Gravity and Strings

TOMÁS ORTÍN

CAMBRIDGE MONOGRAPHS
ON MATHEMATICAL PHYSICS

CAMBRIDGE www.cambridge.org/9780521824750

Gravity and Strings

One appealing feature of string theory is that it provides a theory of quantum gravity. *Gravity and Strings* is a self-contained, pedagogical exposition of this theory, its foundations, and its basic results.

In Part I, the foundations are traced back to the very early special-relativistic field theories of gravity, showing how such theories, which are associated with the concept of the graviton, lead to general relativity. Gauge theories of gravity are then discussed and used to introduce supergravity theories.

Part II covers some of the most interesting solutions of general relativity and its generalizations. These include Schwarzschild and Reissner–Nordström black holes, the Taub–NUT solution, gravitational instantons, and gravitational waves. Kaluza–Klein theories and the uses of residual supersymmetries are discussed in detail.

Part III presents string theory from the effective-action point of view, using the results found earlier in the book as background. The supergravity theories associated with superstrings and M theory are thoroughly studied, and used to describe dualities and classical solutions related to non-pertubative states of these theories. A brief account of extreme black-hole entropy calculations is also given.

This unique book will be useful as a reference for graduate students and researchers, as well as a complementary textbook for courses on gravity, supergravity, and string theory.

TOMÁS ORTÍN completed his graduate studies and obtained his Ph.D. at the Universidad Autónoma de Madrid. He then worked as a postdoctoral student in the Physics Department of Stanford University supported by a Spanish Government grant. Between 1993 and 1995 he was EU Marie Curie postdoctoral fellow in the String Theory Group of the Physics Department of Queen Mary College, University of London, and from 1995 to 1997, he was a Fellow in the Theory Division of CERN. He is currently a Staff Scientist at the Spanish Research Council and a member of the Institute for Theoretical Physics of the Universidad Autónoma de Madrid. Dr Ortín has taught several graduate courses on advanced general relativity, supergravity, and strings. His research interests lie in string theory, gravity, quantum gravity, and black-hole physics.

CAMBRIDGE MONOGRAPHS ON
MATHEMATICAL PHYSICS

General editors: P. V. Landshoff, D. R. Nelson, S. Weinberg

Gravity and Strings

TOMÁS ORTÍN
Spanish Research Council and Universidad Autónoma de Madrid
(CSIC)

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521824750

First published in print format 2004

ISBN-13 978-0-511-18476-5 eBook (NetLibrary)
ISBN-10 0-511-18476-X eBook (NetLibrary)

ISBN-13 978-0-521-82475-0 hardback
ISBN-10 0-521-82475-3 hardback

To Marimar, Diego, and Tomás,
the sweet strings that tie me to
the real world

Contents

Preface

String theory has lived for the past few years during a golden era in which a tremendous upsurge of new ideas, techniques, and results has proliferated. In what form they will contribute to our collective enterprise (theoretical physics) only time can tell, but it is clear that many of them have started to have an impact on closely related areas of physics and mathematics and, even if string theory does not reach its ultimate goal of becoming a theory of everything, it will have played a crucial, inspiring role.

There are many interesting things that have been learned and achieved in this field that we feel can (and perhaps should) be taught to graduate students. However, we have found that this is impossible without the introduction of many ideas, techniques, and results that are not normally taught together in standard courses on general relativity, field theory or string theory, but which have become everyday tools for researchers in this field: black holes, strings, membranes, solitons, instantons, unbroken supersymmetry, Hawking radiation .... They can, of course, be found in various textbooks and research papers, presented from various viewpoints, but not in a single reference with a consistent organization of the ideas (not to mention a consistent notation).

These are the main reasons for the existence of this book, which tries to fill this gap by covering a wide range of topics related, in one way or another, to what we may call *semiclassical string gravity*. The selection of material is according to the author's taste and personal preferences with the aim of self-consistency and the ultimate goal of creating a basic, pedagogical, reference work in which all the results are written in a consistent set of notations and conventions. Some of the material is new and cannot be found elsewhere.

Precisely because of the blend of topics we have touched upon, although a great deal of background material is (briefly) reviewed here, this cannot be considered a textbook on general relativity, supergravity or string theory. Nevertheless, some chapters can be used in graduate courses on these matters, either providing material for a few lectures on a selected topic or combined (as the author has done with the first part, which is self-contained) into an advanced (and a bit eclectic) course on gravity.

It has not been too difficult to order logically the broad range of topics that had to be discussed, though. We can view string theory as the summit of a pyramid whose building blocks are the theories, results, and data that become more and more fundamental and basic the more we approach the base of the pyramid. At the very bottom (Part I) one can find tools such as differential geometry and the use of symmetry in physics and fundamental theories of gravity such as general relativity and extensions to accommodate fermions such as the CSK theory and supergravity. The rest of the book is supported by it. In particular, we can see string theory as the culmination of long-term efforts to construct a theory of quantum gravity for a spin-2 particle (the graviton) and our approach to general relativity as the only self-consistent classical field theory of the graviton is intended to set the ground for this view.

Part II investigates consequences, results, and extensions of general relativity through some of its simplest and most remarkable solutions, which can be regarded as point-particle-like: the Schwarzschild and Reissner–Nordström solutions, gravitational waves, and the Taub-NUT solution. In the course of this study we introduce the reader to black holes, "no-hair theorems," black-hole thermodynamics, Hawking radiation, gravitational instantons, charge quantization, electric–magnetic duality, the Witten effect etc. We will also explain the essentials of dimensional reduction and will obtain black-hole solutions of the dimensionally reduced theory. To finish Part II we introduce the reader to the idea and implications of residual supersymmetry. We will review all our results on black-hole thermodynamics and other black-hole properties in the light of unbroken supersymmetry.

Part III introduces strings and the string effective action as a particular extension of general relativity and supergravity. String dualities and extended objects will be studied from the string-effective-action (spacetime) point of view, making use of the results of Parts I and II and paying special attention to the relation between worldvolume and spacetime phenomena. This part, and the book, closes with an introduction to the calculation of black-hole entropies using string theory.

During these years, I have received the support of many people to whom this book, and I personally, owe much: Enrique Álvarez, Luis Álvarez-Gaumé, and my long-time collaborators Eric Bergshoeff and Renata Kallosh encouraged me and gave me the opportunity to learn from them. My students Natxo Alonso-Alberca, Ernesto Lozano-Tellechea, and Patrick Meessen used and checked many versions of the manuscript they used to call the *PRC*. Their help and friendship in these years has been invaluable. Roberto Emparan, José Miguel Figueroa-O'Farrill, Yolanda Lozano, Javier Más, Alfonso Vázquez-Ramallo, and Miguel Ángel Vázquez-Mozo read several versions of the manuscript and gave me many valuable comments and advice, which contributed to improving it. I am indebted to Arthur Greenspoon for making an extremely thorough final revision of the manuscript.

Nothing would have been possible without Marimar's continuous and enduring support.

If, in spite of all this help, the book has any shortcomings, the responsibility is entirely mine. Comments and notifications of misprints can be sent to the e-mail address `tomas.ortin@uam.es`. The *errata* will be posted in `http://gesalerico.ft.uam.es/prc/misprints.html`.

This book started as a written version of a review talk on string black holes prepared for the first String Theory Meeting of the Benasque Center for Theoretical Physics, back in 1996, parts of it made a first public appearance in a condensed form as lectures for the charming Escuela de Relatividad, Campos y Cosmología "La Hechicera" organized by the Universidad de Los Andes (Mérida, Venezuela), and it was finished during a long-term visit to the CERN Theory Division. I would like to thank the organizers and members of these institutions for their invitations, hospitality, and economic support.

Part I
Introduction to gravity and supergravity

Let no one ignorant of Mathematics enter here.

Inscription above the doorway of Plato's Academy

1
Differential geometry

The main purpose of this chapter is to fix our notation and to review the ideas and formulae of differential geometry we will make heavy use of. There are many excellent physicist-oriented references on differential geometry. Two that we particularly like are [347] and [715]. Our approach here will be quite pragmatic, ignoring many mathematical details and subtleties that can be found in the many excellent books on the subject.

1.1 World tensors

A *manifold* is a topological space that looks (i.e. it is homeomorphic to) locally (i.e. in a *patch*) like a piece of $\mathbb{R}^d$. d is the dimension of the manifold and the correspondence between the patch and the piece of $\mathbb{R}^n$ can be used to label the points in the patch by Cartesian $\mathbb{R}^n$ coordinates x^μ. In the overlap between different patches the different coordinates are consistently related by a *general coordinate transformation* (GCT) $x'^\mu(x)$. Only objects with good transformation properties under GCTs can be defined globally on the manifold. These objects are *tensors*.

A *contravariant vector field* (or $(1,0)$-type tensor or just "vector") $\xi(x) = \xi^\mu(x)\partial_\mu$ is defined at each point on a d-dimensional smooth manifold by its action on a function

$$\xi : f \longrightarrow \xi f = \xi^\mu \partial_\mu f, \tag{1.1}$$

which defines another function. These objects span a d-dimensional linear vector space at each point of the manifold called the *tangent space* $\mathrm{T}_p^{(1,0)}$. The d functions $\xi^\mu(x)$ are the vector components with respect to the *coordinate basis* $\{\partial_\mu\}$.

A *covariant vector field* (or (0, 1)-type tensor or *differential 1-form*) is an element of the dual vector space (sometimes called the *cotangent space*) $\mathrm{T}_p^{(0,1)}$ and therefore transforms vectors into functions. The elements of the basis dual to the coordinate basis of contravariant vectors are usually denoted by $\{dx^\mu\}$ and, by definition,

$$\langle dx^\mu | \partial_\nu \rangle \equiv \delta^\mu{}_\nu, \tag{1.2}$$

which implies that the action of a form $\omega = \omega_\mu dx^\mu$ on a vector $\xi(x) = \xi^\mu(x)\partial_\mu$ gives the function[1]

$$\langle\omega|\xi\rangle = \omega_\mu \xi^\mu. \tag{1.3}$$

Under a GCT vectors and forms transform as functions, i.e. $\xi'(x') = \xi(x(x'))$ etc., which means for their components in the associated coordinate basis

$$\frac{\partial x'^\rho}{\partial x^\mu}\,\xi^\mu(x(x')) = \xi'^\rho(x'), \qquad \omega_\mu(x(x'))\,\frac{\partial x^\mu}{\partial x'^\rho} = \omega'_\rho(x'). \tag{1.4}$$

More general tensors of type (q, r) can be defined as elements of the space $\mathrm{T}_p^{(q,r)}$ which is the tensor product of q copies of the tangent space and r copies of the cotangent space. Their components $\mathrm{T}^{\mu_1\cdots\mu_q}{}_{\nu_1\cdots\nu_r}$ transform under GCTs in the obvious way.

It is also possible to define *tensor densities of weight* w whose components in a coordinate basis change under a GCT with an extra factor of the Jacobian raised to the power $w/2$. Thus, for weight w, the vector density components $\mathfrak{v}^\mu$ and the form density components $\mathfrak{w}_\mu$ transform according to

$$\begin{aligned} \left|\frac{\partial x'}{\partial x}\right|^{w/2} \frac{\partial x'^\rho}{\partial x^\mu}\,\mathfrak{v}^\mu(x(x')) &= \mathfrak{v}'^\rho(x'), \\ \mathfrak{w}_\mu(x(x'))\,\frac{\partial x^\mu}{\partial x'^\rho}\left|\frac{\partial x'}{\partial x}\right|^{w/2} &= \mathfrak{w}'_\rho(x'), \end{aligned} \tag{1.5}$$

where for the Jacobian we use the notation

$$\left|\frac{\partial x'}{\partial x}\right| \equiv \det\left(\frac{\partial x'^\rho}{\partial x^\mu}\right). \tag{1.6}$$

An infinitesimal GCT[2] can be written as follows:

$$\delta x^\mu = x'^\mu - x^\mu = \epsilon^\mu(x). \tag{1.7}$$

The corresponding infinitesimal transformations of scalars ϕ and contravariant and covariant *world vectors* (an alternative name for components in the coordinate basis) are:[3]

$$\begin{aligned} \delta\phi &= -\epsilon^\lambda\partial_\lambda\phi && \equiv -\mathcal{L}_\epsilon\phi, \\ \delta\xi^\mu &= -\epsilon^\lambda\partial_\lambda\xi^\mu + \partial_\nu\epsilon^\mu\xi^\nu && \equiv -\mathcal{L}_\epsilon\xi^\mu \equiv -[\epsilon, \xi]^\mu, \\ \delta\omega_\mu &= -\epsilon^\lambda\partial_\lambda\omega_\mu - \partial_\mu\epsilon^\nu\omega_\nu && \equiv -\mathcal{L}_\epsilon\omega_\mu, \end{aligned} \tag{1.8}$$

[1] Summation over repeated indices in any position will always be assumed, unless they are in parentheses.

[2] This is an element of a one-parameter group of GCTs (the unit element corresponding to the value 0 of the parameter) with a value of the parameter much smaller than 1.

[3] We use the functional variations $\delta\phi \equiv \phi'(x) - \phi(x)$ which refer to the value of the field ϕ at two different points whose coordinates are equal in the two different coordinate systems. They are denoted in [795] by δ_0. They should be distinguished from the total variations $\tilde{\delta} = \phi'(x') - \phi(x)$ which refer to the values of the field ϕ at the same point in two different coordinate systems. The relation between the two is $\delta\phi = \tilde{\delta}\phi - \epsilon^\mu\partial_\mu\phi$. The piece $-\epsilon^\lambda\partial_\lambda\phi$ that appears in δ variations is the "transport term," which is not present in other kinds of infinitesimal variations. The transformations δ do enjoy a group property (their commutator is another δ transformation), whereas the transformations $\tilde{\delta}$ or the transport terms by themselves don't.

and, for weight-w scalar densities $\mathfrak{f}$, vector density components $\mathfrak{v}^\mu$ and the form density components $\mathfrak{w}_\mu$,

$$
\begin{aligned}
\delta\mathfrak{f} &= -\epsilon^\lambda\partial_\lambda\mathfrak{f} - w\partial_\lambda\epsilon^\lambda\mathfrak{f} &&\equiv -\mathcal{L}_\epsilon\mathfrak{f},\\
\delta\mathfrak{v}^\mu &= -\epsilon^\lambda\partial_\lambda\mathfrak{v}^\mu + \partial_\nu\epsilon^\mu\mathfrak{v}^\nu - w\partial_\lambda\epsilon^\lambda\mathfrak{v}^\mu &&\equiv -\mathcal{L}_\epsilon\mathfrak{v}^\mu,\\
\delta\mathfrak{w}_\mu &= -\epsilon^\lambda\partial_\lambda\mathfrak{w}_\mu - \partial_\mu\epsilon^\nu\mathfrak{w}_\nu - w\partial_\lambda\epsilon^\lambda\mathfrak{w}_\mu &&\equiv -\mathcal{L}_\epsilon\mathfrak{w}_\mu,
\end{aligned}
\tag{1.9}
$$

where $\mathcal{L}_\epsilon$ is the *Lie derivative* with respect to the vector field ϵ and $[\epsilon, \xi]$ is the *Lie bracket* of the vectors ϵ and ξ. The definition of the Lie derivative can be extended to tensors or weight-w tensor densities of any type:

$$
\begin{aligned}
\mathcal{L}_\epsilon T^{\mu_1\cdots\mu_p}{}_{\nu_1\cdots\nu_q} &= -\delta_\epsilon T^{\mu_1\cdots\mu_p}{}_{\nu_1\cdots\nu_q}\\
&= \epsilon^\rho\partial_\rho T^{\mu_1\cdots\mu_p}{}_{\nu_1\cdots\nu_q} - \partial_\rho\epsilon^{\mu_1}T^{\rho\mu_2\cdots\mu_p}{}_{\nu_1\cdots\nu_q} + \cdots\\
&\quad + \partial_{\nu_1}\epsilon^\rho T^{\mu_1\cdots\mu_p}{}_{\rho\nu_2\cdots\nu_q} - w\partial_\lambda\epsilon^\lambda T^{\mu_1\cdots\mu_p}{}_{\nu_1\cdots\nu_q}.
\end{aligned}
\tag{1.10}
$$

In particular the metric (a symmetric (0, 2)-type tensor to be defined later) and r-form (a fully antisymmetric type $(0, r)$ tensor) transform as follows:

$$
\begin{aligned}
\delta g_{\mu\nu} &= -\epsilon^\lambda\partial_\lambda g_{\mu\nu} - 2g_{\lambda(\mu}\partial_{\nu)}\epsilon^\lambda &&= -\mathcal{L}_\epsilon g_{\mu\nu},\\
\delta B_{\mu_1\cdots\mu_r} &= -\epsilon^\lambda\partial_\lambda B_{\mu_1\cdots\mu_r} - r(\partial_{[\mu_1|}\epsilon^\lambda)B_{\lambda|\mu_2\cdots\mu_r]} &&= -\mathcal{L}_\epsilon B_{\mu_1\cdots\mu_r}.
\end{aligned}
\tag{1.11}
$$

The main properties of the Lie derivative are that it transforms tensors of a given type into tensors of the same given type, it obeys the Leibniz rule $\mathcal{L}_\epsilon(T_1T_2) = (\mathcal{L}_\epsilon T_1)T_2 + T_1\mathcal{L}_\epsilon T_2$, it is connection-independent, and it is linear with respect to ϵ. Furthermore, it satisfies the *Jacobi identity*

$$
[\mathcal{L}_{\xi_1}, [\mathcal{L}_{\xi_2}, \mathcal{L}_{\xi_3}]] + [\mathcal{L}_{\xi_2}, [\mathcal{L}_{\xi_3}, \mathcal{L}_{\xi_1}]] + [\mathcal{L}_{\xi_3}, [\mathcal{L}_{\xi_1}, \mathcal{L}_{\xi_2}]] = 0,
\tag{1.12}
$$

where the brackets stand for commutators of differential operators. The relation between the commutator $[\mathcal{L}_\xi, \mathcal{L}_\epsilon]$ and the Lie bracket $[\xi, \epsilon]$ is

$$
[\mathcal{L}_\xi, \mathcal{L}_\epsilon] = \mathcal{L}_{[\xi,\epsilon]}.
\tag{1.13}
$$

Thus, the Lie bracket is an antisymmetric, bilinear product in tangent space that also satisfies the Jacobi identity

$$
[\xi_1, [\xi_2, \xi_3]] + [\xi_2, [\xi_3, \xi_1]] + [\xi_3, [\xi_1, \xi_2]] = 0,
\tag{1.14}
$$

which one can use to give it the structure of *Lie algebra*.

1.2 Affinely connected spacetimes

The *covariant derivative* of world tensors is defined by

$$
\begin{aligned}
\nabla_\mu\phi &= \partial_\mu\phi,\\
\nabla_\mu\xi^\nu &= \partial_\mu\xi^\nu + \Gamma_{\mu\rho}{}^\nu\xi^\rho,\\
\nabla_\mu\omega_\nu &= \partial_\mu\omega_\nu - \omega_\rho\Gamma_{\mu\nu}{}^\rho,
\end{aligned}
\tag{1.15}
$$

and on weight-w tensor densities by

$$\begin{aligned}\nabla_\mu \mathfrak{f} &= \partial_\mu \mathfrak{f} - w\Gamma_{\mu\rho}{}^{\rho}\mathfrak{f},\\ \nabla_\mu \mathfrak{v}^\nu &= \partial_\mu \mathfrak{v}^\nu + \Gamma_{\mu\rho}{}^{\nu}\mathfrak{v}^\rho - w\Gamma_{\mu\rho}{}^{\rho}\mathfrak{v}^\nu,\\ \nabla_\mu \mathfrak{w}_\nu &= \partial_\mu \mathfrak{w}_\nu - \mathfrak{w}_\rho\Gamma_{\mu\nu}{}^{\rho} - w\Gamma_{\mu\rho}{}^{\rho}\mathfrak{w}_\nu,\end{aligned} \tag{1.16}$$

where Γ is the *affine connection*, and is added to the partial derivative so that the covariant derivative of a tensor transforms as a tensor in all indices. This requires the affine connection to transform under infinitesimal GCTs as follows:

$$\delta\Gamma_{\mu\nu}{}^{\rho} = -\mathcal{L}_\epsilon\Gamma_{\mu\nu}{}^{\rho} - \partial_\mu\partial_\nu\epsilon^\rho, \tag{1.17}$$

and therefore it is not a tensor. In principle it can be any field with the above transformation properties and should be understood as structure added to our manifold. A d-dimensional manifold equipped with an affine connection is sometimes called an *affinely connected space* and is denoted by L_d.

The definition of a covariant derivative can be extended to tensors of arbitrary type in the standard fashion. Its main properties are that it is a linear differential operator that transforms type-(p, q) tensors into $(p, q+1)$ tensors (hence the name covariant) and obeys the Leibniz rule and the Jacobi identity.

Let us now decompose the connection into two pieces symmetric and antisymmetric under the exchange of the covariant indices:

$$\Gamma_{\mu\nu}{}^{\rho} = \Gamma_{(\mu\nu)}{}^{\rho} + \Gamma_{[\mu\nu]}{}^{\rho}. \tag{1.18}$$

The antisymmetric part is called the *torsion* and it is a tensor (which the connection is not)

$$T_{\mu\nu}{}^{\rho} = -2\Gamma_{[\mu\nu]}{}^{\rho}. \tag{1.19}$$

As we have said, the Lie derivative transforms tensors into tensors in spite of the fact that it is expressed in terms of partial derivatives. We can rewrite it in terms of covariant derivatives and torsion terms to make evident the fact that the result is indeed a tensor:

$$\begin{aligned}\mathcal{L}_\epsilon\phi &= \epsilon^\lambda\nabla_\lambda\phi,\\ \mathcal{L}_\epsilon\xi^\mu &= \epsilon^\lambda\nabla_\lambda\xi^\mu - \nabla_\nu\epsilon^\mu\xi^\nu + \epsilon^\lambda T_{\lambda\rho}{}^{\mu}\xi^\rho,\\ \mathcal{L}_\epsilon\omega_\mu &= \epsilon^\lambda\nabla_\lambda\omega_\mu + \nabla_\mu\epsilon^\nu\omega_\nu - \epsilon^\lambda\omega_\rho T_{\lambda\mu}{}^{\rho},\end{aligned} \tag{1.20}$$

etc. It should be stressed that this is just a rewriting of the Lie derivative, which is independent of any connection. There are other connection-independent derivatives. Particularly important is the *exterior derivative* defined on *differential forms* (completely antisymmetric tensors) which we will study later in Section 1.7.

The additional structure of an affine connection allows us to define *parallel transport*. In a generic spacetime there is no natural notion of parallelism for two vectors defined at two different points. We need to transport one of them keeping it "parallel to itself" to the point at which the other is defined. Then we can compare the two vectors at the same point. Using the affine connection, we can define an infinitesimal *parallel displacement* of a covariant vector ω_μ in the direction of ϵ^μ by

$$\delta_{\mathrm{P}_\epsilon}\omega_\mu = \epsilon^\nu\Gamma_{\nu\mu}{}^{\rho}\omega_\rho. \tag{1.21}$$

If $\omega_\mu(x)$ is a vector field, we can compare its value at a given point $x^\mu + \epsilon^\mu$ with the value obtained by parallel displacement from x^μ. The difference is precisely given by the covariant derivative in the direction ϵ^μ:

$$\omega_\mu(x') - (\omega_\mu + \delta_{\mathrm{P}_\epsilon}\omega_\mu)(x) = \epsilon^\nu \nabla_\nu \omega_\mu . \tag{1.22}$$

A vector field whose value at every point coincides with the value one would obtain by parallel transport from neighboring points is a *covariantly constant* vector field, $\nabla_\nu \omega_\mu = 0$.

If the vector tangential to a curve[4] $v^\mu = dx^\mu/d\xi \equiv \dot{x}^\mu$ is parallel to itself along the curve (as a straight line in flat spacetime) then

$$v^\nu \nabla_\nu v^\mu = \ddot{x}^\mu + \dot{x}^\rho \dot{x}^\sigma \Gamma_{\rho\sigma}{}^\mu = 0, \tag{1.23}$$

which is the *autoparallel equation*. This is the equation satisfied by an *autoparallel curve*, which is the generalization of a straight line to a general affinely connected spacetime. There is a second possible generalization based on the property of straight lines of being the shortest possible curves joining two given points (*geodesics*), but it requires the notion of length and we will have to wait until the introduction of metrics.

We can understand the meaning of torsion using parallel transport: let us consider two vectors ϵ_1^μ and ϵ_2^μ at a given point of coordinates x^μ. Let us now consider at the point of coordinates $x^\mu + \epsilon_1^\mu$ the vector $\epsilon_2'^\mu$ obtained by parallel-transporting ϵ_2^μ in the direction ϵ_1^μ and, at the point of coordinates $x^\mu + \epsilon_2^\mu$, the vector $\epsilon_1'^\mu$ obtained by parallel-transporting ϵ_1^μ in the direction ϵ_2^μ. In flat spacetime, the vectors $\epsilon_1, \epsilon_2, \epsilon_1'$, and ϵ_2' form an infinitesimal parallelogram since $x^\mu + \epsilon_1^\mu + \epsilon_2'^\mu = x^\mu + \epsilon_2^\mu + \epsilon_1'^\mu$. In a general affinely connected spacetime, the infinitesimal parallelogram does not close and

$$\left(x^\mu + \epsilon_1^\mu + \epsilon_2'^\mu\right) - \left(x^\mu + \epsilon_2^\mu + \epsilon_1'^\mu\right) = \epsilon_1^\rho \epsilon_2^\sigma T_{\rho\sigma}{}^\mu . \tag{1.24}$$

Finite parallel transport along a curve γ depends on the curve, not only on the initial and final points, so, if the curve is closed, the original and the parallel-transported vectors do not coincide. The difference is measured by the *(Riemann) curvature tensor* $R_{\mu\nu\rho}{}^\sigma$: let us consider two vectors ϵ_1^μ and ϵ_2^μ at a given point x^μ and let us parallel-transport the vector ω_μ from x^μ to $x^\mu + \epsilon_1^\mu$ and then to $x^\mu + \epsilon_1^\mu + \epsilon_2^\mu$. The result is

$$\omega_\mu + \left(\epsilon_1^\nu + \epsilon_2^\nu\right)\Gamma_{\nu\mu}{}^\rho \omega_\rho + \epsilon_1^\lambda \epsilon_2^\nu \left(\partial_\lambda \Gamma_{\nu\mu}{}^\rho + \Gamma_{\lambda\delta}{}^\rho \Gamma_{\nu\mu}{}^\delta\right)\omega_\rho + \mathcal{O}(\epsilon^3). \tag{1.25}$$

If we go to the same point along the route x^μ to $x^\mu + \epsilon_2^\mu$ and then to $x^\mu + \epsilon_1^\mu + \epsilon_2^\mu$ we obtain a different value and the difference between the parallel-transported vectors is

$$\Delta\omega_\mu = \epsilon_1^\lambda \epsilon_2^\nu R_{\lambda\nu\mu}{}^\rho \omega_\rho , \tag{1.26}$$

where

$$R_{\mu\nu\rho}{}^\sigma(\Gamma) = 2\partial_{[\mu}\Gamma_{\nu]\rho}{}^\sigma + 2\Gamma_{[\mu|\lambda}{}^\sigma \Gamma_{\nu]\rho}{}^\lambda . \tag{1.27}$$

[4] Here we use the mathematical concept of a curve: a map from the real line $\mathbb{R}$ (or an interval) given as a function of a real parameter $x^\mu(\xi)$, rather than the *image* of the real line in the spacetime. Thus, after a reparametrization $\xi'(\xi)$, we obtain a different curve, although the image is the same and physically we would say that we have the same curve.

We can also define the curvature tensor (and the torsion tensor) through the *Ricci identities* for a scalar ϕ, a vector ξ^{μ}, and a 1-form ω_{μ}:

$$\begin{aligned}\left[\nabla_{\mu}, \nabla_{\nu}\right]\phi &= T_{\mu\nu}{}^{\sigma}\nabla_{\sigma}\phi, \\ \left[\nabla_{\mu}, \nabla_{\nu}\right]\xi^{\rho} &= R_{\mu\nu\sigma}{}^{\rho}\xi^{\sigma} + T_{\mu\nu}{}^{\sigma}\nabla_{\sigma}\xi^{\rho}, \\ \left[\nabla_{\mu}, \nabla_{\nu}\right]\omega_{\rho} &= -\omega_{\sigma}R_{\mu\nu\rho}{}^{\sigma} + T_{\mu\nu}{}^{\sigma}\nabla_{\sigma}\omega_{\rho},\end{aligned} \tag{1.28}$$

or, for a general tensor,

$$\left[\nabla_{\alpha}, \nabla_{\beta}\right]\xi_{\mu_1\cdots}{}^{\nu_1\cdots} = -R_{\alpha\beta\mu_1}{}^{\gamma}\xi_{\gamma\cdots}{}^{\nu_1\cdots} - \cdots + R_{\alpha\beta\gamma}{}^{\nu_1}\xi_{\mu_1\cdots}{}^{\gamma\cdots} + \cdots + T_{\alpha\beta}{}^{\gamma}\nabla_{\gamma}\xi_{\mu_1\cdots}{}^{\nu_1\cdots}. \tag{1.29}$$

and, using the antisymmetry of the commutators of covariant derivatives and the fact that the covariant derivative satisfies the Jacobi identity, one can derive the following *Bianchi identities*:

$$\begin{aligned}R_{(\alpha\beta)\gamma}{}^{\delta} &= 0, \\ R_{[\alpha\beta\gamma]}{}^{\delta} + \nabla_{[\alpha}T_{\beta\gamma]}{}^{\delta} + T_{[\alpha\beta}{}^{\rho}T_{\gamma]\rho}{}^{\delta} &= 0, \\ \nabla_{[\alpha}R_{\beta\gamma]\rho}{}^{\sigma} + T_{[\alpha\beta}{}^{\delta}R_{\gamma]\delta\rho}{}^{\sigma} &= 0.\end{aligned} \tag{1.30}$$

(The last two identities are derived from the Jacobi identity of covariant derivatives acting on a scalar and a vector, respectively.)

In general, if we modify the affine connection by adding an arbitrary tensor[5] $\tau_{\mu\nu}{}^{\rho}$,

$$\Gamma_{\mu\nu}{}^{\rho} \to \tilde{\Gamma}_{\mu\nu}{}^{\rho} = \Gamma_{\mu\nu}{}^{\rho} + \tau_{\mu\nu}{}^{\rho}, \tag{1.31}$$

the curvature is modified as follows:

$$R_{\mu\nu\rho}{}^{\sigma}(\tilde{\Gamma}) = R_{\mu\nu\rho}{}^{\sigma}(\Gamma) - T_{\mu\nu}{}^{\lambda}\tau_{\lambda\rho}{}^{\sigma} + 2\nabla_{[\mu}\tau_{\nu]\rho}{}^{\sigma} + 2\tau_{[\mu|\lambda}{}^{\sigma}\tau_{|\nu]\rho}{}^{\lambda}. \tag{1.32}$$

The *Ricci tensor* is defined by

$$R_{\mu\nu} = R_{\mu\rho\nu}{}^{\rho} = \partial_{\mu}\Gamma_{\rho\nu}{}^{\rho} - \partial_{\rho}\Gamma_{\mu\nu}{}^{\rho} + \Gamma_{\mu\lambda}{}^{\rho}\Gamma_{\rho\nu}{}^{\lambda} - \Gamma_{\rho\lambda}{}^{\rho}\Gamma_{\mu\nu}{}^{\lambda}. \tag{1.33}$$

In general it is not symmetric, but, according to the second Bianchi identity,

$$R_{[\mu\nu]} = \tfrac{1}{2}\overset{*}{\nabla}_{\rho}\overset{*}{T}_{\mu\nu}{}^{\rho}, \tag{1.34}$$

where we have used the *modified divergence* $\overset{*}{\nabla}_{\mu}$ and the *modified torsion tensor* $\overset{*}{T}_{\mu\nu}{}^{\rho}$,

$$\overset{*}{\nabla}_{\mu} = \nabla_{\mu} - T_{\mu\rho}{}^{\rho}, \qquad \overset{*}{T}_{\mu\nu}{}^{\rho} = T_{\mu\nu}{}^{\rho} - 2T_{[\mu|\sigma}{}^{\sigma}\delta_{|\nu]}{}^{\rho}. \tag{1.35}$$

If we modify the connection as in Eq. (1.31), the Ricci tensor is also modified:

$$R_{\mu\rho}(\tilde{\Gamma}) = R_{\mu\rho} - T_{\mu\nu}{}^{\lambda}\tau_{\lambda\rho}{}^{\nu} + 2\nabla_{[\mu}\tau_{\nu]\rho}{}^{\nu} + 2\tau_{[\mu|\lambda}{}^{\nu}\tau_{|\nu]\rho}{}^{\lambda}. \tag{1.36}$$

Another useful formula is the Lie derivative of the torsion tensor which, using the first two Bianchi identities, can be rewritten in the form

$$\mathcal{L}_{\xi}T_{\mu\nu}{}^{\rho} = \nabla_{\mu}\left(\xi^{\lambda}T_{\lambda\nu}{}^{\rho}\right) + \nabla_{\nu}\left(\xi^{\lambda}T_{\mu\lambda}{}^{\rho}\right) - \nabla_{\lambda}\left(\epsilon^{\rho}T_{\mu\nu}{}^{\lambda}\right) - 3\epsilon^{\lambda}R_{[\lambda\mu\nu]}{}^{\rho} + \epsilon^{\rho}\nabla_{\sigma}T_{\mu\nu}{}^{\sigma}. \tag{1.37}$$

[5] Only if τ transforms as a tensor can $\tilde{\Gamma}$ transform as a connection.

1.3 Metric spaces

To go further we need to add structure to a manifold: a *metric* in tangent space, i.e. an inner product for tangent-space vectors (symmetric, bilinear) associating a function $g(\xi, \epsilon)$ with any pair of vectors (ξ, ϵ). This corresponds to a symmetric (0, 2)-type tensor g symmetric in its two covariant components $g_{\mu\nu} = g_{(\mu\nu)}$:

$$\xi \cdot \epsilon \equiv g(\xi, \epsilon) = \xi^\mu \epsilon^\nu g_{\mu\nu}. \tag{1.38}$$

The norm squared of a vector is just the product of the vector with itself, $\xi^2 = \xi \cdot \xi$. The metric will be required to be non-singular, i.e.

$$g \equiv \det(g_{\mu\nu}) \neq 0, \tag{1.39}$$

and locally diagonalizable into $\eta_{\mu\nu} = \mathrm{diag}(+ - \cdots -)$ for physical and conventional reasons. Thus, in d dimensions

$$\mathrm{sign}\, g = \frac{g}{|g|} = (-1)^{d-1}. \tag{1.40}$$

As usual, a metric can be used to establish a correspondence between a vector space and its dual, i.e. between vectors and 1-forms: with each vector ξ^μ we associate a 1-form ω_μ whose action on any other vector η^μ is the product of ξ and η, $\omega(\eta) = \xi^\mu \eta^\nu g_{\mu\nu}$, which means the relation between components $\omega_\nu = \xi^\mu g_{\mu\nu}$. It is customary to denote this 1-form by ξ_μ and the transformation from vector to 1-form is represented by lowering the index.

The inverse metric can be used as a metric in cotangent space and its components are those of the inverse matrix and are denoted with upper indices. The operation of raising indices can be similarly defined and the consistency of all these operations is guaranteed because the dual of the dual is the original vector space. The extension to tensors of higher ranks is straightforward.

The determinant of the metric can also be used to relate tensors and weight w tensor densities, since it transforms as a density of weight $w = 2$ and the product of a tensor and $g^{\frac{w}{2}}$ transforms as a density of weight w.

Furthermore, with a metric we can define the *Ricci scalar* R and the *Einstein tensor* $G_{\mu\nu}$,

$$R = R_\mu{}^\mu, \qquad G_{\mu\nu} = R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R, \tag{1.41}$$

which need not be symmetric (just like the Ricci tensor).

So far we have two independent fields defined on our manifold: the metric and the affine connection. An L_d spacetime equipped with a metric is sometimes denoted by (L_d, g). The affine connection and the metric are related by the *non-metricity tensor* $Q_{\mu\nu\rho}$,

$$Q_{\mu\nu\rho} \equiv -\nabla_\mu g_{\nu\rho}. \tag{1.42}$$

If we take the combination $\nabla_\mu g_{\rho\sigma} + \nabla_\rho g_{\sigma\mu} - \nabla_\sigma g_{\mu\rho}$ and expand it, we find that the connection can be written as follows:

$$\Gamma_{\mu\nu}{}^\rho = \left\{ {\rho \atop \mu\,\nu} \right\} + K_{\mu\nu}{}^\rho + L_{\mu\nu}{}^\rho, \tag{1.43}$$

where

$$\left\{ {\rho \atop \mu\,\nu} \right\} = \tfrac{1}{2} g^{\rho\sigma} \left\{ \partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu} \right\} \tag{1.44}$$

are the *Christoffel symbols*, which are completely determined by the metric, and K is called the *contorsion tensor* and is given in terms of the torsion tensor by

$$\begin{aligned} K_{\mu\nu}{}^{\rho} &= \tfrac{1}{2} g^{\rho\sigma} \left\{ T_{\mu\sigma\nu} + T_{\nu\sigma\mu} - T_{\mu\nu\sigma} \right\}, \\ K_{[\mu\nu]}{}^{\rho} &= -\tfrac{1}{2} T_{\mu\nu}{}^{\rho}, \qquad K_{\mu\nu\rho} = -K_{\mu\rho\nu}. \end{aligned} \tag{1.45}$$

Finally

$$L_{\mu\nu}{}^{\rho} = \tfrac{1}{2} \left\{ Q_{\mu\nu}{}^{\rho} + Q_{\nu\mu}{}^{\rho} - Q^{\rho}{}_{\mu\nu} \right\}. \tag{1.46}$$

Observe that the contorsion tensor depends on the metric whereas the torsion tensor does not. Furthermore, observe that, since the contorsion and non-metricity tensors transform as tensors, the piece responsible for the non-homogeneous term in the transformation of the affine connection is the Christoffel symbol.

With a metric it is also possible to define the length of a curve γ, $x^\mu(\xi)$, by the integral

$$s = \int_\gamma d\xi \, \sqrt{g_{\mu\nu}(x) \dot{x}^\mu \dot{x}^\nu}. \tag{1.47}$$

If we consider the above expression as a functional in the space of all curves joining two given points, we can ask which of those curves minimizes it. The answer is given by the Euler–Lagrange equations, which take the simple form

$$\ddot{x}^\mu + \dot{x}^\rho \dot{x}^\sigma \left\{ {\mu \atop \rho\,\sigma} \right\} = 0, \tag{1.48}$$

if we parametrize the curve by its proper length s. This is the *geodesic equation*, and is different from the autoparallel equation (1.23) whenever there is torsion and non-metricity.

In the standard theory of gravity metric and affine connection are not independent variables since we want to describe only the degrees of freedom corresponding to a massless spin-2 particle. To relate these two fields one imposes the *metric postulate*

$$Q_{\mu\rho\sigma} = -\nabla_\mu g_{\rho\sigma} = 0, \tag{1.49}$$

which makes the operations of raising and lowering of indices commute with the covariant derivative. A connection satisfying the above condition is said to be *metric-compatible* and a spacetime (L_d, g) with a metric-compatible connection is called a *Riemann–Cartan spacetime* and denoted by U_d.

Still, the metric postulate leaves the torsion undetermined. If we want to have a connection completely determined by the metric, left as the only independent field, one has to impose the vanishing of the torsion tensor. The torsionless, metric-compatible connection is called *Levi-Cività connection* and its components are given by the Christoffel symbols.[6] A Riemann–Cartan spacetime U_d with vanishing torsion is a *Riemann spacetime* V_d.

[6] Sometimes (specially in the supergravity context) the Levi-Cività connection is written $\Gamma(g)$ to stress the fact that it is a function of the metric in order to distinguish it from arbitrary connections that are independent of the metric. We will do so only when necessary.

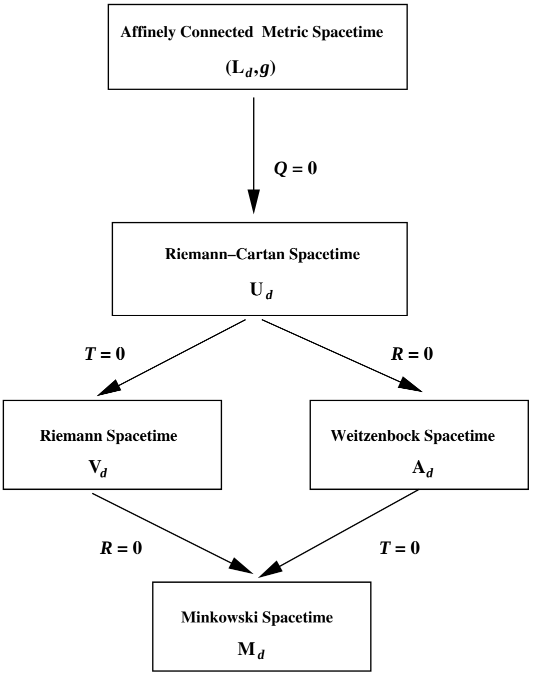

Fig. 1.1. Particular structures in an affinely connected spacetime equipped with a metric (L_d, g).

There is another way of reducing the number of independent fields: by imposing the vanishing of the curvature tensor. In this case, both the metric and the connection are completely determined by the *Vielbein* (to be defined latter). The connection is called *Weitzenböck connection* [944, 945] and has torsion (also determined by the Vielbein). A Riemann–Cartan spacetime with Weitzenböck connection is a *Weitzenböck spacetime* A_d.

If both torsion and curvature vanish, the space has to be *Minkowski spacetime* M_d since the *Minkowski metric* $g_{\mu\nu} = \eta_{\mu\nu}$ is the only one that makes the full Riemann tensor vanish in the absence of torsion.

The diagram in Figure 1.1 summarizes the different particular structures that we can have on an affinely connected manifold equipped with a metric [522, 523].

In the rest of this section we are going to study the particular properties of some of these spacetimes. The Weitzenböck spacetime will be studied after the introduction of Vielbeins in Section 1.4.

1.3.1 *Riemann–Cartan spacetime* U_d

As has been said, this is an affinely connected metric spacetime with a metric-compatible connection, so the non-metricity tensor vanishes, $Q_{\mu\nu\rho} = 0$. According to the general result, a metric-compatible connection of a Riemann–Cartan spacetime always has the form

$$\Gamma_{\mu\nu}{}^{\rho} = \left\{ {\rho \atop \mu\,\nu} \right\} + K_{\mu\nu}{}^{\rho}. \tag{1.50}$$

Observe that the symmetric part of the contorsion tension does not vanish, but

$$K_{(\mu\nu)\,\rho} = \tfrac{1}{2}\left(T_{\mu\rho\nu} + T_{\nu\rho\mu}\right) \neq 0. \tag{1.51}$$

This means that the presence of torsion implies not only that the connection has a non-vanishing antisymmetric part, but also that the symmetric part is not fully determined by the metric but

$$\Gamma_{(\mu\nu)}{}^{\rho} = \left\{ {\rho \atop \mu\,\nu} \right\} + K_{(\mu\nu)}{}^{\rho} \neq \left\{ {\rho \atop \mu\,\nu} \right\}. \tag{1.52}$$

The curvature, Ricci, and Einstein tensors of a metric-compatible connection satisfy further identities. On contracting the γ and σ indices in the third Bianchi identity Eqs. (1.30) and using the metric postulate, we find the so-called *contracted Bianchi identity*

$$\nabla_\alpha G_\mu{}^\alpha + 2T_{\mu\alpha\beta}R^{\beta\alpha} - T_{\alpha\beta\gamma}R_\mu{}^{\gamma\alpha\beta} = 0. \tag{1.53}$$

Furthermore, by applying the Ricci identity to the metric and using the metric postulate, one can prove a fourth Bianchi identity:

$$R_{\alpha\beta(\gamma\delta)} = 0. \tag{1.54}$$

If we modify the connection according to Eq. (1.31) and Γ is metric-compatible, the Ricci scalar is

$$R(\tilde{\Gamma}) = R(\Gamma) - T_{\mu\nu}{}^{\rho}\tau_\rho{}^{\mu\nu} + 2\nabla_\mu \tau_\nu{}^{\mu\nu} + \tau_\mu{}^{\mu\lambda}\tau_\nu{}^\nu{}_\lambda + \tau_\nu{}^{\mu\rho}\tau_{\mu\rho}{}^\nu. \tag{1.55}$$

If $\tilde{\Gamma}$ is not a metric-compatible connection, then τ contains all the contributions of the non-metricity tensor and the above formula allows us to work in the framework of a Riemann–Cartan spacetime with non-metric-compatible connections.

If both $\tilde{\Gamma}$ and Γ are metric-compatible connections and $\tilde{\Gamma}$ has torsion but $\Gamma = \Gamma(g)$, then $\tau = \tilde{K}$, the contorsion tensor of $\tilde{\Gamma}$, and the above formula takes a simpler form:

$$R(\tilde{\Gamma}) = R[\Gamma(g)] + 2\nabla_\mu \tilde{K}_\nu{}^{\mu\nu} + (\tilde{K}_\mu{}^{\mu\lambda})^2 + \tilde{K}_\nu{}^{\mu\rho}\tilde{K}_{\mu\rho}{}^\nu. \tag{1.56}$$

Now, this formula allows us to work with torsion in a Riemann spacetime. Particularly interesting is the case in which the contorsion $\tilde{K}_{\mu\nu\rho}$ is a completely antisymmetric tensor (proportional to the Kalb–Ramond field strength $H_{\mu\nu\rho}$, for instance). Then we have, if

$$\tilde{K}_{\mu\nu\rho} = \frac{1}{\sqrt{12}} H_{\mu\nu\rho}, \tag{1.57}$$

$$\int d^d x \sqrt{|g|}\, R(\tilde{\Gamma}) = \int d^d x \sqrt{|g|} \left\{ R[\Gamma(g)] + \frac{1}{2\cdot 3!} H_{\mu\nu\rho}H^{\mu\nu\rho} \right\}. \tag{1.58}$$

1.3.2 Riemann spacetime V_d

It is defined by the conditions $Q = T = 0$ which determine the connection to be the Levi-Cività connection $\Gamma(g)$ whose components in a coordinate basis are given by the Christoffel symbols. In a Riemann spacetime one can construct infinitesimal parallelograms and autoparallel curves are also geodesics (as in flat spacetime). There are also additional interesting properties. To start with, we can write the transformation of tensors under infinitesimal GCTs (Lie derivatives) in terms of covariant derivatives alone (all torsion terms vanish). In particular, for the metric and r-forms we can write

$$\begin{aligned} \delta_\xi g_{\mu\nu} &= -2\nabla_{(\mu}\xi_{\nu)}, \\ \delta_\xi B_{\mu_1\cdots\mu_r} &= -\xi^\lambda \nabla_\lambda B_{\mu_1\cdots\mu_r} - r(\nabla_{[\mu_1|}\xi^\lambda) B_{\lambda|\mu_2\cdots\mu_r]}. \end{aligned} \tag{1.59}$$

Furthermore, we have the usual identity

$$\Gamma_{\rho\mu}{}^{\rho} = \partial_\mu \ln\left(\sqrt{|g|}\right), \tag{1.60}$$

which allows us to write the Laplacian of a scalar function f in this way:

$$\nabla^2 f = \frac{1}{\sqrt{|g|}}\partial_\mu\left(\sqrt{|g|}\,\partial^\mu f\right), \tag{1.61}$$

and the divergence of a completely antisymmetric tensor (k-form) in this way:[7]

$$\nabla_{\mu_1} F^{\mu_1\mu_2\cdots\mu_k} = \frac{1}{\sqrt{|g|}}\partial_{\mu_1}\left(\sqrt{|g|}\,F^{\mu_1\mu_2\cdots\mu_k}\right). \tag{1.62}$$

The Bianchi identities take the form

$$R_{(\alpha\beta)\gamma}{}^{\delta} = 0, \qquad R_{[\alpha\beta\gamma]}{}^{\delta} = 0, \qquad \nabla_{[\alpha} R_{\beta\gamma]\rho}{}^{\sigma} = 0, \qquad R_{\alpha\beta(\gamma\delta)} = 0. \tag{1.63}$$

The first and fourth identities imply together

$$R_{\alpha\beta\gamma\delta} = R_{\gamma\delta\alpha\beta}, \tag{1.64}$$

which in turn implies that the Ricci and Einstein tensors are symmetric. The contracted Bianchi identity says now that the Einstein tensor is divergence-free:

$$\nabla_\mu G^{\mu\nu} = 0, \tag{1.65}$$

which is a crucial identity in the development of general relativity.

The number of independent components of the curvature in d dimensions after taking into account all these Bianchi identities is $(1/12)\,d^2(d^2-1)$.

The four-dimensional curvature tensor can be split into different pieces which transform irreducibly under the Lorentz group: a scalar piece $D(0, 0)$, which is nothing but the Ricci

[7] Observe that this implies that the second term on the r.h.s. of Eq. (1.56) times $\sqrt{|g|}$ is a total derivative.

scalar R, a two-index symmetric, traceless piece $R_{\mu\nu} - \frac{1}{4} g_{\mu\nu} R$ (corresponding to the representation $D(1,1)$), and a four-index tensor with the same symmetries as the Riemann tensor but traceless: the Weyl tensor $C_{\mu\nu\rho}{}^{\sigma}$ with $C_{\mu\sigma\rho}{}^{\sigma} = 0$:

$$R_{\mu\nu}{}^{\rho\sigma} = C_{\mu\nu}{}^{\rho\sigma} + 2\big(R_{[\mu}{}^{[\rho} - \tfrac{1}{4} R g_{[\mu}{}^{[\rho}\big) g_{\nu]}{}^{\sigma]} + \tfrac{1}{6} R g_{[\mu}{}^{\rho} g_{\nu]}{}^{\sigma}. \tag{1.66}$$

The Weyl tensor can be decomposed into its self-dual and anti-self-dual parts (with respect to the last two indices). These two complex tensors transform in the $D(2,0)$ and $D(0,2)$ representations, respectively.

In d dimensions the Weyl tensor is defined by

$$C_{\mu\nu}{}^{\rho\sigma} = R_{\mu\nu}{}^{\rho\sigma} - \frac{4}{d-2} R_{[\mu}{}^{[\rho} g_{\nu]}{}^{\sigma]} + \frac{2}{(d-1)(d-2)} R\, g_{[\mu}{}^{[\rho} g_{\nu]}{}^{\sigma]}. \tag{1.67}$$

The main property of the Weyl tensor $C_{\mu\nu\rho}{}^{\sigma}$ with the indices in these positions is that it is left invariant by Weyl rescalings of the metric (see Appendix E). Furthermore, just as the Riemann curvature vanishes only for Minkowski spacetime, the Weyl tensor vanishes only for conformally flat (Minkowski) spacetimes, i.e. spacetimes that are related to Minkowski's by a given conformal transformation.

A final property of the Levi-Cività connection that is worth mentioning is the form of its variation under an arbitrary variation of the metric:

$$\delta\Gamma_{\mu\nu}{}^{\rho}(g) = \tfrac{1}{2} g^{\rho\sigma} \left\{ \nabla_{\mu}\delta g_{\nu\sigma} + \nabla_{\nu}\delta g_{\mu\sigma} - \nabla_{\sigma}\delta g_{\mu\nu} \right\}. \tag{1.68}$$

Since $\delta g_{\mu\nu}$ is a tensor, $\delta\Gamma$ is a tensor even though Γ is not.

1.4 Tangent space

So far we have considered, for a given coordinate system, only one basis in tangent space: the coordinate basis. We are now going to consider an arbitrary basis in tangent space. Such a basis is defined by a set of d contravariant vectors labeled by a *tangent-space index a* $\{e_a = e_a{}^{\mu}\partial_{\mu}\}$ and is also referred to as a *frame* or, generically, *Vielbein* basis.[8] The coordinate basis is now a particular case in which $e_a{}^{\mu} = \delta_a{}^{\mu}$. Now we can express any vector in this basis $\xi = \xi^a e_a$ and its components ξ^a will be related to the coordinate basis components by

$$\xi^{\mu} = \xi^a e_a{}^{\mu}. \tag{1.69}$$

We can immediately define the dual basis of 1-forms $\{e_a = e^a{}_{\mu} dx^{\mu}\}$ defined by

$$\langle e^a | e_b \rangle = \delta^a{}_b, \tag{1.70}$$

which implies that the matrix of components $e^a{}_{\mu}$ of the 1-forms in the coordinate basis is the inverse, transposed, of that of the vectors:

$$e^a{}_{\mu} e_b{}^{\mu} = \delta^a{}_b, \quad \Rightarrow e_a{}^{\mu} e^a{}_{\nu} = \delta^{\mu}{}_{\nu}. \tag{1.71}$$

[8] *Einbein* for $d=1$, *Zweibein* for $d=2$, *Dreibein* for $d=3$, *Vierbein* for $d=4$, etc. In four dimensions it is also called a *tetrad*.

We can now relate frame and world indices of any tensor using these two matrices. In particular, we can use the frame components g_{ab} of the metric,

$$g_{ab} = e_a{}^{\mu} e_b{}^{\nu} g_{\mu\nu}, \tag{1.72}$$

that can also be interpreted as the matrix of inner products of the Vielbein basis $g(e_a, e_b) = g_{ab}$. An orthonormal Vielbein basis leads to $g_{ab} = \eta_{ab}$. Frames are usually chosen in such a way as to obtain a particular g_{ab} and orthonormal frames will be particularly important in what follows.

It is easy to see that g_{ab} and its inverse g^{ab} can be consistently used to raise and lower frame indices. In particular,

$$e^a{}_{\mu} = g_{\mu\nu} e_b{}^{\nu} g^{ab}, \qquad g_{\mu\nu} = e^a{}_{\mu} e^b{}_{\nu} g_{ab}. \tag{1.73}$$

A frame is invariant under GCTs (only the components in the coordinate basis change) and, thus, frame components of any tensor are also invariant. However, we can make a change of basis. Any two Vielbein bases are related by a $\mathrm{GL}(d, \mathbb{R})$ transformation $\Lambda^a{}_b$ in tangent space at a given point of the manifold. This transformation can in fact be different at each point and thus we have to consider local frame transformations $\Lambda^a{}_b(x)$. We write their action on vectors and forms as follows:

$$e'_a = e_b \left(\Lambda^{-1}\right)^b{}_a, \qquad e'^a = \Lambda^a{}_b e^b. \tag{1.74}$$

The *Ricci rotation coefficients* (or *anholonomy coefficients*) $\Omega_{ab}{}^c$ are the Lie brackets

$$[e_a, e_b] = -2\Omega_{ab}{}^c e_c, \qquad \Omega_{ab}{}^c = e_a{}^{\mu} e_b{}^{\nu} \partial_{[\mu} e^c{}_{\nu]}. \tag{1.75}$$

A *non-holonomic* frame is one with non-vanishing Ωs. Observe that, given a basis of vectors $\{e_a\}$, we could try to find a new set of coordinates $y^a(x^{\mu})$ such that

$$e_a y^b = e_a{}^{\mu} \partial_{\mu} y^b = \delta_a{}^b. \tag{1.76}$$

The integrability condition for the system of partial differential equations $[e_c, e_a] y^b = 0$ is precisely the vanishing of the anholonomy coefficients $\Omega_{ab}{}^c$. A non-holonomic basis of vectors $\{e_a\}$ is one for which these coefficients vanish and then we can trivialize them ($e_a{}^{\mu} = \delta_a{}^{\mu}$) by a change of coordinates.

Just as we defined a covariant derivative transforming world tensors into world tensors we are now going to define a derivative that transforms tangent-space tensors into tangent-space tensors transforming well under local $\mathrm{GL}(d, \mathbb{R})$ transformations associated with a connection ω. Its action on scalars, vectors, and forms is[9]

$$\begin{aligned} \mathcal{D}_a \phi &= \partial_a \phi, & &(= e_a{}^{\mu} \partial_{\mu} \phi), \\ \mathcal{D}_a \xi^b &= \partial_a \xi^b + \omega_{ac}{}^b \xi^c, & &(= e_a{}^{\mu} \mathcal{D}_{\mu} \xi^b), \\ \mathcal{D}_a \varepsilon_b &= \partial_a \varepsilon_b - \varepsilon_c\, \omega_{ab}{}^c, & &(= e_a{}^{\mu} \mathcal{D}_{\mu} \varepsilon_b). \end{aligned} \tag{1.77}$$

[9] Of course, the formalism we are developing is just that of a $\mathrm{GL}(d, \mathbb{R})$ gauge theory and our notation is basically identical to that of Appendix A. Here we are dealing with vector representations of $\mathrm{GL}(d, \mathbb{R})$. The d^2 generators of its Lie algebra can be labeled by a pair of vector indices ab and they are given, for instance, by $(T_{ab})^c{}_d = -\eta_{ad}\eta_b{}^c$. Thus $A_{\mu}{}^I \Gamma_{\mathrm{v}}(T_I) = \omega_{\mu}{}^{ab} (T_{ab})^c{}_d = -\omega_{\mu\, d}{}^c$. The subgroup $\mathrm{SO}(1, d-1, \mathbb{R})$ will be treated in more detail.

Local $GL(d, \mathbb{R})$ covariance implies the inhomogeneous transformation law for the connection:

$$\omega'_{ab}{}^{c} = \Big[\Lambda^{c}{}_{d}\omega_{ef}{}^{d}\big(\Lambda^{-1}\big)^{f}{}_{b} - \big(\Lambda^{-1}\big)^{c}{}_{d}\partial_{e}\Lambda^{d}{}_{b}\Big](\Lambda^{-1})^{e}{}_{a}. \tag{1.78}$$

The curvature of this connection can be defined through the Ricci identities in the standard fashion (observe that there are no torsion terms here):

$$\begin{aligned}\big[\mathcal{D}_{\mu}, \mathcal{D}_{\nu}\big]\phi &= 0,\\ \big[\mathcal{D}_{\mu}, \mathcal{D}_{\nu}\big]\xi^{a} &= R_{\mu\nu b}{}^{a}\xi^{b},\\ \big[\mathcal{D}_{\mu}, \mathcal{D}_{\nu}\big]\varepsilon_{a} &= -\varepsilon_{b}\, R_{\mu\nu a}{}^{b},\end{aligned} \tag{1.79}$$

and then the curvature is given by[10]

$$R_{\mu\nu a}{}^{b} = 2\partial_{[\mu}\, \omega_{\nu]a}{}^{b} - 2\omega_{[\mu|a}{}^{c}\, \omega_{|\nu]c}{}^{b}. \tag{1.81}$$

At this point we have introduced a new connection ω that is independent of the metric. In the previous section we managed to relate the connection Γ to the metric via the metric postulate. Here we are going to generalize the metric postulate first to relate the two connections (the first Vielbein postulate) and then to relate them to the metric (the second Vielbein postulate). Before we enunciate these postulates we introduce the *total covariant derivative*, covariant with respect to all the indices of the object it acts on. We denote it by ∇ again, and, for instance, acting on Vielbeins it is

$$\nabla_{\mu}e_{a}{}^{\nu} = \partial_{\mu}e_{a}{}^{\nu} + \Gamma_{\mu\rho}{}^{\nu}e_{a}{}^{\rho} - e_{b}{}^{\nu}\omega_{\mu a}{}^{b}. \tag{1.82}$$

We can motivate the first Vielbein postulate as follows: we would like to be able to convert tangent into world indices and vice-versa inside the total covariant derivative, so $e^{a}{}_{\nu}\nabla_{\mu}\xi^{\nu} = \mathcal{D}_{\mu}\xi^{a}$ and $\mathcal{D}$ is just the projection of ∇ onto the Vielbein basis. To have this property we impose the *first Vielbein postulate*,

$$\nabla_{\mu}e_{a}{}^{\nu} = 0. \tag{1.83}$$

It is worth stressing that this does not imply the covariant constancy of the metric $\nabla_{\mu}g_{\nu\rho} = 0$. The above postulate implies the following relation between the connections:

$$\omega_{\mu a}{}^{b} = \Gamma_{\mu a}{}^{b} - e_{a}{}^{\nu}\partial^{b}_{\mu}e_{\nu}. \tag{1.84}$$

Furthermore, the curvatures of the two connections are now related by

$$R_{\mu\nu\rho}{}^{\sigma}(\Gamma) = e^{a}{}_{\rho}e_{b}{}^{\sigma}\, R_{\mu\nu a}{}^{b}(\omega). \tag{1.85}$$

The first Vielbein postulate also gives an important relation between the torsion and the Vielbein: on taking the antisymmetric part of $\nabla_{\mu}e^{a}{}_{\nu} = 0$, we obtain

$$2\mathcal{D}_{[\mu}e^{a}{}_{\nu]} = 2\big(\partial_{[\mu}e^{a}{}_{\nu]} - \omega_{[\mu}{}^{a}{}_{\nu]}\big) = -T_{\mu\nu}{}^{a}. \tag{1.86}$$

[10] Observe that, with all Latin indices, $R_{abc}{}^{d} = e_{a}{}^{\mu}e_{b}{}^{\nu}R_{\mu\nu c}{}^{d}$ and, therefore,

$$R_{abc}{}^{d} = 2\partial_{[a}\, \omega_{b]c}{}^{d} - 2\omega_{[a|c}{}^{e}\, \omega_{|b]e}{}^{d} + 2\Omega_{ab}{}^{e}\omega_{ec}{}^{d}. \tag{1.80}$$

The significance of the torsion in this formalism, from the point of view of the gauge theory of $GL(d, \mathbb{R})$, is unclear. We can provide an interpretation in the framework of the gauge theory of the affine group $IGL(d, \mathbb{R})$ but we will do it in the more restricted context of the Lorentz and Poincaré groups in Section 4.5.

The first Vielbein postulate has allowed us to recover the structure of affinely connected spacetime (L_d, g) with only one (independent) connection, generalized to allow the use of an arbitrary basis in tangent space. Furthermore, we can recover the different particular structures that we defined in the previous section, also generalized to allow the use of arbitrary basis in tangent space. First, if Γ is a completely general connection, it is given by Eq. (1.43) and then ω (which is related to Γ by the first Vielbein postulate) is given by

$$\omega_{ab}{}^{c} = \omega_{ab}{}^{c}(e) + K_{ab}{}^{c} + L_{ab}{}^{c}, \tag{1.87}$$

where $\omega(e)$ is the (*Cartan* or even *Levi-Cività*) connection related to the Levi-Cività connection $\Gamma(g)$ by Eq. (1.84). It is completely determined by the Vielbeins:

$$\omega_{ab}{}^{c}(e) = \left\{ {c \atop a\,b} \right\} + \left\{ -\Omega_{ab}{}^{c} + \Omega_{b}{}^{c}{}_{a} - \Omega^{c}{}_{ab} \right\}, \tag{1.88}$$

where

$$\left\{ {c \atop a\,b} \right\} = \tfrac{1}{2} g^{cd}\{\partial_a g_{bd} + \partial_b g_{ad} - \partial_d g_{ab}\}. \tag{1.89}$$

$K_{ab}{}^{c}$ is nothing but the contorsion tensor expressed in a tangent-space basis, i.e. $K_{ab}{}^{c} = e_a{}^{\mu} e_b{}^{\nu} e^{c}{}_{\rho} K_{\mu\nu}{}^{\rho}$ and, similarly, $L_{ab}{}^{c} = e_a{}^{\mu} e_b{}^{\nu} e^{c}{}_{\rho} L_{\mu\nu}{}^{\rho}$.

Observe that

$$\omega_{a(bc)} = \tfrac{1}{2}(Q_{abc} + \partial_a g_{bc}). \tag{1.90}$$

We can impose the metric-compatibility condition Eq. (1.49), which in this context is known as the *second Vielbein postulate*, and we have a Riemann–Cartan spacetime U_d. The result is that Γ is again given by Eqs. (1.50), (1.44), and (1.45) and ω (which is related to Γ by the first Vielbein postulate) is given by

$$\omega_{ab}{}^{c} = \omega_{ab}{}^{c}(e) + K_{ab}{}^{c}. \tag{1.91}$$

If we now impose the vanishing of torsion, we obtain the Levi-Cività and Cartan connections $\Gamma(g)$ and $\omega(e)$ and we recover a Riemann spacetime V_d.

The two most important cases to which we can apply this general formalism are the following.

1. The case in which we use a coordinate basis $e_a{}^{\mu} = \delta_a{}^{\mu}$, so $g_{ab} = g_{\mu\nu}$, $\Omega = 0$, and the connections Γ and ω are identical.
2. The case in which we use an orthonormal basis $g_{ab} = \eta_{ab}$ in which

$$\left\{ {c \atop a\,b} \right\} = 0$$

and

$$\omega_{abc} = \omega_{abc}(e) + K_{abc} + L_{abc}, \qquad \omega_{abc}(e) = -\Omega_{abc} + \Omega_{bca} - \Omega_{cab}. \tag{1.92}$$

In the second case we would like to restrict ourselves to those changes of frame that preserve the form of the metric in tangent-space indices (here usually referred to as *flat indices* because they are raised and lowered with the flat space metric). By definition, these are transformations of the d-dimensional Lorentz group $\mathrm{SO}(1, d-1)$ whose gauge theory we are now led to consider. This gauge theory is developed in Section 2.3 of Appendix A and the spinorial representations of the Lorentz group are studied in Appendix B. We are simply going to rewrite here the main formulae we have obtained, adapted to the Lorentz subgroup of $\mathrm{GL}(d, \mathbb{R})$.

The main justification for making this step is that the Lorentz group admits spinorial representations, which are necessary in order to describe fermions, whereas the diffeomorphism group of a manifold does not. This is the only known method by which to describe spinors in curved spacetime in arbitrary coordinates and, thus, the only method known to couple fermions to gravity. This formalism was pioneered by Weyl [954].

First of all, the generators M_I of the Lorentz subgroup of $\mathrm{GL}(d, \mathbb{R})$ are just the antisymmetric combinations of those of $\mathrm{GL}(d, \mathbb{R})$ and can be labeled by two antisymmetric vector indices, i.e. M_{ab}. In this notation every generator appears twice and factors of $\frac{1}{2}$ have to be included in the right places. However, in general, the connection $\omega_\mu{}^{ab}$ is not antisymmetric in the "gauge" indices ab unless it is also metric-compatible ($\mathcal{D}\eta_{ab} = 0$), according to Eq. (1.90). We are going to consider only metric-compatible connections that are fully antisymmetric in the gauge indices and we will call them *spin connections*.[11]

Using the explicit form of the infinitesimal Lorentz generators in the vector representation $\Gamma_{\mathrm{v}}(M_{bc})^a{}_d$ given in Eq. (A.60) and in the spinorial representation $\Gamma_{\mathrm{s}}(M_{ab})^\alpha{}_\beta$ (we use temporarily the first few Greek letters $\alpha, \beta, \ldots$ as spinorial indices) given in Eq. (B.3), we find the following expressions for the (total) covariant derivatives of contravariant and covariant vectors and spinors:

$$
\begin{aligned}
\nabla_\mu \xi^a &= \partial_\mu \xi^a - \tfrac{1}{2}\omega_\mu{}^{bc}\Gamma_{\mathrm{v}}(M_{bc})^a_d \xi^d = \partial_\mu \xi^a + \omega_{\mu b}{}^a \xi^b, \\
\nabla_\mu \varepsilon_a &= \partial_\mu \varepsilon_a + \varepsilon_d \tfrac{1}{2}\omega_\mu{}^{bc}\Gamma_{\mathrm{v}}(M_{bc})^d_a = \partial_\mu \varepsilon_a - \varepsilon_b\, \omega_{\mu a}{}^b, \\
\nabla_\mu \psi^\alpha &= \partial_\mu \psi^\alpha - \tfrac{1}{2}\omega_\mu{}^{ab}\Gamma_{\mathrm{s}}(M_{ab})^\alpha{}_\beta \psi^\beta = \partial_\mu \psi^\alpha - \tfrac{1}{4}\omega_\mu{}^{ab}(\Gamma_{ab})^\alpha{}_\beta \psi^\beta, \\
\nabla_\mu \varphi_\alpha &= \partial_\mu \varphi_\alpha + \varphi_\beta \tfrac{1}{2}\omega_\mu{}^{ab}\Gamma_{\mathrm{s}}(M_{ab})^\beta{}_\alpha = \partial_\mu \varphi_\alpha + \varphi_\beta \tfrac{1}{4}\omega_\mu{}^{ab}(\Gamma_{ab})^\beta{}_\alpha.
\end{aligned}
\tag{1.93}
$$

These definitions, once we impose the Vielbein postulates, are consistent with the raising and lowering of vector indices with the Minkowski metric and with Dirac conjugation of the spinors. With the postulates, the spin connection is given by Eqs. (1.92).

The Ricci identities can now be written for the total covariant derivative in this form:

$$
\begin{aligned}
\left[\nabla_\mu, \nabla_\nu\right]\phi &= T_{\mu\nu}{}^\rho \nabla_\rho \phi, \\
\left[\nabla_\mu, \nabla_\nu\right]\xi^a &= R_{\mu\nu b}{}^a(\omega)\xi^b + T_{\mu\nu}{}^\rho \nabla_\rho \xi^a, \\
\left[\nabla_\mu, \nabla_\nu\right]\varepsilon_a &= -\varepsilon_b R_{\mu\nu a}{}^b(\omega) + T_{\mu\nu}{}^\rho \nabla_\rho \varepsilon_a, \\
\left[\nabla_\mu, \nabla_\nu\right]\psi &= -\tfrac{1}{4}R_{\mu\nu}{}^{ab}(\omega)\Gamma_{ab}\psi + T_{\mu\nu}{}^\rho \nabla_\rho \psi, \\
\left[\nabla_\mu, \nabla_\nu\right]\varphi &= +\tfrac{1}{4}\varphi R_{\mu\nu}{}^{ab}(\omega)\Gamma_{ab} + T_{\mu\nu}{}^\rho \nabla_\rho \varphi.
\end{aligned}
\tag{1.94}
$$

[11] If we wanted to have a non-metric-compatible spin connection, we would have to modify the first Vielbein postulate.

For more general tensors one has to add a curvature (ω) term for each flat index and a curvature (Γ) term for each world index. The curvatures have the same form as in Eqs. (1.27) and (1.81) but now $R_{\mu\nu}{}^{ab}$ is antisymmetric in ab.

The following expression is sometimes used:

$$R_{ab} = -\partial_a \omega_c{}^c{}_b - \partial_c \omega_{ab}{}^c + \omega_{cda}\omega^{dc}{}_b + \omega_{abd}\omega_c{}^{cd}. \tag{1.95}$$

The Vielbein formalism allows us to study the Weitzenböck spacetime defined on page 11.

1.4.1 Weitzenböck spacetime A_d

This spacetime is defined by a metric-compatible connection that we denote by $W_{\mu\nu}{}^\rho$ and call the *Weitzenböck connection* [944, 945] whose Riemann curvature is identically zero, $R_{\mu\nu\rho}{}^\sigma(W) = 0$. Trying to solve this equation directly for $W \neq 0$ is a very difficult task. However, we can use the Vielbein formalism to find a solution. Let us denote by $W_{\mathrm{s}\,\mu}{}^{ab}$ the tangent-space connection associated with W via the first Vielbein postulate

$$\nabla_\mu e^a{}_\nu = \partial_\mu e^a{}_\nu - W_{\mu\nu}{}^a + W_{\mathrm{s}\,\mu\nu}{}^a = 0. \tag{1.96}$$

The curvature of W_s is obviously zero on account of Eq. (1.85). Now, however, we can use the trivial solution to the equation $R_{\mu\nu}{}^{ab}(W_\mathrm{s}) = 0$, namely $W_\mathrm{s} = 0$, because, according to the above relation, $W_\mathrm{s} = 0$ does not imply $W = 0$ but

$$W_{\mu\nu}{}^\rho = e_a{}^\rho \partial_\mu e^a{}_\nu. \tag{1.97}$$

This is the Weitzenböck connection whose curvature vanishes identically. It cannot be rewritten in terms of the metric: it is necessary to use the Vielbein formalism. Observe that, using this connection, we can write the relation between any two connections Γ and ω satisfying the first Vielbein postulate in the form

$$\Gamma_{\mu\nu}{}^\rho = W_{\mu\nu}{}^\rho + \omega_{\mu\nu}{}^\rho. \tag{1.98}$$

$\omega_{\mu\nu}{}^\rho$ is a tensor, but $\Gamma_{\mu\nu}{}^\rho$ is not (it is an affine connection), and responsible for this is the Weitzenböck connection $W_{\mu\nu}{}^\rho$. We can also write

$$\omega_\mu{}^{ab} = \Gamma_\mu{}^{ab} - W_\mu{}^{ab}, \qquad W_\mu{}^{ab} = e^{a\nu}\partial_\mu e^b{}_\nu. \tag{1.99}$$

Now $\Gamma_\mu{}^{ab}$ is a $\mathrm{GL}(d,\mathbb{R})$ tensor in the upper two indices whereas $\omega_\mu{}^{ab}$ is not (because it is a $\mathrm{GL}(d,\mathbb{R})$ connection). Again, the Weitzenböck connection $W_\mu{}^{ab}$ is responsible for this.

Even though we have to search explicitly for a metric-compatible connection to find W, it is easy to check that it is indeed metric-compatible. Then, it can be decomposed into the sum of the Levi-Cività connection and the contorsion tensor. The torsion tensor is

$$T_{\mu\nu}{}^\rho = -2\Omega_{\mu\nu}{}^\rho, \tag{1.100}$$

and, therefore, the contorsion tensor is given by

$$K_{\mu\nu\rho}(W) = \Omega_{\mu\nu\rho} - \Omega_{\nu\rho\mu} + \Omega_{\rho\mu\nu} = -\omega_{\mu\nu\rho}(e), \tag{1.101}$$

where $\omega(e)$ is, as usual, the Cartan connection (which is associated via the first Vielbein postulate with the Levi-Cività connection $\Gamma(g)$).

Now, if we use Eqs. (1.31) and (1.32) for $\tilde{\Gamma} = W$, $\Gamma = \Gamma(g)$, and $\tau = K$, we find an expression for the Riemann curvature tensor of the Levi-Cività connection in terms of the contorsion tensor of the Weitzenböck connection:

$$R_{\mu\nu\rho}{}^{\sigma}[\Gamma(g)] = -2\nabla_{[\mu}K_{\nu]\rho}{}^{\sigma} - 2K_{[\mu|\lambda}{}^{\sigma}K_{|\nu]\rho}{}^{\lambda}. \tag{1.102}$$

On contracting indices, and eliminating a total derivative, we find that

$$\int d^dx\,\sqrt{|g|}\,R(g) = -\int d^dx\,\sqrt{|g|}\,\left\{K_{\mu}{}^{\mu\lambda}K^{\mu}{}_{\mu\lambda} + K_{\nu}{}^{\mu\rho}K_{\mu\rho}{}^{\nu}\right\}, \tag{1.103}$$

which can be expressed entirely in terms of the anholonomy coefficients $\Omega_{\mu\nu\rho}$, providing an alternative form of the Einstein–Hilbert action. This is, in fact, an alternative way of deriving the Palatini identity Eq. (D.4).

It is worth stressing here that the building blocks of the Riemann curvature tensor of the Levi-Cività connection in the above expression (the anholonomy coefficients/torsion and contorsion) are tensors, whereas in the standard expression for the curvature the building blocks are the Christoffel symbols, which are not tensors.

