated by adding a new compensating field $A^\mu{}_k$:

$$\nabla_k\phi = \delta^\mu_k\nabla_\mu\phi - A^\mu{}_k\nabla_\mu\phi. \tag{3.8a}$$

Since this expression is homogeneous in $\nabla_\mu\phi$, we can introduce the new variables $h_k{}^\mu = \delta^\mu_k - A^\mu{}_k$, and write

$$\nabla_k\phi = h_k{}^\mu\nabla_\mu\phi. \tag{3.8b}$$

The transformation properties of $h_k{}^\mu$ follow from equations (3.4) and (3.6):

$$\delta_0 h_k{}^\mu = \omega_k{}^s h_s{}^\mu + \xi^\mu{}_{,\lambda}h_k{}^\lambda - \xi^\lambda\partial_\lambda h_k{}^\mu. \tag{3.9}$$

Matter field Lagrangian. Up to now we have found the Lagrangian $\mathcal{L}'_{\mathrm{M}}$ such that $\delta\mathcal{L}'_{\mathrm{M}} = 0$. In the second step of restoring local invariance to the theory, we have to take care of the fact that $\partial_\mu\xi^\mu \neq 0$ in (2.11). This is done by multiplying $\mathcal{L}'_{\mathrm{M}}$ by a suitable function of the fields:

$$\widetilde{\mathcal{L}}_{\mathrm{M}} = \Lambda\mathcal{L}'_{\mathrm{M}}. \tag{3.10}$$

This expression satisfies the invariance condition (2.11) provided that

$$\delta_0\Lambda + \partial_\mu(\xi^\mu\Lambda) = 0.$$

Here, coefficients of the local parameters and their derivatives should vanish. Coefficients of the second derivatives of parameters vanish since Λ does not depend on field derivatives, coefficients of ξ^ν vanish since Λ does not depend explicitly on x, while those of $\omega^{ij}{}_{,\nu}$ vanish if Λ is independent of $A^{ij}{}_\nu$. The remaining two conditions are:

$$\begin{aligned}
\xi^\mu{}_{,\nu}:&\quad \frac{\partial\Lambda}{\partial h_k{}^\mu}h_k{}^\nu + \delta^\nu_\mu\Lambda = 0\\
\omega^{ij}:&\quad \eta_{ik}\frac{\partial\Lambda}{\partial h_k{}^\mu}h_j{}^\mu - \eta_{jk}\frac{\partial\Lambda}{\partial h_k{}^\mu}h_i{}^\mu = 0.
\end{aligned}$$

After multiplying the first equation by $b^s{}_\nu$, the inverse of $h_k{}^\mu$,

$$b^k{}_\mu h_k{}^\nu = \delta^\nu_\mu \qquad b^k{}_\mu h_s{}^\mu = \delta^k_s$$

we easily find a solution for Λ:

$$\Lambda = [\det(h_k{}^\mu)]^{-1} = \det(b^k{}_\mu) \equiv b. \tag{3.11}$$

The solution is defined up to a multiplicative factor, which is chosen so that $\Lambda \to 1$ when $h_k{}^\mu \to \delta^\mu_k$, and the second equation is automatically satisfied.

The final form of the modified Lagrangian for matter fields is

$$\widetilde{\mathcal{L}}_{\mathrm{M}} = b\mathcal{L}_{\mathrm{M}}(\phi, \nabla_k\phi). \tag{3.12}$$

It is obtained from the original Lagrangian $\mathcal{L}_{\mathrm{M}}(\phi, \partial_k\phi)$ in two steps:

- by replacing $\partial_k\phi \to \nabla_k\phi$ (the minimal coupling) and
- multiplying $\mathcal{L}_{\mathrm{M}}$ by b.

The Lagrangian $\widetilde{\mathcal{L}}_{\mathrm{M}}$ satisfies the invariance condition (2.11) by construction, hence it is an invariant density.

Example 1. The free scalar field in Minkowski space is given by the Lagrangian $\mathcal{L}_{\mathrm{S}} = \frac{1}{2}(\eta^{kl}\partial_k\varphi\partial_l\varphi - m^2\varphi^2)$. Localization of Poincaré symmetry leads to

$$\widetilde{\mathcal{L}}_{\mathrm{S}} = \tfrac{1}{2}b(\eta^{kl}\nabla_k\varphi\nabla_l\varphi - m^2\varphi^2) = \tfrac{1}{2}b(g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi - m^2\varphi^2)$$

where $g^{\mu\nu} = \eta^{kl} h_k{}^\mu h_l{}^\nu$ and $\nabla_k \varphi = h_k{}^\mu \partial_\mu \varphi$.

Similarly, the Dirac Lagrangian $\mathcal{L}_\mathrm{D} = \frac{1}{2}(\mathrm{i}\bar\psi\gamma^k \overleftrightarrow{\partial}_k \psi - 2m\bar\psi\psi)$ becomes

$$\widetilde{\mathcal{L}}_\mathrm{D} = \tfrac{1}{2} b(\mathrm{i}\bar\psi\gamma^k \overleftrightarrow{\nabla}_k \psi - 2m\bar\psi\psi) \qquad \overleftrightarrow{\nabla}_k \equiv \overrightarrow{\nabla}_k - \overleftarrow{\nabla}_k$$

where $\nabla_k \psi = h_k{}^\mu(\partial_\mu + A_\mu)\psi$, $\nabla_k \bar\psi = \bar\psi(\overleftarrow{\partial}_\mu - A_\mu)h_k{}^\mu$, $A_\mu \equiv \frac{1}{2} A^{ij}{}_\mu \sigma_{ij}$.

Comments. After localizing the Poincaré transformations, we could think that the Lorentz part can be absorbed into $\xi^\mu(x)$, and lose its independence. However, field transformation law (3.1) shows that the Lorentz part preserves its independence.

The form of transformation law (3.1) motivates us to ask whether we can treat ξ^μ and ω^{ij}, instead of ε^μ and $\omega^{\mu\nu}$, as *independent* parameters. In the case of global Poincaré transformations this is not possible; for instance, the transformation defined by $\xi^\mu = 0$, $\omega^{ij} \neq 0$ is not an allowed one. When the symmetry is localized, the answer to this question is affirmative, due to the transformation laws of compensating fields.

When we go from global to local Poincaré group, the translation generator $P_i = -\partial_i$ should be replaced with $\tilde{P}_i = -\nabla_i$. The group of local Lorentz rotations and local translations has a more general structure than the original Poincaré group, as can be seen from the commutation relations of the new generators.

In order to have a clear geometric interpretation of the local transformations, it is convenient to generalize our previous convention concerning the use of Latin and Greek indices. According to transformation rules (3.7) and (3.9), the use of indices in $A^{ij}{}_\mu$ and $h_k{}^\mu$ is in agreement with this convention: these fields transform as local Lorentz tensors with respect to Latin indices and as world (coordinate) tensors with respect to Greek indices. We can also check that local Lorentz tensors can be transformed into world tensors and vice versa, by multiplication with $h_k{}^\mu$ or $b^k{}_\mu$. The term $-\omega^{ij}{}_{,\mu}$ in (3.7) shows that $A^{ij}{}_\mu$ is a potential (an analogous term appears in $\delta_0 A^\mu{}_k$).

The explicit form of the spin matrix Σ_{ij}, appearing in $\nabla_\mu\phi$, depends only on the Lorentz transformation properties of ϕ. It is, therefore, natural to extend the ω-covariant derivative ∇_μ to any quantity transforming linearly under Lorentz transformations, ignoring its ξ transformations. Thus, for instance,

$$\begin{aligned}
\nabla_\mu h_k{}^\nu &= [\partial_\mu + \tfrac{1}{2} A^{ij}{}_\mu(\Sigma^1_{ij})]_k{}^s h_s{}^\nu = \partial_\mu h_k{}^\nu - A^s{}_{k\mu} h_s{}^\nu \\
\nabla_\mu H^{kl} &= [\partial_\mu + \tfrac{1}{2} A^{ij}{}_\mu(\Sigma^2_{ij})]^{kl}{}_{rs} H^{rs} = \partial_\mu H^{kl} + A^k{}_{s\mu} H^{sl} + A^l{}_{s\mu} H^{ks}
\end{aligned}$$

where Σ^1_{ij} and Σ^2_{ij} are the vector and tensor representations of Σ_{ij}:

$$\begin{aligned}
(\Sigma^1_{ij})^k{}_s &= (\delta^k_i \eta_{js} - \delta^k_j \eta_{is}) \\
(\Sigma^2_{ij})^{kl}{}_{rs} &= (\Sigma^1_{ij})^k{}_r \delta^l_s + (\Sigma^1_{ij})^l{}_s \delta^k_r .
\end{aligned}$$

Field strengths. We succeeded in modifying the original matter Lagrangian by introducing gauge potentials, so that the invariance condition (2.11) also remains true for local Poincaré transformations. In order to construct a free Lagrangian for the new fields $h_k{}^\mu$ and $A^{ij}{}_\mu$, we shall first introduce the corresponding *field strengths*. Let us calculate the commutator of two ω-covariant derivatives:

$$\begin{aligned} [\nabla_\mu, \nabla_\nu]\phi &= (\partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu])\phi = \tfrac{1}{2} F^{ij}{}_{\mu\nu}\Sigma_{ij}\phi \\ F^{ij}{}_{\mu\nu} &= \partial_\mu A^{ij}{}_\nu - \partial_\nu A^{ij}{}_\mu + A^i{}_{s\mu}A^{sj}{}_\nu - A^i{}_{s\nu}A^{sj}{}_\mu . \end{aligned} \tag{3.13}$$

The Lorentz field strength $F^{ij}{}_{\mu\nu}$ transforms as a tensor, in conformity with its index structure:

$$\delta_0 F^{ij}{}_{\mu\nu} = \omega^i{}_s F^{sj}{}_{\mu\nu} + \omega^j{}_s F^{is}{}_{\mu\nu} + \xi^\rho{}_{,\mu} F^{ij}{}_{\rho\nu} + \xi^\rho{}_{,\nu} F^{ij}{}_{\mu\rho} - \xi^\lambda \partial_\lambda F^{ij}{}_{\mu\nu}.$$

Since $\nabla_k\phi = h_k{}^\mu\nabla_\mu\phi$, the commutator of two ∇_k-covariant derivatives will differ from (3.13) by an additional term containing the derivatives of $h_k{}^\mu$:

$$[\nabla_k, \nabla_l]\phi = \tfrac{1}{2} F^{ij}{}_{kl}\Sigma_{ij}\phi - F^s{}_{kl}\nabla_s\phi \tag{3.14a}$$

where

$$\begin{aligned} F^{ij}{}_{kl} &= h_k{}^\mu h_l{}^\nu F^{ij}{}_{\mu\nu} \\ F^s{}_{kl} &= h_k{}^\mu h_l{}^\nu(\nabla_\mu b^s{}_\nu - \nabla_\nu b^s{}_\mu) \equiv h_k{}^\mu h_l{}^\nu F^s{}_{\mu\nu}. \end{aligned} \tag{3.14b}$$

The quantity $F^i{}_{\mu\nu}$ is called the translation field strength. Jacobi identities for the commutators of covariant derivatives imply the following *Bianchi identities*:

$$\begin{aligned} &\text{(first)} \qquad \varepsilon^{\rho\mu\lambda\nu}\nabla_\mu F^s{}_{\lambda\nu} = \varepsilon^{\rho\mu\lambda\nu} F^s{}_{k\lambda\nu} b^k{}_\mu \\ &\text{(second)} \qquad \varepsilon^{\rho\lambda\mu\nu}\nabla_\lambda F^{ij}{}_{\mu\nu} = 0. \end{aligned}$$

The free Lagrangian must be an invariant density depending only on the Lorentz and translation field strengths, so that the complete Lagrangian of matter and gauge fields has the form

$$\widetilde{\mathcal{L}} = b\mathcal{L}_\mathrm{F}(F^{ij}{}_{kl}, F^i{}_{kl}) + b\mathcal{L}_\mathrm{M}(\phi, \nabla_k\phi). \tag{3.15}$$

Conservation laws and field equations

The invariance of the Lagrangian in a gauge theory for an internal symmetry leads, after using the equations of motion, to covariantly generalized differential conservation laws. The same thing also happens in PGT.

1. Let us denote the set of field variables by $Q_A = (\phi, b^k{}_\mu, A^{ij}{}_\mu)$. The invariance condition (2.11) can be written as

$$\Delta\mathcal{L} = \frac{\delta\mathcal{L}}{\delta Q_A}\delta_0 Q_A + \partial_\mu J^\mu = 0 \qquad J^\mu \equiv \frac{\partial\mathcal{L}}{\partial Q_{A,\mu}}\delta_0 Q_A + \mathcal{L}\xi^\mu$$

where $\delta_0 Q_A$ is determined by equations (2.5), (3.7b), and the relation

$$\delta_0 b^k{}_\mu = \omega^k{}_s b^s{}_\mu - \xi^\lambda{}_{,\mu} b^k{}_\lambda - \xi^\lambda \partial_\lambda b^k{}_\mu \tag{3.16}$$

which follows from (3.9). Those terms in $\Delta\mathcal{L}$ that contain derivatives of parameters ω^{ij} and ξ^μ can be transformed so that these derivatives appear only in a form of four-divergences:

$$\begin{aligned}
\frac{\delta\mathcal{L}}{\delta b^k{}_\mu}(-\xi^\nu{}_{,\mu} b^k{}_\nu) &= \xi^\nu \partial_\mu \left(\frac{\delta\mathcal{L}}{\delta b^k{}_\mu} b^k{}_\nu\right) - \partial_\mu \left(\frac{\delta\mathcal{L}}{\delta b^k{}_\mu} b^k{}_\nu \xi^\nu\right) \\
\frac{\delta\mathcal{L}}{\delta A^{ij}{}_\mu}(-\omega^{ij}{}_{,\mu} - \xi^\nu{}_{,\mu} A^{ij}{}_\nu) &= \xi^\nu \partial_\mu \left(\frac{\delta\mathcal{L}}{\delta A^{ij}{}_\mu} A^{ij}{}_\nu\right) + \omega^{ij} \partial_\mu \left(\frac{\delta\mathcal{L}}{\delta A^{ij}{}_\mu}\right) \\
&\quad - \partial_\mu \left[\frac{\delta\mathcal{L}}{\delta A^{ij}{}_\mu}(\omega^{ij} + \xi^\nu A^{ij}{}_\nu)\right].
\end{aligned}$$

After that, the invariance condition takes the form

$$\Delta\mathcal{L} = -\xi^\nu I_\nu + \tfrac{1}{2}\omega^{ij} I_{ij} + \partial_\mu \Lambda^\mu = 0. \tag{3.17}$$

Integrating this expression over a four-dimensional domain Ω, and imposing the requirement that the parameters ξ and ω, together with their first derivatives, vanish on the boundary of Ω, we obtain $-\xi^\nu I_\nu + \frac{1}{2}\omega^{ij} I_{ij} = 0$, so that $\Delta\mathcal{L} = 0$ reduces to

$$\partial_\mu \Lambda^\mu = 0. \tag{3.18a}$$

Since, further, ξ and ω are arbitrary within Ω, it follows that

$$I_\nu = 0 \qquad I_{ij} = 0. \tag{3.18b}$$

These identities completely determine the content of the differential conservation laws.

2. In what follows we shall restrict our discussion to the case $\mathcal{L} = \tilde{\mathcal{L}}_\mathrm{M}$. Let us first introduce *canonical* and *covariant* energy–momentum and spin currents for matter fields,

$$\widetilde{T}^\mu{}_\nu = \frac{\partial\widetilde{\mathcal{L}}_\mathrm{M}}{\partial\phi_{,\mu}}\phi_{,\nu} - \delta^\mu_\nu \widetilde{\mathcal{L}}_\mathrm{M} \qquad \widetilde{S}^\mu{}_{ij} = -\frac{\partial\widetilde{\mathcal{L}}_\mathrm{M}}{\partial\phi_{,\mu}}\Sigma_{ij}\phi \tag{3.19}$$

$$\widetilde{T}'^\mu{}_\nu = \frac{\partial\widetilde{\mathcal{L}}_\mathrm{M}}{\partial\nabla_\mu\phi}\nabla_\nu\phi - \delta^\mu_\nu \widetilde{\mathcal{L}}_\mathrm{M} \qquad \widetilde{S}'^\mu{}_{ij} = -\frac{\partial\widetilde{\mathcal{L}}_\mathrm{M}}{\partial\nabla_\mu\phi}\Sigma_{ij}\phi \tag{3.20}$$

as well as the corresponding *dynamical* quantities:

$$\tau^\mu{}_\nu = h_k{}^\mu \frac{\delta\widetilde{\mathcal{L}}_\mathrm{M}}{\delta h_k{}^\nu} = -\frac{\delta\widetilde{\mathcal{L}}_\mathrm{M}}{\delta b^k{}_\mu} b^k{}_\nu \qquad \sigma^\mu{}_{ij} = -\frac{\delta\widetilde{\mathcal{L}}_\mathrm{M}}{\delta A^{ij}{}_\mu}. \tag{3.21}$$

We assume that the equations of motion for matter fields are satisfied, $\delta\widetilde{\mathcal{L}}_M/\delta\phi = 0$. Demanding that the coefficients of $\partial\xi$ and $\partial\omega$ in (3.18b) vanish, we obtain the equality of the covariant and dynamical currents:

$$\tau^\mu{}_\nu = \widetilde{T}'^\mu{}_\nu \qquad \sigma^\mu{}_{ij} = \widetilde{S}'^\mu{}_{ij}. \tag{3.22}$$

Conditions (3.18a) yield covariantly generalized conservation laws for the energy–momentum and angular momentum currents:

$$\begin{aligned} b^k{}_\mu \nabla_\nu \tau^\nu{}_k &= \tau^\nu{}_k F^k{}_{\mu\nu} + \tfrac{1}{2}\sigma^\nu{}_{ij} F^{ij}{}_{\mu\nu} \\ \nabla_\mu \sigma^\mu{}_{ij} &= \tau_{ij} - \tau_{ji}. \end{aligned} \tag{3.23}$$

Similar analysis can be applied to the complete Lagrangian (3.15).

3. The equations of motion for matter fields, obtained from (3.15), have the covariant form:

$$\delta\phi : \quad \frac{\bar\partial\widetilde{\mathcal{L}}_M}{\partial\phi} - \nabla_\mu \frac{\partial\widetilde{\mathcal{L}}_M}{\partial\nabla_\mu\phi} = 0 \tag{3.24a}$$

where $\bar\partial\widetilde{\mathcal{L}}_M/\partial\phi = [\partial\widetilde{\mathcal{L}}_M(\phi, \nabla_k u)/\partial\phi]_{u=\phi}$.

Example 2. The equation of motion for a free scalar field has the form

$$\delta\varphi : \quad b^{-1}\partial_\mu(bg^{\mu\nu}\partial_\nu\varphi) + m^2\varphi = 0.$$

Since $\nabla_\mu = \partial_\mu$ in the space of the Lorentz scalars (no Latin indices), this equation is obviously covariant. The canonical currents,

$$\widetilde{T}^\mu{}_\nu = b\partial_\mu\varphi\partial_\nu\varphi - \delta^\mu_\nu\widetilde{\mathcal{L}}_S \qquad \widetilde{S}^\mu{}_{ij} = 0$$

are equal to the corresponding covariant and dynamical ones.

For the free Dirac Lagrangian $\tilde{\mathcal{L}}_D$, the equations of motion for $\bar\psi$ are

$$\delta\bar\psi : \quad (\mathrm{i}\gamma^k\nabla_k + \mathrm{i}\gamma^k V_k - m)\psi = 0$$

where $2V_k = b^{-1}\partial_\mu(bh_k{}^\mu) + A_{ks}{}^s = b^{-1}\nabla_\mu(bh_k{}^\mu)$, and similarly for ψ. These equations have the covariant form (3.24a). The canonical and covariant currents are given as

$$\begin{aligned} \widetilde{T}^\mu{}_\nu &= \tfrac{1}{2}\mathrm{i}b\bar\psi\gamma^\mu\overleftrightarrow{\partial}_\nu\psi - \delta^\mu_\nu\widetilde{\mathcal{L}}_D \qquad \widetilde{S}^\mu{}_{ij} = \tfrac{1}{2}\mathrm{i}bh^{k\mu}\varepsilon_{kijs}\bar\psi\gamma_5\gamma^s\psi \\ \widetilde{T}'^\mu{}_\nu &= \tfrac{1}{2}\mathrm{i}b\bar\psi\gamma^\mu\overleftrightarrow{\nabla}_\nu\psi - \delta^\mu_\nu\widetilde{\mathcal{L}}_D \qquad \widetilde{S}'^\mu{}_{ij} = \widetilde{S}^\mu{}_{ij} \end{aligned}$$

while $\tau^\mu{}_\nu = \widetilde{T}'^\mu{}_\nu$ and $\sigma^\mu{}_{ij} = \widetilde{S}'^\mu{}_{ij}$, as expected.

In order to write down the equations of motion for gauge fields, it is useful to introduce the notation

$$f^{\mu}{}_{k} = -\frac{\bar{\partial}\widetilde{\mathcal{L}}_{\mathrm{F}}}{\partial b^{k}{}_{\mu}} \qquad \pi_{k}{}^{\mu\nu} = \frac{\partial\widetilde{\mathcal{L}}_{\mathrm{F}}}{\partial b^{k}{}_{\mu,\nu}} = -\frac{\partial\widetilde{\mathcal{L}}_{\mathrm{F}}}{\partial F^{k}{}_{\mu\nu}}$$
$$f^{\mu}{}_{ij} = -\frac{\bar{\partial}\widetilde{\mathcal{L}}_{\mathrm{F}}}{\partial A^{ij}{}_{\mu}} \qquad \pi_{ij}{}^{\mu\nu} = \frac{\partial\widetilde{\mathcal{L}}_{\mathrm{F}}}{\partial A^{ij}{}_{\mu,\nu}} = -\frac{\partial\widetilde{\mathcal{L}}_{\mathrm{F}}}{\partial F^{ij}{}_{\mu\nu}}$$

where $\bar{\partial}\widetilde{\mathcal{L}}_{\mathrm{F}}/\partial b^{k}{}_{\mu}$ denotes the partial derivative of $\widetilde{\mathcal{L}}_{\mathrm{F}} = b\mathcal{L}_{\mathrm{F}}(F^{i}{}_{kl}, F^{ij}{}_{kl})$, calculated by keeping $F^{i}{}_{\lambda\rho}, F^{ij}{}_{\lambda\rho} = \text{constant}$,

$$\frac{\partial\widetilde{\mathcal{L}}_{\mathrm{F}}}{\partial b^{k}{}_{\mu}} = \frac{\bar{\partial}\widetilde{\mathcal{L}}_{\mathrm{F}}}{\partial b^{k}{}_{\mu}} + \frac{1}{2}\frac{\partial\widetilde{\mathcal{L}}_{\mathrm{F}}}{\partial F^{i}{}_{\lambda\rho}}\frac{\partial F^{i}{}_{\lambda\rho}}{\partial b^{k}{}_{\mu}}$$

and similarly for $\bar{\partial}\widetilde{\mathcal{L}}_{\mathrm{F}}/\partial A^{kl}{}_{\mu}$. After that we find that

$$\begin{aligned} \delta b^{k}{}_{\mu}: &\quad -\nabla_{\nu}(\pi_{k}{}^{\mu\nu}) - f^{\mu}{}_{k} = \tau^{\mu}{}_{k} \\ \delta A^{ij}{}_{\mu}: &\quad -\nabla_{\nu}(\pi_{ij}{}^{\mu\nu}) - f^{\mu}{}_{ij} = \sigma^{\mu}{}_{ij}. \end{aligned} \tag{3.24b}$$

Example 3. Let us consider the gauge field Lagrangian $\widetilde{\mathcal{L}}_{\mathrm{F}} = -abF$, where a is a constant, and $F \equiv h_{i}{}^{\mu}h_{j}{}^{\nu}F^{ij}{}_{\mu\nu}$ (Einstein–Cartan theory). Here, $\pi_{k}{}^{\mu\nu} = f^{\mu}{}_{ij} = 0$, and the equations of motion have the form:

$$2ab(F^{\mu}{}_{k} - \tfrac{1}{2}h_{k}{}^{\mu}F) = \tau^{\mu}{}_{k} \qquad -2a\nabla_{\nu}(H_{ij}^{\mu\nu}) = \sigma^{\mu}{}_{ij}$$

where $H_{ij}^{\mu\nu} \equiv b(h_{i}{}^{\mu}h_{j}{}^{\nu} - h_{j}{}^{\mu}h_{i}{}^{\nu})$.

On the equivalence of different approaches

The idea of local Poincaré symmetry was proposed and developed in the papers by Utiyama (1956) and Kibble (1961) (see also Sciama 1962). Utiyama introduced the fields $A^{ij}{}_{\mu}$ by localizing the Lorentz symmetry, while the $h_{k}{}^{\mu}$ were treated as given functions of x (although at a later stage these functions were regarded as dynamical variables). This rather unsatisfactory procedure was substantially improved by Kibble, who showed that the quantities $h_{k}{}^{\mu}$, just like $A^{ij}{}_{\mu}$, can be introduced as compensating fields if we consider the localization of the full Poincaré group.

To compare our approach with Kibble's, we note that he started by considering Minkowski spacetime, in which infinitesimal transformations (2.4) induce the total field variation $\delta\phi = \frac{1}{2}\omega^{ij}\Sigma_{ij}\phi$, which is equivalent to (2.5). Also, $\xi^{\mu}(x)$ and $\omega^{ij}(x)$ are treated from the very beginning as independent gauge parameters, which makes the geometric interpretation of the theory more direct. The use of the form variation $\delta_0\phi$ instead of the total variation $\delta\phi$ brings our approach closer to the spirit of gauge theories of internal symmetries, since $\delta_0\phi$

realizes the representation of the Poincaré group in the space of fields ϕ. However, the transformation properties of the gauge fields, as well as the form of the covariant derivative, are independent of these details, so that the final structure is the same.

In order to preserve the geometric meaning of translation, the original Poincaré generator $P_\mu = -\partial_\mu$ should be replaced by the covariant derivative $-\nabla_\mu$. If we demand that the matter field transformation $\delta_0\phi$ has the same geometric meaning after localization of the symmetry as it had before, the expression for $\delta_0\phi$ should be changed to

$$\delta_0^*\phi = (\tfrac{1}{2}\omega^{ij}\Sigma_{ij} - \xi^\nu\nabla_\nu)\phi \tag{3.25}$$

(Hehl *et al* 1976). It follows from the relation $\delta_0^*\phi = \delta_0\phi - \frac{1}{2}\xi^\nu A^{ij}{}_\nu\Sigma_{ij}\phi$ that $\delta_0^*\phi$ and $\delta_0\phi$ differ by a local Lorentz rotation with parameter $\Delta\omega^{ij} = -\xi^\nu A^{ij}{}_\nu$. Therefore, invariance under $\delta_0\phi$ implies invariance under $\delta^*\phi$ and vice versa and the two approaches are equivalent.

All three formulations are thus seen to be essentially equivalent approaches to the localization of Poincaré symmetry.

3.2 Geometric interpretation

Up to this point, we have not given any geometric interpretation to the new fields $h_k{}^\mu$ and $A^{ij}{}_\mu$. Such an interpretation is possible and useful, and leads to a new understanding of gravity.

Riemann–Cartan space U_4

In this subsection we give a short exposition of Riemann–Cartan geometry in order to be able to understand the geometric meaning of PGT (Hehl *et al* 1976, Choquet-Bruhat *et al* 1977; see also appendix B).

The differentiable manifold. Spacetime is often described as a 'four-dimensional continuum', since any event in spacetime can be determined by four real numbers: (t, x, y, z). In SR, spacetime has the structure of Minkowski space M_4. In GR, spacetime can be divided into 'small, flat parts' in which SR holds (on the basis of the principle of equivalence), and these pieces are 'sewn together' smoothly. Although spacetime looks like M_4 locally, it may have quite different global properties.

This picture of spacetime can be compared with a two-dimensional surface which is 'smooth', so that a small piece of it, in the neighbourhood of a given point, can be approximated by the tangent plane at that point. On the other hand, the whole surface may be quite different from the plane. Mathematical generalization of a 'smooth' surface leads to the concept of a differentiable manifold.

It is natural to start with the very simple but abstract concept of topological space, which allows a precise formulation of the idea of continuity. A *topological space* consists of a set of points X, and a given collection τ of subsets of X, which are defined as being open. The collection τ defines a topology on X if (a) τ contains the empty set and the whole X and (b) it is closed with respect to the operations of arbitrary union and finite intersection of subsets.

Example 4. A simple example of a topological space is obtained by taking X to be the set of real numbers $\mathcal{R}$, and defining τ as a collection of all subsets of $\mathcal{R}$ which can be expressed as unions of open intervals (a, b), and the empty set. If infinite intersections were allowed in the definition of τ, these sets would not define a topology on $\mathcal{R}$. Indeed, each point in $\mathcal{R}$ can be represented as an infinite intersection of open intervals: for instance, $a = \cap_{n=1}^{\infty}(a - 1/n, a + 1/n)$. However, a point is not a union of open intervals, therefore it is not open. This example represents a historical prototype of a topological space, and explains the terminology 'open set' in the context of abstract topological spaces.

The structure of topological spaces allows the natural introduction of neighbourhoods and continuous mappings. Of special importance for topological spaces are *homeomorphic* mappings f (f is a bijection, f and f^{-1} are continuous).

A topological space X has the structure of a *manifold* if every point in X has a neighbourhood that 'looks like' an open subset of $\mathcal{R}^n$, i.e. if there exists a homeomorphism φ_i of a neighbourhood O_i of a point P in X into an open subset Ω_i of $\mathcal{R}^n$. This mapping defines a *local coordinate system* on X, since the image of P is a set of n real numbers $(x_i^\mu) \equiv (x_i^1, x_i^2, \ldots, x_i^n)$, representing the coordinates of P. The collection of all local coordinate systems is called simply a coordinate system. Since every piece of X 'looks like' $\mathcal{R}^n$, the number n is called the dimension of X.

If a point $P \in X$ lies in the overlap $O_i \cap O_j$, its image in $\mathcal{R}^n$ is given in two local coordinate systems. These two systems are *compatible* if the transition from one to the other, defined by $\varphi_{ij} = \varphi_j \circ \varphi_i^{-1} : (x_i^\mu) \to (x_j^\mu)$, is a smooth function. A manifold is *differentiable* if its local coordinate systems are compatible (figure 3.1).

Example 5. The sphere $S_2 = \{(x, y, z)|x^2 + y^2 + z^2 = 1\}$ has the structure of a differentiable manifold. To see that, let us introduce a neighbourhood O_1 of an arbitrary point $P_1 \in S_2$, and a neighbourhood O_2 of the antipodal point P_2, such that O_1 does not contain P_2 , O_2 does not contain P_1, and $O_1 \cup O_2 = S_2$. Then, we can introduce two local coordinate systems $(\varphi_i, \Omega_i, i = 1, 2)$, where φ_1 is the stereographic projection of O_1 from the point P_1 to the Euclidean plane Ω_1, tangent to the sphere at the point P_2, and similarly for (φ_2, Ω_2).

Since each mapping of a differentiable manifold can be realized in terms of coordinates, we can easily introduce differentiable and smooth mappings.

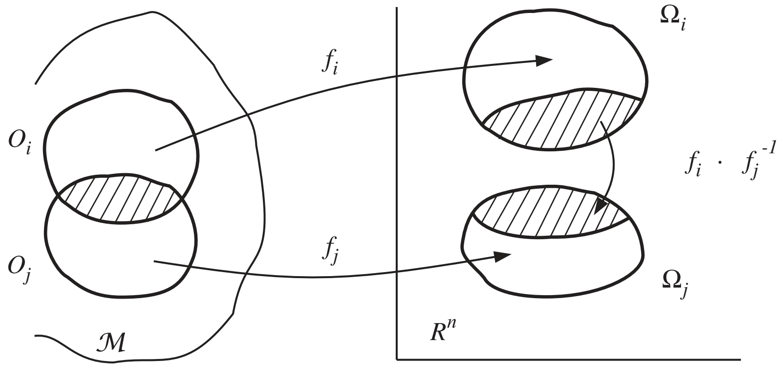

Figure 3.1. Change of coordinates on a differentiable manifold.

Of special importance for differentiable manifolds are *diffeomorphisms* (f is a bijection, f and f^{-1} are smooth).

Tangent space. A differentiable manifold looks like an empty space 'waiting' for something to happen. We believe that the laws of physics must be expressible as relationships between geometric objects. Important objects of this type are vectors and tensors which appear in physical theories as dynamical variables.

In flat space we can think of vectors as finite displacements or arrows extending between two points. This is true, for instance, in Euclidean space E_3, but on a curved, two-dimensional surface the arrow definition breaks down. However, by considering infinitesimal displacements, we can introduce a new structure for the vector space, determined by all tangent vectors, lying in the tangent plane. This structure can be generalized to any n-dimensional manifold X, if we imagine X to be embedded in a higher-dimensional flat space. But such a treatment suggests, falsely, that tangent vectors depend on the embedding. It is most acceptable to define tangent vectors as *directional derivatives* without any reference to embedding.

The set of all tangent vectors at P defines the *tangent space* T_P. The set of vectors tangent to the coordinate lines x^μ defines the coordinate basis $\boldsymbol{e}_\mu = \partial_\mu$ in T_P. The components of a vector $\boldsymbol{v}$ in this basis, defined by $\boldsymbol{v} = v^\mu \boldsymbol{e}_\mu$, transform under the change of coordinates $x \mapsto x'$ according to

$$v'^\mu = \frac{\partial x'^\mu}{\partial x^\nu} v^\nu. \tag{3.26a}$$

Vectors $\boldsymbol{v} = (v^\mu)$ are usually called contravariant vectors.

To each tangent space T_P we can associate the space T_P^* of *dual* vectors (covariant vectors, covectors or 1-forms). The components of a dual vector $\boldsymbol{v}^*$ in the basis $\boldsymbol{\theta}^\mu$, dual to $\boldsymbol{e}_\mu$ ($\boldsymbol{\theta}^\mu(\boldsymbol{e}_\nu) = \delta^\mu_\nu$), transform as

$$v^{*\prime}_\mu = \frac{\partial x^\nu}{\partial x'^\mu} v^*_\nu. \tag{3.26b}$$

It is usual to omit the sign * for dual vectors when using the standard index notation in which components of vectors have upper indices and components of dual vectors—lower indices.

A *tensor* of type (p, q) is characterized by components which transform, under a change of coordinates, in the same way as the product of p vectors and q dual vectors. A tensor t^μ of type $(1, 0)$ is a vector, while a tensor t_μ of type $(0, 1)$ is a dual vector.

Vector and tensor *fields* are important objects in physics. A vector field is a mapping which associates each point in X with a tangent vector, $x \mapsto \boldsymbol{v}(x)$. We can define a tensor field in a similar way.

Parallel transport. We assume that spacetime has the structure of a differentiable manifold X_4. On X_4 we can define differentiable mappings, tensors and various algebraic operations with tensors at a given point (addition, multiplication, contraction). However, comparing tensors at different points requires some additional structure on X_4. Consider, for instance, a vector $\boldsymbol{A}_x$ lying in the tangent space T_x, and a vector $\boldsymbol{A}_{x+\mathrm{d}x}$ lying in $T_{x+\mathrm{d}x}$. In order to compare these two vectors, it is necessary, first, to 'transport' $\boldsymbol{A}_x$ from T_x to $T_{x+\mathrm{d}x}$ and then to compare the resulting object $\boldsymbol{A}_{\mathrm{PT}}$ with $\boldsymbol{A}_{x+\mathrm{d}x}$. This 'transport' procedure generalizes the concept of parallel transport in flat space and bears the same name. The components of $\boldsymbol{A}_{\mathrm{PT}}$ with respect to the coordinate basis in $T_{x+\mathrm{d}x}$ are defined as

$$A^\mu_{\mathrm{PT}}(x + \mathrm{d}x) = A^\mu(x) + \delta A^\mu \qquad \delta A^\mu = -\Gamma^\mu_{\lambda\rho} A^\lambda \,\mathrm{d}x^\rho$$

where the infinitesimal change δA^μ is bilinear in A^μ and $\mathrm{d}x^\rho$. The set of 64 coefficients $\Gamma^\mu_{\lambda\rho}$ defines a linear (or affine) *connection* on X_4. An X_4 equipped with Γ is called a linearly connected space, $L_4 = (X_4, \Gamma)$. If $\boldsymbol{A}_{x+\mathrm{d}x} = \boldsymbol{A}_{\mathrm{PT}}$, we say that $\boldsymbol{A}_{x+\mathrm{d}x}$ and $\boldsymbol{A}_x$ are parallel. Their difference can be expressed in terms of the components as (figure 3.2):

$$\begin{aligned} DA^\mu &= A^\mu(x + \mathrm{d}x) - A^\mu_{\mathrm{PT}}(x + \mathrm{d}x) = \mathrm{d}A^\mu - \delta A^\mu \\ &= (\partial_\rho A^\mu + \Gamma^\mu_{\lambda\rho} A^\lambda)\,\mathrm{d}x^\rho \equiv D_\rho(\Gamma) A^\mu \,\mathrm{d}x^\rho. \end{aligned} \tag{3.27}$$

The expression DA^μ is called the *covariant derivative* of the vector field A^μ, and it represents a generalization of the partial derivative. By convention, the last index in $\Gamma^\mu_{\lambda\rho}$ is the same as the differentiation index.

Parallel transport of a dual vector field A_μ is determined by demanding $\delta(A_\mu B^\mu) = 0$, which implies $\delta A_\mu = \Gamma^\nu_{\mu\lambda} A_\nu \,\mathrm{d}x^\lambda$, i.e.

$$D_\rho(\Gamma) A_\mu = \partial_\rho A_\mu - \Gamma^\nu_{\mu\rho} A_\nu. \tag{3.28}$$

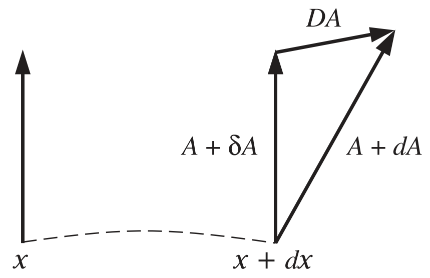

Figure 3.2. Comparison of vectors at different points of X_4 with the help of parallel transport.

The covariant derivative of an arbitrary tensor field is defined as the mapping of a tensor field of type (p, q) into a tensor field of type $(p, q+1)$, with the following properties:

- linearity: $D(\alpha \boldsymbol{t} + \beta \boldsymbol{s}) = \alpha D\boldsymbol{t} + \beta D\boldsymbol{s}$;
- Leibniz rule: $D(\boldsymbol{ts}) = (D\boldsymbol{t})\boldsymbol{s} + \boldsymbol{t}(D\boldsymbol{s})$;
- $Df = \mathrm{d}f$, if f is a scalar function;
- D commutes with contraction.

Torsion and curvature. Using the fact that $D_\rho(\Gamma)A^\mu$ is a tensor, we can derive the transformation law for $\Gamma^\mu_{\lambda\rho}$, which is not a tensor. However, the antisymmetric part of the linear connection,

$$T^\mu{}_{\lambda\rho} = \Gamma^\mu_{\rho\lambda} - \Gamma^\mu_{\lambda\rho} \tag{3.29}$$

transforms as a tensor and is called the *torsion* tensor.

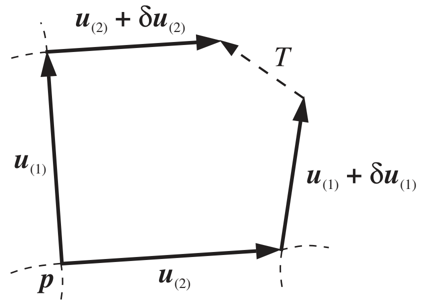

Figure 3.3. Torsion closes the gap in infinitesimal parallelograms.

In order to see the geometric meaning of the torsion, consider two infinitesimal vectors at P, $\boldsymbol{u}_{(1)}$ and $\boldsymbol{u}_{(2)}$, tangent to coordinate lines $x^1(\lambda)$ and $x^2(\lambda)$, respectively (figure 3.3). A change in $\boldsymbol{u}_{(1)}$ under infinitesimal parallel

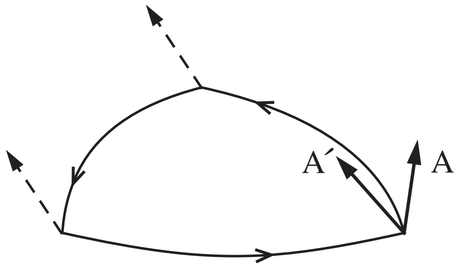

Figure 3.4. Parallel transport depends on path.

transport along $x^2(\lambda)$ has the form $\delta_2 u^\mu_{(1)} = -\Gamma^\mu_{\lambda 2} u^\lambda_{(1)}\,\mathrm{d}x^2$, while a similar change of $\boldsymbol{u}_{(2)}$ along $x^1(\lambda)$ is $\delta_1 u^\mu_{(2)} = -\Gamma^\mu_{\lambda 1} u^\lambda_{(2)}\,\mathrm{d}x^1$. The figure built by the vectors $\boldsymbol{u}_{(1)}, \boldsymbol{u}_{(2)}$, and their parallel transported images, is not a closed parallelogram:

$$\delta_1 u^\mu_{(2)} - \delta_2 u^\mu_{(1)} = -(\Gamma^\mu_{21} - \Gamma^\mu_{12})\,\mathrm{d}x^1\,\mathrm{d}x^2 = T^\mu{}_{21}\,\mathrm{d}x^1\,\mathrm{d}x^2.$$

Parallel transport is a path-dependent concept (figure 3.4). Indeed, if we parallel transport A_ν around an infinitesimal closed path, the result is

$$\Delta A_\nu = \oint \Gamma^\mu_{\nu\rho} A_\mu\,\mathrm{d}x^\rho = \tfrac{1}{2} R^\mu{}_{\nu\lambda\rho} A_\mu \Delta\sigma^{\lambda\rho}$$

where $\Delta\sigma^{\lambda\rho}$ is a surface bounded by the path, and $R^\mu{}_{\nu\lambda\rho}$ is the *Riemann curvature* tensor:

$$R^\mu{}_{\nu\lambda\rho} = \partial_\lambda \Gamma^\mu_{\nu\rho} - \partial_\rho \Gamma^\mu_{\nu\lambda} + \Gamma^\mu_{\sigma\lambda}\Gamma^\sigma_{\nu\rho} - \Gamma^\mu_{\sigma\rho}\Gamma^\sigma_{\nu\lambda}. \tag{3.30}$$

The torsion and curvature tensors in L_4 obey some algebraic and also some differential (Bianchi) identities.

The metric. We assume that on X_4 we can define a *metric tensor* field $\boldsymbol{g}$ as a symmetric, non-degenerate tensor field of type $(0, 2)$. After that, we can introduce the scalar product of two tangent vectors

$$\boldsymbol{u} \cdot \boldsymbol{v} \equiv \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = g_{\mu\nu} u^\mu v^\nu$$

where $g_{\mu\nu} = \boldsymbol{g}(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu)$, and calculate the lengths of curves, angles between vectors, etc. Thus, e.g., the square of the infinitesimal distance between two points x and $x + \mathrm{d}x$ on a curve $C(\lambda)$ is determined by

$$\mathrm{d}s^2 = \boldsymbol{g}(\boldsymbol{t}\,\mathrm{d}\lambda, \boldsymbol{t}\,\mathrm{d}\lambda) = g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu$$

where $\boldsymbol{t}$ is the tangent on $C(\lambda)$ at x. With the help of a metric we can introduce a natural isomorphism between vectors and dual vectors which, in components, has the form

$$u_\mu = g_{\mu\nu} u^\nu \qquad u^\mu = g^{\mu\nu} u_\nu$$

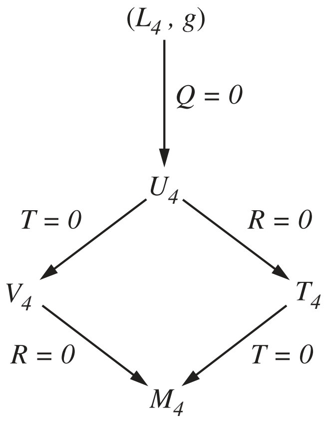

Figure 3.5. Classification of spaces satisfying the metricity condition.

where $g^{\mu\nu}$ is the inverse metric tensor. We also assume that the metric of spacetime has the signature $(+, -, -, -)$.

> *The linear connection and the metric are independent geometric objects.*

After introducing the linear connection and the metric, the differentiable manifold X_4 becomes an (L_4, g) space.

In order to preserve lengths and angles under parallel transport in (L_4, g), we can introduce the *metricity condition*:

$$-Q_{\mu\nu\lambda} \equiv D_\mu(\Gamma)g_{\nu\lambda} = \partial_\mu g_{\nu\lambda} - \Gamma^\rho_{\nu\mu} g_{\rho\lambda} - \Gamma^\rho_{\lambda\mu} g_{\nu\rho} = 0 \tag{3.31}$$

where $Q_{\mu\nu\lambda}$ is called the non-metricity tensor. This requirement establishes a local Minkowskian structure on X_4, and defines a metric-compatible linear connection:

$$\begin{aligned} \Gamma^\mu_{\lambda\nu} &= \left\{ {\mu \atop \lambda\nu} \right\} + K^\mu{}_{\lambda\nu} \\ \left\{ {\mu \atop \lambda\nu} \right\} &\equiv \tfrac{1}{2} g^{\mu\rho}(g_{\nu\rho,\lambda} + g_{\lambda\rho,\nu} - g_{\lambda\nu,\rho}) \end{aligned} \tag{3.32a}$$

where $\left\{ {\mu \atop \lambda\nu} \right\}$ is the Christoffel connection and $K^\mu{}_{\lambda\nu}$ the contortion tensor,

$$K^\mu{}_{\lambda\nu} = -\tfrac{1}{2}(T^\mu{}_{\lambda\nu} - T_\nu{}^\mu{}_\lambda + T_{\lambda\nu}{}^\mu \tag{3.32b}$$

with 24 independent components.

A space (L_4, g) with the most general metric-compatible linear connection is called the *Riemann–Cartan space* U_4. If the torsion vanishes, U_4 becomes the *Riemann space* V_4 of GR; if, alternatively, the curvature vanishes, U_4 becomes Weitzenböck's *teleparallel* space T_4. Finally, the condition $R^\mu{}_{\nu\lambda\rho} = 0$ transforms V_4 into Minkowski space M_4 and $T^\mu{}_{\lambda\rho} = 0$ transforms T_4 into M_4 (figure 3.5).

A curve $x(\lambda)$ in U_4 is an *autoparallel* curve (or affine geodesic) if its tangent vector is parallel to itself. Its equation

$$\frac{d^2x^\mu}{d\lambda^2} + \Gamma^\mu_{\rho\nu}\frac{dx^\rho}{d\lambda}\frac{dx^\nu}{d\lambda} = 0 \tag{3.33}$$

contains only the symmetric part of the connection, $\Gamma^\mu_{(\rho\nu)} = \left\{{\mu \atop \rho\nu}\right\} - T_{(\rho\nu)}{}^\mu$.

An *extremal* curve in U_4 is a curve with an extremal length determined by the condition $\delta \int ds = 0$. Its equation has the same form as that in GR,

$$\frac{d^2x^\mu}{d\lambda^2} + \left\{{\mu \atop \rho\nu}\right\}\frac{dx^\rho}{d\lambda}\frac{dx^\nu}{d\lambda} = 0. \tag{3.34}$$

In a Riemann space autoparallel and extremal curves coincide and are known as geodesic lines.

Example 6. On the unit sphere S_2 the metric in spherical coordinates $(x^1, x^2) = (\theta, \varphi)$ is given by

$$ds^2 = d\theta^2 + \sin^2\theta\, d\varphi^2.$$

The vector $\boldsymbol{A}$ at $(\varphi, \theta) = (0, \pi/4)$ is equal to $\boldsymbol{e}_\theta$. What is $\boldsymbol{A}$ after it has been parallel transported along the circle $\varphi = 0$ to the point $(0, \pi/2)$? Calculating the Christoffel symbols,

$$\left\{{1 \atop 22}\right\} = -\sin\theta\cos\theta \qquad \left\{{2 \atop 12}\right\} = \left\{{2 \atop 21}\right\} = \cot\theta$$

we find that the equations of parallel transport $D_1 A^\mu = 0$ have the form

$$\frac{\partial A^1}{\partial\theta} = 0 \qquad \frac{\partial A^2}{\partial\theta} + \cot\theta A^2 = 0.$$

Solving these equations we obtain $A^1 = C_1$, $A^2 = C_2/\sin\theta$. At $\theta = \pi/2$, $(A^1, A^2) = (1, 0)$, so that $C_1 = 1$, $C_2 = 0$; therefore, after parallel transport the vector is $(A^1, A^2) = (1, 0)$, as we might have expected from the direct geometric picture.

Spin connection. The choice of a basis in a tangent space T_P is not unique. A coordinate frame (C-frame) is determined by a set of four vectors $\boldsymbol{e}_\mu(x)$, tangent to the coordinate lines. In U_4 we can also introduce an orthonormal, Lorentz frame (L-frame), determined by four vectors $\boldsymbol{e}_i(x)$ (vierbein or tetrad), such that

$$\boldsymbol{g}(\boldsymbol{e}_i, \boldsymbol{e}_j) = \eta_{ij}. \tag{3.35}$$

The existence of L-frames is closely related to the principle of equivalence, as we shall see. Every tangent vector $\boldsymbol{u}$ can be expressed in both frames:

$$\boldsymbol{u} = u^\mu \boldsymbol{e}_\mu = u^i \boldsymbol{e}_i.$$

In particular, C- and L-frames can be expressed in terms of each other,

$$\boldsymbol{e}_\mu = e^i{}_\mu \boldsymbol{e}_i \qquad \boldsymbol{e}_i = e_i{}^\mu \boldsymbol{e}_\mu. \tag{3.36}$$

The scalar product of two tangent vectors is defined in terms of the metric (in C-frames we use $g_{\mu\nu}$, and in L-frames η_{ij}):

$$\boldsymbol{u} \cdot \boldsymbol{v} = g_{\mu\nu} u^\mu v^\nu = \eta_{ij} u^i v^j.$$

Consequently,

$$\begin{aligned} \eta_{ij} &= \boldsymbol{e}_i \cdot \boldsymbol{e}_j = g_{\mu\nu} e_i{}^\mu e_j{}^\nu \\ g_{\mu\nu} &= \boldsymbol{e}_\mu \cdot \boldsymbol{e}_\nu = \eta_{ij} e^i{}_\mu e^j{}_\nu. \end{aligned} \tag{3.37}$$

The relation between the coordinate and tetrad components of an arbitrary tangent vector $\boldsymbol{u}$ reads:

$$u^i = e^i{}_\mu u^\mu \qquad u^\mu = e_i{}^\mu u^i. \tag{3.38}$$

Using this, we can associate with each L-frame a local Lorentz coordinate system determined by $\mathrm{d}x^i = e^i{}_\mu \, \mathrm{d}x^\mu$, such that

$$\mathrm{d}s^2 = g_{\mu\nu} \, \mathrm{d}x^\mu \, \mathrm{d}x^\nu = \eta_{ij} \, \mathrm{d}x^i \, \mathrm{d}x^j.$$

The existence of L-frames leads to very important consequences. In particular, they are used to introduce finite spinors into U_4 theory.

If we want to compare vectors $u^i(x)$ and $u^i(x + \mathrm{d}x)$ at points x and $x + \mathrm{d}x$, determined with respect to the L-frames $\boldsymbol{e}_i(x)$ and $\boldsymbol{e}_i(x + \mathrm{d}x)$, respectively, we have to know the rule of parallel transport:

$$\delta u^i = -\omega^i{}_{j\mu} u^j \, \mathrm{d}x^\mu \tag{3.39a}$$

where $\omega^{ij}{}_\mu$ is the so-called *spin connection*, with 64 components. Parallel transport of v_i is determined by requiring $\delta(u^i v_i) = 0$,

$$\delta v_i = \omega^j{}_{i\mu} v_j \, \mathrm{d}x^\mu. \tag{3.39b}$$

The existence of L-frames at each point of X_4 implies the existence of the Lorentz metric at each point of X_4. Demanding that the tensor field η_{ij} be invariant under parallel transport,

$$\delta\eta_{ij} = (\omega^s{}_{i\mu}\eta_{sj} + \omega^s{}_{j\mu}\eta_{si}) \, \mathrm{d}x^\mu = 0 \tag{3.40a}$$

implies that the connection is *antisymmetric*:

$$\omega^{ij}{}_\mu + \omega^{ji}{}_\mu = 0. \tag{3.40b}$$

Having established rule (3.39) of parallel transport, we can define the related ω-covariant derivatives of u^i and v_i:

$$\begin{aligned} Du^i &= (\partial_\mu u^i + \omega^i{}_{s\mu} u^s) \, \mathrm{d}x^\mu \equiv D_\mu(\omega) u^i \, \mathrm{d}x^\mu \\ Dv_i &= (\partial_\mu v_i - \omega^s{}_{i\mu} v_s) \, \mathrm{d}x^\mu \equiv D_\mu(\omega) v_i \, \mathrm{d}x^\mu. \end{aligned} \tag{3.41}$$

Since η is a constant tensor, the condition $\delta\eta = 0$ yields

$$D_\mu(\omega)\eta_{ij} = 0. \tag{3.42}$$

Relation between ω and Γ. Up to now, we have not assumed any relation between the spin connection and Γ. It is natural to demand that the tetrad components of a vector $u^\mu(x)$, parallel transported from x to $x + \mathrm{d}x$, be equal to $u^i + \delta u^i$,

$$u^i + \delta u^i = e^i{}_\mu(x + \mathrm{d}x)(u^\mu + \delta u^\mu)$$

since parallel transport is a unique geometric operation, independent of the choice of frame. In other words, ω and Γ represent one and the same geometric object in two different frames. From this, we obtain the relation

$$D_\mu(\omega + \Gamma)e^i{}_\nu \equiv \partial_\mu e^i{}_\nu + \omega^i{}_{s\mu} e^s{}_\nu - \Gamma^\lambda_{\nu\mu} e^i{}_\lambda = 0 \tag{3.43}$$

which explicitly connects ω and Γ. The operation $D_\mu(\omega + \Gamma)$ can be formally understood as a total covariant derivative. Using equations (3.42) and (3.43) we easily obtain the metricity condition:

$$D_\mu(\Gamma)g_{\nu\lambda} = D_\mu(\omega + \Gamma)g_{\nu\lambda} = D_\mu(\omega + \Gamma)(\eta_{ij} e^i{}_\nu e^j{}_\lambda) = 0.$$

The ω-covariant derivative (3.41) can be generalized to a quantity ϕ belonging to an arbitrary representation of the Lorentz group:

$$D_\mu(\omega)\phi = (\partial_\mu + \omega_\mu)\phi \qquad \omega_\mu \equiv \tfrac{1}{2}\omega^{ij}{}_\mu \Sigma_{ij}. \tag{3.44}$$

The antisymmetry of $\omega^{ij}{}_\mu$ is here clearly connected to the antisymmetry of the spin generator Σ_{ij}. If the spin connection were not antisymmetric, the metricity condition would not be fulfilled and the geometry would be more general than U_4.

Example 7. The relation between ω and Γ implies:

(a) $D_\mu(\omega)v^i = b^i{}_\nu D_\mu(\Gamma)v^\nu$,
(b) $D_i(\omega)v^i = D_\mu(\Gamma)v^\mu = b^{-1}\partial_\mu(bv^\mu)$,
(c) $\Box\varphi \equiv \eta^{ij} D_i(\omega)D_j(\omega)\varphi = g^{\mu\nu} D_\mu(\Gamma)D_\nu(\Gamma)\varphi = b^{-1}\partial_\mu(bg^{\mu\nu}\partial_\nu\varphi)$.

Equation (a) follows from $D_\mu(\omega)v^i = D_\mu(\omega + \Gamma)(b^i{}_\nu v^\nu) = b^i{}_\nu D_\mu(\Gamma)v^\nu$. To prove (b) we use $\Gamma^\rho_{\rho\mu} = \left\{ {\rho \atop \rho\mu} \right\} + K^\rho{}_{\rho\mu} = \left\{ {\rho \atop \rho\mu} \right\} = b^{-1}\partial_\mu b$. Finally, (c) follows from the relation $g^{\mu\nu} D_\mu(\Gamma)D_\nu(\Gamma)\varphi = D_\mu(\Gamma)(g^{\mu\nu}\partial_\nu\varphi)$, in conjunction with (b).

Let us observe some consequences of (3.43). If we find $\Gamma = \Gamma(\omega)$ from (3.43) and replace it in expressions (3.29) and (3.30) for the torsion and curvature, the result is

$$\begin{aligned} T^\mu{}_{\nu\lambda}(\Gamma) &= e_i{}^\mu F^i{}_{\nu\lambda}(\omega) \\ R^\mu{}_{\nu\lambda\rho}(\Gamma) &= e_i{}^\mu e_{j\nu} F^{ij}{}_{\lambda\rho}(\omega) \end{aligned} \tag{3.45}$$

where $F^i{}_{\nu\lambda}$ and $F^{ij}{}_{\nu\lambda}$ are the translation and Lorentz field strengths. On the other hand, after antisymmetrizing equation (3.43) over μ and ν,

$$\begin{gathered} c^i{}_{\mu\nu} + \omega^i{}_{s\mu} e^s{}_\nu - \omega^i{}_{s\nu} e^s{}_\mu = T^\lambda{}_{\mu\nu} e^i{}_\lambda \\ c^i{}_{\mu\nu} \equiv \partial_\mu e^i{}_\nu - \partial_\nu e^i{}_\mu \end{gathered}$$

we find the following solution for ω:

$$\begin{aligned}\omega_{ij\mu} &= \Delta_{ij\mu} + K_{ij\mu} \\ \Delta_{ij\mu} &\equiv \tfrac{1}{2}(c_{ijm} - c_{mij} + c_{jmi})e^m{}_\mu\end{aligned} \tag{3.46}$$

where Δ are the Ricci rotation coefficients and K the contorsion.

Geometric and gauge structure of PGT

The final result from the analysis of PGT is the construction of the invariant Lagrangian (3.15). This is achieved by introducing new fields $b^i{}_\mu$ (or $h_k{}^\nu$) and $A^{ij}{}_\mu$, which are used to construct the covariant derivative $\nabla_k = h_k{}^\nu\nabla_\nu$ and the field strengths $F^{ij}{}_{\mu\nu}$ and $F^i{}_{\mu\nu}$. This theory can be thought of as a field theory in Minkowski space. However, geometric analogies are so strong that it would be unnatural to ignore them.

- Gauge field $A^{ij}{}_\mu$ can be identified with the spin connection $\omega^{ij}{}_\mu$. Indeed, from the definition of $\nabla_\mu\phi$ it follows that
  - the quantity $\nabla_\mu\phi$ has one additional dual index, compared to ϕ; and
  - ∇_μ acts linearly, obeys the Leibniz rule, commutes with contraction and $\nabla_\mu f = \partial_\mu f$ if f is a scalar function.

  Therefore, ∇_μ can be identified with the geometric covariant derivative D_μ. Then, by comparing $\nabla_\mu(A)$ and $D_\mu(\omega)$ in equations (3.5) and (3.44), we can conclude that $A^{ij}{}_\mu$ can be identified with $\omega^{ij}{}_\mu$.
- The field $b^i{}_\mu$ can be identified with the tetrad components on the basis of its transformation law. This ensures the possibility of transforming local Lorentz and coordinate indices into each other.
- The local Lorentz symmetry of PGT implies the metricity condition.

Consequently,

PGT has the geometric structure of the Riemann–Cartan space U_4.

It is not difficult to conclude, by comparing equation (3.45) with (3.13) and (3.14), that the translation field strength $F^i{}_{\mu\nu}$ is nothing but the torsion $T^\lambda{}_{\mu\nu}$, while the Lorentz field strength $F^{ij}{}_{\mu\nu}$ represents the curvature $R^\lambda{}_{\tau\mu\nu}$. Thus, PGT is a specific approach to the theory of gravity in which both mass and spin are sources of the gravitational field.

It is an intriguing fact that PGT does not have the structure of an 'ordinary' gauge theory (McDowell and Mansouri 1977, Regge 1986, Bañados *et al* 1996). To clarify this point, we start from the Poincaré generators P_a, M_{ab} satisfying the Lie algebra (2.6), and define the gauge potential as $A_\mu = e^a{}_\mu P_a + \frac{1}{2}\omega^{ab}{}_\mu M_{ab}$. The infinitesimal gauge transformation

$$\delta_0 A_\mu = -\nabla_\mu\lambda = -\partial_\mu\lambda - [A_\mu, \lambda]$$

where $\lambda = \lambda^a P_a + \frac{1}{2}\lambda^{ab} M_{ab}$, has the following component content:

$$\begin{aligned} \text{Translations:}\quad & \delta_0 e^a{}_\mu = -\nabla'_\mu \lambda^a \qquad & \delta_0 \omega^{ab}{}_\mu = 0 \\ \text{Rotations:}\quad & \delta_0 e^a{}_\mu = \lambda^a{}_b e^b{}_\mu \qquad & \delta_0 \omega^{ab}{}_\mu = -\nabla'_\mu \lambda^{ab} \end{aligned}$$

where $\nabla' = \nabla(\omega)$ is the covariant derivative with respect to the spin connection ω. The resulting gauge transformations are clearly different from those obtained in PGT.

We should observe that although the tetrad field and the spin connection carry a representation of the Poincaré group, the EC action in four dimensions, $I_{\text{EC}} = \frac{1}{4}\int d^4x\, \varepsilon^{\mu\nu\lambda\rho}\varepsilon_{abcd} e^c{}_\lambda e^d{}_\rho R^{ab}{}_{\mu\nu}$, is not invariant under the translational part of the Poincaré group:

$$\delta_{\text{T}} I_{\text{EC}} = \frac{1}{4}\int d^4x\, \varepsilon^{\mu\nu\lambda\rho}\varepsilon_{abcd}\lambda^c T^d{}_{\lambda\rho} R^{ab}{}_{\mu\nu} \neq 0.$$

Thus, I_{EC} is invariant under Lorentz rotations and diffeomorphisms, but *not under translations*. The situation is different in three dimensions where gravity can be represented as a 'true' gauge theory (Witten 1988).

The principle of equivalence in PGT

Minimal coupling. The PE represents a physical basis for understanding the phenomenon of gravity. According to this principle, the effect of a non-inertial reference frame on matter in SR is locally equivalent to the gravitational field (chapter 1).

The dynamical content of the PE is expressed by the minimal coupling of matter to gravity. Consider, in Minkowski space M_4, an inertial frame in which the matter field ϕ is described by the Lagrangian $\mathcal{L}_{\text{M}}(\phi, \partial_k\phi)$. When we go over to a non-inertial frame, $\mathcal{L}_{\text{M}}$ transforms into $\sqrt{-g}\,\mathcal{L}_{\text{M}}(\phi, \nabla_k\phi)$. If ϕ is, for instance, a vector field, $\phi \to \phi^l$, then

$$\nabla_k \phi^l = h_k{}^\mu b^l{}_\nu (\partial_\mu \phi^\nu + \Gamma^\nu_{\lambda\mu}\phi^\lambda) \qquad \Gamma^\nu_{\lambda\mu} = \left\{ {\nu \atop \lambda\mu} \right\}.$$

The pseudo-gravitational field, equivalent to the non-inertial reference frame, is represented by the Christoffel symbol Γ and the factor $\sqrt{-g}$. It is clear that this field can be eliminated on the whole spacetime by simply going back to the global inertial frame, while for real fields in spacetime this is not true—they can only be eliminated locally. For this reason, in the last step of introducing a real gravitational field, Einstein replaced M_4 with a Riemann space V_4. We shall see that this is the correct choice but also that Einstein could have chosen the Riemann–Cartan space U_4.

Let us now recall another formulation of the PE: the effects of gravity on matter in spacetime can be eliminated locally by a suitable choice of reference frame, so that matter behaves as in SR. More precisely, *At any point P in spacetime we can choose an orthonormal reference frame $\boldsymbol{e}_m$, such that (a) $g_{mn} = \eta_{mn}$ and (b) $\Gamma^i{}_{jk} = 0$, at P.*

This statement is correct not only in GR (spacetime = V_4), but also in PGT (spacetime = U_4), as we shall see (von der Heyde 1975).

The principle of equivalence in GR. To prove part (a) of the PE in V_4, we note that a transition to a new reference frame $\boldsymbol{e}_m$ at P, expressed by the local coordinate transformation $x^\mu \to y^m$ ($\mathrm{d}y^m = e^m{}_\mu \,\mathrm{d}x^\mu$), implies

$$g'_{mn}(y) = \frac{\partial x^\lambda}{\partial y^m}\frac{\partial x^\rho}{\partial y^n} g_{\lambda\rho}(x). \tag{3.47a}$$

Since the matrices $L_m{}^\lambda = \partial x^\lambda/\partial y^m$, $g_{\lambda\rho}$ and g'_{mn} are constant at P, this equation can be given a simple matrix form:

$$G' = LGL^T. \tag{3.47b}$$

It is known from the theory of matrices that for any symmetric matrix G, there is a non-singular matrix L such that $G' = LGL^T$ is a diagonal matrix, with diagonal elements equal to 1, 0 or -1. Although the matrix L is not unique, the set of diagonal elements for any given G is unique, and it is called the *signature* of G. Since the metric tensor is non-singular, the diagonal elements are different from 0. If the signature of the metric is $(+1, -1, -1, -1)$, it is clear that G can be transformed to η_{mn} at P. This can be done consistently at every point in the spacetime as a consequence of the metricity condition: $\nabla \boldsymbol{g} = 0$.

This choice of local coordinates does not eliminate the gravitational field at P completely, since $\Gamma \neq 0$. To prove part (b) of the PE we choose the point P as the origin of the coordinate system, and define in its neighbourhood a new coordinate transformation

$$z^m = y^m + \tfrac{1}{2} G^m_{ln} y^l y^n \qquad \left.\frac{\partial z^m}{\partial y^n}\right|_{y=0} = \delta^m_n$$

with $G^m_{ln} = G^m_{nl}$. The connection Γ transforms according to

$$\Gamma'^m_{nl} = \frac{\partial z^m}{\partial y^r}\frac{\partial y^s}{\partial z^n}\frac{\partial y^p}{\partial z^l}\Gamma^r_{sp} + \frac{\partial^2 y^p}{\partial z^n \partial z^l}\frac{\partial z^m}{\partial y^p} = \Gamma^m_{nl} - G^m_{nl} \qquad (\text{at } P). \tag{3.48}$$

Since Γ is symmetric in V_4, the choice $G = \Gamma$ yields $\Gamma' = 0$ at P.

The transition $y \to z$ does not change the value of any tensor at P so that the choice $\Gamma' = 0$ can be realized simultaneously with $\boldsymbol{g} = \boldsymbol{\eta}$. This reveals the local Minkowskian structure of V_4, in accordance with the PE.

The principle of equivalence in U_4. Gravitational theory in V_4 shows certain characteristics that do not necessarily follow from the PE. Namely, the form of the Riemannian connection shows that the relative orientation of the orthonormal frame $\boldsymbol{e}_i(x+\mathrm{d}x)$ with respect to $\boldsymbol{e}_i(x)$ (parallel transported to $x+\mathrm{d}x$) is completely fixed by the metric. Since a change in this orientation is described by Lorentz transformations, it does not induce any gravitational effects; therefore, from the point of view of the PE, *there is no reason to prevent independent Lorentz rotations of local frames*. If we want to use this freedom, the spin connection should contain a part which is independent of the metric, which will realize an independent Lorentz rotation of frames under parallel transport:

$$\omega^i{}_{j\mu} \equiv \Delta^i{}_{j\mu} + K^i{}_{j\mu}. \tag{3.49}$$

In this way, the PE leads to a description of gravity which is not in Riemann space, but in Riemann–Cartan geometry U_4. If all inertial frames at a given point are treated on an equal footing, the spacetime has to have torsion.

We shall now show that locally the Riemann–Cartan space has the structure of M_4, in agreement with the PE. Since $\nabla \boldsymbol{g} = 0$ also holds in U_4, the arguments showing that $\boldsymbol{g}$ can be transformed to $\boldsymbol{\eta}$ at any point P in U_4 are the same as in the case of V_4, while the treatment of the connection must be different: the antisymmetric part of ω can be eliminated only by a suitable choice for the relative orientation of neighbouring tetrads.

Let us choose new local coordinates at P, $\mathrm{d}x^\mu \to \mathrm{d}x^i = e^i{}_\mu\,\mathrm{d}x^\mu$, related to an inertial frame. Then,

$$\begin{aligned} g'_{ij} &= e_i{}^\mu e_j{}^\nu g_{\mu\nu} = \eta_{ij} \\ \Gamma'^i_{jk} &= e^i{}_\mu e_j{}^\nu e_k{}^\lambda (\Delta^\mu{}_{\nu\lambda} + K^\mu{}_{\nu\lambda}) \equiv e_k{}^\lambda \omega^i{}_{j\lambda}. \end{aligned}$$

The metricity condition ensures that this can be done consistently at every point in spacetime.

Suppose that we have a tetrad $\boldsymbol{e}_i(x)$ at the point P, and a tetrad $\boldsymbol{e}_i(x+\mathrm{d}x)$ at another point in a neighbourhood of P; then, we can apply a suitable Lorentz rotation to $\boldsymbol{e}_i(x+\mathrm{d}x)$, so that it becomes parallel to $\boldsymbol{e}_i(x)$. Given a vector $\boldsymbol{v}$ at P, it follows that the components $v_k = \boldsymbol{v}\cdot\boldsymbol{e}_k$ do not change under parallel transport from x to $x+\mathrm{d}x$, provided the metricity condition holds. Hence, the connection coefficients $\omega^{ij}{}_\mu(x)$ at P, defined with respect to this particular tetrad field, vanish: $\omega^{ij}{}_\mu(P) = 0$. This property is compatible with $g'_{ij} = \eta_{ij}$, since Lorentz rotation does not influence the value of the metric at a given point.

Therefore, the existence of torsion does not violate the PE.

The space U_4 locally has the structure of M_4: $g(P) = \eta$, $\omega(P) = 0$. Complete realization of the PE demands, as we have seen, U_4 geometry, with both the metric and the antisymmetric part of the connection as independent geometric objects. In particular, the PE holds in V_4, and also in T_4. In more general geometries, where the symmetry of the tangent space is higher than the Poincaré group, the usual form of the PE is violated and local physics differs from SR.

3.3 Gravitational dynamics

The dynamics of the gravitational field in PGT is determined by the form of the gravitational Lagrangian $\mathcal{L}_\mathrm{G}$. If we demand that the equations of motion are at most of second order in the field derivatives, $\mathcal{L}_\mathrm{G}$ can be, at most, quadratic in torsion and curvature (Hayashi and Shirafuji 1980a):

$$\begin{aligned}
\widetilde{\mathcal{L}}_\mathrm{G} &= b(-\alpha R + \mathcal{L}_\mathrm{T} + \mathcal{L}_R + \lambda) \\
\mathcal{L}_\mathrm{T} &\equiv a(AT_{ijk}T^{ijk} + BT_{ijk}T^{jik} + CT_iT^i) \\
\mathcal{L}_R &\equiv b_1 R_{ijkl}R^{ijkl} + b_2 R_{ijkl}R^{klij} + b_3 R_{ij}R^{ij} \\
&\qquad + b_4 R_{ij}R^{ji} + b_5 R^2 + b_6(\varepsilon_{ijkl}R^{ijkl})^2
\end{aligned} \tag{3.50}$$

where α, A, B, C and b_i are free parameters, λ is a cosmological constant, $a = 1/2\kappa$ (κ is Einstein's gravitational constant), $T_i = T^m{}_{mi}$, and pseudoscalar terms are eliminated by requiring parity invariance. The large number of constants offers many possibilities for the choice of $\mathcal{L}_\mathrm{G}$. In what follows, we shall discuss the dynamical details in two particular cases.

Einstein–Cartan theory

In contrast to Yang–Mills theory, in PGT we can construct an invariant which is *linear* in field derivatives, $R = h_i{}^\mu h_j{}^\nu R^{ij}{}_{\mu\nu}$. The action

$$I_\mathrm{EC} = \int \mathrm{d}^4x\, b(-aR + \mathcal{L}_\mathrm{M}) \tag{3.51}$$

defines the so-called Einstein–Cartan (EC) theory, a direct generalization of GR (Kibble 1961, Sciama 1962).

The matter field equations can be written in the covariant form (3.24a). The tetrad field equations are

$$b(R^\mu{}_k - \tfrac{1}{2}h_k{}^\mu R) = \tau^\mu{}_k/2a \tag{3.52}$$

where τ is the dynamical energy–momentum tensor. Formally, both equations have the same form as Einstein's, but here the connection is not Riemannian and τ_{ij} is not necessarily symmetric. Finally, the equations of motion for $A^{ij}{}_\mu$ are:

$$\begin{aligned}
\nabla_\nu H^{\mu\nu}_{ij} &= bh_m{}^\mu(T^m{}_{ij} + \delta^m_i T^s{}_{js} - \delta^m_j T^s{}_{js}) = -\sigma^\mu{}_{ij}/2a \\
H^{\mu\nu}_{ij} &\equiv b(h_i{}^\mu h_j{}^\nu - h_j{}^\mu h_i{}^\nu)
\end{aligned} \tag{3.53}$$

where σ is the dynamical spin tensor. Thus, both mass and spin appear as sources of the gravitational field.

Equation (3.53) can be formally solved first for the torsion, and then for the spin connection:

$$\begin{aligned} -2abT_{ijk} &= \sigma_{ijk} + \tfrac{1}{2}\eta_{ij}\sigma^m{}_{km} - \tfrac{1}{2}\eta_{ik}\sigma^m{}_{jm} \\ A_{ijk} &= \Delta_{ijk} + K_{ijk}. \end{aligned} \tag{3.54}$$

In the simple case of scalar matter we have $\sigma_{ijk} = 0$, therefore $T_{ijk} = 0$, so that EC theory reduces to GR. If $\mathcal{L}_\mathrm{M}$ is *linear* in the field derivatives, then σ_{ijk} is independent of A_{ijk}, and (3.54) yields an explicit solution for $A^{ij}{}_\mu$. The spin connection is not an independent dynamical variable. To simplify the exposition, we shall assume here the condition of linearity (in particular, this is true for the important case of the Dirac field).

Differential conservation laws. It is interesting to note that equation (3.53) can be written in the form

$$\begin{aligned} \partial_\nu H^{\mu\nu}_{ij} + \gamma^\mu{}_{ij}/2a &= -\sigma^\mu{}_{ij}/2a \\ \gamma^\mu{}_{ij}/2a &\equiv A_i{}^s{}_\nu H^{\mu\nu}_{sj} + A_j{}^s{}_\nu H^{\mu\nu}_{is} \end{aligned}$$

which implies a strict (but not covariant) conservation law:

$$\partial_\mu(\gamma^\mu{}_{ij} + \sigma^\mu{}_{ij}) = 0. \tag{3.55}$$

Since $\gamma^\mu{}_{ij}$ can be expressed in the form $\gamma^\mu{}_{ij} = -\partial\widetilde{\mathcal{L}}_\mathrm{G}/\partial A^{ij}{}_\mu$, which shows that it can be interpreted as the spin current of the gravitational field, this equation appears to be the differential conservation law of the total spin.

Similarly, equation (3.52) implies a strict conservation law

$$\partial_\mu(\gamma^\mu{}_k + \tau^\mu{}_k) = 0 \tag{3.56}$$

where $\gamma^\mu{}_k$ is related to the gravitational energy–momentum current. However, the choice of $\gamma^\mu{}_k$ is not unique. The most natural definition, by analogy with spin, would be $\gamma^\mu{}_k = -\partial\widetilde{\mathcal{L}}_\mathrm{G}/\partial b^k{}_\mu$, but then equation (3.56) would be rather trivial, since $\gamma^\mu{}_k + \tau^\mu{}_k = 0$. The question of the correct form for the conservation laws in gravity will be discussed in chapter 6.

Second-order formalism. The action (3.51) is given in the first-order form, since $b^k{}_\mu$ and $A^{ij}{}_\mu$ are independent field variables. Using (3.54) we can eliminate A_{ijk} from the remaining equations and obtain effective EC equations without A_{ijk}. The same effective equations can be obtained from the action (3.51) in which A_{ijk} is replaced by the expression (3.54). Indeed, using the identities

$$bR(A) = bR(\Delta) + b(\tfrac{1}{4}T_{ijk}T^{ijk} + \tfrac{1}{2}T_{ijk}T^{jik} - T_kT^k) - 2\partial_\nu(bT^\nu) \tag{3.57a}$$

$$\widetilde{\mathcal{L}}_\mathrm{M}(\Delta + K) = \widetilde{\mathcal{L}}_\mathrm{M}(\Delta) - \tfrac{1}{2}\sigma^\mu{}_{ij}K^{ij}{}_\mu \tag{3.57b}$$

we obtain the second-order EC Lagrangian directly:

$$
\begin{aligned}
\widetilde{\mathcal{L}}^{(2)}_{\mathrm{EC}} &= -abR(\Delta) + \widetilde{\mathcal{L}}_{\mathrm{M}}(\Delta) + \widetilde{\mathcal{L}}' \\
\widetilde{\mathcal{L}}' &= \frac{b}{8a}(S_{ijk}S^{ijk} + 2S_{ijk}S^{jki} - 2S^i{}_{ji}S_k{}^{jk})
\end{aligned}
\tag{3.58}
$$

where $\sigma_{ijk} = bS_{ijk}$. Thus we see that the only difference between EC theory and GR is in the term $\widetilde{\mathcal{L}}'$, which represents a 'contact' spin–spin interaction. In the case of the Dirac field this interaction is of the fourth power in matter fields, and looks like the Fermi interaction. Since $\widetilde{\mathcal{L}}'$ contains the factor $G = c^3/16\pi a$, this term is much smaller then the other interaction terms so that, for all practical purposes, EC theory is equivalent to GR.

Generalized Belinfante tensor. After the elimination of $A^{ij}{}_\mu$ from the action, the dynamical energy–momentum tensor takes the form

$$
\theta^\mu{}_k \equiv -\frac{\delta}{\delta b^k{}_\mu}\widetilde{\mathcal{L}}_{\mathrm{M}}(\Delta). \tag{3.59a}
$$

This tensor is known as Einstein's energy–momentum tensor, since it is exactly this expression which appears in GR. How did the transition $\tau^\mu{}_k \to \theta^\mu{}_k$ happen? By observing that Δ in $\widetilde{\mathcal{L}}_{\mathrm{M}}(\Delta)$ depends on $b^k{}_\mu$, we find that

$$
\begin{aligned}
\theta^\mu{}_k &= -\frac{\delta}{\delta b^k{}_\mu}\widetilde{\mathcal{L}}_{\mathrm{M}}(A)\bigg|_{A=\Delta} - \frac{1}{2}\left(\frac{\delta\Delta^{ij}{}_\nu}{\delta b^k{}_\mu}\right)\frac{\delta}{\delta A^{ij}{}_\nu}\widetilde{\mathcal{L}}_{\mathrm{M}}(A)\bigg|_{A=\Delta} \\
&= \tau^\mu{}_k(\Delta) + \frac{1}{2}\left(\frac{\delta\Delta^{ij}{}_\nu}{\delta b^k{}_\mu}\right)\sigma^\nu{}_{ij}(\Delta).
\end{aligned}
$$

In order to evaluate this expression, we note that the variation of the relation $A = \Delta + K$ over $b^k{}_\mu$, under the condition $T = 0$, yields

$$
\begin{aligned}
\delta\Delta^{ij}{}_\mu - \tfrac{1}{2}(h^{j\lambda}\delta T^i{}_{\lambda\mu} - b^s{}_\mu h^{i\lambda}h^{j\rho}\delta T_{s\lambda\rho} + h^{i\rho}\delta T^j{}_{\mu\rho}) &= 0 \\
\delta T^k{}_{\mu\lambda} = \nabla'_\mu(\delta b^k{}_\lambda) - \nabla'_\lambda(\delta b^k{}_\mu)&.
\end{aligned}
$$

Here, $\nabla'_\mu = \nabla_\mu(\Delta)$, since $A = \Delta$ for $T = 0$. The covariant derivative ∇'_μ acts on Latin indices in $\delta b^k{}_\mu$, but it can be easily extended to an operator $\widetilde{\nabla}'_\mu \equiv \nabla_\mu(\Delta + \Gamma)$ which acts on both Latin and Greek indices, since

$$
\Gamma^\rho_{\lambda\mu}\delta b^k{}_\rho - \Gamma^\rho_{\mu\lambda}\delta b^k{}_\rho = 0 \qquad \text{when } T = 0.
$$

After that, explicit calculation leads to the result

$$
\theta^\mu{}_k = \tau^\mu{}_k(\Delta) - \tfrac{1}{2}\widetilde{\nabla}'_\lambda(\sigma^\mu{}_k{}^\lambda + \sigma_k{}^{\mu\lambda} - \sigma^\lambda{}_k{}^\mu) \tag{3.59b}
$$

which represents a covariant generalization of Belinfante's relation (Grensing and Grensing 1983). Since the 'total' covariant derivative $\widetilde{\nabla}'_\lambda$ when acting on $h_k{}^\mu$ or $b^k{}_\mu$ gives zero, it is not difficult to check that $\theta_{mk} = \theta_{km}$. Therefore, during the transition $A \to \Delta$, the role of the dynamical energy–momentum tensor is taken over by Einstein's tensor $\theta^\mu{}_k$.

This result is often used to calculate the symmetrized energy–momentum tensor in M_4, by means of the so-called *Rosenfeld procedure*:

(a) first, the Lagrangian $\mathcal{L}_{\mathrm{M}}$ in M_4 is transformed into the related Lagrangian $\widetilde{\mathcal{L}}_{\mathrm{M}}$ in Riemann space ($\partial_i \to \nabla_i$, $\mathcal{L}_{\mathrm{M}} \to b\mathcal{L}_{\mathrm{M}}$);
(b) then, we define Einstein's energy–momentum tensor $\theta^\mu{}_k$; and
(c) finally, the transition to M_4 reduces $\theta^\mu{}_k$ to the symmetric energy–momentum tensor in M_4.

Teleparallel theory

The general geometric arena of PGT, the Riemann–Cartan space U_4, may be *a priori* restricted by imposing certain conditions on the curvature and the torsion. Thus, Einstein's GR is defined in Riemann space V_4, which is obtained from U_4 by requiring the torsion to vanish. Another interesting limit of PGT is Weitzenböck or *teleparallel* geometry T_4, defined by the requirement of vanishing curvature:

$$R^{ij}{}_{\mu\nu}(A) = 0. \tag{3.60}$$

The vanishing of curvature means that parallel transport is path independent (if some topological restrictions are adopted), hence we have an absolute parallelism. Teleparallel geometry is, in a sense, complementary to Riemannian geometry: the curvature vanishes and the torsion remains to characterize the parallel transport.

Of particular importance for the physical interpretation of the teleparallel geometry is the fact that there is a one-parameter family of teleparallel Lagrangians which is *empirically* equivalent to GR (Hayashi and Shirafuji 1979, Hehl *et al* 1980, Nietsch 1980).

The Lagrangian. The gravitational field in the framework of teleparallel geometry in PGT is described by the tetrad $b^k{}_\mu$ and the Lorentz connection $A^{ij}{}_\mu$, subject to the condition of vanishing curvature. We shall consider here the gravitational dynamics determined by a class of Lagrangians quadratic in the torsion:

$$\begin{aligned} \widetilde{\mathcal{L}} &= b\mathcal{L}_{\mathrm{T}} + \lambda_{ij}{}^{\mu\nu} R^{ij}{}_{\mu\nu} + \widetilde{\mathcal{L}}_{\mathrm{M}} \\ \mathcal{L}_{\mathrm{T}} &= a(AT_{ijk}T^{ijk} + BT_{ijk}T^{jik} + CT_kT^k) \equiv \beta_{ijk}(T)T^{ijk} \end{aligned} \tag{3.61}$$

where $\lambda_{ij}{}^{\mu\nu}$ are Lagrange multipliers introduced to ensure the teleparallelism condition (3.60) in the variational formalism, and the explicit form of β_{ijk} is

$$\beta_{ijk} = a(AT_{ijk} + BT_{[jik]} + C\eta_{i[j}T_{k]}).$$

The parameters A, B, C in the Lagrangian should be determined on physical grounds, so as to obtain a consistent theory which could describe all known gravitational experiments. If we require that the theory (3.61) gives the same results as GR in the linear, weak-field approximation, we can restrict our considerations to the one-parameter family of Lagrangians, defined by the following conditions (Hayashi and Shirafuji 1979, Hehl *et al* 1980, Nietsch 1980):

(i) $2A + B + C = 0$, $C = -1$.

This family represents a viable gravitational theory for macroscopic, spinless matter, empirically indistinguishable from GR. Von der Heyde (1976) and Hehl (1980) have given certain theoretical arguments in favour of the choice $B = 0$. There is, however, another, particularly interesting choice determined by the requirement

(ii) $2A - B = 0$.

In the gravitational sector, this choice leads effectively to the Einstein–Hilbert Lagrangian $\mathcal{L}_{\mathrm{EH}} = -abR(\Delta)$, defined in Riemann spacetime V_4 with Levi-Civita connection $A = \Delta$, via the geometric identity (3.57a):

$$bR(A) = bR(\Delta) + b(\tfrac{1}{4}T_{ijk}T^{ijk} + \tfrac{1}{2}T_{ijk}T^{jik} - T_kT^k) - 2\partial_\nu(bT^\nu).$$

Indeed, in Weitzenböck spacetime, where $R(A) = 0$, this identity implies that the torsion Lagrangian in (3.61) is equivalent to the Einstein–Hilbert Lagrangian, up to a four-divergence, provided that

$$A = \tfrac{1}{4} \qquad B = \tfrac{1}{2} \qquad C = -1 \tag{3.62}$$

which coincides with the conditions (i) and (ii) given earlier. The theory defined by equations (3.61) and (3.62) is called the teleparallel formulation of GR ($\mathrm{GR}_\parallel$).

Field equations. By varying the Lagrangian (3.61) with respect to $b^i{}_\mu$, $A^{ij}{}_\mu$ and $\lambda_{ij}{}^{\mu\nu}$, we obtain the gravitational field equations:

$$4\nabla_\rho(b\beta_i{}^{\mu\rho}) - 4b\beta^{nm\mu}T_{nmi} + h_i{}^\mu b\mathcal{L}_{\mathrm{T}} = \tau^\mu{}_i \tag{3.63a}$$

$$4\nabla_\rho\lambda_{ij}{}^{\mu\rho} - 8b\beta_{[ij]}{}^\mu = \sigma^\mu{}_{ij} \tag{3.63b}$$

$$R^{ij}{}_{\mu\nu} = 0. \tag{3.63c}$$

The third field equation defines the teleparallel geometry in PGT. The first field equation is a dynamical equation for $b^k{}_\mu$ and plays an analogous role to that of Einstein's equation in GR. It can be written in an equivalent form as

$$4\nabla_\rho(b\beta_i{}^{j\rho}) + 2b\beta_{imn}T^{jmn} - 4b\beta^{mnj}T_{mni} + \delta_i^j b\mathcal{L}_{\mathrm{T}} = \tau^j{}_i. \tag{3.63a'}$$

By taking the covariant divergence of the second field equation and using (3.63c), we obtain the following consistency condition:

$$-8\nabla_\mu(b\beta_{[ij]}{}^\mu) \approx \nabla_\mu\sigma^\mu{}_{ij}.$$

This condition is satisfied as a consequence of the first field equation. Indeed, equation (3.63a') implies the relation $4\nabla_\mu(b\beta_{[ij]}{}^\mu) \approx \tau_{[ji]}$, which, in conjunction with the second identity in (3.23), leads exactly to this condition. Thus, the only role of equations (3.63b) is to determine the Lagrange multipliers $\lambda_{ij}{}^{\mu\nu}$. Taking into account the consistency requirements, we conclude that the number of independent equations (3.63b) is $24 - 6 = 18$. Hence, it is clear that these equations alone cannot determine the 36 multipliers $\lambda_{ij}{}^{\mu\nu}$ in a unique way. As we shall see soon, the non-uniqueness of $\lambda_{ij}{}^{\mu\nu}$ is related to the existence of an extra gauge freedom in the theory.

The λ symmetry. The gravitational Lagrangian (3.61) is, by construction, invariant under the local Poincaré transformations:

$$
\begin{aligned}
\delta_0 b^k{}_\mu &= \omega^k{}_s b^s{}_\mu - \xi^\rho{}_{,\mu} b^k{}_\rho - \xi^\rho \partial_\rho b^k{}_\mu \\
\delta_0 A^{ij}{}_\mu &= -\omega^{ij}{}_{,\mu} + \omega^i{}_s A^{sj}{}_\mu + \omega^j{}_s A^{is}{}_\mu - \xi^\rho{}_{,\mu} A^{ij}{}_\rho - \xi^\rho \partial_\rho A^{ij}{}_\mu \\
\delta_0 \lambda_{ij}{}^{\mu\nu} &= \omega_i{}^s \lambda_{sj}{}^{\mu\nu} + \omega_j{}^s \lambda_{is}{}^{\mu\nu} + \xi^\mu{}_{,\rho} \lambda_{ij}{}^{\rho\nu} + \xi^\nu{}_{,\rho} \lambda_{ij}{}^{\mu\rho} - \partial_\rho(\xi^\rho \lambda_{ij}{}^{\mu\nu}).
\end{aligned}
\tag{3.64}
$$

In addition, it is also invariant, up to a four-divergence, under the transformations (Blagojević and Vasilić 2000)

$$
\delta_0 \lambda_{ij}{}^{\mu\nu} = \nabla_\rho \varepsilon_{ij}{}^{\mu\nu\rho} \tag{3.65a}
$$

where the gauge parameter $\varepsilon_{ij}{}^{\mu\nu\rho} = -\varepsilon_{ji}{}^{\mu\nu\rho}$ is completely antisymmetric in its upper indices, and has $6 \times 4 = 24$ components. This invariance is easily verified by using the second Bianchi identity $\varepsilon^{\lambda\rho\mu\nu}\nabla_\rho R^{ij}{}_{\mu\nu} = 0$. On the other hand, the invariance of the field equation (3.63b) follows directly from $R^{ij}{}_{\mu\nu} = 0$. The symmetry (3.65a) will be referred to as the λ symmetry.

It is useful to observe that the λ transformations can be written in the form

$$
\begin{aligned}
\delta_0 \lambda_{ij}{}^{\alpha\beta} &= \nabla_0 \varepsilon_{ij}{}^{\alpha\beta} + \nabla_\gamma \varepsilon_{ij}{}^{\alpha\beta\gamma} \qquad \varepsilon_{ij}{}^{\alpha\beta} \equiv \varepsilon_{ij}{}^{\alpha\beta 0} \\
\delta_0 \lambda_{ij}{}^{0\beta} &= \nabla_\gamma \varepsilon_{ij}{}^{\beta\gamma}.
\end{aligned}
\tag{3.65b}
$$

As we can show by canonical methods (see chapter 5), the only independent parameters of the λ symmetry are $\varepsilon_{ij}{}^{\alpha\beta}$; in other words, the six parameters $\varepsilon_{ij}{}^{\alpha\beta\gamma}$ can be completely discarded. Consequently, the number of *independent* gauge parameters is $24 - 6 = 18$. They can be used to fix 18 multipliers $\lambda_{ij}{}^{\mu\nu}$, whereupon the independent field equations (3.63b) determine the remaining 18 multipliers (at least locally).

It is evident that Poincaré and λ gauge symmetries are always present (*sure* symmetries), independently of the values of parameters A, B and C in the teleparallel theory (3.61). Moreover, it will become clear from canonical analysis that there are *no other* sure gauge symmetries. Specific models, such as $\mathrm{GR}_\parallel$, may have *extra* gauge symmetries which are present only for some special (critical) values of the parameters. The influence of extra gauge symmetries on the existence of a consistent coupling with matter is not completely clear at the moment. We shell assume here that the matter coupling respects all extra gauge symmetries, if they exist.

Orthonormal and teleparallel frames. Teleparallel theories in U_4 are based on the condition of vanishing curvature. Let us choose an arbitrary tetrad at point P of spacetime. Then, by parallel transporting this tetrad to all other points, we generate the tetrad field on the spacetime manifold. If the manifold is parallelizable (which is a strong topological assumption), the vanishing curvature implies that the parallel transport is path independent, so the resulting tetrad field is globally well defined. In such an *orthonormal* and *teleparallel* (OT) frame, the connection coefficients vanish:

$$A^{ij}{}_{\mu} = 0. \tag{3.66}$$

This construction is not unique—it defines a class of OT frames, related to each other by global Lorentz transformations. In an OT frame, the covariant derivative reduces to the partial derivative, and the torsion takes the simple form: $T^i{}_{\mu\nu} = \partial_\mu b^i{}_\nu - \partial_\nu b^i{}_\mu$ (see e.g. Nester 1991).

Equation (3.66) defines a particular solution of the teleparallelism condition $R^{ij}{}_{\mu\nu}(A) = 0$. Since a local Lorentz transformation of the tetrad field, $e'^i{}_\mu = \Lambda^i{}_k e^k{}_\mu$, induces a non-homogeneous change in the connection,

$$A'^{ij}{}_\mu = \Lambda^i{}_m \Lambda^j{}_n A^{mn}{}_\mu + \Lambda^i{}_m \partial_\mu \Lambda^{jm}$$

we can conclude that the general solution of $R^{ij}{}_{\mu\nu}(A) = 0$ has the form $A^{ij}{}_\mu = \Lambda^i{}_m \partial_\mu \Lambda^{jm}$. Thus, the choice (3.66) breaks local Lorentz invariance, and represents a *gauge-fixing condition* in teleparallel theory.

In action (3.61), the teleparallel condition is realized via a Lagrange multiplier. The second field equation, obtained by variation with respect to $A^{ij}{}_\mu$, merely serves to determine the multiplier, while the non-trivial dynamics is completely contained in the first field equation. Hence, teleparallel theory may also be described by imposing gauge condition (3.66) directly on the action. The resulting theory is defined in terms of the tetrad field and may be thought of as the gauge theory of the translational group. This formalism keeps the observable properties of the theory unchanged and is often used in the literature because of its technical simplicity. However, we shall continue to work with the local Lorentz invariant formulation, since it simplifies the canonical treatment of the angular momentum conservation (see chapter 6).

Differential conservation laws. Field equations (3.63a, b) can be written in an equivalent form as

$$\begin{aligned} \partial_\rho \psi_\nu{}^{\mu\rho} &= \tau^\mu{}_\nu + \gamma^\mu{}_\nu \\ 4\partial_\rho \lambda_{ij}{}^{\mu\rho} &= \sigma^\mu{}_{ij} + \gamma^\mu{}_{ij} \end{aligned} \tag{3.67a}$$

where

$$\begin{aligned}\gamma^\mu{}_\nu &= 4b\beta_m{}^{\mu\rho}\partial_\nu b^m{}_\rho + 2\lambda_{mn}{}^{\mu\rho}\partial_\nu A^{mn}{}_\rho - \delta^\mu_\nu b\mathcal{L}_\mathrm{T} \\ \gamma^\mu{}_{ij} &= 8b\beta_{[ij]}{}^\mu + 4(A_i{}^s{}_\rho\lambda_{sj}{}^{\rho\mu} + A_j{}^s{}_\rho\lambda_{is}{}^{\rho\mu})\end{aligned} \qquad (3.67b)$$

and $\psi_\nu{}^{\mu\rho} = 4b\beta_m{}^{\mu\rho}b^m{}_\nu + 2\lambda_{mn}{}^{\mu\rho}A^{mn}{}_\nu$. Equations (3.67*a*) imply a strict conservation laws for the quantities $\tau^\mu{}_\nu + \gamma^\mu{}_\nu$ and $\sigma^\mu{}_{ij} + \gamma^\mu{}_{ij}$, which lead us to interpret $\gamma^\mu{}_\nu$ and $\gamma^\mu{}_{ij}$ as the energy–momentum and spin tensors of the teleparallel gravitational field. The related conservation laws will be studied in chapter 6.

The teleparallel form of GR. For the specific values of parameters (3.62), the gravitational sector of the teleparallel theory is equivalent to GR. Hence, we expect that the tetrad field equation (3.63*a*) in $\mathrm{GR}_\parallel$ coincides with Einstein's equation. To show this, we start from the general identity in U_4,

$$R^{ij}{}_{\mu\nu}(A) = R^{ij}{}_{\mu\nu}(\Delta) + (\nabla_\mu K^{ij}{}_\nu - K^i{}_{s\mu}K^{sj}{}_\nu)$$

multiply it by $aH^{\nu\mu}_{kj}/2$, use the relation $a\nabla_\rho H^{\mu\rho}_{ij} = -4b\beta_{[ij]}{}^\mu$ which holds only in $\mathrm{GR}_\parallel$, and obtain (Blagojević and Nikolić 2000)

$$\begin{aligned}abR^{ik}(A) = abR^{ik}(\Delta) &- 2\nabla_\mu(b\beta^{ik\mu}) + 2b\beta_{mn}{}^kT^{mni} \\ &- b\beta^{imn}T^k{}_{mn} - \eta^{ik}a\partial_\mu(bT^\mu) - 4\nabla_\mu(b\beta^{[ik]\mu}).\end{aligned}$$

The last term on the right-hand side vanishes for $R^{ij}{}_{\mu\nu}(A) = 0$. In this case we find that

$$2ab[R^{ik}(\Delta) - \tfrac{1}{2}\eta^{ik}R(\Delta)] = 4\nabla_\mu(b\beta^{ik\mu}) - 4b\beta_{mn}{}^kT^{mni} + 2b\beta^{imn}T^k{}_{mn} + \eta^{ik}b\mathcal{L}_\mathrm{T}.$$

As a consequence of this identity, the first field equation (3.63*a*′) takes the form of Einstein's equation:

$$b[R^{ik}(\Delta) - \tfrac{1}{2}\eta^{ik}R(\Delta)] = \tau^{ki}/2a. \qquad (3.68a)$$

Here, on the left-hand side we have Einstein's tensor of GR which is a symmetric tensor, so that the dynamical energy–momentum tensor must be also symmetric, $\tau^{ik} = \tau^{ki}$.

The second field equation of $\mathrm{GR}_\parallel$ can be written in the form

$$\nabla_\rho(4\lambda_{ij}{}^{\mu\rho} + 2aH^{\mu\rho}_{ij}) = \sigma^\mu{}_{ij}. \qquad (3.68b)$$

General remarks

To illustrate general dynamical properties of gravity in U_4 spacetime, we give here brief comments on some specific models.

***R* + *T*^{2} theory.** One of the simplest generalizations of the EC theory has the form

$$I_1 = \int \mathrm{d}^4x\, b(-aR + \mathcal{L}_\mathrm{T} + \mathcal{L}_\mathrm{M}).$$

This theory, too, does not contain a kinetic part for the spin connection, so that the related equations of motion are also algebraic:

$$(2A - B)T_{[ij]k} - (B - 1/2)T_{kij} + (C + 1)\eta_{k[i}T^m{}_{j]m} = -\sigma_{kji}/4a.$$

They can be solved for T by means of the irreducible decomposition of torsion with respect to the Lorentz group (Hayashi and Shirafuji 1980b). We find the result

$$\mu_V T^V{}_i = \sigma^V{}_i \qquad \mu_A T^A{}_i = \sigma^A{}_i \qquad \mu_T T^T{}_{ijk} = \sigma^T{}_{ijk}\,,$$

where indices V, A and T denote the vector, axial and tensor components, respectively, and μ_a are mass parameters. If all the $\mu_a \neq 0$, this equation can be solved for T^a (and the connection), and the effective theory differs from GR by contact terms of the type σ^2 . If some $\mu_a = 0$, the related component T^a is absent from the field equation, and the σ^a have to vanish for consistency. This case corresponds to the existence of massless tordions (particles corresponding to non-vanishing torsion).

***R*^{2} + *T*^{2} theory.** In a dynamical sense, EC and $R + T^2$ theories are incomplete. Indeed, the spin connection is introduced as an independent field in U_4 theory, but EC dynamics imposes an algebraic relation between A and the tetrad field. The spin connection acquires a full dynamical content only through the R^2 terms. The following $R^2 + T^2$ model is inspired by some analogies with electrodynamics (von der Heyde 1976):

$$I_2 = \int \mathrm{d}^4x\, b[\alpha R_{ijkl}R^{ijkl} + \beta(-T_{ijk}T^{ijk} + 2T^{ik}{}_i T_{jk}{}^j)].$$

A basic problem for models without a linear curvature term in the action is the question of the macroscopic limit. For the previous model, the following explanation was suggested (Hehl *et al* 1980). In the context of classical field theory, macroscopic matter can be modelled by a scalar field. Since scalar matter does not interact directly with the spin connection, we expect that, effectively, only translation gauge fields should be introduced; hence, $R_{ijkl}(A) = 0$ and I_3 is reduced to a T^2 action in T_4, with $B = 0$. Analysis of this particular teleparallel theory shows that it describes the standard tests of GR correctly. However, if we recall that the macroscopic matter is composed mainly of Dirac particles, these arguments are not sufficiently clear. We would like to have a better understanding of the averaging procedure which completely eliminates the spin effects and enables an effective description of matter in terms of a scalar field.

A more complete insight into the structure of this model is obtained by analysing its classical solutions. An interesting property of the model is that in the linear approximation, in addition to the Newtonian potential, we find a term linear in r. This term is interpreted as part of the 'strong gravity', related to hadron interactions. The same idea is then extended to (L_4, g) theory (Šijački 1982).

General properties. After these comments, we shall now mention some general properties of PGT.

The basic dynamical variables of the theory are $b^k{}_\mu$ and $A^{ij}{}_\mu$. Since the time derivatives of $b^k{}_0$ and $A^{ij}{}_0$ do not appear in the torsion and curvature, their time evolution remains undetermined. The number of remaining variables $b^k{}_\alpha$ and $A^{ij}{}_\alpha$ is $4 \times 3 + 6 \times 3 = 30$. By fixing 10 parameters of the local Poincaré symmetry we can impose 10 gauge conditions, reducing the number of independent variables to $30 - 10 = 20$. Two degrees of freedom correspond to the *graviton*, and 18 degrees of freedom $A^{ij}{}_\alpha$ describe the *tordions*. The decomposition of $A^{ij}{}_\alpha$ into irreducible components of the rotation group gives tordion components of definite spin and parity, $J^P(A) = 2^\pm, 1^\pm, 0^\pm$, the number of which is 2(5+3+1)=18.

Since the scalar curvature and T^2 components contain A^2 terms, tordions may have non-vanishing masses, which are proportional to the mass parameters μ_a. Of course, it is the presence of the R^2 terms which enables the propagation of tordions. A detailed analysis of the mass spectrum of the linearized theory has been performed in the case when all tordions are massive (Hayashi and Shirafuji 1980d, Sezgin and van Nieuwenhuizen 1980, Kuhfuss and Nitsch 1986). The results for massless tordions are not so complete (Battiti and Toller 1985).

Many aspects of the general Hamiltonian structure of the theory are clarified (Blagojević and Nikolić 1983, Nikolić 1984). In the case of massive tordions, a correspondence between the particle spectrum and the nature of the constraints has been found. Namely, whenever a tordion does not propagate (infinite mass), there exists a constraint which can be used to eliminate the related tordion field. In the case of massless tordions we find extra gauge symmetries. The Hamiltonian approach is found to be extremely useful for the analysis of energy, momentum and angular momentum of gravitational systems, as we shall see in chapter 6.

The general structure of the equations of motion has been analysed in order to clarify the evolution of given initial data, i.e. the Cauchy problem (Dimakis 1989). It is found that 10 equations of motion, out of 40, represent the constraints on the initial data, while the remaining 30 define a consistent time evolution of the dynamical variables provided that the parameters of the theory obey certain conditions. These conditions are found to be the same ones that determine the existence of additional constraints in the Hamiltonian analysis (see also Hecht *et al* 1996).

The problem of finding exact solutions of the nonlinear equations of motion in PGT is treated with the help of some well-known methods from Yang–Mills theory. The essence of the method is the so-called double duality anzatz, in which we postulate a linear relation between the generalized momenta $\pi_{ij}{}^{\mu\nu}$ and the double dual image of the curvature tensor. A class of exact solutions has been found for the I_3 action, giving us a deep insight into the physical content of PGT (see, e.g., Mielke 1987).

Although the previous analysis was related to a four-dimensional spacetime, many of the results can be generalized to the case $d > 4$, corresponding to the Kaluza–Klein programme of unification, as well as to the case $d = 2$, related to string theory.

At the very end of this chapter, with a new understanding of gravity, it seems natural to reconsider the following question: What are the basic properties of an acceptable theory of gravity? Since GR is phenomenologically a correct theory, the basic criteria for some alternative theory would be:

- consistency with quantization,
- absence of classical singularities and
- the possibility of unification with the other fundamental interactions.

In the general U_4 theory, the R and T^2 terms contain dimensional constants, which is not an attractive property from the point of view of quantization. On the other hand, the R^2 terms contain dimensionless constants, but the related classical limit is questionable. It is also interesting to mention the possibility of the existence of a repulsive tordion interaction at small distances, which may prevent gravitational collapse and avoid the appearance of infinite matter densities (Minkevich 1980, Blagojević *et al* 1982). Since U_4 theory is based on the gauge principle, it seems to be a promising framework for the unification of gravity with other interactions.

It might be possible that some problems of PGT could be solved by demanding some additional symmetry for the action. Having this in mind, it would be very interesting to study local Weyl theory, which will be our objective in the next chapter.

Exercises

1. Derive the transformation laws of the gauge fields: (a) $A^{ij}{}_\mu$ and (b) $h_k{}^\mu$ and $b^k{}_\mu$.
2. (a) Verify that the Lagrangian $\mathcal{L}'_M = \mathcal{L}_M(\phi, \nabla_k\phi)$ obeys the condition $\delta\mathcal{L}'_M = 0$.
   (b) Show, by using $\delta_0\Lambda + \partial_\mu(\Lambda\xi^\mu) = 0$, that $\Lambda = \Lambda(h, A, x)$ is independent of both x and $A^{ij}{}_\mu$. Then prove that $\Lambda = \det(b^k{}_\mu)$.
3. If v^i is a Lorentz vector, find the transformation law of $v^\mu = h_i{}^\mu v^i$ under local Poincaré transformations.
4. (a) Prove the relation $[\nabla_\mu, \nabla_\nu]b^s{}_\lambda = F^s{}_{\lambda\mu\nu}$.
   (b) Derive the Bianchi identities given in the text, using Jacobi identities for the commutators.
5. (a) Find the transformation law of $K^\mu = \partial\widetilde{\mathcal{L}}_M/\partial\nabla_\mu\phi$, and define $\nabla_\nu K^\mu$.

(b) Show that the equations of motion for matter fields have the Poincaré covariant form (3.24a).

(c) Find the form of these equations for the free electromagnetic field.

6. Show that the equations of motion for the gauge fields $b^k{}_\mu$ and $A^{ij}{}_\mu$ can be written in the form (3.24b).
7. (a) Use the general form of the matter Lagrangian, $\widetilde{\mathcal{L}}_{\mathrm{M}} = b\mathcal{L}_{\mathrm{M}}(\phi, \nabla_k\phi)$, to prove the equality of the covariant and dynamical energy–momentum and spin currents (without using the invariance condition $\Delta\tilde{\mathcal{L}}_{\mathrm{M}} = 0$).

   (b) Verify this result for the free electromagnetic field.
8. Use the matter field equations to prove the generalized conservation laws (3.23).
9. Verify the generalized conservation laws (3.23) for the free (a) scalar, (b) Dirac and (c) electromagnetic fields.
10. Prove that δ_0^* transformations of gauge fields are given as follows:

$$\delta_0^* b^k{}_\mu = \omega^k{}_s b^s{}_\mu - \nabla_\mu \xi^k + \xi^\lambda F^k{}_{\mu\lambda}$$
$$\delta_0^* A^{ij}{}_\mu = -\nabla_\mu \omega^{ij} + \xi^\nu F^{ij}{}_{\mu\nu}.$$

11. (a) Using the fact that $\nabla_\nu(\Gamma)A^\mu$ is a tensor, derive the transformation law of the affine connection $\Gamma^\mu_{\lambda\nu}$. Then show that the torsion $T^\mu{}_{\lambda\nu}$ is a tensor.

    (b) Express the affine connection in terms of the Christoffel symbol and the torsion, using the metricity condition.
12. Find the connection ω from the relation $\nabla_\mu(\omega + \Gamma)e^i{}_\nu = 0$.
13. The metric of a two-dimensional Riemann space is determined by the interval $\mathrm{d}s^2 = \mathrm{d}v^2 - v^2\,\mathrm{d}u^2$. Show that this space is flat.
14. The metric of the two-dimensional Euclidean space E_2 in polar coordinates has the form $\mathrm{d}s^2 = \mathrm{d}\rho^2 + \rho^2\,\mathrm{d}\theta^2$.

    (a) Calculate the Christoffel symbols.

    (b) Check that every straight line in E_2 obeys the geodesic equation.

    (c) Find the components of the vector $\boldsymbol{A} = \boldsymbol{e}_\rho$, given at the point $(\rho, \theta) = (1, 0)$, after parallel transport to $(1, \pi/2)$.
15. Find the transformation law of the spin connection under finite local Lorentz rotations. Show that we can use this freedom to transform the spin connection $\omega^{ij}{}_\mu$ to zero at an arbitrary point P.
16. Calculate the canonical and dynamical energy–momentum and spin tensors of the gravitational field in EC theory.
17. Let $H^{\mu\nu}_{ij} \equiv b(h_i{}^\mu h_j{}^\nu - h_j{}^\mu h_i{}^\nu)$, and $\nabla'_\mu = \nabla_\mu(\Delta)$. Prove the identities:

$$\begin{aligned}
\nabla_\rho H^{\mu\rho}_{ij} &= bh_k{}^\mu(T^k{}_{ij} - \delta^k_i T_j + \delta^k_j T_i)\\
R^{ij}{}_{\mu\nu}(A) &= R^{ij}{}_{\mu\nu}(\Delta) + [\nabla'_\mu K^{ij}{}_\nu + K^i{}_{s\mu}K^{sj}{}_\nu - (\mu \leftrightarrow \nu)]\\
&= R^{ij}{}_{\mu\nu}(\Delta) + [\nabla_\mu K^{ij}{}_\nu - K^i{}_{s\mu}K^{sj}{}_\nu - (\mu \leftrightarrow \nu)].
\end{aligned}$$

18. Prove the identities (3.57a) and (3.57b).

19. Using Rosenfeld's procedure find the energy–momentum tensor for the free (a) scalar, (b) Dirac and (c) electromagnetic fields.
20. Show that equation (3.63a) can be written in the form (3.63a').
21. Calculate the canonical and dynamical energy–momentum and spin tensors of the gravitational field in teleparallel theory.
22. Derive the following identities in $\mathrm{GR}_\parallel$:

$$\begin{aligned} a\nabla_\rho H_{ij}^{\mu\rho} &= -4b\beta_{[ij]}{}^\mu \\ abR^{ik}(A) &= abR^{ik}(\Delta) + 2\nabla_\mu(b\beta^{i\mu k}) + 2b\beta_{mn}{}^k T^{mni} \\ &\quad - b\beta^{imn}T^k{}_{mn} - \eta^{ik}a\partial_\mu(bT^\mu) - 4\nabla_\mu(b\beta^{[ik]\mu}). \end{aligned}$$

23. Consider the theory defined by the action $I = \int \mathrm{d}^4x\, b(\mathcal{L}_\mathrm{T} + \mathcal{L}_\mathrm{M})$, where $\mathcal{L}_\mathrm{M}$ is linear in the field derivatives.
    (a) Write the field equations for $A^{ij}{}_\mu$.
    (b) Assuming conditions (3.62) hold, find T_{ijk} as a function of σ_{ijk}.
24. (a) Show that

$$\beta_{ijk} = \tfrac{1}{2}a(\tau_{jki} - \tau_{ijk} + \tau_{kij}) - \tfrac{1}{4}a(2B-1)\overset{A}{T}{}_{ijk}$$

    where $\tau_{kij} = \eta_{k[i}T_{j]} - \frac{1}{2}T_{kij}$.
    (b) Derive the following identities:

$$\begin{aligned} 2b\tau_{kij} &= -b_{k\lambda}\nabla_\rho H_{ij}^{\lambda\rho} \\ 4b\beta_{(ij)}{}^\mu &= ah^{k\mu}b_{i\lambda}\nabla_\rho H_{jk}^{\lambda\rho} + (i \leftrightarrow j) \\ 4b\beta_{[ij]}{}^\mu &= -a\nabla_\rho H_{ij}^{\mu\rho} - a(2B-1)bh^{\kappa\mu}\overset{A}{T}{}_{ijk}. \end{aligned}$$

    (c) Use the field equations to show that

$$\nabla_\mu[\sigma^\mu{}_{ij} - 2a(2B-1)bh^{k\mu}\overset{A}{T}{}_{ijk}] \approx 0.$$

Chapter 4

Weyl gauge theory

Many attempts to unify gravity with other fundamental interactions are based on the idea of gauge symmetry. The principle of gauge invariance was invoked for the first time in Weyl's unified theory of gravitation and electromagnetism (Weyl 1918). The geometry of Weyl's unified theory represents an extension of the Riemannian structure of spacetime in GR. Weyl introduced spacetime in which the principle of relativity applies not only to the choice of reference frames, but also to the choice of local standards of length.

Weyl spacetime is based on the idea of invariance under local change of the unit of *length* (gauge), which is realized by introducing an additional compensating field. Weyl tried to interpret the new field as the electromagnetic potential but, as further development showed, this was not possible. It turned out that this field interacts in the same manner with both particles and antiparticles, contrary to all experimental evidence about electromagnetic interactions. Passing through a process of change, the idea evolved into a new symmetry principle related to the local change of *phase* (Weyl 1931), the principle that underlies modern understanding of the gauge theories of fundamental interactions (see, e.g., O'Raiffeartaigh 1986).

Although Weyl's original idea was not acceptable for a description of the electromagnetic interaction, it gained new strength after some interesting discoveries in particle physics, in the 1960s. Experimental results on electron–nucleon scattering in the deep inelastic region showed that the scattering amplitudes behaved as if all masses were negligible, thus focusing our attention on physical theories that are invariant under the change of mass (or length) scale. Localization of this symmetry brings us back to the old Weyl theory, which describes, as we shall see, not the electromagnetic but the gravitational interaction (this one is the same for particles and antiparticles, on the basis of the PE).

Weyl's idea can be realized in two complementary ways:

(a) as a gauge field theory, based on the Weyl group $W(1,3)$; and
(b) as a geometric theory, obtained by extending Riemannian structure with local scale invariance.

The first approach gives a new meaning to the old geometric construction.

4.1 Weyl gauge invariance

The theory of gravity based on the local $W(1,3)$ symmetry represents a kind of minimal extension of PGT (Bregman 1973, Charap and Tait 1974, Kasuya 1975). Although the kinematical aspects of Weyl gauge theory (WGT) become more complex due to the presence of new compensating fields, a higher symmetry simplifies the structure of the action.

Localization of Weyl symmetry

Consider a dynamical system of matter fields in M_4, referred to a local Lorentz frame, and described by an action invariant under *global* $W(1,3)$ transformations of coordinates,

$$\delta x^\mu = \varepsilon^\mu + \omega^\mu{}_\nu x^\nu + \rho x^\mu \equiv \xi^\mu \tag{4.1}$$

accompanied by the related transformations of fields:

$$\begin{aligned}\delta_0\phi = (\tfrac{1}{2}\omega\cdot M + \varepsilon\cdot P + \rho D)\phi = (\tfrac{1}{2}\omega\cdot\Sigma + \rho d + \xi\cdot P)\phi \equiv W\phi \\ \delta_0\partial_k\phi = W\partial_k\phi + \omega_k{}^\nu\partial_\nu\phi - \rho\partial_k\phi \equiv W_k{}^\nu\partial_\nu\phi.\end{aligned} \tag{4.2}$$

The invariance condition (2.11), with $\partial_\mu\xi^\mu = 4\rho$, is equivalent to equations (2.12) and (2.32*a*).

If we now generalize these transformations by assuming that the constant parameters are replaced by some *functions* of spacetime points, the invariance condition (2.11) is violated for two reasons. First, the transformation law of $\partial_k\phi$ is changed

$$\begin{aligned}\delta_0\partial_k\phi &= W_k{}^\nu\partial_\nu\phi + \tfrac{1}{2}\omega^{\mu\nu}{}_{,k}M_{\mu\nu}\phi + \varepsilon^\nu{}_{,k}P_\nu\phi + \rho_{,k}D\phi \\ &= W\partial_k\phi - \xi^\nu{}_{,k}\partial_\nu\phi + \tfrac{1}{2}\omega^{ij}{}_{,k}\Sigma_{ij}\phi + \rho_{,k}d\phi\end{aligned} \tag{4.3}$$

and second, $\partial_\mu\xi^\mu \neq 4\rho$. The violation of local invariance takes the form

$$\Delta\mathcal{L}_{\mathrm{M}} = \tfrac{1}{2}\omega^{ij}{}_{,\mu}M^\mu{}_{ij} - \varepsilon^i{}_{,\mu}T^\mu{}_i - \rho_{,\mu}D^\mu \neq 0.$$

The invariance can be restored by introducing the *covariant derivative* $\overset{*}{\nabla}_k\phi$, which transforms according to the 'old rule':

$$\delta_0\overset{*}{\nabla}_k\phi = W\overset{*}{\nabla}_k\phi + \omega_k{}^i\overset{*}{\nabla}_i\phi - \rho\overset{*}{\nabla}_k\phi = W_k{}^i\overset{*}{\nabla}_i\phi \tag{4.4}$$

and by taking care of the fact that $\partial_\mu\xi^\mu \neq 4\rho$.

Covariant derivative. In order to construct the covariant derivative $\overset{*}{\nabla}_k\phi$, we first introduce the (ω, ρ)-covariant derivative,

$$\begin{aligned}\overset{*}{\nabla}_\mu\phi &= \partial_\mu\phi + \delta_0(\omega, \rho, \xi = 0)\phi|_{\omega\to A_\mu, \rho\to B_\mu} \\ &= (\partial_\mu + \overset{*}{A}_\mu)\phi \qquad \overset{*}{A}_\mu \equiv \tfrac{1}{2}A^{ij}{}_\mu\Sigma_{ij} + B_\mu d^*\end{aligned} \tag{4.5}$$

which transforms according to

$$\delta_0\overset{*}{\nabla}_\mu\phi = W\overset{*}{\nabla}_\mu\phi - \xi^\nu{}_{,\mu}\overset{*}{\nabla}_\nu\phi \tag{4.6}$$

and eliminates the terms $\omega^{ij}{}_{,k}$ and $\rho_{,k}$ in (4.3). Note that the Weyl charge d^* is, by assumption, equal to the usual scale dimension,

$$d^* = d.$$

Then, we obtain the following transformation rules for $A^{ij}{}_\mu$ and B_μ:

$$\begin{aligned}\delta_0 A^{ij}{}_\mu &= \delta_0^{\mathrm{P}} A^{ij}{}_\mu \\ \delta_0 B_\mu &= -\xi^\nu{}_{,\mu}B_\nu - \xi\cdot\partial B_\mu - \rho_{,\mu} \equiv \delta_0^{\mathrm{P}}B_\mu - \rho_{,\mu}\end{aligned} \tag{4.7}$$

where δ_0^{P} denotes the Poincaré transformation with new ξ^μ. Rewriting equation (4.6) in the form

$$\delta_0\overset{*}{\nabla}_\mu\phi = W_\mu{}^\nu\overset{*}{\nabla}_\nu\phi - (\xi^\nu{}_{,\mu} - \omega^\nu{}_\mu - \rho\delta^\nu_\mu)\overset{*}{\nabla}_\nu\phi$$

it becomes clear that the last term, homogeneous in $\overset{*}{\nabla}_\nu\phi$, can be eliminated by adding a new compensating field,

$$\overset{*}{\nabla}_k\phi = \delta_k^\nu\overset{*}{\nabla}_\nu\phi - A^\nu{}_k\overset{*}{\nabla}_\nu\phi \equiv h_k{}^\nu\overset{*}{\nabla}_\nu\phi \tag{4.8}$$

with a transformation law which follows from equations (4.4) and (4.6):

$$\delta_0 h_k{}^\mu = \delta_0^{\mathrm{P}} h_k{}^\mu - \rho h_k{}^\mu. \tag{4.9}$$

The Weylean extension of PGT is accompanied by two effects:

(a) a new compensating field (B) appears; and
(b) the transformation rules for 'old' fields A and h are changed.

Therefore, A and h have *different* properties from those in PGT.

Since the covariant derivative (4.5) depends only on the Lorentz and dilatation transformation properties of matter fields, it can be naturally extended to an arbitrary field. Thus, for instance,

$$\overset{*}{\nabla}_\mu h_i{}^\nu = (\partial_\mu + \tfrac{1}{2}A^{mn}{}_\mu\Sigma^1_{mn} - B_\mu)_i{}^s h_s{}^\nu = \partial_\mu h_i{}^\nu - A^s{}_{i\mu}h_s{}^\nu - B_\mu h_i{}^\nu.$$

Matter field Lagrangian. After introducing the covariant derivative $\overset{*}{\nabla}_k \phi$ we can define the Lagrangian

$$\mathcal{L}'_{\mathrm{M}} = \mathcal{L}_{\mathrm{M}}(\phi, \overset{*}{\nabla}_k \phi) \tag{4.10}$$

obeying the invariance condition

$$\delta_0 \mathcal{L}'_{\mathrm{M}} + \xi \cdot \partial \mathcal{L}'_{\mathrm{M}} + 4\rho \mathcal{L}'_{\mathrm{M}} = 0$$

which is different from (2.11). In order to compensate for the property $\partial_\mu \xi^\mu \neq 4\rho$, we perform another modification of the Lagrangian:

$$\widetilde{\mathcal{L}}_{\mathrm{M}} = \Lambda \mathcal{L}'_{\mathrm{M}} \tag{4.11}$$

where Λ is a function of the new fields. Condition (2.11) implies

$$\delta_0 \Lambda + \partial_\mu (\Lambda \xi^\mu) - 4\rho\Lambda = 0.$$

This equation contains local parameters and their derivatives. Coefficients of second derivatives of parameters vanish since Λ does not depend on field derivatives. Coefficients of $\omega^{ij}{}_{,\mu}$ and $\rho_{,\mu}$ vanish if Λ is independent of $A^{ij}{}_\mu$ and B_μ, while coefficients of ξ^μ vanish since Λ does not depend explicitly on x. Finally, if the coefficients of $\xi^\mu{}_{,\nu}$ and ω^{ij} vanish, this gives the same conditions as those in PGT, while if the coefficients of ρ vanish:

$$\rho: \quad \frac{\partial \Lambda}{\partial h_k{}^\mu} h_k{}^\mu + 4\rho\Lambda = 0.$$

The solution to this equation reads as $\Lambda = \det(b^k{}_\mu)$, so that the final matter field Lagrangian takes the form

$$\widetilde{\mathcal{L}}_{\mathrm{M}} \equiv b\mathcal{L}_{\mathrm{M}}(\phi, \overset{*}{\nabla}_k \phi). \tag{4.12}$$

Example 1. The Weyl covariant derivative of the scalar field φ has the form $\overset{*}{\nabla}_k \varphi = h_k{}^\mu \overset{*}{\nabla}_\mu \varphi = h_k{}^\mu (\partial_\mu - B_\mu)\varphi$, since $d(\varphi) = -1$. After localizing the Weyl symmetry, the Lagrangian of the massless φ^4 theory becomes

$$\widetilde{\mathcal{L}}_{\mathrm{S}} = b(\tfrac{1}{2}\eta^{ij} \overset{*}{\nabla}_i \varphi \overset{*}{\nabla}_j \varphi + f\varphi^4) = b(\tfrac{1}{2} g^{\mu\nu} \overset{*}{\nabla}_\mu \varphi \overset{*}{\nabla}_\nu \varphi + f\varphi^4).$$

Analogously, the WGT of the massless Dirac field is determined by

$$\widetilde{\mathcal{L}}_{\mathrm{D}} = \tfrac{1}{2} \mathrm{i} b [\bar{\psi}\gamma^k \overset{*}{\nabla}_k \psi - (\overset{*}{\nabla}_k \bar{\psi})\gamma^k \psi]$$

where $\overset{*}{\nabla}_k \psi = h_k{}^\mu (\partial_\mu + \frac{1}{2} A^{ij}{}_\mu \sigma_{ij} - \frac{3}{2} B_\mu)\psi$, and similarly for $\overset{*}{\nabla}_k \bar{\psi}$, and we use $d(\psi) = d(\bar{\psi}) = -\frac{3}{2}$. Since the field B_μ interacts in the same manner with both

ψ and $\bar{\psi}$, it cannot be interpreted as the electromagnetic potential. In fact, B_μ completely disappears from $\widetilde{\mathcal{L}}_{\rm D}$.

Example 2. The free electromagnetic field in M_4 has a dilatation symmetry with $d(A_i) = -1$, so that

$$\overset{*}{\nabla}_\mu A_i = (\partial_\mu + \tfrac{1}{2}A^{mn}{}_\mu \Sigma^1_{mn} - B_\mu)_i{}^s A_s.$$

It should be observed that we are using the potential A_i, not A_μ. This is so because matter fields belong to representations of $W(1,3)$ which 'live' in the *tangent* space of spacetime. The gauge-invariant Lagrangian has the form

$$\widetilde{\mathcal{L}}_{\rm EM} = -\tfrac{1}{4}b\eta^{ik}\eta^{jl}G_{ij}G_{kl} \qquad G_{ij} \equiv \overset{*}{\nabla}_i A_j - \overset{*}{\nabla}_j A_i.$$

If we go over to the coordinate basis we can define the potential $A_\mu = b^i{}_\mu A_i$, such that $d(A_\mu) = 0$. Sometimes, if A_μ is not distinguished from A_i, we can reach the erroneous conclusion that d^* is always equal to d, except for the electromagnetic potential. The relation $d^* = d$ is correct for *all fields* defined with respect to a local tangent frame, where $W(1,3)$ operates. The transition to the coordinate frame will be discussed in the next subsection.

Field strengths. In order to find the form of the free Lagrangian for the new fields (A, h, B), we shall first define the related field strengths. The commutator of two (ω, ρ)-covariant derivatives has the form

$$[\overset{*}{\nabla}_\mu, \overset{*}{\nabla}_\nu]\phi = \tfrac{1}{2}F^{ij}{}_{\mu\nu}\Sigma_{ij}\phi + F_{\mu\nu}d\phi \tag{4.13}$$

where $F^{ij}{}_{\mu\nu} = F^{ij}{}_{\mu\nu}(A)$ and $F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$ are the Lorentz and dilatation field strengths, respectively. The transition to $\overset{*}{\nabla}_k$ yields

$$[\overset{*}{\nabla}_k, \overset{*}{\nabla}_l]\phi = \tfrac{1}{2}F^{ij}{}_{kl}\Sigma_{ij}\phi + F_{kl}d\phi - \overset{*}{F}{}^s{}_{kl}\overset{*}{\nabla}_s\phi \tag{4.14}$$

where $\overset{*}{F}{}^i{}_{\mu\nu} = \overset{*}{\nabla}_\mu b^i{}_\nu - \overset{*}{\nabla}_\nu b^i{}_\mu$ is the translation field strength, and

$$\begin{aligned}
F^{ij}{}_{kl} &= h_k{}^\mu h_l{}^\nu F^{ij}{}_{\mu\nu} \\
F_{kl} &= h_k{}^\mu h_l{}^\nu F_{\mu\nu} \\
\overset{*}{F}{}^i{}_{kl} &= h_k{}^\mu h_l{}^\nu \overset{*}{F}{}^i{}_{\mu\nu}.
\end{aligned}$$

The field strengths transform under Weyl transformations as follows:

$$\begin{aligned}
\delta_0 F^{ij}{}_{kl} &= \delta_0^{\rm P} F^{ij}{}_{kl} - 2\rho F^{ij}{}_{kl} \\
\delta_0 F_{kl} &= \delta_0^{\rm P} F_{kl} - 2\rho F_{kl} \\
\delta_0 \overset{*}{F}{}^i{}_{kl} &= \delta_0^{\rm P} \overset{*}{F}{}^i{}_{kl} - \rho \overset{*}{F}{}^i{}_{kl}.
\end{aligned}$$

If a quantity X transforms according to the rule

$$\delta_0 X = \delta_0^{\mathrm{P}} X + w\rho X$$

we say that it has *weight* w. Thus, $w(\phi) = d$, $w(F^{ij}{}_{kl}) = w(F_{kl}) = -2$, and $w(\overset{*}{F}{}^{i}{}_{kl}) = -1$. The covariant derivative of a scalar field φ of weight d is

$$\overset{*}{\nabla}_\mu \varphi = (\partial_\mu + dB_\mu)\varphi \equiv \partial^*_\mu \varphi$$

while for a vector V^i of weight d we have

$$\begin{aligned}
\overset{*}{\nabla}_\mu V^i = \partial^*_\mu V^i + A^i{}_{j\mu} V^j &= \partial_\mu V^i + \overset{*}{A}{}^i{}_{j\mu} V^j \\
\overset{*}{A}{}^i{}_{j\mu} &\equiv A^i{}_{j\mu} + d\delta^i_j B_\mu .
\end{aligned}$$

If we define $\overset{*}{F}{}^{ij}{}_{\mu\nu} \equiv F^{ij}{}_{\mu\nu}(\overset{*}{A})$, it is easy to show that

$$\begin{aligned}
\overset{*}{F}{}^{ij}{}_{\mu\nu} &= F^{ij}{}_{\mu\nu}(A) + d\eta^{ij} F_{\mu\nu} \\
\overset{*}{F}{}^{i}{}_{\mu\nu} &= F^{i}{}_{\mu\nu}(A) + (B_\mu b^i{}_\nu - B_\nu b^i{}_\mu).
\end{aligned}$$

The quantity $\overset{*}{F}{}^{ij}{}_{\mu\nu}$ is not antisymmetric with respect to (i, j) as in PGT, which has a definite geometric meaning.

The complete Lagrangian of the matter and gauge fields is given by

$$\mathcal{L} = b\mathcal{L}_{\mathrm{F}}(F^{ij}{}_{kl}, \overset{*}{F}{}^{i}{}_{kl}, F_{kl}) + b\mathcal{L}_{\mathrm{M}}(\phi, \overset{*}{\nabla}_k \phi). \tag{4.15}$$

The free Lagrangian $\mathcal{L}_{\mathrm{F}}$ is an invariant density of weight -4, which means that it can be quadratic in F_{ijkl} and F_{kl}, while terms linear in $F^{ij}{}_{ij}$ and quadratic in $\overset{*}{F}{}^{i}{}_{kl}$ are not allowed. Thus, Weyl invariance restricts some possibilities that exist in PGT: those terms that contain dimensional constants are forbidden in $\mathcal{L}_{\mathrm{F}}$.

Conservation laws and field equations

We shall now discuss differential conservation laws in WGT.

1. Gauge transformations of the fields $Q_A = (\phi, b^k{}_\mu, A^{ij}{}_\mu, B_\mu)$ are determined by equations (4.2*a*), (4.7) and

$$\delta_0 b^k{}_\mu = \delta_0^{\mathrm{P}} b^k{}_\mu + \rho b^k{}_\mu . \tag{4.16}$$

Using the method given in chapter 3, the gauge invariance of the Lagrangian can be expressed as

$$\Delta\mathcal{L} = -\xi^\nu I^*_\nu + \tfrac{1}{2}\omega^{ij} I^*_{ij} + \rho I^* + \partial_\mu \Lambda^* = 0. \tag{4.17}$$

This relation implies that

$$I^*_\nu = 0 \qquad I^*_{ij} = 0 \qquad I^* = 0 \tag{4.18a}$$

$$\partial_\mu \Lambda^{*\mu} = 0. \tag{4.18b}$$

Here, as in PGT, we restrict ourselves to the case $\mathcal{L} = \widetilde{\mathcal{L}}_{\mathrm{M}}$.

2. *Canonical* and *covariant* energy–momentum and spin currents are determined by equations (3.19) and (3.20), in which $\nabla\phi \to \overset{*}{\nabla}\phi$, while the related *dynamical* currents are defined as in (3.21). The dilatation currents are

$$\widetilde{D}^\mu = \frac{\partial \widetilde{\mathcal{L}}_{\mathrm{M}}}{\partial \phi_{,\mu}}\,\mathrm{d}\phi \qquad \widetilde{D}'^\mu = \frac{\partial \widetilde{\mathcal{L}}_{\mathrm{M}}}{\partial \overset{*}{\nabla}_\mu \phi}\,\mathrm{d}\phi \qquad \delta^\mu = \frac{\delta \widetilde{\mathcal{L}}_{\mathrm{M}}}{\delta B_\mu}. \tag{4.19}$$

We assume the matter field equations to be fulfilled. Then, demanding that the coefficients of $\partial\xi$, $\partial\omega$ and $\partial\rho$ in (4.18b) vanish, we obtain equality of the covariant and dynamical currents:

$$\tau^\mu{}_\nu = \widetilde{T}'^\mu{}_\nu \qquad \sigma^\mu{}_{ij} = \widetilde{S}'^\mu{}_{ij} \qquad \delta^\mu = \widetilde{D}'^\mu. \tag{4.20}$$

Of course, $\widetilde{T}'$ and $\widetilde{S}'$ are different from the related PGT expressions.

Conditions (4.18a) lead to the following differential conservation laws:

$$\begin{aligned} b^k{}_\mu \overset{*}{\nabla}_\nu \tau^\nu{}_k &= \tau^\nu{}_k F^k{}_{\mu\nu} + \tfrac{1}{2}\sigma^\nu{}_{ij} F^{ij}{}_{\mu\nu} + \delta^\nu F_{\mu\nu} \\ \overset{*}{\nabla}_\mu \sigma^\mu{}_{ij} &= \tau_{ij} - \tau_{ji} \\ \overset{*}{\nabla}_\mu \delta^\mu &= \tau^\mu{}_\mu. \end{aligned} \tag{4.21}$$

The last equation should be compared with (2.41).

3. The equations of motion for the matter fields, obtained from the action (4.15), can be written in the Weyl-covariant form:

$$\delta\phi: \quad \frac{\bar{\partial}\widetilde{\mathcal{L}}_{\mathrm{M}}}{\partial\phi} - \overset{*}{\nabla}_\mu \frac{\partial \widetilde{\mathcal{L}}_{\mathrm{M}}}{\partial \overset{*}{\nabla}_\mu \phi} = 0 \tag{4.22}$$

where $\bar{\partial}\widetilde{\mathcal{L}}_{\mathrm{M}}/\partial\phi = [\partial\widetilde{\mathcal{L}}_{\mathrm{M}}(\phi, \overset{*}{\nabla}_\mu u)/\partial\phi]_{u=\phi}$.

Example 3. The equations of motion for the massless φ^4 theory have the Weyl-covariant form:

$$\begin{gathered} -K\varphi + 4f\varphi^3 = 0 \\ K\varphi \equiv b^{-1}\overset{*}{\nabla}_\mu (b g^{\mu\nu} \overset{*}{\nabla}_\nu \varphi) = b^{-1}(\partial_\mu + B_\mu)(b g^{\mu\nu} \overset{*}{\nabla}_\nu \varphi). \end{gathered}$$

Using these equations and the expressions for currents,

$$\widetilde{D}^{\mu} = \widetilde{D}'^{\mu} = \delta^{\mu} = -bg^{\mu\nu}\varphi\overset{*}{\nabla}_{\nu}\varphi$$

$$\tau_{\mu\nu} = b\overset{*}{\nabla}_{\mu}\varphi\overset{*}{\nabla}_{\nu}\varphi - g_{\mu\nu}\widetilde{\mathcal{L}}_{\mathrm{S}} \qquad \tau^{\mu}{}_{\mu} = -bg^{\mu\nu}\overset{*}{\nabla}_{\mu}\varphi\overset{*}{\nabla}_{\nu}\varphi - 4bf\varphi^{4}$$

we come to the result

$$\overset{*}{\nabla}_{\mu}\delta^{\mu} = -bg^{\mu\nu}\overset{*}{\nabla}_{\mu}\varphi\overset{*}{\nabla}_{\nu}\varphi - \varphi\overset{*}{\nabla}(bg^{\mu\nu}\overset{*}{\nabla}_{\nu}\varphi) = \tau^{\mu}{}_{\mu}$$

in conformity with (4.21).

For the Dirac field we have $\widetilde{D}^{\mu} = \widetilde{D}'^{\mu} = \delta^{\mu} = 0$. This unusual result stems from the fact that there is no interaction of B_{μ} with the Dirac field (example 1). The equations of motion imply $\tau^{\mu}{}_{\mu} = 0$, in agreement with $\delta^{\mu} = 0$.

The typical action of WGT is quadratic in field strengths; the gravitational field equations are obtained in the usual way.

Conformal versus Weyl gauge symmetry

In order to localize the conformal symmetry, we can start from the global conformal field transformations,

$$\delta_0\phi = (\tfrac{1}{2}\omega \cdot \Sigma + \rho\Delta + c \cdot K + \xi \cdot P)\phi \equiv K\phi$$

define the (ω, ρ, c)-covariant derivative,

$$\overset{c}{\nabla}_{\mu}\phi = (\partial_{\mu} + \tfrac{1}{2}A^{ij}{}_{\mu}\Sigma_{ij} + B_{\mu}\Delta + C^{i}{}_{\mu}\kappa_{i})\phi \tag{4.23}$$

find the transformation laws of the compensating fields, etc. However, by observing that the generator of SCT can be expressed in terms of D, P_{μ} and $\Sigma_{\mu\nu}$ as $K_{\mu} = -2x_{\mu}D + x^{2}P_{\mu} + 2x^{\nu}\Sigma_{\mu\nu}$, we can rewrite $\delta_0\phi$ in the form

$$\delta_0\phi = (\tfrac{1}{2}\bar{\omega} \cdot \Sigma + \bar{\rho}\Delta + \xi \cdot P)\phi$$

where $\bar{\omega}^{ij} = \omega^{ij} + 2(c^{i}x^{j} - c^{j}x^{i})$ and $\bar{\rho} = \rho - 2c \cdot x$.

Surprisingly, $\delta_0\phi$ contains only 11 independent parameters: $\bar{\omega}^{ij}$, $\bar{\rho}$ and ξ^{ν}. Local SCTs are no longer independent, they are reduced to local Lorentz rotations, dilatations and translations. If this is so, is it then necessary to introduce the compensating field $C^{i}{}_{\mu}$, corresponding to local SCT? And if we introduce this field, is it really independent from other fields? To answer these questions, let us recall that, basically, compensating fields are introduced to compensate for the 'extra' terms appearing in $\delta_0\partial_k\phi$ after the symmetry is localized (usually each local parameter demands one compensating field). Now, under global conformal transformations, we have

$$\delta_0\partial_k\phi = K\partial_k\phi + [2(c^{i}\Sigma_{ik} - c_k\Delta) + (\bar{\omega}_{k}{}^{\nu} + \bar{\rho}\delta_{k}^{\nu})P_{\nu}]\phi.$$

After the symmetry has been localized, $\delta_0 \partial_k \phi$ will be expressed in terms of the derivatives of $\bar{\omega}^{ij}$, $\bar{\rho}$, ξ^ν and c^i. Since the number of these parameters is not 11 but 15, it follows that all 15 compensating fields are needed, so that conformal gauge theory is essentially different from WGT.

4.2 Weyl–Cartan geometry

In his attempts to incorporate electromagnetism into geometry, Weyl extended the Riemannian structure of spacetime by introducing a new geometric assumption—the assumption that the length of a vector has no absolute geometric meaning. As an introduction to this idea, we first give an overview of conformal transformations in Riemann space; then, we introduce the basic elements of Weyl geometry and clarify its relation to WGT.

Conformal transformations in Riemann space

Conformal transformations in M_4 define the conformal group $C(1,3)$. Let us now see how these transformations can be generalized when we go over to Riemann space V_4 (Fulton *et al* 1962).

Conformal transformations in V_4. Different formulations of conformal transformations may have different geometric interpretations.

Consider first the *conformal mapping* $f : V_4 \to V_4$, which can be interpreted as the conformal 'movement' of points in V_4 (active interpretation). If (P, Q) are two points in V_4 with coordinates $(x, x + \mathrm{d}x)$ in a local coordinate system S, then

$$\mathrm{d}s^2(P, Q) = g_{\mu\nu}(x)\, \mathrm{d}x^\mu\, \mathrm{d}x^\nu.$$

Under the action of f the points (P, Q) transform into $(\bar{P}, \bar{Q})$, the coordinates of which, in the *same* coordinate system S, are $(\bar{x}, \bar{x} + \mathrm{d}\bar{x})$. The mapping f is conformal if

$$\mathrm{d}s^2(\bar{P}, \bar{Q}) = s(P)\, \mathrm{d}s^2(P, Q) \tag{4.24a}$$

or, equivalently,

$$g_{\mu\nu}(\bar{x})\, \mathrm{d}\bar{x}^\mu\, \mathrm{d}\bar{x}^\nu = s(x) g_{\mu\nu}(x)\, \mathrm{d}x^\mu\, \mathrm{d}x^\nu \tag{4.24b}$$

where $s(x) > 0$.

Now we shall consider the description of points (P, Q) in *two* local coordinate systems, S and S'. Let $x' = F(x)$ be a coordinate transformation, such that the distance between (P, Q) in S' 'looks like' the distance between $(\bar{P}, \bar{Q})$ in S. In other words,

$$g'_{\mu\nu}(x')\, \mathrm{d}x'^\mu\, \mathrm{d}x'^\nu = s(x') g_{\mu\nu}(x')\, \mathrm{d}x'^\mu\, \mathrm{d}x'^\nu \tag{4.25a}$$

or

$$g'_{\mu\nu}(x') = s(x') g_{\mu\nu}(x'). \tag{4.25b}$$

This condition defines the group $\widetilde{C}(1,3)$ of conformal coordinate transformation in V_4 (passive interpretation).

Weyl rescaling. The set of Weyl (conformal) rescalings of the metric,

$$g_{\mu\nu} \to g^{\mathrm{r}}_{\mu\nu} = s(x) g_{\mu\nu} \equiv \mathrm{e}^{2\lambda(x)} g_{\mu\nu} \tag{4.26a}$$

together with the related transformations of all other dynamical variables,

$$\phi \to \phi^{\mathrm{r}} = [s(x)]^{w/2}\phi \equiv \mathrm{e}^{w\lambda}\phi \tag{4.26b}$$

defines the group W_{r} in V_4. As in M_4, the real number w is called the weight, or Weyl dimension, of the field. These transformations do not involve any change of coordinates and are completely different from $\widetilde{C}$. In M_4, the Weyl dimension is just another name for the scale dimension.

Example 4. Consider the following interaction between the scalar and Dirac field in V_4:

$$\widetilde{\mathcal{L}}_{\mathrm{I}} = f\sqrt{-g}\bar{\psi}\psi\varphi.$$

Since $w(g_{\mu\nu}) = 2$, $w(g^{\mu\nu}) = -2$, and $w(\sqrt{-g}) = 4$, it follows that Weyl rescaling with $w(\varphi) = -1$, $w(\psi) = w(\bar{\psi}) = -\frac{3}{2}$, is a symmetry of the interaction Lagrangian. However, the kinetic term for the scalar field,

$$\widetilde{\mathcal{L}}_{\mathrm{S}} = \tfrac{1}{2}\sqrt{-g}g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi$$

is not invariant, since W_{r} transformations are local. The symmetry can be restored by introducing gauge fields, as usual.

Example 5. The electromagnetic field in V_4 is described by the Lagrangian

$$\widetilde{\mathcal{L}}_{\mathrm{EM}} = -\tfrac{1}{4}\sqrt{-g}g^{\mu\rho}g^{\nu\lambda}G_{\mu\nu}G_{\rho\lambda} \qquad G_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu.$$

The replacement $\partial \to \nabla$ (with a symmetric connection) does not change the antisymmetric tensor $G_{\mu\nu}$. A simple counting of weights shows that $w(A_\mu) = 0$, which implies the invariance under W_{r}. The consistency of such an assignment of weights should be checked by considering the electromagnetic interaction with other fields. Thus, for instance, the interaction with a complex scalar field,

$$\widetilde{\mathcal{L}}_{\mathrm{I}} = \tfrac{1}{2}\sqrt{-g}g^{\mu\nu}\{e^2A_\mu A_\nu\varphi^*\varphi + \mathrm{i}eA_\mu[\varphi^*\partial_\nu\varphi - (\partial_\nu\varphi^*)\varphi]\}$$

is W_{r} invariant for $w(A_\mu) = 0$.

Weyl rescaling of the metric transforms a Riemann space into another Riemann space: $(V_4, g_{\mu\nu}) \mapsto (V_4, g^{\mathrm{r}}_{\mu\nu})$. This mapping defines a collection of Riemann spaces, mutually connected by metric rescalings:

$$V_4^{\mathrm{r}} = \{(V_4, g^{\mathrm{r}_i}_{\mu\nu})\}.$$

The length of a vector is not well defined in V_4^{r}; however, the *ratio* of the lengths of two vectors at the same point, as well as the *angle* between them, are well-defined concepts.

The Christoffel connection in $(V_4, g_{\mu\nu})$ transforms under Weyl rescaling according to the law

$$\left\{{\mu \atop \nu\rho}\right\}^{\mathrm{r}} = \left\{{\mu \atop \nu\rho}\right\} + \tfrac{1}{2}(\delta^\mu_\nu s_\rho + \delta^\mu_\rho s_\nu - g_{\nu\rho}s^\mu) \tag{4.27}$$

where $s_\mu = \partial_\mu \ln s$, $s^\mu = g^{\mu\nu}s_\nu$. Consequently, the Riemann curvature tensor also changes under W_{r}.

Weyl introduced the *conformal tensor*,

$$C_{ijkl} = R_{ijkl} - \tfrac{1}{2}(\eta_{ik}R_{jl} - \eta_{il}R_{jk} - \eta_{jk}R_{il} + \eta_{jl}R_{ik}) - \tfrac{1}{6}(\eta_{il}\eta_{jk} - \eta_{ik}\eta_{jl})R$$

which has the following important properties:

(a) $C^\mu{}_{\nu\lambda\rho}$ is invariant under W_{r}; and
(b) the trace of Weyl tensor is zero: $C^k{}_{jkl} = 0$.

A space with a vanishing Weyl tensor is called a conformally flat space. Since the Minkowski space is conformally flat, it follows from (a) that any Riemann space with metric $g_{\mu\nu} = s\eta_{\mu\nu}$ is conformally flat.

Conformal transformations in M_4. Let us now return to the group of conformal transformations $\widetilde{C}$ and denote by $\widetilde{C}_0$ the subgroup of $\widetilde{C}$ transforming the flat space (M_4, η) into the *flat* space $(M_4, s\eta)$:

$$\widetilde{C}_0: \quad R^\mu{}_{\nu\lambda\rho}(s\eta) = 0. \tag{4.28a}$$

Since $(M_4, s\eta)$ is conformally flat, $C^\mu{}_{\nu\lambda\rho}(s\eta) = 0$, the previous condition reduces to

$$R_{\nu\rho}(s\eta) = -s_{\nu,\rho} + \tfrac{1}{2}s_\nu s_\rho - \tfrac{1}{2}\eta_{\nu\rho}s^\lambda{}_{,\lambda} - \tfrac{1}{2}\eta_{\nu\rho}s^\lambda s_\lambda = 0. \tag{4.28b}$$

This is a restriction on $s(x)$, and on related coordinate transformations. For infinitesimal transformations we have $s(x) = 1 - \frac{1}{2}\partial \cdot \xi$, and equation (4.28*b*) is equivalent to the conformal Killing equation (2.20*b*) in M_4. The group $\widetilde{C}_0$ reduces to the group of global conformal transformations $C(1, 3)$, which we already know from chapter 2.

Connection between $\widetilde{C}$ and W_{r}. Although $\widetilde{C}$ and W_{r} are completely different groups, there is an interesting connection between them, as suggested by the related considerations in M_4 (section 2.2). The precise form of this connection is given by the following theorem (Fulton *et al* 1962):

> *If an action is invariant under general coordinate transformations, then W_{r} invariance implies $\widetilde{C}$ invariance.*

The content of this theorem can be seen more clearly from the following, simpler version (Zumino 1970):

If an action, given in Riemann space, is W_{r} invariant, then that action, restricted to M_4, is invariant under $C(1,3)$.

Let us illustrate this statement by considering scalar field theory. Since the action is defined in Riemann space, it is invariant under general coordinate transformations. In particular, the action is invariant under the global dilatations:

$$\xi^\mu = \rho x^\mu \qquad \delta_0 g_{\mu\nu} = -2\rho g_{\mu\nu} - \rho x\cdot\partial g_{\mu\nu} \qquad \delta_0\varphi = -\rho x\cdot\partial\varphi.$$

On the other hand, W_{r} invariance implies invariance under Weyl rescaling with $s(x) = 1 + 2\rho$,

$$\delta_0 g_{\mu\nu} = 2\rho g_{\mu\nu} \qquad \delta_0\varphi = -\rho\varphi.$$

Combining these transformations we conclude that the action must be invariant under

$$\xi^\mu = \rho x^\mu \qquad \delta_0 g_{\mu\nu} = -\rho x\cdot\partial g_{\mu\nu} \qquad \delta_0\varphi = -\rho(x\cdot\partial + 1)\varphi.$$

Going now to the limit $g\to\eta$ we find $\delta_0 g = 0$, therefore the action restricted to M_4 is invariant under global dilatations.

In a similar way we can deduce the invariance under SCT, in the limit $V_4\to M_4$, by choosing

$$\xi^\mu = c^\mu x^2 - 2c\cdot x x^\mu \qquad s(x) = 1 - 4c\cdot x.$$

Thus, instead of investigating the conformal symmetry in M_4, we can carry the action over to V_4 and study the restrictions imposed by W_{r} invariance.

Weyl space W_4

Riemann space V_4 is a manifold with a linear connection and metric such that

(a) the connection Γ is symmetric and
(b) the metricity condition holds true, $D_\mu(\Gamma)g_{\nu\lambda} = 0$.

It follows from here, as we have shown earlier, that the Riemannian connection is equal to the Christoffel symbol. If a vector is parallel transported along a closed curve in V_4, its *orientation* changes, while the *lengths* of the vectors and the *angles* between them remain fixed.

Weyl geometry. In his attempts to accommodate the electromagnetic field into the geometric structure of spacetime, Weyl generalized Riemannian geometry by allowing a greater freedom in the choice of metric. He assumed that the *lengths* of vectors can also be changed under parallel transport (Fulton *et al* 1962, Adler *et al* 1965). The idea is realized by demanding that an infinitesimal change of length is proportional to the length itself, i.e.

$$D(\Gamma)V^2 = (\varphi_\rho \,\mathrm{d}x^\rho)V^2 \tag{4.29a}$$

where $V^2 \equiv g_{\mu\nu}V^\mu V^\nu$, and the covariant vector φ_ρ defines the rule by which the length is changed. Since $D(\Gamma)V^\mu = 0$ under parallel transport, the previous condition is equivalent to the relation

$$D_\rho(\Gamma)g_{\mu\nu} = \varphi_\rho g_{\mu\nu} \tag{4.29b}$$

which violates the metricity postulate and is called the *semi-metricity* condition. This equation can be easily solved for the connection

$$\Gamma^\mu_{\nu\rho} = \left\{ {\mu \atop \nu\rho} \right\} - \tfrac{1}{2}(\delta^\mu_\nu \varphi_\rho + \delta^\mu_\rho \varphi_\nu - g_{\nu\rho}\varphi^\mu) \tag{4.30}$$

where we assumed that the connection is symmetric (no torsion). Such a connection defines *Weyl space* W_4. A Weyl space with torsion will be considered in the next section.

Parallel transport of a vector along a closed curve in W_4 defines the curvature in terms of the connection (4.30). The curvature tensor is no longer antisymmetric in the first two indices: $R_{(\mu\nu)\lambda\rho} = -g_{\mu\nu}D_{[\lambda}\varphi_{\rho]}$. Therefore, the parallel transport has the form (Hehl *et al* 1988)

$$\begin{aligned}
\Delta V_\mu &= (\Delta V_\mu)_{\mathrm{rot}} + (\Delta V_\mu)_{\mathrm{dil}} \\
(\Delta V_\mu)_{\mathrm{rot}} &= \tfrac{1}{2}R_{[\nu\mu]\lambda\rho}\Delta\sigma^{\lambda\rho}V^\nu \\
(\Delta V_\mu)_{\mathrm{dil}} &= \tfrac{1}{2}R_{(\nu\mu)\lambda\rho}\Delta\sigma^{\lambda\rho}V^\nu = -\tfrac{1}{2}\Delta\sigma^{\lambda\rho}(D_\lambda\varphi_\rho)V_\mu .
\end{aligned}$$

Each vector rotates and changes the length but the angles between vectors remain the same. In the case of affine manifolds, where $g_{\mu\nu}$ and $\Gamma^\mu_{\nu\lambda}$ are completely independent, the angles also change (Hehl and Šijački 1980).

Weyl geometry allows us to consider metric rescalings without changing the connection. Indeed, it follows from equations (4.27) and (4.30) that if a metric rescaling is followed by the gradient transformation of φ_ρ,

$$g_{\mu\nu}(x) \to s(x)g_{\mu\nu}(x) \qquad \varphi_\rho(x) \to \varphi_\rho(x) + \partial_\rho \ln s(x) \tag{4.31}$$

the connection remains the same. A conformal change of metric can be interpreted as a change in the length scale at every point of the manifold. In Weyl theory this change is treated as a local symmetry transformation, which we call, as before, the Weyl rescaling W_r.

If the Weyl vector is a pure gradient, $\varphi_\rho = -\partial_\rho\beta$, the conformal transformation

$$g_{\mu\nu} \to \bar{g}_{\mu\nu} = e^\beta g_{\mu\nu} \qquad \varphi_\rho \to \bar{\varphi}_\rho = \varphi_\rho + \partial_\rho\beta = 0$$

transforms Weyl space $W_4(\varphi_\rho, g_{\mu\nu})$ into a Riemannian one: $V_4(\bar{g}_{\mu\nu}) = W_4(0, \bar{g}_{\mu\nu})$. More generally, a Weyl geometry may be reduced to a Riemannian geometry if and only if the length of the vector does not change after parallel transport along an arbitrary closed curve,

$$\oint \frac{DV^2}{V^2} = \oint \varphi_\rho \, \mathrm{d}x^\rho = -\tfrac{1}{2} \int F_{\mu\nu} \, \mathrm{d}\sigma^{\mu\nu} = 0$$

where $F_{\mu\nu} \equiv \partial_\mu \varphi_\nu - \partial_\nu \varphi_\mu$. Therefore, the condition $F_{\mu\nu} = 0$ guarantees that W_4 may be reduced to V_4 by a suitable rescaling transformation (in simply connected regions). The tensor $F_{\mu\nu}$ is W_r invariant.

Weyl-covariant derivative. Since rescaling transformations are of particular importance in Weyl geometry, it is useful to define quantities having well-defined transformation properties under (4.31). This leads us to the concept of weight, as in (4.26). Thus, the weight of $g_{\mu\nu}$ is 2, $w(\mathrm{d}x^\mu) = 0$, $w(\mathrm{d}s) = 1$, etc. Observe, however, that we cannot ascribe a weight to the Weyl vector itself. Tensors of weight $w \neq 0$ are also called tensor densities or pseudotensors.

The covariant derivative in W_4 is introduced so that the lengths of vectors are changed under parallel transport according to the rule (4.29a). If V^ν is a vector of weight zero, than $D_\mu V^\nu$ is also of weight zero, i.e. invariant under Weyl rescaling (4.31). However, if, for instance, V^ν is of weight 2, then DV does not have a well-defined weight:

$$D_\mu V^\nu \to D_\mu (sV^\nu) = sD_\mu V^\nu + (\partial_\mu \ln s)(sV^\nu).$$

It is therefore useful to extend the definition of a covariant derivative in such a way that it does not change the weight of the object on which it acts. To illustrate the idea, consider a scalar ϕ of weight $w = 2$, $\phi^r = s\phi$, and define a new, Weyl-covariant derivative:

$$\overset{*}{D}_\mu \phi = (\partial_\mu - \tfrac{1}{2} w \varphi_\mu)\phi \equiv \partial^*_\mu \phi. \tag{4.32}$$

Under transformations (4.31) $\overset{*}{D}_\mu \phi$ changes according to

$$\overset{*}{D}_\mu \phi \to (\partial^*_\mu - \partial_\mu \ln s)s\phi = s\overset{*}{D}_\mu \phi$$

i.e. $\overset{*}{D}_\mu \phi$ also has weight 2. Generalization to a vector of weight 2 is simple:

$$\begin{aligned} \overset{*}{D}_\mu V^\nu &= \partial^*_\mu V^\nu + \Gamma^\nu_{\lambda\mu} V^\lambda = \partial_\mu V^\nu + \overset{*}{\Gamma}{}^\nu_{\lambda\mu} V^\lambda \\ \overset{*}{\Gamma}{}^\nu_{\lambda\mu} &\equiv \Gamma^\nu_{\lambda\mu} - \tfrac{1}{2} w \delta^\nu_\lambda \varphi_\mu. \end{aligned} \tag{4.33}$$

The connection $\overset{*}{\Gamma}$ ensures the W_r-covariance of the new covariant derivative.

It is interesting to observe that the connection $\Gamma^{\nu}_{\lambda\mu}$ in (4.30) is obtained from $\left\{{}^{\nu}_{\lambda\mu}\right\}$ by the replacement $\partial \to \partial^*$:

$$\Gamma^{\nu}_{\lambda\mu} = \left\{{}^{\nu}_{\lambda\mu}\right\}^* \equiv \left\{{}^{\nu}_{\lambda\mu}\right\}|_{\partial\to\partial^*} \qquad \partial^*_{\mu} g_{\lambda\rho} = (\partial_{\mu} - \varphi_{\mu}) g_{\lambda\rho}.$$

The operation $\overset{*}{D}_{\mu}$ is also covariant with respect to general coordinate transformations, since $\overset{*}{D}_{\mu}$ differs from D_{μ} by a four-vector. Note that Weyl's semi-metricity condition (4.29) can be rewritten as a W_{r}-covariant condition:

$$\overset{*}{D}_{\rho} g_{\mu\nu} = 0. \tag{4.34}$$

This structure of W_4 is sufficient to describe gravitation and tensor matter fields. Spinor matter can be introduced with the help of tetrads in a Weyl space with torsion.

Weyl–Cartan space Y_4

Starting with the space (L_4, g), in which the connection and metric are completely independent of each other, we can introduce the metricity condition and define Riemann–Cartan space U_4. If the metricity postulate is replaced by a weaker, semi-metricity condition (4.29), we obtain the *Weyl–Cartan space* Y_4 (a Weyl space with torsion) (Hayashi and Kugo 1979, Hehl *et al* 1988).

If the torsion vanishes, Y_4 becomes the Weyl space W_4, while $\varphi_{\mu} = 0$ transforms Y_4 into Riemann–Cartan space U_4; finally, $W_4 \to V_4$ if $\varphi_{\mu} \to 0$ (figure 4.1).

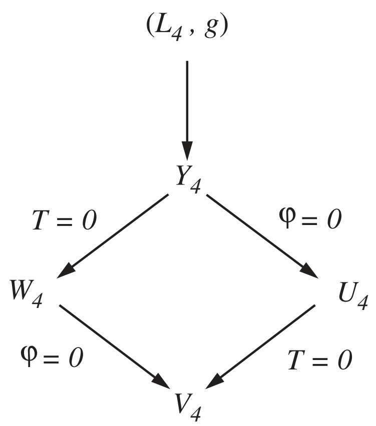

Figure 4.1. Weyl space does not satisfy the metricity condition.

The basic relations in Y_4 are obtained similarly to those in W_4. Using the semi-metricity condition, we find the following expression for the connection:

$$\Gamma^{\mu}_{\nu\rho} = \left\{{}^{\mu}_{\nu\rho}\right\} - \tfrac{1}{2}(\delta^{\mu}_{\nu}\varphi_{\rho} + \delta^{\mu}_{\rho}\varphi_{\nu} - g_{\nu\rho}\varphi^{\mu}) + K^{\mu}{}_{\nu\rho}. \tag{4.35}$$

The covariant derivative can also be extended to a $W_{\rm r}$-covariant form.

The spin connection. As we have already mentioned, the spinor matter is described with the help of the tetrads. From the relation $g_{\mu\nu} = \eta_{ij}e^{i}{}_{\mu}e^{j}{}_{\nu}$, we see that the property $w(g_{\mu\nu}) = 2$ can be obtained by demanding that

$$w(e^{i}{}_{\mu}) = 1 \qquad w(\eta_{ij}) = 0.$$

Let us now consider what the geometric properties of Y_4 look like in the tetrad basis. If the vector u^{μ} has weight 0, its tetrad components $u^{i} = e^{i}{}_{\mu}u^{\mu}$ have weight 1. Weyl's semi-metricity condition takes the form

$$D(\omega)(\eta_{ij}u^{i}u^{j}) = (\varphi_{\rho}dx^{\rho})(\eta_{ij}u^{i}u^{j}) \tag{4.36a}$$

where ω is the spin connection in Y_4. This relation, together with $D(\omega)\eta = 0$, yields

$$D_{\mu}(\omega)u^{i} = \tfrac{1}{2}\varphi_{\mu}u^{i} \tag{4.36b}$$

where $D_{\mu}(\omega)u^{i} \equiv \partial_{\mu}u^{i} + \omega^{i}{}_{j\mu}u^{j}$. From $D\eta = 0$ and the constancy of η it follows that the spin connection is *antisymmetric*:

$$\omega^{i}{}_{s\mu}\eta^{sj} + \omega^{j}{}_{s\mu}\eta^{is} = 0.$$

The Weyl-covariant derivative of the vector u^{i} of weight $w = 1$ has the form

$$\begin{aligned} \overset{*}{D}_{\mu}(\omega)u^{i} &= \partial^{*}_{\mu}u^{i} + \omega^{i}{}_{s\mu}u^{s} = \partial_{\mu}u^{i} + \overset{*}{\omega}{}^{i}{}_{s\mu}u^{s} \\ \overset{*}{\omega}_{ij\mu} &\equiv \omega_{ij\mu} - \tfrac{1}{2}w\eta_{ij}\varphi_{\mu} \end{aligned} \tag{4.37}$$

and is of the same weight as u^{i}. This can be easily extended to an arbitrary representation ϕ of the Lorentz group, with weight w:

$$\overset{*}{D}_{\mu}\phi = (\partial_{\mu} + \omega_{\mu} - \tfrac{1}{2}w\varphi_{\mu})\phi \qquad \omega_{\mu} \equiv \tfrac{1}{2}\omega^{ij}{}_{\mu}\Sigma_{ij}.$$

The relation between $\boldsymbol{\omega}$ and $\boldsymbol{\Gamma}$. Since $D(\Gamma)u^{\mu} = 0$ under parallel transport, Weyl's semi-metricity condition implies

$$D_{\mu}(\omega + \Gamma)e^{i}{}_{\nu} = \tfrac{1}{2}\varphi_{\mu}e^{i}{}_{\nu}. \tag{4.38a}$$

Thus, the vector lengths change under parallel transport because the 'standard of length' $e^{i}{}_{\mu}$ changes. Expressed in terms of $\overset{*}{D}$ this condition becomes

$$\overset{*}{D}_{\mu}(\omega + \Gamma)e^{i}{}_{\nu} \equiv \partial^{*}_{\mu}e^{i}{}_{\nu} + \omega^{i}{}_{s\mu}e^{s}{}_{\nu} - \Gamma^{\lambda}_{\nu\mu}e^{i}{}_{\lambda} = 0. \tag{4.38b}$$

It relates the connections ω and Γ and can be interpreted as the vanishing of the 'total' covariant derivative. Solving this equation for ω we obtain

$$\begin{aligned} \omega_{ij\mu} &= \overset{*}{\Delta}_{ij\mu} + K_{ij\mu} \\ \overset{*}{\Delta}_{ij\mu} &= \Delta_{ij\mu}|_{\partial\to\partial^*} = \Delta_{ij\mu} + \tfrac{1}{2}(\varphi_i b_{j\mu} - \varphi_j b_{i\mu}). \end{aligned} \tag{4.39a}$$

Denoting the Poincaré part of the connection by $\omega^{\mathrm{P}}_{ij\mu}$, we have

$$\omega_{ij\mu} = \omega^{\mathrm{P}}_{ij\mu} + \tfrac{1}{2}(\varphi_i b_{j\mu} - \varphi_j b_{i\mu}) = -\omega_{ji\mu}. \tag{4.39b}$$

The spin connection ω is W_{r} invariant, $w(\omega) = 0$, since it is expressed in terms of $\overset{*}{\Delta}$.

Geometric interpretation of WGT. Weyl gauge theory has the geometric structure of a Y_4 space. This is clearly seen from the previous discussion by establishing the following correspondence:

$$b^i{}_\mu \to e^i{}_\mu \qquad A^{ij}{}_\mu \to \omega^{ij}{}_\mu \qquad B_\mu \to -\tfrac{1}{2}\varphi_\mu \qquad d \to w. \tag{4.40}$$

Note that the curvature and torsion in Y_4 are not the same as in U_4, because $\overset{*}{\omega} \neq \omega$ and $\overset{*}{D} \neq D$.

Example 6. The local Weyl theory of the electromagnetic field is described by the Lagrangian

$$\widetilde{\mathcal{L}}_{\mathrm{EM}} = -\tfrac{1}{4} b\eta^{ik}\eta^{jl} G_{ij} G_{kl} \qquad G_{ij} \equiv \overset{*}{D}_i A_j - \overset{*}{D}_j A_i.$$

Using $\overset{*}{D}_i(\omega)A_j = h_i{}^\mu h_j{}^\nu \overset{*}{D}_\mu(\Gamma)A_\nu$, which follows from $\overset{*}{D}_\mu(\omega+\Gamma)h_i{}^\mu = 0$, and going to the coordinate basis, we obtain

$$\widetilde{\mathcal{L}}_{\mathrm{EM}} = -\tfrac{1}{4} b g^{\mu\rho} g^{\nu\lambda} G_{\mu\nu} G_{\rho\lambda} \qquad G_{\mu\nu} = (\partial_\mu A_\nu - \partial_\nu A_\mu) + 2K^\rho{}_{[\nu\mu]} A_\rho.$$

While the symmetric part of the connection (4.35) is cancelled in $G_{\mu\nu}$, the contribution of the torsion survives. If $K = 0$, then $G_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$.

4.3 Dynamics

The dynamical content of a theory is determined by the choice of the action. There are several ways to realize Weyl invariance dynamically.

(a) WGT is, in general, defined in Weyl–Cartan space Y_4.
(b) If the action is $W(1, 3)$ gauge invariant but the torsion vanishes, the geometry of the theory is described by Weyl space W_4.

(c) The action may possess $W(1,3)$ gauge invariance even in Riemann space V_4, in which case the gauge symmetry has an 'accidental' character.

All three cases can be defined, equivalently, by combining Poincaré gauge symmetry and Weyl rescaling.

The general gravitational action in WGT has the form

$$I = \int \mathrm{d}^4x\, b\mathcal{L}_{\mathrm{G}}(R^{ij}{}_{kl}, F_{kl}, \overset{*}{T}{}^{i}{}_{kl}). \tag{4.41a}$$

Since $w(b) = 4$, the Lagrangian $\mathcal{L}_{\mathrm{G}}$ has to be a scalar of weight $w = -4$. Moreover, the form of Weyl transformations of the field strengths implies $w(R_{ijkl}) = w(F_{kl}) = -2$, $w(\overset{*}{T}_{ikl}) = -1$, so that the general Weyl action is of an $R^2 + F^2$ type:

$$\begin{aligned}\mathcal{L}_{\mathrm{G}} = b_1 R_{ijkl}R^{ijkl} + b_2 R_{ijkl}R^{klij} + b_3 R_{ij}R^{ij} \\ + b_4 R_{ij}R^{ji} + b_5 R^2 + b_6(\varepsilon_{ijkl}R^{ijkl})^2 + cF_{ij}F^{ij}\end{aligned} \tag{4.41b}$$

where terms of the type R and $\overset{*}{T}{}^2$ are absent. Let us now consider some interesting cases of Weyl-invariant theories.

Weyl's theory of gravity and electrodynamics

At the beginning of the 20th century, gravity and electrodynamics were the only known basic interactions in nature. Weyl tried to unify these physical interactions on purely geometric grounds.

The electrodynamics in Riemann space V_4 is described by the action

$$I_V = \int \mathrm{d}^4x\, b(-aR + \alpha G_{\mu\nu}G^{\mu\nu}) \qquad G_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

where the electromagnetic field is not part of the geometry.

Weyl studied purely geometric action in W_4,

$$I_W = \int \mathrm{d}^4x\, b(-R^2 + \beta F_{\mu\nu}F^{\mu\nu}) \qquad F_{\mu\nu} = \partial_\mu\varphi_\nu - \partial_\nu\varphi_\mu \tag{4.42}$$

which seems to be the best possible analogy to I_V. Varying I_W we obtain

$$\delta I_W = \int \mathrm{d}^4x\, [-2bR\delta R - \delta b R^2 + \beta\delta(bF^2)] = 0.$$

At this point, we can violate W_{r} invariance by introducing a local scale of length, such that $R = \lambda$. Of course, this gauge condition will be destroyed by an arbitrary field variation, so that $\delta R \neq 0$ in general. Now, we can bring the previous equation into the form

$$\delta \int \mathrm{d}^4x\, b\left(-R + \frac{\beta}{2\lambda}F^2 + \frac{\lambda}{2}\right) = 0$$

where λ is the cosmological constant. Using connection (4.30) we can express the Weyl curvature as

$$R(W_4) = R(V_4) - \tfrac{3}{2}\varphi_\mu\varphi^\mu + 3\nabla_\mu\varphi^\mu.$$

Since $b\nabla_\mu\varphi^\mu = \partial_\mu(b\varphi^\mu)$, this term can be discarded as a surface term in the action, so that Weyl theory takes the following effective form:

$$I'_W = \int \mathrm{d}^4x\, b[-R(V_4) + \tfrac{1}{2}\beta\bar{F}^2 + \lambda(\tfrac{1}{2} + \tfrac{3}{2}\bar{\varphi}_\mu\bar{\varphi}^\mu)] \qquad (4.43)$$

where $\bar{\varphi}_\mu = \varphi_\mu/\sqrt{\lambda}$. The first two terms precisely correspond to the action I_V (in units $a = 1$), while the last term is a small correction because of the smallness of the cosmological constant λ. Thus, the structure of Weyl's theory is very similar to the classical electrodynamics in V_4 (Adler *et al* 1965).

What then is wrong with Weyl's unified theory? Although the effective action I'_W looks like a good candidate for the description of *free* electrodynamics, the properties of electromagnetic *interactions* lead to serious problems. Thus, for instance, the interaction of φ_μ with the Dirac fields ψ and $\bar{\psi}$ is the same since both fields have the *same* weight, $w = -\frac{3}{2}$, while the development of quantum theory showed that ψ and $\bar{\psi}$ have *opposite* electric charges and, therefore, different electromagnetic interactions. Similar arguments hold for the interaction with a complex scalar field. The electromagnetic interaction is correctly described by the minimal substitution,

$$\partial_\mu \rightarrow (\partial_\mu + \mathrm{i}eA_\mu)$$

which suggests identifying d with ie (and φ_μ with A_μ).

> *In electromagnetic interactions, local scale invariance is replaced by invariance under a local change of phase in the matter fields.*

Weyl's transition to the 'complex' covariant derivative led to the beginning of the modern development of gauge theories of internal symmetries.

Although Weyl's original idea was abandoned in electrodynamics, it found its natural place in gravitation. The experimental discovery that, in some processes at high energies, particle masses may be practically neglected, brought a revival of interest in scale-invariant theories (of course, at low energies scale symmetry is broken). If we want to include gravity in this structure, it is quite natural to consider gauge theories based on scale symmetry.

Scalar fields and the improved energy–momentum tensor

In Riemann spaces local rescaling symmetry W_r implies conformal symmetry $\widetilde{C}$. Therefore, the conformal properties of scalar field theory can be established by studying its behaviour under the transformations

$$\delta g_{\mu\nu}(x) = 2\lambda(x)g_{\mu\nu}(x) \qquad \delta\varphi(x) = -\lambda(x)\varphi(x). \qquad (4.44)$$

1. Massless scalar theory with φ^4 interaction in V_4,

$$I_a = \int \mathrm{d}^4x\, \sqrt{-g}(\tfrac{1}{2}g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi + f\varphi^4) \tag{4.45}$$

is not invariant under (4.44):

$$\delta I_a = \tfrac{1}{2}\int \mathrm{d}^4x\, \sqrt{-g}\varphi^2\Box\lambda \qquad \Box\lambda \equiv \nabla_\mu(\partial^\mu\lambda).$$

The invariance of the theory can be restored by adding a suitable term to the action. Such a term has the form

$$I_b = \tfrac{1}{12}\int \mathrm{d}^4x\, \sqrt{-g}R\varphi^2.$$

Indeed, starting with the relations

$$g_{\mu\nu} \to s g_{\mu\nu} \qquad R \to s^{-1}\left(R - 3\frac{\Box s}{s} + \frac{3}{2}g^{\mu\nu}\frac{\partial_\mu s\partial_\nu s}{s^2}\right)$$

we find that

$$\delta R = -2\lambda R - 6\Box\lambda \qquad \delta(R\varphi^2) = -4\lambda R\varphi^2 - 6\varphi^2\Box\lambda$$

which implies that

$$I_1 \equiv I_a + I_b = \int \mathrm{d}^4x\, \sqrt{-g}(\tfrac{1}{2}g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi + f\varphi^4 + \tfrac{1}{12}R\varphi^2) \tag{4.46}$$

is a rescaling-invariant extension of the scalar φ^4 theory in V_4.

In Minkowski space M_4, conformal invariance is concisely described by the improved energy–momentum tensor. We shall show that the energy–momentum tensor of the action I_1 yields, after transition to M_4, nothing other than the improved energy–momentum tensor of the scalar φ^4 theory.

The dynamical energy–momentum tensor of scalar theory (4.45) is given by

$$\tau^{(a)}_{\mu\nu} = \frac{2}{\sqrt{-g}}\frac{\delta I_a}{\delta g^{\mu\nu}} = \partial_\mu\varphi\partial_\nu\varphi - g_{\mu\nu}(\tfrac{1}{2}g^{\lambda\rho}\partial_\lambda\varphi\partial_\rho\varphi + f\varphi^4).$$

Here, the factor $(\sqrt{-g})^{-1}$ is, for simplicity, included in the definition of τ. The contribution of the additional term $R\varphi^2$ under the variation of $g_{\mu\nu}$ has the form

$$\delta\int \mathrm{d}^4x\, \sqrt{-g}R\varphi^2 = \int \mathrm{d}^4x\, \sqrt{-g}\varphi^2(G_{\mu\nu}\delta g^{\mu\nu} + g^{\mu\nu}\delta R_{\mu\nu})$$

where $G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R$. Using the relation

$$\sqrt{-g}g^{\mu\nu}\delta R_{\mu\nu} = \partial_\rho(\sqrt{-g}w^\rho) \qquad w^\rho \equiv g^{\mu\nu}\delta\Gamma^\rho_{\mu\nu} - g^{\mu\rho}\delta\Gamma^\tau_{\mu\tau}$$

and performing partial integrations, we obtain

$$\int \mathrm{d}^4x\,\sqrt{-g}g^{\mu\nu}\delta R_{\mu\nu}\varphi^2 = -\int \mathrm{d}^4x\,\sqrt{-g}(\nabla_\mu\nabla_\nu - g_{\mu\nu}\Box)\varphi^2\delta g^{\mu\nu}.$$

Thus, the energy–momentum tensor of the action I_1 takes the form

$$\begin{aligned}\theta^{(1)}_{\mu\nu} &= \tfrac{1}{6}G_{\mu\nu}\varphi^2 + \theta^{(a)}_{\mu\nu}\\ \theta^{(a)}_{\mu\nu} &\equiv \tau^{(a)}_{\mu\nu} - \tfrac{1}{6}(\nabla_\mu\nabla_\nu - g_{\mu\nu}\Box)\varphi^2.\end{aligned} \tag{4.47}$$

After going over to M_4, only the term $\theta^{(a)}_{\mu\nu}$, which represents the improved energy–momentum tensor of theory (4.45), remains.

2. The invariant action I_1 can be considered as a simple theory describing the dynamics of scalar and gravitational fields. Varying I_1 with respect to φ and $g_{\mu\nu}$ yields the following equations of motion:

$$\begin{aligned}-\Box\varphi + \tfrac{1}{6}R\varphi + 4f\varphi^3 &= 0\\ \tfrac{1}{6}G_{\mu\nu}\varphi^2 + \theta^{(a)}_{\mu\nu} &= 0.\end{aligned}$$

These equations possess an unusual property. Namely, the trace of the second equation has the form

$$\varphi(-\Box\varphi + \tfrac{1}{6}R\varphi + 4f\varphi^3) = 0$$

which means that either $\varphi = 0$ or the first equation is a consequence of the second one.

The solution $\varphi = 0$ is not interesting, since the gravitational equation in this case becomes trivial. If we could impose the classical condition $\varphi = v \neq 0$, the gravitational equation would be reduced to Einstein's form. This would solve the problem of a long-range limit for the theory, an important problem in the construction of realistic Weyl theory. However, the equation $\varphi = v$ explicitly violates Weyl invariance, which leads to certain problems at the quantum level. There is another mechanism, known as spontaneous symmetry breaking, which yields the equation $\varphi = v$ dynamically (as a solution of the classical equations of motion) and obeys quantum consistency requirements. Complex scalar fields play an interesting role in this mechanism (Domokos 1976).

This dependence of the equations of motion is not a specific property of this particular model, but is essentially a consequence of gauge invariance. The change of the action $I = I[\varphi, g_{\mu\nu}]$ under (4.44) has the form

$$\frac{\delta I}{\delta\lambda} = \frac{\delta I}{\delta g_{\mu\nu}}2g_{\mu\nu} - \frac{\delta I}{\delta\varphi}\varphi. \tag{4.48a}$$

If the action describes a complete theory, then W_r invariance of I implies that the trace of the equation of motion for $g_{\mu\nu}$ is proportional to the equation of motion for φ. On the other hand, if I describes matter fields, $I = I_{\mathrm{M}}$, then the previous equation becomes

$$\frac{\delta I_{\mathrm{M}}}{\delta\lambda} = \sqrt{-g}\tau^{\mu\nu}g_{\mu\nu} - \frac{\delta I_{\mathrm{M}}}{\delta\varphi}\varphi. \tag{4.48b}$$

Using the invariance of the action and the equations of motion for φ, we find that the trace of the energy–momentum tensor vanishes. These results shed a new light on the related considerations in M_4 (chapter 2).

3. In order to clarify the dynamical role of the improved energy–momentum tensor in gravity, we shall now try to find an answer to the following question:

Can the improved energy–momentum tensor be the source of gravity?

Consider Einstein's gravitation in interaction with massive scalar field theory, described by the action (Coleman 1973a)

$$\begin{aligned} I_2 &= \int \mathrm{d}^4x\,\sqrt{-g}[-aR + (\mathcal{L}_{\mathrm{S}} + \tfrac{1}{12}R\varphi^2)] \\ \mathcal{L}_{\mathrm{S}} &\equiv \tfrac{1}{2}g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi - \tfrac{1}{2}m^2\varphi^2 + f\varphi^4. \end{aligned} \tag{4.49}$$

This theory is *not* Weyl invariant due to the presence of aR and $m^2\varphi^2$ terms. The equations of motion for φ and $g_{\mu\nu}$ are

$$\begin{aligned} -(\Box + m^2)\varphi + \tfrac{1}{6}R\varphi + 4f\varphi^3 &= 0 \\ (-2a + \tfrac{1}{6}\varphi^2)G_{\mu\nu} &= -\theta^{(\mathrm{m})}_{\mu\nu} \end{aligned}$$

where $\theta^{(\mathrm{m})}_{\mu\nu}$ is the improved energy–momentum tensor of the massive, scalar field theory,

$$\theta^{(\mathrm{m})}_{\mu\nu} = \partial_\mu\varphi\partial_\nu\varphi - g_{\mu\nu}\mathcal{L}_{\mathrm{S}} - \tfrac{1}{6}(\nabla_\mu\nabla_\nu - g_{\mu\nu}\Box)\varphi^2.$$

Without the $R\varphi^2$ term in I_2, the equation of motion for φ satisfies the principle of equivalence: locally, in a convenient reference frame, it can be reduced to a form valid in the absence of gravity,

$$-(\Box + m^2)\varphi + 4f\varphi^3 = 0.$$

But then, the source of gravity in the second equation is the Belinfante energy–momentum tensor.

The presence of the $R\varphi^2$ term in I_2 implies that the source of gravity is the improved energy–momentum tensor, but the equation of motion for φ seems to violate the principle of equivalence: the $R\varphi$ term looks like the gravitational effect that cannot be transformed away by any choice of reference frame. However, this effect can be ignored classically. Indeed, the trace of the second equation of motion,

$$(-2a + \tfrac{1}{6}\varphi^2)(-R) = -(2m^2\varphi^2 - 4f\varphi^4 + \varphi\Box\varphi)$$

combined with the first one, leads to the relation $-2aR = m^2\varphi^2$. Substituting this into the first equation, we obtain

$$-(\Box + m^2)\varphi + 4\left(f - \frac{m^2}{48a}\right)\varphi^3 = 0$$

which is equivalent to the original equation for φ. The only effect of the $R\varphi$ term is to change the φ^3 interaction constant slightly, i.e. the short-range interaction of the scalar field with itself ($1/a$ is proportional to the gravitational constant). This change is irrelevant in the classical region.

Thus, in gravitational theory (4.49) the principle of equivalence *effectively* holds in the classical region. In this sense, the conventional theory of gravity interacting with scalar fields can be consistently deformed, so that the source of gravity is given by the improved energy–momentum tensor.

Goldstone bosons as compensators

We shall now show how an action that is not invariant under local rescalings can be extended and become invariant by introducing an additional scalar field—the compensator (Zumino 1970, Bergshoeff 1983, see also chapter 2). This field has a very simple transformation law so that, using the gauge symmetry of the extended theory, it can be easily eliminated by imposing an algebraic gauge condition. There are several reasons for such a modification of the original theory:

(a) the existence of higher symmetries is often more convenient in the process of quantization;
(b) the rules for treating higher symmetries are sometimes better known (conformal symmetry in supegravity); and
(c) the scalar field is helpful in transforming a nonlinear realization of symmetry into a linear one.