The main property of the Weitzenböck spacetime (the vanishing of the curvature) implies that parallel transport is path-independent and it is possible to define parallelism of vectors *at different spacetime points*:[12] two vectors $v^{\mu}(x_1)$, $w^{\mu}(x_2)$ are parallel if their components in the Vielbein basis $\{e_a{}^{\mu}\}$ are proportional. This is a consistent definition because the components in the Vielbein basis are invariant under the parallel transport defined by the W connection associated with that Vielbein basis. Indeed, the vector $v^{\mu}(x)$, parallel-transported to $x^{\mu} + \epsilon^{\mu}$ is

$$v^{\mu}(x+\epsilon) = v^{\mu}(x) - \epsilon^{\nu}v^{\rho}(x)W_{\nu\rho}{}^{\mu}(x), \tag{1.104}$$

and its tangent-space components can be found with the inverse Vielbein basis at $x^{\mu} + \epsilon^{\mu}$:

$$\begin{aligned} e^{a}{}_{\mu}(x+\epsilon)v^{\mu}(x+\epsilon) &= [e^{a}{}_{\mu}(x) + \epsilon^{\nu}\partial_{\nu}e^{a}{}_{\mu}(x)][v^{\mu}(x) - \epsilon^{\nu}v^{\rho}(x)W_{\nu\rho}{}^{\mu}(x)] \\ &= e^{a}{}_{\mu}(x)v^{\mu}(x). \end{aligned} \tag{1.105}$$

Also, it can be shown that the vanishing of the curvature is equivalent to the existence of d vector fields (the Vielbeins) covariantly constant with respect to the W connection,

$$\overset{W}{\nabla}_{\mu}\, e^{a}{}_{\nu} = \partial_{\mu}e^{a}{}_{\nu} - W_{\mu\nu}{}^{\rho}e^{a}{}_{\rho} = 0. \tag{1.106}$$

1.5 Killing vectors

If, given a metric $g_{\mu\nu}$, there exists a vector field k^{μ} such that the Lie derivative of $g_{\mu\nu}$ with respect to it vanishes,

$$\mathcal{L}_k g_{\mu\nu} = -2\nabla_{(\mu}k_{\nu)} = 0, \tag{1.107}$$

[12] Also known as *teleparallelism* or *absolute parallelism*.

we say that $g_{\mu\nu}$ admits the *Killing vector* k^μ. The above equation is the *Killing equation*. It means that the metric does not change along the integral curves of k^μ and it is also said that the metric possesses an *isometry* in the direction k^μ. If the metric does not change along the integral curves of a Killing vector, and we use as a coordinate the parameter of those integral curves (*adapted coordinates*), then the metric does not depend on that coordinate.

The Ricci identity implies the following consistency condition:

$$\nabla_\alpha \nabla_\beta k_\nu = -R^\lambda{}_{\alpha\beta\nu} k_\lambda, \quad \Rightarrow \nabla^2 k^\mu = R^\mu{}_\nu k^\nu. \tag{1.108}$$

A weaker but also interesting property that a metric can have is a *conformal isometry*. This happens when there is a vector field c^μ along whose integral curves the metric changes only by a conformal factor,

$$\mathcal{L}_c g_{\mu\nu} = -2\nabla_{(\mu} c_{\nu)} = 2\lambda g_{\mu\nu}. \tag{1.109}$$

On taking the trace of the above equation, we find in d dimensions

$$\lambda = -\frac{1}{d}\nabla_\mu c^\mu, \quad \Rightarrow \nabla_{(\mu} c_{\nu)} - \frac{1}{d} g_{\mu\nu} \nabla_\rho c^\rho = 0, \tag{1.110}$$

called the *conformal Killing equation*. c^μ is then known as a *conformal Killing vector*.

1.6 Duality operations

The antisymmetric Levi-Cività tensor is defined in d dimensions in tangent space by

$$\epsilon^{01\cdots(d-1)} = +1, \qquad \Rightarrow \epsilon_{01\cdots(d-1)} = (-1)^{d-1}, \tag{1.111}$$

and in curved indices by

$$\epsilon^{\mu_1\cdots\mu_d} = \sqrt{|g|}\, e^{\mu_1}{}_{a_1} \cdots e^{\mu_d}{}_{a_d} \epsilon^{a_1\cdots a_d}, \tag{1.112}$$

so, with upper indices, it is independent of the metric and, in curved indices, which we we underline to distinguish them from the tangent-space ones,

$$\epsilon^{\underline{0}\cdots\underline{(d-1)}} = +1, \qquad \epsilon_{\underline{0}\cdots\underline{(d-1)}} = g = (-1)^{d-1}|g|. \tag{1.113}$$

The contraction of n indices of two ϵ symbols gives

$$\epsilon^{\mu_1\cdots\mu_n\rho_1\cdots\rho_{(d-n)}} \epsilon_{\mu_1\cdots\mu_n\sigma_1\cdots\sigma_{(d-n)}} = n!(d-n)!\; g\; g^{\rho_1\cdots\rho_{(d-n)}}{}_{\sigma_1\cdots\sigma_{(d-n)}}, \tag{1.114}$$

where

$$\begin{aligned} g^{\rho_1\cdots\rho_{(d-n)}}{}_{\sigma_1\cdots\sigma_{(d-n)}} &= g_{[\rho_1}{}^{\sigma_1} \cdots g_{[\rho_{(d-n)}]}{}^{\sigma_{(d-n)}} \\ &= \frac{1}{(d-n)!} \begin{vmatrix} g_{\rho_1}{}^{\sigma_1} & \cdots & g_{\rho_1}{}^{\sigma_{(d-n)}} \\ \vdots & \vdots & \vdots \\ g_{\rho_{(d-n)}}{}^{\sigma_1} & \cdots & g_{\rho_{(d-n)}}{}^{\sigma_{(d-n)}} \end{vmatrix} \end{aligned} \tag{1.115}$$

We define the *dual* (or the *Hodge dual*) of a completely antisymmetric tensor of rank k (a *differential form of rank k* or *k-form*[13]) $F_{(k)}$ as the completely antisymmetric tensor of rank $d-k$ which we denote by ${}^\star F_{(d-k)}$ and whose components are given by

$$ {}^\star F_{(k)}{}^{\mu_1\cdots\mu_{(d-k)}} = \frac{1}{k!\sqrt{|g|}} \epsilon^{\mu_1\cdots\mu_{(d-k)}\mu_{(d-k+1)}\cdots\mu_d} F_{(k)\mu_{(d-k+1)}\cdots\mu_d}. \tag{1.116} $$

The dual of the dual is the original tensor up to a sign that depends both on the dimension and on the rank of the tensor,

$$ {}^{\star\star}F_{(k)} = (-1)^{(d-1)+k(d-k)} F_{(k)}. \tag{1.117} $$

An important case is when the spacetime dimension is even and $k=d/2$, so * (the *Hodge star*) is an operator on the space of rank $d/2$ tensors. Then, we have

$$ \begin{aligned} {}^{\star\star}F_{(d/2)} &= +F_{(d/2)}, \qquad d = 4n+2, \\ {}^{\star\star}F_{(d/2)} &= -F_{(d/2)}, \qquad d = 4n, \end{aligned} \tag{1.118} $$

for n an integer. In the former case * has eigenvalues $+1$ and -1 and in the latter $+i$ and $-i$, and any rank $d/2$ tensor can be decomposed into the sum of its self-dual and anti-self-dual parts F^+ and F^-. For $d=4n+2$

$$ \begin{aligned} F^{\pm}_{(d/2)} &= \tfrac{1}{2}\big(F_{(d/2)} \pm {}^\star F_{(d/2)}\big), \\ {}^\star F^{\pm}_{(d/2)} &= \pm F^{\pm}_{(d/2)}, \end{aligned} \tag{1.119} $$

and, for $d=4n$,

$$ \begin{aligned} F^{\pm}_{(d/2)} &= \tfrac{1}{2}\big(F_{(d/2)} \mp i\, {}^\star F_{(d/2)}\big), \\ {}^\star F^{\pm}_{(d/2)} &= \pm i F^{\pm}_{(d/2)}. \end{aligned} \tag{1.120} $$

Real (as opposed to complex) (anti-)self-duality ${}^\star F = (-)+F$ is therefore consistent only in $d=4n+2$ dimensions.

Another important case is when $k=p+2$ and $F_{(k)}$ is the field strength of the potential $A_{(p+1)}$, so $F_{(p+2)\mu_1\cdots\mu_{(p+2)}} = (p+2)\partial_{[\mu_1} A_{(p+1)\mu_2\cdots\mu_{(p+2)}}$. The kinetic term of its action is normalized as follows:

$$ S_{(p)}[A_{(p+1)}] = \int d^dx\sqrt{|g|}\left[\frac{(-1)^{p+1}}{2\cdot(p+2)!}F^2_{(p+2)}\right], \tag{1.121} $$

and its energy–momentum tensor is given by

$$ T_{\mu\nu}^{A_{(p+1)}} = \frac{-2}{\sqrt{|g|}}\frac{\delta S_{(p)}}{\delta g^{\mu\nu}} = \frac{(-1)^p}{(p+1)!}\left[F_{(p+2)\mu}{}^{\rho_1\cdots\rho_{(p+1)}} F_{(p+2)\nu\rho_1\cdots\rho_{(p+1)}} - \frac{1}{2(p+2)} g_{\mu\nu} F^2_{(p+2)}\right]. \tag{1.122} $$

The rank of its dual tensor is $\tilde{p}+2$, where $\tilde{p} = d-p-4$, and we are interested in rewriting the action and energy–momentum tensor in terms of the dual. We immediately find

$$ S_p[A_{(p+1)}] = -\int d^dx\sqrt{|g|}\left[\frac{(-1)^{\tilde{p}+1}}{2\cdot(\tilde{p}+2)!}F^2_{(\tilde{p}+2)}\right] = -S_{\tilde{p}}[\tilde{A}_{(\tilde{p}+1)}], \tag{1.123} $$

[13] We will introduce the notation specific for differential forms in the next section.

which would be the action of a dual vector field $\tilde{A}_{(\tilde{p}+1)}$ such that

$$^{\star}F_{\mu_1\cdots\mu_{(\tilde{p}+2)}} = (\tilde{d}+2)\partial_{[\mu_1}\tilde{A}_{(\tilde{p}+1)\,\mu_2\cdots\mu_{(\tilde{p}+2)}]}.$$

Using

$$\begin{aligned} &^{\star}F_{(\tilde{p}+2)\mu}{}^{\rho_1\cdots\rho_{(\tilde{p}+1)}}\,{}^{\star}F_{(\tilde{p}+2)\nu\rho_1\cdots\rho_{(\tilde{p}+1)}} \\ &\quad = \frac{(-1)^{d-1}(\tilde{p}+1)!}{(p+2)!}g_{\mu\nu}F^2_{(p+2)} + \frac{(-1)^{d}(\tilde{p}+1)!}{(p+1)!}F_{(p+2)\mu}{}^{\sigma_1\cdots\sigma_{(p+1)}}F_{(p+2)\nu\sigma_1\cdots\sigma_{(p+1)}}, \end{aligned} \tag{1.124}$$

we obtain

$$T^{A_{(p+1)}}_{\mu\nu} = T^{\tilde{A}_{(\tilde{p}+1)}}_{\mu\nu}. \tag{1.125}$$

A useful expression for the energy–momentum tensor is

$$\begin{aligned} T^{A_{(p+1)}}_{\mu\nu} = \frac{1}{2}\Bigg\{ &\frac{(-1)^p}{(p+1)!}F_{(p+2)\mu}{}^{\rho_1\cdots\rho_{(p+1)}}F_{(p+2)\nu\rho_1\cdots\rho_{(p+1)}} \\ &+ \frac{(-1)^{\tilde{p}}}{(\tilde{p}+1)!}{}^{*}F_{(\tilde{p}+2)\mu}{}^{\rho_1\cdots\rho_{(\tilde{p}+1)}}\,{}^{\star}F_{(\tilde{p}+2)\nu\rho_1\cdots\rho_{(\tilde{p}+1)}}\Bigg\}. \end{aligned} \tag{1.126}$$

1.7 Differential forms and integration

As we have said before, a *differential form of rank* k, or k*-form* for short, is nothing but a totally antisymmetric tensor field $\omega_{\mu_1\cdots\mu_k} = \omega_{[\mu_1\cdots\mu_k]}$. We write all k-forms in this way:

$$\omega = \frac{1}{k!}\omega_{\mu_1\cdots\mu_k}dx^{\mu_1}\wedge\cdots\wedge dx^{\mu_k}, \tag{1.127}$$

so the action of the *exterior derivative* d on the components is defined by

$$(d\omega)_{\mu_1\cdots\mu_{k+1}} = (k+1)\partial_{[\mu_1}\omega_{\mu_2\cdots\mu_{k+1}]} = (k+1)(\partial\omega)_{\mu_1\cdots\mu_{k+1}}. \tag{1.128}$$

The *Hodge dual* is defined by[14]

$$(^{\star}\omega)_{\mu_1\cdots\mu_{n-k}} = \frac{1}{k!\sqrt{|g|}}\epsilon_{\mu_1\cdots\mu_{n-k}\nu_1\cdots\nu_k}\omega^{\nu_1\cdots\nu_k}, \tag{1.129}$$

and, as before,

$$(^{\star})^2 = (-1)^{k(d-k)}\,\text{sign}\,g = (-1)^{k(d-k)+d-1}. \tag{1.130}$$

The adjoint of d with respect to the inner product of k-forms,

$$(\alpha_k|\beta_k) = \int_{\mathcal{M}}\alpha_k\wedge{}^{\star}\beta_k, \tag{1.131}$$

is defined by

$$(\alpha_k|d\beta_{k-1}) = (\delta\alpha_k|\beta_{k-1}), \quad \Rightarrow \delta = (-1)^{d(k-1)-1}\,\text{sign}\,g\ {}^{\star}d^{\star} \tag{1.132}$$

[14] Observe that we need a metric to do it and that the dual depends explicitly on that metric.

Since

$$\left({}^\star d\,{}^\star\omega\right)_{\rho_1\cdots\rho_{k-1}} = (-1)^{k(d-k+1)-1}\,\mathrm{sign}\,g\;\nabla_\mu\omega^\mu_{\rho_1\cdots\rho_{k-1}}, \tag{1.133}$$

we find that the relation between δ and the divergence is

$$(\delta\omega)_{\rho_1\cdots\rho_{k-1}} = (-1)^d\nabla_\mu\omega^\mu{}_{\rho_1\cdots\rho_{k-1}}. \tag{1.134}$$

Only k-forms can be integrated on k-dimensional manifolds. If ω is a $(d-1)$-form defined on a d-dimensional manifold M with boundary ∂M, then *Stokes' theorem* states that

$$\int_{\mathrm{M}} d\omega = \int_{\partial\mathrm{M}} \omega. \tag{1.135}$$

It is convenient to define *volume forms* for a manifold and its lower-dimensional submanifolds. Their contraction with other tensors results in differential forms that can be integrated. Thus, we define in a d-dimensional manifold, for $(d-n)$-dimensional submanifolds M^{d-n}, $0 \leq n \leq d$, the volume forms

$$d^{d-n}\Sigma_{\mu_1\cdots\mu_n} \equiv dx^{\nu_1}\cdots dx^{\nu_{d-n}}\frac{1}{(d-n)!\sqrt{|g|}}\epsilon_{\nu_1\cdots\nu_{d-n}\mu_1\cdots\mu_n}. \tag{1.136}$$

Observe that the standard invariant-volume form for the total manifold M^d is just $d^d\Sigma$ up to a sign (we now use the signature $(+-\cdots-)$):

$$d^d\Sigma = (-1)^{d-1}dx^1\wedge\cdots\wedge\; dx^d\sqrt{|g|} \equiv (-1)^{d-1}d^dx\sqrt{|g|}. \tag{1.137}$$

Now, if we have a rank-n completely antisymmetric contravariant tensor $T^{\mu_1\cdots\mu_n}$ and contract it with the volume element $d^{d-n}\Sigma_{\mu_1\cdots\mu_n}$, we have constructed a $(d-n)$-form that can be integrated over a $(d-n)$-dimensional submanifold. Up to numerical factors, that form is the Hodge dual of the n-form that one gets by lowering the indices of $T_{\mu_1\cdots\mu_n}$:

$$\frac{1}{n!}d^{d-n}\Sigma_{\mu_1\cdots\mu_n}T^{\mu_1\cdots\mu_n} = {}^\star T. \tag{1.138}$$

We can also take the divergence of the tensor and contract it with the volume element $d^{d-n-1}\Sigma_{\mu_1\cdots\mu_{n-1}}$. The result is

$$\frac{(-1)^{d-n}}{(n-1)!}d^{d-n+1}\Sigma_{\mu_1\cdots\mu_{n-1}}\nabla_\rho T^{\rho\mu_1\cdots\mu_{n-1}} = d\,{}^\star T. \tag{1.139}$$

Stokes' theorem for the exterior derivative of form ${}^\star T$ integrated over a $(d-n+1)$-dimensional submanifold M^{d-n+1} with $(d-n)$-dimensional boundary $\partial\mathrm{M}^{d-n+1}$ is now

$$\int_{\mathrm{M}^{d-n+1}} d^{d-n+1}\Sigma_{\mu_1\cdots\mu_{n-1}}\nabla_\rho T^{\rho\mu_1\cdots\mu_{n-1}} = \frac{(-1)^{d-n}}{n}\int_{\partial\mathrm{M}^{d-n+1}} d^{d-n}\Sigma_{\mu_1\cdots\mu_n}T^{\mu_1\cdots\mu_n}. \tag{1.140}$$

The $n=1$ case is the *Gauss–Ostrogradski theorem*,

$$\int_{\mathrm{M}^d} d^dx\sqrt{|g|}\,\nabla_\mu v^\mu = (-1)^{d-1}\int_{\partial\mathrm{M}} d^{d-1}\Sigma_\mu v^\mu. \tag{1.141}$$

The Vielbein and spin-connection 1-forms and the torsion 2-form are

$$e^a = e_\mu{}^a dx^\mu, \qquad \omega^{ab} = \omega_\mu{}^{ab} dx^\mu, \qquad T^a = \tfrac{1}{2} T_{\mu\nu}{}^a dx^\mu \wedge dx^\nu, \tag{1.142}$$

These 1-forms are related by the *structure equation*

$$de^a + \omega_b{}^a \wedge e^b + T^a = 0, \tag{1.143}$$

which (in the absence of torsion) gives a convenient way of finding ω. The curvature 2-form and the Ricci-tensor 1-form are given by

$$\begin{aligned} R^{ab} &= \tfrac{1}{2} R^{ab}{}_{\mu\nu} dx^\mu \wedge dx^\nu = d\omega^{ab} + \omega_c{}^a \wedge \omega^{cb}, \\ R^a &= R_\mu{}^a dx^\mu = R_{\mu\lambda}{}^{ab} e^\lambda{}_b dx^\mu. \end{aligned} \tag{1.144}$$

1.8 Extrinsic geometry

Let us consider a hypersurface Σ embedded in a d-dimensional spacetime with metric $g_{\mu\nu}$ and with normal unit vector n^μ:

$$n^\mu n_\mu = \varepsilon, \qquad \begin{cases} \varepsilon = +1, & \Sigma \ \text{spacelike}, \\ \varepsilon = -1, & \Sigma \ \text{timelike}. \end{cases} \tag{1.145}$$

The metric induced on Σ by $g_{\mu\nu}$ is defined by

$$h_{\mu\nu} = g_{\mu\nu} - \varepsilon n_\mu n_\nu. \tag{1.146}$$

$h_{\mu\nu}$ has $(d-1)$-dimensional character but it is written in d-dimensional form and it is evidently singular and cannot be inverted. Its indices are raised and lowered with g. Observe that $h_{\mu\nu} n^\nu = 0$ and thus h can be used to project tensors onto the hypersurface Σ.

A way to measure how Σ is curved inside the spacetime would be to measure the variation of the normal unit vector along it. Mathematically this would be expressed by

$$\mathcal{K}_{\mu\nu} \equiv h_\mu{}^\alpha h_\nu{}^\beta \nabla_{(\alpha} n_{\beta)}, \tag{1.147}$$

where $\mathcal{K}_{\mu\nu}$ is the *extrinsic curvature* or *second fundamental form*.

We can consider a field of unit vectors n^μ defined in the whole spacetime determining a family of hypersurfaces. Then we can calculate the Lie derivative of the induced metrics in the direction of the normal unit vectors. We find that this is twice the extrinsic curvature,

$$\mathcal{K}_{\mu\nu} = \tfrac{1}{2} \mathcal{L}_n h_{\mu\nu}. \tag{1.148}$$

The trace of the extrinsic curvature is denoted by $\mathcal{K}$, and given by

$$\mathcal{K} = h^{\mu\nu} \mathcal{K}_{\mu\nu} = h^{\mu\nu} \nabla_\mu n_\nu. \tag{1.149}$$

2
Noether's theorems

In the next chapter, we are going to introduce general relativity as the result of the construction of a self-consistent special-relativistic field theory (SRFT) of gravity. In this construction, gauge symmetry and the energy–momentum tensor will play a key role. In this chapter we want to review Noether's theorems, the relation between global symmetries and conserved charges, and the relation between local symmetries and gauge identities. We will define the canonical energy–momentum tensor as the conserved Noether current associated with the invariance under constant translations and we will review several ways of improving it that are associated with invariance under other spacetime transformations (Lorentz rotations and rescalings). Finally, we will relate these improved energy–momentum tensors to the energy–momentum tensor used in general relativity.

2.1 Equations of motion

Let us consider an action $S[\varphi]$ for a generic field φ, which may have (spacetime or internal) indices that we do not exhibit for the sake of simplicity. Allowing for Lagrangians containing higher derivatives of φ, we write the action as follows:

$$S[\varphi] = \int_{\Sigma} d^d x \, \mathcal{L}(\varphi, \partial\varphi, \partial^2\varphi, \ldots). \tag{2.1}$$

In most cases, $\mathcal{L}$ is a scalar density under the relevant spacetime transformations (Poincaré transformations in SRFTs and general coordinate transformations in general-covariant theories). It is also possible to use a Lagrangian that is a scalar density up to a total derivative,[1] and thus we will make absolutely no assumptions about the transformation properties of the Lagrangian $\mathcal{L}$.

Under arbitrary infinitesimal variations of the field variable $\delta\varphi$

$$\delta S = \int_{\Sigma} d^d x \, \delta\mathcal{L} = \int_{\Sigma} d^d x \, \left\{ \frac{\partial\mathcal{L}}{\partial\varphi}\delta\varphi + \frac{\partial\mathcal{L}}{\partial\partial_\mu\varphi}\delta\partial_\mu\varphi + \frac{\partial\mathcal{L}}{\partial\partial_\mu\partial_\nu\varphi}\delta\partial_\mu\partial_\nu\varphi + \cdots \right\}. \tag{2.2}$$

[1] For instance, in general relativity one may want to eliminate the piece of the Lagrangian with second derivatives, which is a total derivative, but then the rest is not a scalar density.

The variation of the coordinates is zero by hypothesis. Then the variation of the field commutes with the derivatives. On integrating by parts to obtain an overall factor of $\delta\varphi$, we find

$$\delta S = \int_\Sigma d^d x \left\{ \frac{\delta S}{\delta\varphi}\delta\varphi + \partial_\mu \left[\left(\frac{\partial\mathcal{L}}{\partial\partial_\mu\varphi} - \partial_\nu \frac{\partial\mathcal{L}}{\partial\partial_\mu\partial_\nu\varphi} \right) \delta\varphi + \frac{\partial\mathcal{L}}{\partial\partial_\mu\partial_\nu\varphi}\partial_\nu\delta\varphi + \cdots \right] \right\}, \quad (2.3)$$

where we have defined the first variation of the action $\delta S/\delta\varphi$,

$$\frac{\delta S}{\delta\varphi} \equiv \frac{\partial\mathcal{L}}{\partial\varphi} - \partial_\mu \frac{\partial\mathcal{L}}{\partial\partial_\mu\varphi} + \partial_\mu\partial_\nu \frac{\partial\mathcal{L}}{\partial\partial_\mu\partial_\nu\varphi} + \cdots . \quad (2.4)$$

We now use Stokes' theorem Eq. (1.141) to reexpress the integral of the total derivative as an integral over the boundary $\partial\Sigma$:

$$\delta S = \int_\Sigma d^d x \frac{\delta S}{\delta\varphi}\delta\varphi + (-1)^{d-1} \int_{\partial\Sigma} d^{d-1}\Sigma_\mu \left\{ \left(\frac{\partial\mathcal{L}}{\partial\partial_\mu\varphi} - \partial_\nu \frac{\partial\mathcal{L}}{\partial\partial_\mu\partial_\nu\varphi} \right) \delta\varphi \right.$$
$$\left. + \frac{\partial\mathcal{L}}{\partial\partial_\mu\partial_\nu\varphi}\partial_\nu\delta\varphi + \cdots \right\}. \quad (2.5)$$

In theories without higher derivatives $\mathcal{L}(\varphi, \partial\varphi)$ it is enough to impose that the field variations vanish over the boundary $\delta\varphi|_{\partial\Sigma} = 0$, to see that the boundary term vanishes. Then, requiring that the action is stationary, $\delta S = 0$, under those variations we obtain the usual *Euler–Lagrange equations*

$$\frac{\delta S}{\delta\varphi} = \frac{\partial\mathcal{L}}{\partial\varphi} - \partial^\mu \left(\frac{\partial\mathcal{L}}{\partial\partial^\mu\varphi} \right) = 0. \quad (2.6)$$

If the Lagrangian contains higher derivatives of the field, it is necessary either to impose boundary conditions for derivatives of the variation of the field or to introduce (if possible) into the action boundary terms that do not change the equations of motion but eliminate the $\partial\delta\varphi$ term in the total derivative. In any of these cases we obtain the equations of motion

$$\frac{\delta S}{\delta\varphi} = \frac{\partial\mathcal{L}}{\partial\varphi} - \partial^\mu \left(\frac{\partial\mathcal{L}}{\partial\partial^\mu\varphi} \right) + \partial^\nu\partial^\mu \left(\frac{\partial\mathcal{L}}{\partial\partial^\nu\partial^\mu\varphi} \right) - \cdots = 0. \quad (2.7)$$

As we can see, the equations of motion are of degree higher than two in derivatives of the field. Thus, to solve them completely it is also necessary to give boundary conditions for the field, and for its first and higher derivatives.

If we add a total derivative term $\partial_\mu \mathfrak{k}^\mu(\varphi)$ to the Lagrangian, it is clear that the equations of motion will not be modified *as long as the boundary conditions for $\delta\varphi$ and its derivatives make $\mathfrak{k}^\mu(\varphi) = 0$ on the boundary.*

2.2 Noether's theorems

Let us now consider the infinitesimal transformations of the coordinates and fields $\tilde{\delta}x^\mu$ and $\tilde{\delta}\varphi$:

$$\tilde{\delta}x^\mu = x'^\mu - x^\mu,$$
$$\tilde{\delta}\varphi(x) \equiv \varphi'(x') - \varphi(x), \quad (2.8)$$

where x' and x stand for the coordinates of the same point in the two different coordinate systems. The transformation of the fields may contain terms associated with the coordinate transformations and also with other "internal" transformations (see footnote 3 in Chapter 1).

We want to find the consequences of the invariance, possibly up to a total derivative that depends on the variations, of the action Eq. (2.1) under the above infinitesimal changes of the field and the coordinates (which are, then, symmetry transformations). We express this invariance as follows:

$$\tilde{\delta} S = \int_{\Sigma} d^d x \, \partial_\mu \mathfrak{s}^\mu(\tilde{\delta}). \tag{2.9}$$

Let us now perform directly the variation of the action explicitly,[2]

$$\tilde{\delta} S = \int_{\Sigma} \left[\tilde{\delta} d^d x \, \mathcal{L} + d^d x \, \tilde{\delta} \mathcal{L} \right]. \tag{2.10}$$

We have

$$\begin{aligned} \tilde{\delta} d^d x &= d^d x \, \partial_\mu \tilde{\delta} x^\mu, \\ \tilde{\delta} \mathcal{L} &= \delta \mathcal{L} + \tilde{\delta} x^\mu \partial_\mu \mathcal{L}, \\ \delta \mathcal{L} &= \frac{\partial \mathcal{L}}{\partial \varphi} \delta \varphi + \frac{\partial \mathcal{L}}{\partial \partial_\mu \varphi} \delta \partial_\mu \varphi + \frac{\partial \mathcal{L}}{\partial \partial_\mu \partial_\nu \varphi} \delta \partial_\mu \partial_\nu \varphi + \cdots, \end{aligned} \tag{2.11}$$

where δ stands for the variation of the field at two different points whose coordinates are the same in the two different coordinate systems considered,

$$\delta \varphi(x) \equiv \varphi'(x) - \varphi(x), \tag{2.12}$$

and we have used the field-operator identity

$$\tilde{\delta} = \delta + \tilde{\delta} x^\mu \partial_\mu. \tag{2.13}$$

δ and ∂_μ commute since δ does not involve any change of coordinates. Thus

$$\tilde{\delta} S = \int_{\Sigma} d^d x \, \left\{ \partial_\mu \tilde{\delta} x^\mu \mathcal{L} + \tilde{\delta} x^\mu \partial_\mu \mathcal{L} + \frac{\partial \mathcal{L}}{\partial \varphi} \delta \varphi + \frac{\partial \mathcal{L}}{\partial \partial_\mu \varphi} \partial_\mu \delta \varphi + \frac{\partial \mathcal{L}}{\partial \partial_\mu \partial_\nu \varphi} \partial_\mu \partial_\nu \delta \varphi + \cdots \right\}. \tag{2.14}$$

On integrating by parts as many times as necessary, we obtain

$$\tilde{\delta} S = \int_{\Sigma} d^d x \left\{ \partial_\mu \left[\mathcal{L} \tilde{\delta} x^\mu + \left(\frac{\partial \mathcal{L}}{\partial \partial_\mu \varphi} - \partial_\nu \frac{\partial \mathcal{L}}{\partial \partial_\mu \partial_\nu \varphi} \right) \delta \varphi + \frac{\partial \mathcal{L}}{\partial \partial_\mu \partial_\nu \varphi} \partial_\nu \delta \varphi + \cdots \right] + \frac{\delta S}{\delta \varphi} \delta \varphi \right\}. \tag{2.15}$$

On reexpressing $\delta\varphi$ in terms of $\tilde{\delta}\varphi$ inside the total derivative, and equating the result with Eq. (2.9), we arrive at

$$\int_{\Sigma} d^d x \, \left\{ \partial_\mu \mathfrak{j}^\mu_{\mathrm{N}1}(\tilde{\delta}) + \frac{\delta S}{\delta \varphi} \delta \varphi \right\} = 0, \tag{2.16}$$

[2] We follow here [795].

where

$$\mathfrak{j}^{\mu}_{\mathrm{N1}}(\tilde{\delta}) = -\mathfrak{s}^{\mu}(\tilde{\delta}) + T_{\mathrm{can}}{}^{\mu}{}_{\nu}\tilde{\delta}x^{\nu} - \frac{\partial\mathcal{L}}{\partial\partial_{\mu}\partial_{\nu}\varphi}\partial_{\rho}\varphi\partial_{\nu}\tilde{\delta}x^{\rho}$$
$$+\left[\frac{\partial\mathcal{L}}{\partial\partial_{\mu}\varphi} - \partial_{\nu}\left(\frac{\partial\mathcal{L}}{\partial\partial_{\mu}\partial_{\nu}\varphi}\right)\right]\tilde{\delta}\varphi + \frac{\partial\mathcal{L}}{\partial\partial_{\mu}\partial_{\nu}\varphi}\partial_{\nu}\tilde{\delta}\varphi + \cdots, \tag{2.17}$$

where, in turn,

$$T_{\mathrm{can}}{}^{\mu}{}_{\nu} = \eta^{\mu}{}_{\nu}\mathcal{L} - \frac{\partial\mathcal{L}}{\partial\partial_{\mu}\varphi}\partial_{\nu}\varphi - \frac{\partial\mathcal{L}}{\partial\partial_{\mu}\partial_{\rho}\varphi}\partial_{\nu}\partial_{\rho}\varphi + \partial_{\rho}\left(\frac{\partial\mathcal{L}}{\partial\partial_{\mu}\partial_{\rho}\varphi}\right)\partial_{\nu}\varphi + \cdots. \tag{2.18}$$

$T_{\mathrm{can}}{}^{\mu}{}_{\nu}$ is the *canonical energy–momentum tensor* and is the only piece of $\mathfrak{j}^{\mu}_{\mathrm{N1}}$ that survives (apart from $\mathfrak{s}^{\mu}$) when we consider constant $\tilde{\delta}x^{\mu}$s.

It is worth stressing that the total-derivative term will not vanish in general after use of Stokes' theorem because the variations $\tilde{\delta}x^{\mu}$ and $\tilde{\delta}\varphi$ do not vanish on the boundary.

Now we want to derive conservation laws from this identity. We see that, in the general case, if the equations of motion $\delta S/\delta\varphi = 0$ are satisfied, then we can conclude that $\mathfrak{j}^{\mu}_{\mathrm{N1}}(\tilde{\delta})$ is a *conserved vector current* (*Noether current*), i.e. satisfies the *continuity equation*

$$\partial_{\mu}\mathfrak{j}^{\mu}_{\mathrm{N1}}(\tilde{\delta}) = 0. \tag{2.19}$$

Thus, for a theory that is exactly invariant under constant translations, the canonical energy–momentum tensor is the associated Noether conserved current.

Strictly speaking $\mathfrak{j}^{\mu}_{\mathrm{N1}}(\tilde{\delta})$ is a vector density. In the presence of a metric, we can define a vector current $\mathfrak{j}^{\mu}_{\mathrm{N1}}(\tilde{\delta}) = \sqrt{|g|}\,j^{\mu}_{\mathrm{N1}}(\tilde{\delta})$ and write the continuity equation in general-covariant form:

$$\nabla_{\mu}j^{\mu}_{\mathrm{N1}}(\tilde{\delta}) = 0. \tag{2.20}$$

In Minkowski spacetime this distinction is unnecessary. Such terms are called "conserved" because they are used to define quantities (*charges*) that are conserved in time, as we will see next.

This is the best we can do if the transformations are global, i.e. when they take the form

$$\tilde{\delta}x^{\mu} \equiv \sigma^{I}\tilde{\delta}_{I}x^{\mu}, \qquad \tilde{\delta}\varphi \equiv \sigma^{I}\tilde{\delta}_{I}\varphi, \tag{2.21}$$

where $\tilde{\delta}_{I}x^{\mu}$ and $\tilde{\delta}_{I}\varphi$ are given functions of the coordinates and φ and the σ^{I}, $I = 1, \ldots, n$, are the constant transformation parameters. Then, we find n on-shell conserved currents $\mathfrak{j}^{\mu}_{N1\,I}$ independent of the parameters σ^{I} and they are given by

$$\mathfrak{j}^{\mu}_{\mathrm{N1}\,I} = -s^{\mu}(\tilde{\delta}_{I}) + T_{\mathrm{can}}{}^{\mu}{}_{\nu}\tilde{\delta}_{I}x^{\nu} - \frac{\partial\mathcal{L}}{\partial\partial_{\mu}\partial_{\nu}\varphi}\partial_{\rho}\varphi\partial_{\nu}\tilde{\delta}_{I}x^{\rho}$$
$$+\left[\frac{\partial\mathcal{L}}{\partial\partial_{\mu}\varphi} - \partial_{\nu}\left(\frac{\partial\mathcal{L}}{\partial\partial_{\mu}\partial_{\nu}\varphi}\right)\right]\tilde{\delta}_{I}\varphi + \frac{\partial\mathcal{L}}{\partial\partial_{\mu}\partial_{\nu}\varphi}\partial_{\nu}\tilde{\delta}_{I}\varphi + \cdots. \tag{2.22}$$

If the transformations are local, i.e. they depend on n local parameters $\sigma_{I}(x)$, the generic result of the on-shell conservation of the current $\mathfrak{j}^{\mu}_{\mathrm{N1}}$ is still true,[3] but we can do more.

[3] In principle there is one current for each value of the local parameters. This gives an infinite number of on-shell conserved currents. However, only for a certain asymptotic behavior of the $\sigma_{I}(x)$s will the integrals defining the conserved charges converge. These asymptotic behaviors are usually associated with the global invariances of the vacuum configuration.

First, observe that, in general, in the local case, the transformations contain derivatives of the local parameters. We eliminate these derivatives of the transformation parameters by integration by parts in Eq. (2.16). We obtain an identity of the form[4]

$$\int_\Sigma d^d x \left\{ \partial_\mu \mathrm{j}^\mu_{\mathrm{N2}}(\sigma) + \sigma^I D_I \frac{\delta S}{\delta \varphi} \right\}, \tag{2.23}$$

where D_I are operators containing derivatives acting on the equations of motion. This identity is true for arbitrary parameters. We can choose parameters such that $\mathrm{j}^\mu_{\mathrm{N2}}(\sigma)$ vanishes on the boundary. Then, we obtain the off-shell identities that do not involve the transformation parameters

$$D_I \frac{\delta S}{\delta \varphi} = 0, \tag{2.24}$$

that relate the equations of motion, so not all of them are independent. These identities are called *gauge* or *Bianchi identities*. Since they are identically true for arbitrary values of the parameters, we obtain the off-shell "conservation law"[5]

$$\partial_\mu \mathrm{j}^\mu_{\mathrm{N2}}(\sigma) = 0. \tag{2.25}$$

Since this is an identity that holds independently of the equations of motion, it follows that the current density $\mathrm{j}^\mu_{\mathrm{N2}}(\sigma)$ can always be written as the divergence of a two-index antisymmetric tensor, usually called the *superpotential*, that is,

$$\mathrm{j}^\mu_{\mathrm{N2}}(\sigma) = \partial_\nu \mathrm{j}^{\nu\mu}_{\mathrm{N2}}(\sigma), \qquad \mathrm{j}^{\nu\mu}_{\mathrm{N2}}(\sigma) = -\mathrm{j}^{\mu\nu}_{\mathrm{N2}}(\sigma). \tag{2.26}$$

This identity for the vector densities is written in terms of the vectors $\mathrm{j}^\mu_{\mathrm{N2}} = \sqrt{|g|}\, j^\mu_{\mathrm{N2}}$:

$$j^\mu_{\mathrm{N2}}(\sigma) = \nabla_\nu j^{\nu\mu}_{\mathrm{N2}}(\sigma), \qquad j^{\nu\mu}_{\mathrm{N2}}(\sigma) = -j^{\mu\nu}_{\mathrm{N2}}(\sigma). \tag{2.27}$$

Observe that the difference between $\mathrm{j}^\mu_{\mathrm{N1}}$ and $\mathrm{j}^\mu_{\mathrm{N2}}$ is always a term proportional to the equations of motion, i.e. it vanishes on-shell. Thus, these two currents are identical on-shell. In general we are free to add any term that vanishes on-shell to the current j^μ_{N1} since it is conserved only on-shell. We have just seen that there is a specific on-shell vanishing term that relates j^μ_{N1} to j^μ_{N2}. j^μ_{N2} cannot be modified in this way because its defining property is that it is conserved off-shell. However, we could add to both currents terms of the form $\partial_\nu \Psi^{\nu\mu}$, where $\Psi^{\mu\nu} = \Psi^{[\mu\nu]}$, which would change the superpotential. If $\Psi^{\nu\mu}$ is of the form $\partial_\rho U^{\rho\nu\mu}$, with $U^{\rho\nu\mu} = U^{[\rho\nu]\mu}$, then $\partial_\nu \Psi^{\nu\mu} = 0$ and the change in the superpotential will not change the Noether current.

It is easy to see that Noether currents are sensitive to the addition of total derivatives to the Lagrangian even if these do not modify the equations of motion: on adding to the action (2.1)

$$\Delta S = \int_\Sigma d^d x\, \partial_\mu \mathcal{L}^\mu, \tag{2.28}$$

[4] This expression is just symbolic. We need a more explicit form of the infinitesimal transformations in order to obtain more explicit expressions. We will find several examples in the following chapters.

[5] *Strong conservation law* in the language of [110].

which is also invariant up to a total derivative

$$\tilde{\delta}\Delta S = \int_{\Sigma} d^d x \partial_\mu \Delta \mathfrak{s}^\mu(\tilde{\delta}), \tag{2.29}$$

and repeating the same steps as those we followed to find the Noether currents, we find a correction to the Noether current Eq. (2.17):

$$\begin{aligned} \Delta \mathrm{j}_{\mathrm{N1}}^{\mu}(\tilde{\delta}) &= -\Delta \mathfrak{s}^\mu(\tilde{\delta}) + \tilde{\delta}\mathcal{L}^\mu + \partial_\rho \Psi^{\rho\mu}{}_\nu \tilde{\delta} x^\nu, \\ \Psi^{\rho\mu}{}_\nu &= 2\mathcal{L}^{[\rho}\eta^{\mu]}{}_\nu. \end{aligned} \tag{2.30}$$

If we consider only constant spacetime translations and $\mathcal{L}^\mu$ is a vector density, then $\tilde{\delta}\mathcal{L}^\mu = \Delta\mathfrak{s}^\mu(\tilde{\delta})$ and we simply find a correction to the canonical energy–momentum tensor with the form of a superpotential.

We end this section with an important remark: no change in the superpotential can be related to the addition of a total derivative to the Lagrangian.

2.3 Conserved charges

Given a conserved current (density) j^μ, by taking the integral of its time component j^0 over a piece V_t of a constant-time hypersurface we can define a quantity (*charge*) $Q(\mathrm{V}_t)$,

$$Q(\mathrm{V}_t) = \int_{\mathrm{V}_t} d^{d-1}x\, \mathrm{j}^0. \tag{2.31}$$

If we take the total time derivative of $Q(\mathrm{V}_t)$, since the volume of V_t does not depend on time (the subindex t indicates only that it is in a given constant-t hypersurface, but it is the same spatial volume for all t) the total time derivative "goes through the integral symbol" and becomes a partial time derivative of j^0 ($c = 1$):

$$\frac{d}{dt} Q(\mathrm{V}_t) = \int_{\mathrm{V}_t} d^{d-1}x\, \partial_0 \mathrm{j}^0. \tag{2.32}$$

The continuity equation for the current and Stokes' theorem imply that

$$\frac{d}{dt} Q(\mathrm{V}_t) = \int_{\mathrm{V}_t} d^{d-1}x\, \partial_i \mathrm{j}^i = \int_{\partial \mathrm{V}_t} d^{d-2}\Sigma_i\, j^i, \tag{2.33}$$

which is interpreted as the flux of charge across the boundary of the volume of V_t. Observe that the last integral is performed over j^i rather than over j^i.

This is a *local charge-conservation law*: the charge contained in the volume of V_t is only lost (or gained) by the interchange of charge with the exterior; it does not disappear into nothing and it is not created from nothing. This is what we mean by *conserved charge*.

If we take the boundary of the volume to spatial infinity, and we assume that the currents go to zero at infinity (there are no sources at infinity for the charges), then the flux integral over the boundary vanishes and we see that the total charge contained in space at a given time is conserved in absolute terms. It is usually denoted by Q (all reference to time-dependence has been eliminated).

Sometimes it is convenient to use a more-covariant expression for the charge:

$$Q(\mathrm{V}_t) = \int_{\mathrm{V}_t} d^{d-1}\Sigma_\mu j^\mu. \tag{2.34}$$

If the current can be expressed as the divergence of an antisymmetric two-index tensor $j^\mu = \nabla_\nu j^{\nu\mu}$, $j^{\nu\mu} = -j^{\mu\nu}$, then we can again use Stokes' theorem to express the charge as an integral over the boundary of V_t:

$$Q(\mathrm{V}_t) = \frac{(-1)^{d-2}}{2} \int_{\partial \mathrm{V}_t} d^{d-2}\Sigma_{\mu\nu} j^{\mu\nu}. \tag{2.35}$$

The total charge is found by integrating over the boundary of a constant-time slice, which in general has the topology of an S^{d-2} sphere and lies at spatial infinity. Then, the general expression for the total conserved charge associated with a gauge symmetry is

$$Q = \frac{(-1)^{d-2}}{2} \int_{\mathrm{S}^{d-2}_\infty} d^{d-2}\Sigma_{\mu\nu} j^{\mu\nu}. \tag{2.36}$$

A change in the superpotential $\Psi^{\mu\nu}$ will also change the conserved charge unless the change in the potential vanishes at infinity or unless the change in the superpotential is also of the form $\partial_\rho U^{[\rho\mu]\nu}$ because we can use again Stokes' theorem and reduce the above integral to an integral over the boundary of V_t, which is zero.

2.4 The special-relativistic energy–momentum tensor

In special-relativistic field theories the Lagrangian is, by hypothesis, a scalar under Poincaré transformations, i.e. $\tilde{\delta}\mathcal{L} = 0$. These are translations a^μ and Lorentz transformations $\Lambda^\mu{}_\nu$,

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu + a^\mu, \qquad \Lambda^\mu{}_\rho \eta_{\mu\nu} \Lambda^\nu{}_\sigma = \eta_{\rho\sigma}, \tag{2.37}$$

or, infinitesimally,

$$\tilde{\delta} x^\mu = \sigma^\mu{}_\nu x^\nu + \sigma^\mu, \qquad \sigma^{\mu\nu} = -\sigma^{\nu\mu}. \tag{2.38}$$

The Minkowskian volume element $d^d x$ is also invariant under these transformations $\tilde{\delta} d^d x = 0$ and so the action is also exactly invariant, $\tilde{\delta} S = 0$ ($\mathfrak{s}^\mu = 0$).

Let us first consider infinitesimal translations. In SRFTs all fields are scalars under them, i.e. $\tilde{\delta}\varphi = 0$. Following the standard Noether procedure, we obtain d conserved Noether currents (one for each independent translation) that can be labeled by a subindex (ν),

$$j^\mu_{\mathrm{N1}\,(\nu)} = T_{\mathrm{can}}{}^\mu{}_\rho \tilde{\delta}_\nu x^\rho = T_{\mathrm{can}}{}^\mu{}_\nu, \tag{2.39}$$

since $\tilde{\delta}_\nu x^\rho = \delta_\nu{}^\rho$. The d conserved currents transform as a contravariant vector with respect to the label (ν) and thus they are put together into the canonical energy–momentum tensor given by Eq. (2.18) for higher-derivative theories.

Let us take for example a real scalar field φ. The Lagrangian and equation of motion are

$$\mathcal{L}(\varphi) = \tfrac{1}{2}(\partial\varphi)^2, \qquad \partial^2\varphi = 0, \tag{2.40}$$

and the canonical energy–momentum tensor resulting from the use of the general formula in this case is symmetric and conserved using the above equation of motion:

$$\begin{aligned} T_{\mu\nu}(\varphi) &= -\partial_\mu\varphi\partial_\nu\varphi + \tfrac{1}{2}\eta_{\mu\nu}(\partial\varphi)^2, \\ \partial^\mu T_{\text{matter}\,\mu\nu}(\varphi) &= -\partial_\nu\varphi\partial^2\varphi\big|_{\text{on-shell}} = 0. \end{aligned} \tag{2.41}$$

If we add a total derivative $\partial_\rho(\varphi\partial^\rho\varphi)$ to the above Lagrangian, the equations of motion do not change, as can be seen by using the Euler–Lagrange equations for higher-derivative theories (2.7). According to Eq. (2.18), the energy–momentum tensor acquires the extra term

$$+\partial_\rho\Psi^{\rho\mu\nu}, \qquad \Psi^{\rho\mu\nu} = 2\eta^{\nu[\mu}\varphi\partial^{\rho]}\varphi, \tag{2.42}$$

which is also symmetric but contains second derivatives of the field.

Although the canonical energy–momentum tensor arises as the Noether current associated with invariance under constant translations, we are going to see that it is a much richer object and contains information on the response of a theory to spacetime transformations.

Observe that the canonical energy–momentum tensor is not symmetric in general. In fact, it is symmetric only for scalar fields. However, it can be symmetrized, as we are going to explain when we study the conservation of angular momentum.

For each vector current, we can define the charge $Q_{(\nu)}$,

$$Q_{(\nu)} = \int_{V_t} d^{d-1}x\, j^0_{(\nu)} = \int_{V_t} d^{d-1}x\, T_{\text{can}}{}^0{}_\nu. \tag{2.43}$$

The d conserved charges associated with the energy–momentum tensor are the d components of a contravariant Lorentz vector, which is nothing but the momentum vector and thus we have derived the local conservation laws of energy and momentum. It is customary to write $P^\nu = Q_{(\nu)}$.

2.4.1 Conservation of angular momentum

Let us now consider the infinitesimal Lorentz transformations. The fields appearing in SRFTs transform covariantly or contravariantly in definite representations of the Lorentz group. Let us take, for instance, a field φ^α transforming contravariantly in the representation r of the Lorentz group. The index α goes from 1 to d_r, the dimension of the representation r. If, in the representation r, the generators of the Lorentz group are the $d_r \times d_r$ matrices $\Gamma_r\left(M_{\mu\nu}\right)^\alpha{}_\beta$, then an infinitesimal Lorentz transformation of the field φ can be written in the form

$$\tilde{\delta}\varphi^\alpha = \tfrac{1}{2}\sigma^{\mu\nu}\Gamma_r\left(M_{\mu\nu}\right)^\alpha{}_\beta\varphi^\beta = \tfrac{1}{2}\sigma^{\mu\nu}\tilde{\delta}_{(\mu\nu)}\varphi^\alpha. \tag{2.44}$$

Observe that we can write

$$\tilde{\delta}_{(\rho\sigma)}x^\mu = \Gamma_{\text{v}}\left(M_{\rho\sigma}\right)^\mu{}_\nu x^\nu, \tag{2.45}$$

where Γ_{v} is the vector representation given in Eq. (A.60).

According to the general result Eq. (2.22), we find the following set of $d(d-1)/2$ conserved currents labeled by a pair of antisymmetric indices:[6]

$$j_{N1\,(\rho\sigma)}{}^{\mu} = T_{\mathrm{can}}{}^{\mu}{}_{\lambda}\tilde{\delta}_{(\rho\sigma)}x^{\lambda} + \frac{\partial\mathcal{L}}{\partial\partial_{\mu}\varphi^{\alpha}}\tilde{\delta}_{(\rho\sigma)}\varphi^{\alpha} = 2T_{\mathrm{can}}{}^{\mu}{}_{[\rho}x_{\sigma]} + \frac{\partial\mathcal{L}}{\partial\partial_{\mu}\varphi^{\alpha}}\Gamma_r\left(M_{\rho\sigma}\right)^{\alpha}{}_{\beta}\varphi^{\beta}. \quad (2.46)$$

The first contribution to this current is the *orbital-angular-momentum tensor* and the second is the *spin-angular-momentum tensor*, $S^{\mu}{}_{\rho\sigma}$,

$$S^{\mu}{}_{\rho\sigma} \equiv \frac{1}{2}\frac{\partial\mathcal{L}}{\partial\partial_{\mu}\varphi^{\alpha}}\Gamma_r\left(M_{\rho\sigma}\right)^{\alpha}{}_{\beta}\varphi^{\beta}. \quad (2.47)$$

Only the total angular-momentum current is conserved.

The $d(d-1)/2$ conserved charges are the components of a two-index antisymmetric tensor: the angular-momentum tensor $M_{\mu\nu}$,

$$M_{\mu\nu} = Q_{(\mu\nu)} = \int_{\mathrm{V}_t} d^{d-1}x\; j^{0}_{\mathrm{N1}\,(\mu\nu)}. \quad (2.48)$$

It is instructive to take the divergence of the above current. Since in the theories we are dealing with we always have $\partial_{\mu}T_{\mathrm{can}}{}^{\mu}{}_{\nu} = 0$, one finds

$$\partial_{\mu}j_{\mathrm{N1}\,(\rho\sigma)}{}^{\mu} = -2T_{\mathrm{can}\,[\rho\sigma]} + \partial_{\mu}\left(\frac{\partial\mathcal{L}}{\partial\partial_{\mu}\varphi^{\alpha}}\Gamma_r\left(M_{\rho\sigma}\right)^{\alpha}{}_{\beta}\varphi^{\beta}\right), \quad (2.49)$$

which should vanish on-shell according to the general formalism. This means that, except for scalars, $T_{\mathrm{can}\,\mu\nu}$ is not symmetric and the antisymmetric part is given by

$$T_{\mathrm{can}\,[\rho\sigma]} = \partial_{\mu}S^{\mu}{}_{\rho\sigma}, \quad (2.50)$$

up to terms vanishing on-shell. This formula suggests that we can symmetrize the canonical energy–momentum tensor, exploiting the ambiguities of Noether currents mentioned earlier, i.e. adding to it a term of the form

$$\partial_{\mu}\Psi^{\mu\rho}{}_{\sigma}, \qquad \Psi^{\mu\rho}{}_{\sigma} = -\Psi^{\rho\mu}{}_{\sigma}, \quad (2.51)$$

whose divergence is automatically zero, which in this case would be given by the *spin–energy potential*

$$\Psi^{\mu\rho}{}_{\sigma} = -S^{\mu\rho}{}_{\sigma} + S^{\rho\mu}{}_{\sigma} + S_{\sigma}{}^{\mu\rho}, \quad (2.52)$$

and also removing all the antisymmetric terms that vanish on-shell. The resulting symmetric energy–momentum tensor is usually considered as the energy–momentum tensor to which gravity couples[7] [939] and we will denote it simply by $T^{\mu}{}_{\nu}$. It is also called the *Belinfante tensor* [103]. Using it, the conserved current associated with Lorentz rotations is

$$j_{\mathrm{N1}\,(\rho\sigma)}{}^{\mu} = 2T^{\mu}{}_{[\rho}x_{\sigma]} + \partial_{\lambda}\left(\Psi^{\lambda\mu}{}_{[\rho}x_{\sigma]}\right), \quad (2.53)$$

[6] Here we concentrate on theories without higher derivatives.

[7] This can be justified in the framework of the Cartan–Sciama–Kibble (CSK) theory of gravity. As we will see in Section 4.4, the Belinfante tensor has to coincide with the Rosenfeld energy–momentum tensor, whose definition is based precisely on the coupling to gravity. It is also worth mentioning that, in the CSK theory, the spin–energy potential also couples to gravity through the torsion (the energy–momentum tensor couples through the metric).

again, up to terms that vanish on-shell. The second term in this expression can be eliminated by the usual procedure. The spin-angular-momentum tensor has been absorbed into the new angular-momentum tensor. We are left with the following conserved on-shell currents associated with translations and Lorentz rotations, both of them expressed in terms of the same energy–momentum tensor (the Belinfante tensor):

$$\begin{aligned} j_{\mathrm{N1}\,(\nu)}{}^{\mu} &= T^{\mu}{}_{\rho}\tilde{\delta}_{(\nu)}x^{\rho} \;\;= T^{\mu}{}_{\rho}, \\ j_{\mathrm{N1}\,(\rho\sigma)}{}^{\mu} &= T^{\mu}{}_{\lambda}\tilde{\delta}_{(\rho\sigma)}x^{\lambda} = 2T^{\mu}{}_{[\rho}x_{\sigma]}. \end{aligned} \tag{2.54}$$

It is worth stressing that the existence of these conserved currents is primarily due to the invariance of the Minkowski metric that enters into special-relativistic Lagrangians and of the Minkowski volume element under the Poincaré group or, in other words, to the existence of $d(d+1)/2$ Killing vectors precisely of the form

$$\tilde{\delta}_{(\nu)}x^{\mu}\partial_{\mu} = \partial_{\nu}, \qquad \tilde{\delta}_{(\rho\sigma)}x^{\mu}\partial_{\mu} = -2x_{[\rho}\partial_{\sigma]}. \tag{2.55}$$

A couple of simple examples of the symmetrization of the canonical energy–momentum tensor are in order here.

The energy–momentum tensor of a vector field. The Lagrangian and canonical energy–momentum tensor are given by

$$\mathcal{L} = -\tfrac{1}{4}F^{2}, \qquad F_{\mu\nu} = 2\partial_{[\mu}A_{\nu]}, \qquad T_{\mathrm{can}}{}^{\mu}{}_{\nu} = F^{\mu\rho}\partial_{\nu}A_{\rho} - \tfrac{1}{4}\eta^{\mu}{}_{\nu}F^{2}. \tag{2.56}$$

Under Lorentz rotations we have

$$\tilde{\delta}A_{\mu} = -A_{\nu}\sigma^{\nu}{}_{\mu} \Rightarrow S^{\mu}{}_{\rho\sigma} = F^{\mu}{}_{[\rho}A_{\sigma]} \Rightarrow \Psi^{\rho\mu}{}_{\nu} = F^{\rho\mu}A_{\nu}, \tag{2.57}$$

and, using the equations of motion $\partial_{\rho}F^{\rho\mu} = 0$,

$$T^{\mu}{}_{\nu} = T_{\mathrm{can}}{}^{\mu}{}_{\nu} + \partial_{\rho}\Psi^{\rho\mu}{}_{\nu} = F^{\mu\rho}F_{\nu\rho} - \tfrac{1}{4}\eta^{\mu}{}_{\nu}F^{2}, \tag{2.58}$$

which is the standard, gauge-invariant, energy–momentum tensor of a vector field, coinciding with the one derived via Rosenfeld's prescription, which we are going to introduce in Section 2.4.3, inspired by general relativity.

There is yet another way to obtain this energy–momentum tensor that is worth pointing out: let us consider the transformations

$$\tilde{\delta}x^{\mu} = \epsilon^{\mu}, \qquad \tilde{\delta}A_{\mu} = \epsilon^{\lambda}\partial_{\lambda}A_{\mu} - \mathcal{L}_{\epsilon}A_{\mu} = -\partial_{\mu}\epsilon^{\lambda}A_{\lambda}. \tag{2.59}$$

Following the same steps as those we followed to prove the Noether theorem, we find now

$$\tilde{\delta}S = \int d^{d}x\,\partial_{\mu}\epsilon^{\lambda}T^{\mu}{}_{\nu}, \tag{2.60}$$

with $T^{\mu}{}_{\nu}$ as above (the Belinfante tensor). This variation vanishes if

$$\partial_{(\mu}\epsilon_{\lambda)} = 0, \tag{2.61}$$

which is the Killing equation in Minkowski spacetime. Then, there is invariance under Poincaré transformations whose generators are $\tilde{\delta}x^{\lambda} = \tilde{\delta}_{(\nu)}x^{\lambda}$ and $\tilde{\delta}x^{\lambda} = \tilde{\delta}_{(\rho\sigma)}x^{\lambda}$.

On integrating the above variation by parts, using the fact that it vanishes for Poincaré transformations and using the equation of motion (which implies that $\partial_{\mu}T^{\mu}{}_{\lambda} = 0$), we find

$$\int d^{d}x\, \partial_{\mu}\left(\tilde{\delta}x^{\lambda}T^{\mu}{}_{\nu}\right) = 0, \tag{2.62}$$

and we find automatically the above Noether currents.

This method is clearly inspired by general relativity. We will find more applications for it soon.

The energy–momentum tensor of a Dirac spinor. The Lagrangian of a massive Dirac spinor is[8]

$$\mathcal{L} = \tfrac{1}{2}(i\bar{\psi}\,\partial\!\!\!/\psi - i\bar{\psi}\,\overleftarrow{\partial\!\!\!/}\psi) - m\bar{\psi}\psi, \qquad \bar{\psi}\,\overleftarrow{\partial\!\!\!/} \equiv \partial_{\mu}\bar{\psi}\gamma^{\mu}. \tag{2.63}$$

It is customary to vary ψ and $\bar{\psi}$ as if they were independent. This simplifies somewhat the calculations but we have to bear in mind that they are not independent. The equations of motion of ψ and $\bar{\psi}$ are the Dirac conjugates of each other:

$$(i\,\partial\!\!\!/ - m)\psi = 0, \qquad \bar{\psi}\left(i\,\overleftarrow{\partial\!\!\!/} + m\right) = 0. \tag{2.64}$$

Acting with $\partial\!\!\!/$ on the first equation, we find that ψ satisfies the Klein–Gordon equation

$$\left(\partial^{2} + m^{2}\right)\psi = 0. \tag{2.65}$$

The canonical energy–momentum tensor is

$$\begin{aligned} T_{\mathrm{can}\,\lambda}{}^{\mu} &= -\frac{\partial\mathcal{L}}{\partial\partial_{\mu}\psi}\partial_{\lambda}\psi - \partial_{\lambda}\bar{\psi}\frac{\partial\mathcal{L}}{\partial\partial_{\mu}\bar{\psi}} + \eta_{\lambda}{}^{\mu}\mathcal{L} \\ &= -\frac{i}{2}\bar{\psi}\gamma^{\mu}\partial_{\lambda}\psi + \frac{i}{2}\partial_{\lambda}\bar{\psi}\gamma^{\mu}\psi + \eta_{\lambda}{}^{\mu}\left[\frac{1}{2}(i\bar{\psi}\,\partial\!\!\!/\psi - i\bar{\psi}\,\overleftarrow{\partial\!\!\!/}\,\psi) - 2m\bar{\psi}\psi\right], \end{aligned} \tag{2.66}$$

and it is clearly not symmetric. The spin-angular-momentum tensor is

$$\begin{aligned} S^{\mu}{}_{\rho\sigma} &= \frac{1}{2}\frac{\partial\mathcal{L}}{\partial\partial_{\mu}\psi}\Gamma_{\mathrm{s}}\left(M_{\rho\sigma}\right)\psi + \frac{1}{2}\bar{\psi}\Gamma_{\bar{\mathrm{s}}}\left(M_{\rho\sigma}\right)\frac{\partial\mathcal{L}}{\partial\partial_{\mu}\bar{\psi}} \\ &= \frac{1}{2}\frac{i}{2}\bar{\psi}\gamma^{\mu}\left(\frac{1}{2}\gamma_{\nu\rho}\right)\psi + \frac{1}{2}\bar{\psi}\left(-\frac{1}{2}\gamma_{\nu\rho}\right)\left(-\frac{i}{2}\gamma^{\mu}\psi\right) \\ &= \frac{i}{4}\bar{\psi}\gamma^{\mu}{}_{\nu\rho}\psi, \end{aligned} \tag{2.67}$$

and it is totally antisymmetric. The spin–energy potential is just

$$\Psi^{\mu\nu}{}_{\rho} = -S^{\mu\nu}{}_{\rho}, \tag{2.68}$$

[8] Our conventions for spinors and gamma matrices are explained in Appendix B.

and, after use of the equations of motion, we find the Belinfante tensor

$$T_\lambda{}^\mu = \frac{i}{4}\partial_\nu \bar{\psi}\,(\gamma^\mu \eta^\nu{}_\lambda + \gamma_\lambda \eta^{\nu\mu})\psi - \frac{i}{4}\bar{\psi}(\gamma^\mu \eta^\nu{}_\lambda + \gamma_\lambda \eta^{\nu\mu})\partial_\nu \psi$$
$$+\, \eta_\lambda{}^\mu \left[\tfrac{1}{2}(i\bar{\psi}\,\partial\!\!\!/\psi - i\bar{\psi}\,\overleftarrow{\partial\!\!\!/}\,\psi) - 2m\bar{\psi}\psi\right]. \tag{2.69}$$

In the case of the vector field, we managed to find the Belinfante tensor by a method based on the vector transformation law under GCTs. However, it is not clear how to use this method in the present case. The spinorial character is associated only with Lorentz transformations and it is not clear what the spinor transformation law should be for other GCTs. In fact, the only consistent form of dealing with spinors on curved spacetime is to treat them as scalars under GCTs and to associate the spinorial character with the Lorentz group that acts on the tangent space at each given point. This is the formalism invented by Weyl in [954] which we will study later on.

2.4.2 Dilatations

Let is consider now constant rescalings (dilatations) by a factor $\Omega = e^\sigma$:

$$\begin{array}{r} x'^\mu = \Omega x^\mu, \\ \varphi'(x') = \Omega^\omega \varphi(x), \end{array} \quad \Rightarrow \quad \left\{ \begin{array}{l} \tilde{\delta} x^\mu = \sigma x^\mu \equiv \sigma \tilde{\delta}_{\mathrm{D}} x^\mu, \\ \tilde{\delta}\varphi = \omega\sigma\varphi. \end{array} \right. \tag{2.70}$$

The associated conserved current is

$$j_{\mathrm{N1\,D}}{}^\mu = T_{\mathrm{can}}{}^\mu{}_\nu x^\nu + J^\mu, \qquad J^\mu \equiv \omega \frac{\partial \mathcal{L}}{\partial \partial_\mu \varphi}\varphi. \tag{2.71}$$

If we take the divergence of this current and set it equal to zero, we obtain the identity

$$T_{\mathrm{can}}{}^\mu{}_\mu + \partial_\mu J^\mu = 0. \tag{2.72}$$

It is always possible to find a redefinition of the canonical energy–momentum tensor that is symmetric, divergenceless, and, furthermore, traceless if there is scale invariance (see e.g. [204, 247, 491] and [781] and references therein). This redefined energy–momentum tensor is called the *improved energy–momentum tensor* and can be constructed systematically: on rewriting the dilatation current in the form

$$j_{\mathrm{N1\,D}}{}^\mu = \left[T_{\mathrm{can}}{}^\mu{}_\nu + \frac{2}{d-1}\partial_\rho\left(J^{[\mu}\eta^{\rho]}{}_\nu\right)\right]x^\nu - \frac{2}{d-1}\partial_\nu\left(J^{[\mu}x^{\nu]}\right), \tag{2.73}$$

we observe that

$$T^{\mu\nu} = T_{\mathrm{can}}{}^{\mu\nu} + \frac{2}{d-1}\partial_\rho\left(J^{[\mu}\eta^{\rho]\nu}\right), \tag{2.74}$$

is on-shell traceless on account of the identity Eq. (2.72) and also on-shell divergenceless since the piece that we add to the canonical energy–momentum tensor is of the form $\partial_\rho \Psi^{[\rho\mu]\nu}$. Observe that this term can also be obtained directly from the action if we add to it a total derivative term of the form

$$\Delta S = \frac{\omega}{d-1}\int d^d x\; \partial_\rho\left(\frac{\partial \mathcal{L}}{\partial \partial_\rho \varphi}\varphi\right). \tag{2.75}$$

Furthermore, the second term in the dilatation current is also of the form $\partial_\rho \Psi^{[\rho\mu]}$ and, then, up to this term we can write, as in Eqs. (2.54),

$$j_{\mathrm{N1\,D}}{}^{\mu} = T^{\mu}{}_{\nu}\tilde{\delta}_{\mathrm{D}} x^{\nu} = T^{\mu}{}_{\nu} x^{\nu}. \tag{2.76}$$

This result, and the analogous result for Lorentz rotations, suggest the following general picture: for *any given spacetime symmetry* generated by $\tilde{\delta} x^{\mu}$ it always seems possible to find a redefinition of the canonical energy–momentum tensor $T^{\mu\nu}$ such that it is symmetric and on-shell divergenceless and such that the conserved current associated with the spacetime symmetry is given, up to terms of the form $\partial_\rho \Psi^{[\rho\mu]}$, by

$$j_{\mathrm{N1}}{}^{\mu} = T^{\mu}{}_{\nu}\tilde{\delta} x^{\nu}. \tag{2.77}$$

It is to this energy–momentum tensor that gravity (a gauge theory for all spacetime transformations) couples. This immediately suggests Rosenfeld's prescription for finding the energy–momentum tensor. Before we study it, let us work out a couple of simple examples. First, let us consider a free scalar field φ in d dimensions with Lagrangian

$$\mathcal{L} = \tfrac{1}{2}(\partial\varphi)^2. \tag{2.78}$$

The action is invariant if $\omega = -(d-2)/2$. The canonical energy–momentum tensor and dilatation current are in this case

$$\begin{aligned} T_{\mathrm{can}}{}^{\mu}{}_{\nu} &= -\partial^{\mu}\varphi\partial_{\nu}\varphi + \tfrac{1}{2}\eta^{\mu}{}_{\nu}(\partial\varphi)^2, \\ j_{\mathrm{N1\,D}}{}^{\mu} &= T_{\mathrm{can}}{}^{\mu}{}_{\nu} x^{\nu} + \frac{\omega}{2}\partial^{\mu}\varphi^2. \end{aligned} \tag{2.79}$$

The improved energy–momentum tensor, is written in a form in which it is clear that we are adding a total derivative:

$$T^{\mu\nu} = T_{\mathrm{can}}{}^{\mu\nu} - \frac{\omega}{2(d-1)}\partial_\rho\big(\eta^{\rho(\mu}\partial^{\nu)}\varphi^2 - \eta^{\mu\nu}\partial^{\rho}\varphi^2\big). \tag{2.80}$$

Using the improved energy–momentum tensor, the dilatation current can be written as expected:

$$j_{\mathrm{N1\,D}}{}^{\mu} = T^{\mu}{}_{\nu} x^{\nu} + \frac{\omega}{2(d-1)}\partial_\nu\big(x^{\nu}\partial^{\mu}\varphi^2 - x^{\mu}\partial^{\nu}\varphi^2\big). \tag{2.81}$$

Our second example is a d-dimensional vector field whose action is invariant for the same value[9] of ω. In and only in $d = 4$ is the Belinfante tensor traceless. By the same procedure, we find the on-shell traceless, conserved energy–momentum tensor,

[9] This is true for any free-field theory described by a Lagrangian quadratic in first derivatives of the field.

in any dimension:

$$T^{\mu}{}_{\nu} = \frac{d}{2(d-1)} F^{\mu\rho} \partial_{\nu} A_{\rho} - \frac{1}{4(d-1)} \eta^{\mu}{}_{\nu} F^2 - \frac{d-2}{2(d-1)} \partial_{\nu} F^{\mu\sigma} A_{\sigma}. \tag{2.82}$$

This energy–momentum tensor changes under gauge transformations of the vector field.

2.4.3 Rosenfeld's energy–momentum tensor

Rosenfeld's prescription [813] is precisely based on the minimal coupling to gravity postulated by general relativity that we will study later on: place the matter fields in a curved background substituting everywhere the flat Minkowski metric $\eta_{\mu\nu}$ by a general background metric $\gamma_{\mu\nu}$, partial derivatives by covariant derivatives compatible with the background metric, and the flat volume element $d^d x$ by $d^d x \sqrt{|\gamma|}$. Then the energy–momentum tensor is given by

$$T^{\mu\nu}_{\text{matter}} = 2 \frac{\delta S_{\text{matter}}}{\delta \gamma_{\mu\nu}} \bigg|_{\gamma_{\mu\nu} = \eta_{\mu\nu}}. \tag{2.83}$$

Of course, one has to define first which fields are independent of the metric. For instance, if we have a vector field A_μ, we have to decide which of A_μ and A^μ is fundamental. The other field then depends on the metric used to raise or lower the index. Furthermore, we have to decide whether the fields are tensors or tensor densities, and, depending on our choice, we may have to add factors proportional to the determinant of the auxiliary metric or not and we may have to add additional connection terms in the covariant derivatives or not.

This energy–momentum tensor is symmetric by construction and conserved on-shell due to the Bianchi identity[10] associated with the invariance under GCTs of the action written in the background metric $\gamma_{\mu\nu}$. Furthermore, it can be shown to be always identical up to a term of the form $\partial_\rho \Psi^{[\rho\mu]\nu}$ to the canonical one under very general assumptions [63]. For a scalar and a vector field, the energy–momentum tensor found via Rosenfeld's prescription (the Rosenfeld or *metric* energy–momentum tensor) is identical to the canonical tensor and the Belinfante energy–momentum tensor, respectively. We will see in Chapter 3 that the same is true for a spin-2 field and later on we will see that the same is true in a generalized sense for a Dirac spinor. This identity is not a mere coincidence but it can be justified, as has already been pointed out, in the framework of the Cartan–Sciama–Kibble theory of gravity that we will review in Section 4.4. In general, it is easier to compute the energy–momentum tensor using Rosenfeld's prescription than using the canonical one, especially if we are interested in a *symmetric* energy–momentum tensor.

[10] This will be explained and proven later on in Chapter 3.

The Rosenfeld energy–momentum tensor has the required properties.[11,12]

To illustrate this point, let us go back to the massless vector field of the previous section. Let us consider the effect of conformal transformations on its action. The conformal group consists of transformations that leave the Minkowski metric invariant up to a global (possibly local) factor: (infinitesimal) constant translations $\tilde{\delta}x^\mu = \xi^\mu$, Lorentz rotations $\tilde{\delta}x^\mu = \sigma^\mu{}_\nu x^\nu \equiv \sigma^\mu$ (these two generate the Poincaré group), dilatations $\tilde{\delta}x^\mu = \sigma x^\mu \equiv w^\mu$, and special conformal transformations (or conformal boosts) $\tilde{\delta}x^\mu = 2(\zeta \cdot x)x^\mu - x^2\zeta^\mu \equiv v^\mu$. The vector field transforms under these coordinate transformations according to the general rule for (world) vectors (2.59) with $\epsilon^\mu = \xi^\mu + \sigma^\mu + w^\mu + v^\mu$. The variation of the action is, again, given by Eq. (2.60). Conformal transformations are generated by conformal Killing vectors of Minkowski spacetime that satisfy

$$\partial_{(\mu}\epsilon_{\lambda)} \propto \eta_{\mu\lambda}. \tag{2.87}$$

The proportionality factor is zero for Poincaré transformations but non-zero for dilatations and conformal boosts. Then, the variation of the action will be zero only if the energy–momentum tensor is traceless. This happens only in $d = 4$ dimensions. On integrating by parts, etc., we find that the Noether current has the form Eq. (2.77), always with the same (Rosenfeld's) energy–momentum tensor.

---

[11] However, this is still confusing because we have two different symmetric, on-shell divergenceless energy–momentum tensors for a scalar field (the canonical and the improved, which is traceless) and Rosenfeld's procedure seems to give a unique energy–momentum tensor. This is not true, though: when we covariantize a special-relativistic action introducing a metric the result is unique up to curvature terms that vanish in Minkowski spacetime. In the case of the scalar, a covariantization that preserves the scaling invariance is

$$S[\varphi, \gamma] = \int d^dx\sqrt{|\gamma|}\left[\frac{1}{2}(\partial\varphi)^2 + \frac{\omega}{4(d-1)}\varphi^2 R(\gamma)\right], \tag{2.84}$$

where $R(\gamma)$ is the Ricci scalar of the background metric. This action is invariant, in fact, under local Weyl rescalings of the metric and local rescalings of the scalar, leaving the coordinates untouched:

$$\varphi' = \Omega^{(2-d)/2}(x)\varphi, \qquad \gamma'_{\mu\nu} = \Omega^2(x)\gamma_{\mu\nu}. \tag{2.85}$$

Using the results of Section 4.2, we find

$$2\frac{\delta S[\varphi,\gamma]}{\delta\gamma_{\mu\nu}} = T_{\text{can}}{}^{\mu\nu} - \frac{\omega}{2(d-1)}\left(\nabla^\mu\partial^\nu\varphi^2 - \gamma^{\mu\nu}\nabla^2\varphi\right) + \frac{\omega}{2(d-1)}\varphi G^{\mu\nu}(\gamma), \tag{2.86}$$

where $G^{\mu\nu}(\gamma)$ is the Einstein tensor of the background metric. On setting $\gamma_{\mu\nu} = \eta_{\mu\nu}$, we find precisely the improved energy–momentum tensor Eq. (2.80). Something similar can be said of the vector field in $d \neq 4$. If, in the presence of a curved metric, the vector field scales as in Minkowski spacetime, the vector field is really a vector density and then its covariantization is different from the standard one and should lead to a Rosenfeld energy–momentum tensor identical to the improved one.

[12] When the field theory has a symmetry, it is desirable or necessary to have an energy–momentum tensor that is also invariant under the same transformations. For instance, the Belinfante energy–momentum tensor for the Maxwell field is gauge-invariant, as is the Maxwell action. It can be shown that, in general, symmetries of a theory are also symmetries of the Rosenfeld energy–momentum tensor if the symmetries are also symmetries of the same theory covariantized with an arbitrary background metric. The Maxwell action in a curved background is still gauge-invariant and the gauge-invariance of the Belinfante–Rosenfeld energy–momentum tensor follows.