We begin by considering the mass term of the scalar field in V_4,

$$-\tfrac{1}{2}m^2\sqrt{-g}\varphi^2$$

which is not invariant under W_r since $\delta\sqrt{-g} = 4\lambda$, $\delta\varphi^2 = -2\lambda$. We may, however, introduce a new scalar field $\phi(x)$ of weight $w = -1$, and parametrize it in the form

$$\phi(x) = v\mathrm{e}^{\sigma(x)/v} \tag{4.50a}$$

where we assume $(\phi)_0 = v$, i.e. $\sigma(x)$ is the Goldstone boson. The transformation law of σ under W_r may be inferred from that of the field ϕ:

$$\delta\sigma = -v\lambda. \tag{4.50b}$$

It is now easy to see that the expression

$$-\tfrac{1}{2}m^2\sqrt{-g}\varphi^2\mathrm{e}^{2\sigma/v}$$

is W_r invariant, which justifies calling σ the compensator. Expanding the exponential in σ we obtain the mass term for φ, as well as some interaction terms.

An invariant action for the Goldstone boson σ itself may be obtained from equation (4.46) with $f = 0$, replacing φ by the expression (4.50a). Expanding the exponential in σ we obtain

$$I_\sigma = \int \mathrm{d}^4x\, \sqrt{-g}[\tfrac{1}{2}g^{\mu\nu}\partial_\mu\sigma\partial_\nu\sigma + \tfrac{1}{12}v^2 R(1 + 2\sigma/v + 2\sigma^2/v^2 + \cdots)].$$

The field σ is massless unless W_r symmetry is explicitly broken. We could try to produce the mass term for σ by adding the following W_r-invariant term to I_σ:

$$-\tfrac{1}{16}\mu^2 v^2 \sqrt{-g}\mathrm{e}^{4\sigma/v}.$$

On expanding in σ we find that the quadratic term looks like the mass term for σ. However, this expression is not acceptable since it does not vanish at the spatial infinity where $\sigma \to 0$. The mass term for σ can be generated by adding to the action a term that explicitly breaks W_r symmetry, such as

$$-\tfrac{1}{16}\mu^2 v^2 \sqrt{-g}(\mathrm{e}^{2\sigma/v} - 1)^2.$$

Our next example is the Dirac action in V_4. Although the kinetic term is W_r invariant, the mass term is not, but it can be easily generalized to the W_r-invariant form by adding a suitable compensator:

$$-m\sqrt{-g}\bar{\psi}\psi\mathrm{e}^{\sigma/v}.$$

It is interesting to note that Einstein's GR can also be extended in this way. Indeed, after introducing $\bar{g}_{\mu\nu} = g_{\mu\nu}\mathrm{e}^{2\sigma/v}$, the gauge-invariant extension of GR is given by

$$\bar{I}_\mathrm{E} = I_\mathrm{E}[\bar{g}] = -a\int \mathrm{d}^4x\, \sqrt{-\bar{g}}R(\bar{g}).$$

Returning to the original variables we obtain

$$\bar{I}_\mathrm{E} = -a\int \mathrm{d}^4x\, \sqrt{-g}\mathrm{e}^{\sigma/v}[R(g) - 6\Box]\mathrm{e}^{\sigma/v}.$$

Gauge symmetry allows us to choose the gauge condition $\sigma = 0$, which leads us back to the original theory.

The same construction can be applied to an arbitrary field theory. Let Φ_a be a set of fields, transforming under W_r according to $\Phi_a \to \Phi_a\mathrm{e}^{d_a\lambda}$, and $I[\Phi_a]$ their classical action. Then, the gauge-invariant extension of this theory has the form

$$\bar{I} = I[\bar{\Phi}_a] \qquad \bar{\Phi}_a \equiv \Phi_a\mathrm{e}^{d_a\sigma/v} \tag{4.51}$$

where the $\bar{\Phi}_a$ are new, W_r-invariant fields. Gauge symmetry enables us to return to the original action by choosing $\sigma = 0$. Of course, other gauge conditions are also possible.

A further simple example is the theory of massive vector field in M_4,

$$\mathcal{L}_V = -\tfrac{1}{4}G^2(A) + \tfrac{1}{2}m^2A^2 \qquad G_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{4.52a}$$

which is not invariant under $U(1)$ gauge transformations: $\delta A_\mu = \partial_\mu \Lambda$. After introducing a scalar field $B(x)$ transforming as $\delta B = -m\Lambda$, we can define a new, gauge-invariant field as

$$\bar{A}_\mu = A_\mu + \frac{1}{m}\partial_\mu B \qquad \delta\bar{A}_\mu = 0.$$

Then, a gauge-invariant extension of the original theory is given by

$$\bar{\mathcal{L}}_V = \mathcal{L}_V(\bar{A}) = -\tfrac{1}{4}G^2(A) + \tfrac{1}{2}m^2A^2 - m(\partial\cdot A)B + \tfrac{1}{2}(\partial_\mu B)^2.$$

Choosing $B = 0$ we return to the original formulation. Alternatively, gauge symmetry can be violated by adding to $\bar{\mathcal{L}}_V$ the non-invariant term

$$\mathcal{L}_{GF} = -\tfrac{1}{2}(\partial\cdot A - B)^2.$$

The resulting theory takes the form

$$\bar{\mathcal{L}}_V + \mathcal{L}_{GF} = \tfrac{1}{2}A_\mu(\Box + m^2)A^\mu - \tfrac{1}{2}B(\Box + m^2)B \tag{4.52b}$$

which is very suitable for quantization.

General remarks

Here we give some brief comments on some typical Weyl gauge theories.

1. The massless scalar φ^4 theory can be made Weyl invariant in Riemann space V_4 by adding the term $R\varphi^2$, as in equation (4.46). A more general class of theories is obtained by going to Weyl space W_4:

$$I_3 = \int \mathrm{d}^4x\, b\tfrac{1}{2}(g^{\mu\nu}\partial^*_\mu\varphi\partial^*_\nu\varphi + \omega\varphi^2 R - \tfrac{1}{2}F^2 + f\varphi^4)$$

where ω is a parameter. The Hamiltonian structure of this theory has been studied and used to clarify its relation to GR (Omote and Kasuya 1977).

2. There have been attempts to obtain GR as a long distance effective theory of the following fundamental (quantum) Weyl-invariant action in V_4:

$$I_4 = \int \mathrm{d}^4x\, \sqrt{-g}C_{ijkl}C^{ijkl}.$$

The curvature term R is induced by radiative corrections, breaking thereby the original Weyl gauge symmetry dynamically (Adler 1982, Zee 1983).

3. In Riemann space there are only three invariants quadratic in curvature: $R_{ijkl}R^{ijkl}$, $R_{ij}R^{ij}$ and R^2. Using the Gauss–Bonnet theorem,

$$\int \mathrm{d}^4x\,\sqrt{-g}(R_{ijkl}R^{ijkl} - 4R_{ij}R^{ij} + R^2) = \text{constant}$$

we find that only two of them are independent; for instance, R^2 and

$$C_{ijkl}C^{ijkl} = R_{ijkl}R^{ijkl} - 2R_{ij}R^{ij} + \tfrac{1}{3}R^2 = 2(R_{ij}R^{ij} - \tfrac{1}{3}R^2).$$

The quantity C^2 gives a Weyl gauge-invariant contribution to the action, while R^2 does not. However, introducing a new scalar field we can improve the R^2 term without going to W_4:

$$I_5 = \int \mathrm{d}^4x\,\sqrt{-g}\left(R - \frac{\Box\varphi}{\varphi}\right)^2.$$

The most general action, which is at most quadratic in curvature and has Weyl gauge symmetry in V_4, has the form (Antoniadis and Tsamis 1984, Antoniadis *et al* 1985)

$$I_5' = \alpha I_4 + \beta I_5 + \gamma I_1.$$

4. Weyl-invariant theory can be constructed not only in V_4, but also in U_4. By analysing the massless Dirac field and the gravitational action

$$I_6 = \int \mathrm{d}^4x\, b\varphi^2 R$$

we can show that the U_4 geometry, supplied with some additional (conformal) structure, provides a natural framework for the description of Weyl symmetry of matter and gravity, in which the trace of the torsion plays the role of the Weyl gauge field (Obukhov 1982).

The general structure of the U_4 theory is considerably restricted by demanding the invariance under complex rescalings (Fukui *et al* 1985).

5. There is an interesting discussion, based on the action

$$I_7 = I_1 + \int \mathrm{d}^4x\,\sqrt{-g}\,W_{\mu\nu}W^{\mu\nu} \qquad W_{\mu\nu} \equiv R^{\lambda}{}_{\lambda\mu\nu}$$

in Weyl space W_4, regarding the connection between Weyl invariance and the observed, effective Riemannian structure of the physical spacetime (Hochberg and Plunien 1991).

6. Now, we turn our attention to Weyl–Cartan space Y_4. Weyl's action (4.42) can be directly generalized to the Y_4 space (Charap and Tait 1974),

$$I_8 = \int d^4x\, b(\alpha R^2 + \beta F^2)$$

which is also true for I_3 (Kasuya 1975, Hayashi and Kugo 1979, Nieh 1982).

7. Conformal gauge theory is defined in terms of the gauge fields $A^{ij}{}_\mu$, $b^i{}_\mu$, B_μ and $C^i{}_\mu$, corresponding to the generators M_{ij}, P_i, D and K_i. Consider the action

$$I_9 = \int d^4x\, b E^{\mu\nu\lambda\rho} \varepsilon_{ijkl} R^{ij}{}_{\mu\nu}(M) R^{kl}{}_{\lambda\rho}(M)$$

where $E^{\mu\nu\lambda\rho} = \varepsilon^{\mu\nu\lambda\rho}/b$. If we impose the 'standard' condition $R^i{}_{\mu\nu}(P) = 0$ (no torsion), which relates $A^{ij}{}_\mu$ and $b^i{}_\mu$, and use the equation of motion for $C^i{}_\mu$, then

$$C_{\mu\nu} = -\tfrac{1}{4}(\widehat{R}_{\mu\nu} - \tfrac{1}{6} g_{\mu\nu} \widehat{R}) \qquad \widehat{R}_{\mu\nu} \equiv R_{\mu\nu}(A)|_{B=C=0}.$$

After eliminating gauge potentials $A^{ij}{}_\mu$ and $C^i{}_\mu$, the action takes the form

$$I_9' = \int d^4x\, b \widehat{C}^{\mu\nu\lambda\rho} \widehat{C}_{\mu\nu\lambda\rho}.$$

In this way, the original conformal gauge theory, with 15 gauge potentials, is reduced to the Weyl invariant theory in V_4, with four gauge potentials (De Wit 1981, Kaku 1982, see also Nepomechie 1984).

8. As far as the symmetry is concerned, the term $\varphi^2 T^2$ in the action is as good as $\varphi^2 R$, but it has not been studied in the context of Y_4 theories. The phenomenological value of such a term has been discussed in the framework of more general (L_4, g) geometry (Šijački 1982, Ne'eman and Šijački 1988).

9. In physical theories we find two very different scales, quantum (microscopic) and gravitational (macroscopic), the ratio of which determines the so-called big numbers. Trying to find a deeper explanation of these phenomena, Dirac studied the possibility that the factor of proportionality between quantum and gravitational units may be time dependent (Dirac 1973). This idea can be naturally considered in the framework of WGT, where physical laws are the same for the observers using instruments with different scales (Canuto *et al* 1976).

10. A fundamental problem concerning any Weyl theory is the question of its long-range limit, in which scale invariance should be broken. This problem may be solved by using the mechanism of spontaneous symmetry breaking, which obeys certain quantum consistency requirements. Scalar matter fields are here of particular importance, since they ensure spontaneous breaking of Weyl gauge symmetry. This is completely different from the role of the scalar field in the scalar–tensor theory of gravity (Brans and Dicke 1961, Smalley 1986, Kim 1986).

The vanishing of the cosmological constant can be treated in WGT as a dynamical question (Antoniadis and Tsamis 1984). The presence of R^2 terms in the action improves the quantum renormalizability properties of the theory, but, at the same time, leads to problems with unitarity. These issues, as well as the stability of the ground state, are very difficult problems, the solution of which are necessary to give a complete phenomenological foundation to Weyl theory.

Exercises

1. Formulate a gauge theory based on dilatation symmetry only and derive the transformation laws of the compensating fields.
2. Use the transformation rules for ϕ and $\overset{*}{\nabla}_k\phi$ in WGT to show that
$$\delta_0\mathcal{L}'_\mathrm{M} + \xi\cdot\partial\mathcal{L}'_\mathrm{M} + 4\rho\mathcal{L}'_\mathrm{M} = 0.$$
Then derive the invariance condition for $\widetilde{\mathcal{L}}_\mathrm{M} = b\mathcal{L}'_\mathrm{M}$.
3. Write the Euler–Lagrange field equations for the free (a) massless Dirac and (b) electromagnetic field, in WGT.
4. (a) Find the transformation law for $K^\mu = \partial\widetilde{\mathcal{L}}_\mathrm{M}/\partial\overset{*}{\nabla}_\mu\phi$, and define $\overset{*}{\nabla}_\nu K^\mu$.
(b) Show that the equations of motion for matter fields can be written in the Weyl-covariant form.
(c) Apply this result to the free (i) massless Dirac and (ii) electromagnetic field in WGT, and compare these calculations with those in exercise 3.
5. Find the transformation laws of the field strengths in WGT and determine their weights.
6. Derive the Bianchi identities in WGT.
7. Use the equations of motion to check the identity $\overset{*}{\nabla}_\mu\delta^\mu = \tau^\mu{}_\mu$ for the free (i) massless Dirac field and (ii) electromagnetic field (see example 3).
8. Use the general form of the matter Lagrangian, $\widetilde{\mathcal{L}}_\mathrm{M} = b\mathcal{L}_\mathrm{M}(\phi, \overset{*}{\nabla}_k\phi)$, and the related field equations to prove
(a) the equality of the covariant and dynamical dilatation currents; and
(b) the differential identity $\overset{*}{\nabla}_\mu\delta^\mu = \tau^\mu{}_\mu$.
9. Derive the equations of motion for WGT, defined by the action
$$I = \int \mathrm{d}^4x\, b(\alpha F^2 + \beta F_{ik}F^{ik})$$
where $F = F^{ij}{}_{ij}$, $F_{ik} = h_i{}^\mu h_k{}^\nu(\partial_\mu B_\nu - \partial_\nu B_\mu)$.
10. Consider a W_r-invariant scalar field theory in V_4. Show that after the transition to M_4, this theory becomes invariant under the global SCT.

11. Let $g^r_{\mu\nu} = s g_{\mu\nu}$.
    (a) Express $R^\rho{}_{\mu\lambda\nu}(g^r)$ in terms of $R^\rho{}_{\mu\lambda\nu}(g)$.
    (b) Prove the equation

$$R_{\mu\nu}(g^r) = R_{\mu\nu} - \nabla_\mu s_\nu - \tfrac{1}{2} g_{\mu\nu} \nabla_\lambda s^\lambda - \tfrac{1}{2} g_{\mu\nu} s_\lambda s^\lambda + \tfrac{1}{2} s_\mu s_\nu$$

    where $s_\mu = \partial_\mu \ln s$, $s^\mu = g^{\mu\nu} s_\nu$, and ∇_μ is the covariant derivative in $(V_4, g_{\mu\nu})$.
    (c) Derive the corresponding relation for R.
12. Consider conformal coordinate transformations in M_4, with some conformal factor s: $x' = x + \xi$, $\eta' = s\eta$. Show that the condition $R_{\mu\nu}(s\eta) = 0$ is equivalent to the conformal Killing equation (2.20*b*) in M_4.
13. The action that describes the interaction of the electromagnetic and complex scalar fields in V_4 is given by

$$I = \int \mathrm{d}^4 x \sqrt{-g} [\tfrac{1}{2} g^{\mu\nu} (\partial_\mu + \mathrm{i}eA_\mu)\varphi^* (\partial_\mu - \mathrm{i}eA_\mu)\varphi + \tfrac{1}{12} R\varphi^*\varphi - \lambda(\varphi^*\varphi)^2 - \tfrac{1}{4} g^{\mu\rho} g^{\nu\lambda} F_{\mu\nu} F_{\rho\lambda}].$$

    Show that it is invariant under Weyl rescalings.
14. Show that the Dirac action in V_4 is W_r invariant, except for the mass term.
15. Let $\bar{R}$ be the scalar curvature of Weyl space W_4, with the connection (4.30), and R be the corresponding Riemannian curvature. Prove the following identities:

$$\bar{R}_{(\mu\nu)\lambda\rho} = -g_{\mu\nu} \nabla_{[\lambda} \varphi_{\rho]}$$
$$\bar{R} = R + 3\nabla_\mu \varphi^\mu - \tfrac{3}{2} \varphi_\mu \varphi^\mu$$

    where ∇_μ is Riemannian covariant derivative.
16. Find the transformation rules of the Weyl tensor C_{ijkl} under
    (a) Weyl rescalings and
    (b) $W(1,3)$ gauge transformations.

Chapter 5

Hamiltonian dynamics

Classical dynamics is usually considered to be the first level in the development of our understanding of the physical laws in Nature, which today is based on quantum theory. Despite the many successes of quantum theory in describing basic physical phenomena, it is continually running into difficulties in some specific physical situations. Thus, all attempts to quantize the theory of gravity have encountered serious difficulties. In order to find a solution to these problems, it seems to be useful to reconsider the fundamental principles of classical dynamics. In this context, the principles of Hamiltonian dynamics are seen to be of great importance not only for a basic understanding of classical theory, but also for its quantization.

The theories of basic physical interactions, such as the electroweak theory or GR, are theories with gauge symmetries. In gauge theories the number of dynamical variables in the action is larger than the number of variables required by the physics. The presence of unphysical variables is closely related to the existence of gauge symmetries, which are defined by unphysical transformations of dynamical variables. Dynamical systems of this type are also called singular, and their analysis demands a generalization of the usual methods. In the Hamiltonian formalism they are characterized by the presence of constraints.

A systematic investigation of constrained Hamiltonian systems began more than 50 years ago, with the work of Bergmann, Dirac and others. The Hamiltonian formulation results in a clear picture of the physical degrees of freedom and gauge symmetries and enables a thorough understanding of constrained dynamics. The resulting classical structure had an important influence not only on the foundation of the canonical methods of quantization but also on the development of covariant path-integral quantization. In the first part of this chapter we shall present the basic ideas of Dirac's method (Dirac 1964, Hanson *et al* 1976, Sundermeyer 1982, Henneaux and Teitelboim 1992), and develop a systematic approach to the construction of the gauge generators on the basis of the known Hamiltonian structure (Castellani 1982).

PGT represents a natural extension of the gauge principle to spacetime symmetries. In the second section we present a Hamiltonian analysis of the general U_4 theory of gravity. This leads to a simple form of the gravitational Hamiltonian, representing a generalization of the canonical Arnowitt–Deser–Misner (ADM) Hamiltonian from GR (Arnowitt *et al* 1962, Nikolić 1984) and enables a clear understanding of the interrelation between the dynamical and geometric aspects of the theory. Then, in the third section, we analyse two specific but important examples: Einstein–Cartan theory without matter fields, which is equivalent to GR (Nikolić 1995), and teleparallel theory (Blagojević and Nikolić 2000, Blagojević and Vasilić 2000a). The first model represents the basis for a simple transition to Ashtekar's formulation of GR, in which encouraging results concerning the quantization of gravity are obtained (appendix E).

The results obtained here will be used in the next chapter to construct gauge generators and solve the important problem of the conservation laws in the U_4 theory of gravity.

5.1 Constrained Hamiltonian dynamics

Introduction to Dirac's theory

Primary constraints. In order to simplify the exposition of the Hamiltonian dynamics, we shall start by considering a classical system involving only a *finite* number of degrees of freedom, described by coordinates q_i $(i = 1, 2, \ldots, N)$ and the action

$$I = \int \mathrm{d}t\, L(q, \dot{q}). \tag{5.1}$$

To go over to the Hamiltonian formalism we introduce, in the usual way, the momentum variables:

$$p_i = \frac{\partial L}{\partial \dot{q}_i} \equiv f_i(q, \dot{q}) \qquad (i = 1, 2, \ldots, N). \tag{5.2}$$

In simple dynamical theories these relations can be inverted, so that *all* the velocities can be expressed in terms of the coordinates and momenta, whereupon we can simply define the Hamiltonian function. In this case relations (5.2) define the momenta as *independent* functions of the velocities. However, for many interesting theories, such as non-Abelian gauge theories or gravitational theories, such an assumption would be too restrictive. Therefore, we shall allow for the possibility that momentum variables are not independent functions of velocities, i.e. that there exist *constraints*:

$$\phi_m(q, p) = 0 \qquad (m = 1, 2, \ldots, P). \tag{5.3a}$$

The variables (q, p) are local coordinates of the phase space Γ on which the Hamiltonian dynamics is formulated. Relations (5.3a) are called *primary* *constraints*; they determine a subspace Γ_1 of Γ, in which a dynamical system develops over time.

The geometric structure of the subspace Γ_1 can be very complicated. We shall assume for simplicity that the rank of the matrix $\partial\phi_m/\partial(q,p)$ is *constant* on Γ_1. Moreover, we shall assume that the functions ϕ_m satisfy the following regularity condition:

> *All the constraint functions ϕ_m are independent*†*, i.e. the Jacobian $\mathcal{J} = \partial\phi_m/\partial(q,p)$ is of rank P on Γ_1.*

Accordingly, the dimension of Γ_1 is well defined and equal to $2N - P$.

The choice of functions ϕ_m, defining a given subspace Γ_1, is not unique. The regularity condition on the Jacobian imposes certain restrictions on the form of ϕ_m. Thus, e.g., the surface determined by $p_1 = 0$ can be equivalently written as $p_1^2 = 0$. However, since $\mathrm{rank}(\mathcal{J})$ in these two cases is one and zero, respectively, the second form is not admissible.

Weak and strong equalities. It is now useful to introduce the notions of weak and strong equality. Let $F(q,p)$ be a function which is defined and differentiable in a neighbourhood $\mathcal{O}_1 \subseteq \Gamma$ containing the subspace Γ_1. If the restriction of $F(q,p)$ on Γ_1 vanishes, we say that F is weakly equal to zero. Weak equality will be denoted by the symbol $\approx$:

$$F(q,p) \approx 0 \quad \Longleftrightarrow \quad F(q,p)|_{\Gamma_1} = 0.$$

If the function F and all its first derivatives vanish on Γ_1, then F is strongly equal to zero:

$$F(q,p) = 0 \quad \Longleftrightarrow \quad F,\ \partial F/\partial q,\ \partial F/\partial p|_{\Gamma_1} = 0.$$

For strong equality we shall use the usual equality sign. This definition will be especially useful in the analysis of the equations of motion which contain derivatives of functions on Γ_1.

By using these conventions, relations (5.3*a*), which define Γ_1, can be written as weak equalities:

$$\phi_m(q,p) \approx 0 \qquad (m = 1, 2, \ldots, P). \tag{5.3b}$$

This equality is not the strong one, since there exist non-vanishing derivatives of ϕ_m on Γ_1. Indeed, if we solve the constraints for P momentum variables and writes them as $\phi_m \equiv p_m - g_m(q_i, p_a)$, where $a = P+1, \ldots, N$, then $\partial\phi_m/\partial p_n = \delta_m^n$, and this does not vanish for $n = m$.

It is now interesting to clarify the relation between strong and weak equalities: if a phase-space function F vanishes weakly, $F \approx 0$, what can we

† Independent constraints are called irreducible, otherwise, if $\mathrm{rank}(\mathcal{J}) < P$, they are reducible. Our discussion will be restricted to the irreducible case.

say about the derivatives of F on Γ_1? By varying F on Γ_1 we find that

$$\delta F|_{\Gamma_1} = \left(\frac{\partial F}{\partial q_i}\delta q_i + \frac{\partial F}{\partial p_i}\delta p_i\right)\bigg|_{\Gamma_1} = 0.$$

Here we should take into account that $2N$ variations δq and δp are not independent but satisfy P constraints:

$$\frac{\partial \phi_m}{\partial q_i}\delta q_i + \frac{\partial \phi_m}{\partial p_i}\delta p_i \approx 0.$$

By use of the general method of the calculus of variations with constraints (ter Haar 1971), the previous variational problem leads to the equations

$$\frac{\partial F}{\partial q_i} - \lambda^m \frac{\partial \phi_m}{\partial q_i} \approx 0 \qquad \frac{\partial F}{\partial p_i} - \lambda^m \frac{\partial \phi_m}{\partial p_i} \approx 0$$

where λ^m are some multipliers. These $2N$ equations, together with (5.3), determine the conditions satisfied by $\partial F/\partial q$, $\partial F/\partial p$ and λ^m on Γ_1. They imply the relations

$$\frac{\partial}{\partial q_i}(F - \lambda^m \phi_m) \approx 0 \qquad \frac{\partial}{\partial p_i}(F - \lambda^m \phi_m) \approx 0$$

from which we deduce

$$F - \lambda^m \phi_m \approx \mathcal{O}$$

where $\mathcal{O}$ is a quantity with weakly vanishing derivatives: it can be zero, a constant or a second or higher power of a constraint. It is interesting to note that for theories in which the constraint functions ϕ_m satisfy this regularity condition, we can prove that $\mathcal{O} = 0$ (Henneaux and Teitelboim 1992); in other words,

If $F \approx 0$ *then* $F = \lambda^m \phi_m$.

Total Hamiltonian and the equations of motion. Having introduced these definitions, we now return to a further development of the Hamiltonian formalism. Let us consider the quantity

$$H_\mathrm{c} = p_i \dot{q}_i - L(q, \dot{q}). \tag{5.4}$$

By making variations in q and $\dot{q}$ we find that

$$\delta H_\mathrm{c} = \delta p_i \dot{q}_i + p_i \delta \dot{q}_i - \frac{\partial L}{\partial q_i}\delta q_i - \frac{\partial L}{\partial \dot{q}_i}\delta \dot{q}_i \approx \delta p_i \dot{q}_i - \frac{\partial L}{\partial q_i}\delta q_i \tag{5.5}$$

after using the definition of the momentum variables (5.2). Thus, we see that the velocities enter H_c only through the combination $p = f(q, \dot{q})$, so that H_c can be expressed in terms of qs and ps only and is independent of the velocities. Expressed in this way it becomes the *canonical Hamiltonian*.

Relation (5.5) is a weak equation and it holds only for those variations of qs and ps that are consistent with constraints. The variations with constraints, as we have seen, can be treated with the help of multipliers. Let us, therefore, introduce the *total Hamiltonian*

$$H_{\mathrm{T}} = H_{\mathrm{c}} + u^m \phi_m \tag{5.6}$$

where u^m are arbitrary multipliers. By varying this expression with respect to (u, q, p) we obtain the constraints (5.3) and the equations

$$\begin{aligned} \frac{\partial H_{\mathrm{c}}}{\partial p_i} + u^m \frac{\partial \phi_m}{\partial p_i} &= \dot{q}_i \\ \frac{\partial H_{\mathrm{c}}}{\partial q_i} + u^m \frac{\partial \phi_m}{\partial q_i} &= -\frac{\partial L}{\partial q_i} \approx -\dot{p}_i \end{aligned} \tag{5.7}$$

where the last equality follows from the Euler–Lagrange equations and the definition of momenta. In this way we obtain the Hamiltonian equations of motion involving arbitrary multipliers u^m, as a consequence of the existence of constraints.

Let us now introduce the *Poisson bracket* (PB):

$$\{A, B\} \equiv \frac{\partial A}{\partial q_i}\frac{\partial B}{\partial p_i} - \frac{\partial A}{\partial p_i}\frac{\partial B}{\partial q_i}. \tag{5.8}$$

The equation of motion of an arbitrary dynamical quantity $g(q, p)$,

$$\dot{g} = \frac{\partial g}{\partial q_i}\dot{q}_i + \frac{\partial g}{\partial p_i}\dot{p}_i$$

after substituting for the values q and $\dot{q}$ from the Hamiltonian equations (5.7), can be written in a concise form:

$$\dot{g} = \{g, H_{\mathrm{c}}\} + u^m\{g, \phi_m\} \approx \{g, H_{\mathrm{T}}\}. \tag{5.9}$$

The last step in this relation demands an explanation. Since H_{T} contains arbitrary multipliers that are not functions of qs and ps, the Poisson bracket $\{g, H_{\mathrm{T}}\}$ is not determined in the sense of equation (5.8). However, by using the formal properties of the PB operation applied to the sum and product of functions, we can write

$$\begin{aligned} \{g, H_{\mathrm{T}}\} &= \{g, H_{\mathrm{c}}\} + \{g, u^\mu \phi_m\} \\ &= \{g, H_{\mathrm{c}}\} + u^m\{g, \phi_m\} + \{g, u^m\}\phi_m. \end{aligned}$$

The quantity $\{g, u^m\}$ is not defined, but the last term can be neglected as it vanishes weakly, whereupon we obtain (5.9). Bearing in mind that this term will always be neglected, the weak equality in (5.9) will be replaced by the usual equality, for simplicity.

Equations (5.9) describe the motion of a system in the subspace Γ_1 of dimension $2N - P$. The motion is described by $2N$ coordinates (q, p) satisfying P constraints; as a consequence, P multipliers in the evolution equations appear. Explicit elimination of some coordinates is possible but it often leads to a violation of locality and/or covariance of the formalism.

Consistency conditions. The basic consistency of the theory requires that the primary constraints be conserved during the dynamical evolution of the system. The equations of motion (5.9) lead to the consistency conditions

$$\dot{\phi}_m = \{\phi_m, H_\mathrm{c}\} + u^n\{\phi_m, \phi_n\} \approx 0. \tag{5.10}$$

If the equations of motion are consistent, these conditions reduce to one of the following three cases:

(a) the first type of condition reduces to an identity, 0=0, i.e. it is automatically satisfied after using the primary constraints;
(b) the next reduces to an equation which is independent of the multipliers u^n, yielding a new, *secondary constraint*: $\chi(q, p) \approx 0$; and
(c) finally, an equation in (5.10) may impose a restriction on the u^ns.

If we find some secondary constraints in the theory, they also have to satisfy consistency conditions of the type (5.10). The process continues until all consistency conditions are exhausted. As a final result we are left with a number of secondary constraints and a number of conditions on the multipliers. The secondary constraints are for many purposes treated on the same footing as the primary constraints. Let us denote all constraints in the theory as

$$\varphi_s \equiv (\phi_m, \chi_n) \approx 0 \qquad (s = 1, \ldots, P, P+1, \ldots, P+S) \tag{5.11}$$

where P is the number of primary and S the number of all secondary constraints. These constraints define a subspace Γ_2 of the phase space Γ, such that $\Gamma_2 \subseteq \Gamma_1$. The notions of weak and strong equalities are now defined with respect to Γ_2.

The consistency conditions on the constraints φ_s yield relations

$$\dot{\varphi}_s = \{\varphi_s, H_\mathrm{c}\} + u^m\{\varphi_s, \phi_m\} \approx 0 \qquad (s = 1, \ldots, P+S) \tag{5.12}$$

some of which are identically satisfied, while the others represent non-trivial conditions on u^m. We shall consider (5.12) as a set of linear inhomogeneous equations in the unknown u^m. If $V_a^m(q, p)$ $(a = 1, 2, \ldots, N_1)$ are all the independent solutions of the homogeneous equations, $V_a^m\{\varphi_s, \phi_m\} \approx 0$, and $U^m(q, p)$ are particular solutions of the inhomogeneous equations, then the general solution for u^m takes the form

$$u^m = U^m + v^a V_a^m$$

where $v^a = v^a(t)$ are arbitrary coefficients. After that, the total Hamiltonian becomes

$$H_\mathrm{T} = H' + v^a\phi_a \qquad (a = 1, 2, \ldots, N_1') \tag{5.13}$$

where

$$H' = H_\mathrm{c} + U^m\phi_m \qquad \phi_a = V_a^m\phi_m.$$

Thus, we see that even after all consistency requirements are satisfied, we still have *arbitrary functions of time* in the theory. As a consequence, dynamical variables at some future instant of time are not uniquely determined by their initial values.

First and second class quantities. A dynamical variable $R(q, p)$ is said to be first class (FC) if it has weakly vanishing PBs with all constraints in the theory:

$$\{R, \varphi_s\} \approx 0. \tag{5.14}$$

If R is not first class, it is called second class. While the distinction between primary and secondary constraints is of little importance in the final form of the Hamiltonian theory, we shall see that the property of being FC or second class is essential for the dynamical interpretation of constraints.

We have seen that (in a regular theory) any weakly vanishing quantity is strongly equal to a linear combination of constraints. Therefore, if the quantity $R(q, p)$ is FC, it satisfies the strong equality

$$\{R, \varphi_s\} = R_s{}^r \varphi_r.$$

From this we can infer, by virtue of the Jacobi identity, that the PB of two FC constraints is also FC:

$$\begin{aligned}
\{\{R, S\}, \varphi_s\} &= \{\{R, \varphi_s\}, S\} - \{\{S, \varphi_s\}, R\} \\
&= \{R_s{}^r \varphi_r, S\} - \{S_s{}^r \varphi_r, R\} \\
&= R_s{}^r \{\varphi_r, S\} + \{R_s{}^r, S\}\varphi_r - S_s{}^r \{\varphi_r, R\} - \{S_s{}^r, R\}\varphi_r \approx 0.
\end{aligned}$$

It should be noted that the quantities H' and ϕ_a, which determine H_{T}, are FC. The number of arbitrary functions of time $v^a(t)$ in H_{T} is equal to the number of primary FC (PFC) constraints ϕ_a.

The presence of arbitrary multipliers in the equations of motion (and their solutions) means that the variables $(q(t), p(t))$ cannot be uniquely determined from given initial values $(q(0), p(0))$; therefore, they do not have a direct physical meaning. Physical information about a system can be obtained from functions $A(q, p)$, defined on a constraint surface, that are independent of arbitrary multipliers; such functions are called (classical) *observables*. The physical state of a system at time t is determined by the complete set of observables at that time.

In order to illustrate these ideas, let us consider a general dynamical variable $g(t)$ at $t = 0$, and its change after a short time interval δt. The initial value $g(0)$ is determined by $(q(0), p(0))$. The value of $g(t)$ at time δt can be calculated from the equations of motion:

$$\begin{aligned}
g(\delta t) &= g(0) + \delta t \dot{g} = g(0) + \delta t\{g, H_{\mathrm{T}}\} \\
&= g(0) + \delta t[\{g, H'\} + v^a\{g, \phi_a\}].
\end{aligned}$$

Since the coefficients $v^a(t)$ are completely arbitrary, we can take different values for these coefficients and obtain different values for $g(\delta t)$, the difference being of the form

$$\Delta g(\delta t) = \varepsilon^a\{g, \phi_a\} \tag{5.15}$$

where $\varepsilon^a = \delta t(v_2^a - v_1^a)$. This change of $g(\delta t)$ is unphysical, as $g_1(\delta t)$ and $g_2(\delta t) = g_1(\delta t) + \Delta g(\delta t)$ correspond to the same physical state. We come to the conclusion that *PFC constraints* generate unphysical transformations of dynamical variables, known as *gauge transformations*, that do not change the physical state of our system.

The application of two successive transformations of the type (5.15), with parameters ε_1^a and ε_2^a, yields a result that depends on the order of the transformations. The difference in the two possible results is

$$\begin{aligned}(\Delta_1\Delta_2 - \Delta_2\Delta_1)g(\delta t) &= \varepsilon_1^a\varepsilon_2^b[\{\{g, \phi_b\}, \phi_a\} - \{\{g, \phi_a\}, \phi_b\}] \\ &= \varepsilon_1^a\varepsilon_2^b\{g, \{\phi_a, \phi_b\}\}\end{aligned}$$

where the last equality is obtained by virtue of the Jacobi identity. This leads us to conclude that the quantity $\{\phi_a, \phi_b\}$ is also the generator of unphysical transformations. Since ϕ_a are FC constraints, their PB is strongly equal to a linear combinations of FC constraints. We expect that this linear combination will also contain secondary FC constraints, and this is really seen to be the case in practice. Therefore, *secondary FC constraints* are also generators of unphysical transformations.

These considerations do not allow us to conclude that *all* secondary FC constraints are generators of unphysical transformations. Dirac believed this to be true, but was unable to prove it ('Dirac's conjecture'). The answer to this problem will be given in the next subsection.

Can secondary FC constraints appear in H_{T}? Sometimes certain dynamical variables $q_a = A^a$ play the role of arbitrary multipliers. They appear *linearly* in the canonical Hamiltonian in the form $A^a\chi_a$, while, at the same time, there exist PFC constraints of the form $p_a \approx 0$. The consistency conditions of p_a yield $\chi_a \approx 0$, i.e. the χ_a are secondary constraints. They are FC, since the variables A^a are arbitrary functions of time. Indeed, the PFC constraints $p_a \approx 0$ are present in H_{T} in the form $v^a p_a$, so that the equations of motion for A^a are given as $\dot{A}^a \approx v^a$; therefore, A^a (as well as v^a) are arbitrary functions of time, and H_{T} contains secondary FC constraints. This is the case in electrodynamics, as well as in gravitation.

The Hamiltonian dynamics based on H_{T} is known to be equivalent to the related Lagrangian dynamics. Hence, the maximal number of arbitrary multipliers in H_{T} is determined by the nature of the gauge symmetries in the action.

Example 1. Consider the system described by the Lagrangian

$$L = \tfrac{1}{2}(\dot{q}_1)^2 + q_2\dot{q}_1 + (1-\alpha)q_1\dot{q}_2 + \tfrac{1}{2}\beta(q_1 - q_2)^2.$$

From the definition of momenta, $p_1 = \dot{q}_1 + q_2$, $p_2 = (1-\alpha)q_1$, we find one primary constraint,

$$\phi = p_2 + (\alpha - 1)q_1 \approx 0.$$

The canonical and total Hamiltonian are

$$H_c = \tfrac{1}{2}(p_1 - q_2)^2 - \tfrac{1}{2}\beta(q_1 - q_2)^2$$
$$H_T = H_c + u\phi.$$

The consistency condition of ϕ leads to the relation

$$\chi \equiv \dot{\phi} = \{\phi, H_T\} = \alpha(p_1 - q_2) - \beta(q_1 - q_2) \approx 0$$

the meaning of which depends on the values of constants α and β.

(a) $\alpha = 0$, $\beta = 0$. The consistency of the primary constraint $\phi = p_2 - q_1$ is automatically satisfied, ϕ is FC, and the multiplier u remains arbitrary.
For further analysis it is useful to have a general expression for $\dot{\chi}$:

$$\dot{\chi} = \{\chi, H_T\} = -\beta[(p_1 - q_2) - \alpha(q_1 - q_2)] + (\beta - \alpha^2)u \approx 0.$$

(b) $\alpha = 0$, $\beta \neq 0$. The consistency of the primary constraint $\phi = p_2 - q_1$ yields secondary constraint $\chi = -\beta(q_1 - q_2) \approx 0$. The consistency of χ restricts the multiplier u: $u = p_1 - q_2$. Both ϕ and χ are second class.
(c) $\alpha \neq 0$. From the general consistency condition of χ we can see that further analysis depends essentially on the value of $\beta - \alpha^2$.
($c1$) If $\beta = \alpha^2$, we find that $\dot{\chi} = -(\beta/\alpha)\chi$, so that there are no new constraints. Since $\{\phi, \chi\} = 0$, the constraints ϕ and χ are FC, and u remains undetermined.
($c2$) If, on the other hand, $\beta \neq \alpha^2$, the consistency for χ determines u: $u \approx (\beta/\alpha)(q_1 - q_2)$, after using χ. The constraints ϕ and χ are second class.

The extended Hamiltonian. We have seen that gauge transformations, generated by FC constraints, do not change the physical state of a system. This suggests the possibility of generalizing the equations of motion by allowing *any* evolution of dynamical variables that does not change the physical states. To realize this idea we introduce the extended Hamiltonian,

$$H_E = H' + v^{1a}\phi_{1a} + v^{2b}\phi_{2b} \tag{5.16}$$

containing both primary (ϕ_{1a}) and secondary (ϕ_{2b}) FC constraints, multiplied by arbitrary functions of time. In the formalism based on the extended Hamiltonian all gauge freedoms are manifestly present in the dynamics, and any differences between primary and secondary FC constraints are completely absent.

The equations of motion following from H_E are not equivalent with the Euler–Lagrange equations, but the difference is *unphysical*.

Dirac brackets. Having clarified the meaning and importance of FC constraints, we now turn our attention to second-class constraints. To keep the discussion simple, let us start with a simple example of two second-class constraints:

$$q_1 \approx 0 \qquad p_1 \approx 0.$$

Using second-class constraints as generators of gauge transformations may lead to contradictions, as they do not preserve all the constraints. Thus, e.g., starting with $F \equiv p_1\psi(q) \approx 0$, where $\psi \neq 0$, we can calculate $\delta F = \varepsilon\{q_1, F\} = \psi$, and find that $\delta F \neq 0$. Since the constraints are weak equalities, they should not be used before the calculations of PBs. These equations suggest that the variables (q_1, p_1) are not of any importance, and can be completely eliminated from the theory. We shall do that by introducing a modified PB in which the variables (q_1, p_1) are discarded:

$$\{f, g\}^* = \sum_{i \neq 1} \left(\frac{\partial f}{\partial q_i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q_i} \right).$$

After that, the constraints $q_1 \approx 0$, $p_1 \approx 0$ can be treated as strong equations. The resulting theory is defined only in terms of the variables (q_i, p_i), $i \neq 1$.

The idea can be generalized to arbitrary second-class constraints. The presence of second-class constraints means that there are dynamical degrees of freedom in the theory that are of no importance. In order to be able to eliminate these variables, it is necessary to set up a new PB referring only to dynamically important degrees of freedom.

After finding all FC constraints ϕ_a $(a = 1, 2, \dots, N_1)$, the remaining constraints θ_s $(s = 1, 2, \dots, N_2)$ are second class. The matrix $\Delta_{rs} = \{\theta_r, \theta_s\}$ is non-singular. Indeed, if Δ_{rs} were singular, i.e. $\det(\Delta_{rs}) = 0$, then the equation $\lambda^s\{\theta_r, \theta_s\} = 0$ would have a non-trivial solution for λ^s, and, consequently, the linear combination $\lambda^s\theta_s$ would be FC. However, this is impossible by assumption.

The number of second-class constraints N_2 must be *even*, as the matrix Δ is antisymmetric. This follows from the property that each antisymmetric matrix of odd order has a vanishing determinant, which is not the case with Δ.

Because Δ is non-singular, we can define its inverse Δ^{-1}. The new PB is now defined by

$$\{f, g\}^* = \{f, g\} - \{f, \theta_r\}\Delta^{-1}_{rs}\{\theta_s, g\} \tag{5.17}$$

and is called the *Dirac bracket*. It is easy to check that the Dirac bracket satisfies all the standard properties of a PB: it is antisymmetric, linear, obeys the product law and Jacobi's identity.

The Dirac bracket of any second-class constraint with an arbitrary variable vanishes by construction:

$$\{\theta_m, g\}^* = \{\theta_m, g\} - \{\theta_m, \theta_r\}\Delta^{-1}_{rs}\{\theta_s, g\} = 0$$

because $\{\theta_m, \theta_r\}\Delta^{-1}_{rs} = \delta_{ms}$. This means that after the Dirac brackets are constructed, second-class constraints $\theta_m \approx 0$ can be treated as strong equations. The equations of motion (5.9) can be written in terms of the Dirac brackets:

$$\dot{g} \approx \{g, H_{\rm T}\}^*. \tag{5.18}$$

This follows from $\{\theta_m, H_{\rm T}\} \approx 0$, since $H_{\rm T}$ is FC.

Now, we can clearly see the meaning of and the difference between FC and second-class constraints.

> *FC constraints generate unphysical transformations, while second-class constraints, after the introduction of Dirac brackets, are treated as strong equations.*

The construction of Dirac brackets in theories with a lot of second-class constraints can be simplified by using their *iterative property*:

(i) we first construct preliminary Dirac brackets by using a subset of second-class constraints; then,
(ii) we use a new subset of the remaining constraints, and construct new Dirac brackets, using the preliminary brackets instead of the Poisson ones in (5.17), and so on.

The process continues until all second-class constraints are exhausted.

Example 2. Let us calculate the Dirac brackets in the cases (b) and ($c2$) of example 1. In the first case there are two second-class constraints,

$$\theta_1 = p_2 - q_1 \qquad \theta_2 = -\beta(q_1 - q_2)$$

with the PB is $\{\theta_1, \theta_2\} = -\beta$. From this we easily find the matrices Δ and Δ^{-1}, and then the Dirac brackets:

$$\{A, B\}^* = \{A, B\} - (1/\beta)[\{A, \theta_1\}\{\theta_2, B\} - \{A, \theta_2\}\{\theta_1, B\}].$$

The second case is characterized by the constraints

$$\theta_1 = p_2 + (\alpha - 1)q_1 \qquad \theta_2 = \alpha(p_1 - q_2) - \beta(q_1 - q_2)$$

so that $\{\theta_1, \theta_2\} = \alpha^2 - \beta$. The Dirac brackets follow easily.

Gauge conditions. We have seen that the presence of FC constraints is related to the existence of gauge symmetries, which describe unphysical transformations of dynamical variables. This fact can be used to impose suitable restrictions on the set of dynamical variables, so as to bring them into a one-to-one correspondence with the set of all observables. By means of this procedure we can remove any unobservable gauge freedoms in the description of dynamical variables, without changing any observable property of the theory. The restrictions are realized as a suitable set of *gauge conditions*:

$$\Omega_a(q, p) \approx 0 \qquad (a = 1, 2, \ldots, N_{\text{gc}}). \tag{5.19a}$$

In order for Ω_a to be a set of good gauge conditions, the following two requirements must be satisfied:

(i) Gauge conditions must be *accessible*, i.e. for any value of (q, p) in the phase space there must exist a gauge transformation $(q, p) \to (q', p')$, such that $\Omega_a(q', p') \approx 0$.
(ii) Gauge conditions must fix the gauge freedom *completely*, i.e. there should be no gauge transformation that preserves (5.19a).

If an arbitrary gauge transformation is determined by N_{par} parameters, then the first condition is possible only if $N_{\text{par}} \geq N_{\text{gc}}$, while the second one implies that $N_{\text{gc}} \geq N_{\text{par}}$ (the smaller number of gauge conditions could not fix the gauge completely). Therefore, the number of gauge conditions must be equal to the number of *independent gauge transformations*.

The meaning of these requirements can be made more precise if we specify the form of the gauge transformations. Let us assume, for instance, that the dynamical evolution is described by the extended Hamiltonian (5.16), so that all FC constraints generate gauge transformations, $N_{\text{par}} = N_1$. Then the number of gauge conditions becomes equal to the number of *independent FC constraints*. The second requirement means that there is no gauge transformation such that $v^b\{\Omega_a, \phi_b\} \approx 0$ unless $v^b = 0$, which implies

$$\det\{\Omega_a, \phi_b\} \neq 0 \qquad \phi_b = \text{FC}. \tag{5.19b}$$

This means that FC constraints together with the related gauge conditions form a set of second-class constraints. As a consequence, all arbitrary multipliers v^b in H_{E} can be fixed by the consistency conditions:

$$\dot{\Omega}_a = \{\Omega_a, H_{\text{E}}\} \approx \{\Omega_a, H'\} + v^b\{\Omega_a, \phi_b\} \approx 0.$$

After fixing the gauge we can define Dirac brackets and treat both the constraints and gauge conditions as identities.

These requirements are necessary but not sufficient for the correct gauge fixing. They guarantee only that the gauge is correctly fixed locally, while the question of its global validity remains, in general, open. In some cases the geometry of the problem may be such that global gauge conditions do not even exist.

First and second-class constraints, together with gauge conditions, define a subspace Γ^* of the phase space Γ, having dimension

$$N^* = 2N - (2N_1 + N_2)$$

in which the dynamics of independent degrees of freedom is realized. Since the number of second-class constraints N_2 is even, the dimension N^* is also even. This counting of independent degrees of freedom is gauge independent, hence it also holds in the total Hamiltonian formalism.

A similar analysis may be given for the approach based on the total Hamiltonian. The general construction and some specific features of these gauge transformations are given in the following subsection.

Generators of gauge symmetries

The unphysical transformations of dynamical variables are often referred to as local, or gauge, transformations. The term ‘local’ means that the parameters of the transformations are arbitrary functions of time. Gauge transformations are of special interest only if they represent a symmetry of the theory. We shall now describe an algorithm for constructing the generators of all gauge symmetries of the equations of motion governed by the total Hamiltonian (Castellani 1982). This analysis will also give an answer to Dirac’s old conjecture that all FC constraints generate gauge symmetries.

We begin by considering a theory determined by the total Hamiltonian (5.13) and a complete set of constraints, $\varphi_s \approx 0$. Suppose that we have a trajectory $T_1(t) = (q(t), p(t))$, which starts from a point $T_0 = (q(0), p(0))$ on the constraint surface Γ_2, and satisfies the equations of motion with some fixed functions $v^a(t)$:

$$
\begin{aligned}
\dot{q}_i &= \frac{\partial H'}{\partial p_i} + v^a \frac{\partial \phi_a}{\partial p_i} \\
-\dot{p}_i &= \frac{\partial H'}{\partial q_i} + v^a \frac{\partial \phi_a}{\partial q_i} \\
\varphi_s(q, p) &= 0.
\end{aligned}
\tag{5.20}
$$

Such trajectories will be called dynamical trajectories. Consider now a new, varied trajectory $T_2(t) = (q(t) + \delta_0 q(t), p(t) + \delta_0 p(t))$, which starts from the same point T_0, and satisfies the equations of motion with new functions $v^a(t) + \delta_0 v^a(t)$. Expanding these equations to first order in the small variations $\delta_0 q, \delta_0 p, \delta_0 v^a$ and using (5.20) we obtain

$$
\begin{aligned}
\delta_0 \dot{q}_i &= \left(\delta_0 q_j \frac{\partial}{\partial q_j} + \delta_0 p_j \frac{\partial}{\partial p_j} \right) \frac{\partial H_{\mathrm{T}}}{\partial p_i} + \delta_0 v^a \frac{\partial \phi_a}{\partial p_i} \\
-\delta_0 \dot{p}_i &= \left(\delta_0 q_j \frac{\partial}{\partial q_j} + \delta_0 p_j \frac{\partial}{\partial p_j} \right) \frac{\partial H_{\mathrm{T}}}{\partial q_i} + \delta_0 v^a \frac{\partial \phi_a}{\partial q_i} \\
\frac{\partial \varphi_s}{\partial q_i} \delta_0 q_i &+ \frac{\partial \varphi_s}{\partial p_j} \delta_0 p_j = 0.
\end{aligned}
$$

These are necessary and sufficient conditions for the varied trajectories to be dynamical. Transition from one trajectory to another at the same moment of time

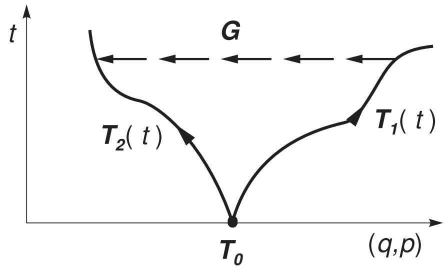

Figure 5.1. The Hamiltonian description of gauge transformations.

represents an unphysical or gauge transformation (figure 5.1).

Now let us assume that the variations of dynamical variables are determined by an arbitrary infinitesimal parameter $\varepsilon(t)$ and have the canonical form

$$\begin{aligned}\delta_0 q_i(t) &= \varepsilon(t)\{q_i, G\} = \varepsilon(t)\frac{\partial G}{\partial p_i}\\ \delta_0 p_i(t) &= \varepsilon(t)\{p_i, G\} = -\varepsilon(t)\frac{\partial G}{\partial q_i}\end{aligned} \tag{5.21}$$

where $G(q, p)$ is the generator of this transformation. Then,

$$\begin{aligned}\delta_0 \dot{q}_i &= \dot{\varepsilon}\frac{\partial G}{\partial p_i} + \varepsilon\left\{\frac{\partial G}{\partial p_i}, H_{\mathrm{T}}\right\}\\ \delta_0 \dot{p}_i &= -\dot{\varepsilon}\frac{\partial G}{\partial q_i} - \varepsilon\left\{\frac{\partial G}{\partial q_i}, H_{\mathrm{T}}\right\}\end{aligned}$$

and the conditions that the varied trajectory be dynamical take the form

$$\begin{aligned}\dot{\varepsilon}\frac{\partial G}{\partial p_i} + \varepsilon\left\{\frac{\partial G}{\partial p_i}, H_{\mathrm{T}}\right\} &= -\varepsilon\left\{G, \frac{\partial H_{\mathrm{T}}}{\partial p_i}\right\} + \frac{\partial \phi_a}{\partial p_i}\delta_0 v^a\\ \dot{\varepsilon}\frac{\partial G}{\partial q_i} + \varepsilon\left\{\frac{\partial G}{\partial q_i}, H_{\mathrm{T}}\right\} &= -\varepsilon\left\{G, \frac{\partial H_{\mathrm{T}}}{\partial q_i}\right\} + \frac{\partial \phi_a}{\partial q_i}\delta_0 v^a\\ \varepsilon\{\varphi_s, G\} &= 0.\end{aligned}$$

These conditions must be fulfilled for every dynamical trajectory $(q(t), p(t))$ at arbitrary time t, i.e. for each point (q, p) on Γ_2, and for every $v^a(t)$. They can be rewritten as

$$\begin{aligned}&\frac{\partial}{\partial p_i}[\dot{\varepsilon}G + \varepsilon\{G, H_{\mathrm{T}}\} - \phi_a\delta_0 v^a] \approx 0\\ &\frac{\partial}{\partial q_i}[\dot{\varepsilon}G + \varepsilon\{G, H_{\mathrm{T}}\} - \phi_a\delta_0 v^a] \approx 0\\ &\varepsilon\{\varphi_s, G\} \approx 0.\end{aligned}$$

From the first two equations we obtain the relation

$$\{F, \dot{\varepsilon}G + \varepsilon\{G, H_{\mathrm{T}}\} - \phi_a \delta_0 v^a\} \approx 0$$

which is valid for every dynamical variable $F(q, p)$. Hence, the expression in braces is equivalent to a trivial generator:

$$\dot{\varepsilon}G + \varepsilon\{G, H_{\mathrm{T}}\} - \phi_a \delta_0 v^a = \mathcal{O}$$

where $\mathcal{O}$ is zero, a constant or the second (or higher) power of a constraint. The conditions of solvability with respect to δv^a, for arbitrary ε, $\dot{\varepsilon}$ and v^a, reduce to

$$G = \lambda^a \phi_a \qquad \{G, H_{\mathrm{T}}\} = \eta^a \phi_a$$

where λ^a, η^a are functions of qs and ps, and the equality of generators is an equality modulo $\mathcal{O}$.

Summarizing, the necessary and sufficient condition for G to be a gauge generator has the form

$$G = C_{\mathrm{PFC}} \qquad \{G, H_{\mathrm{T}}\} = C_{\mathrm{PFC}} \tag{5.22}$$

where C_{PFC} is a PFC constraint. Therefore, if the gauge symmetry has the form (5.21), the generator has to be a PFC constraint, and its PB with H_{T} must also be a PFC constraint.

The assumption that the symmetry transformation has the form (5.21), where a dependence on the derivatives of parameters is absent, is too restrictive for many physical applications. The most interesting case is when the gauge generator has the form

$$G = \varepsilon(t)G_0 + \dot{\varepsilon}(t)G_1. \tag{5.23a}$$

This discussion can now be repeated with

$$\delta_0 q_i(t) = \{q_i, G\} \qquad \delta_0 p_i(t) = \{p_i, G\}$$

and it leads to the following relations:

$$[\ddot{\varepsilon}G_1 + \dot{\varepsilon}(G_0 + \{G_1, H_{\mathrm{T}}\}) + \varepsilon\{G_0, H_{\mathrm{T}}\}] - \phi_a \delta_0 v^a = \mathcal{O}$$
$$\varepsilon\{\varphi_s, G_0\} + \dot{\varepsilon}\{\varphi_s, G_1\} \approx 0$$

The conditions for solvability with respect to δv^a, for arbitrary ε, $\dot{\varepsilon}$ and v^a, have the form:

$$\begin{aligned} G_1 &= C_{\mathrm{PFC}} \\ G_0 + \{G_1, H_{\mathrm{T}}\} &= C_{\mathrm{PFC}} \\ \{G_0, H_{\mathrm{T}}\} &= C_{\mathrm{PFC}}. \end{aligned} \tag{5.23b}$$

Example 2 (continued). Let us find the gauge symmetries for cases (a) and $(c1)$ of example 1. In the first case there exists one PFC constraint, $\phi = p_2 - q_1$, while secondary constraints are absent. From condition (5.22) it follows that the gauge generator has the form $G = \varepsilon\phi$ and the symmetry transformations are

$$\delta_0 p_1 = \varepsilon \qquad \delta_0 q_2 = \varepsilon \qquad \delta_0(\text{rest}) = 0.$$

In the second case ϕ is a primary and χ a secondary FC constraint. Starting from $G_1 = \phi$ in (5.23*b*), the second equation leads to $G_0 + \chi = a\phi$, while the third condition yields $a = -\beta/\alpha = -\alpha$, so that the gauge generator becomes

$$G = -\varepsilon(\chi + \alpha\phi) + \dot{\varepsilon}\phi.$$

From this result we can easily obtain the related transformations of the dynamical variables. That these transformations do not change the form of the equations of motion can be directly checked.

In the general case the gauge generator takes the form

$$G = \varepsilon^{(k)}(t)G_k + \varepsilon^{(k-1)}(t)G_{k-1} + \cdots + \varepsilon G_0 \tag{5.24a}$$

where $\varepsilon^{(n)} = \mathrm{d}^n\varepsilon/\mathrm{d}t^n$, and the conditions for solvability with respect to $\delta_0 v^a$ become

$$\begin{aligned} G_k &= C_{\mathrm{PFC}} \\ G_{k-1} + \{G_k, H_{\mathrm{T}}\} &= C_{\mathrm{PFC}} \\ \cdots \quad & \cdots \\ \{G_0, H_{\mathrm{T}}\} &= C_{\mathrm{PFC}}. \end{aligned} \tag{5.24b}$$

Thus, G_k has to be a PFC constraint and all other G_n $(n < k)$ must be FC constraints.

These conditions clearly define the procedure for constructing the generator. We start with an arbitrary PFC constraint G_k, evaluate its PB with H_{T} and define G_{k-1}, etc. The procedure stops when we obtain the constraint G_0, for which the PB with H_{T} is a PFC constraint. We have assumed that the number of steps is finite, which is equivalent to assuming that the number of generations of secondary constraints is finite.

Conditions (5.24*b*) determine G_n only up to PFC constraints. Therefore, at each stage of the construction of the chain $\{G_n\}$ we should try to add suitable FC constraints, in order to stop the whole process as soon as possible. In this way we can find the *minimal chain* $\{G_n\}$ (the chain that could not have been stopped earlier by any choice of G_n). It should be observed that the chain also stops if, at some stage, we obtain $\{G_i, H_{\mathrm{T}}\} = \chi^n$ $(n \geq 2)$, as the power of constraint χ^n is equivalent to the zero generator. In this case the constraint χ is not present in the gauge generator G.

Dirac's conjecture that all FC constraints generate gauge symmetries is replaced by the following statement:

All FC constraints, except those appearing in the consistency conditions in the form χ^n $(n \geq 2)$, *and following from* $\{\chi, H_T\} \approx 0$, *are parts of the gauge generator.*

We should note that in all relevant physical applications Dirac's conjecture remains true. This is of particular importance for the standard quantization methods, which are based on the assumption that all FC constraints are gauge generators.

When the gauge generator (5.24a) contains all FC constraints, we can simply realize the transition to the extended formalism by replacing time derivatives $\varepsilon^{(k)} = \mathrm{d}^k\varepsilon/\mathrm{d}t^k$ with independent gauge parameters.

Electrodynamics

Field theory can be thought of as a mechanical system in which dynamical variables are defined at each point $\boldsymbol{x}$ of a three-dimensional space, $(q_i, p_i) \to (q_{ix}, p_{ix}) \equiv (q_i(\boldsymbol{x}), p_i(\boldsymbol{x}))$, i.e. where each index also takes on continuous values, $i \equiv (i, \boldsymbol{x})$. Then, a formal generalization of the previous analysis to the case of field theory becomes rather direct: the sum goes over into an integral (and sum), partial derivatives into functional derivatives, δ^i_k becomes $\delta^i_k\delta(\boldsymbol{x} - \boldsymbol{y})$, etc.

We shall now study the important example of electrodynamics, the analysis of which not only represents a fine illustration of the general theory, but also shows in many aspects a great similarity with the theory of gravity. The dynamics of the free electromagnetic field $A^\mu(x)$ is described by the Lagrangian

$$L = -\tfrac{1}{4}\int \mathrm{d}^3x\, F_{\mu\nu}F^{\mu\nu} \qquad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{5.25}$$

Varying this Lagrangian with respect to $\dot{A}^\mu$ we obtain the momenta

$$\pi_\mu(x) = \frac{\delta L}{\delta \dot{A}^\mu(x)} = -F_{0\mu}(x)$$

where $x \equiv (t, \boldsymbol{x})$. Since $F_{\mu\nu}$ is defined as the antisymmetric derivative of A_μ, it does not depend on the velocities $\dot{A}^0$, so that the related momentum vanishes. Thus, we obtain the primary constraint

$$\varphi_1 \equiv \pi_0 \approx 0.$$

(The number of these constraints is, in fact, $1 \times \infty^3$, i.e. one constraint at each point $\boldsymbol{x}$ of the three-dimensional space, but we shall continue to talk about one constraint, for simplicity.) The remaining three momenta are equal to the components of the electric field, $-F_{0\alpha}$. The canonical Hamiltonian has the form

$$H_\mathrm{c} = \int \mathrm{d}^3x\, (\pi_\mu \dot{A}^\mu - \mathcal{L}) = \int \mathrm{d}^3x\, (\tfrac{1}{4}F_{\alpha\beta}F^{\alpha\beta} - \tfrac{1}{2}\pi_\alpha\pi^\alpha - A^0\partial^\alpha\pi_\alpha)$$

where the last term is obtained in that form after performing a partial integration‡. This expression does not depend on velocities, as it contains only the spatial derivatives of coordinates and momenta. With one primary constraint present, the total Hamiltonian takes the form

$$H_{\mathrm{T}} = H_{\mathrm{c}} + \int \mathrm{d}^3x\, u(x)\pi_0(x). \tag{5.26}$$

By use of the basic PBs,

$$\{A^\mu(x), \pi_\nu(x')\} = \delta^\mu_\nu \delta(\boldsymbol{x} - \boldsymbol{x}') \qquad t = t'$$

the consistency condition for φ_1, $\dot{\pi}_0(x) = \{\pi_0(x), H_{\mathrm{T}}\} \approx 0$, leads to a secondary constraint:

$$\varphi_2 \equiv \partial^\alpha \pi_\alpha \approx 0.$$

Further consistency requirements reduce to $\dot{\varphi}_2 = 0$, i.e. no new condition is produced. The constraints φ_1 and φ_2 are FC, as $\{\pi_0, \partial^\alpha \pi_\alpha\} = 0$.

The variable A^0 appears linearly in H_{c}. Its equation of motion

$$\dot{A}^0 = \{A^0, H_{\mathrm{T}}\} = u$$

gives a definite meaning to the arbitrary multiplier u. It follows from this that A^0 is also an arbitrary function of time. We observe that secondary FC constraints are already present in H_{c} in the form $A^0\varphi_2$, showing that A^0 has the role of another arbitrary multiplier. In this way here, as well as in gravitation, we find an interesting situation that all FC constraints are present in the total Hamiltonian. The transition to the extended Hamiltonian,

$$H_{\mathrm{E}} = H_{\mathrm{T}} + \lambda \partial_\alpha \pi^\alpha$$

is equivalent to the replacement $A^0 \to A^0 - \lambda$ in H_{T}.

Let us now look for the generator of gauge symmetries in the form (5.24). Starting with $G_1 = \pi_0$ we obtain

$$G = \int \mathrm{d}^3x\, (\dot{\varepsilon}\pi_0 - \varepsilon \partial^\alpha \pi_\alpha) \tag{5.27}$$

and the related gauge transformations are

$$\delta_0 A^\mu = \partial^\mu \varepsilon \qquad \delta_0 \pi_\mu = 0.$$

The result has the form we know from the Lagrangian analysis.

‡ We assume that the asymptotic behaviour of dynamical variables is such that the surface terms, obtained by partial integration, vanish. The possibility of the appearance of surface terms is an essential property of field theory.

The previous treatment fully respects the gauge symmetry of the theory. One of the standard ways for fixing this symmetry is to choose, e.g., the radiation gauge condition:

$$\Omega_1 \equiv A^0 \approx 0 \qquad \Omega_2 \equiv \partial_\alpha A^\alpha \approx 0.$$

The number of gauge conditions is equal to the number of FC constraints, and the matrix $\{\Omega_a, \varphi_b\}$ is non-singular. In addition, the gauge conditions remain preserved during the time evolution, as can be explicitly checked.

Gauge conditions can be used as strong equations after constructing the related Dirac brackets. Let us first construct preliminary Dirac brackets corresponding to the pair $(\Omega_1, \varphi_1) = (A^0, \pi_0)$. By using the relation $\{A^0(t, \boldsymbol{x}), \pi_0(t, \boldsymbol{u})\} = \delta(\boldsymbol{x} - \boldsymbol{u})$, we find that

$$\Delta(\boldsymbol{u}, \boldsymbol{v}) = \begin{pmatrix} 0 & \delta \\ -\delta & 0 \end{pmatrix} \qquad \Delta^{-1}(\boldsymbol{u}, \boldsymbol{v}) = \begin{pmatrix} 0 & -\delta \\ \delta & 0 \end{pmatrix}$$

where $\delta \equiv \delta(\boldsymbol{u} - \boldsymbol{v})$. From this we have

$$\begin{aligned}\{M(x), N(y)\}^* = \{M(x), N(y)\} &+ \int \mathrm{d}^3u\, \{M(x), A^0(u)\}\{\pi_0(u), N(y)\} \\ &- \int \mathrm{d}^3u\, \{M(x), \pi_0(u)\}\{A^0(u), N(y)\}.\end{aligned}$$

For all other variables, except π_0 and A^0, the Dirac brackets reduce to the usual PBs. On the other hand, we know that the Dirac bracket of π_0 or A^0 with any other variable vanishes. Therefore, π_0 and A^0 can be simply eliminated from the theory by using the conditions $\pi_0 \approx 0$ and $A^0 \approx 0$ as strong equalities, while for the remaining variables we should use the usual PBs. Taking into account the other pair (φ_2, Ω_2) yields the final Dirac brackets on the subspace of transverse variables $(A^\alpha_T, \pi^T_\alpha)$:

$$\{A^\alpha_T(x), \pi^T_\beta(x')\} = \delta^\alpha_\beta \delta(\boldsymbol{x} - \boldsymbol{x}') - \partial^\alpha \partial_\beta \mathcal{D}(\boldsymbol{x} - \boldsymbol{x}')$$

where $\mathcal{D}(\boldsymbol{x}) = (4\pi|\boldsymbol{x}|)^{-1}$ is a solution of the equation $\partial_\alpha \partial^\alpha \mathcal{D}(\boldsymbol{x}) = \delta(\boldsymbol{x})$.

The dimension of the phase space $\Gamma(A^\mu, \pi_\nu)$ is eight (twice larger than the number of Lagrangian variables); after fixing two gauge conditions we come to the phase space Γ^* of dimension $N^* = 8 - 2 \times 2 = 4$, defined by the transverse variables $(A^\alpha_T, \pi^T_\beta)$.

5.2 The gravitational Hamiltonian

Covariance and Hamiltonian dynamics

The existence of spacetime symmetries implies some general characteristics of the Hamiltonian dynamics, knowledge of which facilitates analysis of the theory.

If we study the dynamics of PGT, we see that the Lagrangian equations of motion are *covariant*, i.e. they do not change the form under the local Poincaré transformations. On the other hand, the Hamiltonian equations of motion, $\dot{g} \approx \{g, H_{\mathrm{T}}\}$, contain time t of a specific reference frame, so that they are *not manifestly* covariant. It is interesting to clarify how the covariance of the theory can be seen in the *Hamiltonian* analysis. The covariance demands a formulation in which the spacetime variables are referred to an arbitrary non-inertial reference frame. Such a formulation is sometimes automatically realized by a suitable choice of the action. If this is not the case, the original action has to be conveniently reformulated.

As an illustration of the problem of covariance, consider a simple case of a theory defined by the Lagrangian $L(q, \mathrm{d}q/\mathrm{d}t)$, in which the time parameter refers to a specific reference frame. If we introduce a new time τ as a monotonic function of t, the action can be rewritten as

$$I = \int \mathrm{d}t\, L(q, \mathrm{d}q/\mathrm{d}t) = \int \mathrm{d}\tau\, L_\tau \qquad L_\tau \equiv \frac{\mathrm{d}q_0}{\mathrm{d}\tau} L\left(q, \frac{\mathrm{d}q/\mathrm{d}\tau}{\mathrm{d}q_0/\mathrm{d}\tau}\right)$$

where $q_0 \equiv t$, and $\mathrm{d}q_0/\mathrm{d}\tau \neq 0$. Let us now take the time $t = q_0$ as an extra coordinate and consider the new Lagrangian L_τ as a function of the two coordinates (q_0, q) and the time τ. It should be noted that L_τ is a homogeneous function of the first degree in the velocities $(\mathrm{d}q_0/\mathrm{d}\tau, \mathrm{d}q/\mathrm{d}\tau)$. By introducing the corresponding momenta

$$p_0 = \frac{\partial L_\tau}{\partial(\mathrm{d}q_0/\mathrm{d}\tau)} \qquad p = \frac{\partial L_\tau}{\partial(\mathrm{d}q/\mathrm{d}\tau)}$$

we find, with the help of Euler's theorem for homogeneous functions or by a direct calculation, that the canonical Hamiltonian vanishes:

$$H_{\mathrm{c}} \equiv p_0 \frac{\mathrm{d}q_0}{\mathrm{d}\tau} + p \frac{\mathrm{d}q}{\mathrm{d}\tau} - L_\tau = 0.$$

In this case there exists at least *one* primary constraint. Indeed, p_0 and p are homogeneous functions of the zero degree so that they can depend only on the ratio $(\mathrm{d}q/\mathrm{d}\tau)/(\mathrm{d}q_0/\mathrm{d}\tau)$, from which we deduce the existence of one primary constraint ϕ. If ϕ is the only constraint, it must be FC, the total Hamiltonian is of the form

$$H_{\mathrm{T}} = v\phi$$

where v is an arbitrary multiplier and the equations of motions are

$$\frac{\mathrm{d}g}{\mathrm{d}\tau} = v\{g, \phi\}.$$

These equations contain an *arbitrary time scale* τ. A transition to a new time parameter τ' does not change the form of the equations:

$$\frac{\mathrm{d}g}{\mathrm{d}\tau'} = v'\{g, \phi\}$$

where $v' = v\,\mathrm{d}\tau/\mathrm{d}\tau'$ is a new arbitrary function of time τ'.

Example 3. The dynamics of the free non-relativistic particle is determined by the Lagrangian $L = \frac{1}{2}(\mathrm{d}q/\mathrm{d}t)^2$. Going over to an arbitrary time τ, we find that

$$L_\tau = \tfrac{1}{2}(\dot{q}^2/\dot{q}_0)$$

where the dot means the derivative over τ. From the definitions of momenta, $p = \dot{q}/\dot{q}_0$, $p_0 = -\frac{1}{2}(\dot{q}/\dot{q}_0)^2$, we obtain the primary constraint

$$\phi = p_0 + \tfrac{1}{2}p^2 \approx 0$$

while $H_\mathrm{T} = v(\tau)\phi$. The consistency condition of the constraint ϕ is automatically satisfied, and it is FC. The equations of motion are

$$\dot{q}_0 = v\{q_0, \phi\} = v \qquad \dot{q} = v\{q, \phi\} = vp \qquad \dot{p}_0 = \dot{p} = 0.$$

They are covariant with respect to the choice of time, i.e. they do not change the form under the transformations $\delta_0 q_0 = \varepsilon$, $\delta_0 q = \varepsilon p$, $\delta_0 p_0 = \delta_0 p = 0$.

We can return to time t by choosing the gauge condition $\Omega \equiv q_0 - \tau \approx 0$. In that case the equations of motion take the standard form: $v = 1$, $\dot{q} = p$.

Thus, the introduction of an arbitrary time scale into Hamiltonian theory implies the existence of *one* FC constraint, and $H_\mathrm{T} = v\phi \approx 0$. The method described here represents a general procedure by which we can transform any theory, taking the time parameter as an extra dynamical variable q_0, into the form in which the time scale is arbitrary. If the action is 'already covariant', it contains the variable q_0 from the very beginning, so that $H_\mathrm{T} \approx 0$ automatically.

In field theory, we should have the freedom not only to choose time, but also to choose three spatial coordinates. The corresponding Hamiltonian theory should have at least *four* FC constraints. The number of FC constraints may be larger if other gauge symmetries are also present. In a similar manner we can study the general features of the covariant Hamiltonian formulation of gravitation. The following property characterizes standard gravitational theories§:

> *To each parameter of the gauge symmetry there corresponds one FC constraint ϕ_a; the dynamical evolution of the system is described by weakly vanishing Hamiltonian, $H_\mathrm{T} = v^a\phi_a \approx 0$.*

The existence of gauge symmetries, as well as the vanishing of H_T, may cause a certain confusion in our understanding of the gravitational dynamics. This may arise in connection with two basic questions: the definition of gravitational energy and the dynamical nature of time. Leaving the question of energy for the next chapter, we mention here the *problem of time*. Going back to example 3 we

§ It should be noted that, in general, covariance does not imply a zero Hamiltonian (Henneaux and Teitelboim 1992).

observe that in order to have a physical interpretation of the evolution parameter τ it is necessary to fix the gauge. A natural gauge choice is of the form $\Omega_1 \equiv q_0 - \tau \approx 0$. It satisfies $\{\phi, \Omega_1\} \neq 0$ and means that out of all possible times τ we have chosen the one that is equal to q_0 $(= t)$. An interesting situation arises if we 'forget' about the origin of L_τ, and try to impose the gauge condition $\Omega_2 \equiv q_0 \approx 0$. Formally, this choice also satisfies the necessary condition $\{\Omega_2, \phi\} \neq 0$. However, physically this means that $t = 0$, i.e. the time t 'does not flow'. If the final gauge-invariant theory is obtained from a theory in which the role of time was played by t, then Ω_2 is not the correct gauge choice, since it violates the monotony of the mapping $t \to \tau$. On the other hand, if the theory defined by L_τ is given *a priori*, without any reference to an 'original' theory, additional criteria for a correct choice of gauge become less transparent. The choice of time may be related to the asymptotic structure of spacetime. At this point we shall leave this problem, limiting our further discussion to situations in which the usual interpretation of time exist.

Now we are ready to begin a more detailed exposition of the form of the gravitational Hamiltonian (Nikolić 1984).

Primary constraints

The geometric framework for PGT is defined by the Riemann–Cartan spacetime U_4, while the general Lagrangian has the form (3.15),

$$\widetilde{\mathcal{L}} = b\mathcal{L}_\mathrm{G}(R^{ij}{}_{kl}, T^i{}_{kl}) + b\mathcal{L}_\mathrm{M}(\Psi, \nabla_k\Psi)$$

where Ψ denotes the matter fields and $\mathcal{L}_\mathrm{G}$ is given by (3.50). The basic Lagrangian dynamical variables are $(b^k{}_\mu, A^{ij}{}_\mu, \Psi)$, and the corresponding momenta $(\pi_k{}^\mu, \pi_{ij}{}^\mu, \pi)$ are

$$\pi_k{}^\mu = \frac{\partial\widetilde{\mathcal{L}}}{\partial b^k{}_{\mu,0}} \qquad \pi_{ij}{}^\mu = \frac{\partial\widetilde{\mathcal{L}}}{\partial A^{ij}{}_{\mu,0}} \qquad \pi = \frac{\partial\widetilde{\mathcal{L}}}{\partial \Psi_{,0}}. \tag{5.28}$$

Due to the fact that the curvature and the torsion are defined through the antisymmetric derivatives of $b^k{}_\mu$ and $A^{ij}{}_\mu$, respectively, they do not involve the velocities of $b^k{}_0$ and $A^{ij}{}_0$. As a consequence, we immediately obtain the following set of the so-called *sure* primary constraints:

$$\phi_k{}^0 \equiv \pi_k{}^0 \approx 0 \qquad \phi_{ij}{}^0 \equiv \pi_{ij}{}^0 \approx 0. \tag{5.29}$$

These constraints are always present, independently of the values of parameters in the Lagrangian. They are particularly important for the structure of the theory. Depending on the specific form of the Lagrangian, we may also have *additional* primary constraints in the theory.

The *canonical Hamiltonian* has the standard form:

$$H_\mathrm{c} \equiv \int \mathrm{d}^3x\, \mathcal{H}_\mathrm{c} \qquad \mathcal{H}_\mathrm{c} = \mathcal{H}_\mathrm{M} + \mathcal{H}_\mathrm{G} \tag{5.30}$$

where

$$\mathcal{H}_M = \pi\,\partial_0\Psi - \widetilde{\mathcal{L}}_M$$
$$\mathcal{H}_G = \pi_k{}^\alpha\partial_0 b^k{}_\alpha + \tfrac{1}{2}\pi_{ij}{}^\alpha\partial_0 A^{ij}{}_\alpha - \widetilde{\mathcal{L}}_G.$$

The *total Hamiltonian* is defined by the expression

$$\mathcal{H}_T = \mathcal{H}_c + u^k{}_0\phi_k{}^0 + \tfrac{1}{2}u^{ij}{}_0\phi_{ij}{}^0 + (u\cdot\phi) \tag{5.31}$$

where ϕ symbolically denotes all additional primary constraints, if they exist (if constraints), and $H_T = \int d^3x\,\mathcal{H}_T$.

The evaluation of the consistency conditions of the sure primary constraints, $\dot{\phi}_k{}^0 = \{\pi_k{}^0, H_T\} \approx 0$ and $\dot{\phi}_{ij}{}^0 = \{\pi_{ij}{}^0, H_T\} \approx 0$, is essentially simplified if we previously find the dependence of the Hamiltonian on the unphysical variables $b^k{}_0$ and $A^{ij}{}_0$. We shall show that $\mathcal{H}_c$ is linear in $b^k{}_0$ and $A^{ij}{}_0$,

$$\mathcal{H}_c = b^k{}_0\mathcal{H}_k - \tfrac{1}{2}A^{ij}{}_0\mathcal{H}_{ij} + \partial_\alpha D^\alpha \tag{5.32}$$

where $\partial_\alpha D^\alpha$ is a three-divergence term, while the possible extra primary constraints ϕ are independent of $b^k{}_0$ and $A^{ij}{}_0$. Consequently, the consistency conditions (5.31) will result in the following *secondary* constraints:

$$\mathcal{H}_k \approx 0 \qquad \mathcal{H}_{ij} \approx 0. \tag{5.33}$$

Relations (5.32) will be the basic result of this section.

The (3 + 1) decomposition of spacetime

The investigation of the dependence of the Hamiltonian on $b^k{}_0$ will lead us to the so-called (3+1) decomposition of spacetime.

In order to find the dependence of the inverse tetrad $h_k{}^\mu$ on $b^k{}_0$, we observe that the orthogonality conditions $b^k{}_\mu h_k{}^\nu = \delta^\nu_\mu$ and $b^k{}_\mu h_l{}^\mu = \delta^k_l$ lead to the relations

$$\begin{aligned} {}^3h_a{}^\alpha b^a{}_\beta &\equiv (h_a{}^\alpha - h_a{}^0 h_0{}^\alpha/h_0{}^0)b^a{}_\beta = \delta^\alpha{}_\beta \\ h_a &\equiv h_a{}^0/h_0{}^0 = -{}^3h_a{}^\alpha b^0{}_\alpha. \end{aligned} \tag{5.34}$$

Since ${}^3h_a{}^\alpha$ is the inverse of $b^a{}_\beta$, both ${}^3h_a{}^\alpha$ and h_a are independent of $b^k{}_0$.

Let us now construct the normal $\boldsymbol{n}$ to the hypersurface $\Sigma_0 : x^0 = \text{constant}$. By noting that the vector $\boldsymbol{l} = (h_k{}^0)$ is orthogonal to the three basis vectors $\boldsymbol{e}_\alpha$ lying in Σ_0, $\boldsymbol{l}\cdot\boldsymbol{e}_\alpha = h_k{}^0 b^k{}_\alpha = 0$, we find that

$$\boldsymbol{n} = \frac{\boldsymbol{l}}{\sqrt{\boldsymbol{l}\cdot\boldsymbol{l}}} \qquad n_k = \frac{h_k{}^0}{\sqrt{g^{00}}}. \tag{5.35}$$

The independence of $\boldsymbol{n}$ on $b^k{}_0$ follows from $\boldsymbol{l} \sim (1, h_a)$.

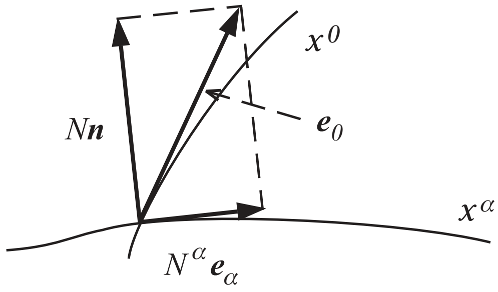

Figure 5.2. The ADM decomposition of the time displacement vector.

The four vectors $(\boldsymbol{n}, \boldsymbol{e}_\alpha)$ define the so-called ADM basis. Introducing the projectors on $\boldsymbol{n}$ and Σ_0,

$$(P_\perp)^l_k = n_k n^l \qquad (P_\parallel)^l_k = \delta^l_k - n_k n^l$$

we can express any vector in terms of its parallel and orthogonal components:

$$\begin{aligned} V_k &= n_k V_\perp + V_{\bar{k}} \\ V_\perp = n_k V^k \qquad V_{\bar{k}} &\equiv (V_\parallel)_k = (P_\parallel)^l_k V_l. \end{aligned} \tag{5.36}$$

Here, by convention, a bar over an index 'k' in $V_{\bar{k}}$ denotes the fact that the contraction of $V_{\bar{k}}$ with n^k vanishes. An analogous decomposition can be defined for any tensor.

The parallel component of V_k can be written in the form

$$V_{\bar{k}} = (P_\parallel)^l_k h_l{}^\mu V_\mu \equiv h_{\bar{k}}{}^\mu V_\mu$$

where $h_{\bar{k}}{}^\mu$ does not depend on $b^k{}_0$. Indeed, the quantity

$$h_{\bar{k}}{}^\mu = (P_\parallel)^l_k h_l{}^\mu = (P_\parallel)^l_k\, {}^3h_l{}^\mu$$

where ${}^3h_l{}^\mu \equiv h_l{}^\mu - h_l{}^0 h_0{}^\mu / h_0{}^0$, is a formal generalization of ${}^3h_a{}^\alpha$, having the zero value whenever at least one of its indices is zero. Therefore, both n_k and $h_{\bar{k}}{}^\mu$ are independent of $b^k{}_0$.

The decomposition of the vector $\boldsymbol{e}_0$ in the ADM basis yields

$$\boldsymbol{e}_0 = N\boldsymbol{n} + N^\alpha \boldsymbol{e}_\alpha \tag{5.37a}$$

where N and N^α are called the lapse and shift functions, respectively. Multiplying this equation with $\mathrm{d}x^0$ we see that, geometrically, N determines the projection of the time displacement vector $\mathrm{d}x^0\boldsymbol{e}_0$ on the normal, while N^α measures the deviation of this vector from $\boldsymbol{n}$ (figure 5.2). The lapse and shift functions are linear in $b^k{}_0$:

$$N = \boldsymbol{e}_0 \cdot \boldsymbol{n} = n_k b^k{}_0 = 1/\sqrt{g^{00}}$$
$$N^\alpha = \boldsymbol{e}_0 \cdot \boldsymbol{e}_\beta\, {}^3g^{\beta\alpha} = h_{\bar{k}}{}^\alpha b^k{}_0 = -g^{0\alpha}/g^{00} \tag{5.37b}$$

where ${}^3g^{\beta\alpha}$ is the inverse of $g_{\alpha\beta}$.

Thus, we have found 16 variables $(n_k, h_{\bar{k}}{}^\mu, N, N^\alpha)$ with a clear geometric meaning and simple dependence on $b^k{}_0$, which can always be used instead of $h_k{}^\mu$.

Using the fact that N and N^α are linear in $b^k{}_0$, the canonical Hamiltonian (5.32) can easily be brought into an equivalent form:

$$\mathcal{H}_\mathrm{c} = N\mathcal{H}_\perp + N^\alpha \mathcal{H}_\alpha - \tfrac{1}{2}A^{ij}{}_0\mathcal{H}_{ij} + \partial_\alpha D^\alpha \tag{5.38a}$$

where

$$\mathcal{H}_k = n_k\mathcal{H}_\perp + h_{\bar{k}}{}^\alpha\mathcal{H}_\alpha$$
$$\mathcal{H}_\perp = n^k\mathcal{H}_k \qquad \mathcal{H}_\alpha = b^k{}_\alpha\mathcal{H}_k. \tag{5.38b}$$

Construction of the Hamiltonian

The matter Hamiltonian. Let us now turn to the proof of (5.38) for the matter Hamiltonian. The Lagrangian of the matter fields depends on the time derivative $\partial_0\Psi$ only through the covariant derivative $\nabla_k\Psi$. It is convenient to decompose $\nabla_k\Psi$ into the orthogonal and parallel components,

$$\nabla_k\Psi = n_k\nabla_\perp\Psi + \nabla_{\bar{k}}\Psi \equiv n_k h_\perp{}^\mu\nabla_\mu\Psi + h_{\bar{k}}{}^\alpha\nabla_\alpha\Psi \tag{5.39}$$

because $\nabla_{\bar{k}}\Psi$ does not depend either on velocities or on unphysical variables $(b^k{}_0, A^{ij}{}_0)$, as follows from $h_{\bar{k}}{}^0 = 0$. Replacing this decomposition into the matter Lagrangian leads to the relation

$$\mathcal{L}_\mathrm{M} = \overline{\mathcal{L}}_\mathrm{M}(\Psi, \nabla_{\bar{k}}\Psi; \nabla_\perp\Psi, n^k)$$

where the complete dependence on velocities and unphysical variables is contained in $\nabla_\perp\Psi$. Then, using the factorization of the determinant

$$b = \det(b^k{}_\mu) = NJ \tag{5.40}$$

where J does not depend on $b^k{}_0$, the expression for momentum can be written as

$$\pi \equiv \frac{\partial(b\mathcal{L}_\mathrm{M})}{\partial\Psi_{,0}} = J\frac{\partial\overline{\mathcal{L}}_\mathrm{M}}{\partial\nabla_\perp\Psi}.$$

Finally, using the relation

$$\nabla_0\Psi \equiv N\nabla_\perp\Psi + N^\alpha\nabla_\alpha\Psi = \partial_0\Psi + \tfrac{1}{2}A^{ij}{}_0\Sigma_{ij}\Psi$$

to express the velocities $\partial_0\Psi$, the canonical Hamiltonian for the matter field takes the form (5.38), where

$$\mathcal{H}_\alpha^{\rm M} = \pi\nabla_\alpha\Psi \qquad \mathcal{H}_{ij}^{\rm M} = \pi\Sigma_{ij}\Psi$$
$$\mathcal{H}_\perp^{\rm M} = \pi\nabla_\perp\Psi - J\overline{\mathcal{L}}_{\rm M} = J\left(\frac{\partial\overline{\mathcal{L}}_{\rm M}}{\partial\nabla_\perp\Psi}\nabla_\perp\Psi - \overline{\mathcal{L}}_{\rm M}\right) \tag{5.41}$$

and $D_{\rm M}^\alpha = 0$.

The expressions for $\mathcal{H}_\alpha^{\rm M}$ and $\mathcal{H}_{ij}^{\rm M}$ are independent of the unphysical variables, as $\nabla_\alpha\Psi$ is independent of them. They do not depend on the specific form of the original Lagrangian $\mathcal{L}_{\rm M}$, but only on the transformation properties of fields. They are called the *kinematical* terms of the Hamiltonian.

The term $\mathcal{H}_\perp^{\rm M}$ is *dynamical*, as it depends on the choice of $\mathcal{L}_{\rm M}$. It represents the Legendre transformation of the function $\mathcal{L}_{\rm M}$ with respect to the 'velocity' $\nabla_\perp\Psi$. After eliminating $\nabla_\perp\Psi$ with the help of the relation defining π, the dynamical Hamiltonian takes the form

$$\mathcal{H}_\perp^{\rm M} = \mathcal{H}_\perp^{\rm M}(\Psi, \nabla_{\bar k}\Psi; \pi/J, n^k)$$

from which it follows that $\mathcal{H}_\perp^{\rm M}$ does not depend on unphysical variables.

If the matter Lagrangian is singular, the equations for momenta give rise to additional primary constraints, which are again independent of unphysical variables.

Example 4. Let us consider the case of the spin-$\frac{1}{2}$ matter field. This is an important case as we believe that most of the matter in the Universe (quarks and leptons) is described by the Dirac field, the Lagrangian of which is given by $\widetilde{\mathcal{L}}_{\rm D} = \frac{1}{2}b[\mathrm{i}\bar\psi\gamma^k\overleftrightarrow{\nabla}_k\psi - 2m\bar\psi\psi]$. If we take ψ and $\bar\psi$ as basic Lagrangian variables, and denote their momenta by $\bar\pi$ and $\mathrm{i}\pi$, it follows that additional primary constraints exist:

$$\phi \equiv \pi + \tfrac{1}{2}\mathrm{i}J\gamma^\perp\psi \approx 0 \qquad \bar\phi \equiv \bar\pi - \tfrac{1}{2}\mathrm{i}J\bar\psi\gamma^\perp \approx 0.$$

These constraints are second class, since

$$\{\phi(\boldsymbol{x}), \bar\phi(\boldsymbol{x}')\} = \mathrm{i}J\gamma^\perp\delta(\boldsymbol{x}-\boldsymbol{x}') \qquad \det\gamma^\perp = 1.$$

Following the general procedure we find that the canonical Hamiltonian is of the form (5.38*a*), where

$$\mathcal{H}_{ij}^{\rm M} = \bar\pi\Sigma_{ij}\psi - \bar\psi\Sigma_{ij}\pi \qquad \mathcal{H}_\alpha^{\rm M} = \bar\pi\nabla_\alpha\psi + (\nabla_\alpha\bar\psi)\pi$$
$$\mathcal{H}_\perp^{\rm M} = -J\tfrac{1}{2}[\mathrm{i}\bar\psi\gamma^k\overleftrightarrow{\nabla}_{\bar k}\psi - 2m\bar\psi\psi].$$

Since $\overline{\mathcal{L}}_{\rm D}$ is linear in velocities, we have $\mathcal{H}_\perp^{\rm M} = -J\overline{\mathcal{L}}_{\rm D}(\nabla_\perp\psi, \nabla_\perp\bar\psi = 0)$, leading to the last equation.

The term $(u \cdot \phi)$ in the total Hamiltonian, stemming from additional primary constraints, has the form $(u \cdot \phi)_{\rm M} = \bar{u}\phi + \bar{\phi}u$. Since $(\phi, \bar{\phi})$ are second-class constraints, their consistency conditions fix the multipliers $(\bar{u}, u)$, and no new constraints appear. After the construction of the Dirac brackets, the constraints $(\phi, \bar{\phi})$ can be used as strong equalities.

Gravitational Hamiltonian. The construction of the gravitational Hamiltonian can be performed in a very similar way, the role of $\nabla_k \Psi$ being taken over by $T^i{}_{km}$ and $R^{ij}{}_{km}$. In the first step we decompose the torsion and curvature, in the last two indices, into orthogonal and parallel components:

$$
\begin{aligned}
T^i{}_{km} &= T^i{}_{\bar{k}\bar{m}} + n_k T^i{}_{\perp\bar{m}} + n_m T^i{}_{\bar{k}\perp} \\
R^{ij}{}_{km} &= R^{ij}{}_{\bar{k}\bar{m}} + n_k R^{ij}{}_{\perp\bar{m}} + n_m R^{ij}{}_{\bar{k}\perp}.
\end{aligned}
\tag{5.42}
$$

The parallel components $T^i{}_{\bar{k}\bar{m}}$ and $R^{ij}{}_{\bar{k}\bar{m}}$ are independent of the velocities and unphysical variables. The replacement in the gravitational Lagrangian yields the relation

$$
\mathcal{L}_{\rm G} = \overline{\mathcal{L}}_{\rm G}(T^i{}_{\bar{k}\bar{m}}, R^{ij}{}_{\bar{k}\bar{m}}; T^i{}_{\perp\bar{k}}, R^{ij}{}_{\perp\bar{k}}, n^k).
$$

Using the factorization of the determinant (5.40), the relations defining the momenta take the form

$$
\hat{\pi}_i{}^{\bar{k}} = J \frac{\partial \overline{\mathcal{L}}_{\rm G}}{\partial T^i{}_{\perp\bar{k}}} \qquad \hat{\pi}_{ij}{}^{\bar{k}} = J \frac{\partial \overline{\mathcal{L}}_{\rm G}}{\partial R^{ij}{}_{\perp\bar{k}}}
$$

where $\hat{\pi}_i{}^{\bar{k}} \equiv \pi_i{}^\alpha b^k{}_\alpha$ and $\hat{\pi}_{ij}{}^{\bar{k}} \equiv \pi_{ij}{}^\alpha b^k{}_\alpha$ are 'parallel' gravitational momenta. The velocities $\partial_0 b^i{}_\alpha$ and $\partial_0 A^{ij}{}_\alpha$ can be calculated from the relations

$$
\begin{aligned}
N(T^i{}_{\perp\alpha} + A^i{}_{\perp\alpha}) + N^\beta(T^i{}_{\beta\alpha} + A^i{}_{\beta\alpha}) &= \partial_0 b^i{}_\alpha - \partial_\alpha b^i{}_0 + \tfrac{1}{2} A^{mn}{}_0 (\Sigma^1_{mn})^i{}_j b^j{}_\alpha \\
N R^{ij}{}_{\perp\alpha} + N^\beta R^{ij}{}_{\beta\alpha} &= \partial_0 A^{ij}{}_\alpha - \partial_\alpha A^{ij}{}_0 + \tfrac{1}{2} A^{mn}{}_0 (\Sigma^2_{mn})^{ij}{}_{kl} A^{kl}{}_\alpha
\end{aligned}
$$

obtained from the definitions of $T^i{}_{0\alpha}$ and $R^{ij}{}_{0\alpha}$. After a simple algebra, the canonical Hamiltonian takes the form (5.38), where

$$
\begin{aligned}
\mathcal{H}^{\rm G}_{ij} &= 2\pi_{[i}{}^\alpha b_{j]\alpha} + 2\pi_{k[i}{}^\alpha A^k{}_{j]\alpha} + \partial_\alpha \pi_{ij}{}^\alpha \\
\mathcal{H}^{\rm G}_\alpha &= \pi_i{}^\beta T^i{}_{\alpha\beta} + \tfrac{1}{2}\pi_{ij}{}^\beta R^{ij}{}_{\alpha\beta} - b^k{}_\alpha \nabla_\beta \pi_k{}^\beta \\
\mathcal{H}^{\rm G}_\perp &= J\left(\frac{1}{J}\hat{\pi}_i{}^{\bar{m}} T^i{}_{\perp\bar{m}} + \frac{1}{2J}\hat{\pi}_{ij}{}^{\bar{m}} R^{ij}{}_{\perp\bar{m}} - \overline{\mathcal{L}}_{\rm G}\right) - n^k \nabla_\beta \pi_k{}^\beta \\
D^\alpha_{\rm G} &= b^i{}_0 \pi_i{}^\alpha + \tfrac{1}{2} A^{ij}{}_0 \pi_{ij}{}^\alpha.
\end{aligned}
\tag{5.43}
$$

The expressions $T^i{}_{\perp\bar{m}}$ and $R^{ij}{}_{\perp\bar{m}}$ in $\mathcal{H}^{\rm G}_\perp$ should be eliminated with the help of the equations defining momenta $\hat{\pi}_i{}^{\bar{m}}$ and $\hat{\pi}_{ij}{}^{\bar{m}}$.

Example 5. As an illustration of this procedure for constructing the canonical Hamiltonian, we shall consider the case of Einstein–Cartan theory. From the form of the Lagrangian $\widetilde{\mathcal{L}} = -abR$, we find additional primary constraints:

$$\pi_i{}^\alpha \approx 0 \qquad \phi_{ij}{}^\alpha \equiv \pi_{ij}{}^\alpha + 4aJn_{[i}h_{j]}{}^\alpha \approx 0.$$

Since the Lagrangian does not depend on $\partial_0 b^i{}_\alpha$, the only way to eliminate these velocities from the Hamiltonian (5.29*b*) is to use the constraints $\pi_i{}^\alpha \approx 0$. In this way, at the same time, we eliminate the terms proportional to $\pi_i{}^\alpha$ (or $\hat{\pi}_i{}^{\bar{m}}$) in (5.43). After that we easily obtain

$$\begin{aligned} \mathcal{H}_{ij} &= \nabla_\alpha \pi_{ij}{}^\alpha & \mathcal{H}_\alpha &= \tfrac{1}{2}\pi_{ij}{}^\beta R^{ij}{}_{\alpha\beta} \\ \mathcal{H}_\perp &= aJR^{\bar{m}\bar{n}}{}_{\bar{m}\bar{n}} & D^\alpha &= \tfrac{1}{2}A^{ij}{}_0 \pi_{ij}{}^\alpha \end{aligned}$$

where, in calculating $\mathcal{H}_\perp$, we have used the relation $\overline{\mathcal{L}}_{\mathrm{G}} = -a(R^{\bar{m}\bar{n}}{}_{\bar{m}\bar{n}} + 2R^{\perp\bar{n}}{}_{\perp\bar{n}})$ and the constraint $\phi_{ij}{}^\alpha$.

It should be stressed that these results describe the *canonical* Hamiltonian, while the presence of the determined multipliers in the *total* Hamiltonian will bring certain modifications to these expressions, as we shall soon see.

Consistency of the theory and gauge conditions

1. The result of the preceding exposition is the conclusion that the total Hamiltonian of the theory can be written in the form

$$\begin{aligned} \mathcal{H}_{\mathrm{T}} &= \widehat{\mathcal{H}}_{\mathrm{T}} + \partial_\alpha D^\alpha \\ \widehat{\mathcal{H}}_{\mathrm{T}} &\equiv \widehat{\mathcal{H}}_{\mathrm{c}} + u^i{}_0 \pi_i{}^0 + \tfrac{1}{2}u^{ij}{}_0 \pi_{ij}{}^0 + (u \cdot \phi) \end{aligned} \tag{5.44a}$$

where $\pi_i{}^0$, $\pi_{ij}{}^\alpha$ and ϕ are primary constraints, the us are the related multipliers, the three-divergence $\partial_\alpha D^\alpha$ is written as a separate term and the canonical Hamiltonian has the form

$$\begin{aligned} \widehat{\mathcal{H}}_{\mathrm{c}} &= b^k{}_0 \mathcal{H}_k - \tfrac{1}{2}A^{ij}{}_0 \mathcal{H}_{ij} \\ &= N\mathcal{H}_\perp + N^\alpha \mathcal{H}_\alpha - \tfrac{1}{2}A^{ij}{}_0 \mathcal{H}_{ij}. \end{aligned} \tag{5.44b}$$

The fact that $\widehat{\mathcal{H}}_{\mathrm{c}}$ is linear in unphysical variables enables us to easily find the consistency conditions of the sure primary constraints (5.28):

$$\mathcal{H}_\perp \approx 0 \qquad \mathcal{H}_\alpha \approx 0 \qquad \mathcal{H}_{ij} \approx 0. \tag{5.45}$$

PGT is invariant under the ten-parameter group of local Poincaré transformations. From the general considerations given at the beginning of this section it follows that *ten* FC constraints should exist in the theory. It is not difficult to guess that these ten constraints are, in fact, *four* constraints $(\mathcal{H}_\perp, \mathcal{H}_\alpha)$, describing local translations, and *six* constraints $\mathcal{H}_{ij}$, related to local Lorentz rotations. An explicit proof of this statement is left for the next chapter. As a consequence, the consistency conditions of the secondary constraints (5.45) are *automatically* satisfied, as their PBs with the Hamiltonian are given as linear combinations of ten FC constraints.

2. Although gauge-invariant formulation of the theory is very important for analysis of its general structure, practical calculations are often simplified if we fix the gauge.

If the number of gauge conditions is smaller than the number of FC constraints, the gauge symmetry is fixed only partially. This is the case when we choose the orientation of the local Lorentz frame $\boldsymbol{h}_k = \boldsymbol{e}_k$ so that the time direction coincides with the normal to Σ_0:

$$\boldsymbol{h}_0 = \boldsymbol{n}. \tag{5.46a}$$

This choice defines the *time gauge*, which violates the local Lorentz symmetry: mutual rotations of the vectors $\boldsymbol{h}_a$ are still allowed (spatial rotations), but any change of $\boldsymbol{h}_0$ (boost transformations) is now forbidden.

In the basis (5.46a) the normal $\boldsymbol{n}$ has the components

$$n_k = \boldsymbol{n} \cdot \boldsymbol{h}_k = (1, 0, 0, 0)$$

so that the projection of $\boldsymbol{V} = (V_k)$ on $\boldsymbol{n}$ becomes $\boldsymbol{V} \cdot \boldsymbol{n} = V_0$. Using the relation $\boldsymbol{n} \cdot \boldsymbol{h}_a = h_a{}^0 \boldsymbol{n} \cdot \boldsymbol{e}_0$, the time gauge condition can be expressed as

$$h_a{}^0 = 0 \tag{5.46b}$$

while from (5.34) we find an equivalent form:

$$b^0{}_\alpha = 0. \tag{5.46c}$$

Since the time gauge fixes the boost symmetry, it affects the form of the Hamiltonian by fixing the multiplier $A^{0c}{}_0$, related to the constraint $\mathcal{H}_{0c}$. This is easily seen to be a consequence of the consistency of (5.46c),

$$\dot{b}^0{}_\alpha = \{b^0{}_\alpha, H_\mathrm{T}\} = \int \mathrm{d}^3x' \, \{b^0{}_\alpha, N'\mathcal{H}'_\perp + N'^\alpha \mathcal{H}'_\alpha\} + A^{0c}{}_0 b_{c\alpha} \approx 0$$

where we used $\{b^0{}_\alpha, \mathcal{H}_{0c}\} = b_{c\alpha}\delta$.

After choosing the time gauge, the constraints $\mathcal{H}_{0c} \approx 0$ and $b^0{}_\alpha \approx 0$ can be treated as second-class constraints. The construction of the preliminary Dirac brackets is very simple. From the definition

$$\{A, B\}^* = \{A, B\} + \int [\{A, b^0{}_\alpha\} h^{c\alpha} \{\mathcal{H}_{0c}, B\} - \{A, \mathcal{H}_{0c}\} h^{c\alpha} \{b^0{}_\alpha, B\}]$$

it follows that for all variables except $(b^0{}_\alpha, \pi_0{}^\beta)$ the Dirac brackets reduce to the Poisson ones, while $(b^0{}_\alpha, \pi_0{}^\beta)$ can be eliminated from the theory by using $b^0{}_\alpha \approx 0$ and $\mathcal{H}_{0c} \approx 0$.

This procedure illustrates the standard method for imposing gauge condition at the end of the construction of the Hamiltonian formulation of the theory. Sometimes, gauge conditions are used directly in the action integral, thus reducing the number of dynamical degrees of freedom from the very beginning. This is not allowed in principle, as by doing so we are losing some of the equations of motion of the original theory.

5.3 Specific models

In this section, we shall use the general Hamiltonian method developed so far to study two specific but important dynamical models: EC theory and teleparallel theory.

Einstein–Cartan theory

The Einstein–Cartan theory (3.51) represents a direct extension of GR to the U_4 spacetime. We shall investigate here the Hamiltonian structure of this theory when matter fields are absent (Nikolić 1995). In this case the EC theory reduces to GR, and the action takes the form $I_{\mathrm{HP}} = -a\int \mathrm{d}^4x\, bR$, known from GR as the Hilbert–Palatini form.

Using the relation

$$bR = -\tfrac{1}{4}\varepsilon^{\mu\nu\lambda\rho}_{mnkl} b^k{}_\lambda b^l{}_\rho R^{mn}{}_{\mu\nu} \qquad \varepsilon^{\mu\nu\lambda\rho}_{mnkl} \equiv \varepsilon^{\mu\nu\lambda\rho}\varepsilon_{mnkl}$$

the action can be written as

$$I_{\mathrm{HP}} = \tfrac{1}{2}a\int \mathrm{d}^4x\, \varepsilon^{\mu\nu\lambda\rho}_{mnkl} b^k{}_\lambda b^l{}_\rho(\partial_\mu A^{mn}{}_\nu + A^m{}_{s\mu}A^{sn}{}_\nu). \tag{5.47}$$

The Hamiltonian and constraints. In addition to the sure primary constraints, $\pi_i{}^0 \approx 0$ i $\pi_{ij}{}^0 \approx 0$, in this case extra primary constraints also exist,

$$\pi_i{}^\alpha \approx 0 \qquad \phi_{ij}{}^\alpha \equiv \pi_{ij}{}^\alpha - a\varepsilon^{0\alpha\beta\gamma}_{ijmn} b^m{}_\beta b^n{}_\gamma \approx 0 \tag{5.48}$$

in agreement with the result of example 5.

Since the Lagrangian is linear in velocities $\dot{A}$, the canonical Hamiltonian is given by $\mathcal{H}_{\mathrm{c}} = -\mathcal{L}(\dot{A} = 0)$. It can be written as a linear function of unphysical variables, up to a three-divergence,

$$\mathcal{H}_{\mathrm{c}} = b^i{}_0\mathcal{H}_i - \tfrac{1}{2}A^{ij}{}_0\mathcal{H}_{ij} + \partial_\alpha D^\alpha \tag{5.49}$$

where

$$\mathcal{H}_i = -\tfrac{1}{2}a\varepsilon^{0\alpha\beta\gamma}_{ijmn} b^j{}_\alpha R^{mn}{}_{\beta\gamma} \qquad \mathcal{H}_{ij} = -a\varepsilon^{0\alpha\beta\gamma}_{ijmn} b^m{}_\alpha T^n{}_{\beta\gamma}$$
$$D^\alpha = \tfrac{1}{2}a\varepsilon^{0\alpha\beta\gamma}_{ijmn} b^m{}_\beta b^n{}_\gamma A^{ij}{}_0.$$

It is not difficult to see, using $\phi_{ij}{}^\alpha$, that the quantities $\mathcal{H}_i$, $\mathcal{H}_{ij}$ and D^α are equivalent to the corresponding expressions in example 5. The form we use here is more suitable for a direct comparison with the Lagrangian formalism, since after eliminating the momenta the constraints are given as functions of the fields.

The total Hamiltonian has the form

$$\mathcal{H}_{\mathrm{T}} = \mathcal{H}_{\mathrm{c}} + u^i{}_0\pi_i{}^0 + \tfrac{1}{2}u^{ij}{}_0\pi_{ij}{}^0 + u^i{}_\alpha\pi_i{}^\alpha + \tfrac{1}{2}u^{ij}{}_\alpha\phi_{ij}{}^\alpha.$$

The consistency conditions of the sure primary constraints result in ten secondary constraints:

$$\mathcal{H}_i \approx 0 \qquad \mathcal{H}_{ij} \approx 0. \tag{5.50a}$$

The consistency conditions for $\pi_i{}^\alpha$ have the form

$$\chi_i{}^\alpha \equiv \tfrac{1}{2} a \varepsilon^{\alpha 0 \beta\gamma}_{ijmn} (b^j{}_0 R^{mn}{}_{\beta\gamma} - 2b^j{}_\beta \underline{R}^{mn}{}_{0\gamma}) \approx 0 \tag{5.50b}$$

where an underbar in $\underline{R}^{mn}{}_{0\gamma}$ denotes that the term $\dot{A}^{mn}{}_\gamma$ is replaced with $u^{mn}{}_\gamma$ (the equation of motion for $A^{mn}{}_\gamma$ has the form $\dot{A}^{mn}{}_\gamma = u^{mn}{}_\gamma$). These relations are used to determine the multipliers $u^{mn}{}_\gamma$. However, the 12 conditions $\chi_i{}^\alpha \approx 0$ are insufficient for a complete determination of $u^{mn}{}_\alpha$, but six additional conditions will be found later.

Finally, the consistency of $\phi_{ij}{}^\alpha$ leads to

$$\chi_{ij}{}^\alpha \equiv -a \varepsilon^{\alpha 0 \beta\gamma}_{ijmn} (b^m{}_0 T^n{}_{\beta\gamma} - 2b^m{}_\beta \underline{T}^n{}_{0\gamma}) \approx 0 \tag{5.50c}$$

where the underbar in $\underline{T}^n{}_{0\gamma}$ means that the term $\dot{b}^n{}_\gamma$ is replaced with $u^n{}_\gamma$ ($\dot{b}^n{}_\gamma = u^n{}_\gamma$ on the basis of the equations of motion). Among the 18 relations $\chi_{ij}{}^\alpha \approx 0$ there are 12 conditions determining the multipliers $u^m{}_\gamma$, and six constraints. The complete number of secondary constraints is $10 + 6 = 16$. In the following exposition we shall omit writing underbars in $\underline{R}^{mn}{}_{0\gamma}$ and $\underline{T}^n{}_{0\gamma}$ for simplicity.

The previous consistency conditions can be rewritten in the 'covariant' form:

$$\begin{aligned} (-\mathcal{H}_i, \chi_i{}^\alpha) : \quad & \chi_i{}^\mu \equiv \tfrac{1}{2} a \varepsilon^{\mu\nu\rho\sigma}_{ijmn} b^j{}_\nu R^{mn}{}_{\rho\sigma} \approx 0 \\ (\mathcal{H}_{ij}, \chi_{ij}{}^\alpha) : \quad & \chi_{ij}{}^\mu \equiv -a \varepsilon^{\mu\nu\rho\sigma}_{ijmn} b^m{}_\nu T^n{}_{\rho\sigma} \approx 0. \end{aligned}$$

The first set is equivalent to

$$h_k{}^\mu R^k{}_i - \tfrac{1}{2} h_i{}^\mu R \approx 0 \tag{5.51a}$$

which we recognize as Einstein's equations in vacuum (they can be transformed into $R_{ij} = 0$). Here, among 16 conditions we have four constraints ($\mathcal{H}_i \approx 0$), and 12 relations which serve to determine the multipliers $u^{ij}{}_\alpha$ ($\chi_i{}^\alpha \approx 0$). The second set of conditions implies that the torsion vanishes,

$$T^k{}_{\mu\nu} \approx 0 \tag{5.51b}$$

which is a consequence of the absence of matter. These 24 equations contain 12 constraints ($T^k{}_{\alpha\beta} \approx 0$) and 12 conditions on multipliers $u^k{}_\alpha$ ($T^k{}_{0\alpha} \approx 0$).

The consistency conditions for secondary constraints do not produce new constraints. Indeed, the consistency of $\mathcal{H}_i$ is automatically satisfied,

$$\dot{\mathcal{H}}_i = \nabla_\alpha \chi_i{}^\alpha + A^m{}_{i0} \mathcal{H}_m \approx 0 \tag{5.52a}$$

while the consistency of $\varepsilon^{0\alpha\beta\gamma}T^k{}_{\alpha\beta} \approx 0$,

$$\varepsilon^{0\alpha\beta\gamma}(b^m{}_0 R^i{}_{m\alpha\beta} - 2b^m{}_\alpha R^i{}_{m0\beta}) \approx 0 \qquad (5.52b)$$

yields additional requirements on the multipliers $u^{ij}{}_\alpha$ (see exercises 15 and 16). The explicit form of all determined multipliers will be given later.

In the set of primary constraints, 10 of them are FC ($\pi_i{}^0, \pi_{ij}{}^0$) and the remaining 30 are second class ($\pi_i{}^\alpha, \phi_{ij}{}^\alpha$); from 16 secondary constraints, 10 are FC ($\mathcal{H}_i, \mathcal{H}_{ij}$) and six are second class. Therefore, the total number of FC and second-class constraints is $N_1 = 20$, $N_2 = 36$, respectively. Taking into account that there is one gauge condition corresponding to each FC constraint and that the number of independent field components is $N = 40$, we find that the total number of physical degrees of freedom is

$$N^* = 2N - (2N_1 + N_2) = 80 - 76 = 4.$$

Thus, the number of independent gravitational degrees of freedom is four in phase space or, equivalently, two in configuration space.

Determined multipliers. The existence of 20 FC constraints is related to the fact that the variables ($b^i{}_0, A^{ij}{}_0$) and the multipliers ($u^i{}_0, u^{ij}{}_0$) are arbitrary functions of time; the existence of 30 second-class constraints is reflected in the fact that the multipliers ($u^i{}_\alpha, u^{ij}{}_\alpha$) are determined functions of the remaining phase-space variables. From the relation $T^i{}_{0\alpha} \approx 0$ and the definition of $R^{ij}{}_{0\alpha}$ it follows that

$$\begin{aligned} u^k{}_\alpha &= \nabla_\alpha b^k{}_0 - A^k{}_{m0} b^m{}_\alpha \\ u^{ij}{}_\alpha &\equiv R^{ij}{}_{0\alpha} + \nabla_\alpha A^{ij}{}_0 \end{aligned}$$

so that the determination of $u^{ij}{}_\alpha$ reduces to the determination of $R^{ij}{}_{0\alpha}$. The 'multipliers' $R^{mn}{}_{\perp\bar{r}} = (R^{\bar{m}\bar{n}}{}_{\perp\bar{r}}, R^{\perp\bar{n}}{}_{\perp\bar{r}})$ can be determined from

$$R^{\bar{m}\bar{n}}{}_{\perp\bar{r}} = R_{\perp\bar{r}}{}^{\bar{m}\bar{n}} \qquad R^{\bar{n}}{}_{\bar{r}} \equiv R^{\perp\bar{n}}{}_{\perp\bar{r}} + R^{\bar{m}\bar{n}}{}_{\bar{m}\bar{r}} = 0.$$

The contribution of the determined multipliers to the total Hamiltonian can be found starting from the relations

$$\begin{aligned} u^i{}_\alpha \pi_i{}^\alpha &= \partial_\alpha(b^i{}_0 \pi_i{}^0) - b^i{}_0 \nabla_\alpha \pi_i{}^\alpha - A^{ij}{}_0 \pi_{[i}{}^\alpha b_{j]\alpha} \\ \tfrac{1}{2} u^{ij}{}_\alpha \phi_{ij}{}^\alpha &= \partial_\alpha(\tfrac{1}{2} A^{ij}{}_0 \phi_{ij}{}^\alpha) - \tfrac{1}{2} A^{ij}{}_0 \nabla_\alpha \phi_{ij}{}^\alpha + \tfrac{1}{2} R^{ij}{}_{0\alpha} \phi_{ij}{}^\alpha . \end{aligned}$$

Using the decomposition

$$R^{ij}{}_{0\alpha}\phi_{ij}{}^\alpha = R^{ij}{}_{0\alpha}\pi_{ij}{}^\alpha + 4aJR^{\perp\bar{n}}{}_{0\bar{n}} = b^i{}_0(R^{mn}{}_{i\alpha}\pi_{mn}{}^\alpha + 4aJR^{\perp\bar{n}}{}_{i\bar{n}})$$

the total Hamiltonian takes the form

$$\mathcal{H}_{\mathrm{T}} \equiv b^i{}_0 \overline{\mathcal{H}}_i - \tfrac{1}{2} A^{ij}{}_0 \overline{\mathcal{H}}_{ij} + \partial_\alpha \bar{D}^\alpha + u^i{}_0 \pi_i{}^0 + \tfrac{1}{2} u^{ij}{}_0 \pi_{ij}{}^0 \qquad (5.53)$$

where the components $\overline{\mathcal{H}}_i, \overline{\mathcal{H}}_{ij}$ and $\bar{D}^\alpha$ contain the contributions of the determined multipliers:

$$\begin{aligned}
\overline{\mathcal{H}}_{ij} &= \mathcal{H}_{ij} + 2\pi_{[i}{}^\alpha b_{j]\alpha} + \nabla_\alpha \phi_{ij}{}^\alpha = 2\pi_{[i}{}^\alpha b_{j]\alpha} + \nabla_\alpha \pi_{ij}{}^\alpha \\
\overline{\mathcal{H}}_i &= \mathcal{H}_i + 2aJ R^{\perp\bar{n}}{}_{i\bar{n}} + \tfrac{1}{2} R^{mn}{}_{i\beta}\pi_{mn}{}^\beta - \nabla_\alpha \pi_i{}^\alpha \\
&= \tfrac{1}{2} R^{mn}{}_{i\beta}\pi_{mn}{}^\beta - n_i J\overline{\mathcal{L}} - \nabla_\alpha \pi_i{}^\alpha \\
\bar{D}^\alpha &= D^\alpha + b^i{}_0 \pi_i{}^\alpha + \tfrac{1}{2} A^{ij}{}_0 \phi_{ij}{}^\alpha = b^i{}_0 \pi_i{}^\alpha + \tfrac{1}{2} A^{ij}{}_0 \pi_{ij}{}^\alpha .
\end{aligned}$$

In calculating $\overline{\mathcal{H}}_i$ we used the relation $\mathcal{H}_i = aJ(n_i R^{\bar{m}\bar{n}}{}_{\bar{m}\bar{n}} - 2h_{\bar{i}}{}^\alpha R^{\perp\bar{n}}{}_{\alpha\bar{n}})$, from which we find

$$\mathcal{H}_i + 2aJ R^{\perp\bar{n}}{}_{i\bar{n}} = aJn_i(R^{\bar{m}\bar{n}}{}_{\bar{m}\bar{n}} + 2R^{\perp\bar{n}}{}_{\perp\bar{n}}) \equiv -n_i J\overline{\mathcal{L}}.$$

It is interesting to note that the expressions for $\overline{\mathcal{H}}_{ij}$ and $\bar{D}^\alpha$ completely coincide with the form (5.43), as a consequence of the contribution of the determined multipliers. On the other hand, the form of $\overline{\mathcal{H}}_i$ implies

$$\begin{aligned}
\overline{\mathcal{H}}_\alpha &= \tfrac{1}{2}\pi_{ij}{}^\beta R^{ij}{}_{\alpha\beta} - b^k{}_\alpha \nabla_\beta \pi_k{}^\beta \\
\overline{\mathcal{H}}_\perp &= \tfrac{1}{2}\hat{\pi}_{ij}{}^{\bar{m}} R^{ij}{}_{\perp\bar{m}} - J\overline{\mathcal{L}} - n^k \nabla_\beta \pi_k{}^\beta
\end{aligned}$$

where we can observe the absence of the πT terms, compared to (5.43). Let us recall that in all expressions the multipliers $u^i{}_\alpha$ and $u^{ij}{}_\alpha$ are determined from the corresponding conditions. Therefore, we conclude that the term $\pi_i{}^\beta T^i{}_{0\beta}$ is absent from the Hamiltonian, since it is exactly the equation $T^i{}_{0\beta} = 0$ which is used to determine the multipliers. Using the values of the 'multipliers' $R^{mn}{}_{\perp\bar{r}} = (R^{\bar{m}\bar{n}}{}_{\perp\bar{r}}, R^{\perp\bar{n}}{}_{\perp\bar{r}})$ found previously, and replacing $R^{\perp\bar{n}}{}_{\perp\bar{n}} \to -R^{\bar{m}\bar{n}}{}_{\bar{m}\bar{r}}$ in $\overline{\mathcal{L}}$, we finally obtain

$$\overline{\mathcal{H}}_\perp = \tfrac{1}{2}\hat{\pi}_{\bar{m}\bar{n}}{}^{\bar{r}} R_{\perp\bar{r}}{}^{\bar{m}\bar{n}} - \hat{\pi}_{\perp\bar{n}}{}^{\bar{r}} R^{\bar{m}\bar{n}}{}_{\bar{m}\bar{r}} - aJ R^{\bar{m}\bar{n}}{}_{\bar{m}\bar{n}} - n^k \nabla_\beta \pi_k{}^\beta .$$

Comments. (1) Hamiltonian (5.49) differs from the result in example 5 as a consequence of the fact that $\mathcal{H}_\mathrm{c}$ is defined only up to primary constraints. What are the consequences of a different choice of $\mathcal{H}_\mathrm{c}$ on the structure of $\mathcal{H}_\mathrm{T}$? If we denote the canonical Hamiltonian from example 5 by $\mathcal{H}'_\mathrm{c}$, its relation to the expression (5.49) is given by

$$\mathcal{H}'_\mathrm{c} = \mathcal{H}_\mathrm{c} + t^{ij}{}_\alpha \phi_{ij}{}^\alpha$$

where t are known coefficients. Hence, $\mathcal{H}'_\mathrm{T}$ is obtained from $\mathcal{H}_\mathrm{T}$ by the replacement $u^{ij}{}_\alpha \to u'^{ij}{}_\alpha = u^{ij}{}_\alpha - t^{ij}{}_\alpha$. Consistency requirements lead to the same form of secondary constraints, and the same conditions on $u^i{}_\alpha$ and $u'^{ij}{}_\alpha$. Therefore, the final total Hamiltonian $\mathcal{H}'_\mathrm{T}$, containing the fixed multipliers, has the same form as $\mathcal{H}_\mathrm{T}$.

(2) Since EC theory without matter fields is equivalent to GR, we shall try to see whether we can use the 36 second-class constraints to eliminate the 36

variables $(A^{ij}{}_\alpha, \pi_{ij}{}^\alpha)$ and obtain the canonical formulation in terms of the tetrads and their momenta $(b^i{}_\mu, \pi_i{}^\mu)$. The analysis is simplified if we choose the time gauge, $b^0{}_\alpha \approx 0$, which effectively means:

– the transition $\perp \to 0$ and $\bar{k} \to a$,
– the elimination of the pair $(b^0{}_\alpha, \pi_0{}^\beta)$ with the help of constraints and
– the use of PBs for the remaining variables.

By imposing this condition, the constraint $\mathcal{H}_{a0}$ becomes effectively second class, giving rise (together with $b^0{}_\alpha \approx 0$) to a total number of $36 + 6$ second-class constraints, which can be used for the elimination of 42 variables $(A^{ij}{}_\alpha, \pi_{ij}{}^\alpha; b^0{}_\alpha, \pi_0{}^\beta)$. Using the $6 + 18$ equations $b^0{}_\alpha = 0$, $\pi_0{}^\alpha = 0$, $T^a{}_{\alpha\beta} = 0$ and $\pi_{ab}{}^\alpha = 0$, we can easily eliminate $b^0{}_\alpha$, $A^{ab}{}_\alpha$ and their momenta:

$$
\begin{aligned}
b^0{}_\alpha &= 0 \qquad \pi_0{}^\alpha = 0 \\
A^{ab}{}_\alpha &= \Delta^{ab}{}_\alpha \qquad \pi_{ab}{}^\alpha = 0.
\end{aligned}
$$

The structure of the remaining 18 equations, $\pi_a{}^\alpha = 0$ and $\phi_{a0}{}^\alpha = 0$, is such that the variables $(A^{a0}{}_\alpha, \pi_{a0}{}^\alpha)$ cannot be eliminated, as we wanted.

Our aim can be realized by introducing a suitable canonical transformation of variables. A simple way to define this transformation is by going over to the action

$$
I'_{\mathrm{HP}} = a \int \mathrm{d}^4x\, \tfrac{1}{2}\varepsilon^{\mu\nu\lambda\rho}_{mnkl}[-\partial_\mu(b^k{}_\lambda b^l{}_\rho)A^{mn}{}_\nu + b^k{}_\lambda b^l{}_\rho A^m{}_{s\mu}A^{sn}{}_\nu]
$$

which differs from the original one by a four-divergence. The elimination of ∂A from the action gives rise to a set of second-class constraints which can be solved to express $(A^{ij}{}_\alpha, \pi_{ij}{}^\alpha)$ in terms of the other variables. The new action is used in appendix E not only to analyse the tetrad formulation of the theory but also to introduce Ashtekar's complex variables.

(3) The appearance of the term $\partial_\alpha D^\alpha$ in the total Hamiltonian seems to be an accident. It ensures that the Hamiltonian does not depend on the derivatives of momenta. The form of this term changes if the Lagrangian is changed by a four-divergence. Its presence 'violates' the statement that H_{T} is equal to a linear combination of FC constraints. Does this term has any physical meaning? Let us assume that the tetrad fields have an asymptotic behaviour at spatial infinity corresponding to the Schwarzschild solution. Then, starting from the action I'_{HP} we find the result $\int \mathrm{d}^3x\, \mathcal{H}_{\mathrm{T}} \approx \int \mathrm{d}^3x\, \partial_\alpha D^\alpha = M$, suggesting a very interesting interpretation. If we look at the Hamiltonian H_{T} as the generator of time evolution which acts on dynamical variables via PBs, then the term $\int \mathrm{d}^3x\, \partial_\alpha D^\alpha$, transformed to a surface integral, can be ignored, being equivalent to a zero generator; the Hamiltonian effectively becomes a linear combination of FC constraints, in agreement with the general considerations. On the other hand, the presence of $\partial_\alpha D^\alpha$ in $\mathcal{H}_{\mathrm{T}}$ enables us to interpret the Hamiltonian as the energy of the system.

In a certain way, this result is accidental, since, until now, we have not considered the surface terms in a systematic manner. We shall see in chapter 6 that there exist general principles that determine the form of the surface terms and relate them to the energy and other conserved quantities in gravitation.

The teleparallel theory

The teleparallel approach to gravity is an interesting alternative to the Riemannian GR. Canonical analysis of teleparallel theory (3.61) slightly differs from the general description given in the previous section, due to the presence of Lagrange multipliers (Blagojević and Nikolić 2000, Blagojević and Vasilić 2000a).

Primary constraints. The basic Lagrangian dynamical variables of the teleparallel theory (3.61) are $(b^i{}_\mu, A^{ij}{}_\mu, \lambda_{ij}{}^{\mu\nu})$ and the corresponding momenta are denoted by $(\pi_i{}^\mu, \pi_{ij}{}^\mu, \pi^{ij}{}_{\mu\nu})$. Since the Lagrangian does not involve velocities $\dot{b}^k{}_0$ and $\dot{A}^{ij}{}_0$, we obtain the following primary constraints:

$$\phi_k{}^0 \equiv \pi_k{}^0 \approx 0 \qquad \phi_{ij}{}^0 \equiv \pi_{ij}{}^0 \approx 0. \tag{5.54}$$

Similarly, the absence of the time derivative of $\lambda_{ij}{}^{\mu\nu}$ implies

$$\phi^{ij}{}_{\mu\nu} \equiv \pi^{ij}{}_{\mu\nu} \approx 0. \tag{5.55}$$

The next set of constraints follows from the linearity of the curvature in $\dot{A}^{ij}{}_\alpha$:

$$\phi_{ij}{}^\alpha \equiv \pi_{ij}{}^\alpha - 4\lambda_{ij}{}^{0\alpha} \approx 0. \tag{5.56}$$

Now we turn our attention to the remaining momenta $\pi_i{}^\alpha$. The relations defining $\pi_i{}^\alpha$ can be written in the form

$$\hat{\pi}_i{}^{\bar{k}} = J\frac{\partial\bar{\mathcal{L}}_\mathrm{T}}{\partial T^i{}_{\perp\bar{k}}} = 4J\beta_i{}^{\perp\bar{k}}(T).$$

Using the fact that β is a linear function of T, we can make the expansion $\beta(T) = \beta(0) + \beta(1)$, where $\beta(0)$ does not depend on 'velocities' $T^i{}_{\perp\bar{k}}$ and $\beta(1)$ is linear in them, and rewrite this equation in the form

$$P_{i\bar{k}} \equiv \hat{\pi}_{i\bar{k}}/J - 4\beta_{i\perp\bar{k}}(0) = 4\beta_{i\perp\bar{k}}(1).$$

Here, the so-called 'generalized momenta' $P_{i\bar{k}}$ do not depend on the velocities, which appear only on the right-hand side of the equation. Explicit calculation leads to the result

$$\begin{aligned} P_{i\bar{k}} &\equiv \hat{\pi}_{i\bar{k}}/J - 4a[\tfrac{1}{2}BT_{\perp\bar{\imath}\bar{k}} + \tfrac{1}{2}Cn_iT^{\bar{m}}{}_{\bar{m}\bar{k}}] \\ &= 4a[AT_{i\perp\bar{k}} + \tfrac{1}{2}BT_{\bar{k}\perp\bar{\imath}} + \tfrac{1}{2}C\eta_{\bar{\imath}\bar{k}}T^{\bar{m}}{}_{\perp\bar{m}} + \tfrac{1}{2}(B+C)n_iT_{\perp\perp\bar{k}}]. \end{aligned}$$

This system of equations can be decomposed into irreducible parts with respect to the group of three-dimensional rotations in Σ_0:

$$
\begin{aligned}
P_{\perp\bar{k}} &\equiv \hat{\pi}_{\perp\bar{k}}/J - 2aCT^{\bar{m}}{}_{\bar{m}\bar{k}} = 2a(2A+B+C)T_{\perp\perp\bar{k}} \\
P^{A}_{\bar{\imath}\bar{k}} &\equiv \hat{\pi}^{A}_{\bar{\imath}\bar{k}}/J - 2aBT_{\perp\bar{\imath}\bar{k}} = 2a(2A-B)T^{A}_{\bar{\imath}\perp\bar{k}} \\
P^{T}_{\bar{\imath}\bar{k}} &\equiv \hat{\pi}^{T}_{\bar{\imath}\bar{k}}/J = 2a(2A+B)T^{T}_{\bar{\imath}\perp\bar{k}} \\
P^{\bar{m}}{}_{\bar{m}} &\equiv \hat{\pi}^{\bar{m}}{}_{\bar{m}}/J = 2a(2A+B+3C)T^{\bar{m}}{}_{\perp\bar{m}}
\end{aligned}
\qquad (5.57)
$$

where $X^{A}_{\bar{\imath}\bar{k}} = X_{[\bar{\imath}\bar{k}]}$, $X^{T}_{\bar{\imath}\bar{k}} = X_{(\bar{\imath}\bar{k})} - \eta_{\bar{\imath}\bar{k}}X^{\bar{n}}{}_{\bar{n}}/3$.

In general, if some of the coefficients on the right-hand sides of equations (5.57) vanish, these relations lead to additional primary constraints ϕ_A (if constraints). Instead of going into a general discussion of various possibilities, we shall analyse, later in this subsection, the important specific example of $\mathrm{GR}_{\parallel}$.

The Hamiltonian. After the primary constraints have been found, we proceed to construct the canonical Hamiltonian density. Following the general procedure developed in section 5.2, we find

$$
\mathcal{H}_{\mathrm{c}} = N\mathcal{H}_{\perp} + N^{\alpha}\mathcal{H}_{\alpha} - \tfrac{1}{2}A^{ij}{}_{0}\mathcal{H}_{ij} - \lambda_{ij}{}^{\alpha\beta}R^{ij}{}_{\alpha\beta} + \partial_{\alpha}D^{\alpha} \qquad (5.58a)
$$

where

$$
\begin{aligned}
\mathcal{H}_{ij} &= 2\pi_{[i}{}^{\beta}b_{j]\beta} + \nabla_{\alpha}\pi_{ij}{}^{\alpha} \\
\mathcal{H}_{\alpha} &= \pi_{k}{}^{\beta}T^{k}{}_{\alpha\beta} - b^{k}{}_{\alpha}\nabla_{\beta}\pi_{k}{}^{\beta} \\
\mathcal{H}_{\perp} &= \hat{\pi}_{i}{}^{\bar{k}}T^{i}{}_{\perp\bar{k}} - J\bar{\mathcal{L}}_{\mathrm{T}} - n^{k}\nabla_{\beta}\pi_{k}{}^{\beta} \\
D^{\alpha} &= b^{k}{}_{0}\pi_{k}{}^{\alpha} + \tfrac{1}{2}A^{ij}{}_{0}\pi_{ij}{}^{\alpha}.
\end{aligned}
\qquad (5.58b)
$$

The explicit form of $\mathcal{H}_{\perp}$ can be obtained by eliminating 'velocities' $T_{i\perp\bar{k}}$ with the help of relations (5.57) defining momenta $\pi_{i\bar{k}}$. Note the minor changes in these expressions compared to equations (5.43) and (5.44), which are caused by the presence of Lagrange multipliers. The canonical Hamiltonian is now linear in unphysical variables $A^{ij}{}_{\mu}, b^{i}{}_{\mu}$ *and* $\lambda_{ij}{}^{\alpha\beta}$.

The general Hamiltonian dynamics of the system is described by the total Hamiltonian,

$$
\mathcal{H}_{\mathrm{T}} = \mathcal{H}_{\mathrm{c}} + u^{i}{}_{0}\pi_{i}{}^{0} + \tfrac{1}{2}u^{ij}{}_{0}\pi_{ij}{}^{0} + \tfrac{1}{4}u_{ij}{}^{\mu\nu}\pi^{ij}{}_{\mu\nu} + u^{ij}{}_{\alpha}\phi_{ij}{}^{\alpha} + (u\cdot\phi) \qquad (5.59)
$$

where $(u\cdot\phi) = u^{A}\phi_{A}$ denotes the contribution of extra primary constraints, if they exist.

Secondary constraints. Having found the form of the sure primary constraints (5.54), (5.55) and (5.56), we now consider the requirements for their consistency.

The consistency conditions of the primary constraints (5.54) yield the standard secondary constraints:

$$\mathcal{H}_\perp \approx 0 \qquad \mathcal{H}_\alpha \approx 0 \qquad \mathcal{H}_{ij} \approx 0. \tag{5.60}$$

Similarly,

$$\dot{\pi}^{ij}{}_{\alpha\beta} \approx 0 \quad \Rightarrow \quad R^{ij}{}_{\alpha\beta} \approx 0 \tag{5.61a}$$

$$\dot{\pi}^{ij}{}_{0\beta} \approx 0 \quad \Rightarrow \quad u^{ij}{}_\beta \approx 0. \tag{5.61b}$$

Since the equation of motion for $A^{ij}{}_\beta$ implies $R^{ij}{}_{0\beta} \approx u^{ij}{}_\beta$, all components of the curvature tensor weakly vanish, as we could have expected.

The consistency condition for $\phi_{ij}{}^\alpha$ can be used to determine $u_{ij}{}^{0\alpha}$:

$$\begin{aligned} 4\bar{u}_{ij}{}^{0\alpha} \approx{}& N'\{\pi_{ij}{}^\alpha, \mathcal{H}'_\perp\} - N^\alpha(\hat{\pi}_{i\bar{j}} - \hat{\pi}_{j\bar{i}}) - A^{kl}{}_0(\eta_{ik}\pi_{lj}{}^\alpha - \eta_{jk}\pi_{li}{}^\alpha) \\ &- 4\nabla_\beta \lambda_{ij}{}^{\beta\alpha} + u'^A\{\pi_{ij}{}^\alpha, \phi'_A\} \end{aligned} \tag{5.62}$$

where a bar over u is used to denote the determined multiplier. Thus, the total Hamiltonian can be written in the form (5.59) with $u_{ij}{}^{0\alpha} \to \bar{u}_{ij}{}^{0\alpha}$ and $u^{ij}{}_\alpha = 0$. Since $\bar{u}_{ij}{}^{0\alpha}$ is linear in the multipliers ($N, N^\alpha, A^{ij}{}_0, \lambda_{ij}{}^{0\alpha}, u^A$), the total Hamiltonian takes the form

$$\begin{aligned} \mathcal{H}_\mathrm{T} &= \hat{\mathcal{H}}_\mathrm{T} + \partial_\alpha \bar{D}^\alpha \\ \hat{\mathcal{H}}_\mathrm{T} &\equiv \overline{\mathcal{H}}_\mathrm{c} + u^i{}_0\pi_i{}^0 + \tfrac{1}{2}u^{ij}{}_0\pi_{ij}{}^0 + \tfrac{1}{4}u_{ij}{}^{\alpha\beta}\pi^{ij}{}_{\alpha\beta} + u^A\bar{\phi}_A \\ \bar{D}^\alpha &= b^k{}_0\pi_k{}^\alpha + \tfrac{1}{2}A^{ij}{}_0\pi_{ij}{}^\alpha - \tfrac{1}{2}\lambda_{ij}{}^{\alpha\beta}\pi^{ij}{}_{0\beta} \end{aligned} \tag{5.63a}$$

where $\overline{\mathcal{H}}_\mathrm{c}$ is the modified canonical Hamiltonian,

$$\overline{\mathcal{H}}_\mathrm{c} = N\overline{\mathcal{H}}_\perp + N^\alpha\overline{\mathcal{H}}_\alpha - \tfrac{1}{2}A^{ij}{}_0\overline{\mathcal{H}}_{ij} - \lambda_{ij}{}^{\alpha\beta}\overline{\mathcal{H}}^{ij}{}_{\alpha\beta} \tag{5.63b}$$

the components of which are given by

$$\begin{aligned} \overline{\mathcal{H}}_\perp &= \mathcal{H}_\perp - \tfrac{1}{8}(\partial\mathcal{H}_\perp/\partial A^{ij}{}_\alpha)\pi^{ij}{}_{0\alpha} \\ \overline{\mathcal{H}}_\alpha &= \mathcal{H}_\alpha - \tfrac{1}{8}(\hat{\pi}_{i\bar{j}} - \hat{\pi}_{j\bar{i}})\pi^{ij}{}_{0\alpha} \\ \overline{\mathcal{H}}_{ij} &= \mathcal{H}_{ij} + \tfrac{1}{2}\pi_{[i}{}^s{}_{0\alpha}\pi_{j]s}{}^\alpha \\ \overline{\mathcal{H}}^{ij}{}_{\alpha\beta} &= R^{ij}{}_{\alpha\beta} - \tfrac{1}{2}\nabla_{[\alpha}\pi^{ij}{}_{0\beta]} \end{aligned} \tag{5.63c}$$

and

$$\bar{\phi}_A = \phi_A + \tfrac{1}{8}\pi^{ij}{}_{0\alpha}\{\pi_{ij}{}^\alpha, \phi_A\}. \tag{5.63d}$$

The total Hamiltonian $\mathcal{H}_\mathrm{T}$, in contrast to $\hat{\mathcal{H}}_\mathrm{T}$, does not contain the derivatives of momentum variables. The only components of $\mathcal{H}_\mathrm{T}$ that depend on the specific form of the Lagrangian are $\overline{\mathcal{H}}_\perp$ and $\bar{\phi}_A$.

Thus, the only sure secondary constraints are (5.60) and $R^{ij}{}_{\alpha\beta} \approx 0$. Before continuing our general analysis, we shall now illustrate the nature of extra constraints by considering the specific case of $\mathrm{GR}_\parallel$.

Canonical structure of $\mathrm{GR}_\parallel$. With a special choice of parameters (3.62), equations (5.57) contain two sets of relations: the first set represents extra primary constraints:

$$
\begin{aligned}
P_{\perp\bar{k}} &= \hat{\pi}_{\perp\bar{k}}/J + 2aT^{\bar{m}}{}_{\bar{m}\bar{k}} \approx 0 \\
P^{A}_{\bar{\imath}\bar{k}} &= \hat{\pi}^{A}_{\bar{\imath}\bar{k}}/J - aT_{\perp\bar{\imath}\bar{k}} \approx 0
\end{aligned}
\qquad (5.64a)
$$

while the second set gives non-singular equations,

$$
\begin{aligned}
P^{T}_{\bar{\imath}\bar{k}} &\equiv \hat{\pi}^{T}_{\bar{\imath}\bar{k}}/J = 2aT^{T}_{\bar{\imath}\perp\bar{k}} \\
P^{\bar{m}}{}_{\bar{m}} &\equiv \hat{\pi}^{\bar{m}}{}_{\bar{m}}/J = -4aT^{\bar{m}}{}_{\perp\bar{m}}
\end{aligned}
\qquad (5.64b)
$$

which can be solved for the velocities.

Further calculations are greatly simplified by observing that extra constraints $(5.64a)$ can be represented in a unified manner as

$$
\phi_{ik} = \pi_{i\bar{k}} - \pi_{k\bar{\imath}} + a\nabla_\alpha B^{0\alpha}_{ik} \qquad B^{0\alpha}_{ik} \equiv \varepsilon^{0\alpha\beta\gamma}_{ikmn} b^m_\beta b^n_\gamma . \qquad (5.65)
$$

This can be seen from the fact that relations $(5.64a)$ can be equivalently written as

$$
\pi_{i\bar{k}} - \pi_{k\bar{\imath}} \approx 2aJ(T_{\perp\bar{\imath}\bar{k}} - n_i T^{\bar{m}}{}_{\bar{m}\bar{k}} + n_k T^{\bar{m}}{}_{\bar{m}\bar{\imath}}) = 2a\nabla_\alpha H^{0\alpha}_{ik}
$$

and the identity $2H^{\mu\nu}_{ik} = -B^{\mu\nu}_{ik}$.

In order to find the explicit form of $\mathcal{H}_\perp$, we first rewrite the first two terms of $\mathcal{H}_\perp$ in the form

$$
\hat{\pi}^{i\bar{k}} T_{i\perp\bar{k}} - J\bar{\mathcal{L}}_{\mathrm{T}} = \tfrac{1}{2} J P^{i\bar{k}} T_{i\perp\bar{k}} - J\bar{\mathcal{L}}_{\mathrm{T}}(\bar{T})
$$

where $\bar{T}_{ikl} = T_{i\bar{k}\bar{l}}$. Then, taking constraints $(5.64a)$ into account we find that $T_{\perp\perp\bar{k}}$ and $T^{A}_{\bar{\imath}\perp\bar{k}}$ are absent from $\mathcal{H}_\perp$, whereupon the relations $(5.64b)$ can be used to eliminate the remaining 'velocities' $T^{T}_{\bar{\imath}\perp\bar{k}}$ and $T^{\bar{m}}{}_{\perp\bar{m}}$, leading directly to

$$
\mathcal{H}_\perp = \tfrac{1}{2} P^2_{\mathrm{T}} - J\bar{\mathcal{L}}_{\mathrm{T}}(\bar{T}) - n^k \nabla_\beta \pi_k{}^\beta \qquad (5.66a)
$$

where

$$
\begin{aligned}
P^2_{\mathrm{T}} &= \frac{1}{2aJ}\left(\pi_{(\bar{\imath}\bar{k})}\pi^{(\bar{\imath}\bar{k})} - \frac{1}{2}\pi^{\bar{m}}{}_{\bar{m}}\pi^{\bar{n}}{}_{\bar{n}}\right) \\
\bar{\mathcal{L}}_{\mathrm{T}}(\bar{T}) &= a(\tfrac{1}{4}T_{m\bar{n}\bar{k}}T^{m\bar{n}\bar{k}} + \tfrac{1}{2}T_{\bar{m}\bar{n}\bar{k}}T^{\bar{n}\bar{m}\bar{k}} - T^{\bar{m}}{}_{\bar{m}\bar{k}}T_{\bar{n}}{}^{\bar{n}\bar{k}}).
\end{aligned}
\qquad (5.66b)
$$

Calculating the modified constraint $\bar{\phi}_{ij}$, equation $(5.63d)$, we obtain

$$
\bar{\phi}_{ij} = \phi_{ij} - \tfrac{1}{4}a(\pi_i{}^s{}_{0\alpha}B^{0\alpha}_{sj} + \pi_j{}^s{}_{0\alpha}B^{0\alpha}_{is}). \qquad (5.67)
$$

The consistency conditions for the extra primary constraints ϕ_{ij} are found to be automatically fulfilled: $\dot{\phi}_{ij} = \{\phi_{ij}, \mathcal{H}_{\mathrm{T}}\} \approx 0$; moreover, we can show that $\bar{\phi}_{ij}$ are FC.

The existence of extra FC constraints $\bar{\phi}_{ij}$ may be interpreted as a consequence of the fact that the velocities contained in $T^{\perp\perp\bar{k}}$ and $T_A^{\bar{\imath}\perp\bar{k}}$ appear at most linear in the Lagrangian and, consequently, remain arbitrary functions of time. Although the torsion components $T^{\perp\perp\bar{k}}$ and $T_A^{\bar{\imath}\perp\bar{k}}$ are absent from the canonical Hamiltonian, they re-appear in the total Hamiltonian as the non-dynamical Hamiltonian multipliers ($u^{\perp\bar{k}}/N$ and $u^{\bar{\imath}\bar{k}}/N$). The presence of non-dynamical torsion components has a very clear interpretation via the gauge structure of the theory: it is related to the existence of additional FC constraints $\bar{\phi}_{ij}$.

On the consistency algorithm. In the previous analysis, we have found that the general teleparallel theory is characterized by the following set of sure constraints:

Primary: $\pi_i{}^0, \pi_{ij}{}^0, \pi^{ij}{}_{\alpha\beta}, \phi_{ij}{}^\alpha, \pi^{ij}{}_{0\beta}$;
Secondary: $\mathcal{H}_\perp, \mathcal{H}_\alpha, \mathcal{H}_{ij}, R^{ij}{}_{\alpha\beta}$.

Now it would be natural to continue the analysis by verifying the consistency of the secondary constraints.

Consider, first, the consistency condition of $R^{ij}{}_{\alpha\beta}$. Since $R^{ij}{}_{\alpha\beta}$ depends only on $A^{ij}{}_\alpha$, we can express $\mathrm{d}R^{ij}{}_{\alpha\beta}/\mathrm{d}t$ in terms of $\mathrm{d}A^{ij}{}_\alpha/\mathrm{d}t$, use the equation of motion for $A^{ij}{}_\alpha$, and rewrite the result in the form

$$\nabla_0 R^{ij}{}_{\alpha\beta} \approx \nabla_\alpha u^{ij}{}_\beta - \nabla_\beta u^{ij}{}_\alpha. \tag{5.68}$$

Hence, the consistency condition for $R^{ij}{}_{\alpha\beta}$ is identically satisfied. This relation has a very interesting geometric interpretation. Indeed, using the equation $R^{ij}{}_{0\beta} \approx u^{ij}{}_\beta$, we see that it represents a weak consequence of the second Bianchi identity.

General arguments in PGT show that the secondary constraints $\overline{\mathcal{H}}_{ij}, \overline{\mathcal{H}}_\alpha$ and $\overline{\mathcal{H}}_\perp$ are related to Poincaré gauge symmetry and, consequently, have to be FC. Hence, their consistency conditions must be automatically satisfied. However, an explicit proof of this property demands knowledge of the algebra of constraints, which is a non-trivial (although straightforward) calculational task by itself. Instead of using this argument, we shall postpone the completion of the consistency algorithm until we find the form of the Poincaré gauge generator (section 6.3), which will provide us with the necessary information concerning the nature of $\overline{\mathcal{H}}_{ij}, \overline{\mathcal{H}}_\alpha$ and $\overline{\mathcal{H}}_\perp$. Here, we only state the result:

The consistency conditions of all the secondary constraints are automatically satisfied.

The relevant dynamical classification of the sure constraints is presented in table 5.1.

The constraints $\phi_{ij}{}^\alpha$ and $\pi^{ij}{}_{0\beta}$ are second class since $\{\phi_{ij}{}^\alpha, \pi^{kl}{}_{0\beta}\} \not\approx 0$. They can be used as strong equalities to eliminate $\lambda_{ij}{}^{0\alpha}$ and $\pi^{ij}{}_{0\beta}$ from the theory

Table 5.1.

	First class	Second class
Primary	$\pi_i{}^0, \pi_{ij}{}^0, \pi^{ij}{}_{\alpha\beta}$	$\phi_{ij}{}^\alpha, \pi^{ij}{}_{0\beta}$
Secondary	$\bar{\mathcal{H}}_\perp, \bar{\mathcal{H}}_\alpha, \bar{\mathcal{H}}_{ij}, \bar{\mathcal{H}}^{ij}{}_{\alpha\beta}$	

and simplify the calculations. Note that all FC constraints appear multiplied by arbitrary multipliers in the total Hamiltonian (5.63).

The algebra of FC constraints plays an important role not only in the classical canonical structure of the theory, but also in studying its quantum properties. These important subjects deserve further investigation. We display here, for later convenience, the part of the Poisson bracket algebra of constraints involving $\bar{\mathcal{H}}^{ij}{}_{\alpha\beta}$:

$$\begin{aligned}
\{\bar{\mathcal{H}}^{ij}{}_{\alpha\beta}, \bar{\mathcal{H}}'_{kl}\} &= (\delta^i_k \bar{\mathcal{H}}_l{}^j{}_{\alpha\beta} + \delta^j_k \bar{\mathcal{H}}^i{}_{l\alpha\beta})\delta - (k \leftrightarrow l) \\
\{\bar{\mathcal{H}}^{ij}{}_{\alpha\beta}, \bar{\mathcal{H}}'_\gamma\} &= \{\bar{\mathcal{H}}^{ij}{}_{\alpha\beta}, \bar{\mathcal{H}}'_\perp\} = \{\bar{\mathcal{H}}^{ij}{}_{\alpha\beta}, \bar{\mathcal{H}}'^{kl}{}_{\gamma\delta}\} = 0.
\end{aligned} \tag{5.69}$$

The equations $\{\bar{\mathcal{H}}^{ij}{}_{\alpha\beta}, \bar{\mathcal{H}}'_\gamma\} = 0$ hold up to squares of constraints, which are always ignored in on-shell analysis.

The λ symmetry. We are now going to construct the canonical generator of the λ symmetry (3.65). If gauge transformations are given in terms of arbitrary parameters $\varepsilon(t)$ and their first time derivatives $\dot\varepsilon(t)$, as is the case with the symmetries of our Lagrangian (3.61), the gauge generators have the form (5.23*a*),

$$G = \varepsilon(t)G^{(0)} + \dot\varepsilon(t)G^{(1)}$$

where the phase-space functions $G^{(0)}$ and $G^{(1)}$ satisfy the conditions (5.23*b*). These conditions clearly define the procedure for constructing the generator: we start with an arbitrary PFC constraint $G^{(1)}$, evaluate its Poisson bracket with H_T, and define $G^{(0)}$ so that $\{G^{(0)}, H_T\} = C_{PFC}$.

The only PFC constraint that acts on the Lagrange multipliers $\lambda_{ij}{}^{\mu\nu}$ is $\pi_{ij}{}^{\alpha\beta}$. Starting with $\pi^{ij}{}_{\alpha\beta}$ as our $G^{(1)}$, we look for the generator in the form

$$G_A(\varepsilon) = \tfrac{1}{4}\dot\varepsilon_{ij}{}^{\alpha\beta}\pi^{ij}{}_{\alpha\beta} + \tfrac{1}{4}\varepsilon_{ij}{}^{\alpha\beta}S^{ij}{}_{\alpha\beta} \tag{5.70a}$$

where the integration symbol $\int d^3x$ is omitted for simplicity. The phase-space function $S^{ij}{}_{\alpha\beta}$ can be found from (5.23*b*). In the first step, we obtain the $G^{(0)}$ part of the generator up to PFC constraints:

$$S^{ij}{}_{\alpha\beta} = -4\bar{\mathcal{H}}^{ij}{}_{\alpha\beta} + C_{PFC}.$$

Then, using the algebra of constraints involving $\overline{\mathcal{H}}^{ij}{}_{\alpha\beta}$, given in (5.69), and the third condition in (5.23*b*), we find

$$S^{ij}{}_{\alpha\beta} = -4\overline{\mathcal{H}}^{ij}{}_{\alpha\beta} + 2A^{[i}{}_{k0}\pi^{j]k}{}_{\alpha\beta}. \tag{5.70b}$$

This completely defines the generator $G_A(\varepsilon)$ for which we were looking.

Using the rule $\delta_0 X = \int \mathrm{d}^3x' \{X, G'\}$, we apply the generator (5.70) to the fields, and find

$$\delta_0^A \lambda_{ij}{}^{0\alpha} = \nabla_\beta \varepsilon_{ij}{}^{\alpha\beta} \qquad \delta_0^A \lambda_{ij}{}^{\alpha\beta} = \nabla_0 \varepsilon_{ij}{}^{\alpha\beta} \tag{5.71}$$

as the only non-trivial field transformations. Surprisingly, this result does not agree with the form of the λ symmetry (3.65*b*), which contains an additional component, $\nabla_\gamma \varepsilon_{ij}{}^{\alpha\beta\gamma}$, in the expression for $\delta_0\lambda_{ij}{}^{\alpha\beta}$. Since there are no other PFC constraints that could produce the transformation of $\lambda_{ij}{}^{\alpha\beta}$, the canonical origin of the additional term seems somewhat puzzling.

The solution of the problem is, however, quite simple: if we consider *independent* gauge transformations only, this term is not needed, since it is not independent of what we already have in (5.71). To prove this statement, consider the following PFC constraint:

$$\Pi^{ij}{}_{\alpha\beta\gamma} = \nabla_\alpha \pi^{ij}{}_{\beta\gamma} + \nabla_\gamma \pi^{ij}{}_{\alpha\beta} + \nabla_\beta \pi^{ij}{}_{\gamma\alpha}.$$

This constraint is essentially a linear combination of the $\pi^{ij}{}_{\alpha\beta}$; hence, the related gauge generator will not be truly independent of the general expression (5.70). Furthermore, using the second Bianchi identity for $R^{ij}{}_{\alpha\beta}$, we find the relation

$$\nabla_\alpha \overline{\mathcal{H}}^{ij}{}_{\beta\gamma} + \nabla_\beta \overline{\mathcal{H}}^{ij}{}_{\gamma\alpha} + \nabla_\gamma \overline{\mathcal{H}}^{ij}{}_{\alpha\beta} = 0$$

which holds up to squares of constraints. As a consequence, $\Pi^{ij}{}_{\alpha\beta\gamma}$ has a vanishing PB with the total Hamiltonian, up to PFC constraints, and is, therefore, a correct gauge generator by itself. Hence, we can introduce a new gauge generator,

$$G_B(\varepsilon) = -\tfrac{1}{4}\, \varepsilon_{ij}{}^{\alpha\beta\gamma} \nabla_\alpha \pi^{ij}{}_{\beta\gamma} \tag{5.72}$$

where the parameter $\varepsilon_{ij}{}^{\alpha\beta\gamma}$ is totally antisymmetric with respect to its upper indices. The only non-trivial field transformation produced by this generator is

$$\delta_0^B \lambda_{ij}{}^{\alpha\beta} = \nabla_\gamma \varepsilon_{ij}{}^{\alpha\beta\gamma}$$

and it coincides with the missing term in equation (5.71). This concludes the proof that the *six parameters* $\varepsilon_{ij}{}^{\alpha\beta\gamma}$ in the λ transformations (3.65*b*) can be completely *discarded* if we are interested only in the *independent* λ transformations.

Although the generator G_B is not truly independent of G_A, it is convenient to define

$$G(\varepsilon) \equiv G_A(\varepsilon) + G_B(\varepsilon) \tag{5.73}$$

as an overcomplete gauge generator, since it automatically generates the covariant Lagrangian form of the λ symmetry.

We can prove that the action of the generator (5.73) on momenta is also correct, in the sense that it yields the result in agreement with the defining relations $\pi_A = \partial\mathcal{L}/\partial\dot{\varphi}^A$. In particular, the only non-trivial transformation law for the momenta, $\delta_0\pi_{ij}{}^\alpha = 4\nabla_\beta\,\varepsilon_{ij}{}^{\alpha\beta}$, agrees with (3.65*b*) through the conservation of the primary constraint $\phi_{ij}{}^\alpha \approx 0$.

This construction is based on using the first class constraints $\pi^{ij}{}_{\alpha\beta}, \overline{\mathcal{H}}^{ij}{}_{\alpha\beta}$, the part of the Poisson bracket algebra involving these constraints, and the second-class constraints $\phi_{ij}{}^\alpha$. All these constraints and their properties are independent of the values of parameters in the theory; hence, we can conclude that

> *$G(\varepsilon)$ is the correct generator of λ symmetry in the general teleparallel theory.*

Exercises

1. For each of the point-particle theories (a)–(f) given here, (i) derive the constraints and the Hamiltonian; and (ii) construct the gauge generators and check the gauge invariance of the equations of motion.

   (a) $L = \frac{1}{2}\mathrm{e}^{q_1}(\dot{q}_2)^2$

   (b) $L = \frac{1}{2}(q_1 - \dot{q}_2)^2 + \frac{1}{2}(q_2 - \dot{q}_3)^2$

   (c) $L = \frac{1}{4}\mathrm{e}^{q_1}(\dot{q}_2)^2 + \frac{1}{4}(\dot{q}_3 - q_2)^2 - \alpha\mathrm{e}^{q_4}(q_2)^2 \qquad (\alpha = 1 \text{ or } 0)$

   (d) $L = \frac{1}{2}(\dot{q}_3 - q_5 - e^{q_1}\dot{q}_2)^2 + \frac{1}{2}(\dot{q}_5 - q_4)^2$

   (e) $L = \dot{q}_1\dot{q}_3 + \frac{1}{2}q_2(q_3)^2$

   (f) $L = \dfrac{1}{2}\left(\dfrac{\dot{q}_2}{q_1}\right)^2 + (q_1)^2 q_2 \qquad q_1 > 0.$

2. Calculate the Dirac brackets for free electrodynamics in the gauge $\Omega_1 = A^3 \approx 0$ (choose an appropriate additional gauge condition).
3. (a) Find the constraints and the Hamiltonian for the non-Abelian gauge theory for which the action is given in appendix A.
   (b) Construct the gauge generators.
4. The theory of the Abelian antisymmetric field is defined by the action

$$I = \tfrac{1}{8}\int \mathrm{d}^4x\,(-\varepsilon^{\mu\nu\lambda\rho}B_{\mu\nu}F_{\lambda\rho} + A_\mu A^\mu)$$

   where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$.
   (a) Show that the following constraints are present in the theory:

$$\text{first class:}\quad \phi^{\alpha 0} \equiv \pi^{\alpha 0} \qquad \widetilde{\chi}^{0\beta} \equiv -\tfrac{1}{2}{}^*F^{0\beta} + \partial_\alpha\pi^{\alpha\beta}$$

$$\text{second class:}\quad \phi^\mu \equiv \pi^\mu + \tfrac{1}{2}{}^*B^{0\mu} \qquad \phi^{\alpha\beta} \equiv \pi^{\alpha\beta}$$
$$\widetilde{\chi}^0 \equiv \partial_\alpha \pi^\alpha + \tfrac{1}{4}A^0.$$

(b) Derive the form of gauge symmetries.
(c) Construct the Dirac brackets.

5. The theory of the non-Abelian antisymmetric field is defined by the action

$$I = \tfrac{1}{8}\int \mathrm{d}^4x\, (-\varepsilon^{\mu\nu\lambda\rho} B^a{}_{\mu\nu} F^a{}_{\lambda\rho} + A^a{}_\mu A^{a\mu})$$

where $F^a{}_{\mu\nu} = \partial_\mu A^a{}_\nu - \partial_\nu A^a{}_\mu + f_{bc}{}^a A^b{}_\mu A^c{}_\nu$, and $f_{bc}{}^a$ are structure constants of a non-Abelian Lie group G. Find the constraints and gauge symmetries of this theory.

6. The Born–Infeld electrodynamics is defined by the action

$$I = \int \mathrm{d}^4x \left[\sqrt{-\det(\eta_{\mu\nu} + F_{\mu\nu})} - 1\right].$$

Derive the Hamiltonian and the constraints of this theory.

7. The relativistic free particle is described by the action

$$I = -m \int \mathrm{d}\tau\, \sqrt{\eta_{\mu\nu}\dot{x}^\mu \dot{x}^\nu}.$$

(a) Show that the action is invariant under the time reparametrization.
(b) Find the Hamiltonian of the theory.
(c) Derive the Hamiltonian equations of motion in the gauge $\Omega \equiv x^0 - \tau \approx 0$, and construct the related Dirac brackets.

8. Suppose that the total Hamiltonian of a dynamical system has the form $H_\mathrm{T} = \int \mathrm{d}t\, (\mathcal{H}_0 + u^m G_m)$, where $G_m(q, p)$ are FC constraints satisfying the conditions

$$\{G_m, G_n\} = U_{mn}{}^r G_r \qquad \{G_m, \mathcal{H}_0\} = V_m{}^n G_n$$

and $U_{mn}{}^r$ and $V_m{}^n$ are functions on the phase space. Show that the action

$$I[q, p] = \int \mathrm{d}t\, (p_a \dot{q}_a - \mathcal{H}_0 - u^m G_m)$$

is invariant under the following gauge transformations:

$$\delta q_a = \varepsilon^m \{q_a, G_m\} \qquad \delta p_a = \varepsilon^m \{p_a, G_m\}$$
$$\delta u^m = \dot{\varepsilon}^m + \varepsilon^r u^s U_{sr}{}^m + \varepsilon^r V_r{}^m.$$

9. Construct the action $I[A, \pi]$ for free electrodynamics. Then find the action $I[A]$ obtained from $I[A, \pi]$ by eliminating π with the help of the equations of motion.

10. (a) Prove that the lapse and shift functions are linear in $b^k{}_0$.
 (b) Show that $\det(b^k{}_\mu)$ has the form $b = NJ$, where J is independent of $b^k{}_0$.
 (c) Check the orthogonality relations: $b^{\bar k}{}_\alpha h_{\bar k}{}^\beta = \delta^\beta_\alpha$, $b^{\bar k}{}_\alpha h_{\bar m}{}^\alpha = \delta^{\bar k}_{\bar m}$.
11. Prove the following identities:

$$\varepsilon^{\mu\nu\rho\sigma} b_{i\mu} b_{j\nu} b_{k\rho} b_{l\sigma} = b\varepsilon_{ijkl} \qquad b \equiv \det(b^i{}_\mu)$$
$$\varepsilon^{\mu\nu\rho\sigma}_{ijkl} b^l{}_\sigma = -2b(h_i{}^\mu h_{[j}{}^\nu h_{k]}{}^\rho + h_k{}^\mu h_{[i}{}^\nu h_{j]}{}^\rho + h_j{}^\mu h_{[k}{}^\nu h_{i]}{}^\rho)$$
$$\varepsilon^{\mu\nu\rho\sigma}_{ijkl} b^k{}_\rho b^l{}_\sigma = -4b h_{[i}{}^\mu h_{j]}{}^\nu .$$

12. Prove the following relations:

$$\varepsilon_{nijk} X_l{}^n + \varepsilon_{lnij} X_k{}^n + \varepsilon_{klni} X_j{}^n + \varepsilon_{jkln} X_i{}^n = 0 \qquad (X_{in} = -X_{ni})$$
$$\varepsilon^{0\alpha\beta\gamma}_{mnil} b_{j\gamma} b^l{}_\beta - (i \leftrightarrow j) = \varepsilon^{0\alpha\beta\gamma}_{mjil} b_{n\gamma} b^l{}_\beta - (m \leftrightarrow n)$$
$$\nabla_\mu \varepsilon_{ijkl} = 0.$$

13. (a) Verify that the constraint $\phi_{ij}{}^\alpha$ from equation (5.48) has the same form as in example 5.
 (b) Show that $\mathcal{H}_i$ from (5.49) has the form $\mathcal{H}_i = aJ(n_i R^{\bar m\bar n}{}_{\bar m\bar n} - 2h_{\bar i}{}^\alpha R^{\perp\bar n}{}_{\alpha\bar n})$, which is equivalent to the result of example 5, up to $\phi_{ij}{}^\alpha$.
 (c) Compare $\mathcal{H}_{ij}$ and D^α from (5.49) to the result of example 5.
 (d) Use the constraint $\phi_{ij}{}^\alpha$ to show that $\pi_{im}{}^\alpha \pi^m{}_j{}^\beta R^{ij}{}_{\alpha\beta}$ is proportional to $\mathcal{H}_\perp$.
14. Using $[\nabla_\alpha, \nabla_\beta] b^i{}_\gamma = R^i{}_{j\alpha\beta} b^j{}_\gamma$ and the condition $T^i{}_{\alpha\beta} = \nabla_\gamma T^i{}_{\alpha\beta} = 0$, prove the relations

$$\text{(i)} \quad R^i{}_{\alpha\beta\gamma} + R^i{}_{\gamma\alpha\beta} + R^i{}_{\beta\gamma\alpha} = 0 \qquad \text{(ii)} \quad R_{\alpha\beta\gamma\delta} = R_{\gamma\delta\alpha\beta}.$$

15. (a) In EC theory without matter fields derive the relation

$$\dot T^i{}_{\alpha\beta} = 2\nabla_{[\alpha} u^i{}_{\beta]} + 2u^{im}{}_{[\alpha} b_{m\beta]}.$$

 (b) By eliminating the multipliers $u^i{}_\alpha$ and $u^{ij}{}_\alpha$ with the help of the equations $T^i{}_{0\alpha} = 0$ and $R^{ij}{}_{0\alpha} = u^{ij}{}_\alpha - \nabla_\alpha A^{ij}{}_0$, prove the relation

$$\dot T^i{}_{\alpha\beta} + A^i{}_{m0} T^m{}_{\alpha\beta} = b^m{}_0 R^i{}_{m\alpha\beta} + b^m{}_\beta R^i{}_{m0\alpha} + b^m{}_\alpha R^i{}_{m\beta 0}$$

 which implies (5.52*b*).
 (c) Using the conditions $T^i{}_{\alpha\beta} = \dot T^i{}_{\alpha\beta} = 0$ prove that

$$\text{(i)} \quad R^i{}_{0\beta\gamma} + R^i{}_{\gamma 0\beta} + R^i{}_{\beta\gamma 0} = 0 \qquad \text{(ii)} \quad R_{\alpha\beta\gamma 0} = R_{\gamma 0\alpha\beta}.$$

16. Use the Bianchi identity $\varepsilon^{0\alpha\beta\gamma} \nabla_\alpha R^{ij}{}_{\beta\gamma} = 0$ and the equations which determine $u^i{}_\alpha$ and $u^{ij}{}_\alpha$, given in the previous exercise, to prove the relation (5.52*a*).

17. In EC theory without matter fields, derive the Bianchi identities

$$\nabla_0 R^{ij}{}_{\beta\gamma} + \nabla_\gamma R^{ij}{}_{0\beta} + \nabla_\beta R^{ij}{}_{\gamma 0} = 0$$

using the Hamiltonian equations of motion.
18. Use the consistency condition of the constraint $\phi_{ij}{}^{\alpha}$ in teleparallel theory to determine the multiplier $u_{ij}{}^{0\alpha}$.
19. Show that the Hamiltonian equations of motion for $A^{ij}{}_{\mu}$, determined by the total Hamiltonian (5.59), imply $R^{ij}{}_{0\beta} \approx u^{ij}{}_{\beta}$. Then verify the relation (5.68).
20. Derive the form of $\mathcal{H}_\perp$ and $\bar{\phi}_{ij}$ in $\mathrm{GR}_\parallel$.
21. Consider the simple Lagrangian $\widetilde{\mathcal{L}}_0 = \lambda_{ij}{}^{\mu\nu} R^{ij}{}_{\mu\nu}$.
    (a) Find the form of the total Hamiltonian.
    (b) Derive the algebra of the related Hamiltonian constraints $\widetilde{\mathcal{H}}_{ij}$ and $\widetilde{\mathcal{H}}^{ij}{}_{\alpha\beta}$.
22. Find the form of the PB algebra of constraints involving $\overline{\mathcal{H}}^{ij}{}_{\alpha\beta}$, in general teleparallel theory.

Chapter 6

Symmetries and conservation laws

Gauge symmetries in the theory of gravity can be naturally analysed within the Hamiltonian formalism for constrained dynamical systems. The existence of gauge symmetries is observed by the presence of arbitrary multipliers (or, equivalently, FC constraints) in the total Hamiltonian. The old question about the relation between the nature of constraints and the form of gauge generators has been resolved by Castellani, who developed an algorithm for constructing all the canonical gauge generators (Castellani 1982). The method demands knowledge of the Poisson bracket algebra of FC constraints, and gives the gauge generators acting on both physical and unphysical phase-space variables.

We shall begin this chapter by studying the structure of the Poincaré gauge generators in the general U_4 theory of gravity. For their construction it is essential to have complete knowledge of the algebra of FC constraints. An interesting supplement to these considerations is given in appendix F, where a reversed argument is presented: using the known structure of the gauge generators, we can obtain very precise information about the algebra of FC constraints.

A continuous symmetry of an action leads, via Noether's theorem, to a differentially conserved current. The conservation of the corresponding charge, which is an integral quantity, can be proved only under certain assumptions about the asymptotic behaviour of the basic dynamical variables, which are usually fulfilled in standard flat-space field theories. The situation in gravitational theories is more complex. A clear and consistent picture of the gravitational energy and other conserved charges emerged only after the role and importance of *boundary conditions* and their *symmetries* had been fully recognized (Arnowitt *et al* 1962, DeWitt 1967, Regge and Teitelboim 1974).

A field theory is defined by both the field equations and the boundary conditions. The concept of asymptotic or boundary symmetry is of fundamental importance for understanding the conservation laws in gravity. It is defined by the gauge transformations that leave a chosen set of boundary conditions invariant. Assuming that the symmetry in the asymptotic region is given by the global Poincaré symmetry, we shall use the Regge–Teitelboim approach to find the form of the *improved* canonical generators for two specific dynamical models: EC theory and teleparallel theory (sections 6.2 and 6.3). Since the generators act on dynamical variables via PBs, they should have well-defined functional derivatives. The fact that the global Poincaré generators do not satisfy this requirement will lead us to improve their form by adding certain *surface terms*, which turn out to define the energy, momentum and angular momentum of the gravitating system (Regge and Teitelboim 1974, Blagojević and Vasilić 1988, 2000b).

The global group of asymptotically flat four-dimensional gravity is given by the Poincaré group, the isometry group of M_4. In section 6.4, we shall show that asymptotic symmetries may have a more complex structure. Studying three-dimensional Chern–Simons gauge theory, we shall find that its boundary symmetry is given by the two-dimensional conformal group (Bañados 1994, 1999a, b), which is not the isometry group of any background geometry. Since three-dimensional gravity with negative cosmological constant can be represented as a Chern–Simons gauge theory (Witten 1988, appendix L), we have here an interesting example of the so called anti de Sitter/conformal field theory correspondence.

6.1 Gauge symmetries

A canonical description of gauge symmetries in PGT is given in terms of the gauge generators, which act on the basic dynamical variables via PBs. The construction of these generators clarifies the relationship between gauge symmetries and the algebra of FC constraints, and enables us to understand the role of asymptotic conditions in formulating the conservation laws better.

Constraint algebra

Explicit knowledge of the algebra of FC constraints in the U_4 theory is necessary for both the construction of the gauge generators and to investigate the consistency of the theory.

We have seen earlier that the consistency conditions of the sure primary constraints lead to the secondary constraints $\mathcal{H}_\perp \approx 0$, $\mathcal{H}_\alpha \approx 0$,$\mathcal{H}_{ij} \approx 0$, where the quantities $\mathcal{H}_\perp$, $\mathcal{H}_\alpha$ and $\mathcal{H}_{ij}$ are given as sums of contributions of the matter and gravitational fields, defined in equations (5.41) and (5.43).

When no extra constraints are present in the theory, we can show that the constraint algebra takes the form (Nikolić 1986, 1992):

$$\begin{aligned}
\{\mathcal{H}_{ij}, \mathcal{H}'_{kl}\} &= \tfrac{1}{2} f_{ij}{}^{mn}{}_{kl}\mathcal{H}_{mn}\delta \\
\{\mathcal{H}_{ij}, \mathcal{H}'_\alpha\} &= 0 \\
\{\mathcal{H}_\alpha, \mathcal{H}'_\beta\} &= (\mathcal{H}'_\alpha \partial_\beta + \mathcal{H}_\beta \partial_\alpha - \tfrac{1}{2} R^{ij}{}_{\alpha\beta}\mathcal{H}_{ij})\delta
\end{aligned} \tag{6.1a}$$

$$
\begin{aligned}
\{\mathcal{H}_{ij}, \mathcal{H}'_\perp\} &= 0 \\
\{\mathcal{H}_\alpha, \mathcal{H}'_\perp\} &= (\mathcal{H}_\perp \partial_\alpha - \tfrac{1}{2} R^{ij}{}_{\alpha\perp} \mathcal{H}_{ij})\delta \qquad (6.1b) \\
\{\mathcal{H}_\perp, \mathcal{H}'_\perp\} &= -({}^3g^{\alpha\beta}\mathcal{H}_\alpha + {}^3g'^{\alpha\beta}\mathcal{H}'_\alpha)\partial_\beta\delta.
\end{aligned}
$$

The first three relations represent PBs between kinematical constraints $\mathcal{H}_{ij}$ and $\mathcal{H}_\alpha$, the form of which does not depend of the choice of the action. Their derivation can be considerably simplified if $\mathcal{H}_{ij}$ and $\mathcal{H}_\alpha$ are written as

$$
\begin{aligned}
\mathcal{H}_{ij} &= \pi_A \Sigma_{ij} \varphi^A + \partial X_{ij} \\
\mathcal{H}_\alpha &= \pi_A \partial_\alpha \varphi^A + \tfrac{1}{2} A^{ij}{}_\alpha \mathcal{H}_{ij} + \partial X_\alpha
\end{aligned}
$$

where $\varphi^A = (\Psi, b^i{}_\alpha, A^{ij}{}_\alpha)$, $\pi_A = (\pi, \pi_i{}^\alpha, \pi_{ij}{}^\alpha)$, Σ_{ij} are the Lorentz generators in the related representation, and ∂X is a three-divergence.

The relations (6.1b) contain the dynamical part of the Hamiltonian $\mathcal{H}_\perp$; their calculation is rather involved, and will not be discussed here.

Using the decomposition (5.38b) for $\mathcal{H}_k$, equations (6.1) in the local Lorentz basis take the form

$$
\begin{aligned}
\{\mathcal{H}_{ij}, \mathcal{H}'_{kl}\} &= \tfrac{1}{2} f_{ij}{}^{mn}{}_{kl} \mathcal{H}_{mn}\delta \\
\{\mathcal{H}_{ij}, \mathcal{H}'_k\} &= -2\eta_{k[i}\mathcal{H}_{j]}\delta \\
\{\mathcal{H}_k, \mathcal{H}'_m\} &= -(n_{[m} R^{ij}{}_{\bar{k}]\perp} + \tfrac{1}{2} R^{ij}{}_{\bar{k}\bar{m}})\mathcal{H}_{ij}\delta + 2(n_{[m} T^i{}_{\bar{k}]\perp} + \tfrac{1}{2} T^i{}_{\bar{k}\bar{m}})\mathcal{H}_i\delta
\end{aligned} \qquad (6.2)
$$

featuring a visible analogy with the standard Poincaré algebra.

These considerations refer to the case when no extra constraints are present in the theory. When extra constraints exist, the whole analysis becomes much more involved, but the results essentially coincide with those in (6.1) (appendix F):

(a) the dynamical Hamiltonian $\mathcal{H}_\perp$ becomes a redefined expression $\overline{\mathcal{H}}_\perp$, which includes the contributions of all primary second-class constraints; and
(b) the algebra may contain terms of the type C_{PFC}.

Hence, the consistency conditions of the secondary constraints are automatically satisfied.

The specific properties of the constraint algebra in EC theory have been discussed by Henneaux (1983) and Nikolić (1995).

Gauge generators

The generators of the Poincaré gauge symmetry take the form

$$
G = \dot{\varepsilon}(t)G^{(1)} + \varepsilon(t)G^{(0)}
$$

where $G^{(0)}, G^{(1)}$ are phase-space functions satisfying conditions (5.23b). It is clear from these conditions that the construction of the gauge generators is based on the algebra of FC constraints. Since the Poincaré gauge symmetry is always present, independently of the specific form of the action, we naturally expect that all the essential features of the Poincaré gauge generators could be obtained by considering a simple case of the theory, in which no extra constraints are present. After that, the obtained result will be easily generalized (Blagojević *et al* 1988).

Construction of the generator. When no extra constraints exist, the primary constraints $\pi_k{}^0$ and $\pi_{ij}{}^0$ are FC. Starting with $G_k^{(1)} = -\pi_k{}^0$ and $G_{ij}^{(1)} = -\pi_{ij}{}^0$, conditions (5.23*b*) yield the following expression for the Poincaré gauge generator:

$$G = -\int \mathrm{d}^3x\,[\dot{\xi}^k\pi_k{}^0 + \xi^k(\mathcal{H}_k + \phi_k) + \tfrac{1}{2}\dot{\varepsilon}^{ij}\pi_{ij}{}^0 + \tfrac{1}{2}\varepsilon^{ij}(-\mathcal{H}_{ij} + \phi_{ij})] \tag{6.3a}$$

where ϕ_k and ϕ_{ij} are PFC constraints, which are to be determined from the relations

$$\begin{aligned}\{\mathcal{H}_k + \phi_k, H_\mathrm{T}\} &= C_\mathrm{PFC}\\ \{-\mathcal{H}_{ij} + \phi_{ij}, H_\mathrm{T}\} &= C_\mathrm{PFC}.\end{aligned}$$

Using the constraint algebra in the form (6.2), we obtain the following expressions for ϕ_k and ϕ_{ij}:

$$\begin{aligned}\phi_k &= [A^i{}_{k0} - b^m{}_0(T^i{}_{\bar{m}\bar{k}} + 2n_{[k}T^i{}_{\bar{m}]\perp})]\pi_i{}^0\\ &\quad + \tfrac{1}{2}b^m{}_0(R^{ij}{}_{\bar{k}\bar{m}} + 2n_{[m}R^{ij}{}_{\bar{k}]\perp})\pi_{ij}{}^0\\ \phi_{ij} &= 2b_{[i0}\pi_{j]}{}^0 + 2A^s{}_{[i0}\pi_{sj]}{}^0.\end{aligned} \tag{6.3b}$$

This completely defines the Poincaré gauge generator.

In order to check whether it generates the correct gauge transformations, we shall introduce a more convenient set of parameters:

$$\xi^k = \xi^\mu b^k{}_\mu \qquad \varepsilon^{ij} = \omega^{ij} + \xi^\nu A^{ij}{}_\nu$$

whereupon the generator G becomes (we omit $\int \mathrm{d}^3x$ for simplicity):

$$\begin{aligned}G &= G(\omega) + G(\xi)\\ G(\omega) &= -\tfrac{1}{2}\dot{\omega}^{ij}\pi_{ij}{}^0 - \tfrac{1}{2}\omega^{ij}S_{ij}\\ G(\xi) &= -\dot{\xi}^\mu(b^k{}_\mu\pi_k{}^0 + \tfrac{1}{2}A^{ij}{}_\mu\pi_{ij}{}^0) - \xi^\mu\mathcal{P}_\mu\end{aligned} \tag{6.4a}$$

where

$$\begin{aligned}\mathcal{P}_\mu &= b^k{}_\mu\mathcal{H}_k - \tfrac{1}{2}A^{ij}{}_\mu\mathcal{H}_{ij} + b^k{}_{0,\mu}\pi_k{}^0 + \tfrac{1}{2}A^{ij}{}_{0,\mu}\pi_{ij}{}^0\\ S_{ij} &= -\mathcal{H}_{ij} + 2b_{[i0}\pi_{j]}{}^0 + 2A^s{}_{[i0}\pi_{sj]}{}^0.\end{aligned} \tag{6.4b}$$

Note that $\mathcal{P}_0$ is equal to the total Hamiltonian (up to a three-divergence),

$$\mathcal{P}_0 = \widehat{\mathcal{H}}_\mathrm{T} \equiv \mathcal{H}_\mathrm{T} - \partial_\alpha D^\alpha$$

since $\dot{b}^k{}_0$ and $\dot{A}^{ij}{}_0$ are nothing but the arbitrary multipliers $u^k{}_0$ and $u^{ij}{}_0$, respectively: $\dot{b}^k{}_0 = \{b^k{}_0, H_\mathrm{T}\} = u^k{}_0$, $\dot{A}^{ij}{}_0 = \{A^{ij}{}_0, H_\mathrm{T}\} = u^{ij}{}_0$.

Action on the fields. We are now going to show that the action of the generators (6.4) on the fields $\Psi, b^k{}_\mu$ and $A^{ij}{}_\mu$ produces the standard Poincaré gauge transformations:

$$\begin{aligned}
\delta_0\Psi &= \tfrac{1}{2}\omega^{ij}\Sigma_{ij}\Psi - \xi^\nu\partial_\nu\Psi \\
\delta_0 b^k{}_\mu &= \omega^k{}_s b^s{}_\mu - \xi^\lambda{}_{,\mu} b^k{}_\lambda - \xi^\lambda\partial_\lambda b^k{}_\mu \\
\delta_0 A^{ij}{}_\mu &= \omega^i{}_s A^{sj}{}_\mu + \omega^j{}_s A^{is}{}_\mu - \omega^{ij}{}_{,\mu} - \xi^\lambda{}_{,\mu} A^{ij}{}_\lambda - \xi^\lambda\partial_\lambda A^{ij}{}_\mu
\end{aligned} \tag{6.5}$$

where $\delta_0 X \equiv \{X, G\}$. To do that, we first rewrite $\mathcal{P}_\mu$ and S_{ij} in the form

$$\begin{aligned}
\mathcal{P}_0 &= \widehat{\mathcal{H}}_\mathrm{T} \\
\mathcal{P}_\alpha &= \pi_i{}^\mu\partial_\alpha b^i{}_\mu + \tfrac{1}{2}\pi_{ij}{}^\mu\partial_\alpha A^{ij}{}_\mu + \pi\partial_\alpha\Psi - \partial_\beta(\pi_i{}^\beta b^i{}_\alpha + \tfrac{1}{2}\pi_{ij}{}^\beta A^{ij}{}_\alpha) \\
S_{ij} &= -2\pi_{[i}{}^\mu b_{j]\mu} - 2\pi_{s[i}{}^\mu A^s{}_{j]\mu} - \pi\Sigma_{ij}\Psi - \partial_\alpha\pi_{ij}{}^\alpha .
\end{aligned} \tag{6.6}$$

Then, it becomes straightforward to verify the ω^{ij} and ξ^α transformations. To derive ξ^0 transformations, we shall use the fact that $\mathcal{H}_\mathrm{T}$ does not depend on the derivatives of the momentum variables on the constraint surface, i.e. $\partial\mathcal{H}_\mathrm{T}/\partial\pi_{,\alpha} \approx 0$ (this is correct for Lagrangians that are, at most, quadratic in velocities). Let us first consider the transformations of $b^k{}_\mu$:

$$\begin{aligned}
\delta_0(\xi^0) b^k{}_\mu &= -\int \mathrm{d}^3x' \, [\dot{\xi}'^0\{b^k{}_\mu, b'^s{}_0\pi'_s{}^0\} + \xi'^0\{b^k{}_\mu, \mathcal{H}'_\mathrm{T} - D'^\alpha{}_{,\alpha}\}] \\
&= -\dot{\xi}^0 b^k{}_0\delta^0_\mu - \xi^0\Lambda^k{}_\mu - \xi^0{}_{,\alpha} b^k{}_0\delta^\alpha_\mu
\end{aligned}$$

where $\Lambda^k{}_\mu$ is defined by the relation $\{b^k{}_\mu, \mathcal{H}'_\mathrm{T}\} \approx \Lambda^k{}_\mu\delta$ ($\mathcal{H}_\mathrm{T}$ does not depend on $\partial\pi^A$, hence there is no $\partial\delta$ term on the right-hand side). Consequently,

$$\delta_0(\xi^0) b^k{}_\mu = -\xi^0{}_{,\mu} b^k{}_0 - \xi^0\{b^k{}_\mu, \mathcal{H}_\mathrm{T}\} \approx -\xi^0{}_{,\mu} b^k{}_0 - \xi^0\dot{b}^k{}_\mu$$

in accordance with (6.5).

The result is valid only on-shell. Note that the only properties of the total Hamiltonian used in the derivation are the following ones:

(a) $\mathcal{H}_\mathrm{T}$ does not depend on $\partial\pi_A$; and
(b) it governs the time evolution of dynamical variables: $\dot{\varphi}^A = \{\varphi^A, H_\mathrm{T}\}$.

In a similar way we can check the transformation rules for the fields $A^{ij}{}_\mu$ and Ψ. Finally, we should also check whether the generator (6.4) produces the correct transformations of momenta; that will be done soon.

We note that the field transformations (6.5) are symmetry transformations not only in the simple case characterized by the absence of any extra constraint, but also in the general case when extra constraints exist, i.e. for an arbitrary choice of parameters in the action. This fact motivates us to assume that the gauge generator (6.4), in which the term P_0 is *replaced* by the new $\widehat{\mathcal{H}}_{\mathrm{T}}$, is the correct gauge generator also in the general case.

The general case. Consider, now, the general theory in which extra constraints are allowed to exist. The Poincaré gauge generator is assumed to have the form

$$G = -\tfrac{1}{2}\dot{\omega}^{ij} G_{ij}^{(1)} - \tfrac{1}{2}\omega^{ij} G_{ij}^{(0)} - \dot{\xi}^{\mu} G_{\mu}^{(1)} - \xi^{\mu} G_{\mu}^{(0)} \tag{6.7a}$$

where

$$\begin{aligned} G_{ij}^{(1)} &= \pi_{ij}{}^0 \qquad G_{ij}^{(0)} = S_{ij} \\ G_{\mu}^{(1)} &= b^k{}_{\mu}\pi_k{}^0 + \tfrac{1}{2}A^{ij}{}_{\mu}\pi_{ij}{}^0 \qquad G_0^{(0)} = \widehat{\mathcal{H}}_{\mathrm{T}} \qquad G_{\alpha}^{(0)} = \mathcal{P}_{\alpha}. \end{aligned} \tag{6.7b}$$

The component $G_0^{(0)} = \widehat{\mathcal{H}}_{\mathrm{T}}$ now differs from the previous case by the presence of $(u \cdot \phi)$ terms; the part of $(u \cdot \phi)$ describing second-class constraints is included in the redefined Hamiltonian $\overline{\mathcal{H}}_{\perp}$.

It is clear that the ω^{ij} and ξ^{α} transformations *on fields* are again of the same form as in (6.5). To discuss the ξ^0 transformations, it is essential to observe that even in the general case $\mathcal{H}_{\mathrm{T}}$ does not depend on the derivatives of momenta on shell. The rest of the derivation leads to the same result as before.

To complete the proof, we have to show that the generator G produces the correct symmetry transformations of *momenta*. These transformations are determined by the defining relation $\pi_A = \partial\mathcal{L}/\partial\dot{\varphi}^A$, and the known transformation laws for φ^A and $\mathcal{L}$. We can show that G also acts correctly on momenta. Therefore:

> *the expression (6.7) is the correct Poincaré gauge generator for any choice of parameters in the action.*

This result enables us to study one of the most important problems of the classical theory of gravity—the definition of the gravitational energy and other conserved quantities.

6.2 Conservation laws—EC theory

Assuming that the asymptotic symmetry of U_4 theory is the global Poincaré symmetry, we shall now discuss the form of the related generators. The general idea is illustrated on the specific case of EC theory. A careful analysis of boundary conditions leads to the appearance of certain *surface terms* in the expressions for the generators. The improved generators enable a correct treatment of the conservation laws of the energy, momentum and angular momentum (Blagojević and Vasilić 1988, 2000b).

Asymptotic structure of spacetime

The asymptotic Poincaré symmetry. The global Poincaré transformations of fields can be obtained from the corresponding gauge transformations by the following replacements of parameters:

$$\begin{aligned} \omega^{ij}(x) &\to \omega^{ij} \\ \xi^\mu(x) &\to \omega^\mu{}_\nu x^\nu + \varepsilon^\nu \equiv \xi^\mu \end{aligned} \tag{6.8}$$

where ω^{ij} and ε^ν are constants, and $\omega^\mu{}_\nu = \delta^\mu_i \omega^{ij}\eta_{j\nu}$. The indices of quantities related to the asymptotic spacetime are treated as in M_4: they are raised and lowered by the Minkowski metric η_{ij}, while the transition between the local Lorentz and coordinate basis is realized with the help of the Kronecker symbols δ^μ_i and δ^j_ν. These replacements have been chosen so as to obtain the standard global Poincaré transformations of fields:

$$\begin{aligned} \delta_0\Psi &= \tfrac{1}{2}\omega^{ij}\Sigma_{ij}\Psi - \xi^\nu\partial_\nu\Psi \\ \delta_0 b^i{}_\mu &= \omega^i{}_s b^s{}_\mu - \omega^\nu{}_\mu b^i{}_\nu - \xi^\nu\partial_\nu b^i{}_\mu \\ \delta_0 A^{ij}{}_\mu &= \omega^i{}_s A^{sj}{}_\mu + \omega^j{}_s A^{is}{}_\mu - \omega^\nu{}_\mu A^{ij}{}_\nu - \xi^\nu\partial_\nu A^{ij}{}_\mu. \end{aligned} \tag{6.9}$$

The generator of these transformations can be obtained from the gauge generator (6.7) in the same manner, leading to

$$G = \tfrac{1}{2}\omega^{ij}M_{ij} - \varepsilon^\nu P_\nu \tag{6.10a}$$

where

$$\begin{aligned} P_\mu = \int \mathrm{d}^3x\, \mathcal{P}_\mu \qquad M_{\mu\nu} &= \int \mathrm{d}^3x\, \mathcal{M}_{\mu\nu} \\ \mathcal{M}_{\alpha\beta} &= x_\alpha\mathcal{P}_\beta - x_\beta\mathcal{P}_\alpha - S_{\alpha\beta} \\ \mathcal{M}_{0\beta} &= x_0\mathcal{P}_\beta - x_\beta\mathcal{P}_0 - S_{0\beta} + b^k{}_\beta\pi_k{}^0 + \tfrac{1}{2}A^{ij}{}_\beta\pi_{ij}{}^0. \end{aligned} \tag{6.10b}$$

As the symmetry generators act on basic dynamical variables via PBs, they are required to have *well-defined functional derivatives*. When the parameters decrease sufficiently fast at spatial infinity, all partial integrations in G are characterized by vanishing surface terms, and the differentiability of G does not present any problem. The parameters of the global Poincaré symmetry are not of that type, so that the surface terms must be treated more carefully. We shall therefore try to improve the form of the generators (6.10) so as to obtain expressions with well-defined functional derivatives. The first step in that direction is to define the phase space in which the generators (6.10) act precisely, by an appropriate choice of boundary conditions.

The boundary conditions. The choice of boundary conditions becomes clearer if we express the asymptotic structure of spacetime in certain geometric terms. Here, we shall be concerned with *finite* gravitational sources, characterized by matter fields that decrease sufficiently rapidly at large distances, so that their contribution to surface integrals vanishes. In that case, we can assume that the spacetime is *asymptotically flat*, i.e. that the following two conditions are satisfied:

(a) There exists a coordinate system in which the metric tensor becomes Minkowskian at large distances: $g_{\mu\nu} = \eta_{\mu\nu} + \mathcal{O}_1$, where $\mathcal{O}_n = \mathcal{O}(r^{-n})$ denotes a term which decreases like r^{-n} or faster for large r, i.e. $r^n\mathcal{O}_n$ remains finite when $r \to \infty$, and $r^2 = (x^1)^2 + (x^2)^2 + (x^3)^2$.

(b) The Lorentz field strength satisfies the condition of the absolute parallelism for large r: $R^{ij}{}_{\mu\nu} = \mathcal{O}_{2+\alpha}$ $(\alpha > 0)$.

The first condition is consistent with the asymptotic global Poincaré symmetry. The second condition can be easily satisfied by demanding

(b′) $A^{ij}{}_{\mu} = \mathcal{O}_{1+\alpha}$.

In EC theory, the connection behaves as the derivative of the metric, so that $A = \mathcal{O}_2$. The same law holds in general U_4 theory when the field A is massive, while massless A can have a slower decrease. Here, we shall limit ourselves, for simplicity, to the EC theory, i.e. we shall assume that the boundary conditions of the *gravitational field* have the form

$$b^k{}_{\mu} = \delta^k_{\mu} + \mathcal{O}_1 \qquad A^{ij}{}_{\mu} = \mathcal{O}_2. \tag{6.11a}$$

The vacuum values of the fields, $b^i{}_{\mu} = \delta^i_{\mu}$ and $A^{ij}{}_{\mu} = 0$, are invariant under the action of the global Poincaré group. Demanding that the conditions (6.11a) remain unchanged under the global Poincaré transformations, we obtain the following conditions on the field derivatives:

$$\begin{aligned} b^k{}_{\mu,\nu} &= \mathcal{O}_2 \qquad & b^k{}_{\mu,\nu\lambda} &= \mathcal{O}_3 \\ A^{ij}{}_{\mu,\nu} &= \mathcal{O}_3 \qquad & A^{ij}{}_{\mu,\nu\lambda} &= \mathcal{O}_4. \end{aligned} \tag{6.11b}$$

These relations impose serious restrictions on the gravitational field in the asymptotic region, and define an *isolated* gravitational system (characterized, in particular, by the absence of gravitational waves).

The requirements (6.11) are *minimal* in the sense that some additional arguments may lead to better asymptotics, i.e. to a faster or more precisely defined decrease in the fields and their derivatives. In addition to the boundary conditions (6.11), we shall adopt the principle that all the expressions that vanish on-shell have an *arbitrarily fast* asymptotic decrease, as no solutions of the equations of motion are thereby lost. In particular, all the constraints of the theory are assumed to decrease arbitrarily fast.

In accordance with this principle, the asymptotic behaviour of the *momenta* is determined by requiring

$$\pi_A - \frac{\partial \widetilde{\mathcal{L}}}{\partial \dot{\varphi}^A} = \widehat{\mathcal{O}}$$

where $\widehat{\mathcal{O}}$ denotes a term with an arbitrarily fast asymptotic decrease. By using the definitions of the gravitational momentum variables, equations (5.28) and (5.48), we find that

$$\begin{aligned}\pi_k{}^0, \pi_{ij}{}^0 &= \widehat{\mathcal{O}}\\ \pi_i{}^\alpha &= \widehat{\mathcal{O}}\\ \pi_{ij}{}^\alpha &= -4aJn_{[i}h_{j]}{}^\alpha + \widehat{\mathcal{O}}.\end{aligned} \tag{6.12}$$

Similar arguments lead to the consistent determination of the asymptotic behaviour of the Hamiltonian multipliers. From $\dot{b}^k{}_0 = u^k{}_0, \dot{A}^{ij}{}_0 = u^{ij}{}_0$ we find that

$$u^k{}_0 = \mathcal{O}_2 \qquad u^{ij}{}_0 = \mathcal{O}_3. \tag{6.13a}$$

The other multipliers are also seen to be related to the velocities, which results in the following asymptotic behaviour:

$$u_{\perp\bar{m}}, u_{\bar{k}\bar{m}} = \mathcal{O}_2 \qquad u_{i\bar{k}\bar{m}} = \mathcal{O}_3. \tag{6.13b}$$

Improving the Poincaré generators

The canonical generators act on dynamical variables via the PB operation, which is defined in terms of functional derivatives. A functional

$$F[\varphi, \pi] = \int \mathrm{d}^3x\, f(\varphi(x), \partial_\mu\varphi(x), \pi(x), \partial_\nu\pi(x))$$

has well-defined functional derivatives if its variation can be written in the form

$$\delta F = \int \mathrm{d}^3x\, [A(x)\delta\varphi(x) + B(x)\delta\pi(x)] \tag{6.14}$$

where the terms $\delta\varphi_{,\mu}$ and $\delta\pi_{,\mu}$ are *absent*.

The global Poincaré generators do not satisfy this requirement, as we shall see. This will lead us to redefine them by adding certain surface terms, which turn out to be the energy, momentum and angular momentum of the physical system.

Spatial translation. Let us demonstrate how this procedure works in the case of global spatial translation. The variation of P_α yields

$$\begin{aligned}\delta P_\alpha &= \int \mathrm{d}^3x\, \delta\mathcal{P}_\alpha\\ \delta\mathcal{P}_\alpha &= \pi_i{}^\mu\delta b^i{}_{\mu,\alpha} + \tfrac{1}{2}\pi_{ij}{}^\mu\delta A^{ij}{}_{\mu,\alpha} - \delta(\pi_i{}^\beta b^i{}_\alpha + \tfrac{1}{2}\pi_{ij}{}^\beta A^{ij}{}_\alpha)_{,\beta} + R.\end{aligned}$$

Here, we have explicitly displayed those terms that contain the unwanted variations $\delta\varphi_{,\mu}$ and $\delta\pi_{,\mu}$, while the remaining terms of the correct, regular form (6.14) are denoted by R. A simple formula

$$\pi_i{}^\mu \delta b^i{}_{\mu,\alpha} = (\pi_i{}^\mu \delta b^i{}_\mu)_{,\alpha} + R$$

allows us to conclude that

$$\pi_i{}^\mu \delta b^i{}_{\mu,\alpha} = \partial\widehat{\mathcal{O}} + R$$

according to the boundary conditions (6.12). Continuing with the same reasoning, we obtain

$$\delta\mathcal{P}_\alpha = (\tfrac{1}{2}\pi_{ij}{}^\beta \delta A^{ij}{}_\beta)_{,\alpha} - \delta(\tfrac{1}{2}\pi_{ij}{}^\beta A^{ij}{}_\alpha)_{,\beta} + R + \partial\widehat{\mathcal{O}}.$$

After taking into account the relation $\delta\pi_{ij}{}^\beta A^{ij}{}_\beta = \mathcal{O}_3$, it follows that

$$\delta\mathcal{P}_\alpha = -\delta(\pi_{ij}{}^\beta A^{ij}{}_{[\alpha}\delta_{\beta]}{}^\gamma)_{,\gamma} + R + \partial\mathcal{O}_3.$$

As a consequence, the variation of P_α can be written in the simple form:

$$\begin{aligned} \delta P_\alpha &= -\delta E_\alpha + R \\ E_\alpha &\equiv \oint \mathrm{d}S_\gamma\, (\pi_{ij}{}^\beta A^{ij}{}_{[\alpha}\delta_{\beta]}{}^\gamma) \end{aligned} \tag{6.15a}$$

where E_α is defined as a surface integral over the boundary of the three-dimensional space. This allows us to redefine the generator P_α,

$$P_\alpha \to \widetilde{P}_\alpha \equiv P_\alpha + E_\alpha \tag{6.15b}$$

so that the new, improved expression $\tilde{P}_\alpha$ has *well-defined functional derivatives*. We can verify that the assumed boundary conditions ensure *finiteness* of E_α.

While the old generator P_α vanishes on-shell (as an integral of a linear combination of constraints), $\tilde{P}_\alpha$ does not—its on-shell value is E_α. Since $\tilde{P}_\alpha$ is the generator of the asymptotic spatial translations, we expect E_α to be the value of the related conserved charge—linear momentum; this will be proved in the next subsection.

Time translation. In a similar, way we can improve the form of the time translation generator. Let us start with

$$\begin{aligned} \delta P_0 &= \int \mathrm{d}^3x\, \delta\widehat{\mathcal{H}}_\mathrm{T} \\ \delta\widehat{\mathcal{H}}_\mathrm{T} &= \delta\widehat{\mathcal{H}}_\mathrm{c} + R = N\delta\overline{\mathcal{H}}_\perp + N^\alpha \delta\mathcal{H}_\alpha - \tfrac{1}{2}A^{ij}{}_0\delta\mathcal{H}_{ij} + R. \end{aligned}$$

Using the formulae

$$\begin{aligned}
\delta\mathcal{H}_{ij} &= (\delta\pi_{ij}{}^{\alpha})_{,\alpha} + R \\
\delta\mathcal{H}_{\alpha} &= (\pi_{ij}{}^{\beta}\delta A^{ij}{}_{[\beta}\delta^{\gamma}_{\alpha]})_{,\gamma} + R + \partial\widehat{\mathcal{O}} \\
\delta\overline{\mathcal{H}}_{\perp} &= -2aJh_{\bar{k}}{}^{\alpha}h_{\bar{l}}{}^{\beta}\delta A^{kl}{}_{\alpha,\beta} + R + \partial\widehat{\mathcal{O}}
\end{aligned}$$

we obtain the result

$$\begin{aligned}
\delta\widehat{\mathcal{H}}_{\mathrm{T}} &= 2aNJh_a{}^{\beta}h_b{}^{\alpha}\delta A^{ab}{}_{\beta,\alpha} + R \\
&= [2aJh_a{}^{\beta}h_b{}^{\alpha}\delta A^{ab}{}_{\beta}]_{,\alpha} + R + \partial\mathcal{O}_3
\end{aligned}$$

with the help of $\delta hA = \mathcal{O}_3$. Hence,

$$\begin{aligned}
&\delta P_0 = -\delta E_0 + R \\
&E_0 \equiv \oint \mathrm{d}S_{\gamma}\,(-2aJh_a{}^{\alpha}h_b{}^{\gamma}A^{ab}{}_{\alpha})
\end{aligned} \tag{6.16a}$$

and the correctly defined generator P_0 has the form

$$\widetilde{P}_0 \equiv P_0 + E_0. \tag{6.16b}$$

The surface term E_0 is finite on account of the adopted boundary conditions. As we shall see, the on-shell value E_0 of $\tilde{P}_0$ represents the value of the energy of the gravitating system.

Rotation. To find the correct definition of the rotation generator, let us look at the expression

$$\begin{aligned}
\delta M_{\alpha\beta} &= \int \mathrm{d}^3x\, \delta\mathcal{M}_{\alpha\beta} \\
\delta\mathcal{M}_{\alpha\beta} &= x_{\alpha}\delta\mathcal{P}_{\beta} - x_{\beta}\delta\mathcal{P}_{\alpha} + \delta\pi_{\alpha\beta}{}^{\gamma}{}_{,\gamma} + R.
\end{aligned}$$

Using the known form of $\delta\mathcal{P}_{\alpha}$, we find that

$$\begin{aligned}
&\delta M_{\alpha\beta} = -\delta E_{\alpha\beta} + R \\
&E_{\alpha\beta} \equiv \oint \mathrm{d}S_{\gamma}\,[-\pi_{\alpha\beta}{}^{\gamma} + x_{[\alpha}(\pi_{ij}{}^{\gamma}A^{ij}{}_{\beta]})].
\end{aligned} \tag{6.17a}$$

In the course of the calculation, the term $\int \mathrm{d}s_{[\alpha}x_{\beta]}\,X$, with $X \equiv \pi_{ij}{}^{\beta}A^{ij}{}_{\beta}$, has been discarded as $\mathrm{d}s_{\alpha} \sim x_{\alpha}$ on the integration sphere. The corresponding improved rotation generator is given by

$$\widetilde{M}_{\alpha\beta} \equiv M_{\alpha\beta} + E_{\alpha\beta}. \tag{6.17b}$$

A detailed analysis shows that the adopted boundary conditions do not guarantee the finiteness of the surface term $E_{\alpha\beta}$, as the integrand contains $\mathcal{O}_1$ terms. These troublesome terms are seen to vanish if we impose additional requirements on the phase space, consisting of conveniently chosen parity conditions (Regge and Teitelboim 1974, Beig and Murchadha 1978, Blagojević and Vasilić 1988). After that, the surface term $E_{\alpha\beta}$ is found to be finite and, consequently, $\widetilde{M}_{\alpha\beta}$ is well defined.

Boost. By varying the boost generator, we find that

$$\delta M_{0\beta} = \int \mathrm{d}^3x\, \delta\mathcal{M}_{0\beta}$$
$$\delta\mathcal{M}_{0\beta} = x_0\delta\mathcal{P}_\beta - x_\beta\delta\mathcal{H}_\mathrm{T} + \delta\pi_{0\beta}{}^{\gamma}{}_{,\gamma} + R.$$

Then, using the known expressions for $\delta\mathcal{P}_\beta$ and $\delta\mathcal{H}_\mathrm{T}$, we obtain

$$\delta M_{0\beta} = -\delta E_{0\beta} + R$$

$$E_{0\beta} \equiv \oint \mathrm{d}S_\gamma\, [-\pi_{0\beta}{}^{\gamma} + x_0(\pi_{ij}{}^{\alpha}A^{ij}{}_{[\beta}\delta_{\alpha]}{}^{\gamma}) - x_\beta(2aJh_a{}^{\alpha}h_b{}^{\gamma}A^{ab}{}_{\alpha})] \tag{6.18a}$$

so that the correct boost generator has the form

$$\widetilde{M}_{0\beta} \equiv M_{0\beta} + E_{0\beta}. \tag{6.18b}$$

Additional asymptotic parity conditions guarantee the finiteness of the surface term $E_{0\beta}$.

All these results refer to EC theory. The general $R + T^2 + R^2$ theory can be treated in an analogous manner.

Asymptotic symmetries and conservation laws

In the previous considerations, we obtained the improved Poincaré generators, acting on the phase space with given asymptotic properties. To clarify the physical meaning of these quantities, we now study their conservation laws. After transforming the canonical generators into the Lagrangian form, we shall make a comparison with the related GR results.

The algebra of the generators. The improved Poincaré generators are defined as the volume integrals of constraints plus certain surface integrals. Their action on the fields and momenta is the same as before, since surface terms act trivially on local quantities. Once we know that the generators $\widetilde{P}_\mu$ and $\widetilde{M}_{\mu\nu}$ have the standard action on the whole phase space, we can easily deduce their algebra to be that of the Poincaré group:

$$
\begin{aligned}
\{\widetilde{P}_\mu, \widetilde{P}_\nu\} &= 0 \\
\{\widetilde{P}_\mu, \widetilde{M}_{\nu\lambda}\} &= \eta_{\mu\nu}\widetilde{P}_\lambda - \eta_{\mu\lambda}\widetilde{P}_\nu \\
\{\widetilde{M}_{\mu\nu}, \widetilde{M}_{\lambda\rho}\} &= \eta_{\mu\rho}\widetilde{M}_{\nu\lambda} - \eta_{\mu\lambda}\widetilde{M}_{\nu\rho} - (\mu \leftrightarrow \nu).
\end{aligned}
\tag{6.19}
$$

This result is in agreement with the general theorem, which states that the PB of two well-defined (differentiable) generators is necessarily a well-defined generator (Brown and Henneaux 1986a). Note, however, that the line of reasoning that leads to this result does not guarantee the strong equalities in (6.19); they are rather equalities up to trivial generators (such as squares of constraints and surface terms). In what follows, we shall explicitly verify the absence of such terms in the part of the algebra involving $\widetilde{P}_0$, which will be sufficient to prove the conservation of all the symmetry generators.

Conservation laws. After a slight modification, Castellani's method can also be applied to study global symmetries. We can show that the necessary and sufficient conditions for a phase-space functional $G(q, \pi, t)$ to be a generator of global symmetries take the form

$$\{G, \widetilde{H}_\mathrm{T}\} + \frac{\partial G}{\partial t} = C_\mathrm{PFC} \tag{6.20}$$

$$\{G, \varphi_s\} \approx 0 \tag{6.21}$$

where $\widetilde{H}_\mathrm{T}$ is the improved Hamiltonian, $\varphi_s \approx 0$ are all the constraints in the theory and, as before, the equality sign means an equality up to the zero generators. The improved Poincaré generators are easily seen to satisfy the second condition, as they are given, up to surface terms, by volume integrals of FC constraints. On the other hand, having in mind that $\widetilde{H}_\mathrm{T} = \widetilde{P}_0$, we can verify that the first condition is also satisfied, as a consequence of the part of the algebra (6.19) involving $\widetilde{P}_0$.

The first condition is the canonical form of the conservation law. Indeed, it implies a weak equality

$$\frac{\mathrm{d}G}{\mathrm{d}t} \equiv \{G, \widetilde{P}_0\} + \frac{\partial G}{\partial t} \approx Z \tag{6.22}$$

wherefrom we see that the generator G is conserved only if the zero generator term Z is *absent*. We shall now explicitly evaluate $\mathrm{d}G/\mathrm{d}t$ for each of the generators $G = \widetilde{P}_\mu, \widetilde{M}_{\mu\nu}$, and check their conservation.

As is clear from equation (6.22), in order to verify the conservation of the improved Poincaré generators, we need their PBs with $\widetilde{P}_0$. Writing E_0 in the form $E_0 = -2a \oint \mathrm{d}S_\gamma\, A^\gamma$, where $A^\gamma \equiv \delta^\alpha_a A^{ac}{}_\alpha \delta^\gamma_c$, we see that the essential part of these brackets is the expression $\{A^\gamma, G\}$, representing the action of the Poincaré generator on the local quantity A^γ. Combining the relation $\delta_0 A^\gamma = \{A^\gamma, G\}$

with the known transformation laws for the fields, equation (6.9), we can read the relevant PBs:

$$\begin{aligned}
\{A^\gamma, \widetilde{P}_\nu\} &\approx -\partial_\nu A^\gamma \\
\{A^\gamma, \widetilde{M}_{\alpha\beta}\} &\approx -(\delta^\gamma_\alpha A_\beta + x_\alpha \partial_\beta A^\gamma) - (\alpha \leftrightarrow \beta) \\
\{A^\gamma, \widetilde{M}_{0\beta}\} &\approx -(A_0{}^\gamma{}_\beta - \delta^\gamma_\beta A_0{}^e{}_e) \\
&\qquad - A_\beta{}^\gamma{}_0 - x_\beta \partial_0 A^\gamma + x_0 \partial_\beta A^\gamma + \mathcal{O}_3.
\end{aligned} \tag{6.23}$$

The last of these equations can be further simplified by using the equations of motion. Thus, $R_{c0} \approx 0$ implies $\partial_0 A^\gamma \approx \partial_\alpha A^{\alpha\gamma}{}_0$, so that

$$A_\beta{}^\gamma{}_0 + x_\beta \partial_0 A^\gamma \approx \partial_\alpha (x_\beta A^{\alpha\gamma}{}_0).$$

(1) Let us begin with the conservation of energy. First, we note that $\widetilde{P}_0$, being a well-defined functional, must commute with itself: $\{\widetilde{P}_0, \widetilde{P}_0\} = 0$. Furthermore, $\partial \widetilde{P}_0/\partial t = \partial \widetilde{H}_\mathrm{T}/\partial t = C_\mathrm{PFC}$, since the only explicit time dependence of the total Hamiltonian is due to the presence of arbitrary multipliers, which are always multiplied by the PFC constraints. Therefore,

$$\frac{\mathrm{d}\widetilde{P}_0}{\mathrm{d}t} \approx \frac{\mathrm{d}E_0}{\mathrm{d}t} \approx 0 \tag{6.24a}$$

and we conclude that the surface term E_0, representing the value of the energy, is a conserved quantity.

(2) The linear momentum and the spatial angular momentum have no explicit time dependence. To evaluate their PBs with $\widetilde{P}_0$, we shall use the following procedure. Our improved generators are (non-local) functionals having the form of integrals of some local densities. The PB of two such generators can be calculated by acting with one of them on the integrand of the other. In the case of linear momentum, we have

$$\{\widetilde{P}_0, \widetilde{P}_\alpha\} = \int \mathrm{d}^3x\, \{\hat{\mathcal{H}}_\mathrm{T} - 2a\partial_\gamma A^\gamma, \widetilde{P}_\alpha\} \approx -2a \int \mathrm{d}^3x\, \partial_\gamma \{A^\gamma, \widetilde{P}_\alpha\}$$

because $\hat{\mathcal{H}}_\mathrm{T} \approx 0$ is a constraint in the theory, and therefore, weakly commutes with all the symmetry generators. The last term in this formula is easily evaluated with the help of equation (6.23), with the final result

$$\{\widetilde{P}_0, \widetilde{P}_\alpha\} \approx 2a \oint \mathrm{d}S_\gamma\, \partial_\alpha A^\gamma = 0$$

as a consequence of $\partial_\alpha A^\gamma = \mathcal{O}_3$. Therefore, no constant term appears in equation (6.22) for $G = \widetilde{P}_\alpha$, and we have the conservation law

$$\frac{\mathrm{d}\widetilde{P}_\alpha}{\mathrm{d}t} \approx \frac{\mathrm{d}E_\alpha}{\mathrm{d}t} \approx 0. \tag{6.24b}$$

(3) In a similar way, we can check the conservation of the rotation generator. Using equation (6.23), we find

$$\{\widetilde{P}_0, \widetilde{M}_{\alpha\beta}\} \approx -2a \int d^3x\, \partial_\gamma \{A^\gamma, \widetilde{M}_{\alpha\beta}\} \approx \oint (dS_\alpha\, x_\beta - dS_\beta\, x_\alpha)\partial_\gamma A^\gamma = 0.$$

The last equality is obtained with the help of the constraint

$$\mathcal{H}_\perp = aJR^{\bar{m}\bar{n}}{}_{\bar{m}\bar{n}} = -2a\partial_\gamma A^\gamma + \mathcal{O}_4$$

which implies $\partial_\gamma A^\gamma = \mathcal{O}_4$. Therefore, the Z term in (6.22) is absent for $G = \widetilde{M}_{\alpha\beta}$, and the value of the rotation generator is conserved:

$$\frac{d\widetilde{M}_{\alpha\beta}}{dt} \approx \frac{dE_{\alpha\beta}}{dt} \approx 0. \tag{6.24c}$$

(4) Finally, the boost generator has an explicit, linear dependence on time, and satisfies the relation $\partial\widetilde{M}_{0\beta}/\delta t = \widetilde{P}_\beta \approx E_\beta$. On the other hand,

$$\begin{aligned}\{\widetilde{P}_0, \widetilde{M}_{0\beta}\} \approx & -2a \int d^3x\, \partial_\gamma[-(A_0{}^\gamma{}_\beta - \delta^\gamma_\beta A_0{}^e{}_e) - \partial_\alpha(x_\beta A^{\alpha\gamma}{}_0)] \\ & = 2a \int d^3x\, \partial_\gamma(A_0{}^\gamma{}_\beta - \delta^\gamma_\beta A_0{}^e{}_e) = E_\beta\end{aligned}$$

where we have used the PBs in (6.23). In contrast to all other generators, the PB of the boost generator with $\widetilde{P}_0$ does not vanish: its on-shell value is precisely the value of the linear momentum E_β. Substitution of these results back into (6.22) yields the boost conservation law:

$$\frac{d\widetilde{M}_{0\beta}}{dt} \approx \frac{dE_{0\beta}}{dt} \approx 0. \tag{6.24d}$$

In conclusion,

All ten Poincaré generators in EC theory are conserved charges in the phase space with appropriate asymptotic conditions.

Comparison with GR. We now wish to transform our Hamiltonian expressions for the conserved charges into the Lagrangian form, and compare the obtained results with the related GR expressions. This will be achieved by expressing all momentum variables in terms of the fields and their derivatives, using primary constraints and the equations of motion.

To transform the expression for linear momentum, we start with the relation

$$\pi_{ij}{}^\beta A^{ij}{}_{[\alpha}\delta^\gamma_{\beta]} \approx 2a(A^{0\gamma}{}_\alpha - \delta^\gamma_\alpha A^{0c}{}_c) + \mathcal{O}_3$$

following from (6.12). Using equation (3.46) in the form

$$A_{ijk} = -(b_{[ij],k} - b_{(ik),j} + b_{(jk),i}) + \mathcal{O}_3$$

the right-hand side of this equation becomes

$$R \approx -2a[\eta^{\beta\gamma}(b_{(0\alpha),\beta} - b_{(\alpha\beta),0} - b_{(0\beta),\alpha}) + \delta^\gamma_\alpha b^\beta{}_{\beta 0}] + 4a(b^{[\beta}{}_0\delta^{\gamma]}_\alpha)_{,\beta} + \mathcal{O}_3.$$

A direct calculation shows that this expression can be rewritten in terms of the metric, i.e.

$$R = \eta_{\alpha\mu}h^{\mu 0\gamma} + \Lambda_\alpha{}^{\gamma\beta}{}_{,\beta} + \mathcal{O}_3$$

where the term $\Lambda_\alpha{}^{\gamma\beta}$, which is antisymmetric in (γ, β), does not contribute to the integral E_α, and

$$h^{\mu\nu\lambda} \equiv a[(-g)(g^{\mu\nu}g^{\lambda\rho} - g^{\mu\lambda}g^{\nu\rho})]_{,\rho} \equiv H^{\mu\nu\lambda\rho}{}_{,\rho}.$$

The final result is

$$E_\alpha = \eta_{\alpha\mu}\oint \mathrm{d}S_\gamma\, h^{\mu 0\gamma}.$$

In a similar way the energy expression can be brought to the form

$$E_0 = \oint \mathrm{d}S_\gamma\, h^{00\gamma}.$$

Energy and momentum can be written in a Lorentz covariant form as

$$E^\mu = \int \mathrm{d}^3x\, \theta^{\mu 0} \qquad \theta^{\mu\nu} \equiv h^{\mu\nu\lambda}{}_{,\lambda} \tag{6.25}$$

where $\theta^{\mu\nu}$ is equal to the symmetric energy–momentum complex in GR.

Energy and momentum in the EC theory are given by the same expressions as in GR.

For the rotation and boost, the calculation is similar but slightly more complicated, and it gives the following result:

$$\begin{aligned} E^{\mu\nu} &= \oint \mathrm{d}S_\gamma\, K^{\mu\nu 0\gamma} \\ K^{\mu\nu\lambda\rho} &\equiv x^\mu h^{\nu\lambda\rho} - x^\nu h^{\mu\lambda\rho} + H^{\lambda\mu\nu\rho} \end{aligned} \tag{6.26}$$

where K satisfies the relation $\partial_\rho K^{\mu\nu\lambda\rho} = x^\mu\theta^{\nu\lambda} - x^\nu\theta^{\mu\lambda}$.

The angular momentum also coincides with the GR expression.

It should be noted that in the general $R + T^2 + R^2$ theory we obtain the same results only when all tordions are massive. The results obtained in this chapter can be used to study the important problem of stability of the Minkowski vacuum in PGT.

6.3 Conservation laws—the teleparallel theory

Our previous treatment of the relation between gauge symmetries and conservation laws refers to EC theory, which is defined in a U_4 space. Teleparallel theory is defined in a T_4 space, so that the corresponding analysis must be changed in some technical details, due to the presence of Lagrange multipliers in the action (Blagojević and Vasilić 2000a, b).

As we are interested in the conservation laws of the energy–momentum and angular momentum, we shall begin our analysis with the construction of the Poincaré gauge generator.

A simple model

The T^2 part of the teleparallel Lagrangian (3.61) is a special case of the general $R + T^2 + R^2$ theory, the Poincaré gauge generator of which has already been constructed in section 6.1. In teleparallel theory, this result is to be corrected with the terms stemming from the $\lambda_{ij}{}^{\mu\nu} R^{ij}{}_{\mu\nu}$ part of the Lagrangian. Since the general construction procedure is rather complicated, we give here a detailed analysis of the simple special case defined by the full absence of the torsion part from (3.61). The results obtained will provide a clear suggestion for the construction of the Poincaré gauge generator in the complete teleparallel theory.

The simple Lagrangian we are going to study reads:

$$\widetilde{\mathcal{L}}_0 = \lambda_{ij}{}^{\mu\nu} R^{ij}{}_{\mu\nu}. \tag{6.27}$$

Clearly, the tetrad variables are absent, but $\widetilde{\mathcal{L}}_0$ possesses all the gauge symmetries of the general teleparallel theory. The field equations are:

$$\nabla_\mu \lambda_{ij}{}^{\mu\nu} = 0 \qquad R^{ij}{}_{\mu\nu} = 0.$$

A straightforward Hamiltonian analysis gives the total Hamiltonian which coincides with $\hat{\mathcal{H}}_{\mathrm{T}}$, equation (5.63), up to the tetrad related terms:

$$\widetilde{\mathcal{H}}_{\mathrm{T}} = -\tfrac{1}{2} A^{ij}{}_0 \widetilde{\mathcal{H}}_{ij} - \lambda_{ij}{}^{\alpha\beta} \widetilde{\mathcal{H}}^{ij}{}_{\alpha\beta} + \tfrac{1}{2} u^{ij}{}_0 \pi_{ij}{}^0 + \tfrac{1}{4} u_{ij}{}^{\alpha\beta} \pi^{ij}{}_{\alpha\beta} \tag{6.28a}$$

where

$$\begin{aligned} \widetilde{\mathcal{H}}_{ij} &\equiv \nabla_\alpha \pi_{ij}{}^\alpha + 2\pi^s{}_{[i0\alpha} \lambda_{sj]}{}^{0\alpha} \\ \widetilde{\mathcal{H}}^{ij}{}_{\alpha\beta} &\equiv R^{ij}{}_{\alpha\beta} - \tfrac{1}{2} \nabla_{[\alpha} \pi^{ij}{}_{0\beta]} = \overline{\mathcal{H}}^{ij}{}_{\alpha\beta}. \end{aligned} \tag{6.28b}$$

Among the primary constraints, $\pi_{ij}{}^0$ and $\pi^{ij}{}_{\alpha\beta}$ are FC and, consequently, responsible for the existence of gauge symmetries, while $\phi_{ij}{}^\alpha \equiv \pi_{ij}{}^\alpha - 4\lambda_{ij}{}^{0\alpha}$ and $\pi^{ij}{}_{0\beta}$ are second class. The PB algebra of the Hamiltonian constraints has the form

$$\begin{aligned} \{\widetilde{\mathcal{H}}_{ij}, \widetilde{\mathcal{H}}'_{kl}\} &= (\eta_{ik} \widetilde{\mathcal{H}}_{lj} + \eta_{jk} \widetilde{\mathcal{H}}_{il})\delta - (k \leftrightarrow l) \\ \{\widetilde{\mathcal{H}}^{ij}{}_{\alpha\beta}, \widetilde{\mathcal{H}}'_{kl}\} &= (\delta^i_k \widetilde{\mathcal{H}}_l{}^j{}_{\alpha\beta} + \delta^j_k \widetilde{\mathcal{H}}^i{}_{l\alpha\beta})\delta - (k \leftrightarrow l) \\ \{\widetilde{\mathcal{H}}^{ij}{}_{\alpha\beta}, \widetilde{\mathcal{H}}'^{kl}{}_{\gamma\delta}\} &= 0. \end{aligned} \tag{6.29}$$

The Lorentz generator. We begin with the PFC constraint $\pi_{ij}{}^0$, and define

$$\widetilde{G}(\omega) = -\tfrac{1}{2}\dot{\omega}^{ij}\pi_{ij}{}^0 - \tfrac{1}{2}\omega^{ij}\widetilde{S}_{ij}. \tag{6.30a}$$

From the second condition in (5.23b), we obtain $\widetilde{S}_{ij} = -\widetilde{\mathcal{H}}_{ij} + C_{\text{PFC}}$. Then, using the constraint algebra (6.29) and the third condition in (5.23b), we find the complete function $\widetilde{S}_{ij}$ to read:

$$\widetilde{S}_{ij} = -\widetilde{\mathcal{H}}_{ij} + 2A^s{}_{[i0}\pi_{sj]}{}^0 + \lambda_{s[i}{}^{\alpha\beta}\pi^s{}_{j]\alpha\beta}. \tag{6.30b}$$

It is easy to verify that the action of the Lorentz gauge generator $\widetilde{G}(\omega)$ on the fields $A^{ij}{}_\mu$ and $\lambda_{ij}{}^{\mu\nu}$ has the expected form, coinciding with the ω part of the Poincaré transformations (3.64).

The λ generator. The λ gauge generator is obtained by starting Castellani's algorithm with the PFC constraint $\pi^{ij}{}_{\alpha\beta}$. All the steps of the construction, the analysis and the final result are the same as in section 5.3. Thus,

$$\widetilde{G}(\varepsilon) = \tfrac{1}{4}\,\dot{\varepsilon}_{ij}{}^{0\alpha\beta}\pi^{ij}{}_{\alpha\beta} + \tfrac{1}{4}\,\varepsilon_{ij}{}^{0\alpha\beta}S^{ij}{}_{\alpha\beta} - \tfrac{1}{4}\,\varepsilon_{ij}{}^{\alpha\beta\gamma}\nabla_\alpha\pi^{ij}{}_{\beta\gamma} \tag{6.31}$$

with $S^{ij}{}_{\alpha\beta}$ given by equation (5.73b).

The action of $\widetilde{G}(\varepsilon)$ on the fields is the same as in (3.65), while the action of the complete generator $\widetilde{G}(\omega) + \widetilde{G}(\varepsilon)$ yields:

$$\begin{aligned}\delta_0 A^{ij}{}_\mu &= \omega^i{}_k\, A^{kj}{}_\mu + \omega^j{}_k\, A^{ik}{}_\mu - \omega^{ij}{}_{,\mu} \\ \delta_0\lambda_{ij}{}^{\mu\nu} &= \omega_i{}^k\,\lambda_{kj}{}^{\mu\nu} + \omega_j{}^k\,\lambda_{ik}{}^{\mu\nu} + \nabla_\lambda\,\varepsilon_{ij}{}^{\mu\nu\lambda}.\end{aligned} \tag{6.32}$$

These transformation laws exhaust the gauge symmetries of our simple model. Note, however, that the Lagrangian $\widetilde{\mathcal{L}}_0$ also possesses the local translational symmetry, which has not been obtained so far. If Castellani's algorithm is an exhaustive one, then the translational symmetry must be somehow hidden in this result. In what follows, we shall demonstrate that this is really true, namely that the translational symmetry emerges from a simple redefinition of the gauge parameters in (6.32).

The Poincaré generator. Let us consider the following replacement of the parameters ω^{ij} and $\varepsilon_{ij}{}^{\mu\nu\lambda}$ in equation (6.32):

$$\begin{aligned}\omega^{ij} &\to \omega^{ij} + \xi^\mu A^{ij}{}_\mu \\ \varepsilon_{ij}{}^{\mu\nu\lambda} &\to -(\xi^\mu\lambda_{ij}{}^{\nu\lambda} + \xi^\nu\lambda_{ij}{}^{\lambda\mu} + \xi^\lambda\lambda_{ij}{}^{\mu\nu}).\end{aligned} \tag{6.33}$$

The resulting *on-shell* transformations of the fields $A^{ij}{}_\mu$ and $\lambda_{ij}{}^{\mu\nu}$ are easily found to be exactly the Poincaré gauge transformations (3.64). As we can see, the local translations are not obtained as independent gauge transformations, but rather emerge as part of the λ and Lorentz symmetries. The corresponding Poincaré generator is obtained by using the same replacement in the gauge generator $\widetilde{G}(\omega) + \widetilde{G}(\varepsilon)$. Thus, we find:

$$\widetilde{G} = \widetilde{G}(\omega) + \widetilde{G}(\xi) \tag{6.34a}$$

where the first term, describing local Lorentz rotations, has the form (6.30), while the second term, describing local translations, is given by

$$\begin{aligned}\tilde{G}(\xi) = & -\dot{\xi}^0(\tfrac{1}{2}A^{ij}{}_0\pi_{ij}{}^0 + \tfrac{1}{4}\lambda_{ij}{}^{\alpha\beta}\pi^{ij}{}_{\alpha\beta}) - \xi^0\widetilde{\mathcal{H}}_{\mathrm{T}} \\ & -\dot{\xi}^\alpha(\tfrac{1}{2}A^{ij}{}_\alpha\pi_{ij}{}^0 - \tfrac{1}{2}\lambda_{ij}{}^{0\beta}\pi^{ij}{}_{\alpha\beta}) \\ & -\xi^\alpha[\tilde{\mathcal{P}}_\alpha - \tfrac{1}{4}\lambda_{ij}{}^{\beta\gamma}\partial_\alpha\pi^{ij}{}_{\beta\gamma} - \tfrac{1}{2}\partial_\gamma(\lambda_{ij}{}^{\beta\gamma}\pi^{ij}{}_{\alpha\beta})].\end{aligned} \tag{6.34b}$$

In these expressions, we used the following notation:

$$\begin{aligned}\tilde{\mathcal{P}}_\alpha &\equiv \widetilde{\mathcal{H}}_\alpha - \tfrac{1}{2}A^{ij}{}_\alpha\widetilde{\mathcal{H}}_{ij} + 2\lambda_{ij}{}^{0\beta}\widetilde{\mathcal{H}}^{ij}{}_{\alpha\beta} + \tfrac{1}{2}\pi_{ij}{}^0\partial_\alpha A^{ij}{}_0 \\ \widetilde{\mathcal{H}}_\alpha &\equiv \tfrac{1}{2}\pi^{ij}{}_{0\alpha}\nabla_\beta\lambda_{ij}{}^{0\beta}.\end{aligned} \tag{6.34c}$$

Note that the term $\widetilde{\mathcal{H}}_\alpha$ in $\tilde{G}(\xi)$ has the structure of squares of constraints, and therefore, does not contribute to the non-trivial field transformations. Nevertheless, we shall retain it in the generator because it makes the field transformations practically off-shell (up to $R^{ij}{}_{\alpha\beta} \approx 0$). This will help us to find the form of the extension of $\tilde{G}$ to general teleparallel theory straightforwardly.

The Poincaré gauge generators

Starting from the Poincaré gauge generator (6.34) of the simple model (6.27), and comparing it with the earlier result (6.4), it is almost clear how its modification to include the tetrad sector should be defined. The complete Poincaré gauge generator of general teleparallel theory (3.61) is expected to have the form

$$G = G(\omega) + G(\xi) \tag{6.35a}$$

where the first term describes local Lorentz rotations,

$$G(\omega) = -\tfrac{1}{2}\dot{\omega}^{ij}\pi_{ij}{}^0 - \tfrac{1}{2}\omega^{ij}S_{ij} \tag{6.35b}$$

while the second term describes local translations,

$$\begin{aligned}G(\xi) = & -\dot{\xi}^0(b^k{}_0\pi_k{}^0 + \tfrac{1}{2}A^{ij}{}_0\pi_{ij}{}^0 + \tfrac{1}{4}\lambda_{ij}{}^{\alpha\beta}\pi^{ij}{}_{\alpha\beta}) - \xi^0\mathcal{P}_0 \\ & -\dot{\xi}^\alpha(b^k{}_\alpha\pi_k{}^0 + \tfrac{1}{2}A^{ij}{}_\alpha\pi_{ij}{}^0 - \tfrac{1}{2}\lambda_{ij}{}^{0\beta}\pi^{ij}{}_{\alpha\beta}) \\ & -\xi^\alpha[\mathcal{P}_\alpha - \tfrac{1}{4}\lambda_{ij}{}^{\beta\gamma}\partial_\alpha\pi^{ij}{}_{\beta\gamma} - \tfrac{1}{2}\partial_\gamma(\lambda_{ij}{}^{\beta\gamma}\pi^{ij}{}_{\alpha\beta})].\end{aligned} \tag{6.35c}$$

Here, we used the following notation:

$$S_{ij} = -\overline{\mathcal{H}}_{ij} + 2b_{[i0}\pi_{j]}{}^0 + 2A^s{}_{[i0}\pi_{sj]}{}^0 + 2\lambda_{s[i}{}^{\alpha\beta}\pi^s{}_{j]\alpha\beta}$$

$$\mathcal{P}_0 \equiv \widehat{\mathcal{H}}_\mathrm{T} = \mathcal{H}_\mathrm{T} - \partial_\alpha \bar{D}^\alpha \tag{6.35d}$$

$$\mathcal{P}_\alpha = \overline{\mathcal{H}}_\alpha - \tfrac{1}{2}A^{ij}{}_\alpha\overline{\mathcal{H}}_{ij} + 2\lambda_{ij}{}^{0\beta}\overline{\mathcal{H}}^{ij}{}_{\alpha\beta} + \pi_k{}^0\partial_\alpha b^k{}_0 + \tfrac{1}{2}\pi_{ij}{}^0\partial_\alpha A^{ij}{}_0$$

where $\overline{\mathcal{H}}_\alpha, \overline{\mathcal{H}}_{ij}, \overline{\mathcal{H}}^{ij}{}_{\alpha\beta}, \bar{D}^\alpha$ and $\mathcal{H}_\mathrm{T}$ are defined in equations (5.63). The Poincaré generator G is obtained from the simplified expression (6.34) by a natural process of extension, which consists of

- the replacements $(\widetilde{\mathcal{H}}_\alpha, \widetilde{\mathcal{H}}_{ij}, \widetilde{\mathcal{H}}^{ij}{}_{\alpha\beta}, \widetilde{\mathcal{H}}_\mathrm{T}) \to (\overline{\mathcal{H}}_\alpha, \overline{\mathcal{H}}_{ij}, \overline{\mathcal{H}}^{ij}{}_{\alpha\beta}, \widehat{\mathcal{H}}_\mathrm{T})$; and
- the addition of various $\pi_k{}^0$ terms.

This amounts to completing the Poincaré gauge generator so that it also acts correctly in the tetrad sector.

The proof that the Poincaré gauge generator has the form (6.35) is realized by showing that G produces the correct Poincaré gauge transformations on the complete phase space, i.e. on all the fields and momenta. Explicit demonstration of this statement can be carried out in analogy with the EC case, and by using only those relations that characterize an arbitrary teleparallel theory. Hence,

The Poincaré gauge generator has the form (6.35) for any choice of parameters in the teleparallel theory (3.61).

Asymptotic conditions

Now we turn our attention to the asymptotic behaviour of the dynamical variables and the related form of the symmetry generators.

Asymptotic Poincaré invariance. We restrict our discussion to gravitating systems that are characterized by the global Poincaré symmetry at large distances. The global Poincaré transformations of fields are obtained from the corresponding local transformations (3.64) by the replacements (6.8). In the same manner, the global Poincaré generators are obtained from the corresponding local expressions (6.35):

$$G = \tfrac{1}{2}\omega^{ij}M_{ij} - \varepsilon^\mu P_\mu \tag{6.36a}$$

where

$$P_\mu = \int \mathrm{d}^3x\,\mathcal{P}_\mu \qquad M_{\mu\nu} = \int \mathrm{d}^3x\,\mathcal{M}_{\mu\nu}$$

$$\mathcal{M}_{\alpha\beta} = x_\alpha\mathcal{P}_\beta - x_\beta\mathcal{P}_\alpha - S_{\alpha\beta}$$

$$\mathcal{M}_{0\beta} = x_0\mathcal{P}_\beta - x_\beta\mathcal{P}_0 - S_{0\beta} + b^k{}_\beta\pi_k{}^0 + \tfrac{1}{2}A^{ij}{}_\beta\pi_{ij}{}^0 - \tfrac{1}{2}\lambda_{ij}{}^{0\gamma}\pi^{ij}{}_{\beta\gamma}. \tag{6.36b}$$

The quantities $\mathcal{P}_\mu$ and $S_{\mu\nu} = \delta^i_\mu\delta^j_\nu S_{ij}$ are defined in (6.35).

Boundary conditions. We assume that dynamical variables satisfy the boundary conditions

(a) $g_{\mu\nu} = \eta_{\mu\nu} + \mathcal{O}_1$ and
(b) $R^{ij}{}_{\mu\nu} = \mathcal{O}_{2+\alpha}$ $(\alpha > 0)$

which correspond to asymptotically flat spacetime.

The first condition is self-evident, and the second is trivially satisfied in teleparallel theory, where $R^{ij}{}_{\mu\nu}(A) = 0$. The vanishing of the curvature means that the Lorentz gauge potential $A^{ij}{}_{\mu}$ is a pure gauge, hence it can be transformed to zero (at least locally) by a suitable local Lorentz transformation. Therefore, we can adopt the following boundary conditions:

$$b^i{}_\mu = \delta^i_\mu + \mathcal{O}_1 \qquad A^{ij}{}_\mu = \widehat{\mathcal{O}}. \tag{6.37a}$$

There is one more Lagrangian variable in teleparallel theory, the Lagrange multiplier $\lambda_{ij}{}^{\mu\nu}$, which is not directly related to the geometric conditions (a) and (b). Its asymptotic behaviour can be determined with the help of the second field equation, leading to

$$\lambda_{ij}{}^{\mu\nu} = \text{constant} + \mathcal{O}_1. \tag{6.37b}$$

The vacuum values of the fields, $b^i{}_\mu = \delta^i_\mu$, $A^{ij}{}_\mu = 0$ and $\lambda_{ij}{}^{\mu\nu} = \text{constant}$, are invariant under the global Poincaré transformations. Demanding also that the boundary conditions (6.37) be invariant under these transformations, we obtain the additional requirements:

$$\begin{aligned} b^k{}_{\mu,\nu} &= \mathcal{O}_2 \qquad b^k{}_{\mu,\nu\rho} = \mathcal{O}_3 \\ \lambda_{ij}{}^{\mu\nu}{}_{,\rho} &= \mathcal{O}_2 \qquad \lambda_{ij}{}^{\mu\nu}{}_{,\rho\lambda} = \mathcal{O}_3. \end{aligned} \tag{6.38}$$

Boundary conditions (6.37) and (6.38) are compatible not only with global Poincaré transformations, but also with a restricted set of local Poincaré transformations, the gauge parameters of which decrease sufficiently fast for large r.

The asymptotic behaviour of momentum variables is determined using the relations $\pi_A = \partial\widetilde{\mathcal{L}}/\partial\varphi^A$. Thus,

$$\begin{aligned} \pi_i{}^0, \pi_{ij}{}^0, \pi^{ij}{}_{\mu\nu} &= \widehat{\mathcal{O}} \\ \pi_i{}^\alpha &= \mathcal{O}_2 \\ \pi_{ij}{}^\alpha &= 4\lambda_{ij}{}^{0\alpha} + \widehat{\mathcal{O}}. \end{aligned} \tag{6.39}$$

In a similar manner, we can determine the asymptotic behaviour of the Hamiltonian multipliers.

Now we wish to check whether the global Poincaré generators (6.35) are well defined in the phase space characterized by these boundary conditions.

The improved Poincaré generators

The Poincaré generators (6.35), which define the asymptotic symmetry of spacetime, are not differentiable. In the process of improving their form by adding certain surface terms, we shall obtain the conserved charges of the gravitating system, energy-momentum and angular momentum.

Spatial translation. The variation of the spatial translation generator P_α has the form

$$\begin{aligned}\delta P_\alpha &= \int \mathrm{d}^3x\, \delta\mathcal{P}_\alpha \\ \delta\mathcal{P}_\alpha &= \delta\overline{\mathcal{H}}_\alpha - \tfrac{1}{2}A^{ij}{}_\alpha \delta\overline{\mathcal{H}}_{ij} + 2\lambda_{ij}{}^{0\beta}\delta\overline{\mathcal{H}}^{ij}{}_{\alpha\beta} + \pi_k{}^0 \delta b^k{}_{0,\alpha} \\ &\quad + \tfrac{1}{2}\pi_{ij}{}^0 \delta A^{ij}{}_{0,\alpha} - \tfrac{1}{4}\lambda_{ij}{}^{\beta\gamma}\delta\pi^{ij}{}_{\beta\gamma,\alpha} - \tfrac{1}{2}\delta(\lambda_{ij}{}^{\beta\gamma}\pi^{ij}{}_{\alpha\beta})_{,\gamma} + R.\end{aligned}$$

The relation $\pi_k{}^0\delta b^k{}_{0,\alpha} = (\pi_k{}^0\delta b^k{}_0)_{,\alpha} + R$, in conjunction with the boundary condition $\pi_k{}^0 = \widehat{\mathcal{O}}$, leads to $\pi_k{}^0\delta b^k{}_{0,\alpha} = \partial\widehat{\mathcal{O}} + R$. Applying the same reasoning to the terms proportional to $\pi_{ij}{}^0$, $\pi^{ij}{}_{\alpha\beta}$, $\pi^{ij}{}_{0\beta}$ (present in $\overline{\mathcal{H}}^{ij}{}_{0\beta}$) and $A^{ij}{}_\mu$, we find that

$$\delta\mathcal{P}_\alpha = \delta\overline{\mathcal{H}}_\alpha + \partial\widehat{\mathcal{O}} + R.$$

Finally, using the explicit form (5.63c) of $\overline{\mathcal{H}}_\alpha$, we obtain

$$\begin{aligned}\delta\mathcal{P}_\alpha &= -\delta(b^i{}_\alpha\pi_i{}^\gamma)_{,\gamma} + (\pi_i{}^\gamma\delta b^i{}_\gamma)_{,\alpha} + \partial\widehat{\mathcal{O}} + R \\ &= -\delta(b^i{}_\alpha\pi_i{}^\gamma)_{,\gamma} + \partial\mathcal{O}_3 + R\end{aligned}$$

where the last equality follows from $\pi_i{}^\gamma\delta b^i{}_\gamma = \mathcal{O}_3$. Therefore,

$$\begin{aligned}\delta P_\alpha &= -\delta E_\alpha + R \\ E_\alpha &\equiv \oint \mathrm{d}S_\gamma\,(b^k{}_\alpha\pi_k{}^\gamma).\end{aligned} \tag{6.40a}$$

Now, if we redefine the translation generator P_α by

$$P_\alpha \to \widetilde{P}_\alpha = P_\alpha + E_\alpha \tag{6.40b}$$

the new, improved expression $\widetilde{P}_\alpha$ has well-defined functional derivatives.

The surface integral for E_α is finite since $b^k{}_\alpha\pi_k{}^\gamma = \mathcal{O}_2$, in view of boundary conditions (6.39). The on-shell value E_α of $\widetilde{P}_\alpha$ represents the value of the linear momentum of the system.

Time translation. A similar procedure can be applied to the time translation generator P_0:

$$\begin{aligned}\delta P_0 &= \int \mathrm{d}^3x\,\delta\mathcal{P}_0 \\ \delta\mathcal{P}_0 &= \delta\mathcal{H}_{\mathrm{T}} - \delta(b^k{}_0\pi_k{}^\gamma)_{,\gamma} + \partial\widehat{\mathcal{O}}\end{aligned}$$

where we have used the adopted boundary conditions for $A^{ij}{}_0$ and $\pi^{ij}{}_{0\beta}$. Since $\mathcal{H}_{\mathrm{T}}$ does not depend on the derivatives of momenta (on shell), we can write

$$\begin{aligned}\delta\mathcal{H}_{\mathrm{T}} &= \frac{\partial\mathcal{H}_{\mathrm{T}}}{\partial b^k{}_{\mu,\alpha}}\delta b^k{}_{\mu,\alpha} + \frac{1}{2}\frac{\partial\mathcal{H}_{\mathrm{T}}}{\partial A^{ij}{}_{\mu,\alpha}}\delta A^{ij}{}_{\mu,\alpha} + R \\ &\approx -\frac{\partial\mathcal{L}}{\partial b^k{}_{\mu,\alpha}}\delta b^k{}_{\mu,\alpha} - \frac{1}{2}\frac{\partial\mathcal{L}}{\partial A^{ij}{}_{\mu,\alpha}}\delta A^{ij}{}_{\mu,\alpha} + R.\end{aligned}$$

The second term has the form $\partial\widehat{\mathcal{O}} + R$, so that

$$\delta\mathcal{H}_{\mathrm{T}} \approx -\partial_\alpha\left(\frac{\partial\mathcal{L}}{\partial b^k{}_{\mu,\alpha}}\delta b^k{}_\mu\right) + \partial\widehat{\mathcal{O}} + R = \partial\mathcal{O}_3 + R$$

and we find

$$\delta\mathcal{P}_0 = -\delta(b^k{}_0\pi_k{}^\gamma)_{,\gamma} + \partial\mathcal{O}_3 + R.$$

Hence,

$$\begin{aligned}\delta P_0 &= -\delta E_0 + R \\ E_0 &= \oint \mathrm{d}S_\gamma\,(b^k{}_0\pi_k{}^\gamma).\end{aligned} \tag{6.41a}$$

The improved time translation generator

$$\widetilde{P}_0 = P_0 + E_0 \tag{6.41b}$$

has well-defined functional derivatives, and the surface term E_0 is finite. The on-shell value E_0 of $\widetilde{P}_0$ is the energy of the gravitating system.

These expressions for the energy and momentum can be written in a Lorentz covariant form:

$$E_\mu = \oint \mathrm{d}S_\gamma\,(b^k{}_\mu\pi_k{}^\gamma). \tag{6.42}$$

Rotation. Next, we want to check whether the rotation generator $M_{\alpha\beta}$ has well-defined functional derivatives:

$$\begin{aligned}\delta M_{\alpha\beta} &= \int \mathrm{d}^3x\,\delta\mathcal{M}_{\alpha\beta} \\ \delta\mathcal{M}_{\alpha\beta} &= x_\alpha\delta\mathcal{P}_\beta - x_\beta\delta\mathcal{P}_\alpha + \delta\pi_{\alpha\beta}{}^\gamma{}_{,\gamma} + R\end{aligned}$$

where $\pi_{\mu\nu}{}^\rho = \delta^i_\mu\delta^j_\nu\pi_{ij}{}^\rho$. Using the known form of $\delta\mathcal{P}_\alpha$, we find that

$$\begin{aligned}\delta M_{\alpha\beta} &= -\delta E_{\alpha\beta} + R \\ E_{\alpha\beta} &= \oint \mathrm{d}S_\gamma\,[x_\alpha(b^k{}_\beta\pi_k{}^\gamma) - x_\beta(b^k{}_\alpha\pi_k{}^\gamma) - \pi_{\alpha\beta}{}^\gamma]\end{aligned} \tag{6.43}$$

where we have used the fact that $\oint \mathrm{d}S_{[\alpha}\, x_{\beta]}\, \phi = 0$ ($\phi \equiv \pi_k{}^\gamma \delta b^k{}_\gamma$). The improved rotation generator has the form

$$\widetilde{M}_{\alpha\beta} \equiv M_{\alpha\beta} + E_{\alpha\beta}. \tag{6.44}$$

Although $\widetilde{M}_{\alpha\beta}$ has well-defined functional derivatives, the assumed boundary conditions (6.37)–(6.39) do not ensure the finiteness of $E_{\alpha\beta}$ owing to the presence of $\mathcal{O}_1$ terms. Note, however, that the actual asymptotics is refined by the principle of arbitrarily fast decrease of all on-shell vanishing expressions. Thus, we can use the constraints $\overline{\mathcal{H}}_\alpha$ and $\overline{\mathcal{H}}_{ij}$, in the lowest order in r^{-1}, to conclude that

$$\pi_\alpha{}^\beta{}_{,\beta} = \mathcal{O}_4 \qquad 2\pi_{[\alpha\beta]} + \pi_{\alpha\beta}{}^\gamma{}_{,\gamma} = \mathcal{O}_3 \tag{6.45}$$

where $\pi_{\mu\nu} = \delta^i_\mu \eta_{\nu\rho} \pi_i{}^\rho$. As a consequence, the angular momentum density decreases like $\mathcal{O}_3$:

$$(2x_{[\alpha}\pi_{\beta]}{}^\gamma - \pi_{\alpha\beta}{}^\gamma)_{,\gamma} = \mathcal{O}_3.$$

As all variables in the theory are assumed to have asymptotically polynomial behaviour in r^{-1}, the integrand of $E_{\alpha\beta}$ must essentially be of an $\mathcal{O}_2$ type to agree with this constraint. The possible $\mathcal{O}_1$ terms are divergenceless, and do not contribute to the corresponding surface integral (exercise 20). This ensures the finiteness of the rotation generator.

Boost. The variation of the boost generator is given by

$$\delta M_{0\beta} = \int \mathrm{d}^3x\, \delta\mathcal{M}_{0\beta}$$
$$\delta\mathcal{M}_{0\beta} = x_0 \delta\mathcal{P}_\beta - x_\beta \delta\mathcal{P}_0 + \delta\pi_{0\beta}{}^\gamma{}_{,\gamma} + R.$$

Here, we need to calculate $\delta\mathcal{P}_0$ up to terms $\partial\mathcal{O}_4$. A simple calculation yields

$$\delta\mathcal{M}_{0\beta} = \delta(\pi_{0\beta}{}^\gamma - x_0 b^k{}_\beta \pi_k{}^\gamma + x_\beta b^k{}_0 \pi_k{}^\gamma)_{,\gamma} + (x_\beta X^\gamma)_{,\gamma} + \partial\mathcal{O}_3 + R$$

where the term $X^\gamma \equiv (\partial\mathcal{L}/\partial b^k{}_{\mu,\gamma})\delta b^k{}_\mu$ is not a total variation, nor does it vanish on account of the adopted boundary conditions. If we further restrict the phase space by adopting suitable parity conditions, this term vanishes (Blagojević and Vasilić 2000b). After that, the improved boost generator reduces to the form

$$\begin{aligned} &\widetilde{M}_{0\beta} \equiv M_{0\beta} + E_{0\beta} \\ &E_{0\beta} = \oint \mathrm{d}S_\gamma\, [x_0(b^k{}_\beta \pi_k{}^\gamma) - x_\beta(b^k{}_0 \pi_k{}^\gamma) - \pi_{0\beta}{}^\gamma]. \end{aligned} \tag{6.46}$$

It remains to be shown that the adopted boundary conditions ensure the finiteness of $E_{0\beta}$. That this is indeed true can be seen by analysing the constraints $\widehat{\mathcal{H}}_\mathrm{T}$ and $\overline{\mathcal{H}}_{ij}$ at spatial infinity, which leads to

$$\pi_0{}^\beta{}_{,\beta} = \mathcal{O}_4 \qquad \pi_{0\beta} + \pi_{0\beta}{}^\gamma{}_{,\gamma} = \mathcal{O}_3. \tag{6.47}$$

The first equation follows from $\hat{\mathcal{H}}_T \equiv \mathcal{H}_T - \partial_\alpha \bar{D}^\alpha \approx 0$, which implies

$$\pi_0{}^\beta{}_{,\beta} \approx \mathcal{H}_T + \mathcal{O}_4 \approx \pi_A \dot{\varphi}^A - \mathcal{L} + \mathcal{O}_4 = \mathcal{O}_4.$$

Now, it follows that the boost density decreases like $\mathcal{O}_3$ for large r. Then, the same line of reasoning as the one used in discussing rotations leads us to conclude that $E_{0\beta}$ is finite. The improved boost generator $\widetilde{M}_{0\beta}$ is a well-defined functional on the phase space defined by the boundary conditions (6.37)–(6.39) together with the adopted parity conditions.

Using the Lorentz four-notation, we can write:

$$\begin{aligned} \tilde{M}_{\mu\nu} &\equiv M_{\mu\nu} + E_{\mu\nu} \\ E_{\mu\nu} &= \oint \mathrm{d}S_\gamma \, [x_\mu(b^k{}_\nu \pi_k{}^\gamma) - x_\nu(b^k{}_\mu \pi_k{}^\gamma) - \pi_{\mu\nu}{}^\gamma]. \end{aligned} \tag{6.48}$$

Conserved charges

The improved Poincaré generators satisfy the PB algebra (6.19), and their conservation laws are determined by the general conditions, the essential content of which is expressed by equation (6.22).

To verify the conservation of the improved generators, we need their PBs with $\widetilde{P}_0 \approx \oint \mathrm{d}S_\gamma \, \pi_0{}^\gamma$. Using the Poincaré transformation law for $\pi_k{}^\gamma$,

$$\delta_0 \pi_k{}^\gamma \approx \varepsilon_k{}^s \pi_s{}^\gamma + \varepsilon^\gamma{}_\beta \pi_k{}^\beta + \varepsilon^0{}_\beta \frac{\partial \mathcal{L}}{\partial b^k{}_{\gamma,\beta}} - (\varepsilon^\mu{}_\nu x^\nu + \varepsilon^\mu)\partial_\mu \pi_k{}^\gamma$$

and writing it in the form $\delta_0 \pi_k{}^\gamma = \{\pi_k{}^\gamma, G\}$, we find that

$$\begin{aligned} \{\pi_0{}^\gamma, \widetilde{P}_\mu\} &\approx \partial_\mu \pi_0{}^\gamma \\ \{\pi_0{}^\gamma, \widetilde{M}_{\alpha\beta}\} &\approx (\delta^\gamma_\alpha \pi_{0\beta} + x_\alpha \partial_\beta \pi_0{}^\gamma) - (\alpha \leftrightarrow \beta) \\ \{\pi_0{}^\gamma, \widetilde{M}_{0\beta}\} &\approx \pi_\beta{}^\gamma + \eta_{\alpha\beta} \tfrac{\partial \mathcal{L}}{\partial b^0{}_{\gamma,\alpha}} - x_\beta \partial_0 \pi_0{}^\gamma + x_0 \partial_\beta \pi_0{}^\gamma. \end{aligned} \tag{6.49}$$

Furthermore, the field equations

$$\partial_0 \pi_0{}^\gamma \approx \partial_0 \frac{\partial \mathcal{L}}{\partial b^0{}_{\gamma,0}} \approx \partial_\alpha \frac{\partial \mathcal{L}}{\partial b^0{}_{\alpha,\gamma}} + \mathcal{O}_4$$

imply the relation

$$x_\beta \partial_0 \pi_0{}^\gamma - \eta_{\alpha\beta} \frac{\partial \mathcal{L}}{\partial b^0{}_{\gamma,\alpha}} = \partial_\alpha \left(x_\beta \frac{\partial \mathcal{L}}{\partial b^0{}_{\alpha,\gamma}} \right) + \mathcal{O}_3$$

which can be used to simplify the last equation in (6.49).

(1) The conservation of energy follows from the same type of arguments as those used in discussing EC theory. Therefore,

$$\frac{\mathrm{d}\widetilde{P}_0}{\mathrm{d}t} \approx \frac{\mathrm{d}E_0}{\mathrm{d}t} \approx 0 \tag{6.50a}$$

and the surface term E_0, representing the value of the energy, is conserved.

(2) The linear momentum has no explicit time dependence, and its PB with $\widetilde{P}_0$ is given by

$$\{\widetilde{P}_0, \widetilde{P}_\alpha\} \approx \int \mathrm{d}^3x\, \{\partial_\gamma \pi_0{}^\gamma, \widetilde{P}_\alpha\}.$$

This expression is easily evaluated using equations (6.49), with the final result

$$\{\widetilde{P}_0, \widetilde{P}_\alpha\} \approx \oint \mathrm{d}S_\gamma\, \partial_\alpha \pi_0{}^\gamma = 0$$

as a consequence of $\partial_\alpha \pi_0{}^\gamma = \mathcal{O}_3$. Therefore, there is no Z term in (6.22) for $G = \widetilde{P}_\alpha$, and we have the conservation law

$$\frac{\mathrm{d}\widetilde{P}_\alpha}{\mathrm{d}t} \approx \frac{\mathrm{d}E_\alpha}{\mathrm{d}t} \approx 0. \tag{6.50b}$$

(3) In a similar way, we can check the conservation of the rotation generator. Using the formulae (6.49), we find:

$$\{\widetilde{P}_0, \widetilde{M}_{\alpha\beta}\} \approx \int \mathrm{d}^3x\, \{\partial_\gamma \pi_0{}^\gamma, \widetilde{M}_{\alpha\beta}\} \approx \oint (x_\alpha\, \mathrm{d}S_\beta - x_\beta\, \mathrm{d}S_\alpha)\partial_\gamma \pi_0{}^\gamma = 0$$

because $\partial_\gamma \pi_0{}^\gamma = \mathcal{O}_4$, according to (6.47). Therefore, equation (6.22) for $G = \widetilde{M}_{\alpha\beta}$ implies

$$\frac{\mathrm{d}\widetilde{M}_{\alpha\beta}}{\mathrm{d}t} \approx \frac{\mathrm{d}E_{\alpha\beta}}{\mathrm{d}t} \approx 0. \tag{6.50c}$$

(4) Finally, the boost generator has an explicit, linear dependence on time and satisfies

$$\frac{\partial \widetilde{M}_{0\beta}}{\partial t} = \widetilde{P}_\beta \approx E_\beta.$$

The evaluation of $\{\widetilde{P}_0, \widetilde{M}_{0\beta}\}$ is done with the help of equations (6.49):

$$\{\widetilde{P}_0, \widetilde{M}_{0\beta}\} \approx \int \mathrm{d}^3x\, \partial_\gamma \left[\pi_\beta{}^\gamma - \partial_\alpha \left(x_\beta \frac{\partial \mathcal{L}}{\partial b^0{}_{\alpha,\gamma}}\right)\right] = \int \mathrm{d}^3x\, \partial_\gamma \pi_\beta{}^\gamma = E_\beta$$

where we used the antisymmetry of $\partial\mathcal{L}/\partial b^0{}_{\alpha,\gamma}$ in (α, γ). These results imply the boost conservation law:

$$\frac{\mathrm{d}\widetilde{M}_{0\beta}}{\mathrm{d}t} \approx \frac{\mathrm{d}E_{0\beta}}{\mathrm{d}t} \approx 0. \tag{6.50d}$$

Thus, all ten Poincaré generators of the general teleparallel theory are conserved.

Comparison with GR. Now we shall transform our results into Lagrangian form, and compare them with GR. All calculations refer to the one-parameter teleparallel theory ($2A + B = 1$, $C = -1$).

We begin by noting that the momentum $\pi_i{}^\gamma$ is defined by the formula

$$\pi_i{}^\gamma = -4bh^{j\gamma}\beta_{ij}{}^0 \tag{6.51a}$$

where $\beta_{ij}{}^0$ can be conveniently written as a sum of two terms:

$$\begin{aligned}\beta_{ij}{}^0 &= (\beta_{ij}{}^0)_{(1)} + (\beta_{ij}{}^0)_{(2)} \\ 4b(\beta_{ij}{}^0)_{(1)} &= ah^{k0}(b_{i\lambda}\nabla_\rho H^{\lambda\rho}_{jk} + b_{j\lambda}\nabla_\rho H^{\lambda\rho}_{ik}) - a\nabla_\rho H^{0\rho}_{ij} \\ 4b(\beta_{ij}{}^0)_{(2)} &= -a(2B-1)bh^{k0}\overset{A}{T}_{ijk}.\end{aligned} \tag{6.51b}$$

The energy–momentum expression (6.42) can be transformed into the Lagrangian form with the help of relation (6.51a) that defines $\pi_i{}^\gamma$:

$$E_\mu = \oint \mathrm{d}S_\gamma\, H_\mu{}^{0\gamma} \qquad H_\mu{}^{0\gamma} \equiv -4bb^i{}_\mu h^{j\gamma}\beta_{ij}{}^0. \tag{6.52}$$

In order to compare this result with GR, we use the decomposition of $f^i{}_\mu \equiv b^i{}_\mu - \delta^i_\mu$ into symmetric and antisymmetric parts, $f_{i\mu} = s_{i\mu} + a_{i\mu}$, apply the asymptotic conditions (6.37)–(6.39) and obtain

$$H_0{}^{0\gamma} = 2a(s_c{}^{c,\gamma} - s_c{}^{\gamma,c}) - 2aa^{c\gamma}{}_{,c} + \mathcal{O}_3.$$

Note that the second term in $\beta_{ij}{}^0$, proportional to $(2B-1)$ and given by equation (6.51b), does not contribute to this result. Now, using the quantity $h^{\mu\nu\lambda}$, defined in the previous section, we easily verify that

$$h^{00\gamma} = 2a(s_c{}^{c,\gamma} - s_c{}^{\gamma,c}) + \mathcal{O}_3.$$

Then, after discarding the inessential divergence of the antisymmetric tensor in $H_0{}^{0\gamma}$, we find the following Lagrangian expression for E_0:

$$E_0 = \oint \mathrm{d}S_\gamma\, h^{00\gamma}. \tag{6.53}$$

Thus, the energy of one-parameter teleparallel theory is given by the *same* expression as in GR.

In a similar manner, we can transform the expression for E_α. Starting with equation (6.51b), we note that the first term in $\beta_{ij}{}^0$, which corresponds to $\mathrm{GR}_\parallel$ ($B = 1/2$), gives the contribution

$$\begin{aligned}(H_\alpha{}^{0\gamma})_{(1)} = {}& 2a[\eta^{\gamma\beta}(s_{\alpha\beta,0} - s_{0\alpha,\beta}) + \delta^\gamma_\alpha(s^c{}_{0,c} - s^c{}_{c,0})] \\ &+ 4a(\delta^{[\gamma}_\alpha a^{\beta]}{}_0)_{,\beta} + \mathcal{O}_3\end{aligned}$$

which, after dropping the irrelevant divergence of the antisymmetric tensor, can be identified with $h_\alpha{}^{0\gamma}$. The contribution of the second term has the form

$$
\begin{aligned}
(H_\alpha{}^{0\gamma})_{(2)} &= a(2B-1)bb^i{}_\alpha h^{j\gamma}h^{k0}\overset{A}{T}_{ijk} \\
&= a(2B-1)\overset{A}{T}_\alpha{}^{\gamma 0} + \mathcal{O}_3
\end{aligned}
$$

where $\overset{A}{T}_{ijk} = T_{ijk} + T_{kij} + T_{jki}$. Thus, the complete linear momentum takes the form

$$
E_\alpha = \eta_{\alpha\mu}\oint \mathrm{d}S_\gamma\,[h^{\mu 0\gamma} - a(2B-1)\overset{A}{T}{}^{\mu 0\gamma}] \tag{6.54}
$$

which is *different* from what we have in GR.

Energy and momentum expressions (6.53) and (6.54) can be written in a Lorentz covariant form as

$$
\begin{gathered}
E^\mu = \oint \mathrm{d}S_\gamma\,\bar{h}^{\mu 0\gamma} = \int \mathrm{d}^3x\,\bar{\theta}^{\mu 0} \\
\bar{\theta}^{\mu\nu} \equiv \bar{h}^{\mu\nu\rho}{}_{,\rho} \qquad \bar{h}^{\mu\nu\rho} \equiv h^{\mu\nu\rho} - a(2B-1)\overset{A}{T}{}^{\mu\nu\rho}.
\end{gathered} \tag{6.55}
$$

In the case $2B - 1 = 0$, corresponding to $\mathrm{GR}_\parallel$, we see that $\bar{\theta}^{\mu\nu}$ coincides with $\theta^{\mu\nu}$, the energy–momentum complex of GR. When $2B - 1 \neq 0$, the momentum acquires a *correction* proportional to the totally antisymmetric part of the torsion.

The elimination of momenta from equation (6.48) leads to the following Lagrangian form of the angular momentum:

$$
E_{\mu\nu} = \oint \mathrm{d}S_\gamma\,(x_\mu H_\nu{}^{0\gamma} - x_\nu H_\mu{}^{0\gamma} - 4\lambda_{\mu\nu}{}^{0\gamma}) \tag{6.56a}
$$

where we have used expression (6.51a) for $\pi_k{}^\alpha$, and $\pi_{\mu\nu}{}^\gamma \approx 4\delta^i_\mu\delta^j_\nu\lambda_{ij}{}^{0\gamma}$.

In order to compare this result with GR, we should first eliminate λ. Omitting technical details, we display here an equivalent form of the complete angular momentum, suitable for comparisson with GR:

$$
\begin{aligned}
E^{\mu\nu} = \int \mathrm{d}^3x\,\{&(x^\mu\bar{\theta}^{\nu 0} - x^\nu\bar{\theta}^{\mu 0}) + 2a(2B-1)\overset{A}{T}{}^{\nu\mu 0} \\
&+ \eta^{i\mu}\eta^{j\nu}[2a(2B-1)bh^{k0}\overset{A}{T}_{ijk} - \sigma^0{}_{ij}]\}.
\end{aligned} \tag{6.56b}
$$

For $2B - 1 = 0$, we obtain the angular momentum of $\mathrm{GR}_\parallel$, which reduces to the GR form if $\sigma^0{}_{ij} = 0$. The terms proportional to $2B - 1$ and $\sigma^0{}_{ij}$ give a *correction* to the GR result.

6.4 Chern–Simons gauge theory in $D = 3$

The standard dynamics of gauge theories is determined by the Lagrangian (A.11), which is quadratic in the field strength. Another possibility is defined by the

Chern–Simons action; it became particularly interesting in the 1980s, after it was realized that Einstein's GR in *three dimensions* can be interpreted as a Chern–Simons gauge theory (Witten 1988; see also appendix L). Although pure Chern–Simons theory in $D = 3$ has no genuine propagating modes, it gives rise to unusual dynamical phenomena at the boundary, characterized by a two-dimensional conformal symmetry (Bañados 1994, 1999a, b).

Chern–Simons action

To construct the Chern–Simons action in $D = 3$, we start from a gauge theory on a four-dimensional manifold $\mathcal{M}_4$, with an action of the form (wedge product between forms is understood):

$$\begin{aligned} I_P &= \int p\,\mathrm{Tr}(FF) = \int_{\mathcal{M}_4} F^a F^b \gamma_{ab} \\ &= \tfrac{1}{4}\int_{\mathcal{M}_4} \mathrm{d}^4x\, \varepsilon^{\mu\nu\lambda\rho} F^a{}_{\mu\nu} F^b{}_{\lambda\rho}\gamma_{ab}. \end{aligned}$$

Here, $F = \mathrm{d}A + AA = T_a F^a$ is the field strength 2-form, $A = T_a A^a$ is the gauge potential 1-form ($A^a = A^a{}_\mu\, \mathrm{d}x^\mu$, $F^a = \frac{1}{2}F^a{}_{\mu\nu}\, \mathrm{d}x^\mu\, \mathrm{d}x^\nu$), T_a is a basis of the Lie algebra of the gauge group G, $\gamma_{ab} = p\,\mathrm{Tr}(T_a T_b)$ represents an invariant bilinear form (metric) on the Lie algebra, and p is a normalization factor depending on the representation. We assume that γ_{ab} is *non-degenerate*, so that the action contains a kinetic term for all the gauge fields. This is always true for semisimple Lie algebras, but may be also fulfilled in some other cases. The action I_P is a topological invariant. Denoting the integrand of I_P by P (the Pontryagin form), we have

$$\begin{aligned} P &= (\mathrm{d}A^a + \tfrac{1}{2}f_{cd}{}^a A^c A^d)(\mathrm{d}A^b + \tfrac{1}{2}f_{ef}{}^b A^e A^f)\gamma_{ab} \\ &= (\mathrm{d}A^a\, \mathrm{d}A^b + f_{ef}{}^a A^e A^f\, \mathrm{d}A^b)\gamma_{ab}. \end{aligned}$$

The last equality is obtained by noting that the product of the four As vanishes on account of the Jacobi identity for the structure constants $f_{ab}{}^c$. It is now easy to see that P is a total divergence,

$$P = \mathrm{d}L_{\mathrm{CS}} \qquad L_{\mathrm{CS}} = (A^a\, \mathrm{d}A^b + \tfrac{1}{3}f_{ce}{}^a A^c A^e A^b)\gamma_{ab}$$

where L_{CS} is the Chern–Simons Lagrangian. Applying Stokes's theorem to the action I_P we can transform it to an integral of L_{CS} over the boundary of $\mathcal{M}_4$. The Chern–Simons action is defined by replacing $\partial\mathcal{M}_4$ with an arbitrary three-dimensional manifold $\mathcal{M}$:

$$I_{\mathrm{CS}} = \int_{\mathcal{M}} L_{\mathrm{CS}} = \int_{\mathcal{M}} \mathrm{d}^3x\, \varepsilon^{\mu\nu\rho}(A^a{}_\mu \partial_\nu A^b{}_\rho + \tfrac{1}{3}f_{ce}{}^a A^c{}_\mu A^e{}_\nu A^b{}_\rho)\gamma_{ab}. \qquad (6.57)$$

Example 1. For $G = SO(1,2)$, we choose the generators in the fundamental representation to be $T_0 = \frac{1}{2}\mathrm{i}\sigma_2$, $T_1 = \frac{1}{2}\sigma_3$, $T_2 = \frac{1}{2}\sigma_1$ (σ_a are the Pauli matrices), i.e.

$$T_0 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad T_1 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad T_2 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

With the normalization $p = -2$, the metric is given by

$$\gamma_{ab} = -2\,\mathrm{Tr}(T_a T_b) = \eta_{ab} \qquad \eta_{ab} \equiv (+1, -1, -1)$$

and the form of Lie algebra, $[T_a, T_b] = \varepsilon_{ab}{}^c T_c$, implies $f_{ab}{}^c = \varepsilon_{ab}{}^c$ ($\varepsilon_{012} = 1$). In these conventions, the gauge potential can be represented as

$$A = \frac{1}{2}\begin{pmatrix} A^{(1)} & A^{(+)} \\ -A^{(-)} & -A^{(1)} \end{pmatrix} \qquad A^{(\pm)} = A^{(0)} \pm A^{(2)}.$$

Gauge invariance. Writing the finite gauge transformations in the form

$$A'_\mu = g^{-1}(A_\mu + \partial_\mu)g \qquad F'_{\mu\nu} = g^{-1}F_{\mu\nu}g \tag{6.58a}$$

where $g = g(u) \in G$, we can easily show that the Pontryagin form P is gauge invariant (using the cyclic property of the trace). Since $P = \mathrm{d}L_{\mathrm{CS}}$, $\mathrm{d}L_{\mathrm{CS}}$ is also gauge invariant, $\delta_0\,\mathrm{d}L_{\mathrm{CS}} = 0$; this implies $\mathrm{d}\delta_0\,L_{\mathrm{CS}} = 0$, and consequently, $\delta_0 L_{\mathrm{CS}} = \mathrm{d}v$. Thus, the Chern–Simons action changes by a term $\delta_0 \int L_{\mathrm{CS}} = \int \mathrm{d}v$, which does not alter the field equations (in particular, this term vanishes for $\partial\mathcal{M} = 0$). Explicit calculation leads to the result

$$I_{\mathrm{CS}}[A'] = I_{\mathrm{CS}}[A] + \int_{\partial\mathcal{M}} p\,\mathrm{Tr}(g^{-1}\,\mathrm{d}g\,A) - \tfrac{1}{3}\int_{\mathcal{M}} p\,\mathrm{Tr}(g^{-1}\,\mathrm{d}g)^3.$$

Since in any variational problem the fields have to satisfy appropriate boundary conditions, which necessarily restrict the form of the allowed gauge transformations, the second term can be made to vanish by demanding that $g \to 1$ sufficiently fast when we approach the boundary of $\mathcal{M}$. The value of the last term depends on the topological properties of the mapping $g : \mathcal{M} \to G$. If we restrict ourselves to topologically trivial gauge transformations (such as the infinitesimal gauge transformations, for instance), the last term also vanishes. Thus, the Chern–Simons action is gauge invariant if (Zanelli 2000)

- $g \to 1$ sufficiently fast when we approach the boundary of $\mathcal{M}$; and
- gauge transformations are topologically trivial.

These conditions are often fulfilled in practical applications. In particular, I_{CS} is invariant under the infinitesimal gauge transformations:

$$\delta_0(u)A^a{}_\mu = -\nabla_\mu u^a = -(\partial_\mu u^a + f_{bc}{}^a A^b{}_\mu u^c). \tag{6.58b}$$

The Chern–Simons action is also invariant under the diffeomorphisms of $\mathcal{M}$, but this is not an independent symmetry, as we shall see.

Field equations. The variation of the action (6.57) yields

$$\delta I_{\mathrm{CS}} = \int_{\mathcal{M}} \delta A^a F^b \gamma_{ab} + \int_{\partial\mathcal{M}} A^a \delta A^b \gamma_{ab}.$$

If δA^a is an arbitrary variation on $\mathcal{M}$, and its value on the boundary is chosen so that the second term vanishes, we obtain the field equations

$$F^a{}_{\mu\nu} = 0 \tag{6.59a}$$

provided γ_{ab} is non-degenerate. The field equations imply, at least locally (on some open sets in $\mathcal{M}$), that A_μ is a gauge transform of the trivial field configuration $A_\mu = 0$:

$$A_\mu = g^{-1}\partial_\mu g \qquad \text{(locally).} \tag{6.59b}$$

Thus, there are no truly propagating degrees of freedom in this theory. The whole dynamics is contained in the non-trivial topology, which prevents the relation (6.59*b*) to be true on the entire manifold $\mathcal{M}$.

The action and boundary conditions. Although the Chern–Simons theory has no local excitations, its dynamical content is by no means trivial. We shall begin the analysis of the Chern–Simons dynamics by choosing appropriate boundary conditions. Then, using the canonical approach, we shall find an infinite number of degrees of freedom living at the boundary.

Boundary conditions are necessary in order to have a well-defined variation of the action, but they are not unique. Dynamical properties of the theory are very sensitive to the choice of boundary conditions. Here, we are interested in Chern–Simons theory on a manifold $\mathcal{M}$ having the topology $R \times \Sigma$, where R is interpreted as time, and Σ is a spatial manifold with a boundary $\partial\Sigma$ that is topologically a circle (which may be located at infinity). The simple examples for Σ are manifolds with the topology of a disc or R^2. The theory may have interesting topological properties related to the existence of holes in Σ, but these effects will not be discussed here. We assume that there exist coordinates $x^\mu = (t, r, \varphi)$ $(\mu = 0, 1, 2)$ on $\mathcal{M}$ such that the boundary $\partial\Sigma$ is described by the standard angular coordinate φ. Then, the boundary term in δI_{CS} takes the form

$$\delta B_{\mathrm{CS}} = \int_{\partial\mathcal{M}} A^a \delta A_a = \int \mathrm{d}t\, \mathrm{d}\varphi\, (A^a{}_t \delta A_{a\varphi} - A^a{}_\varphi \delta A_{at}).$$

A simple way to cancel δB_{CS} is by imposing $A_t = 0$ at $\partial\Sigma$. In this case the symmetry of the theory (the subgroup of gauge transformations which leave the boundary conditions invariant) is given by the gauge transformations that are independent of time at the boundary. Another interesting possibility is to choose one of the following two conditions:

$$A_t = -A_\varphi \qquad \text{or} \qquad A_t = A_\varphi \qquad \text{(at the boundary).} \tag{6.60a}$$

In this case, the gauge parameters at the boundary can depend on time, but this dependence is not arbitrary.

Boundary conditions can be thought of as a mechanism for selecting a general class of field configurations which have a specific dynamical importance. Conditions (6.60a) include some very interesting solutions known from $D = 3$ gravity; they will be adopted as a typical choice in our subsequent discussion.

Thus, our intention is to study the Chern–Simons theory defined by the action (6.57) supplemented with the boundary conditions (6.60a), which produces the field equations (6.59a). We should note that there exist boundary conditions for which $\delta B_{\mathrm{CS}} \neq 0$; the related variational problem can be improved by adding an extra boundary term to the action.

Residual gauge symmetries. Our next step is to determine the most general set of gauge transformations that preserve the boundary conditions (6.60a). Introducing the light-cone basis notation, $x^{\pm} = t \pm \varphi$, the boundary conditions can be written in the form

$$A_+ = 0 \qquad \text{or} \qquad A_- = 0 \qquad \text{(at the boundary).} \tag{6.60b}$$

The invariance of the first condition, $\delta_0 A_+ = -\nabla_+ u = 0$, implies $\partial_+ u = 0$ at $\partial\Sigma$. Thus, the residual gauge symmetry corresponding to the boundary condition $A_+ = 0$ is defined by the gauge transformations the parameters of which are chiral, i.e. $u = u(x^-)$, at the boundary. The resulting symmetry is infinite dimensional. Similarly, the second condition $A_- = 0$ is preserved if the gauge parameters are antichiral, $u = u(x^+)$, at $\partial\Sigma$. In both cases, the gauge parameters are functions of one coordinate (x^- or x^+) at the boundary. This fact will have a significant influence on the structure of the canonical gauge generators.

Since Chern–Simons theory has no propagating modes, the non-trivial dynamics occurs only at the boundary (holes in Σ are ignored). In this context, the residual symmetry at the boundary is usually called global (or asymptotic) symmetry. Its content can be analysed most clearly by fixing the gauge, as will be done in the canonical approach.

Canonical analysis

In order to precisely count the dynamical degrees of freedom in the Chern–Simons theory, we turn to Hamiltonian analysis.

We write the Chern–Simons action in the form

$$I[A] = k \int_{\mathcal{M}} \mathrm{d}^3x\, \varepsilon^{\mu\nu\rho} (A^a{}_\mu \partial_\nu A^b{}_\rho + \tfrac{1}{3} f_{ef}{}^a A^e{}_\mu A^f{}_\nu A^b{}_\rho) \gamma_{ab} \tag{6.61}$$

where k is a normalization constant, and $\varepsilon^{012} = 1$. The action is invariant under the usual gauge transformations (6.58b) and the diffeomorphisms will be discussed later.

The Hamiltonian and constraints. The relations defining $\pi_a{}^\rho$ lead to the primary constraints

$$\phi_a{}^0 \equiv \pi_a{}^0 \approx 0 \qquad \phi_a{}^\alpha \equiv \pi_a{}^\alpha - k\varepsilon^{0\alpha\beta} A_{a\beta} \approx 0.$$

Since the Lagrangian is linear in velocities, we easily find first the canonical, and then the total Hamiltonian:

$$\begin{aligned} \mathcal{H}_\mathrm{T} &= A^a{}_0 \mathcal{H}_a + u_{a0}\phi^{a0} + u_{a\alpha}\phi^{a\alpha} + \partial_\alpha D^\alpha \\ \mathcal{H}_a &= -k\varepsilon^{0\alpha\beta} F_{a\alpha\beta} \\ D^\alpha &= k\varepsilon^{0\alpha\beta} A^a{}_0 A_{a\beta}. \end{aligned}$$

The consistency conditions of the primary constraints have the form

$$\begin{aligned} \dot\phi_a{}^0 \approx 0 \quad &\Rightarrow \quad \mathcal{H}_a \approx 0 \\ \dot\phi_a{}^\alpha \approx 0 \quad &\Rightarrow \quad \nabla_\beta A^a{}_0 - u^a{}_\beta \approx 0. \end{aligned}$$

The first relation yields a secondary constraint and the second determines $u^a{}_\beta$. Taking into account that $\dot A^a{}_\beta \approx u^a{}_\beta$, the second relation can be written in the form $F^a{}_{0\beta} \approx 0$.

Replacing this expression for $u_{a\beta}$ into $\mathcal{H}_\mathrm{T}$, we obtain

$$\begin{aligned} \mathcal{H}_\mathrm{T} &= A^a{}_0 \overline{\mathcal{H}}_a + u_{a0}\phi^{a0} + \partial_\alpha \bar D^\alpha \\ \overline{\mathcal{H}}_a &= \mathcal{H}_a - \nabla_\beta \phi_a{}^\beta = -k\varepsilon^{0\alpha\beta}\partial_\alpha A_{a\beta} - \nabla_\alpha \pi_a{}^\alpha \\ \bar D^\alpha &= A^a{}_0 \pi_a{}^\alpha. \end{aligned} \tag{6.62}$$

Before going on with the consistency analysis, we display here the algebra of constraints:

$$\begin{aligned} \{\phi_a{}^\alpha, \phi_b'{}^\beta\} &= -2k\gamma_{ab}\varepsilon^{0\alpha\beta}\delta \\ \{\phi_a{}^\alpha, \overline{\mathcal{H}}_b'\} &= -f_{abc}\phi^{c\alpha}\delta \\ \{\overline{\mathcal{H}}_a, \overline{\mathcal{H}}_b'\} &= -f_{abc}\overline{\mathcal{H}}^c\delta \end{aligned}$$

while all PBs involving ϕ^{a0} vanish. It is now clear that further consistency requirements produce nothing new.

Thus, the theory is defined by the total Hamiltonian $\mathcal{H}_\mathrm{T}$ and the following set of constraints:

- first class: $\pi_a{}^0, \overline{\mathcal{H}}_b$; and
- second class: $\phi_a{}^\alpha$.

While the FC constraints generate gauge transformations, the second-class constraints $\phi_a{}^\alpha$ can be used to restrict the phase space by eliminating the momenta $\pi_a{}^\alpha$ from the theory. The construction of the preliminary Dirac brackets is straightforward, and this leads to

$$\{A^a{}_\alpha, A^b{}_\beta\}^\bullet = \frac{1}{2k}\varepsilon_{0\alpha\beta}\gamma^{ab} \tag{6.63}$$

as the only non-trivial result in the restricted phase space $(A^a{}_\alpha, A^b{}_0, \pi_c{}^0)$.

Gauge generator. We now apply Castellani's algorithm to construct the gauge generator:

$$G[\varepsilon] = \int_\Sigma \mathrm{d}^2x \, [(\nabla_0 \varepsilon^a)\pi_a{}^0 + \varepsilon^a \overline{\mathcal{H}}_a]. \tag{6.64}$$

It produces the following gauge transformations on the phase space:

$$\begin{array}{cc} \delta_0 A^a{}_0 = \nabla_0 \varepsilon^a & \delta_0 A^a{}_\alpha = \nabla_\alpha \varepsilon^a \\ \delta_o \pi_a{}^0 = -f_{abc}\varepsilon^b \pi^{c0} & \delta_o \pi_a{}^\alpha = k\varepsilon^{0\alpha\beta}\partial_\beta \varepsilon_a - f_{abc}\varepsilon^b \pi^{c\alpha}. \end{array}$$

We can explicitly verify that the equations of motion are invariant under these transformations. In particular, $\delta_0 \phi_a{}^\alpha = -f_{abc}\varepsilon^b \phi^{c\alpha} \approx 0$.

In this construction, the standard gauge transformations are correctly reproduced, but the diffeomorphisms are not found. However, by introducing the new gauge parameters according to

$$\varepsilon^a = \xi^\rho A^a{}_\rho \tag{6.65a}$$

we easily obtain

$$\delta_0(\xi) A^a{}_\mu = \xi^\rho{}_{,\mu} A^a{}_\rho - \xi \cdot \partial A^a{}_\mu + \xi^\rho F^a{}_{\mu\rho}. \tag{6.65b}$$

Thus, the diffeomorphisms are *not an independent symmetry*; they are contained in the ordinary gauge group as an on-shell symmetry (Witten 1988).

Fixing the gauge. Since there are no local degrees of freedom in the Chern–Simons theory, the only relevant symmetries are those associated with the boundary dynamics. Their content becomes clearer if we fix the gauge, whereupon only the degrees of freedom at the boundary are left.

We have found two sets of FC constraints, $\pi_a{}^0$ and $\overline{\mathcal{H}}_a$, hence we are free to impose two sets of gauge conditions. A simple and natural choice for the first gauge condition is:

$$A^a{}_0 = -A^a{}_2 \qquad \text{or} \qquad A^a{}_0 = A^a{}_2 \tag{6.66}$$

whereby our choice of the boundary conditions (6.60) is extended to the whole spacetime, without affecting the physical content of the theory. After this step has been taken, we can construct the related preliminary Dirac brackets and eliminate the variables $(A^a{}_0, \pi_a{}^0)$ from the theory; the PBs of the remaining variables remain unchanged.

An appropriate choice for the second gauge condition is obtained by restricting A_r to be a function of the radial coordinate only:

$$A_r = b^{-1}(r)\partial_r b(r) \tag{6.67}$$

where $b(r)$ is an element of G. By a suitable choice of the radial coordinate, we can write

$$b(r) = \mathrm{e}^{r\alpha} \quad \Rightarrow \quad A_r = \alpha \tag{6.68a}$$

where $\alpha = \alpha^a T_a$ is a constant element of the Lie algebra. The condition (6.67), together with the constraint $F_{r\varphi} \approx 0$, implies that the group element in $A_\alpha = g^{-1}\partial_\alpha g$ can be factorized as $g = h(t,\varphi)b(r)$, which leads to

$$A_\varphi = b^{-1}\hat{A}_\varphi(t,\varphi)b \tag{6.68b}$$

where $\hat{A}_\varphi = h^{-1}\partial_\varphi h$. Demanding that the gauge condition be preserved in time, we obtain the consistency condition $\nabla_r(\partial_\varphi A_0) \approx 0$. The general solution of this equation, compatible with $F = 0$, is given by $A_0 = b^{-1}\hat{A}_0(t,\varphi)b$. In these expressions, $\hat{A}_\varphi$ and $\hat{A}_0$ are arbitrary functions of (t,φ).

Now, we can combine (6.67) with one of the gauge conditions (6.66), say $A_+ = 0$, to eliminate the arbitrary function appearing in A_0: $\hat{A}_0(t,\varphi) = -\hat{A}_\varphi(t,\varphi)$. Moreover, the field equations $F = 0$ imply that $\hat{A}_\varphi = \hat{A}_\varphi(x^-)$. Thus, the canonical gauge conditions do not fix the gauge completely. We are, instead, left with one arbitrary function at the boundary,

$$\hat{A}_-(x^-) = \tfrac{1}{2}[\hat{A}_0(x^-) - \hat{A}_\varphi(x^-)] = -\hat{A}_\varphi(x^-)$$

which is related to the chiral residual gauge symmetry, defined by $g_u = g_u(x^-)$. The action of g_u does not change A_r, while $\hat{A}_-(x^-)$ transforms in the usual way:

$$\hat{A}_-(x^-) \to \hat{A}'(x^-) = g_u^{-1}[\hat{A}_-(x^-) + \partial_-]g_u.$$

Analogous considerations are valid for the second, antichiral gauge choice in (6.66).

The residual gauge symmetry at the boundary is defined by a chiral/antichiral gauge parameter $u(x^\mp)$; it is an infinite dimensional symmetry, the structure of which will be determined by studying the PB algebra of the related generators.

The improved generators. We now turn our attention to the question of the functional differentiability of the gauge generators. After adopting one of the gauge conditions (6.66), the effective form of the gauge generator (6.64) is given by

$$\bar{G}[\varepsilon] = \int_\Sigma \mathrm{d}^2x\, \varepsilon^a \overline{\mathcal{H}}_a. \tag{6.69}$$

The generator acts on the restricted phase space $(A^a{}_\alpha, \pi_a{}^\alpha)$ with the usual PB operation. For the gauge parameters that are independent of field derivatives, the variation of $\bar{G}$ takes the form

$$\delta\bar{G}[\varepsilon] = \int \mathrm{d}^2x\, \varepsilon^a \delta\overline{\mathcal{H}}_a + R = -\delta E[\varepsilon] + R \tag{6.70a}$$

$$\delta E[\varepsilon] = \oint \mathrm{d}S_\alpha\, \varepsilon^a(k\varepsilon^{0\alpha\beta}\delta A_{a\beta} + \delta\pi_a{}^\alpha) \tag{6.70b}$$

where $\mathrm{d}S_\alpha \varepsilon^{0\alpha\beta} = \mathrm{d}x^\beta$. Therefore, if we can integrate δE to find $E[\varepsilon]$, the improved gauge generator

$$\widetilde{G}[\varepsilon] = \bar{G}[\varepsilon] + E[\varepsilon] \tag{6.71a}$$

will have well-defined functional derivatives:

$$\frac{\delta\widetilde{G}[\tau]}{\delta A_{a\beta}} = k\varepsilon^{0\alpha\beta}\nabla_\alpha\tau^a - \tau_e f^{eaf}\phi_f{}^\beta + \frac{\partial\tau^c}{\partial A_{a\beta}}\overline{\mathcal{H}}_c$$
$$\frac{\delta\widetilde{G}[\tau]}{\delta\pi^{a\beta}} = \nabla_\beta\tau_a. \tag{6.71b}$$

Consequently, the PB of two $\widetilde{G}$s can be written in the form

$$\{\widetilde{G}[\tau], \widetilde{G}[\lambda]\} = W_1 + W_2 + W_3 \tag{6.72a}$$

where

$$W_1 = 2k\int \mathrm{d}^2x\, \varepsilon^{0\alpha\beta}\nabla_\alpha\tau^a\nabla_\beta\lambda_a$$
$$W_2 = -\int \mathrm{d}^2x\, \phi_c{}^\beta\nabla_\beta(f^{cef}\tau_e\lambda_f) \tag{6.72b}$$
$$W_3 = \int \mathrm{d}^2x\, \left(\frac{\partial\tau^c}{\partial A_{a\beta}}\nabla_\beta\lambda_a\right)\overline{\mathcal{H}}_c - (\tau \leftrightarrow \lambda).$$

We expect that the complete improved generator $\widetilde{G} = \bar{G} + E$ will appear on the right-hand side of the PB algebra (6.72*a*), but to show that we need to know the explicit form of the surface term E, defined by (6.70*b*). This will be possible only after we impose some additional conditions on the behaviour of gauge parameters at the boundary.

In what follows, we shall simplify the calculation of the right-hand side in (6.72*a*) by using the second-class constraint $\phi_a{}^\alpha \approx 0$, which leads to $W_2 \approx 0$. The structure of the PB algebra can be further simplified by imposing the second gauge condition (6.67).

Symmetries at the boundary

The affine symmetry. We begin this section by studying the simplest boundary condition which allows us to find an explicit form of the surface term $E[\tau]$. We assume that

- gauge parameters are independent of the fields at the boundary.

This choice describes a residual gauge symmetry on the boundary stemming from the standard gauge symmetry, in which the parameters $\varepsilon^a = \tau^a$ are field independent on the whole spacetime. In that case, the expression $\delta E[\tau]$ in (6.70*b*) is easily integrated to give

$$E[\tau] = \oint \mathrm{d}S_\alpha\, \tau^a(2k\varepsilon^{0\alpha\beta}A_{a\beta} + \phi_a{}^\alpha) \tag{6.73a}$$

and the improved gauge generator takes the form

$$\widetilde{G}[\tau] = \int \mathrm{d}^2x\, [\partial_\alpha \tau^a (k\varepsilon^{0\alpha\beta} A_{a\beta} + \pi_a{}^\alpha) - \tau^c f_{cef} A^e{}_\alpha \pi^{f\alpha}]. \qquad (6.73b)$$

The functional derivatives are easily checked to have the form (6.71*b*) with $W_3 = 0$, since τ^c does not depend on the fields.

Let us now return to the PB algebra (6.72) with $W_2 \approx 0$ and $W_3 = 0$. Integrating by parts in W_1 we obtain

$$W_1 = -\int \mathrm{d}^2x\, (f^{aef}\tau_e\lambda_f)\mathcal{H}_a + 2k \oint \mathrm{d}x^\beta\, \tau^a \nabla_\beta \lambda_a.$$

After introducing $\sigma^a \equiv f^{aef}\tau_e\lambda_f$, the first term is easily recognized to have the expected form $-\bar{G}[\sigma]$ (weakly). In order to find the surface term $E[\sigma]$, we use the relation

$$\tau^a \nabla_\beta \lambda_a = -\sigma^a A_{a\beta} + \tau^a \partial_\beta \lambda_a$$

to transform the second term in W_1. This leads directly to

$$\{\widetilde{G}[\tau], \widetilde{G}[\lambda]\} \approx -\widetilde{G}[\sigma] + K[\tau, \lambda] \qquad (6.74a)$$

where K is a field independent term, called the *central charge*:

$$K[\tau, \lambda] \equiv 2k \oint \mathrm{d}x^\beta\, \tau^a \partial_\beta \lambda_a. \qquad (6.74b)$$

The occurrence of central charges is an unusual feature of the *classical* PB algebra; it is a consequence of the fact that $\widetilde{G}$ is an improved generator, which differs from a linear combination of FC constraints by a boundary term.

We shall now fix the gauge in order to have a simpler description of the degrees of freedom at the boundary. The chiral gauge condition is already used to eliminate $(A^a{}_0, \pi_a{}^0)$, and the second gauge condition (6.67) ensures $F^a{}_{r\varphi} \approx 0$. Hence, we have $\mathcal{H}_a \approx 0$, and the whole generator $\widetilde{G}[\tau]$ reduces just to the boundary term:

$$\widetilde{G}[\tau] \approx E[\tau] \approx 2k \oint \mathrm{d}x^\beta\, \tau^a A_{a\beta}.$$

The field equations $F = 0$ together with the gauge conditions (6.66) and (6.67) ensure that τ^a and A_a are functions of x^- or x^+ only. Thus, if we use the Dirac brackets defined by

- the second-class constraints $\phi_a{}^\alpha$ and
- the gauge condition (6.66) and (6.67)

the PB algebra of the gauge fixed generators can be written in the form

$$\{E[\tau], E[\lambda]\}^* = -E[\sigma] + K[\tau, \lambda]. \qquad (6.75a)$$

It is interesting to observe that the Dirac bracket $\{E[\tau], E[\lambda]\}^*$ can be interpreted as $\delta_\lambda E[\tau]$. Indeed, we find that $\delta_\lambda E[\tau] = 2k \oint \mathrm{d}x^\beta\, \tau^a \nabla_\beta \lambda_a$, in complete agreement with the right-hand side of equation $(6.75a)$.

The Lie algebra $(6.75a)$ can be given a more familiar form by representing it in terms of the Fourier modes. Since the boundary of Σ is a circle, we have

$$E[\tau] = 2k \int \mathrm{d}\varphi\, \tau^a(t,\varphi)\hat{A}_a(t,\varphi)$$

where $\hat{A} \equiv \hat{A}_\varphi$. If we now Fourier decompose $\hat{A}^a$ and τ^a at the boundary,

$$\hat{A}^a(t,\varphi) = -\frac{1}{4\pi k}\sum A^a_n \mathrm{e}^{\mathrm{i}n\varphi} \qquad \tau^a(t,\varphi) = \sum \tau^a_m \mathrm{e}^{\mathrm{i}m\varphi}$$

we find $E[\tau] = -\sum \tau^a_{-n} A_{an}$, and the Fourier modes A^a_n satisfy the affine algebra:

$$\{A^a_n, A^b_m\}^* = f^{ab}{}_c A^c_{n+m} + 4\pi k \mathrm{i}\gamma^{ab} n \delta_{n+m,0}. \qquad (6.75b)$$

Thus, if we adopt the boundary condition that the gauge parameters do not depend on the fields, the residual gauge symmetry is described by an infinite set of charges, which satisfy the affine extension of the original PB algebra.

The Virasoro symmetry. In Chern–Simons theory, the diffeomorphisms are contained in ordinary gauge transformations as an on-shell symmetry. We now wish to study boundary conditions that correspond to the freedom of making diffeomorphisms at the boundary. This is achieved by assuming that

- gauge parameters have the form $\varepsilon^a = \xi^\mu A^a{}_\mu$ at the boundary.

After imposing the chiral gauge (6.66), the expression $(6.70b)$ for $\delta E[\varepsilon]$ can be written as

$$\delta E[\xi] \approx 2k \oint \mathrm{d}x^\beta\, \varepsilon^a \delta A_{a\beta} = 2k \int \mathrm{d}\varphi\, [\xi^r A^a{}_r - \xi^- A^a{}_\varphi]\delta A_{a\varphi}.$$

The integration of δE can be easily performed if we have $\delta A^a{}_r = 0$ at the boundary. This property is ensured by the second gauge condition (6.67), which asserts that A_r can be adjusted so as to become a constant element of the Lie algebra: $A^a{}_r = \alpha^a$. Thus, the integration of δE leads to

$$E[\xi] = 2k \int \mathrm{d}\varphi\, [\xi^r \alpha \cdot \hat{A}(t,\varphi) - \tfrac{1}{2}\xi^- \hat{A}^2(t,\varphi)] \qquad (6.76)$$

where the integration constant is set to zero.

Returning now to the calculation of the PB algebra $(6.72a)$, we introduce the notation $\tau^a = \xi^\mu A^a{}_\mu$, $\lambda^a = \eta^\mu A^a{}_\mu$, and note that the term $f_{aef}\tau^e\lambda^f A^a{}_\alpha$ in W_1 vanishes. In order to simplify the calculation,

– we use $\phi_a{}^\alpha \approx 0$ and
– impose the gauge conditions (6.66) and (6.67).

This implies $W_2 \approx 0 \approx W_3$, so that the only non-vanishing contribution comes from W_1:

$$\{\widetilde{G}[\tau], \widetilde{G}[\lambda]\} \approx 2k \int \mathrm{d}\varphi\, \tau^a \partial_\varphi \lambda_a. \tag{6.77a}$$

The boundary term (6.76) is discovered with the help of the identity

$$\int \mathrm{d}\varphi\, \tau^a \partial_\varphi \lambda_a = \int \mathrm{d}\varphi\, (\sigma^r \hat{A} \cdot \alpha - \tfrac{1}{2}\sigma^- \hat{A}^2) + \alpha^2 \int \mathrm{d}\varphi\, \xi^r \partial_\varphi \eta^r$$

where $\sigma^\mu = \xi \cdot \partial \eta^\mu - \eta \cdot \partial \xi^\mu$, so that the final result can be written as

$$\{E[\xi], E[\eta]\}^* = E[\sigma] + 2k\alpha^2 \int \mathrm{d}\varphi\, \xi^r \partial_\varphi \eta^r. \tag{6.77b}$$

Geometric interpretation of the Chern–Simons theory motivates us to consider some specific solutions for $\hat{A}$. In the case $A_+ = 0$, these solutions are defined by the requirements $\hat{A}^{(-)} = 1$, $\hat{A}^{(1)} = 0$ [while for $A_- = 0$, we have $\hat{A}^{(+)} = -1$, $\hat{A}^{(1)} = 0$] (Coussaert *et al* 1995). Gauge invariance of these conditions restricts the values of the gauge parameters in the following way:

$$(\xi^r, \xi^+) = (-\beta \partial_\varphi \xi, \xi) \qquad (\eta^r, \eta^+) = (-\beta \partial_\varphi \eta, \eta)$$

where β can be normalized so that $\beta^2 = 1$, and ξ and η are arbitrary functions of x^- (or x^+). In that case,

$$E[\xi] = 2k \int \mathrm{d}\varphi\, \xi (\beta \alpha^a \partial_\varphi \hat{A}_a - \tfrac{1}{2}\hat{A}^2) \tag{6.78}$$

$$\{E[\xi], E[\eta]\}^* = E[\sigma] + 2k\alpha^2 \int \mathrm{d}\varphi\, (\partial_\varphi \xi\, \partial_\varphi^2 \eta). \tag{6.79}$$

Then, introducing the notation

$$-\frac{1}{4\pi k} L(t, \varphi) = \beta \alpha^a \partial_\varphi \hat{A}_a - \tfrac{1}{2}\hat{A}^2$$

we can write

$$E[\xi] = -\frac{1}{2\pi} \int \mathrm{d}\varphi\, \xi(t, \varphi) L(t, \varphi) = -\sum \xi_{-n} L_n$$

where $L(t, \varphi) = \sum L_n \mathrm{e}^{\mathrm{i}n\varphi}$, $\xi(t, \varphi) = \sum \xi_n \mathrm{e}^{\mathrm{i}n\varphi}$. Written in terms of the Fourier modes, this algebra takes the form

$$\{L_n, L_m\}^* = -\mathrm{i}(n - m) L_{n+m} + 4\pi k \alpha^2 \mathrm{i} n^3 \delta_{n+m,0}. \tag{6.80a}$$

After shifting the zero mode of L_n, $L_n \to L_n + 2\pi k\alpha^2\delta_{n,0}$, we find the standard form of the Virasoro algebra with a *classical* central charge. In the standard string theory normalization for the central charge, we have

$$c_0 = -12 \cdot 4\pi k\alpha^2. \tag{6.80b}$$

Using the known Fourier expansion for $\hat{A}(t,\varphi)$, we obtain

$$L_n = \frac{1}{8\pi k}\sum_m \gamma_{ab}A^a_m A^b_{n-m} + \mathrm{i}n\gamma_{ab}\beta\alpha^a A^b_n \tag{6.81}$$

which is known as the modified Sugawara construction. The Virasoro generators L_n generate diffeomorphisms, while the affine generators A^a_n generate the standard gauge transformations at the boundary. This relation reflects the fact that the diffeomorphisms in Chern–Simons theory can be expressed in terms of the usual gauge symmetry. The Virasoro generators are the generators of the two-dimensional conformal symmetry, which acts in the space of the residual dynamical variables.

> *The boundary dynamics of Chern–Simons theory is described by a conformally invariant field theory with the central charge c_0.*

The derivation of the Virasoro algebra given here is valid for the general gauge group. The appearance of the classical central charge in the canonical realization of the residual symmetry is a consequence of the specific choice of boundary conditions. In the Chern–Simons description of the three-dimensional gravity we have $4\pi k = l/8G$, $\alpha^2 = -1$, hence $c_0 = 3l/2G$.

Exercises

1. Find PBs between the kinematical constraints $\mathcal{H}_{ij}$ and $\mathcal{H}_\alpha$ in the U_4 space.
2. Derive the following relations in U_4:

$$\begin{aligned}
\{n_k, \mathcal{H}'_{ij}\} &= 2\eta_{k[i}n_{j]}\delta \\
\{h_{\bar{k}}{}^\alpha, \mathcal{H}'_{ij}\} &= 2\eta_{k[i}h_{\bar{j}]}{}^\alpha\delta \\
\{n_k, \mathcal{H}'_\beta\} &= -(\nabla_\beta b^s{}_\alpha)n_s h_{\bar{k}}{}^\alpha\delta \\
\{h_{\bar{k}}{}^\alpha, \mathcal{H}'_\beta\} &= (\nabla_\beta b^s{}_\gamma)(n_s n_k\, {}^3g^{\alpha\gamma} - h_{\bar{k}}{}^\gamma h_{\bar{s}}{}^\alpha)\delta - \delta^\alpha_\beta h_{\bar{k}}{}^\gamma\partial_\gamma\delta \\
\{A^{ij}{}_\alpha, \mathcal{H}'_\beta\} &= -R^{ij}{}_{\alpha\beta}\delta.
\end{aligned}$$

3. Prove the following relations in EC theory:

$$\begin{aligned}
\{A^{ij}{}_\alpha, \mathcal{H}'_\perp\} &= -R^{ij}{}_{\alpha\perp}\delta \\
\{n_k, \mathcal{H}'_\perp\} &= (T_{\perp\bar{k}\perp} - h_{\bar{k}}{}^\alpha\partial_\alpha)\delta \\
\{h_{\bar{k}}{}^\alpha, \mathcal{H}'_\perp\} &= h^{\bar{m}\alpha}(-h_{\bar{k}}{}^\beta\nabla_\beta n_m + T_{\bar{m}\bar{k}\perp} - n_k T_{\perp\bar{m}\perp})\delta + {}^3g^{\alpha\beta}n_k\partial_\beta\delta.
\end{aligned}$$

4. Use the result (6.1) to show that the algebra of FC constraints in the local Lorentz frame has the form (6.2).
5. Consider EC theory with the Dirac matter field. Show that the multipliers u and $\bar{u}$, corresponding to the primary constraints $\bar{\phi}$ and ϕ, are determined by the relations

$$N\{\phi, \mathcal{H}_\perp\} + \mathrm{i}J\gamma^\perp u \approx 0 \qquad N\{\bar{\phi}, \mathcal{H}_\perp\} - \mathrm{i}J\bar{u}\gamma^\perp \approx 0.$$

i.e. that they have the form $N\Lambda_\perp(b^k{}_\alpha, A^{ij}{}_\alpha, \pi_k{}^\alpha, \pi_{ij}{}^\alpha)$.
6. Consider the Lagrangians that are at most quadratic in velocities. Show that the total Hamiltonian can depend on the derivatives of momenta only through the determined multipliers, and that $\partial\mathcal{H}_\mathrm{T}/\partial\pi^A_{,\alpha} \approx 0$.
7. Show that the generator of the Poincaré gauge symmetry (6.7) produces the correct transformations of the fields $b^k{}_\mu$, $A^{ij}{}_\mu$ and Ψ.
8. Assuming that $\widetilde{\mathcal{L}}$ is an invariant density under Poincaré gauge transformations, derive the transformation rules for momentum variables:

$$\delta\pi_A = -\pi_B\frac{\partial\delta\varphi^B}{\partial\varphi^A} + \pi^\alpha_A\xi^0{}_{,\alpha} - \pi_A\xi^\alpha{}_{,\alpha} \qquad \pi^\mu_A \equiv \frac{\partial\widetilde{\mathcal{L}}}{\partial\varphi^A{}_{,\mu}}.$$

9. Let the asymptotic conditions (6.11a) be invariant with respect to the global Poincaré transformations. Show, then, that the field derivatives satisfy (6.11b).
10. Let G be the generator of global symmetries of the Hamiltonian equations of motion. Show that G satisfies conditions (6.20)–(6.21).
11. Find PBs between $A^{ij}{}_\mu$ and the improved Poincaré generators $\widetilde{P}_\nu$ and M_{kl} in EC theory. Then, derive the relations (6.23).
12. Give a detailed proof of the boost conservation law (6.24d).
13. Construct the gauge generators for the system of interacting electromagnetic and Dirac fields:

$$I = \int \mathrm{d}^4x\,\{\bar{\psi}[\gamma^\mu(\mathrm{i}\partial_\mu - eA_\mu) - m]\psi - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}\}.$$

Find the form of the surface term, and discuss the meaning of the conservation law in the case when the gauge parameter behaves as a constant at spatial infinity.
14. Consider the $SU(2)$ gauge theory given by

$$I = \int \mathrm{d}^4x\,\{\bar{\Psi}[\gamma^\mu\mathrm{i}\nabla_\mu - m]\Psi - \tfrac{1}{4}F^a{}_{\mu\nu}F^{a\mu\nu}\}$$

where Ψ is the $SU(2)$ doublet of Dirac fields (see appendix A). Find the conserved charges and calculate their PB algebra in the case when the gauge parameters are constant at spatial infinity.

15. On the basis of the adopted boundary conditions (6.11a) for the tetrads, derive the relations:

$$n_0 = 1 + \mathcal{O}_1 \qquad n_a = \mathcal{O}_1$$
$$N = 1 + \mathcal{O}_1 \qquad N^\alpha = \mathcal{O}_1 \qquad J = 1 + \mathcal{O}_1$$
$$\eta_{\bar{a}\bar{b}} = \eta_{ab} + \mathcal{O}_2 \qquad \eta_{\bar{0}\bar{b}} = \mathcal{O}_1 \qquad \eta_{\bar{0}\bar{0}} = \mathcal{O}_1.$$

16. Find the form of the field transformations (6.32) after introducing the new parameters according to (6.33).
17. Find the ξ^0 transformations of $\lambda_{ij}{}^{0\beta}$ under the action of the Poincaré gauge generator (6.35).
18. Derive the asymptotic relations (6.45) and (6.47).
19. Consider the vector field $\boldsymbol{A}$ with the following asymptotic behaviour:

$$\boldsymbol{A} = \frac{\boldsymbol{a}}{r} + \mathcal{O}_2 \qquad \operatorname{div} \boldsymbol{A} = \mathcal{O}_3$$

where $\boldsymbol{a} = \boldsymbol{a}(\boldsymbol{n})$.
   (a) Show that $\operatorname{div}(\boldsymbol{a}/r) = 0$ for all r.
   (b) Integrate $\operatorname{div}(\boldsymbol{a}/r)$ over the region V_R outside the sphere S_R of radius R and use Stokes's theorem to prove that the integral $\oint \boldsymbol{a} \cdot \boldsymbol{n}\, \mathrm{d}\Omega$ over S_∞ vanishes.
   (c) Finally, show that the integral $\oint \boldsymbol{A} \cdot \mathrm{d}\boldsymbol{S}$ over S_∞ is finite.
20. Use the general transformation rule for momentum variables (exercise 8) to find the form of $\delta_0 \pi_k{}^\gamma$. Then, derive the relations (6.49).
21. (a) Show that the Chern–Simons action can be written in the form:

$$I_{\mathrm{CS}} = k \int \mathrm{d}t \int_\Sigma \mathrm{d}^2 x\, \varepsilon^{0\alpha\beta} (\dot{A}^a{}_\alpha A^b{}_\beta + A^a{}_0 F^b{}_{\alpha\beta}) \gamma_{ab} + B$$

   where B is a boundary term.
   (b) Find the transformation law of the Chern–Simons action under the finite gauge transformations.
22. (a) Construct the preliminary Dirac brackets in the Chern–Simons theory, using the second-class constraints $\phi^{i\alpha}$.
   (b) Transform the gauge generator (6.73b) into a functional of A, using $\phi_i{}^\alpha \approx 0$. Then, use the related Dirac brackets to calculate $\{\widetilde{G}[\tau], \widetilde{G}[\lambda]\}^\bullet$.
23. Combine the field equations of the Chern–Simons theory with the gauge conditions $A_+ = 0$ and $A_r = b^{-1}\partial_r b$, $b = b(r)$, to show that $A_- = b^{-1}\hat{A}_-(x^-)b$.
24. Consider the Chern–Simons gauge field satisfying $A_+ = 0$, $\hat{A}^{(-)} = 1$, $\hat{A}^{(1)} = 0$. Find the restricted form of the gauge parameters (ξ^r, ξ^-).
25. Consider the affine PB algebra corresponding to the $SO(1, 2)$ group:

$$\{A^a_n, A^b_m\} = \varepsilon^{ab}{}_c A^c_{n+m} + \mathrm{i} Z \eta^{ab} n \delta^0_{n+m}.$$

(a) Write this algebra in the light-cone basis, and prove its invariance under the exchange $A^{(+)} \leftrightarrow A^{(-)}$, $A^{(1)} \leftrightarrow -A^{(1)}$.

(b) Calculate the Dirac brackets corresponding to the following extra conditions: $A_n^{(-)} = 2Z\gamma\delta_n^0$, $A_n^{(1)} = 0$ (which can be treated as second-class constraints).

(c) Use these brackets to show that, for $\gamma = 1$, the remaining components $A_n^{(+)}$ satisfy the Virasoro algebra.

Chapter 7

Gravity in flat spacetime

There are many attempts to understand the structure of gravity, but all of them can be roughly classified into two categories according to whether they are focused on the *geometric* or the *particle* aspects of gravity.

The standard geometric approach is based on Einstein's GR, a theory in which gravitational phenomena are connected with the geometry of spacetime. Notable analogies between electromagnetic and gravitational interactions have inspired many attempts to unify these two theories into a single structure. Nowadays we know that any programme of geometric unification must necessarily be more general, since the world of fundamental interactions contains much more than just electromagnetism and gravity.

There is, however, another approach to unification, which is based on the idea that gravity can be described as a relativistic quantum field theory in *flat* spacetime, like all the other fundamental interactions. The construction must be such that it can reproduce the following results of experimental observations:

- gravity has a long range (force proportional to r^{-2}); then,
- it is attractive,
- acts in the same way on all kinds of matter (PE) and
- satisfies the standard classical tests (the gravitational redshift, the precession of Mercury's orbit, the deflection of starlight passing near the sun and the time delay of radar signals emitted from Earth).

If we adopt the standard quantum field theory approach to gravity, the gravitational force is explained by the exchange of a particle called a *graviton* (figure 7.1). Which properties should the graviton have in order to reproduce the basic experimental requirements?

In a field theory, long range forces are produced by the exchange of *massless* particles, as in electrodynamics. Can we exclude the possibility of a very small but finite graviton mass? The answer is affirmative, as we shall see.

Relativistic particles are characterized by definite values of mass and spin. In order to determine the spin of the graviton, we shall analyse different possibilities

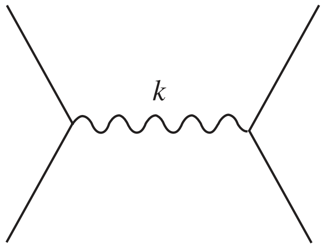

Figure 7.1. The gravitational force is produced by the exchange of a graviton.

and try to identify those which are in agreement with the previous experimental facts.

If the graviton were a fermion, i.e. a particle with half-integer spin, the gravitational force could not be produced by the exchange of a single graviton. A particle emitting or absorbing such a graviton would not remain in the same state as it was initially. Hence, such an exchange would not result in a static force, but rather in a scattering process. The exchange of two spin-$\frac{1}{2}$ gravitons at a time is found to give a force with the incorrect radial dependence. We can continue this way and consider an exchange of a pair of gravitons between three masses (so that any two objects also interact with a third one, which is far away and plays the role of some effective average mass in the universe), but no satisfactory results are obtained (Feynman *et al* 1995).

Thus, we limit our attention to bosons, particles with integer spin. The most appropriate candidate for the graviton will be chosen on the basis of the experimental properties of gravity.

7.1 Theories of long range forces

We begin with a review of some classical properties of massless fields of spin 0, 1 and 2, which will be useful in our attempts to construct a realistic theory of gravity in flat spacetime (Feynman *et al* 1995, Kibble 1965, Van Dam 1974, Duff 1973).

Scalar field

The dynamical properties of free real scalar field of zero mass are defined by the Lagrangian

$$\mathcal{L}_{\mathrm{S}} = \tfrac{1}{2}\partial_\mu\varphi\partial^\mu\varphi. \tag{7.1}$$

The general sign of this expression is determined so as to have a positive Hamiltonian. The choice of the overall numerical factor influences the normalization of φ, and the reality of the field implies its neutrality.

When the coupling to other fields ('matter') is introduced, the field φ obeys the inhomogeneous Klein–Gordon equation:

$$K\varphi = J \equiv -\frac{\delta \mathcal{L}_{\mathrm{I}}}{\delta \varphi} \tag{7.2}$$

where $K \equiv -\Box$, and $\mathcal{L}_{\mathrm{I}}$ is the interaction Lagrangian. If we define the general propagator (Green function) D by the equation

$$K_x D(x-y) = \delta(x-y)$$

we can easily verify that the construct

$$\varphi(x) = \int \mathrm{d}^4 y\, D(x-y) J(y) \tag{7.3}$$

is a particular solution of equation (7.2). If we take a solution $D(x-y)$ which obeys a specific boundary condition, then $\varphi(x)$ satisfies the same boundary condition. In the momentum space the propagator is given by

$$D(x) = \int \frac{\mathrm{d}^4 k}{(2\pi)^4} D(k) \mathrm{e}^{\mathrm{i}kx} \qquad D(k) \equiv \frac{1}{k^2}. \tag{7.4}$$

We note that $D(k)$ has poles at $k_0 = \pm\omega_k$, $\omega_k \equiv \sqrt{\boldsymbol{k}^2}$, so that the Fourier integral for $D(x)$ is not well defined. Choosing different contours to avoid poles at the real k_0-axis, we get well-defined expressions, which obey specific boundary conditions (the retarded propagator D_{R}, the Feynman propagator D_{F}, etc). For classical radiation problems, the appropriate physical boundary conditions are those of retardation.

If the source J is static, we can integrate over y_0 in (7.3) and obtain $\varphi(\boldsymbol{x}) = \int \mathrm{d}^3 y\, D(\boldsymbol{x}-\boldsymbol{y}) J(\boldsymbol{y})$, where

$$D(\boldsymbol{x}) = \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \frac{1}{-\boldsymbol{k}^2} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\boldsymbol{x}} = -\frac{1}{4\pi|\boldsymbol{x}|} \qquad \nabla^2 D(\boldsymbol{x}) = \delta(\boldsymbol{x}).$$

For a field independent source we have $\mathcal{L}_{\mathrm{I}} = -J\varphi$, and the interaction Hamiltonian is $H_I = -\int \mathrm{d}^3 x\, \mathcal{L}_{\mathrm{I}}$. Hence, the interaction energy of a source J_1 with the field φ_2 produced by J_2, is given as

$$E(x^0) = \int \mathrm{d}^3 x\, J_1(x)\varphi_2(x) = \int \mathrm{d}^3 x\, \mathrm{d}^4 y\, J_1(x) D(x-y) J_2(y).$$

For static sources this energy is Coulombic.

Vector field

The free real vector field of mass $\mu = 0$ is described by the Lagrangian

$$\mathcal{L}_{\mathrm{V}} = -\tfrac{1}{4}(\partial_\mu \varphi_\nu - \partial_\nu \varphi_\mu)(\partial^\mu \varphi^\nu - \partial^\nu \varphi^\mu) \tag{7.5}$$

which is invariant under gauge transformations

$$\varphi_\mu \to \varphi'_\mu = \varphi_\mu + \partial_\mu \lambda.$$

This invariance is directly related to the masslessness of the field. To clarify the meaning of this statement we shall first discuss some relevant properties of the massive vector field, and analyse the relation between massless theory and the limiting case $\mu^2 \to 0$ of massive theory.

Massive vector field. The Lagrangian for a free massive vector field,

$$\widetilde{\mathcal{L}}_\mathrm{V} = -\tfrac{1}{4}(\partial_\mu \varphi_\nu - \partial_\nu \varphi_\mu)(\partial^\mu \varphi^\nu - \partial^\nu \varphi^\mu) + \tfrac{1}{2}\mu^2 \varphi_\mu \varphi^\mu \tag{7.6}$$

yields the following field equation:

$$\partial_\mu(\partial^\mu \varphi^\nu - \partial^\nu \varphi^\mu) + \mu^2 \varphi^\nu = 0.$$

By differentiating this equation we obtain the *extra condition* $\partial_\nu \varphi^\nu = 0$ as a consistency condition of the field equation. Note that this is true as long as $\mu^2 \neq 0$. This extra condition simplifies the field equation,

$$(\Box + \mu^2)\varphi^\nu = 0$$

and removes the spin-0 part of φ^ν.

The spin structure of the vector field is conveniently described by introducing polarization vectors e_μ. Consider the plane wave,

$$\varphi_\mu(x) = e_\mu \mathrm{e}^{\mathrm{i}k\cdot x} + e^*_\mu \mathrm{e}^{-\mathrm{i}k\cdot x}$$

which obeys the field equation and the additional condition provided

$$k^2 - \mu^2 = 0 \qquad k \cdot e = 0.$$

The second requirement reduces the number of independent polarization vectors to three: $e^\mu_{(\lambda)}$, $\lambda = 1, 2, 3$. Since k is timelike, the polarization vectors are spacelike, and may be chosen to be orthonormal:

$$e^*_{(\lambda)} \cdot e_{(\lambda')} = -\delta_{\lambda\lambda'}.$$

The completeness relation for the basis $(k^\nu/\mu, e^\nu_{(\lambda)})$ reads:

$$\sum_{\lambda=1}^{3} e^\mu_{(\lambda)} e^{\nu *}_{(\lambda)} + \frac{k^\mu k^\nu}{\mu^2} = \eta^{\mu\nu}. \tag{7.7}$$

Now, we take the momentum $\boldsymbol{k}$ along the z-axis, $k = (k^0, 0, 0, k^3)$, and choose $e_{(1)}$ and $e_{(2)}$ to be orthogonal to $\boldsymbol{k}$:

$$e_{(1)} = (0, 1, 0, 0) \qquad e_{(2)} = (0, 0, 1, 0)$$
$$e_{(3)} = (k^3/\mu, 0, 0, k^0/\mu).$$

The distinction between different polarization vectors is clearly seen from their behaviour with respect to rotations around the z-axis:

$$e'^{\mu} = R^{\mu}{}_{\nu} e^{\nu} \qquad (R^{\mu}{}_{\nu}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta & 0 \\ 0 & \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Such a transformation leaves k^{μ} invariant, and transforms $e_{(\lambda)}$ as follows:

$$e'_{(\pm 1)} = \mathrm{e}^{\pm \mathrm{i}\theta} e_{(\pm 1)} \qquad e'_{(0)} = e_{(0)} \tag{7.8}$$

where

$$e_{(\pm 1)} = \frac{1}{\sqrt{2}}(e_{(1)} \pm \mathrm{i} e_{(2)}) \qquad e_{(0)} = e_{(3)}.$$

Thus, the vectors $e_{(+1)}, e_{(0)}, e_{(-1)}$ describe polarization states of unit spin, $s = 1$, with spin projections $s_3 = +1, 0, -1$.

When φ^{μ} is coupled to the matter fields, the field equation takes the form

$$\partial_{\mu}(\partial^{\mu}\varphi^{\nu} - \partial^{\nu}\varphi^{\mu}) + \mu^2 \varphi^{\nu} = J^{\nu} \equiv -\frac{\delta \mathcal{L}_{\mathrm{I}}}{\delta \varphi_{\nu}}. \tag{7.9a}$$

Using here the consistency requirement $\mu^2 \partial \cdot \varphi = \partial \cdot J$, we find

$$(\Box + \mu^2)\varphi^{\mu} = \left(\eta^{\mu\nu} + \frac{\partial^{\mu}\partial^{\nu}}{\mu^2} \right) J_{\nu}. \tag{7.9b}$$

A particular solution of this equation may be written as

$$\varphi^{\mu}(x) = \int \mathrm{d}^4 y \, D^{\mu\nu}(x - y) J_{\nu}(y) \tag{7.10}$$

where $D^{\mu\nu}$ is the propagator, the Fourier transform of which is

$$D^{\mu\nu}(k, \mu^2) = -\frac{P^{\mu\nu}}{k^2 - \mu^2} \qquad P^{\mu\nu} \equiv \eta^{\mu\nu} - k^{\mu}k^{\nu}/\mu^2 = \sum_{\lambda=1}^{3} e^{\mu}_{(\lambda)} e^{\nu *}_{(\lambda)}. \tag{7.11}$$

The last equality follows from the completeness relation (7.7). In the rest frame and for $k^2 = \mu^2$, the tensor $P^{\mu\nu}$ is the projector on the three-dimensional space,

$$\eta^{\mu\nu} - k^{\mu}k^{\nu}/\mu^2 = \begin{cases} 0 & \text{for } \mu = 0 \text{ or } \nu = 0 \\ -\delta^{\alpha\beta} & \text{for } \mu = \alpha,\ \nu = \beta \end{cases}$$

hence, φ^{μ} in (7.10) has only three independent components.

The massless vector field. The field equation for a massless vector field coupled to matter has the form

$$K_{\mu\nu}\varphi^\nu = J_\mu \qquad K_{\mu\nu} \equiv \Box\eta_{\mu\nu} - \partial_\mu\partial_\nu. \tag{7.12}$$

The consistency of this equation requires the conservation law $\partial \cdot J = 0$.

Owing to the gauge invariance, the field equation has no unique solution. Indeed, gauge freedom reflects itself in the singularity of the kinetic operator $K_{\mu\nu}$: $K_{\mu\nu}\partial^\nu\lambda = 0$. Hence, we cannot define its inverse—the propagator—and every solution of equation (7.12) is defined only up to a gauge transformation.

A unique solution of gauge invariant field equations may be found by imposing convenient *gauge conditions*. The number of these conditions is equal to the number of gauge parameters. For the vector field we can choose, for instance, the covariant Lorentz gauge condition,

$$\partial \cdot \varphi = 0 \tag{7.13}$$

but now it cannot be deduced from the field equation. This condition does not fix the gauge freedom completely: gauge transformations with parameters $\lambda(x)$, such that $\Box\lambda(x) = 0$, are still allowed. The field equation is significantly simplified,

$$\Box\varphi^\mu = J^\mu \tag{7.14}$$

and can be solved for φ^μ. In the space of functions which obey the Lorentz condition, the operator $K_{\mu\nu}$ is not singular ($K_{\mu\nu} = \Box\eta_{\mu\nu}$), and we can define the propagator as the inverse of $K_{\mu\nu}$:

$$D^{\mu\nu} = -\eta^{\mu\nu}D. \tag{7.15}$$

The propagator with a given boundary condition defines a particular solution of the field equation (7.14): $\varphi^\mu(x) = \int \mathrm{d}^4y\, D^{\mu\nu}(x-y)J_\nu(y)$.

Any gauge-fixing condition must be accessible, i.e. it must be possible to choose the gauge parameter $\lambda(x)$ which will transform an arbitrary field configuration into a given, gauge fixed configuration. The Lorentz condition is locally correct. Indeed, for any φ^μ we can choose the gauge parameter such that $\partial \cdot \varphi + \Box\lambda = 0$, whereupon φ'^μ obeys the Lorentz gauge (7.13).

Instead of imposing a gauge condition, we can break the gauge symmetry by adding a suitable term to the original Lagrangian. Although the form of the propagator depends on this term, we can show that physical predictions of the theory remain the same. With the gauge breaking term $-\frac{\alpha}{2}(\partial\cdot\varphi)^2$, the Lagrangian (7.5) and the related propagator are given as

$$\mathcal{L}'_{\mathrm{V}} = \mathcal{L}_{\mathrm{V}} - \frac{\alpha}{2}(\partial\cdot\varphi)^2 \qquad D'_{\mu\nu} = -\frac{1}{k^2}\left[\eta_{\mu\nu} + \left(\frac{1}{\alpha} - 1\right)\frac{k^\mu k^\nu}{k^2}\right].$$

For $\alpha = 1$ we get the propagator (7.15). Another (transverse) propagator is defined in the limit $\alpha \to \infty$. The limit $\alpha \to 0$ is singular, since the additional term in $\mathcal{L}'_{\mathrm{V}}$ vanishes, and the Lagrangian is again gauge invariant.

The plane wave

$$\varphi_\mu = \epsilon_\mu e^{ik\cdot x} + \epsilon_\mu^* e^{-ik\cdot x}$$

is a vacuum solution of the field equation (7.14) in the Lorentz gauge if

$$k^2 = 0 \qquad k\cdot\epsilon = 0. \tag{7.16}$$

The second condition reduces the number of independent components of ϵ_μ from four to three. Furthermore, without leaving the Lorentz gauge we have the freedom of gauge transformations with the parameter λ which obeys $\Box\lambda = 0$. Choosing

$$\lambda(x) = i\eta e^{ik\cdot x} - i\eta^* e^{-ik\cdot x}$$

the residual gauge transformation may be written as

$$\epsilon'_\mu = \epsilon_\mu - \eta k_\mu .$$

Since η is an arbitrary parameter, only two of three components of ϵ_μ are physically significant. This is a direct consequence of the gauge invariance.

If the plane wave is travelling in the z-direction, the condition $k\cdot\epsilon = 0$ gives $\epsilon^0 = \epsilon^3$, and the choice $\eta = \epsilon_3/k_3$ yields $\epsilon^3 = 0$. Thus, there are only two polarization vectors that carry physical significance:

$$\epsilon_{(1)} = (0,1,0,0) \qquad \epsilon_{(2)} = (0,0,1,0).$$

Considering now the action of rotations around the z-axis, we deduce that $\epsilon_{(\pm 1)}$ are the polarization states with helicities $\lambda = \pm 1$, while the vector $\epsilon_{(3)} = (1,0,0,1)$ has no physical relevance (in contrast to the case $\mu^2 \neq 0$), since it may be transformed to zero by a gauge transformation.

The completeness relation takes a new form, different from (7.7), since now $\mu^2 = 0$. Let us introduce the vector $\bar{k} = (k^0, -\boldsymbol{k})$, which obeys, for $k^2 = 0$, the conditions

$$\bar{k}^2 = 0 \qquad \bar{k}\cdot\epsilon_{(1)} = \bar{k}\cdot\epsilon_{(2)} = 0.$$

The vectors $k + \bar{k}$ and $k - \bar{k}$ are timelike and spacelike, respectively, and the completeness relation is given as

$$\sum_{\lambda=1}^{2} \epsilon^\mu_{(\lambda)}\epsilon^{\nu *}_{(\lambda)} + \frac{k^\mu \bar{k}^\nu + \bar{k}^\mu k^\nu}{k\cdot\bar{k}} = \eta^{\mu\nu}. \tag{7.17}$$

Can the massless theory be understood as a limiting case of the massive theory? The answer to this question is not so direct. In the limit $\mu^2 \to 0$ the Lagrangian (7.6) becomes gauge invariant, and the propagator becomes singular. What is then the meaning of this limit? If we consider the massive vector field coupled to a *conserved* current ($k_\mu J^\mu = 0$), the limiting case $\mu^2 \to 0$ represents a sensible theory. The interaction between two currents J_1 and J_2 is described by the amplitude,

$$M_{21} = J_2^\mu D^{\mathrm{E}}_{\mu\nu}(k,\mu^2) J_1^\nu .$$

Here, D^{E} is the effective propagator which is obtained from the expression (7.11) by discarding all the terms proportional to k_ν. When $\mu^2 \to 0$ the effective propagator is the same as the propagator (7.15) of massless theory. This result is related to the fact that the third polarization state of the massive theory $e_{(3)} = (k^3/\mu, 0, 0, k^0/\mu)$ is decoupled from the conserved current J^μ in the limit $\mu^2 \to 0$.

The symmetric tensor field

We now wish to find a Lagrangian for the massless symmetric tensor field $\varphi_{\mu\nu}$, requiring from the very beginning invariance under the gauge transformations

$$\varphi_{\mu\nu} \to \varphi'_{\mu\nu} = \varphi_{\mu\nu} + \partial_\mu \lambda_\nu + \partial_\nu \lambda_\mu. \tag{7.18}$$

We begin by writing down all the possible quadratic terms that can be present in a free Lagrangian:

$$\mathcal{L}_1 = \varphi_{\mu\nu,\sigma}\varphi^{\mu\nu,\sigma} \qquad \mathcal{L}_2 = \varphi_{\mu\nu,\sigma}\varphi^{\mu\sigma,\nu}$$
$$\mathcal{L}_3 = \varphi_{\mu\nu}{}^{,\lambda}\varphi^{\mu\sigma}{}_{,\sigma} \qquad \mathcal{L}_4 = \varphi_{\mu\lambda}{}^{,\lambda}\varphi^{,\mu}$$
$$\mathcal{L}_5 = \varphi_{,\mu}\varphi^{,\mu}$$

where $\varphi \equiv \varphi_\nu{}^\nu$. Not all of these invariants are necessary: $\mathcal{L}_3$ can be converted to $\mathcal{L}_2$ by integration by parts, leaving only four independent terms. Thus, we assume a Lagrangian of the form

$$\mathcal{L}_{\mathrm{T}} = a\mathcal{L}_1 + b\mathcal{L}_2 + c\mathcal{L}_4 + d\mathcal{L}_5.$$

Varying $\mathcal{L}_{\mathrm{T}}$ with respect to $\varphi^{\mu\nu}$ we obtain the field equations:

$$-2a\Box\varphi_{\mu\nu} - b(\varphi_{\mu\sigma,\nu}{}^\sigma + \varphi_{\nu\sigma,\mu}{}^\sigma) - c(\varphi_{,\mu\nu} + \eta_{\mu\nu}\varphi_{\lambda\sigma}{}^{,\lambda\sigma}) - 2d\eta_{\mu\nu}\Box\varphi = 0.$$

We now perform a gauge transformation, and the requirement of gauge invariance gives

$$(2a+b)\Box(\lambda_{\mu,\nu} + \lambda_{\nu,\mu}) + (b+c)2\lambda_{\sigma,\mu\nu}{}^\sigma + (c+2d)2\eta_{\mu\nu}\Box\lambda_\sigma{}^{,\sigma} = 0.$$

Choosing a general scale in $\mathcal{L}_{\mathrm{T}}$ such that $a = 1/2$, we obtain $b = -1$, $c = 1$, $d = -1/2$, and the Lagrangian takes the final form (Fierz and Pauli 1939, Feynman *et al* 1995).

$$\mathcal{L}_{\mathrm{T}} = \tfrac{1}{2}\varphi_{\mu\nu,\sigma}\varphi^{\mu\nu,\sigma} - \varphi_{\mu\nu,\sigma}\varphi^{\mu\sigma,\nu} + \varphi_{\mu\sigma}{}^{,\sigma}\varphi^{,\mu} - \tfrac{1}{2}\varphi_{,\nu}\varphi^{,\nu}. \tag{7.19a}$$

It is useful to introduce another set of variables,

$$\bar{\varphi}_{\mu\nu} = \varphi_{\mu\nu} - \tfrac{1}{2}\eta_{\mu\nu}\varphi$$

which simplifies the form of $\mathcal{L}_T$:

$$\mathcal{L}_T = \tfrac{1}{2}\varphi_{\mu\nu,\sigma}\varphi^{\mu\nu,\sigma} - \bar{\varphi}_{\mu\nu,\sigma}\bar{\varphi}^{\mu\sigma,\nu} - \tfrac{1}{4}\bar{\varphi}_{,\nu}\bar{\varphi}^{,\nu}. \tag{7.19b}$$

If $\varphi_{\mu\nu}$ is coupled to matter, the field equations are of the form

$$\begin{aligned} K_{\mu\nu,\sigma\rho}\varphi^{\sigma\rho} &= J_{\mu\nu} \equiv -\frac{\delta\mathcal{L}_I}{\delta\varphi^{\mu\nu}} \\ K_{\mu\nu,\sigma\rho} &\equiv -\eta_{\mu(\sigma}\eta_{\nu\rho)}\Box + \eta_{\mu(\sigma}\partial_\nu\partial_{\rho)} + \eta_{\nu(\sigma}\partial_\mu\partial_{\rho)} \\ &\quad -\eta_{\sigma\rho}\partial_\mu\partial_\nu - \eta_{\mu\nu}(\partial_\sigma\partial_\rho - \eta_{\sigma\rho}\Box). \end{aligned} \tag{7.20}$$

They are consistent only if the current is conserved: $\partial^\mu J_{\mu\nu} = 0$. Owing to gauge invariance the operator K is singular ($K_{\mu\nu,\sigma\rho}\partial^\sigma\lambda^\rho = 0$) and cannot be inverted, so that there is no unique solution of the field equations.

We may choose the covariant gauge condition for $\varphi_{\mu\nu}$,

$$\partial_\mu\bar{\varphi}^{\mu\nu} \equiv \partial_\mu\varphi^{\mu\nu} - \tfrac{1}{2}\partial^\nu\varphi = 0 \tag{7.21}$$

called the Hilbert gauge. This condition, like Lorentz's, is locally accessible. On the other hand, the condition $\partial_\mu\varphi^{\mu\nu} = 0$, which seems at first sight to be as natural as (7.21), is actually not a correct gauge choice (Kibble 1965).

The Hilbert gauge considerably simplifies the field equations,

$$-\tfrac{1}{2}(\eta_{\mu\lambda}\eta_{\nu\rho} + \eta_{\mu\rho}\eta_{\nu\lambda} - \eta_{\mu\nu}\eta_{\lambda\rho})\Box\varphi^{\lambda\rho} = J^{\mu\nu} \tag{7.22}$$

and leads to the propagator

$$D^{\mu\nu,\lambda\rho} = \tfrac{1}{2}(\eta^{\mu\lambda}\eta^{\nu\rho} + \eta^{\nu\lambda}\eta^{\mu\rho} - \eta^{\mu\nu}\eta^{\lambda\rho})D. \tag{7.23}$$

In order to overcome the difficulty with the singularity of K, it is also possible to modify the Lagrangian $\mathcal{L}_T$ by adding a gauge-breaking term. If we choose, for instance, $\mathcal{L}'_T = \mathcal{L}_T - \frac{\alpha}{2}(\partial_\mu\bar{\varphi}^{\mu\nu})^2$, the corresponding propagator for $\alpha = 1$ goes over into expression (7.23).

A particular solution of equation (7.22) is obtained using a well-defined propagator:

$$\varphi^{\mu\nu}(x) = \int \mathrm{d}^4y\, D^{\mu\nu,\lambda\rho}(x-y)J_{\lambda\rho}(y).$$

If the current $J_{\lambda\rho}$ does not depend on $\varphi_{\mu\nu}$, the interaction energy between currents 1 and 2 is

$$E(x^0) = \int \mathrm{d}^3x\,\mathrm{d}^4y\, J_1^{\mu\nu}(x)D_{\mu\nu,\lambda\rho}(x-y)J_2^{\lambda\rho}(y).$$

Polarization states. The plane wave

$$\varphi_{\mu\nu} = \epsilon_{\mu\nu} e^{ik\cdot x} + \epsilon^*_{\mu\nu} e^{-ik\cdot x}$$

satisfies the field equations in vacuum and the Hilbert gauge if

$$k^2 = 0 \tag{7.24a}$$
$$k^\mu(\epsilon_{\mu\nu} - \tfrac{1}{2}\eta_{\mu\nu}\epsilon_\lambda{}^\lambda) = 0. \tag{7.24b}$$

Four relations (7.24*b*) lower the number of ten independent components of $\epsilon_{\mu\nu}$ to six, but of these six only two are physically significant. Indeed, the condition (7.24*b*) allows the residual gauge transformations

$$\epsilon'_{\mu\nu} = \epsilon_{\mu\nu} - k_\mu\eta_\nu - k_\nu\eta_\mu$$

with four arbitrary parameters η_ν, so the number of physically significant polarization components is reduced to $6 - 4 = 2$.

When the plane wave is travelling along the z-axis, the conditions (7.24*b*) can be used to express $\epsilon_{01}, \epsilon_{02}, \epsilon_{03}$ and ϵ_{22} in terms of the other six components,

$$\epsilon_{01} = -\epsilon_{31} \qquad \epsilon_{03} = -\tfrac{1}{2}(\epsilon_{33} + \epsilon_{00})$$
$$\epsilon_{02} = -\epsilon_{32} \qquad \epsilon_{22} = -\epsilon_{11}$$

while a suitable choice of η_ν leads to

$$\epsilon'_{13} = \epsilon'_{23} = \epsilon'_{33} = \epsilon'_{00} = 0.$$

Thus, only $\epsilon_{11} = -\epsilon_{22}$ and ϵ_{12} have a definite physical significance, which implies that there are only two physical polarization states (Van Dam 1974),

$$\epsilon_{(1)} = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \epsilon_{(2)} = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{7.25}$$

Here, $\epsilon_{(\lambda)}$ denotes the matrix $(\epsilon^{\mu\nu}_{(\lambda)})$.

Under rotations around the z-axis, $\epsilon'^{\mu\nu} = R^\mu{}_\sigma R^\nu{}_\rho \epsilon^{\sigma\rho}$, we find:

$$\epsilon'_{(\pm 2)} = e^{\pm 2i\theta}\epsilon_{(\pm 2)} \qquad \epsilon_{(\pm 2)} \equiv \frac{1}{\sqrt{2}}(\epsilon_{(1)} \pm i\epsilon_{(2)}). \tag{7.26}$$

Thus, $\epsilon_{(\pm 2)}$ are polarization states with helicities $\lambda = \pm 2$.

Currents. The currents carrying helicities $\lambda = \pm 2$ are given as

$$J_{(\pm 2)} = \epsilon^{\mu\nu}_{(\pm 2)} J_{\mu\nu} = \tfrac{1}{2}(J_{11} - J_{22} \pm 2iJ_{12}).$$

Their interaction has the form

$$\sum_{\lambda=\pm 2} J'_{(\lambda)} J^*_{(\lambda)} = \tfrac{1}{2}(J'_{11} - J'_{22})(J^*_{11} - J^*_{22}) + 2J'_{12}J^*_{12}$$
$$= (J'_{11}J^*_{11} + J'_{22}J^*_{22} + 2J'_{12}J^*_{12}) - \tfrac{1}{2}(J'_{11} + J'_{22})(J^*_{11} + J^*_{22}). \tag{7.27}$$

Since, on the other hand,

$$\sum_{\lambda=\pm 2} J'_{(\lambda)} J^*_{(\lambda)} = J'_{\mu\nu}\left(\sum_{\lambda=\pm 2} \epsilon^{\mu\nu}_{(\lambda)}\epsilon^{\sigma\rho *}_{(\lambda)}\right)J^*_{\sigma\rho} \equiv J'_{\mu\nu}\Pi^{\mu\nu,\sigma\rho}J^*_{\sigma\rho}$$

we easily find the expression for the polarization sum (Schwinger 1970):

$$\Pi^{\mu\nu,\sigma\rho} = \begin{cases} \tfrac{1}{2}(\eta^{\mu\sigma}\eta^{\nu\rho} + \eta^{\mu\rho}\eta^{\nu\sigma}) - \tfrac{1}{2}\eta^{\mu\nu}\eta^{\sigma\rho} & \text{for } \mu, \nu, \sigma, \rho \neq 0 \\ 0 & \text{at least one index} = 0 \text{ or } 3. \end{cases}$$

For general momentum k the result is obtained from the first line by the replacement $\eta^{\mu\nu} \to \Pi^{\mu\nu} \equiv \eta^{\mu\nu} - (k^\mu \bar{k}^\nu + \bar{k}^\mu k^\nu)/k \cdot \bar{k}$:

$$\Pi^{\mu\nu,\sigma\rho} = \tfrac{1}{2}(\eta^{\mu\sigma}\eta^{\nu\rho} + \eta^{\mu\rho}\eta^{\nu\sigma}) - \tfrac{1}{2}\eta^{\mu\nu}\eta^{\sigma\rho}$$
$$+ k\text{-dependent terms.} \tag{7.28}$$

The k-dependent terms may be discarded when we consider the interaction between conserved currents.

The field equation (7.22) can be easily brought to the form

$$-\Box\varphi^{\mu\nu} = J^{\mu\nu} - \tfrac{1}{2}\eta^{\mu\nu}J_\lambda{}^\lambda \equiv \bar{J}^{\mu\nu}. \tag{7.29}$$

In order to clarify the role of the second term on the right-hand side, let us consider the equation

$$-\Box\varphi^{\mu\nu} = J^{\mu\nu}.$$

If we choose a particular solution $\varphi^{\mu\nu} = J^{\mu\nu}/k^2$, the interaction between $\varphi^{\mu\nu}$ and $J'_{\mu\nu}$ may be written as

$$J'_{\mu\nu}\varphi^{\mu\nu} = J'_{\mu\nu}\frac{1}{k^2}J^{\mu\nu}.$$

In the frame where $k = (k^0, 0, 0, k^3)$ the current conservation reads $k^0 J_{0\nu} = -k^3 J_{3\nu}$. When we use this relation to eliminate all the terms with index 3 on the right-hand side, the interaction amplitude separates into two parts: an instantaneous part, proportional to $1/(k^3)^2$, and a retarded part, proportional to $1/k^2$. For the retarded part we get the form

$$\frac{1}{k^2}(J'_{11}J_{11} + J'_{22}J_{22} + 2J'_{12}J_{12})$$

which represents the sum of *three* independent terms or three polarizations. To eliminate the spin-0 part we must add a term

$$-\frac{1}{2}J'_{\mu}{}^{\mu}\frac{1}{k^2}J_{\nu}{}^{\nu}$$

whereupon the retarded term reduces to the form (7.27), containing only a sum of two polarizations. Adding this term is equivalent to a modification of the starting equation,

$$-\Box\varphi^{\mu\nu} = J^{\mu\nu} - \tfrac{1}{2}\eta^{\mu\nu}J_{\lambda}{}^{\lambda}$$

which becomes the same as (7.29). Thus, we see that the field equation (7.29) really describe massless tensor field of helicity $\lambda = \pm 2$.

The sign of the static interaction

It is interesting to observe the alternating change of signs in expressions (7.4), (7.15) and (7.23), defining the propagators of massless fields with spins 0, 1 and 2, respectively. These signs come from the signs of the Lagrangians (7.1), (7.5) and (7.19), and are determined by the requirement of the positivity of the respective Hamiltonians. They have an important influence on the structure of static interactions between particles. In the static case it is only the quantity J and the components J^0 and J^{00} that do not vanish (an analogous property holds for fields of arbitrary spin). This follows from the fact that the only vector we can use to construct the currents is p^{μ}. The energy of the static interaction between particles 1 and 2 takes the form

$$J_1 D J_2 \qquad J_1^{\mu} D_{\mu\nu} J_2^{\nu} \qquad J_1^{\mu\nu} D_{\mu\nu,\lambda\rho} J_2^{\lambda\rho}$$

with all indices equal to zero. It is positive or negative depending on whether the spin of the field is even or odd.

> *The static force between like particles is, therefore, attractive only for gravitons of even spin.*

This is the main reason for rejecting the spin-1 field as a candidate for the description of gravity.

At this point we wish to make a comment on the possibility of the existence of *negative* masses. The interaction between a large positive mass M and a small negative mass m would have the opposite sign of the static interaction (force). However, in Newton's theory we have

$$-m_{\mathrm{g}}\frac{GM}{\boldsymbol{x}^2}\hat{\boldsymbol{x}} = m_{\mathrm{i}}\ddot{\boldsymbol{x}}$$

so that if the PE holds ($m_{\mathrm{i}} = m_{\mathrm{g}}$), the *acceleration* remains the same, irrespectively of the change of sign of the force. In other words, Newton's apple would fall on Earth even if it had negative mass. Gravitational repulsion could occur in the case $m_\mathrm{i}/m_\mathrm{g} < 0$. There existed an idea that $m_\mathrm{g} < 0$ for antiparticles, which would violate the PE (since $m_\mathrm{i} > 0$). However, all the experimental evidence shows that matter and antimatter behave identically with respect to the gravitational interaction. For an idea of antigravity in the context of supergravity, see Scherk (1979).

7.2 Attempts to build a realistic theory

Previous arguments on the possible nature of the graviton indicate that

- the mass of the graviton is zero, as follows from the long range of the gravitational interaction; and
- the spin of the graviton is even, otherwise the static gravitational force would have the wrong sign.

After having discussed relevant classical properties of massless fields of spin 0 and 2, we might now continue to study field theories with higher spins $s > 2$. However, there are general arguments from covariant quantum field theory (appendix G) that lead to the following conclusion:

- massless fields of spin $s > 2$ are not suitable candidates for describing gravity.

Thus, for instance, the interaction energy of a test particle of spin-3 with a static point like source has the form that violates the PE (Van Dam 1974). Hence,

massless fields of spin 0 or 2 are the only candidates for describing the graviton.

In order to find out the most realistic possibility we shall now calculate some observational consequences of these theories. The possibility of a small but finite graviton mass will also be discussed (Van Dam and Veltman 1970, Boulware and Deser 1972).

Scalar gravitational field

We begin our discussion with a scalar theory of gravity,

$$\mathcal{L} = \mathcal{L}_\mathrm{S} + \mathcal{L}_\mathrm{M} + \mathcal{L}_\mathrm{I} \tag{7.30}$$

where $\mathcal{L}_\mathrm{S}$ is defined by equation (7.1), $\mathcal{L}_\mathrm{M}$ describes free matter and $\mathcal{L}_\mathrm{I}$ is the interaction term,

$$\mathcal{L}_\mathrm{I} = -\lambda\varphi T \tag{7.31a}$$

with T being the trace of the matter energy–momentum tensor.

Since $T = 0$ for the electromagnetic field, the scalar gravity does not interact with electromagnetism. As a consequence, this theory predicts that gravity does not affect the motion of an electromagnetic signal: there is no bending of light by the sun, nor the time delay of radar echo, in contradiction to the experimental data.

1. We assume that matter is described as a single particle with world line $z^\mu = z^\mu(s)$:

$$I_\mathrm{M} = \int \mathrm{d}^4x\, \mathcal{L}_\mathrm{M}(x) = -M \int \mathrm{d}s \tag{7.31b}$$

$$T^{\mu\nu}(x) = M \int \frac{\mathrm{d}z^\mu}{\mathrm{d}s}\frac{\mathrm{d}z^\nu}{\mathrm{d}s}\delta(x - z(s))\,\mathrm{d}s \tag{7.31c}$$

where $\mathrm{d}s^2 = \eta_{\mu\nu}\,\mathrm{d}z^\mu\,\mathrm{d}z^\nu$, and M is the mass of the particle.

Let us find a field generated by a static particle. From the form of $T^{\mu\nu}$ it follows $T(x) = M \int \delta(x - z(s))\,\mathrm{d}s$, and the integration over $z^0(s) = s$ yields $T(x) = M\delta(\boldsymbol{x})$. The field equation for φ, $-\Box\varphi = \lambda T$, has a solution

$$\varphi(\boldsymbol{x}) = \lambda M D(\boldsymbol{x}) = -\frac{\lambda}{4\pi}\frac{M}{r}. \tag{7.32}$$

The interaction energy of a test particle of mass m with the static field is given by

$$E(x^0) = \lambda \int \mathrm{d}^3x\, T'(x)\varphi(x) = -\frac{\lambda^2}{4\pi}\frac{M}{r}\frac{m}{u^0} \tag{7.33}$$

where $u^0 = \mathrm{d}x^0/\mathrm{d}s$. For a slowly moving test particle we have $u^0 \approx 1$, so that the previous expression represents *Newton's law* if $\lambda^2/4\pi = G$.

The quantity λ is a universal constant, in agreement with the PE.

The gravitational *redshift* may be derived from the law of conservation of energy, applied to a photon in the context of Newton's law of gravity. Consider an atom in the static gravitational field (7.32), at a distance r from the pointlike source. When $r \to \infty$, the energy of the ground state is E_0, the energy of the first excited state is E_1, and the energy of the photon emitted in the transition between these two states is given by $2\pi\nu_0 = E_1 - E_0$. For finite r the energy of the ground state is $E_0(1 - 2MG/r)$, and similarly for the excited state. Hence, the energy of the emitted photon, as measured by an observer at infinity, is given by

$$2\pi\nu = (E_1 - E_0)(1 - GM/r) = 2\pi\nu_0(1 - GM/r). \tag{7.34}$$

The redshift predicted by this formula has been verified experimentally with an accuracy of 1% (Misner *et al* 1970).

In order to find the equation of motion of a test particle in the static field (7.32), we write the part of the action depending on $z^\mu(s)$ as

$$I_\mathrm{MI} = -M \int [1 + \lambda\varphi(z)]\,\mathrm{d}s. \tag{7.35}$$

Introducing

$$g_{\mu\nu} = (1 + \lambda\varphi)^2 \eta_{\mu\nu} \tag{7.36}$$

this expression can be geometrically interpreted as the action for a point particle in a spacetime with metric $g_{\mu\nu}$. Varying with respect to $z(s)$,

$$\frac{\mathrm{d}}{\mathrm{d}s}\left(g_{\mu\nu}\frac{\mathrm{d}z^\nu}{\mathrm{d}s}\right) = \frac{1}{2}\frac{\partial g_{\lambda\sigma}}{\partial z^\mu}\frac{\mathrm{d}z^\lambda}{\mathrm{d}s}\frac{\mathrm{d}z^\sigma}{\mathrm{d}s}$$

we obtain the geodesic equation in Riemann space with metric $g_{\mu\nu}$.

2. Now we treat the gravitational field (7.32) as the field of our sun, and calculate the precession of perihelia of planetary orbits using the Hamilton–Jacobi equation (Landau and Lifshitz 1975)

$$g^{\mu\nu}\frac{\partial I}{\partial x^\mu}\frac{\partial I}{\partial x^\nu} - m^2 = 0 \tag{7.37}$$

where I is the action, and m the mass of the planet we are studying. The metric (7.36) is static and spherically symmetric, and in spherical coordinates $x^\mu = (t, r, \theta, \phi)$ it does not depend on t and ϕ. As a consequence, we can derive, from the geodesic equations, two constants of the motion: the energy E and the angular momentum L:

$$E = (1 + \lambda\varphi)^2\frac{\mathrm{d}t}{\mathrm{d}s} \qquad L = (1 + \lambda\varphi)^2 r^2 \frac{\mathrm{d}\phi}{\mathrm{d}s}. \tag{7.38}$$

The motion is restricted to a plane, which we choose to be the equatorial plane $\theta = \pi/2$, whereupon the Hamilton–Jacobi equation takes the form

$$(1 + \lambda\varphi)^{-2}\left[\left(\frac{\partial I}{\partial t}\right)^2 - \left(\frac{\partial I}{\partial r}\right)^2 - \frac{1}{r^2}\left(\frac{\partial I}{\partial \phi}\right)^2\right] - m^2 = 0.$$

Looking for a solution of this equation in the form

$$I = -Et + L\phi + I_r \tag{7.39}$$

we find that

$$\begin{aligned} I_r &= \int [E^2 - L^2/r^2 - m^2(1 + \lambda\varphi)^2]^{1/2}\,\mathrm{d}r \\ &= \int \left[E^2 - m^2 + \frac{2GMm^2}{r} - \frac{1}{r^2}(L^2 + G^2M^2m^2)\right]^{1/2} \mathrm{d}r. \end{aligned}$$

The change of the coefficient multiplying the factor r^{-2} (with respect to Newton's value L^2) leads to a systematic precession of planetary orbits. Since the equation of an orbit is $\phi + \partial I_r/\partial L =$ constant, the total change in ϕ per revolution is

$$\Delta\phi = -\frac{\partial}{\partial L}\Delta I_r$$

where ΔI_r is the related change in I_r. Expanding I_r in powers of the small parameter $\eta = G^2M^2m^2$ we find that

$$\Delta I_r = \Delta I_r^{(0)} + \frac{\eta}{2L}\frac{\partial}{\partial L}\Delta I_r^{(0)}$$

where $I_r^{(0)}$ corresponds to the elliptical motion. Differentiation of this equation with respect to L, with $-\partial\Delta I_r^{(0)}/\partial L = \Delta\phi^{(0)} = 2\pi$, leads to

$$\Delta\phi = 2\pi - \frac{\eta}{L^2}\pi.$$

The second term gives the value of the planetary precession per revolution,

$$\delta\phi = -\frac{G^2M^2m^2}{L^2}\pi \tag{7.40}$$

and it is equal to $-1/6$ times the prediction of GR.

Thus, the scalar theory is not satisfying, as it leads to no bending of light and no radar echo delay, and the perihelion precession of Mercury's orbit does not agree with observations.

Symmetric tensor gravitational field

The tensor theory of gravity is based on

$$\mathcal{L} = \mathcal{L}_{\mathrm{T}} + \mathcal{L}_{\mathrm{M}} + \mathcal{L}_{\mathrm{I}} \tag{7.41}$$

where $\mathcal{L}_{\mathrm{T}}$ is the Lagrangian (7.19), and the interaction term is given as

$$\mathcal{L}_{\mathrm{I}} = -\lambda\varphi_{\mu\nu}T^{\mu\nu}. \tag{7.42}$$

1. The field equations for $\varphi^{\mu\nu}$ in the Hilbert gauge have the form (7.22), with the replacement $J_{\mu\nu} \to \lambda T_{\mu\nu}$. A particular solution for $\varphi^{\mu\nu}$ is given by

$$\begin{aligned}\varphi^{\mu\nu}(x) &= \lambda\int \mathrm{d}^4x'\, D^{\mu\nu,\lambda\rho}(x-x')T_{\lambda\rho}(x')\\ &= \lambda\int \mathrm{d}^4x'\, D(x-x')\overline{T}^{\mu\nu}(x')\end{aligned}$$

where $\overline{T}^{\mu\nu} = T^{\mu\nu} - \frac{1}{2}\eta^{\mu\nu}T$. For a static source, $T^{\mu\nu}$ does not depend on x'_0 and $T^{\mu\alpha} = T^{\alpha\beta} = 0$, so that

$$\begin{aligned}\varphi^{00}(\boldsymbol{x}) &= \lambda\int \mathrm{d}^3x'\, D(\boldsymbol{x}-\boldsymbol{x}')\tfrac{1}{2}T(\boldsymbol{x}')\\ \varphi^{\alpha\beta}(\boldsymbol{x}) &= -\eta^{\alpha\beta}\varphi^{00}(\boldsymbol{x})\\ \varphi^{\mu 0}(\boldsymbol{x}) &= 0.\end{aligned} \tag{7.43a}$$

If the source is a single particle of mass M, $T(\boldsymbol{x}) = M\delta(\boldsymbol{x})$, then

$$\varphi^{00}(\boldsymbol{x}) = -\frac{\lambda}{8\pi}\frac{M}{r}. \tag{7.43b}$$

The interaction energy of a test particle of mass m with the static field is

$$E(x^0) = \lambda \int \mathrm{d}^3x\, T'_{\mu\nu}(x)\varphi^{\mu\nu}(x) = -\frac{\lambda^2}{4\pi}\frac{M}{r}\left(p_0 - \frac{m^2}{2p_0}\right). \tag{7.44}$$

For a non-relativistic motion $p_0 \approx m$, and we obtain *Newton's law* provided that $\lambda^2/8\pi = G$.

The universality of λ is in accordance with the PE, and the correct *redshift* formula is obtained from Newton's law, as before.

The part of the action containing matter variables can be written as

$$\begin{aligned} I_{\mathrm{MI}} &= -m \int [\eta_{\mu\nu} + \lambda\varphi_{\mu\nu}(z)]\frac{\mathrm{d}z^\mu}{\mathrm{d}s}\frac{\mathrm{d}z^\nu}{\mathrm{d}s}\,\mathrm{d}s \\ &= -m \int [(\eta_{\mu\nu} + 2\lambda\varphi_{\mu\nu})\,\mathrm{d}z^\mu\,\mathrm{d}z^\nu]^{1/2} + \mathcal{O}(\lambda^2). \end{aligned} \tag{7.45}$$

If we introduce the quantity

$$g_{\mu\nu} = \eta_{\mu\nu} + 2\lambda\varphi_{\mu\nu} \tag{7.46}$$

the matter equations of motion become geodesic equations in Riemann space with metric $g_{\mu\nu}$.

2. To calculate the *precession* of planetary orbits we again use the Hamilton–Jacobi equation (7.37), with the metric produced by a static pointlike source. It has the form, for $\theta = \pi/2$,

$$(1 - r_g/r)^{-1}\left(\frac{\partial I}{\partial t}\right)^2 - (1 + r_g/r)^{-1}\left[\left(\frac{\partial I}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial I}{\partial \phi}\right)^2\right] - m^2 = 0$$

where $r_g = 2GM$. Searching for I in the form (7.39) we find that

$$I_r = \int \mathrm{d}r\left[\frac{1 + r_g/r}{1 - r_g/r}E^2 - \frac{L^2}{r^2} - m^2(1 + r_g/r)\right]^{1/2}.$$

Expanding I_r in powers of r_g/r we can identify the term of the type r^{-2} as

$$-\frac{1}{r^2}(L^2 - 2m^2r_g^2) = -\frac{1}{r^2}(L^2 - 8G^2m^2M^2).$$

This leads directly to the value of the planetary perihelion precession per revolution,

$$\delta\phi = \frac{8G^2M^2m^2}{L^2}\pi \tag{7.47}$$

which is 4/3 of the result from GR, and does not agree with the observational data (for Mercury, $\delta\phi = 43.03''$/century).

3. The motion of *light signals* in the gravitational field (7.43) can be described by the Hamilton–Jacobi equation for an action ψ,

$$g^{\mu\nu}\frac{\partial\psi}{\partial x^\mu}\frac{\partial\psi}{\partial x^\nu} = 0 \tag{7.48}$$

which is deduced from equation (7.37) in the limit $m^2 \to 0$. The radial part of the action ψ_r follows directly from I_r in the same limit:

$$\psi_r = \int \mathrm{d}r \left(\frac{1 + r_g/r}{1 - r_g/r}\omega^2 - \frac{L^2}{r^2}\right)^{1/2}$$

where the energy of the light signal is denoted by ω.

The time dependence $r = r(t)$ for the trajectory of a light ray is defined by the relation $-t + \partial\psi_r/\partial\omega = \text{constant}$, and has the form

$$t = \int \mathrm{d}r \, \frac{1 + r_g/r}{1 - r_g/r}\left(\frac{1 + r_g/r}{1 - r_g/r} - \frac{\rho^2}{r^2}\right)^{-1/2}$$

where $\rho = L/\omega$. Furthermore, the radial velocity $\mathrm{d}r/\mathrm{d}t$ must vanish at the distance $r = b$ of closest approach to the sun, so that

$$\rho^2 = b^2\frac{1 + r_g/b}{1 - r_g/b}.$$

Thus, ρ is equal to b only in the lowest order approximation in r_g/b.

(i) We first study the *deflection of light* in the field of the sun. At great distances we can ignore the gravitational interaction in ψ_r ($r_g \to 0$), and the equation of motion, $\phi + \partial\psi_r/\partial L = \text{constant}$, predicts motion on a straight line $r = \rho/\cos\phi$, passing at distance $r = \rho$ from the sun (figure 7.2).

In the first approximation in r_g/b the radial action is given by

$$\psi_r = \omega \int \mathrm{d}r \left(1 + \frac{2r_g}{r} - \frac{\rho^2}{r^2}\right)^{1/2}.$$

Expanding the integrand in powers of r_g/r gives

$$\psi_r \simeq \psi_r^{(0)} + \omega r_g \,\mathrm{Arcosh}(r/\rho)$$

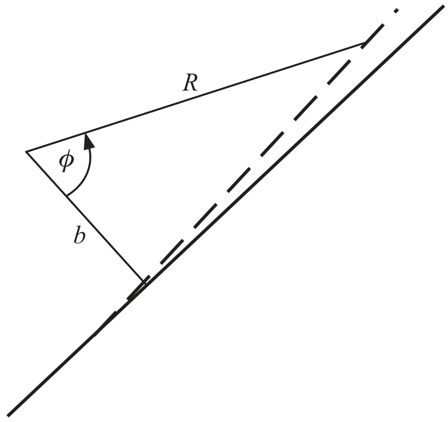

Figure 7.2. The motion of the electromagnetic signal in the central gravitational field.

where $\psi_r^{(0)}$ corresponds to a rectilinear motion of the light ray:

$$\psi_r^{(0)} = \omega \int \mathrm{d}r\, \sqrt{1 - \rho^2/r^2} = \omega\sqrt{r^2 - \rho^2} - \omega\rho\, \mathrm{Arccos}(\rho/r).$$

The total change in ψ_r as the light ray goes from a distance $r = R$ ($\phi < 0$) to the point $r = b$, and then to $r = R$ ($\phi > 0$), is given by

$$\Delta\psi_r = \Delta\psi_r^{(0)} + 2\omega r_g\, \mathrm{Arcosh}(R/\rho).$$

The related change in ϕ is obtained by differentiating $\Delta\psi_r$ with respect to $L = \rho\omega$:

$$\Delta\phi = -\frac{\partial}{\partial L}\Delta\psi_r = \Delta\phi^{(0)} + \frac{2r_g R}{\rho\sqrt{R^2 - \rho^2}}$$

where $\Delta\phi^{(0)} = -\partial\Delta\psi_r^{(0)}/\partial L$. In the limit $R \to \infty$ we recover the same result as in GR:

$$\Delta\phi = \pi + \frac{2r_g}{\rho}.$$

Since $\Delta\phi > \pi$, the trajectory is bent toward the sun by an angle

$$\delta\phi = \frac{2r_g}{b} = \frac{4MG}{b} \tag{7.49}$$

where we used $\rho \approx b$. For a light ray just touching the sun's disc ($b = R_\mathrm{S}$), the deflection is $\delta\phi = 1.75''$, in agreement with observations.

(ii) We now want to calculate the time interval needed for a radar signal to travel from $r = b$ to $r = R$ ($\phi > 0$) (figure 7.2). From the change

$$\Delta\psi_r = \Delta\psi_r^{(0)} + \omega r_g\, \mathrm{Arcosh}(R/\rho)$$

the related time interval is given as

$$\Delta t = \frac{\partial}{\partial\omega}\Delta\psi_r = \frac{\partial}{\partial\omega}\Delta\psi_r^{(0)} + r_g \operatorname{Arcosh}(R/\rho) + r_g\frac{R-\rho}{\sqrt{R^2-\rho^2}}$$
$$= \Delta t^{(0)} + \delta t.$$

Here, the term $\Delta t^{(0)} = \sqrt{R^2-\rho^2}$ corresponds to a rectilinear motion of the radar signal, and the remaining terms represent the gravitational delay δt. A direct comparison with Einstein's theory gives

$$\delta t = \delta t^E + \frac{1}{2}r_g\left(\frac{R-\rho}{R+\rho}\right)^{1/2}. \tag{7.50}$$

When we consider a radar signal travelling from Earth to the sun (the distance R_1), and then to Mercury (or some other planet, at the distance R_2 from the sun), the complete delay is given as ($\rho \approx b$)

$$\delta t_1 + \delta t_2 = \delta t_1^{\mathrm{E}} + \delta t_2^{\mathrm{E}} + \frac{1}{2}r_g\left[\left(\frac{R_1-b}{R_1+b}\right)^{1/2} + \left(\frac{R_2-b}{R_2+b}\right)^{1/2}\right].$$

Einstein's part is at a maximum when Mercury is 'behind' the sun, and the radar signal just grazes the sun ($b = R_{\mathrm{S}}$). The total round-trip excess in this case is 72 km $\simeq 240\ \mu$s, and it agrees with the radar data (although the realization and interpretation of these experiments is extraordinarily difficult). The remaining two terms are just a small correction ($b \ll R_1, R_2$), which is of no significance for the basic conclusion.

Can the graviton have a mass?

The previous results show that a massless field of spin 2 describes standard gravitational effects with significant success. Can a graviton have a very small but still finite mass μ? We shall see that this is not possible without a notable violation of the agreement with experiments. The reason for this somewhat unexpected result lies in the fact that the propagator of the massive graviton differs from the propagator of massless graviton even in the limit $\mu^2 \to 0$ (Van Dam and Veltman 1970, Boulware and Deser 1972, Schwinger 1970).

The theory of massive field of spin 2 is defined by the Lagrangian

$$\mathcal{L} = \mathcal{L}_{\mathrm{T}} - \tfrac{1}{2}\mu^2(\varphi_{\mu\nu}\varphi^{\mu\nu} - \varphi^2). \tag{7.51}$$

Note the form of the massive term, which contains not only $\varphi_{\mu\nu}\varphi^{\mu\nu}$, but also φ^2. The field equation has the form

$$(-\Box - \mu^2)\varphi_{\mu\nu} + \varphi_{\mu\rho,\nu}{}^{\rho} + \varphi_{\nu\rho,\mu}{}^{\rho} - \varphi_{,\mu\nu} - \eta_{\mu\nu}[\varphi_{\sigma\rho}{}^{,\sigma\rho} + (-\Box-\mu^2)\varphi] = \lambda T_{\mu\nu}. \tag{7.52a}$$

Taking the trace and differentiating with respect to x^μ we find that

$$\varphi = \frac{\lambda}{3\mu^2}\left(T - \frac{2}{\mu^2}T_{\sigma\rho}{}^{,\sigma\rho}\right)$$
$$\partial^\mu\varphi_{\mu\nu} = -\tfrac{\lambda}{\mu^2}\partial^\mu T_{\mu\nu} + \tfrac{\lambda}{3\mu^2}\partial_\nu\left(T - \tfrac{2}{\mu^2}T_{\sigma\rho}{}^{,\sigma\rho}\right). \tag{7.52b}$$

The field $\varphi_{\mu\nu}$ in vacuum satisfies the field equations characterizing the massive field of spin 2,

$$(-\Box - \mu^2)\varphi_{\mu\nu} = 0 \qquad \partial^\mu\varphi_{\mu\nu} = 0 \qquad \varphi = 0$$

which justifies the form of the Lagrangian (7.51).

Using the expressions (7.52b) the field equation becomes

$$(-\Box - \mu^2)\varphi_{\mu\nu} = \lambda T_{\mu\nu} + \frac{\lambda}{\mu^2}(\partial_\mu T_{\lambda\nu}{}^{,\lambda} + \partial_\nu T_{\lambda\mu}{}^{,\lambda})$$
$$- \frac{\lambda}{\mu^2}\eta_{\mu\nu}T_{\sigma\rho}{}^{,\sigma\rho} - \frac{\lambda}{3}\left(\eta_{\mu\nu} + \frac{\partial_\mu\partial_\nu}{\mu^2}\right)\left(T - \frac{2}{\mu^2}T_{\sigma\rho}{}^{,\sigma\rho}\right).$$

Going now to momentum space, we find a solution

$$\varphi_{\mu\nu}(k) = D_{\mu\nu,\lambda\rho}(k,\mu^2)\lambda T^{\lambda\rho} \tag{7.53a}$$

where $D(k,\mu^2)$ is the propagator of the massive field of spin 2:

$$D_{\mu\nu,\lambda\rho}(k,\mu^2) = \frac{1}{k^2-\mu^2}[\tfrac{1}{2}(P_{\mu\lambda}P_{\nu\rho} + P_{\mu\rho}P_{\nu\lambda}) - \tfrac{1}{3}P_{\mu\nu}P_{\lambda\rho}]. \tag{7.53b}$$

Since $P_{\mu\nu}$ is the projector on the three-dimensional space of spin-1 states, the field $\varphi_{\mu\nu}$ has six independent components, and the condition $\varphi = 0$ for $k^2 = \mu^2$ reduces that number to five, which corresponds to the massive field of spin 2.