We may expect that this is completely general. We need to know only how the fields transform under general coordinate transformations,[13] which determines completely the coupling to gravity in general relativity.

Just as it is possible to give a prescription for how to find the energy–momentum tensor on the basis of its coupling to gravity through the metric in general relativity, it is possible to give a definition of the spin–energy potential $\Psi^{\mu\nu}{}_{\rho}$ based on its coupling to gravity (maybe we should say geometry instead of gravity) through the torsion tensor in the framework of the Cartan–Sciama–Kibble (CSK) theory:

$$\Psi^{\mu\nu}{}_{\rho} = -\frac{2}{\sqrt{|g|}} \frac{\delta S}{\delta T_{\mu\nu}{}^{\rho}}\bigg|_{\gamma=T=0}. \tag{2.88}$$

The equivalence of this definition and the definition we gave in terms of the spin-angular-momentum tensor $S^{\mu}{}_{\rho\sigma}$ can also be proven in the CSK theory. In fact, the above definition is the main characteristic of that theory in which intrinsic (i.e. not orbital) angular momentum is the source of another field that has a geometrical interpretation (torsion).

2.5 The Noether method

There is a useful recipe for how to find the Noether current associated with global symmetry transformations of the fields $\delta\varphi$: if the action is invariant under transformations with constant parameters, then, if we use local parameters, upon use of the equations of motion, the variation of the action would be proportional to the derivative of the parameters:

$$\delta S = -\int d^d x \partial_\mu \sigma^I \, j_I^\mu, \tag{2.89}$$

because, by hypothesis, it has to vanish for constant σ^I. Up to a total derivative, this is

$$\delta S = \int d^d x \sigma^I \, \partial_\mu j_I^\mu, \tag{2.90}$$

that vanishes for constant σ^I only if $\partial_\mu j_I^\mu = 0$. Thus the currents j_I^μ are the Noether currents associated with the global symmetry.

The observation that the variation of the action must be of the above form is the basis of the so-called *Noether method* which is used to couple fields in a symmetric way. The simplest example of how this method works is the coupling of a complex scalar field Φ to the electromagnetic field A_μ. The Lagrangian of the electromagnetic field Eq. (2.56) is invariant under the transformations with local parameter Λ,

$$\delta A_\mu = \partial_\mu \Lambda, \tag{2.91}$$

while the Lagrangian for the complex scalar,

$$\mathcal{L} = \tfrac{1}{2}\partial_\mu \Phi \partial^\mu \bar{\Phi}, \tag{2.92}$$

[13] A conformal scalar of weight ω is nothing but a scalar density of weight ω/d.

is invariant under phase transformations with a constant parameter σ and a constant g that infinitesimally look like this:

$$\delta\Phi = ig\sigma\Phi. \tag{2.93}$$

These transformations constitute a U(1) symmetry group. g labels the representation of U(1) corresponding to Φ. If σ takes values in the interval $[0, 2\pi]$, then g can be any integer.

If the conserved current of the scalar Lagrangian is seen as an electric current, it is natural to couple it to the electromagnetic vector field to obtain the Maxwell equation with sources:

$$\partial_\nu F^{\nu\mu} = g j_{\mathrm{N}}^{\mu}. \tag{2.94}$$

From a Lagrangian point of view, this equation can be obtained by adding to the free Lagrangians of the vector and scalar a coupling of the form $gA_\mu j_{\mathrm{N}}^{\mu}$. However, this term modifies the equation of motion of the scalar, so the electric current j_{N}^{μ} is not conserved on-shell. This renders the above equation inconsistent since the l.h.s. is automatically divergenceless. Clearly, the addition of a new term to the Lagrangian modifies the Noether current. The modified Noether current should be conserved on-shell upon use of the modified equations of motion. It is easy to see that the vector field contributes to it. This is the Noether current that we should use in the Lagrangian now, and this induces new modifications. This may go on indefinitely until the new correction does not contribute to the new Noether current. Observe that the modified Noether current is found using a local phase transformation according to the above general observation. It should also be stressed that the physical reason why there was inconsistency is that we did not take into account the contribution of the vector field to the electric current. Only the total electric current should be consistently conserved.

The Noether method is essentially a systematic way of performing these iterations emphasizing the role of symmetry. In the case at hand, the basic idea is that one has to identify σ with Λ and one has to make the whole system invariant under transformations of the same form with Λ local. We start by calling $\mathcal{L}_0$ the Lagrangian which is the sum of the free electromagnetic and scalar Lagrangians and using the above general observation: under a local Λ transformation ($\sigma = \Lambda$), and up to total derivatives,

$$\delta\mathcal{L}_0 = g\Lambda\partial_\mu j_{\mathrm{N}}^{\mu}, \qquad j_{\mathrm{N}}^{\mu} = -\frac{i}{2}\left(\Phi\partial^\mu\bar{\Phi} - \bar{\Phi}\partial^\mu\Phi\right). \tag{2.95}$$

j^μ is the on-shell conserved current associated with the global invariance of the Lagrangian.

The Noether method consists in the addition to $\mathcal{L}_0$ of terms that will be of higher order in the constant g to compensate for the above non-vanishing variation. Typically the first correction will be of the form

$$\mathcal{L}_1 = \mathcal{L}_0 + gA_\mu j_{\mathrm{N}}^{\mu}. \tag{2.96}$$

The additional term cancels out the variation of $\mathcal{L}_0$ but generates, due to the variation of the Noether current itself, another term of order $\mathcal{O}(g^2)$. Up to total derivatives

$$\delta\mathcal{L}_1 = -g^2|\Phi|^2 A_\mu\partial^\mu\Lambda. \tag{2.97}$$

This variation can be exactly canceled out by

$$\mathcal{L}_2 = \mathcal{L}_1 + \tfrac{1}{2}g^2|\Phi|^2 A^2, \tag{2.98}$$

which can be rewritten in the standard, manifestly gauge-invariant form

$$\begin{aligned} \mathcal{L}_2 &= -\tfrac{1}{4}F^2 + \tfrac{1}{2}\mathcal{D}_\mu\Phi\mathcal{D}^\mu\bar{\Phi}, \\ \mathcal{D}_\mu\Phi &= (\partial_\mu - igA_\mu)\Phi. \end{aligned} \tag{2.99}$$

A more interesting example is provided by a set of r vector fields $A^I{}_\mu$ with Lagrangian

$$\mathcal{L}_0 = +\tfrac{1}{4}g_{IJ}f^I{}_{\mu\nu}f^{J\,\mu\nu}, \qquad f^I{}_{\mu\nu} = 2\partial_{[\mu}A^I{}_{\nu]}. \tag{2.100}$$

Here g_{IJ} is a negative-definite constant metric. This Lagrangian is evidently invariant under r local gauge transformations,

$$\delta_\Lambda A^I{}_\mu = \partial_\mu\Lambda^I, \tag{2.101}$$

because the field strengths are. It is less evident, but equally true, that the above Lagrangian is invariant under global transformations that form a group G of dimension d such that the Killing metric of the associated Lie algebra[14] is precisely g_{IJ}. Under these transformations, the vector fields transform in the adjoint representation, that is, infinitesimally:

$$\delta_\sigma A^I{}_\mu = g\sigma^K\Gamma_{\text{Adj}}(T_K)^I{}_J A^J{}_\mu, \qquad \Gamma_{\text{Adj}}(T_K)^I{}_J = f_{KJ}{}^I. \tag{2.102}$$

These two symmetries form a closed symmetry algebra:

$$[\delta_\Lambda, \delta_\sigma] = \delta_{\Lambda'}, \qquad \Lambda^{I\,\prime} = \sigma^K\Gamma_{\text{Adj}}(T_K)^I{}_J\Lambda^J. \tag{2.103}$$

There are d conserved Noether currents $j^\mu_{\text{N}\,I}$ associated with the global invariance of G. Following the general argument, they can be found by performing a local G transformation:

$$\begin{aligned} \delta_{\sigma(x)}\mathcal{L}_0 &= g\sigma^K\partial_\mu j^\mu_{\text{N}\,I}, \\ j^\mu_{\text{N}\,I} &= f_{IJK}f_{K\,\mu}{}^\nu A^J{}_\nu. \end{aligned} \tag{2.104}$$

Let us now consider the coupling of d conserved currents $j_I{}^\mu$ associated with some other set of matter fields invariant under global G transformations to the vector fields. As in the Maxwell case, we add to the action the terms $gA^I{}_\mu j_I{}^\mu$ and find that the currents are no longer conserved on-shell because the equations of motion of the fields have changed due to the new coupling term. As in the Maxwell case, the problem is that we have not taken into account all the sources of charge, since only the total charge associated with invariance of G will be conserved once the coupling has been introduced. Thus, we should couple the vector fields to their own Noether currents. We can forget about the matter fields now and try to solve the self-consistency problem of the coupling of the vector fields to themselves by use of the Noether method.

Since Noether currents our found via local G transformations, we look for invariance under local G transformations of $\mathcal{L}_0$. To cancel out $\delta_{\sigma(x)}\mathcal{L}_0$ we have to do two things: first, we have to identify $\Lambda^I = \sigma^I$ and then we have to introduce a correction that is of first order in g into the Lagrangian that takes the characteristic form

$$\mathcal{L}_1 = \mathcal{L}_0 + \frac{g}{2}A^I{}_\mu j^\mu_{\text{N}\,I}. \tag{2.105}$$

[14] See Appendix A for notation and conventions.

In this way, by enforcing local symmetry we arrive at the same conclusion as before by physical arguments: we have to add the self-coupling term. This makes sense if the algebra of the new transformations,

$$\delta_\sigma A^I{}_\mu = \partial_\mu \sigma^I + g\sigma^K \Gamma_{\mathrm{Adj}}(T_K)^I{}_J A^J{}_\mu, \tag{2.106}$$

closes, as is the case. The new term in the Lagrangian produces a new term of second order in g in the transformation:

$$\delta_\sigma \mathcal{L}_1 = g^2 f_{IJK} f_{ML}{}^K A^I_\mu A^J{}_\nu A^{L\,\nu} \partial^\mu \sigma^M, \tag{2.107}$$

which can be exactly canceled out by the addition of an $\mathcal{O}(g^2)$ term that finishes the iterative procedure:

$$\mathcal{L}_2 = \mathcal{L}_1 - \frac{g^2}{4} f_{IJK} f_{ML}{}^K A^I{}_\mu A^J{}_\nu A^{L\,\nu} A^{M\,\mu}. \tag{2.108}$$

This Lagrangian can be written in the standard, manifestly gauge-invariant form

$$\begin{aligned} \mathcal{L}_2 &= \tfrac{1}{4} g_{IJ} F^I{}_{\mu\nu} F^{J\,\mu\nu}, \\ F^I{}_{\mu\nu} &= f^I{}_{\mu\nu} + g f_{JK}{}^I A^J{}_\mu A^K{}_\nu. \end{aligned} \tag{2.109}$$

It is customary to use dimensionless gauge parameters. On rescaling $\sigma^I \to \sigma^I/g$ we recover the gauge transformations in the conventions of Appendix A.

In more complicated cases the Noether procedure will require the addition of more corrections both to the Lagrangian and to the field-transformation rules (see e.g. [912]). The procedure is simplified considerably by using first-order actions [299, 300]. Only in this way is it possible to find all the corrections to the Fierz–Pauli Lagrangian. This is explained in Section 3.2.7.

3

A perturbative introduction to general relativity

The standard approach to general relativity (GR) is purely geometrical: spacetime is curved by its energy content according to Einstein's equation and test particles move along geodesics. This point of view is what makes GR a theory completely different from the theories that describe all the other known interactions that are special-relativistic field theories (SRFTs) that, after quantization, explain the interaction between two charged bodies as the interchange of quanta of the field.

The enormous success of relativistic quantum field theories with a gauge principle made it unavoidable to try to find a theory of that kind to describe gravitational interactions at a classical and quantum level. This path was followed by many people and it was found that such a theory, whose starting point is the linear perturbation theory of GR (the Fierz–Pauli theory for a free, massless spin-2 particle), would be self-consistent only after the introduction of an infinite number of non-linear terms whose summation should be equivalent to the full non-linear GR theory.[1] Thus, this approach may lead to a different justification of Einstein's theory and provides an alternative interpretation of it that is worth studying.[2] Some of the predictions of GR can be obtained at leading or next to leading order in this approach. Since this is not the standard approach, there are only a few complete treatments in the literature: the book [386], based on Feynman's lectures on gravitation, that also contains many references, some of which we will follow in Section 3.2; and also Deser's lectures on the gravitational field [300]. Reference [30] is also an excellent review with many references.

In this chapter, as a warm-up exercise, we are first going to study the construction of SRFTs of gravity based on a scalar field. This is the simplest possibility in the search for a SRFT of the gravitational interaction and it will offer us the possibility of studying, in a simple setting, problems that we will find later on.

As is well known, scalar theories of gravity predict no global bending of light rays (in contrast to observation) and a value for the precession of the perihelion of Mercury which

[1] There are other alternative special-relativistic field theories for spin-2 particles. See, for example, [659] in which gravity is based on a *massive* (with extremely small mass) spin-2 field.

[2] Some string theories have a massless spin-2 particle in their spectra. If these string theories are consistent, the argument we will develop will imply that they contain gravity, which, to the lowest order, will be described by Einstein's theory.

is also wrong (in magnitude and sign) and thus we will have to consider the next logical possibility: a spin-2 field. First, we will have to find a SRFT (the Fierz–Pauli theory) for the free spin-2 field. Gauge invariance plays a crucial role in the construction of this theory and we will emphasize it. We will then proceed to introduce the interaction with matter fields and find the gravitational field produced by a massive point particle. We will immediately show that the interacting theory will be consistent (at the classical level) only if the gravitational field couples to itself in the same form as that in which it couples to matter: through the energy–momentum tensor. Making this self-coupling consistent requires an infinite number of corrections to the Fierz–Pauli theory. We will try to find the first correction via the Noether method, meeting the first difficulties in the definition of the gravitational energy–momentum tensor, of which we will have more to say in Chapter 6. The choice of energy–momentum tensor, which is usually defined up to the divergence of an antisymmetric tensor or up to the addition of on-shell-vanishing terms, is crucial in this context, because different choices lead to different theories with different predictions of the value for the precession of the perihelion of Mercury [725].

These problems are avoided by the use of Deser's argument that allows one to find in just one step both the right energy–momentum tensor for the gravitational field at lowest order and all the corrections to the Fierz–Pauli theory that convert it into a self-consistent theory for a self-interacting massless spin-2 particle. This theory is just GR. We will discuss whether this is the only possible solution to our problem, since Deser's result shows the existence of a solution but not its uniqueness.

In any case, this is how we are going to introduce the Einstein equations and the Einstein–Hilbert action that will be studied in more detail in Chapter 4 and also the action for point-particles moving in a curved background. We will conclude the chapter by studying the perturbative expansion of GR (i.e. the interacting Fierz–Pauli theory consistent to a certain order in the coupling constant) in flat and curved backgrounds for later use.

3.1 Scalar SRFTs of gravity

If we were particle physicists in the pre-Yang–Mills[3] era wanting to describe gravity, we would certainly try to do it (Feynman in [386] or Thirring in [888]) with a relativistic field theory of a bosonic massless particle (to provide long-range interactions) propagating in Minkowski spacetime whose interchange would be responsible for the gravitational interaction between massive bodies. Which particle? The simplest possibility is that of a scalar particle (after all, in Newtonian physics, gravity is described by the Newtonian gravitostatic potential ϕ alone and there was no hint of the existence of any *gravitomagnetic* field) and, for this reason and considering the attractive nature of scalar-mediated interactions (see, for instance, [867]), scalar SRFTs were the first candidates used to describe relativistic gravitation.[4]

[3] A different approach to gravity based on the gauge theory of the Poincaré and (anti-)de Sitter groups is also possible and is described in Chapter 4.5.

[4] Scalar theories of gravity were first proposed by Abraham [4–11], Nordström [729–34], and Einstein [352–4]. (Some old reviews are [12, 635, 645], and a modern review is [736].) They played an important role in the developments that led Einstein to GR. Our interest in them is purely pedagogical.

A free scalar propagating in Minkowski spacetime is described by the action

$$S = \int d^d x \ \tfrac{1}{2}(\partial\phi)^2, \qquad (\partial\phi)^2 \equiv \eta^{\mu\nu}\partial_\mu\phi\partial_\nu\phi, \tag{3.1}$$

and has as equation of motion

$$\partial^2\phi = 0, \qquad \partial^2 \equiv \eta^{\mu\nu}\partial_\mu\partial_\nu. \tag{3.2}$$

The source for the Newtonian gravitational field is the *gravitational mass of matter* which is experimentally found to be proportional (equal in appropriate units) to the inertial mass for all material bodies. In special relativity the inertial mass, the energy, and the momentum of a physical system are combined into the energy–momentum tensor $T^{\mu\nu}$ and, therefore, the source for the gravitational field will be the matter energy–momentum tensor. This is an object of utmost importance and was studied in some detail in Chapter 2.

3.1.1 Scalar gravity coupled to matter

From our previous discussion, the source of the scalar gravitational field (the r.h.s. of Eq. (3.2)) must be a scalar built out of the energy–momentum tensor of the matter fields. The simplest scalar is the trace $T_{\text{matter}} \equiv T_{\text{matter}}{}^{\mu}{}_{\mu}$, and using it, and taking into account all factors of c, we arrive at the action for matter coupled to scalar gravity

$$S = \frac{1}{c}\int d^d x \left\{\frac{1}{2Cc^2}(\partial\phi)^2 + \frac{\phi}{c^2}T_{\text{matter}} + \mathcal{L}_{\text{matter}}\right\}, \tag{3.3}$$

where C is a proportionality constant to be determined. From this action we can derive the equation of motion for the scalar gravitational field,

$$\partial^2\phi = C T_{\text{matter}}, \tag{3.4}$$

and the equation of motion for matter in the gravitational field.

Observe that the conservation of the matter energy–momentum tensor plays no role whatsoever in the construction of this theory. In fact, if it was required in some sense for consistency, we would be in trouble because, after the coupling to the gravitational field, the matter energy–momentum tensor is no longer conserved: only the *total* energy–momentum tensor of the above Lagrangian (the matter energy–momentum tensor, plus the gravitational energy–momentum tensor, plus an interaction term) is conserved. However, the equation of motion that we have obtained is perfectly consistent as it stands.

Observe also that nowhere is it required that the energy–momentum tensor is symmetric (although only its symmetric part contributes to the trace). In fact, there are no conditions that we can impose on the energy–momentum tensor to select only one out of the infinitely many possible energy–momentum tensors that we can obtain by adding terms proportional to the equations of motion or superpotential terms. We can view this as a weakness of scalar SRFTs of gravity. In the cases that we are going to consider, we will simply take the canonical energy–momentum tensor obtained from the matter action in its simplest form.

Now, to determine the constant C, we can require ϕ to be identical to the Newtonian gravitational potential in the static, non-relativistic limit in which only the $T_{\text{matter}\,00} = -\rho c^2$

component contributes to the trace, ρ being the mass density.[5] In this case, the above equation becomes the Poisson equation,

$$\partial_i \partial_i \phi = C c^2 \rho, \quad \Rightarrow C = \frac{(d-3)8\pi G_N^{(d)}}{(d-2)c^2}, \tag{3.5}$$

where $G_N^{(d)}$ is the d-dimensional Newton constant.[6] For a point-particle of mass M at rest at the origin

$$\rho = M \delta^{(d-1)}(\vec{x}_{d-1}), \tag{3.6}$$

and

$$\phi = -\frac{16\pi G_N^{(d)} M}{2(d-2)\omega_{(d-2)}} \frac{1}{|\vec{x}_{d-1}|^{d-3}}. \tag{3.7}$$

This identification will be completely justified if, in the limit considered, ϕ affects the motion of matter just as the Newtonian gravitational potential does. Let us consider the motion of a massive particle in the gravitational field ϕ. The coupling is given by the above action. All we need is the action for the free special-relativistic massive point-particle. Since we are going to make extensive use of this action, we start by reviewing it.

3.1.2 The action for a relativistic massive point-particle

The special-relativistic action for a point-particle of mass M can be written as follows:

$$\boxed{S_{\text{pp}}[X^\mu(\xi)] = -Mc \int d\xi \sqrt{\eta_{\mu\nu} \dot{X}^\mu \dot{X}^\nu}, \qquad \dot{X}^\mu \equiv \frac{dX^\mu}{d\xi},} \tag{3.8}$$

where ξ is a general parameter for the particle's worldline. The reality of the action is related to the fact that usual massive particles move along timelike curves, $\dot{X}^\mu \dot{X}_\mu > 0$. The equations of motion that one derives from it simply express the conservation of the d components of the linear momentum:

$$\frac{dP_\mu}{d\xi} = 0, \qquad P_\mu \equiv \frac{\partial L}{\partial \dot{X}^\mu} = -Mc \frac{\eta_{\mu\nu} \dot{X}^\nu}{\sqrt{\eta_{\rho\sigma} \dot{X}^\rho \dot{X}^\sigma}}. \tag{3.9}$$

The conservation of the $d(d-1)/2$ components of the angular momentum,

$$M_{\mu\nu} = 2X_{[\mu} P_{\nu]}, \tag{3.10}$$

follows. The $d(d+1)/2$ conserved quantities are, as is well known, associated with the invariance of the action under global Poincaré transformations of the spacetime coordinates

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu + a^\mu, \qquad \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta \eta_{\mu\nu} = \eta_{\alpha\beta}, \tag{3.11}$$

[5] Notice the minus sign in our conventions.

[6] This is an unfortunate convention in the literature in which the factor 4π, which is appropriate for rationalized units in four dimensions, is indiscriminately used in all dimensions.

via the Noether theorem for global transformations: using the infinitesimal form of the Poincaré transformations,

$$\delta x^\mu = \sigma^\mu{}_\nu x^\nu + \sigma^\mu, \qquad \sigma^{\mu\nu} = -\sigma^{\nu\mu}, \tag{3.12}$$

we obtain the conservation law

$$\frac{dJ(\sigma)}{d\xi} = 0, \qquad J(\sigma) = P_\mu \delta X^\mu. \tag{3.13}$$

The conserved quantity associated with translations is the linear momentum $J(\sigma^\mu) \sim P^\mu$ and the conserved quantity associated with Lorentz transformations is the angular momentum $J(\sigma^{\mu\nu}) \sim M^{\mu\nu}$.

Observe that the invariance of the action is due to the fact that it depends only on the derivatives of the coordinates. In particular, the Minkowski metric does not depend on the coordinates. A better way to express this fact is to say that the Minkowski metric has $d(d+1)/2$ independent isometries that generate the d-dimensional Poincaré group. This association between spacetime isometries and conserved quantities will still hold in more complicated spacetimes.

This action is also invariant under non-singular reparametrizations of the worldline $\xi'(\xi)$. These are local (gauge) transformations that infinitesimally can be written $\delta\xi = \epsilon(\xi)$. Taking into account that the X^μs are scalars with respect to these transformations, we find

$$\delta\xi = \epsilon(\xi), \qquad \delta d\xi = \dot{\epsilon} d\xi, \qquad \tilde{\delta} X^\mu = 0, \qquad \delta \dot{X}^\mu = -\dot{\epsilon}\dot{X}^\mu, \tag{3.14}$$

and it is a simple exercise to check that $\tilde{\delta} S = 0$ identically. If we now consider the variation of the action under just

$$\delta X^\mu = -\epsilon \dot{X}^\mu, \tag{3.15}$$

we find that it is invariant only up to a total derivative

$$\delta S = \int d\xi \, \frac{d}{d\xi}\left(Mc\epsilon \sqrt{\eta_{\mu\nu}\dot{X}^\mu \dot{X}^\nu} \right). \tag{3.16}$$

On varying the action with respect to general variations of the coordinates first and integrating by parts, we obtain

$$\delta S = \int d\xi \, \left\{ \epsilon \frac{\delta S}{\delta X^\nu} \dot{X}^\nu + \frac{d}{d\xi}\left(Mc\epsilon \sqrt{\eta_{\mu\nu}\dot{X}^\mu \dot{X}^\nu} \right) \right\}. \tag{3.17}$$

By equating the two results and taking into account that the equation is valid for arbitrary functions $\epsilon(\xi)$, we obtain the gauge identity

$$\frac{\delta S}{\delta X^\nu}\dot{X}^\nu = 0, \qquad \Rightarrow \dot{P}_\nu \dot{X}^\nu = 0, \tag{3.18}$$

which is satisfied off-shell (trivially on-shell). Since $\dot{X}^\nu$ is proportional to the momentum, this identity is proportional to

$$\frac{d(P^\mu P_\mu)}{d\xi} = 0. \tag{3.19}$$

Indeed, $P^\mu P_\mu$ is a constant: using the definition of momentum, we find, without using the equations of motion, the mass-shell condition

$$P^\mu P_\mu = M^2c^2, \tag{3.20}$$

and we have just shown that this constraint can be understood as a consequence of reparametrization invariance.

There are two special parameters one can use.[7] One is the particle's proper time (or length) $\xi = s$, defined by the property

$$\eta_{\mu\nu}\dot{X}^\mu \dot{X}^\nu = 1. \tag{3.21}$$

Owing to this definition, the action is usually written as

$$S_{\text{pp}}[X(s)] = -Mc\int ds. \tag{3.22}$$

Although this form is unsuitable for finding the equations of motion, it tells us that the action of a massive point-particle is proportional to its worldline's proper length, and the minimal-action principle tells us that the particle moves along worldlines of minimal proper length. Observe that, from the quantum mechanics point of view, since the measure in the path integral is the exponential of

$$\frac{i}{\hbar}S = i\frac{Mc}{\hbar}\int ds = \frac{i}{\lambda\!\!\!^{-}_{\text{Compton}}}\int ds, \tag{3.23}$$

the proper length is measured in units of the particle's reduced Compton wavelength.

The second special parameter that we can use is the coordinate time $\xi = X^0 = cT$. This choice of gauge fixes one of the particle's coordinates $X^0(\xi) = \xi$. In this gauge (*The physical or static gauge*) one can study the non-relativistic limit $\dot{X}^i\dot{X}^i = (v/c)^2 \ll 1$. In this limit the action (3.8) becomes, up to a total derivative, the non-relativistic action of a particle:

$$S[X^i(t)] = \int dt\left[\tfrac{1}{2}Mv^2 - Mc^2\right]. \tag{3.24}$$

3.1.3 The massive point-particle coupled to scalar gravity

The coupling to the scalar gravitational field is dictated by the action Eq. (3.3). We compute the energy–momentum tensor using Rosenfeld's prescription (Section 2.4.3):

$$T_{\text{pp}}^{\mu\nu}(x) = -Mc^2\int d\xi \frac{\dot{X}^\mu\dot{X}^\nu}{\sqrt{\eta_{\rho\sigma}\dot{X}^\rho\dot{X}^\sigma}}\delta^{(d)}[X(\xi) - x], \tag{3.25}$$

which is conserved, as one can prove by using the equations of motion. The trace is identical to the Lagrangian,[8] and thus the action for the coupled particle-plus-gravity system

[7] Purists call the same curve with two different parametrizations different curves, but from a physical point of view they are clearly the same object.

[8] Observe that, in the static gauge, the 00 component of this tensor gives Eq. (3.6).

Eq. (3.3) becomes

$$S[\phi(x), X^{\mu}(\xi)] = \frac{1}{Cc^3}\int d^dx\, \tfrac{1}{2}(\partial\phi)^2 - Mc\int d\xi \left(1+\frac{\phi(X)}{c^2}\right)\sqrt{\eta_{\mu\nu}\dot{X}^{\mu}\dot{X}^{\nu}}. \qquad (3.26)$$

For low speeds, in the static gauge, the second term is

$$\sim \int dt\, \{\tfrac{1}{2}Mv^2 - M\phi - Mc^2\}, \qquad (3.27)$$

which confirms the consistency of our identification of ϕ with the Newtonian potential in this limit. The complete relativistic action predicts corrections to the Newtonian theory. The next two terms in the expansion of the relativistic action are

$$\int dt\, \{-\tfrac{1}{4}Mv^2(v/c)^2 + \tfrac{1}{2}Mv^2\phi/c^2\}. \qquad (3.28)$$

The second term is there also for free particles, but the third represents a relativistic correction to the Newtonian coupling to the gravitational field. Owing to its sign, if the particle that acts as source for the scalar gravitational field moves, the kinetic energy contributes to T_{pp} with sign opposite to the rest mass and a particle in motion produces (and, therefore, feels) a weaker gravitational field than when it is at rest. The gravitational field, in fact, would vanish in the limit in which the particle moves at the speed of light. This also means that the gravitational field will not affect the motion of particles moving at the speed of light.

Let us now consider the motion of a second massive particle in the scalar gravitational field produced by the first particle. Although ϕ is identical to the Newtonian potential, the action (just the last term in Eq. (3.26) with ϕ given by Eq. (3.7)) also predicts corrections to the Newtonian motion. We will not enter into details, but it can be shown [108] that the lowest-order correction to the Newtonian orbits of planets is a precession of their perihelion which is a factor $-\frac{1}{6}$ of that predicted by GR (which is experimentally confirmed). This is a clear drawback for the scalar SRFT of gravity.

With a SRFT of gravity we can also study the effect of gravity on massless particles or the gravitational field produced by massless particles, which is impossible in Newtonian gravity. Thus, there is no non-relativistic limit for this problem. First, we need to find an action for a massless particle.

3.1.4 The action for a massless point-particle

Clearly, the action (3.8) (from now on referred to as a *Nambu–Goto-type action*[9]) is not well suited to take the $M \to 0$ limit. Furthermore, in spite of the straightforward physical interpretation of the Nambu–Goto-type action, the square root makes it highly non-linear

[9] The origin of this action can be traced back to Planck. However, the generalization of this action to one-dimensional objects was proposed by Nambu and Goto in [714] and [463], respectively, and has inspired further generalizations for higher-dimensional objects. Hence it has become customary to refer to these kinds of actions as *Nambu–Goto-type* actions.

and it would be desirable to have a different, more linear, action giving the same equations of motion.

Thus, we are going to propose an equivalent action that we will call a *Polyakov-type action* with a new, independent, dimensionless, auxiliary "field[10]" γ that can be interpreted as a metric on the worldline. This action is

$$\boxed{S_{\mathrm{pp}}[X^\mu(\xi), \gamma(\xi)] = -\tfrac{1}{2}Mc\int d\xi \sqrt{\gamma}\left[\gamma^{-1}\eta_{\mu\nu}\dot{X}^\mu\dot{X}^\nu + 1\right].} \tag{3.29}$$

This action is, yet again, invariant under Poincaré transformations of the spacetime coordinates and invariant under reparametrizations of the worldline $\xi \to \xi'(\xi)$ under which γ transforms as follows:

$$\gamma(\xi) = \gamma'\left[\xi'(\xi)\right]\left(\frac{d\xi'}{d\xi}\right)^2. \tag{3.30}$$

The equation of motion of γ is a constraint that simply tells us that γ is, on-shell, the *induced metric* on the worldline,

$$\gamma = \eta_{\mu\nu}\dot{X}^\mu\dot{X}^\nu. \tag{3.31}$$

This equation is purely algebraic and can be substituted into the action to eliminate[11] γ, resulting in the Nambu–Goto-type action Eq. (3.8).

Although equivalent, this action is, however, more versatile: we can obtain from it an action for a massless particle. For this we first have to rescale γ to $\gamma' = M^{-2}c^{-2}\gamma$ and then we can take the limit $M \to 0$. We rescale back to obtain a dimensionless worldline metric $\gamma' = p^{-2}\tilde{\gamma}$ (obviously $\tilde{\gamma}$ cannot be identified with the original γ), giving

$$\boxed{S[X^\mu(\xi), \tilde{\gamma}(\xi)] = -\frac{p}{2}\int d\xi \sqrt{\tilde{\gamma}}\,\tilde{\gamma}^{-1}\eta_{\mu\nu}\dot{X}^\mu\dot{X}^\nu,} \tag{3.32}$$

where p is a constant with dimensions of momentum. In the path integral now the action (which is no longer the proper length) is measured in de Broglie's wavelength units $p/\hbar = 1/\lambda\!\!\!^{-}_{\mathrm{deBroglie}}$ associated with the characteristic momentum p.

Now the equation of motion for $\tilde{\gamma}$ states that the particle's worldline is light-like:

$$\eta_{\mu\nu}\dot{X}^\mu\dot{X}^\nu = 0, \tag{3.33}$$

but this equation cannot be used to eliminate $\tilde{\gamma}$ from the action as in the massive case.

By definition, the proper length of a massless particle's worldline is always zero and cannot be used to parametrize it, but the coordinate time can be used for that purpose.

[10] Just a dynamical variable (not a field) of the worldline parameter in the zero-dimensional (point-like) case. This was first done in [189, 316] for strings. Our discussion follows closely those of standard string-theory references. See e.g. [39, 473, 609, 673, 779] and also Section 14.1.

[11] It is guaranteed that, under these conditions, the equations of motion derived from the resulting action are the same equations as those one would obtain from the elimination of γ from the original equations of motion.

3.1.5 The massless point-particle coupled to scalar gravity

We can now try to couple this action to gravity, which is impossible in the Newtonian theory. The energy–momentum tensor is

$$T_{\rm pp}^{\mu\nu} = -pc \int d\xi \sqrt{\gamma}\, \gamma^{-1} \dot{X}^{\mu} \dot{X}^{\nu} \delta^{(d)}[X(\xi) - x]. \tag{3.34}$$

On taking the trace and substituting into Eq. (3.3) we immediately realize that we can make the coupling to gravity disappear by rescaling the worldline auxiliary metric γ with a factor $(1 + \phi(X)/c^2)^{-\frac{1}{2}}$. In other words: there is no coupling of a massless particle to scalar gravity. This was to be expected: we have already mentioned the weakening of the scalar gravitational interaction of a massive particle when we increase the speed. On the other hand, the trace of the energy–momentum tensor of a massless particle above vanishes on-shell.

We know, however, that the light of stars passing near the Sun is bent by its gravitational field. This is the second drawback of this theory.

We could also have used the Maxwell action and the energy–momentum tensor

$$S_{\rm matter} = \frac{1}{c} \int d^d x \,\{-\tfrac{1}{4} F^2\}, \qquad T_{{\rm matter}\,\mu\nu} = F_{\mu}{}^{\rho} F_{\nu\rho} - \tfrac{1}{4} \eta_{\mu\nu} F^2, \tag{3.35}$$

to study the coupling of the scalar gravitational field to massless particles (fields). On taking the trace and substituting into Eq. (3.3) we find the action

$$S = \frac{1}{c} \int d^d x \left\{ \frac{1}{2Cc^2} (\partial\phi)^2 - \frac{1}{4} \left[1 + \frac{d-4}{4} \phi/c^2 \right] F^2 \right\}. \tag{3.36}$$

In $d = 4$ (but only in $d = 4$!) the Maxwell energy–momentum tensor is traceless and there is no coupling to the scalar gravitational field, as expected. In other dimensions, though, there is interaction, in contradiction with the absence of gravitational interaction for massless particles. This apparent paradox can be avoided by the use of the traceless energy–momentum tensor Eq. (2.82). This energy–momentum tensor is not invariant under gauge transformations of the vector field, but, since only its trace enters the Lagrangian, the whole theory is gauge-invariant and, simply, there is no interaction.

3.1.6 Self-coupled scalar gravity

So far, we have found several serious problems hindering this theory from describing gravity realistically and we could simply abandon scalar theories of gravity as hopeless and try the next candidate for a SRFT of gravity. However, before we do, we want to introduce, for illustrative purposes, a possible modification of this theory that cannot fix most of the problems encountered, but is the answer to a legitimate question: does gravity couple to all forms of matter/energy including gravitational energy or only to non-gravitational energies? In the theory we have constructed, gravity does not couple to itself. However, since gravitational energy can be transformed into other forms of energy and vice-versa, it would be reasonable to expect that gravity couples to all forms of energy equally. Can we modify our theory so as to fulfill this expectation?

We are looking for a theory with the equation of motion

$$\partial^2\phi = CT, \tag{3.37}$$

where T is the trace of the *total* energy–momentum tensor, which should include contributions from the scalar gravitational field, matter fields, and interaction terms. The energy–momentum tensor of ϕ in the free theory is quadratic in $\partial\phi$. To obtain it on the r.h.s. of the equation of motion, we must add to the Lagrangian a term of the form $\phi(\partial\phi)^2$. However, this term will also contribute to the new energy–momentum tensor, and, to produce it on the r.h.s. of the new equation of motion, we need a term $\phi^2(\partial\phi)^2$ in the Lagrangian, and so on. Thus, we need to introduce an infinite number of corrections to the scalar Lagrangian.

As for the interaction terms, they contain the trace of the matter energy–momentum tensor, and thus we need to make some assumption about the form of the matter Lagrangian in order to make some progress: we will take it to be of the form

$$\mathcal{L}_{\text{matter}} = K - V, \tag{3.38}$$

where K is quadratic in the first partial derivatives of the matter fields and V is just a function of the fields. This implies that

$$T_{\text{matter}} = (d-2)K - dV, \tag{3.39}$$

and the action Eq. (3.3), which we can consider the lowest order in an expansion in small ϕ, takes the form

$$S = \frac{1}{c}\int d^dx \left\{\frac{1}{2Cc^2}(\partial\phi)^2 + \left(1 + \frac{d-2}{c^2}\phi\right)K - \left(1 + \frac{d\phi}{c^2}\right)V\right\}. \tag{3.40}$$

It is reasonable to expect that the full action, with all the ϕ corrections, takes the form

$$S = \frac{1}{c}\int d^dx \left\{\frac{1}{2Cc^2}f(\phi)(\partial\phi)^2 + g(\phi)K - h(\phi)V\right\}, \tag{3.41}$$

where f, g, and h are functions of ϕ to be found by imposing the condition that the equation of motion of ϕ can be written in the form Eq. (3.37), where T is the trace of the total energy–momentum tensor of the above Lagrangian, which is easily found to be

$$T = (d-2)\frac{1}{2Cc^2}f(\phi)(\partial\phi)^2 + (d-2)g(\phi)K - dh(\phi)V. \tag{3.42}$$

The ϕ equation of motion coming from Eq. (3.41) is

$$\partial^2\phi = -\tfrac{1}{2}(f'/f)(\partial\phi)^2 + Cg'/(fK) - Ch'/(fV), \tag{3.43}$$

and, on comparing this with Eqs. (3.37) and (3.42), one finds

$$f = \frac{1}{a + [(d-2)/c^2]\phi}, \qquad g = f/b, \qquad h = (f/e)^{\frac{d}{d-2}}, \tag{3.44}$$

where a, b, and e are integration constants. If we want to recover Eq. (3.40) in the weak-field limit, we have to take $a = b = e = 1$. Then, we have succeeded and we have found the action

$$S = \frac{1}{c}\int d^dx \left\{\frac{1}{2Cc^2}\frac{(\partial\phi)^2}{1+[(d-2)/c^2]\phi} + \left[1+\frac{d-2}{c^2}\phi\right]K - \left[1+\frac{d-2}{c^2}\phi\right]^{\frac{d}{d-2}} V\right\}, \tag{3.45}$$

that gives rise to the equation of motion Eq. (3.37) with T, the trace of the total energy–momentum tensor corresponding to the above action, given by

$$T = \frac{d-2}{2Cc^2}\frac{(\partial\phi)^2}{1+[(d-2)/c^2]\phi} + (d-2)\left[1+\frac{d-2}{c^2}\phi\right]K - d\left[1+\frac{d-2}{c^2}\phi\right]^{\frac{d}{d-2}} V. \tag{3.46}$$

This result was presented in [405] and [306], but the theory obtained is the one proposed by Nordström back in 1913 in [730, 731] in terms of different variables: on introducing

$$\Phi \equiv c^2\left[1+\frac{d-2}{c^2}\phi\right]^{\frac{1}{2}}, \tag{3.47}$$

the action Eq. (3.45) takes the form

$$S = \frac{1}{c}\int d^dx \left\{\frac{2}{(d-2)^2Cc^2}(\partial\Phi)^2 + [\Phi/c^2]^2K - [\Phi/c^2]^{\frac{2d}{d-2}} V\right\}. \tag{3.48}$$

In the case in which $V = 0$, taking into account Eq. (3.5), the equation of motion can be written in the standard form

$$\partial^2\Phi = \frac{(d-3)4\pi G_{\rm N}^{(d)}}{c^2}\,\Phi\, T^{(0)}_{\rm matter}, \tag{3.49}$$

where $T^{(0)}_{\rm matter}$ is the trace of the matter energy–momentum tensor obtained from the uncoupled $\mathcal{L}_{\rm matter}$. In Nordström's theory, this is the equation valid in all cases ($V \neq 0$).

In this form it is very difficult to see that the theory has the property we wanted (that the source for the gravitational scalar field is the trace of the total energy–momentum tensor).

There is yet another way of rewriting this theory, which was found by Einstein and Fokker [365]. This was one of Einstein's first attempts at building a relativistic theory of gravity in which the gravitational field is represented by a metric, as suggested by Grossmann.

3.1.7 *The geometrical Einstein–Fokker theory*

The Einstein–Fokker theory is based on a conformally flat metric,

$$g_{\mu\nu} \equiv [\Phi/c^2]^{\frac{4}{d-2}}\eta_{\mu\nu}. \tag{3.50}$$

Only the conformal factor Φ is dynamical. The equation of motion for the metric (i.e. for Φ) is

$$R(g) = \frac{(d-1)(d-3)}{d-2} \frac{16\pi G_{\mathrm{N}}^{(d)}}{c^2} T_{\mathrm{matter}}, \tag{3.51}$$

where $R(g)$ is the Ricci scalar for the metric $g_{\mu\nu}$ and T_{matter} is calculated from the canonical, special-relativistic fully covariant energy–momentum tensor $T_{\mathrm{matter}\,\mu\nu}$, by contracting both indices with $g^{\mu\nu}$.

Alternatively, the Einstein–Fokker theory can be formulated by giving the above equation for an arbitrary metric, but adding another equation,

$$C_{\mu\nu}{}^{\rho\sigma}(g) = 0, \tag{3.52}$$

where $C_{\mu\nu}{}^{\rho\sigma}$ is the Weyl tensor. This equation implies that the metric is conformally flat and can be written, in appropriate coordinates, in the form (3.50).

Using the formulae in Appendix E, we find

$$R(g) = \frac{4(d-1)}{d-2} [\Phi/c^2]^{-\frac{d+2}{d-2}} \partial^2 [\Phi/c^2]. \tag{3.53}$$

This, together with

$$T_{\mathrm{matter}} = [\Phi/c^2]^{-\frac{4}{d-2}} T^{(0)}_{\mathrm{matter}}, \tag{3.54}$$

gives Eq. (3.49).

Einstein and Fokker did not give a Lagrangian for gravity coupled to matter, and therefore they had to postulate how gravity affects the motion of matter. Here, the power of the Einstein–Fokker formulation of Nordström's theory becomes manifest: Einstein and Fokker suggested replacing the flat spacetime metric $\eta_{\mu\nu}$ by the conformally flat metric $g_{\mu\nu}$ everywhere in the matter Lagrangian. This prescription can be used in most matter Lagrangians (not involving spinors). For instance, for the massive particle, it leads to

$$\begin{aligned} S_{\mathrm{pp}}[X^\mu(\xi)] &= -Mc \int d\xi \sqrt{g_{\mu\nu}(X)\dot{X}^\mu \dot{X}^\nu} \\ &= -Mc \int d\xi \, [\Phi(X)/c^2]^{\frac{2}{d-2}} \sqrt{\eta_{\mu\nu}\dot{X}^\mu \dot{X}^\nu} \\ &\sim -Mc \int d\xi \, [1 + \phi(X)/c^2 + \cdots] \sqrt{\eta_{\mu\nu}\dot{X}^\mu \dot{X}^\nu}, \end{aligned} \tag{3.55}$$

which is, to lowest order in ϕ, our old result. In general, the equation of motion simply tells us that massive particles move along timelike geodesics with respect to the metric $g_{\mu\nu}$. This is a very powerful statement that goes far beyond Nordström's original theory.

For the massless particle, we also find that the coupling can again be absorbed into the worldline auxiliary metric. There is no bending of light in this theory. However, one can argue [349] that, although there is no *global* bending, there is *local* bending of light rays. As explained in [349], local bending is a kinematical effect associated with accelerating reference frames and occurs, via Einstein's equivalence principle of gravitation and inertia (to be discussed in Section 3.3), in any theory, independently of any equation of motion. Global bending is an integral of local bending, depending on the conformal spacetime structure, which depends on the specific equations of motion of each theory. The contribution of local bending to global bending is just half the value predicted by GR and is experimentally confirmed. In scalar gravity, this contribution is canceled out.

3.2 Gravity as a self-consistent massless spin-2 SRFT

In the previous section we have seen that the simplest possible SRFT of gravity, scalar gravity, is not a good candidate since it does not pass two of the classical tests: bending of light and precession of the perihelion of Mercury. Apart from this, the theory did not have consistency problems regarding coupling to matter[12] or to the gravity field itself but, precisely because of this, there was a lot of freedom in choosing the energy–momentum tensor which could be the matter energy–momentum tensor or the total energy–momentum tensor. We argued that this could be considered a weakness of the theory.

Now we have to try the next simplest possibility. Excluding a vector field (a spin-1 particle) because it leads to repulsion between like charges, the next possibility is that gravity is mediated by a massless spin-2 particle (the graviton).

The field that describes a spin-2 particle is a symmetric two-index *Lorentz* tensor $h_{\mu\nu}$ whose indices are raised and lowered with the Minkowski metric $\eta_{\mu\nu}$ (this is a SRFT). For the free field $h_{\mu\nu}$, one can try the equation of motion [151, 152] (see also [83, 84, 153, 711]

$$\partial^2 h^{\mu\nu} = 0. \tag{3.56}$$

Things are, however, not that simple. On the one hand, this theory does not have positive-definite energy unless one imposes a consistency condition:

$$\partial^\mu\left(h_{\mu\nu} - \tfrac{1}{2}\eta_{\mu\nu}h^\rho{}_\rho\right) = 0, \tag{3.57}$$

as pointed out by Weyl in [951]. On the other hand, the field $h_{\mu\nu}$ describes many more helicity states than those of a massless spin-2 particle (a symmetric $h_{\mu\nu}$ has $d(d+1)/2$ independent components, some of which describe spin-1 and spin-0 helicity states) and therefore the equations of motion of this field should be such that, on-shell, it describes only the $d(d-3)/2$ helicity states that a massless spin-2 particle has in d dimensions (two in four dimensions: $s_z = -2, +2$).

These two problems are related since the negative contribution to the energy comes precisely from some of the unwanted helicities which are eliminated when one imposes the above condition (which we will later call the *De Donder*[13] gauge condition [296]). To eliminate all the helicities not corresponding to the spin-2 particle we want to describe, we have to impose another condition,

$$h^\mu{}_\mu = 0. \tag{3.58}$$

---

[12] We saw, however, that there was some disagreement between the effect of gravity on massless fields and the effect on massless particles.

[13] Also known in the literature as the *harmonic* or *Hilbert* [888], *Hilbert–Lorentz* [739], or *Einstein* gauge condition.

Actually, the correct way of arriving at these two conditions is to introduce into the theory some kind of gauge freedom so that $\partial^\mu\big(h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h^\rho{}_\rho\big)$ and $h^\mu{}_\mu$ can take arbitrary values, in particular zero. However, let us accept for the moment the theory given by Eq. (3.56) supplemented by the conditions Eqs. (3.57) and (3.58) and let us now consider the coupling to matter. As in any SRFT of gravitation, matter must couple to gravity through the energy–momentum tensor. The l.h.s. of the equation of motion has two free indices and, therefore, it is natural to expect the matter energy–momentum tensor on the r.h.s., that is[14]

$$\partial^2 h^{\mu\nu} = \chi T^{\mu\nu}_{\text{matter}}, \tag{3.59}$$

where χ is a coupling constant whose dimensions and value we will discuss later. As opposed to the scalar case, this equation (which still has to be supplemented by Eqs. (3.57) and (3.58)) does impose consistency conditions on the matter energy–momentum tensor. First, it has to be symmetric because the l.h.s. is. Second, it has to be divergence-free (conserved), because the l.h.s. is, as a result of the supplementary conditions imposed on $h_{\mu\nu}$. Both conditions are satisfied by the Belinfante or Rosenfeld energy–momentum tensors and by an infinite number of tensors obtained from these by adding a superpotential correction that does not modify their symmetry. Nevertheless, it is clear that this is a theory with a structure tighter than the scalar one and it is encouraging to find that the consistency of the theory imposes physically meaningful conditions on the energy–momentum tensor. All this makes it worth studying.

Of course, we want to find the gauge-invariant equations of motion (or Lagrangian) and the gauge transformations which allow us to impose the conditions Eqs. (3.57) and (3.58) and arrive at Eq. (3.59). These equations of motion must necessarily be of the form

$$\mathcal{D}^{\mu\nu}(h) = \chi T^{\mu\nu}_{\text{matter}}, \tag{3.60}$$

where, now, by consistency with the conservation of the matter energy–momentum tensor, the wave operator $\mathcal{D}^{\mu\nu}(h)$ should also be divergenceless, viz.

$$\partial_\mu \mathcal{D}^{\mu\nu}(h) = 0, \tag{3.61}$$

off-shell, i.e. independently of the equations of motion (which, in vacuum, should have the form $\mathcal{D}_{\mu\nu}(h) = 0$). In other words, the theory has to have the above property as a *Bianchi* or *gauge identity*. This kind of identity can be derived from theories with a gauge symmetry according to the general procedure outlined in Chapter 2 and, if we obtain a theory with this property (which is easier to do), we will most surely have obtained a theory with the gauge symmetry needed to remove the unwanted degrees of freedom.

The problem of finding a theory with these properties, a theory for a massless spin-2 particle, was solved by Fierz and Pauli in [388] and it was studied again by Ogievetsky and Polubarinov in [739] in a more general setting, including possible self-interactions of the gravitational field.

The matter energy–momentum tensor in Eq. (3.60) is calculated from the free matter field theory. When it is coupled to gravity, only the *total* (matter plus gravity)

[14] Certainly, there are other possibilities: we can add to the r.h.s. terms like $\eta^{\mu\nu}T^{\rho}_{\text{matter}\,\rho}$. However, these possibilities are inconsistent with the supplementary conditions Eqs. (3.57) and (3.58).

energy–momentum tensor is conserved. This is the inconsistency problem of this SRFT of gravity (see, for instance, [707]). Then, we should add, at least, the gravitational energy–momentum tensor calculated from the Lagrangian from which we derived Eq. (3.60) to the r.h.s. of Eq. (3.60), for consistency. However, if we want to derive the new equation of motion from a Lagrangian, we need to add to the old Lagrangian a cubic term, which, in turn, will introduce a correction to the gravitational energy–momentum tensor. If we add this correction to the r.h.s. of Eq. (3.60), we will have to add a further correction to the Lagrangian, and so on. The coupling to matter requires an infinite number of corrections to the free spin-2 (Fierz–Pauli) theory.

The problem of consistent self-interaction of the gravitational field is of great importance and was studied in [488, 489, 638, 639], where Gupta and Kraichnan pointed it out for the first time; in the classical works of Feynman and Thirring [386, 888] in which the first correction to the free equation of motion was found and used to calculate the precession of the perihelion of Mercury;[15] in [965]; in the works of Weinberg [941, 942], in which it was shown that a quantum theory of a massless spin-2 particle can have a Lorentz-invariant quantum S matrix only if it couples to the total energy–momentum tensor; in Deser's paper [299], in which it was shown that GR can be seen as the result of adding this infinite number of corrections;[16] in Boulware and Deser's paper [176], in which Weinberg's result was completed by a determination of the form of the gravitational energy–momentum tensor to which gravity itself would couple in a consistent quantum theory, which was found to be, *in the long-wavelength limit*, the one predicted by GR; in [270, 378, 525, 526, 933, 934], in which general, consistent, non-linear theories of a spin-2 particle were investigated with the conclusion that the only possible symmetries of these theories were "normal spin-2 gauge invariance" (to be defined later) and general covariance and, more recently, in [174], in which an alternative theory for a $d = 3$ spin-2 particle was found.

In this section we are going to study the Fierz–Pauli theory and its gauge symmetry. Then, we will couple it to matter and we will find the predictions for the bending of light by gravity and the precession of the perihelion of Mercury. The latter will come out with the wrong value and we will see the need to introduce corrections into the theory, as the inconsistency problem suggests. We will try to envisage a systematic way of introducing these corrections on the basis of the Noether method explained in the previous chapter. Then, we will spend some time trying to find the first correction (i.e. the gravitational energy–momentum tensor) for various methods and we will calculate the corresponding correction to the precession of the perihelion of Mercury, discovering that the Belinfante–Rosenfeld energy–momentum tensor (employed by Thirring in [888], does not give the right result, whereas the one used in GR does. We will then use Deser's procedure to find a theory that is consistent to all orders. This theory that is will turn out to be GR, which we will introduce in the following section.

Before proceeding to the construction of the Fierz–Pauli theory, it is worth studying a simpler example of the relation among gauge symmetry, Bianchi (gauge) identities, and conserved charges in the SRFT of a spin-1 particle.

---

[15] Thirring's result is actually wrong [725], as we will see.
[16] This result was extended in [301] to general vacua.

3.2.1 Gauge invariance, gauge identities, and charge conservation in the SRFT of a spin-1 particle

A massive or massless spin-1 particle is described by a vector field A^μ. The simplest relativistic wave equation we could imagine for it would be

$$\left(\partial^2 + m^2\right)A^\mu = 0. \tag{3.62}$$

However, the energy density of this theory is not positive-definite unless one imposes the *Lorentz* or *transversality* condition

$$\partial_\mu A^\mu = 0. \tag{3.63}$$

Furthermore, just as in the spin-2-particle case, the vector A^μ describes spin-1 helicity states but also spin-0 helicity states. A d-dimensional vector field has d independent components, but a massive spin-1 particle in d dimensions has $d-1$ states (three in $d=4$: $s_z = -1, 0, 1$) and a massless spin-1 particle has $d-2$ helicity states (two in $d=4$: $s_z = -1, +1$). It is precisely the unwanted spin-0 helicity states that contribute negatively to the energy and the Lorentz condition projects them out.

If we couple the massless theory to charged matter, by Lorentz covariance, this has to be described by a vector current j^μ, so we have

$$\partial^2 A^\mu = j^\mu \tag{3.64}$$

and, by consistency with the Lorentz condition, the vector current has to be conserved, $\partial_\mu j^\mu = 0$, which is, again, a physically meaningful condition that coincides with our experience with electric charges and currents.

We would like to construct a theory in which the Lorentz condition arises as a consequence of the equation of motion in the massive case and in which $\partial_\mu A^\mu$ is completely arbitrary in the massless case. These conditions guarantee the removal of the unwanted helicities. We expect the equation of motion to be of the form

$$\mathcal{D}^\mu(A) + m^2 A^\mu = j^\mu, \tag{3.65}$$

where, now, by consistency, the massless wave operator $\mathcal{D}^\mu(A)$ has to satisfy off-shell the identity

$$\partial_\mu \mathcal{D}^\mu(A) = 0, \tag{3.66}$$

which should arise as the gauge identity associated with some gauge symmetry.

We could proceed as in [739], translating these conditions into a gauge identity for a general Lagrangian and then trying to find, with as much generality as possible, a gauge symmetry (forming a group) leading to that gauge identity. As is well known, the result is the *Proca Lagrangian* and equation of motion,

$$\begin{aligned} S[A] &= \int d^d x \left[-\tfrac{1}{4}F^2 + \frac{m^2}{2}A^2\right], \\ \mathcal{D}^\mu_{(m)}(A) &= \mathcal{D}^\mu(A) + m^2 A^\mu = 0, \end{aligned} \tag{3.67}$$

where

$$\mathcal{D}^\mu(A) \equiv \partial_\mu F^{\mu\nu}, \qquad F_{\mu\nu} = 2\partial_{[\mu}A_{\nu]}, \tag{3.68}$$

which, in the $m = 0$ limit, reduce to Maxwell's Lagrangian and Maxwell's equation. Owing to the antisymmetry of $F_{\mu\nu}$, the massless wave operator does indeed have the off-shell property Eq. (3.66), which implies, in turn, Eq. (3.63) in the massive case, as we needed in order to obtain a positive-definite energy and to eliminate the spin-0 degree of freedom. In turn, the massless theory is easily seen to be invariant under gauge transformations,

$$\delta A_\mu = \partial_\mu \Lambda(x). \tag{3.69}$$

Given any A^μ, we can gauge-transform it into another A'^μ satisfying Lorentz's condition: it is enough to choose a gauge parameter Λ that is a solution of $\partial^2\Lambda = -\partial_\mu A^\mu$. Lorentz's condition does not completely fix the gauge: there are many potentials A_μ that satisfy that gauge condition and are related by non-trivial gauge transformations (those with parameter satisfying $\partial^2\Lambda = 0$). To fix the gauge invariance completely, it is necessary to impose another gauge condition. This is why this gauge symmetry reduces to $d-2$ the number of degrees of freedom described by a massless vector field, just as we needed in order to describe just the spin-1 case.

Furthermore, as we expected, the identity Eq. (3.66) is related to the above gauge symmetry via Noether's theorems. Let us follow Chapter 2: we know that the Maxwell action

$$S[A] = \int d^dx \left\{-\tfrac{1}{4}F^2\right\} \tag{3.70}$$

is exactly invariant under gauge transformations because $F_{\mu\nu}$ is. Thus,[17]

$$\begin{aligned}
\delta S &= \int d^dx \left\{-F^{\mu\nu}\partial_\mu \delta A_\nu\right\} \\
&= \int d^dx \left\{\mathcal{D}^\mu(A)\delta A_\nu - \partial_\mu(F^{\mu\nu}\delta A_\nu)\right\} \\
&= \int d^dx \left\{-\partial_\mu \mathcal{D}^\mu(A)\Lambda - \partial_\mu(F^{\mu\nu}\partial_\nu\Lambda - \mathcal{D}^\mu(A)\Lambda)\right\}.
\end{aligned} \tag{3.71}$$

Now we argue as follows: if the gauge parameter $\Lambda(x)$ and its derivatives vanish on the boundary, the integral of the total derivative term is zero. Since the variation is zero for any Λ, then $\partial_\mu \mathcal{D}^\mu(A) = 0$. This is the gauge identity. Now that we know it always holds, we can consider more general gauge parameters and the invariance of the action implies that

$$\partial_\mu j^\mu_{\mathrm{N2}}(\Lambda) = 0, \qquad j^\mu_{\mathrm{N2}}(\Lambda) = j^\mu_{\mathrm{N1}}(\Lambda) - \mathcal{D}^\mu(A)\Lambda, \qquad j^\mu_{\mathrm{N1}}(\Lambda) = F^{\mu\nu}\partial_\nu\Lambda. \tag{3.72}$$

$j^\mu_{\mathrm{N1}}(\Lambda)$ and $j^\mu_{\mathrm{N2}}(\Lambda)$ are *Noether currents* associated with the gauge parameter Λ. $j^\mu_{\mathrm{N1}}(\Lambda)$ is conserved only on-shell but $j^\mu_{\mathrm{N2}}(\Lambda)$ is automatically conserved (i.e. off-shell). On-shell they are evidently identical. Furthermore, as can easily be checked in this case, the Noether current $j^\mu_{\mathrm{N2}}(\Lambda)$ associated with a gauge symmetry enjoys another property [110]: it is always the divergence of an antisymmetric tensor. In this case

$$j^\mu_{\mathrm{N2}}(\Lambda) = \partial_\nu j^{\nu\mu}_{\mathrm{N2}}(\Lambda), \qquad j^{\nu\mu}_{\mathrm{N2}}(\Lambda) = -F^{\nu\mu}\Lambda. \tag{3.73}$$

[17] We write here δ instead of $\tilde{\delta}$ because these transformations do not involve any coordinate transformation and the two variations are identical.

The conserved charge, which can be written covariantly, up to normalization, as

$$q(\Lambda) \sim \int_{V_t} d^{d-1}\Sigma_\mu j^\mu_{N2}(\Lambda), \tag{3.74}$$

where V_t is a spacelike hypersurface (a constant time slice for some time coordinate), can be reexpressed as an integral over the boundary of V_t, i.e. a surface integral over an S^{d-2} sphere at infinity in d-dimensional Minkowski spacetime:

$$q(\Lambda) \sim \frac{1}{2}\int_{S^{d-2}_\infty} d^{d-2}\Sigma_{\mu\nu} j^{\nu\mu}_{N2}(\Lambda) = -\tfrac{1}{2}\int_{S^{d-2}_\infty} d^{d-2}\Sigma_{\mu\nu} F^{\nu\mu}\Lambda. \tag{3.75}$$

For $\Lambda = 1$ or any $\Lambda(x)$ that goes to 1 at spatial infinity, $q(\Lambda)$ is just the electric charge. In differential-form language

$$q = \int_{\partial\Sigma} {}^\star F. \tag{3.76}$$

For later use, it should be noted that, as a matter of fact, the massless theory could have been found by this simple procedure: write the most general Lorentz-invariant Lagrangian quadratic in derivatives of A^μ with arbitrary coefficients a and b:

$$S[A] = \int d^dx\,\{a\partial_\mu A_\nu \partial^\mu A^\nu + b\partial_\mu A_\nu \partial^\nu A^\mu\}, \tag{3.77}$$

and impose on the equations of motion the gauge identity Eq. (3.66). This fixes $a = -b$ and, on choosing the overall normalization suitably, one obtains Maxwell's Lagrangian. Then we can immediately find the gauge symmetry that leaves it invariant.

How would the presence of sources modify these results? Essentially in no way, but we have to be a bit more careful. First of all, under a gauge transformation, the first variation of the action with sources

$$S_j[A] = \int d^dx\,\left\{-\tfrac{1}{4}F^2 - A_\mu j^\mu\right\} \tag{3.78}$$

is

$$\delta S_j = \int d^dx\,\left\{\Lambda\partial_\mu j^\mu - \partial_\mu\left(\Lambda j^\mu\right)\right\}, \tag{3.79}$$

and we have invariance up to a total derivative only if the source current is conserved. Conservation is also required by consistency of the equation of motion

$$\mathcal{D}^\mu(A) = j^\mu. \tag{3.80}$$

On the other hand, we can vary the action as before: first under a general variation δA_μ and then using the form of the gauge transformation:

$$\begin{aligned}\delta S_j &= \int d^dx\,\left\{-F^{\mu\nu}\partial_\mu\delta A_\nu - \delta A_\nu j^\nu\right\}\\ &= \int d^dx\,\left\{\left[\mathcal{D}^\mu(A) - j^\mu\right]\delta A_\nu - \partial_\mu(F^{\mu\nu}\delta A_\nu)\right\}\\ &= \int d^dx\,\left\{-\partial_\mu\left[\mathcal{D}^\mu(A) - j^\mu\right]\Lambda - \partial_\mu\left\{F^{\mu\nu}\partial_\nu\Lambda - \left[\mathcal{D}^\mu(A) - j^\mu\right]\Lambda\right\}\right\}.\end{aligned} \tag{3.81}$$

The two forms of the variation of the action are identical. By identifying them we arrive at the same results as in the sourceless case because the source terms cancel each other out.

3.2.2 Gauge invariance, gauge identities, and charge conservation in the SRFT of a spin-2 particle

Inspired by the lessons learned in finding the SRFT of a spin-1 particle, we return to the spin-2 theory. We could follow [739] and try to determine the most general theory with the required properties, including non-linear couplings and transformations. Instead, since we want to start with a linear theory (which will be adequate for a free spin-2 particle), we are going to use the shortcut we used in the massless spin-1 case: construct the most general (up to total derivatives) Lorentz-invariant action that is quadratic in $\partial_\rho h_{\mu\nu}$ and impose the gauge identity Eq. (3.61). This should determine the action for the massless theory up to total derivatives and overall normalization and we can then search for the gauge invariance which the theory surely enjoys and prove that it is enough to eliminate the unwanted degrees of freedom. Then we can add terms polynomial in $h_{\mu\nu}$ in order to find the action for the massive theory.

There are only four different possible terms in the Lagrangian up to total derivatives. We can write all of them with unknown coefficients,

$$S = \int d^dx \left\{a\partial^\rho h^{\mu\nu}\partial_\rho h_{\mu\nu} + b\partial^\mu h^{\nu\rho}\partial_\nu h_{\mu\rho} + c\partial^\mu h\partial^\lambda h_{\lambda\mu} + d\partial^\mu h\partial_\mu h\right\}, \tag{3.82}$$

where we use the standard notation h for the trace of $h_{\mu\nu}$,

$$h \equiv h_\mu{}^\mu. \tag{3.83}$$

We normalize the kinetic term canonically[18] by setting $a = +\frac{1}{4}$, and then easily find that the equations of motion will satisfy Eq. (3.61) if $b = -\frac{1}{2}$, $c = \frac{1}{2}$, and $d = -\frac{1}{4}$, so the action we are looking for is the Fierz–Pauli action [388]

$$\boxed{S = \int d^dx \left\{\tfrac{1}{4}\partial^\mu h^{\nu\rho}\partial_\mu h_{\nu\rho} - \tfrac{1}{2}\partial^\mu h^{\nu\rho}\partial_\nu h_{\mu\rho} + \tfrac{1}{2}\partial^\mu h\partial^\lambda h_{\lambda\mu} - \tfrac{1}{4}\partial^\mu h\partial_\mu h\right\}.} \tag{3.84}$$

We want the above action to be dimensionless in natural units $\hbar = c = 1$. The field $h_{\mu\nu}$ has to have the dimensions of $L^{-\frac{d-2}{2}}$. Then, since the energy–momentum tensor has the same dimensions as the Lagrangian, Eq. (3.60) implies that χ has the inverse dimensions of $h_{\mu\nu}$, so $\chi h_{\mu\nu}$ is dimensionless.

The corresponding divergenceless equations of motion are

$$\begin{aligned}\frac{\delta S}{\delta h_{\mu\nu}} &\equiv -\tfrac{1}{2}\mathcal{D}^{\mu\nu}(h),\\ \mathcal{D}_{\mu\nu}(h) &= \partial^2 h_{\mu\nu} + \partial_\mu\partial_\nu h - 2\partial^\lambda\partial_{(\mu}h_{\nu)\lambda} - \eta_{\mu\nu}\left(\partial^2 h - \partial_\lambda\partial_\sigma h^{\lambda\sigma}\right) = 0.\end{aligned} \tag{3.85}$$

[18] That is, $S_{\mathrm{FP}} = \int d^dx \left\{+\frac{1}{4}\partial_t h_{ij}\partial_t h_{ij} + \cdots\right\}$.

By subtracting the trace of this equation we can simplify it without any loss of information:

$$\hat{\mathcal{D}}_{\mu\nu}(h) \equiv \mathcal{D}_{\mu\nu}(h) - \frac{1}{d-2}\eta_{\mu\nu}\mathcal{D}_\rho{}^\rho(h) = \partial^2 h_{\mu\nu} + \partial_\mu\partial_\nu h - 2\partial^\lambda\partial_{(\mu}h_{\nu)\lambda} = 0. \quad (3.86)$$

Sometimes the equation of motion (3.85) is written in terms of the convenient variable $\bar{h}_{\mu\nu}$:

$$\mathcal{D}_{\mu\nu}(\bar{h}) = \partial^2\bar{h}_{\mu\nu} - 2\partial^\lambda\partial_{(\mu}\bar{h}_{\nu)\lambda} + \eta_{\mu\nu}\partial_\lambda\partial_\sigma\bar{h}^{\lambda\sigma} = 0, \quad (3.87)$$

where

$$\bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \tfrac{1}{2}\eta_{\mu\nu}h. \quad (3.88)$$

Finally, we can write the Fierz–Pauli wave operator as the divergence of a tensor $\eta^{\mu\nu\rho}$,

$$\mathcal{D}^{\nu\rho}(h) = 2\partial_\mu\eta^{\nu\rho\mu}, \quad (3.89)$$

but the tensor $\eta^{\mu\nu\rho}$ is not uniquely defined. Some possible candidates are

$$\begin{aligned}
\eta_{\mathrm{T}}^{\nu\rho\mu} &= \eta_{\mathrm{T}}^{(\nu\rho)\mu} &&= -\partial_\sigma H^{\mu\sigma\nu\rho}, \\
\eta_{\mathrm{LL}}^{\nu\rho\mu} &= \eta_{\mathrm{LL}}^{\nu[\rho\mu]} &&= -\partial_\sigma K^{\nu\sigma\rho\mu}, \\
\eta_{\mathrm{AD}}^{\nu\rho\mu} &= \eta_{\mathrm{AD}}^{[\nu|\rho|\mu]} &&= -\partial_\sigma K^{\nu\mu\rho\sigma},
\end{aligned} \quad (3.90)$$

where

$$\begin{aligned}
K^{\mu\nu\rho\sigma} &= \tfrac{1}{2}\left[\eta^{\mu\sigma}\bar{h}^{\nu\rho} + \eta^{\nu\rho}\bar{h}^{\mu\sigma} - \eta^{\mu\rho}\bar{h}^{\nu\sigma} - \eta^{\nu\sigma}\bar{h}^{\mu\rho}\right], \\
H^{\mu\sigma\nu\rho} &= \tfrac{1}{2}\left[\eta^{\sigma\rho}\bar{h}^{\mu\nu} + \eta^{\sigma\nu}\bar{h}^{\mu\rho} - \eta^{\nu\rho}\bar{h}^{\mu\sigma} - \eta^{\mu\sigma}\bar{h}^{\nu\rho}\right],
\end{aligned} \quad (3.91)$$

H is symmetric in the last two indices and K is antisymmetric. In fact, K has exactly the same symmetries as the Riemann tensor (in the Levi-Cività case).

On the other hand, $\eta_{\mathrm{T}}^{\mu\nu\rho}$ has the defining property

$$\frac{\partial\mathcal{L}_{\mathrm{FP}}}{\partial\partial_\mu h_{\nu\rho}} = \eta_{\mathrm{T}}^{\nu\rho\mu}, \quad (3.92)$$

for the Fierz–Pauli Lagrangian written in Eq. (3.84).

Using any of the last two $\eta^{\nu\rho\mu}$s, the fact that the Fierz–Pauli wave operator $\mathcal{D}^{\mu\nu}(h)$ is divergenceless becomes manifest.

Let us now determine the gauge symmetry of the Fierz–Pauli Lagrangian. Under a general variation of $h_{\mu\nu}$, the variation of the action is, up to a total derivative,

$$\delta S_{\mathrm{FP}} = -\tfrac{1}{2}\int d^dx\,\mathcal{D}^{\mu\nu}\delta h_{\mu\nu}. \quad (3.93)$$

If $\delta h_{\mu\nu}$ is a gauge transformation, we know that, up to total derivatives, the integrand of the variation of the action has to be proportional to the gauge identity Eq. (3.61), i.e.

$$\int d^dx\,\mathcal{D}^{\mu\nu}\delta h_{\mu\nu} \sim \int d^dx\,\partial_\mu\mathcal{D}^{\mu\nu}\epsilon_\nu, \quad (3.94)$$

(the gauge parameter $\epsilon_\mu(x)$ has to be a local Lorentz vector). On integrating the r.h.s. by parts, and choosing a convenient normalization, we find the gauge transformation

$$\delta_\epsilon h_{\mu\nu} = -2\partial_{(\mu}\epsilon_{\nu)}. \tag{3.95}$$

We can now check directly that the Fierz–Pauli Lagrangian is invariant under these transformations:

$$\delta S_{\rm FP} = \int d^d x \frac{\partial \mathcal{L}_{\rm FP}}{\partial \partial_\mu h_{\nu\rho}} \partial_\mu \delta h_{\nu\rho} = -\int d^d x \partial_\sigma H^{\mu\sigma\nu\rho} \partial_\mu \delta h_{\nu\rho}. \tag{3.96}$$

Here we have used Eqs. (3.90) and (3.92). On integrating by parts and using the explicit form of the variation, we have

$$\delta S_{\rm FP} = \int d^d x \frac{\partial \mathcal{L}_{\rm FP}}{\partial \partial_\mu h_{\nu\rho}} \partial_\mu \delta h_{\nu\rho} = \int d^d x \left\{ \partial_\sigma [2H^{\mu\sigma\nu\rho} \partial_\mu \partial_\nu \epsilon_\rho] - 2H^{\mu\sigma\nu\rho} \partial_\sigma \partial_\mu \partial_\nu \epsilon_\rho \right\}. \tag{3.97}$$

The second term vanishes identically and the action turns out to be invariant up to a total derivative (the first term).

To complete our program for the massless spin-2 theory, it remains only to show that, using this gauge symmetry, we can remove $2d$ of the $d(d+1)/2$ independent components of $h_{\mu\nu}$ to leave only the $d(d-3)/2$ degrees of freedom of a massless spin-2 particle in d dimensions. The counting of degrees of freedom in a gauge theory is not straightforward. See e.g. [530] for simple rules, but one can show that, using the gauge transformations (3.95), one can indeed eliminate $2d$ components (set them to a given value by fixing the gauge).

There are two popular gauges: the *transverse, traceless gauge*

$$\partial_\mu h^{\mu\nu} = h = 0, \tag{3.98}$$

which automatically leads to the equation of motion

$$\mathcal{D}_{\mu\nu}(h) = \partial^2 h_{\mu\nu} = 0 \tag{3.99}$$

typical of a massless field, and the *De Donder* or *harmonic* gauge

$$\partial_\mu \bar{h}^{\mu\nu} = 0, \tag{3.100}$$

which leads to

$$\mathcal{D}_{\mu\nu}(h) = \partial^2 \bar{h}_{\mu\nu} = 0. \tag{3.101}$$

The traceless transverse gauge implies the De Donder gauge but not conversely. The transverse, traceless condition does not completely fix the gauge, since it is preserved by gauge transformations with $\epsilon^\mu = \partial^\mu \epsilon$ and $\partial^2 \epsilon = 0$.