Polarization states. A plane wave of momentum k,

$$\varphi_{\mu\nu}(x) = e_{\mu\nu}e^{ik\cdot x} + e^*_{\mu\nu}e^{-ik\cdot x}$$

obeys the field equations in vacuum if

$$k^2 - \mu^2 = 0 \qquad k^\mu e_{\mu\nu} = 0 \qquad \eta^{\mu\nu}e_{\mu\nu} = 0. \tag{7.54}$$

The last two relations in the rest frame take the form

$$e_{0\nu} = 0 \qquad e_{11} + e_{22} + e_{33} = 0.$$

Due to these five conditions there is only $10 - 5 = 5$ independent tensors $e_{\mu\nu}^{(\lambda)}$, and they can be chosen so as to be orthonormal ($e_{\mu\nu}^{(\lambda)} e^{(\lambda')\mu\nu} = \delta^{\lambda\lambda'}$):

$$
\begin{aligned}
e_{(1)} &= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad e_{(2)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
e_{(3)} &= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \qquad e_{(4)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \\
e_{(0)} &= \sqrt{\frac{2}{3}} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}
\end{aligned}
\tag{7.55}
$$

where $e_{(\lambda)}$ denotes the matrix $(e_{(\lambda)}^{\mu\nu})$.

Under rotations around the z-axis we obtain the transformation rules

$$
e'_{(\pm 2)} = \mathrm{e}^{\pm 2\mathrm{i}\theta} e_{(\pm 2)} \qquad e'_{(\pm 1)} = \mathrm{e}^{\pm \mathrm{i}\theta} e_{(\pm 1)} \qquad e'_{(0)} = e_{(0)} \tag{7.56}
$$

where $e_{(\pm 2)}$, $e_{(\pm 1)}$ and $e_{(0)}$ are the polarization states of massive field of spin 2, with spin projections $s_3 = \pm 2, \pm 1$ and 0,

$$
e_{(\pm 2)} = \frac{1}{\sqrt{2}}(e_{(1)} \pm \mathrm{i} e_{(2)}) \qquad e_{(\pm 1)} = \frac{1}{\sqrt{2}}(e_{(3)} \pm \mathrm{i} e_{(4)}).
$$

Currents. Now we introduce currents with spin projections $s_3 = \pm 1$,

$$
T_{(\pm 1)} \equiv e_{(\pm 1)}^{\mu\nu} T_{\mu\nu} = \sqrt{2}(T_{13} \pm \mathrm{i} T_{23})
$$

with an interaction between them

$$
\sum_{\lambda = \pm 1} T'_{(\lambda)} T^*_{(\lambda)} = 2\sqrt{2}(T'_{13} T_{13} + T'_{23} T_{23}).
$$

In the case of conserved currents the components T_{13} and T_{23} can be expressed in terms of T_{10} and T_{20} ($k^0 T_{0\nu} = -k^3 T_{3\nu}$), so that the previous expression does not contribute to the retarded interaction.

Similarly, the current with $s_3 = 0$ is

$$
\begin{aligned}
T_{(0)} &\equiv e_{(0)}^{\mu\nu} T_{\mu\nu} = \sqrt{\tfrac{2}{3}}[\tfrac{1}{2}(T_{11} + T_{22}) - T_{13}] \\
T'_{(0)} T^*_{(0)} &= \tfrac{2}{3}[\tfrac{1}{4}(T'_{11} + T'_{22})(T_{11} + T_{22}) + T'_{13} T_{13} \\
&\quad - \tfrac{1}{2}(T'_{11} + T'_{22}) T_{13} - \tfrac{1}{2} T'_{13}(T_{11} + T_{22})]
\end{aligned}
$$

where the first term corresponds to the retarded interaction,

$$\tfrac{1}{6}(T'_{11}+T'_{22})(T_{11}+T_{22}). \tag{7.57}$$

The polarization sum follows from the equality

$$\sum_{\lambda} T'_{(\lambda)}T^{*}_{(\lambda)} = T'_{\mu\nu}\bigg(\sum_{\lambda=0}^{4} e^{\mu\nu}_{(\lambda)}e^{\sigma\rho *}_{(\lambda)}\bigg)T_{\sigma\rho} \equiv T'_{\mu\nu}P^{\mu\nu,\sigma\rho}T_{\sigma\rho}$$

where the left-hand side is calculated using equations (7.28) and (7.57), that describe the contributions of states with $s_3 = \pm 2$ and 0. The result is

$$P^{\mu\nu,\sigma\rho} = \begin{cases} \frac{1}{2}(\eta^{\mu\sigma}\eta^{\nu\rho}+\eta^{\mu\rho}\eta^{\nu\sigma}) - \frac{1}{3}\eta^{\mu\nu}\eta^{\sigma\rho} & \text{for } \mu,\nu,\sigma,\rho \neq 0,3; \\ 0 & \text{at least one index} = 0 \text{ or } 3. \end{cases}$$

This is the result in the rest frame. In the case of a general momentum, the result follows from the first line by the replacement $\eta^{\mu\nu} \to P^{\mu\nu}$:

$$\begin{aligned} P^{\mu\nu,\sigma\rho} &= \tfrac{1}{2}(\eta^{\mu\sigma}\eta^{\nu\rho}+\eta^{\mu\rho}\eta^{\nu\sigma}) - \tfrac{1}{3}\eta^{\mu\nu}\eta^{\sigma\rho} \\ &\quad + k\text{-dependent terms.} \end{aligned} \tag{7.58}$$

Mass discontinuity. How can we relate this theory to the case $\mu^2 = 0$? It is clear that the propagator of the massive field is not defined in the limit $\mu^2 \to 0$, since then the Lagrangian becomes gauge invariant. But, if the massive field is coupled to a conserved current, $k^\mu T_{\mu\nu} = 0$ (which is necessary if the limiting case is to be consistent), then the amplitude of the interaction between two currents has the form

$$M_{21} = T_2^{\mu\nu} D^{\mathrm{E}}_{\mu\nu,\sigma\rho}(k,\mu^2) T_1^{\sigma\rho}$$

where D^{E} is the effective propagator obtained from (7.53*b*) by ignoring all terms proportional to k_μ:

$$D^{\mathrm{E}}_{\mu\nu,\sigma\rho}(k,\mu^2) = \frac{1}{k^2-\mu^2}[\tfrac{1}{2}(\eta_{\mu\sigma}\eta_{\nu\rho}+\eta_{\mu\rho}\eta_{\nu\sigma}) - \tfrac{1}{3}\eta_{\mu\nu}\eta_{\sigma\rho}]. \tag{7.59}$$

This propagator differs from the effective propagator of the massless theory

$$D^{\mathrm{E}}_{\mu\nu,\sigma\rho}(k) = \frac{1}{k^2}[\tfrac{1}{2}(\eta_{\mu\sigma}\eta_{\nu\rho}+\eta_{\mu\rho}\eta_{\nu\sigma}) - \tfrac{1}{2}\eta_{\mu\nu}\eta_{\sigma\rho}]$$

even for $\mu^2 = 0$.

What is the origin of this *discontinuity* in the graviton mass? In order to explain this phenomenon, let us consider what happens with the massive graviton for $\mu^2 \to 0$. The massive graviton has five degrees of freedom. When $\mu^2 \to 0$ it gives massless particles of helicities $\lambda = \pm 2, \pm 1$ and 0. While particles with $\lambda = \pm 1$ are decoupled from the conserved current, this is not the case with the scalar particle. Indeed, it is the contribution of the scalar particle which changes the propagator of the massless graviton just so as to produce the effective propagator of the massive graviton in the limit $\mu^2 = 0$:

$$-\tfrac{1}{2}\eta_{\mu\nu}\eta_{\sigma\rho} + \tfrac{1}{6}\eta_{\mu\nu}\eta_{\sigma\rho} = -\tfrac{1}{3}\eta_{\mu\nu}\eta_{\sigma\rho}.$$

It is evident that experimental consequences of the massless theory and the theory with an arbitrarily small graviton mass differ in a *discrete* way.

Experimental predictions. Let us now calculate the interaction constant of the massive theory by demanding that the static low-energy limit gives Newton's law. The interaction energy of the massless theory is

$$E = \lambda^2 T'_{00}\frac{1}{k^2}(1 - \tfrac{1}{2})T_{00}.$$

Using equation (7.53) we find an analogous expression for the massive theory:

$$E = \lambda^2(\mu) T'_{00}\frac{1}{k^2}(1 - \tfrac{1}{3})T_{00}.$$

If λ^2 is determined correctly ($\lambda^2 = 8\pi G$), then we must have

$$\lambda^2(\mu) = \tfrac{3}{4}\lambda^2 \tag{7.60}$$

if the massive theory is to have a good Newton limit. But with this result for $\lambda^2(\mu)$ other experimental consequences become incorrect.

When we calculated these consequences in the massless theory, the sun was treated as a static pointlike source of the gravitational field $\lambda(\mu)\varphi_{\mu\nu}$, given by equation (7.43). In the massive theory the field $\lambda(\mu)\varphi_{\mu\nu}$ can be easily calculated from the expression (7.53):

$$\begin{aligned}\lambda(\mu)\varphi^{00} &= -\frac{2}{3}\frac{\lambda^2(\mu)}{4\pi}\frac{M}{r}\\ \lambda(\mu)\varphi^{\alpha\beta} &= \eta^{\alpha\beta}\frac{1}{3}\frac{\lambda^2(\mu)}{4\pi}\frac{M}{r}.\end{aligned} \tag{7.61}$$

While $\lambda(\mu)\varphi^{00}$ has the same value as in the massless theory as a consequence of (7.60), $\lambda(\mu)\varphi^{\alpha\beta}$ differs by a factor $\frac{1}{2}$. The calculation of the precession of planetary orbits in the massive theory gives two-thirds of the result of the massless theory.

The results for light deflection and radar echo delay can be obtained in a simpler way. Consider the interaction of an electromagnetic signal with the sun. Since the trace of the electromagnetic energy–momentum tensor vanishes, the last term in the effective propagator $D^{\mathrm{E}}(k, \mu^2)$ does not contribute to the interaction amplitude. Since the same is true for the last term in the propagator of the massless field, the difference in interaction amplitudes stems only from the different coupling constants. Hence, the ratio of the amplitudes is

$$M_{21}(\mu^2)/M_{21} = \lambda^2(\mu)/\lambda^2 = \tfrac{3}{4}.$$

The massive graviton produces light deflection by the sun and radar echo delay which are by factor $\frac{3}{4}$ different from the corresponding predictions of the massless theory.

These results strongly indicate that the existence of an arbitrarily small but finite mass of the graviton of spin 2 is excluded by the observations.

The consistency problem

Previous considerations show that the graviton is most successfully described as a massless particle of spin-2. In studying various alternatives we neglected to check the internal consistency of the theory.

Our theory of gravity (7.41) contains the interaction $\mathcal{L}_{\mathrm{I}} = -\lambda\varphi_{\mu\nu}T^{\mu\nu}$, where $T^{\mu\nu}$ is the energy–momentum tensor of *matter*. We may immediately ask the following question: Why does the energy–momentum not include a contribution from the gravitational field itself? We shall see that the suggestion contained in this question is not only possible, but also necessary for a self-consistent formulation of the theory.

Consider the field equation $K_{\mu\nu,\sigma\rho}\varphi^{\sigma\rho} = \lambda T_{\mu\nu}$. Since the operator $K_{\mu\nu,\sigma\rho}$ is singular, taking the derivative of this equation gives the consistency condition

$$\partial^\mu T_{\mu\nu} = 0 \tag{7.62}$$

which implies a free, rectilinear motion for the particle (on the basis of the momentum conservation). It is clear that this condition is not fulfilled for the *interacting* system of matter and gravity. Indeed, from the expression (7.31*c*) for $T_{\mu\nu}$ we find

$$\partial_\mu T^{\mu\nu}(x) = M\int \frac{\mathrm{d}^2 z^\nu}{\mathrm{d}s^2}\delta(x - z(s))\,\mathrm{d}s.$$

Then, using the equation of motion for matter,

$$\frac{\mathrm{d}^2 z^\nu}{\mathrm{d}s^2} = -\tfrac{1}{2}g^{\nu\sigma}(g_{\sigma\lambda,\rho} + g_{\sigma\rho,\lambda} - g_{\lambda\rho,\sigma})\frac{\mathrm{d}z^\lambda}{\mathrm{d}s}\frac{\mathrm{d}z^\rho}{\mathrm{d}s}$$

where $g_{\mu\nu} = \eta_{m\nu} + 2\lambda\varphi_{\mu\nu}$, we obtain

$$\begin{aligned}(\eta_{\nu\sigma} + 2\lambda\varphi_{\nu\sigma})\partial_\mu T^{\mu\nu} &= -[\lambda\rho,\sigma]T^{\lambda\rho}\\ [\lambda\rho,\sigma] &\equiv \lambda(\varphi_{\sigma\lambda,\rho} + \varphi_{\sigma\rho,\lambda} - \varphi_{\lambda\rho,\sigma}).\end{aligned} \tag{7.63}$$

Thus, condition (7.62) is violated to order λ.

In further exposition we shall attempt to correct this inconsistency by replacing the matter energy–momentum with the complete energy–momentum for the whole system *matter + gravity*. The consistency condition will be satisfied as a consequence of the conservation of the total energy–momentum.

Exercises

1. Show that the three-dimensional propagator of scalar field theory has the form
$$D(\boldsymbol{x}) = -\frac{1}{4\pi|\boldsymbol{x}|}.$$
2. Show that $P_{\mu\nu} = \eta_{\mu\nu} - k_\mu k_\nu/\mu^2$ and $P^\perp_{\mu\nu} = k_\mu k_\nu/\mu^2$ are projectors on the massive states of spin 1 and 0, respectively, i.e. that for $k^2 = \mu^2$ the following relations hold:
$$\begin{aligned} P_{\mu\sigma}P^{\sigma\nu} &= P_\mu{}^\nu \qquad & P^\perp_{\mu\sigma}P^{\perp\sigma\nu} &= P^{\perp\nu}_\mu \\ P_{\mu\nu} + P^\perp_{\mu\nu} &= \eta_{\mu\nu} \qquad & P^\perp_{\mu\sigma}P^{\sigma\nu} &= 0 \\ \eta^{\mu\nu}P_{\mu\nu} &= 3 \qquad & \eta^{\mu\nu}P^\perp_{\mu\nu} &= 1. \end{aligned}$$
3. The quantities $\Pi_{\mu\nu} = \eta_{\mu\nu} - (k_\mu\bar{k}_\nu + k_\nu\bar{k}_\mu)/k\cdot\bar{k}$, $\Pi^\perp_{\mu\nu} = (k_\mu\bar{k}_\nu + k_\nu\bar{k}_\mu)/k\cdot\bar{k}$, where $\bar{k} = (k^0, -\boldsymbol{k})$, are projectors on massless states of helicities ± 1 and 0, respectively. Prove the following relations, for $k^2 = 0$:
$$\begin{aligned} \Pi_{\mu\sigma}\Pi^{\sigma\nu} &= \Pi^\nu_\mu \qquad & \Pi^\perp_{\mu\sigma}\Pi^{\perp\sigma\nu} &= \Pi^{\perp\nu}_\mu \\ \Pi_{\mu\nu} + \Pi^\perp_{\mu\nu} &= \eta_{\mu\nu} \qquad & \Pi^\perp_{\mu\sigma}\Pi^{\sigma\nu} &= 0 \\ \eta^{\mu\nu}\Pi_{\mu\nu} &= 2 \qquad & \eta^{\mu\nu}\Pi^\perp_{\mu\nu} &= 2. \end{aligned}$$
4. Find the propagator $D'_{\mu\nu}(k)$ of the massless vector field, defined by the gauge-breaking term $-\alpha(\partial\cdot\varphi)^2/2$.
5. The Lagrangian
$$\mathcal{L}''_{\mathrm{V}} = \mathcal{L}_{\mathrm{V}} + \frac{1}{2}\mu^2\varphi_\nu\varphi^\nu - \frac{\alpha}{2}(\partial\cdot\varphi)^2 - \varphi_\mu J^\mu$$
where $\partial\cdot J = 0$, is convenient for studying the limit $\mu^2 \to 0$.
   (a) Show that $\partial\cdot\varphi$ obeys the Klein–Gordon equation with mass $m^2 = \mu^2/\alpha$.
   (b) Find the related propagator $D''_{\mu\nu}(k)$ and examine its behaviour in the cases (i) $\alpha \neq 0$, $\mu^2 \to 0$; (ii) $\mu^2 \neq 0$, $\alpha \to 0$.
6. Prove that the polarization state of a massive vector field with $s_3 = 0$ is decoupled from the conserved current J^μ in the limit $\mu^2 \to 0$.
7. Consider a massless field $\varphi_{\mu\nu}$ of helicity ± 2. Prove the following properties:
   (a) The Hilbert gauge condition is locally accessible.
   (b) Free gravitational waves must satisfy the Hilbert condition for $k^2 = 0$.
8. Find the propagator for the massless tensor field, determined by the gauge-breaking term $-\alpha(\partial_\mu\bar{\varphi}^{\mu\nu})^2/2$. Examine the cases $\alpha = 1$ and $\alpha \to \infty$.
9. Show that the projector on massless states of helicity ± 2,
$$\Pi_{\mu\nu,\lambda\rho} = \tfrac{1}{2}(\Pi_{\mu\lambda}\Pi_{\nu\rho} + \Pi_{\mu\rho}\Pi_{\nu\lambda}) - \tfrac{1}{2}\Pi_{\mu\nu}\Pi_{\lambda\rho}$$

satisfies the following relations, for $k^2 = 0$:

$$\Pi_{\mu\nu,\lambda\rho}\Pi^{\lambda\rho,\sigma\tau} = \Pi_{\mu\nu}{}^{,\sigma\tau} \qquad \eta^{\mu\lambda}\eta^{\nu\rho}\Pi_{\mu\nu,\lambda\rho} = 2$$
$$\eta^{\mu\nu}\Pi_{\mu\nu,\lambda\rho} = 0.$$

10. Find the trajectory $u = u(\phi)$, $u \equiv 1/r$, that describes the motion of a test particle in the spherically symmetric scalar gravitational field. Discuss the limit $G \to 0$.
11. Use the result of exercise 10 to calculate the precession of planetary orbits in the scalar theory of gravity.
12. Find constants of the motion of a test particle in the spherically symmetric gravitational field, in both the (a) scalar and (b) tensor theory of gravity.
13. Find the trajectory $u = u(\phi)$, $u \equiv 1/r$, describing the motion of a test particle in the spherically symmetric tensor gravitational field.
14. Use the result of exercise 13 to calculate:
    (a) the change of angle $\Delta\phi$ during the motion of a light ray from $r = R$ $(\phi < 0)$ to $r = R$ $(\phi > 0)$ (figure 7.2);
    (b) the coordinate time Δt required for a light ray to travel from $r = b$ to $r = R$.
15. (a) Calculate the interaction energy of a test particle of mass m with the static tensor gravitational field produced by a pointlike source.
    (b) Show that the interaction energy of a photon in this field is given by

$$E(x^0) = -2G\frac{M}{r}p^0.$$

    Find the corresponding expression in Newton's theory (using the replacement: the mass of a test particle → the total photon energy).
16. A light ray is passing the sun at distance b (figure 7.2). Assume that the sun's gravitational field is the same as that described in exercise 15, and calculate the deflection angle of a light ray as the ratio of the transversal and longitudinal components of the momentum.
17. Prove that the projector on the polarization states of a massive spin-2 field,

$$P_{\mu\nu,\lambda\rho} = \tfrac{1}{2}(P_{\mu\lambda}P_{\nu\rho} + P_{\mu\rho}P_{\nu\lambda}) - \tfrac{1}{3}P_{\mu\nu}P_{\lambda\rho}$$

    obeys the following conditions, for $k^2 = \mu^2$:

$$P_{\mu\nu,\lambda\rho}P^{\lambda\rho,\lambda\tau} = P_{\mu\nu}{}^{,\lambda\tau} \qquad \eta^{\mu\lambda}\eta^{\nu\rho}P_{\mu\nu,\lambda\rho} = 5$$
$$\eta^{\mu\nu}P_{\mu\nu,\lambda\rho} = 0.$$

18. Calculate the contribution of the scalar component of the massive tensor graviton to the effective propagator, in the limit $\mu^2 \to 0$.
19. Show that the precession of planetary orbits in the massive tensor theory of gravity, in the limit $\mu^2 \to 0$, gives two-thirds the result of the massless tensor theory.

Chapter 8

Nonlinear effects in gravity

As we have seen, in ordinary flat spacetime the graviton is best described as a massless field of spin-2, but its interaction with matter is not consistent. A closer inspection shows that this inconsistency stems from the fact that the graviton itself has an energy–momentum, which has to be included in a complete theory. The related correction leads to nonlinear effects: the energy–momentum of matter is a source of the gravitational field, the energy–momentum of which becomes a source of the new field, etc. Can this nonlinear correction account for the small discrepancy in the precession of the perihelion of Mercury? Surprisingly enough, the answer is yes. Moreover, the 'simplest' solution to the nonlinear self-coupling problem leads to a theory which is identical to Einstein's GR. This result gives an unexpected geometric interpretation to the field-theoretic approach.

In order to express the essential features of this problem and its resolution in the most simple manner, we first consider an analogous problem in Yang–Mills theory. Then we go over to the gravitational field by studying, first, scalar and then tensor theory (Feynman *et al* 1995, Van Dam 1974, Okubo 1978). Particular attention is devoted to the first order formalism, which significantly simplifies the whole method (Deser 1970).

8.1 Nonlinear effects in Yang–Mills theory

The nonlinear effects that occur in constructing a consistent Yang–Mills theory are very similar to what happens with gravity. We shall therefore study this case in detail, analysing separately the role of nonlinear effects in the pure Yang–Mills sector and in matter and interaction sectors (Okubo 1978).

Non-Abelian Yang–Mills theory

Consider an $SU(2)$ triplet of massless vector fields $A^a{}_\mu$ $(a = 1, 2, 3)$, and a set of matter fields Ψ belonging to some representation of $SU(2)$. The free Lagrangian of vector fields,

$$\mathcal{L}_{\rm F}^{(0)} = -\tfrac{1}{4}\overset{\circ}{F}{}^{a}{}_{\mu\nu}\overset{\circ}{F}{}^{a\mu\nu} \qquad \overset{\circ}{F}{}^{a}{}_{\mu\nu} \equiv \partial_\mu A^a{}_\nu - \partial_\nu A^a{}_\mu \tag{8.1}$$

is invariant under *global* $SU(2)$ rotations $\delta A^a{}_\mu = -\epsilon^{abc}\theta^b A^c{}_\mu$, and *local* Abelian transformations $\delta A^a{}_\mu = \partial_\mu \lambda^a$. The $SU(2)$ invariance of the matter Lagrangian $\mathcal{L}_{\rm M}$ implies the existence of the canonical (Noether) current J^a_μ.

Now, we assume that the total Lagrangian of the system has the form

$$\mathcal{L}^{(0)} = \mathcal{L}_{\rm F}^{(0)} + \mathcal{L}_{\rm M} + \mathcal{L}_{\rm I} \qquad \mathcal{L}_{\rm I} = -gA^a \cdot J^a \tag{8.2}$$

where A^a is coupled to the canonical current J^a of matter fields.

If the matter is described by an $SU(2)$ doublet of Dirac fields,

$$\Psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \qquad \delta\Psi = {\rm i}\theta^a \tfrac{1}{2}\tau^a\Psi \qquad \delta\bar\Psi = -{\rm i}\theta^a\bar\Psi\tfrac{1}{2}\tau^a \tag{8.3}$$

(τ^a are the Pauli spin matrices, $\psi_{1,2}$ are the Dirac spinors), then

$$\mathcal{L}_{\rm M} = \bar\Psi({\rm i}\gamma\cdot\partial - m)\Psi \qquad J^a_\mu = \bar\Psi\gamma^\mu\tfrac{1}{2}\tau^a\Psi. \tag{8.4}$$

The field equations for A^a have the form

$$\partial^\mu \overset{\circ}{F}{}^{a}{}_{\mu\nu} = gJ^a_\nu \tag{8.5}$$

and their consistency requires $\partial \cdot J^a = 0$. However, this condition is in conflict with the matter field equations. Indeed, the Dirac equations for Ψ imply

$$\partial \cdot J^a = g\epsilon^{abc}A^{b\nu}J^c_\nu \neq 0$$

showing an inconsistency of order g. The reason for this inconsistency lies in the fact that J^a is not the Noether current for the whole theory but only for the part stemming from the matter fields, which cannot be conserved.

In order to resolve this problem we shall try to construct a new theory, on the basis of the following two requirements:

- The new Lagrangian $\mathcal{L} = \mathcal{L}^{(0)} + \Lambda$ should give the field equation

$$\partial^\mu \overset{\circ}{F}{}^{a}{}_{\mu\nu} = g(J^a_\nu + j^a_\nu) \equiv g\mathcal{J}^a_\nu \tag{a}$$

- The dynamical current $\mathcal{J}^a$ is equal to the Noether current generated by the global $SU(2)$ symmetry of $\mathcal{L}$ and is automatically conserved:

$$\partial^\mu \mathcal{J}^a_\mu = 0. \tag{b}$$

Before starting the construction of $\mathcal{L}$, we must note that there is an ambiguity in the definition of $\mathcal{J}^a$: we can replace $\mathcal{J}^a_\nu$ by $\mathcal{J}^a_\nu + \partial^\mu W^a_{\mu\nu}$, where $W^a_{\mu\nu} = -W^a_{\nu\mu}$, without changing the conservation law. Consequently, the requirement (a) may be expressed as

$$\partial^\mu \overset{\circ}{F}{}^a{}_{\mu\nu} = g(\mathcal{J}^a_\nu + \partial^\mu W^a_{\mu\nu}). \tag{a$'$}$$

In order to obtain a self-consistent interaction, we write the new Lagrangian as a series in g, starting from $\mathcal{L}^{(0)}$:

$$\mathcal{L} = \mathcal{L}^{(0)} + g\Lambda^{(1)} + g^2\Lambda^{(2)} + \cdots .$$

Step 1. We start the construction by calculating the Noether current corresponding to the global $SU(2)$ symmetry of $\mathcal{L}^{(0)}$. Using the field equations, the change in $\mathcal{L}^{(0)}$ under the $SU(2)$ transformations takes the form

$$\delta\mathcal{L}^{(0)} = \partial_\mu \left(\delta\bar{\Psi}\frac{\partial\mathcal{L}^{(0)}}{\partial\partial_\mu\bar{\Psi}} + \frac{\partial\mathcal{L}^{(0)}}{\partial\partial_\mu\Psi}\delta\Psi + \delta A^a_\nu \frac{\partial\mathcal{L}^{(0)}}{\partial\partial_\mu A^a_\nu} \right) = -\theta^b \partial\cdot\mathcal{J}^b_{(0)}$$

where the Noether current

$$\mathcal{J}^b_{(0)} = J^b + j^b_{(0)} \tag{8.6}$$

contains two terms: J^b is the contribution of matter fields and $j^b_{(0)}$ the contribution of gauge fields:

$$j^{b\mu}_{(0)} = \epsilon^{abc} A^c{}_\nu \frac{\partial\mathcal{L}^{(0)}_{\mathrm{F}}}{\partial\partial_\mu A^a{}_\nu} = \epsilon^{bca} A^c{}_\nu \overset{\circ}{F}{}^{a\nu\mu}.$$

It is the term $j^b_{(0)}$ that is missing from the right-hand side of equation (8.5).

Now, in accordance with (a$'$), we try to change $\mathcal{L}^{(0)}$ so as to obtain the field equation

$$\partial_\mu \overset{\circ}{F}{}^{a\mu\nu} = g(\mathcal{J}^{a\nu}_{(0)} + \partial_\mu W^{a\mu\nu}). \tag{8.7a}$$

The presence of the current $j^a_{(0)}$ in this equation can be realized by adding to $\mathcal{L}^{(0)}$ a suitable term of order g:

$$\begin{aligned} \mathcal{L}^{(1)} &= \mathcal{L}^{(0)} + g\Lambda^{(1)} \\ \Lambda^{(1)} &= -c_1 A^b\cdot j^b_{(0)} = c_1 A^b{}_\nu(\epsilon^{bca}A^c{}_\lambda \overset{\circ}{F}{}^{a\nu\lambda}). \end{aligned} \tag{8.7b}$$

Varying $\mathcal{L}^{(1)}$ with respect to A^a and comparing the result with (8.7a) gives $c_1 = 1/2$, $W^a_{\lambda\nu} = \epsilon^{abc}A^b{}_\lambda A^c{}_\nu$. Although this change of $\mathcal{L}^{(0)}$ yields the correct field equations, the problem is still not solved: the current $\mathcal{J}^a_{(0)}$ is no longer conserved, since it is not the Noether current of the new Lagrangian $\mathcal{L}^{(1)}$!

Step 2. The new conserved current is calculated from $\mathcal{L}^{(1)}$:

$$\mathcal{J}_{(1)}^{b\mu} = \epsilon^{abc} A^{c\nu} \frac{\partial \mathcal{L}^{(1)}}{\partial \partial_\mu A^a{}_\nu} = \mathcal{J}_{(0)}^{b\mu} + g\epsilon^{abc} A^c{}_\nu (\epsilon^{efa} A^{f\nu} A^{e\mu}). \tag{8.8}$$

In order for the current $\mathcal{J}_{(1)}^b$ to be present on the right-hand side of equation (8.7a), we change the Lagrangian again by adding terms of order g^2:

$$\begin{aligned} \mathcal{L}^{(2)} &= \mathcal{L}^{(0)} + g\Lambda^{(1)} + g^2\Lambda^{(2)} \\ \Lambda^{(2)} &= -c_2[\epsilon^{abc} A^c{}_\nu (\epsilon^{efa} A^{f\nu} A^{e\mu})] A^b{}_\mu \end{aligned} \tag{8.9a}$$

whereupon the field equations for A^a take the form

$$\partial_\mu \overset{\circ}{F}{}^{a\mu\nu} = g(\mathcal{J}_{(1)}^{a\nu} + \partial_\mu W^{a\mu\nu}) \tag{8.9b}$$

if $c_2 = 1/4$, and $W^{a\mu\nu}$ remains the same.

Armed with the previous experience we are ready to face the problem that $\mathcal{J}_{(1)}^a$ is not conserved, since it is not the Noether current of the Lagrangian $\mathcal{L}^{(2)}$. However, the term $\Lambda^{(2)}$ in $\mathcal{L}^{(2)}$ does not contribute to the Noether current, because it does not depend on *field derivatives*; hence $\mathcal{J}_{(2)}^a = \mathcal{J}_{(1)}^a$.

Thus, the iteration procedure stops after the second step, leading to the fulfilment of the requirements (a) and (b). The self-consistent theory is defined by the Lagrangian $\mathcal{L}^{(2)}$, and yields the field equations (8.9b) for A^a. The construction of the theory is characterized by the occurrence of nonlinear effects, since the source of the fields A^a contains the fields A^a themselves.

Gauge symmetry. The initial Lagrangian $\mathcal{L}^{(0)}$ has the global $SU(2)$ symmetry, while $\mathcal{L}_{\mathrm{F}}^{(0)}$ possesses the additional local Abelian symmetry. What is the symmetry of the final, self-consistent Lagrangian $\mathcal{L}^{(2)}$? After introducing the quantity

$$F^a{}_{\mu\nu} = \overset{\circ}{F}{}^a{}_{\mu\nu} - g\epsilon^{abc} A^b{}_\mu A^c{}_\nu$$

we can easily check that

$$\mathcal{L}_{\mathrm{F}}^{(0)} + g\Lambda^{(1)} + g^2\Lambda^{(2)} = -\tfrac{1}{4} F^a{}_{\mu\nu} F^{a\mu\nu} \equiv \mathcal{L}_{\mathrm{F}}. \tag{8.10a}$$

Thus, unexpectedly, we discover that the final theory is invariant under the *local* $SU(2)$ symmetry. In the process of building the self-consistent theory,

> *the original global $SU(2)$ and incomplete local Abelian symmetry merge into the local $SU(2)$ symmetry.*

In the light of this result it is useful to reconsider the form of the conserved current and the field equations. Equations (8.8) and (8.6) imply

$$\mathcal{J}_{(1)}^{b\mu} = J^{b\mu} + \epsilon^{bca} A^c{}_\nu F^{a\nu\mu}$$

where the second term represents the dynamical current corresponding to A^a:

$$-\frac{\partial \mathcal{L}_\mathrm{F}}{\partial A^b{}_\mu} = \epsilon^{bca} A^c{}_\nu F^{a\nu\mu} = j^{b\mu}.$$

On the other hand, using the relation $\overset{\circ}{F}{}^a{}_{\mu\nu} - g W^a_{\mu\nu} = F^a{}_{\mu\nu}$ the field equations (8.9*b*) can be written in the form $\partial_\mu F^{a\mu\nu} = g\mathcal{J}^{a\nu}_{(1)}$ or

$$\nabla_\mu F^{a\mu\nu} = g J^{a\nu} \tag{8.10b}$$

where $\nabla_\mu F^{a\mu\nu} = \partial_\mu F^{a\mu\nu} - g\epsilon^{abc} A^b{}_\mu F^{c\mu\nu}$ is the covariant derivative. The conservation law of the current $\mathcal{J}^b_{(1)}$ takes the form

$$\nabla_\mu J^{a\mu} = 0 \tag{8.10c}$$

which is also, at the same time, the consistency condition of the field equations for A^a. The Lagrangian $\mathcal{L}_\mathrm{F} + \mathcal{L}_\mathrm{M} + \mathcal{L}_\mathrm{I}$ describes the $SU(2)$ gauge theory in the covariant form (appendix A). These considerations can be directly generalized to an arbitrary semisimple Lie group.

The assumption that matter is described by the Dirac field is of no essential importance for the previous analysis. It has the specific consequence that the interaction Lagrangian $\mathcal{L}_\mathrm{I}$, which is independent of the field derivatives, does not contribute to the Noether current $\mathcal{J}^b$, so that the form of the interaction remains *unchanged* in the process of building the self-consistent theory. It is clear that the same result holds for any matter field for which its interaction with A^a does not contain any field derivatives. In that case, the complete change of theory is restricted to the sector of the gauge fields, in which the symmetry makes a transition from local Abelian to local $SU(2)$ symmetry. The Lagrangian $\mathcal{L}_\mathrm{MI}$ remains unchanged in this process; its symmetry is, from the very beginning, local $SU(2)$, as can be clearly seen from the covariant form:

$$\mathcal{L}_\mathrm{MI} = \Psi(\mathrm{i}\gamma \cdot \nabla - m)\Psi \equiv \mathcal{L}_\mathrm{M}(\partial \to \nabla).$$

What happens when the interaction depends on the field derivatives is the subject of the subsequent analysis.

Scalar electrodynamics

Consider now a theory of the electromagnetic field in interaction with a complex scalar field,

$$\begin{aligned}\mathcal{L}^{(0)} &= \mathcal{L}_\mathrm{EM} + \mathcal{L}_\mathrm{M} + \mathcal{L}^{(0)}_\mathrm{I} \\ &= -\tfrac{1}{4} F_{\mu\nu} F^{\mu\nu} + \partial_\mu\phi^* \partial^\mu\phi - m^2\phi^*\phi - eA^\mu J^{(0)}_\mu\end{aligned} \tag{8.11}$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \qquad J^{(0)}_\mu = \mathrm{i}[\phi^*(\partial_\mu\phi) - (\partial_\mu\phi^*)\phi].$$

The total Lagrangian $\mathcal{L}^{(0)}$ possesses the global $U(1)$ symmetry, and $\mathcal{L}_{\mathrm{EM}}$ is invariant under $\delta A_\mu = \partial_\mu \lambda$. The field equations are

$$\begin{aligned}\partial_\mu F^{\mu\nu} &= eJ^\nu_{(0)} \\ (-\Box - m^2)\phi &= \mathrm{i}e[\partial_\mu(A^\mu\phi) + A^\mu\partial_\mu\phi].\end{aligned}$$

The consistency of the first equation requires the conservation of the current. However, since $J^\nu_{(0)}$ is not the complete Noether current, this condition cannot be satisfied. Indeed, $\partial_\mu J^\mu_{(0)} = 2e\partial_\mu(A^\mu\phi^*\phi) \neq 0$. The complete Noether current, defined by the global $U(1)$ symmetry of $\mathcal{L}^{(0)}$, has the form

$$\mathcal{J}^\mu_{(0)} = J^\mu_{(0)} - 2eA^\mu\phi^*\phi$$

where the second term comes from the interaction $\mathcal{L}_{\mathrm{I}}^{(0)}$.

By changing the Lagrangian,

$$\mathcal{L}^{(0)} \to \mathcal{L} = \mathcal{L}^{(0)} + e^2 A_\mu A^\mu \phi^*\phi \tag{8.12a}$$

the field equations become

$$\begin{aligned}\partial_\mu F^{\mu\nu} &= e(J^\nu_{(0)} - 2eA^\nu\phi^*\phi) \equiv e\mathcal{J}^\nu_{(0)} \\ (-\Box - m^2)\phi &= \mathrm{i}e[\partial_\mu(A^\mu\phi) + A^\mu\partial_\mu\phi] - e^2 A_\mu A^\mu\phi.\end{aligned}$$

Since the additional term $e^2A^2\phi^*\phi$ does not depend on the field derivatives, the new Noether current is the same as $\mathcal{J}^\nu_{(0)}$, and the construction of the self-consistent theory is completed.

The symmetry of the resulting theory is most easily recognized by writing the Lagrangian (8.12a) in the form

$$\mathcal{L} = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} + (\partial_\mu + \mathrm{i}eA_\mu)\phi^*(\partial_\mu - \mathrm{i}eA_\mu)\phi. \tag{8.12b}$$

The *global* $U(1)$ symmetry of the original theory and the *local* Abelian symmetry of the electromagnetic sector merge into the *local* $U(1)$ symmetry of the final theory.

This example illustrates the situation which occurs when the original interaction depends on field derivatives: then, $\mathcal{L}_{\mathrm{I}}^{(0)}$ gives a contribution to the Noether current, so that the self-consistent iteration procedure *changes* the form of the interaction. Since the original global symmetry is Abelian, the Noether current does not contain the contribution of the gauge fields.

In the general case the Noether current has contributions from both the interaction term and the free gauge sector. The situation is significantly simpler if the interaction does not depend on gauge field derivatives.

8.2 Scalar theory of gravity

In the previous analysis of Yang–Mills theories the iteration procedure ended after two steps. This will not be the case with gravity, where the iteration procedure is infinite. In order to show how such problems may be treated, we consider here a simpler, scalar theory of gravity, without assuming that the graviton is massless (Okubo 1978).

We start with the Lagrangian

$$\mathcal{L}^{(0)} = \mathcal{L}_{\mathrm{S}}^{(0)} + \mathcal{L}_{\mathrm{M}} + \mathcal{L}_{\mathrm{I}} = \tfrac{1}{2}(\partial_\mu\varphi\partial^\mu\varphi - \mu^2\varphi^2) + \mathcal{L}_{\mathrm{M}} - \lambda\varphi T \qquad (8.13a)$$

where T stands for the trace of the matter energy–momentum tensor. The field equation for φ is

$$-(\Box + \mu^2)\varphi = \lambda T. \qquad (8.13b)$$

Since there are no requirements on the properties of the interaction, even when the scalar graviton is massless (there is no gauge symmetry), this theory is consistent as it stands, without any further modification. Bearing in mind the physical requirement that the scalar graviton should interact with its own energy–momentum tensor, we shall try to build a scalar theory of gravity which is based on the following two requirements:

- The new Lagrangian $\mathcal{L} = \mathcal{L}^{(0)} + \Lambda$ should produce the field equation

$$-(\Box + \mu^2)\varphi = \lambda(T + t) \equiv \lambda\theta. \qquad (\alpha)$$

- The quantity θ is the trace of the total energy–momentum tensor, which is conserved owing to the global translational invariance, $\partial_\mu\theta^\mu{}_\nu = 0$.

The canonical energy–momentum ${}^c\theta^\mu{}_\nu = (\partial\mathcal{L}/\partial\partial_\mu\varphi)\partial_\nu\varphi - \delta^\mu_\nu\mathcal{L}$ is conserved but not uniquely defined. It can be changed, without affecting the conserved charge, by adding a term with vanishing divergence:

$$\begin{aligned} \theta^\mu{}_\nu &= {}^c\theta^\mu\nu + \partial_\rho W^{\mu\rho}{}_\nu \\ W^{\mu\rho}{}_\nu = -W^{\rho\mu}{}_\nu &= (\delta^\mu_\nu\partial^\rho - \delta^\rho_\nu\partial^\mu)W(\varphi) \end{aligned} \qquad (8.14)$$

where $W(\varphi)$ is an arbitrary function of φ. The trace of (8.14),

$$\theta = {}^c\theta + 3\Box W(\varphi) \qquad (\beta)$$

is the expression that should stand on the right-hand side of equation (α).

We are now going to find a new theory using the iteration procedure in powers of λ, starting from the Lagrangian $\mathcal{L}^{(0)}$.

Step 1. The trace of the canonical energy–momentum corresponding to $\mathcal{L}^{(0)}$ has the form

$$^{c}\theta_{(0)} = T + (-\partial_\mu\varphi\partial^\mu\varphi + 2\mu^2\varphi^2). \tag{8.15a}$$

The second term is the contribution of free scalar fields. We now want to find a Lagrangian that will give the field equation

$$-(\Box + \mu^2)\varphi = \lambda[T + (-\partial_\mu\varphi\partial^\mu\varphi + 2\mu^2\varphi^2) + 3\Box W(\varphi)]. \tag{8.15b}$$

The new Lagrangian has the form

$$\mathcal{L}^{(1)} = \mathcal{L}^{(0)} + \lambda\Lambda^{(1)}$$

where $\Lambda^{(1)}$ obeys the condition $\delta\Lambda^{(1)}/\delta\varphi = \partial_\mu\varphi\partial^\mu\varphi - 2\mu^2\varphi^2 - 3\Box W^{(1)}(\varphi)$. A solution for $\Lambda^{(1)}$ reads:

$$\Lambda^{(1)} = \varphi\partial_\mu\varphi\partial^\mu\varphi - \tfrac{2}{3}\mu^2\varphi^3$$

where $W^{(1)}(\varphi) = \frac{1}{2}\varphi^2$. After that $\mathcal{L}^{(1)}$ takes the form

$$\begin{aligned} \mathcal{L}^{(1)} &= \mathcal{L}_{\mathrm{S}}^{(1)} + \mathcal{L}_{\mathrm{M}} - \lambda\varphi T \\ \mathcal{L}_{\mathrm{S}}^{(1)} &= \tfrac{1}{2}(1 + 2\lambda\varphi)\partial_\mu\varphi\partial^\mu\varphi - \tfrac{1}{2}\mu^2(\varphi^2 + \tfrac{4}{3}\lambda\varphi^3). \end{aligned} \tag{8.16}$$

This change of $\mathcal{L}^{(0)}$ yields the field equation (8.15b), but does not solve the problem: the tensor $\theta_{(0)}^{\mu\nu}$, the trace of which occurs in (8.15b), is not conserved, since it corresponds to the old Lagrangian $\mathcal{L}^{(0)}$, and not to $\mathcal{L}^{(1)}$.

The exact solution. Continuing further we can find $\theta_{(1)}^{\mu\nu}$, obtain the new Lagrangian $\mathcal{L}^{(2)}$ which yields the field equation $-(\Box + \mu^2)\varphi = \lambda\theta_{(1)}$, etc. The iteration procedure goes on and on, introducing higher and higher powers of λ. It is clear that this would change only $\mathcal{L}_{\mathrm{S}}$—the gravitational part of the Lagrangian. Since a direct computation of higher order corrections becomes very tedious, we would like to have a more efficient method.

There is a simple approach to finding the complete solution, which is based on the following ansatz:

$$\mathcal{L}_{\mathrm{S}} = \tfrac{1}{2}F(\varphi)\partial_\mu\varphi\partial^\mu\varphi - G(\varphi) \tag{8.17}$$

as suggested by the lowest order results. The functions $F(\varphi)$ and $G(\varphi)$ will be determined directly on the bases of the requirements (α) and (β).

Varying $\mathcal{L}$ with respect to φ we obtain

$$-F\Box\varphi - \tfrac{1}{2}F'\partial_\mu\varphi\partial^\mu\varphi - G' = \lambda T \tag{8.18}$$

where the prime stands for $\mathrm{d}/\mathrm{d}\varphi$. On the other hand, using the expression for the trace of the energy–momentum tensor,

$$\theta = T + (-F + 3W'')\partial^\mu\varphi\partial_\mu\varphi + 3\Box\varphi W' + 4G$$

the general requirement (α) takes the form

$$-(1+3\lambda W')\Box\varphi - \lambda(-F+3W'')\partial^\mu\varphi\partial_\mu\varphi - (\mu^2\varphi^2 + 4\lambda G) = \lambda T. \tag{8.19}$$

The field equation (8.18) will have this form provided

$$\begin{aligned} F(\varphi) &= 1 + 3\lambda W'(\varphi) \\ F'(\varphi) &= 2\lambda[-F(\varphi) + 3W''(\varphi)] \\ G'(\varphi) &= \mu^2\varphi + 4\lambda G(\varphi). \end{aligned} \tag{8.20}$$

The last equation, written in the form $e^{4\lambda\varphi}[e^{-4\lambda\varphi}G(\varphi)]' = \mu^2\varphi$, implies

$$G(\varphi) = -\frac{\mu^2}{(4\lambda)^2}(1+4\lambda\varphi) + a_1 e^{4\lambda\varphi}. \tag{8.21a}$$

Since $G(\varphi) \sim -\frac{1}{2}\mu^2\varphi^2$ as $\varphi \to 0$, we find $a_1 = \mu^2/(4\lambda)^2$.

Eliminating $W(\varphi)$ from the first two equations we obtain

$$F(\varphi) = a_2 e^{2\lambda\varphi} \tag{8.21b}$$

where $a_2 = 1$, from $F(0) = 1$. Then, the first equation gives

$$W(\varphi) = \frac{1}{3\lambda}\left[\frac{1}{2\lambda}(e^{2\lambda\varphi} - a_3) - \varphi\right]. \tag{8.21c}$$

From $W(\varphi) \sim \varphi^2/3$, we find $a_3 = 1$.

Thus, we obtained the self-consistent Lagrangian,

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_S + \mathcal{L}_M - \lambda\varphi T \\ \mathcal{L}_S &\equiv \frac{1}{2}e^{2\lambda\varphi}\partial_\mu\varphi\partial^\mu\varphi - \frac{\mu^2}{(4\lambda)^2}[e^{4\lambda\varphi} - (1+4\lambda\varphi)] \end{aligned} \tag{8.22}$$

which leads to the field equation

$$-e^{2\lambda\varphi}\Box\varphi - \lambda e^{2\lambda\varphi}\partial_\mu\varphi\partial^\mu\varphi - \frac{\mu^2}{4\lambda}(e^{4\lambda\varphi} - 1) = \lambda T.$$

If we introduce $\Phi(x) = e^{\lambda\varphi(x)}$, the Lagrangian $\mathcal{L}_S$ can be rewritten as

$$\mathcal{L}_S = \frac{1}{2\lambda^2}\partial_\mu\Phi\partial^\mu\Phi - \frac{\mu^2}{(4\lambda)^2}(\Phi^4 - 1 - 4\ln\Phi) \tag{8.23}$$

while the field equation becomes

$$-\Box\Phi - \mu^2\frac{\Phi^4 - 1}{4\Phi} = \lambda T.$$

This procedure solves the self-consistency problem of the scalar theory of gravity, in accordance with the requirements (α) and (β).

8.3 Tensor theory of gravity

We are now going to construct a self-consistent theory of massless spin-2 graviton field (Feynman *et al* 1995, Van Dam 1974). The naive formulation (7.41) is not consistent since the matter energy–momentum tensor is not conserved. To make the whole procedure as clear as possible, we start, as usual, from the requirements that the theory should satisfy the following conditions.

- The field equations for the graviton $\varphi_{\mu\nu}$, following from a complete Lagrangian $\mathcal{L} = \mathcal{L}^{(0)} + \Lambda$, should have the form

$$K_{\mu\nu,\sigma\rho}\varphi^{\sigma\rho} = \lambda\theta_{\mu\nu}. \tag{A}$$

- The dynamical current is equal to the symmetric energy–momentum $\theta_{\mu\nu}$, which is conserved:

$$\partial^\mu\theta_{\mu\nu} = 0. \tag{B}$$

The iterative procedure

The starting point of the iterative procedure is the Lagrangian

$$\mathcal{L}^{(0)} = \mathcal{L}_\mathrm{T} + \mathcal{L}_\mathrm{M} - \lambda\varphi^{\mu\nu}T^{(0)}_{\mu\nu} \tag{8.24a}$$

and the corresponding field equations for $\varphi_{\mu\nu}$ are:

$$K_{\mu\nu,\sigma\rho}\varphi^{\sigma\rho} = \lambda T^{(0)}_{\mu\nu}. \tag{8.24b}$$

The term which is missing on the right-hand side is the symmetric energy–momentum tensor of the graviton field $t^{(0)}_{\mu\nu}$. This deficiency may be corrected by modifying the starting Lagrangian as follows:

$$\mathcal{L}^{(0)} \to \mathcal{L}^{(1)} = \mathcal{L}^{(0)} + \lambda\Lambda^{(1)} \tag{8.25a}$$

where the additional term $\Lambda^{(1)}$ obeys the condition

$$-\frac{\delta\Lambda^{(1)}}{\delta\varphi^{\mu\nu}} = t^{(0)}_{\mu\nu}. \tag{8.25b}$$

We search $\Lambda^{(1)}$ in the form of a general sum over all possible independent trilinear products of fields and their derivatives (18 terms). Then, demanding (8.25b), the precise form of this term is determined after a long and hard calculation:

$$\begin{aligned} -\Lambda^{(1)} &= \varphi^{\mu\nu}\bar\varphi^{\sigma\tau}\bar\varphi_{\mu\nu,\sigma\tau} + \varphi_\sigma{}^\nu\varphi^{\sigma\mu}\bar\varphi_{\mu\nu,\tau}{}^\tau - 2\varphi^{\mu\nu}\varphi_{\nu\tau}\bar\varphi_{\mu\sigma}{}^{,\sigma\tau} \\ &\quad + 2\bar\varphi_{\mu\nu}\bar\varphi^{\sigma\mu}{}_{,\sigma}\bar\varphi^{\tau\nu}{}_{,\tau} + (\tfrac{1}{2}\varphi_{\mu\nu}\varphi^{\mu\nu} + \tfrac{1}{4}\varphi^2)\bar\varphi^{\sigma\tau}{}_{,\sigma\tau}. \end{aligned} \tag{8.26}$$

The modification $\mathcal{L}^{(0)} \to \mathcal{L}^{(1)}$ leads to the field equations

$$K_{\mu\nu,\sigma\rho}\varphi^{\sigma\rho} = \lambda(T^{(0)}_{\mu\nu} + t^{(0)}_{\mu\nu}) \equiv \lambda\theta^{(0)}_{\mu\nu}.$$

Here, $\theta^{(0)}_{\mu\nu}$ is not conserved in the new theory, as it corresponds to the old Lagrangian $\mathcal{L}^{(0)}$. Since $\Lambda^{(1)}$ contains field derivatives, $\theta^{(1)}_{\mu\nu}$ differs from $\theta^{(0)}_{\mu\nu}$ by terms of the order λ, so that the inconsistency in the field equations is of the order λ^2.

The term $\Lambda^{(1)}$ can be found by another, equivalent method which is, however, more convenient for generalization. The tensor $T^{\mu\nu}_{(0)}$ obeys the relation (7.63),

$$(\eta_{\nu\sigma} + 2\lambda\varphi_{\nu\sigma})\partial_\mu T^{\mu\nu}_{(0)} = -[\lambda\rho, \sigma]T^{\lambda\rho}_{(0)}$$

which is found upon using the matter field equations. Noting that $\partial_\mu T^{\mu\nu}_{(0)} = \mathcal{O}(\lambda)$, this equation can be written in the form

$$\eta_{\nu\sigma}\partial_\mu T^{\mu\nu}_{(0)} = -[\lambda\rho, \sigma]T^{\lambda\rho}_{(0)} + \mathcal{O}(\lambda^2).$$

On the other hand, since the inconsistency of the field equations is of order λ^2, we have $\partial_\mu\theta^{\mu\nu}_{(0)} = \mathcal{O}(\lambda^2)$ and, consequently,

$$\eta_{\nu\sigma}\partial_\mu t^{\mu\nu}_{(0)} = [\lambda\rho, \sigma]T^{\lambda\rho}_{(0)} + \mathcal{O}(\lambda^2).$$

This relation can be used to calculate $\Lambda^{(1)}$. Indeed, starting from the general cubic expression for $\Lambda^{(1)}$, the last relation, written in the form

$$-\eta_{\nu\sigma}\partial_\mu \frac{\delta\Lambda^{(1)}}{\delta\varphi^{\mu\nu}} = [\lambda\rho, \sigma]T^{\lambda\rho}_{(0)} + \mathcal{O}(\lambda^2) \tag{8.27}$$

determines all the constants in $\Lambda^{(1)}$, leading to the result (8.26).

The realization of the first step in the iterative procedure is followed by a lot of tiresome calculations. It is obvious that the second step would be extremely complicated, so that it is natural to ask whether there is some more efficient approach to this problem. The following considerations are devoted to just this question.

Formulation of a complete theory

It may be exceedingly difficult to calculate higher order corrections in a particular expansion of some quantity but yet it might be possible to construct a complete solution which effectively sums all higher order terms. This is the case with the scalar theory of gravity, where the calculation of functions F and G corresponds to a summation of all higher order corrections to the starting Lagrangian $\mathcal{L}^{(0)}$. Following this idea we now make an attempt to deduce the complete form of the theory based on the requirements (A) and (B).

Consistency requirements. We search for a complete gravitational action

$$I_{\mathrm{G}} = \int (\mathcal{L}_{\mathrm{T}} + \Lambda)\,\mathrm{d}^4x$$

which is determined by the requirement that the resulting field equation (A) may be written in the form

$$\frac{\delta I_{\rm G}}{\delta\varphi_{\mu\nu}} = \lambda\widetilde{T}^{\mu\nu} \tag{8.28}$$

where $\widetilde{T}^{\mu\nu}$ is the part of $\theta^{\mu\nu}$ that comes from the interacting matter fields. We allow that in a consistent theory $\widetilde{T}^{\mu\nu}$ may be different from $T^{\mu\nu}_{(0)}$. The consistency of the gravitational field equation (8.28) and the equation of motion of the matter may be checked by taking the divergence of (8.28),

$$\partial_\mu\frac{\delta I_{\rm G}}{\delta\varphi_{\mu\nu}} = \lambda\partial_\mu\widetilde{T}^{\mu\nu}$$

and evaluating the right-hand side using the matter field equation. In order to be able to realize this procedure we need to have some information about the nature of the gravitational interaction. We *assume* that $\widetilde{T}^{\mu\nu}$ satisfies the same divergence property as $T^{\mu\nu}_{(0)}$:

$$g_{\nu\sigma}\partial_\mu\widetilde{T}^{\mu\nu} = -[\mu\nu,\sigma]\widetilde{T}^{\mu\nu}. \tag{8.29}$$

This property of $T^{\mu\nu}_{(0)}$ is derived from its form (7.31c) and the corresponding matter equation. For equations (8.28) and (8.29) to be consistent, the action $I_{\rm G}$ should satisfy the following functional differential equation:

$$g_{\nu\sigma}\partial_\mu\frac{\delta I_{\rm G}}{\delta\varphi_{\mu\nu}} + [\mu\nu,\sigma]\frac{\delta I_{\rm G}}{\delta\varphi_{\mu\nu}} = 0. \tag{8.30}$$

Solving this equation is an exceedingly difficult problem, since there is no general procedure for generating the solutions. There is *no unique* solution of this problem, even if we demand that for small $\varphi_{\mu\nu}$ the solution should behave as described by equation (8.24a). There is, however, a 'simplest' solution, which yields a theory identical to Einstein's GR.

In order to find a deeper meaning to equation (8.30), we shall convert it to an equivalent but more transparent form. We multiply (8.30) with an arbitrary vector $2\xi^\sigma(x)$ and integrate over all space, and then integrate the first term by parts, which leads to

$$\int \mathrm{d}^4x\,\frac{\delta I_{\rm G}}{\delta\varphi_{\mu\nu}}\{-(\xi^\sigma g_{\sigma\nu})_{,\mu} + [\mu\nu,\sigma]\xi^\sigma\}2 = 0. \tag{8.31}$$

This form of the consistency requirement can be interpreted in the following way. Let us introduce an infinitesimal transformation of $\varphi_{\mu\nu}$,

$$\begin{aligned}\delta_0\varphi_{\mu\nu} &= 2\{(\xi^\sigma g_{\sigma\nu})_{,\mu} - [\mu\nu,\sigma]\xi^\sigma\}\\ &= \xi_{\mu,\nu} + \xi_{\nu,\mu} + 2\lambda(\varphi_{\mu\sigma}\xi^\sigma{}_{,\nu} + \varphi_{\nu\sigma}\xi^\sigma{}_{,\mu} + \varphi_{\mu\nu,\sigma}\xi^\sigma)\end{aligned}$$

which can be written as

$$\delta_0 g_{\mu\nu} = 2\lambda(g_{\mu\sigma}\xi^\sigma{}_{,\nu} + g_{\nu\sigma}\xi^\sigma{}_{,\mu} + g_{\mu\nu,\sigma}\xi^\sigma). \tag{8.32}$$

Equation (8.31) is an invariance condition, telling us that the action I_G *is invariant under infinitesimal local transformations (8.32).*

The transformation (8.32) can be recognized as the transformation of a tensor field $g_{\mu\nu}$ under infinitesimal *coordinate transformations*:

$$x'^{\mu} = x^{\mu} - 2\lambda\xi^{\mu}(x). \tag{8.33}$$

However, the invariance condition (8.31) involves *only* the transformation (8.32), *without* (8.33). For this reason we leave the possibility of a geometric interpretation, which is based on *both* (8.32) and (8.33), for later discussion.

Pure gravity. Having understood that our central consistency requirement (8.30) has a form of an invariance condition, we now proceed to examine the form of various invariants under the transformation (8.32).

From the transformation properties of the determinant $g = \det(g_{\mu\nu})$ under (8.32), $\delta_0 g = 2\lambda g(2\xi^{\sigma}{}_{,\sigma} + \xi^{\sigma} g^{\mu\nu} g_{\mu\nu,\sigma})$, it follows that

$$\sqrt{-g'} = \sqrt{-g} + 2\lambda(\sqrt{-g}\xi^{\sigma})_{,\sigma}.$$

When we integrate this equation over all space, the integral of the total divergence is irrelevant. Thus, we find one invariant solution for the action:

$$I_G^{(0)} = \int \mathrm{d}^4x\, \sqrt{-g}. \tag{8.34}$$

This solution is not satisfying dynamically, since it involves no field derivatives. But using an analogous method we shall be able to construct more satisfying solutions.

Let us note that the scalar field is defined by the condition $\sigma'(x') = \sigma(x)$, which in the case of infinitesimal transformations reads as $\delta_0\sigma(x) = 2\lambda\xi^{\tau}\partial_{\tau}\sigma(x)$. Combining this with the previous results we obtain

$$\sqrt{-g'}\sigma' = \sqrt{-g}\sigma + 2\lambda(\sqrt{-g}\sigma\xi^{\tau})_{,\tau}.$$

Thus the product of $\sqrt{-g}$ with a scalar changes only by a total divergence. Hence,

$$I_G = \int \mathrm{d}^4x\, \sqrt{-g}\sigma(x) \tag{8.35}$$

is also a solution for the gravitational action.

Is there a combination of the fields $g_{\mu\nu}$ and their derivatives which is a scalar? Bearing in mind the hidden geometric meaning of the transformations (8.32), it is not difficult to find the answer to this question using well-known results from differential geometry. In the framework of Riemannian geometry, the only tensor which can be constructed from the metric tensor and its derivatives is the Riemann curvature tensor. The construction of the action is now straightforward, and the 'simplest' solution coincides with *Einstein's GR*:

$$I_{\mathrm{G}} = -\frac{1}{2\lambda^2}\int \mathrm{d}^4x\,\sqrt{-g}R \tag{8.36}$$

where R is the scalar curvature of Riemann space with metric $g_{\mu\nu}$. The existence of other solutions (such as the cosmological term (8.34), or invariants quadratic in the curvature tensor) shows that there is no unique solution of the problem. The solution (8.36) is normalized so that in the weak field approximation it reduces to the Pauli–Fierz action I_{T}. This is the gravitational part of the action which is correct to all orders.

Matter and interaction. We now focus our attention on the part I_{MI} that describes matter and its interaction with gravity. The matter energy–momentum $\widetilde{T}^{\mu\nu}$ in equation (8.28) is defined by

$$-\lambda\widetilde{T}^{\mu\nu} \equiv \frac{\delta I_{\mathrm{MI}}}{\delta\varphi_{\mu\nu}}. \tag{8.37}$$

It is clear from the previous discussion that the assumption (8.29), concerning the behaviour of $\partial_\mu\widetilde{T}^{\mu\nu}$, leads to the conclusion that I_{MI} is *invariant* under (8.32):

$$\int \mathrm{d}^4x\,\frac{\delta I_{\mathrm{MI}}}{\delta g_{\mu\nu}}\delta_0 g_{\mu\nu} = 0. \tag{8.38}$$

The verification of condition (8.29), and, consequently, (8.38), demands the explicit use of the matter field equations.

When matter is described by point particle variables, the action is defined, up to terms of order λ, by the expression

$$I^{(1)} = \int \mathrm{d}^4x\,(\mathcal{L}_{\mathrm{T}} + \mathcal{L}_{\mathrm{M}} - \lambda\varphi_{\mu\nu}T^{\mu\nu}_{(0)}) \equiv I_{\mathrm{T}} + I^{(1)}_{\mathrm{MI}}.$$

The part I_{T} is invariant under local transformations $\delta_0\varphi_{\mu\nu} = \partial_\mu\xi_\nu + \partial_\nu\xi_\mu$, and the complete action is invariant under global translations $\delta x^\mu = a^\mu$, which implies the conservation of the total energy–momentum. This theory is inconsistent, since condition (8.38) for $I^{(1)}_{\mathrm{MI}}$,

$$\frac{\delta I^{(1)}_{\mathrm{MI}}}{\delta g_{\mu\nu}}\delta_0 g_{\mu\nu} = -\lambda T^{\mu\nu}_{(0)}(\partial_\mu\xi_\nu + \partial_\nu\xi_\mu) = 0 \tag{8.39}$$

requires the conservation of $T^{\mu\nu}_{(0)}$. However, if we limit our our considerations to the accuracy of order λ, this condition is correct. Indeed, using the geodesic equation we find that $\partial_\mu T^{\mu\nu}_{(0)} = \mathcal{O}(\lambda)$, which brings in an error of the order λ^2 in (8.39).

The inefficiency of the method based on explicit equations of motion motivates us to to ask ourselves whether condition (8.39) can be checked in some other, more practical way. The answer is: yes. The automatic validity of this condition may be ensured by *extending* the concept of local transformation (8.32) to the matter sector.

Assume that the action $I_{\mathrm{MI}}^{(1)}$ is invariant under extended local transformations, which also act on the point-particle variables:

$$\begin{aligned}\delta_0\varphi_{\mu\nu} &= \partial_\mu\xi_\nu + \partial_\nu\xi_\mu \\ \delta x^\mu &= a^\mu[\xi].\end{aligned}$$

The invariance of $I_{\mathrm{MI}}^{(1)}$ leads to

$$\frac{\delta I_{\mathrm{MI}}^{(1)}}{\delta x^\mu}a^\mu[\xi] + \frac{\delta I_{\mathrm{MI}}^{(1)}}{\delta g_{\mu\nu}}2\lambda(\partial_\mu\xi_\nu + \partial_\nu\xi_\mu) = 0. \tag{8.40}$$

From the form (7.45) for $I_{\mathrm{MI}}^{(1)}$ it follows that the invariance condition is satisfied provided $a^\mu[\xi] = -2\lambda\xi^\mu(x)$. On the other hand, after using the equations of motion for matter, the invariance requirement (8.40) automatically carries over into (8.39).

The invariance requirement (8.40) is a very efficient method for building a consistent theory. The generalized form of the local transformations is given by

$$\delta_0 g_{\mu\nu} = 2\lambda(g_{\mu\sigma}\xi^\sigma{}_{,\nu} + g_{\nu\sigma}\xi^\sigma{}_{,\mu} + g_{\mu\nu,\sigma}\xi^\sigma) \tag{8.41a}$$

$$\delta x^\mu = -2\lambda\xi^\mu \tag{8.41b}$$

while the invariance of the action is ensured by the choice

$$I_{\mathrm{MI}} = -m\int \mathrm{d}\tau\,\sqrt{g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu}. \tag{8.42}$$

Local transformations of the point-particle variables (8.41*b*) can be realized in a simple way by assuming that all points of *spacetime* are transformed in the same way. In that case:

> *Generalized local transformations are nothing else but the complete infinitesimal coordinate transformations.*

The introduction of the concept of coordinate transformations becomes necessary only when we consider the interaction between gravity and matter.

It is important at this stage to clarify the following two questions:

(i) Does the generalized symmetry influence the invariance of the gravitational action I_{G}?
(ii) Does it have the form (8.41) for all types of matter?

The answer to the first question follows from the form of the gravitational action (8.35). Since $\sigma'(x') = \sigma(x)$, and $\mathrm{d}^4x\,\sqrt{-g}$ is an invariant measure, it is evident that the action I_G is also invariant under the generalized transformations (8.41).

If matter is described not by particle variables, but by a field ϕ, then its energy–momentum tensor contains $\partial\phi$. Without going into a detailed analysis, it is clear that the construction of I_MI requires an extension of the concept of local transformations not only to matter fields ϕ, but also to the derivatives $\partial\phi$. The result is, again, the infinitesimal coordinate transformation (8.41) followed, this time, by the transformation rule for ϕ.

In the process of building a self-consistent theory the symmetry structure has changed, just as in the Yang–Mills case.

> *The original global translations and local Abelian symmetry of the gravitational sector merge into the local translations (8.41).*

It is useful to clarify here one simple property of $\widetilde{T}^{\mu\nu}$. Equation (8.29) tells us that $\widetilde{T}^{\mu\nu}$ is not a tensor, but a tensor density. The true energy–momentum tensor is introduced by $T^{\mu\nu} = \widetilde{T}^{\mu\nu}/\sqrt{-g}$. Relation (8.29) is the covariant conservation law of $T^{\mu\nu}$.

The action $I = I_\mathrm{G} + I_\mathrm{MI}$ solves the problem of constructing a self-consistent theory of gravity. It yields Einstein's field equations, which are in agreement with all experimental observations.

We should note finally that the essential step in this construction was the assumption of divergence condition (8.29). In the next subsection we shall attempt to obtain the final result in a simpler way, without additional assumptions of this type.

8.4 The first order formalism

There is no general method for the transition from a linear, non-consistent theory to a consistent but nonlinear formulation. The iterative formalism in the theory of gravity requires an infinite number of steps and the use of an ansatz is not always possible. Feynman's approach is very elegant, but it is based on assumption (8.29), the justification of which is seen only after we have solved the problem.

Now we shall see how a transition to the first order formalism significantly simplifies the iterative procedure, so that the whole path to a consistent theory can be shortened to just *one* step (Deser 1970).

Yang–Mills theory

We begin by applying the new formalism to Yang–Mills theory.

Free Yang–Mills theory. The linearized theory of the free $SU(2)$ Yang–Mills field in the first order formalism is given by

$$\mathcal{L}_{\mathrm{F}}^{(0)} = -\tfrac{1}{2}F^{a\mu\nu}(\partial_\mu A^a{}_\nu - \partial_\nu A^a{}_\mu) + \tfrac{1}{4}F^{a\mu\nu}F^a{}_{\mu\nu} \tag{8.43}$$

where A^a and F^a are *independent* dynamical variables. The field equations, obtained by varying this with respect to A^a and F^a,

$$\partial_\mu F^{a\mu\nu} = 0$$
$$F^a_{\mu\nu} = \partial_\mu A^a{}_\nu - \partial_\nu A^a{}_\mu \equiv \overset{\circ}{F}{}^a_{\mu\nu}$$

are equivalent to the field equations $\partial_\mu \overset{\circ}{F}{}^{a\mu\nu} = 0$ of the second order formalism. The theory is invariant under global $SU(2)$ and local Abelian transformations.

- As before, we want to change the starting Lagrangian $\mathcal{L}_{\mathrm{F}}^{(0)}$ so as to obtain the following field equations:

$$\partial^\mu \overset{\circ}{F}{}^a{}_{\mu\nu} = g(j^a_\nu + \partial^\mu W^a_{\mu\nu}) \tag{YM}$$

  where j^a is the Noether current associated with the new Lagrangian.

In the lowest order approximation, the Noether current is obtained from the Lagrangian $\mathcal{L}_{\mathrm{F}}^{(0)}$, and has the form

$$j^{a\nu}_{(0)} = \epsilon^{abc}A^b{}_\mu F^{c\mu\nu}$$

which, after we have eliminated $F^{c\mu\nu}$, coincides with our earlier result. Following the usual iterative procedure we shall search for a new Lagrangian in the form

$$\mathcal{L}_{\mathrm{F}}^{(1)} = \mathcal{L}_{\mathrm{F}}^{(0)} - \tfrac{1}{2}gA^a \cdot j_a^{(0)}$$

where the factor $\frac{1}{2}$ is chosen since $A \cdot j$ is quadratic in A. Thus, we obtain

$$\mathcal{L}_{\mathrm{F}}^{(1)} = -\tfrac{1}{2}F^{a\mu\nu}(\partial_\mu A^a{}_\nu - \partial_\nu A^a{}_\mu - g\epsilon^{abc}A^b{}_\mu A^c{}_\nu) + \tfrac{1}{4}F^{a\mu\nu}F^a{}_{\mu\nu} \equiv \mathcal{L}_{\mathrm{F}}. \tag{8.44}$$

We are slightly surprised that this expression essentially coincides with the Yang–Mills Lagrangian (8.10a). Since $A \cdot j$ does not contain the field derivatives, the Noether current of this Lagrangian is the same as $j^a_{(0)}$, hence the construction of a consistent theory is completed in just *one* step.

The new field equations can indeed be written in the form (YM):

$$\partial_\mu \overset{\circ}{F}{}^{a\mu\nu} = g[j^{a\nu} + \partial_\mu(\epsilon^{abc}A^{b\mu}A^{c\nu})].$$

Interaction with matter. Now we want to find the interaction Lagrangian corresponding to a general matter field ϕ. The total Lagrangian of a Yang–Mills + matter system yields the field equations

$$\nabla_\mu F_a{}^{\mu\nu} = gJ_a^\nu \equiv -\frac{\delta\mathcal{L}_{\mathrm{MI}}}{\delta A^a{}_\nu}$$

or, equivalently,

$$\partial_\mu \overset{\circ}{F}{}^{a\mu\nu} = g[J^{a\nu} + j^{a\nu} + \partial_\mu(\epsilon^{abc}A^{b\mu}A^{c\nu})]. \qquad \text{(YMm)}$$

We will assume that

- $\mathcal{L}_{\mathrm{MI}}$ does not depend on ∂A^a, hence it does not influence the construction of the current j^a; and
- the dynamical current J_a is equal to the Noether current of $\mathcal{L}_{\mathrm{MI}}$, generated by the global $SU(2)$ symmetry.

Hence,

$$-\frac{\partial\mathcal{L}_{\mathrm{MI}}}{\partial\partial_\mu\phi}T_a\phi = -\frac{1}{g}\frac{\delta\mathcal{L}_{\mathrm{MI}}}{\delta A^a{}_\mu}$$

where the T_a are the $SU(2)$ generators in the representation corresponding to the field ϕ. The solution of this equation reads:

$$\mathcal{L}_{\mathrm{MI}}(\phi, \partial_\mu\phi, A^a{}_\mu) = \mathcal{L}_{\mathrm{M}}(\phi, \nabla_\mu\phi) \qquad (8.45)$$

where $\nabla_\mu\phi \equiv (\partial_\mu + gA^a{}_\mu T_a)\phi$.

> *The consistency requirement in the sector of matter fields and interaction leads to the so-called minimal interaction.*

Einstein's theory

The first order formalism. The transition to the first order formalism in symmetric tensor theory can be realized with the help of the Palatini formalism in GR. In this formalism the action for GR has the form

$$I_P = -\frac{1}{2\lambda^2}\int \mathrm{d}^4x\,\sqrt{-g}g^{\mu\nu}R_{\mu\nu}(\Gamma) \qquad (8.46)$$

where the symmetric connection $\Gamma^\mu_{\nu\lambda}$ is defined as an *independent* dynamical variable. Varying this with respect to $g_{\mu\nu}$ and $\Gamma^\mu_{\nu\lambda}$ we find the field equations

$$R_{\mu\nu}(\Gamma) - \tfrac{1}{2}g_{\mu\nu}R(\Gamma) = 0$$
$$\Gamma^\mu_{\nu\rho} = \tfrac{1}{2}g^{\mu\sigma}(g_{\nu\sigma,\rho} + g_{\rho\sigma,\nu} - g_{\nu\rho,\sigma}) \equiv \left\{{\mu \atop \nu\rho}\right\}$$

which, after eliminating Γ, coincide with the standard Einstein's equations. The Palatini formalism is, in fact, the first order formalism in GR.

Replacing $g_{\mu\nu} = \eta_{\mu\nu} + 2\lambda\varphi_{\mu\nu}$ in I_P and keeping only the terms that are quadratic in the fields and their derivatives, we find the action

$$I_{\mathrm{T}} = -\frac{1}{2\lambda^2}\int \mathrm{d}^4x\,[-2\lambda\bar{\varphi}^{\mu\nu}R^{\mathrm{L}}_{\mu\nu}(\Gamma) + \eta^{\mu\nu}R^{\mathrm{Q}}_{\mu\nu}(\Gamma)] \tag{8.47}$$

where R^{L} and R^{Q} are the linear and quadratic parts of the curvature,

$$R^{\mathrm{L}}_{\mu\nu} = \partial_\rho\Gamma^\rho_{\mu\nu} - \partial_\nu\Gamma_\mu \qquad R^{\mathrm{Q}}_{\mu\nu} = \Gamma_\rho\Gamma^\rho_{\mu\nu} - \Gamma^\lambda_{\mu\sigma}\Gamma^\sigma_{\nu\lambda}$$

and $\Gamma_\rho \equiv \Gamma^\lambda_{\rho\lambda}$. This action describes the tensor theory (7.19) in the first order formalism. Indeed, varying this with respect to $\bar{\varphi}^{\mu\nu}$ and $\Gamma^\mu_{\nu\rho}$ we obtain the field equations:

$$\begin{aligned} &\frac{1}{2\lambda}R^{\mathrm{L}}_{(\mu\nu)} = 0 \\ &\Gamma^\mu_{\nu\rho} = \lambda\eta^{\mu\lambda}(\varphi_{\nu\lambda,\rho} + \varphi_{\rho\lambda,\nu} - \varphi_{\nu\rho,\lambda}). \end{aligned} \tag{8.48a}$$

Putting the second equation into the first and going over to $\bar{\varphi}^{\mu\nu}$, leads to

$$\frac{1}{2\lambda}R^{\mathrm{L}}_{(\mu\nu)}(\bar{\varphi}) = -\Box\bar{\varphi}_{\mu\nu} + 2\bar{\varphi}^\sigma{}_{(\mu,\nu)\sigma} + \tfrac{1}{2}\eta_{\mu\nu}\Box\bar{\varphi} = 0. \tag{8.48b}$$

If we rewrite equation (7.20) for $\varphi_{\mu\nu}$ in the form

$$-\Box\bar{\varphi}_{\mu\nu} + 2\bar{\varphi}^\sigma{}_{(\mu,\nu)\sigma} + \tfrac{1}{2}\eta_{\mu\nu}\Box\bar{\varphi} = J_{\mu\nu} - \tfrac{1}{2}\eta_{\mu\nu}J_\lambda{}^\lambda$$

it is clear that (8.48b) is the field equation of the tensor field in vacuum.

The consistency of pure gravity. A consistent theory of pure gravity can be obtained by modifying the action I_{T}.

- A new action should yield the field equations

$$\frac{1}{2\lambda}R^{\mathrm{L}}_{(\mu\nu)}(\bar{\varphi}) = \lambda(\bar{t}_{\mu\nu} + \overline{W}^\rho{}_{\mu\nu,\rho}) \tag{G}$$

where $\bar{t}_{\mu\nu} \equiv t_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}t$, and $t_{\mu\nu}$ is the conserved symmetric energy–momentum of the gravitational field.

In the first approximation, the symmetric energy–momentum tensor can be calculated from $\mathcal{L}_{\mathrm{T}}$ using the Rosenfeld prescription:

$$t_{\mu\nu} \equiv \frac{2}{\sqrt{-\gamma}}\frac{\delta\mathcal{L}_{\mathrm{T}}(\gamma)}{\delta\gamma^{\mu\nu}}\bigg|_{\gamma=\eta}.$$

Here, $\mathcal{L}_{\mathrm{T}}(\gamma)$ is the covariant Lagrangian obtained from $\mathcal{L}_{\mathrm{T}}$ by introducing an arbitrary curvilinear coordinate system with metric $\gamma_{\mu\nu}$:

$$\mathcal{L}_{\mathrm{T}}(\gamma) = -\frac{1}{2\lambda^2}\sqrt{-\gamma}[-2\lambda\bar{\varphi}^{\mu\nu}\overset{\gamma}{R}{}^{\mathrm{L}}_{\mu\nu}(\Gamma) + \gamma^{\mu\nu}\overset{\gamma}{R}{}^{\mathrm{Q}}_{\mu\nu}(\Gamma)] \tag{8.49a}$$

and $\overset{\gamma}{R}_{\mu\nu}(\Gamma)$ is obtained from $R_{\mu\nu}(\Gamma)$ by the replacement $\partial \to \nabla(G)$, where $G = G(\gamma)$ is the Christoffel connection for the metric $\gamma_{\mu\nu}$.

Further considerations can be greatly simplified if we assume that $\bar{\varphi}^{\mu\nu}$ is a *tensor density*, rather than a tensor. In that case the covariant Lagrangian is given as

$$\mathcal{L}'_{\mathrm{T}}(\gamma) = -\frac{1}{2\lambda^2}[-2\lambda\bar{\varphi}^{\mu\nu}\overset{\gamma}{R}{}^{\mathrm{L}}_{\mu\nu}(\Gamma) + \sqrt{-\gamma}\gamma^{\mu\nu}\overset{\gamma}{R}{}^{\mathrm{Q}}_{\mu\nu}(\Gamma)]. \tag{8.49b}$$

This change is irrelevant for the form of the energy–momentum tensor. Indeed, energy–momentum tensors obtained from the covariant Lagrangians $\mathcal{L}_{\mathrm{T}}$ and $\mathcal{L}'_{\mathrm{T}}$ are the *same*, up to the field equations for $\bar{\varphi}^{\mu\nu}$:

$$t'_{\mu\nu} = t_{\mu\nu} - \eta_{\mu\nu}\bar{\varphi}^{\sigma\rho}\frac{\delta\mathcal{L}_{\mathrm{T}}(\gamma)}{\delta\bar{\varphi}_{\sigma\rho}}\bigg|_{\gamma=\eta}.$$

Since the consistency requirement (G) is expressed as a statement about the field equations, the tensors $t_{\mu\nu}$ and $t'_{\mu\nu}$ are equivalent as regards the construction of a consistent theory. In what follows we shall work with $t'_{\mu\nu}$, which will be shown to coincide with the energy–momentum tensor of the final theory.

When we calculate the Rosenfeld energy–momentum tensor by varying with respect to the metric,

$$\frac{\delta\mathcal{L}}{\delta\gamma^{\mu\nu}}\bigg|_{\gamma=\eta} = \left[\frac{\partial\mathcal{L}}{\partial\gamma^{\mu\nu}} + \frac{\partial\mathcal{L}}{\partial G^{\lambda}_{\sigma\tau}}\frac{\partial G^{\lambda}_{\sigma\tau}}{\partial\gamma^{\mu\nu}} - \partial_\rho\left(\frac{\partial\mathcal{L}}{\partial G^{\lambda}_{\sigma\tau}}\frac{\partial G^{\lambda}_{\sigma\tau}}{\partial\partial_\rho\gamma^{\mu\nu}}\right)\right]_{\gamma=\eta}$$

we should note that the second term does not give a contribution for $\gamma = \eta$. After rearranging this expression and separating the trace, we obtain

$$\bar{t}'_{\mu\nu} \equiv t'_{\mu\nu} - \tfrac{1}{2}\eta_{\mu\nu}t' = -\frac{1}{2\lambda^2}[R^{\mathrm{Q}}_{\mu\nu}(\Gamma) - 2\lambda\sigma_{\mu\nu}] \tag{8.50}$$

where

$$\begin{aligned}\sigma_{\mu\nu} = \partial^\rho[&-\eta_{\mu\nu}(\bar{\varphi}_\lambda{}^\tau\Gamma^\lambda_{\rho\tau} - \tfrac{1}{2}\bar{\varphi}\Gamma_\rho) - 2\bar{\varphi}_\rho{}^\lambda\Gamma_{(\mu\nu)\lambda}\\ &- 2\bar{\varphi}_{(\mu}{}^\lambda\Gamma_{\rho\lambda\nu)} + 2\bar{\varphi}_{(\mu}{}^\lambda\Gamma_{\nu)\rho\lambda} - \bar{\varphi}_{\mu\nu}\Gamma_\rho + 2\bar{\varphi}_{\rho(\mu}\Gamma_{\nu)}].\end{aligned}$$

The indices of φ and Γ are raised and lowered with the help of η, and $\Gamma_{\mu\nu\lambda} = \eta_{\mu\rho}\Gamma^\rho_{\nu\lambda}$.

Since $\bar{t}'_{\mu\nu}$ is a complicated function of the fields, it is not clear whether the new Lagrangian can be found, as in the Yang–Mills case, by simply adding the term $-\lambda\bar{\varphi}^{\mu\nu}(\alpha\bar{t}'_{\mu\nu})$. A direct verification shows that we should first remove the term $\sigma_{\mu\nu}$ from $\bar{t}'_{\mu\nu}$, so that

$$\mathcal{L}^{(1)}_{\mathrm{T}} = \mathcal{L}_{\mathrm{T}} + \lambda\bar{\varphi}^{\mu\nu}\frac{1}{2\lambda^2}\alpha R^{\mathrm{Q}}_{\mu\nu}$$

whereupon the correct result is deduced for $\alpha = 2$:

$$\begin{aligned}\mathcal{L}_{\mathrm{T}}^{(1)} &= -\frac{1}{2\lambda^2}[-2\lambda\bar{\varphi}^{\mu\nu}(R^{\mathrm{L}}_{\mu\nu} + R^{\mathrm{Q}}_{\mu\nu}) + \eta^{\mu\nu}R^{\mathrm{Q}}_{\mu\nu}] \\ &= -\frac{1}{2\lambda^2}[(\eta^{\mu\nu} - 2\lambda\bar{\varphi}^{\mu\nu})(R^{\mathrm{L}}_{\mu\nu} + R^{\mathrm{Q}}_{\mu\nu}) - \eta^{\mu\nu}R^{\mathrm{L}}_{\mu\nu}].\end{aligned} \tag{8.51}$$

If we define

$$\tilde{g}^{\mu\nu} \equiv \sqrt{-g}g^{\mu\nu} = (\eta^{\mu\nu} - 2\lambda\bar{\varphi}^{\mu\nu}) \tag{8.52}$$

and discard the divergence $\eta^{\mu\nu}R^{\mathrm{L}}_{\mu\nu} = \partial_\mu(\Gamma^\mu - \Gamma^{\mu\nu}{}_\nu)$, the Lagrangian $\mathcal{L}_{\mathrm{T}}^{(1)}$ can be easily identified with Einstein's Lagrangian in the first order formalism.

Note that the term $\bar{\varphi}^{\mu\nu}R^{\mathrm{Q}}_{\mu\nu}$ does not depend on γ, so that the energy–momentum tensor associated with $\mathcal{L}_{\mathrm{T}}^{(1)}$ remains the same as $t'_{\mu\nu}$, whereby the construction of a self-consistent theory is completed in a *single* step. The expression $\bar{\varphi}^{\mu\nu}R^{\mathrm{Q}}_{\mu\nu}$ remains γ-independent because we chose $\bar{\varphi}^{\mu\nu}$ to be a tensor density. Were it not so, the corresponding energy–momentum tensors would be equal only after using the field equations following from $\mathcal{L}_{\mathrm{T}}$ and $\mathcal{L}_{\mathrm{T}}^{(1)}$.

In order to verify condition (G) explicitly, we start from the equation $\delta\mathcal{L}_{\mathrm{T}}^{(1)}/\delta\Gamma = 0$ written in the form

$$-2\lambda\bar{\varphi}^{\mu\nu}{}_{,\lambda} = \tilde{g}^{\mu\nu}\Gamma_\lambda - 2\tilde{g}^{(\mu\rho}\Gamma^{\nu)}_{\lambda\rho}.$$

Differentiating this equation we can form on the left-hand side the expressions $\Box\bar{\varphi}^{\mu\nu}$, $\bar{\varphi}^{(\mu\rho,\nu)}{}_\rho$ and $\Box\bar{\varphi}$, and combine them into $R^{\mathrm{L}}_{(\mu\nu)}$. On the right-hand side we should use $\tilde{g}^{\mu\nu} = \eta^{\mu\nu} - 2\lambda\bar{\varphi}^{\mu\nu}$. The result of this calculation is the correct equation in the form (G):

$$\frac{1}{2\lambda}R^{\mathrm{L}}_{(\mu\nu)}(\bar{\varphi}) = \lambda\bar{t}'_{\mu\nu}.$$

Note that this equation is also correct in the strong gravitational field when $\bar{\varphi}^{\mu\nu}$ is not a small field.

The assumption that $\bar{\varphi}^{\mu\nu}$ is a tensor density reduces the generality and the power of this method to some extent, but the simplicity and clarity of the resulting structure is more than sufficient compensation.

Interaction with matter. We now want to find the form of the consistent interaction between gravity and matter. In the presence of matter, the right-hand side of equation (G) should contain a contribution from both the *gravity* and *matter fields*:

$$\frac{1}{2\lambda}R^{\mathrm{L}}_{\mu\nu}(\bar{\varphi}) = \lambda(\bar{t}'_{\mu\nu} + \overline{T}_{\mu\nu}). \tag{Gm}$$

We now introduce the following two assumptions:

- $\mathcal{L}_{\rm MI}$ does not influence the construction of the gravitational energy–momentum $\bar{t}'_{\mu\nu}$, since it does not depend on $\partial\varphi_{\mu\nu}$.
- The dynamical tensor $\overline{T}_{\mu\nu}$ is equal to the symmetrized canonical tensor.

Using the Rosenfeld definition of $\overline{T}_{\mu\nu}$ the second assumption may be expressed by the relation

$$2\frac{\delta\mathcal{L}_{\rm M}(\gamma)}{\delta\gamma^{\mu\nu}}\bigg|_{\gamma=\eta} = -\frac{1}{\lambda}\frac{\delta\mathcal{L}_{\rm MI}}{\delta\bar{\varphi}^{\mu\nu}}$$

which implies

$$\mathcal{L}_{\rm MI} = \mathcal{L}_{\rm M}(\eta^{\mu\nu} - 2\lambda\bar{\varphi}^{\mu\nu}). \tag{8.53}$$

In other words, $\mathcal{L}_{\rm MI}$ is equal to the covariantized matter Lagrangian, obtained from $\mathcal{L}_{\rm M}$ by the replacement $\eta^{\mu\nu} \to \tilde{g}^{\mu\nu} \equiv \eta^{\mu\nu} - 2\lambda\bar{\varphi}^{\mu\nu}$ and $\partial_\mu \to \nabla_\mu$.

> *The gravitational field is introduced in the sector of matter fields by the minimal interaction, in accordance with the principle of equivalence.*

The geometric interpretation of GR is seen to be a consequence of the self-consistency requirement.

In conclusion, here is one final comment concerning the power of the first order formalism. Qualitatively small changes in the first order Lagrangian lead to significant changes in the theory, because they change not only the field equation for $\bar{\varphi}^{\mu\nu}$, but also the dependence $\Gamma = \Gamma(\varphi)$; this is why it becomes possible to introduce the complete nonlinear correction in only one step.

Exercises

1. Consider the following $SU(2)$ invariant field theory:

$$\mathcal{L}^{(0)} = -\tfrac{1}{4}\overset{\circ}{F}{}^{a}{}_{\mu\nu}\overset{\circ}{F}{}^{a\mu\nu} + \tfrac{1}{2}(\partial_\mu\phi^a\partial^\mu\phi^a - m^2\phi^a\phi^a) - gA^a\cdot J^a_{(0)}$$

   where the matter fields ϕ^a belong to the triplet representation of $SU(2)$ and $J^{b\mu}_{(0)} = -\epsilon^{bca}\phi^c\partial^\mu\phi^a$ is the canonical matter current. Construct the related self-consistent theory.

2. The self-consistent action for the theory (8.2) has the form $I = I_{\rm F} + I_{\rm MI}$.
   (a) Show that the matter current J^μ_a satisfies the relation $\partial_\nu J^\nu_a - g\epsilon^{abc}A^b{}_\nu J^\nu_c = 0$, and derive the condition

$$\partial_\nu\frac{\delta I_{\rm F}}{\delta A^a_\nu} - g\epsilon^{abc}A^b_\nu\frac{\delta I_{\rm F}}{\delta A^c_\nu} = 0.$$

   (b) Show that this condition is equivalent to the $SU(2)$ gauge invariance of $I_{\rm F}$.

3. Find the symmetric energy–momentum tensor $t_{\mu\nu}^{(0)}$ of the tensor gravitational field, described by the Lagrangian $\mathcal{L}_{\mathrm{T}}$. Show that the first correction $\Lambda^{(1)}$ of $\mathcal{L}_{\mathrm{T}}$, given by equation (8.26), satisfies the condition $-\delta\Lambda/\delta\varphi^{\mu\nu} = t_{\mu\nu}^{(0)}$.
4. Show that the divergence condition (8.29), expressed in terms of the true tensor $T^{\mu\nu} = \widetilde{T}^{\mu\nu}/\sqrt{-g}$, has the covariant form: $\nabla_\mu T^{\mu\nu} = 0$.
5. Consider the tensor theory of gravity interacting with a massless scalar matter field, described by $\mathcal{L}_{\mathrm{M}} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi$. Check explicitly whether the matter energy–momentum tensor $T_{(0)}^{\mu\nu}$ satisfies the divergence condition (8.29).
6. Derive the transformation law of $g = \det(g_{\mu\nu})$ under the transformations (8.32). Show that the quantity $\int \mathrm{d}^4x\,\sqrt{-g}\sigma(x)$, where $\sigma(x)$ is a scalar field, is invariant under infinitesimal coordinate transformations (8.41).
7. Show that the action of the relativistic particle I_{MI}, given by equation (8.42), is invariant under the generalized gauge transformations (8.41).
8. Prove the relation

$$\bar{t}_{\mu\nu} \equiv t_{\mu\nu} - \tfrac{1}{2}\eta_{\mu\nu}t = 2\frac{\delta\mathcal{L}(\gamma)}{\delta\tilde{\gamma}^{\mu\nu}}\bigg|_{\gamma=\eta}.$$

9. Use the relation $\tilde{g}^{\mu\nu} = \eta^{\mu\nu} - 2\lambda\bar{\varphi}^{\mu\nu}$ to show that for a weak gravitational field $g_{\mu\nu} = \eta_{\mu\nu} + 2\lambda\varphi_{\mu\nu}$.
10. Derive expression (8.50) for the Rosenfeld energy–momentum tensor of the gravitational field.
11. Show that the equation of motion for Γ, obtained from the action $I_{\mathrm{T}}^{(1)}$, has the form

$$2\tilde{g}^{(\mu\rho}\Gamma^{\nu)}_{\lambda\rho} - \tilde{g}^{\mu\nu}\Gamma^{\rho}_{\lambda\rho} - \tilde{g}^{\rho\sigma}\Gamma^{(\mu}_{\rho\sigma}\delta^{\nu)}_{\lambda} - \tilde{g}^{\mu\nu}{}_{,\lambda} - \tilde{g}^{(\mu\rho}{}_{,\rho}\delta^{\nu)}_{\lambda} = 0.$$

Derive the relation $\tilde{g}^{\mu\nu}{}_{,\lambda} = \tilde{g}^{\mu\nu}\Gamma_\lambda - 2\tilde{g}^{(\mu\rho}\Gamma^{\nu)}_{\lambda\rho}$, and solve it for Γ.
12. Construct a self-consistent tensor theory of gravity starting from the Lagrangian $\mathcal{L}_{\mathrm{T}}(\gamma)$ in which $\bar{\varphi}^{\mu\nu}$ is a tensor, rather than a tensor density.

Chapter 9

Supersymmetry and supergravity

The concept of the unified nature of all particles and their basic interactions, which started with Maxwell's unification of electricity and magnetism, led in the 1970s to a successful unification of the weak, electromagnetic and, to some extent, strong interactions. One of the central goals of physics at the present time is the unification of gravity with the other basic interactions, within a consistent quantum theory.

The structure of spacetime at low energies is, to a high degree of accuracy, determined by the Poincaré group. In the 1960s, there was conviction that the underlying unity of particle interactions might have been described by a non-trivial unification of the Poincaré group with some internal symmetries (isospin, $SU(3)$, etc). Many attempts to realize this idea were based on Lie groups, in which particles in a given multiplet have the same statistics—they are either bosons or fermions. The failure of these attempts led finally to a number of no-go theorems, which showed that such constructions are, in fact, not possible in the framework of relativistic field theories and standard Lie groups (Coleman and Mandula 1967).

Supersymmetry is a symmetry that relates *bosons* and *fermions* in a way which is consistent with the basic principles of quantum field theory. It is characterized by both the *commutation* and *anticommutation* relations between the symmetry generators, in contrast to the standard Lie group structure. The results obtained in the 1970s show that supersymmetric theories are the only ones which can lead to a non-trivial unification of spacetime and the internal symmetries within a relativistic quantum field theory. One more point that singles out supersymmetric theories is the problem of quantum divergences. Early investigations of the quantum properties of supersymmetric field theories led to impressive results: some well-known perturbative divergences stemming from bosons and fermions 'cancelled' out just because of supersymmetry (see, e.g., Sohnius 1985).

Since the concept of gauge invariance has been established as the basis for our understanding of particle physics, it was natural to elevate the idea of supersymmetry to the level of gauge symmetry, thus introducing the gravitational interaction into the world of supersymmetry. The investigation of globally supersymmetric models induced a great optimism that the new theory, supergravity, might lead to a consistent formulation of quantum gravity. Quantum supergravity is found to be more finite than ordinary GR. To what extent these results might be extended to a satisfying quantum theory of gravity remains a question for the future (see, e.g., van Nieuwenhuizen 1981a).

There is no firm experimental evidence that supersymmetry is realized in nature. However, on the theoretical side, there is hope that supersymmetry could provide a consistent unification of gravity with the other basic interactions, in the framework of a consistent quantum theory. If this were the case, supersymmetry would relate physical phenomena at two very distinct scales: the Planck and the weak interaction scale. For the phenomenological applications of supersymmetry and the structure of quantum theory, the reader should consult the existing literature (see, e.g., West 1986, Srivastava 1986, Bailin and Love 1994).

This chapter deals with the basic aspects of supersymmetry (SS) and supergravity in the context of classical field theories. Technical appendix J is an integral part of this exposition.

9.1 Supersymmetry

Fermi–Bose symmetry

Consider a field theory in which there is a symmetry between the bosons and fermions, by which a boson field and a fermion field can 'rotate' into each other by an 'angle' ε (van Nieuwenhuizen 1981a, Sohnius 1985, West 1986, Srivastava 1986, Bailin and Love 1994). The general transformation law which 'rotates' a boson into a fermion has the form†

$$\delta A(x) = \varepsilon\psi(x)$$

where all indices are suppressed. This relation implies several interesting consequences.

Spin. Since bosons have integer, and fermions half-integer spin, the parameters ε must have *half-integer* spin. The simplest choice is $s = \frac{1}{2}$; we assume that $\varepsilon = (\varepsilon_\alpha)$ is a four-dimensional Dirac spinor.

Statistics. In quantum field theory bosons are commuting, and fermions anticommuting objects, according to the spin-statistics theorem. By taking the $\hbar \to 0$ limit, we can retain these unusual (anti)commuting properties even in the classical limit. Thus, we consider here a classical field theory in which

$$\{\varepsilon_\alpha, \varepsilon_\beta\} = \{\varepsilon_\alpha, \psi_\beta\} = \{\psi_\alpha, \psi_\beta\} = 0 \qquad [\varepsilon_\alpha, A] = [\psi_\beta, A] = 0.$$

† In this chapter only, the usual symbol of the form variation δ_0 is replaced by δ.

Dimension. The canonical dimensions of the boson and fermion fields are given by $d(A) = 1,\ d(\psi) = 3/2$, in mass units, as follows from the fact that the action is dimensionless. Hence, the transformation law $\delta A = \varepsilon\psi$ implies $d(\varepsilon) = -1/2$. The inverse transformation cannot have the form $\delta\psi \sim \varepsilon A$, since there is one missing unit of dimension on the right-hand side. If the transformation law does not depend on mass or coupling constants, the only object that can fill this dimensional gap is a spacetime derivative.

Object:	I	d^4x	$\partial/\partial x^\mu$	A	ψ	ε
dimension:	0	-4	1	1	$3/2$	$-1/2$

Thus, purely on dimensional grounds we find

$$\delta\psi(x) = \varepsilon\partial A(x).$$

Reality. If we restrict ourselves to the simplest case and assume that bosonic fields are real scalars and fermionic fields are Majorana spinors, then the reality of δA implies that ε is a Majorana spinors.

Taking into account all these properties, we can write the exact formula for the transformation law, with all indices and in completely covariant form:

$$\begin{aligned}\delta A &= \bar\varepsilon^\alpha\psi_\alpha \\ \delta\psi_\alpha &= -\mathrm{i}(\gamma^\mu\varepsilon)_\alpha\partial_\mu A.\end{aligned}$$

Algebra. If these transformations are to represent a symmetry operation, they have to obey a closed algebra. Consider the commutator of two infinitesimal supersymmetric (SS) transformations on A,

$$[\delta(\varepsilon_1), \delta(\varepsilon_2)]A = 2\mathrm{i}\bar\varepsilon_1\gamma^\mu\varepsilon_2\partial_\mu A$$

obtained by using the identity $\bar\varepsilon_1\gamma^\mu\varepsilon_2 = -\bar\varepsilon_2\gamma^\mu\varepsilon_1$. It shows that two global SS transformations lead to *spacetime translations*. This is why global supersymmetry is said to be the 'square root' of translations, while local supersymmetry is expected to be the 'square root' of gravity. A discussion of the complete SS algebra is given later.

Degrees of freedom. As we shall see, an important consequence of this basic relation of the SS algebra is that a representation of supersymmetry contains an equal number of boson and fermion components. A massless Majorana spinor satisfying one complex field equation has two independent degrees of freedom. Knowing this we might anticipate an SS field theory involving a Majorana spinor and two boson fields.

Wess–Zumino model. A simple example of an SS theory in M_4 is the Wess–Zumino model. We start with a Majorana spinor ψ and two boson fields, a scalar A and a pseudoscalar B. The action for the free massless fields is given by

$$I_0 = \int d^4x\, [\tfrac{1}{2}\partial_\mu A \partial^\mu A + \tfrac{1}{2}\partial_\mu B \partial^\mu B + \tfrac{1}{2}\mathrm{i}\bar{\psi}\gamma \cdot \partial\psi]. \tag{9.1}$$

On the grounds of linearity, covariance, dimension and parity we find the following set of SS transformations:

$$\begin{aligned} \delta A &= \bar{\varepsilon}\psi \qquad \delta B = \bar{\varepsilon}\gamma_5\psi \\ \delta\psi &= -\mathrm{i}\gamma^\mu\partial_\mu(aA + b\gamma_5 B)\varepsilon \end{aligned} \tag{9.2a}$$

where a and b are dimensionless parameters, and ε is a Majorana spinor. We can easily verify that these transformations are symmetries of the equations of motion:

$$\Box A = 0 \qquad \Box B = 0 \qquad \mathrm{i}\gamma \cdot \partial\psi = 0.$$

The variation of the action yields, up to a four-divergence,

$$\delta I_0 = \int d^4x\, [\partial_\mu A \partial^\mu(\bar{\varepsilon}\psi) + \partial_\mu B \partial^\mu(\bar{\varepsilon}\gamma_5\psi) + \bar{\psi}\Box(aA + b\gamma_5 B)\varepsilon]$$

showing that the invariance holds provided $a = b = 1$. Thus, the final form of SS transformations is:

$$\begin{aligned} &\delta A = \bar{\varepsilon}\psi \qquad \delta B = \bar{\varepsilon}\gamma_5\psi \\ &\delta\psi = -\mathrm{i}\gamma \cdot \partial(A + \gamma_5 B)\varepsilon \qquad [\delta\bar{\psi} = \mathrm{i}\bar{\varepsilon}\gamma^\mu\partial_\mu(A - \gamma_5 B)]. \end{aligned} \tag{9.2b}$$

The commutator of two supersymmetries, with parameters ε_1 and ε_2, on the fields A, B and ψ, has the form

$$\begin{aligned} [\delta_1, \delta_2]A &= 2\mathrm{i}\bar{\varepsilon}_1\gamma^\mu\varepsilon_2\partial_\mu A \\ [\delta_1, \delta_2]B &= 2\mathrm{i}\bar{\varepsilon}_1\gamma^\mu\varepsilon_2\partial_\mu B \\ [\delta_1, \delta_2]\psi &= 2\mathrm{i}\bar{\varepsilon}_1\gamma^\mu\varepsilon_2\partial_\mu\psi - \bar{\varepsilon}_1\gamma^\mu\varepsilon_2\gamma_\mu F_{\bar{\psi}} \end{aligned} \tag{9.3}$$

where $F_{\bar{\psi}} \equiv \mathrm{i}\gamma \cdot \partial\psi$ is the equation of motion for ψ. In these calculations we make use of the Fierz identity:

$$(\bar{\varepsilon}_2\psi)(\bar{\chi}\varepsilon_1) + (\bar{\varepsilon}_2\gamma_5\psi)(\bar{\chi}\gamma_5\varepsilon_1) - (\varepsilon_1 \leftrightarrow \varepsilon_2) = -(\bar{\varepsilon}_2\gamma_\mu\varepsilon_1)(\bar{\chi}\gamma^\mu\psi).$$

The invariance of the action holds without using the equations of motion (off-shell), while the SS algebra closes only when the fields are subject to these equations (on-shell). This can be explained by noting that the fields A, B and ψ carry an on-shell representation of SS, since the number of boson and fermion components are only equal on-shell. The boson–fermion balance can be restored off-shell by adding two more boson fields, as we shall see soon.

There is a formulation of this model based on the fact that the kinetic term for a Majorana field can be expressed in terms of its left (or right) chiral component, as follows from the relation

$$\bar\psi\gamma\cdot\partial\psi = \bar\psi_-\gamma\cdot\partial\psi_- + \bar\psi_+\gamma\cdot\partial\psi_+ = 2\bar\psi_-\gamma\cdot\partial\psi_- - \partial\cdot(\bar\psi_-\gamma\psi_-)$$

where we have used $\bar\psi_+\gamma\cdot\partial\psi_+ = -(\partial\bar\psi_-)\cdot\gamma\psi_-$. Introducing the complex field $\mathcal{A} = A + \mathrm{i}B$, the action (9.1) can be written as

$$I_0' = \int \mathrm{d}^4x\,[\tfrac{1}{2}\partial_\mu\mathcal{A}^*\partial^\mu\mathcal{A} + \mathrm{i}\bar\psi_-\gamma\cdot\partial\psi_-] \tag{9.4a}$$

while the SS transformations (9.2b) take the form

$$\begin{aligned}\delta\mathcal{A} &= \bar\varepsilon(1+\mathrm{i}\gamma_5)\psi = 2\bar\varepsilon_+\psi_- \\ \delta\psi_- &= P_-\delta\psi = -\mathrm{i}\gamma^\mu\varepsilon_+\partial_\mu\mathcal{A}\end{aligned} \tag{9.4b}$$

where $P_- \equiv \frac{1}{2}(1+\mathrm{i}\gamma_5)$ is the chiral projector. The expression for $\delta\psi_-$ is obtained using $P_-\gamma\cdot\partial(A+\gamma_5 B) = \gamma\cdot\partial\mathcal{A}P_+$. An attempt to reduce the number of bosons to one by setting $B = 0$ is not consistent, since it leads to real $\mathcal{A}$ while $\delta\mathcal{A}$ remains complex ($\bar\varepsilon_+\psi_-$ is not a real quantity). We see that infinitesimal SS transformations can be realized on the massless doublet of complex fields $(\mathcal{A}, \psi_-)$ with helicity $(0, \frac{1}{2})$. This doublet constitutes an irreducible representation of a new kind of algebra—the *superalgebra*.

We also note that the action (9.1) is invariant under two additional global transformations—chiral and phase transformations:

$$\psi \to \mathrm{e}^{\alpha\gamma_5}\psi \qquad \mathcal{A} \to \mathrm{e}^{-\mathrm{i}\beta}\mathcal{A}. \tag{9.5}$$

Super-Poincaré algebra. The field transformations (9.2) can be expressed in terms of the infinitesimal SS generators Q:

$$\delta = \bar\varepsilon^\alpha Q_\alpha$$

where Q is, like ε, a Majorana spinor, which acts on the fields A, B and ψ according to

$$\begin{aligned}&Q_\alpha(A) = \psi_\alpha \qquad Q_\alpha(B) = \gamma_5\psi_\alpha \\ &Q_\alpha(\psi_\beta) = -\mathrm{i}[\gamma\cdot\partial(A+\gamma_5 B)C]_{\beta\alpha}.\end{aligned}$$

Here, $Q(\phi)$ denotes the result of the action of Q on ϕ, and can be equivalently represented as the commutator $[Q, \phi]$.

If we discard the term proportional to the equations of motion in the algebra (9.3), it can be written as

$$[\delta_1, \delta_2] = 2\mathrm{i}\bar{\varepsilon}_1\gamma^\mu\varepsilon_2\partial_\mu .$$

The left-hand side of this equation can be expressed in terms of the SS generators as $[\delta_1, \delta_2] = -\bar{\varepsilon}_1^\alpha\bar{\varepsilon}_2^\beta\{Q_\alpha, Q_\beta\}$, which implies

$$\{Q_\alpha, Q_\beta\} = 2\mathrm{i}(\gamma^\mu C)_{\alpha\beta}P_\mu \tag{9.6a}$$

where $P_\mu = -\partial_\mu$. The generators Q_α, together with P_μ and $M_{\mu\nu}$, give rise to a closed algebra called super-Poincaré algebra. The transformation properties of the constant Dirac spinor Q_α under Lorentz transformations and translations are expressed by

$$[M_{\mu\nu}, Q_\alpha] = -(\sigma_{\mu\nu})_\alpha{}^\beta Q_\beta \tag{9.6b}$$
$$[P_\mu, Q_\alpha] = 0 \tag{9.6c}$$

where $\sigma_{\mu\nu}$ are Lorentz generators in the appropriate representation.

The relations (9.6a, b, c), together with the commutation relations of the Poincaré algebra, express the (on-shell) SS structure of the Wess–Zumino model in terms of the super-Poincaré algebra:

$$\begin{aligned}
&[M_{\mu\nu}, M_{\lambda\rho}] = \tfrac{1}{2}f_{\mu\nu,\lambda\rho}{}^{\tau\sigma}M_{\tau\sigma}\\
&[M_{\mu\nu}, P_\lambda] = \eta_{\nu\lambda}P_\mu - \eta_{\mu\lambda}P_\nu \qquad [P_\mu, P_\nu] = 0\\
&[M_{\mu\nu}, Q_\alpha] = -(\sigma_{\mu\nu})_\alpha{}^\beta Q_\beta \qquad [P_\mu, Q_\alpha] = 0\\
&\{Q_\alpha, Q_\beta\} = 2\mathrm{i}(\gamma^\mu C)_{\alpha\beta}P_\mu \qquad \{Q_\alpha, \bar{Q}_\beta\} = -2\mathrm{i}(\gamma^\mu)_{\alpha\beta}P_\mu .
\end{aligned} \tag{9.7a}$$

Since Q_α is a Majorana spinor, the expression $\{Q_\alpha, \bar{Q}_\beta\}$ is not independent of $\{Q_\alpha, Q_\beta\}$, but is displayed here for completeness.

It is not difficult to see that this algebra remains unaltered under chiral transformations $Q \to \mathrm{e}^{\alpha\gamma_5}Q$. The invariance of $\{Q_\alpha, Q_\beta\}$ follows from $\gamma_5(\gamma^\mu C) + (\gamma^\mu C)\gamma_5^T = 0$. The super-Poincaré algebra may be enlarged by adding the chiral generator R satisfying

$$[R, Q_\alpha] = (\gamma_5 Q)_\alpha \tag{9.7b}$$

while R commutes with P and M.

Relation (9.6a) is the basic relation of the SS algebra. It clearly shows that supersymmetry is a spacetime symmetry, the 'square root' of translations ($Q^2 \sim P$). It is also clear that localization of supersymmetry will lead to gravity.

Supersymmetric extension of the Poincaré algebra

As we have already mentioned at the beginning of this chapter, in the 1960s physicists attempted to find a symmetry that would unify the Poincaré group with an internal symmetry group in a *non-trivial* manner. After much effort it became clear that such a unity could not be achieved in the context of relativistic field theory and a Lie group of symmetry.

Coleman and Mandula (1967) showed, with very general assumptions, that any Lie group of symmetries of the S matrix in relativistic field theory must be a *direct product* of the Poincaré group with an internal symmetry group. If G is an internal symmetry group with generators T_m, such that $[T_m, T_n] = f_{mn}{}^r T_r$, the direct product structure is expressed by the relations

$$[T_m, P_\mu] = [T_m, M_{\mu\nu}] = 0.$$

It follows from this that the generators T_m commute with the Casimir operators of the Poincaré group:

$$[T_m, P^2] = 0 \qquad [T_m, W^2] = 0.$$

In other words, all members of an irreducible multiplet of the internal symmetry group must have the same mass and the same spin.

> *Only the inclusion of fermionic symmetry generators opens the possibility of a non-trivial extension of the Poincaré symmetry.*

Supersymmetry algebra. The structure of a Lie group in the neighbourhood of the identity is determined entirely by its Lie algebra, which is based on the *commutation* relations between the generators. The essential property of an SS algebra is the existence of fermionic generators which satisfy *anticommutation* relations. Every Lie algebra is characterized by the Jacobi identities being the consistency conditions on the algebra. Similar, generalized Jacobi identities also exist in SS algebras. If we denote bosonic generators by B and fermionic generators by F, an SS algebra has the form

$$[B_1, B_2] = B_3 \qquad [B_1, F_2] = F_3 \qquad \{F_1, F_2\} = B_3 \tag{9.8a}$$

and the super-Jacobi identities read:

$$\begin{aligned}
&[[B_1, B_2], B_3] + [[B_3, B_1], B_2] + [[B_2, B_3], B_1] = 0 \\
&[[B_1, B_2], F_3] + [[F_3, B_1], B_2] + [[B_2, F_3], B_1] = 0 \\
&\{[B_1, F_2], F_3\} + \{[F_3, B_1], F_2\} + [\{F_2, F_3\}, B_1] = 0 \\
&[\{F_1, F_2\}, F_3] + [\{F_3, F_1\}, F_2] + [\{F_2, F_3\}, F_1] = 0.
\end{aligned} \tag{9.8b}$$

We can verify that these relations are indeed identities by explicitly expanding each (anti)commutator.

An algebra of the type (9.8) is in mathematics known as a Z_2 *graded Lie algebra*, and can be defined by the following general requirements.

Grading. To each generator (an element of a vector space) we associate a grading g from $Z_2 = (0, 1)$: $g(B) = 0$, $g(F)$=1. For each two generators we define a product rule $\circ$, such that $g(G_1 \circ G_2) = g(G_1) + g(G_2)$ (mod 2). This defines a Z_2 graded algebra.
Supersymmetry. The product rule satisfies the condition of SS or graded antisymmetry: $G_1 \circ G_2 = -(-)^{g_1 g_2} G_2 \circ G_1$.
Super-Jacobi identity. The product rule satisfies the consistency condition:

$$(-1)^{g_1 g_3} G_1 \circ (G_2 \circ G_3) + \text{cyclic}(1, 2, 3) = 0.$$

The first two properties are realized by choosing $G_1 \circ G_2 = [G_1, G_2]$ or $\{G_1, G_2\}$, in accordance with (9.8a), while the super-Jacobi identity reduces to (9.8b).

Consider now an SS extension of the Poincaré algebra. Let us introduce an SS generator Q_α which is, by assumption, a constant (translation invariant) Dirac spinor. These assumptions can be expressed by the relations (9.6b, c). We now check various super-Jacobi identities. Identities for (P, P, Q) and (M, P, Q) are easily verified. To prove the identity for (M, M, Q) it is essential to observe that the matrices $\sigma_{\mu\nu}$ form a representation of the Lorentz algebra. Indeed, the relation

$$[[M_{\mu\nu}, M_{\lambda\rho}], Q_\alpha] + [[Q_\alpha, M_{\mu\nu}], M_{\lambda\rho}] + [[M_{\lambda\rho}, Q_\alpha], M_{\mu\nu}] = 0$$

after using the expressions (9.7a) for $[M, M]$ and $[M, Q]$, reduces to

$$\tfrac{1}{2} f_{\mu\nu,\lambda\rho}{}^{\sigma\tau}\, \sigma_{\sigma\tau} - [\sigma_{\mu\nu}, \sigma_{\lambda\rho}] = 0.$$

This is a correct result since the matrices $\sigma_{\mu\nu}$ satisfy the Lorentz algebra.

The anticommutator of two SS generators is, in general, given as a linear combination of bosonic generators,

$$\{Q_\alpha, Q_\beta\} = \mathrm{i}a(\gamma^\mu C)_{\alpha\beta} P_\mu + \mathrm{i}b(\sigma^{\mu\nu} C)_{\alpha\beta} M_{\mu\nu}$$

where the right-hand side is symmetric under the exchange of α and β, as follows from the properties of gamma matrices, and a and b are constants. The Jacobi identity for (P, Q, Q) implies $b = 0$, while the identities for (M, Q, Q) and (Q, Q, Q) are automatically satisfied.

We can impose some additional restrictions on Q_α without changing the structure of the algebra. Four complex operators Q_α have eight real components. If we assume that Q_α is a Majorana spinor, then $\bar{Q}_\alpha$ and Q_α are no longer independent. We can show that this assumption does not represent a loss of generality in the present case (West 1986).

Up to now, the coefficient a has remained arbitrary. We shall see that the positivity of energy implies $a > 0$, and the standard choice $a = 2$ is achieved by rescaling Q_α. This yields the standard SS extension of the Poincaré algebra, as in equation (9.7*a*).

The previous discussion assumes that there is only one SS generator Q_α and that the internal symmetry is trivial. A non-trivial internal symmetry may be easily incorporated into the preceding analysis. Consider a set of SS generators Q^m_α ($m = 1, 2, \ldots, N$) belonging to an irreducible representation of some internal symmetry group with generators T_m:

$$\begin{aligned} [T_m, T_n] &= f_{mn}{}^k T_k \qquad [T_m, Q^n_\alpha] = t_m{}^n{}_k Q^k_\alpha \\ [M_{\mu\nu}, T_m] &= 0 \qquad [P_\mu, T_m] = 0. \end{aligned} \tag{9.9a}$$

In this case we find the following result:

$$\{Q^m_\alpha, Q^n_\beta\} = 2\delta^{mn}(\gamma^\mu C)_{\alpha\beta} P_\mu + C_{\alpha\beta} Z^{mn}_1 + (\gamma_5 C)_{\alpha\beta} Z^{mn}_2 \tag{9.9b}$$

where the antisymmetric objects Z^{mn}_1 and Z^{mn}_2 commute with all the generators and represent the so-called *central charges* of the algebra.

If there exists only one SS generator Q_α, we have a *simple supersymmetry*, while the case $N > 1$ describes an *extended supersymmetry*. Even in the case of simple supersymmetry we may have a non-trivial internal symmetry, but only of the $U(1)$ type; this symmetry, known as the chiral symmetry, is expressed by equation (9.7*b*).

This concludes our exposition of possible SS extensions of the Poincaré algebra. If all masses are zero, the Poincaré symmetry carries over into the conformal symmetry, and the related SS extension is called the conformal supersymmetry. We limit our discussion entirely to the super-Poincaré symmetry, often using the more general term supersymmetry to mean the same for simplicity.

Consequences. There are several immediate consequences of the SS algebra, which have important physical implications.

Mass degeneracy. The relation $[P_\mu, Q_\alpha] = 0$ implies $[P^2, Q_\alpha] = 0$, i.e. P^2 is a Casimir operator of the SS algebra. Hence,

all particles in a multiplet of supersymmetry have the same mass.

However, there is no experimental evidence that elementary particles come in mass-degenerate SS multiplets. Hence, if supersymmetry has any relevance in nature, it must be realized as a broken symmetry.

Positivity of energy. From the relation $\{Q_\alpha, Q_\beta\} = \mathrm{i}a(\gamma^\mu C)_{\alpha\beta} P_\mu$ and using the fact that Q_α is a Majorana spinor, we find that

$$\{Q_\alpha, \bar{Q}_\beta\} = -\mathrm{i}a(\gamma^\mu)_{\alpha\beta} P_\mu.$$

Replacing here $P_\mu \to \mathrm{i}(E, -\boldsymbol{p})$, multiplying with γ^0 and performing a trace over spinor indices, we find that

$$\sum (Q_\alpha Q_\alpha^+ + Q_\alpha^+ Q_\alpha) = 4aE.$$

If the space in which SS generators act has a positive definite norm, the left-hand side of this equality is always ≥ 0, hence, for $a > 0$, the following statement holds:

The energy in SS theories cannot take negative values: $E \geq 0$.

The boson–fermion balance. We have seen that SS theories can be realized on a set of fields of definite mass and spin. Every representation of the SS algebra can be decomposed into representations of the Poincaré subalgebra. We can use relation (9.6a) to derive a useful theorem that relates the number of bosonic and fermionic degrees of freedom in each representation of supersymmetry.

Since SS transformations relate bosons and fermions, each representation space of the SS algebra can be divided into a bosonic and a fermionic subspace, $\mathcal{B}$ and $\mathcal{F}$. The bosonic generators map each of these subspaces into itself, while the fermionic generators map $\mathcal{B}$ into $\mathcal{F}$, and vice versa. Note that these mappings are, in general, not onto. Consider the action of a composition of two SS generators on the bosonic subspace $\mathcal{B}$ (figure 9.1):

$$Q_\alpha : \mathcal{B} \to \mathcal{F}_1 \subseteq \mathcal{F} \qquad Q_\beta : \mathcal{F}_1 \to \mathcal{B}_2 \subseteq \mathcal{B}.$$

The anticommutator $\{Q_\alpha, Q_\beta\}$ has a similar effect. If the mapping $P_\mu : \mathcal{B} \to \mathcal{B}$ is onto and 1–1, then, as follows from the relation (9.6a), the mapping $\{Q_\alpha, Q_\beta\} : \mathcal{B} \to \mathcal{B}$ is also onto and 1–1, which implies that Q_α must be onto and 1–1; consequently, $\mathcal{B}_2 = \mathcal{B}$. Similar arguments applied to the mapping $\mathcal{F} \to \mathcal{B} \to \mathcal{F}$ lead to the conclusion $\mathcal{F}_1 = \mathcal{F}$. In other words, the bosonic and fermionic subspaces $\mathcal{B}$ and $\mathcal{F}$ have the same number of elements or the same dimension.

What exactly is the meaning of the 'number of elements' for a given representation remains to be seen in more precise considerations. In the case of finite-dimensional representations on fields, this number is determined by the number of independent real field components (the number of degrees of freedom),

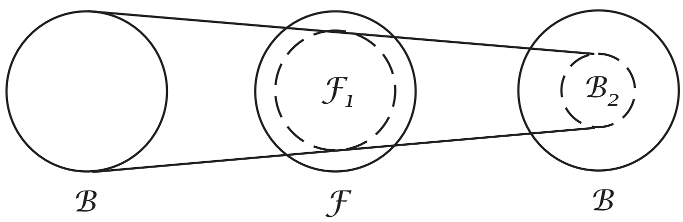

Figure 9.1. Dimensions of the bosonic and fermionic subspace.

as we could have expected. As we shall see later, this number may be calculated with or without using the equations of motion. Therefore, for a large class of representations, in which the mapping P_μ is *onto and 1–1*, we have the following theorem:

> *The number of fermionic and bosonic degrees of freedom in each representation space of the SS algebra is equal.*

There is a couple of cases when the conditions of this theorem are not fulfilled (Sohnius 1985).

The free Wess–Zumino model

The previous theorem on boson–fermion balance tells us that the fields A, B and ψ in the Wess–Zumino model cannot carry a representation of the SS algebra off-shell. Indeed, without use of the equations of motion the Majorana spinor ψ has four real components, while the number of bosonic components is two.

In order to establish the boson–fermion balance off-shell, we introduce two additional scalar fields, F and G. These two bosons should give rise to no on-shell degrees of freedom, since an on-shell balance already exists (two bosons and two fermions). Such fields are called *auxiliary fields*. Since the free action is bilinear in the fields, the contribution of new boson fields has the form $F^2 + G^2$ and, consequently, their mass dimension is *two*.

On dimensional grounds it follows that the SS transformations of the new fields have the form

$$\delta F = -\mathrm{i}\bar{\varepsilon}\gamma \cdot \partial\psi \qquad \delta G = -\mathrm{i}\bar{\varepsilon}\gamma_5\gamma \cdot \partial\psi$$

where we have assumed that F is a scalar and G a pseudoscalar. Similarly, we conclude that F and G cannot occur in δA and δB, but their contribution to $\delta\psi$ is

$$\delta\psi = (\delta\psi)_0 + (aF + b\gamma_5 G)\varepsilon$$

where $(\delta\psi)_0$ denotes the previous result, obtained in (9.2b). The modification is of such a form that on-shell, for $F = G = 0$, we regain the old transformation law (9.2b).

For $a = b = 1$ these new transformations form a realization of the SS algebra, and their complete form reads:

$$\begin{gathered} \delta A = \bar{\varepsilon}\psi \qquad \delta B = \bar{\varepsilon}\gamma_5\psi \\ \delta\psi = -\mathrm{i}\gamma \cdot \partial(A + \gamma_5 B)\varepsilon + (F + \gamma_5 G)\varepsilon \\ \delta F = -\mathrm{i}\bar{\varepsilon}\gamma \cdot \partial\psi \qquad \delta G = -\mathrm{i}\bar{\varepsilon}\gamma_5\gamma \cdot \partial\psi. \end{gathered} \tag{9.10}$$

The invariant action has the form

$$I^0_{\mathrm{WZ}} = \int \mathrm{d}^4x\, [\tfrac{1}{2}\partial_\mu A\partial^\mu A + \tfrac{1}{2}\partial_\mu B\partial^\mu B + \tfrac{1}{2}\mathrm{i}\bar{\psi}\gamma \cdot \partial\psi + \tfrac{1}{2}(F^2 + G^2)]. \tag{9.11}$$

This procedure is typical for the construction of free SS theories. We started with a set of fields (A, B, ψ_α), transforming according to an on-shell representation of the SS algebra, and constructed the action (9.1), which is invariant under these transformations off-shell. In the next step we introduced auxiliary fields F and G which ensure the closure of the algebra off-shell. Finally, we constructed the new invariant action (9.11).

The first step is the standard one since, as we shall see, there is a systematic way to determine on-shell representations of the SS algebra, and construct the related invariant action. The second step, in which we try to find a set of auxiliary fields that lead to an off-shell algebra, is the most difficult, as there is no general rule as to how to do that. The reason lies in the fact that the number of degrees of freedom changes in a complicated way when we turn off the equations of motion, so that it is not easy to control the boson–fermion balance by introducing auxiliary fields. For many SS theories the auxiliary fields are unknown.

The situation becomes even more complicated when we consider interacting theories, which are the subject of our real interest. The form of the interaction is often dictated by some additional principles such as the gauge invariance or general covariance.

Because of great problems which we encounter when trying to introduce auxiliary fields, it is natural to ask the question whether we really need them. Without auxiliary fields the SS algebra closes only on-shell. It is usually believed that this situation may lead to problems; for instance, in the functional integral all field configurations are important, not only those which satisfy the classical equations of motion. However, there are very powerful methods of quantization, such as the Becchi–Rouet–Stora–Tyutin (BRST) approach, which can be applied even when the algebra does not close off-shell. Nevertheless, it is true that a theory with a closed algebra is easier to quantize.

With or without auxiliary fields, an SS theory is characterized by a symmetry in which the generators obey certain algebra. This is of particular importance for the construction of interacting theories. As an illustration of this statement, consider the simple model

$$I = \int \mathrm{d}^4x \, (\tfrac{1}{2}\partial_\mu A \partial^\mu A + \tfrac{1}{2}\mathrm{i}\bar{\psi}\gamma \cdot \partial\psi)$$

which is invariant under the transformations $\delta A = \bar{\varepsilon}\psi$, $\delta\psi = -\mathrm{i}\gamma \cdot \partial A \varepsilon$ that mix fermion and boson fields. In spite of some similarity with the Wess–Zumino model, this symmetry has nothing in common with supersymmetry. Indeed, (a) the algebra of these transformations cannot be generalized to become the symmetry algebra of an interacting theory; and (b) there is no boson–fermion balance on-shell, which is necessary for SS representations. The symmetry structure of an SS theory relies on the existence of the corresponding SS algebra (West 1986).

Chiral symmetry. Consider now the realization of the chiral symmetry (9.7b) on the Wess–Zumino multiplet. Introducing the fields $\mathcal{A} = A + \mathrm{i}B$ and $\mathcal{F} = F - \mathrm{i}G$, the SS transformations (9.10) take the form

$$\delta\mathcal{A} = 2\bar{\varepsilon}\psi_- \qquad \delta\mathcal{F} = -2\mathrm{i}\bar{\varepsilon}\gamma\cdot\partial\psi_-$$
$$\delta\psi_- = -\mathrm{i}\gamma\varepsilon_+\cdot\partial\mathcal{A} + \mathcal{F}\varepsilon_- .$$

We define the chiral transformation of ψ_- by $\delta_R\psi_- = \alpha q\gamma_5\psi_- = \alpha(-\mathrm{i}q)\psi_-$, where q is the corresponding chiral charge. The chiral charge for the other members of the multiplet can be found with the help of (9.7b). By applying this relation to $\mathcal{A}$ we find

$$\delta_R\delta_\mathrm{S}\mathcal{A} - \delta_\mathrm{S}\delta_R\mathcal{A} = \alpha\bar{\varepsilon}\gamma_5 Q(\mathcal{A}) = \alpha(-\mathrm{i})\delta_\mathrm{S}\mathcal{A}.$$

If q' denotes the chiral charge of $\mathcal{A}$, $\delta_R\mathcal{A} = \alpha(-\mathrm{i}q')\mathcal{A}$, we obtain from the previous relation

$$\alpha(-\mathrm{i}q)\delta_\mathrm{S}\mathcal{A} - \alpha(-\mathrm{i}q')\delta_\mathrm{S}\mathcal{A} = \alpha(-\mathrm{i})\delta_\mathrm{S}\mathcal{A}$$

which yields $q' = q - 1$. Analogously we find that the chiral charge of $\mathcal{F}$ is $q'' = q + 1$. Therefore, the chiral transformation acts on the multiplet $(\mathcal{A}, \psi_-, \mathcal{F})$ as

$$(\mathcal{A}, \psi_-, \mathcal{F}) \to (\mathcal{A}', \psi'_-, \mathcal{F}') = \mathrm{e}^{-\mathrm{i}(q-1)\alpha}(\mathcal{A}, \psi_-\mathrm{e}^{-\mathrm{i}\alpha}, \mathcal{F}\mathrm{e}^{-2\mathrm{i}\alpha}). \tag{9.12}$$

The SS transformations of the new fields take the same form, but with the chirally rotated parameters: $\varepsilon' = \mathrm{e}^{-\alpha\gamma_5}\varepsilon = (\mathrm{e}^{\mathrm{i}\alpha}\varepsilon_-, \mathrm{e}^{-\mathrm{i}\alpha}\varepsilon_+)$.

Supersymmetric electrodynamics

We now examine the possibility of constructing a free SS theory based on the massless multiplet (ψ, A_μ), with spin $s = (\frac{1}{2}, 1)$. On-shell, the Majorana spinor carries two degrees of freedom, and the massless boson A_μ three. However, taking into account that a massless four-vector must have an additional freedom of Abelian gauge transformations, the bosonic number of degrees of freedom reduces to two, and the boson–fermion balance is satisfied. The construction of this model leads to an SS generalization of the free electrodynamics.

The local $U(1)$ symmetry on the set of fields (ψ, A_μ) has the form

$$\delta A_\mu = \partial_\mu\lambda \qquad \delta\psi = 0.$$

As a result, the Majorana spinor carries no charge.

On dimensional grounds and Lorentz covariance, an SS transformation has the following general form:

$$\delta A_\mu = \mathrm{i}\bar{\varepsilon}\gamma_\mu\psi$$
$$\delta\psi = (a\sigma^{\mu\nu}F_{\mu\nu} + b\partial\cdot A)\varepsilon$$

where a and b are constants. The group of symmetries consists of the supersymmetry and local $U(1)$ symmetry. The commutator of an SS transformation and a gauge transformation on ψ has the form

$$[\delta(\varepsilon), \delta(\lambda)]\psi = -b(\Box\lambda)\varepsilon.$$

Since the result is neither a supersymmetry nor a gauge transformation, it follows that we must take $b = 0$. Next, the commutator of two SS transformations on A_μ is given by

$$\begin{aligned}[\delta_1, \delta_2]A_\mu &= \mathrm{i}a\bar{\varepsilon}_2\gamma_\mu\sigma^{\lambda\rho}\varepsilon_1 F_{\lambda\rho} - (\varepsilon_1 \leftrightarrow \varepsilon_2) = 2\mathrm{i}a(\bar{\varepsilon}_2\gamma^\nu\varepsilon_1)F_{\mu\nu} \\ &= -2\mathrm{i}a(\bar{\varepsilon}_2\gamma^\nu\varepsilon_1)\partial_\nu A_\mu + \partial_\mu(2\mathrm{i}a\bar{\varepsilon}_2\hat{A}\varepsilon_1).\end{aligned}$$

The first term is a translation if $a = 1$, and the second term represents a gauge transformation with parameter $2\mathrm{i}\bar{\varepsilon}_2\hat{A}\varepsilon_1$, which depends on A. The commutator of two SS transformations on ψ,

$$[\delta_1, \delta_2]\psi = 2\mathrm{i}(\bar{\varepsilon}_1\gamma_\nu\partial_\mu\psi)\sigma^{\mu\nu}\varepsilon_2 - (\varepsilon_1 \leftrightarrow \varepsilon_2)$$

shows that the supersymmetry is closed on ψ on-shell, i.e. for $\gamma \cdot \partial\psi = 0$. Therefore, the final form of the SS and gauge transformations is

$$\begin{aligned}\delta A_\mu &= \mathrm{i}\bar{\varepsilon}\gamma_\mu\psi + \partial_\mu\lambda \\ \delta\psi &= \sigma^{\mu\nu}F_{\mu\nu}\varepsilon.\end{aligned} \tag{9.13a}$$

An action invariant under these transformations has the form

$$I_0 = \int \mathrm{d}^4x\, (-\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} + \tfrac{1}{2}\mathrm{i}\bar{\psi}\gamma\cdot\partial\psi). \tag{9.13b}$$

Indeed,

$$\delta I_0 = \int \mathrm{d}^4x\, [-F^{\mu\nu}\partial_\mu(\mathrm{i}\bar{\varepsilon}\gamma_\nu\psi) - \mathrm{i}F^{\mu\nu}\partial_\rho(\bar{\psi}\gamma_\rho\sigma_{\mu\nu}\varepsilon)] = 0$$

where we have used $\partial_\mu{}^*F^{\mu\nu} = 0$.

The algebra of the SS transformations is given by

$$\begin{aligned}[\delta_1, \delta_2]A_\mu &= -2\mathrm{i}(\bar{\varepsilon}_2\gamma^\nu\varepsilon_1)\partial_\nu A_\mu + \partial_\mu(2\mathrm{i}\bar{\varepsilon}_2\hat{A}\varepsilon_1) \\ [\delta_1, \delta_2]\psi &= -2\mathrm{i}(\bar{\varepsilon}_2\gamma^\mu\varepsilon_1)\partial_\mu\psi + -(\bar{\varepsilon}_2\sigma^{\lambda\rho}\varepsilon_1\sigma_{\lambda\rho} + \tfrac{1}{2}\bar{\varepsilon}_2\gamma^\rho\varepsilon_1\gamma_\rho)F_{\bar{\psi}}.\end{aligned} \tag{9.13c}$$

We now wish to find an off-shell formulation of the model. The counting of bosons and fermions off-shell should be applied only to gauge independent degrees of freedom. Four components of A_μ minus one gauge degree of freedom yields three boson components off-shell, while Majorana spinor ψ has four degrees of freedom. The simplest restoration of the boson–fermion balance is achieved by adding one boson field D. If D is a pseudoscalar, we find the following generalized form of the SS and gauge transformations:

$$\begin{aligned}
\delta A_\mu &= \mathrm{i}\bar{\varepsilon}\gamma_\mu\psi + \partial_\mu\lambda \\
\delta\psi &= (\sigma^{\mu\nu}F_{\mu\nu} - a\gamma_5 D)\varepsilon \\
\delta D &= \mathrm{i}\bar{\varepsilon}\gamma_5\gamma\cdot\partial\psi.
\end{aligned} \tag{9.14a}$$

The closure of the algebra requires $a = 1$ and the invariant action is given by

$$I^0_{\mathrm{ED}} = \int \mathrm{d}^4x\,(-\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} + \tfrac{1}{2}\mathrm{i}\bar{\psi}\gamma\cdot\partial\psi + \tfrac{1}{2}D^2). \tag{9.14b}$$

9.2 Representations of supersymmetry

It is clear from the previous discussion of *free* SS theories without auxiliary fields that their construction is based on the existence of on-shell representations of the SS algebra, that is multiplets consisting of a set of fields, or particle states, on which the SS algebra is realized only on the equations of motion. These representations give a clear picture of the particle content of the theory. However, it is also desirable to have multiplets on which the SS algebra is realized independently of any equations of motion. This will allow us to develop a tensor calculus, which plays an important role in efficient constructions of *interacting* theories and facilitates the building of *quantum* dynamics where the fields must be taken off-shell. These two types of the SS representations are the subject of the present section (Sohnius 1985, West 1986, Srivastava 1986, Müller-Kirsten and Wiedermann 1987).

Invariants of the super-Poincaré algebra

For $m^2 > 0$, the irreducible representations of the Poincaré algebra are specified by the values of the Casimir operators P^2 and W^2, where

$$W_\mu = \tfrac{1}{2}\varepsilon_{\mu\nu\lambda\rho}M^{\nu\lambda}P^\rho.$$

In the case of the super-Poincaré algebra, P^2 is still an invariant operator, and the field components (or states) in a supermultiplet have the same mass. However, W^2 is not an invariant operator and a supermultiplet contains fields (states) of different spins. Indeed,

$$\begin{aligned}
[W_\mu, Q_\alpha] &= -\tfrac{1}{2}(\hat{P}\gamma_\mu\gamma_5 Q)_\alpha + \tfrac{1}{2}(\gamma_5 Q)_\alpha P_\mu \\
[W^2, Q_\alpha] &= -W^\mu(\hat{P}\gamma_\mu\gamma_5 Q)_\alpha + \tfrac{3}{4}P^2 Q_\alpha.
\end{aligned}$$

In order to find an SS generalization of W^2 we first introduce an axial vector $N_\mu = \frac{1}{8}\mathrm{i}\bar{Q}\gamma_\mu\gamma_5 Q$, such that $[N_\mu, Q_\alpha] = -\frac{1}{2}(\hat{P}\gamma_\mu\gamma_5 Q)_\alpha$. Then, we define another vector,

$$X_\mu \equiv W_\mu - N_\mu = W_\mu - \tfrac{1}{8}\mathrm{i}\bar{Q}\gamma_\mu\gamma_5 Q$$

which obeys the relation $[X_\mu, Q_\alpha] = \frac{1}{2}(\gamma_5 Q)_\alpha P_\mu$. It is now easy to see that the tensor

$$C_{\mu\nu} \equiv X_\mu P_\nu - X_\nu P_\mu$$

commutes with all Q_α, and that the square $C_{\mu\nu}C^{\mu\nu} = 2X^2P^2 - 2(X \cdot P)^2$ represents the generalization of W^2 we were looking for, since it commutes with all the generators of the super-Poincaré algebra.

The irreducible representations of the simple ($N = 1$) super-Poincaré algebra are characterized by the values of P^2 and C^2. The meaning of the invariant C^2 is clearly seen in the rest frame where $P^\mu = \mathrm{i}(m, 0, 0, 0)$‡:

$$C^2 = 2m^2\boldsymbol{X}^2 \qquad X^a = -\mathrm{i}mM^a - \tfrac{1}{8}\mathrm{i}\bar{Q}\gamma^a\gamma_5 Q \equiv -\mathrm{i}mY^a.$$

The operator Y^a obeys the commutation rules of the $SO(3)$ group,

$$[Y^a, Y^b] = \varepsilon^{abc}Y^c$$

and represents an SS generalization of the angular momentum M^a. Its eigenvalues define the *superspin* y:

$$\boldsymbol{Y}^2 = -y(y+1) \qquad y = 0, \tfrac{1}{2}, 1, \ldots.$$

Since Y^a does not commute with M^a, an SS multiplet must contain components of different spin. Explicit particle content of these SS multiplets will be discussed later.

Similar considerations for the massless case lead to the concept of superhelicity.

The irreducible representations of the Poincaré group can be found using the Wigner method of induced representations (appendix I). The method consists of finding a representation of a subgroup of the Poincaré group corresponding to some standard momentum $\overset{\circ}{p}$ (the little group), whereupon a representation of the full group, for an arbitrary momentum p, is found by an explicit construction. The physical interpretation of the method is very simple: a representation corresponding to momentum p is obtained by 'boosting' a representation for $\overset{\circ}{p}$.

This procedure can be generalized to the whole super-Poincaré group. In what follows we shall consider only the irreducible super-Poincaré representations corresponding to the standard frame, having in mind that all the other representations are thereby uniquely determined.

We also assume that the central charge is absent. In this case the part of the super-Poincaré algebra that contains Q_α has the following form, in the two-component notation:

$$\begin{gathered}\{Q^m_a, \bar{Q}^n_{\dot{b}}\} = -2\mathrm{i}\delta^{mn}(\sigma^\mu)_{a\dot{b}}P_\mu \\ [M_{\mu\nu}, Q^m_a] = -(\sigma_{\mu\nu})_a{}^b Q^m_b \qquad [M_{\mu\nu}, \bar{Q}^m_{\dot{a}}] = \bar{Q}^m_{\dot{b}}(\bar{\sigma}_{\mu\nu})^{\dot{b}}{}_{\dot{a}}.\end{gathered} \tag{9.15}$$

‡ Since we are using anti-Hermitian generators, the corresponding eigenvalues in unitary representations are imaginary: $P^\mu = \mathrm{i}p^\mu$, $M^{\mu\nu} = \mathrm{i}m^{\mu\nu}$, etc.

Massless states

We begin by considering the representations of the super-Poincaré algebra on one-particle massless states. They are particularly interesting since the most important SS models are generalizations of non-Abelian gauge theories or gravitation, which contain complete massless supermultiplets.

For a massless state, we can always choose the standard momentum in the form $\overset{\circ}{p}{}^{\mu} = \omega(1, 0, 0, 1)$, so that the little group contains the generators Q^m_α, P_μ and T_m, since they all commute with P_μ and leave $\overset{\circ}{p}$ unchanged. Moreover, the momentum $\overset{\circ}{p}$ is invariant under the infinitesimal Lorentz transformations provided the parameters obey the conditions $\omega^{01} = \omega^{31}$, $\omega^{02} = \omega^{32}$, $\omega^{03} = 0$. Hence, the Lorentz generators appear in the combinations

$$E_1 = M_{01} + M_{31} \qquad E_2 = M_{02} + M_{32} \qquad M_{12}.$$

These generators form the Lie algebra of $E(2)$, the group of translations and rotations in the Euclidean plane. In finite-dimensional unitary representations of $E(2)$ the generators E_1 and E_2 are realized trivially (with zero eigenvalues), so that physically relevant representations are determined by the generator M_{12} alone. These representations are one-dimensional and we have

$$M_{12}|\omega,\lambda\rangle = \mathrm{i}\lambda|\omega,\lambda\rangle \qquad W_\mu|\omega,\lambda\rangle = \lambda p_\mu|\omega,\lambda\rangle \tag{9.16}$$

where λ is the helicity of the state $|\omega,\lambda\rangle$, $\lambda = 0, \pm\frac{1}{2}, \pm 1, \ldots$.

Since M_{12} does not commute with Q_a, every supermultiplet contains states with different λ. The action of Q_a on the standard state $|\omega,\lambda\rangle$ leaves the energy and momentum unchanged since $[P_\mu, Q_\alpha] = 0$. The helicity of the state $Q_a|\omega,\lambda\rangle$ is determined from

$$\begin{aligned} M_{12}Q_a|\omega,\lambda\rangle &= (Q_a M_{12} + [M_{12}, Q_a])|\omega,\lambda\rangle \\ &= \mathrm{i}(\lambda + \tfrac{1}{2}\sigma^3)_a{}^b Q_b|\omega,\lambda\rangle. \end{aligned} \tag{9.17}$$

Substituting the explicit form of σ^3, we see that Q_1 raises the helicity for $\frac{1}{2}$, while Q_2 lowers it. A similar calculation for $\bar{Q}_{\dot{a}}$ shows that $\bar{Q}_{\dot{1}}$ lowers and $\bar{Q}_{\dot{2}}$ raises the helicity for $\frac{1}{2}$ (the commutators of Q_a and $\bar{Q}_{\dot{a}}$ with M_{12} have opposite signs).

The SS algebra on the standard states reduces to

$$\{Q^m_a, \bar{Q}^n{}_{\dot{b}}\} = 2\delta^{mn}\omega(1-\sigma^3)_{a\dot{b}} = 4\delta^{mn}\omega\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}_{a\dot{b}}$$

and we find

$$\begin{gathered} \{Q^m_1, \bar{Q}^n{}_{\dot{1}}\} = 0 \qquad \{Q^m_2, \bar{Q}^n{}_{\dot{2}}\} = 4\omega\delta^{mn} \\ \{Q, Q\} = \{\bar{Q}, \bar{Q}\} = 0. \end{gathered} \tag{9.18}$$

The first relation implies

$$\langle\omega,\lambda|Q^m_1(Q^n_1)^* + (Q^n_1)^* Q^m_1|\omega,\lambda\rangle = 0.$$

Assuming that the norm of physical states is positive definite, we have

$$Q_1^m|\omega,\lambda\rangle = \bar{Q}_{\dot{1}}^n|\omega,\lambda\rangle = 0. \tag{9.19}$$

The remaining SS generators define the *Clifford algebra* for N fermionic degrees of freedom,

$$\bar{q}^m \equiv (4\omega)^{-1/2}\bar{Q}^m{}_{\dot{2}} \qquad q^m \equiv (4\omega)^{-1/2}Q_2^m.$$

In each irreducible representation of this algebra there is a Clifford ground state defined by

$$q^m|\Omega\rangle = 0 \qquad |\Omega\rangle \equiv |\omega,\lambda_0\rangle. \tag{9.20}$$

This ground state should not be confused with the concept of a vacuum as the lowest energy state. The existence of the ground state follows from the following simple argument: if $|\beta\rangle$ is not the ground state, since, for instance, $q^1|\beta\rangle \neq 0$, then $q^1|\beta\rangle$ is the ground state, as follows from $q^1(q^1|\beta\rangle) = 0$, and similarly for other modes. All the other states are generated by successive application of the operators $\bar{q}^m$:

$$\bar{q}^n|\Omega\rangle = |\omega,\lambda_0+\tfrac{1}{2},n\rangle$$
$$(\bar{q}^m)(\bar{q}^n)|\Omega\rangle = |\omega,\lambda_0+1,mn\rangle$$

etc. We should note the double role of $\bar{q}$ and q: they are not only the creation and annihilation operators in the Clifford algebra but also the operators that change the helicity for $\pm\frac{1}{2}$. The Clifford states are totally antisymmetric in the internal labels $m, n, \ldots$ carried by the operators $\bar{q}^m$. In the set of states generated over $|\Omega\rangle$, there is the highest state

$$(\bar{q}^1)(\bar{q}^2)\cdots(\bar{q}^N)|\Omega\rangle = |\omega,\lambda_0+\tfrac{1}{2}N,1,2,\ldots,N\rangle$$

on which any further application of $\bar{q}^m$ produces zero. Each application of $\bar{q}$ raises the helicity for $\frac{1}{2}$, and the multiplicities of states are given by the binomial coefficients:

Helicity:	λ_0	$\lambda_0+\frac{1}{2}$	$\cdots$	$\lambda_0+\frac{1}{2}N$
Multiplicity:	$\binom{N}{0}=1$	$\binom{N}{1}=N$	$\cdots$	$\binom{N}{N}=1.$

For a given N it is not difficult to calculate the total number of states, and check the balance of bosons and fermions. Using the notation

$$n = \sum_{k=0}^{N}\binom{N}{k} = 2^N \qquad n_1 = \sum_{k=0}^{[N/2]}\binom{N}{2k} \qquad n_2 = \sum_{k=0}^{[(N-1)/2]}\binom{N}{2k+1}.$$

we find that the binomial expansion of $(1-1)^N$ implies $n_1 - n_2 = 0$, showing that each irreducible representation of supersymmetry contains an equal number of bosonic and fermionic states.

If we include the spatial inversion I_P as the symmetry operation, then to each state with helicity λ we should add a similar state with helicity $-\lambda$ (except for the so-called PCT self-conjugate multiplets, which automatically contain both sets of states).

This is the structure of irreducible representations for the standard momentum $\overset{\circ}{p}$. General representations are obtained in accordance with the Wigner method of induced representations.

The number N of independent supersymmetries is restricted by physical requirements. Spin-$\frac{3}{2}$ fields do not allow renormalizable coupling in quantum field theory. Also, spin-$\frac{5}{2}$ fields are believed to have no consistent interaction with gravity. Therefore:

Renormalizability of quantum theory requires $N_{max} = 4$;
consistency of the gravitational interaction requires $N_{max} = 8$.

For theories in which $N > 4$ (or $N > 8$), particles of spin higher than $\lambda = \frac{3}{2}$ (or $\frac{5}{2}$) will occur in the physical spectrum of the model.

Examples. The simplest $N = 1$ supermultiplet (with I_P symmetry included) has the following structure:

Chiral multiplet $\lambda_0 = -\frac{1}{2}$				
Helicities:	$-\frac{1}{2}$	0	0	$\frac{1}{2}$
States:	1	1	1	1

This corresponds to the massless Wess–Zumino model, which is realized as the theory of massless fields (A, B, ψ) (scalar, pseudoscalar and Majorana spinor) and is called the chiral multiplet.

We display here two more $N = 1$ multiplets:

Gauge multiplet $\lambda_0 = -1$				
Helicities:	-1	$-\frac{1}{2}$	$\frac{1}{2}$	1
States:	1	1	1	1

Supergravity multiplet $\lambda_0 = -2$				
Helicities:	-2	$-\frac{3}{2}$	$\frac{3}{2}$	2
States:	1	1	1	1

The first one is used in SS electrodynamics and is realized on the set of massless fields (A_μ, ψ) (photon and photino); the second describes the particle structure of

$N = 1$ supergravity and corresponds to massless fields $(\varphi_{\mu\nu}, \psi_\mu)$ (graviton and gravitino).

Comment. Limiting our attention to the case $N = 1$, we write the operator N^μ (the SS contribution to the angular momentum) in the two-component notation as

$$N^\mu = \tfrac{1}{8}(\bar{Q}\bar{\sigma}^\mu Q - Q\sigma^\mu \bar{Q}).$$

In the space of states with standard momentum, built over the Clifford ground state $|\Omega\rangle$, it has the form

$$N^\mu = \tfrac{1}{8}(\bar{Q}_{\dot{2}}Q_2 - Q_2\bar{Q}_{\dot{2}})(1, 0, 0, 1) = \tfrac{1}{2}(\bar{q}q - q\bar{q})\overset{\circ}{p}{}^\mu$$

and satisfies the conditions $N^1 = N^2 = 0$, $N^\mu N_\mu = 0$. Therefore, the operator Y^μ is proportional to the standard momentum, which enables us to define the SS generalization of helicity (Srivastava 1986).

Massive states

Going now to the irreducible representations on massive one-particle states, we choose the standard momentum to be the rest-frame momentum: $\overset{\circ}{p}{}^\mu = (m, 0, 0, 0)$. Finite-dimensional unitary representations of the Poincaré group are defined by the representations of the little group—$SO(3)$. The states are labelled by the mass m, the spin j and the spin projection j_3, and obey the conditions

$$\begin{aligned} \boldsymbol{M}^2|j, j_3\rangle &= -j(j+1)|j, j_3\rangle \qquad j = 0, \tfrac{1}{2}, 1, \ldots \\ M^3|j, j_3\rangle &= ij_3|j, j_3\rangle \qquad j_3 = -j, -j+1, \ldots, j. \end{aligned} \tag{9.21}$$

In the absence of central charges, the little group of supersymmetry is generated by $P_\mu, M^a, Q_\alpha, T^m$. The algebra of the supercharges in the rest frame reduces to

$$\{Q^m_a, \bar{Q}^n{}_{\dot{b}}\} = 2m\delta^{mn}(\sigma^0)_{a\dot{b}} = 2m\delta^{mn}\delta_{a\dot{b}}$$

while $\{Q, Q\} = \{\bar{Q}, \bar{Q}\} = 0$. After a rescaling, this algebra becomes the Clifford algebra for $2N$ fermionic degrees of freedom:

$$\{q^m_a, \bar{q}^n{}_{\dot{b}}\} = \delta^{mn}\delta_{a\dot{b}} \qquad \{q, q\} = \{\bar{q}, \bar{q}\} = 0 \tag{9.22}$$

where $(q, \bar{q}) \equiv (2m)^{-1/2}(Q, \bar{Q})$. Unlike the massless case, none of the supercharges is realized trivially, so that the Clifford algebra has $4N$ elements. The irreducible representations of this algebra are found in the usual way. Starting from the ground state $|\Omega\rangle$,

$$q^m_a|\Omega\rangle = 0 \qquad |\Omega\rangle \equiv |j, j_3\rangle \tag{9.23}$$

the representation is carried by the states

$$\bar{q}^n{}_{\dot{b}}|\Omega\rangle, \qquad \bar{q}^m_{\dot{a}}\bar{q}^n{}_{\dot{b}}|\Omega\rangle, \quad \ldots.$$

The maximal spin state is obtained by the application of $2N$ different operators $\bar{q}$. Each state is totally antisymmetric under interchange of the pairs of labels $(m, \dot{a}) \leftrightarrow (n, \dot{b})$.

The particle content of a given irreducible representation is not particularly clear, as we do not see how many states of a given spin are present. More details will be given for the simple case $N = 1$.

Consider, first, the value of the SS spin on the Clifford ground state. In the standard (rest) frame we have

$$Y^a = M^a + \frac{1}{8m}\bar{Q}\gamma^a\gamma_5 Q = M^a - \frac{1}{8m}\mathrm{i}(\bar{Q}\bar{\sigma}^a Q - Q\sigma^a\bar{Q})$$

which implies that the operator Y^a has the same value as the standard spin operator M^a: $y = j$, $y_3 = j_3$.

The effect of the action of Q_a on the ground state is seen from the relation

$$M_{12}Q_a|\Omega\rangle = (Q_a M_{12} + [M_{12}, Q_a])|\Omega\rangle = \mathrm{i}(j_3 + \tfrac{1}{2}\sigma^3)_a{}^b Q_b|\Omega\rangle. \qquad (9.24a)$$

Each q_1 raises the spin projection for $\frac{1}{2}$, and q_2 lowers it, while for $\bar{q}$ the situation is reverse: $\bar{q}_{\dot{1}}$ lowers and $\bar{q}_{\dot{2}}$ raises the spin projection for $\frac{1}{2}$. Thus, the states $\bar{q}_{\dot{1}}|\Omega\rangle$ and $\bar{q}_{\dot{2}}|\Omega\rangle$ have spin projections $j_3 - \frac{1}{2}$ and $j_3 + \frac{1}{2}$, respectively.

Applying two different operators $\bar{q}$ to $|\Omega\rangle$ we obtain the state with spin projection j_3:

$$\begin{aligned} M_{12}\bar{q}_{\dot{1}}\bar{q}_{\dot{2}}|\Omega\rangle &= [\bar{q}_{\dot{1}}M_{12} - \tfrac{1}{2}\mathrm{i}(\bar{\sigma}^3)_{\dot{1}}{}^{\dot{b}}\bar{q}_{\dot{b}}]\bar{q}_{\dot{2}}|\Omega\rangle \\ &= \bar{q}_{\dot{1}}\bar{q}_{\dot{2}}M_{12}|\Omega\rangle = \mathrm{i}j_3\bar{q}_{\dot{1}}\bar{q}_{\dot{2}}|\Omega\rangle. \end{aligned} \qquad (9.24b)$$

Summarizing these results, we see that for each pair of values (m, y) of the Casimir operators the space of an irreducible representation is split into $2y + 1$ subspaces, according to the possible values of the superspin projection $y_3 = -y, -y + 1, \ldots, y$; each of these subspaces contains the states with four projections of the physical spin: $j_3 = y_3, y_3 - \frac{1}{2}, y_3 + \frac{1}{2}, y_3$.

Examples. Here are two simple examples. The smallest representation corresponds to the ground state $|\Omega_1\rangle$ with $y = y_3 = 0$. This state is bosonic, since $j = j_3 = 0$. The representation is four-dimensional and contains the following states:

$y = 0,\ y_3 = 0.$				
States:	$\lvert\Omega_1\rangle$	$\bar{q}_{\dot{1}}\lvert\Omega_1\rangle$	$\bar{q}_{\dot{2}}\lvert\Omega_1\rangle$	$\bar{q}_{\dot{1}}\bar{q}_{\dot{2}}\lvert\Omega_1\rangle$
j_3:	0	$-\frac{1}{2}$	$\frac{1}{2}$	0

From the representation of I_P in the spinor space it follows that two supercharges carry negative parity. Thus, if $|\Omega_1\rangle$ is a scalar then $\bar{q}_{\dot{1}}\bar{q}_{\dot{2}}|\Omega_1\rangle$ is a pseudoscalar, and vice versa. This representation is used in the massive Wess–Zumino model (scalar, pseudoscalar and spin-$\frac{1}{2}$ particle).

The second example is related to the ground state $|\Omega_2\rangle$ with $y = \frac{1}{2}$. This state is fermion, since it has $j = \frac{1}{2}$. In this case there are two subspaces with $y_3 = \pm\frac{1}{2}$, and each of them contains states with four spin projections j_3. The structure of these subspaces is given in the following tables:

$y = \frac{1}{2},\ y_3 = \frac{1}{2}$.				
States:	$\|\Omega_2\rangle$	$\bar{q}_{\dot 1}\|\Omega_2\rangle$	$\bar{q}_{\dot 2}\|\Omega_2\rangle$	$\bar{q}_{\dot 1}\bar{q}_{\dot 2}\|\Omega_2\rangle$
j_3:	$\frac{1}{2}$	0	1	$\frac{1}{2}$
j:	$\frac{1}{2}$	1, 0	1	$\frac{1}{2}$

$y = \frac{1}{2},\ y_3 = -\frac{1}{2}$.				
States:	$\|\Omega_2\rangle$	$\bar{q}_{\dot 1}\|\Omega_2\rangle$	$\bar{q}_{\dot 2}\|\Omega_2\rangle$	$\bar{q}_{\dot 1}\bar{q}_{\dot 2}\|\Omega_2\rangle$
j_3:	$-\frac{1}{2}$	-1	0	$-\frac{1}{2}$
j:	$\frac{1}{2}$	1	1, 0	$\frac{1}{2}$

The fermionic ground state $|\Omega_2\rangle$ has effectively one spinor index: it transforms like $\bar{Q}_{\dot a}|B\rangle$, where $|B\rangle$ is a boson. The states $|\Omega_2\rangle$ and $\bar{q}_{\dot 1}\bar{q}_{\dot 2}|\Omega_2\rangle$ describe two particles of spin $\frac{1}{2}$; linear combinations of the states $\bar{q}_{\dot 1}|\Omega_2\rangle$ and $\bar{q}_{\dot 2}|\Omega_2\rangle$ describe a spin-1 particle and a spin-0 particle (pseudoscalar). The vector state and the pseudoscalar state arise from the product of two spinors:

$$(0, \tfrac{1}{2}) \otimes (0, \tfrac{1}{2}) = (0, 1) + (0, 0).$$

The fermion–boson balance. In the simple examples considered previously, we easily found an equal number of boson and fermion degrees of freedom. Since this is not so clear in more complicated cases, we present here a general proof of this statement, which follows from the basic relation (9.6*a*) of the SS algebra. We first introduce the fermion number operator:

$$N_{\mathrm{F}} = \begin{cases} 0 & \text{on bosonic states} \\ 1 & \text{on fermionic states.} \end{cases}$$

This operator anticommutes with Q_a. For any finite-dimensional representation of the SS algebra in which the trace operation is well defined, we have

$$\mathrm{Tr}[(-1)^{N_{\mathrm{F}}}\{Q_a, \bar{Q}_{\dot b}\}] = \mathrm{Tr}[(-1)^{N_{\mathrm{F}}} Q_a \bar{Q}_{\dot b}] + \mathrm{Tr}[Q_a(-1)^{N_{\mathrm{F}}} \bar{Q}_{\dot b}] = 0$$

where the last equality follows from $(-1)^{N_{\mathrm{F}}} Q_a = Q_a(-1)^{N_{\mathrm{F}}-1}$. On the other hand, using relation (9.15) it follows that the previous result implies

$\mathrm{Tr}[(-1)^{N_\mathrm{F}}] = 0$, i.e.

$$\sum_B \langle B|(-1)^{N_\mathrm{F}}|B\rangle + \sum_F \langle F|(-1)^{N_\mathrm{F}}|F\rangle = n_\mathrm{B} - n_\mathrm{F} = 0. \tag{9.25}$$

Thus, in each SS representation the number of boson and fermion states (degrees of freedom) is found to be equal.

Supermultiplets of fields

Particularly important SS representations are those on fields, as they are used in the construction of interacting field theories. In this subsection we consider their off-shell structure for $N = 1$.

The chiral multiplet. We present here the general method for constructing off-shell SS field multiplets, step by step, on the simple example of the $N = 1$ chiral multiplet (Sohnius 1985).

1. We begin by choosing some complex, scalar $\mathcal{A}(x)$ as the 'ground state' of the representation.
2. We define the fields ψ_α and $\mathcal{F}$ by the following transformation laws:

   $$\delta\mathcal{A} = 2\bar{\varepsilon}\psi \qquad \delta\psi = -a\mathrm{i}\gamma^\mu\varepsilon\partial_\mu\mathcal{A} + \mathcal{F}\varepsilon.$$

   The dimensions of the fields are: $d(\mathcal{A}) = 1$, $d(\psi) = 3/2$, $d(\mathcal{F}) = 2$.
3. Then we impose the constraint that ψ is a chiral spinor: $\psi = \psi_-$. This constraint defines what we know as the chiral multiplet, and means that ψ is a massless field (Weyl spinor). It can be written equivalently as $Q_+(\mathcal{A}) \equiv [Q_+, \mathcal{A}] = 0$.
4. Demanding the closure of the algebra on $\mathcal{A}$ we find the coefficient a in $\delta\psi$: $a = 1$.
5. The transformation law of $\mathcal{F}$ does not introduce new fields, since their dimension would be higher than two; hence,

   $$\delta\mathcal{F} = -2b\mathrm{i}\bar{\varepsilon}\gamma^\mu\partial_\mu\psi.$$

6. Enforcing the closure of the algebra on ψ determines the coefficient b in $\delta\mathcal{F}$: $b = 1$.
7. Finally, we check the algebra on $\mathcal{F}$ and find that it closes.

Thus, we have constructed the multiplet $\phi = (\mathcal{A}, \psi, \mathcal{F})$, known as the chiral multiplet, for which the SS transformation laws are:

$$\begin{aligned} \delta\mathcal{A} &= 2\bar{\varepsilon}\psi_- \\ \delta\psi_- &= -\mathrm{i}\gamma^\mu\varepsilon_+\partial_\mu\mathcal{A} + \mathcal{F}\varepsilon_- \\ \delta\mathcal{F} &= -2\mathrm{i}\bar{\varepsilon}\gamma^\mu\partial_\mu\psi_-. \end{aligned} \tag{9.26a}$$

The component $\mathcal{F}$ is transformed into a divergence, which is always the case with the highest-dimensional component in any multiplet.

The degrees of freedom are counted as unconstrained real field components in the multiplet. There are two complex scalar fields $\mathcal{A}$ and $\mathcal{F}$ with four boson degrees of freedom, and a chiral spinor ψ with four fermion degrees of freedom. The number of $4 + 4$ degrees of freedom is the *smallest* possible number, since each multiplet must contain at least one spinor, and every spinor has at least two complex or four real components. Therefore, the chiral multiplet is *irreducible*.

Instead of the complex fields $\mathcal{A}$, $\mathcal{F}$, we can introduce the corresponding real components,

$$\mathcal{A} = A + \mathrm{i}B \qquad \mathcal{F} = F - \mathrm{i}G$$

while the chiral spinor ψ_- can be replaced by the related Majorana spinor ψ. The transformations laws for the chiral multiplet $\phi = (A, B, \psi, F, G)$ follow from the complex form (9.26*a*):

$$\begin{gathered}\delta A = \bar{\varepsilon}\psi \qquad \delta B = \bar{\varepsilon}\gamma_5\psi \\ \delta\psi = -\mathrm{i}\gamma\cdot\partial(A + \gamma_5 B)\varepsilon + (F + \gamma_5 G)\varepsilon \\ \delta F = -\mathrm{i}\bar{\varepsilon}\gamma\cdot\partial\psi \equiv -\bar{\varepsilon}F_{\bar{\psi}} \qquad \delta G = -\mathrm{i}\bar{\varepsilon}\gamma_5\gamma\cdot\partial\psi \equiv -\bar{\varepsilon}\gamma_5 F_{\bar{\psi}}.\end{gathered} \tag{9.26b}$$

The essential property of the multiplet, namely that ψ is a Majorana spinor, is reflected in the fact that the spinor in δB is equal to γ_5 times the spinor in δA.

In a similar way, by replacing the constraint $\psi = \psi_-$ with $\psi = \psi_+$ in step 3, we can obtain an antichiral multiplet $\bar{\phi}$. This method can also be applied to more complicated multiplets, as we shall see.

The general multiplet. If we repeat the previous construction starting, again, from some complex field $C(x)$, but without imposing the chirality constraint, the result is a larger multiplet,

$$V = (C, \chi_\alpha, M, N, A_\mu, \psi_\alpha, D) \tag{9.27a}$$

with the following transformation rules:

$$\begin{gathered}\delta C = \bar{\varepsilon}\gamma_5\chi \\ \delta\chi = (M + \gamma_5 N)\varepsilon - \mathrm{i}\gamma^\mu(A_\mu + \gamma_5\partial_\mu C)\varepsilon \\ \delta M = \bar{\varepsilon}(\psi - \mathrm{i}\gamma\cdot\partial\chi) \\ \delta N = \bar{\varepsilon}\gamma_5(\psi - \mathrm{i}\gamma\cdot\partial\chi) \\ \delta A_\mu = \bar{\varepsilon}(\mathrm{i}\gamma_\mu\psi + \partial_\mu\chi) \\ \delta\psi = (\sigma^{\mu\nu}F_{\mu\nu} - \gamma_5 D)\varepsilon \\ \delta D = \mathrm{i}\bar{\varepsilon}\gamma_5\gamma\cdot\partial\psi\end{gathered} \tag{9.27b}$$

where ε is a Majorana spinor. The scalar M, the pseudoscalars C, N and D, and the vector A_μ are complex fields, and χ and ψ are Dirac spinors. We note that the component D with the highest dimension transforms into a divergence. This multiplet has $16 + 16$ degrees of freedom, and is called the general multiplet.

The general multiplet is a reducible representation of supersymmetry. Imposing a reality condition

$$V = V^+ \tag{9.27c}$$

(all components are real or Majorana), we obtain the *real* general multiplet, with $8 + 8$ degrees of freedom.

The most general SS multiplets can be obtained from a general multiplet by attaching an additional Lorentz index to each of its components.

Reducibility and submultiplets. A multiplet is reducible if it contains a subset of fields which is closed under SS transformations. We assume that none of these fields satisfy their equations of motion, although some of them may obey some other differential conditions.

The real general multiplet is reducible. Indeed, the fields ψ, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ and D transform among themselves, and form a submultiplet—the curl multiplet dV:

$$dV = (\psi, F_{\mu\nu}, D) \tag{9.28a}$$

which has $4 + 4$ real components. The transformation laws

$$\begin{aligned} \delta F_{\mu\nu} &= \mathrm{i}\bar{\varepsilon}(\gamma_\nu \partial_\mu - \gamma_\mu \partial_\nu)\psi \\ \delta\psi &= (\sigma^{\mu\nu} F_{\mu\nu} - \gamma_5 D)\varepsilon \\ \delta D &= \mathrm{i}\bar{\varepsilon}\gamma_5 \gamma \cdot \partial\psi \end{aligned} \tag{9.28b}$$

represent the algebra provided $F_{\lambda\rho}$ is subject to the condition $\partial_\mu {}^*F^{\mu\nu} = 0$.

We can also write this multiplet in the form $dV = (\psi, A_\mu, D)$, where A_μ realizes the local $U(1)$ symmetry: $A_\mu \to A_\mu + \partial_\mu \lambda$.

If we constrain V by demanding $dV = 0$, the remaining components of V form the chiral multiplet,

$$\phi_1 = (A, C, \chi_\alpha, M, N) \tag{9.29}$$

where the scalar A is defined by the solution $A_\mu = \partial_\mu A$ of the condition $F_{\mu\nu} = 0$.

Tensor calculus and invariants

For any symmetry group there are rules of how to combine two multiplets to obtain a third one. These rules are an analogue of combining two vectors into a tensor or a scalar. They can be used, among other things, to build invariant objects, which is of particular importance for the systematic construction of actions. A similar analysis also exists for the super-Poincaré group, and is called the SS tensor calculus.

A. We can combine two chiral multiplets, $\phi_1 = (\mathcal{A}_1, \psi_1, \mathcal{F}_1)$ and $\phi_2 = (\mathcal{A}_2, \psi_2, \mathcal{F}_2)$, and form a third multiplet which is based on the lowest component

$$\mathcal{A}_3 \equiv \mathcal{A}_1\mathcal{A}_2 = (A_1A_2 - B_1B_2) + \mathrm{i}(A_1B_2 + A_2B_1).$$

The multiplet ϕ_3, obtained from $\mathcal{A}_3$ and the known transformations of ϕ_1 and ϕ_2, is also chiral. Explicit calculations show that the components of the product multiplet

$$\phi_3 = \phi_1 \cdot \phi_2 \tag{9.30a}$$

are expressed in terms of the components of ϕ_1 and ϕ_2 in the following way:

$$\begin{aligned}
A_3 &= A_1A_2 - B_1B_2 \\
B_3 &= A_1B_2 + A_2B_1 \\
\psi_3 &= (A_1 - \gamma_5 B_1)\psi_2 + (A_2 - \gamma_5 B_2)\psi_1 \\
F_3 &= A_1F_2 + B_1G_2 + A_2F_1 + B_2G_1 - \bar{\psi}_1\psi_2 \\
G_3 &= A_1G_2 - B_1F_2 + A_2G_1 - B_2F_1 + \bar{\psi}_1\gamma_5\psi_2.
\end{aligned} \tag{9.30b}$$

This product is symmetric and associative,

$$\phi_1 \cdot \phi_2 = \phi_2 \cdot \phi_1 \qquad (\phi_1 \cdot \phi_2) \cdot \phi_3 = \phi_1 \cdot (\phi_2 \cdot \phi_3)$$

hence, the product of any number of chiral multiplets is well defined. Multiplying ϕ with the two constant chiral multiplets $1_+ = (1, 0, 0, 0, 0)$ and $1_- = (0, 1, 0, 0, 0)$, we obtain

$$\phi \cdot 1_+ = \phi \qquad \phi \cdot 1_- = (-B, A, -\gamma_5\psi, G, -F).$$

A similar product can be defined for two real general multiplets, $V = V_1 \cdot V_2$, starting from the lowest state $C = C_1C_2$.

B. We can now combine two chiral multiplets in another symmetric product,

$$V = \phi_1 \times \phi_2 \tag{9.31a}$$

which starts from $A_1A_2 + B_1B_2$ as the lowest component and represents the real general multiplet:

$$\begin{aligned}
C &= A_1A_2 + B_1B_2 \\
\chi &= (B_1 - \gamma_5 A_1)\psi_2 + (B_2 - \gamma_5 A_2)\psi_1 \\
M &= B_1F_2 + A_1G_2 + B_2F_1 + A_2G_1 \\
N &= B_1G_2 - A_1F_2 + B_2G_1 - A_2F_1 \\
A_\mu &= B_1\overleftrightarrow{\partial}_\mu A_2 + B_2\overleftrightarrow{\partial}_\mu A_1 + \mathrm{i}\bar{\psi}_1\gamma_\mu\gamma_5\psi_2 \\
\psi &= [G_1 + \gamma_5 F_1 + \mathrm{i}\gamma \cdot \partial(B_1 - \gamma_5 A_1)]\psi_2 + (1 \leftrightarrow 2) \\
D &= -2(F_1F_2 + G_1G_2 + \partial A_1 \cdot \partial A_2 + \partial B_1 \cdot \partial B_2) - \mathrm{i}\bar{\psi}_1\gamma \cdot \overleftrightarrow{\partial}\psi_2.
\end{aligned} \tag{9.31b}$$

C. The particular combination

$$\phi_1 \wedge \phi_2 = (\phi_1 \cdot 1_-) \times \phi_2$$

antisymmetric in ϕ_1 and ϕ_2, has the lowest component $A_1B_2 - A_2B_1$, and represents a real general multiplet.

D. The chiral multiplet $\phi = (A, B, \psi, F, G)$ can be used to form another chiral multiplet, based on F as the lowest component. It is called the *kinetic multiplet* and has the form

$$T\phi = (F, G, -\mathrm{i}\gamma \cdot \partial\psi, -\Box A, -\Box B). \tag{9.32a}$$

Doing this we again find the relation

$$TT\phi = -\Box\phi \tag{9.32b}$$

which explains the name kinetic. The operator T is an SS generalization of the Dirac operator $\mathrm{i}\gamma \cdot \partial$.

Invariants. Having found the structure of various SS multiplets and the rules for their multiplication, it remains to clarify the structure of invariant quantities, in particular invariant actions.

The highest component F of the chiral multiplet ϕ varies into a divergence. This is not an accident, but follows from the fact that F is a component of the highest dimension, so that, on dimensional grounds, its SS variation must be of the form ∂(other fields). Therefore, the integral

$$I_1 = \int \mathrm{d}^4x\, [\phi]_F \tag{9.33a}$$

is invariant under SS transformations. If the chiral multiplet ϕ is a product of other multiplets, for instance $\phi = \phi_1 \cdot \phi_2$, then I_1 contains products of the usual fields, and can be taken as an action integral.

In a similar way we conclude that the highest component D of a real general multiplet V may be also used to define an action:

$$I_2 = \int \mathrm{d}^4x\, [V]_D. \tag{9.33b}$$

The interacting Wess–Zumino model

The most general SS action for a single chiral multiplet, which contains no more than two derivatives and no coupling constants of negative dimensions, has the form

$$I_{\mathrm{WZ}} = \int \mathrm{d}^4x\, (\tfrac{1}{2}\phi \cdot T\phi + \tfrac{1}{2}m\phi \cdot \phi - \tfrac{1}{3}g\phi \cdot \phi \cdot \phi)_F \tag{9.34a}$$

and defines the massive Wess–Zumino model with interactions. The term $[\phi \times \phi]_D$ does not give anything new, since it differs from $-2[\phi \cdot T\phi]_\mathrm{F}$ by a divergence. Using the rules of tensor calculus, we can find the component form of the Lagrangian:

$$\begin{gathered}
\mathcal{L}_{\mathrm{WZ}} = \mathcal{L}_0 + \mathcal{L}_m + \mathcal{L}_g \\
\mathcal{L}_0 \equiv \tfrac{1}{2}\partial_\mu A \partial^\mu A + \tfrac{1}{2}\partial_\mu B \partial^\mu B + \tfrac{1}{2}\mathrm{i}\bar{\psi}\gamma\cdot\partial\psi + \tfrac{1}{2}(F^2 + G^2) \\
\mathcal{L}_m \equiv m(AF + BG - \tfrac{1}{2}\bar{\psi}\psi) \\
\mathcal{L}_g \equiv -g[(A^2 - B^2)F + 2ABG - \bar{\psi}(A - \gamma_5 B)\psi].
\end{gathered} \tag{9.34b}$$

This Lagrangian is invariant under SS transformations (9.26b). The related equations of motion can be written most compactly in the super-covariant form:

$$T\phi = -m\phi + g\phi\cdot\phi.$$

The field equations for F and G do not describe propagation in spacetime, but are purely *algebraic*,

$$F = -mA + g(A^2 - B^2) \qquad G = -mB + 2gAB$$

which can be expressed equivalently by saying that F and G are auxiliary fields. The elimination of these fields leads to the Lagrangian

$$\begin{aligned}
\mathcal{L}'_{\mathrm{WZ}} = {} & \tfrac{1}{2}(\partial_\mu A \partial^\mu A - m^2A^2) + \tfrac{1}{2}(\partial_\mu B \partial^\mu B - m^2B^2) \\
& + \tfrac{1}{2}\bar{\psi}(\mathrm{i}\gamma\cdot\partial - m)\psi + mgA(A^2 + B^2) \\
& - \tfrac{1}{2}g^2(A^2 + B^2)^2 + g\bar{\psi}(A - \gamma_5 B)\psi
\end{aligned} \tag{9.35a}$$

which represents a generalization of expression (9.1), containing the mass and interaction terms. We should observe that the masses of all the fields are equal, and that all interactions (A^3, AB^2, A^4, B^4, A^2B^2, $A\bar{\psi}\psi$ and $B\bar{\psi}\gamma_5\psi$) are determined by two parameters: the mass m and the coupling constant g. The Lagrangian (9.35a) is invariant under the SS transformations:

$$\begin{gathered}
\delta A = \bar{\varepsilon}\psi \qquad \delta B = \bar{\varepsilon}\gamma_5\psi \\
\delta\psi = [-(\mathrm{i}\gamma\cdot\partial + m) + g(A + \gamma_5 B)](A + \gamma_5 B)\varepsilon
\end{gathered} \tag{9.35b}$$

obtained from (9.26b) by eliminating F and G.

The role of auxiliary fields is to establish the boson–fermion balance off-shell. Transformations (9.26b) obey the SS algebra without any additional conditions and their form does not depend on the dynamics, i.e. coupling constants, which is not the case with (9.35b). It is interesting to observe that the set of field equations is also an SS multiplet. The elimination of auxiliary fields, which is realized by enforcing their field equations in both the action and the transformation laws, is not an SS invariant procedure. This is why the resulting SS algebra closes only on-shell.

9.3 Supergravity

When the supersymmetry is localized, the SS algebra contains a local translation, which is an indication that gravity has to be an intrinsic part of a locally SS theory. Such a theory, called supergravity, represents a natural framework for the unified treatment of gravity and other basic interactions. Its quantum properties are much better than those of Einstein's gravity. The advantage of supergravity with respect to ordinary SS theories is that it allows a more natural introduction of spontaneously broken modes, which are necessary for phenomenological applications (see the end of this section).

In this section, we want to discuss the SS theory of gravity or the gauge SS theory, following the analogy with SS electrodynamics. We start with the linearized theory, in order to be able to recognize the essential features of the full supergravity in the simpler context. In the linearized approximation, the graviton is described as a massless spin-2 field. In an irreducible representation of supersymmetry, the graviton may be joined with either a spin-$\frac{3}{2}$ or a spin-$\frac{5}{2}$ fermion. If we exclude the spin-$\frac{5}{2}$ field as it seems to have no consistent gravitational interaction, we are left with the supergravity multiplet $(2, \frac{3}{2})$. The related particle content is described by a symmetric tensor $\varphi_{\mu\nu}$ (the graviton) and a Majorana vector spinor $\psi_{\alpha\mu}$ (the gravitino), subject to their field equations. After giving a short review of the Rarita–Schwinger theory of gravitino, we examine the construction of first the linearized and then full supergravity and clarify the role and structure of the auxiliary fields (Schwinger 1970, van Nieuwenhuizen 1981a, West 1986, Srivastava 1986).

The Rarita–Schwinger field

The free-field Lagrangian for a massless gravitino in M_4 can be constructed starting from the requirement of invariance under gauge transformations

$$\psi_\mu \to \psi'_\mu = \psi + \partial_\mu \theta \tag{9.36}$$

where the parameter θ is a Majorana spinor. The most general Lagrangian with first field derivatives has the form

$$\mathcal{L} = \mathrm{i}\bar{\psi}_\mu(a\gamma^\mu\partial^\nu + b\partial^\mu\gamma^\nu + c\eta^{\mu\nu}\gamma\cdot\partial + d\gamma^\mu\gamma^\nu\gamma\cdot\partial)\psi_\nu.$$

Varying $\mathcal{L}$ with respect to ψ_μ yields the equations of motion

$$a\gamma^\mu\partial\cdot\psi + b\partial^\mu\gamma\cdot\psi + c\gamma\cdot\partial\psi^\mu + d\gamma^\mu\gamma^\nu\gamma\cdot\partial\psi_\nu = 0.$$

The requirement of gauge invariance for these equations implies $a + d = 0$, $b + c = 0$. The invariance of the action is achieved if $a + b = 0$. Choosing $a = -\frac{1}{2}$ as the normalization condition, and using the identity

$$\gamma^\mu\eta^{\nu\lambda} - \gamma^\nu\eta^{\lambda\mu} + \gamma^\lambda\eta^{\mu\nu} - \gamma^\mu\gamma^\nu\gamma^\lambda = \varepsilon^{\mu\nu\lambda\rho}\gamma_5\gamma_\rho$$

we end up with the Rarita–Schwinger Lagrangian:

$$\mathcal{L} = \tfrac{1}{2}\mathrm{i}\varepsilon^{\mu\nu\rho\lambda}\bar{\psi}_\mu\gamma_5\gamma_\nu\partial_\rho\psi_\lambda \equiv \mathcal{L}_{\mathrm{RS}}. \tag{9.37}$$

We display here several equivalent forms of the gravitino field equations:

$$\begin{aligned} f^\mu &\equiv \mathrm{i}\varepsilon^{\mu\nu\rho\lambda}\gamma_5\gamma_\nu\partial_\rho\psi_\lambda = 0 \\ -\mathrm{i}[\gamma\cdot\partial(\gamma\cdot\psi) - \partial\cdot\psi] &= \tfrac{1}{2}\gamma\cdot f \\ \mathrm{i}\gamma^j\psi_{j\mu} &= -(f_\mu - \tfrac{1}{2}\gamma_\mu\gamma\cdot f) \\ \mathrm{i}(\gamma_i\psi_{j\mu} + \gamma_\mu\psi_{ij} + \gamma_j\psi_{\mu i}) &= \varepsilon_{\rho ij\mu}\gamma_5 f^\rho \end{aligned} \tag{9.38}$$

where $\psi_{\mu\nu} \equiv \partial_\mu\psi_\nu - \partial_\nu\psi_\mu$.

Mass and helicity. Let us choose the gauge condition

$$\Omega \equiv \gamma\cdot\psi = 0 \tag{9.39a}$$

which is locally admissible since $\delta(\gamma\cdot\psi) = \gamma\cdot\partial\theta$ can be solved for θ. Then, the field equations imply

$$\mathrm{i}\gamma\cdot\partial\psi^\mu = 0 \qquad \mathrm{i}\partial\cdot\psi = 0 \tag{9.39b}$$

showing that the gravitino field is *massless*, as we expected.

We now show that ψ_μ describes two physical modes with helicities $\lambda = \pm\frac{3}{2}$. Consider the plane wave $\psi^\mu = u^\mu(k)\mathrm{e}^{-\mathrm{i}k\cdot x}$, moving along the z-axis, $k = (k^0, 0, 0, k^3)$. Using the complete set of vectors $(\epsilon_{(-)}, \epsilon_{(+)}, k, \bar{k})$, where $\epsilon_\pm = (0, 1, \pm\mathrm{i}, 0, 0)/\sqrt{2}$ and $\bar{k} = (k^0, 0, 0, -k^3)$, we expand $u^\mu(k)$ as

$$u^\mu = \epsilon^\mu_{(-)}u_- + \epsilon^\mu_{(+)}u_+ + k^\mu u_0 + \bar{k}^\mu u_3.$$

Here, u_-, u_+, u_0 and u_3 are Majorana spinors, and $k^2 = 0$. The first condition in (9.39b) means that all spinors are massless:

$$\hat{k}u_- = \hat{k}u_+ = \hat{k}u_0 = \hat{k}u_3 = 0$$

while the second one implies $(k\cdot\bar{k})u_3 = 0$, i.e. $u_3 = 0$. The gauge condition (9.39a) allows the residual gauge symmetry $\psi_\mu \to \psi_\mu - ku$, which may be used to eliminate u_0. Thus, after fixing the gauge we have

$$u^\mu = \epsilon^\mu_{(-)}u_- + \epsilon^\mu_{(+)}u_+. \tag{9.40}$$

Consider now the meaning of the gauge condition for the remaining degrees of freedom. Multiplying this condition with $\hat{\epsilon}_{(-)}$, and using $\hat{\epsilon}_{(-)}\hat{\epsilon}_{(-)} = 0$, we obtain

$$\hat{\epsilon}_{(-)}\hat{\epsilon}_{(+)}u_+ = -(1 - 2\Sigma^3)u_+ = 0$$

where $\Sigma^3 = \frac{1}{2}\mathrm{i}\gamma^1\gamma^2$ is the spin projection operator. This relation means that u_+ has helicity $+\frac{1}{2}$. Similarly, helicity of u_- is $-\frac{1}{2}$. Since $\epsilon_{(\pm)}$ carries helicity ± 1, we conclude that u^μ describes a helicity $\pm\frac{3}{2}$ field.

The propagator. The field equations for ψ_μ in interaction with an external source are

$$K^{\mu\nu}\psi_\nu = J^\mu$$
$$K^{\mu\nu} \equiv \mathrm{i}[\gamma^\mu\partial^\nu + \partial^\mu\gamma^\nu - \gamma\cdot\partial\eta^{\mu\nu} - \gamma^\mu\gamma\cdot\partial\gamma^\nu]. \tag{9.41a}$$

Gauge invariance implies that the operator $K^{\mu\nu}$ is singular, $\partial_\mu K^{\mu\nu} = 0$, hence the consistency requires $\partial_\mu J^\mu = 0$. In order to solve these equations it is convenient to go to the momentum space ($\mathrm{i}\partial \to k$):

$$(-\gamma^\mu k^\nu - k^\mu\gamma^\nu + \hat{k}\eta^{\mu\nu} + \gamma^\mu\hat{k}\gamma^\nu)\psi_\nu = -J^\mu. \tag{9.41b}$$

It follows from this that

$$\hat{k}\psi^\mu - k^\mu\gamma\cdot\psi = -(J^\mu - \tfrac{1}{2}\gamma^\mu\gamma\cdot J)$$

whereupon the application of the gauge condition yields

$$\psi^\mu = G^{\mu\nu}J_\nu \qquad G^{\mu\nu} \equiv -\hat{k}(\eta^{\mu\nu} - \tfrac{1}{2}\gamma^\mu\gamma^\nu)D \tag{9.42a}$$

where $D = 1/k^2$. The previous result for G can be written equivalently as

$$G^{\mu\nu} = -\hat{k}\tfrac{1}{2}\gamma^\nu\gamma^\mu D = \tfrac{1}{2}\gamma^\nu\hat{k}\gamma^\mu D \tag{9.42b}$$

where, in the last step, we have ignored the term proportional to k^ν, which yields zero when acting on the conserved current.

This procedure for finding the propagator is clarified by introducing

$$P_{\mu\nu} = \Pi_{\mu\nu} - \tfrac{1}{3}\bar{\gamma}_\mu\bar{\gamma}_\nu \qquad L_{\mu\nu} = \tfrac{1}{3}\bar{\gamma}_\mu\bar{\gamma}_\nu \qquad \Lambda_{\mu\nu} = k_\mu k_\nu/k^2 \tag{9.43}$$

where $\Pi_{\mu\nu} \equiv \eta_{\mu\nu} - k_\mu k_\nu/k^2$, $\bar{\gamma}_\mu \equiv \gamma_\mu - k_\mu\hat{k}/k^2$. These objects are projectors in the set of solutions of the field equations for ψ_μ. Equation (9.41b) may be rewritten in terms of the projectors as

$$(P_{\mu\nu} - 2L_{\mu\nu})\hat{k}\psi^\nu = -(P_{\mu\nu} + L_{\mu\nu} + \Lambda_{\mu\nu})J^\nu$$

which implies $\Lambda_{\mu\nu}J^\nu = 0$, i.e. $k\cdot J = 0$. The projections of this equation with P and L take the form

$$(P_{\mu\nu} + L_{\mu\nu})\hat{k}\psi^\nu = -(P_{\mu\nu} - \tfrac{1}{2}L_{\mu\nu})J^\nu$$
$$\hat{k}\psi_\mu - \hat{k}\Lambda_{\mu\nu}\psi^\nu = -(\eta_{\mu\nu} - \tfrac{1}{2}\bar{\gamma}_\mu\gamma_\nu)J^\nu.$$

Now, multiplying the last equation with $\bar{\gamma}^\mu$ we find $-\hat{k}(\gamma\cdot\psi) + k\cdot\psi = \gamma\cdot J/2$. Then, after fixing the gauge we easily obtain the solution (9.42a).

Energy. The interaction energy between two currents is given by

$$E = \bar{J}^\mu G_{\mu\nu} J^\nu. \tag{9.44}$$

We shall show that this energy is positive, thus justifying the choice of sign for the Lagrangian. The residue of the expression for E may be written as

$$R = \bar{J}_\mu \eta^{\lambda\mu} \gamma_\sigma (\tfrac{1}{2}\hat{k}) \gamma_\lambda \eta^{\nu\sigma} J_\nu.$$

Consider now the following identities:

$$\eta^{\mu\nu} = \epsilon^\mu_{(-)} \epsilon^{\nu *}_{(-)} + \epsilon^\nu_{(+)} \epsilon^{\mu *}_{(+)} + (k^\mu \bar{k}^\nu + k^\nu \bar{k}^\mu)/(k \cdot \bar{k})$$
$$\hat{k} = u_- \bar{u}_- + u_+ \bar{u}_+.$$

The first identity represents the completeness relation for the set of vectors $(\epsilon_{(-)}, \epsilon_{(+)}, k, \bar{k})$, valid on-shell where $k^2 = 0$, and the second one holds in the space of massless spinors, normalized according to $u_-^+ u_- = u_+^+ u_+ = 2k^0$. The terms in R proportional to k^μ can be ignored, since $k \cdot J = 0$ and $\hat{k} u_\mp = 0$. Using, further, the relations

$$\hat{\epsilon}_{(+)} u_+ = \hat{\epsilon}_{(-)} u_- = 0$$

which tell us that the helicity of u_+ (u_-) cannot be raised (lowered), we find that the residue is given by

$$R = \tfrac{1}{2} |\bar{u}_- \hat{\epsilon}_{(-)}\, \epsilon_{(-)} \cdot J|^2 + \tfrac{1}{2} |\bar{u}_+ \hat{\epsilon}_{(+)}\, \epsilon_{(+)} \cdot J|^2 \tag{9.45}$$

where $|A|^2 \equiv A^+ A$, which proves its positivity.

Linearized theory

The field equations for a graviton and a gravitino are invariant under the gauge transformations

$$\delta_\xi \varphi_{\mu\nu} = \partial_\mu \xi_\nu + \partial_\nu \xi_\mu \qquad \delta_\theta \psi_\mu = \partial_\mu \theta. \tag{9.46}$$

The linearized on-shell simple supergravity can be constructed in analogy with SS electrodynamics, so that its full symmetry is supersymmetry combined with the gauge symmetry (9.46).

SS transformations. On dimensional grounds the general SS transformations are

$$\delta_\varepsilon \varphi_{\mu\nu} = \mathrm{i}\tfrac{1}{2}(\bar{\varepsilon}\gamma_\mu \psi_\nu + \bar{\varepsilon}\gamma_\nu \psi_\mu) + c_1 \eta_{\mu\nu} \varepsilon\gamma \cdot \psi$$
$$\delta_\varepsilon \psi_\mu = 2c_2\, \partial_j \varphi_{i\mu} \sigma^{ij} \varepsilon + c_3 \partial^\rho \varphi_{\rho\mu} \varepsilon$$

where ε is a constant Majorana spinor, and the constants c_i will be determined by demanding that the algebra of the SS and gauge transformations closes on-shell.

For the commutator of a local θ symmetry and a global supersymmetry on $\varphi_{\mu\nu}$ we find

$$[\delta_\theta, \delta_\varepsilon]\varphi_{\mu\nu} = \mathrm{i}\tfrac{1}{2}(\bar{\varepsilon}\gamma_\mu\partial_\nu\theta + \bar{\varepsilon}\gamma_\nu\partial_\mu\theta) + c_1\eta_{\mu\nu}\varepsilon\gamma\cdot\partial\theta.$$

The result is a gauge transformation with the parameter $\xi_\mu = \mathrm{i}\bar{\varepsilon}\gamma_\mu\theta/2$, provided $c_1 = 0$. The commutators $[\delta_\theta, \delta_\varepsilon]\psi_\mu$ and $[\delta_\xi, \delta_\varepsilon]\varphi_{\mu\nu}$ automatically vanish, while

$$[\delta_\xi, \delta_\varepsilon]\psi_\mu = 2c_2\partial_\mu(\partial_j\xi_i)\sigma^{ij}\varepsilon + c_3\partial^\rho(\partial_\rho\xi_\mu + \partial_\mu\xi_\rho)\varepsilon$$

is a gauge transformation with parameter $\theta = 2c_2\partial_j\xi_i\sigma^{ij}\varepsilon$, provided $c_3 = 0$. We now test the commutator of two supersymmetries on $\varphi_{\mu\nu}$:

$$\begin{aligned}[\delta_{\varepsilon_1}, \delta_{\varepsilon_2}]\varphi_{\mu\nu} &= \mathrm{i}c_2[\bar{\varepsilon}_2\gamma_\mu\sigma^{ij}\varepsilon_1\partial_j\varphi_{i\nu} + (\mu\leftrightarrow\nu)] - (\varepsilon_1\leftrightarrow\varepsilon_2)\\ &= \mathrm{i}c_2[\bar{\varepsilon}_2\gamma^j\varepsilon_1\partial_j\varphi_{\mu\nu} - \partial_\mu(\bar{\varepsilon}_2\gamma^i\varepsilon_1\varphi_{i\nu})] + (\mu\leftrightarrow\nu).\end{aligned} \tag{9.47}$$

The result is a gauge transformation with parameter $\xi_\nu = -\mathrm{i}c_2\bar{\varepsilon}_2\gamma^j\varepsilon_1\varphi_{j\nu}$, and a translation. The translation has the correct form if $c_2 = -1$.

Thus, the final form of the SS transformations reads:

$$\begin{aligned}\delta\varphi_{\mu\nu} &= \mathrm{i}\tfrac{1}{2}(\bar{\varepsilon}\gamma_\mu\psi_\nu + \bar{\varepsilon}\gamma_\nu\psi_\mu)\\ \delta\psi_\mu &= -2\partial_j\varphi_{i\mu}\sigma^{ij}\varepsilon.\end{aligned} \tag{9.48}$$

The algebra is closed only by taking into account both the global SS transformations and gauge transformations, in analogy with the electrodynamic case.

Finally, it remains to verify the algebra of the transformations on ψ_μ:

$$\begin{aligned}[\delta_{\varepsilon_1}, \delta_{\varepsilon_2}]\psi_\mu &= \mathrm{i}\partial_i(\bar{\varepsilon}_1\gamma_j\psi_\mu + \bar{\varepsilon}_1\gamma_\mu\psi_j)\sigma^{ij}\varepsilon_2 - (\varepsilon_1\leftrightarrow\varepsilon_2)\\ &= -\tfrac{1}{4}(\bar{\varepsilon}_1\Gamma_A\varepsilon_2)\sigma^{ij}\Gamma^A\mathrm{i}\partial_i(\gamma_j\psi_\mu + \gamma_\mu\psi_j) - (\varepsilon_1\leftrightarrow\varepsilon_2)\\ &= -\tfrac{1}{2}\mathrm{i}(\bar{\varepsilon}_1\Gamma_a\varepsilon_2)\sigma^{ij}\Gamma^a(\gamma_j\psi_{i\mu} + \tfrac{1}{2}\gamma_\mu\psi_{ij})\\ &\quad - \tfrac{1}{2}\mathrm{i}\partial_\mu[(\bar{\varepsilon}_1\Gamma_a\varepsilon_2)\sigma^{ij}\Gamma^a\gamma_j\psi_i].\end{aligned}$$

Here we have used the Fierz identity in which, due to antisymmetry with respect to $(\varepsilon_1\leftrightarrow\varepsilon_2)$, the only terms that survive are $\Gamma_a = \{\gamma_m, 2\mathrm{i}\sigma_{mn}|_{m<n}\}$, and also the relation $\sigma^{ij}\partial_i(\gamma_j\psi_\mu + \gamma_\mu\psi_j) = \sigma^{ij}[\gamma_j\psi_{i\mu} + \frac{1}{2}\gamma_\mu\psi_{ij} + \partial_\mu(\gamma_j\psi_i)]$. The second term has the form of a gauge transformation on ψ_μ, while the first one, on the field equations, becomes $-\mathrm{i}\bar{\varepsilon}_1\Gamma_a\varepsilon_2\sigma^{ij}\Gamma^a\gamma_j\psi_{i\mu} + \mathcal{O}(f)$. Using, further, the identities

$$\sigma^{ij}\sigma^{mn}\gamma_j = -\tfrac{1}{2}\sigma^{mn}\gamma^i \qquad \sigma^{ij}\gamma^m\gamma_j = \tfrac{1}{2}\gamma^m\gamma^i - 2\eta^{im}$$

we can see that the contribution of $\Gamma^a = \sigma^{mn}$ has the form $\mathcal{O}(f)$, and the evaluation of the contribution of $\Gamma^a = \gamma^m$ leads to the final result:

$$\begin{aligned}[\delta_{\varepsilon_1}, \delta_{\varepsilon_2}]\psi_\mu &= 2\mathrm{i}(\bar{\varepsilon}_1\gamma^\rho\varepsilon_2)\partial_\rho\psi_\mu\\ &\quad - \partial_\mu[2\mathrm{i}(\bar{\varepsilon}_1\gamma^m\varepsilon_2)\psi_m + \tfrac{1}{2}\mathrm{i}(\bar{\varepsilon}_1\Gamma_a\varepsilon_2)\sigma^{ij}\Gamma^a\gamma_j\psi_i] + \mathcal{O}(f).\end{aligned} \tag{9.49}$$

The action. We have found an on-shell irreducible representation of supersymmetry combined with an Abelian gauge symmetry, carried by the fields $\varphi_{\mu\nu}$ and ψ_μ. The action invariant under the SS transformations (9.48) and gauge transformations (9.46) represents the linearized supergravity theory, and is given as the sum of the free Fierz–Pauli and Rarita–Schwinger actions for the graviton and gravitino, respectively:

$$I_{\mathrm{SG}}^{\mathrm{L}} = \int \mathrm{d}^4x\,(\tfrac{1}{2}\varphi^{\mu\nu}G_{\mu\nu}^{\mathrm{L}} + \tfrac{1}{2}\bar{\psi}_\mu f^\mu) = I_{\mathrm{FP}} + I_{\mathrm{RS}} \tag{9.50}$$

where

$$G_{\mu\nu}^{\mathrm{L}} = R_{\mu\nu}^{\mathrm{L}} - \tfrac{1}{2}\eta_{\mu\nu}R^{\mathrm{L}}$$
$$R_{\mu\nu}^{\mathrm{L}} = -\Box\varphi_{\mu\nu} + \varphi^\sigma{}_{\mu,\nu\sigma} + \varphi^\sigma{}_{\nu,\mu\sigma} - \partial_\mu\partial_\nu\varphi \qquad R^{\mathrm{L}} = \eta^{\mu\nu}R_{\mu\nu}^{\mathrm{L}}.$$

The invariance holds without use of the equations of motion.

The gauge invariance of the action implies the following differential identities:

$$\partial^\mu G_{\mu\nu}^{\mathrm{L}} = 0 \qquad \partial^\mu f_\mu = 0.$$

Complete supergravity

Full supergravity may be obtained from the linearized theory by introducing nonlinear gravitational effects, and this has to be done in accordance with supersymmetry. The linearized theory possesses an Abelian *gauge* symmetry and global *supersymmetry*. The transition to a consistent self-interacting theory may be realized by applying the Noether coupling method, which leads to a theory with *gauge supersymmetry* or supergravity (West 1986). There is, however, another approach to the complete supergravity, in which the results of the linearized supergravity are *super-covariantized*, i.e. covariantized so as to ensure the validity of both the Poincaré gauge symmetry and supersymmetry. This leads to a supersymmetrized Poincaré gauge theory or gauged super-Poincaré theory.

From the point of view of Poincaré gauge theory, it is natural to expect the full supergravity action to have the form of an Einstein–Cartan theory with a massless Rarita–Schwinger matter field§:

$$I_{\mathrm{SG}} = \int \mathrm{d}^4x\,\left(-\frac{1}{2\kappa^2}bR + \frac{\mathrm{i}}{2}\varepsilon^{\mu\nu\rho\lambda}\bar{\psi}_\mu\gamma_5\gamma_\nu\nabla_\rho\psi_\lambda\right) \tag{9.51a}$$

where $\nabla_\rho\psi_\lambda = (\partial_\rho + \frac{1}{2}A^{ij}{}_\rho\sigma_{ij})\psi_\lambda$. Using the Riemannian connection $A = \Delta$ would be inconsistent, since the presence of spinor fields implies a non-trivial torsion. In the gravitino action there is no factor b, since $\varepsilon^{\mu\nu\rho\lambda}$ is not a tensor, but

§ The usual gravitational constant $\kappa = 8\pi G/c^2$ has dimension -2. In order to simplify the form of the SS gauge transformations we shall use, in this section, the redefined gravitational constant: $\kappa \to \kappa^2$.

a tensor density. The first term in I_{SG} may be rewritten in a more convenient form as

$$\mathcal{L}_{\mathrm{G}} = -\frac{1}{2\kappa^2} bR = \frac{1}{8\kappa^2} \varepsilon^{\mu\nu\rho\lambda}_{mnkl} b^k{}_\rho b^l{}_\lambda R^{mn}{}_{\mu\nu}(A) \tag{9.51b}$$

while the second term represents the covariant Rarita–Schwinger action. In the forthcoming discussion we shall see that the action (9.51a) is in agreement with supersymmetry without any additional modification.

Field equations. The variation of $\mathcal{L}_{\mathrm{G}}$ with respect to A, by making use of the Palatini identity $\delta_A R^{mn}{}_{\mu\nu} = \nabla_\mu(\delta A^{mn}{}_\nu) - \nabla_\nu(\delta A^{mn}{}_\mu)$, yields (up to a four-divergence)

$$\delta_A \mathcal{L}_{\mathrm{G}} = \frac{1}{2\kappa^2} \varepsilon^{\mu\nu\rho\lambda}_{mnkl} (\nabla_\nu b^k{}_\rho) b^l{}_\lambda \delta A^{mn}{}_\mu$$

while the variation of the Rarita–Schwinger Lagrangian has the form

$$\delta_A \widetilde{\mathcal{L}}_{\mathrm{RS}} = -\frac{\mathrm{i}}{8} \varepsilon^{\mu\nu\rho\lambda}_{kmnl} b^k{}_\nu \bar{\psi}_\rho \gamma^l \psi_\lambda \delta A^{mn}{}_\mu$$

where we have used $\gamma_5 \gamma_k \sigma_{mn} \to \frac{1}{2} \varepsilon_{kmnl} \gamma^l$. This implies the field equation for A,

$$T^k{}_{\nu\rho} \equiv \nabla_\nu b^k{}_\rho - \nabla_\rho b^k{}_\nu = \frac{\kappa^2}{2} \mathrm{i} \bar{\psi}_\rho \gamma^k \psi_\nu \tag{9.52a}$$

which can be solved for A:

$$A_{ij\mu} = \Delta_{ij\mu} - \tfrac{1}{2}(T_{ijm} - T_{mij} + T_{jmi}) b^m{}_\mu \equiv \bar{A}_{ij\mu}. \tag{9.52b}$$

Thus, the dynamics implies that the spacetime has a non-vanishing torsion.

The variation of the action with respect to $b^i{}_\nu$ and $\bar{\psi}_\mu$ yields the field equations

$$\begin{aligned} F_i{}^\nu &= \frac{b}{\kappa^2} G^\nu{}_i + \tfrac{1}{2} \mathrm{i} \varepsilon^{\mu\nu\rho\lambda} \bar{\psi}_\mu \gamma_5 \gamma_i \nabla_\rho \psi_\lambda \\ F^\mu &= \mathrm{i} \varepsilon^{\mu\nu\rho\lambda} \gamma_5 \gamma_\nu \nabla_\rho \psi_\lambda \end{aligned} \tag{9.53}$$

where $G^\nu{}_i = R^\nu{}_i - \frac{1}{2} h_i{}^\nu R$.

Gauge invariance of the theory implies that these two equations are not independent, but obey certain differential identities: their covariant derivatives vanish on-shell. Consider, for instance, an explicit proof of the identity $\nabla_\mu F^\mu = 0$. We start from the equality

$$\begin{aligned} \nabla_\mu F^\mu &= \mathrm{i} \varepsilon^{\mu\nu\rho\lambda} \gamma_5 [(\nabla_\mu b^i{}_\nu) \gamma_i \nabla_\rho \psi_\lambda + \gamma_\nu \nabla_\mu \nabla_\rho \psi_\lambda] \\ &= \tfrac{1}{4} \mathrm{i} \varepsilon^{\mu\nu\rho\lambda} \gamma_5 [\mathrm{i}\kappa^2 (\bar{\psi}_\nu \gamma^i \psi_\mu) \gamma_i \nabla_\rho \psi_\lambda + R^{ij}{}_{\mu\rho} \gamma_\nu \sigma_{ij} \psi_\lambda]. \end{aligned}$$

where we have used torsion equation (9.52a). In the second term we expand the product $\gamma_\nu \sigma_{ij}$ in terms of γ_r and $\gamma_5 \gamma_r$, using the cyclic identities

$$\begin{aligned} \varepsilon^{\mu\nu\rho\lambda} \gamma_5 \gamma_\nu \sigma_{ij} \psi_\lambda R^{ij}{}_{\mu\rho} &= \varepsilon^{\mu\nu\rho\lambda} b_{i\nu} R^{ij}{}_{\mu\rho} \gamma_5 \gamma_j \psi_\lambda - 2bG^\lambda{}_l \gamma^l \psi_\lambda \\ \varepsilon^{\mu\nu\rho\lambda} R^j{}_{\nu\mu\rho} &= \varepsilon^{\mu\nu\rho\lambda} \nabla_\rho T^j{}_{\nu\mu} \end{aligned}$$

and express $G^{\lambda}{}_{l}$ from the equation $F_l{}^{\lambda} = 0$. That the expression we have obtained vanishes follows from the Fierz identity for undifferentiated gravitino fields.

Different forms of the gravitino field equations can be easily obtained from the related linearized equations in M_4.

Local supersymmetry. We now demonstrate that action (9.51), in which the connection A is treated as an independent variable, is invariant under local SS transformations (Deser and Zumino 1976):

$$\begin{aligned} \delta b^i{}_\mu &= \kappa \mathrm{i}\bar{\varepsilon}\gamma^i \psi_\mu \qquad (\delta h_i{}^\mu = -\kappa \mathrm{i}\bar{\varepsilon}\gamma^\mu \psi_i) \\ \delta\psi_\mu &= -\frac{2}{\kappa}\nabla_\mu \varepsilon \end{aligned} \tag{9.54a}$$

and

$$\begin{aligned} \delta A^{mn}{}_\mu &= B^{mn}{}_\mu - \tfrac{1}{2} b^m{}_\mu B^{sn}{}_s + \tfrac{1}{2} b^n{}_\mu B^{sm}{}_s \\ B^{mn}{}_\mu &\equiv \kappa \mathrm{i}\bar{\varepsilon}\gamma_5\gamma_\mu \nabla_\nu \psi_\rho \varepsilon^{mn\nu\rho}. \end{aligned} \tag{9.54b}$$

In further exposition we shall often use $\kappa = 1$ for simplicity.

Let us first consider the variation with respect to $b^i{}_\mu$, ψ_μ and $\bar{\psi}_\mu$, which we denote by δ'. The variation of the gravitational action yields

$$\delta'\mathcal{L}_\mathrm{G} = -b(R^i{}_\mu - \tfrac{1}{2} b^i{}_\mu R)(-\mathrm{i}\bar{\varepsilon}\gamma^\mu \psi_i).$$

Variation of the gravitino action yields three terms, produced by varying with respect to ψ_λ, $\bar{\psi}_\mu$ and $\gamma_\nu = b^k{}_\nu \gamma_k$. The first term is given by

$$\delta'_1 \widetilde{\mathcal{L}}_\mathrm{RS} = -\mathrm{i}\varepsilon^{\mu\nu\rho\lambda} \bar{\psi}_\mu \gamma_5 \gamma_\nu \nabla_\rho \nabla_\lambda \varepsilon = -\tfrac{1}{4}\mathrm{i}\varepsilon^{\mu\nu\rho\lambda} \bar{\psi}_\mu \gamma_5 \gamma_\nu \sigma_{ij} \varepsilon R^{ij}{}_{\rho\lambda}.$$

The second term reads:

$$\begin{aligned} \delta'_2 \widetilde{\mathcal{L}}_\mathrm{RS} &= -\mathrm{i}\varepsilon^{\mu\nu\rho\lambda} (\nabla_\mu \bar{\varepsilon}) \gamma_5 \gamma_\nu \nabla_\rho \psi_\lambda \\ &= \mathrm{i}\varepsilon^{\mu\nu\rho\lambda} \bar{\varepsilon}\gamma_5 [(\nabla_\mu \gamma_\nu) \nabla_\rho \psi_\lambda + \gamma_\nu \nabla_\mu \nabla_\rho \psi_\lambda] \equiv A + B. \end{aligned}$$

The second part of this expression, taken together with $\delta'\mathcal{L}_\mathrm{G}$ and $\delta'_1 \widetilde{\mathcal{L}}_\mathrm{RS}$, yields zero. Indeed, the relation

$$\begin{aligned} \delta'_1 \widetilde{\mathcal{L}}_\mathrm{RS} + B &= -\tfrac{1}{4}\mathrm{i}\varepsilon^{\mu\nu\rho\lambda} R^{ij}{}_{\rho\lambda} [\bar{\psi}_\mu \gamma_5 \gamma_\nu \sigma_{ij} \varepsilon - \bar{\varepsilon}\gamma_5 \gamma_\nu \sigma_{ij} \psi_\mu] \\ &= -\tfrac{1}{4}\mathrm{i} R^{ij}{}_{\rho\lambda} \varepsilon^{\mu\nu\rho\lambda} b^k{}_\nu \varepsilon_{kijl} \bar{\psi}_\mu \gamma^l \varepsilon = \mathrm{i}b(R^i{}_\lambda - \tfrac{1}{2} b^i{}_\lambda R) \bar{\psi}_i \gamma^\lambda \varepsilon \end{aligned}$$

directly implies that $\delta'_1 \widetilde{\mathcal{L}}_\mathrm{RS} + B + \delta'\mathcal{L}_\mathrm{G} = 0$.

Thus, we are effectively left with the first part of $\delta'_2 \widetilde{\mathcal{L}}_\mathrm{RS}$, and $\delta'_3 \widetilde{\mathcal{L}}_\mathrm{RS}$:

$$\begin{aligned} A &= \tfrac{1}{2}\mathrm{i}\varepsilon^{\mu\nu\rho\lambda} T^m{}_{\mu\nu} \bar{\varepsilon}\gamma_5 \gamma_m \nabla_\rho \psi_\lambda \\ \delta'_3 \widetilde{\mathcal{L}}_\mathrm{RS} &= -\tfrac{1}{2}\varepsilon^{\mu\nu\rho\lambda} (\bar{\psi}_\mu \gamma_5 \gamma_m \nabla_\rho \psi_\lambda)(\bar{\varepsilon}\gamma^m \psi_\nu). \end{aligned}$$

The last term can be further transformed using the Fierz identity:

$$
\begin{aligned}
&(\bar\varepsilon\gamma^m\psi_\nu)(\bar\psi_\mu\gamma_5\gamma_m\nabla_\rho\psi_\lambda) - (\mu\leftrightarrow\nu)\\
&\qquad = -\tfrac{1}{4}(\bar\varepsilon\gamma^m\Gamma_A\gamma_5\gamma_m\nabla_\rho\psi_\lambda)(\bar\psi_\mu\Gamma^A\psi_\nu) - (\mu\leftrightarrow\nu).
\end{aligned}
$$

Here, due to the antisymmetry in (μ,ν), only those terms that contain $\Gamma_a \in \{\gamma_k, 2i\sigma_{kl}|_{k<l}\}$ may give a non-vanishing contribution. In addition, σ_{kl} should also be discarded since $\gamma^m\sigma_{kl}\gamma_m = 0$. The calculation with $\Gamma_a = \gamma_k$ leads to

$$
(\bar\varepsilon\gamma^m\psi_\nu)(\bar\psi_\mu\gamma_5\gamma_m\nabla_\rho\psi_\lambda) \to \tfrac{1}{2}(\bar\varepsilon\gamma_5\gamma_k\nabla_\rho\psi_\lambda)(\bar\psi_\mu\gamma^k\psi_\nu)
$$

which implies

$$
A + \delta_3'\widetilde{\mathcal{L}}_{\mathrm{RS}} = \tfrac{1}{2}\mathrm{i}\varepsilon^{\mu\nu\rho\lambda}(T^m{}_{\mu\nu} - \tfrac{1}{2}\mathrm{i}\bar\psi_\nu\gamma^m\psi_\mu)\bar\varepsilon\gamma_5\gamma_m\nabla_\rho\psi_\lambda.
$$

In the next step, we calculate the variation with respect to A:

$$
\delta_A(\mathcal{L}_{\mathrm{G}} + \mathcal{L}_{\mathrm{RS}}) = \tfrac{1}{4}\varepsilon^{\mu\nu\rho\lambda}_{mnkl}(T^k{}_{\mu\nu} - \tfrac{1}{2}\mathrm{i}\bar\psi_\nu\gamma^k\psi_\mu)b^l_\lambda\delta A^{mn}{}_\rho.
$$

That this term cancels the contribution of $A + \delta_3'\widetilde{\mathcal{L}}_{\mathrm{RS}}$ follows from the identity

$$
\varepsilon_{klmn}b^l_{[\lambda}\delta A^{mn}{}_{\rho]} = \mathrm{i}\bar\varepsilon\gamma_5\gamma_k\psi_{\lambda\rho}.
$$

This completes the proof of the local supersymmetry of action (9.51), in the first order formalism.

Algebra of local supersymmetries

In the rest of this chapter we shall replace the variable $A_{ij\mu}$ by the expression (9.52*b*), going thereby to the second order formulation of supergravity. The resulting action is invariant under the *local* super-Poincaré transformations with parameters $(a^\mu, \omega^{ij}, \varepsilon_\alpha)$:

$$
\begin{aligned}
\delta b^i{}_\mu &= -a^\rho\partial_\rho b^i{}_\mu - a^\rho{}_{,\mu}b^i{}_\rho + \omega^i{}_s b^s{}_\mu + \kappa\mathrm{i}\bar\varepsilon\gamma^i\psi_\mu\\
\delta\psi_\mu &= -a^\rho\partial_\rho\psi_\mu - a^\rho{}_{,\mu}\psi_\rho + \tfrac{1}{2}\omega\cdot\sigma\psi_\mu - \frac{2}{\kappa}\nabla_\mu\varepsilon.
\end{aligned}
\tag{9.55}
$$

The connection is not now an independent variable, and its on-shell transformation law is derived from the identification $A_{ij\mu} = \bar A_{ij\mu}$:

$$
\begin{aligned}
\delta_{\mathrm{T}}A^{ij}{}_\mu &= -a^\rho\partial_\rho A^{ij}{}_\mu - a^\rho{}_{,\mu}A^{ij}{}_\rho\\
\delta_{\mathrm{L}}A^{ij}{}_\mu &= \omega^i{}_s A^{sj}{}_\mu + \omega^j{}_s A^{is}{}_\mu - \partial_\mu\omega^{ij}\\
\delta_{\mathrm{S}}A^{ij}{}_\mu &= \tfrac{1}{2}\mathrm{i}\kappa(-\bar\varepsilon\gamma^i\psi_\mu{}^j + \bar\varepsilon\gamma^j\psi_\mu{}^i - \bar\varepsilon\gamma_\mu\psi^{ij}).
\end{aligned}
$$

where $\delta_{\mathrm{T}} = \delta_a$, $\delta_{\mathrm{L}} = \delta_\omega$ and $\delta_{\mathrm{S}} = \delta_\varepsilon$.

The Poincaré subalgebra has the standard form,

$$
\begin{aligned}
[\delta_{\mathrm{T}}(a_1), \delta_{\mathrm{T}}(a_2)] &= \delta_{\mathrm{T}}(a_1 \cdot \partial a_2 - a_2 \cdot \partial a_1) \\
[\delta_{\mathrm{L}}(\omega), \delta_{\mathrm{T}}(a)] &= \delta_{\mathrm{L}}(-a \cdot \partial \omega) \\
[\delta_{\mathrm{L}}(\omega_1), \delta_{\mathrm{L}}(\omega_2)] &= \delta_{\mathrm{L}}(\omega^i_{2s}\omega^{sj}_1 - \omega^i_{1s}\omega^{sj}_2)
\end{aligned}
\tag{9.56}
$$

while the commutator of a supersymmetry transformation and a Poincaré transformation reads:

$$
\begin{aligned}
[\delta_{\mathrm{S}}(\bar{\varepsilon}), \delta_{\mathrm{T}}(a)] &= \delta_{\mathrm{S}}(a \cdot \partial \bar{\varepsilon}) \\
[\delta_{\mathrm{S}}(\bar{\varepsilon}), \delta_{\mathrm{L}}(\omega)] &= \delta_{\mathrm{S}}(\tfrac{1}{2}\bar{\varepsilon}\sigma \cdot \omega).
\end{aligned}
\tag{9.57}
$$

The commutator of two supersymmetries on $b^i{}_\mu$ is given by

$$
[\delta_{\mathrm{S}}(\bar{\varepsilon}_1), \delta_{\mathrm{S}}(\bar{\varepsilon}_2)]b^i{}_\mu = [\delta_{\mathrm{T}}(-a^\rho) + \delta_{\mathrm{L}}(a^\rho A^{mn}{}_\rho) + \delta_{\mathrm{S}}(-a^\rho \bar{\psi}_\rho)]b^i{}_\mu \tag{9.58a}
$$

where $a^\rho = 2\mathrm{i}\bar{\varepsilon}_1\gamma^\rho\varepsilon_2$. The result follows from

$$
\begin{aligned}
-2\mathrm{i}\bar{\varepsilon}_2\gamma^i\nabla_\mu\varepsilon_1 - (\varepsilon_1 \leftrightarrow \varepsilon_2) &= -2\mathrm{i}\partial_\mu(\bar{\varepsilon}_2\gamma^i\varepsilon_1) - 2\mathrm{i}A^{in}{}_\mu\bar{\varepsilon}_2\gamma_i\varepsilon_1 \\
&= -2\mathrm{i}\partial_\mu(\bar{\varepsilon}_2\gamma^\lambda\varepsilon_1)b^i{}_\lambda - 2\mathrm{i}(\bar{\varepsilon}_2\gamma^\lambda\varepsilon_1)\nabla_\mu b^i{}_\lambda
\end{aligned}
$$

and the torsion equation.

The commutator of two supersymmetries on ψ_μ is more complicated:

$$
[\delta_1, \delta_2]\psi_\mu = -2\mathrm{i}\bar{\varepsilon}_1\gamma^m\varepsilon_2\psi_{\mu m} - \bar{\varepsilon}_1\gamma^\alpha\varepsilon_2 V_{\alpha\mu\rho}F^\rho + \bar{\varepsilon}_1\sigma^{\alpha\beta}\varepsilon_2 T_{\alpha\beta\mu\rho}F^\rho \tag{9.58b}
$$

where the structure functions V and T are given by

$$
\begin{aligned}
4bV_{\alpha\mu\rho} &= \gamma_\alpha\eta_{\mu\rho} + 2b\varepsilon_{\alpha\mu\rho\lambda}\gamma_5\gamma^\lambda \\
2bT_{\alpha\beta\mu\rho} &= (\eta_{\alpha\rho}\eta_{\beta\mu} - \eta_{\beta\rho}\eta_{\alpha\mu}) - \sigma_{\alpha\beta}\eta_{\mu\rho} - b\varepsilon_{\alpha\beta\mu\rho}\gamma_5.
\end{aligned}
\tag{9.58c}
$$

We note that the first term has the same form as the commutator on $b^k{}_\mu$,

$$
-a^\lambda\psi_{\mu\lambda} = [\delta_{\mathrm{T}}(-a^\rho) + \delta_{\mathrm{L}}(a^\rho A^{mn}{}_\rho) + \delta_{\mathrm{S}}(-a^\rho\bar{\psi}_\rho)]\psi_\mu
$$

while the structure functions V and T describe terms proportional to the equations of motion.

In order to prove the result (9.58b) we start from the equality

$$
[\delta_1, \delta_2]\psi_\mu = \mathrm{i}(\bar{\varepsilon}_1\gamma^i\psi_\mu{}^j)\sigma_{ij}\varepsilon_2 + \tfrac{1}{2}\mathrm{i}(\bar{\varepsilon}_1\gamma_\mu\psi^{ij})\sigma_{ij}\varepsilon_2 - (\varepsilon_1 \leftrightarrow \varepsilon_2)
$$

which is obtained after using the transformation law for $A^{ij}{}_\mu$. The right-hand side is transformed by means of the Fierz identity into

$$
R_\mu = -\tfrac{1}{2}\mathrm{i}(\bar{\varepsilon}_1\Gamma_a\varepsilon_2)\sigma^{ij}\Gamma^a(\gamma_i\psi_{\mu j} + \tfrac{1}{2}\gamma_\mu\psi_{ij})
$$

where, because of the antisymmetry with respect to $(\varepsilon_1 \leftrightarrow \varepsilon_2)$, the only non-vanishing contribution comes from $\Gamma_a \in \{\gamma_m, 2\mathrm{i}\sigma_{mn}|_{m<n}\}$.

The term with $\Gamma_a = \gamma_m$ is calculated by a suitable transformation of the expressions $\sigma^{ij}\gamma^m\gamma_i$ and $\sigma^{ij}\gamma^m\gamma_\mu\psi_{ij}$, and using different forms of the gravitino field equation:

$$R_\mu^{(1)} = -2\mathrm{i}\bar{\varepsilon}_1\gamma^m\varepsilon_2\psi_{\mu m} - \bar{\varepsilon}_1\gamma^\alpha\varepsilon_2 V_{\alpha\mu\rho}F^\rho .$$

The term containing $\Gamma_a = (2\mathrm{i}\sigma_{mn}|_{m<n})$ is also proportional to the field equation. After a straightforward calculation we obtain

$$R_\mu^{(2)} = \bar{\varepsilon}_1\sigma^{mn}\varepsilon_2 T_{mn\mu\rho}F^\rho$$

which proves the final result (9.58*b*).

Auxiliary fields

We now wish to find a formulation of supergravity with auxiliary fields, in which the algebra closes without imposing any field equations.

Linearized theory. The general idea for the structure of the auxiliary fields comes from the boson–fermion counting rule. Off-shell the symmetric tensor $\varphi_{\mu\nu}$ contributes 10 degrees of freedom minus four gauge degrees of freedom (local ξ symmetry), giving six independent bosonic degrees of freedom. On the other hand, four Majorana spinors ψ_μ contribute 16 degrees of freedom, which, after subtracting four gauge degrees of freedom (local θ symmetry), yields 12 fermionic degrees of freedom. Thus, the auxiliary fields must be chosen so as to compensate for the six missing bosonic degrees of freedom.

The choice of auxiliary fields is not unique. We shall try to construct a minimal formulation, in which there are no auxiliary spinors. We assume that the set of six bosonic fields consists of a scalar S, a pseudoscalar P and a pseudo-vector A_μ, and that these fields occur in the action as squares, and are of dimension $d = 2$. If the auxiliary fields are chosen so as to vanish on-shell, they must transform into the field equations. Then, on dimensional grounds we find the general form of the new SS transformations:

$$\begin{aligned}
\delta\varphi_{\mu\nu} &= \tfrac{1}{2}\mathrm{i}(\bar{\varepsilon}\gamma_\mu\psi_\nu + \bar{\varepsilon}\gamma_\nu\psi_\mu)\\
\delta\psi_\mu &= -2\partial_j\varphi_{i\mu}\sigma^{ij}\varepsilon + zA_\mu\gamma_5\varepsilon + \tfrac{1}{3}\mathrm{i}\gamma_\mu(S+\gamma_5 P)\varepsilon + c_1\gamma_\mu\gamma_5\hat{A}\varepsilon\\
\delta S &= c_2\mathrm{i}\bar{\varepsilon}\gamma\cdot F\\
\delta P &= c_3\mathrm{i}\bar{\varepsilon}\gamma_5\gamma\cdot F\\
\delta A_\mu &= c_4\bar{\varepsilon}\gamma_5 F_\mu + c_5\bar{\varepsilon}\gamma_5\gamma_\mu\gamma\cdot F.
\end{aligned}$$

These transformations reduce on-shell ($S = P = A_\mu = 0$) to the form (9.48). The constants c_a are determined by requiring that these transformations, together with the gauge transformations (9.46), form a closed algebra.

The commutator of two SS transformations on S gives

$$[\delta_1, \delta_2]S = -4\mathrm{i}c_2\bar{\varepsilon}_1\gamma^\rho\varepsilon_2\partial_\rho S + 4c_2(z - 3c_1)[\bar{\varepsilon}_2\gamma_5\sigma^{\rho\lambda}\varepsilon_1 - (\varepsilon_1 \leftrightarrow \varepsilon_2)]\partial_\rho A_\lambda$$

which implies $c_1 = z/3$, $c_2 = -1/2$. Similarly, the correct form of the commutator on P is obtained provided $c_3 = -1/2$. Finally, carrying out the commutator on A,

$$[\delta_1, \delta_2]A_\mu = -4\mathrm{i}c_5 z\bar{\varepsilon}_1\gamma^\rho\varepsilon_2\partial_\rho A_\mu + 4\mathrm{i}z(c_5 + \tfrac{1}{3}c_4)[\bar{\varepsilon}_2\sigma_{\mu\rho}\partial^\rho\hat{A}\varepsilon_1 - (\varepsilon_1 \leftrightarrow \varepsilon_2)]$$

we find $zc_5 = -1/2$, $zc_4 = 3/2$. Thus, all constants except z are determined, and z can be reduced to one by a rescaling of A_μ.

Using the relation

$$\delta\bar{\psi}_\mu = 2\partial_j\varphi_{i\mu}\bar{\varepsilon}\sigma^{ij} + zA_\mu\bar{\varepsilon}\gamma_5 - \tfrac{1}{3}\mathrm{i}\bar{\varepsilon}(S + \gamma_5 P)\gamma_\mu + c_1\bar{\varepsilon}\hat{A}\gamma_5\gamma_\mu$$

in the variation of $\widetilde{\mathcal{L}}_{\mathrm{RS}}$, we obtain

$$\delta\widetilde{\mathcal{L}}_{\mathrm{RS}} = -\tfrac{1}{3}\mathrm{i}\bar{\varepsilon}(S + \gamma_5 P)\gamma\cdot F + z\bar{\varepsilon}\gamma_5(F_\mu - \tfrac{1}{3}\gamma_\mu\gamma\cdot F)A^\mu.$$

Furthermore, it follows from $3\delta\widetilde{\mathcal{L}}_{\mathrm{RS}} = \delta(S^2 + P^2 + z^2A^2)$, with $z = 1$, that the invariant action is of the form

$$I^{\mathrm{L}} = I^{\mathrm{L}}_{\mathrm{SG}} - \int \mathrm{d}^4x\, \tfrac{1}{3}(S^2 + P^2 + A^2). \tag{9.59}$$

Upon eliminating the auxiliary fields S, P and A_μ, this action reduces to the form (9.50). The final form of the SS transformations reads:

$$\begin{aligned}
\delta\varphi_{\mu\nu} &= \tfrac{1}{2}\mathrm{i}(\bar{\varepsilon}\gamma_\mu\psi_\nu + \bar{\varepsilon}\gamma_\nu\psi_\mu) \\
\delta\psi_\mu &= -2\partial_j\varphi_{i\mu}\sigma^{ij}\varepsilon + A_\mu\gamma_5\varepsilon + \tfrac{1}{3}\mathrm{i}\gamma_\mu(S + \gamma_5 P - \mathrm{i}\gamma_5\hat{A})\varepsilon \\
\delta S &= -\tfrac{1}{2}\mathrm{i}\bar{\varepsilon}\gamma\cdot F \\
\delta P &= -\tfrac{1}{2}\mathrm{i}\bar{\varepsilon}\gamma_5\gamma\cdot F \\
\delta A^\mu &= \tfrac{1}{2}\bar{\varepsilon}\gamma_5(3F^\mu - \gamma^\mu\gamma\cdot F).
\end{aligned} \tag{9.60}$$

The complete theory. Analysis of the linearized theory suggests that the action of the full nonlinear supergravity with auxiliary fields will have the form

$$I = I_{\mathrm{SG}} - \int \mathrm{d}^4x\, \tfrac{1}{3}b(S^2 + P^2 + A^2). \tag{9.61}$$

The factor b gives an additional contribution to the variation of the action, and complicates the construction of SS transformations.

From the linearized transformation laws for $b^k{}_\mu$ and ψ_μ we find the following covariant expressions:

$$\begin{aligned}\delta b^k{}_\mu &= \mathrm{i}\bar{\varepsilon}\gamma^k\psi_\mu \\ \delta\psi_\mu &= -2\nabla_\mu\varepsilon + A_\mu\gamma_5\varepsilon + \tfrac{1}{3}\mathrm{i}\gamma_\mu(S+\gamma_5 P - \mathrm{i}\gamma_5\hat{A})\varepsilon.\end{aligned} \tag{9.62}$$

Let us now check, by calculating the commutator of two transformations on $b^k{}_\mu$, whether this result is in agreement with supersymmetry. The contribution of the term $-2\nabla_\mu\varepsilon$ is the same as before, while the contribution of the auxiliary fields is changed by the term

$$[\delta_1, \delta_2]b^i_\mu|_{\mathrm{AF}} = [\tfrac{2}{3}\mathrm{i}\varepsilon_2\gamma^\rho\varepsilon_1\varepsilon^{ikm}{}_\rho A_m - \tfrac{4}{3}\bar{\varepsilon}_2\sigma^{ik}(S+\gamma_5 P)\varepsilon_1]b_{k\mu}$$

which represents an additional Lorentz rotation. The complete result for the commutator of two supersymmetries on $b^i{}_\mu$ is:

$$\begin{aligned}[][\delta_1, \delta_2] &= \delta_{\mathrm{T}}(-a^\rho) + \delta_{\mathrm{L}}(a^\rho\hat{A}^{ik}{}_\rho) + \delta_{\mathrm{S}}(-a^\rho\bar{\psi}_\rho) \\ \hat{A}^{ik}{}_\rho &\equiv A^{ik}{}_\rho - \tfrac{1}{3}\varepsilon^{ikm}{}_\rho A_m\end{aligned} \tag{9.63}$$

where $a^\rho = 2\mathrm{i}\varepsilon_1\gamma^\rho\varepsilon_2$. The algebra closes without use of the field equations, and the covariance is in agreement with supersymmetry.

The covariant transformation of the auxiliary field S is given by

$$\delta S = -\tfrac{1}{2}\bar{\varepsilon}\gamma\cdot F + \cdots$$

(and similarly for P and A_μ) where dots indicate possible corrections due to supersymmetry. Applying this transformation rule to the action yields a non-vanishing result, due to the presence of the term $(\delta b)S^2$. Therefore, the expression for δS must contain an additional term proportional to $(\delta b/b)S$, in order to compensate for the contribution stemming from $(\delta b)S^2$ (a similar mechanism operates for P and A_μ).

In order to find the exact form of the additional terms, we observe that the transformation law for ψ_μ can be written as

$$\delta\psi_\mu = -2\nabla^{\mathrm{C}}_\mu\varepsilon \qquad \nabla^{\mathrm{C}}_\mu \equiv \nabla_\mu - \tfrac{1}{2}A_\mu\gamma_5 - \tfrac{1}{2}\mathrm{i}\gamma_\mu\eta$$

where $\eta = (S+\gamma_5 P - \mathrm{i}\gamma_5\hat{A})/3$, and ∇^{C}_μ is the so-called super-covariant derivative which contains, in addition to ∇_μ, some extra terms that reconcile *covariance* and *supersymmetry*. Define, further, the object

$$F^\mu_{\mathrm{C}} \equiv F^\mu|_{\nabla\to\nabla^{\mathrm{C}}} = \mathrm{i}\varepsilon^{\mu\nu\rho\lambda}\gamma_5\gamma_\nu(\nabla_\rho\psi_\lambda - \tfrac{1}{2}A_\rho\gamma_5\psi_\lambda - \tfrac{1}{2}\mathrm{i}\gamma_\rho\eta\psi_\lambda)$$

which represents the super-covariant extension of F^μ.

We determine the transformation laws for the auxiliary fields in the same way as for $\delta\psi_\mu$: by the replacement $\nabla_\mu \to \nabla^{\mathrm{C}}_\mu$.

In this way we obtain:

$$\begin{aligned}\delta S &= -\tfrac{1}{2}\mathrm{i}\bar{\varepsilon}\gamma\cdot F_{\mathrm{C}} \\ \delta P &= -\tfrac{1}{2}\mathrm{i}\bar{\varepsilon}\gamma_5\gamma\cdot F_{\mathrm{C}} \\ \delta A^m &= \tfrac{1}{2}\bar{\varepsilon}\gamma_5(3F^m_{\mathrm{C}} - \gamma^m\gamma\cdot F_{\mathrm{C}}).\end{aligned} \tag{9.64a}$$

Explicit verification shows that these transformations, together with (9.62), really represent the symmetry of action (9.61). Using the equalities

$$\begin{aligned}\gamma\cdot F_{\mathrm{C}} &= \gamma\cdot F + \mathrm{i}\gamma_5 A\cdot\psi + \gamma^\lambda(S+\gamma_5 P)\psi_\lambda \\ 3F^\mu_{\mathrm{C}} &= -\tfrac{1}{2}\mathrm{i}\varepsilon^{\mu\nu\rho\lambda}\gamma_\nu A_\rho\psi_\lambda + 2\sigma^{\mu\lambda}(S+\gamma_5 P)\psi_\lambda + \mathrm{i}\gamma_5 A^\mu\gamma\cdot\psi - \mathrm{i}\gamma_5\gamma^\mu A\cdot\psi\end{aligned}$$

the transformation laws for the auxiliary fields may be written in a more explicit form:

$$\begin{aligned}\delta S &= -\tfrac{1}{2}\mathrm{i}\bar{\varepsilon}\gamma\cdot F + \tfrac{1}{2}\bar{\varepsilon}\gamma_5 A\cdot\psi - \tfrac{1}{2}\mathrm{i}\bar{\varepsilon}\gamma_\rho(S+\gamma_5 P)\psi^\rho \\ \delta P &= -\tfrac{1}{2}\mathrm{i}\bar{\varepsilon}\gamma_5\gamma\cdot F - \tfrac{1}{2}\bar{\varepsilon}A\cdot\psi - \tfrac{1}{2}\mathrm{i}\bar{\varepsilon}\gamma_5\gamma_\rho(S+\gamma_5 P)\psi^\rho \\ \delta A^m &= \tfrac{1}{2}\bar{\varepsilon}\gamma_5(3F^m - \gamma^m\gamma\cdot F) - \tfrac{1}{2}\bar{\varepsilon}\gamma_5(S+\gamma_5 P)\psi^m \\ &\quad -\tfrac{1}{2}\mathrm{i}\bar{\varepsilon}\gamma\cdot\psi A^m + \tfrac{1}{2}\mathrm{i}\varepsilon^{mnks}(\bar{\varepsilon}\gamma_5\gamma_n\psi_k)A_s.\end{aligned} \tag{9.64b}$$

Dimensional analysis of the right-hand sides of equations (9.62) and (9.64) shows that the contributions of the auxiliary fields carry one unit of dimension more than is necessary. This dimensional excess is a consequence of our choice of units in which $\kappa = 1$, and disappears if we reintroduce the gravitational constant κ, with dimension $d(\kappa) = -1$.

General remarks

Bearing in mind that the exposition in this chapter aims to introduce only the basic aspects of gravity in the context of supersymmetry, we want to mention here several closely related and important topics. More details can be found in the additional literature (see, e.g., van Nieuwenhuizen 1981a, Sohnius 1985, West 1986, Srivastava 1986, Bailin and Love 1994).

Matter coupling. The interaction of supergravity with matter fields is determined by the rules that follow from the structure of local supersymmetry multiplets and the related local tensor calculus. The supergravity multiplet $(b^k{}_\mu, \psi_\mu, P, S, A^m)$ is defined by the transformation laws that lead to the *local* SS algebra (9.63). The local multiplets of matter fields must also have SS transformations that realize the same algebra. Thus, the first problem is to define local analogues of the general multiplet, chiral multiplet, etc. The most transparent method to construct a local multiplet is to apply the Noether coupling procedure to the algebra of SS transformations: we start with a global multiplet and the related SS algebra, and modify both the transformation law and the SS algebra, order by order in κ, so as to achieve the local algebra (9.63). The essential step in the construction consists in replacing all partial derivatives with supercovariant derivatives. If a field F transforms under supersymmetry according to $\delta F = \bar{\varepsilon} f$, then its supercovariant derivative is defined by the relation

$$\nabla^{\mathrm{C}}_{\mu} F = \partial_{\mu} F + \tfrac{1}{2}\bar{\psi}_{\mu} f.$$

The transformation law of $\nabla^{\mathrm{C}}_{\mu} F$ does not contain $\partial\varepsilon$. In the next step, we find the rules for combining two local supermultiplets into a third one; we can define the products $\phi_1 \cdot \phi_2$, $\phi_1 \times \phi_2$ and $\phi_1 \wedge \phi_2$, with a structure that parallels the related global results, with the replacement $\partial_{\mu} \to \nabla^{C}_{\mu}$. The last part of the local tensor calculus consists in finding formulae for the construction of invariant actions, which generalize the related formulae of global supersymmetry. After that we can construct realistic models of supergravity in the interaction with matter fields, and study their physical implications.

Spontaneous supersymmetry breaking. Exact supersymmetry implies that bosonic and fermionic states in every supermultiplet have the same mass, so that all the known elementary particles should have superpartners with the same mass. Since no such mass degeneracy is observed in nature, supersymmetry must be broken. In realistic applications supersymmetry is broken spontaneously. This mechanism can make the unobserved superpartners highly massive, ‘explaining’ thereby their unobservability, whereas the standard particles acquire different masses.

Particle	Spin	Superparticle	Spin
electron	$\frac{1}{2}$	selectron	0
quark	$\frac{1}{2}$	squark	0
photon	1	photino	$\frac{1}{2}$
gluon	1	gluino	$\frac{1}{2}$
$W^{\pm}$, Z^0	1	W-ino, Z-ino	$\frac{1}{2}$
graviton	2	gravitino	$\frac{3}{2}$

As we have seen, in a theory with global simple supersymmetry we have $E \geq 0$ for every state. The vacuum state $|0\rangle$ has vanishing energy, $E_0 = 0$, if and only if all the SS generators annihilate it. The state with $E_0 = 0$ is the supersymmetric vacuum, while $E_0 > 0$ means that supersymmetry is spontaneously broken. Thus, the vacuum energy is an indicator for spontaneously broken supersymmetry.

For a given field φ we have $\delta\varphi = [\bar{\varepsilon} Q, \varphi]$, so that supersymmetry is spontaneously broken if and only if $\langle\delta\varphi\rangle \neq 0$ for some $\delta\varphi$. Consider as an illustration the interacting Wess–Zumino model. The Lorentz invariance of the vacuum implies $\langle\psi_\alpha\rangle = \langle\partial_\mu A\rangle = \langle\partial_\mu B\rangle = 0$, hence the only field variation that may have a non-vanishing vacuum expectation value is $\langle\delta\psi\rangle = (\langle F\rangle + \gamma^5\langle G\rangle)\varepsilon$. A closer inspection of the Wess–Zumino model shows that both $\langle F\rangle$ and $\langle G\rangle$ vanish, so that supersymmetry is not broken.

The Wess–Zumino model may be modified so as to have $\langle F\rangle = \lambda$. In that case, we find an inhomogeneous term in the transformation law for ψ, $\langle\delta\psi\rangle = \lambda\varepsilon$, which implies that ψ is a Goldstone fermion, or *goldstino*. This massless particle is an analogue of the usual Goldstone boson.

The fact that the goldstino is not found in nature represents a problem for theories with broken global supersymmetry. The local supersymmetry resolves this problem through the super-Higgs mechanism: the massless goldstino is 'gauged away' and absorbed by the gauge field ψ_μ, the gravitino, which becomes massive.

Superspace and superfields. A Poincaré algebra may be realized by a set of transformations of points in M_4. Analysing the behaviour of ordinary fields $\phi(x)$ in M_4 under Poincaré transformations, we can define representations of the Poincaré (and Lorentz) group on fields and use them to construct relativistic field theories.

We have seen that an SS algebra may be represented on supermultiplets of fields. A very useful technique for dealing with SS theories is to represent a supermultiplet by a *superfield*, which is defined on a *superspace* with coordinates

$$z^M = (x^\mu, \theta_a, \bar\theta^{\dot a})$$

where θ and $\bar\theta$ are anticommuting spinors. Supersymmetry transformations may be conveniently represented as transformations of points in superspace.

In order to indicate the role of superspace, we begin by considering Poincaré transformations in M_4. Let ϕ be a scalar field on M_4, and $g(\Lambda, a)$ an element of the Poincaré group. Acting with the transformation $g(\Lambda, a)$ on $\phi_x = g(1, x)\phi(0)$ and using the group composition rules we obtain $g(\Lambda, a)\phi_x = g(1, \Lambda x + a)\phi(0) = \phi_{\Lambda x+a}$. This shows that the Poincaré transformation $g(\Lambda, a)$ may be realized as the coordinate transformation $x' = \Lambda x + a$ of points in M_4.

In a similar way we can obtain a representations of supersymmetry as a transformation of points in superspace. Let $g(x, \theta, \bar\theta)$ be an element of the super-Poincaré group generated by $(P, Q, \bar Q)$, which describes an ordinary translation and a super-translation in superspace. We define a scalar field Φ at $(x, \theta, \bar\theta)$ as

$$\Phi_{(x,\theta,\bar\theta)} = g(x, \theta, \bar\theta)\Phi(0) \qquad g(x, \theta, \bar\theta) \equiv \exp(x\cdot P - \theta Q - \bar\theta\bar Q).$$

Then, $g(a, \xi, \bar\xi)\Phi_{(x,\theta,\bar\theta)} = \Phi_{(x',\theta',\bar\theta')}$, where

$$\begin{aligned} x'^\mu &= x^\mu + a^\mu - \mathrm{i}(\xi\sigma^\mu\bar\theta - \theta\sigma^\mu\bar\xi) \\ \theta'_a &= \theta_a + \xi_a \qquad \bar\theta'_{\dot a} = \bar\theta_{\dot a} + \bar\xi_{\dot a}. \end{aligned}$$

This defines ordinary and SS translations on superspace. In this representation, the SS generators have the form

$$Q_a = \partial_a - \mathrm{i}(\sigma^\mu \bar{\theta})_a \partial_\mu \qquad \bar{Q}^{\dot{a}} = \partial^{\dot{a}} + \mathrm{i}(\theta\sigma^\mu)^{\dot{a}} \partial_\mu$$

where $\partial_a = \partial/\partial\theta^a$, $\partial^{\dot{a}} = \partial/\partial\bar{\theta}_{\dot{a}}$, and $P_\mu = -\partial_\mu$, as before.

Superfields are constructed so as to carry representations of the super-Poincaré group, in analogy with ordinary fields in M_4 which carry representations of the Poincaré group. They contain just those components that form supermultiplets. The use of superfields leads to a very elegant formalism that keeps the supersymmetry manifest. Field theories defined in terms of superfields become very compact compared with component formulations. Superfields are particularly useful for calculating quantum corrections, where a superfield Feynman diagram gives a compact description of many component–field diagrams.

Quantum effects. In perturbative calculations of physical processes in quantum field theory, we find divergent integrals with divergences that originate from integrations over small distances or large (ultraviolet) momenta. In renormalizable field theories, the problem of ultraviolet divergences is treated by redefining a finite number of physical parameters, such as masses, coupling constants, etc, in such a way that physical predictions of the theory remain finite. The procedure by itself is mathematically not well founded, but has a clear physical interpretation and leads to theoretical predictions that are in good agreement with experiments. The gauge theories of electroweak and strong interactions are renormalizable, but all attempts to construct a consistent, renormalizable quantum theory of gravity have failed. This is certainly one of the most important problems of present day elementary particle physics.

In addition to unifying bosons and fermions, SS theories are also attractive because of their renormalizability properties. Quantum corrections come from integrations over loop diagrams. Since fermions are anticommuting objects, fermion loops contribute with an opposite sign compared to boson loops. If the parameters of the theory (masses and coupling constants) are mutually related in a convenient way, the contributions of boson and fermion loops may cancel each other. This is just what happens in SS theories, to some extent. One of the most impressive cases of this kind is the $N = 4$ non-Abelian gauge theory, which is not only renormalizable, but actually finite (to any order of the perturbation theory): ultraviolet divergences coming from different types of particles 'cancel' each other, and we are left only with finite contributions. Physically, however, this theory seems irrelevant, since its particle spectrum is not realistic.

Einstein's theory of gravity is finite at the one-loop level. Analysing a large number of different diagrams at the two-loop level it has been found that the theory is divergent (Goroff and Sagnotti 1985, 1986). If matter fields are present, the renormalizability is lost even at the one-loop level.

Experience with globally SS theories raised the hope that quantum supergravity might be a renormalizable theory. Explicit calculations in some supergravity models have proved finiteness at one and two loops, while possible problems arise at the three-loop level. Although supersymmetry significantly improves the renormalizability properties of gravity, it seems that standard supergravity does not represent an appropriate framework for a renormalizable quantum gravity.

Exercises

1. Consider the relations

$$\{Q_\alpha, Q_\beta\} = \mathrm{i}a(\gamma^\mu C)_{\alpha\beta} P_\mu + \mathrm{i}b(\sigma^{\mu\nu}C)_{\alpha\beta} M_{\mu\nu}$$
$$[P_\mu, Q_\alpha] = c(\gamma_\mu)_\alpha{}^\beta Q_\beta$$
$$[M_{\mu\nu}, Q_\alpha] = d(\sigma_{\mu\nu})_\alpha{}^\beta Q_\beta.$$

Use the Jacobi identities to find the values of the parameters b, c and d.

2. (a) Find the algebra of the SS transformations (9.10) in the Wess–Zumino model.
   (b) Calculate the corresponding Noether current.
3. Derive the algebra of the transformations (9.14a) in SS electrodynamics.
4. (a) Show that the simple super-Poincaré algebra is invariant under the chiral transformation $Q \to \mathrm{e}^{\alpha\gamma_5} Q$.
   (b) Find the realization of chiral transformations on the Wess–Zumino multiplet (A, B, ψ, F, G).
5. Derive the following relations, using the super-Poincaré algebra (9.6):

$$\{Q_\alpha, \bar{Q}_\beta\} = -2\mathrm{i}(\gamma^\mu)_{\alpha\beta} P_\mu \qquad \{\bar{Q}_\alpha, \bar{Q}_\beta\} = -2\mathrm{i}(C^{-1}\gamma^\mu)_{\alpha\beta} P_\mu$$
$$[M_{\mu\nu}, \bar{Q}_\alpha] = \bar{Q}_\beta (\sigma_{\mu\nu})^\beta{}_\alpha.$$

6. Verify the two-component form (9.15) of the simple super-Poincaré algebra.
7. (a) Show that for $m^2 > 0$ the operator $X^\mu = W^\mu - \frac{1}{8}\mathrm{i}\bar{Q}\gamma^\mu\gamma_5 Q$ obeys the relations:

$$[X^\mu, Q_\alpha] = \tfrac{1}{2}(\gamma_5 Q)_\alpha P^\mu$$
$$[X^1, X^2] = -\mathrm{i}mX^3 \qquad \text{etc} \qquad \text{(in the rest frame)}.$$

   (b) Prove that $N^\mu \equiv \frac{1}{8}\mathrm{i}\bar{Q}\gamma^\mu\gamma_5 Q = \frac{1}{8}(\bar{Q}\bar{\sigma}^\mu Q - Q\sigma^\mu \bar{Q})$.
8. Find the helicity structure of the SS multiplets of massless states:
   (a) $N = 4$, $\lambda_0 = -1$;
   (b) $N = 8$, $\lambda_0 = -2$.
9. (a) Show that in the Clifford ground state $|\Omega\rangle$ for $m^2 > 0$, the superspin operator in the rest frame has the same value as the usual spin operator: $y = j$, $y_3 = j_3$.

(b) Calculate the spin and superspin projections (j_3, y_3) for the members of the massive multiplets built over the ground state $|\Omega\rangle$ for which (i) $y = 0$, $y_3 = 0$; (ii) $y = 1/2$, $y_3 = 1/2$; (iii) $y = 1/2$, $y_3 = -1/2$.

10. (a) Show that the complex form of the transformation law for the chiral multiplet (9.26*a*) realizes the SS algebra.
    (b) Derive the transformation law for the chiral multiplet $\phi = (A, B, \psi, F, G)$ from its complex form.
11. (a) Prove that the transformations (9.27*b*) of the general multiplet realize the SS algebra.
    (b) Derive the transformation law of the curl multiplet dV, and verify the related SS algebra.
    (c) Show that the real general multiplet constrained by $dV = 0$ becomes a chiral multiplet.
12. (a) Find the components of the antisymmetric product $\phi_1 \wedge \phi_2$ of two chiral multiplets.
    (b) Prove that the kinetic multiplet $T\phi$ is a chiral multiplet.
    (c) Show that $[\phi \times \phi]_D$ differs from $-2[\phi \cdot T\phi]_F$ by a divergence.
13. Prove the following relation:

$$\begin{aligned}&(\bar{\varepsilon}\psi_1)(\bar{u}\psi_2) - (\bar{\varepsilon}\gamma_5\psi_1)(\bar{u}\gamma_5\psi_2) + (\psi_1 \leftrightarrow \psi_2)\\ &\quad = -(\bar{\psi}_1\psi_2)(\bar{u}\varepsilon) + (\bar{\psi}_1\gamma_5\psi_2)(\bar{u}\gamma_5\varepsilon).\end{aligned}$$

14. Express the interacting Wess–Zumino Lagrangian in terms of the variables $\widetilde{F} = F + [mA - g(A^2 - B^2)]$ and $\widetilde{G} = G + [mB - 2gAB]$, and derive their SS transformations.
15. Find the Noether current corresponding to the SS symmetry of the interacting Wess–Zumino model.
16. Show that the quantities $P_{\mu\nu}$, $L_{\mu\nu}$ and $\Lambda_{\mu\nu}$, defined in equation (9.43), are projectors, i.e. that they obey the relations:

$$\begin{gathered}P_{\mu\nu} + L_{\mu\nu} + \Lambda_{\mu\nu} = \eta_{\mu\nu}\\ P_{\mu\nu}L^{\nu\lambda} = 0 \qquad P_{\mu\nu}\Lambda^{\nu\lambda} = 0 \qquad L_{\mu\nu}\Lambda^{\nu\lambda} = 0\\ P_{\mu\nu}P^{\nu\lambda} = P_\mu{}^\lambda \qquad L_{\mu\nu}L^{\nu\lambda} = L_\mu{}^\lambda \qquad \Lambda_{\mu\nu}\Lambda^{\nu\lambda} = \Lambda_\mu{}^\lambda.\end{gathered}$$

17. Find the explicit form of the term $\mathcal{O}(f)$ in the commutator (9.49), and calculate the related structure functions defined in (9.58*b*).
18. (a) Prove the different forms of the gravitino field equations (9.38) in M_4.
    (b) Find the related gravitino field equations in full on-shell supergravity.
19. Verify the consistency condition $\nabla_\mu F^\mu = 0$ of the gravitino field equation in full on-shell supergravity.
20. Prove that $R^i{}_{\lambda\mu} = C^i{}_{\lambda\mu} - T^i{}_{\lambda\mu}$ has the following transformation laws:

$$\begin{aligned}\delta_{\mathrm{T}}R^i{}_{\lambda\mu} &= -a^\rho{}_{,\mu}R^i{}_{\lambda\rho} - a^\rho{}_{,\lambda}R^i{}_{\rho\mu} - a^\rho\partial_\rho R^i{}_{\lambda\mu}\\ \delta_{\mathrm{L}}R^i{}_{\lambda\mu} &= \omega^i{}_s R^s{}_{\lambda\mu} + \omega^i{}_{s,\lambda}b^s{}_\mu - \omega^i{}_{s,\mu}b^s{}_\lambda\\ \delta_{\mathrm{S}}R^i{}_{\lambda\mu} &= \mathrm{i}\kappa[-\bar{\varepsilon}\gamma^i(\nabla_\mu\psi_\lambda - \nabla_\lambda\psi_\mu) + A^{is}{}_\mu\bar{\varepsilon}\gamma_s\psi_\lambda - A^{is}{}_\lambda\bar{\varepsilon}\gamma_s\psi_\mu].\end{aligned}$$

Chapter 10

Kaluza–Klein theory

The spacetime in GR is a Riemannian four-dimensional continuum V_4. In accordance with the principle of equivalence, gravity is determined by the geometry of spacetime. Very soon after the discovery and experimental verification of GR, Kaluza (1921) proposed that the four-dimensional spacetime be supplemented with a *fifth dimension*, in order to give a unified account of the gravitational and electromagnetic interactions, the only two basic interactions known at the time. Although the physical effects of gravity and electromagnetism are seemingly very distinct (all particles are subject to gravity, but only charged particles are subject to electromagnetism), Kaluza showed that both can emerge as different manifestations of a five-dimensional GR. In this theory point particles move along geodesic lines in a five-dimensional Riemann spacetime V_5, and these trajectories are seen in V_4 as trajectories of particles subject to both gravitational and electromagnetic forces.

Kaluza studied only the classical structure of the five-dimensional gravity. The first analysis of the compatibility of this theory with quantum mechanics was given by Klein (1926). Later investigations generalized Kaluza's idea to a spacetime having *more than five* dimensions, giving rise to a unification of gravity with non-Abelian gauge theories. All higher-dimensional theories that attempt to unify gauge theories with gravity are now called Kaluza–Klein (KK) theories.

In the papers by Kaluza and Klein it is not clear whether the fifth dimension should be taken seriously or merely as a useful mathematical device necessary to obtain a unified four-dimensional theory, whereupon its physical meaning is completely lost. Nowadays, it is widely accepted that the fifth dimension should be considered as a true, physical dimension. The explanation of why the new dimension has not been detected up to now is 'found' in the assumption that the spacetime along the fifth dimension is curled up into an exceedingly small circle, so small that it could not yet be observed. This is also true for small gravitational fields, so that the ground state structure of the five-dimensional space V_5 has to be essentially different from that of M_5. It can be pictured as an ordinary spacetime V_4, with a tiny circle 'attached' to every point of it.

Significant advances in our understanding of the fundamental interactions in the 1970s brought a revival of interest in Kaluza's ideas. There are several reasons for this. First, since the electroweak and strong interactions can be successfully described as gauge theories, the KK approach may serve as a framework for studying their unification with gravity. Then, the existence of the ground state which does not have the form of M_5 can be best understood as an effect of spontaneous symmetry breaking. Finally, using the idea of supergravity in KK theory it is possible to have a geometric description of both gauge fields and spinor matter.

In spite of many advances, the realization of these ideas has been accompanied by many difficulties, so that even today, more than seven decades after its birth, there is no realistic KK theory. Nevertheless, having a feeling that there must be at least a part of the truth in these intriguing ideas, physicists continue to study them, since 'it appears hard to believe that those relations, hardly to be surpassed in their formal correspondence, are nothing but an alluring play of whimsical chance' (Kaluza 1921).

10.1 Basic ideas

We begin our exposition of KK theories with the theory in five dimensions. The basic motive of the five-dimensional KK theory is the unification of gravity and electromagnetism. The realization of this goal demands a deeper understanding of the role of ground state in gravitational theories (see, for instance, Orzalesi 1981, Witten 1981a, 1982, Mecklenburg 1983, Freedmann and van Nieuwenhuizen 1985, Bailin and Love 1987).

Gravity in five dimensions

Five-dimensional KK theory is defined in a five-dimensional Riemann space V_5 with metric $\hat{g}_{MN}$ of the signature $(+, -, -, -; -)$. The upper-case Latin letters $(M, N, \ldots = 0, 1, 2, 3, 5)$ denote coordinate indices in V_5 and the lower-case Greek letters $(\mu, \nu, \ldots = 0, 1, 2, 3)$ denote, as before, the coordinate indices in four-dimensional Riemann space V_4. The original KK theory is simply five-dimensional GR, determined by the action

$$I_{\mathrm{G}} = -\frac{1}{2\hat{\kappa}} \int \mathrm{d}^5 z \sqrt{\hat{g}}\widehat{R} \tag{10.1}$$

where $z^M = (x^\mu, y)$, $\hat{g} = \det(\hat{g}_{MN})$, $\widehat{R}$ is the scalar curvature and $\hat{\kappa}$ the gravitational constant in V_5. The possibility of having additional matter fields will be considered later. The equations of motion are

$$\widehat{R}_{MN} = 0.$$

The action (10.1) is invariant under general coordinate transformations:

$$z'^M = z'^M(z) \qquad \hat{g}'_{MN}(z') = \frac{\partial z^R}{\partial z'^M}\frac{\partial z^L}{\partial z'^N}\hat{g}_{RL}(z).$$

Kaluza's mechanism. In order to ensure the unobservability of the fifth dimension, Kaluza introduced the assumption that there exists a coordinate system in V_5 such that the metric tensor is independent of the extra coordinate:

$$\partial_y \hat{g}_{MN} = 0. \tag{10.2a}$$

The so-called *cylinder condition* (10.2a) is not generally covariant, but there is a wide class of coordinate systems in which it holds true. Going over to new coordinates we find that condition (10.2a) remains fulfilled provided the partial derivatives ($\partial x'^\mu/\partial x^\lambda$, $\partial x'^\mu/\partial y$, $\partial y'/\partial x^\lambda$, $\partial y'/\partial y$) do not depend on y. It follows that the permitted coordinate transformations have the form $x'^\mu = x'^\mu(x)$, $y' = \rho y + \varepsilon(x)$, where ρ is a constant.

The metric tensor $\hat{g}_{MN}$ has 15 components, which can be naturally grouped in the following way: $\hat{g}_{MN} = (\hat{g}_{\mu\nu}, \hat{g}_{\mu 5}, \hat{g}_{55})$. Starting from the idea that 10 components $\hat{g}_{\mu\nu}$ describe *gravity* and four components $\hat{g}_{\mu 5}$ the *electromagnetic field*, the component $\hat{g}_{55}$ appears as a redundant degree of freedom. Following the ideas from the early period, we eliminate this component by imposing the *additional condition*

$$\hat{g}_{55} = -1. \tag{10.2b}$$

The negative sign of $\hat{g}_{55}$ corresponds to having y spacelike. Condition (10.2b) restricts the value of ρ to $\rho = 1$, so that the residual coordinate transformations have the form:

$$x' = x'(x) \tag{10.3a}$$

$$y' = y + \varepsilon(x). \tag{10.3b}$$

The metric. The previous suggestion concerning the physical interpretation of the metric components can be checked by looking at their transformation laws. Under transformation (10.3a) the components $\hat{g}_{\mu\nu}$ and $\hat{g}_{\mu 5}$ transform according to

$$\hat{g}'_{\mu\nu} = \frac{\partial x^\lambda}{\partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\nu}\hat{g}_{\lambda\rho} \qquad \hat{g}'_{\mu 5} = \frac{\partial x^\lambda}{\partial x'^\mu}\hat{g}_{\lambda 5}$$

i.e. as the usual four-dimensional tensors. Next, consider the transformation (10.3b):

$$\hat{g}'_{\mu\nu} = \hat{g}_{\mu\nu} - \partial_\mu\varepsilon\hat{g}_{\nu 5} - \partial_\nu\varepsilon\hat{g}_{\mu 5} - \partial_\mu\varepsilon\partial_\nu\varepsilon$$

$$\hat{g}'_{\mu 5} = \hat{g}_{\mu 5} + \partial_\mu\varepsilon.$$

Formally, $\hat{g}_{\mu 5}$ transforms like the electromagnetic potential. However, if we recall that gauge transformations in electrodynamics are, in fact, local $U(1)$

transformations, we see that the identification of $\hat{g}_{\mu 5}$ with the electromagnetic potential could be accommodated by postulating the extra dimension to be geometrically a *circle*. In that case any movement along y in V_5 may be interpreted as an Abelian gauge transformation of $\hat{g}_{\mu 5}$.

On the other hand, it seems natural to demand that the metric of the physical four-dimensional space be invariant under translations along y, which is not the case with $\hat{g}_{\mu\nu}$. What, then, is the metric of the physical spacetime V_4? By observing that the combination

$$g_{\mu\nu} \equiv \hat{g}_{\mu\nu} + \hat{g}_{\mu 5}\hat{g}_{\nu 5}$$

is invariant under (10.3b), we conclude that $g_{\mu\nu}$ is the correct metric of V_4.

Using the quantities $g_{\mu\nu}$ and $\hat{g}_{\mu 5} \equiv -B_\mu$, which have a direct physical interpretation, the original five-dimensional metric takes the form

$$\hat{g}_{MN} = \begin{pmatrix} g_{\mu\nu} - B_\mu B_\nu & -B_\mu \\ -B_\nu & -1 \end{pmatrix}. \tag{10.4a}$$

The inverse metric has the form

$$\hat{g}^{MN} = \begin{pmatrix} g^{\mu\nu} & -B^\mu \\ -B^\nu & -1 + B_\lambda B^\lambda \end{pmatrix} \tag{10.4b}$$

where $g^{\mu\nu}$ is the inverse of $g_{\mu\nu}$, and $B^\mu = g^{\mu\nu}B_\nu$. Four-dimensional indices are raised and lowered with the help of $g^{\mu\nu}$ and $g_{\mu\nu}$.

The reduced action. A physical interpretation of the theory derives from the effective four-dimensional form of the action (10.1), restricted by conditions (10.2a, b). The Christoffel connection $\hat{\Gamma}^M_{NR}$ is given by

$$\begin{aligned} \hat{\Gamma}^\mu_{\nu\rho} &= \Gamma^\mu_{\nu\rho} + \tfrac{1}{2}(F^\mu{}_\nu B_\rho + F^\mu{}_\rho B_\nu) \\ \hat{\Gamma}^5_{\nu\rho} &= \tfrac{1}{2}(\nabla_\nu B_\rho + \nabla_\rho B_\nu) - \tfrac{1}{2}B^\lambda(F_{\lambda\nu}B_\rho + F_{\lambda\rho}B_\nu) \\ \hat{\Gamma}^5_{5\rho} &= -\tfrac{1}{2}B^\lambda F_{\lambda\rho} \qquad \hat{\Gamma}^\mu_{5\rho} = \tfrac{1}{2}F^\mu{}_\rho \qquad \hat{\Gamma}^\mu_{55} = \hat{\Gamma}^5_{55} = 0 \end{aligned} \tag{10.5a}$$

where $F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$. Then a direct calculation yields

$$\begin{aligned} \widehat{R}_{\mu\nu} &= R_{\mu\nu} + \tfrac{1}{2}(B_\mu \nabla_\rho F^\rho{}_\nu + B_\nu \nabla_\rho F^\rho{}_\mu) + \tfrac{1}{4}B_\mu B_\nu F^2 + \tfrac{1}{2}F^\rho{}_\mu F_{\rho\nu} \\ \widehat{R}_{\mu 5} &= \tfrac{1}{2}\nabla_\rho F^\rho{}_\mu + \tfrac{1}{4}B_\mu F^2 \qquad \widehat{R}_{55} = \tfrac{1}{4}F^2 \end{aligned} \tag{10.5b}$$

where $R_{\mu\nu} = R_{\mu\nu}(V_4)$, and $F^2 = F_{\mu\nu}F^{\mu\nu}$. This implies

$$\widehat{R} = R + \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{10.5c}$$

Now, using the equality $\sqrt{\hat{g}} = \sqrt{-g}$ we see that the integrand in (10.1) does not depend on y. Therefore, in order to have a finite action the domain of the fifth coordinate has to have a finite measure. This is true, in particular, if the extra spatial dimension is compact, with the geometry of a circle, $0 \leq y \leq L$, in accordance with the physical interpretation of $\hat{g}_{\mu 5}$. Then, the integration over y in (10.1) leads to the following, reduced four-dimensional action:

$$I_{\mathrm{G}}^{(0)} = -\frac{1}{2\kappa}\int \mathrm{d}^4x\,\sqrt{-g}(R + \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}) \qquad \kappa \equiv \hat{\kappa}/L. \tag{10.6a}$$

The form of the first term in the action shows that κ is the usual four-dimensional gravitational constant, $\sqrt{\kappa\hbar/8\pi c} = 1.6 \times 10^{-33}$ cm. Moreover, it is evident that the Maxwell action is not well normalized. Since B_μ is dimensionless and the electromagnetic potential has the dimension of mass, we define

$$B_\mu = f A_\mu \qquad f^2 = 2\kappa \tag{10.6b}$$

where the condition $f^2/2\kappa = 1$ ensures the standard form of the Maxwell action: $-\frac{1}{4}F^2(A)$. Had the extra dimension been timelike, $\hat{g}_{55} = +1$, we would have obtained the wrong sign here.

While the original action (10.1) is invariant under five-dimensional general coordinate transformations, transition to the Kaluza metric reduces this symmetry to the form (10.3). Because of the cylinder condition, all that remains from the use of the fifth dimension is the increased number of fields in $d = 4$: the reduced four-dimensional theory describes gravity and electromagnetism. This procedure, in which the resulting four-dimensional theory is obtained starting from a higher-dimensional theory in such a way, that its properties can be studied without attributing physical existence to the extra dimension, is called a *dimensional reduction*.

On the other hand, we can accept the fifth dimension as a physical reality, and obtain all geometric properties of V_5, including the compactness of the extra dimension and the cylinder condition, by relying on the five-dimensional field equations. This approach, based on the modern concept of spontaneous symmetry breaking, is known as *spontaneous compactification*. It makes the idea of higher-dimensional gravity more natural and attractive.

Now, it is useful to identify several shortcomings in Kaluza's original approach.

- The physical meaning of the fifth dimension is not clear (Is it a genuine physical dimension or merely a mathematical device necessary to obtain unification in four dimensions?).
- The cylinder condition that accounts for unobservability of the fifth dimension does not have a natural explanation.
- The metric component $\hat{g}_{55}$ is eliminated from the action by imposing an *ad hoc* condition. This is not consistent from the point of view of the original theory, as we are thereby lacking the field equation $\widehat{R}_{55} = 0$.
- The theory does not account for the electroweak and strong interactions.

– The approach is purely classical; it does not address itself to the quantum nature of basic interactions.
– The obtained geometric unification of gravity and electromagnetism does not include matter.

Some of these objections are given from the point of view of our present understanding of the basic physical interactions. In the subsequent exposition of the five-dimensional theory we shall show how the first three objections can be eliminated. Inclusion of the weak and strong interactions demands a transition to higher dimensions, and will be discussed in section 10.3. The real explanation of the last two objections lies outside the scope of this book, although some aspects of the problems will be mentioned in passing.

Ground state and stability

An important step in studying physical properties of a dynamical system is to find its *ground state* or *vacuum*. Then, the low-energy physics is obtained by introducing physical fields as excitations around the ground state.

It might seem natural to assume that the ground state of the theory (10.1) is the five-dimensional Minkowski space M_5. However, such an assumption, if true, shows that the theory (10.1) is wrong, since we certainly know that the space we live in is not even close to M_5. Kaluza's basic idea can be expressed by stating that the true vacuum state is the product of four-dimensional Minkowski space M_4 with a circle S_1:

$$(V_5)_0 = M_4 \times S_1 .$$

It is assumed that the circumference of this circle is very small so that the fifth dimension is unobservable in standard experiments.

In a large class of field theories the ground state is determined as the stable solution of the field equations with the *lowest energy*. However, this definition is not relevant for gravitational theories where the concept of energy is very subtle and depends on the boundary conditions. We shall now discuss this problem in more detail.

Classical vacuum. In classical field theories, the positivity of the energy plays an important role for the stability of the vacuum. The energy is usually of the form $E = \int d^3x\, T_{00}$, where $T_{00} = 0$ in the vacuum, and $T_{00} > 0$ for all other configurations. Conservation of energy then implies the stability of the vacuum: there is no other configuration to which it could make a transition, respecting the conservation of energy. Besides the true vacuum state, a field theory may have other classically stable solutions, for which the energy has only a *local minimum*.

Consider now, as an illustrative example, the simple scalar field theory in 1+1 dimensions: $\mathcal{L} = \frac{1}{2}(\partial_t\phi)^2 - \frac{1}{2}(\partial_x\phi)^2 - V(\phi)$. The Hamiltonian is

$$H = \int_{-\infty}^{+\infty} dx\, [\tfrac{1}{2}(\partial_t\phi)^2 + \tfrac{1}{2}(\partial_x\phi)^2 + V(\phi)].$$

Without loss of generality, the condition that the energy is bounded from below may be replaced by the *positive energy* condition: $E \geq 0$. This means that the potential has to be a non-negative function of ϕ, $V(\phi) \geq 0$. Moreover, we restrict ourselves to *finite energy* configurations, which implies the following *boundary conditions*:

$$\partial_t \phi, \quad \partial_x \phi, \quad V(\phi) \to 0 \qquad \text{as } x \to \pm\infty.$$

Consequently, the field ϕ tends to a constant value at infinity, and this value must be one of the absolute minima of the potential $V(\phi)$:

$$x \to \pm\infty : \qquad \phi \to \phi_\pm \quad V(\phi_\pm) = 0.$$

The constants ϕ_- and ϕ_+ need not be equal.

The classical vacuum ϕ_0 is the solution of the field equations that has the lowest possible energy, $E_0 = 0$. It follows from the form of the Hamiltonian that ϕ_0 must satisfy the relations

$$\partial_t \phi_0 = 0 \qquad \partial_x \phi_0 = 0 \qquad V(\phi_0) = 0.$$

Thus, ϕ_0 is a constant field for which $V(\phi)$ has an absolute minimum. In general, the classical vacuum is not unique.

The classical vacuum is a stable, static solution of the field equations. Its dynamical role becomes clearer if we analyse all stable and static solutions of the theory. A static solution $\phi_0(x)$ is a field configuration for which the potential $W[\phi] = \int \mathrm{d}x \, [\frac{1}{2}(\partial_x \phi)^2 + V(\phi)]$ has an extremum:

$$\frac{\delta W[\phi]}{\delta \phi} = -\partial_x^2 \phi + V'(\phi) = 0.$$

We should note that for static solutions $W[\phi] = H[\phi]$. *Classically stable* static solutions correspond to the minima of $W[\phi]$, i.e. to the minima of the static energy functional. Thus, the classical stability of a solution is tested by its behaviour under small perturbations (perturbative stability).

When $W[\phi]$ has a *single* minimum, there is only one static solution, the unique classical vacuum. More interesting situations arise when $W[\phi]$ has *several* minima.

When $W[\phi]$ has an *absolute* but degenerate minimum, it is called the degenerate classical vacuum. A simple example is given by the non-trivial ϕ^4 theory: $V(\phi) = (\lambda/4)(\phi^2 - v^2)^2$. The potential $W[\phi]$ has two absolute minima, $\phi_0 = \pm v$, so that the classical vacuum is degenerate.

A solution ϕ_0 may only be a *local* minimum of $W[\phi]$. Returning to the previous example, we note that for any finite energy solution there are four possible types of boundary conditions: (ϕ_-, ϕ_+) may take the values $(-v, -v)$, $(+v, +v)$, $(-v, +v)$ and $(+v, -v)$. Accordingly, the set of all finite energy solutions can be broken into four sectors. Classical vacuum solutions, $\phi_0 = -v$ and $\phi_0 = +v$, belong to the first two sectors (vacuum sectors). The remaining two sectors are topologically non-trivial (non-perturbative sectors), and the related solutions are called the kink and antikink, respectively. They correspond to the local minima of $W[\phi]$, and each of them has the lowest energy in its own sector (Felsager 1981, Rajaraman 1982).

Classical stability. Now, we are going to consider small fluctuations around a static solution $\phi_0(x)$:

$$\phi(t,x) = \phi_0(x) + \eta(t,x) \qquad |\eta(t,x)| \ll |\phi_0(x)|.$$

Using the series expansion $\eta(t,x) = \sum_k \eta_k(x)\exp(-\mathrm{i}\omega_k t)$ over a complete set of static solutions $\eta_k(x)$, the field equation for $\phi(t,x)$ takes the form

$$[-\partial_x^2 + V''(\phi_0)]\eta_k(x) = \omega_k^2 \eta_k(x).$$

Since the operator $[-\partial_x^2 + V''(\phi_0)]$ is Hermitian, the modes η_k are orthogonal, and ω_k^2 is real. For the stability of ϕ_0 we must have $\omega_k^2 \geq 0$. Indeed, the existence of a negative ω_k^2 (imaginary ω_k) would imply the appearance of the exponential factors $\exp(\pm|\omega_k|t)$ which, for some value of time, become large, and contradict the stability.

The stability condition can be expressed as a condition on the energy of the system. Let $\phi_0(x)$ be a static solution for which the Hamiltonian $H[\phi] = \int \mathrm{d}x\,[\frac{1}{2}(\partial_x\phi)^2 + V(\phi)]$ has a minimum. By varying $\phi_0(x)$ in the set of static functions, $\delta\phi_0(x) = \varepsilon\eta(x)$, the variation of energy takes the form

$$\delta H[\phi_0] = \tfrac{1}{2}\varepsilon^2 \int \mathrm{d}x\, \eta(x)[-\partial_x^2 + V''(\phi_0)]\eta(x).$$

Using, now, the series expansion of $\eta(x)$ in terms of the orthogonal modes, $\eta(x) = \sum_k a_k\eta_k(x)$, this relation reads as

$$\delta H[\phi_0] = \tfrac{1}{2}\varepsilon^2 \sum_k a_k^2\omega_k^2.$$

Thus, the requirement $\omega_k^2 \geq 0$ can be translated into $\delta H[\phi_0] \geq 0$.

The condition $\omega_k^2 \geq 0$ is necessary but not sufficient for stability. If all ω_k are real and not zero, which means that all $\omega_k^2 > 0$ (there is a strong local minimum of the energy), then ϕ_0 is a classically stable solution. If all ω_k are real but some of them are zero, the solution ϕ_0 may be either stable or unstable. Thus, the appearance of zero modes complicates the formalism, and demands additional criteria for stability (Mecklenburg 1981).

Zero modes are solutions $\eta_k(x)$ with $\omega_k = 0$. They are often related to the existence of different static configurations which have the same energy and are connected by a continuous symmetry. In that case any symmetry transformation of the state $\phi_0(x)$ is the zero mode deformation. Consider, for example, the translational invariance of the scalar field theory. It implies that configurations $\phi_0(x)$ and $\phi_0(x+\varepsilon) = \phi_0(x) + \varepsilon\partial_x\phi_0(x)$ have the same energy, i.e. the perturbation along $\partial_x\phi_0(x)$ does not change the energy of the state. To see that explicitly, we should differentiate the equation of motion for $\phi_0(x)$ with respect to x:

$$[-\partial_x^2\phi_0(x) + V''(\phi_0)]\partial_x\phi_0(x) = 0.$$

For stability, the translation mode should be the lowest one. We can often use general arguments to show that the eigenvalue $\omega_k = 0$ cannot be the lowest, which implies instability.

This analysis applies to all static solutions, including both local and absolute minima of the potential, and can be generalized to more complex dynamical systems (Rajaraman 1982, Mecklenburg 1983).

Semiclassical instability. Standard treatments of quantum field theory are usually based on perturbation theory. However, perturbation theory is not an ideal small-coupling approximation, as there are a lot of interesting physical phenomena in the small-coupling regime that are not covered by perturbation theory. A typical phenomenon of this kind is quantum tunnelling, which is not seen in any order of perturbation theory.

There are computational methods based on the *semiclassical* (small $\hbar$) approximation, in which the functional integral is dominated by the stationary points of the action. Many nonlinear classical field theories possess non-trivial stable solutions corresponding to the local minima of the energy functional. If two local minima are separated by a finite potential barrier, semiclassical analysis shows that the system can *tunnel* from one configuration into the other. The tunnelling is associated with a localized solution of the corresponding Euclidean field theory that interpolates between the two local minima, called the instanton (Coleman 1985).

Semiclassical instability in field theory may be investigated by instanton methods in the following way:

(a) First, we find a local minimum solution of the Euclidean field theory.
(b) Then, we look for a bounce solution—a solution of the Euclidean theory that asymptotically approaches the local minimum configuration.
(c) Finally, we check the stability of the bounce solution against small perturbations. If the bounce solution is unstable, its contribution to the energy of the ground state is imaginary, which indicates the instability of the ground state (Mecklenburg 1983).

The KK ground state. We now want to apply these considerations to gravity and to find whether the space $M_4 \times S_1$ could be the ground state of the five-dimensional KK theory. The first step would be to compare the energy of $M_4 \times S_1$ to the energy of M_5. However, the definition of the gravitational energy depends on the boundary conditions, so that this comparison would be meaningless.

A field theory is defined not only by an action but also by boundary conditions. General covariance describes the symmetry of the action but the physical symmetry is determined by the symmetry of the boundary conditions. The choice of boundary conditions defines the asymptotic symmetry of spacetime and, consequently, the associated concept of time; this time is, then, used to introduce the energy for generally covariant dynamical systems (see chapter 6). Therefore,

> *we can compare energetically only those solutions that have the same boundary conditions.*

Thus, for instance, we may compare the energy of all solutions that are asymptotically Minkowskian but no more.

With these limitations in mind, we can now understand the real meaning of a typical positive energy theorem. Consider, for instance, a class of solutions of Einstein's equations that are asymptotically Minkowskian. If we can show that all solutions in this class, which are different from M_4, have positive energy and only M_4 has zero energy, then it follows that the Minkowski space is stable (there are several proofs of this theorem in GR).

Since the definition of the gravitational energy depends on the boundary conditions, and the boundary conditions for M_5 and $M_4 \times S_1$ are different, 'a comparison between them is meaningless, like comparing zero apples and zero oranges' (Witten 1981a).

> *The only criterion for the choice of the gravitational ground state remains stability.*

Consequently, in KK theory we should impose the requirement that $M_4 \times S_1$ should be stable, both classically and semiclassically.

The space $M_4 \times S_1$ is *classically stable* configuration. As we shall see, the spectrum of small excitations consists of a finite number of massless modes (a graviton, a photon and a scalar field), expected for symmetry reasons and an infinite number of massive modes. All frequencies are real and exponentially growing modes are absent.

In order to explore the semiclassical stability of $M_4 \times S_1$, we first define the Euclidean KK ground state by $\mathrm{d}s^2 = \mathrm{d}\tau^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 + \mathrm{d}\phi^2$, or, after introducing spherical coordinates, $\mathrm{d}s^2 = \mathrm{d}\rho^2 + \rho^2\,\mathrm{d}\Omega^2 + \mathrm{d}\phi^2$. Here, ϕ is a periodic variable, $0 \leq \phi \leq 2\pi r$, and $\mathrm{d}\Omega^2$ is the line element on the three-dimensional sphere S_3. Then, we observe that there is a bounce solution specified by

$$\mathrm{d}s^2 = \frac{\mathrm{d}\rho^2}{1 - a^2/\rho^2} + \rho^2\,\mathrm{d}\Omega^2 + (1 - a^2/\rho^2)\,\mathrm{d}\phi^2$$

where $a > 0$ is a constant. For large values of ρ this solution approaches the KK vacuum. To obtain a non-singular behaviour at $\rho = a$, ϕ has to be periodic with period $2\pi/a$, and ρ is restricted to run from a to infinity. Studying small fluctuations around this solution we find that it represents an instability of $M_4 \times S_1$.

By an appropriate Minkowskian continuation, we can find a solution into which the KK vacuum decays. That space (a) contains a hole in it and differs topologically from $M_4 \times S_1$, (b) has the same asymptotic behaviour and (c) its energy is zero. Therefore, the KK vacuum decays in the process of the spontaneous formation of a hole, going over into a final state with the same energy. However, the interpretation that this bounce solution describes an instability of the KK vacuum is correct only if spaces with topologies different from $M_4 \times S_1$ are allowed dynamically, which is not quite clear. If such spaces are not considered, the positive energy theorem also holds for the KK vacuum. Moreover, the effect of the instability could be eliminated by adding elementary fermions (Witten 1982).

In our further exposition we shall leave aside these interesting aspects of the five-dimensional KK theory and continue to analyse its classical properties. It should be stressed, however, that the question of the stability of the ground state has to be clarified in every serious attempt to construct a realistic KK theory.

10.2 Five-dimensional KK theory

In the preceding section we introduced the basic ideas of five-dimensional KK theory and shed light on the important role of the ground state for its physical interpretation. Now, we are going to give a systematic account of the geometric and physical structure of the theory (see, e.g., Zee 1981, Salam and Strathdee 1982, Gross and Perry 1983, Mecklenburg 1983, Duff *et al* 1986, Bailin and Love 1987).

Five-dimensional gravity and effective theory

We begin by discussing the general structure of the five-dimensional gravity and its effective form in $d = 4$.

Gravity in five dimensions. Let X_5 be a five-dimensional differentiable manifold with local coordinates z^M. At each point of this manifold we define the tangent space T_5 with a local Lorentz basis (pentad) $\hat{\boldsymbol{e}}_I$: $\hat{\boldsymbol{e}}_I \hat{\boldsymbol{e}}_J = \eta_{IJ}$, where $\eta_{IJ} = (+, -, -, -; -)$. Here, $(I, J, \ldots)$ are local Lorentz indices, and $(M, N, \ldots)$ are coordinate indices. An arbitrary vector of the coordinate basis $\hat{\boldsymbol{e}}_M$ can be expressed in the Lorentz basis as $\hat{\boldsymbol{e}}_M = b^I{}_M \hat{\boldsymbol{e}}_I$, and the inverse relation reads as $\hat{\boldsymbol{e}}_I = h_I{}^M \hat{\boldsymbol{e}}_M$.

In each tangent space T_5 we can define the metric in the usual way:

$$\hat{g}_{MN} \equiv \hat{\boldsymbol{e}}_M \hat{\boldsymbol{e}}_N = b^I{}_M b^J{}_N \eta_{IJ}.$$

If, in addition, we introduce an antisymmetric, metric compatible connection $\hat{A}^{IJ}{}_M = -\hat{A}^{JI}{}_M$, the theory takes the form of PGT in $d = 5$, and $(X_5, \hat{g}, \hat{A})$ is a Riemann–Cartan space U_5.

The theory of gravity in U_5 is invariant with respect to local translations (general coordinate transformations) and local Lorentz rotations $SO(1,4)$. From gauge potentials $b^I{}_M$ and $\hat{A}^{IJ}{}_M$ we can construct the corresponding field strengths: the torsion $\widehat{T}^I{}_{MN}$ and the curvature $\widehat{R}^{IJ}{}_{MN}$. We assume that the dynamics of the gravitational field is determined by the simple action

$$I_{\mathrm{G}} = -\frac{1}{2\hat{\kappa}}\int \mathrm{d}^5z\, \hat{b}\widehat{R}(\hat{A}) \tag{10.7}$$

where $\hat{b} = \det(b^I{}_M) = \sqrt{\hat{g}}$. Matter fields can be coupled to gravity by generalizing the well-known four-dimensional structures (at the expense of losing the simplicity of Kaluza's original idea).

Example 1. The real massless scalar field in U_5 is described by the action

$$I_{\mathrm{S}} = \int \mathrm{d}^5z\, \sqrt{\hat{g}}(-\tfrac{1}{2}\varphi\widehat{\Box}\varphi).$$

By varying $I_{\mathrm{G}} + I_{\mathrm{S}}$ with respect to $\hat{A}$ we obtain $\widehat{T}^I{}_{MN} \equiv \nabla_M b^I{}_N - \nabla_N b^I{}_M = 0$. This algebraic equation can be explicitly solved for $\hat{A}$:

$$\hat{A}^{IJ}{}_M = \Delta^{IJ}{}_M \equiv \tfrac{1}{2}(C^{IJK} - C^{KIJ} + C^{JKI})b_{KM}$$

where $C^I{}_{MN} = \partial_M b^I{}_N - \partial_N b^I{}_M$. Going into coordinate indices $\Delta^{IJ}{}_M$ becomes the Christoffel connection $\hat{\Gamma}^M_{NR}$. Thus, the condition $\widehat{T} = 0$ converts U_5 into the Riemann space V_5. By replacing $\hat{A} = \Delta$ into the original action we obtain an equivalent formulation of the theory (the second order formalism).

In five dimensions the Dirac field is just a four-component spinor. The dynamics of the massless Dirac field in U_5 is determined by

$$I_{\mathrm{D}} = \int \mathrm{d}^5z\, \tfrac{1}{2}\mathrm{i}\hat{b}\bar{\psi}\gamma^K h_K{}^M \nabla_M\psi + \mathrm{H.C.}$$

where $\gamma^K = (\gamma^k, \gamma^5)$ are the $d = 5$ Dirac matrices, $\nabla_M\psi = (\partial_M + \frac{1}{2}A^{IJ}{}_M\sigma_{IJ})\psi$, and H.C. denotes the Hermitian conjugate term. The equations of motion obtained from $I_{\mathrm{G}} + I_{\mathrm{D}}$ can be solved for $\hat{A}$:

$$\hat{A}^{IJ}{}_M = \Delta^{IJ}{}_M + K^{IJ}{}_M$$

where the contortion tensor K is bilinear in $\bar{\psi}$ and ψ. The connection is not Riemannian, and we have here an analogue of EC theory in U_4. The replacement $\hat{A} = \Delta$ in the original action does not give an equivalent theory. However, this theory by itself represents a consistent theory of the Dirac field in Riemann space V_5.

Ground state. As we have seen, it makes no sense to compare asymptotically different gravitational solutions on energetic grounds, hence there is no satisfying criterion for the choice of the ground state of U_5. We shall, therefore, adopt Kaluza's ansatz by itself,

$$(U_5)_0 = M_4 \times S_1 \tag{10.8a}$$

and explore the consequences.

After introducing local coordinates $z^M = (x^\mu, y)$, with y being an angle that parametrizes S_1, the metric of the ground state takes the form

$$\hat{g}^0_{MN} = \eta_{MN} = \begin{pmatrix} \eta_{\mu\nu} & 0 \\ 0 & -1 \end{pmatrix}. \tag{10.8b}$$

Note that it has the same form as the five-dimensional Minkowskian metric, since M_5 and $M_4 \times S_1$ are locally isometric. This metric is a solution of the classical field equations, $\widehat{R}_{MN} = 0$. Studying small oscillations around (10.8) we find that there are no exponentially growing modes, hence the ground state $M_4 \times S_1$ is classically stable. Semiclassical analysis indicates that $M_4 \times S_1$ may be unstable.

The form of the ground state $M_4 \times S_1$ can be understood as a sort of spontaneous symmetry breaking. The symmetry of M_5 is the five-dimensional Poincaré group P_5, while the symmetry of $M_4 \times S_1$ is only $P_4 \times U(1)$. The possible maximal symmetry of the ground state P_5 is spontaneously broken to $P_4 \times U(1)$. Since S_1 is a compact manifold, the emergence of the ground state $M_4 \times S_1$ is called *spontaneous compactification*.

We know that GR in $d = 5$ with the ground state M_5 can be understood as a theory based on the local P_5 symmetry. In a similar way, KK theory with the ground state $M_4 \times S_1$ is a theory based on the local $P_4 \times U(1)$ symmetry. The local P_4 is responsible for the presence of gravity in $d = 4$, while local $U(1)$ describes electromagnetism. This is why KK theory can unify gravity and electromagnetism.

The harmonic expansion. Having found the ground state, we can now expand every dynamical variable around its ground state value and determine the spectrum of excitations. Harmonic expansion is introduced to describe these excitations. Since the fifth dimension is a circle, we have $y = r\theta$, where r is a constant and θ an angle, $0 \le \theta \le 2\pi$. Any dynamical variable defined on $M_4 \times S_1$ can be expanded in terms of the complete set of harmonics on S_1:

$$\varphi(x, y) = \sum_n \varphi_n(x) Y_n(y) \tag{10.9a}$$

where the Y_n are orthonormal eigenfunctions of the operator $\Box_y = -\partial_y^2$:

$$\begin{aligned} &Y_n(y) = (2\pi r)^{-1/2} \mathrm{e}^{-\mathrm{i}ny/r} \\ &\Box_y Y_n(y) = (n^2/r^2) Y_n(y) \qquad \int \mathrm{d}y\, Y_n^*(y) Y_m(y) = \delta_{nm}. \end{aligned} \tag{10.9b}$$

In five dimensions this is the Fourier expansion of the function $\varphi(x, y)$ periodic in $y \in S_1$, with the coefficients that depend upon x.

Example 2. The physical spectrum of excitations of the five-dimensional metric can be determined by expanding $\hat{g}_{MN}$ around its ground state value,

$$\hat{g}_{MN}(x, y) = \eta_{MN} + h_{MN}(x, y)$$

and using the equations of motion. If we choose an appropriate gauge condition the field equations $\widehat{R}_{MN} = 0$ take the form

$$\widehat{\Box} h_{MN}(x, y) = 0$$

where $\widehat{\Box} \equiv \eta^{MN}\partial_M\partial_N = \Box_x + \Box_y$. Thus, h_{MN} are massless fields in five dimensions. Now, we can make the Fourier expansion $h_{MN}(x, y) = \sum_n h^n_{MN}(x)Y_n(y)$ and obtain the equations for the Fourier components:

$$(\Box_x + n^2/r^2)h^n_{MN}(x) = 0.$$

We have here an infinite number of four-dimensional fields $h^n_{MN}(x)$. The nth mode of the metric has the (four-dimensional) mass squared $m_n^2 = (n\hbar c/r)^2$, that stems from the kinetic energy of the motion along the fifth dimension. All the modes with $n \neq 0$ are massive and $h^0_{MN}(x)$ is the only massless mode. The massless mode reflects the existence of local symmetries (to be discussed later) and $m_n^2 > 0$, so that there are no exponentially growing modes.

If r is of the order of the Planck length, $l_P = (\hbar G/c^3)^{1/2} = 1.6 \times 10^{-33}$ cm, the modes with $n \neq 0$ have masses of order $\hbar c/r \approx 10^{19}$ GeV. At energies well below the Planck energy, these modes are very hard to excite. If we restrict our considerations to the low energy sector of the theory, $E \ll \hbar c/r$, and neglect all massive modes, the metric $\hat{g}_{MN}$ becomes *independent* of the fifth coordinate. This gives a clear meaning to the cylinder condition.

In order to express the complete five-dimensional theory in an *equivalent* four-dimensional form, we should Fourier expand all the dynamical variables in the action, and integrate over the fifth coordinate.

> *The information about the fifth dimension is completely contained in the presence of all modes of the original fields in the effective four-dimensional action.*

Very often, we limit our considerations to the zero mode contribution, which is independent of the fifth dimension. In doing so, we expect the higher excitations to be too massive to be accessible at present energies. However, more detailed investigations show that discarding the massive modes is not always consistent with the higher-dimensional field equations (Duff *et al* 1986).

Choosing dynamical variables

We now introduce a suitable parametrization of the five-dimensional metric, so that the usual gravity and electromagnetism are directly included in the massless sector of the effective four-dimensional theory (Zee 1981, Salam and Strathdee 1982).

The metric. We want to investigate the general structure of the space U_5, assuming that the ground state is $M_4 \times S_1$. Let $\hat{\boldsymbol{e}}_M = (\hat{\boldsymbol{e}}_\mu, \hat{\boldsymbol{e}}_5)$ be the coordinate basis of the tangent vectors, associated with local coordinates $z^M = (x^\mu, y)$. The metric $\hat{g}_{MN} = \hat{\boldsymbol{e}}_M \hat{\boldsymbol{e}}_N$ has the general form

$$\hat{g}_{MN}(x, y) = \begin{pmatrix} \hat{g}_{\mu\nu} & \hat{g}_{\mu 5} \\ \hat{g}_{5\nu} & \hat{g}_{55} \end{pmatrix}.$$

We assume that the space U_5 has a *layered structure*, as represented in figure 10.1: for each fixed point x^μ in V_4 there is a (one-dimensional) manifold V_1 (layer or fibre), and the coordinate y labels the points in V_1. This means that U_5 locally has the structure of the Cartesian product $U_4 \times V_1$, which is not necessarily true globally. Two neighbouring points (x, y) and $(x, y + \mathrm{d}y)$ in a layer V_1 define an infinitesimal displacement $(0, \mathrm{d}y) = \mathrm{d}y\hat{\boldsymbol{e}}_5$, the square of which, $\mathrm{d}s^2 = \hat{g}_{55}\,\mathrm{d}y^2$, defines the metric in V_1: $g_{55} = \hat{g}_{55}$.

Next, we define the spacetime U_4 by demanding that each displacement in it should be *orthogonal* to a given layer V_1. The motivation for this choice is found in a simple physical interpretation: any displacement in spacetime should be orthogonal to the internal space. A displacement $(\mathrm{d}x, 0) = \mathrm{d}x^\mu \hat{\boldsymbol{e}}_\mu$ is not orthogonal to $(0, \mathrm{d}y)$ since $\hat{\boldsymbol{e}}_\mu \hat{\boldsymbol{e}}_5 = \hat{g}_{\mu 5} \neq 0$. Hence, the four-dimensional space the points of which are $(x, y = \text{constant})$ cannot be the physical spacetime. The desired displacement orthogonal to V_1 has the form $(\mathrm{d}x, \Delta y) = \mathrm{d}x^\mu \hat{\boldsymbol{e}}_\mu + \Delta y \hat{\boldsymbol{e}}_5$, where Δy is determined by the orthogonality condition:

$$(\mathrm{d}x^\mu \hat{\boldsymbol{e}}_\mu + \Delta y \hat{\boldsymbol{e}}_5)\hat{\boldsymbol{e}}_5 = 0 \quad \Rightarrow \quad \Delta y = -g^{55}\hat{g}_{5\mu}\,\mathrm{d}x^\mu.$$

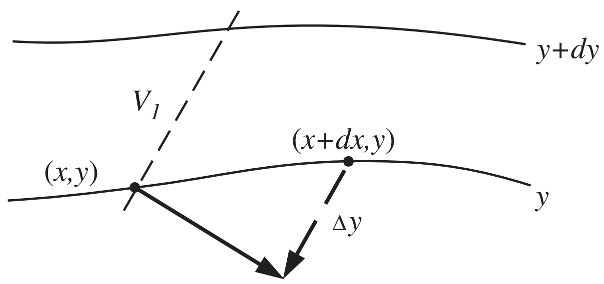

Figure 10.1. The layered structure of U_5.

Here, the quantity g^{55} is the inverse of $g_{55} = \hat{g}_{55}$, and should be clearly distinguished from $\hat{g}^{55}$. The length of the displacement vector $(\mathrm{d}x, \Delta y)$ can be computed to be

$$(\mathrm{d}x^\mu, \Delta y)\hat{g}_{MN}(\mathrm{d}x^\nu, \Delta y) = (\hat{g}_{\mu\nu} - g^{55}\hat{g}_{\mu 5}\hat{g}_{\nu 5})\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu \equiv g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu$$

where the last line defines the metric $g_{\mu\nu}$ of spacetime U_4.

If we introduce the notation

$$\hat{g}_{55} = \phi_{55} \qquad \hat{g}_{\mu 5} = \hat{g}_{55}B^5_\mu$$

and then, for simplicity, omit the indices 5 in ϕ_{55} and B^5_μ, the metric of U_5 takes the form

$$\hat{g}_{MN}(x,y) = \begin{pmatrix} g_{\mu\nu} + \phi B_\mu B_\nu & \phi B_\mu \\ \phi B_\nu & \phi \end{pmatrix} \tag{10.10a}$$

whereupon we find

$$\hat{g}^{MN}(x,y) = \begin{pmatrix} g^{\mu\nu} & -B^\mu \\ -B^\nu & \phi^{-1} + B_\lambda B^\lambda \end{pmatrix}. \tag{10.10b}$$

This result is an expression of the local orthogonality of V_1 and U_4, and leads to the following form of the interval:

$$\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu + g_{55}(\mathrm{d}y + B_\mu\,\mathrm{d}x^\mu)(\mathrm{d}y + B_\nu\,\mathrm{d}x^\nu). \tag{10.10c}$$

The pentad. These considerations can be expressed in a suitable form by using the formalism of orthonormal frames. We observe that the first four vectors of a Lorentz frame $\hat{\boldsymbol{e}}_I = (\hat{\boldsymbol{e}}_i, \hat{\boldsymbol{e}}_{\bar{5}})$ define the Lorentz frame in T_4, the tangent space of U_4. Each vector of the coordinate basis $\hat{\boldsymbol{e}}_M$ can be expressed as $\hat{\boldsymbol{e}}_M = b^i{}_M\hat{\boldsymbol{e}}_i + b^{\bar{5}}{}_M\hat{\boldsymbol{e}}_{\bar{5}}$. Applying this expansion to the vector $\hat{\boldsymbol{e}}_5$ in the layer V_1, and using the local orthogonality of V_1 and U_4, we conclude that the expansion of $\hat{\boldsymbol{e}}_5$ cannot contain any of the vectors $\hat{\boldsymbol{e}}_i$ in U_4:

$$\hat{\boldsymbol{e}}_\mu = b^i{}_\mu\hat{\boldsymbol{e}}_i + b^{\bar{5}}{}_\mu\hat{\boldsymbol{e}}_{\bar{5}} \qquad \hat{\boldsymbol{e}}_5 = b^{\bar{5}}{}_5\hat{\boldsymbol{e}}_{\bar{5}}.$$

Thus, the local orthogonality condition takes the form

$$b^i{}_5 = 0 \quad \Longleftrightarrow \quad b^I{}_M(x,y) = \begin{pmatrix} b^i{}_\mu & 0 \\ b^{\bar{5}}{}_\mu & b^{\bar{5}}{}_5 \end{pmatrix}. \tag{10.11a}$$

From this we easily find the inverse pentad:

$$h_I{}^M(x,y) = \begin{pmatrix} h_i{}^\mu & h_i{}^5 \\ 0 & h_{\bar{5}}{}^5 \end{pmatrix} \tag{10.11b}$$

where $h_i{}^\mu$ is the inverse of $b^i{}_\mu$, $h_{\bar{5}}{}^5 b^{\bar{5}}{}_5 = 1$, and $h_i{}^5$ satisfies the condition $b^i{}_\mu h_i{}^5 + b^{\bar{5}}{}_\mu h_{\bar{5}}{}^5 = 0$.

In terms of the pentad components, we can construct the metric of U_5 in the standard way:

$$
\begin{aligned}
\hat{g}_{\mu\nu} &= b^i{}_\mu b^j{}_\nu \eta_{ij} + b^{\bar{5}}{}_\mu b^{\bar{5}}{}_\nu \eta_{55} = g_{\mu\nu} - b^{\bar{5}}{}_\mu b^{\bar{5}}{}_\nu \\
\hat{g}_{\mu 5} &= b^{\bar{5}}{}_\mu b^{\bar{5}}{}_5 \eta_{55} = -b^{\bar{5}}{}_\mu b^{\bar{5}}{}_5 \\
\hat{g}_{55} &= b^{\bar{5}}{}_5 b^{\bar{5}}{}_5 \eta_{55} = -b^{\bar{5}}{}_5 b^{\bar{5}}{}_5 .
\end{aligned}
$$

After making the identification $B_\mu = h_{\bar{5}}{}^5 b^{\bar{5}}{}_\mu$, we recognize here the result (10.10*a*).

Using the basis $(\hat{\boldsymbol{e}}_i, \hat{\boldsymbol{e}}_{\bar{5}})$, we can define the local Lorentz coordinate system $(\xi^i, \xi^{\bar{5}})$:

$$
\mathrm{d}\xi^i = b^i{}_\mu \,\mathrm{d}x^\mu \qquad \mathrm{d}\xi^{\bar{5}} = b^{\bar{5}}{}_5 \,\mathrm{d}y + b^{\bar{5}}{}_\mu \,\mathrm{d}x^\mu = b^{\bar{5}}{}_5(\mathrm{d}y + B_\mu \,\mathrm{d}x^\mu)
$$

in which the interval has the form $\mathrm{d}s^2 = \eta_{ij}\,\mathrm{d}\xi^i\,\mathrm{d}\xi^j + \eta_{55}(\mathrm{d}\xi^{\bar{5}})^2$.

There is another, very convenient, choice of basis obtained by taking $(\boldsymbol{E}_\mu, \boldsymbol{E}_5) = (b^i{}_\mu \hat{\boldsymbol{e}}_i, \hat{\boldsymbol{e}}_5)$, i.e.

$$
\hat{\boldsymbol{e}}_\mu = \boldsymbol{E}_\mu + B_\mu \boldsymbol{E}_5 \qquad \hat{\boldsymbol{e}}_5 = \boldsymbol{E}_5
$$

which is naturally related to the local orthogonality of V_1 and U_4. This basis defines the local coordinate system (X^μ, Y):

$$
\mathrm{d}X^\mu = \mathrm{d}x^\mu \qquad \mathrm{d}Y = \mathrm{d}y + B_\mu \,\mathrm{d}x^\mu
$$

in which the metric is block-diagonal:

$$
\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}X^\mu\,\mathrm{d}X^\nu + g_{55}\,\mathrm{d}Y^2 .
$$

It is important to realize that this is not the coordinate basis, which influences the calculations of various geometric objects (Toms 1984).

After introducing the metric structure of U_5 by the condition (10.11*a*), we can also obtain the related $(4+1)$ splitting of the Christoffel connection.

Residual symmetry. The original five-dimensional action (10.7) is invariant under general coordinate transformations and local Lorentz rotations. The transformation law of $b^i{}_\mu$ under general coordinate transformations has the form

$$
b'^i{}_\mu = \frac{\partial x^\nu}{\partial x'^\mu} b^i{}_\nu + \frac{\partial y}{\partial x'^\mu} b^i{}_5 .
$$

If we wish to preserve the fundamental condition $b^i{}_5 = 0$ which defines the layered structure of U_5, we have to allow only those coordinate transformations in which x' does not depend on y:

$$
x' = x'(x) \qquad y' = y'(x, y). \tag{10.12}
$$

The related transformation rules of the metric components $(g_{\mu\nu}, B_\mu, \phi)$ are:

$$
\begin{aligned}
g'_{\mu\nu} &= \frac{\partial x^\lambda}{\partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\nu} g_{\lambda\rho} \qquad \phi' = \frac{\partial y}{\partial y'}\frac{\partial y}{\partial y'}\phi \\
B'_\mu &= \frac{\partial x^\lambda}{\partial x'^\mu}\left(\frac{\partial y'}{\partial y}B_\lambda - \frac{\partial y'}{\partial x^\lambda}\right).
\end{aligned}
\tag{10.13}
$$

If we recall the omitted indices 5 in ϕ and B_μ, these rules become much clearer:

(a) $g_{\mu\nu}$ is a tensor with respect to the indices (μ, ν);
(b) $\phi = \phi_{55}$ is a tensor with respect to $(5, 5)$;
(c) $B_\mu = B^5_\mu$ is not a tensor, as it transforms inhomogeneously, but in the special case $y' = y'(y)$ it becomes a tensor with respect to $(5, \mu)$.

The coordinate transformations (10.12) reduce to the form (10.3) if we demand that $\phi = \phi(x)$.

The massless sector of the effective theory

In our previous considerations, we defined the general geometric structure of the space U_5, in which the metric components are functions of all coordinates $z^M = (x^\mu, y)$. Now, we focus our attention on the truncated theory that contains only y-independent, zero modes.

Instead of discussing the general Riemann–Cartan space U_5, we restrict our attention to the important case $T = 0$, corresponding to the Riemann space V_5. Using metric (10.10) and assuming its independence of y, we find the following result for the Christoffel connection:

$$
\begin{aligned}
\hat\Gamma^\mu_{\nu\rho} &= \Gamma^\mu_{\nu\rho} - \tfrac12 (F^\mu{}_\nu B_\rho\phi + F^\mu{}_\rho B_\nu\phi + B_\nu B_\rho \partial^\mu\phi) \\
\hat\Gamma^\mu_{5\rho} &= -\tfrac12 (F^\mu{}_\rho\phi + B_\rho\partial^\mu\phi) \qquad \hat\Gamma^\mu_{55} = -\tfrac12\partial^\mu\phi \\
\hat\Gamma^5_{\nu\rho} &= \tfrac12(\nabla_\nu B_\rho + \nabla_\rho B_\nu) + \tfrac12 B^\lambda (F_{\lambda\nu}B_\rho\phi + F_{\lambda\rho}B_\nu\phi + B_\nu B_\rho\partial_\lambda\phi) \\
&\quad + \tfrac12\phi^{-1}(B_\rho\partial_\nu\phi + B_\nu\partial_\rho\phi) \\
\hat\Gamma^5_{5\rho} &= \tfrac12(B^\lambda F_{\lambda\rho}\phi + B_\rho B^\lambda\partial_\lambda\phi + \phi^{-1}\partial_\rho\phi) \qquad \hat\Gamma^5_{55} = \tfrac12 B^\lambda\partial_\lambda\phi.
\end{aligned}
\tag{10.14a}
$$

A straightforward (but lengthy) calculation leads to the curvature scalar:

$$
\begin{aligned}
\widehat{R} &= R - \tfrac14\phi F_{\mu\nu}F^{\mu\nu} - \phi^{-1}\Box_x\phi + \tfrac12\phi^{-2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi \\
&= R - \tfrac14\phi F_{\mu\nu}F^{\mu\nu} - \frac{2}{\sqrt{-\phi}}\Box_x\sqrt{-\phi}.
\end{aligned}
\tag{10.14b}
$$

Since $\sqrt{\hat g} = \sqrt{-g}\sqrt{-\phi}$, the expression $\sqrt{-\hat g}\widehat{R}$ does not depend on the fifth coordinate, and the integration over y in action (10.7) yields the following effective low-energy theory:

$$I_{\mathrm{G}}^{(0)} = -\frac{1}{2\kappa}\int \mathrm{d}^4x\,\sqrt{-g}\sqrt{-\phi}(R - \tfrac{1}{4}\phi F_{\mu\nu}F^{\mu\nu}) \qquad \kappa \equiv \hat{\kappa}/L \tag{10.15}$$

where we have dropped a four-divergence term. This result represents the zero mode contribution to the effective four-dimensional action, and should be compared to expression (10.6), which holds for $\phi = -1$.

This theory can be recognized as a variant of the Brans–Dicke theory, with $\sigma \equiv \sqrt{-\phi}$ identified as the massless scalar field coupled to gravity and electromagnetism. The scalar field effectively defines the strength of the gravitational interaction. While Brans–Dicke theory contains a term $\omega\partial_\mu\sigma\partial^\mu\sigma/\sigma$, with $\omega \geq 6$, here $\omega = 0$. Also, the coupling of the scalar σ to matter is completely fixed by the five-dimensional covariance, whereas in the Brans–Dicke theory this coupling is absent.

The presence of the multiplicative factor $\sqrt{-\phi}$ in the action does not correspond to the standard formulation, in which the coefficient of $\sqrt{-g}R$ is constant. The theory can be transformed into the usual form by a local Weyl rescaling of $g_{\mu\nu}$ and ϕ. Recall that the scalar curvature in V_4 transforms under Weyl rescaling as

$$g_{\mu\nu} = \lambda^{-1}\bar{g}_{\mu\nu} \qquad R(g) = \lambda[R(\bar{g}) + 3\bar{\Box}\ln\lambda - \tfrac{3}{2}\lambda^{-2}\bar{g}^{\mu\nu}\partial_\mu\lambda\partial_\nu\lambda].$$

Thus, if we define

$$g_{\mu\nu} = \lambda^{-1}\bar{g}_{\mu\nu} \qquad \phi = \lambda^{-1}\bar{\phi} \qquad \lambda \equiv (-\bar{\phi})^{1/3} \tag{10.16a}$$

the effective four-dimensional theory assumes the usual form

$$I_{\mathrm{G}}^{(0)} = -\frac{1}{2\kappa}\int \mathrm{d}^4x\,\sqrt{-\bar{g}}(\bar{R} - \tfrac{1}{4}\bar{\phi}\,F_{\mu\nu}F^{\mu\nu} - \tfrac{1}{6}\bar{\phi}^{-2}\bar{g}^{\mu\nu}\partial_\mu\bar{\phi}\partial_\nu\bar{\phi}) \tag{10.16b}$$

where we have dropped another four-divergence.

Note that the electromagnetic part of the action is not properly normalized. This is easily corrected by introducing

$$B_\mu = fA_\mu \qquad f^2 = 2\kappa \tag{10.16c}$$

where the condition $f^2/2\kappa = 1$ ensures the standard form of the Maxwell action for $\bar{\phi} = -1$. We choose f to be negative, by convention.

The final form of action (10.16b) is obtained by a local Weyl rescaling of $g_{\mu\nu}$ and ϕ, which is equivalent to the rescaling of the complete five-dimensional metric (10.10a):

$$\hat{g}_{MN}(x) = (-\bar{\phi})^{-1/3}\begin{pmatrix} \bar{g}_{\mu\nu} + \bar{\phi}B_\mu B_\nu & \bar{\phi}B_\mu \\ \bar{\phi}B_\nu & \bar{\phi} \end{pmatrix}. \tag{10.17}$$

This rescaling changes the form of the Christoffel symbol $\bar{\Gamma}$, as compared to result (10.14*a*) for Γ.

Thus, KK theory at low energies describes the massless scalar field $s = \ln(-\bar{\phi})/\sqrt{6\kappa}$ coupled to gravity and electromagnetism. While the gravitational coupling is constant, the electromagnetic coupling is proportional to $(-\bar{\phi})^{1/2} = \exp(\sqrt{3\kappa/2}\, s)$.

Always bearing in mind that the final form of the metric is given by equation (10.17), we shall omit the overbars for simplicity.

Dynamics of matter and the fifth dimension

The coupling of matter to five-dimensional gravity is fixed by the invariance properties of the theory. The dynamical behaviour of matter gives us a new insight into the physics associated with the fifth dimension (Salam and Strathdee 1982, Gross and Perry 1983, Mecklenburg 1983).

Classical particle. A classical point-particle in the five-dimensional gravitational field (10.17) has the action $I_{\mathrm{M}} = -m \int \mathrm{d}s$, and the equations of motion are five-dimensional geodesics:

$$\frac{\mathrm{d}^2 z^M}{\mathrm{d}\tau^2} + \hat{\Gamma}^M_{NR} \frac{\mathrm{d}z^N}{\mathrm{d}\tau} \frac{\mathrm{d}z^R}{\mathrm{d}\tau} = 0.$$

Very interesting information concerning the motion along the fifth dimension may be obtain by studying the metric restricted by the conditions $\phi = -1$ and $g_{\mu\nu} = \eta_{\mu\nu}$ (Vasilić 1989). By these assumptions we are neglecting the effects of the scalar and gravitational fields, so that the motion of the test particle is due merely to electromagnetism. If we also ignore every y dependence, we come to the following form of the metric:

$$\hat{g}_{MN}(x) = \begin{pmatrix} \eta_{\mu\nu} - B_\mu B_\nu & -B_\mu \\ -B_\nu & -1 \end{pmatrix} \tag{10.18}$$

where $B_\mu = f A_\mu$. The Christoffel connection is easily calculated from the relations (10.5*a*). Keeping only first order terms in the electromagnetic coupling for simplicity, we find

$$\begin{aligned} &\hat{\Gamma}^\mu_{\nu\rho} = \mathcal{O}_2 \qquad &&\hat{\Gamma}^5_{\nu\rho} = \tfrac{1}{2}(\partial_\nu B_\rho + \partial_\rho B_\nu) + \mathcal{O}_2 \\ &\hat{\Gamma}^\mu_{5\rho} = \tfrac{1}{2} F^\mu{}_\rho \qquad &&\text{remaining components} = \mathcal{O}_2 \end{aligned}$$

where $\mathcal{O}_2$ denotes second or higher order terms in B_μ. Using these expressions the geodesic equation takes the form (Vasilić 1989)

$$\begin{aligned} \frac{\mathrm{d}u^\mu}{\mathrm{d}\tau} + f F^\mu{}_\rho(A) u^\rho u^5 &= \mathcal{O}_2 \\ \frac{\mathrm{d}u^5}{\mathrm{d}\tau} + \tfrac{1}{2} f (\partial_\mu A_\nu + \partial_\nu A_\mu) u^\mu u^\nu &= \mathcal{O}_2. \end{aligned}$$

The second equation implies $u^5 = (u^5)_0 + \mathcal{O}_1$, where the constant $(u^5)_0$ represents the initial value of the particle's velocity in the direction of the fifth dimension. After this, the first equation yields

$$\frac{\mathrm{d}u^\mu}{\mathrm{d}\tau} + f(u^5)_0 F^\mu{}_\rho(A)u^\rho = \mathcal{O}_2.$$

Here $f = -\sqrt{2\kappa}$, and we recognize the well-known equation for the motion of an electrically charged particle in the electromagnetic field, if the charge of the particle q is identified with

$$q/m = \sqrt{2\kappa}(u^5)_0. \tag{10.19a}$$

Thus, we deduce a surprising conclusion concerning the nature of the electric charge:

> *The electric charge of a particle is a manifestation of its motion along the fifth dimension.*

Now, we use very simple arguments from quantum theory to show that the electric charge is *quantized*, i.e. that q is a multiple of some elementary charge e. If we apply the old Bohr–Sommerfeld quantization rule to the periodic motion along the fifth dimension, $(p_5)_0 \cdot 2\pi r = n \cdot 2\pi\hbar$, we deduce that $(p^5)_0 = (n/r)\hbar$, i.e.

$$q_n = ne \equiv n(\hbar\sqrt{2\kappa}/r). \tag{10.19b}$$

The radius of the fifth dimension is thus fixed by the elementary electric charge. From the known value of the elementary charge, we find that r is of the order of the Planck length:

$$\alpha = \frac{e^2}{4\pi\hbar c} \approx \frac{1}{137} \qquad r = \frac{2}{\sqrt{\alpha}}\sqrt{\frac{\kappa\hbar}{8\pi c}} \approx 3.7 \times 10^{-32}\ \text{cm}.$$

If we could find the radius r from some other considerations, this relation might be used to *calculate* the electric charge. The idea of calculating the elementary electric charge has attracted physicists' attention for a long time, but no satisfying solution has been found.

The real scalar field. We are now going to examine the fifth dimension using a real, scalar field $\varphi(x, y)$. If we apply the coordinate transformation $y' = y + \varepsilon$, the scalar field changes according to $\varphi'(x, y) = \varphi(x, y - \varepsilon)$. Making the Fourier expansion $\varphi(x, y) = \sum \varphi_n(x)Y_n(y)$, we find that the transformation law of the nth mode is

$$\varphi'_n(x) = \varphi_n(x)\mathrm{e}^{\mathrm{i}(n/r)\varepsilon}.$$

This means that $\varphi_n(x)$ has an electric charge $\sim n/r$.

The same result follows from the form of the effective electromagnetic interaction in V_4. The action of the free, real, scalar field in V_5 is given by

$$I_{\mathrm{S}} = \int \mathrm{d}^5 z \sqrt{\hat{g}}(-\tfrac{1}{2}\varphi\widehat{\Box}\varphi - \tfrac{1}{2}m^2\varphi^2). \tag{10.20a}$$

Using the metric (10.18) we find that

$$\widehat{\Box} = \eta^{\mu\nu}(\partial_\mu - B_\mu\partial_y)(\partial_\nu - B_\nu\partial_y) + \Box_y.$$

Substituting the Fourier expansion for $\varphi(x, y)$ in the action, and making use of the orthonormality of the harmonics Y_n to integrate over y, leads to

$$I_{\mathrm{S}}^{(0)} = \sum_n \int \mathrm{d}^4 x \sqrt{-g}\tfrac{1}{2}\varphi_n^*(-\nabla_n^2 + m_n^2)\varphi_n \tag{10.20b}$$

where

$$\nabla_n^2 = g^{\mu\nu}(\partial_\mu - \mathrm{i}q_n A_\mu)(\partial_\nu - \mathrm{i}q_n A_\nu)$$
$$q_n = n(\sqrt{2\kappa}/r) \qquad m_n^2 = m^2 + n^2/r^2.$$

The effective four-dimensional theory is seen to consist of one real field φ_0 of mass m, and an infinite number of complex scalar modes φ_n with masses m_n. If the original field in $d = 5$ is massless, the masses of the higher modes are $m_n = n/r$. Each higher mode has charge q_n and minimal coupling. The connection between the electric charge and the radius of the fifth dimension is the same as that predicted by point-particle dynamics. In the limit $r \to 0$ all modes but the zero mode become very heavy. It is usual to assume that all the massive modes can be discarded at sufficiently low energies, where the zero mode is dominant. We should note, however, that this may lead to inconsistencies (Duff *et al* 1986).

The Dirac field. The action for the massless Dirac field is defined as in example 1:

$$I_{\mathrm{D}} = \int \mathrm{d}^5 z\, \tfrac{1}{2}\mathrm{i}\hat{b}\bar{\psi}\gamma^K h_K{}^M \nabla_M \psi + \mathrm{H.C.} \tag{10.21a}$$

More explicitly, we have

$$I_{\mathrm{D}} = \int \mathrm{d}^5 z\, \tfrac{1}{2}\mathrm{i}\hat{b}\bar{\psi}[\gamma^k(h_k{}^\mu\partial_\mu + h_k{}^5\partial_y) + \gamma^{\bar{5}}h_{\bar{5}}{}^5\partial_y + \gamma^K\omega_K]\psi + \mathrm{H.C.}$$

where $\omega_K \equiv \frac{1}{2}A^{IJ}{}_K\sigma_{IJ}$. Now, we neglect four-dimensional gravity and set $\phi = -1$, so that the inverse pentad takes the form

$$h_K{}^M(x) = \begin{pmatrix} \delta_k{}^\mu & h_k{}^5 \\ 0 & 1 \end{pmatrix} \qquad h_k{}^5 = -\delta_k^\mu B_\mu. \tag{10.22}$$

We assume that the connection is y independent and use the Fourier expansion for the Dirac field $\psi(x, y) = \sum \psi_n(x)Y_n(y)$, whereupon the integration over y in I_{D} yields

$$I_{\mathrm{D}}^{(0)} = \sum_n \int \mathrm{d}^4x\, \tfrac{1}{2}\mathrm{i}\bar{\psi}_n[\gamma^k\delta_k^\mu\nabla_{n\mu} - \mathrm{i}\gamma^{\bar{5}}m_n + \gamma^K\omega_K]\psi_n + \mathrm{H.C.} \qquad (10.21b)$$

where $\nabla_{n\mu} = \partial_\mu - \mathrm{i}q_nA_\mu$, $m_n = n/r$.

Similar results are also obtained in the presence of four-dimensional gravity. This five-dimensional theory of the Dirac field in interaction with gravity is defined in Riemann–Cartan space U_5, in which the connection $\hat{A}^{IJ}{}_M$ is not Riemannian. The equations of motion for the connection are algebraic, and can be solved to yield $\hat{A} = \Delta + K$. Hence, the replacement $\hat{A} = \Delta$ is *not consistent* from the point of view of the theory in U_5.

We may, alternatively, start from action (10.21a) with $\hat{A} = \Delta$, representing a theory of the Dirac field in Riemann space V_5. It can be shown that $\gamma^K\Delta_K$ gives rise to a non-minimal coupling, proportional to the Pauli term $\gamma^{\bar{5}}F_{ij}\sigma^{ij}$.

Symmetries and the particle spectrum

The nature of the mass spectrum of the effective four-dimensional theory is best characterized by the related symmetry properties (Dollan and Duff 1984).

The massless sector. Five-dimensional KK theory is invariant under general coordinate transformations and local Lorentz rotations. The condition of the local orthogonality between V_1 and V_4 restricts the coordinate transformation to the form (10.12), whereas the related transformations of the metric components $g_{\mu\nu}(x, y)$, $B_\mu(x, y)$ and $\phi(x, y)$ are given by the rule (10.13). In the truncated theory (10.15), the metric components are independent of y. The condition $g_{\mu\nu} = g_{\mu\nu}(x)$ does not imply any restriction on the coordinate transformations (10.12), while the conditions $B_\mu = B_\mu(x)$ and $\phi = \phi(x)$ imply that $\partial y'/\partial y$ and $\partial y'/\partial x$ do not depend on y. Therefore, the truncated theory (10.15) is invariant under the restricted general coordinate transformations,

$$x' = x'(x) \qquad y' = \rho y + \xi^5(x) \qquad (10.23a)$$

where $\rho =$ constant, and, correspondingly,

$$g'_{\mu\nu} = \frac{\partial x^\lambda}{\partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\nu}g_{\lambda\rho} \qquad \phi' = \frac{\partial y}{\partial y'}\frac{\partial y}{\partial y'}\phi$$
$$B'_\mu = \frac{\partial x^\lambda}{\partial x'^\mu}\rho B_\lambda - \frac{\partial \xi^5}{\partial x'^\mu}. \qquad (10.23b)$$

Bearing in mind the physical interpretation of the effective low-energy four-dimensional theory (10.15), it is convenient to divide these symmetry transformations into three groups.

(a) Transformations

$$x' = x'(x) \qquad y' = y \tag{10.24a}$$

are general coordinate transformations in V_4, under which the fields $g_{\mu\nu}$, B_μ and ϕ transform as second order tensors, vectors and scalars, respectively. Weyl rescaling by a factor $(-\bar{\phi})^{1/3}$ defines new variables $\bar{g}_{\mu\nu}$, $\bar{B}_\mu$ and $\bar{\phi}$. Since $\bar{\phi}$ is a scalar, the transformation rules for the new variables $\bar{g}_{\mu\nu}$ and $\bar{B}_\mu$ remain the same as for the old ones.

(b) Next, we consider

$$\begin{aligned} &x' = x \qquad y' = y' + \xi^5(x) \\ &\delta_0\phi = 0 \qquad \delta_0 g_{\mu\nu} = 0 \qquad \delta_0 B_\mu = -\partial_\mu \xi^5(x). \end{aligned} \tag{10.24b}$$

Originally, the parameter ξ^5 corresponded to the change in the *periodic* fifth coordinate y, that represented local $U(1)$ symmetry. However, since the truncated, y-independent action (10.15) has no memory of the periodicity in y, these transformations are now realization of the local (non-compact) T_1 symmetry, which corresponds to the *real* coordinate y. After Weyl rescaling, the transformation rule remains the same since $\delta_0\phi = 0$.

(c) The remaining component of the residual symmetry consists of global dilatations of the fifth dimension:

$$\begin{aligned} &x' = x \qquad y' = \rho y \\ &\phi' = \rho^{-2}\phi \qquad B'_\mu = \rho B_\mu \qquad g'_{\mu\nu} = g_{\mu\nu}. \end{aligned} \tag{10.25}$$

Direct application of this transformation to the action (10.15) yields a slightly unexpected result—the action is not invariant, $I_G \to \rho I_G$. However, if we recall that the four-dimensional coupling constant κ is defined by $\kappa^{-1} = \hat{\kappa}^{-1}\int \mathrm{d}y$, it becomes clear that global dilatations induce the change $\kappa^{-1} \to \rho\kappa^{-1}$, whereupon the invariance of the action is recovered.

Symmetry transformations that involve transformations of the coupling constant are not customary in four-dimensional theory. Observe, however, that action (10.15) possesses another, completely standard symmetry—symmetry under global Weyl rescalings W:

$$\phi' = \lambda^{-4/3}\phi \qquad B'_\mu = \lambda B_\mu \qquad g'_{\mu\nu} = \lambda^{2/3} g_{\mu\nu}. \tag{10.26a}$$

Expressed in terms of the variables $(\bar{\phi}, \bar{B}_\mu, \bar{g}_{\mu\nu})$, it takes the form

$$\bar{\phi}' = \lambda^{-2}\bar{\phi} \qquad \bar{B}_\mu = \lambda \bar{B}_\mu \qquad \bar{g}'_{\mu\nu} = \bar{g}_{\mu\nu}. \tag{10.26b}$$

The geometric meaning of this symmetry can be illustrated by looking at the interval along the fifth dimension, $\mathrm{d}s_5^2 = \phi\,\mathrm{d}y^2$: since $M_4 \times S_1$ is a flat space, it satisfies field equations for every value of the radius $r\sqrt{-\phi}$.

The global Weyl invariance of action (10.15) is *spontaneously broken* by the KK vacuum ($\phi_0 = -1$), giving rise to the appearance of the Goldstone boson ϕ, called the *dilaton*. However, W symmetry is only an 'accidental' symmetry, valid in the massless sector of the classical theory; the inclusion of higher, massive modes and/or quantum corrections breaks this symmetry. Therefore,

> *the field* $\phi^0 = \phi(x)$ *is actually a pseudo-Goldstone boson of global rescalings.*

We expect massless modes of an approximate theory to remain massless in a complete theory if there is a symmetry that ensures their masslessness. This is indeed the case with the graviton $g_{\mu\nu}(x)$ and the electromagnetic potential $B_\mu(x)$, since four-dimensional coordinate transformations and local $U(1)$ transformations are the symmetries of the full theory. Certain aspects of the symmetry of the full theory are discussed later.

In the complete theory the field ϕ becomes massive and the radius of the fifth dimension is determined. In the massless sector, however, the radius is usually fixed by hand, so as to reproduce the experimental values of the electric charge and the gravitational constant. If there were some dynamical mechanism determining the radius of the fifth dimension, pseudo-Goldstone bosons would not exist and the electric charge would be calculable.

Kac–Moody symmetry. In order to analyse the symmetry of the complete effective theory in four dimensions, we write the general coordinate transformations in the form

$$\delta z^M = \xi^M(x,y) = \sum \xi_n^M(x) u_n(\theta) \qquad u_n(\theta) \equiv \sqrt{2\pi r}\, Y_n(y) = \mathrm{e}^{-\mathrm{i}n\theta}$$

where the choice $y = r\theta$ reflects the geometry of the ground state.

Whereas the general coordinate transformations (10.24a) and the local $U(1)$ transformations (10.24b) correspond to the zero modes of $\xi^M(x,y)$, the global rescaling (10.26) is no more a symmetry of the complete theory. Indeed, any W transformation can be regarded as a composition of (i) the global rescaling of the metric: $\hat{g}'_{MN} = \lambda^{2/3}\hat{g}_{MN}$, i.e.

$$\phi' = \lambda^{2/3}\phi \qquad B'_\mu = B_\mu \qquad g'_{\mu\nu} = \lambda^{2/3} g_{\mu\nu}$$

and (ii) the general coordinate transformation of a special kind:

$$\begin{gathered} y' = \lambda y \\ \phi' = \lambda^{-2}\phi \qquad B'_\mu = \lambda B_\mu \qquad g'_{\mu\nu} = g_{\mu\nu}. \end{gathered}$$

Under transformation (i) the action I_G goes over into λI_G, so that the classical equations of motion remain invariant, but transformation (ii) does not satisfy the periodicity requirement for y. Hence, W is not a symmetry of the field equations of the full theory.

Global Poincaré invariance in four dimensions may be regarded as a special case of the general covariance, $\delta x^\mu = \xi^\mu(x)$, in which the parameter ξ^μ is restricted to the linear form: $\xi^\mu = \varepsilon^\mu + \omega^{\mu\nu}x_\nu$, $\omega^{\mu\nu} = -\omega^{\nu\mu}$. An analogue of this symmetry in five dimensions is obtained by linearizing the parameter $\xi_n^M(x)$:

$$\xi_n^\mu(x) = \varepsilon_n^\mu + \omega_n^{\mu\nu}x_\nu \qquad \xi_n^5(x) = c_n r \tag{10.27}$$

where ε_n^μ, $\omega_n^{\mu\nu}$ and c_n are constants. If $\varphi(x, y)$ is a scalar function, it transforms under(10.27) according to

$$\delta_0\varphi = -\sum u_n(\xi_n^\mu\partial_\mu + \xi_n^5\partial_y)\varphi = \sum(\varepsilon_n^\mu P_\mu^n + \tfrac{1}{2}\omega_n^{\mu\nu}M_{\mu\nu}^n + c_n L_n)\varphi$$

where P_μ^n, $L_{\mu\nu}^n$ and L_n are the corresponding generators:

$$\begin{aligned} P_\mu^n &\equiv -u_n(\theta)\partial_\mu \qquad M_{\mu\nu}^n \equiv u_n(\theta)(x_\mu\partial_\nu - x_\nu\partial_\mu) \\ L_n &= u_n(\theta)r\partial_y = u_n(\theta)\partial_\theta. \end{aligned} \tag{10.28}$$

These generators define the following infinite-parameter global Lie algebra:

$$\begin{aligned} [M_{\mu\nu}^n, M_{\lambda\rho}^m] &= \eta_{\nu\lambda}M_{\mu\rho}^{n+m} - \eta_{\mu\lambda}M_{\nu\rho}^{n+m} - \eta_{\nu\rho}M_{\mu\lambda}^{n+m} + \eta_{\mu\rho}M_{\nu\lambda}^{n+m} \\ [M_{\mu\nu}^n, P_\lambda^m] &= \eta_{\nu\lambda}P_\mu^{n+m} - \eta_{\mu\lambda}P_\nu^{n+m} \qquad [P_\mu^m, P_\nu^n] = 0 \end{aligned} \tag{10.29a}$$

$$\begin{aligned} [L_n, L_m] &= \mathrm{i}(n-m)L_{n+m} \\ [L_n, P_\mu^m] &= -imP_\mu^{n+m} \qquad [L_n, M_{\mu\nu}^m] = -imM_{\mu\nu}^{n+m}. \end{aligned} \tag{10.29b}$$

Equations (10.29*a*) define the Kac–Moody extension of the Poincaré algebra and equation (10.29*b*) represent the Virasoro algebra without a central charge, extended by the set of spacetime generators P_μ^n and $M_{\mu\nu}^n$. If we restrict our attention to the sector $n = m = 0$, we find the usual $P_4 \times U(1)$ algebra, where $P_4 \equiv P(1, 3)$. By adding the generators L_1 and L_{-1}, this finite-dimensional subalgebra can be enlarged to $P_4 \times SO(1, 2)$.

The algebra (10.29) describes the symmetry of the complete effective four-dimensional action, containing all the modes of dynamical variables.

Spontaneous symmetry breaking. Since the KK ground state has only $P_4 \times U(1)$ symmetry, the Kac–Moody symmetry (10.29) is spontaneously broken. The parameters $\xi_n^\mu(x)$ and $\xi_n^5(x)$, given by equation (10.27), describe a global (infinite-dimensional) symmetry. The zero modes ξ_0^μ and ξ_0^5 correspond to the ground state symmetry, while all higher modes ξ_n^μ and ξ_n^5, $n > 0$, describe spontaneously broken Kac–Moody generators (table 10.1). As a consequence, the corresponding fields are expected to be Goldstone bosons. In order to identify Goldstone bosons, we shall look at the transformation laws of all higher modes of the fields, and find those that transform *inhomogeneously* under the action of the broken generators.

Table 10.1. Symmetries of the five-dimensional KK theory.

Symmetry	$P_4 \times U(1)$	$P_4 \times SO(1,2)$	Kac–Moody
Parameters	ξ_0^μ, ξ_0^5	$\xi_0^\mu, \xi_0^5, \xi_1^5, \xi_{-1}^5$	ξ_n^μ, ξ_n^5

Example 3. To illustrate the transformation rule of Goldstone bosons, we return to example 2.5. There we considered a triplet of scalar fields φ^a, with an action invariant under three-dimensional internal rotations: $\varphi^a \to R^a{}_b\varphi^b$. Since the third component has a non-vanishing value in the ground state, $(\varphi^3)_0 = v$, it is convenient to define new fields η^a,

$$\varphi^1 = \eta^1 \qquad \varphi^2 = \eta^2 \qquad \varphi^3 = v + \eta^3$$

having the vanishing ground state values. The fields η^1 and η^2 are Goldstone bosons corresponding to spontaneously broken symmetry generators. The new fields transform under rotations as

$$\begin{aligned} \eta^3 &\to R^3{}_1\eta^1 + R^3{}_2\eta^2 \\ \eta^\alpha &\to R^\alpha{}_\beta\eta^\beta + R^\alpha{}_3(v + \eta^3) \qquad \alpha, \beta = 1, 2. \end{aligned}$$

Thus, the field η^3 transforms homogeneously whereas Goldstone bosons transform inhomogeneously.

In the same way, starting from the transformation rules of the fields $g_{\mu\nu}(x, y)$, $B_\mu(x, y)$ and $\phi(x, y)$ with respect to the general coordinate transformations in $d = 5$, and the form of the ground state, we can determine the transformation properties of $g^n_{\mu\nu}$, B^n_μ and ϕ^n under the transformations (10.29), and identify the Goldstone modes. The result of this analysis is the following:

> *The fields B^n_μ and ϕ^n, $n > 0$, are Goldstone modes of the broken Kac–Moody symmetry.*

These considerations suggest that the fields B^n_μ and ϕ^n, as well as $g^0_{\mu\nu}$ and B^0_μ, are massless, while ϕ^0 is massive as the pseudo-Goldstone boson. What happens with $g^n_{\mu\nu}$? A complete and clear picture of the mass spectrum can be obtained only after the presence of gauge symmetries in the theory is taken into account.

The Higgs mechanism. In theories with spontaneously broken symmetries which also possess gauge invariance, the Goldstone theorem does not hold, and the physical mass spectrum is determined by the so-called Higgs mechanism:

Table 10.2. Particle spectrum of five-dimensional KK theory.

Fields	ϕ^0	B^0_μ	$g^0_{\mu\nu}$	$g^n_{\mu\nu}$	B^n_μ	ϕ^n
Degrees of freedom	1	2	2	5	—	—
Type	$m \neq 0$	$m = 0$	$m = 0$	$m \neq 0$	Gb	Gb

there are no physical particles corresponding to Goldstone bosons—they are 'gauged away' and absorbed by the corresponding gauge fields, which become massive. This mechanism is also responsible for the particle spectrum of the KK theory.

(a) Five-dimensional KK theory possesses a gauge symmetry described by the parameters $\xi^\mu_n(x)$ and the gauge fields ($g^0_{\mu\nu}$; $g^n_{\mu\nu}, n > 0$). The Higgs mechanism acts in the following way:
   - the gauge field $g^0_{\mu\nu}$ remains massless, as a consequence of the local P_4 symmetry [with parameters $\xi^\mu_0(x)$];
   - the gauge fields $g^n_{\mu\nu}$ ($n > 0$) (with two degrees of freedom), corresponding to the broken generators, absorb the potential Goldstone bosons B^n_μ and ϕ^n (with $2 + 1 = 3$ degrees of freedom) and become massive fields ($2 + 3 = 5$ degrees of freedom).

   This is why the full theory, which is invariant under general coordinate transformations in $d = 5$, has an infinite tower of massive fields.

(b) Local $U(1)$ symmetry ensures the masslessness of the gauge field B^0_μ.

(c) Global rescalings do not represent a symmetry of the full theory, hence ϕ^0 is merely a pseudo-Goldstone boson, i.e. a massive field.

In conclusion, the particle spectrum of KK theory contains massive scalar ϕ^0, massless gauge field B^0_μ and massless graviton $g^0_{\mu\nu}$, with the whole tower of massive modes $g^n_{\mu\nu}$ (table 10.2).

10.3 Higher-dimensional KK theory

Kaluza and Klein suggested the unification of gravity and electromagnetism starting from five-dimensional GR. By generalizing this idea to higher dimensions, we can account for the unification of gravity and non-Abelian gauge theory (see, e.g., Zee 1981, Salam and Strathdee 1982, Mecklenburg 1983, Duff *et al* 1986, Bailin and Love 1987).

General structure of higher-dimensional gravity

We assume, for simplicity, that a non-Abelian generalization of KK theory is realized in a Riemann space V_d, $d = 4 + D$.

Table 10.3. Properties of maximally symmetric spacetimes.

λ	Spacetime	Symmetry	$E > 0$?	SUSY?
< 0	de Sitter	$SO(1,4)$	No	No
$= 0$	Minkowski	Poincaré	Yes	Yes
> 0	anti de Sitter	$SO(2,3)$	Yes	Yes

The ground state. Let V_d be a d-dimensional Riemann space with local coordinates $z^M = (x^\mu, y^\alpha)$. At each point of V_d there is a metric $\hat{g}_{MN}$ of signature $(+, -, -, -; -, \ldots)$. If matter fields are absent, the dynamics is determined by the covariant action

$$I_{\mathrm{G}} = -\frac{1}{2\hat{\kappa}} \int \mathrm{d}z \sqrt{|\hat{g}|}(\widehat{R} + \Lambda) \qquad (10.30a)$$

where Λ is a cosmological constant. The equations of motion are

$$\widehat{R}_{MN} - \tfrac{1}{2}\hat{g}_{MN}(\widehat{R} + \Lambda) = 0. \qquad (10.30b)$$

We now look for the ground state solution of the form

$$(V_d)_0 = V_4 \times B_D \qquad \text{i.e.} \quad \hat{g}^0_{MN}(x,y) = \begin{pmatrix} g^0_{\mu\nu}(x) & 0 \\ 0 & g^0_{\alpha\beta}(y) \end{pmatrix} \qquad (10.31)$$

where V_4 is Riemannian spacetime with the usual signature $(+, -, -, -)$, and B_D is a D-dimensional space with Euclidean signature $(-, -, \ldots)$. It is a standard assumption that B_D is a compact space (it may be visualized as a closed bounded subset of a Euclidean space; see appendix K). Since a compact space in $D \geq 2$ is, in general, not flat, the metric $g^0_{\alpha\beta}$ depends on y. The field equations for the ground state metric (10.31) are

$$R^4_{\mu\nu} - \tfrac{1}{2}g^0_{\mu\nu}(R^4 + R^D + \Lambda) = 0 \qquad R^D_{\alpha\beta} - \tfrac{1}{2}g^0_{\alpha\beta}(R^4 + R^D + \Lambda) = 0$$

where $R^4 = R(V_4)$, $R^D = R(B_D)$.

It should be noted that in this theory it is not easy to obtain an acceptable ground state solution with a flat spacetime V_4. Indeed, from $R^4_{\mu\nu} = 0$ and the field equations it follows $R^D_{\alpha\beta} = 0$, which represents, as we shall see, too strong a limitation on B_D.

Thus, we restrict ourselves to *maximally symmetric* spacetimes V_4, hence to spaces of constant curvature, $R^4_{\mu\nu\lambda\rho} = \frac{1}{3}\lambda(g^0_{\mu\lambda}g^0_{\nu\rho} - g^0_{\mu\rho}g^0_{\nu\lambda})$, according to the possibilities shown in table 10.3 (Duff *et al* 1986).

Constant curvature spaces are Einstein spaces: $R^4_{\mu\lambda} = \lambda g^0_{\mu\lambda}$.

As far as extra dimensions are concerned, we assume that

- B_D yields physically important *non-Abelian gauge symmetries*; and

- B_D is compact, in order to guarantee a *discrete spectrum* in four dimensions.

To check the consistency of these assumptions, we start from the relation $R^4_{\mu\nu} = \lambda g^0_{\mu\nu}$ for V_4, and use the field equations to obtain that B_D is also Einstein: $R^D_{\alpha\beta} = \rho g^0_{\alpha\beta}$, where $\rho = \lambda$. The following theorem holds for Einstein spaces (Yano 1970):

> *Compact Einstein spaces with the Euclidean signature* $(-, -, \ldots)$ *and* $\rho > 0$ *have no continuous symmetries.*

Thus, an Einstein space B_D may have non-Abelian symmetries if $\lambda \leq 0$. On the other hand, in order for the space V_4 to admit a positive energy theorem (stability) and supersymmetry, we must have $\lambda \geq 0$. Both of these requirements are satisfied only for $\lambda = 0$, i.e. $V_4 = M_4$. Field equations then imply $R^D_{\alpha\beta} = 0$.

A simple example of a space that admits a spontaneous compactification with $R^D_{\alpha\beta} = 0$ is a D-torus: $B_D = S_1 \times S_1 \times \cdots \times S_1$ (D times). However, this solution yields the Abelian gauge symmetry $U(1) \times U(1) \times \cdots \times U(1)$.

A very economical and simple mechanism for realizing non-Abelian gauge symmetries is obtained when B_D is a coset space. Physically interesting spaces of this type are Einstein spaces with $\rho \neq 0$, but then V_4 cannot be flat.

Therefore, the ground state $M_4 \times$ (coset) cannot be obtained without introducing additional matter fields. In that case, of course, the original simplicity of the KK idea is lost. More detailed considerations concerning the mechanism for spontaneous compactification are left for the end of this section.

The layered structure. In order to simplify the physical interpretation of massless modes, it is convenient to express the metric of V_d in locally orthogonal form. We assume that V_d has a layered structure: for each fixed point x^μ in V_4 there is a (D-dimensional) hypersurface V_D with coordinates y^α. Thus, locally, V_d can be viewed as $V_4 \times V_D$. If $\hat{\boldsymbol{e}}_M = (\hat{\boldsymbol{e}}_\mu, \hat{\boldsymbol{e}}_\alpha)$ is a coordinate frame in V_d, the set of vectors $\hat{\boldsymbol{e}}_\alpha$ defines the coordinate frame in V_D. Furthermore, let $\hat{\boldsymbol{e}}_I = (\hat{\boldsymbol{e}}_i, \hat{\boldsymbol{e}}_a)$ be a Lorentz frame in T_d (vielbein); then, $\hat{\boldsymbol{e}}_i$ and $\hat{\boldsymbol{e}}_a$ are Lorentz frames in T_4 and T_D, respectively. Every coordinate frame $\hat{\boldsymbol{e}}_M$ can be expressed in terms of the Lorentz frame: $\hat{\boldsymbol{e}}_M = b^i{}_M \hat{\boldsymbol{e}}_i + b^a{}_M \hat{\boldsymbol{e}}_a$. The condition of local orthogonality of V_4 and V_D means that the expansion of $\hat{\boldsymbol{e}}_\alpha$ in V_D does not contain any vector $\hat{\boldsymbol{e}}_i$ in V_4, i.e.

$$b^i{}_\alpha = 0 \quad\Longleftrightarrow\quad b^I{}_M(x, y) = \begin{pmatrix} b^i{}_\mu & 0 \\ b^a{}_\mu & b^a{}_\alpha \end{pmatrix}. \tag{10.32a}$$

The inverse matrix $h_I{}^M$ has the form

$$h_I{}^M(x, y) = \begin{pmatrix} h_i{}^\mu & h_i{}^\alpha \\ 0 & h_a{}^\alpha \end{pmatrix}. \tag{10.32b}$$

The usual construction of the metric in V_d yields

$$\hat{g}_{MN}(x, y) = \begin{pmatrix} g_{\mu\nu} + \phi_{\alpha\beta} B^\alpha_\mu B^\beta_\nu & B^\beta_\mu \phi_{\beta\alpha} \\ \phi_{\alpha\beta} B^\beta_\nu & \phi_{\alpha\beta} \end{pmatrix} \tag{10.33a}$$

where we have introduced the notation

$$g_{\mu\nu} \equiv b^i{}_\mu b^j{}_\nu \eta_{ij} \qquad \phi_{\alpha\beta} \equiv \hat{g}_{\alpha\beta} = b^a{}_\alpha b^b{}_\beta \eta_{ab}$$
$$B^\alpha_\mu \equiv h_a{}^\alpha b^a{}_\mu \qquad \text{or} \qquad b^a{}_\mu \equiv b^a{}_\alpha B^\alpha_\mu .$$

The inverse metric is given by

$$\hat{g}^{MN}(x,y) = \begin{pmatrix} g^{\mu\nu} & -g^{\mu\rho} B^\alpha_\rho \\ -B^\alpha_\rho g^{\rho\nu} & \phi^{\alpha\beta} + g^{\lambda\rho} B^\alpha_\lambda B^\beta_\rho \end{pmatrix} \tag{10.33b}$$

where $g^{\mu\nu}$ and $\phi^{\alpha\beta}$ are the inverses of $g_{\mu\nu}$ and $\phi_{\alpha\beta}$, respectively.

This form of the metric is directly related to the local orthogonality of V_4 and V_D. The components $g_{\mu\nu}$ and $\phi_{\alpha\beta}$ have a clear geometric meaning: $\phi_{\alpha\beta}$ is the metric in V_D, while $g_{\mu\nu}$ defines the distance between neighbouring layers. The interpretation of B^α_μ follows from the symmetry structure of B_D.

The residual symmetry. General coordinate transformations in V_d that do not violate its layered structure, expressed by condition (10.32a), have the form

$$x' = x'(x) \qquad y' = y'(x,y). \tag{10.34a}$$

Under these transformations the components of the metric transform according to

$$g'_{\mu\nu} = \frac{\partial x^\lambda}{\partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\nu} g_{\lambda\rho} \qquad \phi'_{\alpha\beta} = \frac{\partial y^\gamma}{\partial y'^\alpha}\frac{\partial y^\delta}{\partial y'^\beta}\phi_{\gamma\delta}$$
$$B'^\alpha_\mu = \frac{\partial x^\nu}{\partial x'^\mu}\left(\frac{\partial y'^\alpha}{\partial y^\gamma} B^\gamma_\nu - \frac{\partial y'^\alpha}{\partial x^\nu}\right). \tag{10.34b}$$

Therefore,

(a) $g_{\mu\nu}(x,y)$ is a tensor in V_d,
(b) $\phi_{\alpha\beta}(x,y)$ is a tensor in V_d and
(c) $B^\alpha_\mu(x,y)$ transforms inhomogeneously, except in the case $y' = y'(y)$ when it becomes a tensor in V_d.

In particular, we note that the quantities $g_{\mu\nu}$, B^α_μ and $\phi_{\alpha\beta}$ are V_4 tensors (i.e. tensors under $x' = x'(x)$, $y' = y$), and also B_D tensors ($x' = x$, $y' = y'(y)$).

Isometries and harmonic expansion. We expect that geometric symmetries of the space B_D will lead to non-Abelian gauge symmetries of the massless sector in four dimensions. The symmetry properties of B_D are precisely expressed using the concept of isometries (appendix K).

Consider, first, an infinitesimal coordinate transformation on B_D, which can be written in the form

$$\delta y^\alpha = \varepsilon^a E^\alpha_a(y) = \varepsilon^a \Gamma_a y^\alpha \qquad \Gamma_a \equiv E^\alpha_a \partial_\alpha \qquad a = 1,2,\ldots,m \tag{10.35}$$

where the ε^a are constant parameters and the Γ_a the generators of the transformations. The number of parameters m is, in general, different from the dimension of B_D. These transformations represent a faithful realization of the group G on B_D if the generators Γ_a satisfy the commutation rules

$$[\Gamma_a, \Gamma_b] = f_{ab}{}^c \Gamma_c \quad \Longleftrightarrow \quad E_a^\alpha \partial_\alpha E_b^\beta - E_b^\alpha \partial_\alpha E_a^\beta = f_{ab}{}^c E_c^\beta \tag{10.36a}$$

coinciding with the Lie algebra of G.

Of particular importance are those transformations that do not change the form of the metric: $\phi'_{\alpha\beta}(y) = \phi_{\alpha\beta}(y)$; they are called *isometries* of B_D. The condition that the infinitesimal transformations (10.35) are isometries of the space B_D, $\delta_0\phi_{\alpha\beta} = 0$, can be written in the form

$$\nabla_\alpha E_{a\beta} + \nabla_\beta E_{a\alpha} = 0 \qquad E_{a\alpha} \equiv \phi_{\alpha\beta} E_a^\beta \tag{10.36b}$$

which is known as the *Killing equation*. Isometry transformations of a given space B_D are determined by the solutions E_a^α of the Killing equation (Killing vectors). They form a group G, called the isometry group of B_D, which gives rise to the gauge group of the massless sector, as we shall see.

A space B_D is said to be *homogeneous* if for every two points P and P_1 in B_D there exists an isometry transformation that moves P into P_1. Thus, every tangent vector at P is a Killing vector. The number of linearly independent Killing vectors at any point P of a homogeneous space B_D is D, and they can be taken as a basis for T_P.

A space B_D is said to be *isotropic* about a point P if there is a subgroup H_P of the isometry group G that leaves the point invariant. Isotropy transformations around P may be imagined as 'rotations' around an axis through P. The number of linearly independent Killing vectors corresponding to the isotropy transformations is $D(D-1)/2$.

The number of independent Killing vectors may be higher than the dimension of B_D. The maximal number of linearly independent Killing vectors in a D-dimensional space is $D(D+1)/2$. A space with a maximal number of Killing vectors is said to be *maximally symmetric*. Homogeneous and isotropic spaces are maximally symmetric: counting the Killing vectors yields $D + D(D-1)/2 = D(D+1)/2$. The specific structure of maximally symmetric spaces enables us to express every vector at P in terms of the set of $D(D+1)/2$ linearly independent Killing vectors at P.

A simple illustration for these concepts is the sphere S_2, which represents maximally symmetric space of the rotation group $SO(3)$ and has three Killing vectors. The action of $SO(3)$ on any point P in S_2 is realized in such a way that only two rotations of $SO(3)$ act non-trivially on P, whereas the third one, the rotation about the normal to S_2 at P, leaves P invariant. Thus, S_2 is the coset space of $SO(3)$ by $SO(2)$: $S_2 = SO(3)/SO(2)$, where $H_P = SO(2)$ is the isotropy subgroup of $SO(3)$ at P (see example 4, appendix K).

Knowledge of the symmetry structure of B_D enables a simple transition to the effective four-dimensional theory, with the help of the so-called harmonic expansion. If a compact space B_D is homogeneous, a complete, orthonormal set of eigenfunctions $Y_{[n]}(y)$ of the Laplacian $\Box_y$ exists on B_D, so that every function $\Phi(x, y)$ can be expanded in terms of $Y_{[n]}$:

$$\Phi(x, y) = \sum_{[n]} \Phi_{[n]}(x) Y_{[n]}(y). \tag{10.37}$$

This expansion is known as the harmonic expansion on B_D; it is a generalization of the Fourier expansion on S_1 (Salam and Strathdee 1982, Viswanatan 1984).

Example 4. Consider a real scalar field φ in a six-dimensional KK theory, in which the space B_2 is a sphere S_2 of radius r. The Laplacian $\widehat{\Box}$ on V_6 may be shown to reduce as

$$\widehat{\Box} = g^{\mu\nu}(\nabla_\nu - B^\alpha_\mu \partial_\alpha)(\nabla_\nu - B^\beta_\nu \partial_\beta) + \Box_y$$

where ∇ is the covariant derivative on V_4 and $\Box_y = (\sqrt{|\phi|})^{-1}\partial_\alpha(\sqrt{|\phi|}\phi^{\alpha\beta}\partial_\beta)$ is the Laplacian on B_2. In spherical coordinates (θ, φ) the metric on B_2 has the form $\mathrm{d}s^2 = -r^2(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2)$, and the operator $\Box_y$ becomes

$$\Box_y = -\frac{1}{r^2}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}\right].$$

Its eigenfunctions are the spherical harmonics $Y_{lm}(\theta, \varphi)$:

$$\Box_y Y_{lm} = \frac{l(l+1)}{r^2} Y_{lm} \qquad (l = 0, 1, 2, \ldots \infty;\ m = -l, -l+1, \ldots, +l).$$

Using the expansion of the scalar field $\varphi(x, \theta, \varphi) = \sum_{l,m} \varphi_{lm}(x) Y_{lm}(\theta, \varphi)$ in the six-dimensional action and integrating over S_2, we obtain the effective four-dimensional theory. It contains an infinite number of modes $\varphi_{lm}(x)$ with masses given by $m_l^2 = l(l+1)/r^2$. Every mode $\varphi_{lm}(x)$ carries an irreducible representations of $SO(3)$, the isometry group of the sphere S_2.

Every dynamical variable in a given d-dimensional theory can be expanded in terms of B_D harmonics, equation (10.37). Replacing this expansion in action (10.30) and integrating over y (using the orthonormality of the harmonics $Y_{[n]}$) we obtain an effective four-dimensional action, in which the effect of extra dimensions is seen in the presence of an infinite number of higher modes $\Phi_{[n]}$. The mass spectrum of the effective four-dimensional theory depends on the choice of B_D.

The massless sector of the effective theory

After discussing the structure of the effective four-dimensional theory, we now turn our attention to the zero modes that characterize the low-energy properties of the theory.

The metric. In contrast to the full five-dimensional theory, the metric of the massless sector contains a reduced y-dependence:

$$g_{\mu\nu} = g_{\mu\nu}(x) \qquad \phi_{\alpha\beta} = \phi_{\alpha\beta}(y) \qquad B^\alpha_\mu = E^\alpha_a(y)B^a_\mu(x) \tag{10.38}$$

where E^α_a is the Killing vector of B_D. Let us now discuss in more detail the meaning of this ansatz.

The condition $g_{\mu\nu} = g_{\mu\nu}(x)$ means that all higher modes in the complete metric $g_{\mu\nu}(x, y)$ in V_4 are discarded.

In general, the space B_D is not flat, and the metric $\phi_{\alpha\beta}$ cannot be constant, but must depend on y. The condition $\phi_{\alpha\beta} = \phi_{\alpha\beta}(y)$ means that all excitations around B_D are neglected.

The structure of the field B^α_μ is based on the symmetry properties of B_D. The space B_D admits an isometry group G with Killing vectors E^α_a. The quantity B^α_μ, which is a vector on B_D, can be expressed in terms of the Killing basis: $B^\alpha_\mu = E^\alpha_a B^a_\mu$. Now, we can impose the restriction $B^a_\mu = B^a_\mu(x)$, thus obtaining the last relation in equation (10.38).

Non-Abelian gauge symmetry. Now, we come to the most exciting point of the KK theory: the isometry group of B_D is seen as the non-Abelian gauge structure in the massless sector of the theory. To see why this is true, consider an infinitesimal coordinate transformation in B_D:

$$x'^\mu = x^\mu \qquad y'^\alpha = y^\alpha + \xi^\alpha(x, y) \equiv y^\alpha + E^\alpha_a(y)\varepsilon^a(x, y) \tag{10.39a}$$

where the parameters ξ^α are expressed in the Killing basis.

Under these transformations $g_{\mu\nu}(x, y)$ is a scalar. The same is true if we assume that $g_{\mu\nu} = g_{\mu\nu}(x)$, so that this condition does not impose any restrictions on the parameters ξ^α.

If we use the expansions $B^\alpha_\mu = E^\alpha_a(y)B^a_\mu(x, y)$ and $\xi^\alpha = E^\alpha_a(y)\varepsilon^a(x, y)$ in the transformation law (10.34b) for B^α_μ, we obtain

$$E^\alpha_a \delta_0 B^a_\mu = [-\partial_\mu \varepsilon^a + f_{bc}{}^a B^b_\mu \varepsilon^c + (\varepsilon^a_{,\gamma} B^c_\mu - \varepsilon^c B^a_{\mu,\gamma})E^\gamma_c]E^\alpha_a.$$

The limitation $B^a_\mu = B^a_\mu(x)$ now implies $\varepsilon^a = \varepsilon^a(x)$, so that

$$\delta_0 B^a_\mu = -\partial_\mu \varepsilon^a + f_{bc}{}^a B^b_\mu \varepsilon^c. \tag{10.40}$$

This is precisely the transformation law of a *gauge field* with respect to the gauge transformations $\delta y^\alpha = E^\alpha_a(y)\varepsilon^a(x)$ of the group G, as defined in equation (10.35).

The metric $\phi_{\alpha\beta}(y)$ is invariant under these transformations, since E^a_α are the Killing vectors of B_D:

$$\delta_0 \phi_{\alpha\beta} = -\nabla_\beta \xi_\alpha - \nabla_\alpha \xi_\beta = -\varepsilon^a(\nabla_\beta E_{a\alpha} + \nabla_\alpha E_{a\beta}) = 0.$$

It is not difficult to see that the transformations $x' = x'(x)$, $y' = y$ are general coordinate transformations in four dimensions. Thus, we conclude that the adopted form of the metric (10.38) implies the invariance of the theory under the reduced general coordinate transformations

$$x'^{\mu} = x'^{\mu}(x) \qquad y'^{\alpha} = y^{\alpha} + \varepsilon^{a}(x)E_{a}^{\alpha}(y). \tag{10.39b}$$

The field $g_{\mu\nu}(x)$ is the metric of the four-dimensional spacetime, and $B_{\mu}^{a}(x)$ is the gauge field associated with the isometries of B_D.

These results give a simple interpretation of the zero modes in terms of the basic physical fields—the graviton and the non-Abelian gauge field.

The effective action. After a straightforward but rather lengthy calculation we obtain from the metric (10.38) the following relation:

$$\widehat{R} = R^4 + R^D - \tfrac{1}{4}\phi_{\alpha\beta}E_{a}^{\alpha}E_{b}^{\beta}F_{\mu\nu}^{a}F^{b\mu\nu} \tag{10.41}$$

where $F_{\mu\nu}^{a} \equiv \partial_{\mu}B_{\nu}^{a} - \partial_{\nu}B_{\mu}^{a} - f_{bc}{}^{a}B_{\mu}^{b}B_{\nu}^{c}$, and $F^{a\mu\nu} = g^{\mu\lambda}g^{\nu\rho}F_{\lambda\rho}^{a}$. Using, further, the factorization of the determinant, $\hat{g} = g\phi$, where $\phi = \det(\phi_{\alpha\beta})$, action (10.30) takes the form

$$I_{\mathrm{G}}^{(0)} = -\frac{1}{2\hat{\kappa}}\int \mathrm{d}^4x\,\mathrm{d}^Dy\,\sqrt{-g}\sqrt{|\phi|}(R^4 + R^D + \Lambda - \tfrac{1}{4}\phi_{\alpha\beta}E_{a}^{\alpha}E_{b}^{\beta}F_{\mu\nu}^{a}F^{b\mu\nu}).$$

We now choose the constants $\hat{\kappa}$ and Λ in the form

$$\hat{\kappa} = \kappa\int \mathrm{d}^Dy\,\sqrt{|\phi(y)|} \qquad \Lambda = -\frac{\int \mathrm{d}^Dy\,\sqrt{|\phi(y)|}R^D(y)}{\int \mathrm{d}^Dy\,\sqrt{|\phi(y)|}}$$

where κ is the four-dimensional gravitational constant, and Λ is chosen so as to cancel the contribution of R^D after the integration over y. If we normalize the Killing vectors according to the rule

$$-\frac{1}{2\hat{\kappa}}\int \mathrm{d}^Dy\,\sqrt{|\phi(y)|}\phi_{\alpha\beta}E_{a}^{\alpha}(y)E_{b}^{\beta}(y) = \delta_{ab} \tag{10.42}$$

the effective four-dimensional action takes the form

$$I_{\mathrm{G}}^{(0)} = \int \mathrm{d}^4x\,\sqrt{-g}\left(-\frac{1}{2\kappa}R^4(x) - \tfrac{1}{4}F_{\mu\nu}^{a}(x)F^{a\mu\nu}(x)\right) \tag{10.43}$$

corresponding to the standard theory of gravity in interaction with non-Abelian gauge fields.

The graviton–scalar sector. In the previous discussion we assumed that $\phi_{\alpha\beta}$ is independent of x. Consider now the ansatz

$$\hat{g}_{MN}(x,y) = \begin{pmatrix} g_{\mu\nu}(x) & 0 \\ 0 & \phi_{\alpha\beta}(x,y) \end{pmatrix} \tag{10.44a}$$

which allows the x dependence of $\phi_{\alpha\beta}$, while the gauge fields are neglected for simplicity. This metric leads to the following $(4+D)$-dimensional action (Cho and Freund 1975)

$$\begin{aligned} I_{\mathrm{G}}^{(0)} = &-\frac{1}{2\hat{\kappa}}\int \mathrm{d}^4x\,\mathrm{d}^Dy\,\sqrt{-g}\sqrt{|\phi|}(R^4 + R^D + \Lambda - \phi^{\alpha\beta}\nabla^\mu\nabla_\mu\phi_{\alpha\beta} \\ &-\tfrac{1}{2}\nabla^\mu\phi_{\alpha\beta}\nabla_\mu\phi^{\alpha\beta} - \tfrac{1}{4}\phi^{\alpha\beta}\nabla^\mu\phi_{\alpha\beta}\phi^{\gamma\delta}\nabla_\mu\phi_{\gamma\delta} \\ &+\tfrac{1}{4}\phi^{\alpha\beta}\phi^{\gamma\delta}\nabla^\mu\phi_{\alpha\gamma}\nabla_\mu\phi_{\beta\delta}). \end{aligned} \tag{10.44b}$$

Using the harmonic expansion of $\phi_{\alpha\beta}$ and integrating over y we can obtain the effective theory in four dimensions.

The coupling constants. The normalization of the Killing vectors by equation (10.42) determines the gauge coupling constant. Indeed, if the Killing vectors are not normalized in accordance with (10.42), the change in their norm implies the change of the coupling constant contained in $f_{ab}{}^c$:

$$E_a^\alpha \to gE_a^\alpha \quad\Longrightarrow\quad f_{ab}{}^c \to gf_{ab}{}^c.$$

Example 5. We illustrate this effect in the case when B_D is the sphere S_2 with radius r. The metric of S_2 in spherical coordinates $y^\alpha = (\theta,\varphi)$ is determined by $\mathrm{d}s^2 = -r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2)$. Solving the Killing equation we obtain the generators

$$\begin{aligned} \Gamma_1 &= \sin\varphi\partial_\theta + \cot\theta\cos\varphi\partial_\varphi \\ \Gamma_2 &= \cos\varphi\partial_\theta - \cot\theta\sin\varphi\partial_\varphi \\ \Gamma_3 &= \partial_\varphi \end{aligned}$$

that satisfy the commutation relations $[\Gamma_a,\Gamma_b] = -\varepsilon_{abc}\Gamma_c$. Using the formulae

$$\int \mathrm{d}^2y\,\sqrt{|\phi|} = 4\pi r^2 \qquad -\int \mathrm{d}^2y\,\sqrt{|\phi|}\phi_{\alpha\beta}E_a^\alpha E_b^\beta = \frac{8\pi r^4}{3}\delta_{ab}$$

it follows that $\hat{\kappa} = \kappa 4\pi r^2$, and the Killing vectors are normalized as

$$-\frac{1}{2\hat{\kappa}}\int \mathrm{d}^Dy\,\sqrt{|\phi(y)|}\phi_{\alpha\beta}E_a^\alpha(y)E_b^\beta(y) = \frac{r^2}{3\kappa}\delta_{ab}.$$

In order to reconcile this relation with (10.42), we have to rescale both the Killing vectors and the structure constants:

$$E_a^\alpha \to g E_a^\alpha \qquad \varepsilon_{abc} \to g\varepsilon_{abc} \qquad g \equiv \frac{1}{r}\sqrt{3\kappa}$$

where g is the coupling constant of $SO(3)$. If the value of g is close to unity, the radius r has to be very small—of the order of the Planck length.

When the internal space is a D-dimensional sphere, the isometry group is $SO(D+1)$, and the coupling constant has the value

$$g = \frac{1}{r}\sqrt{\kappa(D+1)}. \tag{10.45a}$$

In this case the quantity $r/\sqrt{D}$ has to be small, which can be achieved not only by having small r but also by increasing D. Since the sphere S_D is Einstein space, $R^D_{\alpha\beta} = \rho\phi_{\alpha\beta}$ with $\rho = -(D-1)/r^2$, we find that

$$g = \sqrt{\kappa(-\rho)(D+1)/(D-1)} \tag{10.45b}$$

which implies that the constant $(-\rho)$ must be large.

A similar connection between the coupling constant and the diameter of B_D also exists for the general B_D (Weinberg 1983). If B_D is not isotropic, there are several coupling constants. Going back to isotropic B_D we note that after rescaling $f_{ab}{}^c \to g f_{ab}{}^c$, the field strength takes the standard form: $F^a_{\mu\nu} = \partial_\mu B^a_\nu - \partial_\nu B^a_\mu - g f_{bc}{}^a B^b_\mu B^c_\nu$.

Spontaneous compactification

The field equations following from action (10.30) do not have classical solutions of the form $M_4 \times B_D$, where B_D is a compact space. The compactification of the D extra dimensions can be realized by introducing matter fields (for other possibilities see, e.g., Bailin and Love (1987)).

The KK dynamics of gravity plus additional matter fields is described by the action

$$I = -\frac{1}{2\hat\kappa}\int \mathrm{d}z\,\sqrt{|\hat g|}(\widehat R + \Lambda) + I_{\mathrm{M}}. \tag{10.46}$$

The gravitational field equations are

$$\widehat R_{MN} - \tfrac{1}{2}\hat g_{MN}(\widehat R + \Lambda) = \hat\kappa T_{MN} \tag{10.47}$$

where T_{MN} is defined by the relation $\delta I_{\mathrm{M}} \equiv \frac{1}{2}\int \mathrm{d}z\sqrt{|\hat g|}T_{MN}\delta\hat g^{MN}$. If the ground state is of the form $M_4 \times B_D$, then $\widehat R_{\mu\nu} = 0$ implies

$$T_{\mu\nu} = \frac{C_4}{\hat\kappa}\eta_{\mu\nu} \qquad C_4 = -\tfrac{1}{2}(R^D + \Lambda) = \text{constant}.$$

Similarly, demanding that B_D be an Einstein space, $\widehat{R}_{\alpha\beta} = \rho\phi_{\alpha\beta}$, it follows that

$$T_{\alpha\beta} = \frac{C_D}{\hat{\kappa}}\phi_{\alpha\beta} \qquad C_D = -\tfrac{1}{2}(R^D + \Lambda - 2\rho) = \text{constant}.$$

It then follows from the field equations that

$$R^D_{\alpha\beta} = (C_D - C_4)\phi_{\alpha\beta} \qquad \Lambda = (D-2)C_4 - DC_D.$$

This shows that compactification occurs provided

$$\rho = C_D - C_4 < 0. \tag{10.48}$$

This condition ensures that a compact space B_D with signature $(-, -, \ldots)$ has a continuous, non-Abelian group of symmetry. In addition, we shall demand that $C_4 - C_D \approx 1/\kappa$, which follows from $g \approx 1$ and (10.45b).

Freund–Rubin compactification. There is a mechanism for compactification that arises naturally in 11-dimensional supergravity. It is based on a third-rank antisymmetric tensor field A_{MNP}, with field strength $F_{KLMN} = \partial_K A_{LMN} - \partial_N A_{KLM} + \partial_M A_{NKL} - \partial_L A_{MNK}$ and the action (Freund and Rubin 1980)

$$I_{\mathrm{M}} = -\tfrac{1}{48}\int \mathrm{d}z\,\sqrt{|\hat{g}|}\,F_{KLMN}F^{KLMN}. \tag{10.49}$$

The field equation for A_{KLMN} is $\partial_K(\sqrt{|\hat{g}|}F^{KLMN}) = 0$, and the energy–momentum tensor is

$$T_{IJ} = -\tfrac{1}{6}(F_{KLMI}F^{KLM}{}_J - \tfrac{1}{8}\hat{g}_{IJ}F_{KLMN}F^{KLMN}).$$

The field equation has a solution

$$F^{\mu\nu\lambda\rho} = F\varepsilon^{\mu\nu\lambda\rho}/\sqrt{-g} \qquad \text{all other components} = 0 \tag{10.50}$$

where F is a constant. The related energy–momentum tensor takes the form

$$T_{\mu\nu} = \tfrac{1}{2}F^2\eta_{\mu\nu} \qquad T_{\alpha\beta} = -\tfrac{1}{2}F^2\phi_{\alpha\beta}.$$

Thus, $C_4 = -C_D = \hat{\kappa}F^2/2$, and the compactification condition (10.48) is satisfied. The cosmological constant takes the value $\Lambda = \hat{\kappa}(D-1)F^2$.

It should be noted that in 11-dimensional supergravity this mechanism is not satisfying, since a non-vanishing cosmological constant destroys the supersymmetry.

Extra gauge fields. The presence of extra gauge fields contradicts the spirit of the original KK approach, where *all* gauge fields are expected to arise from the metric $\hat{g}_{MN}$. Nevertheless, we shall consider this possibility as an interesting illustration of the compactification mechanism (Luciani 1978).