After the identification of the gauge symmetry of the massless theory, the next step in our program is finding the massive theory. We need to modify the massless equation of motion so that it gives the equation of motion

$$\left(\partial^2 + m^2\right) h^{\mu\nu} = 0, \tag{3.102}$$

plus the De Donder and traceless conditions. These $d+1$ constraints leave only the $(d-2)(d+1)/2$ degrees of freedom of the massive spin-2 particle in d dimensions (five in

$d = 4$: $s_s = -2, -1, 0, +1, +2$). We know that the massless wave operator is transverse due to a Bianchi identity. Thus, we know that we have to add a term $-m^2 h_{\mu\nu}$ to it. This is not enough, though: if we take the trace we find

$$\partial^2 h - \frac{m^2}{d-2} h - \partial_\mu \partial_\nu h^{\mu\nu} = 0. \tag{3.103}$$

This equation would give $h = 0$ if, instead of having just transversality, we had $\partial^\mu h_{\mu\nu} = \partial_\nu h$ and then we would recover transversality. Thus, we add a term $+m^2(h_{\mu\nu} - \eta_{\mu\nu} h)$ and obtain the massive Fierz–Pauli action and equation [388]

$$\boxed{\begin{aligned} S = \int d^d x \, \big\{ & \tfrac{1}{4} \partial^\rho h^{\mu\nu} \partial_\rho h_{\mu\nu} - \tfrac{1}{2} \partial^\rho h^{\mu\nu} \partial_\mu h_{\rho\nu} + \tfrac{1}{2} \partial^\mu h \partial^\lambda h_{\lambda\mu} - \tfrac{1}{4} \partial^\mu h \partial_\mu h \\ & - \tfrac{1}{4} m^2 \big(h^{\mu\nu} h_{\mu\nu} - h^2 \big) \big\}, \end{aligned}} \tag{3.104}$$

$$\mathcal{D}^{\mu\nu}_{(m)}(h) = \mathcal{D}^{\mu\nu}(h) + m^2 \big(h_{\mu\nu} - \eta_{\mu\nu} h \big) = 0, \tag{3.105}$$

from which we obtain, as expected,

$$h = 0, \qquad \partial_\mu h^{\mu\nu} = 0, \qquad \big(\partial^2 + m^2 \big) h_{\mu\nu} = 0. \tag{3.106}$$

To finalize our study of the free Fierz–Pauli theory, we can use the gauge symmetry to derive conserved currents along the path set out in Chapter 2. We have already calculated the direct variation of the Fierz–Pauli action under gauge transformations and have found invariance up to a total derivative. We now calculate the variation of the Fierz–Pauli action by performing first a general variation $\delta h_{\mu\nu}$, obtaining (after integration by parts) a total derivative term and the term proportional to $\delta h_{\mu\nu}$ whose coefficient is the equation of motion:

$$\begin{aligned} \delta S_{\mathrm{FP}} &= \int d^d x \left\{ \partial_\mu \left[\frac{\partial \mathcal{L}_{\mathrm{FP}}}{\partial \partial_\mu h_{\nu\rho}} \delta h_{\nu\rho} \right] - \partial_\mu \frac{\partial \mathcal{L}_{\mathrm{FP}}}{\partial \partial_\mu h_{\nu\rho}} \delta h_{\nu\rho} \right\} \\ &= \int d^d x \, \big\{ \partial_\mu \big[2 \partial_\sigma H^{\mu\sigma\nu\rho} \partial_\nu \epsilon_\rho \big] - \mathcal{D}^{\nu\rho}(h) \partial_\nu \epsilon_\rho \big\}, \end{aligned} \tag{3.107}$$

where we have used the explicit form of the gauge transformation. On integrating again by parts, we obtain the second form of the variation of the action,

$$\delta S_{\mathrm{FP}} = \int d^d x \, \big\{ \partial_\mu \big[2 \partial_\sigma H^{\mu\sigma\nu\rho} \partial_\nu \epsilon_\rho + \mathcal{D}^{\mu\rho}(h) \epsilon_\rho \big] - \partial_\nu \mathcal{D}^{\nu\rho}(h) \epsilon_\rho \big\}. \tag{3.108}$$

By identifying the two forms of the variation of the action and reasoning as in the Maxwell theory, we find the Bianchi identity (the terms proportional to the gauge-transformation parameter) Eq. (3.61) and the conserved current:

$$\begin{aligned} j^\mu_{\mathrm{N2}}(\epsilon) &= j^\mu_{\mathrm{N1}}(\epsilon) + \mathcal{D}^{\mu\nu}(h) \epsilon_\nu, \\ j^\mu_{\mathrm{N1}}(\epsilon) &= 2 \partial_\sigma H^{\mu\sigma\nu\rho} \partial_\nu \epsilon_\rho - 2 H^{\sigma\mu\nu\rho} \partial_\sigma \partial_\nu \epsilon_\rho. \end{aligned} \tag{3.109}$$

Using $\eta_{\rm AD}^{\mu\nu\rho}$ in Eq. (3.90), we can write the Fierz–Pauli wave operator as

$$\mathcal{D}^{\mu\nu}(h) = -2\partial_\sigma\partial_\lambda K^{\mu\lambda\rho\sigma}, \tag{3.110}$$

and, on substituting this into $j^{\mu}_{\rm N2}(\epsilon)$ above and making the obvious manipulations, it takes the form

$$\begin{aligned} j^{\mu}_{\rm N2}(\epsilon) = \partial_\nu\big\{-2\partial_\sigma K^{\mu\nu\rho\sigma}\epsilon_\rho + 2(H^{\mu\nu\sigma\rho} + K^{\mu\sigma\rho\nu})\partial_\sigma\epsilon_\rho\big\} \\ -2\big\{K^{\mu\nu\rho\delta} + H^{\mu\delta\nu\rho} + H^{\delta\mu\nu\rho}\big\}\,\partial_\delta\partial_\nu\epsilon_\rho. \end{aligned} \tag{3.111}$$

The second term in brackets is antisymmetric in $\delta\nu$ and vanishes. Then, we can write[19]

$$\begin{aligned} j^{\mu}_{\rm N2}(\epsilon) &= \partial_\nu j^{\nu\mu}_{\rm N2}(\epsilon), \\ j^{\nu\mu}_{\rm N2}(\epsilon) &= -2\partial_\sigma K^{\mu\nu\rho\sigma}\epsilon_\rho + 2(H^{\mu\nu\sigma\rho} + K^{\mu\sigma\rho\nu})\partial_\sigma\epsilon_\rho. \end{aligned} \tag{3.112}$$

We can now use this expression to calculate conserved charges associated with the gauge parameters ϵ_μ. Observe that the term proportional to H vanishes for ϵ that are Killing vectors of the Minkowski spacetime. We will come back to this point in Chapter 6.

The interpretation of the corresponding conserved charges is more complicated. In the cases in which ϵ is a Killing vector, a symmetry of Minkowski spacetime, we can associate these charges with momenta in the directions associated with those Killing vectors (linear or angular momenta). We will also discuss this point in Chapter 6.

3.2.3 Coupling to matter

As we discussed at the beginning of this section, the coupling of the Fierz–Pauli theory to matter is described by Eq. (3.60). To obtain this equation of motion from a Lagrangian, we will have to add to the Fierz–Pauli Lagrangian $\mathcal{L}_{\rm FP}(h)$ the matter Lagrangian $\mathcal{L}_{\rm matter}(\varphi)$ and a coupling term weighted by the gravitational coupling constant χ combined into a modified matter Lagrangian $\mathcal{L}_{\rm matter}(\varphi, h)$:

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_{\rm FP}(h) + \mathcal{L}_{\rm matter}(\varphi, h), \\ \mathcal{L}_{\rm matter}(\varphi, h) &= \mathcal{L}_{\rm matter}(\varphi) + \tfrac{1}{2}\chi h_{\mu\nu}T^{\mu\nu}_{\rm matter}(\varphi). \end{aligned} \tag{3.113}$$

From this Lagrangian, Eq. (3.60) follows. We also obtain an equation of motion for φ modified by the coupling to $h_{\mu\nu}$. The gauge identity implies that $T^{\mu\nu}_{\rm matter}(\varphi)$ has to be conserved, $\partial_\mu T^{\mu\nu}_{\rm matter}(\varphi) = 0$, for consistency. Furthermore, this Lagrangian is invariant (up to total derivatives) under the gauge transformations $\delta_\epsilon h_{\mu\nu} = -2\partial_{(\mu}\epsilon_{\nu)}$ only if $\partial_\mu T^{\mu\nu}_{\rm matter}(\varphi) = 0$.

Two questions now arise:

1. Which $T^{\mu\nu}_{\rm matter}(\varphi)$ should we use?
2. Is the conservation of $T^{\mu\nu}_{\rm matter}(\varphi)$ consistent with the modifications to the φ equations of motion introduced by the coupling to $h_{\mu\nu}$?

---

[19] As was explained in Chapter 2, this rewriting is not unique.

Let us first address the first question. The energy–momentum tensor on the r.h.s. of Eq. (3.60) has to be symmetric and divergenceless. These two properties are enjoyed by the *Belinfante energy–momentum tensor* of the *free* matter field theory, which, as explained in Chapter 2, is a symmetrization of the canonical energy–momentum tensor obtained by the addition of superpotential terms (which are identically divergenceless) and on-shell-vanishing terms. The Belinfante energy–momentum tensor is generally considered the energy–momentum tensor to which gravity couples minimally (see e.g. [939]).

There are many other symmetric energy–momentum tensors (in fact, an infinite number of them), such as the *improved* energy–momentum tensor associated with some scale-invariant theories. It can be argued that the improved energy–momentum tensor is in general associated with non-minimal couplings to gravity. The example discussed in Chapter 2 (a conformal scalar) should illuminate this point. In the simplest cases (scalar and vector field) the canonical and the Belinfante tensor are just what we need.[20] In Chapter 2 we also discussed an alternative prescription for how to find a symmetric, conserved, energy–momentum tensor that does not consist in finding some symmetric modification of the canonical energy–momentum tensor, viz. Rosenfeld's. In the scalar, vector, and symmetric-tensor cases that we are going to consider, the Rosenfeld and Belinfante energy–momentum tensors are going to be identical, and, therefore, this is the energy–momentum tensor that we are going to use.

Although the ultimate justification for Rosenfeld's prescription, whose logical connection to the physical concept of an energy–momentum tensor is obscure, relies on the final formulation of GR we are tied to, we can already see that the inclusion of the coupling to gravity in the matter action,

$$S_{\text{matter}}[\varphi, \eta_{\mu\nu}] + \int d^d x \chi h_{\mu\nu} \left. \frac{\delta S_{\text{matter}}[\varphi, \gamma_{\mu\nu}]}{\delta \gamma_{\mu\nu}} \right|_{\gamma_{\mu\nu}=\eta_{\mu\nu}}, \tag{3.114}$$

suggests that this is the beginning of a functional series expansion of the action functional $S_{\text{matter}}[\varphi, \gamma_{\mu\nu}]$ of a metric $\gamma_{\mu\nu} = \eta_{\mu\nu} + \chi h_{\mu\nu}$ around the vacuum metric $\eta_{\mu\nu}$,

$$\begin{aligned} S_{\text{matter}}[\varphi, \gamma_{\mu\nu}] = S_{\text{matter}}[\varphi, \eta_{\mu\nu}] &+ \int d^d x \chi h_{\mu\nu} \left. \frac{\delta S_{\text{matter}}[\varphi, \gamma_{\mu\nu}]}{\delta \gamma_{\mu\nu}} \right|_{\gamma_{\mu\nu}=\eta_{\mu\nu}} \\ &+ \int d^d x d^d x' \chi^2 h_{\mu\nu}(x) h_{\rho\sigma}(x') \left. \frac{\delta^2 S_{\text{matter}}[\varphi, \gamma_{\mu\nu}]}{\delta \gamma_{\mu\nu} \delta \gamma_{\rho\sigma}} \right|_{\gamma_{\mu\nu}=\eta_{\mu\nu}} + \cdots, \end{aligned} \tag{3.115}$$

truncated at first order.

As to the answer to the second question, we postpone it until we work out a simple example to show that the theory we have obtained indeed describes a SRFT of gravity that is compatible with our experience.

The gravitational field of a massive point-particle. Just as we did to derive the simplest predictions of the scalar SRFT of gravity, we are going to find the gravitational field produced

[20] A vector field in four dimensions is also invariant under dilatations and, in fact, under the whole conformal group. The improved energy–momentum tensor is, however, nothing but the Belinfante tensor.

by a massive point-particle of mass M placed at rest at the origin of coordinates in some inertial frame. In this calculation we are going to write all the factors of c that we usually omit in order to find the value of χ and have perhaps more-familiar expressions.

The action and energy–momentum tensor for a massive point-particle are given, respectively, by Eqs. (3.8) and (3.25), and the modified action that includes the coupling to gravity is, after the mutual elimination of the spacetime integral and the d-dimensional Dirac δ-function,

$$S_{\mathrm{pp}}[X^\mu] = -Mc\int d\xi \frac{1}{\sqrt{\eta_{\rho\sigma}\dot{X}^\rho\dot{X}^\sigma}}(\eta_{\mu\nu} + \tfrac{1}{2}\chi h_{\mu\nu}(X))\dot{X}^\mu\dot{X}^\nu. \tag{3.116}$$

Before solving any equations, we want to make the following two important observations [888]. First, as happens in general, this action is not invariant under the gauge transformations unless $\partial_\mu T^{\mu\nu}_{\mathrm{matter}}(\varphi) = 0$. However, it is invariant to lowest order in the coupling constant χ without this assumption if we transform the particle coordinates according to

$$\delta_\epsilon X^\mu = \chi\epsilon^\mu(X), \tag{3.117}$$

which is precisely the form of an infinitesimal GCT. This is the first sign of a relation between the gauge symmetry of the Fierz–Pauli field and spacetime transformations.

Second, there are fields $h_{\mu\nu}$ that are gauge-equivalent to zero, for instance [888]

$$h_{\mu\nu} = b_{\mu\nu} + a_{\mu\nu\rho}x^\rho, \tag{3.118}$$

with $b_{\mu\nu}$ and $a_{\mu\nu\rho}$ constants, can be canceled out by a gauge transformation,

$$\epsilon_\mu = \tfrac{1}{2}\big(b_{\mu\nu}x^\nu + a_{\mu\nu\rho}x^\nu x^\rho\big). \tag{3.119}$$

Combined with the previous observation, this means that, by a change of coordinates, we can remove certain gravitational fields. This fact is contained in the principle of equivalence of gravitation and inertia that was one of the basic postulates on which Einstein founded GR.

Now, let us consider the gravitational field equation[21]

$$\mathcal{D}^{\mu\nu}(h) = (\chi/c)T^{\mu\nu}_{\mathrm{pp}}. \tag{3.120}$$

The energy–momentum tensor has to be calculated on a solution of the equations of motion of a free particle $\dot{P}^\mu = 0$ plus $P_\mu P^\mu = M^2c^2$. A solution describing the particle at rest at the origin of coordinates is given by $X^i = 0$ and $\xi = X^0 = cT$. We can perform the integral over ξ eliminating the $\delta(X^0 - x^0)$. The energy–momentum tensor becomes[22]

$$T^{\mu\nu}_{\mathrm{pp}} = -Mc^2\eta^\mu{}_0\eta^\nu{}_0\delta^{(d-1)}(\vec{x}_{d-1}), \qquad \vec{x}_{d-1} = (x^1, \ldots, x^{d-1}), \tag{3.121}$$

and the gravitational field equations are

$$\mathcal{D}^{00}(h) = -\chi Mc\delta^{(d-1)}(\vec{x}_{d-1}), \qquad \mathcal{D}^{ij}(h) = 0. \tag{3.122}$$

---

[21] Let us recall that $S = (1/c)\int d^dx\,\mathcal{L}$. The Fierz–Pauli action does not acquire any factor of c; that is, $c^{-1}\mathcal{L}_{\mathrm{FP}} = \frac{1}{4}\partial_\mu h_{\nu\rho}\partial^\mu h^{\nu\rho} - \cdots$.

[22] We stress again that, with our conventions, T^{00} is negative-definite (it is minus the energy).

It is convenient to use the variable $\bar{h}_{\mu\nu}$ and the De Donder gauge[23] $\partial^\mu \bar{h}_{\mu\nu} = 0$. Then a solution can be immediately obtained for $d \geq 4$:

$$\bar{h}_{\mu\nu} = -\eta_{\mu 0}\eta_{\nu 0} \frac{\chi M c}{(d-3)\omega_{(d-2)}} \frac{1}{|\vec{x}_{d-1}|^{d-3}}, \tag{3.123}$$

and the non-vanishing components of $h_{\mu\nu}$ are[24]

$$h_{00} \equiv \frac{2}{\chi c^2}\phi, \quad h_{ii} = \frac{2}{(d-3)\chi c^2}\phi, \quad \phi = -\frac{\chi^2 M c^3}{2(d-2)\omega_{(d-2)}} \frac{1}{|\vec{x}_{d-1}|^{d-3}}. \tag{3.124}$$

The notation we have chosen suggests, correctly, that ϕ can be identified with the Newtonian potential as in the scalar SRFT of gravity (Eq. (3.7)). Also, as in the case of the scalar SRFT of gravity, we have to see how it affects the motion of test particles in order to confirm it.

The gravitational field of a massless point-particle. The action and energy–momentum tensor for a free massless particle moving in Minkowski spacetime are given, respectively, by Eqs. (3.32) and (3.34). After coupling to the gravitational field $h_{\mu\nu}$, the modified action is

$$S[X^\mu(\xi), \gamma(\xi)] = -\frac{p}{2}\int d\xi \sqrt{\gamma}\, \gamma^{-1} \left[\eta_{\mu\nu} + \chi h_{\mu\nu}(X)\right] \dot{X}^\mu \dot{X}^\nu. \tag{3.125}$$

This time the gravitational field cannot be absorbed into a redefinition of the worldline metric γ (unless $h_{\mu\nu} \propto \eta_{\mu\nu}$) and a massless particle interacts with the gravitational field.

Let us first find the gravitational field produced by a massless particle by solving the equation $\mathcal{D}^{\mu\nu}(h) = (\chi/c) T_{\rm pp}^{\mu\nu}$, where the energy–momentum tensor has to be calculated for a solution of the equations of motion of the free massless particle $\dot{P}^\mu = P^\mu P_\mu = 0$. It is convenient to use light-cone coordinates u, v, and $\vec{x}_{d-2}$ defined by

$$u = \frac{1}{\sqrt{2}}(t - z), \qquad v = \frac{1}{\sqrt{2}}(t + z), \qquad (\vec{x}_{d-2}) = (x^1, \ldots, x^{d-2}), \tag{3.126}$$

where $z \equiv x^{d-1}$, in which the Minkowski metric takes the form

$$\left(\eta_{\mu\nu}\right) = \left(\begin{array}{cc|c} 0 & 1 & \\ 1 & 0 & \\ \hline & & -\mathbb{I}_{(d-2)\times(d-2)} \end{array}\right). \tag{3.127}$$

A solution describing the particle moving at the speed of light along the z axis toward $+\infty$ is given by

$$U = \vec{X}_{d-2} = 0, \qquad V = \xi, \qquad \gamma = 1. \tag{3.128}$$

---

[23] In this case we cannot impose a traceless gauge because the particle's energy–momentum tensor itself is not traceless.

[24] To compare this with Thirring's results [888] it has to be taken into account that Thirring's energy–momentum tensor is twice ours and that its coupling constant $f = \chi/2$.

For this solution, the energy–momentum tensor (3.34) takes, after integration of one of the Dirac delta-function components, the form

$$T_{\rm pp}^{\mu\nu} = -pc\ell^{\mu}\ell^{\nu}\int d\xi\,\delta(\sqrt{2}\xi - u)\delta(u)\delta^{(d-2)}(x^i), \qquad \ell^{\mu} = \delta^{\mu}{}_{v}. \tag{3.129}$$

On integrating over ξ and substituting this into the gravitational equation with $h_{\mu\nu}$ in the transverse, traceless gauge, we arrive at the equation

$$\partial^2 h^{\mu\nu} = -\sqrt{2}p\chi\,\ell^{\mu}\ell^{\nu}\delta(u)\delta^{(d-2)}(x^i). \tag{3.130}$$

Only one component of $h^{\mu\nu}$ will be non-trivial. We define the function $K(u,\vec{x}_{d-2})$ by

$$\chi h^{\mu\nu} = 2K(u,\vec{x}_{d-2})\ell^{\mu}\ell^{\nu}, \tag{3.131}$$

which satisfies

$$\vec{\partial}^{\,2}_{d-2}K(u,\vec{x}_{d-2}) = \frac{p\chi^2}{\sqrt{2}}\delta(u)\delta^{(d-2)}(\vec{x}_{d-2}). \tag{3.132}$$

A solution can immediately be found. For $d \geq 5$ we have

$$K(u,\vec{x}_{d-2}) = \frac{p\chi^2}{\sqrt{2}(d-4)\omega_{(d-3)}}\frac{1}{|\vec{x}_{d-2}|^{d-4}}\delta(u), \tag{3.133}$$

and, for $d = 4$,

$$K(u,\vec{x}_2) = -\frac{p\chi^2}{\sqrt{2}2\pi}\ln|\vec{x}_2|\delta(u). \tag{3.134}$$

This solution describes a sort of gravitational shock wave. We will see in Chapter 10 that this result, which was found in a linear theory, is actually exact in GR and corresponds to the Aichelburg–Sexl solution found in [24] by completely different means.

Motion of massive and massless test particles in a gravitational field. We can now plug any of the two solutions we have found into the actions (3.116) and (3.125) to find the dynamics of a second test particle of mass m or of a second test massless particle in the gravitational field created by the first particle.[25] Clearly, the most important case is the one corresponding to motion in the field of a massive particle, whose mass we will denote by M. We first study the massive case, since it is the one that has a non-relativistic limit. Using the static gauge $\xi = X^0 = cT$ (we write t instead of T), we find

$$S_{\rm pp}[X] = -mc^2\int dt\left\{\sqrt{1-(v/c)^2} + \frac{1}{\sqrt{1-(v/c)^2}}\left[1 + \frac{1}{d-3}\left(\frac{v}{c}\right)^2\right]\frac{\phi}{c^2}\right\}, \tag{3.135}$$

and, in the non-relativistic limit in which we ignore terms of order higher than $\mathcal{O}[(v/c)^4]$ and the constant term, we find

$$S_{\rm pp}[X] = \int dt\left\{\tfrac{1}{2}mv^2 - m\phi - \tfrac{1}{4}mv^2\left(\frac{v}{c}\right)^2 + \frac{d-1}{2(d-3)}mv^2\frac{\phi}{c^2}\right\}. \tag{3.136}$$

[25] Test particle meaning that the effect of its own gravitational field on the first particle (and, correspondingly, on the gravitational field created by the first particle) can be ignored.

The first term is the kinetic energy of a particle of inertial mass m and the second term is (minus) the potential energy of a particle of *gravitational mass m* moving in a Newtonian gravitational potential ϕ (confirming the definition of ϕ). In this scheme the gravitational and inertial masses of a particle are identical. This is essentially the content of the *principle of equivalence of gravitation and inertia* in its weak form, as we will see in Section 3.3. This was also the case for the scalar SRFT of gravity and it is the consequence of taking the energy–momentum tensor (or its trace) as the source for the gravitational field.

There are also two correction terms. One is the standard relativistic correction to the kinetic energy of a free particle and the other correction represents the contribution of the kinetic energy to the gravitational interaction. A similar term was present in the scalar SRFT of gravity (compare with Eq. (3.28)), but with a coefficient that is different in absolute value and sign. Thus, in this case, all gravitational effects will not vanish in the $v \to c$ limit. On the contrary, we see that, due to the sign of the fourth term, the kinetic energy also feels and is a source of gravity just like the (inertial/gravitational) rest mass.

We can now check that the value of ϕ that we have obtained from our relativistic gravitational theory is correct (i.e. coincides with the Newtonian potential created by a mass M). In $d = 4$

$$\phi = -\frac{\chi^2 c^3}{16\pi}\frac{M}{|\vec{x}_3|}, \quad \Rightarrow \chi^2 = \frac{16\pi G_N^{(4)}}{c^3}, \tag{3.137}$$

where $G_N^{(4)}$ is the Newton constant. The force between the masses m and M is then

$$\vec{F} = -m\vec{\nabla}\phi = -G_N^{(4)} m M \frac{\vec{x}_3}{|\vec{x}_3|^3}. \tag{3.138}$$

For higher dimensions the functional form of ϕ is correct. It is (unfortunately) customary in the literature to define in any dimension d

$$\chi^2 = 16\pi G_N^{(d)}/c^3, \tag{3.139}$$

even though the rational definition would have been

$$\chi^2 = 2(d-2)\omega_{(d-2)} G_N^{(d)}/c^3.$$

With these conventions the force between the masses m and M is

$$\vec{F} = -m\vec{\nabla}\phi = -\frac{8(d-3)\pi G_N^{(d)} m M}{(d-2)\omega_{(d-2)}}\frac{\vec{x}_{d-1}}{|\vec{x}_{d-1}|^{d-1}}. \tag{3.140}$$

Before we use the fully relativistic action to find corrections to Keplerian orbits, etc., there is one more point worth discussing. We have learned how the Newtonian gravitational field is encoded in the relativistic field $h_{\mu\nu}$. Of course, the relativistic field has more components and at least one more degree of freedom. We can compare this situation with that of the electrostatic field: to build a relativistic theory of the electrostatic field we would have had to use a vector field (with a scalar field we would never have been able to describe attraction between opposite charges and repulsion between like charges) that has more components. Then we could have discovered the magnetic field as part of the electromagnetic field and we would have discovered electromagnetic radiation. Thus, just to see what other non-relativistic terms the full action for general $h_{\mu\nu}$ produces, let us go back to the action Eq. (3.125), choose the static gauge again, and, instead of substituting the $h_{\mu\nu}$ we obtained for a static point-like charge, let us consider a general background gravitational field and let us make the definition

$$h_{0i} = \frac{1}{\chi c^2} A_i . \tag{3.141}$$

Then, in the non-relativistic limit and ignoring $\mathcal{O}(hv^2)$ terms, we have

$$S_{\rm pp} \sim \int dt \left\{ \tfrac{1}{2} m v^2 - m\phi + \frac{m}{c} \vec{A} \cdot \vec{v} - mc^2 \right\} . \tag{3.142}$$

The new term is a non-Newtonian velocity-dependent interaction. The whole action is identical to the action of a charged particle in an electromagnetic field (8.55). Then, by analogy, the last term describes the interaction of the particle with the *gravitomagnetic* field, whose existence is one of the main predictions of any relativistic theory of gravitation (including GR) but has not yet been detected (see e.g. [242]). The Newtonian term is also called, by analogy, the *gravitostatic* potential.

We are now ready to calculate the corrections to Keplerian orbits of planets predicted by this theory. The main effect will be the precession of the perihelion of planets, a secular, cumulative effect that was known before Einstein's construction of GR and whose explanation by this theory was one of its early successes.

Our starting point will be Eq. (3.116) (with M replaced by m). We consider only the $d = 4$ case. First, we rewrite this action in terms of an action for a particle moving in the background of an effective metric field $g_{\mu\nu}$:

$$S_{\rm pp}[X^\mu] = -mc \int d\xi \sqrt{g_{\mu\nu}(X) \dot{X}^\rho \dot{X}^\sigma}, \qquad g_{\mu\nu} \equiv \eta_{\mu\nu} + \chi h_{\mu\nu}(X), \tag{3.143}$$

which is equivalent to our original action Eq. (3.116). As we explained, this can always be done and it is the basis of Rosenfeld's prescription for calculating a symmetric energy–momentum tensor. The Hamilton–Jacobi equation associated with this action is [644]

$$g^{\mu\nu}(X) \frac{\partial S_{\rm pp}}{\partial X^\mu} \frac{\partial S_{\rm pp}}{\partial X^\nu} - m^2 c^2 = 0, \tag{3.144}$$

and, to first order in χ, it is valid also for our original action. Let us now consider a general static, spherically symmetric metric written as follows:

$$ds^2 = \lambda(r) c^2 dt^2 - \mu(r) dr^2 - R^2(r) d\Omega^2, \qquad d\Omega^2 = d\theta^2 + \sin^2\theta \, d\varphi^2, \tag{3.145}$$

and, knowing that all the dynamics will take place in a plane, let us set $\theta = \pi/2$ from now on. The Hamilton–Jacobi equation takes the form

$$\frac{1}{\lambda c^2} \left(\partial_t S_{\rm pp}\right)^2 - \frac{1}{\mu} \left(\partial_r S_{\rm pp}\right)^2 - \frac{1}{\mu R^2} \left(\partial_\varphi S_{\rm pp}\right)^2 - m^2 c^2 = 0. \tag{3.146}$$

$S_{\rm pp}$ has the form

$$S_{\rm pp} = -Et + l\varphi + W, \tag{3.147}$$

where W is a function only of r. On substituting into the above equation, we find that W is given by

$$W = \int dr \sqrt{\mu\lambda^{-1}\left(\frac{E}{c}\right)^2 - \frac{l^2}{R^2} - m^2c^2\mu}. \tag{3.148}$$

In the absence of the gravitational field $\lambda = \mu = 1$, and $R = r$. On defining the non-relativistic energy $E' = E - mc^2$, assuming that $E' \ll mc^2$ so that

$$\left(\frac{E}{c}\right)^2 - m^2c^2 = m^2c^2\left[\left(\frac{E}{mc^2}\right)^2 - 1\right] = m^2c^2\left[\left(\frac{E'}{mc^2}\right)^2 + 2\frac{E'}{mc^2}\right] \sim 2mE', \tag{3.149}$$

and substituting in the integrand, we obtain W for a classical free particle of energy E'. In the presence of a spherically symmetric gravitational field, vanishing at infinity, on making the same approximation $E' \ll mc^2$, expanding

$$\lambda \sim 1 + \frac{\lambda_1}{r} + \frac{\lambda_2}{r^2} + \cdots, \qquad \mu \sim 1 + \frac{\mu_1}{r} + \frac{\mu_2}{r^2} + \cdots,$$
$$R^2 \sim r^2\left(1 + \frac{R_1}{r} + \cdots\right), \tag{3.150}$$

and expanding the expression under the square root to order $\mathcal{O}(1/r^2)$, we find

$$W \sim \int dr \sqrt{2mE' - \frac{\lambda_1 m^2c^2}{r} - \frac{l^2 - [\lambda_1(\lambda_1 - \mu_1) - \lambda_2]m^2c^2}{r^2}}. \tag{3.151}$$

For the solution Eq. (3.124)

$$\mu_1 = -\lambda_1 = R_S \equiv 2MG_N^{(4)}/c^2, \qquad R_1 = 0, \tag{3.152}$$

where we have introduced R_S, the *Schwarzschild* or *gravitational* radius of an object of mass M, and we obtain from Eq. (3.151)

$$W \sim \int dr \sqrt{2mE' + \frac{R_S m^2c^2}{r} - \frac{l^2 - 2R_S^2 m^2c^2}{r^2}}. \tag{3.153}$$

We should first compare this expression with the Newtonian expression[26]

$$W_{\text{Newtonian}} = \int dr \sqrt{2mE' + \frac{R_S m^2c^2}{r} - \frac{l^2}{r^2}}. \tag{3.154}$$

The second term is the Newtonian potential energy. We see in Eq. (3.153) that there is an $\mathcal{O}(1/r^2)$ relativistic correction to the Newtonian potential. The main consequence will be that the orbits will not be closed and the perihelions will shift. To evaluate the angular difference between two consecutive perihelions we reason, following [644], as follows. The equation for the orbit can be found from

$$\varphi = \beta_\varphi - \frac{\partial W}{\partial l}. \tag{3.155}$$

[26] We assume that $M \gg m$ so that the reduced mass can be approximated by m.

In a complete revolution

$$\Delta\varphi = -\frac{\partial}{\partial l}\Delta W. \tag{3.156}$$

By expanding W around the Newtonian $W_{\text{Newtonian}}$ as a power series in the relativistic correction $\delta = 2R_{\text{S}}^2 m^2 c^2$ and observing that

$$\left.\frac{\partial W}{\partial \delta}\right|_{\delta=0} = -\frac{\partial W}{\partial l^2}, \tag{3.157}$$

we obtain

$$\begin{aligned} W \sim\ W|_{\delta=0} + \delta\left.\frac{\partial W}{\partial \delta}\right|_{\delta=0} &= W_{\text{Newtonian}} - \delta\frac{1}{2l}\frac{\partial W_{\text{Newtonian}}}{\partial l} \\ &= W_{\text{Newtonian}} - \frac{R_{\text{S}}^2 m^2 c^2}{l}\frac{\partial W_{\text{Newtonian}}}{\partial l}, \end{aligned} \tag{3.158}$$

and

$$\begin{aligned} \Delta W &= \Delta W_{\text{Newtonian}} - \frac{R_{\text{S}}^2 m^2 c^2}{l}\frac{\partial \Delta W_{\text{Newtonian}}}{\partial l} \\ &= \Delta W_{\text{Newtonian}} + \frac{R_{\text{S}}^2 m^2 c^2}{l}\Delta\varphi_{\text{Newtonian}}, \end{aligned} \tag{3.159}$$

where we have used Eq. (3.156) for $W_{\text{Newtonian}}$. On substituting this into Eq. (3.156) we find

$$\Delta\varphi = \Delta\varphi_{\text{Newtonian}} + \frac{R_{\text{S}}^2 m^2 c^2}{l^2}\Delta\varphi_{\text{Newtonian}}. \tag{3.160}$$

Newtonian orbits are closed, so in one revolution $\Delta\varphi_{\text{Newtonian}} = 2\pi$ and the deviation from the Newtonian value is, according to this theory

$$\delta\varphi = \frac{2\pi R_{\text{S}}^2 m^2 c^2}{l^2}. \tag{3.161}$$

This result is $\frac{4}{3}$ of the actual value; that is, it is close (better than the value given by the scalar SRFT of gravity) but not quite right. We will have to find a correction to our theory in order to obtain the right value.

The second effect that we want to calculate is the deflection of a light ray (or a massless particle) by the central gravitational field of a massive body, given by Eq. (3.124). To first order in χ we can simply take the Hamilton–Jacobi equation for a relativistic massive particle, Eq. (3.144), and set $m = 0$ [644]. The resulting equation can be solved as in the massive case with the replacement of $E = -\partial_t S$ by $\omega = -\partial_t S$. For W we obtain the equation

$$W = \int dr\sqrt{\mu\lambda^{-1}\left(\frac{\omega}{c}\right)^2 - \frac{l^2}{R^2}}. \tag{3.162}$$

On expanding μ and λ in powers of $1/r$, we obtain, for the solution Eq. (3.124),

$$W \sim \int dr\sqrt{\left(\frac{\omega}{c}\right)^2 + 2R_{\text{S}}\left(\frac{\omega}{c}\right)^2\frac{1}{r} - \frac{l^2}{r^2}}. \tag{3.163}$$

The $1/r$ term is not present[27] in the Newtonian case[28] and, as we did before, we expand W around its Newtonian value,

$$
\begin{aligned}
W &\sim \int dr \sqrt{\left(\frac{\omega}{c}\right)^2 - \frac{l^2}{r^2}} + 2R_S \frac{\partial}{\partial x} \int dr \sqrt{\left(\frac{\omega}{c}\right)^2 + \left(\frac{\omega}{c}\right)^2 \frac{x}{r} - \frac{l^2}{r^2}} \Bigg|_{x=0} \\
&\sim W_{\text{Newtonian}} + \frac{R_S \omega}{c} \int dr \frac{1}{\sqrt{r^2 - \rho^2}} \\
&\sim W_{\text{Newtonian}} + \frac{R_S \omega}{c} \operatorname{arccosh}\left(\frac{r}{\rho}\right),
\end{aligned}
\tag{3.164}
$$

where $\rho = cl/\omega$ is clearly the minimal value of r in the path of the massless particle. Following [644], the variation of W when the particle starts from $r = R \gg \rho$, goes through $r = \rho$, and again reaches $r = R$ is

$$
\Delta W \sim \Delta W_{\text{Newtonian}} + \frac{2R_S \omega}{c} \operatorname{arccosh}\left(\frac{R}{\rho}\right),
\tag{3.165}
$$

and, according to Eq. (3.156),

$$
\Delta\varphi \sim \frac{\partial}{\partial l} \Delta W_{\text{Newtonian}} + \frac{2R}{\rho} \frac{1}{\sqrt{1 - \rho/R}} \overset{R\to\infty}{\longrightarrow} \pi + \frac{2R}{\rho},
\tag{3.166}
$$

and we find that the deviation from the Newtonian value $\Delta\varphi = \pi$ (which means simply no bending of the light ray) is $\delta\varphi = 2R/\rho$, in good agreement with observation. This is an encouraging result, which indicates that we have found a reasonable relativistic theory of gravitation worth studying in more detail.

At this point, we remember that we still have to answer the second question posed on page 67. The answer will prompt us to seek and introduce into our theory corrections that will make the prediction for the precession of the perihelion of Mercury agree completely with observations.

3.2.4 The consistency problem

The answer to the second question formulated on page 67 is that, in general, the matter energy–momentum tensor derived from the free-matter Lagrangian is no longer conserved. As explained in Chapter 2, the divergence of the energy–momentum tensor is proportional to the equations of motion derived from the same Lagrangian, but the coupling to gravity changes these equations. This can be seen in the modified massive-particle action of the above example but the real scalar field which we studied in Chapter 2 will, however, make a better example.

---

[27] There are also $1/r^2$ corrections, but we take only the most important one.

[28] The Newtonian case corresponds to a *free* massive particle (i.e. vanishing gravitational potential energy) moving at the speed of light with $2mE' = (\omega/c)^2$.

The modified matter Lagrangian and equation of motion are

$$\begin{aligned}\mathcal{L}_{\text{matter}}(\varphi, h) &= \tfrac{1}{2}(\partial\varphi)^2 + \tfrac{1}{2}\chi h_{\mu\nu} T_{\text{matter}}{}^{\mu\nu}(\varphi) \\ &= \tfrac{1}{2}\left(\eta^{\mu\nu} - \chi \bar{h}^{\mu\nu}\right)\partial_\mu\varphi\partial_\nu\varphi, \\ 0 &= \partial_\mu\left[\left(\eta^{\mu\nu} - \chi\bar{h}^{\mu\nu}\right)\partial_\nu\varphi\right].\end{aligned} \tag{3.167}$$

Using the new equation of motion

$$\partial_\mu T_{\text{matter}}{}^{\mu\nu}(\varphi) = -\partial_\mu\left(\bar{h}^{\mu\rho}\partial_\rho\varphi\right)\partial^\nu\varphi, \tag{3.168}$$

which is not zero, implying that the first-order matter–gravity coupled system is inconsistent. This is the essence of the consistency problem of the Fierz–Pauli theory.

How could we overcome this problem? One solution is to modify the equation of motion Eq. (3.60) by adding a term on the r.h.s. to make it divergenceless again, consistently with the new equation of motion for matter.[29] In fact, since we have modified the matter Lagrangian to include the coupling, the energy–momentum tensor has also been modified and we should replace $T^{\mu\nu}_{\text{matter}}(\varphi)$ by $T^{\mu\nu}_{\text{matter}}(\varphi, h)$ calculated from $\mathcal{L}_{\text{matter}}(\varphi, h)$. This, however, does not work because, if we include the coupling term in the calculation of the energy–momentum tensor, we should also include the Fierz–Pauli Lagrangian: only the *total* energy–momentum tensor (matter plus gravity plus interactions) is conserved. Clearly this is the physical principle behind our problem.

The situation is not too different from the ones encountered in Section 2.5 in the coupling of Abelian and non-Abelian vector fields to matter. There one also has to take into account the contribution of the vector fields themselves to the full Noether currents, since only then are these conserved.

It is reasonable to expect that full consistency can be achieved only if we can derive the new equation of motion from a Lagrangian. However, to make the correction to the energy–momentum tensor appear in the equation of motion, we have to add new terms to the Lagrangian, which introduce new modifications into the energy–momentum tensor, and so on. This problem is present in the pure-gravity system once we accept that it has to

[29] There is another possibility, proposed and studied in [308], in which consistency without addition of extra terms is recovered at the expense of locality: use on the r.h.s. of the gravitational equation a divergence-free projection of the matter energy–momentum tensor $J_{\mu\nu}$ obtained by applying the manifestly divergence-free Lorentz-covariant projection operator

$$P_{\mu\nu} \equiv \eta_{\mu\nu} - \frac{\partial_\mu\partial_\nu}{\partial^2}. \tag{3.169}$$

The most general divergence-free definition of $J_{\mu\nu}$ is

$$J_{\mu\nu} = \left(P_{\mu\alpha}P_{\nu\beta} + pP_{\mu\nu}\eta_{\alpha\beta} + qP_{\mu\nu}P_{\alpha\beta}\right)T^{\alpha\beta}_{\text{matter}}. \tag{3.170}$$

Thus, the gravitational field couples only to matter, but in this consistent way. The constants p and q are fixed so as to obtain the right predictions for the classical tests of GR. Only two sets of values of p and q are admissible (all the classical tests are passed by the theory) and for one of them, $q = -p = 1$, the theory can be written in a local form with the introduction of auxiliary fields. In this form, it is shown that there are propagating spin-0 degrees of freedom in the theory. Clearly, this theory cannot pass tests in which the self-coupling of the gravitational field (the *strong form of the principle of equivalence*) is probed and it will predict, for instance, a finite value for the Nordtvedt effect (see, for instance, Chapter 3 in [242] and references therein).

couple to itself through its own energy–momentum *tensor*,[30] customarily denoted by $t^{\mu\nu}$, in the same form and with the same strength as it does to matter. This coupling encodes the *strong form of the principle of equivalence*.

In conclusion, we can say that we have achieved consistency if the equations of motion

$$\mathcal{D}^{\mu\nu}(h) = \chi\left[T^{\mu\nu}_{\text{matter}}(\varphi, h) + t^{\mu\nu}\right] \tag{3.171}$$

are consistent with the equations of motion for matter, i.e.

$$\partial_\mu\left[T^{\mu\nu}_{\text{matter}}(\varphi, h) + t^{\mu\nu}\right] = 0, \tag{3.172}$$

on-shell. Equivalently, we can say that the corrected theory is consistent if we can derive the above equation of motion from a Lagrangian and derive the total energy–momentum tensor $T^{\mu\nu}_{\text{matter}}(\varphi, h) + t^{\mu\nu}$ from the same Lagrangian.

It is interesting to try to find at least the first correction.[31] We can follow an iterative procedure that stresses the importance of symmetry: the Noether method, explained in Section 2.5 and applied there to the problem of finding consistent coupling of Abelian and non-Abelian vector fields to charged matter. This case will be much more complex but what we will learn will be worth the effort. In the following section we will give a very elegant and economic argument due to Deser [299] to prove that GR is a self-consistent extension of the Fierz–Pauli theory. In this setup, only one iteration will be necessary.

When a solution to a problem is found, the problem of the uniqueness of that solution arises. The results of Weinberg [941, 942] and Boulware and Deser [176], mentioned at the beginning of this section, indicate that a quantum massless spin-2 theory can have a Lorentz-invariant quantum S matrix only if it couples to the *total* energy–momentum tensor, including the gravitational energy–momentum tensor whose form, *in the long-wavelength limit*, is the one predicted by GR, Eq. (3.200). Thus, *any interacting quantum theory of a spin-2 particle coincides with GR in the infrared limit*.[32]

The approach that we are going to follow stresses the importance of the conservation of the total energy–momentum tensor and its relation to gauge symmetry and it is motivated by the hypothesis of the coupling of the spin-2 field to the matter energy–momentum tensor.[33] Other approaches have tried to determine the most general self-interacting classical SRFT of a spin-2 particle, not using as input the coupling to matter and trying to derive the gauge invariance from the requirement of self-consistency of the equations of motion. We will discuss this approach and its results at the end.

3.2.5 The Noether method for gravity

We start with the Fierz–Pauli Lagrangian Eq. (3.84) plus the Lagrangian for a real scalar

$$\mathcal{L}^{(0)} = \mathcal{L}_{\text{FP}} + \mathcal{L}_{\text{matter}}(\varphi), \qquad \mathcal{L}_{\text{matter}}(\varphi) = \tfrac{1}{2}(\partial\varphi)^2. \tag{3.173}$$

---

[30] It is a Lorentz tensor. In the full GR theory it will still be a Lorentz tensor but not a tensor under GCTs and that is why it will be called in that context the *gravity energy–momentum pseudotensor.*

[31] It is enough to obtain the correct value for the precession of the perihelion of Mercury. The derivation of all the corrections is sometimes called *Gupta's program* [378].

[32] Actually, string theory contains corrections to GR in the ultraviolet limit.

[33] The non-interacting theory is perfectly consistent as it stands.

This Lagrangian is invariant under the local gauge transformations given in Eq. (3.95) with parameter $\epsilon^\mu(x)$ and *global* translations with constant parameter ξ^μ (just like any SRFT):

$$\begin{aligned}\tilde{\delta} x^\mu &= \chi \xi^\mu, \\ \delta h_{\mu\nu} &= -2\partial_{(\mu}\epsilon_{\nu)} - \chi \xi^\lambda \partial_\lambda h_{\mu\nu}, \\ \delta\varphi &= -\chi\xi^\lambda\partial_\lambda\varphi.\end{aligned} \tag{3.174}$$

Both symmetries are Abelian. The conserved current associated with the global symmetry can be found by performing a local transformation of the same form in the Lagrangian, as explained in Section 2.5. Up to total derivatives

$$\begin{aligned}\delta_{\xi(x)}\mathcal{L}_{\text{matter}}(\varphi) &= -\chi\xi^\sigma(x)\partial_\mu T_{\text{can}}{}^\mu{}_\sigma(\varphi), \\ \delta_{\xi(x)}\mathcal{L}_{\text{FP}} &= -\chi\xi^\sigma(x)\partial_\mu + t^{(0)\mu}_{\text{can}}{}_\sigma(h), \\ T_{\text{can}}{}^\mu{}_\sigma(\varphi) &= -\partial^\mu\varphi\partial_\sigma\varphi + \tfrac{1}{2}\eta^\mu{}_\sigma(\partial\varphi)^2, \\ t^{(0)\mu}_{\text{can}}{}_\sigma(h) &= -\tfrac{1}{2}\partial^\mu h^{\nu\rho}\partial_\sigma h_{\nu\rho} + \partial^\nu h^{\mu\rho}\partial_\sigma h_{\nu\rho} - \tfrac{1}{2}\partial_\lambda h^{\lambda\mu}\partial_\sigma h \\ &\quad - \tfrac{1}{2}\partial_\sigma h^{\mu\rho}\partial_\rho h + \tfrac{1}{2}\partial^\mu h\partial_\sigma h + \eta^\mu{}_\sigma\mathcal{L}_{\text{FP}}.\end{aligned} \tag{3.175}$$

Here $T_{\text{can}}{}^\mu{}_\sigma(\varphi)$ is the canonical energy–momentum tensor of the real scalar field and $t^{(0)\mu}_{\text{can}}{}_\sigma(h)$ that of the gravitational field. The latter is not symmetric. Both are separately conserved on-shell. In particular

$$\partial_\mu t^{(0)\mu}_{\text{can}}{}_\sigma(h) = -\tfrac{1}{2}\partial_\sigma h_{\nu\rho}\mathcal{D}^{\nu\rho}(h). \tag{3.176}$$

Our physical problem is to couple consistently these two fields, which requires the self-coupling of the gravity field. From the symmetry point of view, following the Noether philosophy, we will have a consistent theory if we manage to construct a theory that is invariant under the *local* versions of these two symmetries. Since, under local transformations, the Lagrangian transforms as above (up to total derivatives that we will systematically ignore here) it is reasonable to expect that we will have invariance to first order in the coupling constant χ if we introduce an interaction term of the typical form

$$\mathcal{L}^{(1)} = \mathcal{L}^{(0)} + \tfrac{1}{2}\chi h^{\mu\sigma}\left[T_{\text{can}\,\mu\sigma}(\varphi) + t^{(0)}_{\text{can}\,\mu\sigma}(h)\right] \tag{3.177}$$

and we identify the two local parameters $\xi^\mu(x) = \epsilon^\mu(x)$. This identification is also suggested by the observation that the point-particle action coupled to gravity is gauge-invariant only if we complement the gauge transformation of $h_{\mu\nu}$ with a local transformation of the particle's coordinates. It is clear, however, that this is too naive: from the above Lagrangian one cannot obtain the consistent equation of motion (3.171) because the variation of the interaction term with respect to $h_{\mu\nu}$ does give $\chi T^{(0)}_{\text{can}\,\mu\sigma}(\varphi)$ on the r.h.s. but not the corresponding term for the gravitational field (unless some miracle happens, which it does not). Thus, we will have to look for a term quadratic in derivatives of h, symbolically $\mathcal{L}^{(1)}{}_{\mu\nu}(\partial h\partial h)$, and different from the energy–momentum tensor such that the Lagrangian

$$\mathcal{L}^{(1)} = \mathcal{L}^{(0)} + \tfrac{1}{2}\chi h^{\mu\sigma}[T_{\text{can}\,\mu\sigma}(\varphi) + \mathcal{L}^{(1)}{}_{\mu\sigma}] \tag{3.178}$$

produces the wanted equations of motion and is invariant up to $\mathcal{O}(\chi^2)$ under the corrected transformations with local parameter ϵ^μ,

$$\begin{aligned} \delta_\epsilon^{(1)} h_{\mu\nu} &= -2\partial_{(\mu}\epsilon_{\nu)} - \chi\epsilon^\lambda\partial_\lambda h_{\mu\nu}, \\ \delta_\epsilon^{(1)}\varphi &= -\chi\epsilon^\lambda\partial_\lambda\varphi. \end{aligned} \tag{3.179}$$

This is just the simplest possibility. Clearly there are infinitely many local transformations that reduce to some given global transformations, all of them different by terms proportional to the derivatives of the gauge parameters. We need additional criteria in order to find the right $\partial\epsilon$ terms here. The main property that gauge transformations have to enjoy is that they must generate one and the same algebra both on φ and on $h_{\mu\nu}$. Then, given two transformations $\delta^{(1)}$ in Eqs. (3.179) with infinitesimal parameters ϵ_1 and ϵ_2, their commutator, applied to φ and $h_{\mu\nu}$, must give another transformation $\delta^{(1)}$ with an infinitesimal parameter ϵ_3 that should be a function of ϵ_1 and ϵ_2; that is,

$$[\delta_{\epsilon_1}^{(1)}, \delta_{\epsilon_2}^{(1)}] = \delta_{\epsilon_3(\epsilon_1,\epsilon_2)}^{(1)}. \tag{3.180}$$

The simple transformations Eqs. (3.179) do not have this property. The problem of finding the most general gauge transformations which have this property, and reduce at order zero in χ to the normal spin-2 gauge transformations of $h_{\mu\nu}$, was considered by Ogievetsky and Polubarinov in [739]. Their conclusion, which is similar (in spite of the different setup) to Wald's in [933], is that, apart from the $\chi = 0$ Abelian transformations, the only gauge transformations with the required properties are

$$\begin{aligned} \delta_\epsilon^{(1)} h_{\mu\nu} &= -2\partial_{(\mu}\epsilon_{\nu)} - \chi\left[\epsilon^\lambda\partial_\lambda h_{\mu\nu} + 2\partial_{(\mu}\epsilon^\lambda h_{\nu)\lambda}\right] = -2\partial_{(\mu}\epsilon_{\nu)} - \chi\mathcal{L}_\epsilon h_{\mu\nu}, \\ \delta_\epsilon^{(1)}\varphi &= -\chi\epsilon^\lambda\partial_\lambda\varphi = -\chi\mathcal{L}_\epsilon\varphi, \end{aligned} \tag{3.181}$$

and similarly for matter tensor fields of other ranks. These transformations have the algebra of infinitesimal GCTs

$$[\delta_{\epsilon_1}^{(1)}, \delta_{\epsilon_2}^{(1)}] = \delta_{[\epsilon_1,\epsilon_2]}^{(1)}, \tag{3.182}$$

where $[\epsilon_1, \epsilon_2]$ is the Lie bracket of the two vector fields.

For a scalar field, the Noether current associated with these transformations (which are not symmetries of the action) is the canonical energy–momentum tensor, as in Eqs. (3.175). However, for a vector field with action Eq. (2.56) the Noether current is not the canonical energy–momentum tensor, but the symmetric Belinfante–Rosenfeld energy–momentum tensor Eq. (2.58),

$$\delta_\epsilon^{(1)}\mathcal{L}_{\text{matter}}(A) = -\chi\epsilon^\sigma\partial_\mu T^\mu{}_\sigma(A). \tag{3.183}$$

This sounds promising, because we need symmetric energy–momentum tensors. For the gravitational field, we have (as usual, up to total derivatives)

$$\delta_\epsilon^{(1)}\mathcal{L}_{\text{FP}} = -\chi\epsilon^\sigma\partial_\mu\left[t_{\text{can}}^{(0)\,\mu}{}_\sigma(h) + \mathcal{D}^\mu{}_\rho(h)h^\rho{}_\sigma\right]. \tag{3.184}$$

The additional term that we obtain vanishes on-shell. In general it is possible to add to a Noether current any term that vanishes on-shell and so we may understand the additional term as a redefinition of the canonical energy–momentum tensor. This redefinition is, however, important. On the one hand, the equations of motion are going to be corrected and therefore the addition of terms vanishing on-shell to first order is going to become meaningful at higher orders and should be considered with care. On the other hand, if we obtain an action that is invariant under the above gauge symmetry, the equations of motion are going to satisfy a gauge identity that is, at the same time, the condition for the invariance of the action. By varying directly the Lagrangian Eq. (3.178) under $\delta_\epsilon^{(1)}$, we find that it will be invariant up to $\mathcal{O}(\chi^2)$ if the gravitational energy–momentum tensor that appears in the equations of motion (3.171) satisfies, to first order in χ,

$$\partial_\mu t^{(0)\,\mu}{}_\sigma(h) = \partial_\mu\left[t^{(0)\,\mu}_{\mathrm{can}\,\sigma}(h) + \mathcal{D}^\mu{}_\rho(h)h^\rho{}_\sigma\right]. \tag{3.185}$$

On taking explicitly the derivative on the r.h.s., we obtain the gauge identities[34] associated with invariance under $\delta_\epsilon^{(1)}$:

$$\partial_\mu t^{(0)\mu}{}_\sigma(h) = \gamma_{\nu\rho\sigma}\mathcal{D}^{\nu\rho}(h), \qquad \gamma_{\nu\rho\sigma} = \tfrac{1}{2}\left\{\partial_\nu h_{\rho\sigma} + \partial_\rho h_{\nu\sigma} - \partial_\sigma h_{\nu\rho}\right\}, \tag{3.186}$$

Thus, if we look for invariance under the gauge transformations $\delta_\epsilon^{(1)}$, the gravitational energy–momentum tensor that we will put on the r.h.s. of Eq. (3.171) has to be of the form

$$t^{(0)}{}_{\mu\sigma}(h) = t^{(0)}_{\mathrm{can}\,\mu\sigma}(h) + \mathcal{D}^\mu{}_\rho(h)h^\rho{}_\sigma + \partial_\rho\Psi^\rho{}_{\mu\sigma}, \tag{3.187}$$

but we can no longer add on-shell-vanishing terms proportional to $\mathcal{D}^\mu{}_\rho(h)$ because then the above gauge identities would not be satisfied. Here we see how the requirement of gauge symmetry constrains the possible energy–momentum tensors. Comparing this situation with our construction of the scalar theory of gravity in which the energy–momentum tensor could be asymmetric and did not have to satisfy any kind of conditions, we are much better off.

Still, the redefined canonical energy–momentum tensor

$$t^{(0)}_{\mathrm{can}\,\mu\sigma}(h) + \mathcal{D}^\mu{}_\rho(h)h^\rho{}_\sigma$$

is not symmetric as we had hoped and we have to find additional terms $\partial_\rho\Psi^\rho{}_{\mu\sigma}$ that cancel out exactly the antisymmetric part of our energy–momentum tensors. There is only one systematic procedure for doing this and only for the canonical one: the Belinfante method explained in Chapter 2 which, unfortunately, requires the addition of on-shell-vanishing terms. Let us, nevertheless, see where we are taken by this method. It is straightforward (but long and tedious) to find

$$\Psi^{\rho\mu}{}_\sigma = -2\partial^{[\rho}h^{\mu]}{}_\beta h^\beta{}_\sigma - 2\partial_\beta h^{[\rho}{}_\sigma h^{\mu]\beta} + \partial^{[\rho}hh^{\mu]}{}_\sigma + \eta_\sigma{}^{[\rho}\partial_\beta hh^{\mu]\beta}. \tag{3.188}$$

The antisymmetric part of the modified canonical tensor $t^{(0)}_{\mathrm{can}\,\mu\sigma} + \partial_\rho\Psi^\rho{}_{\mu\sigma}$ is $-\mathcal{D}_{\rho[\mu}h^\rho{}_{\sigma]}$,

[34] We are using the zeroth-order gauge identity $\partial_\mu\mathcal{D}^{\mu\nu}(h) = 0$, which is, obviously, valid.

and, therefore, on discarding it, we obtain the Belinfante tensor

$$\begin{aligned}
t^{(0)}_{\text{Bel}\,\mu\sigma} &\equiv t^{(0)}_{\text{can}\,\mu\sigma} + \partial_\rho \Psi^\rho{}_{\mu\sigma} + \mathcal{D}_{\rho[\mu}(h) h^\rho{}_{\sigma]} \\
&= -\tfrac{1}{2}\partial_\mu h^{\nu\rho}\partial_\sigma h_{\nu\rho} - \partial_\nu h_{\rho\mu}\partial^\nu h^\rho{}_\sigma - \partial_\nu h_{\rho\mu}\partial^\rho h^\nu{}_\sigma + 2\partial_{(\mu|} h_{\nu\rho}\partial^\nu h^\rho{}_{|\sigma)} \\
&\quad + \partial_\nu h_{\mu\sigma}\partial_\rho h^{\rho\nu} - \partial_\nu h^\nu{}_{(\mu}\partial_{\sigma)} h + \tfrac{1}{2}\partial^\nu h_{\mu\sigma}\partial_\nu h + \tfrac{1}{2}\partial_\mu h \partial_\sigma h \\
&\quad + \tfrac{1}{2}\eta_{\mu\sigma}\left[\tfrac{1}{2}\partial_\lambda h_{\nu\rho}\partial^\lambda h^{\nu\rho} - \partial_\lambda h_{\nu\rho}\partial^\nu h^{\lambda\rho} - \tfrac{1}{2}(\partial h)^2\right] \\
&\quad + h_{(\mu}{}^\nu \partial_{\sigma)}\partial_\rho h^\rho{}_\nu - h^\nu{}_{(\mu|}\partial_\nu\partial_\rho h_{|\sigma)}{}^\rho - h^\nu{}_{(\mu}\partial^2 h_{\sigma)\nu} \\
&\quad + h^{\lambda\nu}\partial_\lambda\partial_\nu h_{\mu\sigma} - \tfrac{1}{2}\eta_{\mu\sigma} h^{\lambda\nu}\partial_\lambda\partial_\nu h + \tfrac{1}{2} h_{\mu\sigma}\partial^2 h.
\end{aligned} \tag{3.189}$$

As expected, this tensor does not satisfy the gauge identities required because of the addition of on-shell-vanishing terms. However,

$$t^{(0)}_{\text{can}\,\mu\sigma} + \mathcal{D}_{\rho\mu}(h) h^\rho{}_\sigma + \partial_\rho \Psi^\rho{}_{\mu\sigma} = t^{(0)}_{\text{Bel}\,\mu\sigma} + \mathcal{D}_{\rho(\mu}(h) h^\rho{}_{\sigma)} \tag{3.190}$$

is symmetric and evidently satisfies the gauge identities associated with $\delta^{(1)}_\epsilon$.

Thus, using (more or less) the Belinfante method, we have been able to symmetrize the energy–momentum tensor associated with the gauge transformations $\delta^{(1)}_\epsilon$. This is basically the energy–momentum tensor used by Thirring in [888], although he expressed it in the harmonic gauge. As we are going to see, it is unacceptable from several points of view.

We can now try to find the Lagrangian correction from which to derive the above energy–momentum tensor as the r.h.s. of the gravitational equation of motion. It should be a term linear in $h_{\mu\nu}$ and quadratic in $\partial_\mu h_{\nu\rho}$. Unfortunately no such term can be found.[35] This means that further modifications $\partial_\rho \Psi^\rho{}_{\mu\sigma}$ are required, but this time they have to be symmetric in the two free indices and they have to lead to a term derivable from a Lagrangian, which is a difficult problem with no guaranteed unique solution.

As an act of desperation we can try to see whether Rosenfeld's energy–momentum tensor has the properties that we are looking for (even if it is not evidently associated with any Noether current). We first rewrite the Fierz–Pauli action in a background metric $\gamma_{\mu\nu}$:

$$S = \int d^dx \sqrt{|\gamma|}\, \left\{\tfrac{1}{4}\nabla^\rho h^{\mu\nu}\nabla_\rho h_{\mu\nu} - \tfrac{1}{2}\nabla^\rho h^{\mu\nu}\nabla_\mu h_{\rho\nu} + \tfrac{1}{2}\nabla^\mu h \nabla^\lambda h_{\lambda\mu} - \tfrac{1}{4}\nabla^\mu h \nabla_\mu h\right\}, \tag{3.191}$$

where $\gamma = \det(\gamma_{\mu\nu})$ and ∇_μ is the covariant derivative with respect to the Levi-Cività connection $C_{\mu\nu}{}^\rho(\gamma)$ associated with $\gamma_{\mu\nu}$. Now we vary this with respect to the background metric, taking into account that $h_{\mu\nu}$ is assumed to be metric-independent. By varying separately the terms without and with partial derivatives of the background metric, we

---

[35] It is easy to see that, to reproduce the term $\eta_{\mu\sigma}(\partial_\rho h \partial^\rho h)$ in the energy–momentum tensor, we need a term of the form $h(\partial_\rho h \partial^\rho h)$ in the Lagrangian, but this term produces another term of the form $\eta_{\mu\sigma} h(\partial_\rho \partial^\rho h)$, which is not present in the energy–momentum tensor.

obtain

$$\delta S = \int d^d x \sqrt{|\gamma|} \delta\gamma_{\alpha\beta} \Big\{ -\tfrac{1}{4}\nabla^\alpha h_{\nu\rho}\nabla^\beta h^{\nu\rho} - \tfrac{1}{2}\nabla_\nu h_\rho{}^\alpha \nabla^\nu h^{\rho\beta}$$
$$+ \nabla^{(\alpha|} h_{\nu\rho}\nabla^\nu h^{\rho|\beta)} - \tfrac{1}{2}\nabla_\nu h^{\nu\rho}\nabla_\rho h^{\alpha\beta} - \tfrac{1}{2}\nabla^{(\alpha} h^{\beta)\rho}\nabla_\rho h$$
$$- \tfrac{1}{2}\nabla_\nu h^{\nu(\alpha}\nabla^{\beta)} h + \tfrac{1}{2}\nabla^\nu h^{\alpha\beta}\nabla_\nu h + \tfrac{1}{4}\nabla^\alpha h \nabla^\beta h + \tfrac{1}{2}\gamma^{\alpha\beta}\mathcal{L}_{\rm FP} + \frac{\delta C_{\mu\nu}{}^\lambda}{\delta\gamma_{\alpha\beta}} f_\lambda{}^{\mu\nu} \Big\},$$
$$f_\lambda{}^{\mu\nu} = h_{\lambda\rho}\nabla^\rho h^{\mu\nu} - h_\lambda{}^{(\mu}\nabla_\sigma h^{\nu)\sigma} + \tfrac{1}{2} h_\lambda{}^{(\mu}\nabla^{\nu)} h - \tfrac{1}{2}\gamma^{\mu\nu} h_{\lambda\rho}\nabla^\rho h. \tag{3.192}$$

Using now

$$\delta C_{\mu\nu}{}^\lambda = \tfrac{1}{2}\gamma^{\lambda\tau}\{\nabla_\mu \delta\gamma_{\nu\tau} + \nabla_\nu \delta\gamma_{\mu\tau} - \nabla_\tau \delta\gamma_{\mu\nu}\} \tag{3.193}$$

in the last term and integrating it by parts, it becomes

$$\int d^d x \sqrt{|\gamma|}\, \delta\gamma_{\alpha\beta} \big\{ -\tfrac{1}{2}\nabla_\mu f^{(\alpha|\mu|\beta)} - \tfrac{1}{2}\nabla_\nu f^{(\alpha\beta)\nu} + \tfrac{1}{2}\nabla_\tau f^{\tau\alpha\beta} \big\}. \tag{3.194}$$

By expanding all the terms and setting $\gamma_{\alpha\beta} = \eta_{\alpha\beta}$, we obtain the Rosenfeld energy–momentum tensor, which turns out to be identical to the symmetrized one in Eq. (3.190).

At this point it looks impossible, without any other guiding principle, to find the right symmetric energy–momentum tensor satisfying the gauge identities and leading to an equation of motion derivable from a Lagrangian. However, we can try to solve our problem starting from the end; that is, by writing down the most general $\mathcal{L}^{(1)}{}_{\mu\sigma}$ quadratic in $\partial_\alpha h_{\beta\gamma}$ and imposing gauge invariance of the total Lagrangian $\mathcal{L}^{(1)} = \mathcal{L}_{\rm FP} + \frac{1}{2}\chi h^{\mu\sigma}\mathcal{L}^{(1)}{}_{\mu\sigma}$ to first order in χ, or, equivalently, using the fact that the equation of motion derived from it satisfies the gauge identities Eq. (3.186). Up to total derivatives, the most general $\mathcal{L}^{(1)}{}_{\mu\sigma}$ is

$$\mathcal{L}^{(1)}{}_{\alpha\beta} = a\partial_\alpha h_{\lambda\delta}\partial_\beta h^{\lambda\delta} + b\partial_{(\alpha|} h_{\lambda\delta}\partial^\lambda h^\delta{}_{|\beta)} + c\partial_\lambda h_{\delta\alpha}\partial^\lambda h^\delta{}_\beta$$
$$+ q\partial_\lambda h^\lambda{}_\alpha \partial_\delta h^\delta{}_\beta + d\partial_\lambda h_{\delta\alpha}\partial^\delta h^\lambda{}_\beta + e\partial_{(\alpha} h_{\beta)\lambda}\partial_\delta h^{\delta\lambda}$$
$$+ f\partial_\lambda h_{\alpha\beta}\partial_\delta h^{\delta\lambda} + g\partial_{(\alpha} h_{\beta)\lambda}\partial^\lambda h + i\partial_\lambda h_{\alpha\beta}\partial^\lambda h + j\partial_\lambda h^\lambda{}_{(\alpha}\partial_{\beta)} h$$
$$+ m\partial_\alpha h \partial_\beta h + \eta_{\alpha\beta}\big[k\partial_\gamma h_{\delta\lambda}\partial^\gamma h^{\delta\lambda} + l\partial_\gamma h_{\delta\lambda}\partial^\delta h^{\gamma\lambda}$$
$$+ r\partial_\lambda h^{\lambda\delta}\partial_\gamma h^\gamma{}_\delta + n\partial_\gamma h^{\gamma\delta}\partial_\delta h + p(\partial h)^2\big]. \tag{3.195}$$

Now we substitute this expression into $\mathcal{L}^{(1)}$, we find the equation of motion, identify the gravitational energy–momentum tensor

$$t^{(0)\,\alpha\beta} = \mathcal{L}^{(1)\,\alpha\beta} - \partial_\lambda\left(h^{\mu\sigma}\frac{\partial\mathcal{L}^{(1)}{}_{\mu\sigma}}{\partial\partial_\lambda h_{\alpha\beta}} \right), \tag{3.196}$$

and substitute this into the gauge identity Eq. (3.186) to arrive at the condition

$$\partial_\alpha\mathcal{L}^{(1)\,\alpha\beta} - \partial_\alpha\partial_\lambda\left(h^{\mu\sigma}\frac{\partial\mathcal{L}^{(1)}{}_{\mu\sigma}}{\partial\partial_\lambda h_{\alpha\beta}} \right) = \gamma_{\mu\sigma}{}^\beta\mathcal{D}^{\mu\sigma}(h). \tag{3.197}$$

This is an equation in the constant coefficients $a, b, c, d, \ldots$. To solve it, we first observe that all the terms with the structure $h\partial\partial\partial h$ on the l.h.s. must vanish because they do not occur on the r.h.s. Then, we also impose the vanishing of all the terms with the structure $\partial h(\partial\partial h)^\beta$ on the l.h.s. for the same reason. Finally, we identify the terms with structures $\partial^\beta h(\partial\partial h)$ and $\partial h^\beta(\partial\partial h)$ on both sides of the above equation. The result can be expressed in terms of two parameters x and y, which are left undetermined:

$$
\begin{aligned}
&a=-\tfrac{1}{2}, \quad b=2-y, \quad c=-1, \quad d=1-y, \quad e=y, \quad f=-1, \\
&g=-1-x, \quad i=1, \quad j=-1+x, \quad k=\tfrac{1}{4}, \quad l=-\tfrac{1}{2}-x, \\
&m=\tfrac{1}{2}, \quad n=\tfrac{1}{2}, \quad p=-\tfrac{1}{4}, \quad q=y, \quad r=x.
\end{aligned} \tag{3.198}
$$

On substituting these into the general expression for $\mathcal{L}^{(1)}{}_{\mu\sigma}$ and collecting all the terms proportional to the two parameters x and y, we obtain

$$
\begin{aligned}
\mathcal{L}^{(1)}{}_{\mu\sigma} &= \mathcal{L}^{(1)}_{\mathrm{GR}\;\mu\sigma} + \text{total derivatives}, \\
\mathcal{L}^{(1)}_{\mathrm{GR}\;\mu\sigma} &= -\tfrac{1}{2}\partial_\mu h^{\nu\rho}\partial_\sigma h_{\nu\rho} - \partial^\nu h^\rho{}_\mu \partial_\nu h_{\rho\sigma} + \partial^\nu h^\rho{}_{(\mu|}\partial_\rho h_{\nu|\sigma)} \\
&\quad + 2\partial^\nu h^\rho{}_{(\mu}\partial_{\sigma)} h_{\nu\rho} - \partial_{(\mu} h_{\sigma)}{}^\nu \partial_\nu h - \partial^\nu h_{\mu\sigma}\partial_\rho h^\rho{}_\nu \\
&\quad - \partial_\nu h^\nu{}_{(\mu}\partial_{\sigma)} h + \partial_\nu h_{\mu\sigma}\partial^\nu h + \tfrac{1}{2}\partial_\mu h \partial_\sigma h + \eta_{\mu\sigma}\mathcal{L}_{\mathrm{FP}},
\end{aligned} \tag{3.199}
$$

an unambiguous, unique, answer (up to total derivatives), which leads to a Lagrangian $\mathcal{L}^{(1)}$ that is invariant to first order in χ under the gauge transformations Eq. (3.181). The equations of motion are fully determined and the gravitational energy–momentum tensor is the piece of these equations of motion that is proportional to χ, given by Eq. (3.196), or, more explicitly, by

$$
\begin{aligned}
t^{(0)\,\mu\sigma}_{\mathrm{GR}} &= \tfrac{1}{2}\partial^\mu h_{\lambda\delta}\partial^\sigma h^{\lambda\delta} + \partial_\lambda h_\delta{}^\mu \partial^\lambda h^{\delta\sigma} - \partial_\lambda h_\delta{}^\mu \partial^\delta h^{\lambda\sigma} + \partial_\lambda h^{\mu\sigma}\partial_\delta h^{\delta\lambda} \\
&\quad - 2\partial^{(\mu} h^{\sigma)}{}_\delta \partial_\lambda h^{\lambda\delta} - \tfrac{1}{2}\partial_\lambda h^{\mu\sigma}\partial^\lambda h + \partial^{(\mu} h^{\sigma)\lambda}\partial_\lambda h \\
&\quad + \eta^{\mu\sigma}\big[-\tfrac{3}{4}\partial_\alpha h_{\beta\gamma}\partial^\alpha h^{\beta\gamma} + \tfrac{1}{2}\partial_\alpha h_{\beta\gamma}\partial^\beta h^{\alpha\gamma} + \partial_\lambda h^{\lambda\alpha}\partial_\delta h^\delta{}_\alpha \\
&\qquad - \partial_\lambda h^{\lambda\alpha}\partial_\alpha h + \tfrac{1}{4}\partial_\lambda h \partial^\lambda h\big] \\
&\quad + h^{\alpha\beta}\Big[\partial_\alpha\partial_\beta h^{\mu\sigma} - 2\partial_\alpha\partial^{(\mu} h^{\sigma)}{}_\beta + \partial^\mu\partial^\sigma h_{\alpha\beta} + 2\eta^{(\sigma}{}_\alpha \hat{\mathcal{D}}^{\mu)}{}_\beta(h) \\
&\qquad - \tfrac{1}{2}\eta^\mu{}_\alpha \eta^\sigma{}_\beta \hat{\mathcal{D}}^\rho{}_\rho(h) - \tfrac{1}{2}\eta_{\alpha\beta}\mathcal{D}^{\mu\sigma}(h) - \eta^{\mu\sigma}\hat{\mathcal{D}}_{\alpha\beta}(h)\Big].
\end{aligned} \tag{3.200}
$$

This is clearly the energy–momentum tensor we were looking for. It is related to the Rosenfeld energy–momentum tensor Eq. (3.190) by

$$
\begin{aligned}
t^{(0)\,\mu\sigma}_{\mathrm{GR}} - \big(t^{(0)\,\mu\sigma}_{\mathrm{can}} + \mathcal{D}^{\rho\mu}(h)h_\rho{}^\sigma + \partial_\rho \Psi^{\rho\mu\sigma}\big) &\equiv \partial_\rho \Psi^{\rho\mu\sigma}_{\mathrm{GR-Ros}}, \\
\Psi^{\rho\mu\sigma}_{\mathrm{GR-Ros}} = \partial_\nu\big[\eta^{\sigma[\rho}\eta^{\mu]\nu}h^{\lambda\delta}h_{\lambda\delta} &+ 2\eta^{\nu[\rho}h^{\mu]}{}_\lambda h^{\lambda\sigma} - 2\eta^{\sigma[\rho}h^{\mu]\lambda}h_\lambda{}^\nu \\
&+ \eta^{\sigma[\rho}h^{\mu]\nu}h + \eta^{\nu[\mu}h^{\rho]\sigma}h - \tfrac{1}{2}\eta^{\nu[\mu}\eta^{\rho]\sigma}h^2\big].
\end{aligned} \tag{3.201}
$$

Summarizing: the Noether procedure allows us to find corrections to the free Fierz–Pauli theory order by order in the parameter χ, making it self-consistent to that given order.

The procedure seems to be unambiguous and not complicated but is tedious and time-consuming since there is no systematic way of finding the next correction for the energy–momentum tensor (the Belinfante and Rosenfeld prescriptions have proved to be inadequate in this problem). For instance, at second order, we would have to find the second-order corrections to the gauge-transformation rules Eq. (3.181) (quadratic in $h_{\mu\nu}$, linear in the gauge parameter ϵ^{μ}, with two partial derivatives and satisfying the group property), the second-order gauge identities associated with the invariance of the Lagrangian under those gauge transformations at the given order, and the second-order corrections to the Lagrangian (these would be of the form $h^{\mu\nu}h^{\rho\sigma}\mathcal{L}^{(2)}{}_{\mu\nu\rho\sigma}(\partial h\partial h)$) and then we would have to write the most general $\mathcal{L}^{(2)}{}_{\mu\nu\rho\sigma}(\partial h\partial h)$ symmetric in the pairs $(\mu\nu)$ and $(\rho\sigma)$ and then impose on the corresponding term in the Lagrangian the second-order gauge identity. A more efficient way of finding these corrections, like Deser's, is necessary but, before we study it, it is worth checking that the correction to the equations of motion implied by this gravitational energy–momentum tensor leads to the right value of the precession of the perihelion of Mercury. We also study some other properties of $t_{\mathrm{GR}}^{(0)\,\mu\sigma}$.

3.2.6 Properties of the gravitational energy–momentum tensor $t_{\mathrm{GR}}^{(0)\,\mu\sigma}$

Our first observation concerns the gauge-transformation properties of $t_{\mathrm{GR}}^{(0)\,\mu\sigma}$. The first-order Lagrangian $\mathcal{L}^{(1)}$ is invariant under the gauge transformations Eq. (3.181) to first order in χ. This implies the invariance of the first-order equations of motion

$$\mathcal{D}^{\mu\sigma}(h) = \chi t_{\mathrm{GR}}^{(0)\,\mu\sigma}(h). \tag{3.202}$$

The l.h.s. is invariant under the zeroth-order gauge transformations and this implies that the zeroth-order variation of the r.h.s. does not vanish and is identical to the first-order variation of the l.h.s.

From the point of view of the linear (Fierz–Pauli) theory we can say that the energy–momentum tensor $t_{\mathrm{GR}}^{(0)\,\mu\sigma}$ is not invariant under the same (zeroth-order) gauge transformations as those that leave the Lagrangian invariant. Rosenfeld's [46] and other energy–momentum tensors defined in the literature also lack this invariance. In the case of the Rosenfeld energy–momentum tensor, it can be shown that it is not gauge-invariant because the Fierz–Pauli theory is not invariant under the zeroth-order gauge transformations (or their covariantization) when it is written in an arbitrary curved background as in Eq. (3.191). This invariance cannot be recovered by adding terms proportional to the Riemann tensor of the background metric [47].