The action for extra gauge fields has the form

$$I_{\mathrm{M}} = -\tfrac{1}{4}\int \mathrm{d}z\,\sqrt{|\hat{g}|}F^a_{MN}F^{MN}_a \tag{10.51}$$

where F^a_{MN} is the usual field strength and the energy–momentum tensor is given by

$$T_{MN} = -(F^a_{KM}F^K_{a\,N} - \tfrac{1}{4}\hat{g}_{MN}F^a_{KL}F^{KL}_a).$$

We assume that the non-Abelian gauge group G coincides with the isometry group of B_D. There is a solution to the equations of motion that has the form

$$\begin{aligned} A^a_\mu &= 0 \qquad A^a_\alpha = aE^a_\alpha \\ \phi^{\alpha\beta} &= bE^\alpha_a E^\beta_b g^{ab} \end{aligned} \tag{10.52}$$

where a and b are constants, E^a_α are the Killing vectors of B_D, and g^{ab} the Cartan metric of G. Since only the components $F^a_{\alpha\beta}$ have a non-vanishing value, the substitution in T_{MN} yields

$$\frac{C_4}{\hat{\kappa}} = \tfrac{1}{4}F^a_{\alpha\beta}F^{a\alpha\beta} \qquad \frac{C_D}{\hat{\kappa}} = \frac{C_4}{\hat{\kappa}} - \frac{1}{D}F^a_{\alpha\beta}F^{a\alpha\beta}.$$

This is consistent with the compactification condition (10.48). The quantity $F^a_{\alpha\beta}F^{a\alpha\beta}$ is positive and can be made sufficiently large by a suitable choice of a.

General remarks

The dimension of B_D. The basic idea of KK theory is that four-dimensional internal symmetries are, in fact, spacetime symmetries in the extra dimensions. One of the first problems that we face when trying to construct a realistic KK theory is the choice of the dimension and symmetry structure of B_D. If we want our $(4+D)$-dimensional theory to describe known (electroweak and strong) particle interactions, B_D must contain $SU(3)\times SU(2)\times U(1)$ as a symmetry group. A simple choice for B_D is $CP_2\times S_2\times S_1$: $SU(3)$ is the symmetry group of the complex projective space CP_2 $(= SU(3)/SU(2))$, $SU(2)$ is the symmetry of the sphere S_2, and $U(1)$ is the symmetry of the circle S_1. This space has $4+2+1=7$ dimensions. We should note that $CP_2\times S_2\times S_1$ is not the only seven-dimensional spaces with $SU(3)\times SU(2)\times U(1)$ symmetry, but no space with smaller dimensionality has this symmetry. These seven extra dimensions together with the usual four dimensions of spacetime make the total dimensionality of a realistic KK theory to be at least $4+7=11$.

On the other hand, there are convincing arguments that $d = 11$ is the maximal dimension in which we can consistently formulate a supergravity theory. These arguments are based on the fact that a supergravity theory in $d > 11$ has to contain massless fields of spin greater than two, which are believed to have no consistent coupling to gravity. This intriguing numerical coincidence, that the maximal dimension for supergravity is the minimal dimension in which one can obtain $SU(3) \times SU(2) \times U(1)$, suggests that the dimension of the world we live in is just *11* (Witten 1981a).

Fermions. To have a realistic description of nature we have to overcome a number of difficulties, such as the incorporation of fermions (quarks and leptons) and Higgs bosons (necessary for $SU(2) \times U(1)$ breaking). On the KK scale, the fermions should have essentially zero mass. To illustrate the situation, let us consider massless Dirac field in $d = 4 + D$ dimensions:

$$\mathrm{i}\gamma^M \nabla_M \Psi = \mathrm{i}\gamma^\mu \nabla_\mu \Psi + \mathrm{i}\gamma^\alpha \nabla_\alpha \Psi = 0.$$

The observed fermion masses are determined as the eigenvalues of the mass operator $M \equiv \mathrm{i}\gamma^\alpha \nabla_\alpha$, $M\Psi = \lambda\Psi$. The operator M on a compact space B_D has discrete eigenvalues, with the spacing set by the size of B_D: $\lambda = 0$ or $\sim 1/r$. Since $1/r$ is of the order of the Planck mass, the observed fermions must correspond to the zero modes $\lambda = 0$ of the mass operator.

However, in many interesting cases the mass operator has no zero eigenvalues. If B_D is Riemannian space, then

$$M^2\Psi = (-\nabla^2 + \tfrac{1}{4}R^D)\Psi \qquad \nabla^2 = \phi^{\alpha\beta}\nabla_\alpha\nabla_\beta.$$

The operator $-\nabla^2$ is positive definite, as can be seen from the relation

$$\int \mathrm{d}^D y\, \sqrt{|\phi|}\Psi^+(-\nabla^2)\Psi = \int \mathrm{d}^D y\, \sqrt{|\phi|}(\nabla_\alpha\Psi^+)(\nabla^\alpha\Psi)$$

where we have used the metricity condition $\nabla_\alpha \phi_{\beta\gamma} = 0$. Thus, the Dirac operator has no zero eigenmodes on a manifold B_D with positive scalar curvature R^D (Zee 1981). We can overcome this negative result by introducing torsion on B_D (Wu and Zee 1984). However, it seems that the massless fermions so obtained do not have realistic quantum numbers. Another possibility is to introduce extra gauge fields, which would modify the form of the Dirac operator.

An explicit analysis of the effective Dirac theory in four dimensions shows that parity violation only occurs in the massive sector. From the physical point of view, however, we would like to have parity violation in the zero mode sector. Thus, the problems of zero modes and parity violation should be treated simultaneously.

The zero modes of the operator M, if they exist, form representations of the (internal) symmetry group $G_0 = SU(3) \times SU(2) \times U(1)$. These representations do not depend on the helicity of fermions. This result, however, contradicts some of the basic facts about the observed quarks and leptons. We know that the fermions of a given helicity belong to a complex representation of G, i.e. left chiral components of the fermions transform differently from the right chiral components. The interesting possibility of obtaining massless fermions with left–right asymmetry is suggested by the following consequence of the index theorem: the zero modes of the Dirac operator with a definite helicity form a real representation if extra gauge fields are absent. Thus, in order to have the chance for the zero modes of fermions to form a complex representation of G we should introduce extra gauge fields (in topologically non-trivial configuration). Another solution of the same problem would be to start from a set of chiral fermions in D dimensions, which contains parts of both chiralities in four dimensions. Then, of course, the left–right asymmetry in four dimensions would not be a derived property, but rather a separate assumption.

Similar problems occur for the Rarita–Schwinger field of spin $\frac{3}{2}$. To obtain a realistic structure for the fermions is a very difficult problem in KK theory (Witten 1982, Mecklenburg 1983, Bailin and Love 1987).

Anomalies. An important property of chiral fermions is that their interaction with gauge bosons depends on the chirality, i.e. the coupling of gauge bosons to left and right chiral fermions is different. This feature of chiral fermions leads, in general, to anomalies (the quantum effects generate a breakdown of the classical gauge invariance). Since theories with anomalies cannot be quantized in the standard manner, we usually solve the problem by constructions in which the anomalies cancel. The requirement of anomaly cancellation imposes additional restrictions on the structure of fermion representations. In particular, the gauge group of extra gauge fields must be much larger and, as a consequence, we find a lot of additional fermions, most of which do not correspond to the physical reality (Bailin and Love 1987).

Super KK. Apart from the matter fields that are introduced in order to have spontaneous compactification, fermions are also introduced as additional matter fields, so that the unified picture for all fields is definitely lost. In the supersymmetric version of KK theory bosons and fermions are treated in a unified way. This is why, according to some opinions, 'the most attractive Kaluza–Klein theories are the supersymmetric ones' (Duff *et al* 1986).

A supersymmetric KK theory is based on supergravity formulated in more than four dimensions. As we have already mentioned, the structure of supergravity gives a restriction on the dimension of space: $d \leq 11$. In $d = 11$ the Lorentz group is $SO(1, 10)$, and the Dirac spinor has $2^5 = 32$ components, which corresponds to eight four-dimensional spinors in M_4. Thus, in four dimensions there are eight supersymmetry generators ($N = 8$), each of which can change the helicity for $\frac{1}{2}$, so that the helicity in a supersymmetric multiplet can take the values $\lambda = -2, -\frac{3}{2}, \ldots, \frac{3}{2}, 2$. In other words, the statements—(a) $\lambda \leq 2$ (in $d = 4$), (b) $N \leq 8$ (in $d = 4$) and (c) $d \leq 11$ (with $N = 1$)—are equivalent.

The form of supergravity in $d = 11$ is uniquely determined by the structure of supersymmetry (Cremmer *et al* 1978). Consistency requirements imply that it must have only one supersymmetry generator ($N = 1$). The theory contains the following fields: the vielbein $b^I{}_M$ (the graviton), the Rarita–Schwinger field Ψ_M (the gravitino) and the antisymmetric tensor field A_{MNR} (connection is not an independent dynamical degree of freedom). The symmetries of the action are: general covariance in $d = 11$, local $SO(1, 10)$ Lorentz symmetry, local $N = 1$ supersymmetry and local Abelian symmetry of the antisymmetric tensor field.

An antisymmetric tensor field with non-zero F^2 enables a natural mechanism for spontaneous compactification of $d = 11$ supergravity into $AdS_4 \times B_7$, where AdS_4 is a four-dimensional anti de Sitter space and B_7 is a compact space which may have several different forms (seven-sphere, squashed seven-sphere, etc). The separation of four physical dimensions of spacetime is a direct consequence of the existence of four indices of the field F^{MNRL} (Freund and Rubin 1980). If F^2 is zero, the compactification may give the ground state of the form $M_4 \times T_7$, where $T_7 = (S_1)^7$ is the seven-torus. This leads to $N = 8$ supergravity in $d = 4$.

The important questions of unbroken supersymmetries in four dimensions, the value of the cosmological constant and the effective symmetry of the massless sector are related to the structure of B_7.

Cosmology. Extra dimensions in KK theory have little direct influence on particle physics at low energies (compared to the Planck scale). In cosmology, where the energy scale is much higher, the situation might be different. In the very early Universe, at extremely high energies, we expect all spatial dimensions to have been of the same scale. How can we explain the evolution of the Universe to its present form $V_4 \times B_D$, with two spatial scales? In order to understand this problem dynamically, it is natural to start with the metric

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a(t)g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu - b(t)g_{\alpha\beta}\,\mathrm{d}x^\alpha\,\mathrm{d}x^\beta$$

where $a(t)$ is the scale of the physical three-space and $b(t)$ the scale of the compact space B_D. There are explicit models which describe the mechanisms by which the two scales could have been developed in the dynamical evolution of the Universe. According to these models, at some moment in the past scale $a(t)$ started to increase more rapidly than scale $b(t)$. In many of these considerations we use fine tuning of the initial conditions, which is regarded as an unattractive explanation. Although many interesting effects are found, the importance of KK models in cosmology seems not to have been completely investigated (Bailin and Love 1987).

Instead of a conclusion. KK theory is 80 years old and it has retained its original charming attractiveness but its physical predictions were never sufficiently realistic. Why is this so? Looking backwards from today's perspective, we can note that the original five-dimensional theory was *premature*. The structure of the fundamental interactions was insufficiently known and the idea of spontaneous symmetry breaking had not yet found its place in physics.

What we do not know is whether the time is finally ripe for Kaluza–Klein theory, whether there still are crucial things we do not know, or whether the idea is completely wrong. Time will tell. (Witten 1981b).

Exercises

1. (a) Find the Christoffel connection $\hat{\Gamma}^M_{NR}$ corresponding to the metric $(10.4a)$.
   (b) Then, prove the relation $\widehat{R} = R + \frac{1}{4}F_{\mu\nu}F^{\mu\nu}$.
2. Derive the factorization property for the determinant of metric $(10.4a)$: $\hat{g} = -g$, where $\hat{g} = \det(\hat{g}_{MN})$, $g = \det(g_{\mu\nu})$.
3. Find the field equations following from the reduced action $(10.6a)$, and compare them with the five-dimensional equations $\widehat{R}_{\mu\nu} = 0$, $\widehat{R}_{\mu 5} = 0$ and $\widehat{R}_{55} = 0$.
4. Consider the scalar field theory in $1+1$ dimensions:

$$\mathcal{L} = \tfrac{1}{2}(\partial_t\phi)^2 - \tfrac{1}{2}(\partial_x\phi)^2 - V(\phi) \qquad V(\phi) = (\lambda/4)(\phi^2 - v^2)^2.$$

   Prove the following statements:
   (a) Excitations around the classical vacuum $\phi_0 = \pm v$ are physical fields of mass $m^2 = 2\lambda v^2$.
   (b) Functions $\phi_K(z) = \pm v\tanh(z)$, $z = m(x-c)/2$, where $c =$ constant, are static solutions of the field equations (the kink and antikink, respectively).
   (c) Both solutions have the same static energy (mass): $M_K = m^3/3\lambda$.
   (d) $\eta_0(z) = \pm v/\cosh^2(z)$ is the zero mode perturbation around ϕ_K.
5. (a) Find the Christoffel connection corresponding to metric $(10.10a)$.
   (b) Calculate the five-dimensional scalar curvature $\widehat{R}$ when the four-dimensional space V_4 is flat, $g_{\mu\nu} = \eta_{\mu\nu}$. Then find the general form of $\widehat{R}$.
6. Show that under Weyl rescalings $g_{\mu\nu} = \rho\bar{g}_{\mu\nu}$, the Christoffel connection of a d-dimensional Riemann space transforms according to

$$\Gamma^\mu_{\nu\lambda} = \bar{\Gamma}^\mu_{\nu\lambda} + \tfrac{1}{2}(\delta^\mu_\nu\rho_\lambda + \delta^\mu_\lambda\rho_\nu - \bar{g}_{\nu\lambda}\rho^\mu)$$

   where $\rho_\mu = \partial_\mu \ln\rho$, $\rho^\mu = \bar{g}^{\mu\nu}\rho_\nu$. Then derive the transformation rule of the scalar curvature:

$$R(g) = \rho^{-1}[R(\bar{g}) + (1-d)\rho^{-1}\bar{\Box}\rho + \tfrac{1}{4}(7d - 6 - d^2)\rho^{-2}\bar{g}^{\mu\nu}\partial_\mu\rho\partial_\nu\rho].$$

7. (a) Assuming the rescaling of variables according to $(10.16a)$, prove the following relations:

$$\sqrt{-g}\sqrt{-\phi}R = \sqrt{-\bar{g}}[\bar{R} + \bar{\Box}\ln(-\bar{\phi}) - \tfrac{1}{6}\bar{\phi}^{-2}\bar{g}^{\mu\nu}\partial_\mu\bar{\phi}\partial_\nu\bar{\phi}]$$
$$\sqrt{-g}\sqrt{-\phi}(-\phi)F_{\mu\nu}F^{\mu\nu} = \sqrt{-\bar{g}}(-\bar{\phi})F_{\mu\nu}F^{\mu\nu}.$$

Then derive action (10.16b) from (10.15).

(b) Express Christoffel connection (10.14a) in terms of the new variables.

8. A point test particle of mass m moves along the geodesics of the Riemann space V_5 with metric (10.18).
   (a) Show that the quantity $(B_\mu u^\mu + u^5) \equiv \hat{q}/m$ is a constant of the motion.
   (b) Derive the equation describing the effective four-dimensional motion of the particle.
9. A point test particle of mass m moves along the geodesics of the Riemann space V_5 with metric (10.10a).
   (a) Show that $\xi_M = \hat{g}_{M5} \equiv (\phi B_\mu, \phi)$ is the Killing vector of V_5.
   (b) Show that the quantity $\xi_M u^M = \phi(B_\mu u^\mu + u^5) \equiv -\hat{q}/m$, where $u^M = \mathrm{d}z^M/\mathrm{d}\tau$, is a constant of the motion.
   (c) Derive the equation describing the effective four-dimensional motion of the particle.
10. Find the effective four-dimensional action of the scalar field theory (10.20a), assuming that the metric has the form (10.10a) with $g_{\mu\nu} = \eta_{\mu\nu}$, and without y dependence.
11. Consider the Dirac action in Riemann space V_5, assuming that the pentad is given by equation (10.22).
   (a) Show that the non-vanishing connection coefficients Δ_{IJM} are proportional to F_{ij}, and calculate the form of the term $\gamma^K\omega_K$.
   (b) Find the effective four-dimensional theory.
12. (a) Derive the Kac–Moody algebra (10.29).
   (b) Show that the generators P^0_μ, $M^0_{\mu\nu}$, L_1, L_0, and L_{-1} define the $P_4 \times SO(1,2)$ algebra.
13. Show that if the ground state metric is of the form (10.31), then $\widehat{R} = R_4 + R_D$.
14. (a) Find the displacement vector $(\mathrm{d}x, \Delta y)$ in the spacetime V_4, orthogonal to the layer V_D. Then determine the metric $g_{\mu\nu}$ of V_4.
   (b) Using vielbein (10.32) derive the form of the metric (10.33a).
   (c) Find the connection between local coordinates (X, Y) associated with the basis $(\boldsymbol{E}_\mu, \boldsymbol{E}_\alpha) = (b^i{}_\mu \hat{\boldsymbol{e}}_i, \hat{\boldsymbol{e}}_\alpha)$, and the original coordinates (x, y). Show that in these coordinates $\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}X^\mu\,\mathrm{d}X^\nu + \phi_{\alpha\beta}\,\mathrm{d}Y^\alpha\,\mathrm{d}Y^\beta$.
15. Show that the Laplacian $\widehat{\Box} = \hat{g}^{MN}\nabla_M\nabla_N$ corresponding to the metric (10.33) with $g_{\mu\nu} = \eta_{\mu\nu}$, can be expressed as
$$\widehat{\Box} = \eta^{\mu\nu}(\partial_\mu - B^\alpha_\mu\partial_\alpha)(\partial_\nu - B^\beta_\nu\partial_\beta) + \Box_y$$
   where $\Box_y$ is the Laplacian on B_D.
16. Find the transformation of the field B^α_μ under the infinitesimal coordinate transformations: $\delta x^\mu = 0$, $\delta y^\alpha = \varepsilon^a(x, y)E^\alpha_a(y)$. What happens if $\varepsilon^a = \varepsilon^a(x)$?
17. (a) Derive the factorization property for the determinant of metric (10.38): $\hat{g} = g\phi$, where $\hat{g} = \det(\hat{g}_{MN})$, $g = \det(g_{\mu\nu})$, $\phi = \det(\phi_{\alpha\beta})$.

(b) Find the scalar curvature $\widehat{R}$ for the metric (10.38) in the case $g_{\mu\nu} = \eta_{\mu\nu}$.

18. Starting from metric (10.44a) with $g_{\mu\nu} = \eta_{\mu\nu}$, find the $(4+D)$-dimensional action for the scalar field $\phi_{\alpha\beta}(x, y)$.
19. (a) Derive field equations and the energy–momentum tensor for the action (10.51), describing extra gauge fields.

    (b) Check that (10.52) is a solution to the equations of motion, and calculate the value of the cosmological constant in the related KK theory.

Chapter 11

String theory

The description of hadronic processes in the 1960s was based on models with an infinite number of states, lying on linear Regge trajectories. An important theoretical realization of these ideas was given by the Veneziano (1968) model, which actually describes the scattering of one-dimensional objects, *strings*. The inclusion of fermions into the theory laid the ground for supersymmetry (Ramond 1971, Neveu and Schwarz 1971). A radical change in the interpretation of the model has been suggested in the 1970s: the hadronic scale is replaced with the Planck scale; and the string model is interpreted as a framework for the unification of all basic interactions including gravity (Neveu and Scherk 1972, Scherk and Schwarz 1974). The importance of these ideas became completely clear in 1980s, after the discovery that string theory was free of quantum anomalies (Green and Schwarz 1984, 1985).

Theories of fundamental interactions are usually formulated starting from certain symmetry principles, which are then used to construct an invariant action and derive the S-matrix describing physical processes. In complete contrast, the first step in string theory was made by Veneziano, who guessed the form of the hadronic scattering amplitude. In a certain sense, that was an 'answer looking for a question' (Kaku 1985).

The significant success of string theory in providing a consistent treatment of *quantum gravity* led to a renewed interest for identifying the underlying symmetry principles of strings, and construct the related *covariant* field theory. Although, in principle, we can discover all the relevant properties of a theory by using a given gauge, this may be difficult in practice. Our understanding of non-perturbative semiclassical phenomena and spontaneous symmetry breaking has been developed in the gauge-invariant framework. Let us mention that standard quantum consistency requirements imply that strings live in the spacetimes of critical dimensions $D = 26$ or 10. The transition to the effective four-dimensional theory is a non-perturbative effect. Thus, the identification of gauge symmetries plays an important role in the formulation of realistic models. We also expect the quantum properties of the theory to be more transparent in a covariant formulation.

We should note, however, that the construction of the covariant field theory is only a first step towards a full understanding of the string *geometry*. The situation could be compared to one in which a field theorist would study the covariant GR perturbatively, without having any idea about the Riemannian geometry behind the theory.

It is our intention here to present an introduction to the *covariant field theory* of *free bosonic strings*, which will cover the main features of the classical theory. Our exposition is based on the gauge-invariant Hamiltonian formalism, which clearly relates the two-dimensional reparametrization invariance, given in the form of the Virasoro conditions, with gauge invariances in field theory. The bosonic string is used to illustrate the common features of all string models, without too many technical complications. In particular, we shall see how the gauge fields, the *photon* and the *graviton*, are obtained from string dynamics. Interacting models and supersymmetric formulations can be described in a similar way but are technically more involved.

11.1 Classical bosonic strings

The fundamental constituents of standard field theories are point particles. The replacement of point particles by one-dimensional extended objects—strings—leads to important changes in our understanding of the nature of the basic interactions. The root of these changes can be traced back to the classical dynamics of strings (Scherk 1975, Sundermeyer 1982, West 1986b, Brink and Henneaux 1988, Bailin and Love 1994).

The relativistic point particle

Many of the concepts used in discussing string theory are already present in the simpler case of relativistic point particles. We start, therefore, by discussing this model first.

Classical mechanics. A point particle moving in spacetime M_4 describes the world line $x^\mu = x^\mu(\tau)$, where τ is the evolution parameter. The relativistic action that describes the motion of a point particle is proportional to the length of its world line:

$$I = -m \int \mathrm{d}\tau \, \sqrt{\eta_{\mu\nu}\dot{x}^\mu \dot{x}^\nu} \equiv \int \mathrm{d}\tau \, L \tag{11.1}$$

where $\dot{x}^\mu = \mathrm{d}x^\mu/\mathrm{d}\tau$. The action is invariant under reparametrizations of the world line: $\tau \to \tau' = \tau + \varepsilon(\tau)$, $\delta x^\mu = x^\mu(\tau') - x^\mu(\tau) = \varepsilon(\tau)\dot{x}^\mu$.

In order to define the Hamiltonian formalism, which is the first step in the quantization procedure, we introduce the canonical momenta by†

$$p_\mu = \frac{\partial L}{\partial \dot{x}^\mu} = -\frac{m\dot{x}_\mu}{\sqrt{\dot{x}^2}} \qquad \dot{x}^2 \equiv \eta_{\mu\nu}\dot{x}^\mu \dot{x}^\nu .$$

Since the momenta are homogeneous functions of velocities, they satisfy the constraint

$$\phi \equiv p^2 - m^2 \approx 0$$

and the canonical Hamiltonian vanishes: $H_\mathrm{c} \equiv p_\mu \dot{x}^\mu - L \approx 0$.

In the presence of constraints, the equations of motion are generated by the total Hamiltonian,

$$H_\mathrm{T} \equiv H_\mathrm{c} + v(\tau)\phi \approx v(\tau)\phi$$

where $v(\tau)$ is an arbitrary multiplier. Using the basic Poisson brackets $\{x^\mu, p_\nu\} = \delta^\mu_\nu$, we find that the consistency of the primary constraint ϕ is automatically satisfied for every $v(\tau)$, $\dot{\phi} \equiv \{\phi, H_\mathrm{T}\} \approx 0$, and no further constraints are generated in the theory.

The vanishing of the canonical Hamiltonian and the existence of an arbitrary multiplier in H_T are directly related to the *reparametrization invariance* of the theory, which is clearly seen from the Hamiltonian equations of motion:

$$\begin{aligned} \dot{x}^\mu &= \{x^\mu, H_\mathrm{T}\} = 2v(\tau)p^\mu \\ \dot{p} &= \{p^\mu, H_\mathrm{T}\} = 0. \end{aligned}$$

The presence of $v(\tau)$ is a sign of the arbitrariness in the choice of τ (in order for a non-trivial motion to exist, the multiplier $v(\tau)$ must be different from zero).

The reparametrization invariance can be easily broken by imposing a suitable gauge condition, such as $\Omega \equiv x^0(\tau) - \tau \approx 0$, for instance. The consistency requirement on Ω determines $v(\tau)$:

$$\dot{\Omega} \equiv \{\Omega, H_\mathrm{T}\} + \frac{\partial \Omega}{\partial \tau} = 2v(\tau)p^0 - 1 = 0$$

whereupon the reparametrization invariance is lost.

The first quantization. The canonical transition from the classical to *quantum* mechanics in a gauge invariant theory can be realized in the following way (Dirac 1964):

(a) the basic dynamical variables are elevated to operators:

$$x(\tau) \to \hat{x}(\tau) \qquad p(\tau) \to \hat{p}(\tau)$$

(b) the PBs go over into commutators with an appropriate factor of $-\mathrm{i}$:

$$\{A, B\} \to -\mathrm{i}[\hat{A}, \hat{B}]$$

† This choice yields different sign compared to the usual physical momenta.

(c) first class constraints become conditions on the physical states:

$$\hat{\phi}|\psi\rangle = 0$$

(d) the physical state vectors satisfy the 'Schrödinger equation':

$$\mathrm{i}\frac{\partial}{\partial\tau}|\psi\rangle = \hat{H}_{\mathrm{T}}|\psi\rangle.$$

As $\hat{\phi}$ is the only constraint in the theory, it is first class. The quantum condition on the physical states has the form

$$\hat{\phi}|\psi\rangle \equiv (\hat{p}^2 - m^2)|\psi\rangle = 0. \tag{11.2}$$

Since H_{T} is proportional to $\hat{\phi}$, the right-hand side of the 'Schrödinger equation' vanishes, and we find that $|\psi\rangle$ does not depend explicitly on τ. The essential dynamical information is contained in the quantum dynamical variables $\hat{x}(\tau)$ and $\hat{p}(\tau)$, which satisfy the Heisenberg equations of motion.

Classical field theory. If we realize the quantum dynamical variables in the coordinate representation, $\hat{x}^\mu \to x^\mu$, $\hat{p}_\mu \to -\mathrm{i}\partial_\mu$, the state vector $|\psi\rangle$ becomes the wavefunction $\psi(x) = \langle x|\psi\rangle$, which is independent of τ and satisfies the Klein–Gordon equation (11.2): $(-\Box - m^2)\psi(x) = 0$. The action that leads to this field equation is given by

$$I = \int \mathrm{d}^4x\, \psi^+(x)(-\Box - m^2)\psi(x). \tag{11.3}$$

Classical field theory is an infinite system of classical point particles that occupy each point $\boldsymbol{x}$ of the three-dimensional space. It is interesting to observe that the original reparametrization invariance is absent from the field theory; its only remnant is the Klein–Gordon field equation.

The second quantization. The canonical description of the classical field theory (11.3), or one of its generalizations that may include interaction terms, is realized in the phase space $(\psi(\boldsymbol{x}), \pi(\boldsymbol{x}))$. Like any other dynamical system, it can be quantized in the same manner as that described earlier. This procedure is now called the second quantization, and the related quantum field theory describes an infinite system of quantum particles.

We now wish to see how these steps can be repeated for strings.

Action principle for the string

A string in D-dimensional Minkowski space M_D is a one-dimensional extended object that sweeps out in time a two-dimensional world sheet Σ, $x^\mu = x^\mu(\tau, \sigma)$, labelled by local coordinates $\xi^\alpha = (\tau, \sigma)$. The string curve at fixed τ may be

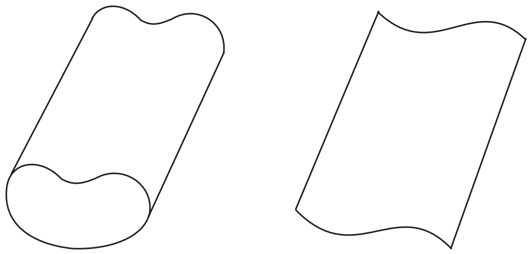

Figure 11.1. World sheets of closed and open strings.

closed or open (closed or open strings as in figure 11.1), and the spatial coordinate σ in both cases takes values in the range $[0, \pi]$, by convention.

The action of the relativistic string is taken to be proportional to the area of its world sheet Σ (Nambu 1970, Gotto 1971),

$$I = -\frac{1}{2\pi\alpha'}\int \mathrm{d}^2\xi\,\sqrt{-\gamma} \equiv \int \mathrm{d}^2\xi\,\mathcal{L}(x(\xi), \dot{x}(\xi)) \tag{11.4a}$$

where γ is the determinant of the world sheet metric $\gamma_{\alpha\beta}$ induced by the embedding of Σ in M_D, $\gamma_{\alpha\beta} = \eta_{\mu\nu}\partial_\alpha x^\mu \partial_\beta x^\nu \equiv \partial_\alpha x \cdot \partial_\beta x$, and α' is a constant of the dimension $[m^{-2}]$.

An alternative expression for the string action is given by

$$I' = -\frac{1}{4\pi\alpha'}\int \mathrm{d}^2\xi\,\sqrt{-g}g^{\alpha\beta}\partial_\alpha x \cdot \partial_\beta x \tag{11.4b}$$

where the metric $g_{\alpha\beta}$ is treated as an independent dynamical variable (Polyakov 1981). This expression, in contrast to I, is quadratic in the derivatives of x^μ, which is more convenient in the path integral quantization. We can think of I' as the action describing D massless, scalar fields $x^\mu(\xi)$ in two dimensions, interacting with the gravitational field $g_{\alpha\beta}$.

The new action I' is classically equivalent to I. Indeed, the equations of motion obtained by varying I' with respect to $g_{\alpha\beta}$ are

$$\gamma_{\alpha\beta} - \tfrac{1}{2}g_{\alpha\beta}g^{\gamma\delta}\gamma_{\gamma\delta} = 0.$$

These equations imply the relation $\sqrt{-\gamma} = \frac{1}{2}\sqrt{-g}g^{\gamma\delta}\gamma_{\gamma\delta}$, which, when substituted back into I', yields exactly I‡.

The action I' is, by construction, *invariant* under

(a) *global* Poincaré transformations in M_D:

$$\begin{aligned} \delta x^\mu &= \omega^\mu{}_\nu x^\nu + a^\mu \\ \delta\xi^\alpha &= 0 \qquad \delta g_{\alpha\beta} = 0 \end{aligned} \tag{11.5a}$$

‡ The proof of the quantum equivalence is not so simple, and holds only for $D = 26$.

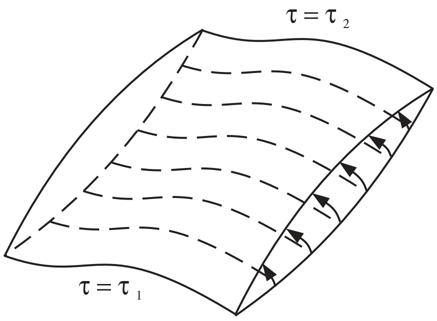

Figure 11.2. The variation of the world sheet for open strings.

(b) *local* world sheet reparametrizations:

$$\begin{aligned} \delta\xi^\alpha &\equiv \xi'^\alpha - \xi^\alpha = -\varepsilon^\alpha(\xi) \\ \delta x^\mu &= 0 \qquad (\delta_0 x^\mu = \varepsilon^\alpha \partial_\alpha x^\mu) \\ \delta_0 g_{\alpha\beta} &= \partial_\alpha \varepsilon^\gamma g_{\gamma\beta} + \partial_\beta \varepsilon^\gamma g_{\alpha\gamma} + \varepsilon^\gamma \partial_\gamma g_{\alpha\beta} \end{aligned} \tag{11.5b}$$

(note that x^μ are scalar fields on Σ) and

(c) *local* Weyl rescalings:

$$\begin{aligned} \delta g_{\alpha\beta} &= \Lambda(\xi) g_{\alpha\beta} \\ \delta x^\mu &= 0, \qquad \delta\xi^\alpha = 0. \end{aligned} \tag{11.5c}$$

Local Weyl invariance is an accidental property of the geometric action in two dimensions, and is absent when we consider higher-dimensional objects, such as membranes. We shall see that it imposes important restrictions on the structure of the theory.

In our further exposition we shall use the action (11.4*a*), although all the results can also be obtained from the quadratic form (11.4*b*).

Consider the action I where the initial and final positions of the string (at $\tau = \tau_1$ and $\tau = \tau_2$) are fixed and $0 \le \sigma \le \pi$ (figure 11.2). Introducing the notation

$$\pi_\mu = \frac{\partial\mathcal{L}}{\partial\dot{x}^\mu} \qquad \pi_\mu^{(\sigma)} = \frac{\partial\mathcal{L}}{\partial x'^\mu}$$

where $\dot{x} = \mathrm{d}x/\mathrm{d}\tau$, $x' = \mathrm{d}x/\mathrm{d}\sigma$, the variation of the action with respect to $x^\mu(\tau,\sigma)$ yields

$$\begin{aligned} \delta I = &\int_0^\pi \mathrm{d}\sigma\, \pi_\mu \delta x^\mu \Big|_{\tau_1}^{\tau_2} + \int_{\tau_1}^{\tau_2} \mathrm{d}\tau\, \pi_\mu^{(\sigma)} \delta x^\mu \Big|_0^\pi \\ &- \int_{\tau_1}^{\tau_2} \mathrm{d}\tau \int_0^\pi \mathrm{d}\sigma\, (\partial_\tau \pi_\mu + \partial_\sigma \pi_\mu^{(\sigma)}) \delta x^\mu = 0. \end{aligned}$$

At all points in the interior of the world sheet, the variations δx^μ are arbitrary and independent, and the last term in δI implies the following *equations of motion*:

$$\partial_\tau \pi_\mu + \partial_\sigma \pi_\mu^{(\sigma)} = 0. \tag{11.6}$$

The consistency of the variational principle depends on the vanishing of the boundary terms, and can be ensured by imposing suitable boundary conditions. Since the initial and final string configurations, $x^\mu(\tau_1, \sigma)$ and $x^\mu(\tau_2, \sigma)$, are held fixed, the first term in δI vanishes. As regards the form of the second term, it is convenient to distinguish between closed and open strings.

(i) For the closed string, i.e.

$$x^\mu(\tau, 0) = x^\mu(\tau, \pi) \tag{11.7a}$$

we have $\delta x^\mu(\tau, 0) = \delta x^\mu(\tau, \pi)$, $\pi_\mu^{(\sigma)}(\tau, 0) = \pi_\mu^{(\sigma)}(\tau, \pi)$, and the second term in δI vanishes.

(ii) For the open string, the situation is more sensitive. If $\delta x^\mu(\tau, 0)$ and $\delta x^\mu(\tau, \pi)$ are independent variations, it is natural to demand the following *boundary conditions*:

$$\pi_\mu^{(\sigma)} = 0 \qquad \text{at } \sigma = 0, \pi. \tag{11.7b}$$

This means that no momentum can flow out of the ends of the string.

Thus, the Lagrangian equations of motion have the usual form for both closed and open strings, provided we impose these boundary conditions.

Hamiltonian formalism and symmetries

We now introduce the constrained Hamiltonian formalism, which is particularly suitable for studying the gauge symmetries of the string theory.

Hamiltonian and its constraints. The canonical momenta

$$\pi_\mu(\sigma) \equiv \frac{\partial \mathcal{L}}{\partial \dot{x}^\mu(\sigma)} = \frac{1}{2\pi\alpha'} \frac{\dot{x}_\mu x'^2 - x'_\mu(\dot{x} \cdot x')}{\sqrt{-\gamma}}$$

are homogeneous functions of the velocities. They satisfy two primary constraints:

$$\begin{aligned} G_0(\sigma) &\equiv \pi^2 + x'^2/(2\pi\alpha')^2 \approx 0 \\ G_1(\sigma) &\equiv \pi \cdot x' \approx 0. \end{aligned} \tag{11.8}$$

The canonical Hamiltonian vanishes, $H_\mathrm{c} = \int \mathrm{d}\sigma\, [\pi_\mu(\sigma)\dot{x}^\mu(\sigma) - \mathcal{L}] \approx 0$, and the total Hamiltonian contains two arbitrary multipliers,

$$H_\mathrm{T} = \int_0^\pi \mathrm{d}\sigma\, [v^0(\sigma)G_0(\sigma) + v^1(\sigma)G_1(\sigma)] \tag{11.9}$$

as we expected it would.

Using the basic PBs, $\{x^\mu(\sigma), \pi_\nu(\sigma')\} = \delta^\mu_\nu \delta$, where $\delta \equiv \delta(\sigma, \sigma')$, we can check that the primary constraints G_0 and G_1 are first class, since

$$\begin{aligned}
\{G_0(\sigma), G_0(\sigma')\} &= \frac{1}{(\pi\alpha')^2}[G_1(\sigma) + G_1(\sigma')]\partial_\sigma\delta \\
\{G_1(\sigma), G_0(\sigma')\} &= [G_0(\sigma) + G_0(\sigma')]\partial_\sigma\delta \\
\{G_1(\sigma), G_1(\sigma')\} &= [G_1(\sigma) + G_1(\sigma')]\partial_\sigma\delta.
\end{aligned} \tag{11.10}$$

The consistency conditions are automatically satisfied, without producing any new constraints.

The presence of two arbitrary multipliers $v^0(\sigma)$ and $v^1(\sigma)$ in H_T is related to the general reparametrization invariance on Σ.

The conformal gauge. The reparametrization invariance can be fixed by imposing convenient gauge conditions. This can be done so as to maintain the manifest Lorentz covariance in M_D, which is of particular interest for the construction of Lorentz covariant field theory.

The Hamiltonian equations of motion read:

$$\dot{x}^\mu \equiv \{x^\mu, H_T\} = 2v^0\pi^\mu + v^1 x'^\mu \tag{11.11a}$$

$$\dot{\pi}^\mu \equiv \{\pi^\mu, H_T\} = \frac{2}{(2\pi\alpha')^2}\partial_\sigma(v^0 x'^\mu) + \partial_\sigma(v^1\pi^\mu). \tag{11.11b}$$

They are complicated and intractable. In order to proceed, we use the fact that v^0 and v^1 are arbitrary multipliers, and choose their forms so as to simplify the equations of motion. One possible choice is

$$v^0 = -\pi\alpha' \qquad v^1 = 0. \tag{11.12}$$

With this choice the Hamiltonian equations yield the well known M_D-covariant wave equation for x^μ:

$$\ddot{x}^\mu - x''^\mu = 0. \tag{11.13}$$

It is clear that any definite choice of arbitrary multipliers in H_T is equivalent to a gauge condition, as it represents a restriction on the form of general reparametrizations. In order to find the geometric meaning of conditions (11.12), let us write the first equation of motion in the form

$$\pi^\mu = -\dot{x}^\mu/2\pi\alpha'. \tag{11.14}$$

Combining this with the constraints G_0 and G_1 we find the relations

$$\begin{aligned}
\gamma_{00} + \gamma_{11} &\equiv \dot{x}^2 + x'^2 \approx 0 \\
\gamma_{01} &\equiv \dot{x}\cdot x' \approx 0
\end{aligned}$$

which imply

$$\mathrm{d}s^2 = \gamma_{\alpha\beta}\,\mathrm{d}\xi^\alpha\,\mathrm{d}\xi^\beta = \gamma_{00}(\mathrm{d}\tau^2 - \mathrm{d}\sigma^2). \tag{11.15}$$

Hence, these gauge conditions do not fix the coordinates on Σ completely. The residual reparametrizations are those that leave the form of interval (11.15) invariant; they are known as the *conformal* transformations in Σ. The total Hamiltonian now has a simpler form:

$$H_\mathrm{T} = -\pi\alpha' \int_0^\pi \mathrm{d}\sigma\, G_0(\sigma). \tag{11.16}$$

The gauge fixing that leads to the metric $\gamma_{\alpha\beta} = \sqrt{-\gamma}\,\eta_{\alpha\beta}$, as in equation (11.15), is called the conformal gauge; it is consistent with Lorentz covariance in M_D. Note that conditions (11.12) are in agreement with $x'^2 < 0$, $\dot{x}^2 > 0$, as can be seen by comparing equation (11.14) with the definition of π^μ.

Poincaré symmetry. Let us now return to global Poincaré symmetry (11.5a), which is related to the fact that the string is embedded in M_D. The quantities $x^\mu(\xi)$ ($\mu = 0, 1, \ldots, D-1$) are scalar fields with respect to the coordinate transformations in Σ. Global Poincaré transformations look like internal symmetry transformations in Σ, as they relate fields $x^\mu(\xi)$ at the same point ξ in Σ.

Starting from the known form of global Poincaré symmetry, we can find the conserved currents associated with the translations and Lorentz rotations. The related conserved charges§ are

$$\begin{aligned} P^\mu &= \int \mathrm{d}\sigma\, \pi^\mu(\sigma) \\ M^{\mu\nu} &= \int \mathrm{d}\sigma\, [x^\mu(\sigma)\pi^\nu(\sigma) - x^\nu(\sigma)\pi^\mu(\sigma)] \end{aligned} \tag{11.17}$$

and satisfy the PB algebra of the Poincaré group.

The dimension. So far, the dimension D of the space M_D in which the string moves has remained completely arbitrary. It is interesting to mention here that the usual consistency requirements in quantum theory lead to $D = 26$. The transition to an effective theory in four dimensions is realized through the mechanism of spontaneous compactification, and represents one of the most difficult problems of realistic string theory.

11.2 Oscillator formalism

Since the constraints are quadratic in the fields, a natural way to analyse the string dynamics is Fourier expansion (Scherk 1975, Sundermeyer 1982, West 1986b, Brink and Henneaux 1988, Bailin and Love 1994).

§ The physical values of the linear and angular momentum have opposite signs.

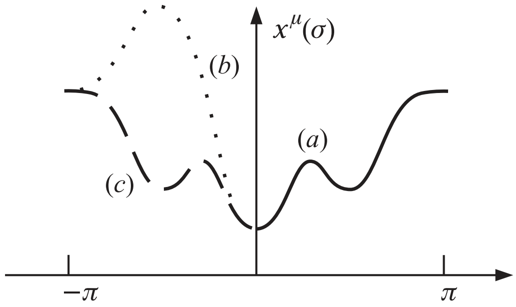

Figure 11.3. Possible extensions of open strings.

Open string

Boundary conditions (11.7b) for an open string in the conformal gauge take the form

$$x'_{\mu} = 0 \qquad \text{at} \quad \sigma = 0, \pi. \tag{11.18}$$

We now want to find a Fourier expansion of the real function $x^{\mu}(\tau, \sigma)$, that is defined on the basic interval $0 \le \sigma \le \pi$ for a fixed value of τ, and satisfies boundary conditions (11.18). This can be done in several equivalent ways:

(a) We can Fourier expand $x^{\mu}(\tau, \sigma)$, defined on $[0, \pi]$, in terms of $\cos 2n\sigma$ and $\sin 2n\sigma$.
(b) We can first extend the definition of $x^{\mu}(\tau, \sigma)$ to the region $[-\pi, \pi]$ in accordance with the boundary conditions, and then expand in $\cos n\sigma$ and $\sin n\sigma$.
(c) If the extension is *symmetric*, the Fourier expansion contains only $\cos n\sigma$.
(d) An antisymmetric extension is only possible if $x(\tau, 0) = 0$.

It is clear that the simplest choice is (c). Hence, the Fourier expansion of the field $x(\tau, \sigma)$ has the form (L.3) (appendix L)

$$x^{\mu}(\tau, \sigma) = \sum_{n} x_n^{\mu} \mathrm{e}^{\mathrm{i}n\sigma} = x_0^{\mu} + 2 \sum_{n \ge 1} x_n^{\mu} \cos n\sigma \tag{11.19a}$$

where $x_n^{\mu} = x_{-n}^{\mu}$ ($x_{-n}^{\mu} = x_n^{\mu *}$). Similarly, the related expansion of the momentum variable $\pi(\tau, \sigma)$ reads

$$\pi^{\mu}(\tau, \sigma) = \sum_{n} \pi_n^{\mu} \mathrm{e}^{\mathrm{i}n\sigma} = \pi_0^{\mu} + 2 \sum_{n \ge 1} \pi_n^{\mu} \cos n\sigma \tag{11.19b}$$

where $\pi_n^{\mu} = \pi_{-n}^{\mu}$ ($\pi_{-n}^{\mu} = \pi_n^{\mu *}$). The time dependence of coefficients x_n^{μ} and π_n^{μ} is determined by the equations of motion.

The basic PBs between $x(\sigma)$ and $\pi(\sigma)$ (the dependence on τ is omitted for simplicity) have the form

$$\{x^\mu(\sigma), \pi_\nu(\sigma')\} = \delta^\mu_\nu \delta(\sigma, \sigma'). \qquad (11.20a)$$

Since $x(\sigma)$ and $\pi(\sigma)$ are symmetric functions on the basic interval $[-\pi, \pi]$, expression $\delta(\sigma, \sigma')$ should be defined as the symmetric delta function, $\delta(\sigma, \sigma') = \delta_S(\sigma, \sigma')$, as in equation (L.7) (appendix L). We shall consider instead the expressions (11.19a, b) on the interval $[0, \pi]$, and, accordingly, change the normalization of the delta function:

$$\delta(\sigma, \sigma') \to 2\delta_S(\sigma, \sigma').$$

This implies

$$\{x^\mu_n, \pi^\nu_m\} = \eta^{\mu\nu} \frac{1}{\pi}(\delta_{n,m} + \delta_{n,-m}) \qquad (n, m \geq 0). \qquad (11.20b)$$

Let us now introduce the quantity

$$\Pi^\mu(\sigma) \equiv \pi^\mu + \frac{x'^\mu}{2\pi\alpha'} \qquad -\pi \leq \sigma \leq \pi. \qquad (11.21a)$$

It is interesting to note that the two constraints $G_0(\sigma)$ and $G_1(\sigma)$ on the interval $[0, \pi]$ can be unified into one constraint on $[-\pi, \pi]$:

$$\Pi^2(\sigma) = G_0(\sigma) + G_1(\sigma)/\pi\alpha'$$

where $G_0(-\sigma) = G_0(\sigma)$, $G_1(-\sigma) = -G_1(\sigma)$. Fourier expansion of Π^μ reads:

$$\Pi^\mu(\sigma) = \frac{-1}{\pi\sqrt{2\alpha'}} \sum_n a^\mu_n \mathrm{e}^{\mathrm{i}n\sigma} \qquad (11.21b)$$

where

$$-a^\mu_n = \left(\pi^\mu_n + \frac{\mathrm{i}n}{2\pi\alpha'} x^\mu_n\right) \pi\sqrt{2\alpha'}$$

or, equivalently,

$$-\pi^\mu_n = \frac{1}{\pi\sqrt{2\alpha'}} \frac{1}{2}(a^\mu_n + a^\mu_{-n})$$
$$x^\mu_n = \frac{\mathrm{i}\sqrt{2\alpha'}}{n} \frac{1}{2}(a^\mu_n - a^\mu_{-n}).$$

The basic PBs can be expressed in terms of a_n as

$$\{a^\mu_n, a^\nu_{-m}\} = \mathrm{i}n\delta_{n,m}\eta^{\mu\nu} \qquad (n, m \geq 1) \qquad (11.22)$$

which suggests that in quantum theory a_n and $a^*_m = a_{-m}$ $(n, m \geq 1)$ will be the *annihilation* and *creation* operators of the string modes (up to the normalization factors $\sqrt{n}$, $\sqrt{m}$, respectively).

The *time dependence* of the string modes is determined by the equations of motion (11.13):

$$\ddot{x}^\mu_n + n^2 x^\mu_n = 0.$$

Using $\pi_n = -\dot{x}_n/2\pi\alpha'$ and the definition of a_n we find that

$$a_n(\tau) = a_n(0)\mathrm{e}^{-\mathrm{i}n\tau} \qquad (n \neq 0)$$

while the time dependence of the zero modes is

$$x^\mu_0 = x^\mu + p^\mu\tau \qquad \pi^\mu_0 = -p^\mu/2\pi\alpha'$$

where the variables (x^μ, p^μ) describe the centre-of-mass motion, and satisfy

$$\{x^\mu, p^\nu\} = -\eta^{\mu\nu}(2\alpha').$$

The general solution of the equations of motion for an open string is given by

$$\begin{aligned} x^\mu(\sigma) &= x^\mu_0 + \mathrm{i}\sqrt{2\alpha'}\sum_{n\geq 1}\frac{1}{n}[a^\mu_n(0)\mathrm{e}^{-\mathrm{i}n\tau} - a^\mu_{-n}(0)\mathrm{e}^{\mathrm{i}n\tau}]\cos n\sigma \\ \pi^\mu(\sigma) &= \pi^\mu_0 - \frac{\sqrt{2\alpha'}}{2\pi\alpha'}\sum_{n\geq 1}[a^\mu_n(0)\mathrm{e}^{-\mathrm{i}n\tau} + a^\mu_{-n}(0)\mathrm{e}^{\mathrm{i}n\tau}]\cos n\sigma. \end{aligned} \tag{11.23}$$

Closed strings

For a closed string defined on $[0, \pi]$, the Fourier expansion is given by

$$\begin{aligned} x^\mu(\sigma) &= x^\mu_0 + \sum_{n\geq 1}(x^\mu_n\mathrm{e}^{2\mathrm{i}n\sigma} + x^{\mu\star}_n\mathrm{e}^{-2\mathrm{i}n\sigma}) = \sum_n x^\mu_n\mathrm{e}^{2\mathrm{i}n\sigma} \\ \pi^\mu(\sigma) &= \pi^\mu_0 + \sum_{n\geq 1}(\pi^\mu_n\mathrm{e}^{2\mathrm{i}n\sigma} + \pi^{\mu\star}_n\mathrm{e}^{-2\mathrm{i}n\sigma}) = \sum_n \pi^\mu_n\mathrm{e}^{2\mathrm{i}n\sigma}. \end{aligned} \tag{11.24}$$

Here, $x_n \neq x_{-n}$ and $\pi_n \neq \pi_{-n}$, in contrast to the open string case (but, as before, $x_{-n} \equiv x^*_n$, $\pi_{-n} \equiv \pi^*_n$).

The basic PBs have the form (11.20a), where the periodic delta function is defined in equation (L.8), and, consequently,

$$\{x^\mu_n, \pi^\nu_{-m}\} = \eta^{\mu\nu}\frac{1}{\pi}\delta_{n,m} \qquad (n, m \geq 0).$$

In analogy with (11.21a, b), we introduce on the interval $[0, \pi]$ the following quantities:

$$\begin{aligned} \Pi^\mu &\equiv \pi^\mu + \frac{x'^\mu}{2\pi\alpha'} = \frac{-1}{\pi\sqrt{2\alpha'}}\sum_n 2a^\mu_n\mathrm{e}^{2\mathrm{i}n\sigma} \\ \tilde{\Pi}^\mu &\equiv \pi^\mu - \frac{x'^\mu}{2\pi\alpha'} = \frac{-1}{\pi\sqrt{2\alpha'}}\sum_n 2b^\mu_n\mathrm{e}^{2\mathrm{i}n\sigma}. \end{aligned} \tag{11.25}$$

Here,

$$-2a_n^\mu = \left(\pi_n^\mu + \frac{\mathrm{i}n}{\pi\alpha'}x_n^\mu\right)\pi\sqrt{2\alpha'}$$
$$-2b_{-n}^\mu = \left(\pi_n^\mu - \frac{\mathrm{i}n}{\pi\alpha'}x_n^\mu\right)\pi\sqrt{2\alpha'}$$

$(a_{-n} = a_n^*,\ b_{-n} = b_n^*)$ and, consequently,

$$-\pi_n^\mu = \frac{1}{\pi\sqrt{2\alpha'}}(a_n^\mu + b_{-n}^\mu)$$
$$x_n^\mu = \frac{\mathrm{i}\sqrt{2\alpha'}}{2n}(a_n^\mu - b_{-n}^\mu).$$

For an open string we would have $a_n = b_n$.

The basic PBs, when expressed in terms of a_n and b_n, become

$$\begin{aligned}\{a_n^\mu, a_{-m}^\nu\} = \mathrm{i}n\delta_{n,m}\eta^{\mu\nu} = \{b_n^\mu, b_{-m}^\nu\} \qquad (n, m \geq 1)\\ \{a_r^\mu, b_s^\nu\} = 0\end{aligned} \tag{11.26}$$

showing clearly the existence of two types of mode.

The *time dependence* of the modes is given by

$$a_n(\tau) = a_n(0)\mathrm{e}^{-2\mathrm{i}n\tau} \qquad b_n(\tau) = b_n(0)\mathrm{e}^{-2\mathrm{i}n\tau} \qquad (n \neq 0)$$
$$x_0^\mu = x^\mu + p^\mu\tau \qquad \pi_0^\mu = -p^\mu/2\pi\alpha'$$

with

$$\{x^\mu, p^\nu\} = -\eta^{\mu\nu}(2\alpha').$$

We can now separate the left-moving and right-moving waves in the expressions (11.24):

$$\begin{aligned}x^\mu(\sigma) &= x_0^\mu + \sqrt{2\alpha'}\frac{\mathrm{i}}{2}\sum_{n\neq 0}\frac{1}{n}[a_n^\mu(0)\mathrm{e}^{2\mathrm{i}n(\sigma-\tau)} - b_n^\mu(0)\mathrm{e}^{-2\mathrm{i}n(\sigma+\tau)}]\\ \pi^\mu(\sigma) &= \pi_0^\mu - \frac{\sqrt{2\alpha'}}{2\pi\alpha'}\sum_{n\neq 0}[a_n^\mu(0)\mathrm{e}^{2\mathrm{i}n(\sigma-\tau)} + b_n^\mu(0)\mathrm{e}^{-2\mathrm{i}n(\sigma+\tau)}].\end{aligned} \tag{11.27}$$

Classical Virasoro algebra

Reparametrization invariance of the classical string action reflects itself in the existence of two first class constraints, $G_0(\sigma)$ and $G_1(\sigma)$, which satisfy the PB algebra (11.10). After fixing the gauge as in equation (11.12), the symmetry of the theory is reduced to *conformal symmetry* in two dimensions. It will play a crucial role in the construction of covariant string field theory. We shall now study the classical form of this symmetry in the oscillator formalism.

Open strings. As we have seen, for an open string the two constraints, $G_0(\sigma)$ and $G_1(\sigma)$, defined on $[0, \pi]$, can be unified into one constraint, $\Pi^2(\sigma)$, on $[-\pi, \pi]$. The Fourier components of $\Pi^2(\sigma)$, at $\tau = 0$, are given as

$$L_n \equiv -\frac{\pi\alpha'}{2}\int_{-\pi}^{\pi} f_n^*(\sigma)\Pi^2(\sigma)\,\mathrm{d}\sigma \tag{11.28a}$$

where $f_n(\sigma) = \exp(\mathrm{i}n\sigma)$. The constraint $\Pi^2(\sigma)$ is completely determined by the set of coefficients L_n, and vice versa. The algebra of $G_o(\sigma)$ and $G_1(\sigma)$ can be easily rewritten in the form

$$\{L_n, L_m\} = -\mathrm{i}(n-m)L_{n+m}. \tag{11.29}$$

The coefficients L_n are referred to as the *Virasoro generators* (Virasoro 1970) and their algebra is the conformal algebra in two dimensions (note that this result is obtained without any gauge fixing).

Using expansion (11.21) the Virasoro generators can be expressed in terms of the Fourier coefficients a_n at $\tau = 0$:

$$L_n = -\tfrac{1}{2}\sum_r a_r \cdot a_{n-r}. \tag{11.28b}$$

This expression, combined with (11.22), can be used to rederive the classical Virasoro algebra (11.29).

The total Hamiltonian H_T is a linear combination of constraints, hence it can be represented as a linear combination of the Virasoro generators. Indeed, the expression (11.9) for H_T can be brought to the form

$$H_\mathrm{T} = \int_{-\pi}^{\pi} \mathrm{d}\sigma\, v(\sigma)\Pi^2(\sigma) \qquad 2v(\sigma) \equiv v^0(\sigma) + \pi\alpha' v^1(\sigma)$$

where $v^0(-\sigma) = v^0(\sigma)$, $v^1(-\sigma) = -v^1(\sigma)$. After that, the Fourier expansion $v(\sigma) = -(\pi\alpha'/2)\sum_n v_n \mathrm{e}^{-\mathrm{i}n\sigma}$ gives the result

$$H_\mathrm{T} = \sum_n v_n L_n. \tag{11.30a}$$

In the conformal gauge we have $v(\sigma) = -\pi\alpha'/2$, and

$$H_\mathrm{T} = L_0. \tag{11.30b}$$

It is interesting to observe that L_0 can be expressed in terms of the 'number of excitations' N:

$$L_0 = -\tfrac{1}{2}a_0 \cdot a_0 - \sum_{r\geq 1} a_r^\star \cdot a_r = -\tfrac{1}{2}a_0 \cdot a_0 + N \tag{11.31}$$

where

$$a_0^\mu = -\pi_0^\mu \pi \sqrt{2\alpha'} = p^\mu/\sqrt{2\alpha'}.$$

We should note that $N = \sum r N_r$, where N_r is the number of excitations of the type a_r, as a consequence of the factor $\sqrt{r}$ in the normalization of a_r. For this reason N is not the true number of string excitations. However, for the lowest level excitations with $r = 0, 1$, which we are going to consider, this difference is irrelevant.

In classical theory the coefficients a_n commute with each other and satisfy the PBs (11.22). In quantum theory they become operators and we have to take care about operator ordering.

Closed strings. The closed string is treated in a similar manner. We have two types of Virasoro generators:

$$\begin{aligned}
L_n &= -\frac{\pi\alpha'}{4}\int_0^\pi f_{2n}^*(\sigma)\Pi^2(\sigma)\,\mathrm{d}\sigma = -\tfrac{1}{2}\sum_r a_r \cdot a_{n-r} \\
\tilde{L}_n &= -\frac{\pi\alpha'}{4}\int_0^\pi f_{2n}^*(\sigma)\tilde{\Pi}^2(\sigma)\,\mathrm{d}\sigma = -\tfrac{1}{2}\sum_r b_r \cdot b_{n-r}
\end{aligned} \tag{11.32}$$

with two independent classical Virasoro algebras:

$$\begin{aligned}
\{L_n, L_m\} &= -\mathrm{i}(n-m)L_{n+m} \\
\{\tilde{L}_n, \tilde{L}_m\} &= -\mathrm{i}(n-m)\tilde{L}_{n+m} \\
\{L_n, \tilde{L}_m\} &= 0.
\end{aligned} \tag{11.33}$$

The total Hamiltonian is given as a linear combination of L_ns and $\tilde{L}_n$s. Indeed, the relation

$$H_\mathrm{T} = \int_0^\pi \mathrm{d}\sigma\,[v\Pi^2 + \tilde{v}\tilde{\Pi}^2]$$

where $2v = v^0 + \pi\alpha' v^1$, $2\tilde{v} = v^0 - \pi\alpha' v^1$, after using the Fourier expansion $v = -(\pi\alpha'/4)\sum_n v_n \mathrm{e}^{2\mathrm{i}n\sigma}$, and similarly for $\tilde{v}$, implies

$$H_\mathrm{T} = \sum_n (v_n L_n + \tilde{v}_n \tilde{L}_n). \tag{11.34a}$$

In the conformal gauge we have

$$H_\mathrm{T} = 2(L_0 + \tilde{L}_0). \tag{11.34b}$$

Introducing the excitation numbers N and $\tilde{N}$ we obtain

$$\begin{aligned}
L_0 &= -\tfrac{1}{2}a_0^2 - \sum_{r\geq 1} a_r^\star \cdot a_r = -\tfrac{1}{2}a_0^2 + N \\
\tilde{L}_0 &= -\tfrac{1}{2}b_0^2 - \sum_{r\geq 1} b_r^\star \cdot b_r = -\tfrac{1}{2}b_0^2 + \tilde{N}
\end{aligned} \tag{11.35}$$

where

$$a_0^\mu = b_0^\mu = p^\mu/2\sqrt{2\alpha'}.$$

Closed strings have an additional invariance, an invariance under global 'translations' of σ, $\sigma \to \sigma + a$. The generator of these transformations is the following first class constraint:

$$L_0 - \tilde{L}_0 \approx 0. \tag{11.36}$$

Unoriented strings are defined by invariance under $\sigma \to -\sigma$. The existence of two independent types of excitations makes the closed string intrinsically oriented, in general. However, there are solutions that are invariant under $a_n \leftrightarrow b_n$; they define an unoriented closed string.

11.3 First quantization

The quantum mechanics of strings is the first step towards a string field theory. The quantization procedure will be carried out using the standard canonical approach for dynamical systems with constraints (Peskin 1985, West 1986b, Green *et al* 1987, Brink and Henneaux 1988).

Quantum mechanics of the string

To quantize the string model we use the covariant approach outlined in discussing the point particle. We begin with the open string case.

(a) The basic dynamical variables $x(\tau, \sigma)$ and $\pi(\tau, \sigma)$ become the operators:

$$x(\tau, \sigma) \to \hat{x}(\tau, \sigma) \qquad \pi(\tau, \sigma) \to \hat{\pi}(\tau, \sigma).$$

(b) Each {PB} is replaced by $(-\mathrm{i}) \times$ [commutator]. In particular,

$$[\hat{x}^\mu(\sigma), \hat{\pi}_\nu(\sigma')] = \mathrm{i}\delta^\mu_\nu \delta(\sigma, \sigma') \qquad \delta(\sigma, \sigma') = 2\delta_\mathrm{S}(\sigma, \sigma'). \tag{11.37a}$$

In the x-representation we have $\hat{x}^\mu \to x^\mu$, $\hat{\pi}_\nu \to -\mathrm{i}\frac{\delta}{\delta x^\nu}$. Transition to the Fourier modes yields

$$\begin{aligned} [\hat{x}^\mu_n, \hat{\pi}^\nu_m] &= \eta^{\mu\nu}\frac{\mathrm{i}}{2\pi}(\delta_{n,m} + \delta_{n,-m}) \qquad (n, m \geq 0) \\ [\hat{a}^\mu_n, \hat{a}^{\nu+}_m] &= -n\delta_{n,m}\eta^{\mu\nu} \qquad (n, m \geq 1) \\ [x^\mu, p^\nu] &= -i\eta^{\mu\nu}(2\alpha'). \end{aligned} \tag{11.37b}$$

The x_n-representation is given by

$$\hat{x}^\mu_n \to x^\mu_n \qquad \hat{\pi}_{n\nu} \to -\frac{\mathrm{i}}{2\pi}\left(\frac{\partial}{\partial x^\nu_n} + \frac{\partial}{\partial x^\nu_{-n}}\right).$$

(c) First class constraints L_n are imposed as quantum conditions on the physical states, the Virasoro conditions:

$$(\hat{L}_0 - \alpha_0)|\psi\rangle = 0 \tag{A}$$

$$\hat{L}_n|\psi\rangle = 0 \qquad (n \geq 1). \tag{B}$$

(d) Physical states satisfy the 'Schródinger equation':

$$\mathrm{i}\frac{\partial}{\partial\tau}|\psi\rangle = H_\mathrm{T}|\psi\rangle \approx 0. \tag{11.38}$$

The right-hand side of this equation vanishes because $H_\mathrm{T} \approx 0$, and physical states do not depend explicitly on time (the time dependence is hidden in dynamical variables).

The form of Virasoro conditions (A) and (B) differs from what we might naively expect, and this deserves several comments.

– In order to give a precise meaning to the Virasoro operators $\hat{L}_n$, we assume that they are determined in the harmonic oscillator representation (11.28b). It is, then, natural to choose a Fock representation for the Hilbert space of states, which is based on the existence of a vacuum state, annihilated by all destruction operators a_n, $n \geq 1$.
– Upon quantization the Virasoro operators suffer from operator ordering ambiguities, since a_{-n} no longer commutes with a_n. A closer inspection shows that it is only $\hat{L}_0$ that is afflicted by this ambiguity. To resolve the problem we *normal order* the a_ns in $\hat{L}_0$:

$$\hat{L}_0 = -\tfrac{1}{2}\sum_r :a_{-r}\cdot a_r: = -\tfrac{1}{2}\sum_r a_{-r}\cdot a_r + \alpha_0 \tag{11.39}$$

where the colons indicate normal ordering with respect to the Fock vacuum, and α_0 is the normal ordering constant appearing in (A).
– Equation (B) tells us that the physical states do not satisfy the condition $\hat{L}_n|\psi\rangle = 0$ for all $n \neq 0$, since this would lead to a conflict with the quantum Virasoro algebra, as we shall see shortly. However, the condition (B) implies

$$\bar{L}_n \equiv \langle\psi|\hat{L}_n|\psi\rangle = 0$$

for every $n \neq 0$ (since $\hat{L}_{-n} = \hat{L}_n^+$), so that the mean value of $\hat{L}_n$ vanishes, in accordance with the classical constraints. Equation (B) is analogous to the Gupta–Bleuler quantization condition in electrodynamics.
– In the covariant approach the Lorentz covariance is maintained in the quantization process. However, the indefiniteness of the Minkowski metric implies the existence of negative norm states, or ghosts, which are forbidden by

unitarity. Ghosts are absent from the physical Hilbert space in two cases only (see, e.g., Green *et al* 1987):

$$\begin{aligned} &\text{(a)} \quad \alpha_0 = 1 \qquad D = 26 \\ &\text{(b)} \quad \alpha_0 \leq 1 \qquad D \leq 25. \end{aligned} \tag{11.40}$$

With the constraints and states as given in (A) and (B), the covariant quantization is possible only in $D = 26$. In $D < 26$ we have to add the so-called Liouville modes (Polyakov 1981). We shall limit our discussion in the remainder of this chapter to the critical dimension $D = 26$ (and $\alpha_0 = 1$).

For a closed string the basic commutation relations in the oscillator formalism have the following form:

$$\begin{gathered} [\hat{x}_n^\mu, \hat{\pi}_{-m}^\nu] = \eta^{\mu\nu} \frac{\mathrm{i}}{\pi} \delta_{n,m} \qquad (n, m \geq 0) \\ [\hat{a}_n^\mu, \hat{a}_m^{\nu+}] = [\hat{b}_n^\mu, \hat{b}_m^{\nu+}] = -n\delta_{n,m}\eta^{\mu\nu} \qquad (n, m \geq 1) \\ [x^\mu, p^\nu] = -\mathrm{i}\eta^{\mu\nu}(2\alpha') \end{gathered}$$

while $[\hat{a}_n^\mu, \hat{b}_m^{\nu+}] = 0$. They are realized in the x_n-representation as follows:

$$\hat{x}_n^\mu \to x_n^\mu \qquad \hat{\pi}_{-n\nu} \to -\frac{\mathrm{i}}{\pi} \frac{\partial}{\partial x_n^\nu}.$$

In the rest of this chapter we shall study the meaning of the quantum Virasoro conditions (A) and (B) in detail and use them to construct a covariant field theory. We shall see that these conditions have an enormously rich content in field theory, a content that we might not have expected.

Quantum Virasoro algebra

The fact that the quantum Virasoro generators are normal ordered brings a change into the Virasoro algebra (11.29). In calculating the commutator $[\hat{L}_n, \hat{L}_m]$ the problem of normal ordering arises only when the result contains the term $\hat{L}_0$, i.e. for $n = -m$. Let us, therefore, concentrate on the expression

$$[\hat{L}_n, \hat{L}_{-n}] = -\tfrac{1}{2} \sum_r [(n-r)a_r \cdot a_{-r} + r a_{-n+r} \cdot a_{n-r}].$$

Both terms in the infinite sum have to be normal ordered and this is a delicate problem. In order to do that in a proper manner, we shall first regularize definition (11.39) (Peskin 1985),

$$\hat{L}_n = \lim_{\Lambda\to\infty} -\tfrac{1}{2} \sum_{m=-\Lambda}^{\Lambda} :a_m \cdot a_{n-m}:$$

and then perform the normal ordering before taking the limit $\Lambda \to \infty$. In this way, the regularized sum for $[\hat{L}_n, \hat{L}_{-n}]$ becomes finite, and c-number terms arising from normal ordering are calculated by a change of the summation variable, with the result

$$[\hat{L}_n, \hat{L}_{-n}] = 2n\hat{L}_0 + nD\sum_{1}^{\Lambda} r - \tfrac{1}{2}D \sum_{\Lambda-n+1}^{\Lambda} (r+n)r.$$

Here, $\hat{L}_0$ is normal ordered, the dimension D originates from the contraction of δ^μ_ν, and at the end we should take the limit $\Lambda \to \infty$. The last step is actually trivial, since the evaluation of the previous expression shows that it does not depend on Λ:

$$[\hat{L}_n, \hat{L}_{-n}] = 2n\hat{L}_0 + \frac{D}{12}n(n^2-1).$$

This result is easily generalized to give

$$[\hat{L}_n, \hat{L}_m] = (n-m)\hat{L}_{n+m} + \frac{D}{12}n(n^2-1)\delta_{n,-m}. \tag{11.41}$$

The new, c-number term on the right-hand side, the so-called *central charge*, arises as a consequence of the normal ordering in $[\hat{L}_n, \hat{L}_m]$, and represents a typical quantum effect.

A change in the algebra of the Virasoro constraints indicates a change in the symmetry structure of the theory after quantization. Indeed, in classical theory all L_ns are first class constraints, which generate the conformal symmetry of the theory, while the presence of the central charge in the quantum algebra implies that only the $SL(2, R)$ subalgebra generated by $(\hat{L}_0, \hat{L}_{\pm 1})$ is first class. The change of symmetry is a direct reflection of the so-called *conformal anomaly* in the theory.

Let us now return to the Virasoro condition (B). It is evident that we cannot consistently demand $\hat{L}_n|\psi\rangle = 0$ for every $n \neq 0$, since in that case algebra (11.41) would imply $|\psi\rangle = 0$, due to the presence of the central charge.

Fock space of states

Since the 'Schródinger equation' only tells us that the physical states are time independent, the most important dynamical informations are contained in the Virasoro conditions (A) and (B). We have formally defined the quantum Virasoro generators by the normal ordering of the a_ns. The true meaning of this prescription can be understood only after we have specified the space of the states on which these operators act. Since $a_n^\dagger$ and a_n have harmonic-oscillator commutation relations, it appears natural to choose a Fock representation for states.

In order to clarify the physical interpretation of the theory, it is useful to introduce the *mass operator* for string excitations. Using the equality $a_0^\mu = p^\mu = \mathrm{i}\partial^\mu$ and relation (11.31), we find (in units $2\alpha' = 1$):

$$\begin{aligned} 2(\hat{L}_0 - 1) &= -p^2 + \mathcal{M}^2 \\ \mathcal{M}^2 &= 2(N-1) \end{aligned} \tag{11.42}$$

where $N \equiv -\sum_{m\geq 1} a_m^+ \cdot a_m$, and $\mathcal{M}^2$ is the mass squared operator, with eigenvalues $2(n-1)$. The state with $n = 0$ has $\mathcal{M}^2 < 0$, and is called the *tachyon*, while the states with $n = 1$ are massless. The tachyonic ground state is a problem for the bosonic string, which can be solved in superstring theory.

In a similar way, using the relation $a_0^\mu = b_0^\mu = p^\mu/2$, we find, for a closed string:

$$\begin{aligned} 4[(\hat{L}_0 - 1) + (\hat{\tilde{L}} - 1)] &= -p^2 + \mathcal{M}^2 \\ \mathcal{M}^2 &= 4(N + \tilde{N} - 2). \end{aligned} \tag{11.43}$$

Here, $n = \tilde{n} = 0$ is the tachyonic ground state and $n = \tilde{n} = 1$ are massless states.

Let us now define the string *ground state* by the condition

$$a_m|0\rangle = 0 \quad (m \geq 1) \qquad p^\mu|0\rangle = 0. \tag{11.44}$$

In the coordinate representation it has the form

$$\langle x(\sigma)|0\rangle \equiv \Phi^{(0)} = \prod_{m\geq 1} c_m \exp\left(\frac{m x_m^2}{2\alpha'}\right).$$

Note that $\Phi^{(0)}$ does not contain x_0, the zero mode of $x(\sigma)$. The set of states obtained by applying a product of the creation operators a_m^+ $(m \geq 1)$ on the ground state constitutes the basis of the *Fock space* of string excitations‖. Each state in this basis is an eigenstate of the occupation number N, and, hence, of the mass square operator $\mathcal{M}^2$.

A state of the string $|\psi\rangle$ can be given in the coordinate representation as a *functional* $\psi[x(\sigma)] = \langle x(\sigma)|\psi\rangle$. The scalar product is formally defined by the functional integral

$$\langle\psi_1|\psi_2\rangle = \int \mathcal{D}x(\sigma)\psi_1^*[x(\sigma)]\psi_2[x(\sigma)]$$

with a suitable measure. Every string state $\psi[x(\sigma)]$ can be expanded in terms of the occupation number basis as

$$\begin{aligned} \psi[x(\sigma)] = [&\phi(x) - \mathrm{i}A_\mu(x)a_1^{\mu+} - h_{\mu\nu}(x)a_1^{\mu+}a_1^{\nu+} - \cdots \\ &- \mathrm{i}B_\mu(x)a_2^{\mu+} - l_{\mu\nu}(x)a_2^{\mu+}a_2^{\nu+} - \cdots]\Phi^{(0)} \end{aligned} \tag{11.45}$$

‖ In general, we can also ‘create’ the centre-of-mass momentum k by application of $\exp(-\mathrm{i}k \cdot x)$: $\Phi^{(k)} = \exp(-\mathrm{i}k \cdot x)\Phi^{(0)}$, $p\Phi^{(k)} = k\Phi^{(k)}$.

where the dependence on the zero mode $x = x_0$ is present in the coefficient functions $\phi(x)$, $A_\mu(x)$, etc. It should be stressed that the a_n^+s are the creation operators of the string modes, not of the physical particles. The information about the physical fields resides in the coefficient functions, which are determined by the Virasoro conditions (A) and (B), and the equation of motion (11.38).

In the case of closed strings, the invariance under $\sigma \to \sigma + a$ can be expressed as the condition

$$(L_0 - \tilde{L}_0)\psi[x(\sigma)] = (n - \tilde{n})\psi[x(\sigma)] = 0 \tag{11.46}$$

where we have omitted the hats over the Virasoro generators for simplicity.

11.4 Covariant field theory

We now turn to the problem of finding a covariant field theory for free strings which would be consistent with the Virasoro constraints (Peskin 1985, West 1986b, Bailin and Love 1994).

Classical string field theory is constructed in the following way:

– Quantum-mechanical single-string state $|\psi\rangle$ is promoted to a classical field $\psi[x(\sigma)]$, which satisfies the same field equation and the same Virasoro conditions.
– Then, the corresponding action for the free field theory is constructed. This step is much more complicated than in point-particle theory.
– Finally, the interactions are introduced in accordance with certain physical principles, and an attempt is made to describe realistic physical processes.

After that, we can proceed to construct *quantum* field theory. We begin by developing the canonical formulation of classical field theory, i.e. we define momenta $\pi[x(\sigma)]$, find the constraints, Hamiltonian and the equations of motion. Then, in the process of *second quantization*, the basic dynamical variables $\psi[x(\sigma)]$ and $\pi[x(\sigma)]$ become the operators, etc. If the classical theory describes the interactions, so does its second-quantized version.

The central point in the forthcoming exposition will be to find a *covariant* action of *classical field theory* for *free* strings. Trying to solve this problem we shall discover some fundamental symmetry principles of the theory. This is of great importance not only for operator quantization, which we are using in the present context, but also for path-integral quantization, where knowledge of the classical action and its symmetries is of crucial importance. The important subject of string interactions lies outside the scope of this book.

Gauge symmetries

Virasoro conditions. In order to construct a covariant action, let us consider the meaning of the Virasoro conditions for open strings:

$$(L_0 - 1)\psi[x(\sigma)] = 0 \tag{A}$$

$$L_n\psi[x(\sigma)] = 0 \qquad (n \geq 1) \tag{B}$$

where we write L instead of $\hat{L}$, for simplicity. A physical state $\psi[x(\sigma)]$ is represented by equation (11.45) in the Fock space, where the components $\phi(x)$, $A_\mu(x)$, $h_{\mu\nu}(x)$, etc, are the standard fields of increasing spin.

Condition (A) represents an equation of the Klein–Gordon type for each component field of $\psi[x(\sigma)]$. It implies that the field $\phi(x)$ is a tachyon, $A_\mu(x)$ is a massless field, $h_{\mu\nu}(x)$ is a field of mass $(\alpha')^{-1/2}$, etc.

In order to discover the meaning of (B), let us recall the definition of the Virasoro generators for $n = 1, 2$:

$$L_1 = -(a_0 \cdot a_1 + a_1^+ \cdot a_2 + a_2^+ \cdot a_3 + \cdots)$$
$$L_2 = -(a_0 \cdot a_2 + \tfrac{1}{2}a_1 \cdot a_1 + a_1^+ \cdot a_3 + \cdots)$$

where $a_0^\mu = p^\mu = \mathrm{i}\partial^\mu$. Acting on $\psi[x(\sigma)]$ with L_1 and L_2 we find that

$$L_1\psi[x(\sigma)] = [\mathrm{i}p^\mu A_\mu + (2p^\mu h_{\mu\nu} + \mathrm{i}B_\nu)a_1^{\nu+} + \cdots]\Phi^{(0)}$$
$$L_2\psi[x(\sigma)] = [2\mathrm{i}p^\mu B_\mu - h^\nu{}_\nu + \cdots]\Phi^{(0)}.$$

From these equations we can read off the form of conditions (B) in terms of the component fields:

$$\begin{gathered} \partial^\mu A_\mu = 0 \qquad 2\partial^\mu h_{\mu\nu} + B_\nu = 0 \\ 2\partial^\mu B_\mu + h^\nu{}_\nu = 0 \qquad \cdots \end{gathered} \tag{11.47}$$

The Virasoro conditions at the massless level lead to the standard gauge condition: $\partial^\mu A_\mu = 0$.

This is rather an unexpected result. In classical theory the Virasoro generators are related to the reparametrization invariance. After choosing the conformal gauge, they describe the conformal symmetry in the theory. In field theory the Virasoro conditions become related to the gauge fixing conditions.

A field theory based on Virasoro conditions (B) would represent a theory with fixed gauge symmetry. We wish to find an action which possesses gauge invariances, for which conditions (B) would be a possible gauge choice and equation (A)—the equation of motion.

Gauge transformations. If Virasoro conditions (B) are interpreted as gauge conditions on $|\psi\rangle$, it is natural to ask: What form do the related gauge transformations take?

Let us try to construct gauge transformations of $\psi[x(\sigma)]$ with the L_{-n} as symmetry generators. Consider the open string and the transformation

$$\delta\psi = -L_{-n}\Lambda \qquad (n \geq 1) \tag{C}$$

where Λ is a functional 'parameter' of the gauge transformation:

$$\Lambda[x(\sigma)] = [\lambda(x) + \mathrm{i}\lambda_\mu(x)a_1^{\mu+} + \mathrm{i}\varepsilon_\mu(x)a_2^{\mu+} + \cdots]\Phi^{(0)}.$$

For $n = 1, 2$, we use

$$\begin{aligned} -L_{-1} &= (a_1^+ \cdot a_0 + a_2^+ \cdot a_1 + a_3^+ \cdot a_2 + \cdots) \\ -L_{-2} &= (a_2^+ \cdot a_0 + \tfrac{1}{2}a_1^+ \cdot a_1^+ + a_3^+ \cdot a_1 + \cdots) \end{aligned}$$

and obtain

$$\begin{aligned} -L_{-1}\Lambda &= (p_\mu \lambda a_1^{\mu+} + \mathrm{i}p_\mu\lambda_\nu a_1^{\mu+}a_1^{\nu+} - \mathrm{i}\lambda_\mu a_2^{\mu+} + \cdots)\Phi^{(0)} \\ -L_{-2}\Lambda' &= (p_\mu\lambda' a_2^{\mu+} + \tfrac{1}{2}\lambda' a_1^+ \cdot a_1^+ + \cdots)\Phi^{(0)}. \end{aligned}$$

From these expressions and equation (C) we can derive the transformation law for component fields. The ground state field is gauge invariant:

$$\delta\phi(x) = 0. \tag{11.48a}$$

The massless field transforms as

$$\delta A_\mu = -\partial_\mu\lambda \tag{11.48b}$$

which is just the gauge transformation of a vector field. At the next mass level we obtain

$$\begin{aligned} \delta h_{\mu\nu} &= \partial_{(\mu}\lambda_{\nu)} - \tfrac{1}{2}\eta_{\mu\nu}\lambda' \\ \delta B_\mu &= \lambda_\mu - \partial_\mu\lambda'. \end{aligned} \tag{11.48c}$$

> *The transformation (C), at the massless level, represents exactly the gauge symmetry for which we have been looking.*

At higher mass levels it yields many other gauge symmetries, the existence of which could have been hardly expected. Similar results are also obtained for closed strings.

Thus, we found that gauge symmetries, at the level of free field theory, have the simple form (C), which is a very important step towards a deeper understanding of the geometric structure of the theory. These symmetries take a complicated form when expressed in terms of the component fields, but their physical interpretation at the massless level becomes very simple.

These considerations shed a new light on Virasoro conditions (A) and (B), and suggested the form gauge symmetry (C) should have in a complete, covariant formulation. Now, we shall try to incorporate these results into an action principle.

The action for the free string field

We seek a free field action for strings in the form

$$I = -\tfrac{1}{2}\int \mathcal{D}x(\sigma)\psi^{\star}[x(\sigma)]\,K\,\psi[x(\sigma)] \equiv -\tfrac{1}{2}(\psi, K\psi). \tag{11.49}$$

The condition that gauge transformations

$$\delta\psi = \mathrm{i}\sum_{n\geq 1} c_n L_{-n}\Lambda$$

are a symmetry of the action leads to

$$\delta I = -\frac{\mathrm{i}}{2}\sum_{n\geq 1} c_n(\psi, KL_{-n}\Lambda) + \text{C.C.} = 0.$$

We also demand that Virasoro conditions (A) are fulfilled as the equations of motion.

The Verma module. It will be helpful for the construction of K to first define some mathematical properties of the Virasoro algebra (Goddard and Olive 1986).

The Virasoro algebra is an expression of conformal symmetry in two dimensions, and has an infinite number of generators. According to (11.41), we conclude that L_0 is the only generator of the commuting subalgebra. Bearing in mind the application to string theory, we impose the requirement that (a) L_0 be a *positive definite* operator:

$$L_0 \geq 0.$$

In addition, we shall restrict our considerations to unitary representations of the algebra, for which (b) a state space is equipped with a positive definite scalar product, and (c) $L_n^+ = L_{-n}$ (hermiticity).

Let $H[x(\sigma)]$ be an eigenstate of L_0 with eigenvalue h. According to the relation

$$L_0 L_n = L_n(L_0 - n) \qquad (n \geq 1)$$

the action of L_n on $H[x(\sigma)]$ *lowers* the eigenvalue of L_0 by n units. In view of the positive definiteness of L_0, there must be states ψ_0 on which this process stops:

$$L_n\psi_0 = 0 \qquad (n \geq 1)$$
$$L_0\psi_0 = h\psi_0.$$

In states ψ_0, the eigenvalue $h \geq 0$ of L_0 is the lowest one possible, hence the eigenvalues of $\mathcal{M}^2$ and N are also the lowest. These states are said to be the states at level $v = 0$. They are also called the physical states, since we recognize here the structure of the quantum Virasoro conditions.

States at level $v = 1$ are formed by applying L_{-1} to a zero-level state ψ_0, states at level $v = 2$ are $L_{-2}\psi_0$ and $L_{-1}L_{-1}\psi_0$, etc. The space of all states constructed from a particular ψ_0 is called a *Verma module*. States at different levels in a Verma module have different eigenvalues of L_0, and are orthogonal to each other; for instance, $(L_{-1}\psi_0, \psi_0) = (\psi_0, L_1\psi_0) = 0$.

Thus, starting from a given state ψ_0, we can build up a representation of the Virasoro algebra level by level. A description of the strings at finite levels of excitation needs only a finite number of generators. This construction is a generalization of the standard techniques in the theory of finite Lie groups.

The action. The condition of gauge invariance for the action can be fulfilled if we find an operator K for which

$$KL_{-n}\Lambda = 0 \qquad (n \geq 1).$$

Since the state $L_{-n}\Lambda$ is at least of level $v = n$, this condition is easily satisfied if K is proportional to the projector P on the states of level $v = 0$:

$$K = K_0 P \qquad PL_{-n} = 0 \qquad (n \geq 1).$$

The construction of projector P is known from mathematics (Feigin and Fuchs 1982). Demanding that Virasoro condition (A) holds on the level-0 states, we conclude that K_0 should be proportional to $L_0 - 1$, so that, finally,

$$K = 2(L_0 - 1)P. \tag{11.50}$$

The action defined by equations (11.49) and (11.50) solves our problem. It defines a covariant field theory for free strings, which yields Virasoro condition (A), and possesses gauge symmetry (C), on the basis of which we can impose Virasoro conditions (B) as gauge conditions. In this way, the conformal symmetry of the classical string carries over, in field theory, into a gauge symmetry which will produce, at the massless level, electrodynamics and (linearized) gravity.

We shall see that action (11.50) is non-local, but this non-locality can be absorbed by introducing a set of auxiliary fields.

Electrodynamics

Up to the first v-level, the projector P for the open string is given by

$$P_{(1)} = 1 - \tfrac{1}{2}L_{-1}L_0^{-1}L_1.$$

Indeed, acting on the level-0 and level-1 states of a Verma module we find

$$P_{(1)}\psi_0 = \psi_0 \qquad P_{(1)}\psi_1 = \psi_1 - \tfrac{1}{2}L_{-1}L_0^{-1} \cdot 2L_0\psi_0 = 0.$$

The operator K is given, with the same accuracy, as

$$K_{(1)} = 2(L_0 - 1)P_{(1)} = 2(L_0 - 1) - L_{-1}L_1$$

and the action takes the form

$$I_{(1)} = -\tfrac{1}{2} \int \mathcal{D}x(\sigma)\psi^*[x(\sigma)]K_{(1)}\psi[x(\sigma)]. \tag{11.51a}$$

Here, the string functional $\psi[x(\sigma)]$ is assumed to contain the lowest and first excited levels ($n = 0, 1$) without any constraints, while higher excitations are ignored. The action is invariant under the gauge transformation

$$\delta\psi = -L_{-1}\Lambda_0. \tag{11.52a}$$

Since $\delta\psi$ must have the same structure as ψ, we take Λ_0 to contain only the lowest, $n = 0$ excitation level, hence $L_1\Lambda_0 = L_2\Lambda_0 = \cdots = 0$. The invariance can be checked explicitly by showing that $P_{(1)}L_{-1}\Lambda_0 = 0$. Expressed in terms of the component fields in a Fock space, the symmetry transformation (11.52a) takes the form of the usual *Abelian gauge transformation*:

$$\delta\phi = 0 \qquad \delta A_\mu = -\partial_\mu\lambda. \tag{11.52b}$$

What is the content of the action $I_{(1)}$ in terms of the component fields? At the lowest level,

$$K_{(1)}\psi|_0 = (-p^2 + m^2)\phi\Phi^{(0)}$$

where $m^2 = -2$ (the tachyon). Using the expression

$$K_{(1)} = -p^2 + 2(N-1) - p_\mu p_\nu a_1^{\mu+} a_1^\nu + \cdots$$

we find, at the first excited (massless) level, that

$$\begin{aligned} K_{(1)}\psi|_1 &= (-p^2 - p_\mu p_\nu a_1^{\mu+} a_1^\nu)(-\mathrm{i}A_\lambda a_1^{\lambda+})\Phi^{(0)} \\ &= \mathrm{i}p^2\Pi_{\mu\nu}A^\mu a_1^{\nu+}\Phi^{(0)} \end{aligned}$$

where $\Pi_{\mu\nu} \equiv \eta_{\mu\nu} - p_\mu p_\nu/p^2$. Let us normalize the integration measure by demanding

$$\int \mathcal{D}x(\sigma)\Phi^{(0)*}\Phi^{(0)} = \int \mathrm{d}^D x \prod_{n\geq 1} \mathcal{D}x_n \Phi^{(0)\star}\Phi^{(0)} = \int \mathrm{d}^D x.$$

Then, at the massless level we find

$$I_{(1)}|_1 = -\tfrac{1}{2}\int \mathrm{d}^D x\, A^\mu p^2 \Pi_{\mu\nu} A^\nu = -\tfrac{1}{4}\int \mathrm{d}^D x\, F_{\mu\nu}F^{\mu\nu} \tag{11.51b}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$.

This is just the action of free electrodynamics, the result that could have been expected on the basis of the gauge invariance (11.52).

On higher levels, we find gauge-invariant theories of higher spin fields. These theories are already *non-local* at the second level, but they can be made local by introducing auxiliary fields. We shall see how this general phenomenon arises at the massless level of closed string field theory.

Gravity

Repeating this procedure for a *closed string*, we shall discover linearized gravitational theory.

We begin by considering the meaning of the Virasoro conditions for closed strings. Closed strings have a doubled spectrum of normal modes, and two sets of commuting Virasoro generators, L_n and $\tilde{L}_n$. String fields in the Fock space have the form

$$\psi[x(\sigma)] = [\phi(x) + t_{\mu\nu}(x)a_1^{\mu+}b_1^{\nu+} + \cdots]\Phi^{(0)}$$

where we have used $n - \tilde{n} = 0$, and here $\Phi^{(0)}$ is the ground state for both types of excitations. The field ϕ is the tachyon, $t_{\mu\nu}$ is the *massless field*, etc, as seen from the form of $\mathcal{M}^2$. The Virasoro constraints become:

$$(L_0 - 1)\psi = (\tilde{L}_0 - 1)\psi = 0 \qquad \text{(A}'\text{)}$$

$$L_n\psi = \tilde{L}_n\psi = 0 \qquad (n \geq 1). \qquad \text{(B}'\text{)}$$

At the massless level, conditions (B′) give the usual gauge condition:

$$\partial^\mu t_{\mu\nu} = \partial^\nu t_{\mu\nu} = 0. \qquad (11.53)$$

Using similar arguments as before, we find that the covariant action for the closed string is given by

$$I = -\tfrac{1}{2}(\psi, K\psi) \qquad K \equiv 4[(L_0 - 1) + (\tilde{L}_0 - 1)]P\widetilde{P} \qquad (11.54)$$

where P and $\widetilde{P}$ are projectors on the respective level-0 states. This action is invariant under the gauge transformations:

$$\delta\psi = -L_{-n}\tilde{L} - \tilde{L}_{-n}\Lambda \qquad (n \geq 1). \qquad \text{(C}'\text{)}$$

Let us consider the form of this symmetry at the massless level. Using the equality $2a_0^\mu = 2b_0^\mu = p^\mu = \mathrm{i}\partial^\mu$, and following the procedure for the open string, we obtain

$$\delta t_{\mu\nu} = -\tfrac{1}{2}(\partial_\mu\tilde{\lambda}_\nu + \partial_\nu\lambda_\mu). \qquad (11.55a)$$

Now, the tensor $t_{\mu\nu}$ can be decomposed into its symmetric and antisymmetric parts, $t_{\mu\nu} = h_{\mu\nu} + b_{\mu\nu}$, with the transformation laws

$$\begin{aligned} \delta h_{\mu\nu} &= \partial_\mu\lambda_\nu^{\mathrm{S}} + \partial_\nu\lambda_\mu^{\mathrm{S}} \qquad & \lambda_\mu^{\mathrm{S}} &\equiv -\tfrac{1}{4}(\lambda_\mu + \tilde{\lambda}_\mu) \\ \delta b_{\mu\nu} &= \partial_\mu\lambda_\nu^{\mathrm{A}} - \partial_\nu\lambda_\nu^{\mathrm{A}} \qquad & \lambda_\mu^{\mathrm{A}} &\equiv -\tfrac{1}{4}(\tilde{\lambda}_\mu - \lambda_\mu). \end{aligned} \qquad (11.55b)$$

We can identify $h_{\mu\nu}$ with the linearized gravitational field, since its transformation law is just a linearized *general coordinate transformation*, while the expression for $\delta b_{\mu\nu}$ represents a natural gauge transformation of the antisymmetric tensor field.

The action for the component fields $b_{\mu\nu}$ and $h_{\mu\nu}$ can be worked out by evaluating expression (11.54) at the massless level. We first note that, at the level $n = \tilde{n} = 1$, the following relation holds:

$$K_{(1\bar{1})} t_{\mu\nu} a_1^{\mu+} b_1^{\nu+} \Phi^{(0)} = -p^2 \Pi^{\lambda\mu} \Pi^{\rho\nu} t_{\mu\nu} a_{1\lambda}^{+} b_{1\rho}^{+} \Phi^{(0)}.$$

This result is obtained with the help of

$$P a_1^{\mu+} \Phi^{(0)}|_1 = \Pi^{\lambda\mu} a_{1\lambda}^{+} \Phi^{(0)} \qquad \widetilde{P} b_1^{\nu+} \Phi^{(0)}|_1 = \Pi^{\rho\nu} b_{1\rho}^{+} \Phi^{(0)}$$
$$4[(L_0 - 1) + (\tilde{L}_0 - 1)]|_{1\bar{1}} \to -p^2.$$

Then, the action for $t^{\mu\nu}$ can be written in the form

$$\begin{aligned} I_{(1)} &= -\tfrac{1}{2}(t^{\sigma\tau} a_{1\sigma}^{+} b_{1\tau}^{+} \Phi^{(0)}, K_{(1\bar{1})} t^{\mu\nu} a_{1\mu}^{+} b_{1\nu}^{+} \Phi^{(0)}) \\ &= \tfrac{1}{2} \int \mathrm{d}^D x \, t_{\lambda\rho} p^2 \Pi^{\lambda\mu} \Pi^{\rho\nu} t_{\mu\nu} \end{aligned} \tag{11.56}$$

where we have used the normalization $\int \mathcal{D}x(\sigma) \Phi^{(0)*} \Phi^{(0)} = \int \mathrm{d}^D x$.

The part of the action which is quadratic in $b_{\mu\nu}$ is found to be

$$I(b) = \tfrac{1}{6} \int \mathrm{d}^D x \, H_{\mu\nu\lambda} H^{\mu\nu\lambda} \tag{11.57}$$

where $H_{\mu\nu\lambda} \equiv \partial_\mu b_{\nu\lambda} + \partial_\lambda b_{\mu\nu} + \partial_\nu b_{\lambda\mu}$ is the field strength for $b_{\mu\nu}$, which is invariant under gauge transformation (11.55).

To obtain the part which is quadratic in $h_{\mu\nu}$ is slightly more complicated. We start with

$$I(h) = \int \mathrm{d}^D x \, h_{\lambda\rho} \, p^2 (\Pi^{\lambda\mu} \Pi^{\rho\nu} + \Pi^{\rho\mu} \Pi^{\lambda\nu}) h_{\mu\nu}.$$

Then, we add and subtract a conveniently chosen term:

$$I(h) = \int \mathrm{d}^D x \, h_{\lambda\rho} p^2 [(\Pi^{\lambda\mu} \Pi^{\rho\nu} + \Pi^{\rho\mu} \Pi^{\lambda\nu}) - 2\Pi^{\lambda\rho} \Pi^{\mu\nu} + 2\Pi^{\lambda\rho} \Pi^{\mu\nu}] h_{\mu\nu}.$$

The first three terms of this expression may be recognized as the Pauli–Fierz action for a massless field of spin 2:

$$I_{\mathrm{PF}} = \int \mathrm{d}^D x \, (\tfrac{1}{2} h_{\mu\nu,\sigma} h^{\mu\nu,\sigma} - h_{\mu\nu,\sigma} h^{\mu\sigma,\nu} + h_{\mu\sigma}{}^{,\sigma} h^{,\mu} - \tfrac{1}{2} h_{,\sigma} h^{,\sigma}) \tag{11.58a}$$

where $h = h^\nu{}_\nu$. This action coincides with the quadratic term of the *Einstein–Hilbert* action,

$$I_{\mathrm{EH}}(h) = -\frac{1}{2\lambda^2} \int \mathrm{d}^D x \, \sqrt{-g} R \tag{11.58b}$$

where R is the scalar curvature of Riemann space with metric $g_{\mu\nu} = \eta_{\mu\nu} + 2\lambda h_{\mu\nu}$ (see chapter 8).

Using $R = 2\lambda(\partial^\mu \partial^\nu h_{\mu\nu} - \Box h) = 2\lambda p^2 \Pi^{\mu\nu} h_{\mu\nu}$, the last term in $I(h)$ can be written as a *non-local* expression:

$$I_{\mathrm{NL}} = -\frac{1}{2\lambda^2} \int \mathrm{d}^D x\, R \frac{1}{\Box} R. \tag{11.59a}$$

This term can be made local by introducing an additional, *auxiliary* scalar field $\varphi(x)$. Indeed, the action

$$I_{\mathrm{S}} = \int \mathrm{d}^D x\, (\tfrac{1}{2} \partial_\mu \varphi \partial^\mu \varphi - \lambda^{-1} \varphi R) \tag{11.59b}$$

reduces to I_{NL} when φ is eliminated with the help of its equation of motion: $\Box\varphi + \lambda^{-1} R = 0$.

Thus, the final action we have obtained is given by

$$I_{(1)} = I(b) + I_{\mathrm{EH}}(h) + I_{\mathrm{S}}(\varphi, h). \tag{11.60}$$

The closed string action at the massless level contains an antisymmetric tensor field $b_{\mu\nu}$, a graviton $h_{\mu\nu}$ and a scalar field φ.

The field φ appears indirectly as a device to remove the non-locality in the action. This is a generic situation at all higher levels of (open and closed) string field theory. All non-localities can be eliminated by introducing auxiliary fields, which are also necessary to complete the particle content of the theory.

11.5 General remarks

To complete the present exposition, we shall mention here several additional topics, which have played important roles in the development of string theory.

The superstring. The bosonic string is not considered to be an entirely satisfactory theory for at least two reasons: first, it does not contain fermionic states; and second, there is a tachyon in its spectrum. A tachyon appears as a consequence of the positivity of the parameter α_0, which is related to the ordering ambiguity in $\hat{L}_0$. It is therefore natural to try to find a supersymmetric generalization of the model, where we could expect a cancellation of the contributions of bosons and fermions, leading to $\alpha_0 = 0$. There are two different ways to introduce supersymmetry into string models.

(a) We can think of the bosonic string as a two-dimensional field theory describing D scalar fields $x^\mu(\xi)$ coupled to gravity. The theory is invariant under two-dimensional reparametrizations, which can be generalized to a local supersymmetry (Neveu and Schwarz 1971, Ramond 1971). The critical

dimension is now $D = 10$, and the parameter α_0 takes the value $\frac{1}{2}$ (Neveu–Schwarz) or 0 (Ramond), depending on the boundary conditions. The potential problem of tachyonic ground state may be avoided by introducing certain restrictions on the states (Gliozzi *et al* 1977).
(b) The bosonic string is invariant under global Poincaré symmetry in M_D, which can be enlarged to global supersymmetry. This leads to the second version of the superstring, which is anomaly free in $D = 10$ (Green and Schwarz 1982, 1984).

Surprisingly enough, the first formulation of the superstring, after appropriate restrictions as mentioned earlier, becomes equivalent to the second one. While the existence of global supersymmetry in M_D is not evident at first glance, two-dimensional local supersymmetry does not appear in an obvious way in the second formulation. There is no formulation in which both types of supersymmetry would be obviously present. The form of supersymmetry in four dimensions depends on the compactification mechanism.

It turned out that the realistic gauge field content in string theory, which is necessary to describe the electroweak and strong interactions, can be obtained by combining elements of both bosonic and supersymmetric strings. Such an approach is known as a heterotic string (the concept 'heterosis' denotes 'increased vigour displayed by crossbred plants or animals') (Gross *et al* 1986)

Covariant quantization. Historically, string theory was first developed in the non-covariant, light-cone gauge (see, e.g., Scherk 1975, Green *et al* 1987), in which only the physical dynamical degrees of freedom are left. This formalism is manifestly ghost-free, but not manifestly covariant. The conditions of Poincaré invariance led to the critical dimension ($D = 26$ or 10), but the gauge symmetries and the geometric structure of the theory remain unclear.

In the present chapter we have used the gauge-invariant canonical approach to explore some basic features of the field theory of free bosonic strings. The method is develop so as to respect Weyl invariance on the world sheet and M_D-covariance, and yields a remarkable correspondence with the usual, point-particle fields, which are components of the string functional $\psi[x(\sigma)]$ in the Fock space of states. The gauge properties of field theory are 'derived' from the Virasoro constraints, which stem from the conformal invariance on the world sheet. A full understanding of string geometry demands a direct geometric interpretation of the 'vector' $\psi[x(\sigma)]$.

An important stimulus to the development of a geometric understanding of string theory came from the modern covariant approach—the BRST formalism (Kato and Ogawa 1983, Hwang 1983, Siegel 1985, Siegel and Zwiebach 1986), based on the experience with non-Abelian gauge theories. In this approach all field components, physical and pure gauge degrees of freedom, are treated as dynamical variables on an equal footing. The unitarity of the theory is ensured by introducing additional fields, 'ghosts', which compensate for the unwanted effects induced by pure gauge fields.

Some problems of the covariant quantization of strings are often studied on the simpler superparticle models.

Interacting strings. String theory is, by its very nature, a theory of quantum gravity. Its importance stems from our belief that it lay at the foundation of a consistent quantum theory of all the basic interactions in nature. Here, of course, we have to include also string interactions. Earlier studies of string field theories were based on the light-cone gauge formalism (Kaku and Kikkawa 1974, Cremmer and Gervais 1974, 1975). Although this formulation is completely acceptable in principle, there are questions for which it cannot give clear answers. There is a hope that covariant field theory will enable a deeper understanding of non-perturbative phenomena, which might be of importance in the process of compactification. In the covariant formulation, the basic principles of the theory become evident. As usual, the action is constructed starting from symmetry principles. First, the structure of the free field theory is clarified (Neveu *et al* 1985, Neveu and West 1985, Banks and Peskin 1986, Siegel and Zwiebach 1986), and then the interacting theories are constructed (Hata *et al* 1986, Neveu and West 1986, Witten 1986). The outstanding power of the BRST method is particularly useful in treating interacting strings (see, e.g., Thorn 1989).

Effective action. The theory of strings embedded in flat space can be generalized to the case of a curved background space. Our discussion of the bosonic string in M_D shows that conformal symmetry is of essential importance for the structure of the theory. Even in the flat space M_D, there exists a conformal anomaly, which can spoil the conformal invariance of the quantum theory; the anomaly is absent only in $D = 26$ (Polyakov 1981). The problem of conformal anomaly is expected to become more severe in a curved background V_D. The requirement that quantum string theory maintains its classical conformal invariance, which can be expressed as the condition that all the β functions vanish, implies the field equations of the background geometry. In the bosonic theory of closed strings, these equations are nothing but the classical field equations for the massless background fields (the graviton, antisymmetric tensor field and dilaton), including the Einstein equations for gravity (Fradkin and Tseytlin 1985, Callan *et al* 1985). These equations can be derived from an effective action as the equations of motion.

If we now identify the background fields that describe the geometry of V_D with the massless string fields, this result can be interpreted as a derivation of the effective action for the massless sector of the closed string. We can think of the background fields as condensates of the string fluctuations. In this way, conformal symmetry, in the form of a consistency condition, yields the covariant field equations for the massless sector of the string. Note that these equations are not the string equations of motion, they are only consistency conditions for the string dynamics. Thus, the general covariance of the background field equations follows directly from the conformal invariance of the string dynamics. It would be very interesting to interpret the meaning of these results in string field theory.

Anomalies and compactification. Classical field theories may have symmetries that are broken by quantum effects known as anomalies. These effects originate from certain Feynman diagrams, which do not admit a regularization compatible with all classical symmetries. Anomalies in local symmetries, such as internal gauge symmetries or general covariance, lead to inconsistencies at the quantum level, which show up as a breakdown of renormalizability and unitarity. Anomalies can be avoided by imposing certain restrictions on gauge couplings, so that contributions from anomalies in different Feynman diagrams cancel each other. Studying these requirements for supersymmetric theories in ten dimensions, it has been found that the cancellation of gauge and gravitational anomalies singles out superstring theories based on $SO(32)$ and $E_8 \times E_8$ gauge groups (see, e.g., Green *et al* 1987). The latter possibility has led to phenomenologically promising heterotic string models.

In any superstring theory we should be able to construct a realistic four-dimensional spacetime, starting from the space of $D = 10$ dimensions in which the superstring 'lives'. As in KK theory, this problem could be solved by spontaneous compactification: we should find a ground state which is 'curled up' in six dimensions at a sufficiently small scale, unobservable at currently available energies, so that the physical spacetime has effectively $10 - 6 = 4$ dimensions. The form of the the true ground state seems to be the central problem of string theory. If we knew the ground state, we might be able to discover whether superstring theory is a consistent unified field theory of all fundamental interactions. However, trying to solve this problem (perturbatively) we discover thousands of possible ground states, without being able to decide which one is correct. Clearly, until this fundamental problem is solved by non-perturbative calculations, we will not be able to build a unique and realistic string theory (Kaku 1991).

Membranes. String theory can be further generalized by considering higher-dimensional extended objects, such as membranes. The problem is mathematically more complex, since the basic equations are highly nonlinear (see, e.g., Collins and Tucker 1976, Howe and Tucker 1978, Taylor 1986, Kikkawa and Yamasaki 1986).

Exercises

1. A relativistic particle is described by the action

$$I[x, g, R] = -\int \mathrm{d}\tau \, [R_\mu \dot{x}^\mu - \tfrac{1}{2} g(R^2 - m^2)].$$

(a) Find the form of the action $I[x, g]$, which is obtained from $I[x, g, R]$ by eliminating R using its equation of motion. Then, in a similar way, eliminate g from $I[x, g]$ to obtain $I[x]$.
(b) Construct the canonical gauge generators for the action $I[x, g, R]$, and find its gauge symmetries.
(c) Repeat the same analysis for $I[x, g]$.

2. Find Noether currents corresponding to the global Poincaré invariance of the classical string. Prove the conservation of the related charges P_μ and $M_{\mu\nu}$, using the equations of motion and the boundary conditions.
3. Calculate the quantity $\pi^{(\sigma)} = \partial\mathcal{L}/\partial x'$, and derive the relations

$$(\pi^{(\sigma)})^2 + \dot{x}^2/(2\pi\alpha')^2 = 0 \qquad \pi^{(\sigma)} \cdot \dot{x} = 0.$$

Show that the end points of the open string move at the speed of light, in the direction orthogonal to the string.

4. A rigid, open string in M_4 rotates uniformly in the x^1–x^2 plane around its middle point located at the origin.

(a) Show that its motion is described by the equations

$$x^0 = \tau \qquad x^3 = 0$$
$$x^1 = A(\sigma - \pi/2)\cos\omega\tau \qquad x^2 = A(\sigma - \pi/2)\sin\omega\tau$$

where $\frac{1}{2}\pi\omega A = 1$.
(b) Prove that a light signal travels from one end of the string to the other for a finite amount of time.
(c) Compute the values of the mass M and angular momentum J of the string, and derive the relation $J = \alpha' M^2$.

5. Consider a closed string in M_4 that is, at time $\tau = 0$, at rest, and has the form of a circle in the x^1–x^2 plane:

$$x^1 = R\cos 2\sigma \qquad x^2 = R\sin 2\sigma \qquad x^0 = x^3 = 0.$$

(a) Find $x^0 = x^0(\tau)$ near $\tau = 0$ by solving the equations of motion and the constraint equations in the conformal gauge.
(b) Calculate the energy of the string.

6. (a) Use the oscillator basis to calculate the Virasoro generators for an open string, and check the form of the classical Virasoro algebra. Derive the form of the Poincaré charges in the same basis.
(b) Repeat the calculations for a closed string.
7. Verify the form of the quantum Virasoro algebra using the regularization

$$\hat{L}_n = \lim_{\Lambda\to\infty} -\frac{1}{2}\sum_{m=-\Lambda}^{\Lambda} : a_m \cdot a_{n-m} : .$$

8. Consider the quantum Virasoro algebra with a central charge:

$$[\hat{L}_n, \hat{L}_m] = (n-m)\hat{L}_{n+m} + C(n)\delta_{n,-m}.$$

   (a) Show that $C(-n) = -C(n)$.
   (b) Prove the relation $C(3n)+5C(n)-4C(2n) = 0$ using the Jacobi identity for $n = 2m \neq 0$.
   (c) If $C(n)$ is a polynomial in n, then $C(n) = c_1 n + c_3 n^3$. Prove it.
   (d) Calculate c_1 and c_3 directly from the values of $[\hat{L}_n, \hat{L}_{-n}]$ in the Fock space ground state, for $n = 1, 2$.

9. Show that the physical state conditions are invariant under Lorentz transformations, i.e. that $[\hat{M}^{\mu\nu}, \hat{L}_n] = 0$.
10. The form of the Hamiltonian in the light-cone gauge is obtained from expression (11.28b) by the replacement $\eta_{\mu\nu} \to -\delta_{\alpha\beta}$: $H = \frac{1}{2}\sum_r a_r^\alpha a_{-r}^\beta \delta_{\alpha\beta}$, where $\alpha, \beta = 1, 2, \ldots, D-2$.
   (a) Prove the relation $H =: H: -\alpha_0$, where $\alpha_0 \equiv -\frac{1}{2}(D-2)\sum_{r>0} r$.
   (b) Evaluate the regularized value of the sum $\sum_{r>0} r$ using a zeta function regularization. Show that $D = 26$ yields $\alpha_0 = 1$.
11. Show that the constraint $L_0 - \tilde{L}_0$, in the closed string theory, generates global spatial 'translations', $\sigma \to \sigma + a$.
12. Show that the implementation of the conditions $\hat{L}_n\psi = 0$, $n \neq 0$, is in conflict with the quantum Virasoro algebra.
13. Consider the physical state $\psi = -\mathrm{i}\varepsilon \cdot a_1^+ \Phi^{(k)}$, where $\varepsilon^\mu(k)$ is a polarization vector, and $\Phi^{(k)} = \exp(-\mathrm{i}k \cdot x)\Phi^{(0)}$ is the Fock vacuum of momentum k.
   (a) Show that $L_1\psi = 0$ implies $\varepsilon \cdot k = 0$.
   (b) Calculate the norm of the state ψ.
   (c) Use $(L_0 - \alpha_0)\psi = 0$ to show that the norm of ψ is negative for $\alpha_0 > 1$.
14. Consider the physical state $\psi = [a_1^+ \cdot a_1^+ + \beta a_0 \cdot a_2^+ + \gamma(a_0 \cdot a_1^+)^2]\Phi^{(k)}$.
   (a) Use the Virasoro conditions and $\alpha_0 = 1$ to find the values of k^2, β and γ.
   (b) Show that the norm of ψ is negative for $D > 26$.
15. Derive, in the field theory of the open string,
   (a) the gauge conditions (11.47); and
   (b) the gauge transformations (11.48).
16. (a) Show that $\Phi^{(k)} = \exp(-\mathrm{i}k \cdot x)\Phi^{(0)}$ can be taken as the zero-level state ψ_0 of a Verma module, in the open string model. Construct the states at levels $v = 1, 2$, and calculate the corresponding eigenvalues of L_0.
   (b) Evaluate the matrix $S_2^{ij} = (\psi_2^{(i)}, \psi_2^{(j)})$, where $\psi_2^{(1)} = L_{-1}^2\psi_0$, $\psi_2^{(2)} = L_{-2}\psi_0$. Can the determinant $\det(S_2^{ij})$ vanish for $D > 1$ and $h > 0$?
17. Verify, in field theory of the closed string,
   (a) the gauge conditions (11.53); and

(b) the gauge transformations (11.55).

18. Check the form of the action (11.56) for the closed string, and show that it can be written as in equation (11.60).

Appendix A

Local internal symmetries

We discuss here certain properties of theories with local internal non-Abelian symmetries, which will be of particular interest in the process of generalization to local spacetime symmetries (Utiyama 1956, Kibble 1961, Abers and Lee 1973).

Localization of internal symmetries. Let us consider a multicomponent matter field $\phi(x) = \{\phi^m(x)\}$, transforming according to some representation of a Lie group of internal symmetries G. An infinitesimal transformation has the form

$$\begin{aligned} \phi'(x) &= \phi(x) + \delta_0\phi(x) \\ \delta_0\phi(x) &= \theta^a T_a \phi(x) \equiv \theta\phi(x) \qquad (a = 1, 2, \ldots, n) \end{aligned} \tag{A.1}$$

where the θ^a are *constant* infinitesimal parameters, the T_a are the group generators satisfying the commutation relations

$$[T_a, T_b] = f_{ab}{}^c T_c \tag{A.2a}$$

and the structure constants $f_{ab}{}^c$ satisfy the Jacobi identity:

$$f_{ae}{}^m f_{bc}{}^e + \text{cyclic}\,(a, b, c) = 0. \tag{A.2b}$$

Since the θ^a are constant parameters, $\partial_\mu\phi$ transforms like the field itself,

$$\delta_0\partial_\mu\phi(x) = \partial_\mu\delta_0\phi(x) = \theta\partial_\mu\phi(x)$$

because δ_0 and ∂_μ commute.

The action integral $I_\mathrm{M} = \int \mathrm{d}^4x\, \mathcal{L}_\mathrm{M}(\phi, \partial_\mu\phi)$ is invariant under the previous transformations if $\mathcal{L}_\mathrm{M}$ satisfies the condition

$$\delta_0\mathcal{L}_\mathrm{M} \equiv \mathcal{L}_\mathrm{M}(\phi'(x), \partial_\mu\phi'(x)) - \mathcal{L}_\mathrm{M}(\phi(x), \partial_\mu\phi(x)) = 0$$

which implies n identities:

$$\frac{\partial\mathcal{L}_\mathrm{M}}{\partial\phi} T_a\phi + \frac{\partial\mathcal{L}_\mathrm{M}}{\partial\partial_\mu\phi} T_a\partial_\mu\phi = 0. \tag{A.3}$$

These identities, with the help of the equations of motion,

$$\frac{\partial \mathcal{L}_{\mathrm{M}}}{\partial \phi} - \partial_\mu \frac{\partial \mathcal{L}_{\mathrm{M}}}{\partial \partial_\mu \phi} = 0$$

lead to the conservation of the canonical current:

$$\partial_\mu J_a^\mu = 0 \qquad J_a^\mu \equiv -\frac{\partial \mathcal{L}_{\mathrm{M}}}{\partial \partial_\mu \phi} T_a \phi. \tag{A.4}$$

Let us now consider these transformations with the constant parameters replaced by arbitrary *functions* of position, $\theta^a = \theta^a(x)$; they are called *local* or *gauge* transformations. The Lagrangian is no longer invariant because the transformation law of $\partial_\mu \phi$ is modified:

$$\delta_0 \partial_\mu \phi = \theta \partial_\mu \phi + \theta_{,\mu} \phi.$$

Indeed, a direct calculation yields

$$\delta_0 \mathcal{L}_{\mathrm{M}} = \frac{\partial \mathcal{L}_{\mathrm{M}}}{\partial \partial_\mu \phi} \theta_{,\mu} \phi = -\theta^a_{,\mu} J^\mu_a.$$

Covariant derivative. The invariance under local transformations can be restored by certain modifications to the original theory. Let us introduce a new Lagrangian

$$\mathcal{L}'_{\mathrm{M}} = \mathcal{L}_{\mathrm{M}}(\phi, \nabla_\mu \phi) \tag{A.5}$$

where $\nabla_\mu \phi$ is the *covariant derivative*, which transforms under local transformations in the same way as $\partial_\mu \phi$ does under the global ones:

$$\delta_0 \nabla_\mu \phi = \theta \nabla_\mu \phi. \tag{A.6}$$

It is now easy to see, with the help of (A.3), that the new Lagrangian is invariant under local transformations:

$$\delta_0 \mathcal{L}'_{\mathrm{M}} = \frac{\bar{\partial} \mathcal{L}_{\mathrm{M}}}{\partial \phi} \theta \phi + \frac{\partial \mathcal{L}_{\mathrm{M}}}{\partial \nabla_\mu \phi} \theta \nabla_\mu \phi = 0$$

where $\bar{\partial} \mathcal{L}_{\mathrm{M}} / \partial \phi = [\partial \mathcal{L}_{\mathrm{M}}(\phi, \nabla u)/\partial \phi]_{u=\phi}$. The covariant derivative is constructed by introducing the *compensating fields* (gauge potentials) A_μ,

$$\nabla_\mu \phi = (\partial_\mu + A_\mu)\phi \qquad A_\mu \equiv T_a A^a{}_\mu \tag{A.7}$$

for which the transformation law follows from (A.6):

$$\delta_0 A^a{}_\mu = (-\partial_\mu \theta - [A_\mu, \theta])^a = -\partial_\mu \theta^a - f_{bc}{}^a A^b{}_\mu \theta^c. \tag{A.8}$$

It is instructive to observe that the form of the covariant derivative $\nabla \phi$ is determined by the transformation rule of ϕ:

$$\nabla_\mu \phi = \partial_\mu \phi + \delta_0 \phi|_{\theta \to A_\mu}.$$

Field strength. The commutator of two covariant derivatives has the form

$$[\nabla_\mu, \nabla_\nu]\phi = F^a{}_{\mu\nu} T_a \phi \equiv F_{\mu\nu}\phi$$

where $F_{\mu\nu}$ is the *field strength*,

$$\begin{aligned} F^a{}_{\mu\nu} &= \partial_\mu A^a{}_\nu - \partial_\nu A^a{}_\mu + f_{bc}{}^a A^b{}_\mu A^c{}_\nu \\ &= (\partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu])^a \end{aligned} \tag{A.9}$$

transforming as

$$\delta_0 F^a{}_{\mu\nu} = f_{bc}{}^a \theta^b F^c{}_{\mu\nu} = [\theta, F_{\mu\nu}]^a.$$

Therefore, we define the related covariant derivative as

$$\nabla_\lambda F^a{}_{\mu\nu} = \partial_\lambda F^a{}_{\mu\nu} + f_{bc}{}^a A^b{}_\lambda F^c{}_{\mu\nu} \equiv (\partial_\lambda F_{\mu\nu} + [A_\lambda, F_{\mu\nu}])^a.$$

The Jacobi identity for the commutator of covariant derivatives implies the Bianchi identity for $F_{\mu\nu}$:

$$\nabla_\lambda F^a{}_{\mu\nu} + \nabla_\nu F^a{}_{\lambda\mu} + \nabla_\mu F^a{}_{\nu\lambda} = 0 \quad \Longleftrightarrow \quad \nabla_\mu {}^*F^{a\mu\nu} = 0$$

where ${}^*F^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\lambda\rho} F_{\lambda\rho}$ is the dual tensor of $F_{\mu\nu}$.

Invariant Lagrangian. The matter Lagrangian is made invariant under local transformations by introducing gauge potentials. The next step is to construct a free Lagrangian $\mathcal{L}_\mathrm{F}(A, \partial A)$ for the new fields, which should also be locally invariant:

$$\delta_0 \mathcal{L}_\mathrm{F} = \frac{\partial \mathcal{L}_\mathrm{F}}{\partial A^a{}_\mu} \delta_0 A^a{}_\mu + \frac{\partial \mathcal{L}_\mathrm{F}}{\partial A^a{}_{\mu,\nu}} \delta_0 A^a{}_{\mu,\nu} = 0$$

Demanding that the coefficients of $\theta^b, \theta^b_{,\mu}$ and $\theta^b_{,\mu\nu}$ vanish, we obtain the following identities:

$$\frac{\partial \mathcal{L}_\mathrm{F}}{\partial A^a{}_\mu} f_{bc}{}^a A^c{}_\mu + \frac{\partial \mathcal{L}_\mathrm{F}}{\partial A^a{}_{\mu,\nu}} f_{bc}{}^a A^c{}_{\mu,\nu} = 0 \tag{A.10a}$$

$$\frac{\partial \mathcal{L}_\mathrm{F}}{\partial A^b{}_\mu} + \frac{\partial \mathcal{L}_\mathrm{F}}{\partial A^a{}_{\mu,\nu}} f_{bc}{}^a A^c{}_\nu = 0 \tag{A.10b}$$

$$\frac{\partial \mathcal{L}_\mathrm{F}}{\partial A^b{}_{\mu,\nu}} + \frac{\partial \mathcal{L}_\mathrm{F}}{\partial A^b{}_{\nu,\mu}} = 0. \tag{A.10c}$$

The last identity implies that $\partial_\mu A_\nu$ occurs in $\mathcal{L}_\mathrm{F}$ only in the antisymmetric combination $\partial_\mu A_\nu - \partial_\nu A_\mu$. The condition (A.10$b$) means that the exact form of that combination is given as the field strength $F_{\mu\nu}$. From the same condition we can conclude that the only dependence on A_μ is through $F_{\mu\nu}$. The first identity means that $\mathcal{L}_\mathrm{F}$ is an invariant function of $F_{\mu\nu}$. Indeed, after the elimination of

$\partial\mathcal{L}_F/\partial A_\mu$ by using (A.10*b*) and the Jacobi identity (A.2*b*), the relation (A.10*a*) implies

$$\frac{\partial\mathcal{L}_F}{\partial F^a{}_{\mu\nu}} f_{cb}{}^a F^b{}_{\mu\nu} = \frac{\partial\mathcal{L}_F}{\partial F^a{}_{\mu\nu}} \delta_0 F^a{}_{\mu\nu} = 0.$$

If we require the equations of motion not to contain derivatives higher than second order ones, $\mathcal{L}_F$ must the quadratic invariant:

$$\mathcal{L}_F = -\frac{1}{4g^2} g_{ab} F^a{}_{\mu\nu} F^{b\mu\nu} \tag{A.11}$$

where g is an interaction constant, and g_{ab} is the Cartan metric of the Lie algebra of G, as we shall see soon. The factor g^{-2} can be easily eliminated by rescaling the gauge potentials, $A^a{}_\mu \to gA^a{}_\mu$, but then g reappears in the covariant derivative.

Conservation laws. The equations of motion for matter fields ϕ can be written in the covariant form:

$$\frac{\bar{\partial}\mathcal{L}_M}{\partial\phi} - \nabla_\mu \frac{\partial\mathcal{L}_M}{\partial\nabla_\mu\phi} = 0. \tag{A.12}$$

Here, the covariant derivative of $K^\mu \equiv \partial\mathcal{L}_M/\partial\nabla_\mu\phi$ is defined in accordance with the transformation rule $\delta_0 K^\mu = -K^\mu\theta$, contragradient with respect to ϕ. Using the equations of motion, the invariance condition yields the 'conservation law' of the *covariant current* J'^μ_a:

$$\begin{aligned} &\nabla_\mu J'^\mu_a \equiv \partial_\mu J'^\mu_a + f_{ab}{}^c A^b{}_\mu J'^\mu_c = 0 \\ &J'^\mu_a \equiv -\frac{\partial\mathcal{L}_M}{\partial\nabla_\mu\phi} T_a\phi = -\frac{\partial\mathcal{L}'_M}{\partial A^a{}_\mu}. \end{aligned} \tag{A.13}$$

Of course, this condition is not a true conservation law, for which the usual four-divergence should vanish. The covariant current transforms according to the rule

$$\delta_0 J'^\mu_a = f_{ab}{}^c \theta^b J'^\mu_c$$

which is used to define its covariant derivative.

The complete Lagrangian of the theory $\mathcal{L} = \mathcal{L}_F + \mathcal{L}'_M$ yields the following equations of motion for A_μ:

$$\nabla_\mu F_a{}^{\mu\nu} = J'^\nu_a. \tag{A.14}$$

Here, $F_a \equiv g_{ab} F^b$, and we use $g = 1$ for simplicity. If we write this equation in the form

$$\partial_\mu F_a{}^{\mu\nu} = J'^\nu_a + j^\nu_a \qquad j^\nu_a \equiv -f_{ab}{}^c A^b{}_\mu F_c{}^{\mu\nu}$$

we obtain, using the antisymmetry of $F^{\mu\nu}$, the *true* conservation law:

$$\partial_\nu(J'^\nu_a + j^\nu_a) = 0.$$

Since j^ν_a is not a covariant quantity, the true conservation law is obtained at the expense of losing the covariance of the conserved quantity.

Construction of the invariant Lagrangian. Now we shall show how we can construct the quadratic invariant from the field strength F^a. First, let us mention two useful facts from the theory of Lie groups. First, from the Jacobi identity for the generators T_a,

$$[T_a, [T_b, T_c]] + \text{cyclic}(a, b, c) = 0$$

we obtain relation $(A.2b)$, which is also called the Jacobi identity. Second, there exists a representation of the Lie algebra, called the adjoint representation, which is completely determined by the structure constants: $(T'_a)^b{}_c = f_{ac}{}^b$. This is easily seen from the relation

$$[T'_a, T'_b]^c{}_d = f_{ab}{}^e (T_e)^c{}_d$$

that follows from (A.2b). If we write the transformation law for F^a in the form

$$\delta_0 F^a = \theta^b (T'_b)^a{}_c F^c = (\theta' F)^a \tag{A.15a}$$

we can see that the field strength F^a transforms cogradiently to ϕ, i.e. according to the same rule as ϕ. From two quantities G_a and F^a (spacetime indices are, for simplicity, omitted) we can construct the bilinear invariant $G_a F^a$, if G_a transforms contragradiently with respect to ϕ, i.e. as

$$\delta_0 G_a = -(G\theta')_a. \tag{A.15b}$$

Following the standard terminology, we call F^a a contravariant vector, and G_a a covariant vector. In order to construct an invariant quadratic in F^a, we choose

$$G_a = F_a \qquad F_a \equiv g_{ab} F^b$$

where g_{ab} is the Cartan metric on the Lie algebra of G:

$$g_{ab} = -\tfrac{1}{2}\,\mathrm{Tr}(T'_a T'_b) = -\tfrac{1}{2} f_{ae}{}^c f_{bc}{}^e. \tag{A.16}$$

With this choice of metric, F_a transforms covariantly. Indeed, if we start with

$$\delta_0 F_a = g_{ab}\delta_0 F^b = g_{ab} f_{cd}{}^b \theta^c F^d$$

and use the fact that $f_{cda} \equiv f_{cd}{}^b g_{ba}$ is totally antisymmetric, we easily obtain $\delta_0 F_a = -\theta^c f_{cad} F^d = -\theta^c f_{ca}{}^d F_d$.

If we consider $f_{ab}{}^c$ as a third rank tensor with its type determined by the position of its indices, then $f_{ab}{}^c$ is a constant and invariant tensor:

$$\delta_0 f_{ab}{}^c = (f_{ed}{}^c f_{ab}{}^d - f_{eb}{}^d f_{ad}{}^c - f_{ea}{}^d f_{db}{}^c)\theta^e = 0$$

as follows from the Jacobi identity. The same is true for g_{ab}.

If the group G is *semisimple*, i.e. it does not contain a non-trivial, invariant, Abelian subgroup, then $\det(g_{ab}) \neq 0$. In that case, the inverse metric g^{ab} exists and we can construct the standard tensor algebra.

A semisimple group G is *compact* if and only if the Cartan metric tensor is positive (negative) definite. In that case, by a suitable choice of basis we can transform the metric g_{ab} to the form of a unit tensor, $g_{ab} = \delta_{ab}$, and the structure constants $f_{ab}{}^c$ become completely antisymmetric.

Example 1. Let us consider two examples of the construction of the Cartan metric tensor. For the rotation group $SO(3)$, the structure constants are completely antisymmetric, $f_{ab}{}^c = -\epsilon_{abc}$, so that

$$g_{ab} = -\tfrac{1}{2}\epsilon_{aef}\epsilon_{bfe} = \delta_{ab} \qquad g_{ab}F^a F^b = \delta_{ab}F^a F^b .$$

Our second example is the Lorentz group $SO(1,3)$, with structure constants as defined in (2.6). The Cartan metric tensor has the form

$$g_{ij,kl} = -\tfrac{1}{8} f_{ij,rm}{}^{sn} f_{kl,sn}{}^{rm} = 2(\eta_{ik}\eta_{jl} - \eta_{il}\eta_{jk})$$

and the quadratic invariant becomes $g_{ij,kl}F^{ij}F^{kl} = 4F^{ij}F_{ij}$.

Exercises

1. (a) Derive gauge transformations for A_μ and $F_{\mu\nu}$.
   (b) Prove the Bianchi identity: $\nabla_\mu {}^*F^{\mu\nu} = 0$.
2. (a) Find the transformation law for $K^\mu = \partial\mathcal{L}_\mathrm{M}/\partial\nabla_\mu\phi$ and define $\nabla_\nu K^\mu$.
   (b) Show that the equations of motion for matter fields ϕ, derived from the Lagrangian $\mathcal{L}'_\mathrm{M}$, have the G-covariant form (A.12).
3. (a) Find the transformation law for the covariant current J'^μ_a, and define $\nabla_\nu J'^\mu_a$.
   (b) Use the equations of motion for matter fields to prove the 'conservation law' $\nabla_\mu J'^\mu_a = 0$.
4. Show that $f_{abc} = f_{ab}{}^e g_{ce}$ is a totally antisymmetric quantity.
5. Show that the Lagrangian $\mathcal{L}_\mathrm{F}$ in (A.11) is invariant under local G-transformations.
6. Let a Lie group G contain a non-trivial, invariant, Abelian subgroup. Show that the Cartan metric tensor is singular.

Appendix B

Differentiable manifolds

In order to make the geometric content of local spacetime symmetries clearer, we shall give here a short overview of the mathematical structure of differentiable manifolds (Misner *et al* 1970, Choquet-Bruhat *et al* 1977).

Topological spaces. One of the basic concepts of mathematical analysis is the limiting process, which is based on the existence of distance on the real line. Many important results are based only on the existence and properties of the distance. By generalizing the notion of real line as a set on which there exists a distance, we arrive at the concept of metric spaces. Metric spaces are a natural generalization of Euclidean spaces. It follows from the study of metric spaces that the essence of the limiting process lies in the existence of neighbourhoods or open sets. A further generalization of metric spaces leads to topological spaces, where open sets are introduced directly; here we have a natural structure for studying continuity.

Let X be a set, and $\tau = \{O_\alpha\}$ a collection of subsets of X. A collection τ defines a *topology* on X if τ contains:

- the empty set $\emptyset$ and the set X itself;
- arbitrary union $\cup_\alpha O_\alpha$ (finite or infinite) of elements in τ; and
- finite intersection $\cap_{\alpha=1}^{n} O_\alpha$ of elements in τ.

The pair (X, τ), often abbreviated to X, is called the *topological space*. The sets O_α in τ are *open sets* of the topological space, while their complements $X \backslash O_\alpha$ are *closed*.

We can introduce different topologies on one and the same set X and thereby define different topological spaces. The topology may be characterized by defining a *basis*, i.e. a collection B of open subsets, such that every open subset of X can be expressed as a union of elements of B. A basis for the usual topology on the real line $\mathcal{R}$ consists of all open intervals (a, b).

A collection $\{U_\alpha\}$ of open subsets of X is an *open covering* if $\cup U_\alpha = X$. We can define the *induced topology* on a subset Y of X by demanding that open sets in Y are of the form $Y \cap O_\alpha$, where O_α are open in X.

The *neighbourhood* (open) of a point P in X is every open set O_P containing P. The notion of a neighbourhood is here based on the existence of open sets (not on the existence of distance as our experience with metric spaces might have suggested). Open sets enable a natural definition of *continuous mappings* (a mapping $f : X \to Y$ is continuous if the inverse image of an open set in Y is open in X), and homeomorphisms.

A mapping $f : X \to Y$ from a topological space X to a topological space Y is called a *homeomorphism* if

- f is a 1–1 correspondence (1–1 and onto); and
- both f and f^{-1} are continuous.

Homeomorphic spaces have identical topological properties.

Although topological spaces are the natural generalization of metric spaces, their structure is, often, too general. In order to single out those topological spaces that are more interesting from the point of view of specific mathematical or physical applications, we usually introduce some additional conditions on their structure, such as connectedness, axioms of separation, etc. Here are several examples:

- a topological space is *connected* if it is not a union of two disjoint (non-empty) open sets;
- a topological space is *Hausdorff* if any two disjoint points possess disjoint neighbourhoods;
- a topological space is *compact* if every open covering has a finite subcovering (a subset which is itself a covering).

Without trying to clarify the nature of these additional conditions, we shall assume that they are always fulfilled to an extent which is sufficient for all our considerations.

Differentiable manifolds. If we think of topological spaces as the natural structure for studying continuity, differentiable manifolds are the natural structure for studying differentiability.

A topological *manifold* X is a (Hausdorff) topological space, such that every point has a neighbourhood homeomorphic to an open set in $\mathcal{R}^n$. The number n is the dimension of a manifold. Thus, manifolds are 'locally Euclidean' topological spaces. Using this property, we can introduce *local coordinate systems* (or charts) on X. Let $\{O_i\}$ be a covering of X, and φ_i homeomorphisms, mapping O_i onto open sets Ω_i in $\mathcal{R}^n$, $\varphi_i : O_i \to \Omega_i$. Then, the image of a point P in O_i is $\varphi_i(P) = (x_i^1, x_i^2, \ldots, x_i^n)$, and x_i^μ are local coordinates of P.

For a given family $\varphi = \{\varphi_i\}$ of homeomorphisms, at each intersection $O_{ij} = O_i \cap O_j$ there are two local coordinate systems, (x_i^μ) and (x_j^μ), and we may question their compatibility. A collection of all local coordinate systems (O_i, φ_i) defines a *coordinate system* (or atlas) on X, if the transition functions

$\varphi_{ij} = \varphi_j \circ \varphi_i^{-1} : (x_i^\mu) \to (x_j^\mu)$ are smooth functions (= of class C^m), for each pair (O_i, O_j). This consideration is made mathematically complete by considering not only one coordinate system, related to a given family φ of homeomorphisms, but the set of all equivalent coordinate systems (two coordinate systems of class C^m are equivalent if their local coordinate systems are C^m compatible). Then, a *differentiable manifold* (or a smooth manifold) is defined as a topological manifold together with the set of all equivalent coordinate systems.

After introducing local coordinates, we can define differentiable mappings of manifolds. Consider a mapping $f : X \to Y$, and denote by x_0^μ and y_0^μ the local coordinates of the points $P \in X$ and $f(P) \in Y$, respectively. Then, f is *differentiable* at P if it is differentiable in coordinates, i.e. if local coordinates y are differentiable functions of local coordinates x at $x = x_0$. In a similar manner we can define *smooth* mappings (of class C^r, $r \leq m$). These definitions do not depend on the choice of local coordinates. The differentiability is used to extend the notion of homeomorphisms to diffeomorphisms.

A mapping $f : X \to Y$ from a differentiable manifold X to a differentiable manifold Y is a *diffeomorphism* if

- f is a 1–1 correspondence; and
- both f and f^{-1} are smooth (of class C^r, $r \leq m$).

Diffeomorphic manifolds have identical differential properties. Diffeomorphisms are for differentiable manifolds the same as homeomorphisms are for topological spaces.

Tangent vectors. Consider a smooth curve $C(\lambda)$ on a differentiable manifold X, defined by a smooth mapping $C : \mathcal{R} \to X$. A *tangent vector* to $C(\lambda)$ at the point $P = C(0)$ is defined by the expression

$$\boldsymbol{v} = \lim_{\lambda \to 0} \frac{C(\lambda) - C(0)}{\lambda} = \left.\frac{\mathrm{d}C(\lambda)}{\mathrm{d}\lambda}\right|_{\lambda=0}. \tag{B.1}$$

In the case of a two-dimensional surface embedded in E_3, tangent vectors, according to this definition, can be visualized as vectors lying in the tangent plane at $C(0)$. But an 'infinitesimal displacement', $C(\lambda) - C(0)$, is not a well-defined geometric object within the manifold itself. If we imagine that the manifold of interest is embedded in a higher dimensional flat space, the geometric interpretation becomes completely clear, but the definition here relies on a specific way of embedding.

In order to define tangent vectors in terms of the internal structure of the manifold, we should abandon the idea of the 'displacement' of a point, and consider only changes in objects that always stay within the manifold; such objects are differentiable functions $f : X \to \mathcal{R}$. In the expression for a change in f, $\mathrm{d}f/\mathrm{d}\lambda$, the part that does not depend on f is the operation $\mathrm{d}/\mathrm{d}\lambda$, so that the tangent vector on $C(\lambda)$ at $P = C(0)$ is defined as the *differential operator*

$$\boldsymbol{v} = \frac{\mathrm{d}}{\mathrm{d}\lambda} \tag{B.2a}$$

calculated at $\lambda = 0$. This operator maps differentiable functions to $\mathcal{R}$ according to

$$\boldsymbol{v}(f) = \frac{\mathrm{d}}{\mathrm{d}\lambda}(f \circ C)\bigg|_{C(0)}. \tag{B.2b}$$

The set of all directional derivatives (B.2) at P has the structure of a vector space (with respect to the usual addition and multiplication by scalars),

$$(a\boldsymbol{v}_1 + b\boldsymbol{v}_2)(f) = a\boldsymbol{v}_1(f) + b\boldsymbol{v}_2(f)$$

and is called the *tangent space* T_P of the manifold X at P. Differential operators $\boldsymbol{v} \in T_P$ are linear and satisfy the Leibniz rule (these two properties can be taken as the defining properties of tangent vectors).

Although the notion of 'infinitesimal displacement' is now given a precise mathematical meaning, it may not be quite clear in which sense the differential operators (B.2) and the tangent vectors (B.1) are the *same* objects. The answer lies in the observation that these two vector spaces are *isomorphic* (an isomorphism between vector spaces is a 1–1 correspondence that 'preserves' the vector space operations). For instance, any linear combination of vectors (B.1) maps into the same linear combination of directional derivatives (B.2) and vice versa:

$$\boldsymbol{u} = a\boldsymbol{v}_1 + b\boldsymbol{v}_2 \quad \Longleftrightarrow \quad \boldsymbol{u}(f) = a\boldsymbol{v}_1(f) + b\boldsymbol{v}_2(f).$$

Thus, directional derivatives represent an abstract realization of the usual notion of tangent vectors. Such a unification of the concepts of analysis and geometry has many far-reaching consequences.

Consider a coordinate line x^μ in $\mathcal{R}^n$, and the curve $\varphi_i^{-1} : x^\mu \to X$. The tangent vector to this curve at a point P is

$$\boldsymbol{e}_\mu(f) = \frac{\partial}{\partial x^\mu}(f \circ \varphi_i^{-1})\bigg|_{x^\mu(P)} \tag{B.3a}$$

Vectors $(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n)$ are linearly independent, so that every vector $\boldsymbol{v}$ in T_P can be expressed as

$$\boldsymbol{v}(f) = v^\mu \boldsymbol{e}_\mu(f) \tag{B.3b}$$

where, by convention, the repeated index denotes a summation. It follows that $\dim(T_P) = n$. The set of vectors $\boldsymbol{e}_\mu$ forms a basis (frame) in T_P which is called the coordinate (or natural) basis, and v^μ are components of $\boldsymbol{v}$ in this basis. The usual notation for $\boldsymbol{e}_\mu$ is $\partial/\partial x^\mu$. A transformation from one coordinate basis to another has the form

$$\boldsymbol{e}'_\mu = \frac{\partial x^\nu}{\partial x'^\mu}\boldsymbol{e}_\nu \tag{B.4a}$$

while, at the same time, the components of $\boldsymbol{v}$ transform by the inverse law:

$$v'^{\mu} = \frac{\partial x'^{\mu}}{\partial x^{\nu}} v^{\nu}. \tag{B.4b}$$

This equation is known as the vector transformation law, and it can be used to define tangent vectors in terms of the local coordinates. Vectors $\boldsymbol{v} = (v^{\mu})$ are often called contravariant vectors.

A *vector field* V_X on X is a mapping that associates a tangent vector $\boldsymbol{v}_P$ with each point $P \in X$. Although the existing structure of a differentiable manifold does not allow the tangent vectors to be compared at different points, there is a natural way to define the smoothness of the vector field when we move from one point to another. Consider a mapping from X to $\mathcal{R}$, defined by $P \mapsto \boldsymbol{v}_P(f)$, where f is a smooth real function on X. Then, a vector field V_X is *smooth* if the mapping $\boldsymbol{v}_P(f)$ is a smooth function on X. In terms of the coordinates, a vector field is smooth if its components $v^{\mu}(x)$ are smooth functions.

Dual vectors. Following the usual ideas of linear algebra, we can associate a dual vector space T_P^* with each tangent space T_P of X. Consider linear mappings from T_P to $\mathcal{R}$, defined by $\boldsymbol{w}^* : \boldsymbol{v} \mapsto \boldsymbol{w}^*(\boldsymbol{v}) \in \mathcal{R}$. If a set of these mappings is equipped with the usual operations of addition and scalar multiplication, we obtain the *dual* vector space T_P^*. Vectors $\boldsymbol{w}^*$ in T_P^* are called dual vectors, covariant vectors (covectors) or differential forms. There is no natural isomorphism between a tangent space and its dual. However, for a given basis $\boldsymbol{e}_b$ in T_P, we can construct its dual basis $\boldsymbol{\theta}^a$ in T_P^* by demanding $\boldsymbol{\theta}^a(\boldsymbol{v}) = v^a$, so that

$$\boldsymbol{\theta}^a(\boldsymbol{e}_b) = \delta^a_b.$$

This implies, in particular, that $\dim(T_P^*) = \dim(T_P)$. The correspondence $\boldsymbol{\theta}^a \leftrightarrow \boldsymbol{e}_a$ is an isomorphism, but it does not relate geometric objects—it depends on the choice of the basis $\boldsymbol{e}_a$. The spaces T_P and T_P^* cannot be identified in a natural (geometric) way without introducing some additional structure on X.

The space T_P^{**} is isomorphic to T_P. This follows from the fact that, with each $\boldsymbol{u}^{**}$ in T_P^{**} we can associate a vector $\boldsymbol{u}$ in T_P, such that $\boldsymbol{u}^{**}(\boldsymbol{w}^*) = \boldsymbol{w}^*(\boldsymbol{u})$, for every $\boldsymbol{w}^* \in T_P^*$.

Let $\boldsymbol{\theta}^{\mu}$ be the basis dual to the coordinate basis $\boldsymbol{e}_{\nu}$. Each form $\boldsymbol{w}^* \in T_P^*$ can be represented as

$$\boldsymbol{w}^* = w^*_{\mu} \boldsymbol{\theta}^{\mu}. \tag{B.5a}$$

A change of coordinates induces the following change in $\boldsymbol{\theta}^{\mu}$ and w^*_{μ}:

$$\boldsymbol{\theta}'^{\mu} = \frac{\partial x'^{\mu}}{\partial x^{\nu}} \boldsymbol{\theta}^{\nu} \qquad w^{*\prime}_{\mu} = \frac{\partial x^{\nu}}{\partial x'^{\mu}} w^*_{\nu}. \tag{B.5b}$$

Using the standard convention we shall omit the sign * for dual vectors, and label dual vectors by subscripts, and vectors by superscripts.

A dual vector field V_X^* on X is defined by analogy with V_X.

Note that 'being dual' is a symmetrical relation. Indeed, just as a dual vector is a mapping from T_P to $\mathcal{R}$, $\boldsymbol{v} \mapsto \boldsymbol{w}^*(\boldsymbol{v}) \in R$, so a tangent vector is a mapping from T_P^* to $\mathcal{R}$, $\boldsymbol{w}^* \mapsto \boldsymbol{v}(\boldsymbol{w}^*) \in \mathcal{R}$, where $\boldsymbol{v}(\boldsymbol{w}^*) = \boldsymbol{w}^*(\boldsymbol{v})$.

Tensors. The concept of a dual vector, as a linear mapping from T_P to $\mathcal{R}$, can be naturally extended to the concept of tensor, as a multilinear mapping. We shall begin our consideration by some simple examples.

A *tensor product* $T_P(0,2) = T_P^* \otimes T_P^*$ of two dual spaces at P is the vector space of all bilinear forms $\boldsymbol{\omega}$, mapping $T_P \times T_P$ to $\mathcal{R}$ by the rule $(\boldsymbol{u}, \boldsymbol{v}) \mapsto \boldsymbol{\omega}(\boldsymbol{u}, \boldsymbol{v})$. The tensor product $T_P(0,2)$ should be clearly distinguished from the Cartesian product $T_P^* \times T_P^*$, defined as a set of pairs $(\boldsymbol{u}^*, \boldsymbol{v}^*)$. A particularly simple, 'factorized' tensor product is determined by mappings $(\boldsymbol{u}, \boldsymbol{v}) \mapsto \boldsymbol{u}^*(\boldsymbol{u}) \cdot \boldsymbol{v}^*(\boldsymbol{v})$. Let $\boldsymbol{\theta}^a$ be a basis in T_P^*, and define $\boldsymbol{\theta}^a \otimes \boldsymbol{\theta}^b$ by $\boldsymbol{\theta}^a \otimes \boldsymbol{\theta}^b(\boldsymbol{u}, \boldsymbol{v}) = u^a v^b$. Then $\boldsymbol{\theta}^a \otimes \boldsymbol{\theta}^b$ is a basis in $T_P(0,2)$:

$$\boldsymbol{\omega} = \omega_{ab}\, \boldsymbol{\theta}^a \otimes \boldsymbol{\theta}^b \quad \Rightarrow \quad \boldsymbol{\omega}(\boldsymbol{u}, \boldsymbol{v}) = \omega_{ab} u^a v^b.$$

In particular, if $\omega_{ab} = u_a^* v_b^*$, the tensor product space is 'factorized'.

A *tensor* of type $(0,2)$ is an element of $T_P(0,2)$; it is a bilinear mapping $\boldsymbol{\omega}$ which maps a pair of vectors $(\boldsymbol{u}, \boldsymbol{v})$ into a real number $\boldsymbol{\omega}(\boldsymbol{u}, \boldsymbol{v})$. Similarly, a tensor $\boldsymbol{\alpha}$ of type $(1,1)$ maps a pair $(\boldsymbol{v}, \boldsymbol{w}^*)$ into a real number $\boldsymbol{\alpha}(\boldsymbol{v}, \boldsymbol{w}^*)$.

After these examples it is not difficult to define a general tensor $\boldsymbol{t}$ of type (p, q). The space of tensors of a given type is a vector space. The components of a tensor $\boldsymbol{t}$ in the coordinate basis transform as the product of p vectors and q dual vectors. We can introduce in the usual way a multiplication between tensors and the operation of contraction.

By associating with each point $P \in X$ a tensor $\boldsymbol{t}_P$, we obtain a tensor field on X.

An important tensor that can be introduced on a differentiable manifold is the *metric tensor*. Intuitively, the metric serves to define the squared distance between any two neighbouring points. Since the displacement between two neighbouring points represents a tangent vector, the metric can be thought of as a rule for defining the squared length of a tangent vector. More precisely, the metric tensor $\boldsymbol{g}$ is a symmetrical, non-degenerate tensor field of type $(0,2)$, that maps a pair of vectors $(\boldsymbol{u}, \boldsymbol{v})$ into a real number $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v})$. The link between this abstract definition and the 'squared distance' point of view is easily exhibited. Let the tangent vector $\boldsymbol{\xi} = \mathrm{d}x^\mu\, \boldsymbol{e}_\mu$ represent the displacement vector between two neighbouring points. Since in the coordinate basis $\boldsymbol{g} = g_{\mu\nu}\, \boldsymbol{\theta}^\mu \otimes \boldsymbol{\theta}^\nu$, the abstract definition gives

$$\boldsymbol{g}(\boldsymbol{\xi}, \boldsymbol{\xi}) = g_{\mu\nu} \boldsymbol{\theta}^\mu \otimes \boldsymbol{\theta}^\nu(\boldsymbol{\xi}, \boldsymbol{\xi}) = g_{\mu\nu}\, \mathrm{d}x^\mu\, \mathrm{d}x^\nu = \mathrm{d}s^2 \tag{B.6}$$

in agreement with the intuitive viewpoint.

For any given vector $\boldsymbol{u}$, the quantity $\boldsymbol{u}^* = \boldsymbol{g}(\boldsymbol{u}, \cdot)$ belongs to T_P^*, since it maps $\boldsymbol{v}$ to a number $\boldsymbol{u}^*(\boldsymbol{v}) \equiv \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v})$. Thus, with the help of the metric we can define a natural isomorphism $\boldsymbol{u} \mapsto \boldsymbol{u}^*$ between T_P and T_P^*. In terms of the components, and after omitting the star symbol, this correspondence takes the form $u_\mu = g_{\mu\nu}u^\nu$. With a metric tensor defined on a manifold, there are automatically dual vectors there.

The metric tensor supplies manifolds with a geometric structure that is of fundamental importance in physical applications.

Differential forms. Totally antisymmetric tensor fields of type $(0, p)$ are particularly important in the study of differentiable manifolds; such objects are called (differential) p-forms, or forms of degree p.

A 1-form $\boldsymbol{\alpha}$ is a dual vector, $\boldsymbol{\alpha} = \alpha_a \boldsymbol{\theta}^a$. A 2-form $\boldsymbol{\beta}$ in the basis $\boldsymbol{\theta}^a \otimes \boldsymbol{\theta}^b$ can be written as

$$\boldsymbol{\beta} = \beta_{ab}\, \boldsymbol{\theta}^a \wedge \boldsymbol{\theta}^b \qquad \boldsymbol{\theta}^a \wedge \boldsymbol{\theta}^b \equiv \boldsymbol{\theta}^a \otimes \boldsymbol{\theta}^b - \boldsymbol{\theta}^b \otimes \boldsymbol{\theta}^a$$

where $\wedge$ denotes the exterior product (wedge product), an antisymmetrized tensor product. In a similar way we can represent an arbitrary p-form $\boldsymbol{\omega}$.

The exterior product of two forms, $\boldsymbol{\omega}_1 \wedge \boldsymbol{\omega}_2$, is associative, but in general not commutative.

In the space of smooth p-forms we can introduce the *exterior derivative* as a differential operator which maps a p-form $\boldsymbol{\alpha}$ into a $(p+1)$-form $\boldsymbol{d\alpha}$. It is defined by the following properties:

- $\boldsymbol{d}$ is linear;
- $\boldsymbol{d}(\boldsymbol{\alpha} \wedge \boldsymbol{\beta}) = \boldsymbol{d\alpha} \wedge \boldsymbol{\beta} + (-1)^p \boldsymbol{\alpha} \wedge \boldsymbol{d\beta}$;
- $\boldsymbol{d}^2 = 0$;
- if f is a 0-form (real function), $\boldsymbol{d}f$ is the ordinary differential of f.

A good feel for how this operator acts can be obtained by using local coordinates. A 1-form $\boldsymbol{d}f$ is defined by the relation $\boldsymbol{d}f(\boldsymbol{u}) = \boldsymbol{u}(f)$, which in local coordinates reads as $\boldsymbol{d}f(\boldsymbol{u}) = u^\mu \partial_\mu f$. If $f(x) = x^\mu$ (a projection), then $\boldsymbol{d}x^\mu(\boldsymbol{u}) = u^\mu$ implies $\boldsymbol{\theta}^\mu = \boldsymbol{d}x^\mu$. Therefore, $\boldsymbol{d}f = \boldsymbol{d}x^\mu\, \partial_\mu f$, i.e. $\boldsymbol{d}f$ is the differential of f.

The exterior derivative acts on a 1-form $\boldsymbol{\alpha} = \alpha_\nu\, \boldsymbol{d}x^\nu$ according to

$$\boldsymbol{d\alpha} = \boldsymbol{d}\alpha_\nu \wedge \boldsymbol{d}x^\nu = \partial_\mu \alpha_\nu\, \boldsymbol{d}x^\mu \wedge \boldsymbol{d}x^\nu.$$

In a similar way we can represent the action of $\boldsymbol{d}$ on an arbitrary form $\boldsymbol{\omega}$. In local coordinates, the property $\boldsymbol{d}^2 = 0$ follows from the fact that ordinary partial derivatives commute.

For an arbitrary basis $\boldsymbol{e}_a$ in T_P, we can define the commutation coefficients (structure functions) $c^a{}_{bc}$ by

$$[\boldsymbol{e}_b, \boldsymbol{e}_c] = c^a{}_{bc}\, \boldsymbol{e}_a. \tag{B.7a}$$

In the coordinate basis $c^a{}_{bc} = 0$. It is interesting to note that the coefficients $c^a{}_{bc}$ determine some properties of the dual basis $\boldsymbol{\theta}^a$. Since $\boldsymbol{d\theta}^a$ is a 2-form, it can be expressed in the basis $\boldsymbol{\theta}^a \wedge \boldsymbol{\theta}^b$, with the result

$$\boldsymbol{d\theta}^a = -\tfrac{1}{2} c^a{}_{bc} \boldsymbol{\theta}^b \wedge \boldsymbol{\theta}^c. \tag{B.7b}$$

Components of $\boldsymbol{d\theta}^a$, up to a factor, coincide with $c^a{}_{bc}$.

Parallel transport. Parallel transport is an extremely important concept not only in differential geometry, but also in theoretical physics (non-Abelian gauge theories, theory of gravity, etc).

In Euclidean spaces, the parallel transport of a vector $\boldsymbol{v}$ from P to P' is the vector at P', the components of which with respect to the basis $\boldsymbol{e}_a(P')$ are the same as the components of the original vector with respect to $\boldsymbol{e}_a(P)$, where the basis at P' is the parallel transported basis $\boldsymbol{e}_a(P)$. Thus, the parallel transport of a vector is defined by the parallel transport of the basis.

In differentiable manifolds, the idea of parallel transport is applied to tangent vectors. Consider a smooth vector field $\boldsymbol{v}(P)$ on X, which in local basis has the form $\boldsymbol{v}(P) = v^a(x)\boldsymbol{e}_a(x)$. Trying to compare tangent vectors at two neighbouring points $P(x)$ and $P'(x + \mathrm{d}x)$, we find the following two contributions:

(a) the components $v^a(x)$ are changed into $v^a(x + \mathrm{d}x)$; and
(b) the basis $\boldsymbol{e}_a(x)$ is changed by the parallel transport rule.

Since the change in $v^a(x)$ is easily calculable, the total change is determined by parallel transport of the basis. The idea of total change is realized by introducing the concept of a covariant derivative.

A *covariant derivative* of a vector on a smooth manifold X is a mapping $\boldsymbol{v} \mapsto \nabla\boldsymbol{v}$ of a smooth vector field into a differentiable tensor field of the type $(1, 1)$, which satisfies the following conditions:

- linearity: $\nabla(\boldsymbol{u} + \boldsymbol{v}) = \nabla\boldsymbol{u} + \nabla\boldsymbol{v}$; and
- the Leibniz rule: $\nabla(f\boldsymbol{v}) = \boldsymbol{d}f \otimes \boldsymbol{v} + f\nabla\boldsymbol{v}$, f is a real function.

The *coefficients of linear connection* $\Gamma^a{}_{bc}$ are defined by the change of basis,

$$\nabla\boldsymbol{e}_b = \Gamma^a{}_{bc}\boldsymbol{\theta}^c \otimes \boldsymbol{e}_a \tag{B.8}$$

so that

$$\begin{aligned}\nabla\boldsymbol{v} &= \nabla(v^a\boldsymbol{e}_a) = \boldsymbol{d}v^a \otimes \boldsymbol{e}_a + v^a\nabla\boldsymbol{e}_a \\ &= (\partial_b v^a + \Gamma^a{}_{cb}v^c)\,\boldsymbol{\theta}^b \otimes \boldsymbol{e}_a \equiv \nabla_b v^a\,\boldsymbol{\theta}^b \otimes \boldsymbol{e}_a.\end{aligned}$$

In terms of the *connection 1-form* $\boldsymbol{\omega}^a{}_c = \Gamma^a{}_{cb}\boldsymbol{\theta}^b$ we obtain

$$\nabla\boldsymbol{v} = (\boldsymbol{d}v^a + \boldsymbol{\omega}^a{}_b v^b) \otimes \boldsymbol{e}_a. \tag{B.9a}$$

The fact that $\nabla \boldsymbol{v}$ is a tensor can be used to find the transformation rules of the connection coefficients.

Demanding that the covariant derivative of a real function $\boldsymbol{w}^*(\boldsymbol{v})$ be a gradient, the covariant derivative of a 1-form $\boldsymbol{w}^*$ must be

$$\nabla \boldsymbol{w}^* = (\boldsymbol{d}w_b - \boldsymbol{\omega}^a{}_b w_a) \otimes \boldsymbol{\theta}^b. \tag{B.9b}$$

The covariant derivative is extended to tensors of arbitrary type by requiring the following properties:

- $\nabla f = \mathrm{d}f$, if f is a real function;
- linearity: $\nabla(\boldsymbol{t} + \boldsymbol{s}) = \nabla \boldsymbol{t} + \nabla \boldsymbol{s}$;
- Leibniz rule: $\nabla(\boldsymbol{t} \otimes \boldsymbol{s}) = \nabla \boldsymbol{t} \otimes \boldsymbol{s} + \boldsymbol{t} \otimes \nabla \boldsymbol{s}$; and
- commutativity with contraction.

The concept of parallel transport (covariant derivative) is *independent* of the existence of the metric.

The covariant derivative of $\boldsymbol{v}$ in the direction of $\boldsymbol{u}$ is a mapping $\boldsymbol{v} \mapsto \nabla_u \boldsymbol{v}$ of a vector field into a vector field, defined by

$$\nabla_u \boldsymbol{v} = (\nabla \boldsymbol{v})(\boldsymbol{u}, \boldsymbol{\theta}^a)\boldsymbol{e}_a = u^b \nabla_b v^a \boldsymbol{e}_a.$$

In particular, $\nabla_b \boldsymbol{e}_a = \Gamma^c{}_{ab} \boldsymbol{e}_c$. A vector $\boldsymbol{v}$ is said to be parallel along a curve $C(\lambda)$ if $\nabla_u \boldsymbol{v} = 0$, where $\boldsymbol{u}$ is the tangent vector on $C(\lambda)$. A curve $C(\lambda)$ is an affine geodesic (autoparallel) if its tangent vector is parallel to itself, $\nabla_u \boldsymbol{u} = 0$.

Introducing *spinors* on a manifold is a more complex problem than introducing tensors. Just as tensors are defined by the transformation law under general coordinate transformations, so we can define (world) spinors by considering nonlinear representations of the group of general coordinate transformations. Spinors introduced in this way are infinite dimensional objects (Ne'eman and Šijački 1985, 1987). Finite spinors can be defined in a much simpler way. Consider a collection of all frames $\mathcal{E}_P = \{\boldsymbol{e}_a\}$ in the tangent space T_P. Each frame can be obtained from some fixed frame by a suitable transformation of the type $GL(n, R)$. If we introduce the Lorentz metric in T_P, we can restrict ourselves to orthonormal frames, so that the related symmetry group becomes $SO(1, n-1)$. Since, now, each tangent space has a Minkowskian structure, we can introduce finite spinors in the standard manner. Spinors can be parallel transported if we define the parallel transport of orthonormal frames, which leads to the concept of *spin connection*.

Torsion and curvature. The torsion and curvature operators are defined by

$$\begin{aligned} T(\boldsymbol{u}, \boldsymbol{v}) &= \nabla_u \boldsymbol{v} - \nabla_v \boldsymbol{u} - [\boldsymbol{u}, \boldsymbol{v}] \\ R(\boldsymbol{u}, \boldsymbol{v}) &= \nabla_u \nabla_v - \nabla_v \nabla_u - \nabla_{[u,v]}. \end{aligned} \tag{B.10}$$

Their components with respect to an arbitrary frame are

$$
\begin{aligned}
T(\boldsymbol{e}_a, \boldsymbol{e}_b) &= (\Gamma^c{}_{ba} - \Gamma^c{}_{ab} - c^c{}_{ab})\boldsymbol{e}_c \equiv T^c{}_{ab}\boldsymbol{e}_c \\
R(\boldsymbol{e}_a, \boldsymbol{e}_b)\boldsymbol{e}_c &= [\partial_a\Gamma^e{}_{cb} + \Gamma^e{}_{da}\Gamma^d{}_{cb} - \partial_b\Gamma^e{}_{ca} + \Gamma^e{}_{db}\Gamma^d{}_{ca} - c^d{}_{ab}\Gamma^e_{cd}]\boldsymbol{e}_e \\
&\equiv R^e{}_{cab}\boldsymbol{e}_e.
\end{aligned}
$$

Due to the antisymmetry of $T^a{}_{bc}$ and $R^a{}_{bcd}$ in the last two indices, it is possible to define the torsion and curvature 2-forms:

$$
\mathcal{T}^a \equiv \tfrac{1}{2}T^a{}_{bc}\boldsymbol{\theta}^b \wedge \boldsymbol{\theta}^c \qquad \mathcal{R}^a{}_b \equiv \tfrac{1}{2}R^a{}_{bcd}\boldsymbol{\theta}^c \wedge \boldsymbol{\theta}^d. \tag{B.11}
$$

These forms obey the *Cartan structure equations*:

$$
\begin{aligned}
&\text{(first)} \qquad \mathcal{T}^a = \boldsymbol{d\theta}^a + \boldsymbol{\omega}^a{}_b \wedge \boldsymbol{\theta}^b \\
&\text{(second)} \qquad \mathcal{R}^a{}_b = \boldsymbol{d\omega}^a{}_b + \boldsymbol{\omega}^a{}_c \wedge \boldsymbol{\omega}^c{}_b.
\end{aligned}
$$

Application of the exterior derivative to the analysis of torsion and curvature requires a minor extension of this concept. Let us define an *extended* exterior derivative $\bar{\boldsymbol{d}}$, which acts on a form as $\boldsymbol{d}$, and on a vector as

$$
\bar{\boldsymbol{d}}\boldsymbol{v} = \nabla\boldsymbol{v} \qquad \nabla = \text{the covariant derivative.} \tag{B.12}
$$

Consider, now, a vector valued 1-form $\boldsymbol{w} = \boldsymbol{\theta}^a\boldsymbol{e}_a (\equiv \boldsymbol{\theta}^a \otimes \boldsymbol{e}_a)$. The action of $\bar{\boldsymbol{d}}$ on $\boldsymbol{w}$ yields the torsion:

$$
\bar{\boldsymbol{d}}\boldsymbol{w} = \boldsymbol{d\theta}^a\boldsymbol{e}_a - \boldsymbol{\theta}^a\nabla\boldsymbol{e}_a = (\boldsymbol{d\theta}^a + \boldsymbol{\omega}^a{}_b \wedge \boldsymbol{\theta}^b)\boldsymbol{e}_a \equiv \mathcal{T}^a\boldsymbol{e}_a. \tag{B.13a}
$$

Similarly, the action of $\bar{\boldsymbol{d}}$ on the relation $\bar{\boldsymbol{d}}\boldsymbol{e}_a \equiv \nabla\boldsymbol{e}_a = \boldsymbol{\omega}^c{}_a\boldsymbol{e}_c$ produces the curvature form:

$$
\bar{\boldsymbol{d}}^2\boldsymbol{e}_a = \boldsymbol{d\omega}^c{}_a\boldsymbol{e}_c - \boldsymbol{\omega}^c{}_a\nabla\boldsymbol{e}_c = (\boldsymbol{d\omega}^b{}_a + \boldsymbol{\omega}^b{}_c \wedge \boldsymbol{\omega}^c{}_a)\boldsymbol{e}_b = \mathcal{R}^b{}_a\boldsymbol{e}_b. \tag{B.13b}
$$

Note that $\bar{\boldsymbol{d}}^2 \neq 0$. The advantage of using $\bar{\boldsymbol{d}}$ lies in the fact that many terms in the calculation of torsion and curvature automatically cancel.

Differentiation of the Cartan structure equations gives the following *Bianchi identities*:

$$
\begin{aligned}
&\text{(first)} \qquad \boldsymbol{d}\mathcal{T}^a + \boldsymbol{\omega}^a{}_b \wedge \mathcal{T}^b = \mathcal{R}^a{}_b \wedge \boldsymbol{\theta}^b \\
&\text{(second)} \qquad \boldsymbol{d}\mathcal{R}^a{}_b + \boldsymbol{\omega}^a{}_c \wedge \mathcal{R}^c{}_b - \boldsymbol{\omega}^c{}_b \wedge \mathcal{R}^a{}_c = 0.
\end{aligned}
$$

We stress again that the metric and the connection are completely independent geometric objects. Imposing certain conditions on these objects leads to special types of differentiable manifolds. Thus, the condition $\nabla\boldsymbol{g} = 0$ defines the Riemann–Cartan space, while the additional requirement $\mathcal{T}^a = 0$ yields the Riemann space (alternatively, $\mathcal{R}^a{}_b = 0$ leads to T_4).

Exercises

1. Show that the following properties define topological, but not Hausdorff spaces:
   (a) $X = \{x, y\}$, and the open sets are $\emptyset$, X, and $\{x\}$;
   (b) $X = [0, 1]$, and the open sets are the empty set, and all sets obtained by removing at most a countable number of points from $[0, 1]$.
2. Show that a circle $S_1 = \{(x, y)|x^2 + y^2 = 1\}$ is a differentiable manifold.
3. Find the components of the tangent vector to curve $C(\lambda)$ in the coordinate basis.
4. Let $\boldsymbol{v}$ be the tangent vector at the point $P \in X$ to a curve $C(\lambda)$, and f a differentiable mapping $X \to Y$. A derivative of f is a linear map f' of tangent spaces $T_P \to T_{f(P)}$, given as $\boldsymbol{v} \mapsto \boldsymbol{u} : \boldsymbol{u}(h) = \boldsymbol{v}(h \circ f)$, where $h : Y \to \mathcal{R}$ is a differentiable function. Show that $\boldsymbol{u}$ is the tangent vector to the curve $f(C(\lambda))$, and find its components in the coordinate basis.
5. Consider a three-dimensional Euclidean space with spherical local coordinates r, θ, ϕ. Define the orthonormal basis of tangent vectors $\{\boldsymbol{e}_r, \boldsymbol{e}_\theta, \boldsymbol{e}_\phi\}$, and find the related basis of dual vectors.
6. Prove the relation (B.7*b*) by using $\boldsymbol{\theta}^a = e^a{}_\nu \boldsymbol{d}x^\nu$, or otherwise. Then, use $\boldsymbol{d}^2\boldsymbol{\theta}^a = 0$ to derive the Jacobi identity if $c^a{}_{bc}$ are constants.
7. Find the component of the torsion and curvature operators in an arbitrary basis.
8. In a Riemann space calculate the antisymmetric part of the connection coefficients $\Gamma^a{}_{bc}$ in an arbitrary basis.
9. Consider the upper half-plane as a Riemann space with the (Poincaré) metric:

$$\mathrm{d}s^2 = \frac{\mathrm{d}x^2 + \mathrm{d}y^2}{y^2} \qquad y > 0.$$

   (a) Find an orthonormal basis of 1-forms $\boldsymbol{\theta}^a$ and calculate the connection 1-form $\boldsymbol{\omega}^a{}_a$.
   (b) Calculate the curvature 2-form $\mathcal{R}^a{}_b$, and find the scalar curvature R.
   (c) Calculate R in the coordinate basis, in terms of Christoffel symbols.
10. Write the Bianchi identities in the coordinate basis.

Appendix C

De Sitter gauge theory

Although PGT leads to a satisfactory classical theory of gravity, the analogy with gauge theories of internal symmetries is not perfect, because of the specific treatment of translations. It is possible, however, to formulate gauge theory of gravity in a way that treats the whole Poincaré group in a more unified way. The approach is based on the de Sitter group and the Lorentz and translation parts are distinguished through the mechanism of spontaneous symmetry breaking (Townsend 1977, Mac Dowell and Mansouri 1977, Stelle and West 1980, Kibble and Stelle 1986).

The de Sitter group has the interesting property that in a special limit, when the parameter a of the group tends to infinity, it reduces to the Poincaré group. The parameter a represents the radius of the mathematical de Sitter space. Since for large a the structures of two groups are very 'close' to each other, we are motivated to study the de Sitter group as an alternative for the description of the spacetime.

The de Sitter theory of gravity can be formulated by analogy with PGT: the spacetime is assumed to have the de Sitter structure, matter fields are described by an action which is invariant under the global de Sitter symmetry and gravity is introduced as a gauge field in the process of localization of this symmetry. Observationally, a would appear to be of the order of the radius of the universe, so that the Poincaré symmetry is a good symmetry at all but cosmological scales.

There is, however, another, more interesting possibility:

(a) the spacetime continues to have a Minkowskian structure; and
(b) at each point of the spacetime we have the de Sitter group acting on the matter fields as the internal gauge symmetry group.

In this approach there is an interesting connection to PGT and the parameter a might be very small—of the order of the Planck length.

The de Sitter group and its contraction. In order to study the structure of the de Sitter group, we consider a flat, five-dimensional space M_5 with metric

$\eta_{ab} = (+, -, -, -, +)$ $(a, b = 0, 1, 2, 3, 5)$,

$$ds^2 = (dy^0)^2 - (dy^1)^2 - (dy^2)^2 - (dy^3)^2 + (dy^5)^2 \equiv \eta_{ij}\, dy^i\, dy^j + (dy^5)^2. \quad \text{(C.1)}$$

A hypersphere H_4 of 'radius' a embedded in M_5,

$$\eta_{ij} y^i y^j + (y^5)^2 = a^2 \quad \text{(C.2)}$$

is the maximally symmetric subspace of M_5 (the de Sitter space). On H_4, the quantity $(dy^5)^2$ takes the form $(dy^5)^2 = (\eta_{ij} y^i\, dy^j)^2/(y^5)^2$, so that the interval becomes

$$ds^2 = \eta_{ij}\, dy^i\, dy^j + \frac{(\eta_{ij} y^i\, dy^j)^2}{a^2 - \eta_{mn} y^m y^n}. \quad \text{(C.3)}$$

This expression defines the metric on H_4 in coordinates y^i $(i = 0, 1, 2, 3)$.

The curvature of any maximally symmetric space is the same at each point. In the vicinity of $y^i = 0$, the metric and the (Riemannian) connection of H_4 have the form

$$g_{ij} = \eta_{ij} + \frac{y_i y_j}{a^2} \qquad \Gamma^i_{jk} = \frac{1}{a^2} y^i \eta_{jk}$$

so that

$$(R_{ijkl})_0 = \frac{1}{a^2}(\eta_{ik}\eta_{jl} - \eta_{il}\eta_{jk}).$$

Therefore, the H_4 space has curvature $R = 12/a^2$.

Instead of y^i, we can introduce the pseudo-spherical coordinates $(t, \rho, \theta, \varphi)$ which cover the whole space and lead to the metric

$$ds^2 = \cosh^2\rho\, dt^2 - d\rho^2 - \sinh^2\rho\,(d\theta^2 + \sin^2\theta\, d\varphi^2).$$

In the conformal coordinates x^i,

$$y^i = \Phi(x^2)x^i \qquad \Phi(x^2) \equiv (1 + x^2/4a^2)^{-1}$$

where $x^2 = \eta_{ij} x^i x^j$, the metric takes the form

$$ds^2 = \Phi^2 \eta_{ij}\, dx^i\, dx^j.$$

The isometry group of the hypersphere H_4, $SO(2,3)$, is called the de Sitter group†. Infinitesimal de Sitter transformations of coordinates are pseudo-rotations in M_5:

$$\delta y^a = \omega^a{}_b y^b \qquad \omega^{ab} = -\omega^{ba}. \quad \text{(C.4)}$$

The generators in the space of scalar fields have the form

$$M_{ab} = y_a \partial_b - y_b \partial_a$$

† In the literature, $SO(2,3)$ is usually called the anti de Sitter group, while the de Sitter group is $SO(1,4)$; we omit that distinction here.

and satisfy the Lie algebra:

$$[M_{ab}, M_{cd}] = \eta_{bc}M_{ad} - \eta_{ac}M_{bd} - \eta_{bd}M_{ac} + \eta_{ad}M_{bc}. \tag{C.5}$$

In the general case of arbitrary fields, the generators contain an additional 'spin' part (Gürsey 1964).

Conformal coordinates lead to a very simple form of the metric, but from the point of view of symmetries, the following implicitly defined coordinates u^i are more interesting:

$$y^i = a\frac{u^i}{u}\sin(u/a) \qquad y^5 = a\cos(u/a) \tag{C.6}$$

where $u = (u^2)^{1/2}$ and $u^2 = \eta_{ij}u^i u^j$. De Sitter transformations of y^a in M_5 induce complicated, nonlinear transformations of u^i in H_4. The infinitesimal form of these transformations can be found by observing that

$$\delta u^i = \bar{\omega}^i{}_j u^j + \varepsilon^i u \qquad \bar{\omega}^{ij} = -\bar{\omega}^{ji} \tag{C.7}$$

implies

$$\delta_\omega y^i = \bar{\omega}^i{}_j y^j \qquad \delta_\omega y^5 = 0$$

$$\delta_\varepsilon y^i = \frac{1}{u}(\varepsilon^i u_j - \varepsilon_j u^i)y^j + \frac{u\cdot\varepsilon}{ua}u^i y^5 \qquad \delta_\varepsilon y^5 = -\frac{u\cdot\varepsilon}{ua}u_i y^i.$$

Therefore, $\delta_\omega u^i$ and $\delta_\varepsilon u^i$ realize de Sitter transformations with parameters $[\omega^{ij} = \bar{\omega}^{ij}, \omega^{i5} = 0]$ and $[\omega^{ij} = (\varepsilon^i u^j - \varepsilon^j u^i)/u, \omega^{i5} = u^i(u\cdot\varepsilon)/ua]$, respectively. The quantities

$$M_{ij}(u) = u_i\partial_j - u_j\partial_i \qquad M_{i5}(u) = u\partial_i$$

are de Sitter generators in the coordinates u^i, that satisfy (C.5).

If we now introduce

$$P_i = \frac{1}{a}M_{i5}$$

the de Sitter algebra takes the form

$$\begin{aligned} [M_{mn}, M_{lr}] &= \eta_{nl}M_{mr} - \eta_{ml}M_{nr} - \eta_{nr}M_{ml} + \eta_{mr}M_{nl} \\ [M_{mn}, P_l] &= \eta_{nl}P_m - \eta_{ml}P_n \\ [P_m, P_n] &= -\frac{1}{a^2}M_{mn}. \end{aligned} \tag{C.8}$$

In the limit $a \to \infty$ this algebra is transformed into the Poincaré form, by the process called *contraction* (Inönü 1964).

If we identify the spacetime with the de Sitter space H_4, it is clear that the parameter a must be large in order for the deviation from Poincaré symmetry to be sufficiently small. Instead of that, we shall consider the alternative possibility in which the spacetime retains the basic structure of M_4.

Localization of de Sitter symmetry. We assume that the spacetime has the structure of Minkowski space M_4, and 'attach' at each point of M_4 a 'tangent' space F_x (a fibre), representing a copy of the de Sitter space H_4. The de Sitter group $SO(2,3)$ acts on the matter fields in F_x as a group of *internal* symmetries.

We also assume that the initial theory of matter fields is invariant under global $SO(2,3)$ transformations. The symmetry is localized by introducing the covariant derivative $\nabla_\mu\phi = (\partial_\mu + \frac{1}{2}A^{ab}{}_\mu\Sigma_{ab})\phi$, the transformation law of which defines the transformation properties of the gauge fields:

$$\delta_0 A^{ab}{}_\mu = -\nabla_\mu\omega^{ab} - \xi^\lambda{}_{,\mu}A^{ab}{}_\lambda - \xi^\lambda\partial_\lambda A^{ab}{}_\mu. \tag{C.9}$$

Here, $\xi^i \equiv \delta u^i = \omega^i{}_j u^j + \varepsilon^i u$ as in (C.7), and $\xi^\lambda = \delta^\lambda_l\xi^l$. The commutator of the covariant derivatives determines the field strength:

$$F^{ab}{}_{\mu\nu} = \partial_\mu A^{ab}{}_\nu - \partial_\nu A^{ab}{}_\mu + A^a{}_{c\mu}A^{cb}{}_\nu - A^a{}_{c\nu}A^{cb}{}_\mu. \tag{C.10}$$

Internal indices $(a, b, \ldots)$ and spacetime indices $(\mu, \nu, \ldots)$ are at this level completely *unrelated*, as there is no quantity analogous to the tetrad that could connect them.

Introducing the notation

$$P_i = \frac{1}{a}M_{i5} \qquad \lambda^i = a\omega^{i5} \qquad B^i{}_\mu = aA^{i5}{}_\mu \tag{C.11}$$

equations (C.9) and (C.10) become

$$\begin{aligned}\delta_0 A^{ij}{}_\mu &= \delta_0^{\mathrm{P}} A^{ij}{}_\mu - \frac{1}{a^2}(\lambda^i B^j{}_\mu - \lambda^j B^i{}_\mu)\\ \delta_0 B^i{}_\mu &= \delta_0^{\mathrm{P}} B^i{}_\mu - \lambda_s A^{is}{}_\mu - \lambda^i{}_{,\mu}\end{aligned} \tag{C.12}$$

$$\begin{aligned}F^{ij}{}_{\mu\nu} &= R^{ij}{}_{\mu\nu} - \frac{1}{a^2}(B^i{}_\mu B^j{}_\nu - B^i{}_\nu B^j{}_\mu)\\ F^{i5}{}_{\mu\nu} &= \frac{1}{a}T^i{}_{\mu\nu} = \frac{1}{a}(\nabla_\mu B^i{}_\nu - \nabla_\nu B^i{}_\mu).\end{aligned} \tag{C.13}$$

Here, δ_0^{P} denotes the Poincaré transformation, $\nabla_\mu B^i{}_\nu = \partial_\mu B^i{}_\nu + A^i{}_{s\mu}B^s{}_\nu$, and $R^{ij}{}_{\mu\nu}$, $T^i{}_{\mu\nu}$ are the curvature and torsion of PGT, respectively. The form of the algebra (C.8) suggests the identification of $B^i{}_\mu$ with the tetrad field:

$$B^i{}_\mu = b^i{}_\mu \qquad ? \tag{C.14}$$

However, the transformation law of $B^i{}_\mu$ shows that such an identification is correct only if $\lambda^i = a\omega^{i5} = 0$, i.e. when the de Sitter symmetry is *broken*.

In constructing an $SO(2,3)$ invariant action we may use, in addition to the field strength F, the completely antisymmetric tensor ε_{abcde}, the tensor density $\varepsilon_{\mu\nu\lambda\rho}$ and the metrics η_{ab} and $\eta_{\mu\nu}$. With these ingredients we cannot construct an action linear in F. The only quadratic action is $\int d^4x\, \varepsilon^{\mu\nu\lambda\rho} F^{ab}{}_{\mu\nu} F_{ab\lambda\rho}$, but it is trivial as a topological invariant. The action $\int d^4x\, \sqrt{g} g^{\mu\lambda} g^{\nu\rho} F^{ab}{}_{\mu\nu} F_{ab\lambda\rho}$, where $g_{\mu\nu}$ is defined by (C.5), is not a polynomial and, moreover, condition (C.5) breaks the de Sitter symmetry.

A simpler solution can be found by using the mechanism of spontaneous symmetry breaking. Let us choose the Lagrangian

$$\mathcal{L} = \frac{f}{a} \varepsilon^{\mu\nu\lambda\rho} F^{ab}{}_{\mu\nu} F^{cd}{}_{\lambda\rho} \varepsilon_{abcde} \phi^e - \lambda(\phi^e \phi_e - a^2) \tag{C.15}$$

where f is a dimensional constant, ϕ^e is an auxiliary field, and λ is a multiplier, imposing the constraint $\phi^e \phi_e = a^2$ as the equation of motion. Now, we can choose the solution

$$\phi^e = (0, 0, 0, 0, a) \tag{C.16}$$

so that the effective Lagrangian becomes

$$\mathcal{L} = f \varepsilon^{\mu\nu\lambda\rho} F^{ij}{}_{\mu\nu} F^{kl}{}_{\lambda\rho} \varepsilon_{ijkl5}. \tag{C.17}$$

The solution (C.16) corresponds to *spontaneous* breaking of the local de Sitter symmetry down to the Lorentz subgroup $SO(1, 3)$, whereupon the identification (C.14) becomes *correct*.

Now, we shall find the connection of this theory with PGT. After using (C.13) and (C.14) the Lagrangian takes the form

$$f^{-1}\mathcal{L} = \mathcal{L}_2 - \frac{4}{a^2} \varepsilon^{\mu\nu\lambda\rho} R^{ij}{}_{\mu\nu} b^k{}_\lambda b^l{}_\rho \varepsilon_{ijkl} + \frac{4}{a^4} \varepsilon^{\mu\nu\lambda\rho} b^i{}_\mu b^j{}_\nu b^k{}_\lambda b^l{}_\rho \varepsilon_{ijkl}$$

where $\varepsilon_{ijkl} \equiv \varepsilon_{ijkl5}$. The term $\mathcal{L}_2$ is quadratic in $R^{ij}{}_{\mu\nu}$, and represents the Gauss–Bonnet topological invariant. It gives no contribution to the equations of motion and can be discarded, at least classically. Using the identity $\varepsilon^{\mu\nu\lambda\rho} \varepsilon_{ijkl} b^k{}_\lambda b^l{}_\rho = -2b(h_i{}^\mu h_j{}^\nu - h_j{}^\mu h_i{}^\nu)$, we finally obtain

$$\mathcal{L} = \frac{16f}{a^2} b(R + \Lambda) \tag{C.18}$$

where $\Lambda = 6/a^2$. Thus, the Lagrangian (C.17), quadratic in F, leads to the Einstein–Cartan theory with a cosmological constant.

We should observe that a^2/f is proportional to the gravitational constant G, so that $\dim(f) = \text{energy} \times \text{time}$. If we identify f with the Planck constant $\hbar$, the de Sitter theory of gravity leads to a natural introduction of the dimensional constant a, which has the value of the Planck length (10^{-33} cm). This scale (together with $\hbar$) sets the scale for the gravitational constant G and the small-scale structure of spacetime is determined by the de Sitter group.

This interpretation leads to a large cosmological constant, which is unacceptable in classical theory. The suggestion that this term might be cancelled by quantum corrections currently remains only a hope. Certainly, we can assume that a is very large, since in the limit $a \to \infty$ the cosmological constant vanishes, and the only term that survives is bR. However, in this case the small-scale structure of the spacetime remains untouched.

Although some aspects of de Sitter gauge theory are not satisfying, this approach yields an inspired insight into the structure of gravity.

Exercises

1. Derive the following relations for Riemann space with the metric (C.3):

$$\Gamma^i_{jk} = \frac{1}{a^2} y^i g_{jk} \qquad R_{ijkl} = \frac{1}{a^2}(g_{ik}g_{jl} - g_{il}g_{jk}).$$

2. Show that the conformal coordinates of the de Sitter space satisfy the following relations:

$$x^i = \frac{2y^i}{1 + y^5/a} \qquad \frac{x^2}{4a^2} = \frac{1 - y^5/a}{1 + y^5/a} \qquad y^5 = (2\Phi - 1)a.$$

3. Show that the contraction of the Lorentz group defines the Galilei group, i.e. the group containing Galilean transformations to a moving reference frame, and spatial rotations.
4. Show that the action $\int d^4x\, \varepsilon^{\mu\nu\lambda\rho} F^{ab}{}_{\mu\nu} F_{ab\lambda\rho}$ can be written in the form of a surface term $\int d^4x\, \partial_\mu K^\mu$.
5. Show that the action $\int d^4x\, \varepsilon^{\mu\nu\lambda\rho} \varepsilon_{ijkl} R^{ij}{}_{\mu\nu} R^{kl}{}_{\lambda\rho}$ is proportional to the Gauss–Bonnet topological invariant, and verify that its variation vanishes identically.
6. Express the non-polynomial action $\int d^4x\, \sqrt{-g} g^{\mu\lambda} g^{\nu\rho} F^{ab}{}_{\mu\nu} F_{ab\lambda\rho}$ in terms of the curvature and the torsion, using (C.14) to identify the tetrad field.

Appendix D

The scalar–tensor theory

Various attempts to modify certain properties of GR led to alternative formulations of gravitational theory, in which new principles appear together with new dynamical variables. Starting from the conviction that Mach's ideas about inertia have found only a limited expression in GR, Brans and Dicke (BD) proposed an alternative theory of gravity in which the metric tensor is accompanied by a scalar field as a new dynamical variable (Brans and Dicke 1961).

The Brans–Dicke theory. According to Mach's ideas, inertial forces that are observed in an accelerated reference frame can be interpreted as a gravitational field, with its origin in distant matter of the universe, accelerated relative to that frame. In GR, the influence of *matter* in defining local inertial frames is sometimes negligible compared to the influence of the *boundary conditions*. Consider a space containing nothing but a single observer in his/her laboratory, which is of standard size and mass. If we use GR and the boundary conditions according to which the space is asymptotically Minkowskian, the effect of the laboratory on the metric is minor and can be calculated in the weak field approximation. The laboratory is practically an inertial reference frame. However, if the observer fires a bullet through an open window tangentially to the wall, the laboratory is set into the state of rotation, which can be registered with the help of the gyroscope. After some time, the bullet, which is almost massless and at great distance from the laboratory, becomes dominant in determining the local inertial frame (orientation of the gyroscope). This situation in GR is much closer to a description of an absolute space in the sense of Newton than a physical space as interpreted by Mach. According to Mach, the influence of the massive, nearby laboratory should be dominant.

The influence of all masses in the universe on the local gravitational field can be simply described by introducing a scalar field ϕ. If ϕ obeys the Poisson equation, its average value can be estimated by calculating the central gravitational potential of a homogeneous sphere, with a radius which is equal to the dimension of the visible universe, $R \sim 10^{28}$ cm, and which contains mass M distributed with the cosmological density $\rho \sim M/R^3 \sim 10^{-29}$ g cm^{-3}. In this way, we obtain

$$\langle\phi\rangle \sim \rho R^2 \sim 10^{27} \text{ g cm}^{-1}.$$

Recalling the value of the gravitational constant, $G = 0.68 \times 10^{-28}$ g^{-1} cm (in units $c = 1$), we find an interesting relation:

$$\langle\phi\rangle \sim \frac{1}{G}. \tag{D.1}$$

It connects the average value of ϕ, representing the influence of the cosmological mass distribution, with the gravitational constant, which defines the locally observed gravitational field.

According to Mach's interpretation of inertia, the local gravitational field (which defines the local inertial reference frame or local standard of inertia) should depend on the distribution of all masses in the universe. If the influence of these masses is realized through a scalar field ϕ, then Mach's idea means that the gravitational 'constant' is a function of ϕ. Relation (D.1) suggests that a true Machian theory of gravity can be obtained from GR by replacing $1/G \to \phi$, and adding the usual kinetic term for ϕ:

$$I_{\mathrm{BD}} = \int \mathrm{d}^4x\, \sqrt{-g}[-\phi R + (\omega/\phi) g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi + 16\pi\mathcal{L}_{\mathrm{M}}]. \tag{D.2}$$

Here, ϕ plays a role analogous to that of G^{-1}, and ω is a dimensionless constant. This constant has to be *positive* in order for the energy of the scalar field to be positive. The matter Lagrangian $\mathcal{L}_{\mathrm{M}}$ does not depend on ϕ and has the same form as in GR, therefore the equations of motion of matter also have the same form as in GR. The gravitational field in (D.2) is described in part by geometry (metric) and in part by a scalar field.

It is interesting to observe that this theory satisfies the 'medium-strong' principle of equivalence: it predicts, like GR, that the laws of motion of matter fields (or particles) in every local inertial frame are the same as in SR; in particular, all dimensionless constants are the same at every point of spacetime. On the other hand, since the gravitational 'constant' varies from point to point, the gravitational effects (e.g. the ratio of electromagnetic and gravitational force between two electrons) also vary from point to point. Thus, the 'very strong' form of the principle of equivalence is not compatible with BD theory.

The equations of motion are obtained from (D.2) in the usual way. Using the definition of the energy–momentum tensor for the scalar field,

$$\tfrac{1}{2}B_{\mu\nu} = (\nabla_\mu\nabla_\nu - g_{\mu\nu}\Box)\phi + (\omega/\phi)[\partial_\mu\phi\partial_\nu\phi - \tfrac{1}{2}g_{\mu\nu}(\partial_\lambda\phi\partial^\lambda\phi)] \tag{D.3}$$

we find the following equations of motion for $g_{\mu\nu}$ and ϕ:

$$\phi G_{\mu\nu} = 8\pi T_{\mu\nu} + \tfrac{1}{2}B_{\mu\nu} \tag{D.4a}$$

$$-2(\omega/\phi)\Box\phi + (\omega/\phi^2)\partial_\lambda\phi\partial^\lambda\phi - R = 0. \tag{D.4b}$$

The first equation is an analogue of Einstein's field equation but with a variable gravitational coupling ϕ^{-1} and the source of gravity is the energy–momentum tensor of matter and scalar field. When $T_{\mu\nu}$ dominates $B_{\mu\nu}$, this equations differs from Einstein's only by the presence of a variable gravitational 'constant'. The covariant divergence of this equation has the form

$$\nabla^\mu T_{\mu\nu} = 0$$

which is the same as in GR (it implies, in particular, that point particles move along geodesic lines). The second equation can be transformed in such a way that the source of the field ϕ is given by the trace of the matter energy–momentum, in agreement with Mach's principle. Indeed, the contraction of equation (D.4*a*), $-\phi R = 8\pi T - 3\Box\phi - (\omega/\phi)\partial_\lambda\phi\partial^\lambda\phi$, combined with (D.4*b*), leads to

$$(2\omega + 3)\Box\phi = 8\pi T. \tag{D.5}$$

The observational consequences of the theory can be found by solving equations (D.4*a*) and (D.5). The results obtained here differ from those in GR, and can be used to test the validity of the theory, and find the allowed values of ω. For large ω, equation (D.5) has a solution

$$\phi = \frac{1}{G} + \mathcal{O}(1/\omega).$$

Using this solution in equation (D.4*a*) we obtain

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} + \mathcal{O}(1/\omega).$$

Thus, in the limit of large ω, the theory carries over to GR.

In order to set the lower limit on ω it is useful to calculate the perihelion precession, which leads to the following result:

$$\frac{3\omega + 4}{3\omega + 6} \times \text{(the value of GR)}.$$

Since the value computed from GR agrees with observations with an accuracy of 8%, this equation implies $(3\omega + 4)/(3\omega + 6) \gtrsim 0.92$, i.e.

$$\omega \gtrsim 6.$$

A detailed comparison of the predictions of BD theory with observations can be found in standard textbooks (see, e.g., Weinberg 1972).

Connection with Weyl theory. It is interesting to clarify the connection between BD theory and Weyl theory in Riemannian space V_4. The scalar field ϕ in (D.2) has weight $w = -2$. It is convenient to define a new scalar field, $\varphi^2 = \phi$, having weight $w = -1$. Then the action (D.2) takes the form

$$I_{\mathrm{BD}} = \int \mathrm{d}^4x\, \sqrt{-g}\, (-\varphi^2 R + 4\omega g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi + 16\pi\mathcal{L}_{\mathrm{M}}). \tag{D.6}$$

In the absence of matter, the action is invariant under local rescalings

$$g_{\mu\nu} \to \mathrm{e}^{2\lambda} g_{\mu\nu} \qquad \varphi \to \mathrm{e}^{-\lambda}\varphi$$

provided the parameter ω has the value

$$\omega = -\tfrac{3}{2} \tag{D.7}$$

as follows from equation (4.46). However, for negative ω the energy of the scalar field becomes negative. Weyl invariance fixes the relative sign of the first two terms, while their overall sign can be fixed by demanding that the energy of the scalar field be positive. This leads to the following modified action as a model for gravitation:

$$I = \int \mathrm{d}^4x\, \sqrt{-g}(\varphi^2 R + 6g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi + 16\pi\mathcal{L}_{\mathrm{M}}). \tag{D.8}$$

(Alternatively, the condition of the positive gravitational energy would change the sign of the first two terms.)

The equations of motion following from (D.8) include the matter field equations together with

$$\begin{aligned} (\Box - \tfrac{1}{6}R)\varphi &= 0 \\ \varphi^2 G_{\mu\nu} + 6\theta_{\mu\nu} + 8\pi T_{\mu\nu} &= 0 \end{aligned} \tag{D.9}$$

where $\theta_{\mu\nu}$ is the improved energy–momentum tensor of the field φ:

$$\theta_{\mu\nu} = \partial_\mu\varphi\partial_\nu\varphi - g_{\mu\nu}\tfrac{1}{2}(\partial\varphi)^2 - \tfrac{1}{6}(\nabla_\mu\nabla_\nu - g_{\mu\nu}\Box)\varphi^2.$$

Taking the trace of the second equation we find

$$6\varphi(\Box - \tfrac{1}{6}R)\varphi + 8\pi T = 0$$

whereupon the first equation implies that the trace of the matter energy–momentum has to vanish. This means that only scale invariant matter can be present in (D.8) (which is not the case with BD theory, where the matter field may be massive).

To allow for the presence of massive matter, Deser (1970) suggested explicitly breaking the scale invariance by adding a mass term for the scalar field. Then, instead of having the vanishing trace of the matter energy–momentum, we get a consistent relation

$$12m^2\varphi^2 + 8\pi T = 0.$$

Another possibility is to withhold scale invariance from the complete theory, retreating thereby from BD's original idea. The algebraic dependence of the equations of motion for $g_{\mu\nu}$ and φ is a consequence of Weyl invariance. In this case, the scalar field is not a true dynamical variable and can be entirely removed from the theory. To show this, consider action (D.8) without the matter component. Applying rescaling transformation with parameter $\lambda = \ln \varphi$, the field φ goes over into $\bar{\varphi} = \mathrm{e}^{-\lambda}\varphi = 1$, so that the complete theory reduces to GR in the space with metric $\bar{g}_{\mu\nu} = \varphi^2 g_{\mu\nu}$,

$$\int \mathrm{d}^4x\, \sqrt{-g}(\varphi^2 R + 6g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi) = \int \mathrm{d}^4x\, \sqrt{-\bar{g}}R(\bar{g}).$$

Thus, the scalar field completely disappears as an independent degree of freedom. The presence of Weyl invariant matter does not change this conclusion (Anderson 1970). Of course, when the Lagrangian $\mathcal{L}_\mathrm{M}$ is not Weyl invariant the situation changes and the equivalence with GR is lost.

If the scalar field is not determined dynamically, as in Weyl invariant theories, we could consider the possibility of fixing its value by experimental conditions. There have been attempts to formulate a theoretical interpretation of the hypothesis of large numbers by making a distinction between the 'cosmological' and 'atomic' scales (Canuto *et al* 1976).

We see that the role of the scalar field in the modified BD theory with $\omega = -\frac{3}{2}$ is essentially different from that in the original BD theory.

Discussion. If we want to investigate matter fields with non-vanishing spin, it is natural to generalize BD theory (D.2) to the case of Riemann–Cartan space U_4 (Kim 1986). An interesting conclusion from this model is that the torsion may be generated even by spinless matter—the source of the torsion is the gradient of the scalar field.

The basic characteristics of the theory can be clearly seen by studying the following simple action:

$$I = -\int \mathrm{d}^4x\, \sqrt{-g}\phi R \qquad w(\phi) = -2. \tag{D.10}$$

The torsion here is also given in terms of $\partial\phi$. It is interesting to note that after eliminating the torsion from the action with the help of the equations of motion, the resulting effective action is Weyl invariant and equivalent to GR with metric $\bar{g}_{\mu\nu} = \phi g_{\mu\nu}$. Weyl invariance is a consequence of the existence of some hidden symmetry in the original theory (German 1985).

Action (D.10) in Weyl space W_4 does not give anything new: using the equations of motion we can express the Weyl vector φ_μ in terms of the gradient of the scalar field, whereupon the elimination of φ_μ from the action leads to a theory which is, again, equivalent to GR (Smalley 1986).

Trying to establish a theoretical foundation for the hypothesis of large numbers, Dirac (1973) studied action (D.8) in Weyl space W_4. This action is invariant under Weyl rescalings. We can show that in this case the scalar field is also non-dynamical, i.e. decoupled from other dynamical variables (Pietenpol *et al* 1974).

Exercises

1. Using the equations of motion of BD theory, show that the energy–momentum of matter obeys the relation $\nabla^{\mu} T_{\mu\nu} = 0$.
2. Find the trace of the energy–momentum of matter in a theory obtained from (D.8) by adding a mass term for the scalar field.
3. Find the change in the GR action under Weyl rescaling $g_{\mu\nu} \rightarrow \phi g_{\mu\nu}$. Rewrite the result in terms of φ, where $\phi = \varphi^2$.
4. Find an expression for the torsion using the equations of motion of the theory (D.10) in the U_4 space.
5. Calculate Weyl vector using the equations of motion of the theory (D.10) in the W_4 space.

Appendix E

Ashtekar's formulation of GR

The discord between quantum theory and gravitation is one of the greatest mysteries of modern physics. The perturbative methods of quantum field theory have been very successful in studying non-gravitational phenomena, but not very adequate for quantum GR. It seems, therefore, useful to try to develop a non-perturbative understanding of this theory. The Hamiltonian method is one of the standard approaches of this type. However, it usually leads to great difficulties in GR, where the connection is expressed in terms of the metric and the constraints are complicated, non-polynomial functions of metric (geometrodynamics). In Ashtekar's formalism, the connection remains one of the basic dynamical variables (connexo-dynamics) and the constraints become simpler by passing to a new set of *complex* variables (Ashtekar 1988, 1991). The main achievement of this approach is that the constraints now become *polynomial* functions of canonical variables.

As a gauge theory, gravity is naturally formulated in terms of tetrads and the connection. Ashtekar's formulation of GR can be understood as a canonical transform of Einstein–Cartan theory without matter, which is equivalent to GR, in the region of complex variables (Kamimura and Fukuyama 1990). In order to show this, we shall first clarify the Hamiltonian structure of a suitable reformulation of the EC theory.

Tetrad formulation of GR. EC theory without matter is defined by the Hilbert–Palatini (HP) action (5.47). As we mentioned in chapter 5, there exists a canonically equivalent formulation defined by

$$I'_{\mathrm{HP}} = a \int \mathrm{d}^4x\, \tfrac{1}{2}\varepsilon^{\mu\nu\lambda\rho}_{mnkl}[-\partial_\mu(b^k{}_\lambda b^l{}_\rho)A^{mn}{}_\nu + b^k{}_\lambda b^l{}_\rho A^m{}_{s\mu}A^{sn}{}_\nu] \tag{E.1}$$

where the ∂A terms are eliminated by adding a suitable four-divergence to the original EC action. In this approach, we can simply eliminate the connection A and the related momenta by using constraints and obtain the so-called tetrad formulation of GR.

In addition to the sure primary constraints, $\pi_i{}^0 \approx 0$ and $\pi_{ij}{}^0 \approx 0$, action (E.1) also yields the following additional primary constraints:

$$
\begin{aligned}
\phi_i{}^\alpha &\equiv \pi_i{}^\alpha + a\varepsilon^{0\alpha\beta\gamma}_{ijmn} b^j{}_\beta A^{mn}{}_\gamma \approx 0 \\
\phi_{ij}{}^\alpha &\equiv \pi_{ij}{}^\alpha \approx 0.
\end{aligned} \tag{E.2}
$$

Comparing these relations with (5.48), we see that this theory can be obtained from the HP formulation by the following canonical transformation:

$$
\begin{aligned}
\pi_i{}^\alpha &\to \pi_i{}^\alpha + a\varepsilon^{0\alpha\beta\gamma}_{ijmn} b^j{}_\beta A^{mn}{}_\gamma \\
\pi_{ij}{}^\alpha &\to \pi_{ij}{}^\alpha + a\varepsilon^{0\alpha\beta\gamma}_{ijmn} b^m{}_\beta b^n{}_\gamma
\end{aligned}
$$

while $b^i{}_\mu$ and $A^{ij}{}_\mu$ remain unchanged.

Since the Lagrangian is linear in velocities $\dot{b}^i{}_\mu$, the canonical Hamiltonian is given as $\mathcal{H}_{\mathrm{c}} = -\mathcal{L}(\dot{b} = 0)$. Explicit calculation leads to

$$
\mathcal{H}_{\mathrm{c}} = b^i{}_0\mathcal{H}_i - \tfrac{1}{2}A^{ij}{}_0\mathcal{H}_{ij} + \partial_\alpha D^\alpha \tag{E.3a}
$$

where

$$
\begin{aligned}
\mathcal{H}_i &= -\tfrac{1}{2}a\varepsilon^{0\alpha\beta\gamma}_{ijmn} b^j{}_\alpha R^{mn}{}_{\beta\gamma} \qquad \mathcal{H}_{ij} = -a\varepsilon^{0\alpha\beta\gamma}_{ijmn} b^m{}_\alpha T^n{}_{\beta\gamma} \\
D^\alpha &= -a\varepsilon^{0\alpha\beta\gamma}_{ijmn} b^m{}_\beta b^n{}_0 A^{ij}{}_\gamma.
\end{aligned} \tag{E.3b}
$$

Note that the expressions for $\mathcal{H}_i$ and $\mathcal{H}_{ij}$ are the same as in the HP formulation. This is not a surprise, since they are invariant under the canonical transformations.

Going over to the total Hamiltonian,

$$
\mathcal{H}_{\mathrm{T}} = \mathcal{H}_{\mathrm{c}} + u^i{}_0\pi_i{}^0 + \tfrac{1}{2}u^{ij}{}_0\pi_{ij}{}^0 + u^i{}_\alpha\phi_i{}^\alpha + \tfrac{1}{2}u^{ij}{}_\alpha\pi_{ij}{}^\alpha
$$

we obtain the consistency conditions of the sure primary constraints:

$$
\mathcal{H}_i \approx 0 \qquad \mathcal{H}_{ij} \approx 0. \tag{E.4a}
$$

Although the constraints $\phi_i{}^\alpha$ and $\phi_{ij}{}^\alpha$ are different from the related HP expressions, their consistency conditions remain of the same form:

$$
\begin{aligned}
\chi_i{}^\alpha &\equiv \tfrac{1}{2}a\varepsilon^{\alpha 0\beta\gamma}_{ijmn}(b^j{}_0 R^{mn}{}_{\beta\gamma} - 2b^j{}_\beta \underline{R}^{mn}{}_{0\gamma}) \approx 0 \\
\chi_{ij}{}^\alpha &\equiv -a\varepsilon^{\alpha 0\beta\gamma}_{ijmn}(b^m{}_0 T^n{}_{\beta\gamma} - 2b^m{}_\beta \underline{T}^n{}_{0\gamma}) \approx 0
\end{aligned} \tag{E.4b}
$$

where the underbars in $\underline{R}$ and $\underline{T}$ have the same meaning as in chapter 5, and will be omitted in further exposition for simplicity.

Combining the conditions $(-\mathcal{H}_i, \chi_i{}^\alpha)$ and $(\mathcal{H}_{ij}, \chi_{ij}{}^\alpha)$ we obtain the relations

$$
h_k{}^\mu R^k{}_i - \tfrac{1}{2}h_i{}^\mu R \approx 0 \qquad T^k{}_{\mu\nu} \approx 0
$$

which can be recognized as the Einstein equations for a gravitational field without matter. Out of $16 + 24$ of these equations, we have $4 + 12$ secondary constraints ($\mathcal{H}_i \approx 0$ and $T^k{}_{\alpha\beta} \approx 0$), while the remaining $12 + 12$ equations are used to fix 30 multipliers $u^{ij}{}_\alpha$ and $u^k{}_\alpha$ ($\chi_i{}^\alpha \approx 0$ and $T^k{}_{0\alpha} \approx 0$).

The consistency of the secondary constraints produces no new constraints. Indeed, the consistency condition for $\mathcal{H}_i$ is automatically fulfilled, while the consistency of $T^k{}_{\alpha\beta} \approx 0$ yields additional equations for the determination of $u^{ij}{}_\alpha$.

Among 40 primary constraints, we find 10 first class constraints ($\pi_i{}^0, \pi_{ij}{}^0$) and 30 second class ($\phi_i{}^\alpha, \pi_{ij}{}^\alpha$) ones; out of 16 secondary constraints, 10 of them are first class ($\mathcal{H}_i, \mathcal{H}_{ij}$), and the remaining six are second class (they are unified with $\mathcal{H}_{ij}$ into $T^k{}_{\alpha\beta} \approx 0$).

In order to study the elimination of the connection variables, we impose the time gauge condition: $b^0{}_\alpha \approx 0$. Then, $\mathcal{H}_{0a}$ effectively becomes second class. We shall now show that the 21 field variables ($b^0{}_\alpha, A^{ij}{}_\alpha$) and the corresponding momenta can be eliminated from the 42 second-class constraints ($\phi_i{}^\alpha, \pi_{ij}{}^\alpha, T^a{}_{\alpha\beta}, b^0{}_\alpha$).

In the time gauge, the constraints $\phi_i{}^\alpha$ are simplified:

$$\phi_0{}^\alpha \approx \pi_0{}^\alpha + a\varepsilon^{\alpha\beta\gamma}_{ebc} b^e{}_\beta A^{bc}{}_\gamma$$
$$\phi_a{}^\alpha \approx \pi_a{}^\alpha - 2a\varepsilon^{\alpha\beta\gamma}_{abc} b^b{}_\beta A^{c0}{}_\gamma$$

where $\varepsilon^{\alpha\beta\gamma} \equiv \varepsilon^{0\alpha\beta\gamma}$, $\varepsilon_{abc} \equiv \varepsilon_{0abc}$. Using 24 constraints $b^0{}_\alpha \approx 0$, $\phi_0{}^\alpha \approx 0$, $T^a{}_{\alpha\beta} \approx 0$ and $\pi_{ab}{}^\alpha \approx 0$, we can easily eliminate ($b^0{}_\alpha, A^{ab}{}_\alpha$) and the related momenta:

$$b^0{}_\alpha \approx 0 \qquad \pi_0{}^\alpha \approx -a\varepsilon^{\alpha\beta\gamma}_{ebc} b^e{}_\beta A^{bc}{}_\gamma$$
$$A^{ab}{}_\alpha \approx \Delta^{ab}{}_\alpha \qquad \pi_{ab}{}^\alpha \approx 0.$$

In order to solve the constraints $\phi_a{}^\alpha$ for $A^{c0}{}_\gamma$, it is useful to introduce the canonical transformation $(b^a{}_\alpha, A^{0b}{}_\beta) \to (E_a{}^\alpha, P^b{}_\beta)$, defined by

$$E_a{}^\alpha \equiv -\tfrac{1}{2}\varepsilon^{\alpha\beta\gamma}_{abc} b^b{}_\beta b^c{}_\gamma \qquad \pi_a{}^\alpha \equiv -\varepsilon^{\alpha\beta\gamma}_{abc} b^b{}_\beta P^c{}_\gamma. \tag{E.5}$$

In terms of the new variables, the constraint $\phi_a{}^\alpha$ has the simple form $P^c{}_\gamma + 2aA^{c0}{}_\gamma \approx 0$, so that the remaining 18 equations $\phi_a{}^\alpha \approx 0$ and $\pi_{a0}{}^\alpha \approx 0$ can be used to eliminate ($A^{0a}{}_\alpha, \pi_{0a}{}^\alpha$):

$$A^{0a}{}_\alpha \approx \frac{1}{2a} P^a{}_\alpha \qquad \pi_{0a}{}^\alpha \approx 0.$$

In the time gauge, the variable $A^{0b}{}_0$ is also fixed. Indeed, the consistency of the time gauge implies $\dot{b}^0{}_\alpha = u^0{}_\alpha \approx 0$, where $u^0{}_\alpha$ is determined by the relation $T^0{}_{0\alpha} \approx 0$. This leads to the additional relations

$$A^0{}_{c0} b^c{}_\alpha \approx \partial_\alpha b^0{}_0 + \frac{1}{2a} P_{c\alpha} b^c{}_0 \qquad \pi_{b0}{}^0 \approx 0$$

which do not influence the form of the FC constraints, since $\mathcal{H}_i$ and $\mathcal{H}_{ab}$ are independent of $A^{0a}{}_0$ ($\mathcal{H}_{0b}$ is effectively second class).

After imposing the time gauge, the FC constraints $\mathcal{H}_i$ and $\mathcal{H}_{ab}$ become simpler:

$$\begin{aligned} M_a &\equiv -\tfrac{1}{2}\varepsilon_{abc}\mathcal{H}^{bc} = a\varepsilon^{\alpha\beta\gamma}T^0{}_{\alpha\beta}b_{a\gamma} \\ -(1/a)\mathcal{H}_0 &= \varepsilon^{\alpha\beta\gamma}_{abc}b^a{}_\alpha(\partial_\beta A^{bc}{}_\gamma + A^b{}_{e\beta}A^{ec}{}_\gamma + A^b{}_{0\beta}A^{0c}{}_\gamma) \\ \mathcal{H}_a &= 2a\varepsilon^{\alpha\beta\gamma}_{abc}b^b{}_\alpha(\partial_\beta A^{c0}{}_\gamma + A^c{}_{d\beta}A^{d0}{}_\gamma). \end{aligned} \tag{E.6}$$

They are polynomial in the basic canonical variables, up to the multiplicative factor J^{-1}. This property is spoiled after the connection is eliminated with the help of the second-class constraints. After eliminating $A^{ab}{}_\alpha$ and $A^{0b}{}_\alpha$, the constraints $\mathcal{H}_{ab}$ and $\mathcal{H}_a$ take the form

$$\begin{aligned} M_a &= -\varepsilon_{abc}P^b{}_\alpha E^{c\alpha} \\ \mathcal{H}_a &= J^{-1}[E_a{}^\gamma E_b{}^\beta(\partial_\beta P^b{}_\gamma - \partial_\gamma P^b{}_\beta) + P_{a\beta}E_b{}^\beta\partial_\alpha E^{b\alpha}]. \end{aligned} \tag{E.7a}$$

Turning now to $\mathcal{H}_0$, we note that its first term can be written in the form $2\partial_\beta(J^{-1}E_b{}^\beta\partial_\alpha E^{b\alpha}) + 2\partial_\beta(E_b{}^\beta h_c{}^\gamma)\Delta^{bc}{}_\gamma$. Half of the second part of this expression is cancelled with the term $\Delta^b{}_{e\beta}\Delta^{ec}{}_\gamma$, so that

$$\begin{aligned} -(1/a)\mathcal{H}_0 = 2\partial_\beta(J^{-1}E_b{}^\beta\partial_\alpha E^{b\alpha}) &+ \frac{1}{4a^2}\varepsilon^{abc}(J^{-1}E_b{}^\beta E_a{}^\gamma)(\varepsilon_{cef}P^e{}_\beta P^f{}_\gamma) \\ &+ \partial_\beta(J^{-1}E_b{}^\beta E_c{}^\gamma)\Delta^{bc}{}_\gamma. \end{aligned} \tag{E.7b}$$

Constraints M_a and $J\mathcal{H}_a$ remain polynomial on the reduced set of new canonical variables. Unfortunately, this is not the case with $\mathcal{H}_0$, the reason being not only the presence of the last term, but also the appearance of the factor J^{-1}. In further analysis, we shall see how the transition to complex variables transforms the FC constraints into the polynomial form, thereby simplifying some aspects of the canonical structure of the theory.

Ashtekar's formalism. The complex canonical formalism can be simply introduced by adding an imaginary total divergence to the Lagrangian in (E.1). Although the action is *complex*, it correctly describes the equations of motion of the real theory. The canonical variables also become complex; however, their structure is not arbitrary, but precisely determined by the form of the imaginary total divergence. This implies the existence of certain relations between canonical variables and their complex conjugates, called the *reality* conditions, which play an important role in the new formalism.

Ashtekar's formulation can be obtained starting from the following complex action:

$$I_\mathrm{A} = I'_\mathrm{HP} - a\int \mathrm{d}^4x\,\mathrm{i}\varepsilon^{\mu\nu\lambda\rho}\partial_\mu(b^k{}_\nu\partial_\lambda b_{k\rho}). \tag{E.8}$$

The imaginary four-divergence implies the appearance of complex momenta, as can be seen from the form of primary constraints:

$$\begin{aligned}\phi_i{}^\alpha &\equiv \pi_i{}^\alpha + a\varepsilon^{0\alpha\beta\gamma}_{ijmn} b^j{}_\beta A^{mn}{}_\gamma + 2\mathrm{i}a\varepsilon^{0\alpha\beta\gamma}\partial_\beta b_{i\gamma} \approx 0 \\ \pi_{ij}{}^\alpha &\approx 0.\end{aligned} \tag{E.9a}$$

Since the imaginary part of the action is linear in the velocities, the canonical Hamiltonian remains essentially of the form (E.3): the only effect is the replacement $D^\alpha \rightarrow D^\alpha + \mathrm{i}\varepsilon^{\alpha\nu\lambda\rho}(b^k{}_\nu\partial_\lambda b_{k\rho})$. The imaginary part of the constraints, which depends only on the fields (and not on the momenta), does not change earlier considerations of the consistency, so that, again, we obtain relations (E.4a, b). These results can be understood by observing that the canonical transformation acts only on the momenta, so that all expressions that are independent of the momenta remain unchanged. Of course, the same results can also be obtained by direct calculation. Therefore, the only important change, compared to the real formulation, is the change in the form of the primary constraints (E.9a).

As before, we shall now impose the time gauge for simplicity, whereupon the primary constraints take the form

$$\begin{aligned}\phi_0{}^\alpha &\approx \pi_0{}^\alpha + a\varepsilon^{\alpha\beta\gamma}_{ebc} b^e{}_\beta A^{bc}{}_\gamma \\ \phi_a{}^\alpha &\approx \pi_a{}^\alpha - 2a\varepsilon^{\alpha\beta\gamma}_{abc} b^b{}_\beta A^{c0}{}_\gamma + 2\mathrm{i}a\varepsilon^{0\alpha\beta\gamma}\partial_\beta b_{a\gamma}.\end{aligned} \tag{E.9b}$$

Let us now introduce the concept of self-dual connection. Omitting the spacetime index, we first define the dual of A^{ij} by ${}^*A^{ij} = \frac{1}{2}\varepsilon^{ijmn}A_{mn}$, so that ${}^{**}A^{ij} = -A^{ij}$. Each Lorentz connection can be decomposed according to $A^{ij} = \frac{1}{2}(\mathcal{A}^{ij}_+ + \mathcal{A}^{ij}_-)$, where

$$\mathcal{A}^{ij}_\pm = A^{ij} \mp \mathrm{i}^*A^{ij} \tag{E.10a}$$

are the complex self-dual (+) and anti self-dual (−) parts of A^{ij}, ${}^*\mathcal{A}^{ij}_\pm = \pm i\mathcal{A}^{ij}_\pm$. These parts are orthogonal in the sense that $(\mathcal{A}_+\mathcal{B}_-)^{ij} \equiv \mathcal{A}^{im}_+\mathcal{B}_{-m}{}^j = 0$, and, as a consequence, $[\mathcal{A},\mathcal{B}]^{ij} = [\mathcal{A}_+,\mathcal{B}_+]^{ij} + [\mathcal{A}_-,\mathcal{B}_-]^{ij}$. The complexified Lorentz algebra can be decomposed into self-dual and anti self-dual subalgebra: $so(1,3,C) = so(3) \oplus so(3)$. In further exposition we shall use only the self-dual connection, omitting its index + for simplicity.

In the time gauge, the complex self-dual connection has the form

$$\mathcal{A}^{a0}{}_\alpha = A^{a0}{}_\alpha + \tfrac{1}{2}\mathrm{i}\varepsilon^{abc}\Delta_{bc\alpha} \tag{E.10b}$$

and the constraints $\phi_a{}^\alpha$ can be expressed as

$$\phi_a{}^\alpha = \pi_a{}^\alpha - 2a\varepsilon^{\alpha\beta\gamma}_{abc} b^b{}_\beta \mathcal{A}^{c0}{}_\gamma.$$

Of course, this is not just a coincidence: the imaginary part of the action has been chosen so that $\phi_a{}^\alpha$ contains the whole complex self-dual connection.

After we apply the canonical transformation (E.5), the constraint $\phi_a{}^\alpha$ acquires the simple form $P^a{}_\alpha + 2a\mathcal{A}^{a0}{}_\alpha \approx 0$, which can be solved for $A^{0a}{}_\alpha$:

$$A^{0a}{}_\alpha = \frac{1}{2a}P^a{}_\alpha + \frac{\mathrm{i}}{2}\varepsilon^{aef}\Delta_{ef\alpha}.$$

Now, we shall find the form of the remaining FC constraints (E.6) in terms of the new variables. Look, first, at M_a. Using the previous expression for $A^{0a}{}_\alpha$ and the definition of $T^0{}_{\alpha\beta}$, we obtain

$$(1/2a)M_a = \mathrm{i}\nabla_\alpha E_a{}^\alpha \qquad \nabla_\alpha E_a{}^\alpha \equiv \partial_\alpha E_a{}^\alpha + \frac{\mathrm{i}}{2a}\varepsilon_{abc}P^b{}_\alpha E^{c\alpha}. \tag{E.11a}$$

After eliminating the connection, the first term on the right-hand side of $-(1/a)\mathcal{H}_0$ takes the form $2\partial_\beta(J^{-1}E_b{}^\beta\partial_\alpha E^{b\alpha}) + 2\partial_\beta(E_b{}^\beta h_c{}^\gamma)\Delta^{bc}{}_\gamma$. Combining the first part of this expression with terms linear and quadratic in $P^c{}_\gamma$ that stem from $A^b{}_{0\beta}A^{0c}{}_\gamma$,

$$\varepsilon^{\alpha\beta\gamma}_{abc}b^a{}_\alpha\left(-\frac{1}{4a^2}P^b{}_\beta P^c{}_\gamma - \frac{\mathrm{i}}{2a}\varepsilon^{bef}\Delta_{ef\beta}P^c{}_\gamma + \frac{1}{4}\varepsilon^{bef}\Delta_{ef\beta}\varepsilon^{cdg}\Delta_{dg\gamma}\right)$$

and taking into account that all the remaining contributions cancel out, we find that

$$\begin{aligned}\mathcal{H}_0 &= \mathrm{i}\partial_\beta(J^{-1}E_b{}^\beta M^b) - \frac{\mathrm{i}}{2}\varepsilon^{abc}(J^{-1}E_a{}^\beta E_b{}^\gamma)F_{c\beta\gamma}\\ F_{c\beta\gamma} &\equiv \partial_\beta P_{c\gamma} - \partial_\gamma P_{c\beta} + \frac{\mathrm{i}}{2a}\varepsilon_{cef}P^e{}_\beta P^f{}_\gamma.\end{aligned} \tag{E.11b}$$

Note the importance of the imaginary part of the connection in the calculations leading to this result.

After we have eliminated the connection in $\mathcal{H}_a$, the related contribution of the terms linear in $P^a{}_\alpha$ is given by

$$\begin{aligned}&J^{-1}[-E_a{}^\beta E_c{}^\gamma F^c{}_\beta\gamma + (P_{a\beta} + \mathrm{i}a\varepsilon_{aef}\Delta^{ef}{}_\beta)E_b{}^\beta(\nabla_\alpha E^{b\alpha})]\\ &\quad - J^{-1}\mathrm{i}a\varepsilon_{aef}\Delta^{ef}{}_\beta E_b{}^\beta\partial_\alpha E^{b\alpha}.\end{aligned}$$

Terms quadratic in $P^a{}_\alpha$ are also present, but they cancel each other out. After the last part of this expression is cancelled out with the remaining contributions from $\mathcal{H}_a$, the final result takes the form

$$\mathcal{H}_a = J^{-1}\left[E_a{}^\gamma E_b{}^\beta F^b{}_{\beta\gamma} - \frac{\mathrm{i}}{2a}(P_{a\beta} + \mathrm{i}a\varepsilon_{aef}\Delta^{ef}{}_\beta)E^{b\beta}M_b\right]. \tag{E.11c}$$

Summarizing, we can conclude that after the elimination of all second-class constraints, the following FC constraints remain:

$$\begin{aligned}G_0 &\equiv -\tfrac{1}{2}\mathrm{i}\varepsilon^{abc}E_a{}^\alpha E_b{}^\beta F_{c\alpha\beta}\\ G_\alpha &\equiv E^{b\beta}F_{b\beta\alpha} \qquad M_a \equiv 2\mathrm{i}a\nabla_\alpha E_a{}^\alpha\end{aligned} \tag{E.12}$$

which are expressed as *polynomials* in the complex variables $E_a{}^\alpha$ and $P^a{}_\alpha$.

Starting from the relations

$$\mathcal{H}_0 = J^{-1}G_0 + \mathrm{i}\partial_\beta(J^{-1}E^{b\beta}M_b)$$
$$\mathcal{H}_a = J^{-1}\left[E_a{}^\alpha G_\alpha - \frac{\mathrm{i}}{2a}(P_{a\beta} + \mathrm{i}a\varepsilon_{aef}\Delta^{ef}{}_\beta)E^{b\beta}M_b\right].$$

we can rewrite the canonical Hamiltonian in the form

$$\mathcal{H}_\mathrm{c} = NG_0 + N^\alpha G_\alpha + \Lambda^a M_a \tag{E.13}$$

where N and N^α are real, and Λ^a is a complex multiplier.

The use of complex variables demands a careful analysis of the *reality conditions* in the theory. The transition to the complex action, in which $b^i{}_\mu$ and $A^{ij}{}_\mu$ are real variables, implies that the momenta $\pi_a{}^\alpha$ become complex. We note that the momenta always appear in constraints in the real combination $\pi_a{}^\alpha + 2\mathrm{i}a\varepsilon^{\alpha\beta\gamma}\partial_\beta b_{a\gamma}$. Consequently, all constraints are real, as is the total Hamiltonian. Real constraints are, however, non-polynomial in real variables. The transition to polynomial constraints $M_a, \mathcal{H}_a$ and $\mathcal{H}_0$ is realized by introducing complex variables. It is important to stress that the Hamiltonian is a real quantity, as a consequence of the fact that complex constraints in (E.13) are not arbitrary complex functions, but satisfy definite reality conditions. Thus, for instance, since the quantity $P^a{}_\alpha + \mathrm{i}a\varepsilon^{aef}\Delta_{ef\alpha}$ is real, the reality condition for $P^a{}_\alpha$ has the form

$$P^a{}_\alpha{}^+ = P^a{}_\alpha + 2\mathrm{i}a\varepsilon^{aef}\Delta_{ef\alpha} \tag{E.14}$$

and similar conditions exist for constraints M_a, $\mathcal{H}_a$, $\mathcal{H}_0$ and multipliers Λ^a. In general, these conditions are non-polynomial.

From the mere fact that the constraints are polynomial, it is not easy to understand the importance of this property for the theory. Studies of quantum physical states, that obey all conditions imposed by the constraints, led to the result that these states can be constructed using certain gauge invariant variables. Variables of this type are known in non-Abelian gauge theories under the name of Wilson loops. Despite great efforts, such a result has been never obtained in the standard formulation of GR.

Polynomial constraints are also obtained in the presence of some matter fields. In particular, supergravity can be successfully studied in this way. However, the construction of physical states is not sufficiently clear here.

Generalization of Ashtekar's approach to higher dimensional theories seems rather difficult, since the concept of self-duality is specific to $d = 4$.

The transition to complex polynomial constraints is accompanied by the corresponding reality conditions, in which all the specific features of the gravitational interaction are reflected. In spite of many encouraging results, there is much work still to be done before we can reach a final conclusion concerning the role and importance of this programme in the search for quantum gravity.

Exercises

1. Prove the following identities in the time gauge:

$$\varepsilon^{\alpha\beta\gamma} b^a{}_\alpha b^b{}_\beta b^c{}_\gamma = J\varepsilon^{abc} \qquad \varepsilon^{abc} h_a{}^\alpha h_b{}^\beta h_c{}^\gamma = J^{-1}\varepsilon^{\alpha\beta\gamma}$$
$$\varepsilon^{\alpha\beta\gamma} b^b{}_\beta b^c{}_\gamma = J\varepsilon^{abc} h_a{}^\alpha \qquad \varepsilon^{\alpha\beta\gamma}_{abc} b^b{}_\beta b^c{}_\gamma = -2Jh_a{}^\alpha$$
$$\varepsilon^{\alpha\beta\gamma} b^c{}_\gamma = J\varepsilon^{abc} h_a{}^\alpha h_b^\beta \qquad \varepsilon^{\alpha\beta\gamma}_{abc} b^c{}_\gamma = -2Jh_{[a}{}^\alpha h_{b]}{}^\beta$$

2. Check whether the following identities are correct in the time gauge:

$$E_a{}^\alpha = Jh_a{}^\alpha \qquad P^a{}_\alpha = J^{-1}(b^a{}_\beta b^b{}_\alpha - \tfrac{1}{2} b^a{}_\alpha b^b{}_\beta)\pi_b{}^\beta$$
$$-2b^a{}_\alpha = J^{-1}\varepsilon^{abc}_{\alpha\beta\gamma} E_b{}^\beta E_c{}^\gamma$$
$$\partial_\alpha E^{a\alpha} = -J\Delta^{ab}{}_b \qquad \varepsilon^{\alpha\beta\gamma}_{abc} b^b{}_\beta(\tfrac{1}{2}\varepsilon^{cef}\Delta_{ef\gamma}) = -\varepsilon^{\alpha\beta\gamma}\partial_\beta b_{a\gamma}.$$

3. Show that the constraints (E.6) of the real formulation (E.1) are polynomial in the variables $E_a{}^\alpha$, $P^b{}_\beta$ and $A^{ab}{}_\gamma$, up to the multiplicative factor J^{-1}.
4. Derive relations (E.11a, c) for constraints M_a and $\mathcal{H}_a$. Taking the variable $P^a{}_\alpha$ as real, find the real parts of constraints M_a and $\mathcal{H}_a$, and compare them with the result (E.7a) for the real formulation.
5. Derive the relation (E.11b) for constraint $\mathcal{H}_0$. Taking the variable $P^a{}_\alpha$ as real, find $\mathrm{Re}(\mathcal{H}_0)$ and compare it with the result (E.7b) for the real formulation. Explain any disagreement.
6. Derive the form of the canonical Hamiltonian (E.13) and find the explicit form of the multipliers N, N^α and Λ^a.
7. Find the reality conditions for the constraints M_a, $\mathcal{H}_a$ and $\mathcal{H}_0$, as well as for the multipliers N, N_α and Λ^a.
8. Find the Hamiltonian equations of motion for $E_a{}^\alpha$ and $P^a{}_\alpha$ when the multipliers N^α and Λ^a vanish ('pure' temporal evolution).

Appendix F

Constraint algebra and gauge symmetries

In the framework of the Hamiltonian formalism, Castellani (1982) developed an algorithm for constructing all gauge generators on the basis of the known Hamiltonian and the algebra of constraints. His arguments can also be used in a *reversed* direction: PGT possesses local symmetry for which the generators have been found in chapter 6, which allows us to obtain very precise information about the algebra of FC constraints (Blagojević and Vasilić 1987).

Dirac found the algebra of FC constraints in any generally covariant theory, such as metric gravity (Dirac 1964). The same result was derived on the basis of the principle of 'path independence' of dynamical evolution (Teitelboim 1973). These 'geometric' methods are very powerful and give us a deep insight into the structure of the theory. However, they are unable to give us complete information on the structure of the constraint algebra, such as, for instance, the information on the presence of squares of constraints, and must be supplemented by specific additional considerations.

When the only constraints in PGT are those related to the Poincaré gauge symmetry, the algebra of constraints has the standard form (6.1), containing no squares of constraints. Here, we shall show that this algebra also has the same form in the general case, up to the presence of PFC constraints and squares or higher powers of constraints. The conclusion follows from certain consistency requirements on the structure of gauge generators. In the course of our exposition, it will become clear that all methods based on 'geometric' arguments have the same degree of uncertainty.

The Hamiltonian of the general theory has the form (5.44), and the gauge generator is given by the expression (6.7), where the phase-space functions $G^{(0)}$, $G^{(1)}$ must satisfy the following consistency requirements:

$$G_1 = C_{\mathrm{PFC}} \tag{F.1a}$$

$$G_0 + \{G_1, H_{\mathrm{T}}\} = C_{\mathrm{PFC}} \tag{F.1b}$$

$$\{G_0, H_{\mathrm{T}}\} = C_{\mathrm{PFC}}. \tag{F.1c}$$

Here, $C_{\rm PFC}$ denotes PFC constraints, and the equality means an equality up to the constraints of the type $\chi^n (n \geq 2)$, which are FC. Therefore, all consequences of these relations have the same ambiguity. Now, we are going to follow the 'inverse' Castellani method, i.e. we shall start with the known Hamiltonian and gauge generator, and examine all consequences of the consistency requirement (F.1) on the algebra of the FC constraints $\mathcal{H}_\perp$, $\mathcal{H}_\alpha$ and $\mathcal{H}_{ij}$.

The first requirement (F.1a) applied to the gauge generator (6.7) leads immediately to

$$\pi_k{}^0, \pi_{ij}{}^0 = C_{\rm PFC}. \tag{F.2}$$

Using the fact that primary constraints do not depend on the unphysical variables $b^k{}_0$ and $A^{ij}{}_0$, we can easily show that

$$\{\pi_k{}^0, C'_{\rm PFC}\} = C_{\rm PFC} \qquad \{\pi_{ij}{}^0, C'_{\rm PFC}\} = C_{\rm PFC}.$$

Let us now denote primary first- and second-class constraints by ϕ_1 and ϕ_2, and the related multipliers by v and u_2, respectively, so that the total Hamiltonian (5.44a) can be written as

$$\widehat{\mathcal{H}}_{\rm T} = \mathcal{H}_{\rm c} + (v\phi_1) + (u_2\phi_2). \tag{F.3a}$$

Taking into account that $G^0_{ij}, G^0_\alpha = C_{\rm PFC}$, the second requirement (F.1b) applied to G_{ij}, G_α and G_0 leads to

$$\int \{\pi_{ij}{}^0, (u_2\phi_2)'\} = C_{\rm PFC}$$
$$\int \{b^k{}_\alpha \pi_k{}^0, (u_2\phi_2)'\} = C_{\rm PFC}$$
$$(u_2\phi_2) + \int \{b^k{}_0 \pi_k{}^0, (u_2\phi_2)'\} = C_{\rm PFC}.$$

By solving these equations, we conclude that all determined multipliers u_2 can be chosen so that (a) they do not depend on $A^{ij}{}_0$, and (b) their dependence on $b^k{}_0$ is given by a proportionality to N. Therefore,

$$u_2 = N\Lambda_\perp(b^k{}_\alpha, A^{ij}{}_\alpha, \pi_k{}^\alpha, \pi_{ij}{}^\alpha).$$

Consequently, we can redefine the canonical Hamiltonian by including the contribution of the term $u_2\phi_2$,

$$\begin{aligned} \overline{\mathcal{H}}_{\rm c} &= N\overline{\mathcal{H}}_\perp + N^\alpha \mathcal{H}_\alpha - \tfrac{1}{2}A^{ij}{}_0\mathcal{H}_{ij} \\ \overline{\mathcal{H}}_\perp &\equiv \mathcal{H}_\perp + \Lambda_\perp\phi_2 \end{aligned} \tag{F.3b}$$

and rewrite the total Hamiltonian as

$$\widehat{\mathcal{H}}_{\rm T} = \overline{\mathcal{H}}_{\rm c} + (v\phi_1). \tag{F.3c}$$

The third requirement (F.1c) results in two sets of conditions. Demanding its validity for an arbitrary multiplier v, the $(v\phi_1)$ part of $\widehat{\mathcal{H}}_{\mathrm{T}}$ gives rise to

$$
\begin{aligned}
\{\mathcal{H}_{ij}, \phi'_1\} &= C_{\mathrm{PFC}} \\
\{\mathcal{H}_\alpha - \tfrac{1}{2}A^{ij}{}_\alpha \mathcal{H}_{ij}, \phi'_1\} &= C_{\mathrm{PFC}} \\
\{\mathcal{H}_{\mathrm{c}} + (u_2\phi_2), \phi'_1\} &= C_{\mathrm{PFC}} \\
\{C_{\mathrm{PFC}}, \phi'_1\} &= C_{\mathrm{PFC}}
\end{aligned}
\tag{F.4}
$$

whereas the $\overline{H}_{\mathrm{c}}$ part produces

$$\{\mathcal{H}_{ij}, \overline{H}_{\mathrm{c}}\} = -2b_{[i0}\overline{\mathcal{H}}_{j]} - 2A^s{}_{[i0}\mathcal{H}_{j]s} + C_{\mathrm{PFC}} \tag{F.5a}$$

$$\{\mathcal{H}_\alpha - \tfrac{1}{2}A^{ij}{}_\alpha \mathcal{H}_{ij}, \overline{H}_{\mathrm{c}}\} = b^k{}_{0,\alpha}\overline{\mathcal{H}}_k - \tfrac{1}{2}A^{ij}{}_{0,\alpha}\mathcal{H}_{ij} + C_{\mathrm{PFC}} \tag{F.5b}$$

$$\{\overline{\mathcal{H}}_{\mathrm{c}}, \overline{H}_{\mathrm{c}}\} = C_{\mathrm{PFC}}. \tag{F.5c}$$

The first set of conditions is trivially satisfied by $\phi'_1 = \pi'_k{}^0, \pi'_{ij}{}^0$. As for extra FC constraints, we could also find some useful relations for them but they would be of no interest here.

An analysis of the second set of conditions gives the most interesting consequences. Indeed, explicitly extracting the dependence of the left-hand sides of equations (F.5) on the unphysical variables $b^k{}_0$ and $A^{ij}{}_0$, and then comparing with the right-hand sides, leads directly to the algebra of constraints.

The left-hand side of (F.5a) can be written as

$$L_{ij} \equiv \{\mathcal{H}_{ij}, \overline{H}_{\mathrm{c}}\} = \int b'^k{}_0 \{\mathcal{H}_{ij}, \overline{\mathcal{H}}'_k\} - \tfrac{1}{2}\int A'^{kl}{}_0 \{\mathcal{H}_{ij}, \mathcal{H}'_{kl}\}$$

since $\mathcal{H}_{ij}$ does not contain the momenta $\pi_k{}^0, \pi_{ij}{}^0$. We search for a solution of the Poisson brackets of constraints in the form

$$
\begin{aligned}
\{\mathcal{H}_{ij}, \overline{\mathcal{H}}'_k\} &= \Lambda_{ijk}\delta + \Lambda_{ijk}{}^\alpha \partial_\alpha \delta + \cdots \\
\{\mathcal{H}_{ij}, \mathcal{H}'_{kl}\} &= \Lambda_{ijkl}\delta + \Lambda_{ijkl}{}^\alpha \partial_\alpha \delta + \cdots
\end{aligned}
$$

where the coefficients Λ do not depend on the unphysical variables. By using these relations in L_{ij}, and comparing with the right-hand side of equation (F.5a), we find that

$$
\begin{aligned}
\Lambda_{ijk} &= -2\eta_{k[i}\overline{\mathcal{H}}_{j]} + C_{\mathrm{PFC}} &\qquad \Lambda_{ijk}{}^\alpha &= C_{\mathrm{PFC}} \\
\Lambda_{ijkl} &= 4\eta_{[i[l}\mathcal{H}_{j]k]} + C_{\mathrm{PFC}} &\qquad \Lambda_{ijkl}{}^\alpha &= C_{\mathrm{PFC}}
\end{aligned}
$$

and, consequently,

$$
\begin{aligned}
\{\mathcal{H}_{ij}, \mathcal{H}'_{kl}\} &= \tfrac{1}{2} f_{ij}{}^{mn}{}_{kl}\mathcal{H}_{mn}\delta + C_{\mathrm{PFC}} \\
\{\mathcal{H}_{ij}, \overline{\mathcal{H}}'_k\} &= -2\eta_{k[i}\overline{\mathcal{H}}_{j]}\delta + C_{\mathrm{PFC}}.
\end{aligned}
$$

Going, now, to the ADM basis $\boldsymbol{e}'_\mu = \{\boldsymbol{n}, \boldsymbol{e}_\alpha\}$, we finally obtain

$$\{\mathcal{H}_{ij}, \mathcal{H}'_{kl}\} = \tfrac{1}{2} f_{ij}{}^{mn}{}_{kl}\mathcal{H}_{mn}\delta + C_{\text{PFC}} \qquad (\text{F.6}a)$$
$$\{\mathcal{H}_{ij}, \mathcal{H}'_\alpha\} = C_{\text{PFC}} \qquad (\text{F.6}b)$$
$$\{\mathcal{H}_{ij}, \overline{\mathcal{H}}'_\perp\} = C_{\text{PFC}}. \qquad (\text{F.6}c)$$

The rest of the algebra of constraints is obtained in a similar way and has the form

$$\{\mathcal{H}_\alpha, \mathcal{H}'_\beta\} = (\mathcal{H}'_\alpha \partial_\beta + \mathcal{H}_\beta \partial_\alpha - \tfrac{1}{2} R^{ij}{}_{\alpha\beta}\mathcal{H}_{ij})\delta + C_{\text{PFC}} \qquad (\text{F.6}d)$$
$$\{\mathcal{H}_\alpha, \overline{\mathcal{H}}'_\perp\} = (\overline{\mathcal{H}}_\perp \partial_\alpha - \tfrac{1}{2} R^{ij}{}_{\alpha\perp}\mathcal{H}_{ij})\delta + C_{\text{PFC}} \qquad (\text{F.6}e)$$
$$\{\overline{\mathcal{H}}_\perp, \overline{\mathcal{H}}'_\perp\} = -(\mathcal{H}^\alpha + \mathcal{H}'^\alpha)\partial_\alpha\delta + C_{\text{PFC}} \qquad (\text{F.6}f)$$

where $\mathcal{H}^\alpha = {}^3g^{\alpha\beta}\mathcal{H}_\beta$. The time derivative $\dot{A}^{ij}{}_\alpha$ u $R^{ij}{}_{\alpha\perp}$ is only short for $\{A^{ij}{}_\alpha, \overline{H}_{\text{c}}\}$, so that these relations do not depend on arbitrary multipliers v appearing in $\widehat{\mathcal{H}}_{\text{T}}$.

Let us observe once more that the equalities in (F.6) are equalities up to the powers of constraints $\chi^n (n \geq 2)$. This is a consequence of our method, which is based on symmetry requirements, as expressed by (F.1).

At the end of these considerations, it is natural to ask the following question: can the form of the constraint algebra (F.6) be improved, with more precise information concerning terms of the type C_{PFC} and χ^n? Because of the simple form of the kinematical generators $\mathcal{H}_{ij}$ and $\mathcal{H}_\alpha$, their Poisson brackets can easily be checked by a direct calculation (as we have seen in chapter 6), and the result is that C_{PFC} and the χ^n terms are in fact *absent*. The remaining relations in (F.6) involve the dynamical part of the Hamiltonian $\overline{\mathcal{H}}_\perp$, so that an explicit calculation becomes much more difficult. Since $\mathcal{H}_{ij}$, $\mathcal{H}_\alpha$ and $\overline{\mathcal{H}}_\perp$ do not depend on $\pi_k{}^0$, $\pi_{ij}{}^0$, we can conclude that C_{PFC} terms in (F.6) are *extra* PFC constraints. Consider, further, relations (F.6c) and (F.6e) that describe the behaviour of $\overline{\mathcal{H}}_\perp$ under Lorentz rotations and space translations. The absence of C_{PFC} and χ^n terms in these relations is equivalent to the statement that $\overline{\mathcal{H}}_\perp$ is a scalar density. The nature of $\overline{\mathcal{H}}_\perp$ can be checked on the basis of the known behaviour of all variables in $\overline{\mathcal{H}}_\perp$ under the Poincaré gauge transformations. Finally, the question of the exact form of the Poisson bracket $\{\overline{\mathcal{H}}_\perp, \overline{\mathcal{H}}'_\perp\}$ is the most difficult one. It cannot be improved by a similar method, as the generators of the time translations are not, and cannot be, off-shell generators. The only thing we can do is calculate C_{PFC} and the χ^n terms in (F.6f), in a particular theory (Nikolić 1995).

It should be noted that there are certain features of the constraint algebra which do not follow from a given theory, but from ambiguities in the process of constructing the Hamiltonian. For instance, we can choose all determined multipliers to be independent of the momentum variables appearing in the constraints, while any other choice would be equivalent to the replacement $u_2 \to u_2 + \lambda_2\phi$. It is clear that this would change not only the form of $\overline{\mathcal{H}}_\perp$, but also the corresponding Poisson brackets (by adding terms of the form $C_{\rm PFC}$ and χ^n). The most natural choice is one that gives the simplest gauge algebra.

Appendix G

Covariance, spin and interaction of massless particles

In relativistic quantum field theory, a process in which a massless particle of spin 1 or 2 is emitted (or absorbed) is described by the scattering (S) matrix, which may be written in the form

$$\epsilon_\mu M^\mu \quad \text{or} \quad \epsilon_{\mu\nu} M^{\mu\nu}.$$

Here, ϵ_μ and $\epsilon_{\mu\nu}$ are the respective polarization states, and M^μ, $M^{\mu\nu}$ are formed from the variables corresponding to the other particles. The polarizations ϵ_μ and $\epsilon_{\mu\nu}$ are not covariant objects (vectors or tensors) with respect to the Poincaré group, but undergo additional gauge transformations (appendix I),

$$\epsilon_\mu \to \epsilon_\mu + k_\mu \eta \qquad \epsilon_{\mu\nu} \to \epsilon_{\mu\nu} + k_\mu \eta_\nu + k_\nu \eta_\mu$$

which are specific for massless particles. Hence, the Poincaré invariance of quantum field theory requires the conditions

$$k_\mu M^\mu = 0, \qquad k_\mu M^{\mu\nu} = 0. \tag{G.1}$$

These conditions, usually referred to as the gauge invariance conditions of the S matrix, give very stringent restrictions on the structure of the theory through the so-called low-energy theorems (Weinberg 1964, Kibble 1965).

In order to illustrate the nature of the gauge-invariance conditions (G.1), we assume that matter is described by a complex scalar field Φ,

$$\mathcal{L}_{\mathrm{M}} = \partial_\mu \Phi^* \partial^\mu \Phi - m^2 \Phi^* \Phi \tag{G.2}$$

in which the propagator has the form $D(k, m^2) = 1/(k^2 - m^2)$.

For massless fields of spin $s > 0$, gauge invariance implies the conservation of the matter current. The matter Lagrangian usually has a global symmetry which automatically leads to a conserved current. This is the case with the standard

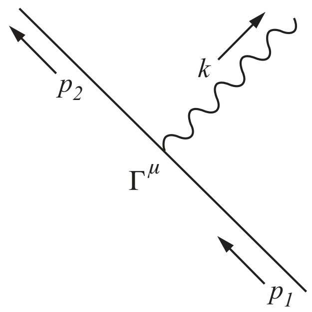

Figure G.1. The vertex diagram.

electrodynamics, where the global $U(1)$ symmetry in the matter sector generates the conserved current J_μ. Similarly, the global translational symmetry leads to the conserved energy–momentum tensor $T_{\mu\nu}$ (to which the massless spin-2 field is coupled).

The vertex of the spin-1 graviton. The massless spin-1 field φ_μ, usually called the photon, is coupled to the conserved matter current J_μ. The complete Lagrangian has the form

$$\mathcal{L} = \mathcal{L}_\mathrm{V} + \mathcal{L}_\mathrm{M} + \mathcal{L}_\mathrm{I} \tag{G.3}$$

where $\mathcal{L}_\mathrm{V}$ and $\mathcal{L}_\mathrm{M}$ are given by equations (7.5) and (G.2), and $\mathcal{L}_\mathrm{I}$ is the interaction term:

$$\begin{aligned} \mathcal{L}_\mathrm{I} &= -\lambda\varphi_\mu J^\mu \\ J_\mu &\equiv -\mathrm{i}[(\partial_\mu\Phi^*)\Phi - \Phi^*(\partial_\mu\Phi)]. \end{aligned} \tag{G.4}$$

Here, J_μ is the conserved current defined by the *global* $U(1)$ symmetry of the matter Lagrangian: $\Phi \to \exp(-\mathrm{i}\alpha)\Phi$, $\Phi^* \to \exp(\mathrm{i}\alpha)\Phi^*$.

Note that the theory defined in this way differs from the usual $U(1)$ gauge theory: it is invariant under gauge transformations $\varphi_\mu \to \varphi_\mu + \partial_\mu\omega$, while, at the same time, the matter field remains *unchanged*.

A coupling of type (G.4) is represented by a vertex diagram, as in figure G.1: the straight lines correspond to the matter field, wavy lines to the photon and the vertex Γ (in the momentum representation) is obtained from λJ_μ by the replacement $\mathrm{i}\partial_\mu\Phi \to p_\mu$, $\mathrm{i}\partial_\mu\Phi^* \to -p_\mu$:

$$\Gamma^\mu = \lambda(p_2^\mu + p_1^\mu) = \lambda(2p_2^\mu + k^\mu). \tag{G.5a}$$

In the soft photon limit, when the photon momentum k is much smaller than the matter field momentum p, the vertex Γ^μ takes the effective form

$$f^\mu = \Gamma^\mu(k=0) = 2\lambda p_2^\mu. \tag{G.5b}$$

This limit is sufficient to define the electric charge as the soft photon coupling constant: $e = \lambda$. On-shell (for $p_2^2 = m^2$), f^μ represents the classical particle current.

The vertex of the spin-2 graviton. The massless spin-2 field $\varphi_{\mu\nu}$ is coupled to the matter energy–momentum tensor $T_{\mu\nu}$, so that the complete Lagrangian is given by

$$\mathcal{L} = \mathcal{L}_\mathrm{T} + \mathcal{L}_\mathrm{M} + \mathcal{L}_\mathrm{I} \tag{G.6}$$

where $\mathcal{L}_\mathrm{T}$ and $\mathcal{L}_\mathrm{M}$ are determined by expressions (7.19) and (G.2), and

$$\begin{aligned} \mathcal{L}_\mathrm{I} &= -\lambda\varphi^{\mu\nu}T_{\mu\nu} \\ T_{\mu\nu} &\equiv \partial_\mu\Phi^*\partial_\nu\Phi + \partial_\nu\Phi^*\partial_\mu\Phi - \eta_{\mu\nu}\mathcal{L}_\mathrm{M}. \end{aligned} \tag{G.7}$$

The vertex $\Gamma^{\mu\nu}$ has the form

$$\Gamma^{\mu\nu} = \lambda[p_2^\mu p_1^\nu + p_2^\nu p_1^\nu - \eta^{\mu\nu}(p_1 \cdot p_2 - m^2)] \tag{G.8a}$$

where $p_2 = p_1 - k$. In the soft graviton limit ($k \to 0$) the coupling of the graviton to the external matter field line ($p_2^2 - m^2 = 0$) yields the effective vertex

$$f^{\mu\nu} = \Gamma^{\mu\nu}(k = 0, p_2^2 = m^2) = 2\lambda p_2^\mu p_2^\nu \tag{G.8b}$$

which corresponds to the classical particle energy–momentum. The form of the soft graviton vertex defines the gravitational charge: $\kappa = \lambda$.

The general structure of the vertex. These properties of the photon and graviton vertex are obtained for the scalar matter, but can be easily generalized to an arbitrary matter field with spin $s > 0$.

The general form of the current for a complex boson field $\Phi_{\lambda\sigma\ldots}$ follows from expression (G.4) by replacing $\Phi \to \Phi_{\lambda\sigma\ldots}$ (up to a sign, which is the same for even, and opposite for odd spins, as a consequence of the alternation of signs of the Lagrangians with growing spin). The transition to the momentum space reproduces result (G.5a), whereupon the soft photon limit yields the effective vertex (G.5b). Similarly, the tensor $T_{\mu\nu}$ is obtained (up to a sign) from (G.7) by $\Phi \to \Phi_{\lambda\sigma\ldots}$. Going over to the momentum space, the term proportional to $\mathcal{L}_\mathrm{M}$ tends to zero as $k \to 0$, $p^2 \to m^2$, and we end up with result (G.8b).

The same result also holds for arbitrary spinor matter fields.

Low-energy theorems and the spin of the graviton. The general considerations in chapter 7 show that the massless graviton can have spin 0 or spin 2, while the possibility of spin 1 is eliminated. What about higher spins? We shall see that the low-energy theorems in quantum field theory eliminate the possibility $s > 2$.

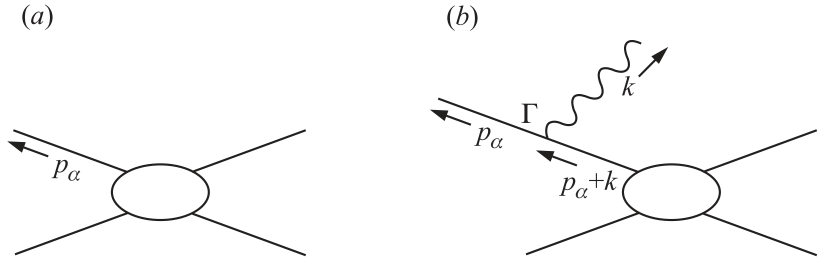

Figure G.2. The modified process (*b*) contains an additional graviton.

Let us consider a scattering process described by an amplitude M (figure G.2(*a*)), and compare it with a slightly modified process, in which an additional graviton of momentum k is emitted (or absorbed) (figure G.2(*b*)). When $k \to 0$, the dominant diagrams are those in which the graviton is 'attached' to an external matter field line, as they possess an additional propagator (between the vertex Γ and the rest of the diagram) which is close to its pole. The additional propagator, corresponding to a line carrying momentum $p_\alpha + k$, becomes, in the limit $k \to 0$,

$$D(p_\alpha + k, m^2) = \frac{1}{(p_\alpha + k)^2 - m^2} \approx \frac{1}{2p_\alpha \cdot k}.$$

We write a detailed form of the original amplitude M, with all spinor indices and external line momenta explicitly indicated, as

$$M \to M_{\sigma_1 \ldots \sigma_\alpha \ldots \sigma_n}(p_1, \ldots, p_\alpha, \ldots, p_n).$$

Consider, first, the case of the *spin-1 graviton*. The modified process is described by the matrix element $\epsilon_\mu M^\mu_{(\alpha)}$, where

$$\begin{aligned} M^\mu_{(\alpha)} &= \Gamma^\mu(p_\alpha, k) D(p_\alpha + k, m^2) M(p_\alpha \to p_\alpha + k) \\ &\equiv \sum_{\sigma_\beta} \Gamma^\mu_{\sigma_\alpha \sigma_\beta}(p_\alpha, k) D(p_\alpha + k, m^2) M_{\sigma_1 \ldots \sigma_\beta \ldots \sigma_n}(p_\alpha \to p_\alpha + k) \end{aligned} \tag{G.9a}$$

and $\Gamma^\mu(p_\alpha, k)$ is a vertex of type (G.5*a*), which describes the coupling of the graviton to the line $(\sigma_\alpha, p_\alpha)$. In the limit $k \to 0$, we find that

$$M^\mu_{(\alpha)} \approx \frac{f^\mu_\alpha}{2p_\alpha \cdot k} M \tag{G.9b}$$

where $f^\mu_{(\alpha)} = \Gamma^\mu(p_\alpha, 0)$. After summing the contributions of the gravitons emitted from all external lines (sum over α), the gauge invariance requirement

(G.1), applied to the complete amplitude $M^\mu = \sum_\alpha M^\mu_{(\alpha)}$, yields

$$\sum_\alpha \frac{k_\mu f^\mu_{(\alpha)}}{2p_\alpha \cdot k} M = 0. \tag{G.10}$$

Since we are considering only soft gravitons, the vertex $f^\mu_{(\alpha)}$ is of the form (G.5*b*), $f^\mu_{(\alpha)} = 2e_\alpha p^\mu_\alpha$, so that this requirement reduces to

$$\sum_\alpha e_\alpha = 0 \tag{G.11}$$

provided $M \neq 0$. This is, of course, the charge conservation law.

Now, we turn our attention to the *spin-2 graviton*. A completely analogous procedure leads to the result

$$\sum_\alpha \frac{k_\mu f^{\mu\nu}_{(\alpha)}}{2p_\alpha \cdot k} M = 0. \tag{G.12}$$

In the limit $k \to 0$, the vertex takes the form (G.8*b*), $f^{\mu\nu}_{(\alpha)} = 2\kappa_\alpha p^\mu_\alpha p^\nu_\alpha$, and the previous condition, for $M \neq 0$, reduces to

$$\sum_\alpha \kappa_\alpha p^\nu_\alpha = 0. \tag{G.13}$$

Taking into account the momentum conservation law, $\sum_\alpha p^\nu_\alpha = 0$ (all external momenta are directed outwards), relation (G.13) can only be satisfied by choosing

$$\kappa_\alpha = \kappa \tag{G.14}$$

where κ is a constant. Thus, all particles have the same gravitational charge. This demonstrates the *universality* of the gravitational coupling, in accordance with the principle of equivalence.

It should be noted that the graviton field $\varphi_{\mu\nu}$ can be coupled to its own energy–momentum tensor. In other words, the graviton carries the gravitational charge (energy–momentum) and can interact with itself. This is not the case with the photon: the electric charge of the photon vanishes and the photon is electrically neutral.

For gravitons with higher spins, the condition of gauge invariance leads to equations that cannot be satisfied at all (for instance, $\sum_\alpha g_\alpha p^\mu_\alpha p^\nu_\alpha = 0$). Thus, we conclude that

the only massless bosons which can be consistently coupled to matter in the limit $k \to 0$ are those of spins $s \leq 2$.

Since the expressions ep^μ and $\kappa p^\mu p^\nu$ should remain invariant if we replace $p \to -p$ and reinterpret the particle line as an antiparticle line, it follows that the particle and antiparticle must have opposite coupling to the spin-1 field, and the same coupling to the spin-2 field.

We should emphasize that these arguments do not show that massless fields of spin $s > 2$ do not exist, but only that their interaction with matter in the limiting case $k \to 0$ must vanish ($M \to 0$), otherwise it will be inconsistent. Such interactions do not generate long-range static forces, hence they are not expected to be of any importance in describing macroscopic gravitational phenomena.

Appendix H

Lorentz group and spinors

In classical relativistic physics, the basic physical quantities are represented by tensors and the covariance of the physical laws is achieved with the help of tensor equations which have the same form in every reference frame. On the other hand, quantum physics is formulated in terms of both tensor and spinor fields. This motivates us to reconsider the role of relativistic spinors in classical field theory (Carruthers 1971, Novozhilov 1972, Wybourne 1973, Barut and Raczka 1977, Berestetskii *et al* 1982). We should note that Lorentz spinors, which are used to build relativistic actions, are not directly related to physical particles, having definite mass and spin; the physical particles are properly described by the representations of the Poincaré group (see appendix I).

Consider two observers who are in different inertial reference frames S and S' of the Minkowski space M_4. We assume that they describe the same physical event by the spacetime coordinates x and x', respectively, which are related to each other by a Lorentz transformation:

$$x'^{\mu} = \Lambda^{\mu}{}_{\nu} x^{\nu}. \tag{H.1}$$

The invariance of the spacetime interval (2.1) leads to the orthogonality conditions on the real matrix $\Lambda = (\Lambda^{\mu}{}_{\nu})$: $\Lambda^T \eta \Lambda = \eta$. This implies

$$\det \Lambda = \pm 1$$
$$\Lambda^{00} \geq 1 \qquad \text{or} \qquad \Lambda^{00} \leq -1.$$

If Λ is any invertible 4×4 matrix with real elements, we introduce the following terminology:

the full Lorentz group: $L \equiv O(1,3) = \{\Lambda \mid \Lambda^T \eta \Lambda = \eta\}$;
the proper Lorentz group: $L_+ \equiv SO(1,3) = \{\Lambda \in L \mid \det \Lambda = +1\}$;
the restricted Lorentz group: $L_0 = \{\Lambda \in L \mid \det \Lambda = +1, \Lambda^{00} \geq +1\}$.

The discrete transformations of space inversion I_P and time inversion I_{T} are improper, i.e. such that $\det \Lambda = -1$. The full Lorentz group contains L_0,

$I_P L_0$, $I_T L_0$ and $I_P I_T L_0$. The proper and restricted Lorentz transformations are subgroups of L. The restricted Lorentz group L_0 (Lorentz group) is the only subgroup in which the elements are continuously connected to the identity; it contains ordinary three-dimensional rotations and boosts, but not space nor time inversions.

Lie algebra. An infinitesimal transformation of the restricted Lorentz group L_0 can be written as $\Lambda^{\mu}{}_{\nu} = \delta^{\mu}_{\nu} + \omega^{\mu}{}_{\nu}$, where $\omega^{\mu\nu} = -\omega^{\nu\mu}$, or, in matrix notation,

$$\Lambda = 1 + \tfrac{1}{2}\omega^{\lambda\rho} M_{\lambda\rho} \qquad (M_{\lambda\rho})^{\mu}{}_{\nu} = \delta^{\mu}_{\lambda}\eta_{\rho\nu} - \delta^{\mu}_{\rho}\eta_{\lambda\nu}.$$

Here, $M_{\lambda\rho}$ are the infinitesimal *generators* of L_0 which satisfy the Lie algebra

$$\begin{aligned} [M_{\mu\nu}, M_{\lambda\rho}] &= \eta_{\nu\lambda}M_{\mu\rho} - \eta_{\mu\lambda}M_{\nu\rho} - \eta_{\nu\rho}M_{\mu\lambda} + \eta_{\mu\rho}M_{\nu\lambda} \\ &\equiv \tfrac{1}{2} f_{\mu\nu,\lambda\rho}{}^{\sigma\tau} M_{\sigma\tau}. \end{aligned} \tag{H.2a}$$

The matrices $M_{\mu\nu}$ are anti-Hermitian, $(M^{+})_{\mu\nu} = -M_{\mu\nu}$, and equation (H.2$a$) *defines* the Lie algebra of the Lorentz group. The problem of finding representations of the Lorentz algebra leads to solutions for $M_{\mu\nu}$ which are different from the particular form from which we started.

We now introduce the angular momentum and boost generators,

$$\begin{aligned} M^{i} &= \tfrac{1}{2}\varepsilon^{ijk}M_{jk} = (M_{23}, M_{31}, M_{12}) \\ K^{i} &= M_{0i} = (M_{01}, M_{02}, M_{03}) \end{aligned}$$

which obey the commutation rules

$$\begin{gathered} [M^{i}, M^{j}] = \varepsilon^{ijk}M^{k} \qquad [M^{i}, K^{j}] = \varepsilon^{ijk}K^{k} \\ [K^{i}, K^{j}] = -\varepsilon^{ijk}K^{k}. \end{gathered} \tag{H.2b}$$

The first equation defines the algebra $so(3)$ of the rotation subgroup $SO(3)$ of L_0, while the second one states that K^{i} is a vector with respect to these rotations. The minus sign in the third equation expresses the difference between the non-compact Lorentz group $SO(1,3)$ and its compact version $SO(4)$. This seemingly small difference in sign leads to essential differences in the structure of two groups, with physically important consequences.

Finite-dimensional representations. The form of the finite Lorentz transformations in M_4 shows that the four-dimensional representation is not unitary: rotations are represented by unitary and boosts by non-unitary matrices. The same is true for every finite-dimensional irreducible representation (IR), in accordance with the following general theorem:

> *All (non-trivial) unitary IRs of a connected, simple, non-compact Lie group must be of infinite dimension.*

In covariant field theories of particle physics, finite-dimensional representations of the Lorentz group are needed to describe the spin degrees of freedom. In order to classify these representations, we introduce the *complex* linear combinations of the generators:

$$J_1^i = \tfrac{1}{2}(M^i - \mathrm{i}K^i) \qquad J_2^i = \tfrac{1}{2}(M^i + \mathrm{i}K^i).$$

Their algebra decomposes into the direct sum of two $so(3)$ algebras:

$$[J_1^i, J_1^j] = \varepsilon^{ijk} J_1^k \qquad [J_1^i, J_2^j] = 0$$
$$[J_2^i, J_2^j] = \varepsilon^{ijk} J_2^k.$$

Linear combinations with complex coefficients are not allowed in $so(1,3)$—in fact they define the complexified Lie algebra $so(1,3)^{\mathrm{c}}$. However, since there is a 1–1 correspondence between representations of a complex Lie algebra and representations of any of its real forms, we can use the classification of representations of $so(3)_1 \oplus so(3)_2$ to classify the representations of $so(1,3)$. Consequently, the finite-dimensional IRs of the Lorentz algebra can be labelled by a pair of half-integer or integer numbers,

$$(j_1, j_2) \qquad j_1, j_2 = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \ldots \tag{H.3a}$$

which correspond to the Casimir operators for the two $so(3)$ subalgebras: $\boldsymbol{J}_1{}^2 = -j_1(j_1+1)$ and $\boldsymbol{J}_2{}^2 = -j_2(j_2+1)$. Since $\boldsymbol{M} = \boldsymbol{J}_1 + \boldsymbol{J}_2$, the total angular momentum, or *spin*, of the representation may take the values

$$j = |j_1 - j_2|, |j_1 - j_2| + 1, \ldots, j_1 + j_2 \tag{H.3b}$$

and the dimension of the representation space is $d = (2j_1+1)(2j_2+1)$. Here are some examples:

$(0, 0)$, the scalar representation, $j = 0, d = 1$;
$(\frac{1}{2}, 0)$, the left-handed spinor representation, $j = \frac{1}{2}, d = 2$;
$(0, \frac{1}{2})$, the right-handed spinor representation, $j = \frac{1}{2}, d = 2$;

The importance of the basic representations $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$ lies in the fact that any other finite-dimensional representation of the Lorentz algebra can be generated from these two. Thus, for instance, the direct product of these representations, $(\frac{1}{2}, 0) \times (0, \frac{1}{2}) = (\frac{1}{2}, \frac{1}{2})$, gives a four-vector representation with spin $j = 0, 1$, and $d = 4$ (any four-vector, such as x^μ, belongs to this representation). Finite-dimensional representations, in general, have no definite spin.

The universal covering group. The Lie algebra only defines the local structure of a given Lie group. For every Lie group there is a Lie algebra, but the inverse is not true. In order to understand the importance of Lie algebras for the structure of Lie groups, we have to clarify the global (topological) properties of Lie groups sharing the same Lie algebra.

Consider a family $\mathcal{G}$ of multiply connected Lie groups which are locally isomorphic (have the same Lie algebra). We can show that in each family $\mathcal{G}$ there is a unique (up to an isomorphism) simply connected group $\bar{G}$, known as the universal covering group of the family $\mathcal{G}$, such that $\bar{G}$ can be homomorphically mapped onto any group G in $\mathcal{G}$. The group $\bar{G}$ contains a discrete invariant subgroup Z such that every group G is locally isomorphic to the factor group $\bar{G}/Z$.

An arbitrary Lorentz transformation Λ in L_0 can be represented as the combination of a boost and a three-dimensional rotation. Since the space of parameters of the rotation group is *double* connected, the same is also true for L_0. We shall show that the universal covering group of L_0 is $SL(2, C)$, the group of complex 2×2 matrices with a unit determinant.

We start by introducing the notation

$$\sigma^\mu = (1, \boldsymbol{\sigma}) \qquad \bar{\sigma}^\mu = (1, -\boldsymbol{\sigma})$$

where $\boldsymbol{\sigma}$ are the Pauli matrices,

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma^2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \qquad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

satisfying $\sigma^a \sigma^b = \mathrm{i}\varepsilon^{abc}\sigma^c + \delta^{ab}$. The matrices σ^μ and $\bar{\sigma}^\mu$ obey the identities

$$\mathrm{Tr}(\sigma^\mu \bar{\sigma}^\nu) = 2\eta^{\mu\nu} \qquad \sigma^\mu \bar{\sigma}^\nu + \sigma^\nu \bar{\sigma}^\mu = 2\eta^{\mu\nu}.$$

Now, we show that L_0 is homomorphic to $SL(2, C)$. First, we construct a mapping from M_4 to the set of Hermitian complex 2×2 matrices:

$$x_\mu \to X = x_\mu \sigma^\mu = \begin{pmatrix} x_0 + x_3 & x_1 - \mathrm{i}x_2 \\ x_1 + \mathrm{i}x_2 & x_0 - x_3 \end{pmatrix} \qquad \det X = x^2.$$

Then, we consider the action of $SL(2, C)$ on X defined by

$$X' = AXA^+ \qquad A \in SL(2, C). \tag{H.4a}$$

This action conserves $x^2 = \det X$ and represents, basically, a Lorentz transformation. Indeed, to each such transformation A in $SL(2, C)$, there corresponds a Lorentz transformation of x^μ defined by

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu \qquad \Lambda^\mu{}_\nu(A) = \tfrac{1}{2}\,\mathrm{Tr}(\bar{\sigma}^\mu A \sigma_\nu A^+) \tag{H.4b}$$

where $\sigma_\nu = \eta_{\nu\lambda}\sigma^\lambda$. The mapping $A \to \Lambda(A)$ is a homomorphism of $SL(2, C)$ on L_0, i.e.

$$\begin{aligned} &\Lambda^\mu{}_\nu(A) \in L_0 \\ &\Lambda^\mu{}_\rho(A_1 A_2) = \Lambda^\mu{}_\nu(A_1)\Lambda^\nu{}_\rho(A_2). \end{aligned}$$

Since $\Lambda(-A) = \Lambda(A)$, the inverse mapping is defined only up to a sign:

$$A = \pm \frac{1}{[\det(\Lambda^\rho{}_\lambda \sigma_\rho \bar{\sigma}^\lambda)]^{1/2}} \Lambda^\mu{}_\nu \sigma_\mu \bar{\sigma}^\nu. \tag{H.4c}$$

This follows from the relation $\Lambda^\rho{}_\nu \sigma_\rho \bar{\sigma}^\nu = A\sigma_\nu A^+ \bar{\sigma}^\nu = 2A\,\mathrm{Tr}(A^+)$.

For the infinitesimal $SL(2,C)$ transformations,

$$A = 1 + \tfrac{1}{2}\omega_{\mu\nu}\sigma^{\mu\nu} \qquad A^{*-1T} = 1 + \tfrac{1}{2}\omega_{\mu\nu}\bar{\sigma}^{\mu\nu} \tag{H.5a}$$

the relation (H.4c) implies

$$\begin{aligned} \sigma^{\mu\nu} &= \tfrac{1}{4}(\sigma^\mu \bar{\sigma}^\nu - \sigma^\nu \bar{\sigma}^\mu) = [-\tfrac{1}{2}\boldsymbol{\sigma}, -\tfrac{1}{2}\mathrm{i}\boldsymbol{\sigma}] \\ \bar{\sigma}^{\mu\nu} &= \tfrac{1}{4}(\bar{\sigma}^\mu \sigma^\nu - \bar{\sigma}^\nu \sigma^\mu) = [\tfrac{1}{2}\boldsymbol{\sigma}, -\tfrac{1}{2}\mathrm{i}\boldsymbol{\sigma}] \end{aligned} \tag{H.5b}$$

where the rotation and boost parts of the generator are displayed in the square brackets: $\sigma^{\mu\nu} = [\frac{1}{2}\varepsilon^{abc}\sigma^{bc}, \sigma^{0a}]$, and similarly for $\bar{\sigma}^{\mu\nu}$. These generators satisfy the Lorentz algebra (H.2).

In conclusion,

(a) there is a homomorphism $SL(2,C) \to L_0$, for which the kernel is the discrete invariant subgroup $Z_2 = (1, -1)$ (the elements $A = \pm 1$ in $SL(2,C)$ are mapped into the identity element in L_0):

$$L_0 = SL(2,C)/Z_2$$

(b) the groups $SL(2,C)$ and L_0 have the same Lie algebra.

The group $SL(2,C)$ is simply connected, and represents the universal covering group $\bar{L}$ of L_0.

The importance of the universal covering group $\bar{L}$ lies in the fact that

- all its IRs are *single*-valued, whereas those of L_0 may be both single- and *double*-valued (L_0 is double connected); and
- every IR of L_0 is a single-valued representations of $\bar{L}$.

Bearing in mind this connection, we continue the study of $\bar{L} = SL(2,C)$.

Two-component spinors. Let us now consider several important representations of $SL(2,C)$. A two-component spinor is a pair of complex numbers ξ_a ($a = 1, 2$) transforming under $SL(2,C)$ according to the rule

$$\xi'_a = A_a{}^b \xi_b. \tag{H.6a}$$

The spinors ξ_a are elements of a complex two-dimensional vector space, the representation space of $SL(2,R)$ such that, to each Lorentz transformation in M_4, there corresponds an $SL(2,C)$ transformation of ξ:

$$\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \qquad \xi' = A\xi.$$

Since the mapping $SL(2, C) \to L_0$ is a two-to-one mapping, the spinors are only defined up to a sign. The spinors ξ_a are called the left-handed Weyl spinors and they belong to the representation ($\frac{1}{2}$,0).

We introduce spinors η^a, dual to ξ_a, by demanding that the bilinear forms $\eta^a \xi_a$ remain invariant under the action of $SL(2, C)$. Consequently,

$$\eta'^a = A^{-1T\,a}{}_b \eta^b. \tag{H.6b}$$

The representations (H.6*a*) and (H.6*b*) are *equivalent* since

$$A = gA^{-1T}g^{-1} \qquad g = -g^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \mathrm{i}\sigma^2$$

or, using the index notation,

$$g_{ab} = \varepsilon_{ab} \qquad g^{-1ab} \equiv g^{ab} = -\varepsilon^{ab} \qquad \varepsilon_{12} = \varepsilon^{12} = 1.$$

The inverse matrix g^{-1} is denoted simply as g^{ab}. Using $\eta^a = g^{ab}\eta_b$ and $\xi_a = g_{ab}\xi^b$ we find that

$$\eta^a \xi_a = g^{ab}\eta_b g_{ac}\xi^c = -\eta_a \xi^a.$$

The matrices g and g^{-1} are used to raise and lower the spinor indices, and play the role of a metric and its inverse in the space of spinors ξ_a and η^a. The metric tensor is an invariant tensor:

$$\varepsilon'_{ab} = A_a{}^c A_b{}^d \varepsilon_{cd} = \varepsilon_{ab} \det A = \varepsilon_{ab}.$$

The complex conjugate spinor ξ^*_a transforms as

$$\xi^{*\prime}_a = A^*{}_a{}^b \xi^*_b$$

where the matrix A^* is the complex-conjugate of A. Writing the components of ξ^*_a as $\bar{\xi}_{\dot{a}}$ and the matrix elements of A^* as $A^*{}_{\dot{a}}{}^{\dot{b}}$, we have

$$\bar{\xi}'_{\dot{a}} = A^*{}_{\dot{a}}{}^{\dot{b}} \bar{\xi}_{\dot{b}}. \tag{H.7a}$$

The components $\bar{\xi}_{\dot{a}}$ define the right-handed Weyl spinor, transforming according to the representation (0,$\frac{1}{2}$). Since, in general, there is no linear connection between A and A^* of the form $A = CA^*C^{-1}$, the spinors ξ_a and $\bar{\xi}_{\dot{a}}$ transform according to *inequivalent* representations. If A is a unitary matrix, as in the case of space rotations, then $A^* = A^{-1T}$, and the spinors $\bar{\xi}_{\dot{a}}$ and ξ^a transform equivalently.

Finally, we introduce spinors $\bar{\eta}^{\dot{a}}$, dual and equivalent to $\bar{\xi}_{\dot{a}}$:

$$\bar{\eta}'^{\dot{a}} = A^{*-1T\,\dot{a}}{}_{\dot{b}} \bar{\eta}^{\dot{b}}. \tag{H.7b}$$

The matrix $\bar{g}$ is of the form

$$\bar{g}_{\dot{a}\dot{b}} = \varepsilon_{\dot{a}\dot{b}} \qquad \bar{g}^{-1\dot{a}\dot{b}} \equiv \bar{g}^{\dot{a}\dot{b}} = -\varepsilon^{\dot{a}\dot{b}} \qquad \varepsilon_{\dot{1}\dot{2}} = \varepsilon^{\dot{1}\dot{2}} = 1$$

and we have $\bar{\eta}^{\dot{a}}\bar{\xi}_{\dot{a}} = -\bar{\eta}_{\dot{a}}\bar{\xi}^{\dot{a}}$.

Thus, we have defined two-dimensional *elementary* spinors $\xi_a, \bar{\xi}_{\dot{a}}$ and the related dual spinors $\eta^a, \bar{\eta}^{\dot{a}}$, which transform according to the $SL(2, C)$ representations A, A^* and $A^{-1T}, A^{*-1T} = A^{-1+}$, respectively. Multiplying these spinors with each other we obtain *higher* spinors, which transform under $SL(2, C)$ as the appropriate product of elementary spinors. The rank of a higher spinor is determined by a pair of numbers (k, l)—the numbers of undotted and dotted indices.

Contraction of the same type of upper and lower indices, with the help of g or $\bar{g}$, lowers the rank for two. The contraction of symmetric indices yields zero, which is related to the fact that symmetric spinors form IRs of $SL(2, R)$. A symmetric spinor of rank (k, l) has $(k+1)(l+1)$ independent components.

Spinors and four-vectors. We now investigate some elements of the spinor calculus, and establish a direct connection between transformations in spinor space and Lorentz transformations in spacetime.

We postulate that the components of spinors are *anticommuting* variables, in accordance with the spin-statistics theorem. Since the matrix elements of A and A^+ are given by $A = (A_a{}^b)$ and $A^+ = (A^{+\dot{a}}{}_{\dot{b}})$, the relations $X' = AXA^+$, $X = x_\mu \sigma^\mu$ and $\mathrm{Tr}(\sigma^\mu \bar{\sigma}^\nu) = 2\eta_{\mu\nu}$ imply that $X = (X_{a\dot{b}})$, $\sigma^\mu = (\sigma^\mu_{a\dot{b}})$ and $\bar{\sigma}^\mu = (\bar{\sigma}^{\mu\dot{b}a})$.

Let us now find x_μ from the equation $X = x_\mu \sigma^\mu$, and replace $X_{a\dot{b}}$ with the quantity $\xi_a \bar{\xi}_{\dot{b}}$, which has the same transformation properties:

$$x^\mu = \tfrac{1}{2}\,\mathrm{Tr}(X\bar{\sigma}^\mu) \sim \tfrac{1}{2}\xi_a \bar{\xi}_{\dot{b}} (\bar{\sigma}^\mu)^{\dot{b}a} = -\tfrac{1}{2}\bar{\xi}\bar{\sigma}^\mu \xi .$$

Then, the completeness relation

$$\sigma^\mu_{a\dot{b}} \bar{\sigma}^{\dot{c}d}_\mu = 2\delta^d_a \delta^{\dot{c}}_{\dot{b}}$$

implies that the right-hand side of this equation transforms exactly as a four-vector under $SL(2, C)$:

$$V^\mu \equiv \bar{\xi}\bar{\sigma}^\mu \xi \qquad \xi \to A\xi \quad \Rightarrow \quad V^\mu \to \Lambda^\mu{}_\nu V^\nu . \tag{H.8}$$

The matrix $\bar{\sigma}^\mu$ serves to convert the spinor product $\bar{\xi}_{\dot{a}}\xi_b$ into a four-vector. Equation (H.8) lies at the root of the whole subject of relating Lorentz transformations in spacetime with spin-space transformations.

In order to simplify the notation, we shall often omit the summed spinor indices and write

$$\xi\eta = \xi^a \eta_a \qquad \bar{\xi}\bar{\eta} = \bar{\xi}_{\dot{a}}\bar{\eta}^{\dot{a}}$$

in agreement with the index structure of σ^μ, $\bar\sigma^\mu$: $\xi\sigma^\mu\bar\eta = \xi^a\sigma^\mu{}_{a\dot b}\bar\eta^{\dot b}$, etc.

Consider now some useful properties of σ^μ and $\bar\sigma^\mu$. Using the result (H.4*a*) we find

$$\Lambda_\mu{}^\nu\sigma^\mu = A\sigma^\nu A^+ \quad\Longrightarrow\quad \sigma^\mu = \Lambda^\mu{}_\nu A\sigma^\nu A^+. \tag{H.9a}$$

This relation shows that $\sigma^\mu_{a\dot b}$ can be interpreted as both a four-vector and a mixed spinor and that, moreover, this quantity is an invariant object (the effects of spinor and four-vector transformations cancel each other).

Raising both spinor indices in $\sigma^\mu_{a\dot b}$, we find

$$\sigma^{\mu a\dot b} \equiv [-g^{-1}(1,\boldsymbol\sigma)\bar g^{-1}]^{a\dot b} = (1,-\boldsymbol\sigma^*)^{a\dot b}.$$

Then, taking the complex conjugation, we obtain

$$(\sigma^{\mu a\dot b})^* = (1,-\boldsymbol\sigma)^{\dot a b} \equiv \bar\sigma^{\mu\dot a b}.$$

The matrix $\bar\sigma^{\mu\dot a b}$ is numerically equal to $\sigma_{\mu a\dot b}$, but has a different spinor structure; it is denoted by $\bar\sigma^\mu$. From $g^{-1}Ag = A^{-1T}$ and our previous considerations, we find one more useful relation:

$$\bar\sigma^\mu = \Lambda^\mu{}_\nu A^{-1+}\bar\sigma^\nu A^{-1}. \tag{H.9b}$$

On the other hand, $\boldsymbol\sigma^* = \boldsymbol\sigma^T$ implies $\sigma^{\mu a\dot b} = \bar\sigma^{\mu\dot b a}$.

Space inversion and Dirac spinors. The space inversion I_P acts on the Lorentz generators according to the rule $M^i \to M^i$, $K^i \to -K^i$, hence it transforms J_1^i into J_2^i and *vice versa*, I_P: $(j_1, j_2) \to (j_2, j_1)$.

Thus, for instance, $I_P : (\frac12, 0) \leftrightarrow (0, \frac12)$, by $\xi_a \to z_p\bar\eta^{\dot a}$ and $\bar\eta \to z_p\xi$, where z_p is a phase factor. Two applications of the space inversion should return a spinor to its original configuration, which may be thought of as a rotation for 0 or 2π. Since basic spinors change sign under the rotation for 2π, there are two possible definitions of the space inversion, with

$$z_p^2 = 1 \qquad\text{or}\qquad z_p^2 = -1.$$

In the first case we may choose $z_p = +1$:

$$I_P\xi_a = \bar\eta^{\dot a} \qquad I_P\bar\eta^{\dot a} = \xi_a$$

which implies $I_P\xi^a = -\bar\eta_{\dot a}$, $I_P\bar\eta_{\dot a} = -\xi^a$. On the other hand, for $z_p^2 = -1$ we may choose $z_p = +i$:

$$I_P\xi_a = \mathrm{i}\bar\eta^{\dot a} \qquad I_P\bar\eta^{\dot a} = \mathrm{i}\xi_a$$

whereas $I_P\xi^a = -\mathrm{i}\bar\eta_{\dot a}$, $I_P\bar\eta_{\dot a} = -\mathrm{i}\xi^a$.

The only difference in physical consequences between the two definitions would occur for truly neutral spin-$\frac12$ fields (Majorana spinors), for which particles are identical with antiparticles: in the case $z_p^2 = 1$, the condition of neutrality would not be consistent with the space inversion symmetry. We choose $z_p = \mathrm{i}$ in order to include Majorana spinors in our discussion.

As we have seen, the space inversion cannot be represented within a single basic spinor representation. In order to have an IR for a symmetry which includes I_P and $SL(2, C)$, it is necessary to consider a pair (the direct sum) of spinors ξ_a and $\bar{\eta}^{\dot{a}}$. Such a pair is called the Dirac spinor or four-spinor and may be written in the form

$$\psi = \begin{pmatrix} \xi \\ \bar{\eta} \end{pmatrix}. \tag{H.10a}$$

It transforms under $SL(2, C)$ as

$$\psi' = S(A)\psi = \begin{pmatrix} A & 0 \\ 0 & A^{-1+} \end{pmatrix} \begin{pmatrix} \xi \\ \bar{\eta} \end{pmatrix} = \begin{pmatrix} A\xi \\ A^{-1+}\bar{\eta} \end{pmatrix} \tag{H.10b}$$

whereas the space inversion is represented by

$$I_P\psi = \mathrm{i}\begin{pmatrix} \bar{\eta} \\ \xi \end{pmatrix} \qquad \text{i.e.} \qquad I_P = \mathrm{i}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{H.10c}$$

The Dirac equation. The concept of two-spinors provides a simple and systematic way of deriving $SL(2, C)$ covariant equations. The basic objects from which typical physical equations are constructed are spinor *fields*, partial derivatives and some parameters. The spinor structure of these equations ensures their covariance.

Consider a simple example of a linear differential spinor equation, which illustrates the general method and, at the same time, represents physically important case of the free Dirac equation. Using the spinor fields $\xi(x)$, $\bar{\eta}(x)$ and the operators $p = \sigma^\mu p_\mu$, $\bar{p} = \bar{\sigma}^\mu p_\mu$, ($p_\mu = \mathrm{i}\partial_\mu$), the conditions of linearity and covariance lead to the pair of coupled equations

$$\bar{p}^{\dot{a}b}\xi_b = m\bar{\eta}^{\dot{a}} \qquad p_{a\dot{b}}\bar{\eta}^{\dot{b}} = m\xi_a \tag{H.11a}$$

where m is the mass parameter. These equations are clearly covariant provided the spinor fields transform as

$$\xi'(x') = A\xi_b(x) \qquad \bar{\eta}'(x') = A^{*-1T}\bar{\eta}(x).$$

In the four-dimensional notation, these equations take the form

$$\begin{pmatrix} 0 & \sigma^\mu p_\mu \\ \bar{\sigma}^\mu p_\mu & 0 \end{pmatrix} \begin{pmatrix} \xi \\ \bar{\eta} \end{pmatrix} = m \begin{pmatrix} \xi \\ \bar{\eta} \end{pmatrix}$$

which is equivalent to the usual Dirac equation:

$$\gamma^\mu p_\mu \psi = m\psi \tag{H.11b}$$

where γ^μ are the Dirac matrices in the spinor representation,

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \bar{\sigma}^\mu & 0 \end{pmatrix}. \tag{H.12a}$$

The Dirac matrices satisfy the algebra

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu} \tag{H.12b}$$

and have different equivalent representations.

Covariant bilinear combinations of the type $\bar{\psi}\psi$ are defined with the help of the Dirac adjoint spinor $\bar{\psi} = \psi^+\gamma^0$, and not $\psi^+ = (\xi^*, \bar{\eta}^*)$. The reason for this lies in the fact that γ^0 interchanges ξ^* and $\bar{\eta}^*$ in $\bar{\psi}$, so that two-spinors in $\bar{\psi} = (\bar{\eta}^*, \xi^*)$ are of the correct type (index structure), which leads to the covariantly defined expression $\bar{\psi}\psi = \bar{\eta}^*\xi + \xi^*\bar{\eta}$.

Multiplying equation (H.11b) with $\gamma \cdot p$, we find that $\psi(x)$ obeys the free Klein–Gordon equation, $(\Box + m^2)\psi = 0$. Thus, the Dirac field, as the direct sum of two basic spinors, describes two physical spin-$\frac{1}{2}$ particles (in fact, a particle and an antiparticle) of equal mass m.

Majorana and Weyl spinors. The Dirac equation makes sense even when η is not distinct from ξ, i.e. when $\bar{\eta}^{\dot{a}} = \bar{\xi}_c^{\dot{a}}$, where $\bar{\xi}_c$ is obtained from ξ by the *charge conjugation* operation:

$$\xi_a \to \bar{\xi}_c^{\dot{a}} = \bar{g}^{\dot{a}\dot{b}}\bar{\xi}_{\dot{b}} \qquad \bar{\xi}_{\dot{b}} \equiv (\xi_b)^*.$$

We similarly define $\bar{\eta}^{\dot{a}} \to \eta_a^c = g_{ab}\eta^b$, $\eta^b \equiv (\bar{\eta}_b)^*$, and, consequently,

$$\psi \to \psi_c = \begin{pmatrix} \eta_a^c \\ \bar{\xi}_c^{\dot{b}} \end{pmatrix} = C\bar{\psi}^T \qquad C \equiv \mathrm{i}\gamma^2\gamma^0 = \begin{pmatrix} \mathrm{i}\sigma^2 & 0 \\ 0 & -\mathrm{i}\sigma^2 \end{pmatrix} \tag{H.13}$$

where C is the charge conjugation matrix.

Charge conjugation is a discrete, non-spacetime operation, which transforms particles into antiparticles and *vice versa*: $\psi \to \psi_c$, $\psi_c \to (\psi_c)_c = \psi$. The condition $\psi = \psi_c$ defines the Majorana spinor, a truly neutral field (with the particles identical to the antiparticles) which has only two independent complex components.

In the case $m = 0$, the Dirac equation splits into separate equations for ξ and $\bar{\eta}$, called the Weyl equations. They describe left- and right-handed massless fields, which transform into each other under the space inversion, and are especially important when treating neutrinos. In the four-component notation Weyl spinors ξ and $\bar{\eta}$ may be represented as chiral projections

$$\psi_\mp = P_\mp\psi \qquad \psi_- = \begin{pmatrix} \xi \\ 0 \end{pmatrix} \qquad \psi_+ = \begin{pmatrix} 0 \\ \bar{\eta} \end{pmatrix} \tag{H.14a}$$

where $P_\mp$ are the projection operators:

$$P_\mp = \tfrac{1}{2}(1 \pm \mathrm{i}\gamma_5) \qquad \gamma_5 = \gamma_0\gamma_1\gamma_2\gamma_3 = -\mathrm{i}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{H.14b}$$

Exercises

1. Show that the Lorentz transformations of coordinates in M_4 that describe a rotation by angle θ around the x^1-axis and a boost with velocity $v = \tanh\varphi$ along the x^1-axis, are described by the following matrices:

$$R_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\theta & \sin\theta \\ 0 & 0 & -\sin\theta & \cos\theta \end{pmatrix}$$

$$L_1 = \begin{pmatrix} \cosh\varphi & -\sinh\varphi & 0 & 0 \\ -\sinh\varphi & \cosh\varphi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

2. Demonstrate that the mapping (H.4*b*) is a homomorphism of $SL(2, C)$ on L_0, and derive equation (H.4*c*).
3. Show that the generators M^i and K^i in the spinor representations $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$ have the form

$$r_1(M^1) = \sigma_{23} = -\tfrac{1}{2}\mathrm{i}\sigma^1 \qquad r_1(K^1) = \sigma_{01} = \tfrac{1}{2}\sigma^1$$
$$r_2(M^1) = \bar\sigma_{23} = -\tfrac{1}{2}\mathrm{i}\sigma^1 \qquad r_2(K^1) = \bar\sigma_{01} = -\tfrac{1}{2}\sigma^1$$

   and similarly for the other components.
4. Show that a finite rotation of the spinor ξ by angle $\omega_{23} = \theta$ around the x^1-axis is described by the matrix

$$A = \mathrm{e}^{-\frac{1}{2}\mathrm{i}\theta\sigma^1} = \cos\frac{\theta}{2} - \mathrm{i}\sigma^1 \sin\frac{\theta}{2}$$

   which is unitary, $A^+ = A^{-1}$, and obeys the conditions $A(0) = 1$, $A(2\pi) = -1$. Find the related transformations of the four-vector x^μ in M_4.
5. Show that a finite boost of the spinor $\bar\xi$ with velocity $v = \tanh\varphi$ ($\omega_{01} = \varphi$) along the x^1-axis is described by the matrix

$$A = \mathrm{e}^{-\frac{1}{2}\varphi\sigma^1} = \cosh\frac{\varphi}{2} - \sigma^1 \sinh\frac{\varphi}{2}$$

   which is Hermitian, $A^+ = A$. Find the related transformations of the four-vector x^μ in M_4.
6. Prove the following relations:

$$\sigma^\mu_{a\dot b}\bar\sigma_\mu^{\dot c d} = 2\delta_a^d\delta_{\dot b}^{\dot c} \qquad \text{(the completeness relation)}$$
$$\mathrm{Tr}(G\sigma^\mu)\,\mathrm{Tr}(H\bar\sigma_\mu) = 2\,\mathrm{Tr}(GH).$$

7. Find the transformation laws of $\xi\sigma^\mu\bar\eta$ and $\xi\sigma^{\mu\nu}\eta$ under $SL(2, C)$.

8. Show that anticommuting spinors obey the following identities:

$$\theta_a\theta_b = \tfrac{1}{2}\varepsilon_{ab}(\theta\theta) \qquad (\xi\eta)^* = (\bar{\eta}\bar{\xi})$$
$$\xi\eta = \eta\xi \qquad \bar{\xi}\bar{\eta} = \bar{\eta}\bar{\xi} \qquad (\xi\sigma^\mu\bar{\eta})^* = (\eta\sigma^\mu\bar{\xi})$$
$$\xi\sigma^\mu\bar{\eta} = -\bar{\eta}\bar{\sigma}^\mu\xi \qquad [(\sigma^\mu\bar{\eta})_a]^* = (\eta\sigma^\mu)_{\dot{a}}.$$

The complex conjugation is defined so as to reverse the order of spinors.

9. Prove the following Fierz identities:

$$(\theta\xi)(\theta\eta) = -\tfrac{1}{2}(\theta\theta)(\xi\eta) \qquad (\xi\sigma^\mu\bar{\eta})(\bar{\sigma}_\mu\eta)^{\dot{a}} = -2(\xi\eta)\bar{\eta}^{\dot{a}}$$
$$2\xi^a\bar{\eta}^{\dot{b}} = (\xi\sigma^\mu\bar{\eta})\bar{\sigma}_\mu^{\dot{b}a} \qquad (\xi_1\xi_2)(\bar{\eta}_1\bar{\eta}_2) = -\tfrac{1}{2}(\xi_1\sigma^\mu\bar{\eta}_1)(\bar{\eta}_2\bar{\sigma}_\mu\xi_2).$$

10. Show that the Majorana condition $\psi_c = \psi$ is consistent with the space inversion symmetry provided we choose $I_P = \mathrm{i}\gamma^0$. What happens if $I_P = \gamma^0$?

Appendix I

Poincaré group and massless particles

The finite dimensional IRs of the Lorentz group studied in appendix H are non-unitary. These representations cannot be associated with elementary physical objects, because the invariants of an IR do not correspond directly to the physical invariants of elementary particles, the mass and the spin (the same conclusion also holds for infinite dimensional unitary representations). Only when we add translations and enlarge the Lorentz group to the whole Poincaré group, do we obtain proper description of the elementary particles.

In this appendix we give a short review of the Poincaré group, with an emphasis on the properties which characterize the massless particles and gauge symmetries in field theory (Carruthers 1971, Novozhilov 1972, Van Dam 1974, Barut and Raczka 1977).

The most general linear transformation of coordinates between two inertial frames in M_4 is obtained by composing a Lorentz transformation and a translation:

$$x'^{\mu} = \Lambda^{\mu}{}_{\nu} x^{\nu} + a^{\mu}. \tag{I.1a}$$

The resulting ten-parameter group is called the Poincaré group P. Denoting the elements of P by $g = (\Lambda, a)$, the group multiplication law reads:

$$(\Lambda_2, a_2)(\Lambda_1, a_1) = (\Lambda_2\Lambda_1, a_2 + \Lambda_2 a_1). \tag{I.1b}$$

This shows that P is not a direct product of the Lorentz transformations and translations (this structure is known as the semidirect product).

The Lie algebra. In order to find the Lie algebra of the Poincaré group, consider a scalar field $\phi(x)$ which transforms under P according to

$$\phi'(x') = \phi(x) \qquad \text{or} \qquad \phi'(x) = \phi(\Lambda^{-1}(x-a)) \equiv U(g)\phi(x)$$

where $U(g)$ represents the action of P on scalar fields. For infinitesimal Poincaré transformations the operator U takes the form

$$U(g) = 1 + \tfrac{1}{2}\omega^{\mu\nu} L_{\mu\nu} + a^{\mu} P_{\mu}$$

where $M_{\mu\nu} = x_\mu \partial_\nu - x_\nu \partial_\mu$ and $P_\mu = -\partial_\mu$ are Poincaré generators in the particular representation, which satisfy the following Lie algebra:

$$
\begin{aligned}
[M_{\mu\nu}, M_{\lambda\rho}] &= \eta_{\nu\lambda} M_{\mu\rho} - \eta_{\mu\lambda} M_{\nu\rho} - \eta_{\nu\rho} M_{\mu\lambda} + \eta_{\mu\rho} M_{\nu\lambda} \\
[M_{\mu\nu}, P_\lambda] &= \eta_{\nu\lambda} P_\mu - \eta_{\mu\lambda} P_\nu \qquad [P_\mu, P_\nu] = 0.
\end{aligned}
\tag{I.2}
$$

Note that there exist representations in which $M_{\mu\nu}$ and P_μ differ from these expressions, obtained in the space of scalar fields.

Very general considerations in quantum field theory require the symmetry operators, which relate the quantum states in different reference frames, to be either unitary or anti-unitary. If the symmetry group is continuously connected to the identity, the symmetry operators are *unitary*, which is why we are going to consider only unitary representations of P. Since P is simple, connected and non-compact, all its unitary IRs are necessarily infinite-dimensional.

The universal covering group. Let P_0 be the subgroup of the Poincaré group which contains the elements of the restricted Lorentz group L_0 and translations. It is continuously connected to the identity, and also double connected, like L_0. Denote the related universal covering group by $\bar{P}$. Starting from the fact that the universal covering group of L_0 is $SL(2, C)$, we find that a general element of $\bar{P}$ has the form $\bar{g} = (A, a)$, where $A \in SL(2, C)$, $a = a_\mu \sigma^\mu$. The composition law takes the form

$$
(A_2, a_2)(A_1, a_1) = (A_2 A_1, a_2 + A_2 a_1 A_2^+). \tag{I.3a}
$$

From this, we obtain the important relation

$$
(A, a) \equiv (1, a)(A, 0) = (A, 0)(1, A^{-1} a A^{-1+}) \tag{I.3b}
$$

which will be useful for constructing unitary IRs of the Poincaré group in terms of the related representations of translations $(1, a)$ and Lorentz transformations $(A, 0)$.

For infinitesimal transformations $\bar{g}$, the operator $U(\bar{g})$, which represents the group $\bar{P}$ in some representation space, is close to the identity:

$$
U(\bar{g}) = 1 + \tfrac{1}{2} \omega^{\mu\nu} M_{\mu\nu} + a^\mu P_\mu. \tag{I.4}
$$

The generators $M_{\mu\nu}$ and P_μ satisfy the Lie algebra (I.2), in accordance with the local isomorphism of $\bar{P}$ and P_0.

The invariants. Using the generators P_μ and $M_{\mu\nu}$, we now construct invariant operators the eigenvalues of which may be used to characterize the IRs of $\bar{P}$. It is useful to introduce the notation $P_\mu = \mathrm{i} p_\mu$ and $M_{\mu\nu} = \mathrm{i} m_{\mu\nu}$, where p_μ and $m_{\mu\nu}$ are Hermitian in unitary representations and have a more direct physical interpretation.

There is only one general invariant composed from the momentum components:

$$p_\mu p^\mu = m^2. \tag{I.5a}$$

For $m^2 \geq 0$, the quantity m is interpreted as the rest mass of a system. In this case, we can define another invariant function of the momentum components:

$$\epsilon = \text{the sign of } p_0 \tag{I.5b}$$

which is positive on physical states. The invariance of ϵ is a consequence of the fact that P_0 is orthochronous ($\Lambda^0{}_0 > 0$).

From the generators P_μ and $M_{\mu\nu}$, we can construct the four-vector

$$W_\mu \equiv \tfrac{1}{2}\varepsilon_{\mu\nu\lambda\sigma} M^{\nu\lambda} P^\sigma \tag{I.5c}$$

which commutes with P_μ. Its square is also an invariant object. The physical meaning of W^2 is easily seen for $m^2 > 0$: in the rest frame, the invariant W^2/m^2 is equal to the square of the angular momentum $\boldsymbol{M}^2$, i.e. to the square of the intrinsic spin. When $m^2 = 0$, the square W^2/m^2 is replaced by the helicity λ, as we shall see.

The unitary IRs of $\bar{P}$ are labelled by the invariants (m^2, W^2 or λ, ϵ), which describe elementary particle states in quantum field theory.

The little group. The Hilbert space of quantum field theory is a vector space in which the action of $\bar{P}$ is realized in terms of the operators $U(\bar{g})$, composed from quantum fields. Vectors within an IR carry labels which characterize the invariant physical properties of particle states (i.e. mass and spin). The set of all momentum states $\{|\boldsymbol{p}, s\rangle\}$, where s labels different spin states, can be chosen as the basis of an IR. The momentum eigenstates satisfy $\hat{p}_\mu|\boldsymbol{p}, s\rangle = p_\mu|\boldsymbol{p}, s\rangle$, where we have used the hat to denote quantum operators. Every momentum eigenstate realizes a one-dimensional representation of the translations:

$$U(0, a)|\boldsymbol{p}, s\rangle = \exp(-\mathrm{i}p \cdot a)|\boldsymbol{p}, s\rangle. \tag{I.6a}$$

Hence, in order to construct IRs of the whole $\bar{P}$ it is sufficient to find the related representations of the Lorentz transformations $U(A, 0) \equiv U(A)$, as follows from (I.3b).

Since an $SL(2, C)$ transformation A maps p^μ into p'^μ, the action of $U(A)$ on $|\boldsymbol{p}, s\rangle$ produces a state of momentum p'^μ:

$$U(A)|\boldsymbol{p}, s\rangle = \sum_{s'} |\boldsymbol{p}', s'\rangle D_{s's}(A) \qquad p'^\mu = \Lambda^\mu{}_\nu(A) p^\nu \tag{I.6b}$$

where D is a matrix acting on spin indices. We will specify this matrix using the so-called little group method.

A Hermitian matrix $p = p_\mu \sigma^\mu$ transforms according to $p' = ApA^+$. Consider the elements of $SL(2, C)$ that leave p invariant:

$$p = \tilde{A} p \tilde{A}^+. \tag{I.7}$$

The form of $\tilde{A}$ depends on p, $\tilde{A} = \tilde{A}(p)$, and the set of all matrices $\tilde{A}(p)$ forms a subgroup of $SL(2, C)$—the *little group* $L(p)$ associated with the momentum p.

If momenta p and $\mathring{p}$ are related by a Lorentz transformation,

$$p = \alpha(p, \mathring{p})\mathring{p}\,\alpha^+(p, \mathring{p})$$

the little groups $L(p)$ and $L(\mathring{p})$ are isomorphic and

$$\tilde{A}(p) = \alpha(p, \mathring{p})\tilde{A}(\mathring{p})\alpha^{-1}(p, \mathring{p}).$$

Thus, for a given $p^2 = m^2$, it is sufficient to consider the little group of a particular momentum $\mathring{p}$, the *standard momentum*, which may be chosen arbitrarily.

The operator $\alpha(p, \mathring{p})$ is called the *Wigner operator*. Using the relations

$$p' = ApA^+ \qquad p = \alpha(\boldsymbol{p})\mathring{p}\,\alpha^+(\boldsymbol{p}) \qquad p' = \alpha(\boldsymbol{p}')\mathring{p}\,\alpha^+(\boldsymbol{p}')$$

where $\alpha(\boldsymbol{p}) \equiv \alpha(p, \mathring{p})$, we find that an arbitrary matrix A in $SL(2, C)$ may be represented as

$$A = \alpha(\boldsymbol{p}')\tilde{A}(\mathring{p})\alpha^{-1}(\boldsymbol{p}). \tag{I.8}$$

The operator $\alpha(\boldsymbol{p})$ is not uniquely defined, since $\alpha(\boldsymbol{p})$ and $\alpha(\boldsymbol{p})\tilde{A}(\mathring{p})$ produce the same effect on $\mathring{p}$. However, if we fix $\alpha(\boldsymbol{p})$, the relation between A and $\tilde{A}(\mathring{p})$ becomes unique.

Let us now clarify the role of the little group in constructing IRs of the Lorentz group. The relation $|\boldsymbol{p}, s\rangle = U[\alpha(\boldsymbol{p})]|\mathring{\boldsymbol{p}}, s\rangle$ implies

$$U(A)|\boldsymbol{p}, s\rangle = U(A)U[\alpha(\boldsymbol{p})]|\mathring{\boldsymbol{p}}, s\rangle.$$

Since $A(p)\alpha(\boldsymbol{p}) = \alpha(\boldsymbol{p}')\tilde{A}(\mathring{p})$, we find

$$\begin{aligned} U(A)|\boldsymbol{p}, s\rangle &= U[\alpha(\boldsymbol{p}')]U[\tilde{A}(\mathring{p})]|\mathring{\boldsymbol{p}}, s\rangle \\ &= U[\alpha(\boldsymbol{p}')]\sum_{s'}|\mathring{\boldsymbol{p}}, s'\rangle D_{s's}[\tilde{A}(\mathring{p})] = \sum_{s'}|\boldsymbol{p}', s'\rangle D_{s's}[\tilde{A}(\mathring{p})]. \end{aligned}$$

Therefore, in order to find the IRs of the Lorentz transformations $U(A)$, it is sufficient to find the matrix D only for the standard momentum $\mathring{p}$.

It is now easy to conclude that a unitary IR of $\bar{P}$ has the form

$$U(A, a)|\boldsymbol{p}, s\rangle = \exp(-\mathrm{i}p' \cdot a)\sum_{s'}|\boldsymbol{p}', s'\rangle D_{s's}[\tilde{A}(\mathring{p})] \qquad p' = \Lambda(A)p \tag{I.9}$$

where $D_{s's}$ is a unitary IR of the little group for the standard momentum.

Unitarity and irreducibility of the representations of $\bar{P}$ are directly related to the same properties of the little group.

According to the value of p^2, there are four classes of representations of $\bar{P}$: (1) $p^2 > 0$, (2) $p^2 < 0$, (3) $p^2 = 0$ and (4) $p = 0$ (the first and third cases correspond to physical states). We now give a detailed account of the third case, which describes massless particle states.

Unitary representations for $m^2 = 0$. The representations on massless states are particularly interesting for a better understanding of the structure of gauge theories. Since $p^2 = 0$, the standard momentum may be chosen as $\mathring{p}^{\mu} = \omega(1, 0, 0, 1)$, so that $\mathring{p} = \omega(1 - \sigma^3)$. The little group of $\mathring{p}$ is defined by

$$\tilde{A}(1 - \sigma^3)\tilde{A}^{+} = 1 - \sigma^3.$$

For infinitesimal $\tilde{A}$, this equation yields

$$\begin{gathered} \tilde{A} = t(\varepsilon)u(\theta) \\ t(\varepsilon) = 1 + \tfrac{1}{2}(\sigma^1 - \mathrm{i}\sigma^2)\varepsilon \qquad u(\theta) = 1 - \tfrac{1}{2}\mathrm{i}\sigma^3\theta \end{gathered} \tag{I.10}$$

where ε is a complex and θ a real parameter. If we write $\varepsilon = \varepsilon_1 + \mathrm{i}\varepsilon_2$, it follows that $u(\theta)$ and $t(\varepsilon)$ define the Euclidean group $E(2)$—the group of rotations and translations in the Euclidean plane E_2.

The Wigner operator is determined from the relation

$$\alpha(\boldsymbol{p})\omega(1 - \sigma^3)\alpha^{+}(\boldsymbol{p}) = |\boldsymbol{p}|\sigma^0 + p_{\alpha}\sigma^{\alpha} \equiv p_{\mu}\sigma^{\mu}.$$

The meaning of the little group can be clarified by looking at the form of $U(\bar{g})$ for infinitesimal $\bar{g}$. Replacing the infinitesimal transformations (I.10) in (H.4b), we find the following values for $\omega^{\mu}{}_{\nu} = \Lambda^{\mu}{}_{\nu} - \delta^{\mu}{}_{\nu}$:

$$\omega^{12}(\theta) = \theta \qquad \omega^{02}(\varepsilon) = \omega^{32}(\varepsilon) = \varepsilon_2 \qquad \omega^{01}(\varepsilon) = \omega^{31}(\varepsilon) = \varepsilon_1$$

so that $U(\bar{g})$ takes the form

$$\begin{gathered} U(\theta, \varepsilon) = 1 + \theta M_{12} + \varepsilon_1 E_1 + \varepsilon_2 E_2 \\ E_1 \equiv M_{01} + M_{31} \qquad E_2 \equiv M_{02} + M_{32}. \end{gathered} \tag{I.11}$$

The little group algebra generated by E_1, E_2 and M_{12} is isomorphic to the Lie algebra of $E(2)$:

$$[E_1, E_2] = 0 \qquad [M_{12}, E_1] = E_2 \qquad [M_{12}, E_2] = -E_1. \tag{I.12}$$

IRs of the group $E(2)$ are labelled by the eigenvalues $\boldsymbol{t} = (t_1, t_2)$ and λ of the generators $-\mathrm{i}\boldsymbol{E} = -\mathrm{i}(E_1, E_2)$ and $-\mathrm{i}M_{12} = m_{12}$, respectively. Hence, the state vectors are $|\boldsymbol{p}, \boldsymbol{t}, \lambda\rangle$.

For $\boldsymbol{t}^2 > 0$, the 'momentum' of $E(2)$ takes continuous eigenvalues which differ from zero, and we have infinite dimensional IRs. They should be interpreted as particles of infinite spin. Since no such particles are observed in nature, these representations are of no physical interest.

For $\boldsymbol{t}^2 = 0$, the group $E(2)$ effectively acts as a group of rotations in the Euclidean plane, i.e. as $U(1)$. Its representations are *one-dimensional*, and are labelled by real, integer or half-integer eigenvalues λ:

$$U(\bar{g}) \to \exp(\mathrm{i}\lambda\theta) \qquad (\lambda = 0, \pm\tfrac{1}{2}, \pm 1, \ldots).$$

In the representation $|\boldsymbol{p}, \boldsymbol{t}, \lambda\rangle$, the four-vector W_μ can be written as $W_\mu = \mathrm{i}\omega(-M_{12}, -E_2, E_1, M_{12})$. For $\boldsymbol{t}^2 = 0$, we find that W_μ is collinear with p_μ:

$$W_\mu = \lambda p_\mu.$$

The quantity λ is the eigenvalue of the helicity operator $\lambda = \boldsymbol{mp}/|\boldsymbol{p}|$. The change of sign of λ defines, in general, a different particle state (which may not exist physically). The absolute value of λ is called the spin of the zero mass particle.

Thus, we conclude that massless states are classified by the eigenvalues λ of the helicity operator, and have the transformation law

$$U(a, A)|\boldsymbol{p}, \lambda\rangle = \exp(-ip \cdot a)\exp(i\lambda\theta)|\boldsymbol{p}, \lambda\rangle. \tag{I.13}$$

It is known that the photon has two helicity states, $\lambda = \pm 1$. This means that the photon is described by a reducible representation of $\bar{P}$: under space inversion, the states with $\lambda = \pm 1$ transform into one another. These two states form an IR of $\bar{P}$ together with I_P.

Massless states and covariance. If we demand space inversion symmetry, a massless particle of spin 1 is described by two helicity states. Can these two states be considered to be the components of one four-vector? In order to examine this suggestion, let us first write the generators of the little group for $\overset{\circ}{p}{}^{\mu} = \omega(1, 0, 0, 1)$ in the four-vector representation:

$$E_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \qquad E_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$M_{12} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

where we have used $(M_{\mu\nu})^\lambda{}_\rho = \delta^\lambda_\mu \epsilon_{\nu\rho} - \delta^\lambda_\nu \epsilon_{\mu\rho}$. The states of helicity $\lambda = \pm 1$, the eigenstates of $m_{12} = -\mathrm{i}M_{12}$, have the form

$$\epsilon_{(\pm 1)} = \frac{1}{\sqrt{2}}(\epsilon_{(1)} \pm \mathrm{i}\epsilon_{(2)})$$

where $\epsilon_{(1)} = (0, 1, 0, 0)$ and $\epsilon_{(2)} = (0, 0, 1, 0)$. They transform into one another under space inversion. However, the four-vectors $\epsilon_{(1)}$ and $\epsilon_{(2)}$ are not invariant under $E(2)$ translations:

$$\epsilon_{(a)} \to \epsilon_{(a)} + t'_a(1, 0, 0, 1) = \epsilon_{(a)} + t_a \overset{\circ}{p} \qquad a = 1, 2.$$

Hence, they cannot be used to describe two polarization states of the photon. These states, however, can be described as the *equivalence classes* of these four-vectors,

$$\{\epsilon_{(1)} + t_1 \overset{\circ}{p}\} \qquad \{\epsilon_{(2)} + t_2 \overset{\circ}{p}\}$$

which *are* invariant under $E(2)$ translations. For the general momentum p^μ, the polarization states of the photon are described by the equivalence classes

$$\begin{gathered}\{\epsilon_{(1)} + t_1 p\} \qquad \{\epsilon_{(2)} + t_2 p\} \\ p \cdot \epsilon_{(1)} = p \cdot \epsilon_{(2)} = \epsilon_{(1)} \cdot \epsilon_{(2)} = 0.\end{gathered} \tag{I.14}$$

Two four-vectors in the same equivalence class are related to each other by a *gauge* transformation: $\epsilon^\mu \to \epsilon'^\mu = \epsilon^\mu + tp^\mu$.

In the Lagrangian formulation, the dependence of the theory on the classes of equivalence, and not on the individual four-vectors ϵ, is ensured by the gauge invariance:

(a) the Lagrangian of the free electromagnetic field depends only on the gauge invariant combination $p^\mu \epsilon^\nu - p^\nu \epsilon^\mu$; and
(b) the interaction with matter is of the form $\epsilon^\mu J_\mu$, where J_μ is the conserved current.

Let us now try, in analogy with the photon, to describe a massless particle with helicities $\lambda = \pm 2$, the graviton, by a symmetric tensor of rank 2. Choosing the same $\overset{\circ}{p}$, the generators of the little group take the form

$$(E_1)^{ij}{}_{kl} = (E_1)^i{}_k \delta^j_l + (E_1)^j{}_l \delta^i_k$$

and similarly for E_2 and M_{12}. Here, $(E_1)^i{}_k$ is the four-vector representation of E_1. The linear polarization states are described by the tensors

$$\epsilon_{(1)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \epsilon_{(2)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

which are obtained from the symmetrized product $\epsilon^\mu_\lambda \epsilon^\nu_{(\lambda')}$ by subtracting the trace. Demanding the invariance under $E(2)$ translations leads to the following definition of the classes of equivalence:

$$\begin{gathered}\{\epsilon^{\mu\nu}_{(1)}(p) + p^\mu t^\nu_1 + p^\nu t^\mu_1 \mid p \cdot t_1 = 0\} \\ \{\epsilon^{\mu\nu}_{(2)}(p) + p^\mu t^\nu_2 + p^\nu t^\mu_2 \mid p \cdot t_2 = 0\} \\ p_\mu \epsilon^{\mu\nu}_{(\lambda)} = 0 \qquad \epsilon^{\mu\nu}_{(\lambda)} = \epsilon^{\nu\mu}_{(\lambda)} \qquad \eta_{\mu\nu}\epsilon^{\mu\nu}_{(\lambda)} = 0.\end{gathered} \tag{I.15}$$

The gauge invariance of the theory is ensured by demanding that

(a) the Lagrangian of the free graviton is gauge invariant; and
(b) the tensor $\epsilon^{\mu\nu}$ interacts with a symmetric, conserved current $J_{\mu\nu}$.

Fields and states. We have seen in appendix H that Lorentz transformations of the *classical* Dirac field have the form

$$\psi'(x') = S(A)\psi(x) \qquad x' = \Lambda x$$

where the matrix $S(A)$ realizes a four-dimensional *non-unitary* representation of $\bar{L}$. This is a typical situation for an arbitrary spinor field. We should note that since we are dealing with *fields* there is a dependence on x in the transformation law, so that the complete transformation defines, in fact, an infinite dimensional representation. Spinor fields are used to construct covariant field equations, whereby these fields acquire definite mass and spin. On the other hand, we have also seen that infinite dimensional *unitary* representations of the Poincaré group $\bar{P}$ in the Hilbert space of states describe particle states of definite mass and spin. Is there any direct relation between these two results?

In quantum field theory, we have both physical states and quantum field operators. A field operator is a spinor field, an object which, in general, does not carry a definite spin. There is an important result of quantum field theory which gives an algorithm for constructing the field operators of definite spin, such that they are subject to no additional conditions at all. The construction is realized with the help of the creation and annihilation operators corresponding to one-particle states, and provides a direct bridge between the representations on states and on fields.

The same result can also be simply illustrated at the classical level. Consider a four-vector field $A^\mu(x)$, transforming under $\bar{P}$ according to

$$A'^\mu(x') = \Lambda^\mu{}_\nu A^\nu(x) \qquad x' = \Lambda x + a.$$

From the behaviour of the field under Lorentz transformations, we know that it carries spin $j = 0, 1$. Let us now impose the conditions

$$\partial_\mu A^\mu = 0 \qquad (\Box + m^2)A^\mu = 0$$

with $m \neq 0$. The first condition eliminates the spin component $j = 0$, whereas the second one ensures that the field has mass m. Indeed, in the rest frame, the first condition has the form $p_\mu A^\mu = mA^0 = 0$, i.e. $A^0 = 0$, and we are left with a three-vector $\boldsymbol{A}$ which carries spin $j = 1$. In this way, we have constructed a field of spin $j = 1$ and mass m starting from a four-vector A^μ. Both of these conditions can be realized by adopting the following free field equation:

$$\partial_\mu(\partial^\mu A^\nu - \partial^\nu A^\mu) + m^2 A^\nu = 0.$$

Differentiating this equation we obtain $\partial \cdot A = 0$, whereupon the equation itself reduces to $(\Box + m^2)A^\mu = 0$.

The previous example illustrates that

relativistic equations for free fields represent kinematical conditions which ensure that a given field has a definite mass and spin.

Usually field equations are obtained from the action principle.

Exercises

1. (a) Use the multiplication law (I.1b) to prove the relations:

$$g^{-1}(\Lambda, a) = g(\Lambda^{-1}, -\Lambda^{-1}a)$$
$$g(\Lambda)T_a = T_{\Lambda a}g(\Lambda)$$
$$g^{-1}(\Lambda)g(\Lambda', a)g(\Lambda) = g(\Lambda^{-1}\Lambda'\Lambda, \Lambda^{-1}a)$$

   where $T_a \equiv g(1, a)$, $g(\Lambda) \equiv g(\Lambda, 0)$.

   (b) Derive the transformation properties of the Poincaré generators P_μ and $M_{\mu\nu}$ under $g(\Lambda)$, using the last relation with $\Lambda' = 1$ and $a = 0$, respectively.

   (c) Find the algebra of the Poincaré generators.
2. Derive the Lie algebra of the universal covering group $\bar{P}$ using the composition rule (I.3a).
3. Consider the set of scalar fields $\phi(x)$ satisfying the Klein–Gordon equation $(\Box + m^2)\phi = 0$. Show that the operators $L_{\mu\nu} = x_\mu\partial_\nu - x_\nu\partial_\mu$ and $P_\mu = -\partial_\mu$ are anti-Hermitian within the scalar product

$$(\phi_1, \phi_2) = \mathrm{i}\int \mathrm{d}^3x\, [\phi_1^*(x)\partial_0\phi_2(x) - \partial_0\phi_1^*(x)\phi_2(x)].$$

4. Show that the four-vector W_μ, defined in (I.5c), satisfies the following relations:

$$W^2 = M^{\nu\lambda}M_{\nu\rho}P_\lambda P^\rho - \tfrac{1}{2}M^{\nu\lambda}M_{\nu\lambda}P^2$$
$$W_\nu = [P_\nu, C] \qquad C \equiv \tfrac{1}{8}\varepsilon_{\mu\nu\lambda\rho}M^{\mu\nu}M^{\lambda\rho}$$
$$[P_\mu, W_\nu] = 0 \qquad [M_{\mu\nu}, W_\lambda] = \eta_{\nu\lambda}W_\mu - \eta_{\mu\lambda}W_\nu \qquad [M_{\mu\nu}, W^2] = 0.$$

5. Find the form of matrix $\tilde{A}$ that describes infinitesimal transformations of the little group for $m^2 = 0$. Then construct the operator $U(\bar{g})$.
6. Show that in the space of massless states $|\mathring{p}\rangle$, with the standard momentum $\mathring{p}^\mu = \omega(1, 0, 0, 1)$, the components of W_μ satisfy the commutation relations

$$[W_1, W_2] = 0 \qquad [M_{12}, W_1] = W_2 \qquad [M_{12}, W_2] = -W_1.$$

7. (a) Verify the following commutation rules in the space of massless states:

$$[m_{12}, W_\pm] = \mp W_\pm \qquad [W_+, W_-] = 0 \qquad [W^2, m_{12}] = 0$$

where $W_\pm = W_1 \pm i W_2$ and $m_{12} = -i M_{12}$.

(b) Find the spectrum of the operators m_{12} and W^2.

8. Show that in the case $p^2 > 0$, the standard momentum may be chosen in the form $\overset{\circ}{p} = (m, 0, 0, 0)$ and the little group is $SO(3)$.

Appendix J

Dirac matrices and spinors

In this appendix, we review basic conventions and properties of the Dirac matrices and spinors, which are used in discussing supersymmetry and supergravity (Sohnius 1985, Srivastava 1986, van Nieuwenhuizen 1981).

Dirac matrices. Dirac matrices in a flat space of dimension d are defined as IRs of the algebra

$$\{\gamma_i, \gamma_j\} = 2\eta_{ij}. \tag{J.1}$$

> *In even dimensions IRs of the Dirac algebra are complex $n \times n$ matrices, where $n = 2^{d/2}$. All IRs are equivalent, i.e. any two representations γ_i and γ_i' are connected by the relation $\gamma_i = S\gamma_i' S^{-1}$, where S is a non-singular matrix.*

If γ_i is an IR of the Dirac algebra, then

$$\pm\gamma_i \qquad \pm\gamma_i^+ \qquad \pm\gamma_i^T \qquad \pm\gamma_i^*$$

are also IRs. On the basis of this theorem, we can introduce non-singular matrices A and C, such that

$$A\gamma_i A^{-1} = \gamma_i^+ \qquad C^{-1}\gamma_i C = -\gamma_i^T. \tag{J.2}$$

Going now to the four-dimensional Minkowski space ($d = 4, n = 4$), we define the γ_5 matrix,

$$\gamma_5 \equiv \gamma_0\gamma_1\gamma_2\gamma_3 \qquad \{\gamma_5, \gamma_i\} = 0 \qquad \gamma_5^2 = -1 \tag{J.3}$$

which anticommutes with all γ_i. The matrices A, C and γ_5 allow us to transform any two of these representations into each other. Thus, for instance,

$$D^{-1}\gamma_i D = -\gamma_i^* \qquad D \equiv CA^T.$$

Applying the Hermitian conjugation, transposition and complex conjugation to the relations defining the action of A, C and D, respectively, we get

$$A = \alpha A^{+} \qquad C = \eta C^{T} \qquad DD^{*} = \delta$$

where $\alpha\alpha^{*} = \eta^{2} = 1$ and $\delta = \delta^{+}$, for consistency.

By a redefinition $A \to \omega A$, with a suitable choice of phase and modulus of a complex number ω, we can obtain $\alpha = |\delta| = 1$. The signs of δ and η cannot be changed: they are determined by the dimension and the metric of the space. In M_4, the matrix C is necessarily antisymmetric, $C^{T} = -C$, in order for us to have ten symmetric $(\gamma_i C, \sigma_{ij} C)$ and six antisymmetric $(C, \gamma_5 C, \gamma_5\gamma_i C)$, linearly independent 4×4 matrices. Using any particular representation of gamma matrices, we can show that $\delta = 1$.

We summarize here the important properties of A, C and γ_5 in M_4:

$$\begin{aligned} &A = A^{+} \qquad (A\gamma_i)^{+} = A\gamma_i \qquad (A\sigma_{ij})^{+} = -A\sigma_{ij} \\ &C = -C^{T} \qquad (\gamma_i C)^{T} = \gamma_i C \qquad (\sigma_{ij} C)^{T} = \sigma_{ij} C \\ &\gamma_5^2 = -1 \qquad A\gamma_5 A^{-1} = \gamma_5^{+} \qquad C^{-1}\gamma_5 C = \gamma_5^{T} \end{aligned} \tag{J.4}$$

where $\sigma_{ij} \equiv \frac{1}{4}[\gamma_i, \gamma_j]$. Also,

$$\begin{aligned} &\gamma_i C, \sigma_{ij} C && \text{are symmetric} \\ &C, \gamma_5 C, \gamma_5\gamma_i C && \text{are antisymmetric.} \end{aligned}$$

The matrices $A, A\gamma_i, A\gamma_5\gamma_i$ are Hermitian, whereas $A\gamma_5, A\sigma_{ij}$ are anti-Hermitian.

The commutation relations between σ_{mn} and γ_l are the same as those between the Poincaré generators $M_{\mu\nu}$ and P_λ:

$$\begin{aligned} [\sigma_{mn}, \gamma_l] &= \eta_{nl}\gamma_m - \eta_{ml}\gamma_n \\ [\sigma_{mn}, \sigma_{lr}] &= \eta_{nl}\sigma_{mr} - \eta_{ml}\sigma_{nr} + \eta_{mr}\sigma_{nl} - \eta_{nr}\sigma_{ml}. \end{aligned}$$

This is so because γ_i transforms according to the n-dimensional representation $(\frac{1}{2},\frac{1}{2})$ of $SL(2, C)$, just as P_λ, and σ_{mn} are the Lorentz generators in that space.

The Dirac spinors. Dirac matrices act naturally on the space of complex four-spinors ψ_α. Dirac spinors transform under the Lorentz transformations $x' = \Lambda x$ according to the rule

$$\psi'(x') = S(\Lambda)\psi(x) \qquad S(\Lambda) \equiv \exp(\tfrac{1}{2}\omega^{mn}\sigma_{mn}).$$

The space inversion is represented as $\psi'(x') = \eta_P \gamma^0 \psi(x)$ (we choose $\eta_P = \mathrm{i}$). With the help of A, C and D, we can show that the representations S^{-1+}, S^{-1T}, S^{*} and S are equivalent. We also verify that

$$\gamma^m = \Lambda^m{}_n S\gamma^n S^{-1}.$$

The adjoint and charge-conjugate spinors are defined by

$$\bar{\psi} \equiv \psi^+ A \qquad \psi_c \equiv C\bar{\psi}^T = D\psi^*. \tag{J.5}$$

The adjoint spinor $\bar{\psi}$ transforms as $\bar{\psi}'(x') = \bar{\psi}(x)S^{-1}$, whereas ψ_c transforms in the same way as ψ. The matrix D in M_4 can be chosen so that $DD^* = 1$, i.e. $(\psi_c)_c = \psi$. The fields ψ and ψ_c have opposite charges.

For the anticommuting spinors we have the following identities:

$$(\bar{\psi}M\chi)^+ = \begin{cases} \bar{\chi}M\psi & \text{for } M = 1, \gamma_5, \gamma_i \\ -\bar{\chi}M\psi & \text{for } M = \gamma_5\gamma_i, \sigma_{ij} \end{cases} \tag{J.6}$$

as well as $\bar{\psi}_c\psi_c = \bar{\psi}\psi$ and $\bar{\psi}_c\gamma_i\psi_c = -\bar{\psi}\gamma_i\psi$. The Hermitian conjugation is defined such that the order of spinor factors is reversed: $(\bar{\psi}\chi)^+ = \chi^+\bar{\psi}^+$.

The Majorana spinors. In general, the four complex components of a Dirac spinor are independent. The condition

$$\psi_c = \psi \tag{J.7}$$

defines a Majorana spinor, which has only two independent complex components and describes truly neutral field. The consistency of the definition (J.7) requires $(\psi_c)_c = \psi$, and the choice of the space inversion with $z_p^2 = -1$.

The Majorana condition is incompatible with the $U(1)$ gauge symmetry, $\psi \to \psi' = \mathrm{e}^{\mathrm{i}\alpha}\psi$. However, the chiral $U(1)$ gauge symmetry may be consistently defined:

$$\psi' = \mathrm{e}^{\alpha\gamma_5}\psi \qquad \bar{\psi}' = \bar{\psi}\mathrm{e}^{\alpha\gamma_5}\psi_c' = \mathrm{e}^{\alpha\gamma_5}\psi_c \qquad \bar{\psi}_c' = \bar{\psi}_c\mathrm{e}^{\alpha\gamma_5}.$$

We also note that the Majorana spinors satisfy the identities

$$\bar{\psi}M\chi = \begin{cases} \bar{\chi}M\psi & \text{for } M = 1, \gamma_5, \gamma_5\gamma_i \\ -\bar{\chi}M\psi & \text{for } M = \gamma_i, \sigma_{ij}. \end{cases} \tag{J.8}$$

From this we obtain $\bar{\psi}\gamma_i\psi = \bar{\psi}\sigma_{ij}\psi = 0$, as well as the following reality conditions:

$$(\bar{\psi}M\chi)^+ = \begin{cases} \bar{\psi}M\chi & \text{for } M = 1, \gamma_5, \sigma_{ij} \\ -\bar{\psi}M\chi & \text{for } M = \gamma_i, \gamma_5\gamma_i. \end{cases} \tag{J.9}$$

The chiral spinors. The chiral projections $\psi_\mp$ (left/right) are defined by

$$\psi_\mp \equiv P_\mp\psi \qquad P_\mp \equiv \tfrac{1}{2}(1 \pm \mathrm{i}\gamma_5) \tag{J.10}$$

where $P_\mp$ are chiral projectors that satisfy the conditions

$$P_\mp^T = C^{-1}P_\mp C \qquad P_\pm^+ = AP_\mp A^{-1} \qquad P_\mp\gamma_i = \gamma_i P_\pm.$$

This implies $\bar\psi_\mp = \bar\psi P_\pm$ and, consequently,

$$\bar\psi M\chi = \begin{cases} \bar\psi_+ M\chi_- + \bar\psi_- M\chi_+ & \text{for } M = 1, \gamma_5, \sigma_{ij} \\ \bar\psi_+ M\chi_+ + \bar\psi_- M\chi_- & \text{for } M = \gamma_i, \gamma_5\gamma_i. \end{cases} \tag{J.11}$$

Every chiral spinor carries an IR of the Lorentz group: $\psi_- \to S(\Lambda)\psi_-$, $\psi_+ \to S(\Lambda)\psi_+$. To realize the space inversion we need both ψ_- and ψ_+. Indeed, the choice $I_P = \mathrm{i}\gamma^0$ yields $I_P\psi_\mp = \mathrm{i}\psi_\pm$.

Chiral spinors are used to describe *massless* fields of spin $j = \frac{1}{2}$, for which the chirality and helicity stand in a 1–1 correspondence: $\psi_\mp$ have helicities $\lambda = \mp\frac{1}{2}$. Chiral spinors $\psi_\mp$ carry opposite chiral charges:

$$\psi'_- = \mathrm{e}^{-\mathrm{i}\alpha}\psi_- \qquad \psi'_+ = \mathrm{e}^{\mathrm{i}\alpha}\psi_+.$$

The Majorana condition $\psi = \psi_c$, expressed in terms of chiral spinors, reads as $(\psi_c)_\mp = C\bar\psi^T_\pm$.

The Fierz identities. Consider the complete set of complex 4×4 matrices

$$\Gamma^A = \{1, \mathrm{i}\gamma^5, \gamma^m, \gamma^5\gamma^m, 2\mathrm{i}\sigma^{mn}|_{m>n}\} \qquad \gamma^5 \equiv -\gamma_5$$

as well as the set Γ_A with lowered indices $(m, n, 5)$. These matrices are normalized according to

$$\begin{gathered} \mathrm{Tr}(\Gamma^A) = 0 \qquad (\Gamma^A \neq 1) \\ \Gamma^A\Gamma_A = 1 \qquad \mathrm{Tr}(\Gamma^A\Gamma_B) = 4\delta^A_B. \end{gathered}$$

They are linearly independent and form a complete set: any complex 4×4 matrix Γ can be decomposed in terms of Γ^A as

$$\Gamma = \sum c_A\Gamma^A \qquad c_A = \tfrac{1}{4}\,\mathrm{Tr}(\Gamma\Gamma_A)$$

or, more explicitly,

$$\Gamma_{mn} = \tfrac{1}{4}\sum \Gamma_{ik}(\Gamma_A)_{ki}(\Gamma^A)_{mn}.$$

Since Γ is an arbitrary matrix, it follows that

$$\tfrac{1}{4}\sum(\Gamma_A)_{ki}(\Gamma^A)_{mn} = \delta_{im}\delta_{kn}.$$

Multiplying this equation with $\bar\psi^1_k\psi^2_i\bar\psi^3_m\psi^4_n$ we obtain the following Fierz rearrangement formula:

$$(\bar\psi^1\psi^4)(\bar\psi^3\psi^2) = -\tfrac{1}{4}\sum(\bar\psi^1\Gamma_A\psi^2)(\bar\psi^3\Gamma^A\psi^4). \tag{J.12}$$

Another form of this formula is obtained by the replacement $\psi^4 \to \Gamma^B\psi^4$, $\psi^2 \to \Gamma^C\psi^2$.

As an example, we display here several Fierz identities for Majorana spinors:

$$(\bar{\varepsilon}_2\psi)(\bar{\chi}\varepsilon_1) + (\bar{\varepsilon}_2\gamma_5\psi)(\bar{\chi}\gamma_5\varepsilon_1) - (\varepsilon_1 \leftrightarrow \varepsilon_2) = -(\bar{\varepsilon}_2\gamma_\mu\varepsilon_1)(\bar{\chi}\gamma^\mu\psi)$$
$$(\bar{\varepsilon}\gamma^m\psi_\nu)(\bar{\psi}_\mu\gamma_5\gamma_m\partial_\rho\psi_\lambda) - (\mu \leftrightarrow \nu) = 0$$
$$\varepsilon^{\mu\nu\rho\lambda}\bar{\psi}_\mu\gamma^i\psi_\nu\bar{u}\gamma_i\psi_\rho = 0$$

where $\psi_\mu = (\psi_{\mu\alpha})$ is a vector–spinor (Rarita–Schwinger field).

The two-component formalism. The Dirac spinor is a reducible representation of the Lorentz group, which is most easily seen in the spinor representation of the gamma matrices:

$$\gamma^m = \begin{pmatrix} 0 & \sigma^m \\ \bar{\sigma}^m & 0 \end{pmatrix} \qquad \gamma_5 = -\mathrm{i}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

where $\sigma^m \equiv (1, \boldsymbol{\sigma})$, $\bar{\sigma}^m \equiv (1, -\boldsymbol{\sigma})$. In this representation, the matrices A, C and $D = CA^T$ have the form

$$A = \gamma^0 \qquad C = \mathrm{i}\gamma^2\gamma^0 = \begin{pmatrix} \mathrm{i}\sigma^2 & 0 \\ 0 & -\mathrm{i}\sigma^2 \end{pmatrix} \qquad D = \mathrm{i}\gamma^2 = \begin{pmatrix} 0 & \mathrm{i}\sigma^2 \\ -\mathrm{i}\sigma^2 & 0 \end{pmatrix}$$

and satisfy the conditions $C^+ = C^{-1} = -C$, $DD^* = 1$ $(\delta = 1)$.

Chiral spinors ψ_-, ψ_+ only have the two upper/lower components different from zero, and are essentially two-component objects:

$$\psi_- = \begin{pmatrix} \xi \\ 0 \end{pmatrix} \qquad \psi_+ = \begin{pmatrix} 0 \\ \bar{\eta} \end{pmatrix} \qquad \psi = \psi_- + \psi_+ = \begin{pmatrix} \xi \\ \bar{\eta} \end{pmatrix}.$$

Since the Lorentz generators are diagonal,

$$\sigma^{0a} = -\tfrac{1}{2}\begin{pmatrix} \sigma^a & 0 \\ 0 & -\sigma^a \end{pmatrix} \qquad \sigma^{ab} = -\tfrac{1}{2}\mathrm{i}\begin{pmatrix} \sigma^c & 0 \\ 0 & \sigma^c \end{pmatrix} \quad (a, b, c \text{ cyclic})$$

each spinor ψ_-, ψ_+ realizes a two-dimensional representation of the Lorentz group with generators

$$(\sigma^{mn})_- = \tfrac{1}{4}(\sigma^m\bar{\sigma}^n - \sigma^n\bar{\sigma}^m) = [-\tfrac{1}{2}\boldsymbol{\sigma}, -\tfrac{1}{2}\mathrm{i}\boldsymbol{\sigma}]$$
$$(\sigma^{mn})_+ = \tfrac{1}{4}(\bar{\sigma}^m\sigma^n - \bar{\sigma}^n\sigma^m) = [\tfrac{1}{2}\boldsymbol{\sigma}, -\tfrac{1}{2}\mathrm{i}\boldsymbol{\sigma}].$$

The transformation laws under infinitesimal transformations $\omega_{mn} = [\boldsymbol{\beta}, \boldsymbol{\theta}]$ (a rotation by angle $\boldsymbol{\theta}$ and a boost with velocity $\boldsymbol{\beta}$) are:

$$\delta_0\xi = (-\tfrac{1}{2}\boldsymbol{\beta\sigma} - \tfrac{1}{2}\mathrm{i}\boldsymbol{\theta\sigma})\xi \qquad \delta_0\bar{\eta} = (\tfrac{1}{2}\boldsymbol{\beta\sigma} - \tfrac{1}{2}\mathrm{i}\boldsymbol{\theta\sigma})\bar{\eta}.$$

The bar over η is to remind us that the $\bar{\eta}$ have a different transformation law from ξ. The finite Lorentz transformations are

$$S(\Lambda) = \begin{pmatrix} M & 0 \\ 0 & M^{-1+} \end{pmatrix}$$
$$M = \exp[\tfrac{1}{2}\omega^{mn}(\sigma_{mn})_-] \qquad M^{-1+} = \exp[\tfrac{1}{2}\omega^{mn}(\sigma_{mn})_+]$$

where $\det M = \det M^{-1+} = 1$ (this condition defines the group $SL(2, C)$). In the two-component formalism, the quantities $(\sigma_{mn})_-$ and $(\sigma_{mn})_+$ are usually denoted as σ_{mn} and $\bar{\sigma}_{mn}$, respectively.

The charge conjugation operation is defined by

$$\psi_c \equiv \begin{pmatrix} \eta^c \\ \bar{\xi}_c \end{pmatrix} = \begin{pmatrix} \mathrm{i}\sigma^2 \bar{\eta}^* \\ -\mathrm{i}\sigma^2 \xi^* \end{pmatrix}.$$

It is usual to use different types of indices for ξ and $\bar{\eta}$, $\xi = (\xi_a)$, $\bar{\eta} = (\bar{\eta}^{\dot{a}})$; they change under complex conjugation as $\xi^* = (\xi^*_{\dot{a}}) = (\bar{\xi}_{\dot{a}})$, and $\bar{\eta}^* = (\bar{\eta}^{*a}) = (\eta^a)$. Introducing the matrices

$$g_{ab} = \mathrm{i}(\sigma^2)_{ab} = \varepsilon_{ab} \qquad (\bar{g}^{-1})^{\dot{a}\dot{b}} \equiv \bar{g}^{\dot{a}\dot{b}} = -\mathrm{i}(\sigma^2)^{\dot{a}\dot{b}} = -\varepsilon^{\dot{a}\dot{b}}$$

we can write

$$(\eta^c)_a = g_{ab}(\bar{\eta}^*)^b \equiv (\bar{\eta}^*)_a \qquad (\bar{\xi}_c)^{\dot{a}} = \bar{g}^{\dot{a}\dot{b}}(\xi^*)_{\dot{b}} \equiv (\xi^*)^{\dot{a}}$$

which shows that charge conjugation is realized as a complex conjugation followed by raising or lowering of an index.

More details about the two-component formalism may be found in appendix H.

The Weyl equations. If we set $m = 0$ in the Dirac equation, we obtain two decoupled, Weyl equations:

$$(\mathrm{i}\partial_0 - \mathrm{i}\sigma^a\partial_a)\xi = 0 \qquad (\mathrm{i}\partial_0 + \mathrm{i}\sigma^a\partial_a)\bar{\eta} = 0. \tag{J.13a}$$

The general solutions of these equations can be written as linear combinations of plane waves. For the positive and negative frequency waves,

$$\xi_{\pm p} = w_{\pm p}\mathrm{e}^{\mp \mathrm{i} p\cdot x} \qquad \bar{\eta}_{\pm p} = v_{\pm p}\mathrm{e}^{-\mathrm{i} p\cdot x}$$

with $p_0 = |\boldsymbol{p}|$, the Weyl equations yield

$$(p_0 + \boldsymbol{\sigma p})w_{\pm p} = 0 \qquad (p_0 - \boldsymbol{\sigma p})v_{\pm p} = 0. \tag{J.13b}$$

Hence, the spinors $\xi_{\pm p}$, $\bar{\eta}_{\pm p}$ are helicity eigenstates with $\lambda = -\frac{1}{2}, +\frac{1}{2}$, respectively.

The Weyl equations describe massless left-handed and right-handed fermion fields, which transform into each other under space inversion. If only one of these fields exists, space inversion is not a good symmetry.

The charge conjugate field $\bar{\xi}_c = -\mathrm{i}\sigma^2\xi^*$ satisfies the same equation as $\bar{\eta}$. Charge conjugation transforms a left-handed particle to a right-handed antiparticle. Note, however, that $\bar{\xi}_c$ is a new right-handed field, different from $\bar{\eta}$. Choosing $\bar{\xi}_c = \bar{\eta}$, we obtain a massless Majorana spinor $(\xi, \bar{\xi}_c)$, which describes a truly neutral particle with two possible helicities (particles and antiparticles are now identical).

The bilinear form $j^\mu = w^+\bar{\sigma}^\mu w = (w^+w, -w^+\boldsymbol{\sigma}w)$ is a Lorentz four-vector (the current density). Multiplying the equation for w from left with w^+, we obtain the continuity equation: $p \cdot j = 0$.

The spinor w may be normalized by the covariant condition

$$w^+w = 2p^0.$$

Indeed, $j^\mu = w^+(1, -\boldsymbol{\sigma})w = 2(p^0, \boldsymbol{p})$, as follows from the Weyl equations.

Let us introduce the projector on the states of helicity $-\frac{1}{2}$:

$$\rho_{-a\dot{b}} = w_a w^+_{\dot{b}}.$$

From $(p_0 + \boldsymbol{\sigma p})\rho_- = 0$, $\rho_-(p_0 + \boldsymbol{\sigma p}) = 0$ and the normalization condition, it follows

$$\rho_- = p_0 - \boldsymbol{\sigma p} = \sigma \cdot p.$$

The projectors on the states with helicity $+\frac{1}{2}$ are:

$$\rho_+^{\dot{a}b} = v^{\dot{a}} v^{+b} \qquad \rho_+ = (p_0 + \boldsymbol{\sigma p}) = \bar{\sigma} \cdot p.$$

The complete four-dimensional projector is given by

$$\rho = \rho_- \oplus \rho_+ = ww^+ \oplus vv^+ = \begin{pmatrix} \sigma \cdot p & 0 \\ 0 & \bar{\sigma} \cdot p \end{pmatrix} = \hat{p}\gamma^0. \tag{J.14a}$$

The expressions for $\rho_\mp$ may be represented as the chiral components of the complete projector: $\rho_\mp = P_\mp \rho P_\mp$.

The quantities w and v can be easily rewritten as the equivalent four-spinors: $w \to u_- = (w, 0)$, $v \to u_+ = (0, v)$. The complete projector expressed in terms of $u_\mp$ has the form

$$u_- u_-^+ + u_+ u_+^+ = \hat{p}\gamma^0. \tag{J.14b}$$

Exercises

1. (a) Show that the relation $C^{-1}\gamma_i C = -\gamma_i^T$ implies $C = \eta C^T$.
   (b) Prove that $\eta = -1$, demanding that the set of 16 matrices $\Gamma^A C$ in M_4 contains 10 symmetric and six antisymmetric linearly independent matrices.
2. Prove the identities:

$$\begin{aligned} &\gamma^l \gamma_m \gamma_l = -2\gamma_m && \sigma^{mn}\sigma_{mn} = -3 \\ &\gamma^l \gamma_m \gamma_n \gamma_l = 4\eta_{mn} && \sigma^{mn}\gamma_l \sigma_{mn} = 0 \\ &\gamma^l \sigma_{mn} \gamma_l = 0 && \sigma^{lr}\sigma_{mn}\sigma_{lr} = \sigma_{mn} \\ &\sigma^{ij}\sigma^{mn}\gamma_j = -\tfrac{1}{2}\sigma^{mn}\gamma^i && \sigma^{ij}\gamma^m\gamma_j = \tfrac{1}{2}\gamma^m\gamma^i - 2\eta^{im} \end{aligned}$$

3. Prove the identities:

$$\gamma_m\gamma_n\gamma_l = -\varepsilon_{mnlr}\gamma_5\gamma^r + \eta_{mn}\gamma_l - \eta_{ml}\gamma_n + \eta_{nl}\gamma_m$$
$$\{\gamma_m, \sigma_{nl}\} = -\varepsilon_{mnlr}\gamma_5\gamma^r$$
$$2\{\sigma_{mn}, \sigma_{lr}\} = \eta_{mr}\eta_{nl} - \eta_{ml}\eta_{nr} - \varepsilon_{mnlr}\gamma_5$$
$$2\gamma_5\sigma_{mn} = -\varepsilon_{mnlr}\sigma^{lr}$$
$$\varepsilon_{mnrl}\gamma_5\sigma^{lk} = -\delta^k_m\sigma_{nr} + \delta^k_n\sigma_{mr} - \delta^k_r\sigma_{mn}.$$

4. (a) Verify the following relations:

$$C^{-1}SC = S^{-1T} \qquad ASA^{-1} = S^{-1+}$$
$$D^{-1}SD = S^* \qquad \gamma^m = \Lambda^m{}_n S\gamma^n S^{-1}.$$

   (b) Find the transformation laws of the bilinear forms $\bar\psi\Gamma^A\psi$ under Lorentz transformations and space inversion.
5. (a) Show that $\psi_c = C\bar\psi^T$ is equivalent to $\psi = C\bar\psi_c^T$.
   (b) Find the Lorentz transformation law of ψ_c.
6. Find the transformation laws for the bilinear forms $\bar\psi\Gamma^A\psi$ under chiral transformations.
7. Show that Majorana spinors satisfy the relations:

$$\bar\psi_\mp\gamma_i\chi_\mp = -\bar\chi_\pm\gamma_i\psi_\pm \qquad \bar\psi_\mp\chi_\pm = \bar\chi_\mp\psi_\pm$$
$$\bar\psi_\mp\sigma_{ij}\chi_\pm = -\bar\chi_\mp\sigma_{ij}\psi_\pm.$$

   Then prove that $\bar\psi_\mp\gamma_5\gamma_i\chi_\mp = \bar\chi_\pm\gamma_5\gamma_i\psi_\pm$ and $\bar\psi_\mp\gamma_5\chi_\pm = \bar\chi_\mp\gamma_5\psi_\pm$.
8. Show by explicit calculation that the matrices $A = \gamma^0$, $C = \mathrm{i}\gamma^2\gamma^0$ and $D = CA^T$ in the spinor representation obey the relations:

$$A\gamma_m A^{-1} = \gamma_m^+ \qquad C^{-1}\gamma_m C = -\gamma_m^T$$
$$DD^* = 1 \qquad C^+ = C^{-1} = C^T = -C.$$

   Then prove that $(\psi_c)_c = \psi$.
9. Prove the following relations in the spinor representation:

$$\sigma^m = \Lambda^m{}_n M\sigma^n M^+ \qquad \bar\sigma^m = \Lambda^m{}_n M^{-1+}\bar\sigma^n M^{-1}.$$

Appendix K

Symmetry groups and manifolds

In this appendix, we review some aspects of symmetries of Riemann spaces that are important for constructing higher-dimensional Kaluza–Klein theories (Weinberg 1972, Choquet-Bruhat *et al* 1977, Barut and Raczka 1977, Dubrovin *et al* 1979, Zee 1981).

Isometries. Let X be an N-dimensional differentiable manifold. Consider an infinitesimal transformation of the points in X described by the change of coordinates

$$x^\alpha \mapsto x'^\alpha = x^\alpha + tE^\alpha(x) \qquad |t| \ll 1 \tag{K.1}$$

where t is an infinitesimal parameter, and E^α a tangent vector on the curve $x'(t)$ at $t = 0$. Under this transformation, a tensor field $T(x)$ becomes the transformed tensor field $T'(x')$. The *Lie derivative* of $T(x)$ characterizes the change of form of $T(x)$ under (K.1):

$$\mathcal{L}_\mathrm{E} T|_x = \lim_{t\to 0} \frac{-1}{t}[T'(x) - T(x)]. \tag{K.2}$$

The Lie derivative is closely related to the concept of form variation: $\delta_0 T(x) = T'(x) - T(x)$.

For a scalar field we have $\phi'(x') = \phi(x)$, or $\phi'(x) = \phi(x - tE)$; keeping terms to order t, we obtain $\mathcal{L}_\mathrm{E}\phi(x) = E^\alpha\partial_\alpha\phi(x)$. The transformation law of a vector field $u(x)$ has the form

$$u'^\alpha(x') = \frac{\partial x'^\alpha}{\partial x^\beta}u^\beta(x) \approx (\delta^\alpha_\beta + t\partial_\beta E^\alpha)u^\beta(x).$$

If we change the variables according to $x' \to x' - tE = x$, $x \to x - tE$, and drop the second order terms in t, we find

$$\mathcal{L}_\mathrm{E} u^\alpha(x) = E^\beta\partial_\beta u^\alpha - u^\beta\partial_\beta E^\alpha.$$

In a space with the Christoffel connection, the ordinary derivatives in this equation may be replaced by covariant derivatives. Going through similar manipulations we can evaluate the Lie derivative of any tensor field.

Now we restrict our attention to a Riemann space $V(X, \boldsymbol{g})$ and consider a coordinate transformation that leaves the form of the metric invariant: $g'_{\alpha\beta}(y) = g_{\alpha\beta}(y)$. Any such transformation is called an *isometry* of V. An infinitesimal isometry is determined by the condition

$$\mathcal{L}_{\rm E} g_{\alpha\beta} \equiv \partial_\alpha E^\gamma g_{\gamma\beta} + \partial_\beta E^\gamma g_{\alpha\gamma} + E^\gamma \partial_\gamma g_{\alpha\beta} = 0.$$

Replacing ∂ here by ∇ and using $\nabla_\gamma g_{\alpha\beta} = 0$, we deduce the relation

$$\nabla_\alpha E_\beta + \nabla_\beta E_\alpha = 0 \tag{K.3}$$

which is known as the *Killing equation*. Thus, the coordinate transformation (K.1) is an isometry of V if the vector E^α satisfies the Killing equation. The problem of determining all infinitesimal isometries of a given Riemann space V is now reduced to the problem of determining all solutions E^α_a of the related Killing equation. To each set of solutions $\{E^\alpha_a, a = 1, 2, \ldots, n\}$ there corresponds an isometry transformation with n parameters t^a:

$$\delta x^\alpha = t^a E^\alpha_a(x) = t^a \boldsymbol{e}_a x^\beta \qquad \boldsymbol{e}_a \equiv E^\alpha_a \partial_\alpha \tag{K.4}$$

where the differential operators $\boldsymbol{e}_a$ are the generators of the isometry.

All Killing vectors E^α_a are the tangent vectors of V, but their number may be higher than N, the dimension of V. We shall see that the maximal number of independent Killing vectors is $N(N+1)/2$.

In order to gain a deeper understanding of the isometries, it is useful to relate their structure to the concept of a group. After giving a short recapitulation of topological and Lie groups, we shall introduce the concept of a Lie group of transformations on a manifold, and then return again to the isometries of Riemann spaces.

Topological groups. One and the same set G may have the structure of both a group and a topological space. Considered as a group, the set G is supplied with a group multiplication law $*$. The same set may be equipped with a topological structure by specifying a topology τ on G.

A set G is a *group* if there exists an internal binary operation $*$ on G (for any g, h in G, the product $g * h$ belongs to G), such that

- the group operation is associative: $g_1 * (g_2 * g_3) = (g_1 * g_2) * g_3$;
- there is an element I in G, called the identity, such that $I * g = g * I = I$, for all $g \in G$;
- for each $g \in G$ there is an element g^{-1} in G, called the inverse of g, such that $g * g^{-1} = g^{-1} * g = I$.

We shall often omit the group multiplication sign $*$ for simplicity.

A mapping f from a group $(G, *)$ to a group $(H, \cdot)$ is said to be a *homomorphism* if $f(g_1 * g_2) = f(g_1) \cdot f(g_2)$. If f is a 1–1 correspondence, it is called an *isomorphism*.

As far as the group properties are concerned, isomorphic groups may be considered to be identical. Isomorphisms play the same role for groups as homeomorphisms do for topological spaces.

Let $\{G, *\}$ be a group, and $\{G, \tau\}$ a topological space. Then the triple $\{G, *, \tau\}$ is said to be a *topological group* if the operations which define the group structure are continuous, i.e. if

- the inverse mapping $g \mapsto g^{-1}$ is continuous; and
- the multiplication mapping $(g_1, g_2) \mapsto g_1 * g_2$ is continuous.

A set G may be both a group and a topological space without necessarily being a topological group. Let us briefly mention some important properties of topological groups.

(1) One of the most powerful notions in topology is that of compactness. Every closed surface in E_3 with a finite diameter, such as the sphere or the torus, may be described as closed and bounded. The concept of the boundedness of closed sets in E_3 can be related to collections of open sets, known as open coverings, using a generalization of the Heine–Borel theorem: X is a closed and bounded subspace of E_n iff every open covering of X contains a finite subcovering. The description in terms of open coverings has an advantage, as it avoids using the non-topological concept of boundedness.

This result suggests the following definition: a topological space $\{X, \tau\}$ is *compact* if every covering of X contains a finite subcovering. Thus, every bounded and closed subset of E_n is compact. A topological group $\{G, *, \tau\}$ is compact if $\{G, \tau\}$ is a compact topological space.

(2) There is an important property of topological spaces, which represents, as we shall see, an embryo of the idea of symmetry in Riemann spaces. A topological space $\{X, \tau\}$ is said to be *homogeneous* if for each $x \in X$ and $y \in X$ there exists a homeomorphism $f : X \to X$ that maps x into y.

It follows that every topological group is necessarily homogeneous. Indeed, for any two elements g_1 and g_2 in G, there is a mapping $g_1 \mapsto g_2 = \gamma g_1$, where $\gamma \equiv g_2 g_1^{-1}$, which is a homeomorphism of G onto G, as a consequence of the uniqueness and continuity of group multiplication. The homogeneity of a topological group G ensures that any local property determined in the neighbourhood of one point is the same as in the neighbourhood of any other point. For this reason, we usually study the local properties of G in the neighbourhood of the identity element.

(3) The concept of connectedness is fairly clear intuitively: a topological space X is *connected* if it cannot be expressed as a union of two disjoint non-empty (open) sets. A topological space is said to be arcwise connected if, given any two

points x and y in X, there is a path (continuous curve) from x to y. Any arcwise connected topological space is connected.

The union of all connected subspaces of X containing a point x is called the (connected) component of x. A component of identity G_0 of topological group G is a subgroup of G.

(4) The concept of homotopy plays an important role in further development of the idea of connectedness. Let x and y be two points in X connected by a path $P(x, y)$. Two paths $P_0(x, y)$ and $P_1(x, y)$ are *homotopic* if there is a continuous deformation transforming $P_0(x, y)$ to $P_1(x, y)$. A path is closed if its end and origin coincide. A closed path at x is called a null-path at x if the whole path coincides with x.

A topological space X is said to be *simply connected* if every closed path in X is homotopic to a null-path. If there are m homotopy classes, the space is said to be m-fold connected.

For any multiply connected space X we can define a covering space of X which is simply connected—the universal covering space of X (a covering space $\bar{X}$ of X may be regarded as an m-fold wrapping of $\bar{X}$ around X, defined by a continuous mapping $\pi : \bar{X} \to X$).

Example 1. Let us now illustrate the previous exposition by considering the group of rotations of E_3. A rotation of the Euclidean space E_3 is the mapping $R : E_3 \to E_3$ which preserves the Euclidean norm of E_3 and its orientation. An arbitrary rotation of a point $\boldsymbol{x} = (x^1, x^2, x^3)$ is described by an action of a 3×3 orthogonal matrix with unit determinant: $x^\alpha \mapsto x'^\alpha = R^\alpha{}_\beta x^\beta$. The set of these matrices defines the group $SO(3)$.

Each rotation can be represented as a rotation through an angle ω around a unit vector $\boldsymbol{n}$ ($\boldsymbol{n}^2 = 1$):

$$\boldsymbol{x}' \equiv R(\boldsymbol{n}, \omega)\boldsymbol{x} = \boldsymbol{x} \cos\omega + \boldsymbol{n}(\boldsymbol{n}\boldsymbol{x})(1 - \cos\omega) + \boldsymbol{x} \wedge \boldsymbol{n} \sin\omega \qquad (0 \le \omega \le \pi)$$

where the orientation of $\boldsymbol{n}$ is given in terms of the spherical angles θ and φ:

$$\boldsymbol{n} = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta) \qquad (0 \le \varphi \le 2\pi, 0 \le \theta \le \pi).$$

We choose $\boldsymbol{\rho} = \omega\boldsymbol{n}$ to be the set of parameters which identify elements of the group. The set of all rotations is determined by the set of values of $\boldsymbol{\rho}$ belonging to a three-dimensional Euclidean ball with radius π and its centre at $\boldsymbol{\rho} = 0$. However, since the rotations corresponding to $\boldsymbol{\rho} = \pi\boldsymbol{n}$ and $\boldsymbol{\rho} = -\pi\boldsymbol{n}$ are the same, the end points of each diameter should be identified. Thus, the space of parameters is a real, three-dimensional projective space RP_3 (a space of lines in E_4 passing through the origin or a sphere S_3 with antipodal points identified). If a path intersects the boundary of the ball at some point, it is considered to enter the ball at the related antipodal point. The parameter space can be equipped with a natural topology, whereupon, by 'translating' that topology to $SO(3)$, the group $SO(3)$ becomes the topological group. Here are some of the topological properties of $SO(3)$.

(1) The group of rotations is compact since the space of parameters is a compact topological space.
(2) It is also homogeneous as it is a topological group.
(3) The group of rotations is connected: given an arbitrary rotation, there is always a path containing that rotation and the identity element $\boldsymbol{\rho} = 0$.
(4) An arbitrary point $\omega \boldsymbol{n}$ ($\omega \geq 0$) can be reached from the centre by using two different paths:

(i) the first path l_1 is a 'direct' one: $\boldsymbol{\rho}(t) = t\boldsymbol{n}$ $(0 \leq t \leq \omega)$;
(ii) the second path l_2 first goes 'backwards' from the centre to $\boldsymbol{\rho} = -\pi \boldsymbol{n}$, $\boldsymbol{\rho}_1(t') = -t'\boldsymbol{n}$ $(0 \leq t' \leq \pi)$, then it 'jumps' to the antipodal point $\boldsymbol{\rho} = \pi \boldsymbol{n}$, and then continues to follow the path $\boldsymbol{\rho}_2(t'') = (\pi - t'')\boldsymbol{n}$ $(0 \leq t'' \leq \pi - \omega)$.

These two paths are not homotopic. In a similar manner, we can conclude that there are two homotopy classes of closed paths, i.e. $SO(3)$ is double connected.

Lie groups. If we introduce local coordinates into a topological group, and then the concept of differentiability, we come to a differentiable group manifold, or a Lie group.

A group $\{G, *\}$ is a *Lie group* if G is a differentiable manifold such that the differentiable structure is compatible with the group structure, i.e.

- the inverse mapping $g \mapsto g^{-1}$ is differentiable; and
- the multiplication mapping $(g_1, g_2) \mapsto g_1 * g_2$ is differentiable.

Very often the differentiable structure is enlarged by introducing an analytic structure which enables us to use the convergent Taylor expansion in the finite neighbourhood of each point.

Let the parameters t^α $(\alpha = 1, 2, \ldots, n)$ be local coordinates of Lie group G in the neighbourhood of the identity element, $c(t)$ a continuous curve in the space of parameters passing through the null point $c(0) = (0, 0, \ldots, 0)$, and $\boldsymbol{e}_a = E_a^\alpha \boldsymbol{e}_\alpha$ its tangent vector at $c(0)$. With each curve $c(t)$ we can associate the curve $C(t)$ in the group manifold, such that $C(0) = I$. The tangent vector to the curve $C(t)$ at $t = 0$ is given by the expression

$$T_a = E_a^\alpha \partial_\alpha C(t)|_{t=0} \equiv \boldsymbol{e}_a C(t)|_{t=0}$$

and is called the *generator* of the group. With each generator T_a we can associate the operator $\boldsymbol{e}_a = E_a^\alpha \partial_\alpha$—the tangent vector to $c(t)$. If the curve $c(t)$ is chosen so as to coincide with the t^α coordinate line, its tangent vector will be equal to the coordinate tangent vector $\boldsymbol{e}_\alpha$: $\boldsymbol{e}_a = \delta_a^\alpha \boldsymbol{e}_\alpha$. Then, for sufficiently small values of the parameters, each element $C(t)$ lying close to the identity may be represented as $C(t) \approx I + t^a T_a$, where $t^a = \delta^a_\alpha t^\alpha$. We can show that the linearly independent generators of G satisfy the *Lie algebra*,

$$[T_a, T_b] = f_{ab}{}^c T_c \qquad f_{ab}{}^c = -f_{ba}{}^c \tag{K.5}$$

where the structure constants $f_{ab}{}^c$ obey the Jacobi identity.

The study of Lie algebras is of fundamental importance in the study of Lie groups. For each Lie group G we can define its Lie algebra but the inverse statement is not, in general, true. Different Lie groups may have the same Lie algebra, but be very different in the large (such Lie groups are locally isomorphic). However, with each Lie algebra we can associate a *unique* simply connected Lie group, $\bar{G}$. For any (multiply) connected Lie group G with the same Lie algebra, the group $\bar{G}$ is the universal covering group of G. Each group G from this family is locally isomorphic to $\bar{G}/Z$, where Z is a discrete invariant subgroup of $\bar{G}$. The group $\bar{G}$ can be homomorphically mapped onto G. It has an important role in the theory of group representations, since all of its IRs are single-valued.

Example 1 (continued). By introducing suitable local coordinates, the group $SO(3)$ becomes the Lie group of dimension three. The set of parameters $\boldsymbol{\rho} = (\omega, \varphi, \theta)$ cannot be used as a global coordinate system on the group manifold, since it is singular at $R = I$, where $\omega = 0$ but φ and θ are not determined. A similar problem also exists in the familiar Euler parametrization. Singular points arise in any parametrization of the rotation matrices so that the coordinates have to be introduced locally. The appearance of the singularity about $R = I$ makes the $\boldsymbol{\rho}$ parametrization particularly inappropriate for defining the group generators as the tangent vectors at the identity element.

A more appropriate parametrization can be obtained by representing an arbitrary rotation as a composition of the following three elements: a rotation through an angle θ^1 about x^1, a rotation through θ^2 about x^2 and, finally, a rotation through θ^3 about x^3. The first rotation matrix has the form

$$R_1(\theta^1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta^1 & \sin\theta^1 \\ 0 & -\sin\theta^1 & \cos\theta^1 \end{pmatrix}$$

and similarly for $R_2(\theta^2)$ and $R_3(\theta^3)$. An arbitrary rotation is given as

$$R(\theta^1, \theta^2, \theta^3) = R_3(\theta^3) R_2(\theta^2) R_1(\theta^1)$$

where $-\pi \le \theta^1 < \pi$, $-\pi \le \theta^2 < \pi$ and $-\pi/2 \le \theta^3 \le \pi/2$. The singular point no longer occurs at $R = I$, but rather at $\theta^3 = \pm\pi/2$. By differentiating the rotation matrix along θ^a at the point $\theta^a = 0$, we obtain the rotation generators as the matrices $(T_a)^b_c = \varepsilon_{abc}$. Their Lie algebra has the form

$$[T_a, T_b] = -\varepsilon_{abc} T_c.$$

Example 2. The group $SU(2)$ is a group of all unitary 2×2 complex matrices A with unit determinant. Every matrix A in $SU(2)$ can be written in the form

$$A = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix} \qquad |a|^2 + |b|^2 = 1.$$

Setting $a = u_0 + \mathrm{i}u_3$ and $b = u_2 + \mathrm{i}u_1$, we find that the parameters (u_0, u_1, u_2, u_3) satisfy the condition

$$u_0^2 + u_1^2 + u_2^2 + u_3^2 = 1.$$

Therefore, the group manifold of $SU(2)$ is the unit sphere S_3 (if we rewrite this condition in the form $u_1^2 + u_2^2 + u_3^2 = 1 - u_0^2$, we see that the projection of the 'upper' half of this sphere, defined by $u_0 \geq 0$, can be looked upon as the unit ball in E_3). The manifold S_3 is compact and simply connected.

Every $SU(2)$ matrix can be expressed in terms of the Pauli matrices σ^a:

$$A = \begin{pmatrix} u_0 + \mathrm{i}u_3 & u_2 + \mathrm{i}u_1 \\ -u_2 + \mathrm{i}u_1 & u_0 - \mathrm{i}u_3 \end{pmatrix} = u_0 I + \mathrm{i}u_1\sigma^1 + \mathrm{i}u_2\sigma^2 + \mathrm{i}u_3\sigma^3.$$

The group generators $T_a = \mathrm{i}\sigma^a/2$ satisfy the Lie algebra $[T_a, T_b] = -\varepsilon_{abc}T_c$, which is isomorphic to the Lie algebra of $SO(3)$. Thus, $SU(2)$ and $SO(3)$ are locally isomorphic groups. Finite elements of $SU(2)$ can be obtained by integrating infinitesimal transformations $A = I + \mathrm{i}\omega n^a\sigma^a/2$ ($\boldsymbol{n}^2 = 1$):

$$A = \exp(\tfrac{1}{2}\mathrm{i}\omega\boldsymbol{n\sigma}) = I\cos(\omega/2) + \mathrm{i}\boldsymbol{n\sigma}\sin(\omega/2) \qquad -2\pi \leq \omega \leq 2\pi.$$

Now we shall show that there exists a homomorphism $R : SU(2) \to SO(3)$. With each point $\boldsymbol{x}$ in E_3, we associate a traceless Hermitian matrix X, given by

$$X = \mathrm{i}x^a\sigma^a \equiv \mathrm{i}\begin{pmatrix} x^3 & x^1 - \mathrm{i}x^2 \\ x^1 + \mathrm{i}x^2 & -x^3 \end{pmatrix} \qquad \det X = \boldsymbol{x}^2.$$

For every A in $SU(2)$, the mapping $X \mapsto X' = AXA^{-1}$ defines an $SO(3)$ rotation of E_3, $\boldsymbol{x} \to \boldsymbol{x}' = R(A)\boldsymbol{x}$. Indeed, $(\boldsymbol{x})^2 = (\boldsymbol{x}')^2$ ($\det X = \det X'$) and the Jacobian is positive, as can be seen from the explicit formulae:

$$x'^a = R^a{}_b x^b \qquad R^a{}_b(A) = \tfrac{1}{2}\,\mathrm{Tr}(\sigma^a A \sigma^b A^{-1}).$$

The mapping $A \mapsto R(A)$ is a homomorphism, since $R(A_1A_2) = R(A_1)R(A_2)$.

The homomorphism R is 2–1. This follows from the fact that the kernel of R (the set of elements in $SU(2)$ that are mapped into the identity element of $SO(3)$) is $Z_2 = (I, -I)$, the discrete invariant subgroup of $SU(2)$. Therefore,

$$SO(3) = SU(2)/Z_2.$$

The group $SU(2)$ is simply connected and represents the universal covering group of $SO(3)$.

We shall now define the adjoint representation of G and use it to construct the Cartan metric on G. Let G be a Lie group, and $\mathcal{A}_G$ the associated Lie algebra at $g = I$. The isomorphic mapping of G into itself, given by

$$U_\Omega : \quad g \mapsto g_\Omega = \Omega g \Omega^{-1}$$

where Ω is a fixed element of G, is known as an inner automorphism of G. The inner automorphisms form a group. Each inner automorphism of G induces an automorphism of the Lie algebra $\mathcal{A}_G$:

$$Ad_\Omega : \quad T_a \mapsto (T_a)_\Omega = \Omega T_a \Omega^{-1}. \qquad \text{(K.6}a\text{)}$$

The automorphisms Ad_Ω also form a group G_A ($Ad_{\Omega_1} Ad_{\Omega_2} = Ad_{\Omega_1\Omega_2}$, etc.). The (homomorphic) mapping $\Omega \mapsto Ad_\Omega$ is called the *adjoint representation* of the group G on its Lie algebra $\mathcal{A}_G$.

If a group element Ω is close to the identity, $\Omega = I + \omega$ with $\omega = \omega^e T_e$, the adjoint automorphism is given by

$$(T_a)_\Omega = T_a + [\omega, T_a] = T_a + \omega^e f_{ea}{}^c T_c \qquad \text{for each } \Omega \in G. \qquad \text{(K.6}b\text{)}$$

The generator of an element Ad_Ω has the form $(T'_e)^c{}_a = f_{ea}{}^c = [T_e, T_a]^c$. The related mapping $T_e \mapsto T'_e = [T_e, T_a]$ defines the adjoint representation of the Lie algebra $\mathcal{A}_G$ on $\mathcal{A}_G$.

Let $U = u^a T_a$ and $V = v^b T_b$ be elements of a Lie algebra $\mathcal{A}_G$, and define the following 'scalar product' on $\mathcal{A}_G$:

$$(U, V) = -\tfrac{1}{2}\,\mathrm{Tr}(U'V'). \qquad \text{(K.7}a\text{)}$$

This bilinear symmetric form on $\mathcal{A}_G \times \mathcal{A}_G$ is known as the *Killing form*. It is invariant under the adjoint automorphisms G_A, as follows from (K.6a) and the standard properties of the trace operation. The coordinate expression for the Killing form is given by $(U, V) = g_{ab} u^a v^b$, where

$$g_{ab} = -\tfrac{1}{2}\,\mathrm{Tr}(T'_a T'_b) = -\tfrac{1}{2} f_{ae}{}^c f_{bc}{}^e \qquad \text{(K.7}b\text{)}$$

is the *Cartan metric*. The Cartan metric is defined in $\mathcal{A}_G$—the tangent space of G at $g = I$. If the elements U, V from $\mathcal{A}_G$ are replaced with the related tangent vectors $\boldsymbol{u} = u^a \boldsymbol{e}_a$ and $\boldsymbol{v} = v^a \boldsymbol{e}_a$ in the space of parameters, the scalar product of $\boldsymbol{u}$ and $\boldsymbol{v}$ at $t = 0$ is naturally defined by the relation

$$(\boldsymbol{u}, \boldsymbol{v}) = (U, V) = g_{ab} u^a v^b. \qquad \text{(K.7}c\text{)}$$

We shall later define the metric in the tangent space at an arbitrary point of G, which is important for the Riemannian structure of G.