This lack of gauge-invariance is in contrast to the invariance of the Rosenfeld energy–momentum tensor of other gauge fields under the relevant gauge transformations. However, while the lack of gauge-invariance of the energy–momentum tensors of other gauge theories would be a serious problem in its coupling to gravity, it is not a problem for gravity itself since, as we have seen, only in this way can the full equation of motion be gauge-invariant under the full gauge transformations.

On the other hand, the situation is not too different from the one encountered in the Noether procedure for n vector fields in which the Noether current associated with the lowest-order gauge transformations is not invariant under them, and we have to add further corrections.

Once this point has been clarified, we proceed to evaluate the correction to the linear solution for the gravitational field of a point-like massive particle Eq. (3.124) and the gravitational field of a point-like massless particle Eqs. (3.131), (3.133), and (3.134). The general setup used to calculate corrections is the following. From the first-order Lagrangian[36]

$$\mathcal{L}^{(1)} = \mathcal{L}_{\rm FP} + \mathcal{L}_{\rm matter}(\varphi) + \tfrac{1}{2}(\chi/c)h^{\mu\nu}\big[\mathcal{L}^{(1)}{}_{\mu\nu}(h) + T_{\mathrm{matter}\,\mu\nu}(\varphi)\big] \tag{3.203}$$

we obtain the equations of motion

$$\begin{aligned} \mathcal{D}_{\mu\nu}(h) - (\chi/c)\left[t^{(0)}{}_{\mu\nu}(h) + T_{\mathrm{matter}\,\mu\nu}(\varphi)\right] = 0, \\ \mathcal{D}^{(0)}(\varphi) + (\chi/c)\mathcal{D}^{(1)}(\varphi, h) = 0. \end{aligned} \tag{3.204}$$

To find solutions to these equations, we expand the gravitational and matter fields

$$h_{\mu\nu} = h^{(0)}{}_{\mu\nu} + \chi h^{(1)}{}_{\mu\nu} + \cdots, \qquad \varphi = \varphi^{(0)} + \chi\varphi^{(1)} + \cdots, \tag{3.205}$$

around a solution $(h^{(0)}, \varphi^{(0)})$ of the equations

$$\begin{aligned} \mathcal{D}_{\mu\nu}(h^{(0)}) - (\chi/c)T_{\mathrm{matter}\,\mu\nu}(\varphi^{(0)}) = 0, \\ \mathcal{D}^{(0)}(\varphi^{(0)}) = 0. \end{aligned} \tag{3.206}$$

On substituting the expansion into the first-order equations of motion, taking into account that $t^{(0)}{}_{\mu\nu}(h)$ is quadratic in h, $\mathcal{D}^{(0)}(\varphi)$ is linear in φ, and $\mathcal{D}^{(1)}(h, \varphi)$ is linear both in h and in φ, and using the above zeroth-order equations, we find, to lowest order in χ,

$$\begin{aligned} \mathcal{D}_{\mu\nu}(h^{(1)}) - \frac{1}{c}t^{(0)}{}_{\mu\nu}(h^{(0)}) = 0, \\ \mathcal{D}^{(0)}(\varphi^{(1)}) + \frac{1}{c}\mathcal{D}^{(1)}(h^{(0)}, \varphi^{(0)}) = 0. \end{aligned} \tag{3.207}$$

We are interested in $h^{(1)}$ in $d = 4$ and we are going to calculate it by using the Rosenfeld energy–momentum tensor Eq. (3.190) on the linear solution $t^{(0)}_{\mathrm{Ros}\,\mu\nu}(h^{(0)})$ and the energy–momentum tensor Eq. (3.200) we found by imposing $\delta^{(1)}_{\epsilon}$ gauge invariance on the linear solution $t^{(0)}_{\mathrm{GR}\,\mu\nu}(h^{(0)})$.

In $d = 4$ the solution Eq. (3.124) for a massive particle can be written in the simple form

$$h^{(0)}{}_{\mu\nu} = \delta_{\mu\nu}k, \qquad k = -\frac{\chi Mc}{8\pi}\frac{1}{|\vec{x}_3|}. \tag{3.208}$$

On substituting this expression into the energy–momentum tensors, we find

$$\begin{aligned} \frac{1}{c}t^{(0)}_{\mathrm{Ros}\,\mu\nu}(h^{(0)}) &= -\partial_\mu k\partial_\nu k - \big(\tfrac{3}{2}\eta_{\mu\nu} + 2\delta_{\mu\nu}\big)(\partial k)^2 - \big(\eta_{\mu\nu} + \delta_{\mu\nu}\big)k\partial^2 k, \\ \frac{1}{c}t^{(0)}_{\mathrm{GR}\,\mu\nu}(h^{(0)}) &= \partial_\mu k\partial_\nu k - \tfrac{3}{2}(\partial k)^2 + 2k\partial_\mu\partial_\nu k - \big(\eta_{\mu\nu} - \delta_{\mu\nu}\big)k\partial^2 k. \end{aligned} \tag{3.209}$$

[36] We again restore all factors of c.

There are two types of terms: terms of the form $\partial k \partial k$ and of the form $k\partial_\mu \partial_\nu k$, which give finite contributions, and terms of the form $k\partial^2 k$, which give singular contributions ($\partial^2 k \sim \delta^{(3)}(\vec{x}_3)$) but only at the origin $\vec{x}_3 = \vec{0}$ and have to be absorbed into a renormalization of the source. In the Rosenfeld case, it is just a renormalization of the mass, but in the second case the mass is not renormalized and, instead, the source's energy–momentum tensor has singular terms $T_{\text{source}\ ij} \sim \delta_{ij}\delta^{(3)}(\vec{x}_3)$, which do not fit within the concept of a point-particle. Since we are mainly interested in obtaining corrections to the gravitational field of massive, finite-sized bodies of spherical symmetry (the Sun, for instance), we opt for hiding this problem in the closet with the other skeletons for the moment and simply ignore these terms.

By taking the derivatives on the r.h.s. of the above expressions, we find

$$
\begin{aligned}
\frac{1}{c} t^{(0)}_{\text{Ros}\ 00}(h^{(0)}) &= \frac{7}{2}\frac{R_S^2}{\chi^2}\frac{1}{|\vec{x}_3|^4},\\
\frac{1}{c} t^{(0)}_{\text{Ros}\ ij}(h^{(0)}) &= -\frac{R_S^2}{\chi^2}\left(\frac{x^i x^j}{|\vec{x}_3|^6} - \frac{1}{2}\delta_{ij}\frac{1}{|\vec{x}_3|^4}\right),\\
\frac{1}{c} t^{(0)}_{\text{GR}\ 00}(h^{(0)}) &= \frac{3}{2}\frac{R_S^2}{\chi^2}\frac{1}{|\vec{x}_3|^2},\\
\frac{1}{c} t^{(0)}_{\text{GR}\ ij}(h^{(0)}) &= 7\frac{R_S^2}{\chi^2}\left(\frac{x^i x^j}{|\vec{x}_3|^6} - \frac{1}{2}\delta_{ij}\frac{1}{|\vec{x}_3|^4}\right).
\end{aligned}
\tag{3.210}
$$

To solve these equations, we could try to eliminate all the off-diagonal terms in the energy–momentum tensor by a gauge transformation, as did Thirring in [888]. However, as observed in [725], the gauge transformation that one has to use is $\epsilon_\mu \sim \partial_\mu \ln r$, which does not go to zero at infinity and, furthermore, takes us out of the De Donder gauge in which we want to solve the equation. This clearly invalidates Thirring's results.

However, we can solve directly the first of Eqs. (3.207) in the De Donder gauge: observe that the r.h.s. of this equation,

$$
\partial^2 \bar{h}^{(1)}{}_{\mu\nu} = \frac{1}{c} t^{(0)}{}_{\mu\nu}(h^{(0)}),
\tag{3.211}
$$

is divergence-free. For the Rosenfeld energy–momentum tensor we obtain [725]

$$
\bar{h}^{(1)}{}_{00} = -\frac{7}{4}\frac{R_S^2}{\chi^2}\frac{1}{|\vec{x}_3|^2}, \qquad \bar{h}^{(1)}{}_{ij} = -\frac{1}{4}\frac{R_S^2}{\chi^2}\frac{x^i x^j}{|\vec{x}_3|^4},
\tag{3.212}
$$

and, by combining this correction and the linear term into $g_{\mu\nu} = \eta_{\mu\nu} + \chi h^{(0)}{}_{\mu\nu} + \chi^2 h^{(1)}{}_{\mu\nu}$, we obtain the spherically symmetric metric

$$
ds^2_{\text{Ros}} = \left(1 - \frac{R_S}{r} - \frac{R_S^2}{r^2}\right)c^2 dt^2 - \left(1 + \frac{R_S}{r} + \frac{R_S^2}{r^2}\right)dr^2 - \left(1 + \frac{R_S}{r} + \frac{3}{4}\frac{R_S^2}{r^2}\right)r^2 d\Omega^2_{(2)},
\tag{3.213}
$$

where we have defined $r = |\vec{x}_3|$ and used $dr = x^i dx^i/|\vec{x}_3|$, $dx^i dx^i = dr^2 + r^2 d\Omega^2$, etc.

For the GR energy–momentum tensor we obtain [725]

$$\bar{h}^{(1)}{}_{00} = -\frac{3}{4}\frac{R_S^2}{\chi^2}\frac{1}{|\vec{x}_3|^2}, \qquad \bar{h}^{(1)}{}_{ij} = \frac{7}{4}\frac{R_S^2}{\chi^2}\frac{x^i x^j}{|\vec{x}_3|^4}, \tag{3.214}$$

and the metric

$$ds^2_{\rm GR} = \left(1 - \frac{R_S}{r} + \frac{1}{2}\frac{R_S^2}{r^2}\right)c^2 dt^2 - \left(1 + \frac{R_S}{r} - \frac{1}{2}\frac{R_S^2}{r^2}\right)dr^2 - \left(1 + \frac{R_S}{r} + \frac{5}{4}\frac{R_S^2}{r^2}\right)r^2 d\Omega^2_{(2)}. \tag{3.215}$$

It is, however, more convenient to perform a gauge transformation with parameter

$$\epsilon_i = -R_S x^i / r^2, \tag{3.216}$$

which changes the gauge of $h^{(1)}{}_{\mu\nu}$ and leaves the metric in the form

$$ds^2_{\rm GR} = \left(1 - \frac{R_S}{r} + \frac{1}{2}\frac{R_S^2}{r^2}\right)c^2 dt^2 - \left(1 + \frac{R_S}{r} + \frac{1}{2}\frac{R_S^2}{r^2}\right)dr^2 - \left(1 + \frac{R_S}{r} + \frac{1}{4}\frac{R_S^2}{r^2}\right)r^2 d\Omega^2_{(2)}, \tag{3.217}$$

which we will be able to compare later on with the expansion of an exact solution of general relativity (hence the subscript “GR”), Eq. (7.31). In any case, this gauge transformation does not change the coefficient λ_2 in the expansion Eq. (3.150), which is all we need to recalculate the precession of the perihelion of Mercury. Taking into account now the values of λ_2 obtained, and substituting into Eq. (3.151), we obtain

$$\delta\varphi_{\rm Ros} = 3\pi R_S^2 m^2 c^2 / l^2, \qquad \delta\varphi_{\rm GR} = \tfrac{3}{2}\pi R_S^2 m^2 c^2 / l^2. \tag{3.218}$$

The second is in agreement with observations. This result gives us more confidence in the self-consistent spin-2 theory that we are constructing and confirms the importance of gauge symmetry, which is a property not enjoyed by the theory built on Rosenfeld’s energy–momentum tensor.

Now we can do the same for the massless point-like-particle gravitational field given in Eqs. (3.131), (3.133), and (3.134). We can write the solution in this form:

$$h^{\mu\nu} = k\ell^\mu \ell^\nu, \qquad k = k(u, x). \tag{3.219}$$

It is easy to see that all these terms vanish identically:

$$h = 0, \quad h_{\mu\rho}h^{\mu\nu} = 0, \quad \partial_\mu h^{\mu\nu} = 0, \quad h^{\mu\nu}\partial_\nu h_{\alpha\beta} = 0, \tag{3.220}$$

and all terms in $t^{(0)}_{\mathrm{GR}\,\mu\nu}(h^{(0)})$ identically vanish. There is neither renormalization of the source nor corrections to the lowest-order solution. The same must also be true if we consider higher-order corrections to the equations of motion and, therefore, we expect the solution Eqs. (3.131), (3.133), and (3.134) to be an exact solution of the full theory, whatever it is. Actually, we will study this solution in Chapter 10 and we can compare the present solution with the one in Eqs. (10.23) and (10.26).

Now that we have convinced ourselves that the self-consistent spin-2 theory is a good candidate for a theory of gravitation but is at the same time hard to obtain in a perturbative series, we are prepared to use Deser’s argument, which shows that GR is precisely the resummation of the perturbative series we were generating in such a painful way.

3.2.7 Deser's argument

In [299] Deser presented an argument that allows one to see GR as the self-consistent SRFT of a spin-2 particle we were looking for in the sense that, in GR, the gravitational field couples to its own energy–momentum tensor, at least for a certain choice of field variables, Lagrangian, and energy–momentum tensor. The emphasis is on physical consistency rather than on gauge-invariance, and, therefore, the choice of energy–momentum tensor is not based on that criterion, as in our previous discussions about the Noether method. These would be weak points if we wanted to take this work as proof of the uniqueness of GR as a solution to our initial problem, but we should understand Deser's work as a proof that GR is *a* solution to our problem from the physical standpoint.

The starting point in Deser's argument is a first-order version of the Fierz–Pauli action that uses two (off-shell) independent fields $\varphi^{\mu\nu}$ and $\Gamma_{\mu\nu}{}^{\rho}$ (see [841] for a construction of this action),

$$\boxed{S^{(1)}_{\rm FP}[\varphi^{\mu\nu},\Gamma_{\mu\nu}{}^{\rho}]=\frac{1}{\chi^2}\int d^dx\left\{-\chi\varphi^{\mu\nu}2\partial_{[\mu}\Gamma_{\rho]\nu}{}^{\rho}+\eta^{\mu\nu}2\Gamma_{\lambda[\mu}{}^{\rho}\Gamma_{\rho]\nu}{}^{\lambda}\right\},}\tag{3.221}$$

which are Lorentz tensors symmetric in the pair of indices $\mu\nu$. This action is invariant up to a total derivative under the gauge transformations

$$\delta_\epsilon\varphi_{\mu\nu}=-2\partial_{(\mu}\epsilon_{\nu)}+\eta_{\mu\nu}\partial_\rho\epsilon^{\rho},\qquad \delta_\epsilon\Gamma_{\mu\nu\rho}=-\chi\,\partial_\mu\partial_\nu\epsilon_\rho,\tag{3.222}$$

and it is equivalent on-shell to the Fierz–Pauli action because it gives the same equations of motion: the equations of motion of the fields $\varphi^{\mu\nu}$ and $\Gamma_{\mu\nu}{}^{\rho}$ are

$$\begin{aligned}\chi\frac{\delta S^{(1)}}{\delta\varphi^{\mu\nu}}&=-\partial_{(\mu}\Gamma_{\nu)\rho}{}^{\rho}+\partial_\rho\Gamma_{\mu\nu}{}^{\rho}=0,\\ \chi^2\frac{\delta S^{(1)}}{\delta\Gamma_{\mu\nu}{}^{\rho}}&=2\Gamma_{\rho}{}^{(\mu\nu)}-\eta^{\mu\nu}\Gamma_{\rho\lambda}{}^{\lambda}-\eta^{\tau\sigma}\Gamma_{\tau\sigma}{}^{(\mu}\eta^{\nu)}{}_{\rho}-\chi\,\partial_\rho\varphi^{\mu\nu}+\chi\,\eta_{\rho}{}^{(\mu}\partial_\sigma\varphi^{\nu)\sigma}=0.\end{aligned}\tag{3.223}$$

The second equation is just a constraint for $\Gamma_{\mu\nu}{}^{\rho}$. On contracting it with $\eta^{\rho\sigma}$, we obtain

$$\eta^{\rho\sigma}\Gamma_{\rho\sigma}{}^{\nu}=\chi\,\partial_\lambda\varphi^{\lambda\nu},\tag{3.224}$$

and, on contracting instead with $\eta_{\mu\nu}$ and using the last result, we find

$$\Gamma_{\rho\lambda}{}^{\lambda}=-\frac{1}{d-2}\chi\,\partial_\rho\varphi,\qquad \varphi=\varphi_\mu{}^{\mu}.\tag{3.225}$$

Using now these two last equations in the equation for $\Gamma_{\mu\nu}{}^{\rho}$, we obtain

$$\Gamma_{\rho\mu\nu}+\Gamma_{\nu\rho\mu}=\chi\,\partial_\rho h_{\mu\nu},\qquad h_{\mu\nu}=\varphi_{\mu\nu}-\frac{1}{d-2}\eta_{\mu\nu}\varphi.\tag{3.226}$$

In order to solve for Γ, we add to this equation $(\rho\mu\nu)$ the permutation $\mu\nu\rho$ and subtract the permutation $\nu\rho\mu$, obtaining, finally,

$$\Gamma_{\rho\mu\nu}=\tfrac{1}{2}\chi\{\partial_\rho h_{\mu\nu}+\partial_\mu h_{\nu\rho}-\partial_\nu h_{\rho\mu}\}.\tag{3.227}$$

On substituting this into the equation of motion for $\varphi^{\mu\nu}$, we find that, in terms of the variable $h_{\mu\nu}$, it takes the form

$$\frac{\delta S^{(1)}}{\delta\varphi^{\mu\nu}} = -\tfrac{1}{2}\left[\mathcal{D}_{\mu\nu}(h) - \frac{1}{d-2}\eta_{\mu\nu}\mathcal{D}_\rho{}^\rho(h)\right] = 0, \tag{3.228}$$

which is equivalent to the Fierz–Pauli equation.

Now we want to find a correction $S^{(2)}$ such that the equation of motion becomes

$$\mathcal{D}_{\mu\nu}(h) = \chi t_{\mu\nu}, \tag{3.229}$$

for the total action $S^{(1)} + S^{(2)}$, i.e. we have to obtain

$$\frac{\delta S^{(2)}}{\delta\varphi^{\mu\nu}} = \tfrac{1}{2}\chi\left(t_{\mu\nu} - \frac{1}{d-2}\eta_{\mu\nu}t_\rho{}^\rho\right) \equiv \tau_{\mu\nu}, \tag{3.230}$$

where $t_{\mu\nu}$ is the energy–momentum tensor of $\varphi^{\mu\nu}$ in $S^{(1)}$. We first calculate $t_{\mu\nu}$ using Rosenfeld's prescription. In writing the action $S^{(1)}$ in the background metric $\gamma_{\mu\nu}$, we will assume (and this is one of the key points of this argument) that $\varphi^{\mu\nu}$ is a *tensor density* of weight $w = 1$, i.e. it transforms as $\sqrt{|\gamma|}\, f^{\mu\nu}$, where $f^{\mu\nu}$ is an ordinary tensor. Thus, there is no need to introduce a $\sqrt{|\gamma|}$ factor in front of $\varphi^{\mu\nu}$ and, furthermore, $\varphi^{\mu\nu}$ is independent of the background metric. By expanding the covariant derivatives[37] of $\Gamma_{\mu\nu}{}^\rho$, we obtain

$$\begin{aligned} S^{(1)}[\varphi^{\mu\nu}, \Gamma_{\mu\nu}{}^\rho, \gamma_{\mu\nu}] = \frac{1}{\chi^2}\int d^dx\ \Big\{&-\chi\varphi^{\mu\nu}\big[2\partial_{[\mu}\Gamma_{\rho]\nu}{}^\rho - 2C_{\nu[\mu}{}^\sigma\Gamma_{\rho]\sigma}{}^\rho + 2C_{\sigma[\mu}{}^\rho\Gamma_{\rho]\nu}{}^\sigma\big] \\ &+ \sqrt{|\gamma|}\,\gamma^{\mu\nu}2\Gamma_{\lambda[\mu}{}^\rho\Gamma_{\rho]\nu}{}^\lambda\Big\}. \end{aligned} \tag{3.231}$$

A long calculation gives

$$\begin{aligned} \chi^2 t_{\alpha\beta} &= -\frac{2\chi^2}{\sqrt{|\gamma|}}\,\frac{\delta S^{(1)}}{\delta\gamma^{\alpha\beta}}\bigg|_{\gamma_{\alpha\beta}=\eta_{\alpha\beta}} \\ &= -4\Gamma_{\lambda[\alpha}{}^\rho\Gamma_{\rho]\beta}{}^\lambda + 2\eta_{\alpha\beta}\eta^{\kappa\delta}\Gamma_{\lambda[\kappa}{}^\rho\Gamma_{\rho]\delta}{}^\lambda \\ &\quad - \chi\,\partial_\tau\big\{\eta_{\alpha\beta}\varphi^{\mu\nu}\Gamma_{\mu\nu}{}^\tau + 2\varphi^\tau{}_{(\alpha}\Gamma_{\beta)\rho}{}^\rho + \varphi_{\alpha\beta}\Gamma^\tau{}_\rho{}^\rho \\ &\qquad - 2\varphi^{\tau\mu}\Gamma_{\mu(\alpha\beta)} - 2\varphi^\mu_{(\alpha}\left[\Gamma_{\mu|\beta)}{}^\tau - \Gamma_\mu{}^\tau{}_{|\beta)}\right]\big\}, \end{aligned} \tag{3.232}$$

and, thus,

$$\begin{aligned} \tau_{\alpha\beta} = {} & -2\chi^{-1}\Gamma_{\lambda[\alpha}{}^\rho\Gamma_{\rho]\beta}{}^\lambda \\ & + \tfrac{1}{2}\partial_\tau\Bigg\{\frac{1}{d-2}\eta_{\alpha\beta}\big(\varphi^{\mu\nu}\Gamma_\mu{}^\tau{}_\nu - \tfrac{1}{2}\varphi\Gamma^\tau{}_\rho{}^\rho\big) - \big(\varphi^\tau{}_\alpha\Gamma_{\beta\rho}{}^\rho + \varphi^\tau{}_\beta\Gamma_{\alpha\rho}{}^\rho - \varphi_{\alpha\beta}\Gamma^\tau{}_\rho{}^\rho\big) \\ & + \varphi^{\tau\mu}\big(\Gamma_{\mu\alpha\beta} + \Gamma_{\mu\beta\alpha}\big) + \varphi^\mu{}_\alpha\big(\Gamma_{\mu\beta}{}^\tau - \Gamma_\mu{}^\tau{}_\beta\big) + \varphi^\mu{}_\beta\big(\Gamma_{\mu\alpha}{}^\tau - \Gamma_\mu{}^\tau{}_\alpha\big)\Bigg\}. \end{aligned} \tag{3.233}$$

[37] In the end $\Gamma_{\mu\nu}{}^\rho$ will not be a general-covariant tensor. However, it is a Lorentz tensor and Rosenfeld's prescription tells us to replace its partial derivatives by covariant derivatives in order to find Lorentz energy–momentum tensors.

The correction to the action with the property (3.230) is precisely

$$S^{(2)} = \frac{1}{\chi^2}\int d^d x \left\{-2\chi\varphi^{\alpha\beta}\Gamma_{\lambda[\alpha}{}^{\rho}\Gamma_{\rho]\beta}{}^{\lambda}\right\}. \tag{3.234}$$

One could naively think that, with this correction, we can obtain only the first term (that quadratic in $\Gamma_{\mu\nu}{}^{\rho}$) in $\tau_{\alpha\beta}$. However, we have to take into account that the equation for $\Gamma_{\mu\nu}{}^{\rho}$ changes and, hence, substituting its solution into the equation for $\varphi^{\mu\nu}$ will give us all the terms we need. Observe also that this correction is cubic in fields whereas the action we started from is quadratic. Finally, observe that this term will not contribute to the energy–momentum tensor: there are no Minkowski metrics here to be replaced by the background metric and there is no need to introduce $\sqrt{|\gamma|}$ because $\varphi^{\mu\nu}$ is, by hypothesis, a tensor density. Thus, if this term really works, we will not need to introduce any more corrections.

For the total action

$$\boxed{S^{(1)} + S^{(2)} = \frac{1}{\chi^2}\int d^d x \left\{-\chi\varphi^{\mu\nu}2\partial_{[\mu}\Gamma_{\rho]\nu}{}^{\rho} + (\eta^{\mu\nu} - \chi\varphi^{\mu\nu})2\Gamma_{\lambda[\mu}{}^{\rho}\Gamma_{\rho]\nu}{}^{\lambda}\right\},} \tag{3.235}$$

we find the following equations of motion:

$$\begin{aligned}
\chi\frac{\delta(S^{(1)} + S^{(2)})}{\delta\varphi^{\mu\nu}} &= -R_{\mu\nu}(\Gamma) = 0, \\
\chi^2\frac{(\delta S^{(1)} + S^{(2)})}{\delta\Gamma_{\mu\nu}{}^{\rho}} &= 2\Gamma_{\rho}{}^{(\mu\nu)} - \eta^{\mu\nu}\Gamma_{\rho\lambda}{}^{\lambda} - \eta^{\lambda\sigma}\Gamma_{\lambda\sigma}{}^{(\mu}\eta^{\nu)}{}_{\rho} - \chi\partial_{\rho}\varphi^{\mu\nu} + \chi\eta_{\rho}{}^{(\mu}\partial_{\sigma}\varphi^{\nu)\sigma} \\
&\quad - 2\chi\varphi^{\delta(\mu}\Gamma_{\rho\delta}{}^{\nu)} + \chi\varphi^{\mu\nu}\Gamma_{\rho\sigma}{}^{\sigma} + \chi\varphi^{\lambda\sigma}\Gamma_{\lambda\sigma}{}^{(\mu}\eta^{\nu)}{}_{\rho} = 0,
\end{aligned} \tag{3.236}$$

where $R_{\mu\nu}(\Gamma)$ is nothing but the Ricci tensor associated with the connection $\Gamma_{\mu\nu}{}^{\rho}$ given in Eq. (1.33). By defining

$$\mathfrak{g}^{\mu\nu} = \eta^{\mu\nu} - \chi\varphi^{\mu\nu} \tag{3.237}$$

and its inverse $\mathfrak{g}^{\mu\rho}\mathfrak{g}_{\rho\nu} = \mathfrak{g}^{\mu}{}_{\nu} = \delta^{\mu}{}_{\nu}$, which we are going to use as a metric to raise and lower indices, we can write

$$\chi^2\frac{(\delta S^{(1)} + S^{(2)})}{\delta\Gamma_{\mu\nu}{}^{\rho}} = 2\mathfrak{g}^{\delta(\mu}\Gamma_{\rho\delta}{}^{\nu)} - \mathfrak{g}^{\mu\nu}\Gamma_{\rho\delta}{}^{\delta} - \mathfrak{g}^{\lambda\sigma}\Gamma_{\lambda\sigma}{}^{(\mu}\mathfrak{g}^{\nu)}{}_{\rho} + \partial_{\rho}\mathfrak{g}^{\mu\nu} - \mathfrak{g}_{\rho}{}^{(\mu}\partial_{\sigma}\mathfrak{g}^{\nu)\sigma} = 0. \tag{3.238}$$

Now we proceed as before: we contract this equation of motion with $\mathfrak{g}_{\mu}{}^{\rho}$, giving

$$\Gamma_{\lambda}{}^{\lambda\nu} = -\partial_{\sigma}\mathfrak{g}^{\sigma\nu}, \tag{3.239}$$

and then contract with $\mathfrak{g}_{\mu\nu}$, using the last equation, giving

$$\Gamma_{\rho\lambda}{}^{\lambda} = \frac{1}{d-2}\mathfrak{g}_{\mu\nu}\partial_{\rho}\mathfrak{g}^{\mu\nu} = \frac{1}{d-2}\partial_{\rho}\ln|\mathfrak{g}|, \qquad |\mathfrak{g}| \equiv \det\mathfrak{g}^{\mu\nu}. \tag{3.240}$$

We already see here that the expression for $\Gamma_{\mu\nu}{}^{\rho}$ in terms of $\varphi^{\mu\nu}$ involves an infinite series of terms. This is the reason why one iteration will be enough even though we had expected an infinite series of corrections.

On substituting the last two results into the equation for $\Gamma_{\mu\nu}{}^{\rho}$, we find

$$\Gamma_{\rho\mu}{}^{\sigma}\mathfrak{g}_{\sigma\nu}|\mathfrak{g}|^{\frac{1}{d-2}} + \Gamma_{\rho\nu}{}^{\sigma}\mathfrak{g}_{\sigma\mu}|\mathfrak{g}|^{\frac{1}{d-2}} = \partial_{\rho}\Big(|\mathfrak{g}|^{\frac{1}{d-2}}\mathfrak{g}_{\mu\nu}\Big). \tag{3.241}$$

We see that, again, it is convenient to make the following definition:

$$g_{\mu\nu} \equiv |\mathfrak{g}|^{\frac{1}{d-2}}\mathfrak{g}_{\mu\nu}, \qquad \Rightarrow \mathfrak{g}^{\mu\nu} = \sqrt{|g|}\, g^{\mu\nu}. \tag{3.242}$$

In terms of the variable $g_{\mu\nu}$, the above equation can be solved using the same procedure as before. The result is that $\Gamma_{\mu\nu}{}^{\rho}$ is given by the Christoffel symbols associated with the metric $g_{\mu\nu}$ (1.44). The two equations of motion can now be combined into one:

$$R_{\mu\nu}(g) = 0, \tag{3.243}$$

where $R_{\mu\nu}(g)$ is the Ricci tensor associated with the Levi-Cività connection of the metric $g_{\mu\nu}$. This is the *vacuum Einstein equation*, the equation of motion of GR, as we will see.

So far we have not shown that the corrected action has the required self-consistency property. We are now going to do this, and this will allow us to claim that the vacuum Einstein equation is the self-consistent extension of the Fierz–Pauli theory we were looking for, written in terms of the new variable $g_{\mu\nu}$, which turns out to have a geometrical meaning that is really unexpected, given our starting point of view.

We turn back to the equation for $\Gamma_{\mu\nu}{}^{\rho}$ and try to solve it without the use of $\mathfrak{g}^{\mu\nu}$ and its inverse, by raising and lowering indices with the Minkowski metric again. First, we contract it with $\eta_{\mu}{}^{\rho}$, giving

$$(\eta^{\rho\sigma} - \chi\varphi^{\rho\sigma})\Gamma_{\rho\sigma}{}^{\nu} = \chi\,\partial_{\sigma}\varphi^{\sigma\nu}. \tag{3.244}$$

Contracting now with $\eta_{\mu\nu}$ and substituting into it the last result, we obtain

$$\Gamma_{\rho\delta}{}^{\delta} = -\frac{1}{d-2}\chi\big[-\partial_{\rho}\varphi + 2\varphi^{\delta}{}_{\mu}\Gamma_{\rho\delta}{}^{\mu} - \varphi\Gamma_{\rho\delta}{}^{\delta}\big], \tag{3.245}$$

and, on plugging these results into the full equation, we arrive at

$$\begin{aligned} \Gamma_{\rho\mu\nu} + \Gamma_{\nu\rho\mu} &= \chi\,\partial_{\rho}h_{\mu\nu} + f_{\rho\mu\nu}, \\ f_{\rho\mu\nu} &= 2\chi\varphi^{\delta}{}_{(\mu|}\Gamma_{\rho\delta|\nu)} - \chi\varphi_{\mu\nu}\Gamma_{\rho\delta}{}^{\delta} - \frac{1}{d-2}\chi\eta_{\mu\nu}\big[2\varphi^{\delta}{}_{\lambda}\Gamma_{\rho\delta}{}^{\lambda} - \varphi\Gamma_{\rho\delta}{}^{\delta}\big], \end{aligned} \tag{3.246}$$

which can be "solved" in exactly the same way, giving

$$\Gamma_{\rho\mu\nu} = \tfrac{1}{2}\chi\{\partial_{\rho}h_{\mu\nu} + \partial_{\mu}h_{\nu\rho} - \partial_{\nu}h_{\rho\mu}\} + \tfrac{1}{2}\{f_{\rho\mu\nu} + f_{\mu\nu\rho} - f_{\nu\rho\mu}\}. \tag{3.247}$$

There are Γs on the r.h.s. of this equation, but we do not need anything better (neither can we obtain it without inverting the matrix $\varphi^{\mu\nu}$). On substituting into the equation for

$\varphi^{\mu\nu}$, we find

$$-\frac{1}{\chi}R_{\mu\nu}(\Gamma) = -\tfrac{1}{2}\left[\mathcal{D}_{\mu\nu}(h) - \frac{1}{d-2}\eta_{\mu\nu}\mathcal{D}_{\rho}{}^{\rho}(h)\right] - 2\chi^{-1}\Gamma_{\lambda[\mu}{}^{\rho}\Gamma_{\rho]\nu}{}^{\lambda}$$
$$+\frac{1}{2\chi}\partial_{\tau}\left\{f_{\nu\mu}{}^{\tau} + f_{\mu}{}^{\tau}{}_{\nu} - f^{\tau}{}_{\nu\mu} + \frac{2\chi}{d-2}\eta^{\tau}{}_{(\nu|}\left[2\varphi^{\delta}{}_{\lambda}\Gamma_{|\mu)\delta}{}^{\lambda} - \varphi\Gamma_{|\mu)\delta}{}^{\delta}\right]\right\}. \tag{3.248}$$

On expanding the last term we find agreement with Eqs. (3.228), (3.230), and (3.233).

Let us review this result: we have obtained a first-order action for $\varphi^{\mu\nu}$, which, to lowest order in an expansion in the parameter χ, is equivalent to the free Fierz–Pauli action. The full equation of motion is the equation of motion of GR in vacuum and we have shown that it is equivalent to the Fierz–Pauli equation with a source that is precisely the conserved energy–momentum tensor of the $\varphi^{\mu\nu}$ field that one derives directly from the action using the Rosenfeld prescription and without having to add any $\partial_{\rho}\Psi^{\mu\nu\rho}$ term. The action Eq. (3.235) satisfies the physical criterion of self-consistency we asked for and is the action of GR. We have, though, not checked that the Rosenfeld energy–momentum tensor is the Noether current associated with the symmetry of the problem and we have not discussed the gauge invariance of the result.

In this construction we have found that the objects that appear in the self-consistent action have a simple geometrical interpretation: there is a non-linear function of the field $\varphi^{\mu\nu}$, $g_{\mu\nu}(\varphi)$, that we can interpret as a metric tensor and the other field in the first-order action $\Gamma_{\mu\nu}{}^{\rho}$ is the associated Levi-Cività connection on-shell. The equation of motion (the vacuum Einstein equation) states that the metric is Ricci-flat. This equation is covariant under GCTs.

This geometrical interpretation is very powerful because all the infinite non-linear terms that the theory would have when written in terms of $\varphi^{\mu\nu}$ are packaged into objects that can be easily manipulated. However, this new interpretation also goes far beyond the original theory, which was a SRFT in Minkowski spacetime (that is, $\mathbb{R}^d$ equipped with the Minkowski metric). In the original SRFT of gravity, any gravitational field is always defined on $\mathbb{R}^d$ and the Minkowski metric is always there. However, in GR, in many cases it is not possible to find or define a Minkowski metric in the whole spacetime. Furthermore, many metric fields that solve the equations of motion of GR cannot be interpreted as metric fields defined on the whole $\mathbb{R}^d$ but demand spacetime manifolds with different topology. This is particularly true when there are submanifolds on which the metric field is singular. The geometrical theory is therefore much richer because non-trivial topology and causal structures (as in black-hole spacetimes) will be the origin of very interesting phenomena (such as Hawking radiation).

Another strong point of the geometrical interpretation is that it provides us with a simple principle to couple matter to gravity: that of *covariance under general coordinate transformations* ("general covariance"), which encodes the equivalence principle in its stronger form. The matter action has to be a scalar under GCTs and this is achieved by introducing the metric field in the right places (precisely as in Rosenfeld's prescription for how to calculate the energy–momentum tensor). A general-covariant matter energy–momentum tensor arises from this formalism in a natural way. However, in the SRFT approach we would have to find, case by case, the corrections to the lowest-order coupled system, which we know is inconsistent. A weakness of the geometrical point of view is that there is no general-covariant energy–momentum tensor of the gravitational field itself. As we have seen, there is a Lorentz-covariant energy–momentum tensor (or pseudotensor) of the gravitational field embedded in the Ricci tensor together with the wave operator, but it cannot be promoted to a general-covariant tensor, as can be understood from the equivalence principle. This obscures the physical interpretation of vacuum solutions of the geometrical theory, which are not strictly speaking *vacuum* solutions since the whole spacetime is filled by a non-trivial gravitational field that acts as a source for itself. We will come back to this point in Chapter 6.

Where does this principle of general covariance come from? We started from a theory with an Abelian gauge symmetry[38] $\delta_\epsilon^{(0)} h_{\mu\nu} = -2\partial_{(\mu}\epsilon_{\nu)}$. We argued that this symmetry was necessary in order to have a consistent theory of free massless spin-2 particles. Then we coupled this free theory to the conserved energy–momentum tensor of the matter fields, saw the need to introduce a self-coupling of the spin-2 field, and argued that the form of the coupling should be dictated by gauge invariance with respect to the corrected transformations $\delta_\epsilon^{(1)} h_{\mu\nu} = -2\partial_{(\mu}\epsilon_{\nu)} - \chi \mathcal{L}_\epsilon h_{\mu\nu}$, which combined the Abelian gauge symmetry we started from and "localized" translations in such a way that the commutator of two $\delta_\epsilon^{(1)}$ infinitesimal transformations gives another $\delta_\epsilon^{(1)}$ transformation. This is the only possible extension of the Abelian $\delta_\epsilon^{(0)}$ transformations [739, 933] and the algebra is the algebra of infinitesimal GCTs.[39] In fact, we can easily see how the full gauge transformation $\delta_\epsilon^{(1)} h_{\mu\nu}$ arises from the effect of a GCT on the metric $g_{\mu\nu} = \eta_{\mu\nu} + \chi h_{\mu\nu}$, just by substituting and

[38] Any two of these gauge transformations commute because $\delta_{\xi_1}^{(0)} \delta_{\xi_2} h_{\mu\nu} = \delta_{\xi_1+\xi_2}^{(0)} h_{\mu\nu}$.

[39] As shown in [933], it is possible to have a self-coupled spin-2 theory with only "normal spin-2 gauge symmetry" ($\delta_\epsilon^{(0)}$). For instance, we can add to the Fierz–Pauli Lagrangian a term proportional to some (for instance, the third) power of the linearized Ricci scalar

$$\partial^2 h - \partial_\mu \partial_\nu h^{\mu\nu}, \tag{3.249}$$

which is exactly invariant under $\delta_\epsilon^{(0)}$. Of course, the resulting higher-derivative theory cannot have the same interpretation, since the r.h.s. of the equation of motion is not the gravitational energy–momentum tensor. Also, we can couple the linear theory to matter and obtain an interacting theory that is invariant under $\delta_\epsilon^{(0)}$: we just have to add to the free-matter Lagrangian and the Fierz–Pauli Lagrangian an interaction term of the form

$$\int d^d x \, h^{\mu\nu} J_{\mu\nu}(\varphi), \tag{3.250}$$

where $J_{\mu\nu}$ is any symmetric, identically conserved tensor built out of φ and its derivatives. This excludes the matter energy–momentum tensor, which is conserved only on-shell. Since $J_{\mu\nu}$ is identically conserved, the modification introduced into the equations of motion for matter by the coupling to gravity is immaterial. Local $J_{\mu\nu}$s can be constructed from local four-index tensors with the symmetries $J_{\mu\rho\nu\sigma} = J_{[\mu\rho][\nu\sigma]} = J_{\nu\sigma\mu\rho}$, defining

$$J_{\mu\nu} = \partial^\rho \partial^\sigma J_{\mu\rho\nu\sigma}. \tag{3.251}$$

These $J_{\mu\nu}$s are called *Pauli terms* in [943]. It is also possible to define identically conserved nonlocal $J_{\mu\nu}$s, for instance the non-local projection of the energy–momentum tensor Eq. (3.170). In all these cases, we see that the spin-2 field does not couple to the total energy–momentum tensor and the quantum theories are not consistent, according to [942].

expanding in powers of χ the infinitesimal GCT

$$\delta_\epsilon g_{\mu\nu} = -\mathcal{L}_\epsilon g_{\mu\nu} = -2\nabla_{(\mu}\epsilon_{\nu)}.$$

We can consider, then, that the gauge transformations that we found in the Noether method are just the perturbative expansion of GCTs. The self-consistent Fierz–Pauli theory can be considered as a perturbative expansion of the geometrical theory (GR) either in powers of a weak field $\varphi^{\mu\nu}$ or in powers of the dimensional coupling constant χ, which we know from experience is extremely small. From this point of view, the geometrical action is extremely non-perturbative. Thus, the free Fierz–Pauli theory has been the starting point of any attempt to quantize the gravitational interaction in the standard sense (that is, in perturbation theory), as a special-relativistic quantum field theory (SRQFT).[40] Although they were unsuccessful,[41] these attempts have rendered many benefits to the general theory of covariant quantization of gauge field theories,[42] leading, for instance, to Feynman's discovery of ghosts [387].

We know that the theory we have obtained is experimentally correct, although most experiments probe the perturbative regime only to a very low order in χ. However, we have two very different interpretations. Which is the right one? This is a very difficult question which is still open. For many years, the geometrical form of the theory of *general relativity*, which was the first to be obtained (it is clearly easier to obtain) and was proposed by Einstein himself, was accepted as the only possible one. On the other hand, the SRFT form of the theory is necessary in order to study aspects such as the self-coupling of gravity and gravitational waves. Also, any *standard* quantization of GR[43] has to go through the identification of the particles which are going to be the gravitational-field quanta and this takes us to the SRFT. However, the quantization of this theory has been unsuccessful.[44] We are tempted to say that any theory of gravity with the same weak-field limit that we could quantize should be the true theory. Actually, this is the main argument in favor of string theory.

Meanwhile, it is probably healthy to use both aspects of the theory in the appropriate realms. This is what we intend to do here.

There is a final detail we should comment upon: we have obtained geometrical equations of motion, but the action Eq. (3.235) is not fully geometrical in the sense that it is not invariant under GCTs. We need to add a total derivative term to it:

$$S^{(0)} = \frac{1}{\chi^2}\int d^dx \left\{2\eta^{\mu\nu}\partial_{[\mu}\Gamma_{\rho]\nu}{}^{\rho}\right\}. \tag{3.252}$$

---

40 Classical references on this approach are [318, 319, 387, 922]. More can be found in [30, 386].

41 Pure gravity, perturbatively derived from GR, is one-loop convergent but it is divergent at the same order when coupled to matter [311–4, 889, 891] and at two-loop order without coupling to matter [462, 911].

42 See e.g. [914].

43 Other proposals such as Euclidean quantum gravity and loop quantization, which we may consider less standard, do not need the identification of gravitons.

44 In view of the fact that the self-consistency of the theory requires the inclusion of an infinite number of higher-order terms, it is legitimate to wonder whether the lack of success is due to the theory itself, to the method of quantization, or just to our inability to quantize in the standard manner a theory with an infinite number of terms without making truncations that would render it inconsistent even if only to some order in χ.

Then, written in terms of $g_{\mu\nu}$ and $\Gamma_{\mu\nu}{}^{\rho}$, the total action $S^{(0)} + S^{(1)} + S^{(2)}$ becomes the first-order Einstein–Hilbert action

$$\boxed{S_{\mathrm{EH}}[g_{\mu\nu}, \Gamma_{\mu\nu}{}^{\rho}] = \frac{1}{\chi^2}\int d^d x \sqrt{|g|}\, g^{\mu\nu} R_{\mu\nu}(\Gamma),} \tag{3.253}$$

which can be taken as the starting point of GR. Observe that the equation of motion of $g_{\mu\nu}$, looks different from that of $\mathfrak{g}^{\mu\nu}$, although it is completely equivalent. In Chapter 4 we will see in detail that the equation of motion is

$$G^{\mu\nu} = 0, \tag{3.254}$$

where $G^{\mu\nu} = R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R$ is the Einstein tensor.

To end this section, we would like to remark that the addition of the total derivative changes the gravity energy–momentum tensor by a $\partial_\rho \Psi^{\rho\mu\nu}$ term. In any case, we are going to need to add total derivatives to this action for various reasons. The issue of the gravitational-field energy–momentum tensor will be studied in Chapter 6.

3.3 General relativity

The search for self-consistency of the Fierz–Pauli theory has led us to the Einstein–Hilbert action, Eq. (3.253), which has the property of invariance under GCTs (*general covariance*). This property, elevated to the rank of the *principle of general covariance of relativity* (PGR) is the basis of the theory of general relativity which we want to review here in a extremely condensed way.

The PGR can be considered as the generalization of the principle of (special) relativity and states that all laws of physics should be form-invariant (or covariant) under arbitrary changes of reference frame. Since any SRFT requires the use of the standard constant Minkowski metric $(\eta_{\mu\nu}) = \mathrm{diag}(+ - - \cdots -)$ which is invariant only under transformations between inertial frames related by Poincaré transformations, general covariance requires its substitution by a metric field $g_{\mu\nu}(x)$ behaving as a tensor under all GCTs and the substitution of all partial derivatives by (general-)covariant derivatives. If the metric field $g_{\mu\nu}$ is simply the Minkowski metric in a non-Cartesian, non-inertial reference frame, then it will always be possible to perform a GCT to a Cartesian, inertial reference frame in which the metric $g_{\mu\nu}$ takes the constant standard form $\eta_{\mu\nu}$. Later on we will extend this property to more general metrics in a local form. Finally, if we do not want to introduce any new fields in using covariant derivatives, we have to use the Levi-Cività connection $\Gamma(g)$.

To see how far we are taken by this principle, we first apply it to point-particles.

Point-particle actions. Actions for free point-particles moving in spacetime that are consistent with the PGR and reduce to the special-relativistic action can be readily written by replacing $\eta_{\mu\nu}$ by a general metric $g_{\mu\nu}$ in Eqs. (3.8), (3.29), and (3.32). In this way we obtain the Nambu–Goto-type action for a massive particle in a general background metric $g_{\mu\nu}(x)$,

$$\boxed{S[X^\mu(\xi)] = -Mc\int d\xi\, \sqrt{g_{\mu\nu}(X)\dot{X}^\mu\dot{X}^\nu},} \tag{3.255}$$

which is still proportional to the particle's proper time s as in Eq. (3.22), where the proper time is now defined by

$$\frac{ds}{d\xi} = \sqrt{g_{\mu\nu}(X)\dot{X}^\mu\dot{X}^\nu}. \tag{3.256}$$

The Polyakov-type action for a massive particle is

$$\boxed{S[X^\mu(\xi), \gamma(\xi)] = -\frac{Mc}{2}\int d\xi \sqrt{\gamma}\left[\gamma^{-1}g_{\mu\nu}(X)\dot{X}^\mu\dot{X}^\nu + 1\right],} \tag{3.257}$$

which is also equivalent to the Nambu–Goto-type action upon elimination of the world-line metric $\gamma(\xi)$ through its own equation of motion and the Polyakov-type action for a massless particle:

$$\boxed{S[X^\mu(\xi), \gamma'(\xi)] = -\frac{p}{2}\int d\xi \sqrt{\gamma'}\, \gamma'^{\,-1}g_{\mu\nu}(X)\dot{X}^\mu\dot{X}^\nu.} \tag{3.258}$$

These three actions are manifestly invariant under reparametrizations of the worldline as in the Minkowski case. Thus, there is going to be a constraint associated with this invariance and it is going to coincide with the mass-shell condition in each case:

$$P^\mu P_\mu = M^2c^2, \qquad P_\mu \equiv \frac{\partial\mathcal{L}}{\partial\dot{X}^\mu} \tag{3.259}$$

(evidently $M = 0$ in the massless case).

Furthermore, they all are invariant[45] under spacetime GCTs $X^\mu \to X^{\mu\,\prime}(X)$ under which the metric transforms as follows:

$$g_{\mu\nu}(X) = g'_{\rho\sigma}\left[X'(X)\right]\frac{\partial X'^{\,\rho}}{\partial X^\mu}\frac{\partial X'^{\,\sigma}}{\partial X^\nu}, \tag{3.260}$$

so the combination $g_{\mu\nu}(X)\dot{X}^\mu\dot{X}^\nu$ is invariant.

Since the Polyakov-type action is equivalent to the Nambu–Goto-type one, let us find the equations of motion derived from the Nambu–Goto-type action. These are

$$\ddot{X}^\lambda + \Gamma_{\rho\sigma}{}^\lambda\dot{X}^\rho\dot{X}^\sigma - \frac{d}{d\xi}\left(\ln\gamma^{\frac{1}{2}}\right)\dot{X}^\lambda = 0, \tag{3.261}$$

[45] By invariant we mean "form-invariant" or, as it is sometimes put, *covariant*. This is all that the PGR requires.

where we have introduced the *induced metric* on the worldline γ

$$\gamma(\xi) \equiv g_{\mu\nu}(X)\dot{X}^{\mu}\dot{X}^{\nu}. \tag{3.262}$$

We use for it the same symbol as for the auxiliary metric of the Polyakov-type action because the equation of motion for γ says that γ is the induced worldline metric.

We can easily recognize in Eq. (3.261) the *geodesic equation* written in terms of an arbitrary parameter. Curves obeying that equation are called *geodesics* and are the curves of minimal (occasionally maximal) proper length between two given points. When $\xi = s$, the proper time, then $\gamma = 1$ and the third term in Eq. (3.261) vanishes and the standard form of the geodesic equation is recovered:

$$\ddot{X}^{\lambda} + \Gamma_{\rho\sigma}{}^{\lambda}\dot{X}^{\rho}\dot{X}^{\sigma} = 0. \tag{3.263}$$

If the metric is $\eta_{\mu\nu}$, it is clear that we recover all the special-relativistic results. Furthermore, if the metric is related to $\eta_{\mu\nu}$ through a GCT, it is clear that we will be describing the same motion (straight lines in spacetime) in some system of curvilinear coordinates. Thus, even though it is difficult to see, the dynamics of the particle will have the same $d(d+1)/2$ conserved quantities associated with the invariances of the Minkowski metric in Cartesian coordinates. Now that we are dealing with general curvilinear coordinates, it is good to have a better characterization of the invariances of a metric and how they are associated with conserved quantities in the dynamics of a particle.

Let us consider the effect of infinitesimal transformations of the form

$$\begin{aligned} \delta X^{\mu} &= \epsilon^{\mu}(X), \\ \delta g_{\mu\nu} &= \epsilon^{\lambda}\partial_{\lambda} g_{\mu\nu}. \end{aligned} \tag{3.264}$$

It is worth stressing that these transformations are not GCTs in spacetime (the metric does not transform in the required way). We know that the action (3.255) is invariant under arbitrary GCTs. However, under the above transformations

$$\delta S_{\text{pp}} = -Mc \int d\xi \frac{1}{2\sqrt{g_{\mu\nu}\dot{X}^{\mu}\dot{X}^{\nu}}} \dot{X}^{\rho}\dot{X}^{\sigma}\mathcal{L}_{\epsilon} g_{\rho\sigma}, \tag{3.265}$$

and is invariant only if $\epsilon^{\mu} = \epsilon k^{\mu}$, where ϵ is an infinitesimal constant parameter and k^{μ} is a Killing vector satisfying the Killing equation (1.107).

These transformations can be exponentiated, giving a one-dimensional group (for one Killing vector) that leaves the action invariant. There is a conserved quantity associated with it via the Noether theorem for global symmetries,[46]

$$P(k) = -\frac{Mc}{\sqrt{g_{\mu\nu}\dot{X}^{\mu}\dot{X}^{\nu}}} k_{\rho}\dot{X}^{\rho}, \tag{3.266}$$

[46] The components of k^{μ} are fixed functions of the spacetime coordinates and the parameters of the group have to be constant over the worldline; they cannot be arbitrary functions of ξ. Thus, this is a group of global transformations. These transformations can be gauged by the standard method of introducing a gauge vector and a covariant derivative, as will be seen in due course.

which can be interpreted as the components of the momentum vector in the direction of the Killing vector.

This general framework can be applied to any metric in any coordinate system. We can use it to recover the conserved quantities of a free particle moving in Minkowski spacetime. First of all, observe that we can always use coordinates adapted to a given Killing vector k^μ: there is a coordinate z such that $k^\mu \partial_\mu = \partial_z$ and $\partial_z g_{\mu\nu} = 0$. Then, there is always a coordinate system in which the action does not depend on the variable $Z(\xi)$ and hence the momentum associated with it is conserved as usual. Thus, we are simply encoding known facts in coordinate-independent form. Second, we can check that the above general expression gives the usual linear- and angular-momentum components when we use the Killing vectors of the Minkowski metric:

$$k^{(\mu)\,\rho} = \eta^{\mu\rho}, \qquad k^{([\mu\nu])\,\rho} = 2\eta^{\rho[\mu} x^{\nu]}, \tag{3.267}$$

where (μ) and $([\mu\nu])$ are labels for the d translational and $d(d-1)/2$ rotational isometries.

To finish this digression, let us mention that the Polyakov-type actions (3.257) and (3.258) are one-dimensional examples of what is called a *non-linear σ-model*.[47] The non-linearity is associated with the dependence of the metric on the coordinates, which are the dynamical degrees of freedom.

The principle of equivalence. Accepting that, according to the PGR, the action (3.255) gives the dynamics of a massive particle in the background given by the metric $g_{\mu\nu}$, we are led to the discovery of the *principle of equivalence of gravitation and inertia* (PEGI) formulated by Einstein in [350, 351]: consider a near-Minkowskian metric $g_{\mu\nu} = \eta_{\mu\nu} + \chi h_{\mu\nu}$ with $\chi h_{\mu\nu} \ll 1$. It is easy to see that, up to second-order terms, the action is precisely the one given by Eq. (3.116). In particular, we studied the low-velocity (non-relativistic) limit in order to show that the field $h_{\mu\nu}$ describes a gravitational special-relativistic field and how in the non-relativistic limit that action can be interpreted as the non-relativistic action of a particle with potential energy $Mc^2\chi h_{00}/2$ proportional to its inertial mass. This potential energy can be interpreted as a gravitational potential energy, identifying in this way inertial and gravitational masses and χh_{00} with $2\phi/c^2$, where ϕ is the Newtonian gravitational potential.

Thus, a GCT that, applied to an inertial frame, generates a non-trivial h_{00} can be seen as generating a gravitational field. We are identifying the so-called inertial forces with a gravitational field and we are saying that we cannot distinguish between them. Furthermore, all the effects of the gravitational field can be eliminated by going to an inertial frame. This is the essence of the PEGI which we will refine later. One can distinguish among *weak (or Galilean), medium-strong (or Einstein's), and strong* forms of the PEGI [242].

The weak form applies to the dynamics of one particle (precisely our case): one cannot distinguish whether we are describing its motion in a non-inertial frame or whether there is a gravitational field present. This implies that the inertial and gravitational masses of any particle are always proportional, with a universal proportionality constant that, in carefully chosen units, can be made 1. We have seen that, in the action Eq. (3.116), the inertial and

[47] Two useful references on σ-models are [210, 576].

gravitational masses of the particle are identical. We can certainly say that the PGR implies the weak form of the PEGI.

The medium-strong form extends the rank of applicability from the dynamics of one particle to all non-gravitational laws of physics. The introduction of the curved metric $g_{\mu\nu}$ into the actions of all known interactions guarantees that it is also a consequence of the PGR.

The strong form applies to all laws of physics, including gravity itself. There is nothing we can say about this form of the PEGI for the moment, although we already mentioned in the previous section that GR satisfies it, but let us mention that it is not a direct consequence of general covariance, for we can write SRFTs in Minkowski spacetime in general-covariant form.

So far we have considered only $g_{\mu\nu}$s that can be generated by GCTs from $\eta_{\mu\nu}$. Our experience tells us that there are non-trivial gravitational fields in what we would previously have called inertial frames. These gravitational fields must be described by the metric, too. To incorporate them into the theory, we are forced to allow for all kinds of metrics $g_{\mu\nu}$ that cannot be transformed into $\eta_{\mu\nu}$ by a GCT. However, for any arbitrary spacetime metric *at a given point*, there will always be coordinate systems defining *local inertial frames* in which $g_{\mu\nu}$ is equal to $\eta_{\mu\nu}$ at that given spacetime point P and in which the first derivatives of $g_{\mu\nu}$ vanish at that given point[48] and so all the components of the Levi-Cività connection $\Gamma_{\mu\nu}{}^{\rho}(g)$ also vanish at P. One such system is provided by the *Riemann normal coordinates at the point* P (see, for instance, [707]), which have the following properties:

$$\begin{aligned} g_{\mu\nu}(\mathrm{P}) &= \eta_{\mu\nu}, & \partial_\rho g_{\mu\nu}(\mathrm{P}) &= 0, \\ \partial_\rho\partial_\sigma g_{\mu\nu}(\mathrm{P}) &= \tfrac{2}{3} R_{\mu(\rho\sigma)\nu}(\mathrm{P}), & R_{\mu\nu\rho\sigma}(\mathrm{P}) &= 2\partial_\mu\partial_{[\rho} g_{\sigma]\nu}(\mathrm{P}). \end{aligned} \tag{3.268}$$

In this coordinate system, although the first derivatives vanish, the second derivatives do not. In fact, in general, there is no coordinate system in which both first and second derivatives at P vanish, because, otherwise, the Riemann tensor would vanish also at P, which is possible only if it vanishes at P in any coordinate system. This reflects the fact that, although the gravitational field is encoded in the metric tensor, it is actually characterized by the Riemann curvature tensor. The two tensors play a role similar in this respect to those of the vector potential and the field strength in Maxwell electrodynamics. Then, if there is a non-trivial gravitational field at P, the curvature tensor will not vanish at that point and the same will be true in any coordinates, including Riemann normal coordinates. Thus, to what extent is it true that all gravitational effects can be eliminated in the neighborhood of a point as the PEGI states? The point is that *observable gravitational effects* depend on the product of Riemann tensor components and spacetime coordinate intervals that can be made arbitrarily small and the upshot of this discussion is that the equivalence between gravitation and inertia will work only *locally* and for observable effects. The PEGI is only

[48] Any real non-singular metric can be diagonalized at a given point using the appropriate coordinate system, the non-vanishing components being $+1$s and -1s. The number of -1s minus the number of $+1$s cannot be changed by a further coordinate transformation and is an intrinsic property of the metric, an invariant called the *signature*. Continuity of the metric implies that the signature is the same at all points of spacetime. We consider only metrics of signature $d-2$, the signature of $\eta_{\mu\nu}$ in our conventions.

local and we can say that observable effects of the gravitational field can be eliminated locally in a small enough neighborhood of a given point. A longer discussion with examples can be found in [242].

There is an ongoing debate on the validity and interpretation of the PEGI into which we will not enter. Some interesting criticisms can be found in [659].

So far we have seen that the PGR forces us to use general spacetime metrics $g_{\mu\nu}$ and that these encode gravitational and inertial forces on the same footing, implying the PEGI in its medium-strong form. Any theory making use of a metric in this way would do the same. Now we want to find an equation of motion for the metric field which determines the dynamics of the gravitational field.

The PGR tells us that the equation of motion of the metric field must be a general tensor equation, $A^{\alpha\beta} = T^{\alpha\beta}_{\text{matter}}$. We have to find a suitable two-index, symmetric, tensor $A^{\alpha\beta} = A^{(\alpha\beta)}$ that is a function only of the metric and its first and second derivatives, $A^{\alpha\beta} = A^{\alpha\beta}(g_{\mu\nu}, \partial_\rho g_{\mu\nu}, \partial_\sigma\partial_\rho g_{\mu\nu})$. Now comes a very important point: in special relativity the matter energy–momentum tensor is always conserved: $\partial_\mu T^{\mu\nu}_{\text{matter}} = 0$. Now we require that the covariant generalization (as required by the PGR) of this equation

$$\nabla_\alpha T^{\alpha\beta}_{\text{matter}} = 0 \tag{3.269}$$

also holds. The connection is the Levi-Cività connection. It has to be stressed that this equation is no longer a conservation equation, as we will explain in detail in Chapter 6. However, it is the covariant generalization of the special-relativistic continuity equation and reduces to it in locally inertial frames and it seems a plausible requirement. Thus, we have to ask that $A^{\alpha\beta}$ be covariantly divergence-free.

The problem of finding the most general tensor $A^{\alpha\beta}$ satisfying these conditions was solved by Lovelock in [662] and the solution is

$$A^{\alpha}{}_{\beta} = \sum_{p=1}^{p=[\frac{d+1}{2}]} c_p g^{\alpha\gamma_1\cdots\gamma_{2p}}{}_{\beta\delta_1\cdots\delta_{2p}} R_{\gamma_1\gamma_2}{}^{\delta_1\delta_2}\cdots R_{\gamma_{2p-1}\gamma_{2p}}{}^{\delta_{2p-1}\delta_{2p}} + c_0 g^{\alpha}{}_{\beta}, \tag{3.270}$$

where the cs are arbitrary constants and the Riemann tensor is the one associated with the Levi-Cività connection. If we also want to recover the Fierz–Pauli equation in the linear limit $g_{\mu\nu} = \eta_{\mu\nu} + \chi h_{\mu\nu}$, $A^{\alpha\beta}$ has to be linear in second derivatives of the metric. In that case, the only possibility is, as originally proven in [215, 924, 952],

$$A^{\alpha\beta} = aG^{\alpha\beta} + bg^{\alpha\beta}, \tag{3.271}$$

where $G^{\alpha\beta}$ is the Einstein tensor. This is also the only possibility in $d = 4$ even if we do not impose the requirement of linearity in second derivatives of the metric. The vanishing of its covariant divergence is due to the contracted Bianchi identity $\nabla_\mu G^{\mu\nu} = 0$ when the connection is the Levi-Cività connection as we have assumed and to the metric-compatibility of the same connection.

In the Fierz–Pauli theory there is no room for the constant b. Thus, let us set it to zero for the moment. Now we have only to fix the proportionality constant a, which can be inferred from the linearized (Fierz–Pauli) theory. We obtain the *Einstein equation*

$$G_{\mu\nu} = \frac{8\pi G_N^{(d)}}{c^4} T_{\text{matter}\,\mu\nu}. \tag{3.272}$$

As we will see in detail in Chapter 4, this equation can be derived from the following action principle (up to boundary terms that we will find then):

$$\boxed{S[g_{\mu\nu}, \varphi] = \frac{c^3}{16\pi G_{\rm N}^{(d)}} \int d^dx \sqrt{|g|}\, R(g) + S_{\rm matter}[g, \varphi],} \tag{3.273}$$

where $R(g)$ is the Ricci scalar for the Levi-Cività connection[49] and the matter energy–momentum tensor is defined by

$$\boxed{T_{\rm matter}^{\mu\nu} = \frac{2c}{\sqrt{|g|}} \frac{\delta S_{\rm matter}}{\delta g_{\mu\nu}},} \tag{3.274}$$

justifying Rosenfeld's definition of the energy–momentum tensor.

We may wonder whether the contracted Bianchi identity that supports the above equations of motion[50] is associated with some sort of gauge symmetry. Indeed, the group of GCTs can be understood as an infinite-dimensional continuous group of local transformations and from the invariance of the action under this group we will derive a gauge identity (the contracted Bianchi identity) and conserved currents in Chapter 6.

In this quick review we have seen how to use the PGR to construct the theory of GR. We have introduced the minimal number of elements necessary for a general-covariant theory, but there are additional objects that one can introduce. One of them is torsion. We will see that it can be introduced consistently in the presence of fermions without adding further degrees of freedom to the theory. Another object compatible with general covariance that we can add to the theory is a *cosmological constant*, which is basically the constant b that we discarded on the basis of its absence from the Fierz–Pauli theory. It occurs as the constant Λ in the action[51]

$$\boxed{S[g_{\mu\nu}, \varphi] = \frac{c^3}{16\pi G_{\rm N}^{(d)}} \int d^dx \sqrt{|g|}\, [R(g) - (d-2)\Lambda] + S_{\rm matter}[\varphi],} \tag{3.275}$$

leading to the *cosmological Einstein equation*

$$G_{\mu\nu} + \frac{d-2}{2}\Lambda g_{\mu\nu} = \frac{8\pi G_{\rm N}^{(d)}}{c^4}\, T_{\rm matter\,\mu\nu}. \tag{3.276}$$

---

[49] Thus, this is the second-order Einstein–Hilbert action that one obtains from the first-order action (3.253) by eliminating $\Gamma_{\mu\nu}{}^{\rho}$ through its equation of motion. This action is quadratic in first-order derivatives of the metric but contains second-order derivatives, which, however, appear in total derivatives.

[50] Einstein himself proposed first $R_{\mu\nu} = (8\pi G_{\rm N}^{(d)}/c^4) T_{\rm matter\,\mu\nu}$ until he realized the inconsistency of this equation with the covariant "conservation" of the energy–momentum tensor.

[51] The dimension-dependent factor has been chosen in order to have the equation $R_{\mu\nu} = \Lambda g_{\mu\nu}$ in vacuum.

This constant can be understood in various ways: first of all, one may think of some kind of matter distributed in spacetime in such a way that its energy–momentum tensor is precisely $T_{\mu\nu} = -[(d-2)/16\pi G_N^{(d)}]\Lambda g_{\mu\nu}$. It is commonly accepted that the vacuum energy of the quantum fields gives Λ. The value of Λ obtained according to this prescription is many orders of magnitude bigger than the experimental upper bound. This huge disagreement is known as the "cosmological-constant problem" (see e.g. [940]).

One can also understand Λ as a fundamental constant of Nature. Then, the question of its smallness (if it is not zero) need not be such a big problem, at least not bigger than the question of why the values of the other fundamental constants of Nature are what they are, some of them being really small (such as the Planck length).

The main effect of the cosmological constant is to change the vacuum of the theory, which in this context we can define as the maximally symmetric solution of the classical equations of motion with all matter fields set to zero. In the presence of a cosmological constant, Minkowski spacetime is no longer a vacuum solution and the new maximally symmetric solutions are *de Sitter* (dS_d) spacetime for positive Λ and *anti-de Sitter* (AdS_d) spacetime for negative Λ. Now, in the weak-field limit, we should be considering perturbations around the new vacuum $\bar{g}_{\mu\nu}$ as follows: $g_{\mu\nu} = \bar{g}_{\mu\nu} + \chi h_{\mu\nu}$. The theory that one obtains by linearizing the cosmological Einstein theory is not the Fierz–Pauli theory in Minkowski spacetime. This is why there was no room for the constant b in considering that limit. In the next section we are precisely going to study the linearized theory one obtains by expanding the cosmological Einstein equation around a general vacuum metric $\bar{g}_{\mu\nu}$ that can be curved or can even be the Minkowski metric in arbitrary coordinates.

3.4 The Fierz–Pauli theory in a curved background

In the previous sections we have constructed a theory of spin-2 particles moving in the background of Minkowski spacetime in Cartesian coordinates (constant, diagonal $\eta_{\mu\nu}$). In this section we want to try to extend this construction to other backgrounds. As we have seen, the Fierz–Pauli theory can also be considered as the lowest-order perturbation theory of GR over Minkowski spacetime. Here we will construct extensions of the Fierz–Pauli theory by constructing the lowest-order perturbation theory of GR over a given background spacetime metric that is a vacuum solution of the full GR theory.

We may wonder whether it is possible to write the Fierz–Pauli theory (or a generalization thereof) in an arbitrary curved background metric. Such a construction would be necessary, for instance, in order to couple a spin-2 particle to GR in the same way as we couple scalars or vector fields. Such a theory would necessarily contain the same terms as the flat spacetime one but covariantized so that it has the right flat-spacetime limit but can also contain additional terms proportional to the curvature of the background metric that vanish in that limit. The guiding principle determining whether to introduce these terms is gauge invariance: the theory should be invariant under the general-covariantized gauge transformations $\delta^{(0)} h_{\mu\nu} = -2\bar{\nabla}_{(\mu}\epsilon_{\nu)}$. However, it can be shown that it is not possible to write this gauge-invariant theory, no matter what curvature terms one introduces [47]. This is one of the indications of the problems one encounters in trying to couple spin-2 particles to (GR) gravity.

While a Fierz–Pauli theory in a general curved background does not exist, such a theory does exist in backgrounds that solve the vacuum (cosmological) Einstein equations and this is the theory we are going to obtain here. Its construction is useful for many purposes. We will use it in constructing conserved quantities in spacetimes with arbitrary asymptotics and we can use it to work with the Minkowskian Fierz–Pauli theory in arbitrary coordinates. However, apart from these prosaic applications it will also teach us interesting things, e.g. how to define masslessness in curved backgrounds.

To be as general as possible we will include a cosmological-constant term from the beginning as in Eq. (3.275).

3.4.1 Linearized gravity

Let us first describe the setup: we consider a spacetime metric $g_{\mu\nu}$ that solves the d-dimensional cosmological Einstein equations for some matter energy–momentum tensor $T_{\text{matter}}^{\mu\nu}$ (here we set $c=1$, as usual),

$$G_c{}^{\mu\nu} = 8\pi G_{\mathrm{N}}^{(d)} T_{\text{matter}}^{\mu\nu}, \tag{3.277}$$

where $G_c{}^{\mu\nu}$ is the *cosmological Einstein tensor*,

$$G_c{}^{\mu\nu} \equiv G^{\mu\nu} + \frac{d-2}{2}\Lambda g^{\mu\nu}. \tag{3.278}$$

The metric $g_{\mu\nu}$ must be such that we can consider it as produced by a small perturbation of the background metric $\bar{g}_{\mu\nu}$, i.e. we can write

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu}, \tag{3.279}$$

where the perturbation $h_{\mu\nu}$ goes to zero at infinity fast enough that the metric $g_{\mu\nu}$ is asymptotically $\bar{g}_{\mu\nu}$. Furthermore, $h_{\mu\nu}$ and its derivatives are assumed to be small enough that we can ignore higher-order terms.[52]

Usually, the background metric $\bar{g}_{\mu\nu}$ will be the vacuum metric, i.e. a maximally symmetric solution of the vacuum Einstein equations

$$\bar{G}_c{}^{\mu\nu} = 0. \tag{3.280}$$

Therefore, the metrics $g_{\mu\nu}$ that we consider describe in the gravitational language isolated systems. There are no matter sources of the gravitational field at infinity. In the absence of a cosmological constant, the vacuum metric $\bar{g}_{\mu\nu} = \eta_{\mu\nu}$, the Minkowski metric, and the metrics $g_{\mu\nu}$ will be *asymptotically flat*. With positive (negative) cosmological constant, the (maximally symmetric) vacuum solution is the *(anti-)de Sitter* ((A)dS$_d$) spacetime and the metrics $g_{\mu\nu}$ will be *asymptotically (anti-)de Sitter*. However, we will keep the background metric completely general in order to cover other interesting cases in which a solution $g_{\mu\nu}$ goes asymptotically to a $\bar{g}_{\mu\nu}$ that is not the vacuum solution or even a solution of the vacuum Einstein equations. Thus, we will use only Eqs. (3.277) and (3.279) to find the equation satisfied by the perturbation $h_{\mu\nu}$. Later on, we will impose the condition that the background metric solves the Einstein equation (3.280).

---

[52] Here we have absorbed the coupling constant $\chi = \sqrt{16\pi G_{\mathrm{N}}^{(d)}}$ into $h_{\mu\nu}$.

The first thing we have to do is to expand this equation in powers of the perturbation $h_{\mu\nu}$. The perturbation can be treated as a tensor on the background manifold. Then, it is natural to lower and raise its indices (and those of all tensors) with the background metric $\bar{g}_{\mu\nu}$ and its inverse $\bar{g}^{\mu\nu}$. In particular, $h^{\mu\nu} = \bar{g}^{\mu\rho}\bar{g}^{\nu\sigma}h_{\rho\sigma}$ is not the inverse of $h_{\mu\nu}$ (which need not exist) and we also define $h = \bar{g}^{\mu\nu}h_{\mu\nu}$. All barred covariant derivatives are also taken with respect to the background metric's Levi-Cività connection $\bar{\Gamma}_{\mu\nu}{}^{\rho}$. We find

$$\begin{aligned} g^{\mu\nu} &= \bar{g}^{\mu\nu} - h^{\mu\nu} + \mathcal{O}(h^2), \\ \Gamma_{\mu\nu}{}^{\rho} &= \bar{\Gamma}_{\mu\nu}{}^{\rho} + \gamma_{\mu\nu}{}^{\rho} + \mathcal{O}(h^2), \\ R_{\mu\nu\rho}{}^{\sigma} &= \bar{R}_{\mu\nu\rho}{}^{\sigma} + 2\bar{\nabla}_{[\mu}\gamma_{\nu]\rho}{}^{\sigma} + \mathcal{O}(h^2), \end{aligned} \tag{3.281}$$

with

$$\gamma_{\mu\nu}{}^{\rho} = \tfrac{1}{2}\bar{g}^{\rho\sigma}\left\{\bar{\nabla}_{\mu}h_{\sigma\nu} + \bar{\nabla}_{\nu}h_{\mu\sigma} - \bar{\nabla}_{\sigma}h_{\mu\nu}\right\}. \tag{3.282}$$

This equation is essentially the equation that gives the variation of the Levi-Cività connection $\delta\Gamma_{\mu\nu}{}^{\rho}$ ($\equiv \gamma_{\mu\nu}{}^{\rho}$) under an arbitrary variation of the metric[53] $\delta g_{\mu\nu}$ ($\equiv h_{\mu\nu}$),

$$\delta\Gamma_{\mu\nu}{}^{\rho} = \tfrac{1}{2}g^{\rho\sigma}\left\{\nabla_{\mu}\delta g_{\sigma\nu} + \nabla_{\nu}\delta g_{\sigma\rho} - \nabla_{\sigma}\delta g_{\mu\nu}\right\}. \tag{3.284}$$

Now we can find the expansion of $R_{\mu\nu\rho}{}^{\sigma}$ to first order in $h_{\mu\nu}$ using the so-called *Palatini identity* that gives the variation of the curvature tensor under an arbitrary variation of the connection

$$\delta R_{\mu\nu\rho}{}^{\sigma} = +2\nabla_{[\mu}\delta\Gamma_{\nu]\rho}{}^{\sigma}. \tag{3.285}$$

The Palatini identity follows from Eqs. (1.31) and (1.36), on setting the torsion equal to zero, identifying $\tau_{\mu\nu}{}^{\rho}$ with $\delta\Gamma_{\mu\nu}{}^{\rho}$, and keeping only the linear terms. We stress that, unlike $\Gamma_{\mu\nu}{}^{\rho}$, the variation $\delta\Gamma_{\mu\nu}{}^{\rho}$ is a true tensor and its covariant derivative is well defined.[54]
For the variation of $\Gamma_{\mu\nu}{}^{\rho}$ that we have just found we obtain

$$R_{\mu\nu\rho}{}^{\sigma} = \bar{R}_{\mu\nu\rho}{}^{\sigma} + \bar{g}^{\sigma\lambda}\left\{\bar{\nabla}_{[\mu}\bar{\nabla}_{\nu]}h_{\lambda\rho} + \bar{\nabla}_{[\mu|}\bar{\nabla}_{\rho}h_{|\nu]\lambda} - \bar{\nabla}_{[\mu|}\bar{\nabla}_{\lambda}h_{|\nu]\rho}\right\} + \mathcal{O}(h^2), \tag{3.287}$$

and, on contracting the indices σ and ν, we find[55]

$$R_{\mu\rho} = \bar{R}_{\mu\rho} + \tfrac{1}{2}\left\{\bar{\nabla}^2 h_{\mu\rho} - 2\bar{\nabla}^{\lambda}\bar{\nabla}_{(\mu}h_{\rho\lambda} + \bar{\nabla}_{\mu}\bar{\nabla}_{\rho}h\right\} + \mathcal{O}(h^2). \tag{3.288}$$

---

[53] For further use we quote here the generalization of this equation when there is torsion present:

$$\begin{aligned} \delta\Gamma_{\alpha\beta}{}^{\gamma} = {} & \tfrac{1}{2}g^{\gamma\delta}\left\{\nabla_{\alpha}\delta g_{\beta\delta} + \nabla_{\beta}\delta g_{\alpha\delta} - \nabla_{\delta}\delta g_{\alpha\beta}\right\} \\ & + \tfrac{1}{2}\left\{g^{\delta\gamma}g_{\sigma\beta}\delta T_{\alpha\delta}{}^{\sigma} + g^{\delta\gamma}g_{\sigma\alpha}\delta T_{\beta\delta}{}^{\sigma} - \delta T_{\alpha\beta}{}^{\gamma}\right\}. \end{aligned} \tag{3.283}$$

[54] Also for further use, here we quote the formula valid for a general connection:

$$\delta R_{\mu\rho} = \nabla_{\mu}\delta\Gamma_{\nu\rho}{}^{\nu} - \nabla_{\nu}\delta\Gamma_{\mu\rho}{}^{\nu} - T_{\mu\nu}{}^{\lambda}\delta\Gamma_{\lambda\rho}{}^{\nu}. \tag{3.286}$$

[55] Sometimes the subindex L is used to indicate that the object is the part linear in $h_{\mu\nu}$ of the corresponding tensor with the indices in the same position. Observe that for any tensor $T_{\mathrm{L}}{}^{\mu} \neq \bar{g}^{\mu\nu}T_{\mathrm{L}\,\nu}$ and for this reason we try to avoid this notation.