The adjoint automorphisms describe an action of the group G on the tangent space of G. Under the action of G, the tangent vector $\boldsymbol{e}_a$ transforms in the same way as T_a: $\delta'_0 \boldsymbol{e}_a = \varepsilon^e f_{ea}{}^c \boldsymbol{e}_c$. This defines the transformation rule of the contravariant vector u^a, while the transformation of the covariant vector v_a is determined by demanding the invariance of $u^a v_a$:

$$\delta_0' u^a = -\varepsilon^e f_{ec}{}^a u^c \qquad \delta_0' v_a = \varepsilon^e f_{ea}{}^c v_c.$$

Treating $f_{ab}{}^c$ as a third rank tensor, we obtain $\delta_0' f_{ab}{}^c = 0$, as a consequence of the Jacobi identity. This implies that the Cartan tensor is invariant under the action of G (appendix A).

Cartan has shown that a Lie group is semisimple iff $\det(g_{ab}) \neq 0$. This is just a condition that the inverse metric g^{ab} exists, which enables the development of the standard tensor algebra.

Lie groups of transformations. Lie groups appear in physics as groups of continuous transformations of a manifold. To each element g in G there corresponds a transformation T_g on a manifold X, such that gh is mapped into $T_g T_h$ (homomorphism). We should clearly distinguish between the group manifold G and the manifold X on which the group acts.

A Lie group G is realized as a *Lie group of transformations* on a differentiable manifold X, if to each element g in G there corresponds a mapping $T_g : X \to X$, such that

- $T_I(x) = x$ for each $x \in X$,
- $(T_{g_1} T_{g_2})(x) = T_{g_1 g_2}(x)$ and
- the mapping $(g, x) \mapsto T_g(x)$ is differentiable.

It follows that $T_{g^{-1}} = (T_g)^{-1}$. The mapping T_g is called a transformation of the manifold X, and the homomorphism $g \mapsto T_g$ is said to be a *realization* of G. When T_g are linear transformations of a vector space X, the homomorphism $g \to T_g$ is called a *representation* of G.

A group G operates *effectively* on X if $T_g(x) = x$ for every $x \in X$ implies $g = I$. This means that the mapping $g \mapsto T_g$ is a 1–1 correspondence ($g \neq h$ implies $T_g \neq T_h$), and therefore, an isomorphism. When $g \mapsto T_g$ is an isomorphic mapping, it defines a *faithful* realization of G.

A group G operates *transitively* on X if, for every $x \in X$ and $y \in X$, there exists a transformation T_g that moves x into y.

Under an infinitesimal transformation $T_g = I + t^a \boldsymbol{e}_a$ the points in X move by an infinitesimal amount:

$$x^\alpha \mapsto x'^\alpha(x, t) \approx x^\alpha + t^a \frac{\partial x'^\alpha}{\partial t^a}\bigg|_x \equiv x^\alpha + t^a E_a^\alpha(x).$$

If the group of transformations T_g is a faithful realization of G, the generators $\boldsymbol{e}_a = E_a^\alpha \partial_\alpha$ satisfy the commutation relations

$$[\boldsymbol{e}_a, \boldsymbol{e}_b] = f_{ab}{}^c \boldsymbol{e}_c \qquad \text{(K.8}a\text{)}$$

characterizing the Lie algebra of G. The commutation relations can be written as a condition on E^a_α:

$$E^\alpha_a \partial_\alpha E^\beta_b - E^\alpha_b \partial_\alpha E^\beta_a = f_{ab}{}^c E^\beta_c \tag{K.8b}$$

which is known as the *Lie equation*.

The Lie equation can be written very compactly in terms of the Lie derivative:

$$\mathcal{L}_a \boldsymbol{e}_b = f_{ab}{}^c \boldsymbol{e}_c \qquad \text{or} \qquad \delta_0 \boldsymbol{e}_b = -t^a f_{ab}{}^c \boldsymbol{e}_c$$

where $\mathcal{L}_a \equiv \mathcal{L}_{e_a}$. Thus, the change of the tangent vector $\delta_0 \boldsymbol{e}_b$ with respect to the coordinate transformations $\delta x^\alpha = t^a E^\alpha_a$ on X amounts to the change $\delta'_0 \boldsymbol{e}_b$, produced by the action of the group in the tangent space of X. Similarly, we can use the transformation laws of E^α_a and u_α to find the change of $u_a = E^\alpha_a u_\alpha$.

On a group manifold G, we can consider transformations of points in G produced by the action of the group G itself. Transformations of G onto G defined by

L_g: $h \mapsto gh$ (left translation) and
D_g: $h \mapsto hg$ (right translation)

have particularly important roles in studying the structure of G as a manifold. We shall limit our exposition to the left translations L_g, having in mind that the right translations can be treated analogously.

We now introduce left invariant vector fields on G. The left translation L_g moves a curve $C(t)$, passing through the identity element, to the curve $C_g(t) = gC(t)$, passing through the point $g = g(\tau)$. For small t we have $C_g(t^\alpha) = g(\tau^\alpha + t^a E^\alpha_a(\tau))$, where $E^\alpha_a(\tau)$ depends on the group multiplication rule, so that

$$T_a(\tau) = \partial_a C_g(t)|_{t=0} = E^\alpha_a(\tau)\partial_\alpha g(\tau) \equiv \boldsymbol{e}_a(g)$$

where $\partial_a = \partial/\partial t^a$. The transformation L_g induces the transformation L'_g of the tangent vector $\boldsymbol{e}_a$ to the curve $c(t)$ into the corresponding tangent vector $\boldsymbol{e}_a(\tau)$ to the transformed curve $c_g(t)$. Thus, starting from an arbitrary tangent vector $\boldsymbol{u}$ we can generate the *vector field* $\boldsymbol{u}_g = L'_g \boldsymbol{u}$. This vector field is left invariant by construction: $L'_g \boldsymbol{u}_h = \boldsymbol{u}_{gh}$.

The following two statements characterize this structure.

(i) Every left translation maps the tangent space at $g = I$ into the tangent space at the point $g(\tau)$; this mapping is a 1–1 correspondence.
(ii) The commutator structure is left invariant: if $[\boldsymbol{u}, \boldsymbol{v}] = \boldsymbol{w}$ at $g = I$, then $[\boldsymbol{u}_g, \boldsymbol{v}_g] = \boldsymbol{w}_g$, for any g in G.

The first statement implies that a left translation moves a basis $\boldsymbol{e}_a$ at the identity into the set of tangent vectors $\boldsymbol{e}_a(\tau)$, which is a basis at the point $g(\tau)$. The basis $\boldsymbol{e}_a(\tau)$ is related to the coordinate basis $\boldsymbol{e}_\alpha(\tau) = \partial_\alpha$ by the relation

$$\boldsymbol{e}_a(\tau) = E^\alpha_a(\tau)\boldsymbol{e}_\alpha(\tau)$$

where the coefficients E_a^α depend on the point $g(\tau)$; in particular, $E_a^\alpha = \delta_a^\alpha$ at $g = I$. The second statement implies that the basis $\boldsymbol{e}_a(\tau)$ satisfies the same Lie algebra as the basis $\boldsymbol{e}_a$:

$$[\boldsymbol{e}_a(\tau), \boldsymbol{e}_b(\tau)] = f_{ab}{}^c \boldsymbol{e}_c(\tau). \tag{K.9}$$

This Lie algebra can also be transformed into the form of the Lie equation (K.8*b*).

Let us, furthermore, consider the quantity $\boldsymbol{w} = g^{-1}\boldsymbol{d}g$, where $\boldsymbol{d}g$ is the differential of $g(t)$. This quantity is linear in $\boldsymbol{d}t^\alpha$ and represents a 1-form, which is invariant under left translations $g(t) \mapsto hg(t)$, where h is a fixed element in G. Moreover, $g^{-1}\boldsymbol{d}g$ produces infinitesimal transformations of points in G, which implies that $\boldsymbol{w}$ can be expressed in terms of the group generators. Therefore, $\boldsymbol{w}$ is a Lie algebra valued 1-form at the point g:

$$\boldsymbol{w} \equiv g^{-1}\boldsymbol{d}g = \boldsymbol{\theta}^a T_a \qquad \boldsymbol{\theta}^a \equiv \boldsymbol{d}t^\alpha E_\alpha^a. \tag{K.10a}$$

From the definition of $\boldsymbol{w}$ it follows that $\boldsymbol{d}g = g\boldsymbol{w}$, whereupon the application of the exterior derivative $\boldsymbol{d}$ yields the relation

$$\boldsymbol{d}\boldsymbol{w} + \boldsymbol{w} \wedge \boldsymbol{w} = 0. \tag{K.10b}$$

known as the *Maurer–Cartan equation*. This can be written in the following equivalent forms:

$$\begin{aligned} &\boldsymbol{d}\boldsymbol{\theta}^a + \tfrac{1}{2} f_{bc}{}^a \boldsymbol{\theta}^b \wedge \boldsymbol{\theta}^c = 0 \\ &\partial_\alpha E_\beta^a - \partial_\beta E_\alpha^a + f_{bc}{}^a E_\alpha^b E_\beta^c = 0. \end{aligned} \tag{K.10c}$$

Now, we show that the Maurer–Cartan equation has a direct relation to the Lie equation. Let us define the matrix H_a^α which is 'dual' to E_α^a, in the sense that $E_\alpha^b H_a^\alpha = \delta_a^b$ and $E_\alpha^a H_a^\beta = \delta_\alpha^\beta$. Multiplying the last equation in (K.10*c*) by $H_m^\alpha H_n^\beta H_a^\gamma$ we find that H_a^γ satisfies the Lie equation (K.8*b*). In other words, the coefficients E_a^α and E_α^a, appearing in the Lie and Maurer–Cartan equations, respectively, are 'dual' to each other. While the Lie equation defines the commutation properties of the vector fields $\boldsymbol{e}_a = E_a^\alpha \partial_\alpha$, the Maurer–Cartan equation gives the equivalent information in terms of the 1-forms $\boldsymbol{\theta}^a = \boldsymbol{d}t^\alpha E_\alpha^a$. The basis of 1-forms $\boldsymbol{\theta}^a$ is dual to the basis of the tangent vectors $\boldsymbol{e}_a$. Equations (K.8) and (K.10) should be compared with the related equations (B.7*a*, *b*).

Example 3. Let A be an arbitrary element of $SU(2)$ which we parametrize in terms of the Euler angles (θ, ψ, ϕ): $A(\psi, \theta, \phi) = A_3(\psi)A_1(\theta)A_3(\phi)$. Using $A_a(\omega) = I\cos(\omega/2) + \mathrm{i}\boldsymbol{\sigma}^a \sin(\omega/2)$, we find that

$$A(\psi, \theta, \phi) = \begin{pmatrix} \cos(\theta/2)\mathrm{e}^{\mathrm{i}(\psi+\phi)/2} & \mathrm{i}\sin(\theta/2)\mathrm{e}^{\mathrm{i}(\psi-\phi)/2} \\ \mathrm{i}\sin(\theta/2)\mathrm{e}^{-\mathrm{i}(\psi-\phi)/2} & \cos(\theta/2)\mathrm{e}^{-\mathrm{i}(\psi+\phi)/2} \end{pmatrix}.$$

$(0 \le \psi \le 4\pi, 0 \le \theta \le \pi, 0 \le \phi \le 2\pi)$. Let $A_3(t)$ be a path in $SU(2)$ passing through the identity, $A_3(0) = I$, and $C_3(t)$ a path obtained from $A_3(t)$ by a left translation: $C_3(t) = AA_3(t)$. The tangent vector to the curve $C_3(t)$ at $t = 0$ has the form

$$T_{3A} = (\mathrm{d}/\mathrm{d}t)C_3(t)|_{t=0} \equiv \boldsymbol{e}_3(A)$$

where $\boldsymbol{e}_3 = \partial/\partial\phi$ is a left invariant vector field on $SU(2)$. In a similar way, we define the tangent vectors to the curves $C_1(t) = AA_1(t)$ and $C_2(t) = AA_2(t)$, respectively:

$$T_{1A} = (\mathrm{d}/\mathrm{d}t)C_1(t)|_{t=0} \equiv \boldsymbol{e}_1(A) \qquad T_{2A} = (\mathrm{d}/\mathrm{d}t)C_2(t)|_{t=0} \equiv \boldsymbol{e}_2(A).$$

The first equation implies the relation $A(\psi,\theta,\phi)(\mathrm{i}\sigma^1/2) = E_1^\alpha \partial_\alpha A(\psi,\theta,\phi)$, which can be used to determine the components of the vector field $\boldsymbol{e}_1 = E_1^\alpha\partial_\alpha$. Similar manipulations lead to the components of $\boldsymbol{e}_2 = E_2^\alpha\partial_\alpha$. The result can be written in the form

$$\begin{aligned}
\boldsymbol{e}_1 &= \cos\phi\frac{\partial}{\partial\theta} - \sin\phi\left(\cot\theta\frac{\partial}{\partial\phi} - \frac{1}{\sin\theta}\frac{\partial}{\partial\psi}\right)\\
\boldsymbol{e}_2 &= \sin\phi\frac{\partial}{\partial\theta} + \cos\phi\left(\cot\theta\frac{\partial}{\partial\phi} - \frac{1}{\sin\theta}\frac{\partial}{\partial\psi}\right)\\
\boldsymbol{e}_3 &= \frac{\partial}{\partial\phi}.
\end{aligned}$$

The commutation rules of the generators $\boldsymbol{e}_a$ have the form $[\boldsymbol{e}_a, \boldsymbol{e}_b] = -\varepsilon_{abc}\boldsymbol{e}_c$.

Explicit evaluation of the left invariant 1-form $\boldsymbol{w} = A^{-1}\boldsymbol{d}A$ yields

$$\begin{aligned}
\boldsymbol{w} = {}&(\boldsymbol{d}\psi\sin\theta\sin\phi + \boldsymbol{d}\theta\cos\phi)T_1\\
&+ (-\boldsymbol{d}\psi\sin\theta\cos\phi + \boldsymbol{d}\theta\sin\phi)T_2 + (\boldsymbol{d}\psi\cos\theta + \boldsymbol{d}\phi)T_3 \equiv \boldsymbol{d}t^\alpha E^a_\alpha T_a
\end{aligned}$$

where $T_a = (\mathrm{i}/2)\sigma^a$. Now we define the vector fields $\boldsymbol{e}'_a = E_a^\alpha\partial_\alpha$ as the duals of $\boldsymbol{\theta}^a = \boldsymbol{d}t^\alpha E^a_\alpha$, and find that the result coincides with $\boldsymbol{e}_a$: $\boldsymbol{e}'_a = \boldsymbol{e}_a$.

Riemannian structure on *G*. The Lie and Maurer–Cartan equation, together with the Cartan metric at $g = I$, represent important characteristics of group manifolds. By introducing the metric and a suitable connection at each point of G, the group manifold becomes a Riemann space.

Consider two left invariant vector fields on G, $\boldsymbol{u}_g = u^a\boldsymbol{e}_a$ and $\boldsymbol{v}_g = v^b\boldsymbol{e}_b$. The scalar product of $\boldsymbol{u}_g$ and $\boldsymbol{v}_g$ at an arbitrary point g is, by definition, equal to their scalar product at $g = I$:

$$(\boldsymbol{u}_g, \boldsymbol{v}_g) = (\boldsymbol{u}, \boldsymbol{v}). \tag{K.11a}$$

Expressed in terms of the components, this condition yields

$$g_{\alpha\beta}E_a^\alpha E_b^\beta = g_{ab} \qquad g_{\alpha\beta} \equiv (\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta) \tag{K.11b}$$

where $g_{\alpha\beta}$ is the *Killing metric* on G; at the point $g = I$, where $E^a_\alpha = \delta^a_\alpha$, this reduces to g_{ab}.

The Killing metric can also be defined by using the Killing bilinear form $(\boldsymbol{w}, \boldsymbol{w})$ of the 1-form $\boldsymbol{w} = g^{-1}\boldsymbol{d}g$:

$$(\boldsymbol{w}, \boldsymbol{w}) = \boldsymbol{d}t^\alpha \boldsymbol{d}t^\alpha E^a_\alpha E^b_\beta g_{ab} \equiv \boldsymbol{d}t^\alpha \boldsymbol{d}t^\beta g_{\alpha\beta}. \tag{K.12}$$

The definition of $(\boldsymbol{w}, \boldsymbol{w})$ implies that the Killing metric is left invariant.

We define the *connection* on G by introducing the covariant derivative (Dubrovin *et al* 1979):

$$\nabla_u \boldsymbol{v} = \tfrac{1}{2}\mathcal{L}_u \boldsymbol{v} = \tfrac{1}{2}[\boldsymbol{u}, \boldsymbol{v}] \tag{K.13a}$$

where $\boldsymbol{u}$ and $\boldsymbol{v}$ are left invariant vector fields on G. The connection coefficients are determined by the change of basis,

$$\nabla \boldsymbol{e}_a \equiv \boldsymbol{\theta}^e \otimes \nabla_e \boldsymbol{e}_a = \boldsymbol{\omega}^c{}_a \otimes \boldsymbol{e}_c \qquad \boldsymbol{\omega}^c{}_a = \tfrac{1}{2} f_{ea}{}^c \boldsymbol{\theta}^e. \tag{K.13b}$$

We shall show that this connection is Riemannian, i.e. that the torsion vanishes and $\nabla g_{ab} = 0$.

Let $\boldsymbol{w} = \boldsymbol{\theta}^a \boldsymbol{e}_a$ be a vector valued 1-form, i.e. a (1,1) tensor. By acting on $\boldsymbol{w}$ with the generalized exterior derivative (appendix B), we obtain

$$\bar{\boldsymbol{d}}\boldsymbol{w} = \boldsymbol{d}\boldsymbol{\theta}^a \boldsymbol{e}_a - \boldsymbol{\theta}^a \nabla \boldsymbol{e}_a = (\boldsymbol{d}\boldsymbol{\theta}^c + \boldsymbol{\omega}^c{}_a \boldsymbol{\theta}^a)\boldsymbol{e}_c \equiv \mathcal{T}^c \boldsymbol{e}_c$$

where $\mathcal{T}^c$ is the torsion. Assuming the connection (K.13), we see that the Maurer–Cartan equation coincides with the condition $\mathcal{T}^c = 0$. On the other hand,

$$\nabla_c g_{ab} = \nabla_c(\boldsymbol{e}_a, \boldsymbol{e}_b) = \tfrac{1}{2}(f_{ca}{}^e g_{eb} + f_{cb}{}^e g_{ea}) = 0$$

since $f_{abc} = f_{ab}{}^e g_{ec}$ is completely antisymmetric.

The curvature tensor of G is calculated according to (B.10):

$$R(\boldsymbol{u}, \boldsymbol{v})\boldsymbol{w} = \nabla_u \nabla_v \boldsymbol{w} - \nabla_v \nabla_u \boldsymbol{w} - \nabla_{[u,v]}\boldsymbol{w} = -\tfrac{1}{4}\,[[\boldsymbol{u}, \boldsymbol{v}], \boldsymbol{w}]. \tag{K.14a}$$

Using $R(\boldsymbol{e}_a, \boldsymbol{e}_b)\boldsymbol{e}_c = \boldsymbol{e}_e\, R^e{}_{cab}$, we obtain

$$R^e{}_{cab} = -\tfrac{1}{4} f_{ab}{}^d f_{dc}{}^e \qquad R_{cb} = \tfrac{1}{2} g_{cb}. \tag{K.14b}$$

The symmetries of Riemann spaces. The essential property of a symmetry transformation of a given space is that it leaves some important properties of the space unchanged. Thus, $SO(3)$ is a symmetry of the Euclidean space E_3, as it leaves the distance and orientation of E_3 invariant. This brings us back to the idea of isometry transformations in Riemann spaces.

We have seen that the isometry transformations of a Riemann space V are determined by the solutions of the Killing equation. If the Killing vectors satisfy the Lie equation, the isometries have the structure of a Lie group. Each Killing vector connects neighbouring points of V which are, as far as the form of the metric is concerned, equivalent. The number of linearly independent Killing vectors depends on the form of the metric.

A Riemann space V is said to be *homogeneous* if, for every $x \in V$ and $y \in V$, there exists an isometry transformation T_g that moves x into y (the isometry group operates transitively on V). Stated differently, at any point x in V, there exist Killing vectors that take all possible values. In an N-dimensional homogeneous space V, the number of linearly independent Killing vectors at any given point is N. For instance, we can choose a set of N Killing vectors defined by $E_a^\alpha = \delta_a^\alpha$.

A Riemann space V is said to be *isotropic* about a point x if there is an isometry group H_x such that it does not move the point x, $T_H x = x$. As a consequence, any Killing vector must vanish at x, $E_a^\alpha(x) = 0$, but the first derivatives $E_a^{\alpha,\beta}(x)$ at x may take all possible values, respecting, of course, the antisymmetry condition (K.3b). In an N-dimensional isotropic space V, the number of linearly independent Killing vectors is $N(N-1)/2$; they can be defined by choosing a set of coefficients E_{mn}^α satisfying the following conditions:

$$E_{mn}^\alpha(x) = 0 \qquad E_{mn}^\alpha(x) = -E_{nm}^\alpha(x)$$
$$E_{mn}^{\alpha,\beta}(x) = \delta_m^\alpha\delta_n^\beta - \delta_n^\alpha\delta_m^\beta.$$

Using the general formula $[\nabla_\gamma, \nabla_\beta]E_\alpha = E_\delta R^\delta{}_{\alpha\beta\gamma}$ and the cyclic identity for the curvature tensor $R^\delta{}_{\alpha\beta\gamma} + R^\delta{}_{\gamma\alpha\beta} + R^\delta{}_{\beta\gamma\alpha} = 0$, we can show that any Killing vector E_α satisfies the condition

$$\nabla_\alpha\nabla_\beta E_\gamma = E_\epsilon R^\epsilon{}_{\alpha\beta\gamma}.$$

By a repeated differentiation of this equation, we find that any derivative of E_α at y can be expressed in terms of $E_\alpha(y)$ and $\nabla_\beta E_\alpha(y)$. Then, expressing a Killing vector E_α at some point x, lying in a neighbourhood of y, as a Taylor series in $x-y$, we find that $E_\alpha(x)$ is given as a linear combination of $E_\beta(y)$ and $\nabla_\gamma E_\beta(y)$:

$$E_\alpha(x) = A_\alpha{}^\beta(x, y)E_\beta(y) + B_\alpha{}^{\gamma\beta}(x, y)\nabla_\gamma E_\beta(y)$$

where A and B do not depend on $E_\beta(y)$ or $\nabla_\gamma E_\beta(y)$, and $B_\alpha{}^{\gamma\beta} = -B_\alpha{}^{\beta\gamma}$. This equation implies that there can be no more than $N(N+1)/2$ linearly independent Killing vectors in N dimensions. Indeed, since there are N independent quantities $E_\beta(y)$ and $N(N-1)/2$ independent quantities $\nabla_\gamma E_\beta(y)$, the maximal number of linearly independent Killing vectors $E_\alpha(x)$ is equal to the sum $N+N(N-1)/2 = N(N+1)/2$.

A Riemann space V is said to be *maximally symmetric* if it admits the maximal number of $N(N+1)/2$ Killing vectors. In particular, a homogeneous space that is isotropic about some point is maximally symmetric. Isotropy about a given point implies, as a consequence of homogeneity, isotropy about every point; then, summing the number of Killing vectors corresponding to homogeneity and isotropy, $N + (N-1)/2$, we obtain the maximal number of $N(N+1)/2$ linearly independent Killing vectors.

Note that a space that is isotropic about every point is also homogeneous and, consequently, maximally symmetric.

Maximally symmetric spaces have the following important properties (see, e.g., Weinberg 1972):

- If a Riemann space V is maximally symmetric, its curvature tensor must be of the form

$$R_{\alpha\beta\gamma\delta} = \Lambda(g_{\alpha\gamma}g_{\beta\delta} - g_{\alpha\delta}g_{\beta\gamma}) \qquad \Lambda = \text{constant}. \tag{K.15}$$

  A space of this type is called a space of constant curvature.
- Any two constant curvature spaces with the same signature which have the same value of Λ are locally isometric, i.e. there is a coordinate transformation that maps one metric into the other.

Thus, maximally symmetric spaces (of a given signature) are essentially unique and can be determined by constructing particular representatives of constant curvature spaces with all values of Λ.

Example 4. The unit sphere S_2 is a Riemann space with a metric, in local coordinates (θ, φ), $\theta \neq 0, \pi$, given by

$$\mathrm{d}s^s = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2.$$

The Killing equations have the form

$$\begin{gathered} E_{\varphi,\varphi} + \sin\theta\cos\theta\,E_\theta = 0 \\ E_{\theta,\theta} = 0 \\ E_{\theta,\varphi} + E_{\varphi,\theta} - 2\cot\theta\,E_\varphi = 0. \end{gathered}$$

The first two equations yield $E_\theta = f(\varphi)$, $E_\varphi = -F(\varphi)\sin\theta\cos\theta + g(\theta)$, where $F(\varphi) = \int f(\varphi)\,\mathrm{d}\varphi$, and f, g are arbitrary functions. The functions f and g are determined from the third equation: $f = a\sin\varphi + b\cos\varphi$, $g = c\sin^2\theta$, where a, b, c are arbitrary constants. After that, the solution for the Killing vector (E_θ, E_φ) takes the form

$$\begin{gathered} E_\theta = a\sin\varphi + b\cos\varphi = E^\theta \\ E_\varphi = (a\cos\varphi - b\sin\varphi)\sin\theta\cos\theta + c\sin^2\theta = \sin^2\theta\,E^\varphi. \end{gathered}$$

The presence of three arbitrary parameters a, b, c means that there exist three linearly independent solutions. Introducing the notation $\boldsymbol{e} = E^\theta\partial_\theta + E^\varphi\partial_\varphi$ and separating the terms multiplying a, b and c we obtain three independent generators:

$$\begin{aligned}\boldsymbol{e}_1 &= \sin\varphi\partial_\theta + \cot\theta\cos\varphi\partial_\varphi\\ \boldsymbol{e}_2 &= \cos\varphi\partial_\theta - \cot\theta\sin\varphi\partial_\varphi\\ \boldsymbol{e}_3 &= \partial_\varphi.\end{aligned}$$

These generators satisfy the commutation rules $[\boldsymbol{e}_a, \boldsymbol{e}_b] = -\varepsilon_{abc}\boldsymbol{e}_c$. Therefore, the isometry group of the sphere S_2 is the rotation group $SO(3)$. Note that the number of generators is higher than the dimension of the space.

The rotation group operates transitively on S_2. Take, for instance, the point $P \in S_2$ with Cartesian coordinates $x_P^a = (1, 0, 0)$ in E_3. Then, we see that an arbitrary point (x^1, x^2, x^3) on the sphere can be obtained by a suitably chosen rotation of the point P: $x^a = R^a{}_b x_P^b = R^a{}_1$. Similar arguments lead to the same conclusion for any other point P'. Consequently, S_2 is a homogeneous space.

Observe that the choice of the rotation that moves P into (x^1, x^2, x^3) is not unique, since there exists a subgroup $H_P = SO(2)$ of the group $SO(3)$ that leaves the point P fixed:

$$\begin{pmatrix}1 & 0 & 0\\ 0 & R^2{}_2 & R^2{}_3\\ 0 & R^3{}_2 & R^3{}_3\end{pmatrix}\begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}1\\0\\0\end{pmatrix}.$$

The subgroup H_P is the isotropy group of P. If a transformation R moves the point P into (x^1, x^2, x^3), so does the transformation RR_H, $R_H \in H_P$. Isotropy of S_2 about P implies its isotropy about every point.

The sphere S_2 is homogeneous and isotropic, hence it is maximally symmetric. The same conclusion can also be drawn from the existence of three Killing vectors on S_2.

The construction of maximally symmetric spaces. We describe here one specific construction of a three-dimensional maximally symmetric space with the metric signature $(+, +, +)$, having in mind that the procedure can be easily generalized to higher dimensions and different signatures. Consider a four-dimensional Euclidean space E_4 with metric

$$\mathrm{d}s^2 = \delta_{ab}\,\mathrm{d}x^a\,\mathrm{d}x^b + \mathrm{d}z^2. \tag{K.16a}$$

A three-dimensional sphere S_3 embedded in E_4 is described by the equation

$$S_3: \quad \delta_{ab}x^a x^b + z^2 = \kappa^2. \tag{K.16b}$$

The interval $\mathrm{d}s^2$ on S_3 is calculated using the relation $\delta_{ab}x^a\,\mathrm{d}x^b + z\,\mathrm{d}z = 0$, valid on S_3, to eliminate $\mathrm{d}z$ in (K.16a):

$$\mathrm{d}s^2|_{S_3} = \delta_{ab}\,\mathrm{d}x^a\,\mathrm{d}x^b + \frac{(\delta_{ab}x^a\,\mathrm{d}x^b)^2}{\kappa^2 - \delta_{cd}x^c x^d} \equiv g_{ab}\,\mathrm{d}x^a\,\mathrm{d}x^b. \tag{K.17}$$

The metric g_{ab} is the metric on S_3.

It is obvious that the interval (K.16a) of E_4 and the condition (K.16b) that defines S_3 are both invariant under the $SO(4)$ rotations of E_4, which have the following form:

$$x'^a = R^a{}_b x^b + R^a{}_4 z \qquad z' = R^4{}_b x^b + R^4{}_4 z$$

where the Rs are 4×4 orthogonal matrices with unit determinant. The group $SO(4)$ has six generators (this is the number of traceless antisymmetric 4×4 matrices). Thus, the sphere S_3 has the maximal number of six Killing vectors and is, therefore, maximally symmetric.

Since S_3 is maximally symmetric, it is enough to calculate its curvature in the vicinity of $x^a = 0$. A direct calculation yields $g_{ab} \approx \delta_{ab} + x_a x_b/\kappa^2$, $\Gamma^a_{bc} \approx x^a \delta_{bc}/\kappa^2$, so that

$$R_{abcd} = \frac{1}{\kappa^2}(g_{ac}g_{bd} - g_{ad}g_{bc}) \qquad R = \frac{6}{\kappa^2}.$$

Hence, the constant κ^2 introduced in (K.16b) is proportional to the inverse scalar curvature of S_3.

Going over to the spherical coordinates (r, θ, φ), the metric of S_3 takes the form

$$\mathrm{d}s^2 = \frac{\mathrm{d}r^2}{1 - r^2/\kappa^2} + r^2(\sin^2\theta\,\mathrm{d}\varphi^2 + \mathrm{d}\theta^2)$$

which shows some geometric characteristics of this space more clearly.

In the previous analysis we constructed the maximally symmetric space of $SO(4)$. According to the value of $\Lambda = 1/2\kappa^2$ we can distinguish the following three cases:

(a) $\Lambda > 0$, the space of constant positive curvature,
(b) $\Lambda < 0$, the space of constant negative curvature and
(c) $\Lambda = 0$, the flat space.

This construction can be easily generalized to other symmetry groups (in appendix C we give the related construction for the anti de Sitter group $SO(2, 3)$). For a given signature and a given symmetry group G, the only freedom we have is described by the curvature constant Λ.

Coset spaces. Although the structure of maximally symmetric spaces is now clear by itself, their relation to the corresponding symmetry groups remains somewhat obscure. For instance, what is the relation between S_2 and the structure of $SO(3)$? In order to answer these questions, we discuss here the coset spaces.

Let $\{G, *\}$ be a group. A subset H of G is a *subgroup* of G if $h_1, h_2 \in H$ implies: (a) $h_1 h_2 \in H$ and (b) $h_1^{-1} \in H$, i.e. if $\{H, *\}$ is also a group. The group of rotations about the x-axis, $SO(2)$, is the subgroup of all rotations $SO(3)$ of the Euclidean space E_3.

Let H be a subgroup of G. The set of elements $gH = \{gh | h \in H\}$, for a given $g \in G$, is called a *left coset* of H in G. Taking all possible elements $g_i \in G$, we obtain a collection of left cosets $g_i H$. Now we define an equivalence relation in G: $g_1 \sim g_2$ if $g_1 \in g_2 H$. It is obvious that the related equivalence classes coincide with the left cosets in G.

In a similar way, every set $Hg = \{hg | h \in H\}$ is called a *right coset* of H in G. All transformations from $SO(3)$ which have the form RR_x, where R_x is a given rotation about the x-axis and R an arbitrary rotation from $SO(3)$, define the right coset of R_x in $SO(3)$ (see example 4).

A subgroup H of G is *invariant* (or normal) if $gHg^{-1} = H$, for every $g \in G$. The left coset gH and the right coset Hg of an invariant subgroup H are identical.

The collection of all cosets of an invariant subgroup H is a new group G/H, known as the *factor group* (or the quotient group) of G by H. The group operation in G/H is given in terms of the products of cosets:

$$[g_1][g_2] = [g_1 g_2].$$

The coset $[h]$, $h \in H$, plays the role of the identity element in G/H. If there exists a homomorphism $G \to G/H$ defined by $g \mapsto [g]$, it is called the natural or canonical homomorphism.

If G is a topological group, the group G/H can be topologized by introducing a natural topology (the quotient topology) and become a topological space, called the *coset space* (the quotient space). We shall retain the same name when G is a Lie group.

In accordance with the division G/H, all the generators of G will be divided into two sets: the generators of H, denoted by $\Gamma_H = \{H_{\bar{a}}\}$, and the remaining generators, $\Gamma_{\mathrm{M}} = \{M_{a'}\}$; then, symbolically, $\Gamma = \Gamma_H + \Gamma_{\mathrm{M}}$. The set of generators Γ_H constitutes the Lie algebra of H, while the set Γ_{M} is not a Lie algebra. If the group G is semisimple, its Lie algebra has the form

$$\begin{aligned} [H_{\bar{a}}, H_{\bar{b}}] &= f_{\bar{a}\bar{b}}{}^{\bar{c}} H_{\bar{c}} \\ [H_{\bar{a}}, M_{b'}] &= f_{\bar{a}b'}{}^{c'} M_{c'} \\ [M_{a'}, M_{b'}] &= f_{a'b'}{}^{\bar{c}} H_{\bar{c}} + f_{a'b'}{}^{c'} M_{c'}. \end{aligned} \tag{K.18}$$

Indeed, for semisimple groups the structure constants $f_{abc} = f_{ab}{}^{e} g_{ec}$ are completely antisymmetric, so that the existence of the subalgebra $\mathcal{A}_H$ implies $f_{\bar{a}\bar{b}}{}^{c'} = 0$ and $f_{\bar{a}b'}{}^{\bar{c}} = 0$.

Let the group G be parametrized by $t^a = (t^{\bar{a}}, t^{a'})$. Since $[h]$ is the identity element in G/H, and it contains the whole H, we see that the factor group can be described by the set of parameters $t^{a'}$ (ignoring the parameters $t^{\bar{a}}$ that describe H). After introducing the quotient topology, the factor group becomes the coset space of the dimension $\dim(G/H) = \dim(G) - \dim(H)$.

Example 5. In this example, we prove the relation $S_2 = SO(3)/SO(2)$. First, we note that (a) the sphere S_2 is a homogeneous space of the group $SO(3)$ and (b) the group $SO(3)$ contains the invariant subgroup $H = SO(2)_3$, defined by the rotations about the z-axis, which is the isotropy group of the point $x_0 = (0, 0, 1)$ in S_2: $T_H x_0 = x_0$. Then, with each fixed element $R_{\boldsymbol{n}}(\omega)$ of $SO(3)$ (the rotation about $\boldsymbol{n}$ through an angle ω) we associate the coset $[R_{\boldsymbol{n}}(\omega)] = R_{\boldsymbol{n}}(\omega)H$.

Now, with each coset $[R_{\boldsymbol{n}}(\omega)]$ we associate the point $x_0' = T_{\boldsymbol{n}}(\omega)x_0$ in S_2. This mapping is a 1–1 correspondence, since it does not depend on the choice of the representative element of the coset (x_0' is the same for all elements of the coset). Therefore, *the sphere S_2 coincides with the coset space $SO(3)/SO(2)$.*

This result is a special case of the following general theorem (Dubrovin *et al* 1979).

> *There is a 1–1 correspondence between a homogeneous space V of the group G and the coset space G/H, where H is the isotropy group of V.*

The Riemannian structure of coset spaces. Let V be the maximally symmetric space of the group G, $H = H_{x_0}$ the subgroup of isotropy about $x_0 \in V$, and $\boldsymbol{e}_a = E_a^\alpha \partial_\alpha$ the generators of G on V. Since V is essentially identical to the coset space G/H, the problem of constructing its metric can be understood as that of introducing Riemannian structure onto the coset space G/H (see, e.g., Zee 1981).

The metric of V must be such that the transformations of G are isometries of V. We shall show that the quantity

$$g^{\alpha\beta} = E_a^\alpha E_b^\beta g^{ab} \tag{K.19}$$

where g^{ab} is the Cartan metric of G, satisfies this demand. The condition that the E_a^α are Killing vectors of the metric (K.19) has the form

$$\mathcal{L}_c g^{\alpha\beta} = g^{\alpha\varepsilon}\partial_\varepsilon E_c^\beta + g^{\varepsilon\beta}\partial_\varepsilon E_c^\alpha - E_c^\varepsilon \partial_\varepsilon g^{\alpha\beta} = 0.$$

Using the Lie equation, this relation becomes

$$E_a^\alpha E_b^\beta (f_{ec}{}^b g^{ae} + f_{ec}{}^a g^{eb} - E_c^\varepsilon \partial_\varepsilon g^{ab}) = 0.$$

This condition is fulfilled since g^{ab} is the Cartan metric. Therefore, the metric (K.19) is form-invariant under the transformations generated by the Killing vectors E_a^α.

Let us now find the curvature of the Riemann space V with metric (K.19). It is useful to introduce the quantity

$$h_{ab} = E_a^\alpha E_b^\beta g_{\alpha\beta} \tag{K.20}$$

which satisfies the relations $h_{ab}h^{bc} = h_a^c$ and $h_a^b E_b^\alpha = E_a^\alpha$. Thus, h is a projector and acts as the identity on the Killing vectors. If the subgroup H of G is trivial, then h_{ab} is the Cartan metric, otherwise it is a restriction of g_{ab} on the coset space. The group indices (a, α) are raised and lowered with the Cartan metric g_{ab} and the Killing metric $g_{\alpha\beta}$, respectively.

To calculate the curvature we use the Lie equation and the Killing equation, written in the covariant form:

$$E_a^\alpha \nabla_\alpha E_b^\beta - E_b^\alpha \nabla_\alpha E_a^\beta = f_{ab}{}^c E_c^\beta$$
$$\nabla_\alpha E_{b\beta} + \nabla_\beta E_{b\alpha} = 0.$$

Multiplying the Lie equation by $E^{b\gamma}$, using the Killing equation and the condition $\nabla_\gamma g_{\alpha\beta} = 0$, we obtain the relation $h_a^b \nabla^\beta E_b^\gamma - \nabla^\gamma E_a^\beta = f_a{}^{bc} E_b^\gamma E_c^\beta$, which implies

$$\nabla_\beta E_{a\gamma} = f_e{}^{bc}(\delta_a^e - \tfrac{1}{2}h_a^e) E_{b\gamma} E_{c\beta}.$$

A repeated use of this equation yields

$$\begin{aligned}\nabla_\alpha \nabla_\beta E_{a\gamma} &= (f_a{}^{bc} - \tfrac{1}{2} f_e{}^{bc} h_a^e)(\nabla_\alpha E_{b\gamma}) E_{c\beta} - (\beta \leftrightarrow \gamma)\\ &= (f_a{}^{bc} - \tfrac{1}{2} f_e{}^{bc} h_a^e)(f_b{}^{pq} - \tfrac{1}{2} f_m{}^{pq} h_b^m) E_{p\gamma} E_{q\alpha} E_{c\beta} - (\beta \leftrightarrow \gamma).\end{aligned}$$

We now use the maximal symmetry of V to simplify further calculations by going over to the isotropy point x_0. At this point we have $(E_{\bar a}^\alpha, E_{a'}^\alpha) = (0, \delta_{a'}^\alpha)$, so that $h_{ab} = \delta_a^{a'} \delta_b^{b'} g_{a'b'}$ is the restriction of the Cartan metric on the coset space. As a consequence,

$$(\nabla_\alpha \nabla_\beta E_{a\gamma})_0 = \tfrac{1}{2} f_{ab\beta} f^b{}_{\gamma\alpha} - \tfrac{1}{4} f_{ab'\beta} f^{b'}{}_{\gamma\alpha} - (\beta \leftrightarrow \gamma)$$

so that, after antisymmetrizing in α and β and replacing $a \to \varepsilon$, we obtain

$$(R_{\varepsilon\gamma\beta\alpha})_0 = \tfrac{1}{2} f_{\varepsilon\gamma b} f_{\beta\alpha}{}^b - \tfrac{1}{2} f_{\varepsilon\gamma b'} f_{\beta\alpha}{}^{b'} + \tfrac{1}{4}(f_{\varepsilon\beta b'} f_{\gamma\alpha}{}^{b'} - f_{\varepsilon\alpha b'} f_{\gamma\beta}{}^{b'}). \quad \text{(K.21)}$$

This expression further simplifies in those cases in which $f_{a'b'}{}^{c'} = 0$ (involutive algebras), which implies $f_{a'ce} f_{b'}{}^{ce} = 2 f_{a'c'e} f_{b'}{}^{c'e}$. Then,

$$(R_{\alpha\gamma})_0 = \tfrac{1}{2} g_{\alpha\gamma} \qquad R = \tfrac{1}{2} \dim(G/H)$$

so that the scalar curvature of the coset space G/H is completely determined by its dimension.

Exercises

1. Show that in Riemann spaces the Lie derivatives of u^α, $g_{\alpha\beta}$, and the Lie equation can be written in the covariant form:

$$\mathcal{L}_{\mathrm{E}} u^\alpha(x) = E^\beta \nabla_\beta u^\alpha - u^\beta \nabla_\beta E^\alpha$$
$$\mathcal{L}_{\mathrm{E}} g_{\alpha\beta} = \nabla_\alpha E_\beta + \nabla_\beta E_\alpha$$
$$E_a^\alpha \nabla_\alpha E_b^\beta - E_b^\alpha \nabla_\alpha E_a^\beta = f_{ab}{}^c E_c^\beta.$$

2. Prove the following statements:
   (a) The rotation group $SO(2)$ is compact and infinitely connected, while the group of one-dimensional translations T_1 is non-compact and simply connected.
   (b) These groups are locally isomorphic and $SO(2) = T_1/N$, where N is the discrete invariant subgroup of T_1 equal to the additive group of all integers.
3. The generators of $SU(2)$ and $SO(3)$ are given by the expressions $\tau_a = \mathrm{i}\sigma^a/2$ and $(T_a)^b{}_c = \varepsilon_{abc}$, respectively.
   (a) Find the matrix form of the following finite transformations: $A_1 = \exp(\theta\tau_1)$ and $R_1 = \exp(\theta T_1)$.
   (b) Show that the mapping $A_1(\theta) \mapsto R_1(\theta)$ is a 2–1 homomorphism of two one-parameter subgroups.
   (c) Repeat the same analysis for A_3 and R_3.
4. A matrix R of the rotation group $SO(3)$ is given in terms of the Euler angles as $R = R_3(\psi)R_1(\theta)R_3(\phi)$.
   (a) Define the vector $\boldsymbol{e}_3$ at the point R as the tangent vector to the continuous curve $C_3(t) = RR_3(t)$; show that $\boldsymbol{e}_3 = \partial/\partial\phi$.
   (b) Find the tangent vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ to the curves $C_1(t) = RR_1(t)$ and $C_2(t) = RR_2(t)$, respectively.
   (c) Prove the commutation relations $[\boldsymbol{e}_a, \boldsymbol{e}_b] = -\varepsilon_{abc}\boldsymbol{e}_c$.
5. (a) Show that the Maurer–Cartan equation is invariant under the local transformations $\delta_\lambda \boldsymbol{w} = \boldsymbol{d}\boldsymbol{\lambda} + [\boldsymbol{w}, \boldsymbol{\lambda}]$, where $\boldsymbol{\lambda} = \lambda^a T_a$ is a 1-form.
   (b) Introducing the variable ε^α by $\lambda^a = \varepsilon^\alpha E^a_\alpha$, prove the equations:

$$\delta_\varepsilon E^a_\alpha = \varepsilon^\gamma \partial_\gamma E^a_\alpha + E^a_\gamma \partial_\alpha \varepsilon^\gamma$$
$$\delta_\varepsilon g_{\alpha\beta} = \partial_\alpha \varepsilon^\gamma g_{\gamma\beta} + \partial_\beta \varepsilon^\gamma g_{\gamma\alpha} + \varepsilon^\gamma \partial_\gamma g_{\alpha\beta}.$$

6. Use the Cartan's second equation of structure to calculate the curvature tensor of the Lie group.
7. Let V be the homogeneous space of a group G, and H_0 the isotropy group of a point x_0 in V. Show that:
   (a) V is isotropic about every point; and
   (b) isotropy groups H_x of different points x in V are homomorphic to each other.
8. Show that the sphere S_2 is a space of constant curvature (by calculating its curvature tensor).
9. (a) Calculate the quantity h_{ab}, equation (K.20), on the unit sphere S_2. Compare h_{ab} at the point $\theta = 0$ with the Cartan metric g_{ab} for $SO(3)$.
   (b) Calculate the Killing metric (K.19) on the unit sphere S_2.
10. (a) Find the metric of the sphere S_3, equation (K.17), in the usual spherical coordinates (r, θ, φ).

(b) For $\kappa^2 > 0$, find the length of the circumference of the 'circle' $r, \theta =$ constant, and the length of the 'radius' $\theta, \varphi =$ constant. Compare their ratio to the Euclidean value 2π.

(c) Calculate the volume of the sphere S_3.

11. An arbitrary element A of $SU(2)$, given in example 2, can be expressed in terms of the Euler-like coordinates (θ, ψ, ϕ):

$$u_0 + \mathrm{i}u_3 = \cos\theta \mathrm{e}^{\mathrm{i}(\psi+\phi)} \qquad u_2 + \mathrm{i}u_1 = \mathrm{i}\sin\theta \mathrm{e}^{\mathrm{i}(\psi-\phi)}.$$

(a) Calculate the Killing bilinear form $(\boldsymbol{w}, \boldsymbol{w})$, where $\boldsymbol{w} = A^{-1}\boldsymbol{d}A$.

(b) Find the interval $\mathrm{d}s^2$ on S_3 in the same coordinate system using (K.17), and compare the result with $(\boldsymbol{w}, \boldsymbol{w})$.

12. Let G/H be a factor group, with the multiplication rule defined in the usual way. Show that the identity element of G/H is the coset $[h]$ which contains the invariant subgroup H.

13. The projective space RP_3 can be considered as a set of lines in E_4 passing through the origin. Prove the following statements:

(a) RP_3 is a homogeneous space of the group $O(4)$.

(b) Consider the line x_0 in RP_3 determined by the vector $(1, 0, 0, 0)$. Prove that the isotropy group of x_0 is $O(1) \times O(3)$, and $RP_3 = O(4)/O(1) \times O(3)$.

Appendix L

Chern–Simons gravity in three dimensions

In four dimensions, gravity is similar but not equivalent to the standard gauge theory (Regge 1986, Bañados *et al* 1996). In three dimensions, there is an equivalence between gravity and an ordinary gauge theory with a specific interaction of the Chern–Simons type (Witten 1988). This equivalence is helpful for understanding the structure of the classical solutions, which is crucial for the formulation of a quantum theory (Bañados 1999a, b).

Three-dimensional PGT. In three dimensions, PGT can be constructed in a complete analogy with the four-dimensional case. The Poincaré group $P(1,2)$ is the isometry group of the three-dimensional Minkowski space M_3 with metric $\eta = (+,-,-)$. Its generators M_{ij} and P_i satisfy the Lie algebra (2.6). For $D = 3$, it is convenient to replace M_{ij} with $J_i = -\frac{1}{2}\varepsilon_{ijk}M^{jk}$, whereupon the Lie algebra of $P(1,2)$ takes the form

$$[J_i, J_j] = \varepsilon_{ijk}J^k \qquad [J_i, P_j] = \varepsilon_{ijk}P^k \qquad [P_i, P_j] = 0. \tag{L.1}$$

The fact that this is $P(1,2)$, and not $P(3)$, can be seen from the property that the indices are raised and lowered with the Lorentz metric.

For arbitrary dimension D, an invariant bilinear expression in the Poincaré generators is expected to have the general form $W = xM_{ij}M^{ij} + yP_iP^i$. However, it does not commute with P_k unless we set $x = 0$. Hence, the bilinear form on the Poincaré Lie algebra is, in general, degenerate. A special feature of $D = 3$ is that in that case there exists a *non-degenerate* bilinear form

$$W_1 = -\tfrac{1}{2}\varepsilon_{ijk}P^iM^{jk} = P^iJ^k\eta_{ik} \tag{L.2a}$$

which can be used to define the metric on the Lie algebra of $P(1,2)$:

$$\gamma(P_i, J_j) = \eta_{ij} \qquad \gamma(P_i, P_j) = \gamma(J_i, J_j) = 0. \tag{L.2b}$$

The basic gravitational variables in the three-dimensional PGT are the *triad* field $b^i{}_\mu$ and the Lorentz connection $A^{ij}{}_\mu$. Their transformation laws are given by

$$\delta_0 b^k{}_\mu = \theta^k{}_s b^s{}_\mu - \xi^\rho{}_{,\mu} b^k{}_\rho - \xi^\rho \partial_\rho b^k{}_\mu$$
$$\delta_0 A^{ij}{}_\mu = -\nabla_\mu \theta^{ij} - \xi^\rho{}_{,\mu} A^{ij}{}_\rho - \xi^\rho \partial_\rho A^{ij}{}_\mu.$$

The related field strengths $T^i{}_{\mu\nu}$ and $R^{ij}{}_{\mu\nu}$ are geometrically identified with the torsion and the curvature. The geometric content of the three-dimensional PGT is described by the Riemann–Cartan space U_3.

Following the replacement of M_{ij} by J_i, we now introduce

$$\omega_{i\mu} = -\tfrac{1}{2}\varepsilon_{ijk} A^{jk}{}_\mu \qquad \tau_i = -\tfrac{1}{2}\varepsilon_{ijk}\theta^{jk}$$
$$R_{i\mu\nu} = -\tfrac{1}{2}\varepsilon_{ijk} R^{jk}{}_{\mu\nu}. \tag{L.3}$$

Then the transformation laws of the gauge fields take the form

$$\delta_0 b^i{}_\mu = -\varepsilon^{ijk} b_{j\mu}\tau_k - \xi^\rho{}_{,\mu} b^i{}_\rho - \xi \cdot \partial b^i{}_\mu$$
$$\delta_0 \omega_{i\mu} = -\nabla_\mu \tau_i - \xi^\rho{}_{,\mu}\omega_{i\rho} - \xi \cdot \partial \omega_{i\mu} \tag{L.4}$$

where $\nabla_\mu \tau_i = \partial_\mu \tau_i + \varepsilon_{ijk}\omega^j{}_\mu \tau^k$, and the field strengths are given by the expressions

$$R_{i\mu\nu} = \partial_\mu \omega_{i\nu} - \partial_\nu \omega_{i\mu} + \varepsilon_{ijk}\omega^j{}_\mu \omega^k{}_\nu$$
$$T^i{}_{\mu\nu} = \partial_\mu b^i{}_\nu - \partial_\nu b^i{}_\mu + \varepsilon^{ijk}(\omega_{j\mu} b_{k\nu} + b_{j\mu}\omega_{k\nu}). \tag{L.5}$$

GR in three dimensions. In a three-dimensional spacetime manifold $\mathcal{M}_3$, we can use the identity

$$bR = \tfrac{1}{2}\varepsilon^{\mu\nu\rho}_{ijk} b^i{}_\mu R^{jk}{}_{\nu\rho} = -\varepsilon^{\mu\nu\rho} b^i{}_\mu R_{i\mu\nu} \tag{L.6a}$$

to rewrite the Einstein–Hilbert action in the form

$$I_0 = -\int \mathrm{d}^3x\, bR = \int \mathrm{d}^3x\, \varepsilon^{\mu\nu\rho} b^i{}_\mu R_{i\mu\nu}. \tag{L.6b}$$

The related equations of motion are

$$F_i{}^\mu \equiv \varepsilon^{\mu\nu\rho} R_{i\nu\rho} = 0 \qquad G_i{}^\mu \equiv \varepsilon^{\mu\nu\rho} T^i{}_{\nu\rho} = 0.$$

Since both the curvature and the torsion vanish, the spacetime is the flat, Minkowski space M_3.

Since the isometry group of M_3 is $P(1,2)$, we shall now try to discover whether this theory can be described as an *ordinary* gauge theory of $P(1,2)$. We begin by introducing the gauge field as a Lie algebra valued 1-form:

$$A_\mu = b^i{}_\mu P_i + \omega^i{}_\mu J_i. \tag{L.7}$$

The gauge transformation of A_μ is given by

$$\delta_0 A_\mu = -\nabla_\mu u = -\partial_\mu u - [A_\mu, u]$$

where $u = \xi^i P_i + \tau^i J_i$ is an infinitesimal gauge parameter. Upon calculating the previous expression in components, we find

$$\begin{aligned}\delta_0 b^i{}_\mu &= -\nabla_\mu \xi^i - \varepsilon^{ijk} b_{j\mu} \tau_k \\ \delta_0 \omega^i{}_\mu &= -\nabla_\mu \tau^i .\end{aligned} \tag{L.8}$$

Next we calculate the field strength:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu] = P_i T^i{}_{\mu\nu} + J_i R^i{}_{\mu\nu}. \tag{L.9}$$

While the form of the field strength coincides with PGT expressions (L.5), the transformation laws (L.8), at first sight, do not have much in common with the usual transformations (L.4). The terms proportional to τ^k, describing the local Lorentz rotations, have the correct form but there seems to be a problem with the terms proportional to ξ^k, which should describe local translations. However, if we introduce the new parameters ε^i and ξ^ρ,

$$\varepsilon^i = \tau^i - \xi^\rho \omega^i{}_\rho \qquad \xi^i = b^i{}_\rho \xi^\rho \tag{L.10a}$$

then these gauge transformations can be written in the form

$$\begin{aligned}\delta_0 b^i{}_\mu &= -\varepsilon^{ijk} b_{j\mu} \varepsilon_k - \xi^\rho{}_{,\mu} b^i{}_\rho - \xi \cdot \partial b^i{}_\mu - \xi^\rho T^i{}_{\mu\rho} \\ \delta_0 \omega^i{}_\mu &= -\nabla_\mu \varepsilon^i - \xi^\rho{}_{,\mu} \omega^i{}_\rho - \xi \cdot \partial \omega^i{}_\mu - \xi^\rho R^i{}_{\mu\rho}\end{aligned} \tag{L.10b}$$

Thus, we see that the gauge transformations (L.8) are *equivalent* to the usual PGT transformations (L.4) *on shell*. This property is of basic importance for the simplicity of the three-dimensional gravity.

GR as Chern–Simons theory. We now wish to show that three-dimensional GR, as described by action (L.6), is equivalent to the Chern–Simons gauge theory for $P(1,2)$. If the gauge group is $G = P(1,2)$, we start by constructing the Pontyagin topological invariant on a four-dimensional manifold $\mathcal{M}_4$, with the metric defined by the invariant non-degenerate quadratic form (L.2):

$$I_P^0 = \int T^i R^j \eta_{ij} = \tfrac{1}{4} \int_{\mathcal{M}_4} \mathrm{d}^4 x\, \varepsilon^{\mu\nu\lambda\rho} T^i{}_{\mu\nu} R^j{}_{\lambda\rho} \eta_{ij} .$$

The integrand in this equation is a total derivative of the Chern–Simons three-form $L^0_{\mathrm{CS}} = 2\theta^i R_i$, with $\theta^i = b^i{}_\mu \, \mathrm{d}x^\mu$. This form defines the Chern–Simons action on a three-dimensional manifold $\mathcal{M}$:

$$I^0_{\mathrm{CS}} = \int_{\mathcal{M}} L^0_{\mathrm{CS}} = \int_{\mathcal{M}} \mathrm{d}^3 x\, \varepsilon^{\nu\lambda\rho} b^i{}_\nu R_{i\lambda\rho} . \tag{L.11}$$

A direct comparison with equation (L.6) shows that this is precisely the GR action. Thus, we have shown that three-dimensional GR can be interpreted as Chern–Simons gauge theory.

Adding a cosmological constant. We now wish to generalize the previous discussion of three-dimensional GR by including a cosmological constant. The generalized theory, GR_Λ, is defined by the action (in units $16\pi G = 1$)

$$\begin{aligned} I_1 &= -\int d^3x\, b(R + 2\Lambda) \\ &= \int d^3x\, \varepsilon^{\mu\nu\rho}(b^i{}_\mu R_{i\nu\rho} - \tfrac{1}{3}\Lambda\varepsilon_{ijk}b^i{}_\mu b^j{}_\nu b^k{}_\rho). \end{aligned} \tag{L.12}$$

The action is invariant under Poincaré gauge transformations (L.4), and the field equations of the theory are:

$$F_i{}^\mu \equiv \varepsilon^{\mu\nu\rho}(R_{i\nu\rho} - \Lambda\varepsilon_{ijk}b^i{}_\nu b^k{}_\rho) = 0 \qquad G_i{}^\mu \equiv \varepsilon^{\mu\nu\rho}T^i{}_{\nu\rho} = 0.$$

The second equation tells us that the torsion vanishes, so that ω is the Levi-Civita connection. Rewriting the first equation in the form

$$F_i{}^\mu = -\tfrac{1}{2}\varepsilon^{\mu\nu\rho}_{ijk}(R^{kl}{}_{\nu\rho} + 2\Lambda b^k_\nu b^l{}_\rho) = 0$$

we see that the spacetime is not flat but has a constant curvature: $R = -6\Lambda$. Depending on whether the sign of Λ is negative or positive, we have either an anti de Sitter or a de Sitter spacetime, respectively. The isometry groups of these spaces are $SO(2,2)$ and $SO(1,3)$.

If three-dimensional GR without a cosmological constant is related to the gauge theory of $P(1,2)$, we could naturally expect that GR_Λ might be described as the gauge theory of $SO(2,2)$ or $SO(1,3)$. Let us investigate this assumption for the anti de Sitter group. We begin by noting that the Lie algebra of $SO(2,2)$ can be written as a generalization of the $P(1,2)$ algebra $(L.1)$:

$$[J_i, J_j] = \varepsilon_{ijk}J^k \qquad [J_i, P_j] = \varepsilon_{ijk}P^k \qquad [P_i, P_j] = \lambda\varepsilon_{ijk}J^k \tag{L.13}$$

where $\lambda \equiv 1/l^2 > 0$. This form is obtained from the usual $SO(2,2)$ algebra by making the replacements: $J_i = -\frac{1}{2}\varepsilon_{ijk}M^{jk}$, $P_i = M_{i4}/l$ (see appendix C). After introducing the gauge field as in equation (L.7), we find the generalized transformation laws,

$$\begin{aligned} \delta_0 b^i{}_\mu &= -\nabla_\mu\xi^i - \varepsilon^{ijk}b_{j\mu}\tau_k \\ \delta_0\omega^i{}_\mu &= -\nabla_\mu\tau^i - \lambda\varepsilon^{ijk}b_{j\mu}\xi_k \end{aligned} \tag{L.14}$$

and the following formula for the field strength:

$$F_{\mu\nu} = P_i T^i{}_{\mu\nu} + J_i(R^i{}_{\mu\nu} + \lambda\varepsilon^{ijk}b_{j\mu}b_{k\nu}). \tag{L.15}$$

Expression (L.2a) is an invariant quadratic form on the generalized Lie algebra (L.13), which can be used to construct the related Chern–Simons action. Starting from the Pontryagin topological invariant,

$$I_P^1 = \tfrac{1}{4}\int_{\mathcal{M}_4} d^4x\, \varepsilon^{\mu\nu\lambda\rho}T^i{}_{\mu\nu}(R_{i\lambda\rho} + \lambda\varepsilon_{imn}b^m{}_\lambda b^n{}_\rho)$$

we find that the related Chern–Simons action $I^1_{\rm CS}$ is precisely the action (L.12), which includes the cosmological constant, $\Lambda = -\lambda$. Thus, GR_Λ with a negative cosmological constant can be interpreted as the Chern–Simons gauge theory of anti de Sitter group $SO(2,2)$. The field equations of theory (L.12) assert that both components of the generalized field strength $F_{\mu\nu}$ vanish.

More on the Chern–Simons formulation. It should be pointed out that, in addition to (L.2a), a *second* invariant bilinear form exists on the $SO(2,2)$ Lie algebra:

$$W_2 = J^i J^j \eta_{ij} + \lambda P^i P^j \eta_{ij} \tag{L.16a}$$

which defines the metric

$$\gamma(J_i, J_j) = \eta_{ij} \qquad \gamma(J_i, P_j) = 0 \qquad \gamma(P_i, P_j) = \lambda\eta_{ij}. \tag{L.16b}$$

This form exists for general D. For $\lambda = 0$ it becomes degenerate, which is the reason why we discarded it when we considered quadratic bilinear forms on $P(1,2)$. The existence of two non-degenerate quadratic forms for $D = 3$ is a consequence of the isomorphism $SO(2,2) \simeq SO(1,2) \times SO(1,2)$ (note that $SO(1,2) \simeq SL(2,R)$).

Now, we can use the quadratic form (L.16) to construct the Pontryagin topological invariant, and derive the form of the *new* Chern–Simons action:

$$\begin{aligned} I_2 = \int \mathrm{d}^3x\, \varepsilon^{\nu\lambda\rho}[&\omega^i{}_\nu(\partial_\lambda\omega_{i\rho} - \partial_\rho\omega_{i\lambda} + \tfrac{2}{3}\varepsilon_{imn}\omega^m{}_\lambda\omega^n{}_\rho) \\ &+ 2\lambda b^i{}_\nu(\partial_\lambda b_{i\rho} + \lambda\varepsilon_{imn}\omega^m{}_\lambda b^n{}_\rho)]. \end{aligned} \tag{L.17}$$

This action is also invariant under gauge transformations (L.14); hence, we can consider $I_1 + \alpha I_2$ as the general $SO(2,2)$ gauge theory. It is interesting to observe that, for generic values of α, the classical field equations remain the same as for the original action I_1.

The existence of two actions for $SO(2,2)$ gauge theory can be clarified by introducing the new basis of generators,

$$G_{\pm i} = \frac{1}{2}\left(J_i \pm \frac{1}{\sqrt{\lambda}}P_i\right) \tag{L.18a}$$

in terms of which the Lie algebra (L.13) takes the simple form:

$$[G_{\pm i}, G_{\pm j}] = \varepsilon_{ijk}G^k_\pm \qquad [G_{\pm i}, G_{\mp j}] = 0. \tag{L.18b}$$

For $\lambda > 0$, this is the Lie algebra of $SO(1,2)\times SO(1,2)$. Then, we can introduce the related gauge fields,

$$A_\mu = A^{+i}_\mu G_{+i} + A^{-i}_\mu G_{-i} \qquad A^\pm_{i\mu} \equiv \omega_{i\mu} \pm \sqrt{\lambda}b_{i\mu} \tag{L.19a}$$

and find their transformation laws:

$$\delta_0 A^{\pm}_{i\mu} = -\partial_\mu \tau_i - \varepsilon_i{}^{jk} A^{\pm}_{j\mu}\tau_k - \xi^\rho{}_{,\mu} A^{\pm}_{i\rho} - \xi \cdot \partial A^{\pm}_{i\mu}. \tag{L.19b}$$

Two Chern–Simons actions, corresponding to the two commuting $SO(1,2)$ Lie algebras in (L.18), take the simple forms:

$$I_{\mathrm{CS}}[A] = k\int (A^i dA^j + \tfrac{1}{3}\varepsilon_{mn}{}^i A^m A^n A^j)\eta_{ij} \qquad A = A^+, A^-. \tag{L.20}$$

In order to clarify the relation between $I_{\mathrm{CS}}[A^+]$, $I_{\mathrm{CS}}[A^-]$ and the actions I_1, I_2 constructed earlier, we introduce the fields

$$A^i = \omega^i + x\theta^i \qquad \bar{A}^i = \omega^i - x\theta^i$$

where x is a complex number. Then, we can prove the following identity:

$$2\theta^i R_i + \frac{x^2}{3}\varepsilon_{ijk}\theta^i\theta^j\theta^k = \frac{1}{2x}[L_{\mathrm{CS}}(A) - L_{\mathrm{CS}}(\bar{A})] + d\Delta \tag{L.21}$$

where $L_{\mathrm{CS}}(A) = A^i\,\mathrm{d}A_i + \frac{1}{3}\varepsilon_{ijk}A^iA^jA^k$, and $\mathrm{d}\Delta$ is a total derivative term. This relation is true for both positive and negative x^2, and does not depend on the signature of the spacetime metric. For real x we can take $x = 1/l$, so that the left-hand side coincides with action (L.12), which is the generalized Einstein action in units $16\pi G = 1$. Returning to the standard normalization, and continuing to use the notation $(A, \bar{A})$ instead of (A^+, A^-), we can write

$$I_\Lambda \equiv \frac{1}{16\pi G} I_1 = I_{\mathrm{CS}}[A] - I_{\mathrm{CS}}[\bar{A}] \tag{L.22a}$$

where the Chern–Simons coupling constant is given by

$$4\pi k = \frac{l}{8G}. \tag{L.22b}$$

The relations (L.22) are of crucial importance for the Chern–Simons interpretation of GR_Λ. The other 'exotic' action (L.17) has the form $kI_2 = I_{\mathrm{CS}}[A] + I_{\mathrm{CS}}[\bar{A}]$.

The anti de Sitter space. Three-dimensional GR with a cosmological constant is described by action (L.12). The field equations in vacuum,

$$R^{kl}{}_{\nu\rho} + \Lambda(b^k{}_\nu b^l{}_\rho - b^k{}_\rho b^l{}_\nu) = 0$$

describe a symmetric space of constant curvature $R = -6\Lambda$. The space of constant curvature with $R > 0$ ($\Lambda < 0$) is called the anti de Sitter space (or de

Sitter space of the second kind). In three dimensions, it has the topology $S^1 \times R^2$, and can be represented as a hypersurface:

$$H_3: \quad \eta_{\mu\nu} y^\mu y^\nu + z^2 = l^2 \qquad l^2 > 0 \qquad (\mu, \nu = 0, 1, 2)$$

in a four-dimensional flat space M_4 with the metric $\eta = (+, -, -, +)$. The anti de Sitter space has the signature $(+, -, -)$, and $R = 6/l^2$ is positive. By construction, the isometry group of the anti de Sitter space is $SO(2, 2)$ and the Killing vectors are $K_{ab} = y_a \partial_b - y_b \partial_a$, where $y^a = (y^\mu, z)$.

The metric of H_3 can be written as $\mathrm{d}s^2 = \eta_{\mu\nu}\,\mathrm{d}y^\mu\,\mathrm{d}y^\nu + \mathrm{d}z^2$, where (y^μ, z) lies on H_3. Introducing the pseudo-spherical coordinates (τ, χ, ϕ) by

$$\begin{aligned} y^0 &= l \sin\tau & \qquad y^1 &= l \cos\tau \sinh\chi \cos\phi \\ z &= l \cos\tau \cosh\chi & \qquad y^2 &= l \cos\tau \sinh\chi \sin\phi \end{aligned}$$

with $-\pi \le \tau < \pi$, $0 \le \chi < \infty$, $0 \le \phi < 2\pi$, the metric of H_3 takes the form

$$\mathrm{d}s^2 = l^2[\mathrm{d}\tau^2 - \cos^2\tau(\mathrm{d}\chi^2 + \sin^2\chi\,\mathrm{d}\phi^2)].$$

These coordinates do not cover the whole space and the metric has coordinate singularities at $\tau = \pm\pi/2$. Leaving the metric unchanged, we can replace the S^1 time $\tau \in [-\pi, \pi]$ by a new, R^1 time $\tau \in (-\infty, +\infty)$, thus changing the topology from $S^1 \times R^2$ to R^3. The space obtained in this way is the universal covering of anti de Sitter space. Following the common terminology, this universal covering space will be called the anti de Sitter space, and denoted by AdS_3 (Hawking and Ellis 1973).

The whole AdS_3 can be covered by coordinates (t, ρ, φ),

$$\begin{aligned} y^0 &= l \cosh\rho \cos t & \qquad y^1 &= l \sinh\rho \cos\varphi \\ z &= l \cosh\rho \sin t & \qquad y^2 &= l \sinh\rho \sin\varphi \end{aligned}$$

$(-\infty < t < \infty, 0 \le \rho < \infty, 0 \le \varphi < 2\pi)$ in which the metric takes the form

$$\mathrm{d}s^2 = l^2[\mathrm{d}t^2 \cosh^2\rho - (\mathrm{d}\rho^2 + \sinh^2\rho\,\mathrm{d}\varphi^2)]. \tag{L.23a}$$

Introducing $r = l \sinh\rho$ and $t \to t/l$, we obtain

$$\mathrm{d}s^2 = (1 + r^2/l^2)\,\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{1 + r^2/l^2} - r^2\,\mathrm{d}\varphi^2 \tag{L.23b}$$

which is the standard static form of the AdS_3 metric. Another change of coordinates, defined by

$$\theta = 2\,\mathrm{Arctan}(\mathrm{e}^\rho) - \pi/2 \qquad 0 \le \theta < \pi/2$$

is suitable for studying the structure of AdS_3 at infinity. It leads to $\cosh\rho = 1/\cos\theta$, $\sinh\rho = \tan\theta$, and

$$\mathrm{d}s^2 = \frac{1}{\sin^2\theta}[\mathrm{d}t^2 - l^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2)].$$

Thus, the whole AdS_3 space is conformal to the region $0 \le \theta < \pi/2$ of the static 'cylinder'.

Since AdS_3 is a maximally symmetric space, it must be locally isometric to any other solution having the same curvature and signature. As a simple illustration, consider the metric

$$\mathrm{d}s^2 = f^2\,\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{f^2} - r^2\,\mathrm{d}\varphi^2 \qquad f^2 = -8MG + \frac{r^2}{l^2} \tag{L.24}$$

which describes a spherically symmetric *black hole*, with an event horizon at $r = r_0$, where $r_0^2 = 8MGl^2$. For $r > r_0$, we can introduce a coordinate transformation

$$\begin{aligned} y^0 &= \sqrt{A(r)}\cosh(r_0\varphi/l) \qquad & y^2 &= \sqrt{B(r)}\cosh(r_0 t/l^2) \\ y^1 &= \sqrt{A(r)}\sinh(r_0\varphi/l) \qquad & z &= \sqrt{B(r)}\sinh(r_0 t/l^2) \end{aligned}$$

where $A = (l^2/r_0^2)r^2$, $B = (l^2/r_0^2)(r^2 - r_0^2)$, which transforms the AdS_3 metric into the black hole metric. Note, however, that the periodicity in φ requires us to identify points in AdS_3 obtained by $\varphi \to \varphi + 2\pi$. Thus, although AdS_3 and the black hole are isometric solutions, they are topologically distinct: the black hole is obtained from AdS_3 by a process of identification, which does not influence the local properties of the solutions. By identifying two points on a curve in AdS_3, we define new closed curves in the quotient space, that should not be timelike or null. Similar considerations also hold for axially symmetric black holes (Bañados *et al* 1993).

Example 1. To see how the geometric structure of the three-dimensional gravity can be related to the Chern–Simons picture, consider the AdS_3 interval (L.23*a*). A natural choice for the triads is

$$\theta^0 = l\cosh\rho\,\mathrm{d}t \qquad \theta^1 = l\,\mathrm{d}\rho \qquad \theta^2 = l\sinh\rho\,\mathrm{d}\varphi.$$

The vanishing of torsion, $\mathrm{d}\theta^i + \varepsilon^{ijk}\omega_j\theta_k = 0$, yields the connection components:

$$\omega^0 = -\cosh\rho\,\mathrm{d}\varphi \qquad \omega^1 = 0 \qquad \omega^2 = -\sinh\rho\,\mathrm{d}t.$$

The Chern–Simons gauge field $A^i = \omega^i + \theta^i/l$ is given by

$$A = \frac{1}{2}b^{-1}\begin{pmatrix} \mathrm{d}\rho & \mathrm{d}x^- \\ -\mathrm{d}x^- & -\mathrm{d}\rho \end{pmatrix} b \qquad b = \mathrm{e}^{\rho T_1} = \begin{pmatrix} \mathrm{e}^{\rho/2} & 0 \\ 0 & \mathrm{e}^{-\rho/2} \end{pmatrix}.$$

Thus, we have $A_+ = 0$, $\alpha = b^{-1}\partial_\rho b = T_1$ and $\alpha^2 = \eta_{ij}\alpha^i\alpha^j = -1$.

Similarly, $\bar{A}^i = \omega^i - \theta^i/l$ has the form

$$\bar{A} = \frac{1}{2}\bar{b}^{-1}\begin{pmatrix} \mathrm{d}\rho & -\mathrm{d}x^+ \\ \mathrm{d}x^+ & -\mathrm{d}\rho \end{pmatrix}\bar{b} \qquad \bar{b} = \mathrm{e}^{-\rho T_1}$$

hence, $\bar{A}_- = 0$, $\bar{\alpha} = -T_1$, and $\bar{\alpha}^2 = -1$.

Example 2. Consider now the black hole metric (L.24) for $M = 0$ (the black hole vacuum). After introducing the new radial coordinate $r = \mathrm{e}^{\rho}$, and rescaling the time by $t \to t/l$, the interval takes the form

$$\mathrm{d}s^2 = \mathrm{e}^{2\rho}(\mathrm{d}t^2 - \mathrm{d}\varphi^2) - l^2\,\mathrm{d}\rho^2.$$

The triad field and the related connection are given by

$$\begin{aligned} &\theta^0 = \mathrm{e}^{\rho}\,\mathrm{d}t \qquad \theta^1 = l\,\mathrm{d}\rho \qquad \theta^2 = \mathrm{e}^{\rho}\,\mathrm{d}\varphi \\ &\omega^0 = -\frac{1}{l}\mathrm{e}^{\rho}\,\mathrm{d}\varphi \qquad \omega^1 = 0 \qquad \omega^2 = -\frac{1}{l}\mathrm{e}^{\rho}\,\mathrm{d}t. \end{aligned}$$

The Chern–Simons gauge fields are:

$$A = \frac{1}{2}\begin{pmatrix} \mathrm{d}\rho & 0 \\ -(2/l)\mathrm{e}^{\rho}\,\mathrm{d}x^- & -\mathrm{d}\rho \end{pmatrix} \qquad \bar{A} = \frac{1}{2}\begin{pmatrix} -\mathrm{d}\rho & -(2/l)\mathrm{e}^{\rho}\,\mathrm{d}x^+ \\ 0 & \mathrm{d}\rho \end{pmatrix}$$

hence, $b = \mathrm{e}^{\rho T_1}$, $\bar{b} = \mathrm{e}^{-\rho T_1}$, as in the previous example.

Asymptotic symmetry. In $D = 3$, GR_Λ has no local dynamical degrees of freedom, so that the spacetime outside localized matter sources is described by the vacuum solutions of the field equations. Matter has no influence on the local geometry in the source-free regions, it can only change the global properties of spacetime.

For $\Lambda < 0$, the spacetime is locally described as AdS_3, but its global properties may be different. Brown and Henneaux (1986*b*) considered a large class of metrics that have the same leading asymptotic terms as AdS_3:

$$\mathrm{d}s^2 \sim \frac{r^2}{l^2}\,\mathrm{d}t^2 - \frac{l^2}{r^2}\,\mathrm{d}r^2 - r^2\,\mathrm{d}\varphi^2 \qquad \text{(for large } r\text{)}.$$

It seems natural to say that such solutions are asymptotically anti de Sitter. More precisely, a metric $g_{\mu\nu}$ is said to be *asymptotically anti de Sitter* if

(a) $g_{\mu\nu}$ approaches the AdS_3 metric for $r \to \infty$; and
(b) it is invariant under the action of the anti de Sitter group $SO(2,2)$.

The second condition determines the precise way in which $\mathrm{d}s^2$ approaches this asymptotic form, and ensures that we have at least $SO(2,2)$ as an asymptotic symmetry group. Acting with all possible $SO(2,2)$ transformations on this metric, we generate the following boundary conditions:

$$\begin{aligned} &g_{tt} = \frac{r^2}{l^2} + \mathcal{O}_0 \qquad g_{tr} = \mathcal{O}_3 \qquad g_{t\varphi} = \mathcal{O}_0 \\ &g_{rr} = -\frac{l^2}{r^2} + \mathcal{O}_4 \qquad g_{r\varphi} = \mathcal{O}_3 \qquad g_{\varphi\varphi} = -r^2 + \mathcal{O}_0. \end{aligned} \tag{L.25}$$

These are the most general boundary conditions compatible with the asymptotic anti de Sitter symmetry.

Having chosen the boundary conditions, we now wish to find the form of the related asymptotic symmetry transformations, i.e. to construct the infinitesimal diffeomorphisms which leave the boundary conditions invariant. In the process, we discover that the natural asymptotic symmetry is not only $SO(2,2)$, as we could have expected, but the whole group of conformal symmetries in two dimensions. The allowed diffeomorphisms are generated by vector fields $\xi^\mu = (\xi^t, \xi^r, \xi^\varphi)$ which preserve (L.25). They are found to have the form (Brown and Henneaux 1986)

$$\begin{aligned}
\xi^t &= T(t,\varphi) + \frac{1}{r^2}\bar{T}(t,\varphi) + \mathcal{O}_4 \\
\xi^\varphi &= \Phi(t,\varphi) + \frac{1}{r^2}\bar{\Phi}(t,\varphi) + \mathcal{O}_4 \\
\xi^r &= rR(t,\varphi) + \mathcal{O}_1
\end{aligned} \tag{L.26a}$$

where

$$\begin{aligned}
\partial_t T = \partial_\varphi \Phi = -R \qquad & \partial_t \Phi = \partial_\varphi T \\
2\bar{T} = -\partial_t R \qquad & 2\bar{\Phi} = \partial_\varphi R.
\end{aligned} \tag{L.26b}$$

The asymptotic symmetry defined by these vector fields is isomorphic to the *conformal group* in two dimensions. This follows from the fact that the functions (T, Φ) satisfy the conformal Killing equations in $D = 2$ ($\partial_t T = \partial_\varphi \Phi$, $\partial_t \Phi = \partial_\varphi T$), and once these functions are known we can easily calculate R, $\bar{T}$, and $\bar{\Phi}$. The asymptotic symmetry is infinite dimensional and contains $SO(2,2)$ as a subgroup.

The conformal symmetry at the boundary can also be recognized by analysing the spatial infinity. A surface at $r \to \infty$ has the induced metric

$$\mathrm{d}s_\infty^2 = \lim_{r\to\infty} \frac{1}{r^2}\,\mathrm{d}s^2 = \mathrm{d}t^2/l^2 - \mathrm{d}\varphi^2.$$

The isometry of this surface is the conformal group in $D = 2$.

This conformal structure becomes clearer when we express it in the light-cone coordinates $x^\pm = t/l \pm \varphi$. Including the first subleading terms $\gamma_{\mu\nu}$ in the asymptotic expansion (L.25) of the metric, we can write

$$\begin{aligned}
g_{+-} &= \frac{r^2}{2} + \gamma_{+-} + \mathcal{O}_1 \qquad & g_{rr} &= -\frac{l^2}{r^2} + \frac{\gamma_{rr}}{r^4} + \mathcal{O}_5 \\
g_{\pm\pm} &= \gamma_{\pm\pm} + \mathcal{O}_1 \qquad & g_{\pm r} &= \frac{\gamma_{\pm r}}{r^3} + \mathcal{O}_4.
\end{aligned} \tag{L.27a}$$

The field equations $R_{\mu\nu} = (2/l^2) g_{\mu\nu}$ imply

$$\gamma_{\pm r} = (1/4l^2)\gamma_{rr} \qquad \partial_- \gamma^{++} = 0 \qquad \partial_+ \gamma^{--} = 0.$$

After fixing the residual gauge symmetry by $\gamma_{rr} = \gamma_{\pm r} = 0$, the interval takes the form (Navaro-Salas and Navaro 1998)

$$\mathrm{d}s^2 = r^2\,\mathrm{d}x^+\mathrm{d}x^- - \frac{l^2}{r^2}\,\mathrm{d}r^2 + \gamma_{++}(\mathrm{d}x^+)^2 + \gamma_{--}(\mathrm{d}x^-)^2 + \mathcal{O}_1. \qquad \text{(L.27}b\text{)}$$

The vector fields $\xi^\pm = \xi^t/l \pm \xi^\varphi$ are given by (Strominger 1998)

$$\xi^\pm = T^\pm + \frac{l^2}{r^2}\bar{T}^\pm + \mathcal{O}_4$$

where the functions $T^\pm = T/l \pm \Phi$, $\bar{T}^\pm = \bar{T}/l \pm \bar{\Phi}$ satisfy the relations

$$\partial_\mp T^\pm = 0 \qquad \partial_\pm^2 T^\pm = 2\bar{T}^\mp.$$

Hence,

$$\begin{aligned}\xi^+ &= T^+ + \frac{l^2}{2r^2}\partial_-^2 T^- + \mathcal{O}_4\\ \xi^- &= T^- + \frac{l^2}{2r^2}\partial_+^2 T^+ + \mathcal{O}_4\\ \xi^r &= -\frac{r}{2}(\partial_+ T^+ + \partial_- T^-) + \mathcal{O}_1.\end{aligned} \qquad \text{(L.28)}$$

The subleading terms in the metric can be identified with the appropriate conformal structures at the boundary. By studying the conformal transformation properties of $\gamma_{\pm\pm}$, we can show that these fields are proportional to the energy–momentum tensor of the underlying conformal field theory with central charge $c = 3l/2G$ (Navaro-Salas and Navaro 1998).

The asymptotic conformal symmetry is best analysed by studying the PB algebra of the canonical generators. This can be done directly by performing the canonical analysis of action (L.12) but the simpler approach is to represent the action in terms of two Chern–Simons terms, and then to use the results of section 6.4. The classical PB algebra of the canonical generators at the boundary is given by the Virasoro algebra with the classical central charge $c = 3l/2G$ (Brown and Henneaux 1986).

Comments. Two particularly interesting aspects of the three-dimensional gravity are (a) that it can be formulated as a Chern–Simons gauge theory; and (b) that the general solution of the equations of motion, satisfying anti de Sitter boundary conditions, is known. Here, we wish to make some additional comments on these subjects.

(i) General solutions for Chern–Simons gauge fields which are compatible with conformal symmetry at the boundary have the form

$$A_+ = 0 \qquad A_- = b^{-1}\begin{pmatrix} 0 & L(x^-) \\ -1 & 0 \end{pmatrix} b \qquad A_r = T_1$$

and

$$A_- = 0 \qquad \bar{A}_+ = \bar{b}^{-1}\begin{pmatrix} 0 & -1 \\ \bar{L}(x^+) & 0 \end{pmatrix}\bar{b} \qquad \bar{A}_r = -T_1.$$

Here, $(b, \bar{b}) = (\mathrm{e}^{-\rho T_1}, \mathrm{e}^{\rho T_1})$, L and $\bar{L}$ are arbitrary functions of x^- and x^+, respectively. Using these solutions, we can construct first the triad and then the metric:

$$b^i{}_\mu = (l/2)(A^i{}_\mu - \bar{A}^i{}_\mu) \qquad g_{\mu\nu} = \eta_{ij} b^i{}_\mu b^j{}_\nu.$$

The interval $\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu$ represents the most general solution to the three-dimensional GR_Λ, which is asymptotically anti de Sitter. It contains two arbitrary functions, L and $\bar{L}$, so that, effectively, it describes an infinite number of solutions.

(ii) Different values of L and $\bar{L}$ correspond to physically different configurations. When L and $\bar{L}$ are constants, $L = L_0$, $\bar{L} = \bar{L}_0$, the general metric reduces to the axially symmetric black hole:

$$\mathrm{d}s^2 = l^2 N^2\,\mathrm{d}t^2 - N^{-2}\,\mathrm{d}r^2 - r^2(\mathrm{d}\varphi + N^\varphi\,\mathrm{d}t)^2$$

where

$$N^2 = -8MG + \frac{r^2}{l^2} + \frac{16G^2J^2}{r^2} \qquad N^\varphi = \frac{4GJ}{r^2}.$$

The black hole parameters M and J are defined by

$$L_0 + \bar{L}_0 = Ml \qquad L_0 - \bar{L}_0 = J.$$

For $Ml > |J|$ this metric has two horizons, defined as the solutions of the equation $N^2(r) = 0$: $r_+^2 + r_-^2 = 8MGl^2$, $r_-r_+ = 4GJl$. For $J = 0$ and $M = 0$, this metric reduces to the black hole vacuum, while for $M = -1/8G$, it goes into AdS_3 (locally). The AdS_3 solution cannot be deformed continuously into the black hole vacuum, because between $M = 0$ and $M = -1/8G$ there are naked singularities, which are excluded from the physical spectrum. The specific role played by the AdS_3 state has been clearly explained by Coussaert and Henneaux (1994).

(iii) The general solution has asymptotic conformal symmetry with central charge $c = 3l/2G$. The conformal symmetry is not an exact symmetry of any background metric—it maps one solution into the other. The associated Virasoro generators can be regarded as the symmetry generators, related to the underlying conformal field theory. (At the same time, the Virasoro algebra can be thought of as the basic PB algebra of the gauge fixed variables.) The identification of this conformal field theory is of basic importance for three-dimensional quantum gravity. In the classical domain, this theory seems to coincide with Liouville theory (Coussaert *et al* 1995).

(iv) The classical properties of three-dimensional gravity are not only interesting by themselves—understanding them is crucial for the formulation of a quantum gravity (Carlip 1995). In quantum theory, functions L and $\bar{L}$ are promoted to operators. By counting the number of quantum states that induce the classical metric of a black hole of mass M and the angular momentum J, we can obtain the formula for black hole entropy, which is closely related to the quantum properties of gravity (Strominger 1998).

Exercises

1. Prove the following formulae in M_3:

$$\varepsilon^{\mu\nu\rho}b^i{}_\mu b^j{}_\nu b^k{}_\lambda = \varepsilon^{ijk}b \qquad \varepsilon^{\mu\nu\rho}_{ijk}b^i{}_\mu = 2bh_{[j}{}^\nu h_{k]}{}^\rho.$$

2. Construct the Chern–Simons action I^0_{CS} for the group $P(1,2)$, using the Lie algebra metric (L.2*b*).
3. Construct the Chern–Simons action I^1_{CS} for the group $SO(2,2)$, using the Lie algebra metric (L.2*a*).
4. Construct the Chern–Simons action I^2_{CS} for the group $SO(2,2)$, using the Lie algebra metric (L.16*b*). Write down the corresponding field equations.
5. Find the transformation law for Chern–Simons Lagrangians $L_{\mathrm{CS}}[A]$ in equation (L.20), under gauge transformations (L.19*b*).
6. Prove the identity (L.21).
7. Show that the change of coordinates

$$y^0 = \frac{1}{2Z}(Z^2 + X^2 - T^2 + l^2) \qquad y^2 = \frac{lX}{Z}$$
$$y^1 = -\frac{1}{2Z}(Z^2 + X^2 - T^2 - l^2) \qquad z = \frac{lT}{Z}$$

transforms the AdS_3 metric into the Poincaré form:

$$\mathrm{d}s^2 = \frac{l^2}{Z^2}(\mathrm{d}T^2 - \mathrm{d}X^2 - \mathrm{d}Z^2) \qquad Z > 0.$$

8. Einstein's equations for GR_Λ have the form $R_{\mu\nu} = -2\Lambda g_{\mu\nu}$.
   (a) Write down the explicit form of these equations for a static, spherically symmetric metric:

   $$\mathrm{d}s^2 = A(r)\,\mathrm{d}t^2 - B(r)\,\mathrm{d}r^2 - r^2\,\mathrm{d}\varphi^2.$$

   (b) Derive the relation $R_{tt}/A + R_{rr}/B = 0$ and show that it leads to $(AB)' = 0$.
   (c) Use the equation $R_{\varphi\varphi} = -2\Lambda g_{\varphi\varphi}$ to conclude that $\partial_r A = -2\Lambda r(AB)$.
   (d) For $\Lambda = -1/l^2$, show that the AdS_3 metric (L.23*b*) is a solution of the field equations. Repeat the arguments for the black hole metric (L.24).

9. Consider the Chern–Simons gauge fields:

$$A = \frac{1}{2}\begin{pmatrix} \mathrm{d}\rho & 2\mathrm{e}^{-\rho}L(x^-)\,\mathrm{d}x^- \\ -2\mathrm{e}^{\rho}\,\mathrm{d}x^- & -\mathrm{d}\rho \end{pmatrix}$$
$$\bar{A} = \frac{1}{2}\begin{pmatrix} -\mathrm{d}\rho & -2\mathrm{e}^{\rho}\,\mathrm{d}x^+ \\ 0 & \mathrm{d}\rho \end{pmatrix}.$$

Find the related triad field $\theta = (l/2)(A - \bar{A})$, then calculate the metric tensor $g_{\mu\nu} = -2\,\mathrm{Tr}(\theta_\mu\theta_\nu)$, and write the interval $\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu$.

10. Calculate the Chern–Simons fields A and $\bar{A}$ for the metric

$$\mathrm{d}s^2 = r^2\,\mathrm{d}x^+\mathrm{d}x^- - \frac{l^2}{r^2}\,\mathrm{d}r^2 + \gamma(x^+)(\mathrm{d}x^+)^2 \qquad x^\pm = t/l \pm \varphi$$

after introducing $r = \mathrm{e}^\rho$.

Appendix M

Fourier expansion

In this appendix, we give a short review of the basic relations from the theory of Fourier series, which are often used in string theory.

Interval $[-\pi, \pi]$. Let $f(x)$ be a function defined on the basic interval $[-\pi, \pi]$, and periodic with period $T = 2\pi$. Then, its Fourier expansion (when it exists) is defined by

$$\begin{aligned} f(x) &= \frac{a_0}{2} + \sum_{n\geq 1}(a_n \cos nx + b_n \sin nx) \\ a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} \mathrm{d}x\, f(x)\cos nx &\qquad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} \mathrm{d}x\, f(x)\sin nx. \end{aligned} \tag{M.1}$$

This result can be easily transformed into a complex form

$$\begin{aligned} f(x) &= C_0 + \sum_{n\geq 1}(C_n \mathrm{e}^{\mathrm{i}nx} + C_n^* \mathrm{e}^{-\mathrm{i}nx}) \\ C_n &= \tfrac{1}{2}(a_n - \mathrm{i}b_n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{d}x\, f(x)\mathrm{e}^{-\mathrm{i}nx} \end{aligned}$$

which can be simplified by introducing $C_{-n} = C_n^*$:

$$f(x) = \sum_n C_n \mathrm{e}^{\mathrm{i}nx} \tag{M.2}$$

where the sum over n goes from $-\infty$ to $+\infty$.

If the function $f(x)$ is *symmetric* on the basic interval, $f(-x) = f(x)$, then $b_n = 0$ and we have

$$f(x) = \sum_n C_n \mathrm{e}^{\mathrm{i}nx} = C_0 + 2\sum_{n\geq 1} C_n \cos nx \tag{M.3}$$

since $C_n = C_{-n}$.

Interval $[0, \pi]$. Analogous relations are valid for a function $f(x)$ defined on the basic interval $[0, \pi]$, with the period $T = \pi$.

(a) Real form:

$$\begin{gathered} f(x) = \frac{a_0}{2} + \sum_{n \geq 1} (a_n \cos 2nx + b_n \sin 2nx) \\ a_n = \frac{1}{\pi/2} \int_0^\pi \mathrm{d}x\, f(x) \cos 2nx \qquad b_n = \frac{1}{\pi/2} \int_0^\pi \mathrm{d}x\, f(x) \sin 2nx. \end{gathered} \tag{M.4}$$

(b) Complex form:

$$\begin{gathered} f(x) = \sum_n C_n \mathrm{e}^{2\mathrm{i}nx} \\ C_n = \tfrac{1}{2}(a_n - \mathrm{i}b_n) = \frac{1}{\pi} \int_0^\pi \mathrm{d}x\, f(x) \mathrm{e}^{-2\mathrm{i}nx}. \end{gathered} \tag{M.5}$$

Periodic delta function. A periodic delta function can be *formally* defined from the relation of completeness.

Consider, first, the case of the interval $[-\pi, \pi]$, where

$$f(x) = \sum_n C_n f_n(x) \qquad f_n(x) \equiv \mathrm{e}^{\mathrm{i}nx}.$$

Now replacing the expression for C_n we obtain

$$f(x) = \sum_n \frac{1}{2\pi} \int \mathrm{d}x'\, f(x') f_n^*(x') f_n(x).$$

Then after formally exchanging the order of integration and summation, we obtain an expression for the periodic delta function:

$$\begin{aligned} \delta(x - x') &= \frac{1}{2\pi} \sum_n f_n^*(x') f_n(x) \\ &= \frac{1}{2\pi} \sum_n \mathrm{e}^{\mathrm{i}n(x-x')} = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n \geq 1} \cos n(x - x'). \end{aligned} \tag{M.6}$$

For symmetric functions we have $C_n = C_{-n}$, hence

$$\begin{aligned} \delta_{\mathrm{S}}(x, x') &= \tfrac{1}{2}[\delta(x - x') + \delta(x + x')] \\ &= \frac{1}{4\pi} \sum [\mathrm{e}^{\mathrm{i}n(x-x')} + \mathrm{e}^{\mathrm{i}n(x+x')}] = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n \geq 1} \cos nx \cos nx'. \end{aligned} \tag{M.7}$$

An analogous result for the interval $[0, \pi]$ has the form:

$$\delta(x - x') = \frac{1}{\pi} \sum_n \mathrm{e}^{2\mathrm{i}n(x-x')} = \frac{1}{\pi} + \frac{2}{\pi} \sum_{n \geq 1} \cos 2n(x - x'). \tag{M.8}$$

Bibliography

Besides the literature directly related to the topics covered in the present book, I also include here a selection of standard textbooks on gravitation, field theory and differential geometry which the reader may profitably use, whenever necessary, to broaden his or her knowledge on these subjects. This list of references is not complete but is representative as far as the content of the book and the level of exposition are concerned.

For each chapter (except the appendices), several general references, which I consider as the most suitable for further reading, are denoted by the symbol •, thus enabling the reader to use the literature efficiently.

Chapter I

Adler R, Bazin M and Schiffer M 1965 *Introduction to General Relativity* (New York: McGraw-Hill)

Braginsky V B and Panov V I 1971 Verification of the equivalence of inertial and gravitational mass *Sov. Phys.–JETP* **34** 464 (*Zh. Eksp. Teor. Fiz.* **61** 873)

Dicke R H 1964 Remarks on the observational basis of general relativity *Gravitation and Relativity*, ed H Y Chiu and W F Hoffman (New York: Benjamin) p 1

Edington A 1966 *Space, Time and Gravitation* (Cambridge: Cambridge University Press)

Einstein A 1965 *Sobranie Nauchnyh Trudov*, vols 1 and 2, ed I E Tamm, J A Smorodinski and B G Kuznecov (Moscow: Nauka)

Einstein A and Infeld L 1954 *The Evolution of Physics* (New York: Simon and Schuster)

Eötvös R v 1889 Über die anziehung der erde auf verschiedene substanzen *Math. Naturw. Ber. aus Ungarn* **8** 65

Feynman R P, Leighton R B and Sands M 1963 *The Feynman Lectures on Physics* vol 1 (Reading, MA: Addison-Wesley)

• Hoffmann B 1983 *Relativity and Its Roots* (New York: Freeman)

Kittel C, Knight W D and Ruderman M A 1965 *Mechanics—Berkeley Physics Course* vol 1 (New York: McGraw-Hill)

Landau L D and Lifshitz E M 1975 *The Classical Theory of Fields* 4th revised English edn, translated by Morton Hamermesh (Oxford: Pergamon)

Lucas J R 1973 *A Treatise on Space and Time* (London: Methuen)

Mach E (1912) 1960 *The Science of Mechanics* English translation, from German, of the book published in 1912, by T J McCorwick (La Salle, IL: Open Court)

Misner C H, Thorne K S and Wheeler J A 1970 *Gravitation* (San Francisco, CA: Freeman)

Møller C 1972 *The Theory of Relativity* (Oxford: Clarendon)
• Rindler W 1977 *Essential Relativity* 2nd edn (New York: Springer)
Rindler W 1982 *Introduction to Special Relativity* (Oxford: Clarendon)
Roll P G, Krotkov R and Dicke R H 1964 The equivalence of inertial and passive gravitational mass *Ann. Phys., USA* **26** 442
Schlick M 1969 *Space and Time in Contemporary Physics* (New York: Dover)
Schmutzer E 1979 *Relativitätstheorie—Aktuel* (Leipzig: Teubner Verlaggeselschaft)
• Sciama D W 1969 *The Physical Foundations of General Relativity* (New York: Doubleday)
Tolman R C 1969 *Relativity, Thermodynamics and Cosmology* (Oxford: Clarendon)
Wald R 1984 *General Relativity* (Chicago, IL: The University of Chicago Press)
• Weinberg S 1972 *Gravitation and Cosmology* (New York: Wiley)

Chapter II

Abers E S and Lee B W 1973 Gauge theories *Phys. Rep.* C **9** 1
Barut A O and Brittin W E 1971 De Sitter and conformal groups and their applications, Lectures at the University of Colorado Symposium (Boulder, 1970) Colorado University *Preprint*
Barut A O and Raczka R 1977 *Theory of Group Representations and Applications* (Warsaw: PWN—Polish Scientific)
Belinfante F J. 1940 On the current and the density of the electric charge, the energy, the linear momentum and the angular momentum of arbitrary fields *Physica* **7** 449
Bergshoeff E A 1983 Conformal invariance in supergravity *PhD Thesis* Leiden University, RX-1008 (LEIDEN)-mc (microfishe)
Boyanovsky D and Carlos M N 1987 Informal introduction to conformal invariance in statistical mechanics, Lectures presented at the Latin American School of Physics, La Plata *Preprint* PITT-87-12
Callan C, Coleman S and Jackiw R 1970 A new improved energy–momentum tensor *Ann. Phys., NY* **59** 42
Carruthers P 1971 Broken scale invariance in particle physics *Phys. Rep.* C **1** 3
Choquet-Bruhat Y, de Witt-Morette C and Dillard-Bleick M 1977 *Analysis, Manifolds and Physics* (Amsterdam: North-Holland)
• Coleman S 1973a Dilatations *Properties of the Fundamental Interactions, Int. School of Physics Ettore Majorana (Erice, 1971)* ed A Zichichi and G C Wick (Bologna: Editrice Compositori) p 357
• ——1973b An improved energy–momentum tensor *Properties of the Fundamental Interactions, Int. School of Physics Ettore Majorana (Erice, 1971)* ed A Zichichi and G C Wick (Bologna: Editrice Compositori) p 525
Coleman S 1975 Secret symmetry: An introduction to spontaneous symmetry breakdown and gauge fields *Laws of Hadronic Matter, Int. School of Physics Etorre Majorana (Erice, 1973)* ed A Zichichi (New York: Academic) p 138
Dubrovin B, Novikov S and Fomenko A 1979 *Sovremennaya Geometriya* (Moscow: Nauka)
Felsager B 1981 *Geometry, Particles and Fields* (Odense: Odense University Press)
Ferrara S, Gatto R and Grillo A 1973 Conformal Algebra in Spacetime and Operator Product Expansion *Springer Tracts in Modern Physics* **67** 1

• Fulton T, Rohrlich F and Witten L 1962 Conformal invariance in physics *Rev. Mod. Phys.* **34** 442

Ginsparg P 1990 Applied conformal field theory *Fields, Strings, Critical Phenomena, Proc. Les Houches Summer School in Theoretical Physics (Les Houches, France, 1988)* ed E Brezen and J Zin-Justin (Amsterdam: North-Holland) p 1

• Kibble T W B 1961 Lorentz invariance and the gravitational field *J. Math. Phys.* **2** 212

Novozhilov Yu V 1972 *Vvedenie v Teoriyu Elementarnyh Chastitz* (Moscow: Nauka)

Sohnius M 1985 Supersymmetry *Phys. Rep.* C **128** 39

• Treiman S, Jackiw R and Gross D 1972, *Lectures on Current Algebra and Its Applications* (Princeton, NJ: Princeton University Press)

Van Holten J W 1986 $D = 2$ conformal gauge theory *Acta Phys. Pol.* B **18** 163

Wess J 1971 Conformal invariance and the energy–momentum tensor *Springer Tracts in Modern Physics* **60** 1

Zumino B 1970 Effective Lagrangians and broken symmetries *Lectures on Elementary Particles and Quantum Field Theory, Brandeis University Summer Institute* vol 2, ed S Deser *et al* (Cambridge, MA: MIT Press) p 437

Chapter III

Baekler P, Hehl F W and Mielke E W 1986 Nonmetricity and torsion: Facts and fancies in gauge approaches to gravity *Proc. 4th Marcel Grossmann Meeting on General Relativity* ed R Rufini (Amsterdam: Elsevier) p 277

Bañados M, Tronsoco R and Zanelli J 1996 Higher dimensional Chern–Simons supergravity *Phys. Rev.* D **54** 2605

Battiti R and Toller M 1985 Zero-mass normal modes in linearized Poincaré gauge theories *Nuovo Cimento Lett.* **44** 35

Blagojević M 2001 Hamiltonian structure and gauge symmetries of Poincaré gauge theory, Invited talk presented at the Meeting of German Physical Society, Dresden, 20–24 March, 2000; *Ann. Phys., Lpz* **10** [5] (hep-th/0006031)

Blagojević M and Nikolić I 1983 Hamiltonian dynamics of Poincaré gauge theory: General structure in the time gauge *Phys. Rev.* D **28** 2455

——2000 Hamiltonian structure of the teleparallel formulation of general relativity *Phys. Rev.* D **62** 024021 (hep-th/0002022)

Blagojević M, Nikolić I, Popović D and Živanović Dj 1981 Poincaré gauge theory of gravitation and its Hamiltonian formulation *Nuovo Cimento* B **62** 257

Blagojević M, Popović D and Živanović Dj 1982 Gravitational singularity in Poincaré gauge theory *Phys. Lett.* B **109** 431

Blagojević M and Vasilić M 1987 Extra gauge symmetries in a weak-field approximation of an $R + T^2 + R^2$ theory of gravity *Phys. Rev.* D **35** 3748

——2000 Gauge symmetries of the teleparallel theory of gravity *Class. Quantum Grav.* **17** 3785 (hep-th/0006080)

Choquet-Bruhat Y, de Witt-Morette C and Dillard–Bleick M 1977 *Analysis, Manifolds and Physics* (Amsterdam: North-Holland)

de Andrade V C, Guillen L C and Pereira J G 2000a Gravitational energy–momentum density in teleparallel gravity *Phys. Rev. Lett.* **84** 4533

——2000b Teleparallel gravity—an overview *Talk Given at the IX Marcel Grossmann Meeting (Rome, July/2000)* (gr-qc/0011087)

Dimakis A 1989 The initial value problem of the Poincaré gauge theory in vacuum. I.

Second order formalism, II. First order formalism *Ann. Inst. H. Poincaré* A **51** 371, 389

Dubrovin B, Novikov S and Fomenko A 1979 *Sovremennaya Geometriya* (Moscow: Nauka)

Fairchild E E 1976 Gauge theory of gravitation *Phys. Rev.* D **14** 384

Frolov B 1963 Princip lokal'noi invariantnosti i teorema Neter *Vestnik Moskovskogo univerziteta* **6** 48

Grensing D and Grensing G 1983 General relativity as a gauge theory of the Poincaré group, the symmetric momentum tensor of both matter and gravity, and gauge fixing conditions *Phys. Rev.* D **28** 286

Grignani G and Nardelli G 1992 Gravity and the Poincaré group *Phys. Rev.* D **45** 2719

Gronwald F and Hehl F W 1996 On the gauge aspects of gravity *Proc. 14th Course of the School of Cosmology and Gravitation on Quantum Gravity (Erice, Italy, 1995)* ed P G Bergmann, V de Sabata and H J Treder (Singapore: World Scientific) p 148

Hamamoto S 1983 Manifestly covariant canonical formalism of Poincaré gauge theories *Z. Phys.* C **10** 353

Hayashi K 1968 Gauge theories of massive and massless tensor fields *Prog. Theor. Phys.* **39** 495

Hayashi K and Bregman A 1975 Poincaré gauge invariance and the dynamical role of spin in gravitational theory *Ann. Phys., NY* **75** 562

Hayashi K and Shirafuji T 1979 New general relativity *Phys. Rev.* D **19** 3524

• ——1980–81 Gravity from the Poincaré gauge theory of fundamental particles. I. General formulation, II. Equations of motion for test bodies and various limits, III. Weak field approximation, IV. Mass and energy of particle spectrum, V. The extended Bach–Lanczos identity, VI. Scattering amplitudes, VII. The axial-vector model *Prog. Theor. Phys.* **64** 866, 883, 1435, 2222; **65**, 525; **66**, 318, 2258

——1988, Gauge theory of gravitation—A unified formulation of Poincaré and (anti-) de Sitter gauge theories *Prog. Theor. Phys.* **80** 711

Hecht R D, Lemke J and Wallner R P 1991 Can Poincaré gauge theory be saved? *Phys. Rev.* D **44** 2442

Hecht R D, Nester J M and Zhytnikov V V 1996 Some Poincaré gauge theory Lagrangians with well-posed initial value problem *Phys. Lett.* A **222** 37

Hehl F W 1973 Spin and torsion in general relativity: I. Foundations *Gen. Rel. Grav.* **4** 333

——1974 Spin and torsion in general relativity: II. Geometry and field equations *Gen. Rel. Grav.* **5** 491

——1980 Four lectures in Poincaré gauge theory *Proc. 6th Course of the School of Cosmology and Gravitation on Spin, Torsion, and Supergravity (Erice, Italy, 1979)* ed P G Bergmann and V de Sabbata (New York: Plenum) p 5

Hehl F W, McCrea J D, Mielke E W and Ne'eman Y 1995 Metric-affine gauge theory of gravity: field equations, Noether identities, world spinors and breaking of dilation invariance *Phys. Rep.* C **258** 1

Hehl F W, Nietsch J and von der Heyde P 1980 Gravitation and the Poincaré gauge field theory with quadratic Lagrangian *General Relativity and Gravitation—One Hundred Years After the Birth of Albert Einstein* ed A Held (New York: Plenum) p 329

Hehl F W and Šijački Dj 1980 Towards a unified gauge theory of gravitational and strong interactions *Gen. Rel. Grav.* **12** 83

• Hehl F W, von der Heyde P, Kerlick D and Nester J 1976 General relativity with spin and torsion: Foundations and prospects *Rev. Mod. Phys.* **48** 393

Ivanenko D and Sardanashvily G 1983 The gauge treatment of gravity *Phys. Rep.* C **94** 1

——1985 *Gravitaciya* (Kiev: Naukova Dumka)

Katanaev M and Volovich I V 1990 Two-dimensional gravity with dynamical torsion and strings *Ann. Phys., NY* **197** 1

Kawai T 1986 A Poincaré gauge theory of gravity *Gen. Rel. Grav.* **18** 995

Kawai T and Toma N 1991 An extended new general relativity as a reduction of $\overline{\text{Poincaré}}$ gauge theory of gravity *Prog. Theor. Phys.* **85** 901

• Kibble T W B 1961 Lorentz invariance and the gravitational field *J. Math. Phys.* **2** 212

Kopczyński W 1982, Problems with metric-teleparallel theories of gravitation *J. Phys. A: Math. Gen.* **15** 493

Kuhfuss R and Nitsch J 1986 Propagating modes in gauge field theories of gravity *Gen. Rel. Grav.* **18** 1207

Mac Dowell S W and Mansouri F 1977 Unified geometric theory of gravity and supergravity *Phys. Rev. Lett.* **38** 739

Mansouri F and Chang L N 1976 Gravitation as a gauge theory *Phys. Rev.* D **13** 3192

Mielke E W 1987, *Geometrodynamics of Gauge Fields—On the Geometry of Yang–Mills Fields and Gravitational Gauge Theories* (Berlin: Akademie-Verlag)

Minkevich A 1980 Generalized cosmological Friedmann equations without gravitational singularity *Phys. Lett.* A **80** 232

Mishenko A S and Fomenko A T 1980 *Kurs Diferencial'noi Geometrii i Topologii* (Moscow: Moscow University)

Moffat J W 1978 A geometrical gauge theory of gravity and elementary particle forces *Phys. Rev.* D **17** 1965

Ne'eman Y 1978 Gravity is the gauge theory of the parallel transport modification of the Poincaré group *Differential Geometrical Methods in Mathematical Physics (Lecture Notes in Mathematics 676)* ed K Bleuler, H Petry and A Reetz (Berlin: Springer) p 189

Ne'eman Y and Šijački Dj 1979 Unified affine gauge theory of gravity and strong interactions with finite and infinite $GL(4, R)$ spinor fields *Ann. Phys., NY* **120** 292

Nester J 1988 Is there really a problem with the teleparallel theory? *Class. Quantum Grav.* **5** 1003

——1991 Special orthonormal frames and energy localization *Class. Quantum Grav.* **8** L19

Nester J and Yo H-J 1998 Symmetric teleparallel general relativity *Preprint* (gr-qc/9809049)

Nikolić I 1984 Dirac Hamiltonian structure of $R + R^2 + T^2$ Poincaré gauge theory of gravity without gauge fixing *Phys. Rev.* D **30** 2508

Nitsch J 1980 The macroscopic limit of the Poincaré gauge field theory of gravitation *Proc. 6th Course of the School of Cosmology and Gravitation on Spin, Torsion, and Supergravity (Erice, Italy, 1979)* ed P G Bergmann and V de Sabbata (New York: Plenum) p 63

Obukhov Yu N 1983 Spectral geometry of the Riemann–Cartan spacetime *Nucl. Phys.* B **212** 237–54

Obukhov Yu N and Korotky V A 1987 The Weyssenhoff fluid in Einstein–Cartan theory *Class. Quantum Grav.* **4** 1633–57

Okubo S 1991 A BRST-like operator for space with zero curvature but non-zero torsion tensor *Gen. Rel. Grav.* **23** 599

Ponomariev V N, Barvinsky A O and Obukhov Yu N 1985 *Geometro-dinamicheskie*

Metody i Kalibrovochnyi Podhod k Teorii Gravitacionyh Vzaimodeistvii (Moscow: Energoatomizdat)
Prugovečki E 1995 *Principles of Quantum General Relativity* (Singapore: World Scientific)
Puntigam R A 1996 Exact string-type solutions in the Einstein–Cartan theory of gravity *PhD Thesis* University of Cologne
Regge T 1986 On broken symmetries and gravity *Phys. Rep.* **137** 31
Schwinger J 1963 Quantized gravitational field *Phys. Rev.* **130** 1253
Sciama D W 1962 On the analogy between charge and spin in general relativity *Recent Developments in General Relativity* Festschrift for Infeld (Oxford: Pergamon) p 415
——1964 The physical structure of general relativity *Rev. Mod. Phys.* **36** 463 (E) 1103
Sezgin E and van Nieuwenhuizen P 1980 New ghost-free gravity Lagrangians with propagating torsion *Phys. Rev.* D **21** 3269
Shirafuji T and Suzuki M 1988 Gauge theory of gravitation—a unified formulation of Poincaré and (anti-) de Sitter gauge theories *Prog. Theor. Phys.* **80** 711
Šijački Dj 1982 Quark confinement and the short-range component of general affine gauge theory *Phys. Lett.* B **109** 435
Smalley L 1980 Post-Newtonian approximation of the Poincaré gauge theory of gravitation *Phys. Rev.* D **21** 328
Thirring W 1978 Gauge theories of gravitation *Facts and Prospects of Gauge Theories* ed P Urban (Vienna: Springer) p 439
Utiyama R 1956 Invariant theoretical interpretation of interactions *Phys. Rev.* **101** 1597
Von der Heyde P 1975 The equivalence principle in the U_4 theory of gravitation *Nuovo Cimento Lett.* **14** 250
——1976 The field equations of the Poincaré gauge theory of gravitation *Phys. Lett.* A **58** 141; Is gravitation mediated by the torsion of spacetime? *Z. Naturf.* **31a** 1725
Witten E 1988 $2 + 1$ gravity as an exactly soluable system *Nucl. Phys.* B **311** 46
Yang C N and Mills R 1954 Conservation of isotopic spin and isotopic gauge invariance *Phys. Rev.* **96** 191

Chapter IV

Adler S L 1982 Einstein gravity as a symmetry-breaking effect *Rev. Mod. Phys.* **54** 729
Adler R, Bazin M and Schiffer M 1965 *Introduction to General Relativity* (New York: McGraw-Hill)
Antoniadis I, Iliopoulos J and Tomaras T 1985 On the stability of background solutions in conformal gravity *Nucl. Phys.* B **261** 157
Antoniadis I and Tsamis N 1984 On the cosmological constant problem *Preprint* SLAC-PUB-3296; Weyl invariance and the cosmological constant *Preprint* SLAC-PUB-3297
Bergshoeff E A 1983 Conformal invariance in supergravity *PhD Thesis* Leiden University, RX-1008 (LEIDEN)-mc (microfishe)
Brans C and Dicke R 1961 Mach's principle and a relativistic theory of gravitation *Phys. Rev.* **124** 925
Bregman A 1973 Weyl transformations and Poincaré gauge invariance *Prog. Theor. Phys.* **49** 667
Callan C, Coleman S and Jackiw R 1970 A new improved energy–momentum tensor *Ann. Phys., NY* **59** 42

Canuto V, Adams P, Hsieh S H and Tsiang E 1976 Gauge covariant theory of gravitation *Preprint* Institute for Space Studies, New York

• Charap J and Tait W 1974 A gauge theory of the Weyl group *Proc. R. Soc.* A **340** 249

Choquet-Bruhat Y, de Witt-Morette C and Dillard-Bleick M 1977 *Analysis, Manifolds and Physics* (Amsterdam: North-Holland)

• Coleman S 1973a Dilatations *Properties of the Fundamental Interactions, Int. School of Physics Ettore Majorana (Erice, 1971)* ed A Zichichi and G C Wick (Bologna: Editrice Compositori) p 357

• ——1973b An improved energy–momentum tensor *Properties of the Fundamental Interactions, Int. School of Physics Ettore Majorana (Erice, 1971)* ed A Zichichi and G C Wick (Bologna: Editrice Compositori) p 525

——1975 Secret symmetry: An introduction to spontaneous symmetry breakdown and gauge fields *Laws of Hadronic Matter, Int. School of Physics Etorre Majorana (Erice, 1973)* ed A Zichichi (New York: Academic) p 138

de Alfaro V, Fubini S and Furlan G 1980 Small distance behavior in Einstein theory of gravitation *Phys. Lett.* B **97** 67

De Wit B 1981 Conformal invariance in gravity and supergravity *Preprint* NIKHEF-H/81-20, Lectures on the 18th Winter School of Theoretical Physics, Karpacz, Poland, 18 February–3 March, 1981

Dereli T and Tucker W 1982 A note on a generalization of Weyl's theory of gravitation *J. Phys. A: Math. Gen.* **15** L7

——1982 Weyl scalings and spinor matter interactions in scalar–tensor theories of gravitation *Phys. Lett.* B **110** 206

Dirac P A M 1973 Long range forces and broken symmetries *Proc. R. Soc.* A **333** 403

Domokos G 1976 Broken Weyl symmetry *Trieste Internal Report* IC/76/92

Fukui M, Hayashi K and Shirafuji T 1985 Conformal rescalings applied to Poincaré gauge theory *Prog. Theor. Phys.* **74** 852

• Fulton T, Rohrlich F and Witten L 1962 Conformal invariance in physics *Rev. Mod. Phys.* **34** 442

Gronwald F 1996 On the gauge structure and covariant gauge fixing of metric-affine gravity *PhD Thesis* University of Cologne

Hayashi K, Kasuya M and Shirafuji T 1977 Elementary particles and Weyl's gauge field *Prog. Theor. Phys.* **57** 431

Hayashi K and Kugo T 1979 Everything about Weyl's gauge field *Prog. Theor. Phys.* **61** 334

Hehl F W, Kerlick G P and von der Heyde P 1976 On a new metric affine theory of gravitation *Phys. Lett.* B **63** 446

• Hehl F W, McCrea and Mielke 1988 Weyl spacetimes, the dilation current and creation of gravitating mass by symmetry breaking *Proc. Int. H. Weyl Conf. (Kiel, Germany, 1985)* ed W Deppert *et al* (Frankfurt: P. Lang) p 241

Hehl F W, McCrea J D, Mielke E W and Ne'eman Y 1995 Metric-affine gauge theory of gravity: field equations, Noether identities, world spinors and breaking of dilation invariance *Phys. Rep.* C **258** 1

Hehl F W and Šijački Dj 1980 Towards a unified gauge theory of gravitational and strong interactions *Gen. Rel. Grav.* **12** 83

Hochberg D and Plunien G 1991 Theory of matter in Weyl spacetime *Phys. Rev.* D **43** 3358

Kaku M 1982 Quantization of conformal gravity *Nucl. Phys.* B **203** 285

Kasuya M 1975 On the gauge theory in the Einstein–Cartan–Weyl spacetime *Nuovo Cimento* B **28** 127

Kim Sung-Won 1986 Brans–Dicke theory in general spacetime with torsion *Phys. Rev.* D **34** 1011

Maluf J W 1987 Conformal invariance and torsion in general relativity *Gen. Rel. Grav.* **19** 57

Miletić S 1986 Weyl's theory of gravity *Bachelor's Thesis* University of Belgrade (in Serbian)

Minkowski P 1977 On the spontaneous origin of Newton's constant *Phys. Lett.* B **71** 419

Ne'eman Y and Šijački Dj 1988 Gravity from symmetry breakdown of a gauge affine theory *Phys. Lett.* B **200** 489

Nepomechie R 1984 Einstein gravity as the low energy effective theory of Weyl gravity *Phys. Lett.* B **136** 33

Nieh H T 1982 A spontaneously broken conformal gauge theory of gravitation *Phys. Lett.* A **88** 388

Obukhov Yu N 1982 Conformal invariance and spacetime torsion *Phys. Lett.* A **90** 13

Omote M 1971 Scale transformations of the second kind and the Weyl spacetime *Nuovo Cimento Lett.* **2** 58

Omote M and Kasuya M 1977 The Hamiltonian formalism of the local scale invariant gravitational theory *Prog. Theor. Phys.* **58** 1627

O'Raiffeartaigh L 1986 *Group Structure of Gauge Theories* (Cambridge: Cambridge University Press)

Pascual J F 1980 Conformal gauge theory and Weyl invariance *Proc. Int. Conf. GR9—Abstracts of Contributed Papers (Jena, Germany, 1980)* vol 3, ed E Schmutzer (Jena: Friedrich Schiller University) p 639

Ponomariev V N, Barvinsky A O and Obukhov Yu N 1985 *Geometro-dinamicheskie Metody i Kalibrovochnyi Podhod k Teorii Gravitacionyh Vzaimodeistvii* (Moscow: Energoatomizdat)

Ponomariev V N and Obukhov Yu N 1982 The generalized Einstein–Maxwell theory of gravitation *Gen. Rel. Grav.* **14** 309–30

Smalley L 1986 Brans–Dicke-type models with nonmetricity *Phys. Rev.* D **33** 3590

Šijački Dj 1982 Quark confinement and the short-range component of general affine gauge theory *Phys. Lett.* B **109** 435

Utiyama R 1973 On Weyl's gauge field *Prog. Theor. Phys.* **50** 2080

Weyl H 1918 Gravitation und elektrizität *Sitzungsber. Preuss. Akad. Wiss. (Berlin)* 465

——1931 Geometrie und physik *Naturwissenschaften* **19** 49

——1952 (1923) *Spacetime–Matter* (New York: Dover); English translation of the book *Raum, Zeit und Materie* from 1923

Wess J 1971 Conformal invariance and the energy–momentum tensor *Springer Tracts in Modern Physics* **60** 1

Zee A 1983 Einstein gravity emerging from quantum Weyl gravity *Ann. Phys., NY* **151** 431

Zumino B 1970 Effective Lagrangians and broken symmetries *Lectures on Elementary Particles and Quantum Field Theory (Brandeis University Summer Institute)* vol 2, ed S Deser *et al* (Cambridge, MA: MIT Press) p 437

Chapters V and VI

Abott L F and Deser S 1982 Stability of gravity with the cosmological constant *Nucl. Phys.* B **195** 76

——1982 Charge definition in non-Abelian gauge theories *Phys. Lett.* B **116** 259

Antonowicz M and Szcyrba W 1985 The dynamical of gravitational theories with $GL(4, R)$ connections *J. Math. Phys.* **26** 1711

Arnowitt R, Deser S and Misner C W 1962 The dynamics of general relativity *Gravitation—An Introduction to Current Research* ed L Witten (New York: Wiley) p 227

Baekler P, Hecht R, Hehl F W and Shirafuji T 1987 Mass and spin of exact solutions of the Poincaré gauge theory *Prog. Theor. Phys.* **78** 16

Baekler P and Mielke E W 1988 Hamiltonian structure of Poincaré gauge theory and separation of non-dynamical variables in exact torsion solutions *Fortschr. Phys.* **36** 549

• Bañados M 1994 Global charges in Chern–Simons theory and 2 + 1 black hole *Phys. Rev.* D **52** 5861 (hep-th/9405171)

——1999a Three-dimensional quantum geometry and black holes *Invited talk at the Second Meeting 'Trends in Theoretical Physics'* held in Buenos Aires, December, 1998 (hep-th/9901148)

——1999b Notes on black holes and three dimensional gravity *Proc. VIII Mexican School on Particles and Fields* to appear (hep-th/9903244)

Beig R and O'Murchadha N 1987, The Poincaré group as the symmetry group of canonical general relativity *Ann. Phys., NY* **174** 463

Blagojević M 1998 Conservation laws in Poincaré gauge theory *Acta Phys. Pol.* B **29** 881

——2001 Hamiltonian structure and gauge symmetries of Poincaré gauge theory *Invited Talk Presented at the Meeting of German Physical Society (Dresden, 20–24 March, 2000)*; *Ann. Phys., Lpz* **10** [5] (hep-th/0006031)

Blagojević M and Nikolić I 1983 Hamiltonian dynamics of Poincaré gauge theory: General structure in the time gauge *Phys. Rev.* D **28** 2455

——2000 Hamiltonian structure of the teleparallel formulation of general relativity *Phys. Rev.* D **62** 024021 (hep-th/0002022)

Blagojević M, Nikolić I, Popović D and Živanović Dj 1981 Poincaré gauge theory of gravitation and its Hamiltonian formulation *Nuovo Cimento* B **62** 257

Blagojević M, Nikolić I and Vasilić M 1988 Local Poincaré generators in the $R + T^2 + R^2$ theory of gravity *Nuovo Cimento* B **101** 439

Blagojević M, Popović D and Sazdović B 1989 Hamiltonian BRST quantization of antisymmetric tensor gauge theory *Nucl. Phys.* B **322** 587

Blagojević M, Popović D and Živanović Dj 1982 Gravitational singularity in Poincaré gauge theory *Phys. Lett.* B **109** 431

Blagojević M and Vasilić M 1987 Extra gauge symmetries in a weak-field approximation of an $R + T^2 + R^2$ theory of gravity *Phys. Rev.* D **35** 3748

——1987 Constraint algebra in Poincaré gauge theory *Phys. Rev.* D **36** 1679

• ——1988 Asymptotic symmetries and conserved quantities in the Poincaré gauge theory *Class. Quantum Grav.* **5** 1241

——2000a Gauge symmetries of the teleparallel theory of gravity *Class. Quantum Grav.* **17** 3785 (hep-th/0006080)

——2000b Conservation laws in the teleparallel theory of gravity (hep-th/0008159)

Born M and Infeld L 1934 Foundations of the new field theory *Proc. R. Soc.* A **144** 425
Boulware D G, Deser S and Stelle K S 1985 Properties of energy in higher derivative theories *Preprint* DOE-ER-40048-17-P5.
Boulware D G, Horowitz G and Strominger A 1983 Zero-energy for scale-invariant gravity *Phys. Rev. Lett.* **50** 1726
Brill D and Deser S 1968 Positive definiteness of gravitational field energy *Phys. Rev. Lett.* **20** 75
Brown J D and Henneaux M 1986a On the Poisson brackets of differentiable generators in classical field theory *J. Math. Phys.* **27** (2) 489
——1986b Central charges in the canonical realization of assymptrotic symmetries: an example from three dimensional gravity *Commun. Math. Phys.* **104** 207
• Castellani L 1982 Symmetries of constrained Hamiltonian systems *Ann. Phys., NY* **143** 357
Castellani L, van Nieuwenhuizen P and Pilati M 1982 First-order tetrad gravity in Dirac's Hamiltonian formalism *Phys. Rev.* D **26** 352
Cawley R 1979 Determination of the Hamiltonian in the presence of the constraints *Phys. Rev. Lett.* **42** 413
Chang C-C, Nester J and Chen C-M 1999 Pseudotensors and quasilocal energy–momentum *Phys. Rev. Lett.* **83** 1897 (gr-qc/9809040)
Charap J M and Nelson J E 1983 Canonical general relativity: the primary constraint algebra *Preprint* QMC/83/3
——1983 Surface integrals and the gravitational action *J. Phys. A: Math. Gen.* **16** 1661
Charap J M, Henneaux M and Nelson J E 1988 Explicit form of the constraint algebra in tetrad gravity *Class. Quantum Grav.* **5** 1405
Chen H, Nester J M and Yo H-J 1998 Acausal PGT modes and the nonlinear constraint effect *Acta Phys. Pol.* B **29** 961
Cheng W-H, Chern D-C and Nester J M 1988 Canonical analysis of the one-parameter teleparallel theory *Phys. Rev.* D **38** 2656
Coussaert O, Henneaux M and van Driel P 1995 The asymptotic dynamics of three-dimensional Einstein gravity with a negative cosmological constant *Class. Quantum Grav.* **12** 2916 (gr-qc/9506019-v2)
Deser S 1969 Hamiltonian dynamics and positive energy in General Relativity *Int. Symp. on Contemporary Physics (Trieste, 1968)* vol 1, ed A Salam (Vienna: IAEA) p 563
——1981 Stability properties of gravity theories *Proc. Unified Theories of Elementary Particles (München, Germany)* p 152
——1982 Energy in gravitational theories: definition, positivity theorems and stability, Invited talk at 11th Texas Symp. on Relativistic Astrophysics, Austin, TX *Texas Rel. Astro.* **45**
——1983 Positive classical gravitational energy from classical supergravity *Phys. Rev.* D **27** 2805
DeWitt B 1967 Quantum theory of gravity: I, *Phys. Rev.* **160** 1195; Quantum theory of gravity: II. The manifestly covariant theory *Phys. Rev.* **162** 1195; Quantum theory of gravity: III. Application of the covariant theory *Phys. Rev.* **162** 1239
Dirac P A M 1958 The theory of gravitation in Hamiltonian form *Proc. R. Soc.* A **246** 333
——1962 Interacting gravitational and spinor field *Recent Developments in General Relativity* (Festschrift für Infeld) (Oxford: Pergamon; Warsaw: PWN—Polish Scientific) p 191
• ——1964 *Lectures on Quantum Mechanics* (New York: Yeshiva University)

Faddeev L D 1968 Gamil'tonova formulirovka teorii tyagotenia *Proc. 5th Int. Conf. on Gravitation and Relativity (Tbilisi)* p 229

Gitman D M and Tyutin I V 1986 *Kanonicheskoe Kvantovanie Polei so Svyazami* (Moscow: Nauka)

• Hanson A, Regge T and Teitelboim C 1976 *Constrained Hamiltonian Systems* (Rome: Academia Nazionale dei Lincei)

Hartle J B and Kuchar K 1984 The role of time in path integral formulations of parametrized theories *Quantum Theory of Gravity* ed M Christensen (Bristol: Adam Hilger) p 315

Hayashi K and Shirafuji T 1985 Energy, momentum and angular momentum in Poincaré gauge theory *Prog. Theor. Phys.* **73** 54

Hecht R D 1995 Mass and spin of Poincaré gauge theory *Gen. Rel. Grav.* **27** 537

Hecht R D and Nester J M 1993 A new evaluation of PGT mass and spin *Phys. Lett.* A **180** 324

Henneaux M 1983 Poisson brackets of the constraints in the Hamiltonian formulation of the tetrad gravity *Phys. Rev.* D **27** 986

Henneaux M and Teitelboim C 1992 *Quantization of Gauge Systems* (Princeton, NJ: Princeton University Press)

Isenberg J and Nester J 1980 Canonical gravity *General Relativity and Gravitation: One Hundred Years After the Birth of Alberth Einstein* vol I, ed A Held (New York: Plenum) p 23

Isham C J 1992 Canonical quantum gravity and the problem of time, Lectures presented at the NATO Advanced Study Institute *Recent Problems in Mathematical Physics (Salamanca, 1992)* Imperial College preprint Imperial/TP/91-92/25

Kasuya M 1978 The Einstein–Cartan theory of gravitation in a Hamiltonian form *Prog. Theor. Phys.* **60** 167

Kawai T 1986 (3 + 1) structure of spacetime in Poincaré gauge theory of gravity *Prog. Theor. Phys.* **76** 1166

——1986 Energy–momentum and angular momentum in Poincaré gauge theory of gravity *Prog. Theor. Phys.* **79** 920

López-Pinto A, Tiemblo A and Tresguerres R 1997 Hamiltonian Poincaré gauge theory of gravitation *Class. Quantum Grav.* **14** 549

Maluf J W 1994 Hamiltonian formulation of the telepaerallel description of general relativity *J. Math. Phys.* **35** 335

——1995 Localization of energy in general relativity *J. Math. Phys.* **36** 4242

Mielke E W 1992 Ashtekar's complex variables in general relativity and its teleparallel equivalent *Ann. Phys., NY* **219** 78

Mielke E W and Wallner R P 1988 Mass and spin of double dual solutions in Poincaré gauge theory *Nuovo Cimento* B **101** 607

Misner C H, Thorne K S and Wheeler J A 1970 *Gravitation* (San Francisco, CA: Freeman)

Nelson J E and Teitelboim C 1978 Hamiltonian formulation of the theory of interacting gravitational and electron fields *Ann. Phys., NY* **116** 86

Nester J M 1977 Canonical formalism and the ECSK theory *PhD Thesis* University of Maryland

——1989a Positive energy via the teleparallel Hamiltonian *Int. J. Mod. Phys.* A **4** 1755

——1989b A positive gravitational energy proof *Phys. Lett.* A **139** 112

——1993 Some progress in classical canonical gravity *Directions in General Relativity, Proc. 1993 Int. Symposium, Maryland* vol I, papers in honour of Charles Misner, ed B L Hu, M P Ryan and C V Vishveshwara (Cambridge: Cambridge University Press) p 245. This paper gives a review of the covariant Hamiltonian approach based on differential form methods

Nikolić I 1981 Hamiltonian formulation of Einstein–Cartan theory of gravity *MSc Thesis* University of Belgrade (in Serbian)

——1983 Canonical structure of Poincaré gauge theory *PhD Thesis* University of Belgrade (in Serbian)

• ——1984 Dirac Hamiltonian structure of $R + R^2 + T^2$ Poincaré gauge theory of gravity without gauge fixing *Phys. Rev.* D **30** 2508

——1986 Canonical structure of Poincaré gauge-invariant theory of gravity *Fiz. Suppl.* **18** 135

——1990 Schwinger's energy–energy brackets in the U_4 theory of gravity *Class. Quantum Grav.* **7** 1895

——1992 Constraint algebra from local Poincaré symmetry *Gen. Rel. Grav.* **24** 159

——1995 Dirac Hamiltonian formulation and algebra of the constraints in the Einstein–Cartan theory *Class. Quantum Grav.* **12** 3103

Peldan P 1994 Actions for gravity, with generalizations: a review *Class. Quantum Grav.* **11** 1087

Regge T and Teitelboim C 1974 Role of surface integrals in the Hamiltonian formulation of general relativity *Ann. Phys., NY* **88** 286

Rovelli C 1986 Constraint algebra in general relativity *Nuovo Cimento* B **92** 49

Sudarshan E C G and Mukunda N 1974 *Classical Dynamics—A Modern Perspective* (New York: Wiley)

Sundermeyer K 1982 *Constrained Dynamics* (Berlin: Springer)

Szczyrba W 1982 Hamiltonian dynamics of gauge theories of gravity *Phys. Rev.* D **25** 2548

Teitelboim C 1973 How commutators of constraints reflect the spacetime structure *Ann. Phys., NY* **79** 542

——1980 The Hamiltonian structure of spacetime *General Relativity and Gravitation—One Hundred Years After the Birth of Albert Einstein* vol I, ed A Held (New York: Plenum) p 195

——1994 Abandoning prejudices about time reparametrization invariance *Proc. Workshop on Physical Origins of Time Asymmetry (Mazagon, Spain, 1991)* ed J J Halliwell *et al* (Cambridge: Cambridge University Press) p 442

ter Haar D 1971 *Elements of Hamiltonian Mechanics* (Oxford: Pergamon)

Trautman A 1980 Fiber bundles, gauge fields and gravitation *General Relativity and Gravitation—One Hundred Years After the Birth of Albert Einstein* vol I, ed A Held (New York: Plenum) p 287

Tseytlin A A 1982 Poincaré and de Sitter gauge theories of gravity with propagating torsion *Phys. Rev.* D **26** 3327

Vasilić M 1984 Extra gauge symmetries in an $R + T^2$ theory of gravity *MSc Thesis* University of Belgrade (in Serbian)

——1984 Local symmetries in Poincaré gauge theory *PhD Thesis* University of Belgrade (in Serbian)

Witten E 1988 $2 + 1$ gravity as an exactly soluable system *Nucl. Phys.* B **311** 46

York J W Jr 1980 Energy and momentum of the gravitational field *Essays in General Relativity: a festschrift for Abraham Taub* ed F J Tipler (New York: Academic) p 39

Zanelli J 2000 Chern–Simons Gravity: From $2+1$ to $2n+1$ Dimensions, Lecture presented at the XX Encontro de Fisica de Particulas e Campos, Brazil, October 1988, and at the Fifth La Hechicera School, Venezuela, November 1999 *Braz. J. Phys.* **30** 251 (hep-th/0010049)

Chapters VII and VIII

Berestetskii V B, Lifshitz E M and Pitaevskii L P 1982 *Quantum Electrodynamics* 2nd edition, translated by J B Sykes and J S Bell (Oxford: Pergamon)

Bjorken J D and Drell S 1964 *Relativistic Quantum Mechanics* (New York: McGraw-Hill)

——1965 *Relativistic Quantum Fields* (New York: McGraw-Hill)

Boulware D and Deser S 1972 Can gravitation have a finite range? *Phys. Rev.* D **6** 3368

Deser S 1970 Self–interaction and gauge invariance *Gen. Rel. Grav.* **1** 9

• Duff M J 1973 A particle physicist's approach to the theory of gravity *Trieste Preprint* IC/73/70

• Feynman R P, Morinigo F B and Wagner W G 1995 *Feynman Lectures on Gravitation* (Reading, MA: Addison-Wesley)

Fierz M and Pauli W 1939 Relativistic wave equation for particles of arbitrary spin in an electromagnetic field *Proc. R. Soc.* A **173** 211

• Kibble T W B 1965 The quantum theory of gravitation *High Energy Physics and Elementary Particles, Proc. Trieste Symposium (Trieste, 1965)* ed G Fronsdal and A Salam (Vienna: IAEA) p 885

Landau L D and Lifshitz E M 1975 *The Classical Theory of Fields* 4th revised English edn, translated by Morton Hamermesh (Oxford: Pergamon)

Logunov A A and Mestvirishvili M A 1985 Relativistic theory of gravity Institute for High Energy Physics *Preprint* IFVE 85-162

Misner C H, Thorne K S and Wheeler J A 1970 *Gravitation* (San Francisco, CA: Freeman)

Nordström G 1913 Zur Theorie der Gravitation vom Standpunkt des Relativitätsprinzips *Ann. Phys., Germany* **42** 533

• Okubo S 1978 Introduction to general relativity, lecture notes *Preprint* UR-687, the University of Rochester

Peskin M and Schroeder P 1995 *An Introduction to Quantum Field Theory* (Reading, MA: Addison-Wesley)

Roman P 1969 *Introduction to Quantum Field Theory* (New York: Wiley)

Scherk J 1979 A short review of supergravity *Preprint* LPTENS 79/23

Schwinger J 1970 *Particles, Sources and Fields* vol 1 (Reading, MA: Addison-Wesley)

Takahashi Y 1969 *An Introduction to Field Quantization* (Oxford: Pergamon)

• Van Dam H 1974 *Theory of Gravity* (Nijmegen: ITP—University of Nijmegen)

Van Dam H and Veltman M 1970 Massive and massless Yang–Mills and gravitational fields *Nucl. Phys.* B **22** 397

Van Nieuwenhuizen P 1973 Radiation of massive gravitation *Phys. Rev.* D **7** 2300

Veltman M 1976 Quantum theory of gravitation *Methods in Field Theory* ed R Balian and J Zinn-Justin (Amsterdam: North-Holland) p 265

Weinberg S 1964 Photons and gravitons in S-matrix theory: Derivation of charge conservation and equality of gravitational and inertial mass *Phys. Rev.* B **135** 1049

——1972 *Gravitation and Cosmology* (New York: Wiley)

——1995 *The Quantum Theory of Fields* (Cambridge: Cambridge University Press)

Chapter IX

• Bailin D and Love A 1994 *Supersymmetric Gauge Field Theory and String Theory* (Bristol: IOP Publishing)

Barut A O and Raczka R 1977 *Theory of Group Representations and Applications* (Warsaw: PWN—Polish Scientific)

Berestetskii V B, Lifshitz E M and Pitaevskii L P 1982 *Quantum Electrodynamics* 2nd edn, translated by J B Sykes and J S Bell (Oxford: Pergamon)

Carruthers A 1971 *Spin and Isospin in Particle Physics* (New York: Gordon and Breach)

Coleman S and Mandula J 1967 All possible symmetries of the S matrix *Phys. Rev.* **159** 1251

Deser S 1979 Supergravity: a post Einstein unification *On the Path of Albert Einstein, Proc. Coral Gables Conf. (Miami, 1979)* ed A Perlmutter *et al* (New York: Plenum) p 39

——1980 Supergravities: Successes and problems, Invited lecture given at the Europhysics Conference on Unification of Fundamental Interactions, Erice, 1980 *Preprint* Print-80-0388

Deser S and Zumino B 1976 Consistent supergravity *Phys. Lett.* B **62** 335

Fayet P and Ferrara S 1977 Supersymmetry *Phys. Rep.* C **32** 250

Ferrara S and Taylor J G (ed) 1982 *Supergravity '81, School on Supergravity (Trieste, 1981)* (Cambridge: Cambridge University Press)

Ferrara S, Taylor J G and van Nieuwenhuizen P (ed) 1983 *Supersymmetry and Supergravity '82, School on Supersymmetry and Supergravity (Trieste, 1982)* (Singapore: World Scientific)

Ferrara S and van Nieuwenhuizen P 1978 The auxiliary fields of supergravity *Phys. Lett.* B **74** 333

Freedman D Z and van Nieuwenhuizen P (ed) 1979 *Supergravity, Stony Brook Workshop, 1979* (Amsterdam: North-Holland)

Freedman D Z, van Nieuwenhuizen P and Ferrara S 1976 Progress towards a theory of supergravity *Phys. Rev.* D **13** 3214

Freund P G 1986 *Introduction to Supersymmetry* (Cambridge: Cambridge University Press)

Gates S J Jr, Grisaru M T, Roček M and Siegel W 1983 *Superspace—or One Thousand and One Lessons in Supersymmetry* (London: Benjamin-Cummings)

Goroff M and Sagnotti A 1985 Quantum gravity at two loops *Phys. Lett.* B **160** 81

——1986 The ultraviolet behavior of Einstein gravity *Nucl. Phys.* B **266** 709

Haag R J, Lopuszanski J and Sohnius M 1975 All possible generators of supersymmetries of the S-matrix *Nucl. Phys.* B **88** 257

Hawking S W and Roček M (ed) 1981 *Superspace and Supergravity, Cambridge Conference, 1980* (Cambridge: Cambridge University Press)

Levy M and Deser S (ed) 1979 *Recent Developments in Gravitation, Cargese Conference, 1978* (New York: Gordon and Breach)

Müller-Kirsten H J W and Wiedermann A 1987 *Supersymmetry—an Introduction with Conceptual and Calculational Details* (Singapore: World Scientific)

Novozhilov Yu V 1972 *Vvedenie v Teoriyu Elementarnyh Chastitz* (Moscow: Nauka)

Ogievetski V and Mezinchesku L 1975 Simetrii meždu bozonami i fermionami i superpolja *Usp. Fiz. Nauk* **117** 637 (Engl. transl. 1976 Boson fermion symmetries and superfields *Sov. Phys.–Usp.* **18** 960)

Rarita W and Schwinger J 1941 On a theory of particles with half-integer spin *Phys. Rev.* **60** 61

Salam A and Strathdee J 1978 Supersymmetry and superfields *Fortschritte der Physik* **26** 57

Sazdović B 1989 Supergravitacija *Sveske fizičkih nauka* **2** 1 (in Serbian)

Scherk J 1979 A short review of supergravity l'Ecole Normale Superieure *Preprint* LPTENS 79/23

Schwinger J 1970 *Particles, Sources and Fields* vol 1 (Reading, MA: Addison-Wesley)

• Sohnius M F 1985 Introducing supersymmetry *Phys. Rep.* C **128** 39

Srivastava Prem P 1973 Conformal symmetry in Lagrangian field theory *Nucl. Phys.* B **64** 499

• ——1986 *Supersymmetry, Superfields and Supergravity: An Introduction* (Bristol: IOP Publishing)

Stelle K S 1983a Introduction to supersymmetry *Trieste Preprint* ICTP/ 82-83/8

——1983b Supersymmetry, finite theories and quantum gravity *Trieste Preprint* ICTP/82-83/9

Stelle K S and West P C 1978 Minimal auxiliary fields for supergravity *Phys. Lett.* B **74** 330

——1980 Realizing the supersymmetry algebra *Trieste Preprint* ICTP/79-80/37

Takahashi Y 1969 *An Introduction to Field Quantization* (Oxford: Pergamon Press)

Van Nieuwenhuizen P 1980 Four lectures at 1979 Erice school on spin, torsion, rotation and supergravity *Proc. 6th Course of the School of Cosmology and Gravitation on Spin, Torsion, and Supergravity (Erice, Italy, 1979)* ed P G Bergmann and V de Sabbata (New York: Plenum) p 227

• ——1981a Supergravity *Phys. Rep.* C **68** 192

——1981b Six lectures at the Cambridge workshop on supergravity *Proc. Nufield Workshop on Supersymmetry and Supergravity (Cambridge, 1980)* ed S W Hawking and M Roček (Cambridge: Cambridge University Press) p 9

——1981c Six lectures at the Trieste 1981 school on supergravity *Stony Brook Preprint* ITP-SB-81-67

Van Nieuwenhuizen P and Freedman D Z 1978 Supergravity *Sci. Am.* February

Wess J and Bagger J 1982 *Supersymmetry and Supergravity* (Princeton, NJ: Princeton University Press)

Wess J and Zumino B 1974 Supergauge transformations in four dimensions *Nucl. Phys.* B **70** 39

• West P 1986 *Introduction to Supersymmetry and Supergravity* (Singapore: World Scientific)

Wyborn G B 1973 *Classical Groups for Physicists* (New York: Wiley)

Zumino B 1975 Supersymmetry and the vacuum *Nucl. Phys.* B **89** 535

Chapter X

Abott L F and Deser S 1982 Stability of gravity with a cosmological constant *Nucl. Phys.* B **195** 76

• Bailin D and Love A 1987 Kaluza–Klein theories *Rep. Prog. Phys.* **50** 1087

Bergmann P 1942 *Introduction to the Theory of Relativity* (Englewood Cliffs, NJ: Prentice-Hall)

Blagojević M and Vasilić M 1988 Asymptotic symmetries and conserved quantities in the Poincaré gauge theory *Class. Quantum Grav.* **5** 1241

Candelas P and Weinberg S 1984 Calculations of gauge couplings and compact circumferences from self-consistent dimensional reduction *Nucl. Phys.* B **237** 397

Cho Y M 1975 Higher dimensional unifications of gravitation and gauge theories *J. Math. Phys.* **16** 2029

Cho Y M and Freund P G O 1975 Non-Abelian gauge fields as Nambu–Goldstone fields *Phys. Rev.* D **12** 1711

Cho Y M and Yang P S 1975 Unified geometry of internal space with spacetime *Phys. Rev.* D **12** 3789

Chodos A 1984 Kaluza–Klein theories: an overview *Comment. Nucl. Part. Phys.* **13** 171

Coleman S 1985 The uses of instantons *Aspects of Symmetry* (Cambridge: Cambridge University Press) ch 7, p 265

Cremmer E and Scherk J 1977 Spontaneous compactification of extra space dimensions *Nucl. Phys.* B **110** 61

Cremmer E, Julia B and Scherk J 1978 Supergravity theory in 11 dimensions *Phys. Lett.* B **76** 409

de Andrade V C, Guillen L C and Pereira J G 2000 Teleparallel equivalent of the Kaluza–Klein theory *Phys. Rev.* D **61** 084031 (gr-qc/9909004)

Deser S 1983 Positive classical gravitational energy from classical supergravity *Phys. Rev.* D **27** 2805

Deser S and Soldate M 1988/89 Gravitational energy in spaces with compactified dimension *Nucl. Phys.* B **311** 739

De Witt B 1964 Dynamical theory of groups and fields *Relativity, Groups and Topology* ed C de Witt ad B S de Witt (New York: Gordon and Breach)

Dolan L and Duff M J 1984 Kac–Moody symmetries of Kaluza–Klein theories *Phys. Rev. Lett.* **52** 14

• Duff M, Nilsson B and Pope C 1986 Kaluza–Klein supergravity *Phys. Rep.* C **130** 1

Duff M, Nilsson B, Pope C and Warner N 1984 On the consistency of the Kaluza–Klein ansatz *Phys. Lett.* B **149** 60

Eguchi T, Gilkey P and Hanson A 1980 Gravitation, gauge theories and differential geometry *Phys. Rep.* **66** 213

Einstein A and Bergmann P 1938 On a generalization of Kaluza's theory of electricity *Ann. Math.* **39** 683

Englert F and Nikolai H 1983 Supergravity in eleven-dimensional spacetime *CERN Preprint* TH-3711/83

Felsager B 1981 *Geometry, Particles and Fields* (Odense: Odense University Press)

Freund P G O and Rubin M A 1980 Dynamics of dimensional reduction *Phys. Lett.* B **97** 233

Freedmann D and van Nieuwenhuizen P 1985 The hidden dimension of spacetime *Sci. Am.* **252** 62

Freedmann D, van Nieuwenhuizen P and Ferrara S 1976 Progress towards a theory of supergravity *Phys. Rev.* D **13** 3214

Gates J, Grisaru M, Roček M and Siegel W 1983 *Superspace—or One Thousand and One Lessons in Supersymmetry* (London: Benjamin-Cummings)

Gross D and Perry M 1983 Magnetic monopoles in Kaluza–Klein theories *Nucl. Phys.* B **226** 29

Jordan P 1947 Erweiterung der projektiven Relativitätstheorie *Ann. Phys., Lpz* **1** 219

Kaluza T 1921 Zum unitätsproblem in der physik *Sitzungsber. Preuss. Akad. Wiss. Berlin, Math. Phys.* K **1** 966

Karlhede A and Tomaras T 1983 Spontaneous symmetry breaking in Kaluza–Klein theories *Phys. Lett.* **125** 49

Kerner R 1968 Generalization of the Kaluza–Klein theory for an arbitrary non-Abelian gauge group *Ann. Inst. H. Poincaré* A **9** 143

Klein O 1926 Quantentheorie und fünfdimensionale Relativitätstheorie *Z. Phys.* **37** 895

Lee H C (ed) 1984 An introduction to Kaluza–Klein theories *Proc. Workshop on Kaluza–Klein Theories (Chalk River, Canada, 1983)* (Singapore: World Scientific). One can find here an English translation of the original papers of Kaluza and Klein

Luciani J F 1978 Spacetime geometry and symmetry breaking *Nucl. Phys.* B **135** 111

Mecklenburg W 1981 Comment on stability properties of degenerate systems *Phys. Rev.* D **24** 2776

• ——1983 Ten lectures on Kaluza–Klein theory *Lecture Notes* Departmento de Mathematica, Universidade de Brasilia, Brazil

——1984 The Kaluza–Klein idea. Status and prospects *Fortschr. Phys.* **32** 207

Nahm W 1978 Supersymmetries and their representations *Nucl. Phys.* B **135** 149

Orzalesi C 1981 Riemannian geometry of classical gauge field theories *Particle Physics 1980, Proc. 3rd Adriatic Meeting (Dubrovnik, 1980)* ed I Andrić, I Dadić and N Zovko (Amsterdam: North-Holland) p 313

Rajaraman R 1982 *Solitons and Instantons* (Amsterdam: North-Holland)

Randjbar-Daemi S, Salam A and Strathdee J 1983 Spontaneous compactification in six-dimensional Einstein–Maxwell theory *Nucl. Phys.* B **214** 491

Rayski J 1965 A unified description of spacetime and isosphere, I and II *Acta Phys. Pol.* **27** 947; *Acta Phys. Pol.* **28**, 87

• Salam A and Strathdee J 1982 On Kaluza–Klein theory *Ann. Phys., NY* **141** 316

Schwarz A S and Tyupkin Yu S 1981 Dimensional reduction of the gauge field theory *Nucl. Phys.* B **187** 321

Sherk J and Schwarz J 1975 Dual field theory of quarks and gluons *Phys. Lett.* B **57** 463

Toms D 1984 Kaluza–Klein theories *Proc. Workshop on Kaluza–Klein Theories* ed H C Lee (Singapore: World Scientific) p 185

Trautman A 1970 Fiber bundles associated with spacetime *Rep. Math. Phys.* **1** 29

Van Nieuwenhuizen P 1985 An introduction to simple supergravity and the Kaluza–Klein program *Relativity, Groups and Topology, Proc. Les Houches Lectures, Session XI (Les Houches, 1983)* ed B S De Witt and R Stora (Amsterdam: Elsevier) p 823

Vasilić M 1989 Kaluza–Klein theory and unification of interactions *Sveske fizičkih nauka* **II-2** 97 (in Serbian)

Viswanathan K 1984 Dimensional reduction and harmonic expansions on coset spaces *An Introduction to Kaluza–Klein Theories* ed H C Lee (Singapore: World Scientific) p 159

Weinberg S 1972 *Gravitation and Cosmology* (New York: Wiley)

——1983 Charges from extra dimensions *Phys. Lett.* B **125** 265

Wess J and Bagger J 1983 *Supersymmetry and Supergravity* (Princeton, NJ: Princeton University Press)

Wetterich C 1982 Spontaneous compactification in higher dimensional gravity *Phys. Lett.* B **113** 377

Witten E 1981a Search for a realistic Kaluza–Klein theory *Nucl. Phys.* B **186** 412

——1981b Kaluza–Klein theory and the positive energy theorem *Particles and Fields—*

2, Proc. Banff Summer Institute on Particles and Fields (Banff, Canada, 1981) ed A Capri and A N Kamal (New York: Plenum) p 243

——1981c A new proof of the positive energy theorem *Commun. Math. Phys.* **80** 381

——1982 Instability of the Kaluza–Klein vacuum *Nucl. Phys.* B **195** 481

Wu Y S and Zee A 1984 Massless fermions and Kaluza–Klein theories *J. Math. Phys.* **25** 2696

Yano T 1970 Integral formulae in Riemannian geometry (New York: Dekker)

• Zee A 1981 Grand unification and gravity—selected topics *Proc. Fourth Kyoto Summer Institute on Grand Unified Theories and Related Topics (Kyoto, Japan, 1981)* ed M Konuma and T Maskawa (Singapore: World Scientific) p 143

Chapter XI

Bailin D and Love A 1994 *Supersymmetric Gauge Field Theory and String Theory* (Bristol: IOP Publishing)

Banks T and Peskin M 1986 Gauge invariance of string fields *Nucl. Phys.* B **264** 513

Barbashov B M and Nesterenko V V 1987 *Model' Relyativistkoi Struny v Fizike Adronov* (Moscow: Energoatomizdat)

Bardeen W A and White A R (ed) 1985 *Symp. on Anomalies, Geometry, Topology, Proc. Int. Conf. (Chicago, 1985)* (Singapore: World Scientific)

Brink L 1984 Superstrings, Lectures delivered at the 1985 Les Houches summer school; *Goeteborg Preprint* 85-68

——1985 Free strings *Superstrings, Supergravity and Unified Theories* ed G Furlan *et al* (Singapore: World Scientific) p 57

• Brink L and Henneaux M 1988 *Principles of String Theory* (New York: Plenum)

Brink L, di Vecchia P and Howe P S 1976 A locally supersymmetric and reparametrization invariant action for the spinning string *Phys. Lett.* B **65** 471

Callan C, Friedan D, Martinec E and Perry M 1985 Strings in background fields *Nucl. Phys.* B **262** 593

Collins P and Tucker R 1976 Classical and quantum mechanics of free relativistic membranes *Nucl. Phys.* B **112** 150

Cremmer E and Gervais J L 1974 Combining and splitting relativistic strings *Nucl. Phys.* B **76** 209

——1975 Infinite component field theory of interacting relativistic strings and dual theory *Nucl. Phys.* B **90** 410

Deser S and Zumino B 1976, A complete action for the spinning string *Phys. Lett.* B **65** 369

D'Hoker E and Phong D H 1988 The geometry of string perturbation theory *Rev. Mod. Phys.* **60** 917

Dirac P A M 1964 *Lectures on Quantum Mechanics* (New York: Yeshiva University)

Feigin B and Fuchs D 1982 Invariant skew-symmetric differential operators on the line and Verma modules over the Virasoro algebra *Funct. Anal. Appl.* **16** 114

Fradkin E and Tseytlin A 1985 Quantum string theory effective action *Nucl. Phys.* B **261** 1; Effective action approach to superstring theory *Phys. Lett.* B **160** 69

Furlan G, Jengo R, Pati J C, Sciama D W and Shafi Q (ed) 1985 *Superstrings, Supergravity and Unified Theories, Proc. Summer Workshop in High Energy Physics and Cosmology (Trieste, 1985)* (Singapore: World Scientific)

Gervais J L 1985 Basic features of string theories *Superstrings, Supergravity and Unified Theories* ed G Furlan *et al* (Singapore: World Scientific) p 22

Gliozzi F, Scherk J and Olive D 1977 Supersymmetry, supergravity theories and the dual spinor model *Nucl. Phys.* B **122** 253

Goddard P, Goldstone J, Rebbi C and Thorn C 1973 Quantum dynamics of a massless relativistic string *Nucl. Phys.* B **56** 109

Goddard P and Olive D 1986 An introduction to Kac–Moody algebras and their physical interpretation *Unified String Theories* ed M Green and D Gross (Singapore: World Scientific) p 214

——(ed) 1988 *Kac–Moody and Virasoro Algebras* (Singapore: World Scientific)

Gotto T 1971 Relativistic quantum mechanics of one-dimensional mechanical continuum and subsidiary condition of dual resonance model *Prog. Theor. Phys.* **46** 1560

Green M 1983 Supersymmetric dual string theories and their field theory limits—a review *Surv. High Energy Phys.* **3** 127

Green M and Gross D (ed) 1985 *Unified String Theories, Proc. Workshop on Unified String Theories, ITP (University of California, Santa Barbara, 1985)* (Singapore: World Scientific)

Green M and Schwarz J H 1982 Supersymmetric string theories *Phys. Lett.* B **109** 444

——1984 Covariant description of superstrings *Phys. Lett.* B **136** 367; Superstring field theory *Nucl. Phys.* B **243** 475

——1985 Infinity cancellations in $SO(32)$ superstring theory *Phys. Lett.* B **151** 21; The hexagon gauge anomaly in type I superstring theory *Nucl. Phys.* B **255** 93

Green M, Schwarz J and Witten E 1987 *Superstring Theory* (Cambridge: Cambridge University Press)

Gross D, Harvey J, Martinec E and Rohm R 1986 Heterotic string theory. 1. The free heterotic string, 2. The interacting heterotic string *Nucl. Phys.* B **256** 253; *Nucl. Phys.* B **267** 75

Hara O 1971 On origin and physical meaning of Ward-like identity in dual-resonance model *Prog. Theor. Phys.* **46** 1549

Hata H, Itoh K, Kugo T, Kunimoto H and Ogawa H 1986 Manifestly covariant field theory of interacting string, I and II *Phys. Lett.* B **172** 186, 195

Hatfield B 1992 *Quantum Field Theory of Point Particles and Strings* (Redwood City, CA: Addison-Wesley)

Howe P and Tucker R 1978 Local supersymmetry in $(2+1)$ dimensions. II. An action for a spinning membrane *J. Math. Phys.* **19** 981

Hwang S 1983 Covariant quantization of the string in dimensions $D \leq 26$ using a Becchi–Rouet–Stora formulation *Phys. Rev.* D **28** 2614

Itoh K, Kugo T, Kunimoto H and Ouguri H 1986 Gauge invariant local action of string field from BRST *Prog. Theor. Phys.* **75** 162

Kaku M 1985 Introduction to the field theory of strings *Preprint* CCNY-HEP-11

——1986 Geometric derivation of string field theory from first principles. 1. Curvature tensor and the tensor calculus *Nucl. Phys.* B **267** 125

——1988 *Introduction to Superstrings* (New York: Springer)

——1991 *Strings, Conformal Fields, and Topology: An Introduction* (New York: Springer)

Kaku M and Kikkawa K 1974 Field theory of relativistic strings. I. Trees, II. Loops and Pomerons *Phys. Rev.* D **10** 1110, 1823

Kato M and Ogawa K 1983 Covariant quantization of string based on BRS invariance *Nucl. Phys.* B **212** 443

Kazama Y 1986 An introduction to covariant superstring field theory *Preprint* CERN-TH 4518/86

Kikkawa K and Yamasaki M 1986 Can the membrane be a unification model? *Preprint* OU-HET 94, Osaka University

Kugo T and Uehara S 1982 General procedure of gauge fixing based on BRS invariance principle *Nucl. Phys.* B **197** 378

Mandelstam S 1974 Dual-resonance model *Phys. Rep.* C **13** 259

Nambu Y 1970 Lectures at the Copenhagen symposium (unpublished)

Neveu A and Scherk J 1972 Connection between Yang–Mills fields and dual models *Nucl. Phys.* B **36** 155

Neveu A and Schwarz J H 1971 Factorizable dual model of pions *Nucl. Phys.* B **31** 86

Neveu A, Schwarz J H and West P 1985 Gauge symmetries of the free bosonic string field theory *Phys. Lett.* B **164** 51

Neveu A and West P 1985 Gauge symmetries of the free supersymmetric string field theories *Phys. Lett.* B **165** 63

——1986 The interacting gauge covariant bosonic string *Phys. Lett.* B **168** 192

• Peskin M 1985 An introduction to the theory of strings *Preprint* SLAC-PUB-3821

Polyakov A M 1981 Quantum geometry of bosonic string and Quantum geometry of fermionic string *Phys. Lett.* B **103** 207, 211

——1987a Quantum gravity in two dimensions *Mod. Phys. Lett.* A **2** 893

——1987b *Gauge Fields and Strings* (Chur: Harwood)

Ramond P M 1971 Dual theory for free fermions *Phys. Rev.* D **3** 2415

• Scherk J 1975 An introduction to the theory of dual models and strings *Rev. Mod. Phys.* **47** 123

Scherk J and Schwarz J 1974 Dual models for non-hadrons *Nucl. Phys.* B **81** 118

Schwarz J 1982 Superstring theory *Phys. Rep.* **89** 223

——1985a Superstrings *Symp. on Anomalies, Geometry, Topology* ed W Bardeen and A White (Singapore: World Scientific) p 50

——(ed) 1985b *Superstrings. The First 15 Years of Superstring Theory* in 2 volumes (Singapore: World Scientific)

Siegel W 1985 Covariantly second-quantized string II, III *Phys. Lett.* B **151** 391, 396

——1988 *Introduction to String Field Theory* (Singapore: World Scientific)

Siegel W and Zwiebach B 1986 Gauge string fields *Nucl. Phys.* B **263** 105

Stelle K 1977 Renormalization of higher-derivative quantum gravity *Phys. Rev.* D **16** 953

Sundermeyer K 1982 *Constrained Dynamics* (Berlin: Springer)

Taylor J G 1986 A survey of string and superstring theories, Lectures at the 22nd Karpacz Winter School of Theoretical Physics 'On fields and geometry', Karpacz, Poland, 17 February–1 March

Thorn C 1985 Introduction to the theory of relativistic strings *Unified String Theories* ed M Green and D Gross (Singapore: World Scientific) p 5

——1989 String field theory *Phys. Rep.* **175** 1

Van Nieuwenhuizen P 1973 On ghost-free tensor Lagrangians and linearized gravitation *Nucl. Phys.* B **60** 478

Veneziano G 1968 Construction of a crossing-symmetric, Regge-behaved amplitude for linearly rising trajectories *Nuovo Cimento* A **57** 190

Virasoro M A 1970 Subsidiary conditions and ghosts in dual-resonance models *Phys. Rev.* D **1** 2933

Weinberg S 1987 Covariant path-integral approach to string theory, Lectures at the

3rd Jerusalem Winter School of Theoretical Physics, University of Texas *Preprint* UTTG-17-87

West P 1985 Introduction to gauge covariant string field theory *Preprint* CERN-TH 4304/85

——1986a Gauge-covariant string field theory *Preprint* CERN-TH 4460/86

• ——1986b *Introduction to Supersymmetry and Supergravity* (Singapore: World Scientific)

——1989 An introduction to string theory *Preprint* CERN-TH 51565/ 88; *Acta Phys. Pol.* B **20** 471

Witten E 1986 Non-commutative geometry and string field theory *Nucl. Phys.* B **268** 253

——1996 Reflections on the fate of spacetime *Phys. Today* **April** 24

Appendices

Abers E S and Lee B W 1973 Gauge theories *Phys. Rep.* C **9** 1

Achúcarro A and Townsend P 1986 A Chern–Simons action for three-dimensional anti-de Sitter supergravity theories *Phys. Lett.* B **180** 89

Anderson J L 1970 Scale invariance of the second kind and the Brans–Dicke scalar–tensor theory *Phys. Rev.* D **3** 1689

Ashtekar A 1988 *New Perspectives in Canonical Quantum Gravity* (Napoli: Bibliopolis)

——1991 *Non-perturbative Canonical Quantum Gravity* (notes prepared in collaboration with R S Tate) (Singapore: World Scientific)

Bañados M 1994 Global charges in Chern–Simons theory and $2+1$ black hole *Phys. Rev.* D **52** 5861 (hep-th/9405171)

——1999a Three-dimensional quantum geometry and black holes, Invited talk at the Second Meeting 'Trends in Theoretical Physics', held in Buenos Aires, December, 1998 (hep-th/9901148)

——1999b Notes on black holes and three dimensional gravity *Proc. VIII Mexican School on Particles and Fields* to appear (hep-th/9903244)

Bañados M, Brotz T and Ortiz M 1999 Boundary dynamics and the statistical mechanics of the $2+1$ dimensional hole *Nucl. Phys.* B **545** 340 (hep-th/9802076)

Bañados M, Henneaux M, Teitelboim C and Zanelli J 1993 Geometry of $2+1$ black hole *Phys. Rev.* D **48** 1506 (gr-qc/9302012)

Bañados M, Tronsoco R and Zanelli J 1996 Higher dimensional Chern–Simons supergravity *Phys. Rev.* D **54** 2605

Barut A O and Raczka R 1977 *Theory of Group Representations and Applications* (Warsaw: PWN—Polish Scientific)

Berestetskii V B, Lifshitz E M and Pitaevskii L P 1982 *Quantum Electrodynamics* 2nd edn, translated by J B Sykes and J S Bell (Oxford: Pergamon)

Blagojević M and Vasilić M 1987 Constraint algebra in Poincaré gauge theory *Phys. Rev.* D **36** 1679

Brans C and Dicke R 1961 Mach's principle and a relativistic theory of gravitation *Phys. Rev.* **124** 925

Brown J D 1988 *Lower Dimensional Gravity* (Singapore: World Scientific)

Brown J D and Henneaux M 1986 Central charges in the canonical realization of assymptrotic symmetries: an example from three dimensional gravity *Commun. Math. Phys.* **104** 207

Canuto V, Adams P, Hsieh S H and Tsiang E 1976 Gauge covariant theory of gravitation *Preprint* Institute for Space Studies, New York

Carlip S 1995 Lectures on $(2+1)$-dimensional gravity (gr-qc/9503024)

Carruthers A 1971 *Spin and Isospin in Particle Physics* (New York: Gordon and Breach)

Castellani L 1982 Symmetries of constrained Hamiltonian systems *Ann. Phys., NY* **143** 357

Choquet-Bruhat Y, de Witt-Morette C and Dillard-Bleick M 1977 *Analysis, Manifolds and Physics* (Amsterdam: North-Holland)

Coussaert O and Henneaux M 1994 Supersymmetry of the $2+1$ black holes *Phys. Rev. Lett.* **72** 183 (hep-th/9310194)

Coussaert O, Henneaux M and van Driel P 1995 The asymptotic dynamics of three-dimensional Einstein gravity with a negative cosmological constant *Class. Quantum Grav.* **12** 2916 (gr-qc/9506019-v2)

Deser S 1970 Scale invariance and gravitational coupling *Ann. Phys., NY* **59** 248

Dirac P A M 1964 *Lectures on Quantum Mechanics* (New York: Yeshiva University)

——1973 Long range forces and broken symmetries *Proc. R. Soc.* A **333** 403

Dubrovin B, Novikov S and Fomenko A 1979 *Sovremennaya Geometriya* (Moscow: Nauka)

Eguchi T, Gilkey P B and Hanson A J 1980 Gravitation, gauge theories and differential geometry *Phys. Rep.* C **66** 213

Fukujama T 1983 De Sitter invariant gravity coupled to matters and its cosmological consequences, Kyoto University *Preprint* RIMS-442

German G 1985 Brans–Dicke-type models with torsion *Phys. Rev.* D **32** 3307

Gürsey F 1964 Introduction to the de Sitter group *Group Theoretical Concepts and Methods in Elementary Particle Physics* ed F Gürsey (New York: Gordon and Breach) p 365

Horowitz G 1991 Ashtekar's approach to quantum gravity, a talk at the conference *Strings and Symmetries (Stony Brook, 1991)* University of California *Preprint* UCSBTH-91-40

Hawking S and Ellis G F R 1973 *The Large Scale Structure of Spacetime* (Cambridge: Cambridge University Press)

Hsu J P 1979 Gravity as Yang–Mills' spacetime gauge fields Space Sciences Laboratory *Preprint* No 78-103, Huntsville, AL

Inönü E 1964 Contractions of Lie groups and their representations *Group Theoretical Concepts and Methods in Elementary Particle Physics* ed F Gürsey (New York: Gordon and Breach) p 391

Kamimura K and Fukuyama T 1990 Ashtekar's formalism in first-order tetrad form *Phys. Rev.* D **41** 1885

Kibble T W B 1961 Lorentz invariance and the gravitational field *J. Math. Phys.* **2** 212

——1965 The quantum theory of gravitation *High Energy Physics and Elementary Particles, Proc. Trieste Symposium (Trieste, 1965)* ed G Fronsdal and A Salam (Vienna: IAEA) p 885

Kibble T W B and Stelle K S 1986 Gauge theories of gravity and supergravity *Progress in Quantum Field Theory* ed H Ezava and S Kamefuchi (Amsterdam: Elsevier) p 57

Kim Sung-Won 1986 Brans–Dicke theory in general spacetime with torsion *Phys. Rev.* D **34** 1011

Luciani J F 1978, Spacetime geometry and symmetry breaking *Nucl. Phys.* B **135** 111

Mac Dowell S W and Mansouri F 1977 Unified geometric theory of gravity and supergravity *Phys. Rev. Lett.* **38** 739

Mielke E W 1992 Ashtekar's complex variables in general relativity and its teleparallel equivalent *Ann. Phys., NY* **219** 78

——1998 Chern–Simons solution of the Ashtekar constraints for the teleparallelism equivalent of gravity *Acta Phys. Polon.* B **29** 871

Misner C H, Thorne K S and Wheeler J A 1970 *Gravitation* (San Francisco, CA: Freeman)

Mishenko A S and Fomenko A T 1980 *Kurs Diferencial'noi Geometrii i Topologii* (Moscow: Moscow University)

Navaro-Salas J and Navaro P 1998 A note on Einstein gravity on AdS_3 and boundary conformal field theory *Phys. Lett.* B **439** 262 (hep-th/9807019)

Ne'eman Y and Šijački Dj 1985 $SL(4, R)$ world spinors and gravity *Phys. Lett.* B **157** 275

——1987 Group-topology, covariance and curved-space spinors *Int. J. Mod. Phys.* A **2** 1655

Nikolić 1992 Constraint algebra from local Poincaré symmetry *Gen. Rel. Grav.* **24** 159

——1995 Dirac Hamiltonian formulation and algebra of the constraints in the Einstein–Cartan theory *Class. Quantum Grav.* **12** 3103

Novozhilov Yu V 1972 *Vvedenie v Teoriyu Elementarnyh Chastitz* (Moscow: Nauka)

Peldan P 1994 Actions for gravity, with generalizations: a review *Class. Quantum Grav.* **11** 1087

Pietenpol J L, Incoul R and Speiser D 1974 Remarks on Dirac's new theory *Phys. Rev. Lett.* **33** 387

Ponomariev V N, Barvinsky A O and Obukhov Yu N 1985 *Geometro-dinamicheskie Metody i Kalibrovochnyi Podhod k Teorii Gravitacionyh Vzaimodeistvii* (Moscow: Energoatomizdat)

Regge T 1986 On broken symmetries and gravity *Phys. Rep.* **137** 31

Rovelli C 1991 Ashtekar's formulation of general relativity and loop-space non-perturbative quantum gravity: a report *Class. Quantum Grav.* **8** 1613

Šijački Dj 1998 World spinors revisited *Acta Phys. Polon.* B **29** 1089

Smalley L 1986 Brans–Dicke-type models with nonmetricity *Phys. Rev.* D **33** 3590

Smollin L 1992 Recent developments in nonperturbative quantum gravity, Syracuse University *Preprint* SU-GP-92/2-2

Sohnius M 1985 Supersymmetry *Phys. Rep.* C **128** 39

Srivastava Prem P 1986 *Supersymmetry, Superfields and Supergravity: An Introduction* (Bristol: IOP Publishing)

Stelle K S and West P C 1979 De Sitter gauge invariance and the geometry of the Einstein–Cartan theory *J. Phys. A: Math. Gen.* **12** 205

——1980 Spontaneously broken de Sitter geometry and the gravitational holonomy group *Phys. Rev.* D **21** 1466

Strominger A 1998 Black hole entropy from near horizon microstates *JHEP* **9802** 009 (hep-th/9712251)

Teitelboim C 1973 How commutators of constraints reflect the spacetime structure *Ann. Phys., NY* **79** 542

——1981 The Hamiltonian structure of spacetime *General Relativity and Gravitation—One Hundred Years After the Birth of Albert Einstein* vol I, ed A Held (New York: Plenum) p 195

Townsend P K 1977 Small-scale structure of spacetime as the origin of the gravitational constant *Phys. Rev.* D **15** 2795

Utiyama R 1956 Invariant theoretical interpretation of interactions *Phys. Rev.* **101** 1597

Van Dam H 1974 *Theory of Gravity* (Nijmegen: ITP—University of Nijmegen)

van Nieuwenhuizen P 1981 Supergravity *Phys. Rep.* C **68** 192

Weinberg S 1964 Photons and gravitons in S-matrix theory: Derivation of charge conservation and equality of gravitational and inertial mass *Phys. Rev.* B **135** 1049

Weinberg S 1972 *Gravitation and Cosmology* (New York: Wiley)

West P C 1978 A geometric gravity Lagrangian *Phys. Lett.* B **76** 569

Witten E 1988 $2 + 1$ gravity as an exactly soluable system *Nucl. Phys.* B **311** 46

Wybourne G B 1973 *Classical Groups for Physicists* (New York: Wiley)

Yang C N and Mills R 1954 Conservation of isotopic spin and isotopic gauge invariance *Phys. Rev.* **96** 191

Zanelli J 1999 Chern–Simons gravity: from $2+1$ to $2n+1$ dimensions, Lecture presented at the XX Encontro de Fisica de Particulas e Campos, Brazil, October 1988, and at the Fifth La Hechicera School, Venezuela, November 1999 *Braz. J. Phys.* **30** 251 (hep-th/0010049)

Zee A 1981 Grand unification and gravity—selected topics *Proc. Fourth Kyoto Summer Institute on Grand Unified Theories and Related Topics (Kyoto, Japan, 1981)* ed M Konuma and T Maskawa (Singapore: World Scientific) p 143

Notations and conventions

Notations and conventions are defined in the text, at the places where they first appear. The following list contains a set of often used abbreviations, symbols and conventions.

Abbreviations

ADM	Arnowitt–Deser–Misner	PGT	Poincaré gauge theory
EC	Einstein–Cartan	PR	Principle of relativity
FC	First class	SCT	Special conformal transformations
GR	General relativity	SR	Special relativity
KK	Kaluza–Klein	SS	Supersymmetry
PB	Poisson bracket	WGT	Weyl gauge theory
PE	Principle of equivalence		

X_d	d-dimensional differentiable manifold, $d = 4 + D$.
M_d	d-dimensional Minkowski space with metric $\boldsymbol{\eta} = (+1, -1, \dots, -1)$.

Indices

	Repeated indices are summed.
$i, j, k, \dots$	Local Lorentz indices in X_4, run over 0, 1, 2, 3; the related coordinates are x^i.
$\mu, \nu, \lambda, \dots$	Coordinate indices in X_4, run over 0, 1, 2, 3; the coordinates are x^μ.
$a, b, c, \dots$	Spatial local Lorentz indices in X_4, run over 1, 2, 3.
$\alpha, \beta, \gamma, \dots$	Spatial coordinate indices in X_4, run over 1, 2, 3.
$I, J, K, \dots$	Local Lorentz indices in X_5, run over 0, 1, 2, 3, 5; the coordinates are z^I.
$M, N, L, \dots$	Coordinate indices in X_5, run over 0, 1, 2, 3, 5; the coordinates are $z^M = (x^\mu, y)$ (for $d > 5$, $y \to y^\alpha$).

$\alpha, \beta, \gamma, \dots$	Coordinate indices in X_2 (string), run over 0, 1; the coordinates are $\xi^\alpha = (\tau, \sigma)$.
$a, b, c, \dots$	Non-coordinate indices of internal groups, run over $1, 2, \dots, n$.
$a, \dot{a}, \dots$	Two-dimensional spinor indices, run over 1, 2; $\dot{1}, \dot{2}$.
$\alpha, \beta, \gamma, \dots$	Coordinate indices of internal groups, run over $1, 2, \dots, n$.
	Symmetrization and antisymmetrization: $X_{(ij)} = \frac{1}{2}(X_{ij} + X_{ji})$, $X_{[ij]} = \frac{1}{2}(X_{ij} - X_{ji})$.

Tensor and spinor fields

φ, ϕ	Scalar fields.
φ_μ, A_μ	Vector fields.
$\varphi_{\mu\nu}, h_{\mu\nu}$	Symmetric tensor fields.
$b_{\mu\nu}$	Antisymmetric tensor, field strength: $H_{\mu\nu\lambda} = \partial_\mu b_{\nu\lambda}+$ cyclic.
ψ, Ψ	Dirac field, spin $= \frac{1}{2}$.
ψ_μ	Rarita–Schwinger field, spin $= \frac{3}{2}$.
$\sigma^\mu, \bar{\sigma}^\mu$	Pauli spin matrices, $\sigma^\mu = (1, \boldsymbol{\sigma})$, $\bar{\sigma}^\mu = (1, -\boldsymbol{\sigma})$.
γ^μ, γ^M	Dirac matrices in $d = 4$ and $d > 4$.
$\varepsilon_\mu, \varepsilon_{\mu\nu}$	Polarization vector and tensor.

Completely antisymmetric symbols

ε^{ijkl}	$\varepsilon^{0123} = +1$, but $\varepsilon_{0123} = -1$ (in M_4).
ε^{abc}	$\varepsilon^{abc} \equiv \varepsilon^{0abc}$, $\varepsilon_{abc} \equiv \varepsilon_{0abc}$, hence $\varepsilon^{123} = +1$, $\varepsilon_{123} = -1$ (in M_4).
ϵ^{abc}	$\epsilon^{123} = \epsilon_{123} = +1$ (in E_3).

Geometric objects in X_4

$\boldsymbol{e}_\mu, \boldsymbol{e}_i$	The coordinate and Lorentz basis of the tangent space, $\boldsymbol{e}_\mu = e^i{}_\mu \boldsymbol{e}_i$, $\boldsymbol{e}_i = e_i{}^\mu \boldsymbol{e}_\mu$ (alternatively: $e^i{}_\mu \to b^i{}_\mu$, $e_i{}^\mu \to h_i{}^\mu$).
$\boldsymbol{g}$	Metric tensor, $g_{\mu\nu} = \boldsymbol{e}_\mu \cdot \boldsymbol{e}_\nu$, $g_{ij} \equiv \eta_{ij} = \boldsymbol{e}_i \cdot \boldsymbol{e}_j = (+1, -1, -1, -1)$.
$\boldsymbol{u} \cdot \boldsymbol{v}$	The scalar product of two vectors, $\boldsymbol{u} \cdot \boldsymbol{v} = \eta_{ij} u^i v^j = g_{\mu\nu} u^\mu v^\nu$.
$\omega^{ij}{}_\mu$	Spin connection, alternative notation: $A^{ij}{}_\mu$.
$\nabla_\mu(\omega)$	ω-covariant derivative, $\nabla_\mu(\omega) u^i = \partial_\mu u^i + \omega^i{}_{s\mu} u^s$, alternative notation: $D_\mu(\omega)$.
$\nabla_k(\omega)$	$\nabla_k(\omega) = e_k{}^\mu \nabla_\mu(\omega)$, alternative notation: $D_k(\omega)$.
$\Gamma^\lambda_{\rho\mu}$	Connection in the coordinate basis.
$\nabla_\mu(\Gamma)$	Γ-covariant derivative, $\nabla_\mu(\Gamma) u^\lambda = \partial_\mu u^\lambda + \Gamma^\lambda_{\rho\mu} u^\rho$, alternative notation: $D_\mu(\Gamma)$.
$\left\{ {\lambda \atop \rho\mu} \right\}$	Christoffel symbol.

$Q_{\mu\nu\lambda}$	Non-metricity, $Q_{\mu\nu\lambda} = -\nabla_\mu(\Gamma)g_{\nu\lambda}$.
φ_μ	Weyl vector, $\nabla_\mu(\Gamma)g_{\nu\lambda} = \varphi_\mu g_{\nu\lambda}$.
$R^{ij}{}_{\mu\nu}(\omega)$	Curvature, $R^{ij}{}_{\mu\nu}(\omega) = \partial_\mu\omega^{ij}{}_\nu + \omega^i{}_{s\mu}\omega^{sj}{}_\nu - (\mu \leftrightarrow \nu)$.
$T^i{}_{\mu\nu}(e)$	Torsion, $T^i{}_{\mu\nu}(e) = \nabla_\mu(\omega)e^i{}_\nu - \nabla_\nu(\omega)e^i{}_\mu$.
$F_{\mu\nu}(\varphi)$	Weyl curvature, $F_{\mu\nu}(\varphi) = \partial_\mu\varphi_\nu - \partial_\nu\varphi_\mu$.

Gauge objects in X_4

$A^a{}_\mu$	Non-Abelian gauge potential.
$F^a{}_{\mu\nu}$	Non-Abelian field strength, $F^a{}_{\mu\nu} = \partial_\mu A^a{}_\nu - \partial_\nu A^a{}_\mu + f_{bc}{}^a A^b{}_\mu A^c{}_\nu$.
$A^{ij}{}_\mu, b^i{}_\mu$	Compensating fields in PGT.
$F^{ij}{}_{\mu\nu}(A)$	Lorentz field strength.
$F^i{}_{\mu\nu}(b)$	Translation field strength.
$A^{ij}{}_\mu, b^i{}_\mu, B_\mu$	Compensating fields in WGT.
$F_{\mu\nu}(B)$	Dilatation field strength.

Geometric objects in X_d ($d > 4$ or $d = 2$)

$\hat{\boldsymbol{e}}_M, \hat{\boldsymbol{e}}_I$	The coordinate and Lorentz basis of the tangent space, $\hat{\boldsymbol{e}}_M = b^I{}_M\hat{\boldsymbol{e}}_I$, $\hat{\boldsymbol{e}}_I = h_I{}^M\hat{\boldsymbol{e}}_M$.
$\hat{\boldsymbol{g}}$	Metric tensor, $\hat{g}_{MN} = \hat{\boldsymbol{e}}_M \cdot \hat{\boldsymbol{e}}_N$, $\hat{g}_{IJ} \equiv \eta_{IJ} = \hat{\boldsymbol{e}}_I \cdot \hat{\boldsymbol{e}}_J = (+1, -1, \ldots, -1)$.
$\hat{A}^{IJ}{}_M$	Spin connection.
$\hat{\Gamma}^L_{RM}$	Connection in the coordinate basis.
$\hat{R}^{IJ}{}_{MN}$	Curvature.
$\hat{T}^I{}_{MN}$	Torsion.
$\boldsymbol{\gamma}$	Induced metric on X_2 (string), $\gamma_{\alpha\beta}$.

Spaces over X_4

L_4	linearly connected space, $L_4 = (X_4, \Gamma) = (X_4, \omega)$.
Y_4	Weyl–Cartan space.
W_4	Weyl space.
U_4	Riemann–Cartan space.
V_4	Riemann space.
T_4	Weitzenböck (teleparallel) space.
M_4	Minkowski space.

Constants

c	the speed of light, $c = 3.00 \times 10^{10}$ cm s^{-1}.

G	Newton's constant, $G = 6.67 \times 10^{-8}$ cm^3 g^{-1} s^{-2}.
κ	$\kappa = 8\pi G/c^2 \approx 1.86 \times 10^{-27}$ cm g^{-1}.
$\hbar$	the Planck constant, $\hbar = 1.05 \times 10^{-27}$ erg s.
l_P	the Planck length, $l_P = \sqrt{\hbar G/c^3} \approx 1.61 \times 10^{-33}$ cm.
E_P	the Planck energy, $E_P = c^2\sqrt{\hbar c/G} \approx 1.22 \times 10^{19}$ GeV.

Index