On contracting with $g^{\mu\rho} = \bar{g}^{\mu\rho} - h^{\mu\rho}$, we find that the Ricci scalar is given by

$$R = \bar{R} - \bar{R}_{\lambda\sigma} h^{\lambda\sigma} + \bar{\nabla}^2 h - \bar{\nabla}_\lambda \bar{\nabla}_\sigma h^{\lambda\sigma} + \mathcal{O}(h^2). \tag{3.289}$$

Now, to find the cosmological Einstein tensor we use

$$G_{\mathrm{c}}{}^{\alpha\beta} = \left(g^{\alpha\mu} g^{\beta\rho} - \tfrac{1}{2} g^{\alpha\beta} g^{\mu\rho}\right) R_{\mu\rho} + \frac{d-2}{2} \Lambda g^{\alpha\beta}, \tag{3.290}$$

obtaining

$$\begin{aligned}
G_{\mathrm{c}}{}^{\alpha\beta} &= \bar{G}_c{}^{\alpha\beta} + G_{\mathrm{cL}}{}^{\alpha\beta} + \mathcal{O}(h^2), \\
G_{\mathrm{cL}}{}^{\alpha\beta} &= G_{\mathrm{cL1}}{}^{\alpha\beta} + G_{\mathrm{cL2}}{}^{\alpha\beta}, \\
G_{\mathrm{cL1}}{}^{\alpha\beta} &= \tfrac{1}{2}\left\{\bar{\nabla}^2 h^{\alpha\beta} - 2\bar{\nabla}^\lambda \bar{\nabla}^{(\alpha} h_\lambda{}^{\beta)} + \bar{\nabla}^\alpha \bar{\nabla}^\beta h\right\} - \tfrac{1}{2}\bar{g}^{\alpha\beta}\left\{\bar{\nabla}^2 h - \bar{\nabla}^\mu \bar{\nabla}^\nu h_{\mu\nu}\right\}, \\
G_{\mathrm{cL2}}{}^{\alpha\beta} &= -\left\{h^{\alpha\mu}\bar{g}^{\beta\rho} + \bar{g}^{\alpha\mu} h^{\beta\rho} - \tfrac{1}{2} h^{\alpha\beta}\bar{g}^{\mu\rho} - \tfrac{1}{2}\bar{g}^{\alpha\beta} h^{\mu\rho}\right\}\bar{R}_{\mu\rho} - \frac{d-2}{2}\Lambda h^{\alpha\beta}.
\end{aligned} \tag{3.291}$$

On substituting into the cosmological Einstein equation (3.277), we find

$$\bar{G}_{\mathrm{c}}{}^{\mu\nu} + G_{\mathrm{cL}}{}^{\mu\nu} = 8\pi G_{\mathrm{N}}^{(d)}\left(T_{\mathrm{matter}}^{\mu\nu} + t^{\mu\nu}\right), \tag{3.292}$$

where the l.h.s. contains terms up to first order in $h_{\mu\nu}$ and $8\pi G_{\mathrm{N}}^{(d)} t^{\mu\nu}$ stands for all the second- and higher-order terms in $h_{\mu\nu}$ and is referred to as the *gravitational energy–momentum (pseudo-)tensor*. This is the definition we will use in Section 6.1.2, and it is clearly justified by our previous results.

Now we can particularize to the case in which the background metric satisfies the vacuum cosmological Einstein equation (3.280), which, upon subtraction of the trace, implies

$$\bar{R}_{\mu\nu} = \Lambda \bar{g}_{\mu\nu}. \tag{3.293}$$

We find the same expressions as before for $R_{\mu\rho}$ and $G_{\mathrm{cL1}}{}^{\alpha\beta}$ but the expression for $G_{\mathrm{cL2}}{}^{\alpha\beta}$ is considerably simpler,

$$G_{\mathrm{cL2}}{}^{\alpha\beta} = -\Lambda\left(h^{\alpha\beta} - \tfrac{1}{2}\bar{g}^{\alpha\beta} h\right), \tag{3.294}$$

and the l.h.s. of the cosmological Einstein equation is purely linear in $h_{\mu\nu}$,

$$G_{\mathrm{cL}}{}^{\mu\nu} = 8\pi G_{\mathrm{N}}^{(d)}\left(T_{\mathrm{matter}}^{\mu\nu} + t^{\mu\nu}\right). \tag{3.295}$$

This l.h.s. gives us the generalization of the Fierz–Pauli equations wave operator in curved spacetime we were looking for:

$$\begin{aligned}
\bar{\mathcal{D}}^{\alpha\beta}(h) = 2G_{\mathrm{cL}}{}^{\alpha\beta} &= \bar{\nabla}^2 h^{\alpha\beta} - 2\bar{\nabla}^\lambda \bar{\nabla}^{(\alpha} h_\lambda{}^{\beta)} + \bar{\nabla}^\alpha \bar{\nabla}^\beta h \\
&\quad - \bar{g}^{\alpha\beta}\left\{\bar{\nabla}^2 h - \bar{\nabla}^\mu \bar{\nabla}^\nu h_{\mu\nu}\right\} - 2\Lambda\left(h^{\alpha\beta} - \tfrac{1}{2}\bar{g}^{\alpha\beta} h\right) \\
&= 16\pi G_{\mathrm{N}}^{(d)}\left(T_{\mathrm{matter}}^{\alpha\beta} + t^{\alpha\beta}\right),
\end{aligned} \tag{3.296}$$

which justifies the present definition of $t^{\mu\nu}$ which coincides with the one we have used before. In fact, in the previous sections we have found the lowest-order term (quadratic in h) of $t^{\mu\nu}$ in the case $\bar{\mu\nu} = \eta_{\mu\nu}$ ($\Lambda = 0$) which we denoted by $t_{\rm GR}^{(0)\,\mu\nu}$.

This equation, with the r.h.s. set to zero, is the equation of motion of a massless spin-2 field moving on a background spacetime $\bar{g}_{\mu\nu}$, which we are going to study in the next section. We can already see that this equation does not look like the typical wave equation for a massless field because it has mass-like terms proportional to the cosmological constant. However, we are going to argue that precisely those terms are necessary in order to describe massless fields in a spacetime with $\bar{R}_{\mu\nu} = \Lambda \bar{g}_{\mu\nu}$.

Observe that, since $\zeta^{\mu\nu} = h^{\mu\nu} + \mathcal{O}(h^2)$ and $h^{\mu\nu} = \zeta^{\mu\nu} + \mathcal{O}(\zeta^2)$, we could have arrived at the same linear-order results by expanding around the inverse metric

$$g^{\mu\nu} = \bar{g}^{\mu\nu} - \zeta^{\mu\nu}. \tag{3.297}$$

We would like to have an action from which to derive the above equation of motion with vanishing r.h.s. Instead of guessing, we simply expand the integrand of the Einstein–Hilbert action to second order in $h_{\mu\nu}$. Using the matrix identity

$$\sqrt{|M|} = \exp(\tfrac{1}{2}\mathrm{tr}\ln M), \tag{3.298}$$

and the expansions

$$\begin{aligned}
(1+x)^{-1} &= 1 - x + x^2 - x^3 + \cdots,\\
\ln(1+x) &= x - \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 - \tfrac{1}{4}x^4 + \cdots,\\
\exp y &= 1 + y + \frac{1}{2!}y^2 + \frac{1}{3!}y^3 + \cdots,
\end{aligned} \tag{3.299}$$

we can easily calculate second- and higher-order terms:

$$\begin{aligned}
g_{\mu\nu} &= \bar{g}_{\mu\nu} + h_{\mu\nu},\\
g^{\mu\nu} &= \bar{g}^{\mu\nu} - h^{\mu\nu} + h^{\mu}{}_{\sigma}h^{\sigma\nu} - h^{\mu}{}_{\sigma}h^{\sigma\rho}h_{\rho}{}^{\nu} + \mathcal{O}(h^4),\\
\sqrt{|g|} &= \sqrt{|\bar{g}|}\left(1 + \frac{1}{2}h + \frac{1}{8}h^2 - \frac{1}{4}h_{\mu\nu}h^{\mu\nu} + \frac{1}{6}h_{\mu}{}^{\nu}h_{\nu}{}^{\rho}h_{\rho}{}^{\mu}\right.\\
&\qquad \left. - \frac{1}{8}hh_{\mu\nu}h^{\mu\nu} + \frac{1}{48}h^3\right) + \mathcal{O}(h^4).
\end{aligned} \tag{3.300}$$

For the Levi-Cività connection we can write the exact expression

$$\Gamma_{\mu\nu}{}^{\rho} = \bar{\Gamma}_{\mu\nu}{}^{\rho} + g^{\rho\sigma}\gamma_{\mu\nu\sigma}, \qquad \gamma_{\mu\nu\sigma} = \tfrac{1}{2}\left\{\bar{\nabla}_{\mu}h_{\nu\sigma} + \bar{\nabla}_{\nu}h_{\mu\sigma} - \bar{\nabla}_{\sigma}h_{\mu\nu}\right\}, \tag{3.301}$$

and just have to substitute the above expansion of $g^{\rho\sigma}$ to the desired order. For the Riemann curvature tensor and the Ricci tensor we can write also write exact expressions,

$$\begin{aligned}
R_{\mu\nu\rho}{}^{\sigma} &= \bar{R}_{\mu\nu\rho}{}^{\sigma} + 2\bar{\nabla}_{[\mu}\big(g^{\sigma\lambda}\gamma_{\nu]\rho\lambda}\big) + 2g^{\sigma\delta}g^{\lambda\epsilon}\gamma_{[\mu|\lambda\delta}\gamma_{|\nu]\rho\epsilon},\\
R_{\mu\rho} &= \bar{R}_{\mu\rho} + \bar{\nabla}_{\mu}\big(g^{\sigma\lambda}\gamma_{\sigma\rho\lambda}\big) - \bar{\nabla}_{\sigma}\big(g^{\sigma\lambda}\gamma_{\mu\rho\lambda}\big) + g^{\sigma\delta}g^{\lambda\epsilon}\big(\gamma_{\mu\lambda\delta}\gamma_{\sigma\rho\epsilon} - \gamma_{\sigma\lambda\delta}\gamma_{\mu\rho\epsilon}\big),
\end{aligned} \tag{3.302}$$

on which, again, we simply have to expand the inverse metric. A similar expression can immediately be found for the Ricci scalar R and for the scalar density $\sqrt{|g|}\,R$. Then, up to total derivatives and $\mathcal{O}(h^3)$ terms, and using the equations of motion for the background $\bar{R}_{\mu\nu} = \Lambda \bar{g}_{\mu\nu}$ that we have not used so far, the Einstein–Hilbert action (3.275) becomes the Fierz–Pauli action in a curved background

$$\boxed{\begin{aligned} S = \frac{1}{\chi^2}\int d^dx\sqrt{|\bar{g}|}\,\Big\{&\tfrac{1}{4}\bar{\nabla}_\mu h_{\rho\lambda}\bar{\nabla}^\mu h^{\rho\lambda} - \tfrac{1}{2}\bar{\nabla}_\mu h_{\rho\lambda}\bar{\nabla}^\rho h^{\mu\lambda} + \tfrac{1}{2}\bar{\nabla}_\mu h^{\mu\nu}\bar{\nabla}_\nu h \\ &- \tfrac{1}{4}\bar{\nabla}_\mu h\bar{\nabla}^\mu h + \tfrac{1}{2}\Lambda\big(h^{\mu\nu}h_{\mu\nu} - \tfrac{1}{2}h^2\big)\Big\}\,. \end{aligned}} \tag{3.303}$$

In the Minkowski background $\bar{g}_{\mu\nu} = \eta_{\mu\nu}$ ($\Lambda = 0$) it is easier to find higher corrections both to the action and to the equations of motion. A long but straightforward calculation gives as cubic term in the action (up to total derivatives)

$$S^{(3)} = \frac{1}{\chi^2}\int d^dx\, \tfrac{1}{2}h^{\mu\sigma}\mathcal{L}^{(1)}_{\mu\sigma}, \tag{3.304}$$

where $\mathcal{L}^{(1)}_{\mu\sigma}$ is written in Eq. (3.199).[56] The equation of motion that one obtains from the variation of the vacuum Einstein–Hilbert action is

$$\frac{\delta S}{\delta g_{\mu\nu}} - \frac{1}{\chi^2}\sqrt{|g|}\,G^{\mu\nu} = 0. \tag{3.305}$$

Therefore, the linear equation of motion (the Fierz–Pauli equation) is (restoring everywhere χ) obtained from the quadratic term in the action and the quadratic energy–momentum tensor from the cubic term in in the action:

$$\begin{aligned} \frac{\delta S^{(2)}}{\delta h_{\mu\nu}} &= -G^{(1)\,\mu\nu} &&= -\tfrac{1}{2}\mathcal{D}^{\mu\nu}(h), \\ \frac{\delta S^{(3)}}{\delta h_{\mu\nu}} &= \chi\big(G^{(2)\,\mu\nu} + \tfrac{1}{2}hG^{(1)\,\mu\nu}\big) &&= \tfrac{1}{2}\chi\, t_{\mathrm{GR}}^{(0)\,\mu\nu}(h). \end{aligned} \tag{3.306}$$

This is the $t_{\mathrm{GR}}^{(0)\,\mu\nu}(h)$ given in Eq. (3.200). The physical consistency of these results has been discussed at length before.

3.4.2 Massless spin-2 particles in curved backgrounds

We have obtained a generalization of the Fierz–Pauli action for curved backgrounds that are solutions of the vacuum Einstein equations and we want to see whether the theory can describe massless spin-2 particles in those backgrounds.

We should start by saying that the concepts of mass and angular momentum (spin) are in principle associated exclusively with the Poincaré group, which is the isometry group of Minkowski spacetime. In more general spaces one has to study the representations of the

[56] To recover the factors of χ we have to rescale $h_{\mu\nu} \to \chi h_{\mu\nu}$.

isometry group and, in general, there will be no obvious generalizations of these concepts that work in all cases.

Instead of proceeding case by case trying to give definitions of the mass of a field, we are going to adopt a general point of view and give a characterization of the masslessness of a field. The main observation is that massless fields have, as a rule, fewer degrees of freedom (DOF) than do massive fields, the extra DOF being removed by gauge symmetries that appear when the mass parameters are set to zero. At the beginning of this chapter we studied two cases in Minkowski spacetime: a massive vector field has $d-1$ DOF and no gauge symmetries. When we set the mass parameter to zero, the theory has a gauge symmetry and we can remove one more DOF (a total of two) so there are only the $d-2$ DOF of a massless vector. In the spin-2 case, in the presence of mass the field describes $(d-2)(d+1)/2$ DOF. When we switch off the mass parameter, there appears a gauge symmetry that allows us to remove $d-1$ DOF more (a total of $2d$) and we are left with the $d(d-3)/2$ DOF of a massless spin-2 particle.

In conclusion, we are going to characterize masslessness by the occurrence of new gauge symmetries that appear when we switch off the mass parameter.

We have obtained a generalization of the Fierz–Pauli theory to curved backgrounds given by the action Eq. (3.303) and equation of motion Eq. (3.296) (with vanishing r.h.s.). In this theory there are terms proportional to the cosmological constant Λ that have the form of mass terms. To see whether they really are mass terms according to our definition, we look for gauge symmetries. The obvious candidate is the linearization of the invariance under GCTs that generalizes Eq. (3.95) to curved backgrounds:

$$\delta_\epsilon h_{\mu\nu} = -2\bar{\nabla}_{(\mu}\epsilon_{\nu)}, \tag{3.307}$$

Let us first check the invariance of the action under these transformations. First we vary the action as usual. We obtain two types of terms: $\bar{\nabla}h\bar{\nabla}^2\epsilon$ and $\Lambda h\bar{\nabla}\epsilon$ (these arise from the variation of the "mass terms"). We want to move all the derivatives so they act over ϵ. Thus, we integrate by parts all the terms of the first kind, obtaining $h\bar{\nabla}^3\epsilon$-type terms and a total derivative. These terms can be combined into terms of the forms $h\bar{\nabla}[\bar{\nabla},\bar{\nabla}]\epsilon$ and $h[\bar{\nabla},\bar{\nabla}]\bar{\nabla}\epsilon$. Then, the commutators of covariant derivatives can be replaced by curvature terms using the Ricci identity and all these terms become terms of the type $h\bar{R}\bar{\nabla}\epsilon$ and $h\bar{\nabla}\bar{R}\epsilon$. The first cancel out, *upon use of the vacuum cosmological Einstein equation for the background metric* $\bar{R}_{\mu\nu} = \Lambda\bar{g}_{\mu\nu}$, the $\Lambda h\bar{\nabla}\epsilon$ terms. The second cancel out upon use of the background Bianchi identity $\bar{\nabla}_{[\mu}\bar{R}_{\nu\rho]\sigma}{}^{\lambda} = 0$ and we are left with the total derivative:

$$\begin{aligned}
\delta_\epsilon S = \frac{1}{\chi^2}\int d^dx\sqrt{|\bar{g}|}\,\bar{\nabla}_\mu\big\{\tfrac{1}{2}h_{\rho\sigma}\big[&4\bar{\nabla}^{[\mu}\bar{\nabla}^{\rho]}\epsilon^\sigma - 2\bar{\nabla}^\rho\bar{\nabla}^\sigma\epsilon^\mu \\
&+ \bar{g}^{\rho\sigma}\big(\bar{\nabla}^2\epsilon^\mu - \bar{\nabla}_\lambda\bar{\nabla}^\mu\epsilon^\lambda\big) + 2\bar{g}^{\mu\rho}\bar{\nabla}^\sigma\bar{\nabla}_\lambda\epsilon^\lambda - 2\bar{g}^{\rho\sigma}\bar{\nabla}^\mu\bar{\nabla}_\lambda\epsilon^\lambda\big]\big\} \\
\equiv \int d^dx\sqrt{|\bar{g}|}\,\bar{\nabla}_\mu s^\mu(\epsilon).&
\end{aligned} \tag{3.308}$$

The Fierz–Pauli equation of motion Eq. (3.296) is, therefore, invariant for the backgrounds considered. The proof makes crucial use of the Einstein equation satisfied by the background metric. As we remarked in the introduction to this section, in general backgrounds there is no way to construct a gauge-invariant theory by adding curvature terms [47].

Furthermore, in the proof of invariance of the action the presence of the cosmological constant terms is also crucial. Had we tried to prove the invariance of the equation of motion directly, we would have seen the necessity for these terms to cancel out curvature terms coming from the commutators of covariant derivatives. We can conclude that the theory, with those terms, is massless.

It is interesting to see what kind of gauge identity and conserved current we obtain from this invariance. We proceed as usual. We first find the variation of the action under an arbitrary infinitesimal transformation of $\delta h_{\mu\nu}$:

$$\begin{aligned}
\delta S_{\mathrm{FP}} &= \frac{1}{\chi^2}\int d^dx\sqrt{|\bar{g}|}\left\{\frac{S_{\mathrm{FP}}}{\delta h_{\alpha\beta}}\delta h_{\alpha\beta} + \bar{\nabla}_\mu\left(l^{\mu(\alpha\beta)}\delta h_{\alpha\beta}\right)\right\},\\
\frac{S_{\mathrm{FP}}}{\delta h_{\alpha\beta}} &= -\tfrac{1}{2}\bar{\mathcal{D}}^{\alpha\beta}(h),\\
l^{\mu\alpha\beta} &= \tfrac{1}{2}\bar{\nabla}^\mu h^{\alpha\beta} - \bar{\nabla}^\alpha h^{\beta\mu} + \tfrac{1}{2}\bar{g}^{\mu\alpha}\bar{\nabla}^\beta h + \tfrac{1}{2}\bar{g}^{\alpha\beta}\bar{\nabla}_\nu h^{\mu\nu} - \tfrac{1}{2}\bar{g}^{\alpha\beta}\bar{\nabla}^\mu h.
\end{aligned} \tag{3.309}$$

Using now the particular form of the gauge transformation $\delta_\epsilon h_{\mu\nu}$ in the above equation and integrating by parts, we obtain

$$\delta_\epsilon S = \int d^dx\sqrt{|\bar{g}|}\left\{-\frac{1}{\chi^2}\epsilon_\beta\bar{\nabla}_\alpha\bar{\mathcal{D}}^{\alpha\beta}(h) + \bar{\nabla}_\mu\left[\frac{1}{\chi^2}\bar{\mathcal{D}}^{\mu\beta}\epsilon_\beta - \frac{2}{\chi^2}l^{\mu(\alpha\beta)}\bar{\nabla}_\alpha\epsilon_\beta\right]\right\}, \tag{3.310}$$

and, on comparing this with the first form of the variation of the action that we found, we arrive finally at the identity, which is valid for arbitrary ϵ^μs and without the use of any equations of motion,

$$\begin{aligned}
0 &= \int d^dx\sqrt{|\bar{g}|}\left\{-\frac{1}{\chi^2}\epsilon_\beta\bar{\nabla}_\alpha\bar{\mathcal{D}}^{\alpha\beta}(h) + \bar{\nabla}_\mu j^\mu_{\mathrm{N2}}(\epsilon)\right\},\\
j^\mu_{\mathrm{N2}}(\epsilon) &= j^\mu_{\mathrm{N1}}(\epsilon) + \frac{1}{\chi^2}\bar{\mathcal{D}}^{\mu\beta}(h)\epsilon_\beta,\\
j^\mu_{\mathrm{N1}}(\epsilon) &= -\frac{2}{\chi^2}l^{\mu(\alpha\beta)}\bar{\nabla}_\alpha\epsilon_\beta - s^\mu(\epsilon).
\end{aligned} \tag{3.311}$$

From this identity we derive the gauge identity,

$$\bar{\nabla}_\alpha\bar{\mathcal{D}}^{\alpha\beta}(h) = 0, \tag{3.312}$$

and the off-shell covariant conservation of the above Noether current,

$$\bar{\nabla}_\mu j^\mu_{\mathrm{N2}}(\epsilon) = 0 \qquad \left(\partial_\mu \mathrm{j}^\mu_{\mathrm{N2}}(\epsilon) = 0\right). \tag{3.313}$$

We know that this Noether current can always be written as $\mathrm{j}^\mu_{\mathrm{N2}}(\epsilon) = \partial_\nu \mathrm{j}^{\nu\mu}_{\mathrm{N2}}(\epsilon)$ with $\mathrm{j}^{\nu\mu}_{\mathrm{N2}}(\epsilon) = -\mathrm{j}^{\mu\nu}_{\mathrm{N2}}(\epsilon)$. Finding this antisymmetric tensor in the general case is complicated and we are going to do it only for the most interesting case, in which ϵ^μ is a Killing vector of the background metric $\epsilon^\mu \equiv \bar{\xi}^\mu$ with $\bar{\nabla}_{(\mu}\bar{\xi}_{\nu)} = 0$. In this case, $s^\mu(\xi)$ has to vanish identically, because the variations of $h_{\mu\nu}$ also vanish identically, and the first term of $j^\mu_{\mathrm{N1}}(\epsilon)$ also vanishes because of the Killing equation. Then, only the second term in the expression for $j^{\mu}_{\rm N2}(\bar{\xi})$ Eq. (3.311) survives and we are left with

$$j^{\mu}_{\rm N2}(\bar{\xi}) = \frac{1}{\chi^2}\bar{\mathcal{D}}^{\mu\nu}(h)\bar{\xi}_\nu. \tag{3.314}$$

The conservation of this current is easy to check using the Bianchi identity and the Killing equation. To find $j^{\mu\nu}_{\rm N2}(\bar{\xi}) = (1/\sqrt{|g|})\mathfrak{j}^{\mu\nu}_{\rm N2}(\bar{\xi})$ we follow Abbott and Deser in [1]. First we separate $\bar{\mathcal{D}}^{\mu\nu}$ into two pieces:

$$\begin{aligned} \bar{\mathcal{D}}^{\mu\nu}(h) = \text{curvature terms} + [\bar{\nabla}^\nu, \bar{\nabla}_\lambda]h^{\lambda\mu} \quad (\equiv 2X^{\mu\nu}) \\ \text{the rest} - [\bar{\nabla}^\nu, \bar{\nabla}_\lambda]h^{\lambda\mu} \qquad (\equiv 2Y^{\mu\nu}). \end{aligned} \tag{3.315}$$

$Y^{\mu\nu}$ can be written in this form,

$$Y^{\mu\nu} = -\bar{\nabla}_\alpha\bar{\nabla}_\beta K^{\mu\alpha\nu\beta}, \tag{3.316}$$

where $K^{\mu\alpha\nu\beta}$ is as defined in Eq. (3.91) but with a general background metric instead of the Minkowski metric, i.e.

$$K^{\mu\alpha\nu\beta} = \tfrac{1}{2}\left\{\bar{g}^{\mu\beta}\bar{h}^{\nu\alpha} + \bar{g}^{\nu\alpha}\bar{h}^{\mu\beta} - \bar{g}^{\mu\nu}\bar{h}^{\alpha\beta} - \bar{g}^{\alpha\beta}\bar{h}^{\mu\nu}\right\}. \tag{3.317}$$

This tensor has the same symmetries as the Riemann tensor and is sometimes called the *superpotential*. Using $\bar{R}_{\mu\nu} = \Lambda\bar{g}_{\mu\nu}$, we find

$$X^{\mu\nu} = \tfrac{1}{2}\left[\bar{\nabla}^\nu, \bar{\nabla}_\lambda\right]\bar{h}^{\lambda\mu} - \Lambda\bar{h}^{\mu\nu}, \tag{3.318}$$

and, using the Ricci identity, it can be rewritten as follows:

$$X^{\mu\nu} = \tfrac{1}{2}\left[\bar{R}^{\mu}{}_{\lambda\sigma}{}^{\nu}\bar{h}^{\lambda\sigma} - \Lambda\bar{h}^{\mu\nu}\right]. \tag{3.319}$$

Finally, we can also rewrite it as follows:

$$X^{\mu\nu} = \tfrac{1}{2}\bar{R}^{\nu}{}_{\alpha\beta\gamma}K^{\mu\alpha\beta\gamma}. \tag{3.320}$$

Using the expression for Y in terms of the superpotential K,

$$Y^{\mu\nu}\bar{\xi}_\nu = -\bar{\nabla}_\alpha\left[\left(\bar{\nabla}_\beta K^{\mu\alpha\nu\beta}\right)\bar{\xi}_\nu - K^{\mu\beta\nu\alpha}\bar{\nabla}_\beta\bar{\xi}_\nu\right] - K^{\mu\beta\nu\alpha}\bar{\nabla}_\alpha\bar{\nabla}_\beta\bar{\xi}_\nu. \tag{3.321}$$

Using the Killing vector identity Eq. (1.108) for the background Killing vectors and the definition of the superpotential K, we see that

$$Y^{\mu\nu}\bar{\xi}_\nu = -\bar{\nabla}_\alpha\left[\left(\bar{\nabla}_\beta K^{\mu\alpha\nu\beta}\right)\bar{\xi}_\nu - K^{\mu\beta\nu\alpha}\bar{\nabla}_\beta\bar{\xi}_\nu\right] - X^{\mu\nu}\bar{\xi}_\nu, \tag{3.322}$$

and, therefore,

$$j^{\alpha\mu}_{\rm N2}(\bar{\xi}) = -\frac{2}{\chi^2}\left[\left(\bar{\nabla}_\beta K^{\mu\alpha\nu\beta}\right)\bar{\xi}_\nu - K^{\mu\beta\nu\alpha}\bar{\nabla}_\beta\bar{\xi}_\nu\right]. \tag{3.323}$$

3.4.3 Self-consistency

In this chapter we have seen how the consistency of the Fierz–Pauli theory in Minkowski spacetime coupled to matter requires the introduction of an infinite series of higher-order terms whose resummation leads to GR without a cosmological constant. This is evidently consistent with the derivation of the Fierz–Pauli theory from GR as the linear perturbation theory around Minkowski spacetime.

Now we have found a generalization of the Fierz–Pauli theory in an arbitrary background satisfying the cosmological vacuum Einstein equation $\bar{R}_{\mu\nu} = \Lambda \bar{g}_{\mu\nu}$ as the linear perturbation theory around that background and it is natural to ask ourselves whether, by requiring consistency in the coupling of this theory to matter, we are going to arrive at GR with a cosmological constant. (The linear theory coupled to matter is inconsistent for exactly the same reasons as in the Minkowski case.)

As shown by Deser in [301], the answer to this question is affirmative. We are not going to give here all the details of the proof, which follows closely the proof in the Minkowski case, but it is, however, interesting to see the first-order form of the Fierz–Pauli action in curved background that constitutes its starting point:

$$\boxed{\begin{aligned} S^{(1)}_{\mathrm{FP}}[\varphi^{\mu\nu}, \Gamma_{\mu\nu}{}^{\rho}] = \frac{1}{\chi^2}\int d^d x \,\big\{ \chi\, \Gamma_{\mu\nu}{}^{\rho}\big(\delta_{\rho}{}^{\mu}\bar{\nabla}_{\sigma}\varphi^{\sigma\nu} - \bar{\nabla}_{\rho}\varphi^{\mu\nu}\big) \\ + \bar{\mathfrak{g}}^{\mu\nu} 2\Gamma_{\lambda[\mu}{}^{\rho}\Gamma_{\rho]\nu}{}^{\lambda} + \tfrac{1}{2}\Lambda\big(h^{\mu\nu}h_{\mu\nu} - \tfrac{1}{2}h^2\big)\big\}\,. \end{aligned}} \tag{3.324}$$

Here both $\varphi^{\mu\nu}$ and $\mathfrak{g}^{\mu\nu}$ are tensor densities.

3.5 Final Comments

In this chapter we have found a SRFT of gravity (GR) that is very satisfactory from many points of view. First of all, it describes extremely well what is observed. Second, it is a theory with a high degree of internal self-consistency that can be obtained from very few principles (either the principle of equivalence and general covariance or consistent interaction of a massless spin-2 particle).

However, it also has some drawbacks: we wanted to follow the steps that led to the development of the SRQFTs like quantum electrodynamics that we know so well, but we found at the end that the quantum theory based on this consistent classical theory is not consistent. Thus, at the microscopic level, the answer we have obtained is not satisfactory. In fact, at the microscopic level there arise questions like that of the coupling of gravity to fermions that have no answer in the formalism we have developed.

How should GR be modified in order to obtain a consistent quantum theory is a question that has received many tentative answers, the latest being string theory. In string theory, as in some of the alternative theories that have been proposed, there are additional fields, in the presence of which the proofs of uniqueness and self-consistency of GR are no longer valid. Furthermore, there is a prescription for the coupling of all those fields to fermions and some of the additional gravitational fields can be interpreted as torsion. We want to gain some understanding of all these elements that enter into the gravitational part of string theory as well as other alternative theories of gravity. Some of these elements are more or less trivial extensions of GR (for instance, its reformulation in the Vielbein formalism which allows the coupling to spinors) and, in fact, it is always (or usually) possible to see the new theories as GR coupled to different fields. In the next few chapters we are going to review these elements and theories that contain them: the Cartan–Sciama–Kibble theory, non-symmetric theories of gravity, theories of teleparallelism, and supergravity theories. The simplest way to introduce most of them is through a minimal action principle and the formulation of the minimal action principle for GR will be the first step in this direction.

4
Action principles for gravity

A minimal action principle is a basic ingredient of any field theory. With it (with an action) we can systematically find conserved currents and charges, canonically conjugate momenta, and a Hamiltonian (which is necessary for canonical quantization), etc. On the other hand, it is easier to deal with actions than with equations of motion; it is easier to include new fields and couplings in the action respecting certain symmetries than to invent new consistent equations of motion for them and modifications of the equations of motion of the old fields.

In this chapter we are going to study in detail several action principles for GR and for more general theories we will be concerned with later on. First, we will study the standard second-order Einstein–Hilbert action that we found as the result of imposing self-consistency on the Fierz–Pauli theory coupled to matter. We will derive the Einstein equations from it and we will find the right boundary term that will allow us to impose boundary conditions on the variations of the metric $\delta g_{\mu\nu}$ only, not on its derivatives. We will do the same for theories including a scalar and in a conformal frame that is not Einstein's. In these theories, an extra scalar factor K (which could be $e^{-2\phi}$ in the string effective action) appears multiplying the Ricci scalar and obtaining the gravitational equations becomes more involved.

We are also going to study the behavior of the Einstein–Hilbert action under GCTs and we will obtain the Bianchi (gauge) identity and Noether current associated with them and see how they are modified by the addition of boundary terms to the action.

Then we will study the first-order formalism in which the metric $g_{\mu\nu}$ and the connection $\Gamma_{\mu\nu}{}^{\rho}$ are considered as independent variables and the first-order formalism for the Vielbein $e_{\mu}{}^{a}$ and the spin connection $\omega_{\mu}{}^{ab}$, with and without fermions, which will be seen to induce torsion. There is also a *purely affine* formulation of GR in which the only variable is the (symmetric) affine connection $\Gamma_{\mu\nu}{}^{\rho}$ and we will review it briefly.

The first-order formalism and the purely affine formulation are very useful for formulating Einstein's "unified theory," which is based on a non-symmetric "metric" tensor. We take the opportunity to revisit this and other non-symmetric gravity theories (NGTs).

Motivated by the success of the first-order formalism with Vielbein and spin connection, we will review the MacDowell–Mansouri formulation of four-dimensional gravity as the gauge theory of the four-dimensional Poincaré group, which we will obtain by Wigner–Inönü contraction from the AdS_4 case.

Finally, we will briefly review teleparallel formulations and generalizations of GR.

4.1 The Einstein–Hilbert action

In d dimensions, the Einstein–Hilbert action [535] is

$$\boxed{S_{\mathrm{EH}}[g] = \frac{c^3}{16\pi G_{\mathrm{N}}^{(d)}} \int_{\mathcal{M}} d^d x \sqrt{|g|}\, R(g),} \tag{4.1}$$

where $R(g)$ is the Ricci scalar of the metric $g_{\mu\nu}$, $G_{\mathrm{N}}^{(d)}$ is the d-dimensional Newton constant and $\mathcal{M}$ is the d-dimensional manifold we are integrating over. Since we have obtained this action by imposing consistent coupling of the special-relativistic field theory, we know that it is canonically normalized and we also know which expression for the force between two particles it leads to (see Eq. (3.140)). We have introduced here the speed of light in order to find the dimensions of $G_{\mathrm{N}}^{(d)}$ in "unnatural units,": $M^{-1}L^{d-1}T^{-2}$. Recall that the metric $g_{\mu\nu}$ is dimensionless in our conventions. Recall also that the factor of 16π is associated with rationalized units only in $d = 4$.

Observe that what will appear in the path integral

$$\mathcal{Z} = \int Dg\; e^{+iS_{\mathrm{EH}}/\hbar} \tag{4.2}$$

is the dimensionless combination

$$\frac{S_{\mathrm{EH}}}{\hbar} = \frac{2\pi}{\ell_{\mathrm{Planck}}^{d-2}} \int d^d x \cdots, \tag{4.3}$$

where

$$\frac{\ell_{\mathrm{Planck}}^{d-2}}{2\pi} = \frac{\hbar G_{\mathrm{N}}^{(d)}}{c^3}, \tag{4.4}$$

is the d-dimensional *Planck length*.[1] In the absence of any other dimensional quantity this is the only combination of the constants $\hbar$, c, and $G_{\mathrm{N}}^{(d)}$ with dimensions of length. However, if there is an object of mass M, there are two more combinations with dimensions of length: the *Compton wavelength* associated with the object,

$$\lambda\!\!\!^{-}_{\mathrm{Compton}} = \frac{\hbar}{Mc}, \tag{4.6}$$

[1] Sometimes the *reduced Planck length*

$$\bar{\ell}_{\mathrm{Planck}} = \ell_{\mathrm{Planck}}/(2\pi). \tag{4.5}$$

We have also been using the constant χ defined by $\chi^2 = 16\pi G_{\mathrm{N}}^{(d)}/c^3$.

which is of purely quantum-mechanical nature, and the d-dimensional *Schwarzschild radius*,

$$R_{\mathrm{S}} = \left(\frac{16\pi M G_{\mathrm{N}}^{(d)} c^{-2}}{(d-2)\omega_{(d-2)}}\right)^{\frac{1}{d-3}}, \tag{4.7}$$

which is of purely classical, gravitational nature. It occurred naturally in the gravitational field of a massive point-like particle, Eq. (3.124).

With the constants $\hbar$, c, and $G_{\mathrm{N}}^{(d)}$ one can also build a combination with units of mass: the *Planck mass*,

$$M_{\mathrm{Planck}} = \left(\frac{\hbar^{d-3}}{G_{\mathrm{N}}^{(d)} c^{d-5}}\right)^{\frac{1}{d-2}}, \tag{4.8}$$

so the prefactor of the action in the path integral is

$$\frac{c^3}{G_{\mathrm{N}}^{(d)}\hbar} = \left(\frac{M_{\mathrm{Planck}}c}{\hbar}\right)^{d-2}. \tag{4.9}$$

If we consider objects whose masses are of the order of the Planck mass, then it is immediately seen that their Compton wavelengths become of the order of their Schwarzschild radii, which are of the order of the Planck length:

$$M \sim M_{\mathrm{Planck}} \Rightarrow \lambda\!\!\!^{-}_{\mathrm{Compton}} \sim R_{\mathrm{S}} \sim \ell_{\mathrm{Planck}}. \tag{4.10}$$

At that point quantum-mechanical effects will become important.

If we naively try to quantize by standard GR methods (starting from its perturbative expansion), we find that the quantum gravitational coupling constant (Planck length) is dimensional and, by standard arguments, we expect to obtain a non-renormalizable theory. This is indeed the case. As we will see, in string theory there is no unique constant that plays the role of length scale and coupling constant as does the Planck length in GR: there are two constants with dimensions of length: Planck's constant and the *string length* ℓ_{s}. The dimensionless quotient is essentially the *string coupling constant* g_{s}. In that context the Schwarzschild radius has to be compared with ℓ_{s} in order to see when (string) quantum gravity effects become important. On the other hand, we can have better expectations about the perturbative renormalizability of the theory since the expansion is made in the dimensionless parameter g_{s}, instead of ℓ_{Planck} or ℓ_{s}.

The Einstein–Hilbert action Eq. (4.1) contains second derivatives of the metric. However, the terms with second derivatives take the form of a total derivative,[2] symbolically

$$S_{\mathrm{EH}}[g] = \frac{c^3}{16\pi G_{\mathrm{N}}^{(d)}}\int_{\mathcal{M}} d^dx\sqrt{|g|}\,(\partial g)^2 + \frac{c^3}{16\pi G_{\mathrm{N}}^{(d)}}\int_{\mathcal{M}} d^dx\ \partial_\mu\omega^\mu(\partial g). \tag{4.11}$$

This means that the original action Eq. (4.1) can in principle be be used to obtain equations of motion that are of second order in derivatives of the metric. However, we would have to impose conditions on the derivatives of the metric on the boundary. Furthermore, observe

[2] See, for instance, [644] and Appendix D.

that the "vector" $\omega^\mu(\partial g)$ does not transform as such under GCTs. The solution to these problems consists in adding a general-covariant boundary term to the original EH action. We are going to see next how to find the equations of motion and the right boundary term.

4.1.1 Equations of motion

Let us vary the original Einstein–Hilbert action with respect to the metric. For simplicity we temporarily set $\chi = 1$. Bearing in mind that $R(g) = g^{\mu\nu} R_{\mu\nu}(\Gamma(g))$ and $R_{\mu\nu}(\Gamma(g))$ depends on g only through the Levi-Cività connection $\Gamma(g)$ so we can use the Palatini identity Eq. (3.285),

$$\delta R_{\mu\nu} = \nabla_\mu \delta\Gamma_{\rho\nu}{}^\rho - \nabla_\rho \delta\Gamma_{\mu\nu}{}^\rho, \tag{4.12}$$

and using the identities

$$\delta g^{\mu\nu} = -g^{\nu\alpha} g^{\mu\beta} \delta g_{\alpha\beta}, \qquad \delta g = g\, g^{\alpha\beta} \delta g_{\alpha\beta}, \tag{4.13}$$

we immediately find

$$\delta S_{\text{EH}} = \int d^d x \sqrt{|g|} \left\{ -G^{\mu\nu} \delta g_{\mu\nu} + g^{\mu\nu} \left[\nabla_\mu \delta\Gamma_{\rho\nu}{}^\rho - \nabla_\rho \delta\Gamma_{\mu\nu}{}^\rho \right] \right\}. \tag{4.14}$$

Since our covariant derivative is metric-compatible we can absorb the metric in the last term and combine the two terms into a single total derivative,

$$\delta S_{\text{EH}} = -\int_{\mathcal{M}} d^d x \sqrt{|g|}\, G^{\mu\nu} \delta g_{\mu\nu} + \int_{\mathcal{M}} d^d x \sqrt{|g|}\, \nabla_\rho v^\rho, \tag{4.15}$$

where

$$v^\rho = g^{\rho\mu} \delta\Gamma_{\mu\nu}{}^\nu - g^{\mu\nu} \delta\Gamma_{\mu\nu}{}^\rho. \tag{4.16}$$

We now have to use the equation that expresses the variation of the Levi-Cività connection with respect to a variation of the metric in order to find the variation of the action as a function of the variation of the metric. That expression was given in Eq. (3.282) and with it we find

$$v^\rho = g^{\rho\mu} g^{\sigma\nu} \left(\nabla_\mu \delta g_{\sigma\nu} - \nabla_\sigma \delta g_{\mu\nu} \right). \tag{4.17}$$

Using now Stokes' theorem Eq. (1.141), we reexpress the integral of the total derivative terms as an integral over the boundary,

$$\int_{\mathcal{M}} d^d x \sqrt{|g|}\, \nabla_\rho v^\rho = (-1)^{d-1} \int_{\partial\mathcal{M}} d^{d-1}\Sigma_\rho v^\rho = (-1)^{d-1} \int_{\partial\mathcal{M}} d^{d-1}\Sigma\, n_\rho v^\rho, \tag{4.18}$$

where $d^{d-1}\Sigma_\rho$ is defined in Chapter 1,

$$d^{d-1}\Sigma \equiv n^2 d^{d-1}\Sigma_\rho n^\rho, \tag{4.19}$$

and n^μ is the unit vector normal to the boundary hypersurface $\partial\mathcal{M}$ ($n^2 = +1$ for spacelike hypersurfaces with timelike normal unit vector and $n^2 = -1$ for timelike hypersurfaces with spacelike normal unit vector). Finally, we expand the integrand

$$n_\rho v^\rho = n^\mu g^{\sigma\nu} \left(\nabla_\mu \delta g_{\sigma\nu} - \nabla_\sigma \delta g_{\mu\nu} \right) = n^\mu h^{\sigma\nu} \left(\nabla_\mu \delta g_{\sigma\nu} - \nabla_\sigma \delta g_{\mu\nu} \right), \tag{4.20}$$

where $h_{\mu\nu} = g_{\mu\nu} - n^2 n_\mu n_\nu$ is the induced metric on the hypersurface $\partial\mathcal{M}$ (see Section 1.8). Thus, we arrive at

$$\delta S_{\rm EH} = -\int_{\mathcal{M}} d^d x\sqrt{|g|}\, G^{\mu\nu}\delta g_{\mu\nu} + (-1)^{d-1}\int_{\partial\mathcal{M}} d^{d-1}\Sigma\, n^\mu h^{\sigma\nu}\nabla_\mu \delta g_{\sigma\nu} - (-1)^{d-1}\int_{\partial\mathcal{M}} d^{d-1}\Sigma\, n^\mu h^{\sigma\nu}\nabla_\sigma\delta g_{\mu\nu}. \tag{4.21}$$

This is the final form of the variation of the action we were after. Now, we would like to be able to obtain the Einstein equation by requiring the action to be stationary (so $\delta S_{\rm EH} = 0$) under arbitrary variations of the metric *vanishing on the boundary*:

$$\delta g_{\mu\nu}\,\big|_{\partial\mathcal{M}} = 0. \tag{4.22}$$

If $\delta g_{\mu\nu}$ is constant on the boundary, then its covariant derivative projected onto the boundary directions with $h^{\mu\nu}$ must vanish:

$$h^{\sigma\nu}\nabla_\sigma\delta g_{\mu\nu} = 0, \tag{4.23}$$

and the second of the two boundary terms vanishes. However, the first does not vanish unless we impose boundary conditions for the covariant derivative of the variation of the metric. In order to obtain the Einstein equation we must cancel out that boundary term with the variation of another boundary term added to the Einstein–Hilbert action. This boundary term is nothing but the integral over the boundary of the trace of the extrinsic curvature of the boundary given in Eq. (1.149). Observe that

$$\delta\mathcal{K} = \delta h^\mu{}_\nu \nabla_\mu n^\nu + h^\mu{}_\nu \delta\Gamma_{\mu\rho}{}^\nu n^\rho. \tag{4.24}$$

The first term vanishes on the boundary due to our boundary condition (4.22). Using Eq. (3.282) for $\delta\Gamma$, we find

$$\delta\mathcal{K}|_{\partial\mathcal{M}} = \tfrac{1}{2} n^\rho h^{\mu\sigma}\nabla_\rho\delta g_{\mu\sigma}. \tag{4.25}$$

In conclusion, the action that one should use is the following [436, 932]:

$$\boxed{S_{\rm EH}[g] = \frac{1}{\chi^2}\int_{\mathcal{M}} d^d x\sqrt{|g|}\; R + (-1)^d \frac{2}{\chi^2}\int_{\partial\mathcal{M}} d^{d-1}\Sigma\,\mathcal{K}.} \tag{4.26}$$

Under otherwise arbitrary variations of the metric satisfying Eq. (4.22), we have shown that the variation of the Einstein–Hilbert action with boundary term (4.26), is just

$$\delta S_{\rm EH} = -\frac{1}{\chi^2}\int_{\mathcal{M}} d^d x\sqrt{|g|}\, G^{\mu\nu}\delta g_{\mu\nu}, \tag{4.27}$$

and then the vacuum Einstein equation follows, as we wanted.

4.1.2 Gauge identity and Noether current

The Einstein–Hilbert action is invariant under GCTs and we can write $\tilde{\delta}_\xi S_{\rm EH} = 0$. For variations at the same point the action transforms into the integral of a total derivative ($\chi = 1$ again):

$$\delta_\xi S_{\rm EH} = \int_{\mathcal{M}} d^d x \delta_\xi \hat{\mathcal{L}} = -\int_{\mathcal{M}} d^d x \mathcal{L}_\xi \hat{\mathcal{L}} = -\int_{\mathcal{M}} d^d x \partial_\mu (\xi^\mu \hat{\mathcal{L}}), \tag{4.28}$$

because $\hat{\mathcal{L}}$ is a scalar density. This result will be valid for any general-covariant action.

To find the gauge identity associated with the invariance under GCTs we have to find the variation of the action under variations of the metric and then use the explicit form of the variation of the metric under GCTs. For simplicity we will use the original Einstein–Hilbert action with no boundary terms and then we will discuss the effect of the addition of boundary terms. The variation of the action is given by Eqs. (4.15) and (4.16):

$$\delta_\xi S_{\rm EH} = \int_{\mathcal{M}} d^d x \sqrt{|g|} \left\{ -G^{\mu\nu} \delta_\xi g_{\mu\nu} + \nabla_\rho \big(2 g^{\mu\sigma,\rho\nu} \nabla_\mu \delta_\xi g_{\sigma\nu} \big) \right\}, \tag{4.29}$$

and, using the expression for $\delta_\xi g_{\mu\nu}$ in Eq. (1.59) and integrating once by parts, we obtain

$$\delta_\xi S_{\rm EH} = \int_{\mathcal{M}} d^d x \sqrt{|g|} \big\{ -2 \big(\nabla_\mu G^{\mu\nu} \big) \xi_\nu + \nabla_\rho 2 \big(G^{\rho\sigma} \xi_\sigma - 2 g^{\mu\sigma,\rho\nu} \nabla_\mu \nabla_{(\sigma} \xi_{\nu)} \big) \big\}. \tag{4.30}$$

On comparing this with the first form of the variation (4.28) with $\hat{\mathcal{L}} = \sqrt{|g|}\, R$, we obtain the identity

$$\int_{\mathcal{M}} d^d x \sqrt{|g|} \big\{ -2 \big(\nabla_\mu G^{\mu\nu} \big) \xi_\nu + \nabla_\rho \big(2 R^{\rho\sigma} \xi_\sigma - 4 g^{\mu\sigma,\rho\nu} \nabla_\mu \nabla_{(\sigma} \xi_{\nu)} \big) \big\} = 0. \tag{4.31}$$

This equation is true for arbitrary infinitesimal GCTs. If we take ξ^μs such that the total derivative term vanishes on the boundary, then we obtain the contracted Bianchi identity $\nabla_\mu G^{\mu\nu} = 0$ as associated gauge identity. We know that this identity is always true in this context. This, in turn, implies that the total derivative term vanishes identically, i.e. the Noether current

$$j_{\rm N}^\rho(\xi) = 2 R^{\rho\sigma} \xi_\sigma - 4 g^{\mu\sigma,\rho\nu} \nabla_\mu \nabla_{(\sigma} \xi_{\nu)}, \tag{4.32}$$

is covariantly conserved, $\nabla_\rho j_{\rm N}^\rho = 0$. By massaging this expression a bit, we can rewrite it in the form

$$j_{\rm N}^\rho(\xi) = \nabla_\mu j_{\rm N}^{\mu\rho}(\xi), \qquad j_{\rm N}^{\mu\rho}(\xi) = 2 \nabla^{[\mu} \xi^{\rho]}, \tag{4.33}$$

as is always expected in gauge theories. In Chapter 6 we will study the use of this current to define conserved quantities in GR.

Now we want to see the effect of additional total derivatives in the Einstein–Hilbert action

$$\Delta S_{\rm EH} = \int_{\mathcal{M}} d^d x \sqrt{|g|} \nabla_\mu k^\mu. \tag{4.34}$$

We just have to vary this additional piece in two different ways. One of the variations has the general form of the variation of any general-covariant action (4.28), that is,

$$\delta_\xi \Delta S_{\rm EH} = \int_{\mathcal{M}} d^d x \sqrt{|g|} \nabla_\mu \big(-\xi^\mu \nabla_\rho k^\rho \big). \tag{4.35}$$

The variation through the equation of motion gives

$$\delta_\xi \Delta S_{\text{EH}} = \int_{\mathcal{M}} d^d x \sqrt{|g|} \nabla_\mu \big[k^\rho \nabla_\rho \xi^\mu - \nabla_\rho (\xi^\rho k^\mu)\big]. \tag{4.36}$$

On combining these two results we find the additional terms in the Noether current,

$$\Delta j_{\text{N}}^\mu(\xi) = \nabla_\rho \big(2k^{[\rho}\xi^{\mu]}\big), \qquad \Delta j_{\text{N}}^{\mu\rho}(\xi) = 2k^{[\rho}\xi^{\mu]}. \tag{4.37}$$

4.1.3 Coupling to matter

As required by the PEGI, to couple matter to the gravitational field we first rewrite the Minkowskian matter action in the background of the metric that appears in the Einstein–Hilbert action, replacing everywhere $\eta_{\mu\nu}$ by $g_{\mu\nu}$ and the volume element $d^d x$ by the GCT-invariant volume element $d^d x \sqrt{|g|}$ and replacing, if necessary (in the most important cases it is not), partial derivatives by covariant derivatives with the Levi-Cività connection. The total action for the gravity–matter system is simply the sum of the Einstein–Hilbert action and the rewritten matter action Eq. (3.273). It is clear that, in general, we will not have to modify the boundary conditions for $\delta g_{\mu\nu}$ due to the addition of the matter action. Thus, the same boundary term as in the vacuum case should work. By varying this with respect to the metric, we obtain the Einstein equation (3.272), where the energy–momentum tensor is defined in Eq. (3.274), which we rewrite here for convenience:

$$T_{\text{matter}}^{\mu\nu} = \frac{2c}{\sqrt{|g|}} \frac{\delta S_{\text{matter}}}{\delta g_{\mu\nu}}. \tag{4.38}$$

First of all, we may ask ourselves about the consistency of Einstein's equation: we know that the (covariant) divergence of the l.h.s. (Einstein's tensor) vanishes due to the contracted Bianchi identity, which can be seen as a consequence of (or a condition for) the invariance of the Einstein–Hilbert action under GCTs. The r.h.s. (the energy–momentum tensor) should also be covariantly divergenceless. In fact, given *any* general-covariant action $S[\phi, g_{\mu\nu}]$, under a general variation of the fields, up to total derivatives,

$$\delta S = \int d^d x \left\{ \frac{\delta S}{\delta \phi} \delta\phi + \frac{\delta S}{\delta g_{\mu\nu}} \delta g_{\mu\nu} \right\}. \tag{4.39}$$

If the field equations of motion are satisfied and the variations are infinitesimal GCTs, then, on integrating by parts, we immediately realize that the gauge identity associated with the invariance under GCTs is always

$$\nabla_\mu \left(\frac{1}{\sqrt{|g|}} \frac{\delta S}{\delta g_{\mu\nu}} \right) = 0. \tag{4.40}$$

If the action is the Einstein–Hilbert action, this is the contracted Bianchi identity. If it is a matter action, this is the general covariantization of the Minkowskian energy–momentum conservation law $\partial_\mu T_{\text{matter}}^{\mu\nu} = 0$, namely

$$\nabla_\mu T_{\text{matter}}^{\mu\nu} = 0. \tag{4.41}$$

This equation ensures the consistency of the Einstein equations. However, it is not a conservation law. We will explain and discuss this problem in Chapter 6.

4.2 The Einstein–Hilbert action in different conformal frames

The simplest field that a matter Lagrangian added to the Einstein–Hilbert action can have is a scalar. Matter Lagrangians containing scalars appear in many theories, particularly extended $N > 2$ supergravity theories, Kaluza–Klein theories, and string theory. The scalars' kinetic term usually has the form of a non-linear σ-model in which the (real) scalars can be understood as coordinates in some target space, which usually is a homogeneous space. Hence, real scalars can take values in different ranges. If a particular scalar that we will denote by K takes values in $\mathbb{R}^+$ then, we can always rescale the metric in the Einstein–Hilbert action (which we will henceforth refer to as the *Einstein metric*) via a *Weyl* or *conformal* transformation

$$g_{\mu\nu} \to K^\alpha g_{\mu\nu}, \tag{4.42}$$

where α is some number. Sometimes this transformation is called a change of *conformal frame*. The Einstein–Hilbert action is written in the *Einstein (conformal) frame*. The new metric has the same signature and its equation of motion can be derived from the rescaled action (see Appendix E) that we will generically write in this form, ignoring the matter Lagrangian:

$$S[g, K] \sim \int d^d x \sqrt{|g|}\, K R(g). \tag{4.43}$$

In the context of string theory $K = e^{-2\phi}$, where ϕ is the *dilaton field*; then the metric is called the *string metric* and it is usually said that the action is written in the *string (conformal) frame*. In the context of Kaluza–Klein theory, if we reduce over a circle, and $K = k$, where k is the Kaluza–Klein scalar and, in more general compactifications, K is a scalar that measures the volume of the internal manifold, then the metric is called the *Kaluza–Klein metric* and we say that the action is written in the *Kaluza–Klein (conformal) frame*. We will define other conformal frames (*p-brane frames, etc.*) later on.

One important detail that has to be taken into account is the possibility that the vacuum value of the scalar K is not just 1 but some number K_0. In that case, the vacuum of the metric $g_{\mu\nu}$ is rescaled by K_0^α, which is not permissible. We will discuss this important issue at length in Section 11.2.2. In this section we are simply going to explain in detail how to obtain the metric equation of motion by direct variation of the above action (it is obvious that one can always perform the rescaling in the Einstein equation, but, as usual, we expect to obtain more information from the variation of the action).

Using the Palatini identity Eq. (3.285) and Eqs. (4.13), we find

$$\delta S[g, K] = -\int_{\mathcal{M}} d^d x \sqrt{|g|} \left\{ K [G^{\mu\nu} \delta g_{\mu\nu} - \nabla_\mu v^\mu] - R \delta K \right\}, \tag{4.44}$$

where v is given by (4.16). We are going to ignore the piece proportional to δK because, after all, in general, in the full action there will be more terms containing K. Integrating by parts once gives

$$\delta S[g, K] = -\int_{\mathcal{M}} d^d x \sqrt{|g|} \left\{ K [G^{\mu\nu} \delta g_{\mu\nu} + v^\mu \nabla_\mu \ln K] - \nabla_\rho (K v^\rho) \right\}. \tag{4.45}$$

Writing v^μ as

$$v^\mu = 2 g^{\mu\nu,\rho\sigma} \nabla_\rho \delta g_{\sigma\nu} \tag{4.46}$$

and integrating by parts again gives

$$\delta S[g,K] = -\int_{\mathcal{M}} d^d x\sqrt{|g|}K\left\{G^{\mu\nu}-2\left[\nabla_\rho \ln K\nabla_\sigma \ln K+\nabla_\rho\nabla_\sigma \ln K\right]g^{\sigma\mu,\rho\nu}\right\}\delta g_{\mu\nu}$$
$$+\int_{\mathcal{M}} d^d x\sqrt{|g|}\,\nabla_\lambda\left\{2\left[K g^{\lambda\nu,\rho\sigma}\nabla_\rho \delta g_{\sigma\nu} - \nabla_\rho K g^{\rho\nu,\lambda\sigma}\delta g_{\sigma\nu}\right]\right\}. \tag{4.47}$$

If we add the boundary term

$$-2\int d^{d-1}\Sigma\, K\mathcal{K} \tag{4.48}$$

to the action, it is clear that, on imposing the boundary condition Eq. (4.22), we will obtain the following equation for the metric:

$$G^{\alpha\beta}+\left[\partial^\alpha \ln K\partial^\beta \ln K - g^{\alpha\beta}(\partial \ln K)^2\right]+\left[\nabla^\alpha\nabla^\beta \ln K - g^{\alpha\beta}\nabla^2 \ln K\right]=0. \tag{4.49}$$

Observe that one obtains a non-trivial equation of motion for the scalar K (or $\log K$) even though there is (apparently) no kinetic term for it in the action we have considered. This is

$$\left(\nabla^2+\frac{d-2}{2(d-1)}R\right)K=0. \tag{4.50}$$

Otherwise, by going to a conformal frame in which the kinetic term explicitly disappears, one could eliminate a scalar degree of freedom that would be present in any other frame.

Observe also that this scalar K is not a *conformal scalar*. A conformal scalar K_c has the equation of motion

$$\left(\nabla^2+\frac{d-2}{4(d-1)}R\right)K_c=0, \tag{4.51}$$

which, under simultaneous Weyl rescalings of the metric and the scalar,

$$\tilde{g}_{\mu\nu}=\Omega^2 g_{\mu\nu}, \qquad \tilde{K}_c=\Omega^{\frac{2-d}{2}}K_c, \tag{4.52}$$

also rescales (i.e. it is invariant),

$$\left(\tilde{\nabla}^2+\frac{2-d}{4(1-d)}\tilde{R}\right)\tilde{K}=\Omega^{-\frac{(2+d)}{2}}\left(\nabla^2+\frac{(2-d)}{4(1-d)}R\right)K_c=0. \tag{4.53}$$

To construct an action for a conformal scalar, we have to add to the above action a kinetic term with the right coefficient:

$$S_c\sim\int d^d x\sqrt{|g|}K\left[R-\frac{d-1}{d-2}(\partial \ln K)^2\right], \tag{4.54}$$

and then we find that $K=K_c^2$, so the action written in terms of the conformal scalar is

$$S[K_c]\sim\int d^d x\sqrt{|g|}\,K_c^2\left[R-\frac{4(d-1)}{d-2}(\partial \ln K_c)^2\right]. \tag{4.55}$$

Both the trace of the variation with respect to the metric and the variation with respect to K_c lead to the above equation of motion.

When we studied vector and tensor fields living on a general background, we adopted as sign of their masslessness the existence of gauge transformations leaving their equations of motion invariant. If we interpret the above equations as the equations of a scalar field living on a background metric $g_{\mu\nu}$, we may wonder how we can tell whether the scalar field is massless. The only kind of local transformations that we can define for a scalar field are the above Weyl transformations and we can define as a massless field one whose equation of motion is invariant under them. Therefore we could consider the conformal scalar as a massless scalar. This means, in particular, that the equation of motion of a massless scalar in a spacetime satisfying $R_{\mu\nu} = \Lambda g_{\mu\nu}$ is

$$\left(\nabla^2 + \frac{d(d-2)}{4(d-1)}\Lambda\right)K_{\rm c} = 0, \tag{4.56}$$

and, as usual, the Λ term is not a mass term but, on the contrary, its presence ensures the masslessness of the scalar field.

4.3 The first-order (Palatini) formalism

This formalism [752] consists in writing an action in which the metric and the connection (which contains the dependence on the derivatives of the metric) are considered independent variables. The connection is, therefore, not the Levi-Cività connection. It is assumed to be torsion-free, i.e. $\Gamma_{[\mu\nu]}{}^{\rho} = 0$, but no other properties (metric-compatibility, for example) are assumed. The first-order action contains only derivatives of the connection and it is linear in them. To obtain the equations of motion, one now has to vary the metric and the connection independently. The connection equation of motion gives us the standard relation between the connection and the metric and the metric equation is, after substitution of the solution to the other equation, nothing but the Einstein equation.

The first-order action turns out to be essentially the Einstein–Hilbert action:[3]

$$\boxed{S[g_{\mu\nu}, \Gamma_{\mu\nu}{}^{\rho}] = \int d^dx\sqrt{|g|}\, g^{\mu\nu}R_{\mu\nu}(\Gamma).} \tag{4.57}$$

All the dependence on the metric is concentrated in the factor $\sqrt{|g|}\, g^{\mu\nu}$ since the Ricci tensor depends only on the connection and its derivatives as shown in Eq. (1.33).

We stress that, since the connection is here a variable, and it is not the Levi-Cività connection, one cannot use the standard property

$$\int d^dx\sqrt{|g|}\ \nabla_\mu \xi^\mu = \int d^dx\ \partial_\mu\left(\sqrt{|g|}\,\xi^\mu\right). \tag{4.58}$$

The calculations are simpler using as a variable the density

$$\mathfrak{g}^{\mu\nu} = \sqrt{|g|}\, g^{\mu\nu}. \tag{4.59}$$

[3] We set $\chi = 1$ throughout this section.

Furthermore, we are not going to assume in our derivation of the equations of motion either the symmetry of the connection or the symmetry of the "metric," which we will impose at the very end. In this way, we can obtain with a minimum extra work the equations of the Einstein–Straus–Kaufman [358, 364, 367, 369] non-symmetric gravity theory (NGT) which was (unsuccessfully) proposed as a unified relativistic theory of gravitation and electromagnetism in which the antisymmetric part of the "metric" $g^{[\mu\nu]}$ should be identified with the electromagnetic field strength tensor[4] $F^{\mu\nu}$.

In the NGT the inverse "metric" is also denoted by $g^{\mu\nu}$ and satisfies

$$g^{\mu\nu}g_{\nu\rho} = \delta^{\mu}{}_{\rho}, \qquad g_{\alpha\beta}g^{\beta\gamma} = \delta_{\alpha}{}^{\gamma}, \tag{4.60}$$

but $g^{\mu\nu}g_{\mu\rho} \neq \delta^{\nu}{}_{\rho}$. Also, we cannot use it to lower or raise indices.

Let us now vary the above action with respect to the metric and connection. By using Palatini's identity Eq. (3.286), we find

$$\begin{aligned}
\delta S &= \int d^d x \, \left\{\delta\mathfrak{g}^{\alpha\beta}\, R_{\alpha\beta}(\Gamma) + \mathfrak{g}^{\alpha\beta}\left[\nabla_{\alpha}\delta\Gamma_{\rho\beta}{}^{\rho} - \nabla_{\rho}\delta\Gamma_{\alpha\beta}{}^{\rho} - T_{\alpha\rho}{}^{\sigma}\delta\Gamma_{\sigma\beta}{}^{\rho}\right]\right\} \\
&= \int d^d x \, \left\{\delta\mathfrak{g}^{\alpha\beta}\, R_{\alpha\beta}(\Gamma) + \nabla_{\rho}\left[\left(\mathfrak{g}^{\rho\beta}\delta_{\sigma}{}^{\alpha} - \mathfrak{g}^{\alpha\beta}\delta_{\sigma}{}^{\rho}\right)\delta\Gamma_{\alpha\beta}{}^{\sigma}\right]\right. \\
&\qquad\qquad \left. + \left[\nabla_{\sigma}\mathfrak{g}^{\alpha\beta} - \nabla_{\rho}\mathfrak{g}^{\rho\beta}\delta_{\sigma}{}^{\alpha} - \mathfrak{g}^{\lambda\beta}T_{\lambda\sigma}{}^{\alpha}\right]\delta\Gamma_{\alpha\beta}{}^{\sigma}\right\}.
\end{aligned} \tag{4.61}$$

Using now the identity for vector densities

$$\nabla_{\mu}\mathfrak{v}^{\mu} = \partial_{\mu}\mathfrak{v}^{\mu} + \mathfrak{v}^{\mu}T_{\mu\rho}{}^{\rho}, \tag{4.62}$$

and integrating by parts, we obtain, up to a total derivative

$$\begin{aligned}
\delta S = \int d^d x \, &\left\{\delta\mathfrak{g}^{\alpha\beta}R_{\alpha\beta}(\Gamma) + \left[T_{\rho\delta}{}^{\delta}\left(\mathfrak{g}^{\rho\beta}\delta_{\sigma}{}^{\alpha} - \mathfrak{g}^{\alpha\beta}\delta_{\sigma}{}^{\rho}\right)\right.\right. \\
&\left.\left. + \nabla_{\sigma}\mathfrak{g}^{\alpha\beta} - \nabla_{\rho}\mathfrak{g}^{\rho\beta}\delta_{\sigma}{}^{\alpha} - \mathfrak{g}^{\lambda\beta}T_{\lambda\sigma}{}^{\alpha}\right]\delta\Gamma_{\alpha\beta}{}^{\sigma}\right\}.
\end{aligned} \tag{4.63}$$

Since the metric and the connection are independent, we obtain two equations from the minimal action principle:

$$\begin{aligned}
\frac{\delta S}{\delta\mathfrak{g}^{\alpha\beta}} &= R_{\alpha\beta}(\Gamma) = 0, \\
\frac{\delta S}{\delta\Gamma_{\alpha\beta}{}^{\gamma}} &= \nabla_{\gamma}\mathfrak{g}^{\alpha\beta} - \nabla_{\rho}\mathfrak{g}^{\rho\beta}\delta_{\gamma}{}^{\alpha} - \mathfrak{g}^{\lambda\beta}T_{\lambda\gamma}{}^{\alpha} + \mathfrak{g}^{\rho\beta}\delta_{\gamma}{}^{\alpha}T_{\rho\delta}{}^{\delta} - \mathfrak{g}^{\alpha\beta}T_{\gamma\delta}{}^{\delta} = 0.
\end{aligned} \tag{4.64}$$

The first equation would be the Einstein equation if the connection were the Levi-Cività connection. Observe that, if we couple bosonic (scalar or vector) matter minimally to this

[4] See also [654, 837, 838, 895]. A more recent NGT that reinterprets Einstein's theory was proposed in [699]. In it the antisymmetric part of the metric is also considered as a sort of new gravitational interaction. Clearly, the weak-field limit cannot be the Fierz–Pauli theory but contains another field corresponding to the antisymmetric part of the metric. While this suggests a relation with string theory, which also contains a rank-2 antisymmetric tensor (the Kalb–Ramond field), these two fields appear in quite different ways: the Kalb–Ramond field has an extra gauge symmetry, which allows it to be consistently quantized, whereas the antisymmetric part of the NGT "metric" transforms only under GCTs. See [253, 285, 286, 616].

action, we do not have to introduce any term containing the connection. Thus, the equation for the connection would not change and the equation for the metric, would become the Einstein equation with non-vanishing energy–momentum tensor (again, if the connection were the Levi-Cività connection).

To find the relation between the connection and the metric, we have to solve the second equation. It is convenient to use a new connection $\tilde{\Gamma}$, defined by

$$\tilde{\Gamma}_{\mu\nu}{}^{\rho} = \Gamma_{\mu\nu}{}^{\rho} + \frac{1}{d-1} T_{\mu\sigma}{}^{\sigma} \delta_{\nu}{}^{\rho}. \tag{4.65}$$

Observe that the new connection $\tilde{\Gamma}$ does not completely determine the old one, Γ. In fact, if we shift Γ by an arbitrary vector f_μ according to

$$\Gamma_{\mu\nu}{}^{\rho} \to \Gamma_{\mu\nu}{}^{\rho} + f_\mu \delta_\nu{}^{\rho}, \tag{4.66}$$

the connection $\tilde{\Gamma}$ is not modified. Thus, the expression for Γ in terms of $\tilde{\Gamma}$ is

$$\Gamma_{\mu\nu}{}^{\rho} = \tilde{\Gamma}_{\mu\nu}{}^{\rho} + f_\mu \delta_\nu{}^{\rho}, \tag{4.67}$$

where f_μ cannot be determined from $\tilde{\Gamma}$. The new connection allows us to rewrite the second equation in the form

$$\partial_\sigma \mathfrak{g}^{\alpha\beta} + \tilde{\Gamma}_{\delta\sigma}{}^{\alpha} \mathfrak{g}^{\delta\beta} + \mathfrak{g}^{\alpha\delta} \tilde{\Gamma}_{\sigma\delta}{}^{\beta} - \mathfrak{g}^{\alpha\beta} \tilde{\Gamma}_{\sigma\delta}{}^{\delta} = 0. \tag{4.68}$$

On contracting in the above equation the indices σ with α and σ with β, taking the difference, and using the property

$$\tilde{\Gamma}_{\mu\rho}{}^{\rho} = \tilde{\Gamma}_{\rho\mu}{}^{\rho}, \tag{4.69}$$

we arrive at the Maxwell-like equation for the antisymmetric part of $\mathfrak{g}$

$$\partial_\alpha \mathfrak{g}^{[\alpha\beta]} = 0. \tag{4.70}$$

By contracting now Eq. (4.68) with $g_{\alpha\beta}/\sqrt{|g|}$, we obtain

$$\partial_\sigma \ln \sqrt{|g|} = \tilde{\Gamma}_{\sigma\alpha}{}^{\alpha}, \tag{4.71}$$

and, on plugging this back into Eq. (4.68), we obtain an equation for the inverse metric,

$$\partial_\sigma g^{\alpha\beta} + \tilde{\Gamma}_{\sigma\delta}{}^{\beta} g^{\alpha\delta} + \tilde{\Gamma}_{\delta\sigma}{}^{\alpha} g^{\delta\beta} = 0. \tag{4.72}$$

We now multiply by the inverse "metrics" $g_{\gamma\alpha}$ and $g_{\beta\varphi}$ to obtain, at last,

$$\partial_\sigma g_{\gamma\varphi} - \tilde{\Gamma}_{\sigma\gamma}{}^{\beta} g_{\beta\varphi} - \tilde{\Gamma}_{\varphi\sigma}{}^{\alpha} g_{\gamma\alpha} = 0. \tag{4.73}$$

Although we have started with the connection Γ, the above equation allows us only to solve for the connection $\tilde{\Gamma}$ in terms of the metric.

It is easy to particularize this general setup for the case that interests us: a symmetric metric $g^{[\mu\nu]} = 0$ and a torsion-free connection $\Gamma_{[\mu\nu]}{}^{\rho} = 0$. In this case, $R_{\mu\nu}(\Gamma)$ is automatically symmetric, $\Gamma = \tilde{\Gamma}$, and the above equation Eq. (4.73) is the metric-compatibility equation $\nabla_\sigma g_{\gamma\varphi} = 0$ whose solution is (see Chapter 1) the Levi-Cività connection (Christoffel symbols). Then we recover the vacuum Einstein equation.

In the presence of matter, this formalism leads to the standard Einstein equation if the affine connection does not occur in the matter action, which is the case for scalars and gauge fields. Otherwise, the equation for the equation is modified and, in general, the connection has torsion. Actually, this can turn into an advantage of this formalism in certain cases (e.g. supergravity theories), although we develop a formalism to couple fermions to gravity in Section 4.4.

4.3.1 The purely affine theory

We have seen two action principles leading to the Einstein equations. In the first one, the fundamental variables were the components of the metric tensor. In the second one, the fundamental variables were both the components of the metric tensor and the components of the affine connection. For completeness, we are going to see briefly that it is actually possible to write an action leading to the vacuum Einstein equations in the presence of a cosmological constant that is a functional of the components of the affine connection alone.

The simplest tensors that one can construct from the affine connection and its first derivatives are the curvature and Ricci tensors. To write an action, we need to integrate a density. The simplest density constructed from these two tensors alone we can think of is the square root of the determinant of the Ricci tensor, so

$$S \sim \int d^d x \sqrt{|R_{\mu\nu}(\Gamma)|}, \quad \Rightarrow \delta S = \int d^d x \frac{\delta S}{\delta R_{\mu\nu}} \delta R_{\mu\nu}(\Gamma). \tag{4.74}$$

The crucial point in this formalism is the definition

$$\frac{\delta S}{\delta R_{\mu\nu}} \equiv \frac{\alpha}{2} \mathfrak{g}^{\mu\nu}, \tag{4.75}$$

where α is some constant and the metric density is $\sqrt{|g|} g^{\mu\nu}$, which does not need to be symmetric. Actually, it has the same symmetry as the Ricci tensor. Thus, if we want to have a symmetric metric, we have to take the determinant of the symmetric part of the Ricci tensor in the action, but the connection is arbitrary. From the above equation we find an equation with the structure of the cosmological Einstein equation:

$$R_{\mu\nu}(\Gamma) = \Lambda g_{\mu\nu}, \qquad \Lambda = \alpha^{\frac{2}{d-2}}. \tag{4.76}$$

On substituting this into the variation of the action, we obtain

$$\delta S = \frac{\Lambda^{\frac{d-2}{2}}}{2} \int d^d x \, \mathfrak{g}^{\mu\nu} \delta R_{\mu\nu}(\Gamma), \tag{4.77}$$

and, using the Palatini identity, we find the same equation of motion for the connection (4.64) as in the NGT theory. If the metric is symmetric, this equation tells us that the connection is the Levi-Cività connection.

To obtain the Einstein equations in the presence of matter in this formalism, one has to use more complicated techniques.[5]

4.4 The Cartan–Sciama–Kibble theory

The formalism developed so far can be used to couple matter fields that behave as tensors under GCTs. In general, the tensorial character of the matter fields under GCTs is determined from their behavior under Poincaré transformations and the only possible ambiguity is whether the field is just a tensor or a tensor density. However, this identification does not work for spinor fields, because it is based on a relation that exists only between the tensor representations of the Poincaré group and tensor representations of the diffeomorphism group. Thus, to couple fermions to gravity, we must first find out how to define spinors in a general curved spacetime.

In a classical paper,[6] [954], Weyl proposed to define spinors in *tangent space* using an orthonormal Vielbein basis $\{e^a{}_\mu\}$ as fundamental fields instead of the metric and developed a formalism that is invariant under Lorentz transformations of this Vielbein basis even if we perform a different Lorentz transformation in (the tangent space associated with) every spacetime point. Thus, in d spacetime dimensions, the $d(d+1)/2$ *off-shell* degrees of freedom of the metric (the number of independent components of a $d \times d$ symmetric matrix) are replaced by the same number of off-shell degrees of freedom of the Vielbein (the number of independent components of a generic $d \times d$ matrix minus the $d(d-1)/2$ independent local Lorentz transformations). In modern language,[7] this is a gauge theory of the Lorentz group $\mathrm{SO}(1, d-1)$ and requires the introduction of a Lorentz covariant derivative $\mathcal{D}_\mu$ and a Lorentz (spin) connection $\omega_\mu{}^{ab}$. Otherwise, the Vielbeins will describe more degrees of freedom than the metric.

However, if we want to recover GR, we do not want to introduce new fields apart from the metric (Vielbeins) and thus we have to relate the spin connection to the Vielbeins, destroying the similarity with a standard Yang–Mills theory in which the connection is the dynamical field. The natural way to relate connection and Vielbeins is through the first Vielbein postulate Eqs. (1.83) which connects the spin and the affine connections by Eq. (1.84). This does not seem to help much, because the affine connection is completely undetermined. However, metric-compatibility is automatic for spin and affine connections satisfying the first Vielbein postulate, because, by assumption, the spin connection $\omega_\mu{}^{ab}$ is antisymmetric in the indices ab, which implies $\nabla_\mu \eta_{ab} = 0$, which, with the first Vielbein postulate, implies $\nabla_\mu g_{\rho\sigma} = 0$. Therefore, the first Vielbein postulate determines the connection in terms of the Vielbein up to the torsion term. Now if we want to have as fundamental fields the Vielbeins alone, we need to impose the vanishing of torsion. In that case, the affine connection is the Levi-Cività connection $\Gamma(g)$ whose components are the Christoffel symbols Eq. (1.44) and

---

[5] See e.g. [682] and references therein. Further generalizations of the Einstein–Hilbert action are also reviewed there.

[6] A guide to the old literature on this formalism and its generalizations to include torsion is [523]. A pedagogical introduction to this formalism is [818] (see also [817]). A more recent reference is [851].

[7] The basic formalism of Yang–Mills gauge theories is developed in Appendix A and, for the Lorentz group in particular, for the present application, in Section 1.4.

then the relation Eq. (1.84) implies that the spin connection is the Cartan spin connection $\omega(e)$ given in Eq. (1.92). This case will be treated in the next section. The possibility of including torsion will be studied in Section 4.4.2.

The first Vielbein postulate can be imposed from the beginning (the second-order formalism in which the only fundamental fields are the Vielbein components) or via the spin-connection equation of motion (the first-order formalism in which both the Vielbein and the spin-connection components are independent, fundamental fields). In the first-order formalism the theory resembles more a standard Yang–Mills theory, as we will discuss in Section 4.4.4.

4.4.1 The coupling of gravity to fermions

In this section, as a warm-up exercise, we want to study the coupling of fermions to gravity using the torsionless Cartan (Levi-Cività) connection (see e.g. [187]).

Let us first summarize Weyl's recipe: to couple spinors to gravity we now replace all partial derivatives in the special-relativistic action for Lorentz (or total-)covariant derivatives by the Cartan–Levi-Cività derivatives and the Minkowski metric $\eta_{\mu\nu}$ by the general metric $g_{\mu\nu}$ or by the Vielbeins $e^a{}_\mu$ if necessary.[8] Since the Cartan spin connection cannot be expressed in terms of the metric, it is clear that the fundamental variables in this formalism will be the Vielbeins. This does not require any change in the Einstein–Hilbert action since we simply have to use

$$\frac{\delta S_{\rm EH}[e]}{\delta e^a{}_\mu} = 2\frac{\delta S_{\rm EH}[g]}{\delta g_{\rho\sigma}} e_{a\,(\rho} g_{\rho)}{}^\mu = -\frac{2}{\chi^2} e G_a{}^\mu, \tag{4.78}$$

and, correspondingly, redefine the matter energy–momentum tensor

$$T_{\text{matter}\,a}{}^\mu = \frac{c}{e}\frac{\delta S_{\text{matter}}[\varphi, e]}{\delta e^a{}_\mu}, \qquad e = \det(e^a{}_\mu) = \sqrt{|g|}. \tag{4.79}$$

Observe that, with this new definition, the energy–momentum tensor (that we can call the *Vielbein energy–momentum tensor*) does not have to be symmetric. However, we can prove that it is symmetric when the matter equations of motion hold: let us consider the variation of the matter action under a local Lorentz transformation with parameter $\sigma^{ab}(x)$, which we know leaves the Lagrangian invariant. Up to a total derivative

$$\delta_\sigma S_{\text{matter}}[\varphi, e] = \int d^dx \left\{ \frac{\delta S_{\text{matter}}}{\delta\varphi}\delta_\sigma\varphi + \frac{\delta S_{\text{matter}}}{\delta e^a{}_\mu}\delta_\sigma e^a{}_\mu \right\}. \tag{4.80}$$

Using the definition of the energy–momentum tensor and the transformation rules (assuming that φ transforms in the representation r of the Lorentz group)

$$\delta_\sigma\varphi^\alpha = \tfrac{1}{2}\sigma^{ab}\Gamma_r(M_{ab})^\alpha{}_\beta\varphi^\beta, \qquad \delta_\sigma e^a{}_\mu = \tfrac{1}{2}\sigma^{cd}\Gamma_{\rm v}(M_{cd})^a{}_b e^b{}_\mu = \sigma^a{}_\mu, \tag{4.81}$$

[8] We could ask whether this formalism should also be applied to other fields: for instance, whether we should consider the Maxwell field as a tangential vector field. The answer is that we can do it and the choice of torsionless connection that we have made ensures that there is no difference, although we gain more insight if we consider the Maxwell field as a tangential vector field. In the presence of torsion and in more general contexts this will be impossible.

we find the Bianchi identity

$$T_{\text{matter}\,[ab]} = -\frac{1}{2e}\frac{\delta S_{\text{matter}}}{\delta\varphi^\alpha}\Gamma_r(M_{ab})^\alpha{}_\beta\varphi^\beta, \tag{4.82}$$

which vanishes on-shell.

We can also use the invariance under reparametrizations of the matter action to show that the Vielbein energy–momentum tensor is covariantly conserved on-shell:

$$\nabla_\mu T_a{}^\mu = 0. \tag{4.83}$$

As for the Vielbein energy–momentum tensor, we can try to determine its form by assuming the validity of a more or less standard matter Lagrangian, namely, a standard Lagrangian whose dependence on the Vielbeins comes from two sources: the spin connection and the rest. It is easy to convince oneself by looking at simple examples that "the rest," which depends only algebraically on the Vielbeins, gives e times the canonical energy–momentum tensor when $e^a{}_\mu = \delta^a{}_\mu$:

$$eT_{\text{can}\,a}{}^\mu = -\frac{\partial\mathcal{L}_{\text{matter}}}{\partial\nabla_\mu\varphi}\nabla_a\varphi + e_a{}^\mu\mathcal{L}_{\text{matter}}. \tag{4.84}$$

The dependence of the Lagrangian through the spin connection can be computed by observing that the matter Lagrangian depends on derivatives of the Vielbeins only through the Cartan spin connection which appears in covariant derivatives of the field φ^α,

$$\mathcal{D}_\mu\varphi^\alpha = \partial_\mu\varphi^\alpha - \tfrac{1}{2}\omega_\mu{}^{ab}(e)\Gamma_r(M_{ab})^\alpha{}_\beta\varphi^\beta. \tag{4.85}$$

The contribution of these terms to the energy–momentum tensor is given by

$$\frac{\partial\mathcal{L}_{\text{matter}}}{\partial\omega_{\rho bc}}\frac{\partial\omega_{\rho bc}}{\partial e^a{}_\mu} - \partial_\nu\left(\frac{\partial\mathcal{L}_{\text{matter}}}{\partial\omega_{\rho bc}}\frac{\partial\omega_{\rho bc}}{\partial\partial_\nu e^a{}_\mu}\right). \tag{4.86}$$

Using[9]

$$\omega_{\rho bc} = 2\Delta_{\rho bc}{}^{\sigma\tau d}\,\Omega_{\sigma\tau d}, \qquad \Delta_{\rho bc}{}^{\sigma\tau d} = \tfrac{1}{2}\left\{\delta_\rho{}^\sigma e_c{}^\tau\delta_b{}^d + e_b{}^\sigma e_c{}^\tau e^d{}_\rho - \delta_\rho{}^\sigma e_b{}^\tau\delta_c{}^d\right\}, \tag{4.88}$$

we find that the contribution to the energy–momentum tensor of the spin connection is given by

$$2\frac{\partial\mathcal{L}_{\text{matter}}}{\partial\omega_{\rho bc}}\frac{\partial\Delta_{\rho bc}{}^{\sigma\tau d}}{\partial e^a{}_\mu}\Omega_{\sigma\tau d} - 2\partial_\nu\left(\frac{\partial\mathcal{L}_{\text{matter}}}{\partial\omega_{\rho bc}}\Delta_{\rho bc}{}^{\sigma\tau d}\frac{\partial\Omega_{\sigma\tau d}}{\partial\partial_\nu e^a{}_\mu}\right). \tag{4.89}$$

Since the spin connection occurs in the matter Lagrangian only via covariant derivatives of the matter field φ, it is easy to see that

$$\frac{\partial\mathcal{L}_{\text{matter}}}{\partial\omega_{\rho bc}} = -eS^{\rho bc}, \tag{4.90}$$

[9] Two similar useful relations are

$$K_{\mu ab} = \Delta_{\mu ab}{}^{\sigma\tau d}T_{\sigma\tau d}, \qquad \left\{ {\sigma \atop \mu\,\nu} \right\} g_{\sigma\rho} = -\Delta_{\nu\rho\mu}{}^{\alpha\beta\gamma}\partial_\alpha g_{\beta\gamma}. \tag{4.87}$$

and, using

$$S^{\rho bc}\Delta_{\rho bc}{}^{\sigma\tau d} = \tfrac{1}{2}\Psi^{\sigma\tau d}, \tag{4.91}$$

we obtain

$$\begin{aligned} 2\frac{\partial\mathcal{L}_{\text{matter}}}{\partial\omega_{\rho bc}}\frac{\partial\Delta_{\rho bc}{}^{\sigma\tau d}}{\partial e^a{}_\mu}\Omega_{\sigma\tau d} &= -e\Psi^{\rho\mu}{}_b\omega_{\rho a}{}^b, \\ -2\partial_\nu\left(\frac{\partial\mathcal{L}_{\text{matter}}}{\partial\omega_{\rho bc}}\Delta_{\rho bc}{}^{\sigma\tau d}\frac{\partial\Omega_{\sigma\tau d}}{\partial\partial_\nu e^a{}_\mu}\right) &= \partial_\nu(e\Psi^{\nu\mu}{}_a), \end{aligned} \tag{4.92}$$

which add up to

$$e\nabla_\nu\Psi^{\nu\mu}{}_a, \tag{4.93}$$

and so we have (observe the order of indices)

$$T_a{}^\mu = T_{\text{can}\,a}{}^\mu + \nabla_\nu\Psi^{\nu\mu}{}_a, \tag{4.94}$$

which is the relation between the Vielbein energy–momentum tensor and the canonical one. If we substract the antisymmetric part of the Vielbein energy–momentum tensor, we obtain a symmetric tensor that is conserved when the matter equations of motion hold. When $e^a{}_\mu = \delta^a{}_\mu$, this symmetric tensor becomes the Belinfante tensor, proving the relation between the Belinfante tensor and the metric (Rosenfeld) energy–momentum tensor that we mentioned in Section 2.4.1.

Example: a Dirac spinor. Let us now apply this recipe to a Dirac spinor.[10] A Dirac spinor ψ^α has only a spinorial index (which we usually hide). Thus, we are going to assume that it transforms as a spinor in tangent space and as a scalar under GCTs. Thus, the total covariant derivative ∇_μ coincides with the Lorentz-covariant derivative $\mathcal{D}_\mu$ acting on it:

$$\nabla_\mu\psi = \mathcal{D}_\mu\psi = \left(\partial_\mu - \tfrac{1}{4}\omega_\mu{}^{ab}\gamma_{ab}\right)\psi. \tag{4.95}$$

In the special-relativistic Lagrangian of the Dirac spinor Eq. (2.63) the partial derivative appears contracted with a constant gamma matrix. Now we have to distinguish between the derivative index, which is a world-tensor index, and the gamma matrix index, which is a Lorentz (tangent-space) index and, to contract both indices, we have to use a Vielbein

$$\not{\nabla}\psi = e_a{}^\mu\gamma^a\nabla_\mu\psi. \tag{4.96}$$

Finally, we also need the covariant derivative on the Dirac conjugate. The Dirac conjugate $\bar\psi_\alpha$ transforms covariantly (as opposed to the spinor ψ^α, which transforms contravariantly). Then, applying the definitions in Section 1.4,

$$\bar\psi\overleftarrow{\nabla}_\mu \equiv \nabla_\mu\bar\psi = \partial_\mu\bar\psi + \tfrac{1}{4}\omega_\mu{}^{ab}\bar\psi\gamma_{ab}. \tag{4.97}$$

With all these elements we can immediately write the action

$$S_{\text{matter}} = \int d^dx\, e\left\{\tfrac{1}{2}(i\bar\psi\not{\nabla}\psi - i\bar\psi\overleftarrow{\not{\nabla}}\psi) - m\bar\psi\psi\right\}. \tag{4.98}$$

[10] The special-relativistic Dirac spinor was studied in Section 2.4.1.

The equations of motion are the evident covariantization of the flat-space ones:

$$(i \not{\nabla} - m)\psi = 0, \tag{4.99}$$

and the spin-angular-momentum tensor $S^{\mu}{}_{ab}$ and spin–energy potential $\Psi^{\mu\nu}{}_{a}$ are identical to the ones calculated in Section 2.4.1. By varying with respect to the Vielbeins, we find the Vielbein energy–momentum tensor, which has the general form Eq. (4.94) with

$$T_{\mathrm{can}\,a}{}^{\mu} = -\frac{i}{2}\bar{\psi}\gamma^{\mu}\nabla_{a}\psi + \frac{i}{2}\nabla_{a}\bar{\psi}\gamma^{\mu}\psi + e_{a}{}^{\mu}\mathcal{L}_{\mathrm{matter}}, \tag{4.100}$$

giving

$$\begin{aligned} T_{a}{}^{\mu} = {} & -\frac{i}{2}\bar{\psi}(\gamma^{\mu}e_{a}{}^{\nu} + \gamma_{a}g^{\mu\nu})\nabla_{\nu}\psi + \frac{i}{2}\nabla_{\nu}\bar{\psi}(\gamma^{\mu}e_{a}{}^{\nu} + \gamma_{a}g^{\mu\nu})\psi \\ & + e_{a}{}^{\mu}\mathcal{L}_{\mathrm{matter}} - \frac{i}{2}\bar{\psi}\gamma^{\mu}{}_{a}\not{\nabla}\psi + \frac{i}{2}\bar{\psi}\overleftarrow{\not{\nabla}}\gamma_{a}{}^{\mu}\psi, \end{aligned} \tag{4.101}$$

which is not symmetric because of the last two terms, which vanish on-shell, as expected. This is what saves the consistency of the Einstein equation

$$G_{a}{}^{\mu} = \frac{\chi^{2}}{2}T_{a}{}^{\mu}, \tag{4.102}$$

whose l.h.s. is symmetric in the absence of torsion. This is not too different from the way in which consistency is achieved in the standard GR theory in which the l.h.s. is divergenceless (due to the contracted Bianchi identity) and the r.h.s. is divergenceless only when the matter equations of motion are satisfied.

4.4.2 *The coupling to torsion: the CSK theory*

Perhaps the simplest generalization of GR one can think of is the use of a (still metric-compatible) connection with non-vanishing torsion $T_{\mu\nu}{}^{\rho}$. Now, the torsion is a new field whose value we have to determine. The simplest possibility is to consider it a fundamental field and just include it in a generalized Einstein–Hilbert action and in the covariant derivatives acting on matter fields (minimal coupling). Then its equation of motion is determined, as usual, by varying the action with respect to it and imposing the vanishing of the variation. As we are going to see, the resulting equation of motion is algebraic and simply gives the torsion as a function of other fields. In fact, in the torsion equation of motion one can see the matter spin–energy potential $\Psi^{\mu\nu}{}_{a}$ as the source for torsion $T_{\mu\nu}{}^{a}$. This is essentially the definition of the Cartan–Sciama–Kibble (CSK) theory (reviewed in [523]; and, in a more pedagogical form, in [818]; and in the Newman–Penrose formalism in [768]).

Why should we couple intrinsic spin to torsion? The CSK theory is based on Weyl's Vielbein formalism in which there are two distinct gauge symmetries: reparametrizations and local Lorentz transformations in tangent space. Reparametrization invariance leads to the coupling of the energy–momentum tensor to the metric and, similarly, local Lorentz invariance leads to the coupling of the spin–energy potential to torsion.

In the CSK theory, torsion is not a propagating new field. Furthermore, there is no way to couple it to vector gauge potentials without breaking the gauge symmetry, which is inadmissible. However, it is possible to generalize the theory further in such a way as to have propagating torsion. The most popular way of doing it, which occurs naturally in supergravity and string theory [834], is to consider torsion as the 3-form field strength of a 2-form (*Kalb–Ramond*) field $B_{\mu\nu}$:

$$T_{\mu\nu\rho} = 3\partial_{[\mu}B_{\nu\rho]} \equiv H_{\mu\nu\rho}. \tag{4.103}$$

This particular form of torsion can be consistently coupled to gauge vector fields through the addition to the field strength of the gauge-field Chern–Simons 3-form ω_3, Eq. (A.50),

$$H_{\mu\nu\rho} = 3\partial_{[\mu}B_{\nu\rho]} + \omega_{3\,\mu\nu\rho}, \tag{4.104}$$

and modifying the gauge-transformation rule for $B_{\mu\nu}$ to make $H_{\mu\nu\rho}$ gauge-invariant. Since we will encounter this propagating torsion later on, we postpone its discussion until then. One of the reasons for why we are reviewing the CSK theory here is precisely that it constitutes an important link in the evolutionary chain that goes from GR to supergravity and superstring theories. The next link in the chain will be the gauge theories of the Poincaré and (anti-)de Sitter groups that we will also study in this chapter.

Let us first consider the generalization of the Einstein–Hilbert action in the CSK theory,

$$S_{\text{CSK}}[e^{a}{}_{\mu}, T_{\mu\nu}{}^{a}] = \frac{1}{\chi^2}\int d^d x\, eR(e,T), \tag{4.105}$$

where $R(e,T)$ is the Ricci scalar constructed from the curvature associated with the metric-compatible torsionful spin connection Eq. (1.92) or its associated affine connection given in Eq. (1.50) and is, therefore, a function of the Vielbeins and torsion. We have chosen the Vielbeins instead of the metric as the fundamental fields since the CSK theory is relevant only in the coupling of gravity to fermions because, as we have already said, the coupling of torsion to vector fields by substitution of partial derivatives for covariant derivatives necessarily breaks their gauge invariance.

We now vary the above action with respect to the Vielbeins and torsion. First, we vary with respect to the metric and connection. Using Palatini's identity Eq. (3.286), we find

$$\delta S_{\text{CSK}} = \frac{1}{\chi^2}\int d^d x\, e\big\{-G^{\alpha\beta}\delta g_{\alpha\beta} + g^{\alpha\beta}\big[\nabla_{\alpha}\delta\Gamma_{\rho\beta}{}^{\rho} - \nabla_{\rho}\delta\Gamma_{\alpha\beta}{}^{\rho} - T_{\alpha\rho}{}^{\sigma}\delta\Gamma_{\sigma\beta}{}^{\rho}\big]\big\}. \tag{4.106}$$

The covariant derivatives can be split into Levi-Cività covariant derivatives $\overset{\{\}}{\nabla}_{\mu}$, which can be integrated away, and contorsion pieces. After some calculations, we find

$$\delta S_{\text{CSK}} = \frac{1}{\chi^2}\int d^d x\, e\Big\{-G^{\alpha\beta}\delta g_{\alpha\beta} + \delta\Gamma_{\alpha\beta}{}^{\gamma}g^{\beta\delta}\,\overset{*}{T}_{\gamma\delta}{}^{\alpha}\Big\}, \tag{4.107}$$

where $\overset{*}{T}$ is the modified torsion tensor defined in Eq. (1.35). Using now Eq. (3.283), we find, at last,

$$\delta S_{\text{CSK}} = \frac{1}{\chi^2}\int d^d x\; e\Big\{-\Big[G^{\alpha\beta} - \overset{*}{\nabla}_{\mu}\overset{*}{T}{}^{\mu\alpha\beta}\Big]\delta g_{\alpha\beta} + \tfrac{1}{2}\Big[\overset{*}{T}_{\gamma}{}^{\alpha\beta} - \overset{*}{T}_{\gamma}{}^{\beta\alpha} - \overset{*}{T}{}^{\alpha\beta}{}_{\gamma}\Big]\delta T_{\alpha\beta}{}^{\gamma}\Big\}, \tag{4.108}$$

where we have also used the modified divergence $\overset{*}{\nabla}_{\mu}$ defined in Eq. (1.35).

Now we couple the pure gravity Lagrangian to the matter Lagrangian and use the definition of the Vielbein energy–momentum tensor Eq. (4.79) and the following definition of the spin–energy potential, which generalizes Eq. (2.88),

$$\Psi_{\text{matter}}{}^{\mu\nu}{}_a = -\frac{2c}{e}\frac{\delta S_{\text{matter}}}{\delta T_{\mu\nu}{}^a}, \tag{4.109}$$

to obtain the equations of the CSK theory:

$$\begin{aligned} G^{(\alpha\beta)} - \overset{*}{\nabla}_\mu \overset{*}{T}{}^{\mu(\alpha\beta)} &= \frac{\chi^2}{2} T_{\text{matter}}{}^{\alpha\beta}, \\ \tfrac{1}{2}\left[\overset{*}{T}{}_\gamma{}^{\alpha\beta} - \overset{*}{T}{}_\gamma{}^{\beta\alpha} - \overset{*}{T}{}^{\alpha\beta}{}_\gamma\right] &= \frac{\chi^2}{2}\Psi_{\text{matter}}{}^{\alpha\beta}{}_\gamma. \end{aligned} \tag{4.110}$$

We have taken into account in the l.h.s. of the first equation that only the symmetric part contributes to it, even though the r.h.s. (the Vielbein energy–momentum tensor) is not symmetric in general (we have seen that the antisymmetric part vanishes on-shell).

These equations can be rewritten in a more suggestive form: taking the modified divergence of the second equation, we find the equation

$$\overset{*}{\nabla}_\mu \overset{*}{T}{}^{\mu(\alpha\beta)} - \tfrac{1}{2}\overset{*}{\nabla}_\mu \overset{*}{T}{}^{\alpha\beta\mu} = \frac{\chi^2}{2}\overset{*}{\nabla}_\mu \Psi_{\text{matter}}{}^{\mu\alpha\beta}, \tag{4.111}$$

which, when subtracted from the first equation (4.110), gives a more elegant equation,

$$G^{\alpha\beta} = \frac{\chi^2}{2} T_{\text{can}}{}^{\alpha\beta}, \tag{4.112}$$

where we have used Eq. (1.34) and have defined the canonical energy–momentum tensor here by

$$T_{\text{cal}}{}^{\beta\alpha} = T_{\text{matter}}{}^{\alpha\beta} - \overset{*}{\nabla}_\mu \Psi_{\text{matter}}{}^{\mu\alpha\beta}. \tag{4.113}$$

This identification is evidently based on the definition of the Belinfante tensor, but we will prove that this tensor is indeed given by Eq. (4.84).

The second Eq. (4.110) can be simplified by raising the index γ and antisymmetrizing it with β:

$$\overset{*}{T}{}^{\alpha\beta\gamma} = \chi^2 S^{\gamma\beta\alpha}. \tag{4.114}$$

Now we can use this equation to rewrite the Vielbein equation (the first of Eqs. (4.110)) in a general-relativistic form. First, we take the symmetric part of the equation that relates the Einstein tensor of the torsionful connection Γ to the Einstein tensor of the Levi-Cività connection Γ, which is

$$G_{\alpha\beta}(\Gamma) = G_{\alpha\beta}[\Gamma(g)] - \tfrac{1}{2}\overset{*}{\nabla}_\mu\left[\overset{*}{T}{}_\alpha{}^\mu{}_\beta + \overset{*}{T}{}_\beta{}^\mu{}_\alpha - \overset{*}{T}{}_{\alpha\beta}{}^\mu\right] - f(T^2), \tag{4.115}$$

where $f(T^2)$ is a complicated expression that is quadratic in the torsion whose explicit form we do not need.[11] Then the Vielbein equation takes the form

$$G^{\alpha\beta}[\Gamma(g)] = \frac{\chi^2}{2} T_{\text{matter}}{}^{\alpha\beta} + f(T^2). \tag{4.116}$$

Then, by substituting Eq. (4.114) into this, we obtain

$$G^{\alpha\beta}[\Gamma(g)] = \frac{\chi^2}{2} T_{\text{matter}}{}^{\alpha\beta} + \mathcal{O}(\chi^4), \tag{4.117}$$

which coincides with Einstein's equation to order χ^2. In fact, taking into account that the order-χ^4 correction is associated with the density of intrinsic spins, only under the most extreme macroscopic conditions [523] can the CSK theory give predictions different from Einstein's, which is good. At the microscopic level, the CSK theory gives different predictions: for instance, it predicts contact interactions between fermions. These have two origins: the term quadratic in the torsion in the CSK gravity action[12] and the covariant derivatives in the matter action. All of them are of higher order in χ.

Conceptually, the CSK theory offers clear advantages over Einstein's. It allows the coupling to fermions and the relation between the canonical and Vielbein energy–momentum tensors is clarified. As we are going to see, the simplest supergravity theory ($N = 1, d = 4$) has the structure of the CSK theory for a Rarita–Schwinger spinor coupled to gravity (and torsion). Finally, we are going to see that the separation between GCTs (which can be seen as the local generalization of translations) and local Lorentz transformations suggests a reinterpretation of gravity as a gauge theory (in the Yang–Mills sense) of the Poincaré group.

Before we move on to these developments, we want to derive the complete gauge identities and Noether currents for matter coupled to gravity in the CSK theory and study the first-order formalism for it.

4.4.3 Gauge identities and Noether currents

Let us consider the action of matter minimally coupled to Vielbein and torsion $e^a{}_\mu$ and $T_{\mu\nu}{}^a$:

$$S_{\text{matter}} = \frac{1}{c}\int d^dx\, \mathcal{L}_{\text{matter}}(\varphi, \nabla\varphi, e) = \frac{1}{c}\int d^dx\, \mathcal{L}_{\text{matter}}(\varphi, \partial\varphi, e, \partial e, T). \tag{4.119}$$

(According to the minimal coupling prescription, the dependence on torsion is only through the covariant derivative.) We assume that our matter fields, generically denoted by φ, have only Lorentz indices and that only their first derivatives occur in the action. Furthermore, the fundamental fields are assumed to be $e^a{}_\mu$ and $T_{\mu\nu}{}^a$ (not $T_{\mu\nu}{}^\rho$).

[11] Actually, the second term on the l.h.s. of this equation also contains terms quadratic in the torsion that we can include in $f(T^2)$ by replacing the modified divergence $\stackrel{*}{\nabla}_\mu$ by the Levi-Cività covariant derivative $\stackrel{\{\}}{\nabla}_\mu$.

[12] Using Eq. (1.56), we can split the CSK action into a standard Einstein–Hilbert action and a piece quadratic in the torsion plus a total derivative that we can ignore:

$$S_{\text{CSK}}[e^a{}_\mu, T_{\mu\nu}{}^a] = \frac{1}{\chi^2}\int d^dx\, e\Big\{R(e) + K_\mu{}^{\mu\lambda}K_\nu{}^\nu{}_\lambda + K_{\nu\mu\rho}K^{\mu\rho\nu}\Big\}. \tag{4.118}$$

By construction, the action is exactly invariant under local Lorentz transformations and GCTs. Let us now compute the variation of the action through the variation of the fundamental fields. Following the standard procedure developed in Chapter 2, we find[13]

$$\tilde{\delta} S_{\text{matter}} = \frac{1}{c}\int d^dx \,\left\{\partial_\mu\left[\epsilon^a\left(\mathcal{L}_{\text{matter}} e_a{}^\mu - \frac{\partial \mathcal{L}_{\text{matter}}}{\partial\partial_\mu\varphi}\partial_a\varphi\right) + \frac{\partial \mathcal{L}_{\text{matter}}}{\partial\partial_\mu\varphi}\tilde{\delta}\varphi \right.\right.$$
$$\left.\left. + \frac{\partial \mathcal{L}_{\text{matter}}}{\partial\partial_\mu e^a{}_\nu}\tilde{\delta} e^a{}_\nu\right] + c\frac{\delta S_{\text{matter}}}{\delta\varphi}\delta\varphi + c\frac{\delta S_{\text{matter}}}{\delta e^a{}_\mu}\delta e^a{}_\mu + c\frac{\delta S_{\text{matter}}}{\delta T_{\mu\nu}{}^a}\delta T_{\mu\nu}{}^a\right\}. \quad (4.120)$$

The variations of the matter action with respect to the matter fields are the matter equations of motion. The variations of the matter action with respect to the geometric fields are source terms. Now, with our choice of fundamental fields, we define the spin-angular-momentum tensor $S^\mu{}_{ab}$, the spin–energy-potential tensor $\Psi^{\mu\nu}{}_a$ and the Vielbein energy–momentum tensor $T_a{}^\mu$ by

$$\frac{c}{e}\frac{\delta S_{\text{matter}}}{\delta K_\mu{}^{ab}} = -S^\mu{}_{ab}, \qquad \frac{c}{e}\frac{\delta S_{\text{matter}}}{\delta T_{\mu\nu}{}^a} = -\tfrac{1}{2}\Psi^{\mu\nu}{}_a, \qquad \frac{c}{e}\frac{\delta S_{\text{matter}}}{\delta e^a{}_\mu} = T_a{}^\mu. \quad (4.121)$$

The canonical energy–momentum tensor $T_a{}^\mu$ has an extra term due to our choice of fundamental fields:

$$T_{\text{can}\,a}{}^\mu = T_a{}^\mu - \overset{*}{\nabla}_\rho \Psi^{\rho\mu}{}_a - \tfrac{1}{2}\Psi^{\nu\rho}{}_a T_{\nu\rho}{}^\mu. \quad (4.122)$$

Now we substitute the explicit form of the variations of the fundamental fields under GCTs and local Lorentz transformations rewritten in a convenient form,

$$\begin{aligned}
\delta e^a{}_\mu &= -\mathcal{D}_\mu\epsilon^a + 2\epsilon^\nu\mathcal{D}_{[\mu}e^a{}_{\nu]} + \sigma'^a{}_b e^b{}_\mu,\\
\delta T_{\mu\nu}{}^a &= -\nabla_\mu\left(\epsilon^\lambda T_{\lambda\nu}{}^a\right) - \nabla_\nu\left(\epsilon^\lambda T_{\mu\lambda}{}^a\right) - \epsilon^\lambda\left[3R_{[\mu\nu\lambda]}{}^a + T_{\mu\nu}{}^\rho T_{\lambda\rho}{}^a\right] + \sigma'^a{}_b T_{\mu\nu}{}^b,\\
\tilde{\delta}\varphi &= \tfrac{1}{2}\sigma'^{ab}\Gamma_r\left(M_{ab}\right)\varphi + \tfrac{1}{2}\epsilon^\lambda\omega_\lambda{}^{ab}\Gamma_r\left(M_{ab}\right)\varphi,
\end{aligned} \quad (4.123)$$

where

$$\sigma'^{ab} = \sigma^{ab} - \epsilon^\mu\omega_\mu{}^{ab}. \quad (4.124)$$

After some massaging, using the Bianchi identities for the curvature, we arrive at

$$\begin{aligned}
\tilde{\delta} S = \frac{1}{c}\int\Bigg\{&\partial_\mu\Bigg\{\epsilon^a\left[\left(\mathcal{L}_{\text{matter}} e_a{}^\mu - \frac{\partial\mathcal{L}_{\text{matter}}}{\partial\nabla_\mu\varphi}\nabla_a\varphi\right) - e\,T_{\text{can}\,a}{}^\mu\right]\\
&- e\left(\overset{\{\}}{\nabla}_\rho\,\epsilon_\lambda - \epsilon^\sigma K_{\sigma\rho\lambda}\right)\left(\Psi^{\mu\rho\lambda} - \underline{\Psi}^{\mu\rho\lambda}\right)\Bigg\}\\
&+ e\epsilon^\lambda\left[\overset{*}{\nabla}_\mu\,T_{\text{can}\,\lambda}{}^\mu + T_{\lambda\mu}{}^a T_{\text{can}\,a}{}^\mu + S^\mu{}_{ab}R_{\lambda\mu}{}^{ab} - \frac{\delta S_{\text{matter}}}{\delta\varphi}\nabla_\lambda\varphi\right]\\
&+ e\sigma'^{ab}\left[T_{ab} - \frac{1}{2}\Psi^{\rho\sigma}{}_a T_{\rho\sigma b} + \frac{1}{2}\frac{\delta S_{\text{matter}}}{\delta\varphi}\Gamma_r\left(M_{ab}\right)\varphi\right]\Bigg\},
\end{aligned} \quad (4.125)$$

[13] Taking into account $\tilde{\delta}x^\mu = \epsilon^\mu$ and that local Lorentz transformations with parameter σ^{ab} act only on fields.

where we are using the notation

$$\underline{S}^{\mu}{}_{ab} = \frac{\partial \mathcal{L}_{\text{matter}}}{\partial \partial_{\mu}\varphi}\Gamma_r(M_{ab})\,\varphi, \qquad \underline{\Psi}^{\mu\rho}{}_a = -\underline{S}^{\mu\rho}{}_a + \underline{S}^{\rho\mu}{}_a + \underline{S}_a{}^{\mu\rho}. \tag{4.126}$$

Since the above variation of the action vanishes identically for arbitrary GCTs and local Lorentz transformations, we obtain four identities. The first identity just gives the expression for the canonical covariant energy–momentum tensor Eq. (4.84). The second gives the expression for the spin–energy-potential tensor Eq. (2.52). The third is the Bianchi identity associated with the invariance under GCTs,

$$\overset{*}{\nabla}_{\mu}\, T_{\text{can}\,\lambda}{}^{\mu} + T_{\lambda\mu}{}^{a} T_{\text{can}\,a}{}^{\mu} + S^{\mu}{}_{ab} R_{\lambda\mu}{}^{ab} - \frac{\delta S_{\text{matter}}}{\delta\varphi}\nabla_{\lambda}\varphi = 0, \tag{4.127}$$

that in flat, torsionless spacetime is the on-shell conservation of the energy–momentum tensor. The fourth is the Bianchi identity associated with the invariance under local Lorentz transformations,

$$T_{[ab]} - \tfrac{1}{2}\Psi^{\rho\sigma}{}_{[a|} T_{\rho\sigma|b]} + \tfrac{1}{2}\frac{\delta S_{\text{matter}}}{\delta\varphi}\Gamma_r\,(M_{ab})\,\varphi = 0, \tag{4.128}$$

which tells us that the Vielbein energy–momentum tensor in flat, torsionless, spacetime is symmetric on-shell.

As an example, we will study a Dirac spinor coupled to the Vielbein and torsion in the CSK theory, but in first-order form (Section 4.4.4).

4.4.4 *The first-order Vielbein formalism*

As we have seen, the Einstein action written in terms of Vielbeins and the spin connection with the spin connection considered as a function of the Vielbeins provides a second-order action functional of the Vielbeins that is fully equivalent to the one written in terms of the metric.

There is also a first-order action for Vielbeins and the spin connection considered as independent variables. In differential-forms language it takes the form

$$\boxed{S[e^a, \omega^{ab}] = \frac{(-1)^{d-1}}{(d-2)!}\int R^{a_1 a_2}(\omega)\wedge e^{a_3}\wedge\cdots\wedge e^{a_d}\epsilon_{a_1\cdots a_d},} \tag{4.129}$$

where $R^{a_1 a_2}$ is the curvature 2-form associated with the spin connection ω defined in Eqs. (1.81) and (1.144).

This action is equivalent[14] to the first-order Einstein–Hilbert action for the metric and an affine connection Γ related to this spin connection ω via the first Vielbein postulate. This equivalence can be seen by expanding the curvature 2-form in a Vielbein 1-form basis,

$$R^{a_1 a_2} = \tfrac{1}{2} R_{b_1 b_2}{}^{a_1 a_2} e^{b_1}\wedge e^{b_2}, \tag{4.130}$$

[14] In our conventions that action is exactly equivalent to $+\int d^d x\sqrt{|g|}\,R$.

and using

$$e^{b_1} \wedge \cdots \wedge e^{b_d} = d^d x \sqrt{|g|}\, \epsilon^{b_1 \cdots b_d} \tag{4.131}$$

and the relation between the curvatures of ω and Γ Eq. (1.85).

As mentioned before, this theory has some of the elements of a Yang–Mills gauge theory of the Lorentz group SO$(1, d-1)$ introduced in Appendix A.2.3.

1. There is an independent gauge field (the spin connection).
2. The gauge field appears through its gauge field strength (the curvature).

However, there are also very important differences, which make it completely different from a standard Yang–Mills theory.

1. The action is not quadratic in the field strength. Therefore, the equation of motion of the gauge field will be a constraint, as we are going to see. This is necessary in order to obtain Einstein's gravity theory in which the connection is not dynamical and the only degrees of freedom are those contained in the metric (or the Vielbein) which describe a spin-2 particle.
2. It is not clear how the Vielbeins should be considered. They are in principle matter in the vector representation but they do not have a standard kinetic term.
3. To recover the Einstein–Hilbert action, we have assumed the invertibility of the Vielbeins. This geometrical property cannot be explained from the gauge-theory point of view.

It is clear that gravity cannot be considered a pure gauge theory of the Lorentz group. At most, it would be a gauge theory containing "matter," which is conceptually hard to understand. Later on we will see how to overcome some of these problems by considering the gauge theory of the Poincaré group.

It is possible to find the equations of motion using differential-forms language (as in [221]). However, we prefer to reexpress the above action in components

$$\boxed{S[e^a{}_\mu, \omega_\mu{}^{ab}] = \frac{(-1)^{d-1}}{2 \cdot (d-2)!} \int d^d x\, R_{\mu_1\mu_2}{}^{a_1 a_2}(\omega) e^{a_3}{}_{\mu_3} \cdots e^{a_d}{}_{\mu_d} \epsilon_{a_1 \cdots a_d} \epsilon^{\mu_1 \cdots \mu_d}.} \tag{4.132}$$

On varying this action taking into account the analog of Palatini's identity Eq. (3.285) for the Lorentz covariant derivative $\mathcal{D}_\mu$,

$$\delta R_{\mu\nu}{}^{ab} = 2\mathcal{D}_{[\mu} \delta\omega_{\nu]}{}^{ab}, \qquad \mathcal{D}_\mu \delta\omega_\nu{}^{ab} = \partial_\mu \delta\omega_\nu{}^{ab} - \omega_\mu{}^a{}_c \delta\omega_\nu{}^{cb} - \omega_\mu{}^b{}_c \delta\omega_\nu{}^{ac}, \tag{4.133}$$

we find

$$\begin{aligned} \delta S = \frac{(-1)^{d-1}}{2 \cdot (d-2)!} \int d^d x\, \big[& 2\mathcal{D}_{\mu_1} \delta\omega_{\mu_2}{}^{a_1 a_2} e^{a_3}{}_{\mu_3} \\ & + (d-2) R_{\mu_1\mu_2}{}^{a_1 a_2} \delta e^{a_3}{}_{\mu_3} \big] e^{a_4}{}_{\mu_4} \cdots e^{a_d}{}_{\mu_d} \epsilon_{a_1 \cdots a_d} \epsilon^{\mu_1 \cdots \mu_d}. \end{aligned} \tag{4.134}$$

We first analyze the second term:

$$
\begin{aligned}
&(d-2)R_{\mu_1\mu_2}{}^{a_1a_2}\delta e^{a_3}{}_{\mu_3}e^{a_4}{}_{\mu_4}\cdots e^{a_d}{}_{\mu_d}\epsilon_{a_1\cdots a_d}\epsilon^{\mu_1\cdots\mu_d} \\
&\quad = (-1)^{d-1}3!(d-2)!\sqrt{|g|}\,R_{\mu_1\mu_2}{}^{a_1a_2}\delta e^{a_3}{}_{\mu_3}e_{a_1a_2a_3}{}^{\mu_1\mu_2\mu_3} \\
&\quad = (-1)^d 4\cdot(d-2)!\sqrt{|g|}\,G_a{}^\mu\delta e^a{}_\mu .
\end{aligned}
\tag{4.135}
$$

Now we consider the second term. We have to integrate by parts without the use of any special properties of the connection ω. We find

$$
\begin{aligned}
&2\mathcal{D}_{\mu_1}\delta\omega_{\mu_2}{}^{a_1a_2}e^{a_3\cdots a_d}{}_{\mu_3\cdots\mu_d}\epsilon_{a_1\cdots a_d}\epsilon^{\mu_1\cdots\mu_d} \\
&\quad = \big[2(d-2)\delta\omega_{\mu_1}{}^{a_1a_2}\partial_{\mu_2}e^{a_3}{}_{\mu_3} - 4\delta\omega_{\mu_1}{}^{a_1c}\omega_{\mu_2c}{}^{a_2}e^{a_3}{}_{\mu_3}\big]e^{a_4\cdots a_d}{}_{\mu_4\cdots\mu_d}\epsilon_{a_1\cdots a_d}\epsilon^{\mu_1\cdots\mu_d} \\
&\qquad + \partial_{\mu_1}\big[2\delta\omega_{\mu_2}{}^{a_1a_2}e^{a_3\cdots a_d}{}_{\mu_3\cdots\mu_d}\epsilon_{a_1\cdots a_d}\epsilon^{\mu_1\cdots\mu_d}\big] \\
&\quad = (-1)^{d-1}12\cdot(d-2)!\sqrt{|g|}\,e_{a_1a_2a_3}{}^{\mu_1\mu_2\mu_3}\delta\omega_{\mu_1}{}^{a_1a_2}\mathcal{D}_{\mu_2}e^{a_3}{}_{\mu_3} \\
&\qquad + \partial_{\mu_1}\Big[(-1)^{d-1}4(d-2)!\sqrt{|g|}\,\delta\omega_{\mu_2}{}^{a_1a_2}e_{a_1a_2}{}^{\mu_1\mu_2}\Big],
\end{aligned}
\tag{4.136}
$$

where we have used the identities

$$
\begin{aligned}
e_{\mu_3\cdots\mu_d}{}^{a_3\cdots a_d}\epsilon_{a_1\cdots a_d}\epsilon^{\mu_1\cdots\mu_d} &= (-1)^{d-1}2\cdot(d-2)!\sqrt{|g|}\,e_{a_1a_2}{}^{\mu_1\mu_2}, \\
e_{\mu_4\cdots\mu_d}{}^{a_4\cdots a_d}\epsilon_{a_1\cdots a_d}\epsilon^{\mu_1\cdots\mu_d} &= (-1)^{d-1}3!(d-3)!\sqrt{|g|}\,e_{a_1a_2a_3}{}^{\mu_1\mu_2\mu_3}, \\
2e_{[a_3}{}^{\mu_3}e_{a_1]a_2}{}^{\mu_1\mu_2} &= 3e_{a_1a_2a_3}{}^{\mu_1\mu_2\mu_3} - 2e_{[a_2}{}^{\mu_3}e_{a_3]a_1}{}^{\mu_1\mu_2}.
\end{aligned}
\tag{4.137}
$$

Assuming that the variations $\delta e^a{}_\mu$ and $\delta\omega_\mu{}^{ab}$ vanish on the boundary, we obtain the equations of motion

$$
G_a{}^\mu = 0, \qquad \mathcal{D}_{[\mu}e^a{}_{\nu]} = 0.
\tag{4.138}
$$

Now we introduce a connection $\Gamma_{\mu\nu}{}^\rho$ such that the total covariant derivative satisfies the first Vielbein postulate Eq. (1.83). As we stressed before, the connection is automatically metric-compatible and is the sum of a (Cartan) Levi-Cività part that depends only on the Vielbeins and a contorsion part. On comparing this now with Eq. (1.86), we conclude that the connection equation tells us that the torsion vanishes, which implies that the connection is just the (Cartan) Levi-Cività connection $\omega_\mu{}^{ab}(e)$ given by the standard expression Eq. (1.92). On substituting this spin connection into the Einstein tensor, we obtain the standard Einstein equations.

An interesting thing happens in $d=4$: if we replace the connection ω in the action by its self-dual part, one still obtains Einstein's equation. This observation allows one to find new variables (*Ashtekar variables*), which are used in loop quantization of gravity [55, 414].

In coupling bosonic matter (including a cosmological constant) minimally to this action one uses only Vielbeins, but it is usually not necessary to write any term containing spin connections. Therefore, only the Einstein equation would be modified in the expected way. However, if we coupled fermions, we would necessarily have to introduce terms containing the spin connection and its equation would be modified. On applying the definition of torsion, we would find that fermions generate torsion and the solution for the spin connection would be the standard spin connection plus the corresponding contorsion tensor that would be a function of the fermions. This is exactly what happens in the CSK theory[15] and in supergravity theories (see e.g. [912] and [221], where the so-called *rheonomic approach* for constructing supergravity theories which makes use of the first-order formalism is explained), for which the first-order formalism seems especially well suited since it leads to much simpler actions. Furthermore, in the first order formalism, there is an independent connection and a relation of gravity with Yang–Mills theories and a relation of supergravity with gauge theories based on supergroups can be established (see Section 4.5 and Chapter 5).

Now we will study a simple example: a Dirac spinor coupled to gravity in the first-order formalism. We are going to see that the resulting equations of motion are the same as those we would have obtained from the second order CSK theory.

Example: a Dirac spinor. The action for a Dirac spinor coupled to gravity in the first-order formalism is the sum of Eq. (4.132) and Eq. (4.98),

$$\begin{aligned} S[e,\omega,\psi] = {} & \frac{(-1)^{d-1}}{2\cdot(d-2)!\chi^2}\int d^dx\, R_{\mu_1\mu_2}{}^{a_1a_2}(\omega) e^{a_3}{}_{\mu_3}\cdots e^{a_d}{}_{\mu_d}\epsilon_{a_1\cdots a_d}\epsilon^{\mu_1\cdots\mu_d} \\ & + \int d^dx\, e\left\{\tfrac{1}{2}(i\bar\psi \not{\mathcal{D}}\psi - i\bar\psi \overleftarrow{\not{\mathcal{D}}}\psi) - m\bar\psi\psi\right\}, \end{aligned} \tag{4.139}$$

where $\mathcal{D}$ stands for the Lorentz covariant derivative.

By varying the Vierbein, spin connection, and spinor independently in the action we find, after the use of our previous results, up to total derivatives,

$$\begin{aligned} \delta S = \frac{2}{\chi^2}\int d^dx\, e\Bigg\{ & -\left[G_a{}^\mu - \frac{\chi^2}{2}T_{\mathrm{can}\,a}{}^\mu\right]\delta e^a{}_\mu + 3e_{abc}{}^{\mu\nu\rho}\left[\mathcal{D}_\nu e^c{}_\rho - \frac{\chi^2}{2}S^c{}_{\nu\rho}\right]\delta\omega_\mu{}^{ab} \\ & + \frac{\chi^2}{2}\delta\bar\psi\left[i\,\nabla\!\!\!/\,\psi - \frac{i}{2}\left(\Gamma_{\mu\nu}{}^\mu - \left\{ {\mu \atop \mu\,\nu} \right\}\right)\gamma^\nu\psi - m\psi\right] \\ & + \frac{\chi^2}{2}\left[-i\bar\psi\overleftarrow{\not{\mathcal{D}}} + \frac{i}{2}\bar\psi\gamma^\nu\left(\Gamma_{\mu\nu}{}^\mu - \left\{ {\mu \atop \mu\,\nu} \right\}\right) - m\bar\psi\right]\delta\psi\Bigg\}, \end{aligned} \tag{4.140}$$

where we have introduced an affine connection Γ such that the total covariant derivative ∇ satisfies the first Vielbein postulate, which means that it is also metric-compatible as we have explained before. Then,

$$\Gamma_{\mu\nu}{}^\mu - \left\{ {\mu \atop \mu\,\nu} \right\} = K_{\mu\nu}{}^\mu = T_{\nu\mu}{}^\mu. \tag{4.141}$$

[15] Observe that, in the first-order formalism, the Vielbein equation is the full Einstein tensor, whereas in the second-order formalism, it is only the symmetric part of the Einstein tensor. The variation of the matter action will give automatically the canonical energy–momentum tensor, since there will be no contributions from the spin connection. Thus, the first-order formalism gives us the equation $G_a{}^\mu = (\chi^2/2)T_{\mathrm{can}\,a}{}^\mu$ in just one shot.

$T_{\mathrm{can}\,a}{}^{\mu}$ is the Dirac-spinor covariant canonical energy–momentum tensor. It has the same form as in Eq. (4.100) but now the total covariant derivative uses the general connections considered here. As we have already pointed out, in the first-order formalism, the covariant canonical energy–momentum tensor is obtained by direct variation with respect to the Vielbeins:

$$\frac{\delta S_{\mathrm{matter}}}{\delta e^{a}{}_{\mu}} = e\, T_{\mathrm{can}\,a}{}^{\mu}. \tag{4.142}$$

Finally, $S^{\mu}{}_{ab}$ is the spin-angular-momentum tensor, which is totally antisymmetric and given by Eq. (2.67).

The equations of motion are

$$G_{a}{}^{\mu} = \frac{\chi^2}{2} T_{\mathrm{can}\,a}{}^{\mu}, \qquad \mathcal{D}_{\nu} e_{\rho}{}^{c} = \frac{\chi^2}{2} S^{c}{}_{\nu\rho}, \qquad i\not{\nabla}\psi - m\psi = \frac{i}{2} T_{\nu\mu}{}^{\mu} \gamma^{\nu}\psi. \tag{4.143}$$

The second equation has the solution

$$T_{\mu\nu}{}^{a} = -\chi^2 S^{a}{}_{\mu\nu}, \tag{4.144}$$

as in the CSK theory. On account of the complete antisymmetry of S, this equation implies that the r.h.s. of the third equation vanishes identically, so we are left with

$$i\not{\nabla}\psi - m\psi = 0. \tag{4.145}$$

Finally, the first equation is just the Einstein equation one obtains in the CSK theory after several manipulations. We can split it into a Riemannian part and the torsion contributions, which we know are of quartic order in χ.

As we have stressed before, the simplicity of the first-order formalism is related to the previously mentioned fact that this kind of action makes contact with the formulation of gravity as the gauge theory of the Poincaré group which we are going to study next.

4.5 Gravity as a gauge theory

In [674] MacDowell and Mansouri formulated gravity as the gauge theory of the Poincaré group and supergravity as the gauge theory of the super-Poincaré group.[16] This approach was later extended successfully to many other situations and it is interesting enough to review it briefly here because the similarities with and differences of gravity from the gauge theories of internal symmetries (some of which we have already mentioned) are manifest in this formulation. Here we will loosely follow [404, 912].

One of the differences we observed in the previous section between the first-order formalism for gravity using Vielbeins and spin connection and a pure gauge theory is that we did not have an interpretation of the Vielbeins as gauge fields. Furthermore, our intuition tells us that, if gravity can be interpreted as a gauge theory at all, it cannot be a gauge theory of the Lorentz group alone and at least gauge translations should be introduced into the game. We should then consider the gauging of the Poincaré group. It is worth stressing here that we are talking about the "Poincaré group of the tangent space." That is, at each point

[16] The earliest work on this subject is [918].

in the base manifold, which may but need not be invariant under any translational isometry, we consider inhomogeneous transformations of Lorentz vectors preserving the Minkowski metric. The relation between these gauge transformations and GCTs is one of the subtle points of this formulation of gravity.

To find the generators of the Poincaré group and their commutation relations, we can use the representation in position space (as differential operators) or alternatively we can use the following representation by $(d+1) \times (d+1)$ matrices of Poincaré transformations composed of a translation a^a and a Lorentz transformation $\Lambda^a{}_b$:

$$\begin{pmatrix} 1 \\ v'^a \end{pmatrix} = \left(\begin{array}{c|c} 1 & 0 \\ \hline a^a & \Lambda^a{}_b \end{array}\right) \begin{pmatrix} 1 \\ v^b \end{pmatrix}. \tag{4.146}$$

This representation is suggestive because of its $(d+1)$-dimensional homogeneous form. We will later see that there is a reason for its existence.

We give here again the non-vanishing commutators of the generators $\{M_{ab}, P_a\}$:

$$\begin{aligned} [M_{ab}, M_{cd}] &= -M_{eb}\Gamma_{\mathrm{v}}(M_{cd})^e{}_a - M_{ae}\Gamma_{\mathrm{v}}(M_{cd})^e{}_b, \\ [P_c, M_{ab}] &= -P_d\Gamma_{\mathrm{v}}(M_{ab})^d{}_c. \end{aligned} \tag{4.147}$$

Here $\Gamma_{\mathrm{v}}(M_{ab})^d{}_c$ is the matrix corresponding to the generator M_{ab} in the vector representation of the Lorentz group. The last commutator says that the d generators of translations P_a can be understood as the components of a Lorentz vector. Observe that P_a acts trivially on objects with Lorentz indices. It would act non-trivially on objects with a non-trivial "$(d+1)$th" index in the above representation, but by construction they do not exist.

For each generator we would introduce a gauge field: the spin connection $\omega_\mu{}^{ab}$ for the Lorentz subalgebra plus d new gauge fields for the translation subalgebra. Our theory has d Vielbein fields with Lorentz-vector indices and it is natural to try to interpret them as the gauge fields of translations and the gauge field of the Poincaré group would, tentatively, be, in some representation Γ,

$$A_\mu = \tfrac{1}{2}\omega_\mu{}^{ab}\Gamma(M_{ab}) + e_\mu{}^a\Gamma(P_a)\,. \tag{4.148}$$

Observe that, since P_a does not act on objects with Lorentz indices, the covariant derivative contains in practice only the spin connection.

If we can reproduce Einstein's theory with these elements, we could say that Einstein's theory is the pure gauge theory of the Poincaré group. We are going to see whether this is possible. First we determine the effect of gauge transformations using the standard formalism of Appendix A. If σ^{ab} and ξ^a are the infinitesimal gauge parameters of Lorentz rotations and translations, then

$$\delta\omega_\mu{}^{ab} = -\mathcal{D}_\mu\sigma^{ab}, \qquad \delta e_\mu{}^a = -\mathcal{D}_\mu\xi^a + \sigma^a{}_b e_\mu{}^b. \tag{4.149}$$

In both cases $\mathcal{D}$ stands for the gauge covariant derivative (no Levi-Cività connection is contained in it because, for the moment, we have no metric but a gauge field $e_\mu{}^a$). It is useful to compare the last expression with the effect of an infinitesimal GCT generated by the world vector ξ^μ (unrelated in principle to the Lorentz vector ξ^a):

$$\begin{aligned}
&\delta_\xi x^\mu = \xi^\mu, \\
&\delta_\xi e^a{}_\mu = -\xi^\nu \partial_\nu e^a{}_\mu - \partial_\mu \xi^\nu e^a{}_\nu = -\mathcal{D}_\mu\left(\xi^\nu e^a{}_\nu\right) + 2\xi^\nu \mathcal{D}_{[\mu} e^a{}_{\nu]} - (\xi^\nu \omega_\nu{}^a{}_b) e^b{}_\mu .
\end{aligned} \tag{4.150}$$

The covariant derivative is, again, the Poincaré (Lorentz) gauge one. The effect of an infinitesimal reparametrization is identical to the effect of a P_a gauge transformation with parameter $\xi^a = \xi^\mu e^a{}_\mu$ plus a local Lorentz transformation with parameter $\sigma^{ab} = \xi^\mu \omega_\mu{}^{ab}$ *if* $\mathcal{D}_{[\mu} e^a{}_{\nu]}$ *vanishes.*

We know that this condition is equivalent to the vanishing of torsion and we know that this constraint allows us to express $\omega_\mu{}^{ab}$ in terms of $e^a{}_\mu$. If we implement this constraint in our gauge theory, it will automatically become invariant under reparametrizations.[17] It is implemented in the first-order formalism of the previous section, where it appears as the equation of motion of $\omega_\mu{}^{ab}$.

The next step is to construct the gauge field strength:

$$\begin{aligned}
R_{\mu\nu} &= \tfrac{1}{2} R_{\mu\nu}{}^{ab} \Gamma(M_{ab}) + R_{\mu\nu}{}^a \Gamma(P_a)\,, \\
R_{\mu\nu}{}^{ab} &= 2\partial_{[\mu} \omega_{\nu]}{}^{ab} - 2\omega_{[\mu}{}^a{}_c \omega_{\nu]}{}^{cb}, \\
R_{\mu\nu}{}^a &= 2\mathcal{D}_{[\mu} e^a{}_{\nu]}.
\end{aligned} \tag{4.151}$$

The last line is identically equal to $-T_{\mu\nu}{}^a$. Thus, we have just learned that torsion can be interpreted in this formalism as the part of the gauge field strength that is associated with translations.

Now the moment to construct the action arrives. As we mentioned, in order to recover the constraint $R_{\mu\nu}{}^a$, the action has to be linear in the curvature components $R_{\mu\nu}{}^{ab}$. The requirement of Lorentz invariance also makes it very difficult to build quadratic actions (different from $\mathrm{Tr}(R \wedge {}^\star R)$, which is wrong for gravity) that are not trivial (i.e. they do not correspond to topological invariants). We are then led to the action Eq. (4.129), which we know is correct.

What have we learned by considering the gauge theory of the Poincaré group? Essentially we have given a gauge-field interpretation to Vielbeins (although we have not justified why they have to be invertible) and we have found that constraints are necessary in order to relate Poincaré gauge invariance to reparametrization invariance. The construction of the action is still rather *ad hoc*.

A slight improvement of the situation was achieved by MacDowell and Mansouri [674] (see also [227, 231, 864]), who used it to construct supergravity actions [404]. Working in four dimensions, they considered the anti-de Sitter group SO(2, 3). Upon performing a Wigner–Inönü contraction [592] (which is essentially the zero-cosmological-constant limit), this group becomes the Poincaré group ISO(1, 3) and one recovers our previous results.

[17] In supergravity formulated as the gauge theory of a supergroup the problem is how to relate supersymmetry transformations (in general super-reparametrizations in superspace) to gauge transformations associated with the supersymmetry and translation generators.

More precisely, we introduce SO(2, 3) indices $\hat{a}, \hat{b}, \cdots = -1, 0, 1, 2, 3$. The metric is $\hat{\eta}^{\hat{a}\hat{b}} = \mathrm{diag}(++---)$ and the algebra so(2, 3) can be written in the general form

$$\left[\hat{M}_{\hat{a}\hat{b}}, \hat{M}_{\hat{c}\hat{d}}\right] = -\hat{\eta}_{\hat{a}\hat{c}}\hat{M}_{\hat{b}\hat{d}} - \hat{\eta}_{\hat{b}\hat{d}}\hat{M}_{\hat{a}\hat{c}} + \hat{\eta}_{\hat{a}\hat{d}}\hat{M}_{\hat{b}\hat{c}} + \hat{\eta}_{\hat{b}\hat{c}}\hat{M}_{\hat{a}\hat{d}}. \tag{4.152}$$

To perform the contraction, we need to introduce a dimensional parameter. This is, naturally, g, the gauge coupling constant in gauged $d = 4$, $N = 2$ supergravity. g is related to the AdS$_4$ radius R and to the cosmological constant Λ by

$$R = 1/g = \sqrt{-3/\Lambda}. \tag{4.153}$$

We can now perform a $1 + 4$ splitting of the indices $\hat{a} = (-1, a)$, $a = 0, 1, 2, 3$, to interpret this algebra from the point of view of the Lorentz subalgebra so(1, 3). On defining

$$\hat{M}_{ab} = M_{ab}, \qquad \hat{M}_{a-1} = -g^{-1}P_a, \tag{4.154}$$

we can rewrite the AdS$_4$ algebra as follows:

$$\begin{aligned} [M_{ab}, M_{cd}] &= -\eta_{ac}M_{bd} - \eta_{bd}M_{ac} + \eta_{ad}M_{bc} + \eta_{bc}M_{ad}, \\ [P_c, M_{ab}] &= -2P_{[a}\eta_{b]c}, \qquad [P_a, P_b] = -g^2 M_{ab}. \end{aligned} \tag{4.155}$$

Taking now the limit $g \to 0$, we recover the Poincaré algebra.

We could equally well have started with the four-dimensional de Sitter group SO(1, 4). The difference is that, instead of having an extra timelike direction (which we have denoted with a -1 index), we have an extra spacelike direction (which we would denote with a 4 index). The two spaces (and groups) are related by analytic continuation $x^{-1} \to x^4$ and, in the contraction of the extra dimension, we would find that the sign of the cosmological constant is reversed ($g \to ig$). We will use a general notation and point out where differences between the two groups could arise. However, one should keep in mind that only the anti-de Sitter space is a good background for QFT and only its group can consistently be supersymmetrized.

The gauge theory of the AdS$_4$ group is just a particular case of the general construction in Appendix A.2.3. We can also perform the contraction in the gauge field and curvature. First, we split the indices in the connection and then we rescale the gauge fields inversely to the generators:

$$\begin{aligned} \hat{\omega}_\mu &= \tfrac{1}{2}\hat{\omega}_\mu{}^{\hat{a}\hat{b}}\Gamma\left(\hat{M}_{\hat{a}\hat{b}}\right) = \tfrac{1}{2}\hat{\omega}_\mu{}^{ab}\Gamma\left(\hat{M}_{ab}\right) + \hat{\omega}_\mu{}^{a,-1}\Gamma\left(\hat{M}_{a,-1}\right) \\ &= \tfrac{1}{2}\omega_\mu{}^{ab}\Gamma(M_{ab}) + e^a{}_\mu\Gamma(P_a), \end{aligned} \tag{4.156}$$

where

$$\hat{\omega}_\mu{}^{ab} = \omega_\mu{}^{ab}, \qquad \hat{\omega}_\mu{}^{a,-1} = -ge^a{}_\mu. \tag{4.157}$$

In this scheme, linear momentum and Vierbeins are on the same footing as the rest of the generators and gauge fields. This is obviously due to the semisimple nature of the AdS$_4$ group. There is some resemblance between this structure and the idea of grand unification in particle physics, although there are also obvious differences.

We can also split and rescale the curvature components, expressing everything in terms of Lorentz tensors:

$$\hat{R}_{\mu\nu}{}^{ab} = R_{\mu\nu}{}^{ab} + 2g^2 e^{[a}{}_{\mu} e^{b]}{}_{\nu}, \qquad \hat{R}_{\mu\nu}{}^{a,-1} = 2g\mathcal{D}_{[\mu} e^{a}{}_{\nu]}. \tag{4.158}$$

Now we can address again the construction of a quadratic action for this group. To have diffeomorphism invariance, the Lagrangian has to be a 4-form that we can integrate over a four-dimensional manifold and, therefore, the exterior product of two curvature terms $R \wedge R$. We now have to saturate the so(2, 3) indices. If we did it with the Killing metric, we would have manifest SO(2, 3) invariance but the Lagrangian would be a total derivative, as in any Yang–Mills theory. Thus, we have to give up explicit SO(2, 3) invariance. We have to keep Lorentz invariance, though, and with the only two invariant tensors of the Lorentz group ($\eta_{ab}, \epsilon_{abcd}$) we can build two terms:

$$\hat{R}^{ab} \wedge \hat{R}^{cd} \epsilon_{abcd}, \qquad \hat{R}^{a,-1} \wedge \hat{R}^{b,-1} \eta_{ab}. \tag{4.159}$$

The second term is not invariant under parity and for this reason it is discarded. The inclusion of this term would also introduce torsion and it would also lead to the existence of non-invertible Vierbeins (see the discussion in [404]).

The first term can also be given an SO(2,3)-invariant origin [864]: by introducing a constant vector $V^{\hat{a}} = \eta^{\hat{a}}{}_{-1}$ it can be written using the invariant tensor $\hat{\epsilon}$ and we obtain the action

$$S = \alpha \int \hat{R}^{\hat{a}\hat{b}} \wedge \hat{R}^{\hat{c}\hat{d}} V^{\hat{e}} \hat{\epsilon}_{\hat{a}\hat{b}\hat{c}\hat{d}\hat{e}}. \tag{4.160}$$

This is only formally SO(2, 3)-invariant because the vector would change under AdS_4 transformations. Nevertheless, this form of the action is very suggestive.

On expanding this action in terms of Lorentz tensors, we have

$$S = \alpha \int d^4x\, R_{\mu\nu}{}^{ab} R_{\rho\sigma}{}^{cd} \epsilon_{abcd} \epsilon^{\mu\nu\rho\sigma} - 16g^2\alpha \int d^4x\, e\big[R(e,\omega) + 6g^2\big]. \tag{4.161}$$

The first term is a total derivative (proportional to the Euler characteristic, a topological invariant) that does not contribute to the equations of motion and the second term is the first-order Einstein–Hilbert action with cosmological constant $\Lambda = -3g^2$. In the $g \to 0$ limit (provided that $\alpha \sim g^{-2}$) we recover the usual Einstein–Hilbert action plus a topological term. Observe that the variation of the action under P_a gauge transformations is proportional to torsion terms and, thus, vanishes on-shell.

This is a very attractive result, which, however, leaves some questions unanswered, such as the reason for the invertibility of the Vierbein and the value of the vector $V^{\hat{a}}$. A possible solution has recently been proposed by Wilczek in [955].

To finish this section, we should mention that the gauge approach has been extended to larger groups such as the full d-dimensional affine group. A comprehensive review on these developments is [524].

4.6 Teleparallelism

In this section we would like to give a short introduction to relativistic theories of gravity based on *teleparallelism*, i.e. theories in which there is a well-defined notion of parallelism of vectors defined at different points. In GR and other generalizations based on the Riemannian or Riemann–Cartan geometry, gravity, described by the metric or Vielbein fields, is characterized by a curvature and, therefore, parallel transport is path-dependent and there is no such well-defined (path-independent) notion of parallelism. Teleparallelism is based on the Weitzenböck geometry and the Weitzenböck connection $W_{\mu\nu}{}^{\rho}$ described in Section 1.4.1, which has identically vanishing curvature (but non-vanishing torsion[18]).

These theories are interesting for several reasons: first of all, GR can be viewed as a particular theory of teleparallelism and, thus, teleparallelism could be considered at the very least as a different point of view that can lead to the same results. Of course, there are teleparallel theories different from and even inconsistent with GR. Second, in this framework, one can define an energy–momentum tensor for the gravitational field that is a true tensor under all GCTs. This is the reason why teleparallelism was reconsidered[19] by Møller in 1961 [702] when he was studying the problem of defining an energy–momentum *tensor* for the gravitational field [700, 701]. The idea was taken over by Pellegrini and Plebański in [761] that constructed the general Lagrangian for these theories. The third reason why these theories are interesting is that they can be seen as gauge theories of the translation group [237, 521] (not the full Poincaré group) and, thus, they give an alternative interpretation of GR.

The basic field in these theories is the Vielbein $e^{a}{}_{\mu}$. This field has d^2 independent components, while the metric has only $d(d+1)/2$. The extra independent components that the Vielbein field has are those of an antisymmetric $d \times d$ tensor, such as the electromagnetic-field-strength tensor $F_{\mu\nu}$, and that is why Einstein thought that these theories could describe gravitation and electromagnetism in a unified way. In the standard Vielbein formalism (Weyl's), the extra $d(d-1)/2$ independent components of the Vielbein field are removed by introducing local Lorentz invariance, with a Lorentz connection that is not an independent field but is built out of the Vielbeins. Here, we are not interested *a priori* in having this local invariance and, in principle, we will construct only theories that are invariant under GCTs and global Lorentz transformations. Thus, as we will see, these theories describe in general more degrees of freedom than just those of a graviton.

The construction of the Lagrangian of these theories is fairly simple: we look for terms that have the required invariances and are, at most, quadratic in derivatives of the Vielbeins. The elementary building blocks are the Ricci rotation coefficients $\Omega_{\mu\nu}{}^{a} = \partial_{[\mu} e^{a}{}_{\nu]}$ that transform as tensors (2-forms) under GCTs and as vectors under (global) Lorentz transformations. In the context of the Weitzenböck geometry, the $-2\Omega_{\mu\nu}{}^{a}$s are the components of the torsion of Weitzenböck connection and, since there is no curvature tensor available, it is only natural to construct the Lagrangian using them.

In any dimension there are three terms with the required properties (they transform as densities under GCTs and as scalars under Lorentz transformations, being quadratic in first partial derivatives of the Vielbeins): the *Weitzenböck invariants* $I_1, \ldots, I_3$,

$$I_1 = e\,\Omega_{\mu\nu\rho}\Omega^{\mu\nu\rho}, \qquad I_2 = e\,\Omega_{\mu\nu\rho}\Omega^{\rho\nu\mu}, \qquad I_3 = e\,\Omega_{\mu\rho}{}^{\rho}\Omega^{\mu}{}_{\sigma}{}^{\sigma}. \tag{4.162}$$

[18] It is possible to have a non-trivial theory with vanishing curvature and torsion if the non-metricity tensor does not vanish. In [720] a theory equivalent to GR based on this geometry was constructed.

[19] Teleparallelism had originally been considered by Einstein, who studied it as a unified theory of gravitation and electromagnetism in [359–363] (see also [651]) until [368] showed that the particular theory considered by him was inconsistent.

There is another invariant I_4, which is quadratic only in $d=4$:

$$I_4 = \epsilon^{\mu_1\cdots\mu_{d-3}\nu_1\nu_2\nu_3}\Omega_{\mu_1\rho_1}{}^{\rho_1}\Omega_{\mu_2\rho_2}{}^{\rho_2}\cdots\Omega_{\mu_{d-3}\rho_{d-3}}{}^{\rho_{d-3}}\Omega_{\nu_1\nu_2\nu_3}, \tag{4.163}$$

but it is not invariant under parity transformations (a further requirement) and it is usually not considered. Also, e by itself is another invariant (a cosmological-constant term) that we will not consider. Observe that all the Weitzenböck invariants involve the inverse Vielbeins $e_a{}^\mu$ and are, therefore, highly non-linear in the Vielbeins.

The general teleparallel Lagrangian of Pellegrini and Plebański [761] is the integral of a linear combination of the Weitzenböck invariants with arbitrary coefficients:

$$\mathcal{L}_{\mathrm{T}} = \sum_{i=1}^{3} c^i I_i. \tag{4.164}$$

Only two of them are really independent since we can choose the overall normalization.

This general Lagrangian, written in differential-forms language to relate it to the Poincaré gauge theory of gravity which is customarily written in it (see e.g. [524]), is known as the *Rumpf Lagrangian* [815] (see also [482, 704]).

There are other ways to parametrize this Lagrangian, for instance, by splitting Ω_{abc} into several pieces ${}^{(1)}\Omega$, ${}^{(2)}\Omega$, and ${}^{(3)}\Omega$ (*tentor, trator, and axitor*, respectively, [482]). First we define

$$\begin{aligned} v_a &\equiv \Omega_{ab}{}^b, & {}^{(2)}\Omega_{abc} &= \frac{2}{1-d}\eta_{a[b}v_{c]}, \\ {}^{(3)}\Omega_{abc} &= \Omega_{[abc]}, & {}^{(1)}\Omega_{abc} &= \Omega_{abc} - {}^{(2)}\Omega_{abc} - {}^{(3)}\Omega_{abc}. \end{aligned} \tag{4.165}$$

Then, a Lagrangian equivalent to Pellegrini and Plebański's is [482]

$$\mathcal{L}_{\mathrm{T}} = e\,\Omega^{abc}\sum_{i=1}^{3} a_i\,{}^{(i)}\Omega_{abc}. \tag{4.166}$$

The relation between these two parametrizations is

$$a_1 = c_1 + \tfrac{1}{2}c_2, \qquad a_2 = c_1 + \tfrac{1}{2}c_2 + \frac{d-1}{2}c_3, \qquad a_3 = c_1 - c_2. \tag{4.167}$$

Another parametrization based on v_a, the tensors

$$a^{a_1\cdots a_{d-3}} = \frac{1}{3!}\epsilon^{a_1\cdots a_{d-3}b_1b_2b_3}\Omega_{b_1b_2b_3}, \qquad t_{abc} = \Omega_{a(bc)} - {}^{(2)}\Omega_{a(bc)}, \tag{4.168}$$

and the invariants v^2, t^2, and a^2 can be found in [522].

4.6.1 The linearized limit

Now, our goal is to try to understand which kind of theories are those defined by the Lagrangian Eq. (4.164). First, we observe with Møller [702] that, for $c_1 = 1, c_2 = 2$, and $c_4 = -4$, this Lagrangian is identical (up to total derivatives) to the Einstein–Hilbert

Lagrangian, and, therefore, gives the vacuum Einstein equations.[20] The Lagrangian turns out to be invariant under not just global but also local Lorentz transformations and the only degrees of freedom left are (we know it) those of the graviton. For general values of the parameters, the analysis is more complicated and it is convenient to start by studying the linear limit. To this end, we split the Vielbeins into their vacuum (Minkowski) values plus perturbations. Working in Cartesian coordinates for simplicity, we write

$$e^a{}_\mu = \delta^a{}_\mu + A^a{}_\mu. \tag{4.169}$$

For the inverse Vielbeins, we have

$$e_a{}^\mu = \delta_a{}^\mu - \delta_b{}^\mu \delta_a{}^\nu A^b{}_\nu + \mathcal{O}(A^2). \tag{4.170}$$

To this order we can unambiguously trade curved and flat indices and the above formula can be rewritten

$$e_a{}^\mu = \delta_a{}^\mu - A^\mu{}_a + \mathcal{O}(A^2), \qquad A^\mu{}_a \equiv \delta_b{}^\mu \delta_a{}^\nu A^b{}_\nu. \tag{4.171}$$

The metric perturbation that we have called $h_{\mu\nu}$ in previous chapters is given by the symmetric part of A at lowest order:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} + \mathcal{O}(A^2), \quad h_{\mu\nu} \equiv 2A_{(\mu\nu)}, \quad b_{\mu\nu} \equiv 2A_{[\mu\nu]}, \quad A_{\mu\nu} \equiv \delta_{a\mu} A^a{}_\mu. \tag{4.172}$$

With these definitions is straightforward to obtain, up to total derivatives, the linear limit of action for the Lagrangian density Eq. (4.164):

$$\begin{aligned} S_{\mathrm{T}}[h,b] = \int d^dx \; \Big\{ & \frac{1}{16}(2c_1+c_2)\partial_\mu h_{\nu\rho}\partial^\mu h^{\nu\rho} - \frac{1}{16}(2c_1+c_2-c_3)\partial_\mu h_{\nu\rho}\partial^\nu h^{\mu\rho} \\ & - \frac{1}{8}c_3 \partial_\mu h \partial_\nu h^{\nu\mu} + \frac{1}{16}c_3(\partial h)^2 - \frac{1}{16}[4c_1 + 2(c_2+c_3)]\partial_\mu h_{\nu\rho}\partial^\rho b^{\nu\mu} \\ & + \frac{1}{16}\partial_\mu b_{\nu\rho}\partial^\mu b^{\nu\rho} - \frac{1}{16}(2c_1 - 3c_2 - c_3)\partial_\mu b_{\nu\rho}\partial^\rho b^{\nu\mu} \Big\}. \end{aligned} \tag{4.173}$$

The first four terms are familiar to us: up to coefficients, they are the same terms as those that appear in the Fierz–Pauli Lagrangian Eq. (3.84). The last two terms are also well known: up to coefficients, they are exactly those that appear in the Lagrangian of the Kalb–Ramond 2-form field, which we still have not seen. The terms in the third line represent a coupling (already at the linear level) between these two fields.

Now, it is clear that it is not possible to recover solutions of the vacuum Fierz–Pauli theory if the coupling terms have a non-zero coefficient: a non-vanishing h field is a source for a non-vanishing b field and vice-versa. Thus, the only theories which we expect to be phenomenologically viable are those in the family

$$2c_1 + c_2 + c_3 = 0. \tag{4.174}$$

[20] In fact, this theory is sometimes referred to as the *teleparallel equivalent of GR*.

Table 4.1. In this table we give the values of the parameters c_i in the general Lagrangian of Pellegrini and Plebański Eq. (4.164) for several theories: GR, the viable models, the Yang–Mills-type model (YM), and the von der Heyde model (vdH) [926] (which is one of the viable ones with $\lambda = 0$).

	GR	Viable	YM	vdH
c_1	1	$2-\lambda$	2	2
c_2	2	2λ	0	0
c_3	4	-4	0	-4

Furthermore, both in the Kalb–Ramond and in the Fierz–Pauli cases, for certain choices of the coefficients, the action has a gauge invariance,

$$\delta_\epsilon h_{\mu\nu} = -2\partial_{(\mu}\epsilon_{\nu)}, \qquad \delta_\eta b_{\mu\nu} = 2\partial_{[\mu}\eta_{\nu]}, \tag{4.175}$$

whose existence is crucial for its consistent quantization. We also expect that only when these gauge invariances are present will the theory be consistent. It turns out that all these conditions are simultaneously met: let us eliminate c_3 using Eq. (4.174) and then, calling $c_2 = 2\lambda$, the action can be rewritten in the form

$$S_{\mathrm{T}}[h, b] = \frac{c_1+\lambda}{2} S_{\mathrm{FP}}[h] + \frac{c_1-\lambda}{2} S_{\mathrm{KR}}[b], \tag{4.176}$$

where $S_{\mathrm{FP}}[h]$ is the Fierz–Pauli action given in Eq. (3.84) and $S_{\mathrm{KR}}[b]$ is the Kalb–Ramond action

$$S_{\mathrm{KR}}[b] = \int d^d x \, \frac{1}{12} H^2, \qquad H_{\mu\nu\rho} \equiv 3\partial_{[\mu} b_{\nu\rho]}, \qquad H^2 = H_{\mu\nu\rho} H^{\mu\nu\rho}. \tag{4.177}$$

For $c_1 = \lambda$ ($c_2 = 2\lambda,\ c_3 = -4\lambda$), the Kalb–Ramond Lagrangian disappears. Up to an overall normalization constant and a total derivative, this teleparallel Lagrangian is completely equivalent to the Einstein–Hilbert Lagrangian, as we mentioned before. For $c_1 = -\lambda$ the Fierz–Pauli Lagrangian disappears and only the Kalb–Ramond Lagrangian remains. This theory does not describe gravity. If we are always going to keep the Fierz–Pauli Lagrangian, then it makes sense to set $c_1 = 2-\lambda$ ($c_3 = -4,\ c_2 = 2\lambda$) and keep the one-parameter family of actions

$$S_{\mathrm{T}}[h, b] = S_{\mathrm{FP}}[h] + (1-\lambda) S_{\mathrm{KR}}[b], \tag{4.178}$$

which represent viable models of gravity (in the sense that they fulfill the above requirements) based on teleparallelism (see Table. 4.1). The case $\lambda = 0$ is the model proposed in [926].

Of course, we know that the full non-linear theory will be consistent only if additional conditions are satisfied. In particular, we know from our results in Chapter 3 that the quantization of the spin-2 field $h_{\mu\nu}$ will be consistent only if it couples to the total energy–momentum tensor, the sum of the spin-2 energy–momentum tensor and the Kalb–Ramond energy–momentum tensor, although the presence of the Kalb–Ramond field could modify this result. Checking that this is (or not) the case in the above family of theories requires an expansion to order $\mathcal{O}(A^3)$ that would be interesting to do.

It is amusing to compare these results with the linearized limit of the low-energy string effective action (see Chapter 15). The linearized actions are identical, except for the presence of the dilaton in the string case. However, the non-linear actions are quite different: in the string case, we simply have standard gravity coupled to matter (the Kalb–Ramond field) that appears only quadratically (at least to lowest order in α'), whereas, in the teleparallel case, the Kalb–Ramond field should also appear non-linearly in the full action.

It is possible to view the theories of teleparallelism as gauge theories of the group of translations [237, 521] with the Vielbeins playing the role of gauge vectors, but we will not enter into this interesting aspect.

5
$N = 1, 2, d = 4$ supergravities

In the previous chapter, we introduced increasingly complex theories of gravity, starting from GR, to accommodate fermions and we saw that the generalizations of GR that we had to use could be thought of as gauge theories of the symmetries of flat spacetime.

A very important development of the last few decades has been the discovery of supersymmetry and its application to the theory of fundamental particles and interactions. This symmetry relating bosons and fermions can be understood as the generalization of the Poincaré or AdS groups which are the symmetries of our background spacetime to the *super-Poincaré* or *super*-AdS (super-)groups which are the symmetries of our background *superspacetime*, a generalization of standard spacetime that has fermionic coordinates.

It is natural to construct generalizations of the standard gravity theories that can be understood as gauge theories of the (super-)symmetries of the background (vacuum) superspacetime. These generalizations are the *supergravity* (SUGRA) theories. Given that the kind of fermions that one can have depends critically on the spacetime dimension, the SUGRA theories that one can construct also depend critically on the spacetime dimension. Furthermore, one can extend the standard bosonic spacetime in different ways by including more than one (N) set of fermionic coordinates. This gives rise to additional supersymmetries relating them and, therefore, to supersymmetric field theories and SUGRA theories with N supersymmetries. The latter are also known as *extended* SUGRAs (SUEGRAs).

There is, thus, a large variety of supergravities, but not infinitely large, because the gauging of supersymmetries with $N > 8$ in $d = 4$ dimensions or $N = 1$ in $d = 11$ needs the inclusion of more than one graviton and/or fields of spin higher than 2, which we do not know how to couple consistently.

We are going to study SUGRA theories because they provide an interesting extension of the ideas we have reviewed so far and because the effective-field theories that describe the behavior of superstrings at low energies are SUGRA theories.

Supersymmetry and SUGRA have been developed over the last several years and are currently the object of extensive work, so we cannot give here a complete review of any of these subjects. There are excellent books and reviews that cover most of the basic aspects, though, for instance [150, 404, 912, 915, 916, 946, 948]. Reference [828] contains reprints of many of the original articles on SUGRA.

Our goal in this chapter will be to introduce some of the concepts that we will use later on, profiting from and extending the material we have studied so far. Our method will be to construct the simplest four-dimensional SUGRA theories ($N=1, d=4$ Poincaré and AdS supergravities) by gauging the corresponding supergroups and studying them separately. We will then study the two simplest four-dimensional SUEGRA theories ($N=2, d=4$ Poincaré and AdS supergravities) since they illustrate important ideas we will make use of later. Our conventions for tensors, gamma matrices, and spinors are explained in Chapter 1 and Appendix B, respectively.

5.1 Gauging $N=1, d=4$ superalgebras

Just as the $d=4$ Poincaré group can be constructed by exponentiation of the Poincaré algebra, the $N=1, d=4$ super-Poincaré group can be constructed by exponentiation of the $N=1, d=4$ super-Poincaré superalgebra. This superalgebra is an extension of the Poincaré algebra with (bosonic) generators P_a and M_{ab} by one set of anti-Hermitian fermionic generators Q^α (the *supersymmetry generators* or *supersymmetry charges*) that transform as Majorana[1] spinors under Poincaré transformations, so they have four components and

$$[Q^\alpha, M_{ab}] = \Gamma_{\mathrm{s}}(M_{ab})^\alpha{}_\beta Q^\beta, \tag{5.1}$$

while the commutator with P_a vanishes. To complete all the relations of the superalgebra, we need to give the commutator of two Q^αs. Actually, in a superalgebra, one has to give the *anticommutator* of fermionic generators (that is the difference from the bosonic ones) and the (anti)commutation relations have to satisfy a super-Jacobi identity, which takes the same form as the standard Jacobi identity but with commutators replaced by anticommutators whenever two fermionic generators are involved and with a relative sign between the terms related to the permutation of fermionic generators. The anticommutation relation that satisfies the super-Jacobi identities is[2]

$$\{Q^\alpha, Q^\beta\} = i\left(\gamma^a \mathcal{C}^{-1}\right)^{\alpha\beta} P_a. \tag{5.2}$$

The non-vanishing commutation relations for the $N=1, d=4$ superalgebra are

$$\boxed{\begin{aligned}
[M_{ab}, M_{cd}] &= -M_{eb}\Gamma_{\mathrm{v}}(M_{cd})^e{}_a - M_{ae}\Gamma_{\mathrm{v}}(M_{cd})^e{}_b,\\
[P_a, M_{bc}] &= -P_e\Gamma_{\mathrm{v}}(M_{bc})^e{}_a,\\
[Q^\alpha, M_{ab}] &= \Gamma_{\mathrm{s}}(M_{ab})^\alpha{}_\beta Q^\beta,\\
\{Q^\alpha, Q^\beta\} &= i\left(\gamma^a \mathcal{C}^{-1}\right)^{\alpha\beta} P_a.
\end{aligned}} \tag{5.3}$$

[1] The need for Majorana representations is associated with the anti-Hermiticity of the generators. In $d=4$ Majorana and Weyl spinors are equivalent and the superalgebra can be written in terms of Weyl spinors only (see e.g. [946]).

[2] Since our convention for Hermitian conjugation of fermionic objects is $(ab)^\dagger = +b^\dagger a^\dagger$, the structure constants have to be purely imaginary here. We are using a purely imaginary representation of the gamma matrices with a purely imaginary charge-conjugation matrix, hence the factor i.

This is the superalgebra that one has to gauge in order to construct $N = 1, d = 4$ Poincaré supergravity. However, to follow Section 4.5, we prefer to start from the supersymmetrized version of the AdS_4 algebra and then perform a Wigner–Inönü contraction. To supersymmetrize it, we need to add consistently a set of fermionic supersymmetry generators to those of the bosonic algebra $\hat{M}_{\hat{a}\hat{b}}$. To have consistency, the fermionic generators have to transform as AdS_4 Majorana spinors, which, as discussed in Appendix B, have four real (or purely imaginary) components. Denoting them by $\hat{Q}^\alpha$, we find the following (anti)commutation relations for the AdS_4 superalgebra:

$$\boxed{\begin{aligned} \left[\hat{M}_{\hat{a}\hat{b}}, \hat{M}_{\hat{c}\hat{d}}\right] &= -\hat{M}_{\hat{e}\hat{b}}\Gamma_{\mathrm{v}}\left(\hat{M}_{\hat{c}\hat{d}}\right)^{\hat{e}}{}_{\hat{a}} - \hat{M}_{\hat{a}\hat{e}}\Gamma_{\mathrm{v}}\left(\hat{M}_{\hat{c}\hat{d}}\right)^{\hat{e}}{}_{\hat{b}}, \\ \left[\hat{Q}^\alpha, \hat{M}_{\hat{a}\hat{b}}\right] &= \Gamma_{\mathrm{s}}\left(\hat{M}_{\hat{a}\hat{b}}\right)^\alpha{}_\beta \hat{Q}^\beta, \\ \{Q^\alpha, Q^\beta\} &= \left[\Gamma_{\mathrm{s}}\left(\hat{M}^{\hat{a}\hat{b}}\right)\hat{\mathcal{C}}^{-1}\right]^{\alpha\beta}\hat{M}_{\hat{a}\hat{b}}. \end{aligned}} \tag{5.4}$$

An infinitesimal transformation generated by this superalgebra is

$$\hat{\Lambda} \equiv \tfrac{1}{2}\hat{\sigma}^{\hat{a}\hat{b}}\hat{M}_{\hat{a}\hat{b}} + \bar{\hat{\epsilon}}_\alpha \hat{Q}^\alpha, \tag{5.5}$$

where $\hat{\sigma}^{\hat{a}\hat{b}} = -\hat{\sigma}^{\hat{b}\hat{a}}$ is the infinitesimal parameter of an SO(2, 3) transformation and $\hat{\epsilon}^\alpha$, an anticommuting Majorana spinor, is the infinitesimal parameter of a supersymmetry transformation. The bar indicates Dirac conjugation.

To construct theories that are invariant under *local* infinitesimal transformations ($\hat{\sigma}^{\hat{a}\hat{b}} = \hat{\sigma}^{\hat{a}\hat{b}}(x)$, $\bar{\hat{\epsilon}}_\alpha = \bar{\hat{\epsilon}}_\alpha(x)$), we need to introduce a gauge field $\hat{A}_\mu$,

$$\hat{A}_\mu \equiv \tfrac{1}{2}\hat{\omega}_\mu{}^{\hat{a}\hat{b}}\hat{M}_{\hat{a}\hat{b}} + \bar{\hat{\psi}}_{\mu\,\alpha}\hat{Q}^\alpha, \tag{5.6}$$

whose components are the standard bosonic SO(2, 3) connection $\hat{\omega}_\mu{}^{\hat{a}\hat{b}}$ from which we will obtain the Lorentz connection $\omega_\mu{}^{ab}$ and the Vierbein $e^a{}_\mu$ that will describe the graviton. It also contains a new fermionic field: the Rarita–Schwinger field $\bar{\hat{\psi}}_{\mu\,\alpha}$, which has a vector index and a spinor index. This field describes a particle of spin $\frac{3}{2}$; the *gravitino*, which is the supersymmetric partner of the graviton, related to it by supersymmetry transformations, and other excitations, which should be eliminated if there is enough gauge symmetry in its action (as is the case).

By construction, the action of an infinitesimal transformation of the gauge field is the supercovariant derivative of $\hat{\Lambda}(x)$,

$$\delta\hat{A}_\mu = \partial_\mu\hat{\Lambda} + [\hat{\Lambda}, \hat{A}_\mu]. \tag{5.7}$$

On expanding the commutator (which should be understood as the anticommutator between the fermionic generators), we find the following transformation laws for the component fields under local SO(2, 3) transformations and supersymmetry transformations:

$$
\begin{aligned}
\delta_{\hat\sigma}\hat\omega_\mu{}^{\hat a\hat b} &= \hat{\mathcal{D}}_\mu\hat\sigma^{\hat a\hat b}, & \delta_{\hat\sigma}\bar{\hat\psi}_\mu &= -\bar{\hat\psi}_\mu\left[\tfrac{1}{2}\hat\sigma^{\hat a\hat b}\Gamma_{\mathrm{s}}\left(\hat M_{\hat a\hat b}\right)\right],\\
\delta_{\hat\epsilon}\hat\omega_\mu{}^{\hat a\hat b} &= -2\bar{\hat\epsilon}\Gamma_{\mathrm{s}}\left(\hat M_{\hat a\hat b}\right)\hat\psi_\mu, & \delta_{\hat\epsilon}\bar{\hat\psi}_\mu &= \hat{\mathcal{D}}_\mu\hat\epsilon,
\end{aligned}
\tag{5.8}
$$

where $\mathcal{D}$ is the Lorentz SO(1,3) covariant derivative defined in Chapter 1.

The supercurvature is defined by

$$
\hat R_{\mu\nu}(\hat A) \equiv 2\partial_{[\mu}\hat A_{\nu]} - [\hat A_\mu, \hat A_\nu],
\tag{5.9}
$$

and, by expanding it and decomposing it into bosonic and fermionic components, we find

$$
\hat R_{\mu\nu}{}^{\hat a\hat b}(\hat A) = \hat R_{\mu\nu}{}^{\hat a\hat b}(\hat\omega) - 2\bar{\hat\psi}_{[\mu}\Gamma_{\mathrm{s}}\left(\hat M_{\hat a\hat b}\right)\hat\psi_{\nu]}, \qquad \bar{\hat R}_{\mu\nu\,\alpha}(\hat A) = 2\hat{\mathcal{D}}_{[\mu}\bar{\hat\psi}_{\nu]\,\alpha}.
\tag{5.10}
$$

Having the supercurvature components, we can now proceed to construct an action that has to be invariant under GCTs, local Lorentz transformations, parity transformations, and local supersymmetry transformations without the use of any metric. The requirement of invariance under local supersymmetry transformations is more difficult to impose and we will have to check it explicitly afterwards. The other requirements imply that the action has to be of the form

$$
S[\hat A] = \alpha\int d^4x\,\left[\hat R_{\mu\nu}{}^{ab}\hat R_{\rho\sigma}{}^{cd}\epsilon_{abcd} + \beta\bar{\hat R}_{\mu\nu\,\alpha}\,(\gamma_5)^\alpha{}_\beta\hat R_{\rho\sigma}{}^\beta\right]\epsilon^{\mu\nu\rho\sigma}.
\tag{5.11}
$$

We now want to rewrite this action in terms of component Poincaré fields and in terms of the parameter g whose zero limit gives the Wigner–Inönü contraction. First we study it in the superalgebra. Defining

$$
\hat M_{ab} \equiv M_{ab}, \quad \hat M_{a,-1} \equiv -g^{-1}P_a, \quad \hat Q^\alpha \equiv g^{-\frac{1}{2}}Q^\alpha,
\tag{5.12}
$$

the AdS_4 superalgebra takes the form

$$
\begin{aligned}
[M_{ab}, M_{cd}] &= -M_{eb}\Gamma_{\mathrm{v}}(M_{cd})^e{}_a - M_{ae}\Gamma_{\mathrm{v}}(M_{cd})^e{}_b,\\
[P_a, M_{bc}] &= -P_e\Gamma_{\mathrm{v}}(M_{bc})^e{}_a, \qquad [P_a, P_b] = -g^2M_{ab},\\
\{Q^\alpha, Q^\beta\} &= -2\left[\Gamma_{\mathrm{s}}\left(\hat M^{a,-1}\right)\hat{\mathcal{C}}^{-1}\right]^{\alpha\beta}P_a + g\left[\Gamma_{\mathrm{s}}\left(\hat M^{ab}\right)\right]^{\alpha\beta}M_{ab}.\\
[Q^\alpha, M_{ab}] &= \Gamma_{\mathrm{s}}(M_{ab})^\alpha{}_\beta Q^\beta, \qquad [Q^\alpha, P_a] = -g\Gamma_{\mathrm{s}}\left(\hat M_{a,-1}\right)^\alpha{}_\beta Q^\beta.
\end{aligned}
\tag{5.13}
$$

In the $g\to 0$ limit we recover the $N=1, d=4$ Poincaré superalgebra using, for instance, the representation of AdS_4 gamma matrices

$$
\hat\gamma_a = i\gamma_a\gamma_5, \quad \hat\gamma_{-1} = \gamma_5, \quad \hat{\mathcal{C}} = \mathcal{C} = i\gamma_0.
\tag{5.14}
$$

The infinitesimal transformation parameters and gauge fields are also split and rescaled as follows:

$$
\begin{aligned}
\hat\omega_\mu{}^{ab} &= \omega_\mu{}^{ab}, & \hat\omega_\mu{}^{a,-1} &= ge^a{}_\mu, & \hat\psi_\mu &= g^{\frac{1}{2}}\psi_\mu,\\
\hat\sigma^{ab} &= \sigma^{ab}, & \hat\sigma^{a,-1} &= g\sigma^a, & \hat\epsilon &= g^{\frac{1}{2}}\epsilon.
\end{aligned}
\tag{5.15}
$$

In terms of these variables, the SO(2, 3) and supersymmetry transformations take the forms

$$
\begin{aligned}
\delta_\sigma e^a{}_\mu &= \mathcal{D}_\mu \sigma^a + \sigma^a{}_b e^b{}_\mu, & \delta_\epsilon e^a{}_\mu &= -i\bar{\epsilon}\gamma^a\psi_\mu, \\
\delta_\sigma \omega_\mu{}^{ab} &= \mathcal{D}_\mu \sigma^{ab} + 2g^2 e^{[a}{}_\mu \sigma^{b]}, & \delta_\epsilon \omega_\mu{}^{ab} &= -2g\bar{\epsilon}\gamma^{ab}\psi_\mu, \\
\delta_\sigma \bar{\psi}_\mu &= -\bar{\psi}_\mu\left(\tfrac{1}{4}\sigma^{ab}\gamma_{ab}\right) - \frac{ig}{2}\bar{\psi}_\mu\sigma^a\gamma_a, & \delta_\epsilon \bar{\psi}_\mu &= \mathcal{D}_\mu\epsilon - \frac{ig}{2}\gamma_\mu\epsilon,
\end{aligned}
\tag{5.16}
$$

and the components of the supercurvature are given by

$$
\begin{aligned}
\hat{R}_{\mu\nu}{}^{ab} &= R_{\mu\nu}{}^{ab}(\omega) + 2g^2 e^{[a}{}_\mu e^{b]}{}_\nu + g\bar{\psi}_{[\mu}\gamma^{ab}\psi_{\nu]}, \\
\hat{R}^{a,-1}_{\mu\nu} &= -g\left(T_{\mu\nu}{}^a - i\bar{\psi}_{[\mu}\gamma^a\psi_{\nu]}\right), \\
\hat{R}_{\mu\nu} &= 2g^{\frac{1}{2}}\left(\mathcal{D}_{[\mu}\psi_{\nu]} - \frac{ig}{2}\gamma_{[\mu}\psi_{\nu]}\right).
\end{aligned}
\tag{5.17}
$$

On substituting these components into the action, we find that the right normalization of the Einstein–Hilbert term in the action requires $\alpha = -1/(16g^2\chi^2)$. Furthermore, the explicit[3] terms quartic in fermions drop out from the action (after Fierzing and massaging of some terms) if $\beta = -8i$. This is the value that will also make the action supersymmetry-invariant. The result is the action for $N = 1, d = 4$ AdS_4 SUGRA,

$$
\boxed{S[e^a{}_\mu, \omega_\mu{}^{ab}, \psi_\mu] = \frac{1}{\chi^2}\int d^4x\, e\left[R(e,\omega) + 6g^2 + 2e^{-1}\epsilon^{\mu\nu\rho\sigma}\bar{\psi}_\mu\gamma_5\gamma_\nu\hat{\mathcal{D}}_\rho\psi_\sigma\right],}
\tag{5.18}
$$

which, in the $g \to 0$ limit, gives the action for $N=1$, $d=4$ Poincaré SUGRA [315, 403]:

$$
\boxed{S[e^a{}_\mu, \omega_\mu{}^{ab}, \psi_\mu] = \frac{1}{\chi^2}\int d^4x\, e\left[R(e,\omega) + 2e^{-1}\epsilon^{\mu\nu\rho\sigma}\bar{\psi}_\mu\gamma_5\gamma_\nu\mathcal{D}_\rho\psi_\sigma\right].}
\tag{5.19}
$$

These are first-order actions in which, as indicated, the fundamental variables are the Vielbein, spin connection, and gravitino field. Thanks to our experience with the CSK theory,[4] we know that, when we solve the spin connection equation of motion, which is

[3] Later we will see that the on-shell spin connection contains terms quadratic in the fermions, so the action contains implicitly terms quartic in fermions, just as in the CSK theory.

[4] We can interpret these actions as the CSK theory coupled to gravitino fields. However, there is more to it, because the consistency of the gravitino field theory requires its action to be invariant under the gauge transformations (in flat spacetime) $\delta\psi_\mu = \partial_\mu\epsilon(x)$ in order to decouple unwanted spins. When we couple the gravitino to gravity, consistency requires that the Vierbeins also transform under these fermionic transformations (otherwise, that gauge symmetry is broken), which become the local supersymmetry transformations. In this way local supersymmetry does not reduce any further the number of degrees of freedom (graviton plus gravitino). The non-trivial part is the transformation of the Vierbeins under supersymmetry. We could have tried to arrive at the $N = 1, d = 4$ supergravity action from the linearized action which is just the sum of the Fierz–Pauli action and the Rarita–Schwinger action, decoupled, by asking for consistent interaction and following the Noether method as we did in Chapter 3. Then, the full supersymmetry transformations should arise as the consistency requirement.

purely algebraic, we are going to find that there is torsion proportional to some expression quadratic in fermions, making the $\hat{R}_{\mu\nu}{}^{a,-1}(\hat{A})$ components of the supercurvature vanish. Substituting the torsion into the action will give rise to terms that are quartic in fermions.

In what follows we are going to study these actions, their equations of motion, and their symmetries separately. The most efficient way to do it is to treat them in the so-called 1.5-order formalism: we consider that we have solved the equation of motion of the spin connection and we have substituted its solution back into the action, but we do not do it explicitly, keeping the action in its first-order form. Then, in varying over the two remaining fundamental fields (the Vierbein and gravitino), we use the chain rule, varying over the spin connection first. That variation is its equation of motion, which has been solved, and simply vanishes. In this way, many calculations are greatly simplified.

We are going to make this study as self-contained as possible and, thus, we will repeat some of the general points explained in this introductory section.

5.2 $N = 1, d = 4$ (Poincaré) supergravity

The fields of $N = 1, d = 4$ supergravity are the Vierbein and the gravitino $\{e^a{}_\mu, \psi_\mu\}$. The gravitino is a vector of Majorana (real) spinors. The action is written in a first-order form, in which the spin connection $\omega_\mu{}^{ab}$ is also considered as an independent field and the action contains only first derivatives. We rewrite the action here for convenience, setting $\chi = 1$:

$$\boxed{S[e^a{}_\mu, \omega_\mu{}^{ab}, \psi_\mu] = \int d^4x\, e\big[R(e,\omega) + 2e^{-1}\epsilon^{\mu\nu\rho\sigma}\bar{\psi}_\mu\gamma_5\gamma_\nu\mathcal{D}_\rho\psi_\sigma\big].} \tag{5.20}$$

Here $\mathcal{D}_\mu$ is the Lorentz-covariant derivative (rather than the completely covariant derivative, which we denote as usual by ∇_μ),

$$\mathcal{D}_\mu\psi_\nu = \partial_\mu\psi_\nu - \tfrac{1}{4}\omega_\mu{}^{ab}\gamma_{ab}\psi_\nu, \qquad \nabla_\mu\psi_\nu = \mathcal{D}_\mu\psi_\nu - \Gamma_{\mu\nu}{}^\rho\psi_\rho, \tag{5.21}$$

and

$$R(e,\omega) = e_a{}^\mu e_b{}^\nu R_{\mu\nu}{}^{ab}(\omega), \tag{5.22}$$

where $R_{\mu\nu}{}^{ab}(\omega)$ is the Lorentz curvature of the Lorentz connection $\omega_\mu{}^{ab}$, Eq. (1.81).

As usual, to obtain the second-order action we solve the spin-connection equation of motion and substitute the solution for $\omega_\mu{}^{ab}$ in terms of $e^a{}_\mu$ and ψ_μ back into the first-order action. The spin-connection equation of motion is

$$\frac{\delta S}{\delta\omega_\mu{}^{ab}} = 3!e_{abc}{}^{\mu\nu\rho}\left(\mathcal{D}_\nu e^c{}_\rho + \frac{i}{2}\bar{\psi}_\nu\gamma^c\psi_\rho\right) = 0. \tag{5.23}$$

This equation implies that the expression in brackets, antisymmetrized in ν and ρ, is zero. Looking at Eq. (1.86), we see that there is torsion in this theory and it is given by[5]

$$T_{\mu\nu}{}^a = i\bar{\psi}_\mu\gamma^a\psi_\nu. \tag{5.24}$$

[5] The bilinear $\bar{\psi}_\mu\gamma^a\psi_\nu$ is automatically antisymmetric in $\mu\nu$.

Furthermore, we see that the solution to the new equation is just that the Lorentz connection consists of two pieces: the one that solves the standard equation $\mathcal{D}_{[\mu} e^a{}_{\nu]} = 0$, which we denote by $\omega_\mu{}^{ab}(e)$ because it is completely determined by the Vierbein, and the contorsion tensor $K_\mu{}^{ab}$, which depends on the gravitino through the torsion. It is convenient to write the solution as follows:

$$\omega_{abc} = -\Omega_{abc} + \Omega_{bca} - \Omega_{cab}, \quad \Omega_{\mu\nu}{}^a = \Omega_{\mu\nu}{}^a(e) + \tfrac{1}{2} T_{\mu\nu}{}^a, \quad \Omega_{\mu\nu}{}^a(e) = \partial_{[\mu} e^a{}_{\nu]}. \quad (5.25)$$

The other two equations of motion that the first-order action gives are

$$\begin{aligned} \frac{\delta S}{\delta e^a{}_\mu} &= -2e\left[G_a{}^\mu - 2T_{\mathrm{can}\,a}{}^\mu(\psi)\right] = 0, \\ T_{\mathrm{can}\,a}{}^\mu(\psi) &= \frac{1}{2e}\epsilon^{\rho\mu\sigma\nu}\bar{\psi}_\rho\gamma_5\gamma_a\mathcal{D}_\sigma\psi_\nu, \\ \frac{\delta S}{\delta\bar{\psi}_\mu} &= 4\epsilon^{\mu\nu\rho\sigma}\left[\gamma_5\gamma_\nu\mathcal{D}_\rho\psi_\sigma + \tfrac{1}{4}T_{\nu\rho}{}^a\gamma_5\gamma_a\psi_\sigma\right] = 0, \end{aligned} \quad (5.26)$$

where we have used

$$\mathcal{D}_{[\mu}\gamma_{\nu]} = -\tfrac{1}{2} T_{\mu\nu}{}^a \gamma_a. \quad (5.27)$$

The second-order equations of motion follow from the substitution of Eq. (5.25) into the first-order ones.

The action Eq. (5.20) and equations of motion are manifestly invariant under

general coordinate transformations,

$$\delta_\xi x^\mu = \xi^\mu, \qquad \delta_\xi e^a{}_\mu = -\xi^\nu\partial_\nu e^a{}_\mu - \partial_\mu\xi^\nu e^a{}_\nu, \qquad \delta_\xi\psi_\mu = -\xi^\nu\partial_\nu\psi_\mu - \partial_\mu\xi^\nu\psi_\nu, \quad (5.28)$$

and local Lorentz transformations,

$$\delta_\sigma e^a{}_\mu = \sigma^a{}_b e^b{}_\mu, \qquad \delta_\sigma\psi_\mu = \tfrac{1}{2}\sigma^{ab}\gamma_{ab}\psi_\mu, \quad (5.29)$$

where $\sigma^{ab} = -\sigma^{ba}$. On top of this, if we eliminate the spin connection as an independent field by substituting the solution of its equation of motion, there is invariance under

local $N = 1$ supersymmetry transformations:

$$\delta_\epsilon e^a{}_\mu = -i\bar{\epsilon}\gamma^a\psi_\mu, \qquad \delta_\epsilon\psi_\mu = \mathcal{D}_\mu\epsilon. \quad (5.30)$$

This requires some explanation. The first-order action is also invariant under the same transformations supplemented by the supersymmetry transformation of the spin connection. In the second-order formalism, the supersymmetry variation of the spin connection is completely different and can be found by varying Eq. (5.25) with respect to the Vierbein and gravitino:

$$\delta_\epsilon\omega_\mu{}^{ab} = -i\bar{\epsilon}\gamma_\mu\psi^{ab} + i\bar{\epsilon}\gamma^a\psi^b{}_\mu - i\bar{\epsilon}\gamma^b\psi_\mu{}^a, \qquad \psi_{\mu\nu} \equiv \mathcal{D}_{[\mu}\psi_{\nu]}. \quad (5.31)$$

One may think that the gauging of the supersymmetry algebra should give us the first-order supersymmetry transformation rule for the spin connection, but it does not: it just gives $\delta_\epsilon \omega_\mu{}^{ab} = 0$. Nevertheless, to check the invariance of the action in the 1.5-order formalism we do not need this variation, as we are going to see.

Let us check the invariance of the action Eq. (5.20) under these transformations in the 1.5-order formalism. This is not a complicated calculation if we construct the right setup, which is the general setup explained in Chapter 2 for theories that are invariant under local symmetries. There we showed that a given theory would be invariant up to total derivatives under a local transformation if a certain *gauge identity* was satisfied by its equations of motion. Thus, all we have to do is to identify the gauge identity that has to be satisfied in this case by the Vierbein and gravitino equations of motion.

Under a general variation of the fields, the $N = 1, d = 4$ SUGRA action Eq. (5.20) transforms as follows:

$$\delta S = \int d^4x \left[\frac{\delta S}{\delta e^a{}_\mu} \delta e^a{}_\mu + \frac{\delta S}{\delta \omega_\mu{}^{ab}} \delta \omega_\mu{}^{ab} + \delta \bar{\psi}_\mu \frac{\delta S}{\delta \bar{\psi}_\mu} \right]. \tag{5.32}$$

Here the variations are only with respect to explicit appearances of each field in the first-order action. The variation of the second-order action would be obtained by applying the chain rule to the variation with respect to the spin connection, using Eq. (5.25). However, these additional terms are proportional to the equation of motion of the spin connection $\delta S / \delta \omega_\mu{}^{ab}$, which we have assumed is satisfied (the 1.5-order formalism). Thus, the term containing $\delta \omega_\mu{}^{ab}$ will always vanish (for any kind of variation) because it is proportional to that equation of motion and we need only vary *explicit* appearances of the Vierbein and gravitino in the first-order action Eq. (5.20),

$$\delta S = \int d^4x \left[\frac{\delta S}{\delta e^a{}_\mu} \delta e^a{}_\mu + \delta \bar{\psi}_\mu \frac{\delta S}{\delta \bar{\psi}_\mu} \right]. \tag{5.33}$$

Consider now the local supersymmetry transformations Eqs. (5.30). On substituting into the above the explicit form of these transformations and integrating by parts the partial derivative in

$$\mathcal{D}_\mu \bar{\epsilon} = \overline{\mathcal{D}_\mu \epsilon} = \partial_\mu \bar{\epsilon} + \tfrac{1}{4} \bar{\epsilon} \omega_\mu{}^{ab} \gamma_{ab}, \tag{5.34}$$

we obtain, up to total derivatives,

$$\delta_\epsilon S = \int d^4x \, \bar{\epsilon} \left[-i \frac{\delta S}{\delta e^a{}_\mu} \gamma^a \psi_\mu - \mathcal{D}_\mu \frac{\delta S}{\delta \bar{\psi}_\mu} \right]. \tag{5.35}$$

The theory will be locally supersymmetric, then, if

$$\mathcal{D}_\mu \frac{\delta S}{\delta \bar{\psi}_\mu} = -i \frac{\delta S}{\delta e^a{}_\mu} \gamma^a \psi_\mu, \tag{5.36}$$

which will be, at the same time, the supersymmetry gauge identity. Let us prove it:

$$\begin{aligned} \mathcal{D}_\mu \frac{\delta S}{\delta \bar{\psi}_\mu} = {} & 4\epsilon^{\mu\nu\rho\sigma} \gamma_5 (\mathcal{D}_\mu \gamma_\nu) \mathcal{D}_\rho \psi_\sigma + 4\epsilon^{\mu\nu\rho\sigma} \gamma_5 \gamma_\nu \mathcal{D}_\mu \mathcal{D}_\rho \psi_\sigma \\ & + \epsilon^{\mu\nu\rho\sigma} \gamma_5 \gamma_a \mathcal{D}_\mu T_{\nu\rho}{}^a \psi_\sigma + \epsilon^{\mu\nu\rho\sigma} \gamma_5 \gamma_a T_{\nu\rho}{}^a \mathcal{D}_\mu \psi_\sigma, \end{aligned} \tag{5.37}$$

where we have used $\mathcal{D}_\mu \gamma_a = 0$. Using Eq. (5.27) in the first term on the r.h.s. of the above equation, we obtain minus two times the last term. In the second term we first use the Ricci identity for the anticommutator of Lorentz-covariant derivatives, then expand the product of gammas in antisymmetrized products $\gamma^{(3)}$ and $\gamma^{(1)}$, reexpress the $\gamma^{(3)}$ in terms of $\gamma^{(1)}\gamma_5$ and the antisymmetric symbol, and, finally, use the identity

$$G_a{}^\mu = -\tfrac{3}{2} g_{abc}{}^{\mu\nu\rho} R_{\nu\rho}{}^{bc}. \tag{5.38}$$

We keep the third term as it is and obtain the total result

$$\mathcal{D}_\mu \frac{\delta S}{\delta \bar{\psi}_\mu} = 2ei\, G_a{}^\mu \gamma^a \psi_\mu - \epsilon^{\mu\nu\rho\sigma} \gamma_5 \gamma_a T_{\mu\nu}{}^a \mathcal{D}_\rho \psi_\sigma + \epsilon^{\mu\nu\rho\sigma} \left[R_{\mu\nu\rho}{}^a + \mathcal{D}_\mu T_{\nu\rho}{}^a \right] \gamma_5 \gamma_a \psi_\sigma . \tag{5.39}$$

The first term is one of the two we want. The second term is equal to the other term we want, due to the Fierz identity

$$\left(\bar{\psi}_\nu \gamma_5 \gamma_a \mathcal{D}_\rho \psi_\sigma\right)(\gamma^a \psi_\mu) = -\tfrac{1}{2} (\bar{\psi}_\nu \gamma^a \psi_\mu)(\gamma_a \gamma_5 \mathcal{D}_\rho \psi_\sigma). \tag{5.40}$$

The expression in brackets vanishes due to the Bianchi identity[6]

$$R_{[\mu\nu\rho]}{}^a + \mathcal{D}_{[\mu} T_{\nu\rho]}{}^a = 0, \tag{5.44}$$

and this proves the supersymmetry gauge identity.

5.2.1 Local supersymmetry algebra

An important check to be performed is the confirmation that we have on-shell closure of the $N = 1$ supersymmetry algebra on the fields. Let us first consider the Vierbein. Using the supersymmetry rules ($\mathcal{D}_\mu \epsilon = \nabla_\mu \epsilon$), it is easy to obtain

$$[\delta_{\epsilon_1}, \delta_{\epsilon_2}] e^a{}_\mu = -\nabla_\mu \xi^a, \tag{5.45}$$

where ξ^a is the bilinear

$$\xi^a = -i \bar{\epsilon}_1 \gamma^a \epsilon_2. \tag{5.46}$$

The effect of the GCT generated by $\xi^\mu = \xi^a e_a{}^\mu$ can be rewritten in this form:

$$\delta_\xi e^a{}_\mu = -\nabla_\mu \xi^a - \xi^\nu T_{\mu\nu}{}^a - \xi^\nu \omega_\nu{}^a{}_b e^b{}_\mu . \tag{5.47}$$

---

[6] This identity can be related to the standard Bianchi identity as follows. First,

$$\mathcal{D}_\mu T_{\nu\rho}{}^a = \nabla_\mu T_{\nu\rho}{}^a - \Gamma_{\mu\nu}{}^\lambda T_{\rho\lambda}{}^a + \Gamma_{\mu\rho}{}^\lambda T_{\nu\lambda}{}^a . \tag{5.41}$$

Antisymmetrizing and using the definition of torsion $\Gamma_{[\mu\nu]}{}^\rho = -\frac{1}{2} T_{\mu\nu}{}^\rho$ gives

$$\mathcal{D}_\mu T_{\nu\rho}{}^a = \nabla_{[\mu} T_{\nu\rho]}{}^a + T_{[\mu\nu}{}^\lambda T_{\rho]\lambda}{}^a . \tag{5.42}$$

Finally,

$$R_{[\mu\nu\rho]}{}^a + \mathcal{D}_{[\mu} T_{\nu\rho]}{}^a = R_{[\mu\nu\rho]}{}^a + \nabla_{[\mu} T_{\nu\rho]}{}^a + T_{[\mu\nu}{}^\lambda T_{\rho]\lambda}{}^a , \tag{5.43}$$

which vanishes on account of the usual Bianchi identity Eq. (1.30).

Thus, using the value of the torsion field in this theory, we find

$$[\delta_{\epsilon_1}, \delta_{\epsilon_2}]e^a{}_\mu = \left(\delta_\xi + \delta_\sigma + \delta_\epsilon\right)e^a{}_\mu, \tag{5.48}$$

where

$$\sigma^a{}_b = \xi^\nu \omega_\nu{}^a{}_b, \qquad \epsilon = \xi^\mu \psi_\mu. \tag{5.49}$$

The same algebra is realized on all the fields of the theory.

5.3 $N = 1, d = 4$ AdS supergravity

The simplest $N = 1, d = 4$ Poincaré supergravity theory that we have just described can be generalized in essentially two ways: adding $N = 1$ supersymmetric matter or generalizing the Lorentz connection. Adding certain matter supermultiplets sometimes produces enhancement of supersymmetry and in this way one obtains extended supergravities. We will review $N = 2, d = 4$ (*gauged* and *ungauged*) supergravity later.

The only generalizations of the four-dimensional Poincaré group which are usually studied are the four-dimensional (anti-)de Sitter groups $dS_4 = SO(1, 4)$ and $AdS_4 = SO(2, 3)$. Of these, only AdS_4 is compatible with consistent supergravity. We have obtained at the beginning of this chapter the action for $N = 1, d = 4$ AdS supergravity in the first-order form

$$\boxed{S[e^a{}_\mu, \omega_\mu{}^{ab}, \psi_\mu] = \int d^4x\, e\Big[R(e, \omega) + 6g^2 + 2e^{-1}\epsilon^{\mu\nu\rho\sigma}\bar{\psi}_\mu\gamma_5\gamma_\nu\hat{\mathcal{D}}_\rho\psi_\sigma\Big],} \tag{5.50}$$

where

$$\hat{\mathcal{D}}_\mu = \mathcal{D}_\mu - \frac{ig}{2}\gamma_\mu \tag{5.51}$$

is the AdS_4-covariant derivative and $\mathcal{D}_\mu$ is the Lorentz-covariant derivative in the spinor representation.

This theory contains a negative cosmological constant proportional to the square of the Wigner–Inönü parameter g, $\Lambda = -3g^2$. The vacuum will be anti-de Sitter spacetime.

The equation of motion for $\omega_\mu{}^{ab}$ takes the same form as in the $g = 0$ (Poincaré) case and therefore has the same solution, Eq. (5.25). The other two equations of motion suffer g-dependent modifications:

$$\begin{aligned}
\frac{\delta S}{\delta e^a{}_\mu} &= -2e\big[G_a{}^\mu - 3g^2 e_a{}^\mu - 2T_{\mathrm{can}\,a}{}^\mu\big] = 0,\\
T_{\mathrm{can}\,a}{}^\mu &= \frac{1}{2e}\epsilon^{\rho\mu\sigma\nu}\bar{\psi}_\rho\gamma_5\gamma_a\hat{\mathcal{D}}_\sigma\psi_\nu - \frac{ig}{2e}\epsilon^{\mu\nu\rho\sigma}\bar{\psi}_\nu\gamma_5\gamma_{\rho a}\psi_\sigma,\\
\frac{\delta S}{\delta\bar{\psi}_\mu} &= 4\epsilon^{\mu\nu\rho\sigma}\Big[\gamma_5\gamma_\nu\hat{\mathcal{D}}_\rho\psi_\sigma + \tfrac{1}{4}T_{\nu\rho}{}^a\gamma_5\gamma_a\psi_\sigma\Big] = 0.
\end{aligned} \tag{5.52}$$

The torsion term can be shown to vanish on-shell using Fierz identities.[7]

[7] This is also true in the Poincaré case.

This theory is invariant under local Lorentz transformations and GCTs. Furthermore, it is invariant under local supersymmetry transformations,

$$\delta_\epsilon e^a{}_\mu = -i\bar{\epsilon}\gamma^a\psi_\mu, \qquad \delta_\epsilon\psi_\mu = \hat{\mathcal{D}}_\mu\epsilon. \tag{5.53}$$

To prove it, one has to prove the corresponding generalization of the Poincaré supersymmetry gauge identity

$$\hat{\mathcal{D}}_\mu \frac{\delta S}{\delta\bar{\psi}_\mu} = -i\frac{\delta S}{\delta e^a{}_\mu}\gamma^a\psi_\mu. \tag{5.54}$$

We find

$$\begin{aligned}\hat{\mathcal{D}}_\mu \frac{\delta S}{\delta\bar{\psi}_\mu} &= \mathcal{D}_\mu\left(\frac{\delta S}{\delta\bar{\psi}_\mu}\right)_{g=0} - \frac{ig}{2}\gamma_\mu\left(\frac{\delta S}{\delta\bar{\psi}_\mu}\right)_{g=0} \\ &\quad + \mathcal{D}_\mu\left(-2ig\epsilon^{\mu\nu\rho\sigma}\gamma_5\gamma_{\nu\rho}\psi_\sigma\right) - \frac{ig}{2}\gamma_\mu\left(-2ig\epsilon^{\mu\nu\rho\sigma}\gamma_5\gamma_{\nu\rho}\psi_\sigma\right),\end{aligned} \tag{5.55}$$

where we have simplified the gravitino equation of motion by using the fact that, on-shell (*Fierzing*),

$$\epsilon^{\mu\nu\rho\sigma}\gamma_5\gamma_a T_{\nu\rho}{}^a\psi_\sigma = 0. \tag{5.56}$$

The $g = 0$ supersymmetry gauge identity can be used for the first term. The last term gives the cosmological-constant term in the Einstein equation. Thus, we need only check that the second and third terms (linear in g) give the two g-dependent pieces of the gravitino energy–momentum tensor, which can be combined into a single term. By expanding the third term we obtain a term that cancels out the second, a torsion term that vanishes due to the above identity, and a term

$$ig\epsilon^{\mu\nu\rho\sigma}T_{\mu\nu}{}^a\gamma_5\gamma_a\gamma_\rho\psi_\sigma, \tag{5.57}$$

which, upon Fierzing, gives the right result.

5.3.1 *Local supersymmetry algebra*

On-shell we find

$$[\delta_{\epsilon_1}, \delta_{\epsilon_2}] = \delta_\xi + \delta_\sigma + \delta_\epsilon, \tag{5.58}$$

with

$$\xi^a = -i\bar{\epsilon}_1\gamma^a\epsilon_2, \qquad \sigma^a{}_b = \xi^\nu\omega_\nu{}^a{}_b + g\bar{\epsilon}_1\gamma^a{}_b\epsilon_2, \qquad \epsilon = \xi^\mu\psi_\mu. \tag{5.59}$$

5.4 Extended supersymmetry algebras

As we said in the introduction, one can generalize spacetime by adding one or more sets of fermionic coordinates. The corresponding supersymmetry algebras have one or more (N) sets of supersymmetry generators that we denote by adding an index $i = 1, \ldots, N$, $Q^{i\,\alpha}$. For $N > 1$ they are called *extended supersymmetry algebras*. In this section we are going to introduce them in $d = 4$ and in the next two sections we will study two SUEGRA theories based on the simplest extended superalgebras.

It is convenient for our purposes to start by generalizing the $N = 1, d = 4$ AdS superalgebra to $N > 1$. It turns out that to have a consistent superalgebra, one is *forced* to introduce further bosonic generators $T^{ij} = -T^{ji}$, which generate SO(N) rotations between the N supersymmetry charges $\hat{Q}^{i\,\alpha}$. In fact, consistency requires these generators to appear in the anticommutator of two supercharges. The complete superalgebra has the non-vanishing (anti) commutation relations

$$\boxed{\begin{aligned}
\left[T^{ij}, T^{kl}\right] &= \Gamma\left(T^{kl}\right)^{i}{}_{m} T^{mj} + \Gamma\left(T^{kl}\right)^{j}{}_{m} T^{im},\\
\left[\hat{Q}^{k\,\alpha}, T^{ij}\right] &= \Gamma\left(T^{ij}\right)^{k}{}_{m} \hat{Q}^{m\,\alpha},\\
\left[\hat{M}_{\hat{a}\hat{b}}, \hat{M}_{\hat{c}\hat{d}}\right] &= -\hat{M}_{\hat{e}\hat{b}}\, \Gamma_{\mathrm{v}}\left(\hat{M}_{\hat{c}\hat{d}}\right)^{\hat{e}}{}_{\hat{a}} - \hat{M}_{\hat{a}\hat{e}}\, \Gamma_{\mathrm{v}}\left(\hat{M}_{\hat{c}\hat{d}}\right)^{\hat{e}}{}_{\hat{b}},\\
\left[\hat{Q}^{i\,\alpha}, \hat{M}_{\hat{a}\hat{b}}\right] &= \Gamma_{\mathrm{s}}\left(\hat{M}_{\hat{a}\hat{b}}\right)^{\alpha}{}_{\beta}\, \hat{Q}^{i\,\beta},\\
\left\{\hat{Q}^{i\,\alpha}, \hat{Q}^{j\,\beta}\right\} &= \delta^{ij}\left[\Gamma_{\mathrm{s}}\left(\hat{M}^{\hat{a}\hat{b}}\right)\mathcal{C}^{-1}\right]^{\alpha\beta} \hat{M}_{\hat{a}\hat{b}} - \left(\mathcal{C}^{-1}\right)^{\alpha\beta} T^{ij}.
\end{aligned}} \tag{5.60}$$

The new SO(N) generators T^{ij} play a very interesting role. If we gauge the algebra to obtain a supergravity theory based on this algebra, we first have to construct the superconnection $\hat{A}_\mu$, which will have the form

$$\hat{A}_\mu = \tfrac{1}{2}\hat{\omega}_\mu{}^{\hat{a}\hat{b}} \hat{M}_{\hat{a}\hat{b}} + \bar{\hat{\psi}}^{i}_{\mu} \hat{Q}^{i} + \tfrac{1}{2} A^{ij}{}_{\mu} T^{ij}. \tag{5.61}$$

Thus, on general grounds, we expect the supergravity theory to have a Vierbein, N gravitinos, and an SO(N) connection $A^{ij}{}_{\mu}$, and the theory to be invariant under SO(N) gauge transformations. Moreover, since the T^{ij}s rotate the supercharges, we expect the gravitinos to transform under SO(N) gauge transformations and be charged with respect to the SO(N) gauge field. For this reason, these theories are also called *gauged supergravities*. Since they are generalizations of the $N = 1$ case, they should also contain a negative cosmological constant and the vacuum will be anti-de Sitter spacetime. For $N > 1$ the procedure of gauging superalgebras is no longer straightforward and more fields usually occur in the theories, but the general facts we have just discussed remain true.

To obtain N-extended Poincaré superalgebras, we simply have to perform the Wigner–Inönü contraction Eq. (5.12) supplemented with

$$T^{ij} = g^{-1} Z^{ij}. \tag{5.62}$$

The effect of this rescaling (which is the only one that leads to a consistent superalgebra) is that these Z^{ij}s commute with every other generator in the superalgebra and become in fact a set of $N(N-1)/2$ SO(2) generators. Generators of this kind are called *central charges* and we could forget about them if they did not occur in the anticommutator of the supercharges. Before we write the resulting superalgebra, it is instructive to make some general considerations. Now we expect the theory to have $N(N-1)/2$ SO(2) gauge fields that we can still label $A^{ij}{}_{\mu}$. Since the Z^{ij}s are central, we do not expect the gravitinos to be charged under the gauge fields, although they will be invariant under some sort of constant SO(N) rotations. One may want to make the theory invariant under the local version of these SO(N) rotations, *gauging* them, and then one would recover the *gauged supergravities* (hence the name) we obtained by gauging the N-extended AdS superalgebra.

Now, to perform the Wigner–Inönü contraction, we need to choose a spinor representation of SO(2, 3). There are two such representations, which are called *electric* and *magnetic* representations, which are explicitly worked out in Appendix B.2.1. They are equivalent in the sense that they are related by a similarity transformation and, obviously, they are just two of an infinite family of equivalent representations. These two are, however, of special interest. If we contract using the electric representation, we obtain, for the anticommutator of two supercharges,

$$\{Q^{\alpha\, i}, Q^{\beta\, j}\} = i\delta^{ij}\left(\gamma^a \mathcal{C}^{-1}\right)^{\alpha\beta} P_a - i\left(\mathcal{C}^{-1}\right)^{\alpha\beta} Z^{ij}, \tag{5.63}$$

whereas, if we contract using the magnetic representation, we obtain

$$\{Q^{\alpha\, i}, Q^{\beta\, j}\} = i\delta^{ij}\left(\gamma^a \mathcal{C}^{-1}\right)^{\alpha\beta} P_a - \gamma_5\left(\mathcal{C}^{-1}\right)^{\alpha\beta} Z^{ij}. \tag{5.64}$$

As we advanced, the first surprise is that the central charges occur in this anticommutator, but nowhere else. The second surprise is that the central charges occur in two different ways. From the Poincaré point of view, in the electric case the Z^{ij}s are scalars whereas in the magnetic case they are pseudoscalars. How should we interpret these charges? If we construct supergravity theories gauging the "electric" superalgebra, we will have to associate gauge potentials with the Z^{ij}s, which will be, then, interpreted as electric charges, in agreement with their scalar nature. In the magnetic case, the Z^{ij}s should be interpreted as magnetic charges. The similarity transformation that relates the electric and magnetic AdS$_4$ representations becomes a *chiral–dual* transformation that rotates electric into magnetic charges and vice-versa. In fact, we can write the most general anticommutator of the supercharges including both kinds of charges of the most general N-extended Poincaré superalgebra,[8]

$$\boxed{\begin{aligned}
[M_{ab}, M_{cd}] &= -M_{eb}\Gamma_{\mathrm{v}}(M_{cd})^{e}{}_{a} - M_{ae}\Gamma_{\mathrm{v}}(M_{cd})^{e}{}_{b},\\
[P_a, M_{bc}] &= -P_e\Gamma_{\mathrm{v}}(M_{bc})^{e}{}_{a},\\
\left[Q^{\alpha\, i}, M_{ab}\right] &= \Gamma_{\mathrm{s}}(M_{ab})^{\alpha}{}_{\beta}Q^{\beta\, i},\\
\{Q^{\alpha\, i}, Q^{\beta\, j}\} &= i\delta^{ij}\left(\gamma^a\mathcal{C}^{-1}\right)^{\alpha\beta}P_a - i\left(\mathcal{C}^{-1}\right)^{\alpha\beta}Q^{ij} - \gamma_5\left(\mathcal{C}^{-1}\right)^{\alpha\beta}P^{ij},
\end{aligned}} \tag{5.65}$$

and this anticommutator (and the full superalgebra) will be invariant under the chiral–dual (electric–magnetic-duality) transformations which we expect to be symmetries of the N-extended Poincaré supergravity theories, but not of the N-extended AdS supergravities.

[8] In a Weyl basis, the electric and magnetic charges are combined into a single complex central charge matrix.

The main reason for this is that we do not know how to generalize electric–magnetic-duality transformations to the non-Abelian setting and also that, in the gauged supergravity theories, the gravitinos are electrically charged with respect to the gauge vectors but there are no additional fields magnetically charged with respect to them.

The above result opens up the possibility that there are more general central charges in the anticommutator of two supercharges that we have not considered at the beginning. We consider this interesting possibility in the next section.

5.4.1 Central extensions

According to the Haag–Lopuszański–Sohnius theorem, [496], the above anticommutator is the most general allowed if we impose the condition that our theory is Poincaré-invariant. Let us, therefore, not require Poincaré invariance. It turns out that any (Poincaré or AdS) superalgebra can be extended by including "central charges" with n antisymmetric Lorentz indices and two $\mathrm{SO}(N)$ indices $Z^{ij}_{a_1\cdots a_n}$ [538]. Generically, they appear in the anticommutator of two supercharges in the form

$$\frac{1}{n!}\left(\gamma^{a_1\cdots a_n}\mathcal{C}^{-1}\right)^{\alpha\beta} Z^{ij}_{a_1\cdots a_n}, \tag{5.66}$$

with the factor being necessary in order to have the right Hermiticity properties (which can be a γ_5 only in Poincaré superalgebras). These are not central charges in the strict sense because they do not commute with the Lorentz generators. In fact, consistency implies

$$\left[Z^{kl}_{c_1\cdots c_n}, M_{ab}\right] = -n\Gamma_{\mathrm{v}}(M_{ab})^{e}{}_{[c_1} Z^{kl}_{|e|c_2\cdots c_n]}. \tag{5.67}$$

The new central charge will be symmetric or antisymmetric in the $\mathrm{SO}(N)$ indices depending on whether $\left(\gamma^{a_1\cdots a_n}\mathcal{C}^{-1}\right)^{\alpha\beta}$ is symmetric or antisymmetric in $\alpha\beta$ since the full anticommutator has to be symmetric under the simultaneous interchange of $\alpha\beta$ and ij.

In four dimensions (and similarly in any dimensionality) it is easy to determine the symmetry of the possible terms:

$$\mathcal{C}^{-1}, \quad \gamma_5\mathcal{C}^{-1}, \quad \gamma_5\gamma_a\mathcal{C}^{-1}, \quad \gamma_{abc}\mathcal{C}^{-1}, \quad \gamma_{abcd}\mathcal{C}^{-1}, \tag{5.68}$$

are antisymmetric. In fact the second and the fifth and the third and the fourth matrices are related by Eq. (B.94). The symmetric matrices are

$$\gamma_a\mathcal{C}^{-1}, \quad \gamma_{ab}\mathcal{C}^{-1}, \quad \gamma_5\gamma_{ab}\mathcal{C}^{-1}, \quad \gamma_5\gamma_{abc}\mathcal{C}^{-1}. \tag{5.69}$$

The first and the fourth and the second and the third matrices are related by Eq. (B.94).

The most general anticommutator of the two central charges in $d=4$ will, therefore, be

$$\begin{aligned}\{Q^{\alpha\, i}, Q^{\beta\, j}\} = {} & i\delta^{ij}\left(\gamma^a\mathcal{C}^{-1}\right)^{\alpha\beta} P_a + i\left(\mathcal{C}^{-1}\right)^{\alpha\beta} Z^{[ij]} + \gamma_5\left(\mathcal{C}^{-1}\right)^{\alpha\beta} \tilde{Z}^{[ij]} \\ & + \left(\gamma^a\mathcal{C}^{-1}\right)^{\alpha\beta} Z^{(ij)}_a + i\left(\gamma_5\gamma^a\mathcal{C}^{-1}\right)^{\alpha\beta} Z^{[ij]}_a \\ & + i\left(\gamma^{ab}\mathcal{C}^{-1}\right)^{\alpha\beta} Z^{(ij)}_{ab} + \left(\gamma_5\gamma^{ab}\mathcal{C}^{-1}\right)^{\alpha\beta} \tilde{Z}^{(ij)}_{ab}.\end{aligned} \tag{5.70}$$

It is equally easy to determine the most general anticommutator of two supercharges in the AdS case, but in this case the Jacobi identities do not allow for any central charge.

We are now going to study the two simplest examples of extended Poincaré and AdS supergravity.

5.5 $N = 2, d = 4$ (Poincaré) supergravity

As mentioned before, the $N = 1, d = 4$ Poincaré supergravity theory can also be generalized by adding supersymmetric matter, giving, in some cases, theories that are invariant under more supersymmetry transformations.

The simplest case in which this happens is the addition of a supermultiplet containing a second gravitino ψ^2_μ and a vector field A_μ (the original gravitino in the $N = 1$ supergravity multiplet is now denoted by ψ^1_μ) and was studied by Ferrara and van Nieuwenhuizen in [380]. This theory is invariant under the original $N = 1$ local supersymmetry transformation with a parameter that we denote now by ϵ^1 and under a new independent local supersymmetry transformation with parameter ϵ^2. This theory is called for obvious reasons $N = 2, d = 4$ (Poincaré) supergravity and it is sometimes qualified as *ungauged* because it does not contain matter charged under the vector field.

From a different point of view, this SUEGRA is based on the $N = 2$ Poincaré superalgebra which we have just studied and could be derived by a generalization of the gauging of the algebraic procedure that worked for the $N = 1$ case (see [221]). Therefore, the fact that it has an SO(2) gauge vector field under which the gravitinos are not charged fits in the general scheme according to which we also expect the theory to be invariant under some sort of chiral–dual (electric–magnetic-duality) symmetry.

Forgetting the historical way in which the theory was constructed, it can now be described by treating on an equal footing both gravitinos and supersymmetries as follows: the $N = 2, d = 4$ supergravity multiplet consists of the Vierbein, a couple of real gravitinos, and a vector field

$$\left\{ e^a{}_\mu, \psi_\mu = \begin{pmatrix} \psi^1_\mu \\ \psi^2_\mu \end{pmatrix}, A_\mu \right\}, \tag{5.71}$$

respectively. The SO(2) indices $i = 1, 2$ that the fermions (and Pauli matrices) have in this theory will not be shown explicitly unless necessary and will be assumed to be contracted in obvious ways.[9]

The action for $N = 2, d = 4$ Poincaré supergravity is, in the first-order formalism [380],

$$\boxed{\begin{aligned} S = \int d^4x\, e\big\{ & R(e,\omega) + 2e^{-1}\epsilon^{\mu\nu\rho\sigma}\bar{\psi}_\mu\gamma_5\gamma_\nu\mathcal{D}_\rho\psi_\sigma - \mathcal{F}^2 \\ & + \mathcal{J}_{(\mathrm{m})}{}^{\mu\nu}(\mathcal{J}_{(\mathrm{e})\mu\nu} + \mathcal{J}_{(\mathrm{m})\mu\nu})\big\}, \end{aligned}} \tag{5.72}$$

where $\mathcal{D}$ is, as before, the Lorentz-covariant derivative, and

$$\mathcal{F}_{\mu\nu} = \tilde{F}_{\mu\nu} + \mathcal{J}_{(\mathrm{m})\mu\nu}, \quad \tilde{F}_{\mu\nu} = F_{\mu\nu} + \mathcal{J}_{(\mathrm{e})\mu\nu}, \quad F_{\mu\nu} = 2\partial_{[\mu}A_{\nu]}, \tag{5.73}$$

and

$$\mathcal{J}_{(\mathrm{e})\mu\nu} = i\bar{\psi}_\mu\sigma^2\psi_\nu, \qquad \mathcal{J}_{(\mathrm{m})\mu\nu} = -\frac{1}{2e}\epsilon^{\mu\nu\rho\sigma}\bar{\psi}_\rho\gamma_5\sigma^2\psi_\sigma. \tag{5.74}$$

F is the standard vector-field strength, $\tilde{F}$ is the *supercovariant* field strength[10] and, in terms

---

[9] It is possible to combine the two real gravitinos into a single complex gravitino. This has some advantages: the theory looks simpler because there is no need to use Pauli matrices. However, the structure of the supergravity theory is somewhat obscured and we choose the real form for pedagogical reasons.

[10] Whose supersymmetry transformation rule does not contain any derivatives of the gauge parameters.

of $\mathcal{F}$, the ("Maxwell") equation of the vector field is simply

$$\frac{\delta S}{\delta A_\nu} = 4e\nabla_\mu(e)\mathcal{F}^{\mu\nu} = 0. \tag{5.75}$$

The divergences of $\mathcal{J}_{\text{e}}$ and $\mathcal{J}_{\text{m}}$ are two topologically conserved currents that appear as electric-like and magnetic-like sources for the vector field:

$$\partial_\mu(eF^{\mu\nu}) = +\partial_\mu\left(e\mathcal{J}_{\text{e}}^{\nu\mu}\right) + \partial_\mu\left(e\mathcal{J}_{\text{m}}^{\nu\mu}\right). \tag{5.76}$$

They are naturally associated with the electric and magnetic central charges of the $N=2$, $d=4$ Poincaré supersymmetry algebra.

The equation of motion for $\omega_\mu{}^{ab}$ is the same as in the $N=1$ case (except for the SO(2) indices, which we do not show explicitly) and, thus, the solution is the same and, in particular, the torsion is given in terms of the gravitinos by

$$T_{\mu\nu}{}^a = i\bar{\psi}_\mu\gamma^a\psi_\nu \ \ (\equiv i\bar{\psi}_{j\,\mu}\gamma^a\psi^j_\nu). \tag{5.77}$$

The remaining two equations of motion are

$$\begin{aligned} \frac{\delta S}{\delta e^a{}_\mu} &= -2e\Big[G_a{}^\mu - 2T(\psi)_a{}^\mu - 2\tilde{T}(A)_a{}^\mu\Big], \\ \frac{\delta S}{\delta\bar{\psi}_\mu} &= 4\epsilon^{\mu\nu\rho\sigma}\gamma_5\gamma_\nu\hat{\mathcal{D}}_\rho\psi_\sigma - 4i\left(\tilde{F}^{\mu\nu} + i{}^\star\tilde{F}^{\mu\nu}\gamma_5\right)\sigma^2\psi_\nu, \end{aligned} \tag{5.78}$$

where the equation of motion for $\omega_\mu{}^{ab}$ has been used and

$$T(\psi)_a{}^\mu = -\frac{1}{2e}\epsilon^{\mu\nu\rho\sigma}\bar{\psi}_\nu\gamma_5\gamma_a\mathcal{D}_\rho\psi_\sigma, \qquad \tilde{T}(A)_a{}^\mu = \tilde{F}_a{}^\rho\tilde{F}^\mu{}_\rho - \tfrac{1}{4}e_a{}^\mu\tilde{F}^2. \tag{5.79}$$

The action and equations of motion are invariant under

general coordinate transformations,

$$\begin{aligned} \delta_\xi x^\mu &= \xi^\mu, & \delta_\xi e^a{}_\mu &= -\xi^\nu\partial_\nu e^a{}_\mu - \partial_\mu\xi^\nu e^a{}_\nu, \\ \delta_\xi\psi_\mu &= -\xi^\nu\partial_\nu\psi_\mu - \partial_\mu\xi^\nu\psi_\nu, & \delta_\xi A_\mu &= -\xi^\nu\partial_\nu A_\mu - \partial_\mu\xi^\nu A_\nu, \end{aligned} \tag{5.80}$$

local Lorentz transformations,

$$\delta_\sigma e^a{}_\mu = \sigma^a{}_b e^b{}_\mu, \qquad \delta_\sigma\psi_\mu = \tfrac{1}{2}\sigma^{ab}\gamma_{ab}\psi_\mu, \tag{5.81}$$

U(1) **gauge transformations,**

$$\delta_\chi A_\mu = \partial_\mu\chi, \tag{5.82}$$

and internal SO(2) rotations of the gravitinos,

$$\psi'_\mu = e^{i\varphi\sigma^2}\psi_\mu, \tag{5.83}$$

where φ is a constant (not spacetime-dependent) parameter.

The equations of motion (but not the action) are invariant under

chiral–dual (electric–magnetic-duality) SO(2) transformations,

$$\tilde{F}'_{\mu\nu} = \cos\theta\, \tilde{F}_{\mu\nu} + \sin\theta\, {}^\star\tilde{F}_{\mu\nu}, \qquad \psi'_\mu = e^{\frac{i}{2}\theta\gamma_5}\psi_\mu. \tag{5.84}$$

These transformations rotate electric into magnetic components of the supercovariant field strength and, at the same time, multiply by opposite phases the two chiral components of spinors (hence the name):

$$\psi'_\mu = \left[\tfrac{1}{2}e^{\frac{i}{2}\theta}(1+\gamma_5) + \tfrac{1}{2}e^{-\frac{i}{2}\theta}(1-\gamma_5)\right]\psi_\mu. \tag{5.85}$$

These transformations also rotate the two topologically conserved currents,

$$\begin{aligned} \mathcal{J}'_{(\mathrm{e})} &= \cos\theta\, \mathcal{J}_{(\mathrm{e})} - \sin\theta\, {}^\star\mathcal{J}_{(\mathrm{m})}, \\ \mathcal{J}'_{(\mathrm{m})} &= -\sin\theta\, {}^\star\mathcal{J}_{(\mathrm{e})} + \cos\theta\, \mathcal{J}_{(\mathrm{m})}, \end{aligned} \tag{5.86}$$

which helps to prove that these transformations also rotate the Maxwell equation into the Bianchi identity

$$\partial_\mu(e\, {}^\star F^{\mu\nu}) = 0, \tag{5.87}$$

since they are equivalent to

$$\mathcal{F}'_{\mu\nu} = \cos\theta\, \mathcal{F}_{\mu\nu} + \sin\theta\, {}^\star F_{\mu\nu}. \tag{5.88}$$

This is, of course, the same rotation as that which takes place between the two central charges in the $N = 2, d = 4$ Poincaré supersymmetry algebra.
Finally, the theory is invariant under

local $N = 2, d = 4$ **supersymmetry transformations,**

$$\delta_\epsilon e^a{}_\mu = -i\bar{\epsilon}\gamma^a\psi_\mu, \qquad \delta_\epsilon A_\mu = -i\bar{\epsilon}\sigma^2\psi_\mu, \qquad \delta_\epsilon\psi_\mu = \tilde{\mathcal{D}}_\mu\epsilon, \tag{5.89}$$

where

$$\tilde{\mathcal{D}}_\mu = \mathcal{D}_\mu + \tfrac{1}{4}\, \tilde{\not{F}}\gamma_\mu\sigma^2, \tag{5.90}$$

is the supercovariant derivative acting on ϵ.

It is instructive to check the invariance of the action under the above transformations. On varying the whole action, using the equation of motion of the spin connection, using the specific form of the supersymmetry transformation rules, using then

$$\tilde{\mathcal{D}}_\mu\bar{\epsilon} = \overline{\tilde{\mathcal{D}}_\mu\epsilon} = \overline{\mathcal{D}_\mu\epsilon} - \tfrac{1}{4}\bar{\epsilon}\gamma_\mu\, \tilde{\not{F}}\sigma^2, \tag{5.91}$$

and integrating by parts the partial derivative, we find that the invariance of the action depends on the $N = 2, d = 4$ Poincaré supersymmetry gauge identity

$$\tilde{\mathcal{D}}_\mu\frac{\delta S}{\delta\bar{\psi}_\mu} = -i\left(\frac{\delta S}{\delta e^a{}_\mu}\gamma^a + \frac{\delta S}{\delta A_\mu}\sigma^2\right)\psi_\mu, \tag{5.92}$$

where here the supercovariant derivative takes the form

$$\tilde{\mathcal{D}}_\mu \frac{\delta S}{\delta \bar{\psi}_\mu} = \left[\mathcal{D}_\mu + \tfrac{1}{4}\gamma_\mu \, \tilde{\not{F}}\sigma^2\right] \frac{\delta S}{\delta \bar{\psi}_\mu}. \tag{5.93}$$

To prove this gauge identity, one needs to use some of the results we used to prove the $N = 1$ gauge identity, the Bianchi identity for $F_{\mu\nu}$, and the $N = 2$ Fierz identities, with which it is possible to prove two main identities (see Section 5.7):

$$e^{-1}\epsilon^{\mu\nu\rho\sigma} T_{\nu\rho}{}^a \gamma_5 \gamma_a \psi_\sigma = 2\big({}^\star\mathcal{J}^{\mu\nu}_{(\mathrm{e})}\gamma_5 + i\mathcal{J}^{\mu\nu}_{(\mathrm{m})}\big)\sigma^2\psi_\nu \tag{5.94}$$

and

$$\begin{aligned} -e^{-1}\epsilon^{\mu\nu\rho\sigma} T_{\nu\rho}{}^a \gamma_5 \gamma_a \mathcal{D}_\rho \psi_\sigma = {} & 2iT(\psi)_a{}^\mu \gamma^a \psi_\mu - 2\mathcal{D}_\mu\big({}^\star\mathcal{J}^{\mu\nu}_{(\mathrm{e})}\gamma_5 + i\mathcal{J}^{\mu\nu}_{(\mathrm{m})}\big)\sigma^2\psi_\nu \\ & + 2\big({}^\star\mathcal{J}^{\mu\nu}_{(\mathrm{e})}\gamma_5 + i\mathcal{J}^{\mu\nu}_{(\mathrm{m})}\big)\sigma^2\mathcal{D}_\mu\psi_\nu. \end{aligned} \tag{5.95}$$

5.5.1 *The local supersymmetry algebra*

The commutator of two supersymmetry variations closes on shell with

$$[\delta_{\epsilon_1}, \delta_{\epsilon_2}] = \delta_\xi + \delta_\sigma + \delta_\chi + \delta_\epsilon, \tag{5.96}$$

where

$$\begin{aligned} \xi^\mu &= -i\bar{\epsilon}_1\gamma^\mu\epsilon_2, & \sigma^{ab} &= \xi^\mu \omega_\mu{}^{ab} - i\bar{\epsilon}_2\Big(\tilde{F}^{ab} - i\gamma_5{}^\star\tilde{F}^{ab}\Big)\sigma^2\epsilon_1, \\ \chi &= -i\bar{\epsilon}_2\sigma^2\epsilon_1 + \xi^\nu A_\nu, & \epsilon &= \xi^\mu\psi_\mu. \end{aligned} \tag{5.97}$$

5.6 $N = 2, d = 4$ "gauged" (AdS) supergravity

There are two main ways to arrive at this theory, apart from the algebra-gauging procedure. First, we could simply add supersymmetric matter to the $N = 1, d = 4$ AdS supergravity theory. Consistency requires that the pair of gravitini are charged under the vector field with a coupling constant that is equal to the Wigner–Inönü parameter g. For this reason, the theory was first found from the $N = 2, d = 4$ Poincaré theory by a gauging procedure: the internal SO(2) symmetry that rotates the two real gravitinos can be gauged [291, 399], the gauge field being the vector field already present in the theory (the field content is, therefore, the same). The pair of real gravitinos transforms as a complex, charged gravitino with a gauge parameter φ and we have to relate this parameter to the gauge parameter of U(1) transformations of the vector field according to

$$\varphi = -g\chi, \tag{5.98}$$

where g is the gauge coupling constant. The introduction of the minimal coupling between gravitinos and vector field requires, in order to preserve supersymmetry, the introduction of several other g-dependent terms, which can be absorbed into a change of connection from the Lorentz one to the anti-de Sitter one. In the end, the result is obviously the same as that which one obtains by adding supersymmetric matter to the $N = 1, d = 4$ AdS supergravity theory.

In any case, the two main characteristics of the theory are the presence of a negative cosmological constant $\Lambda = -3g^2$ and the fact that the gravitinos are minimally coupled to the vector field with coupling constant g.

We anticipate that there is going to be a third source term in the Maxwell equation, which is going to break the invariance under chiral–dual transformations of the "ungauged" (Poincaré) theory.

The gauged $N = 2, d = 4$ "gauged" supergravity action for these fields in the first-order formalism is, thus,

$$\boxed{\begin{aligned} S = \int d^4x\, e\Big\{& R(e,\omega) + 6g^2 + 2e^{-1}\epsilon^{\mu\nu\rho\sigma}\bar{\psi}_\mu\gamma_5\gamma_\nu\Big(\hat{\mathcal{D}}_\rho + igA_\rho\sigma^2\Big)\psi_\sigma \\ & - \mathcal{F}^2 + \mathcal{J}_{(\mathrm{m})}{}^{\mu\nu}(\mathcal{J}_{(\mathrm{e})\mu\nu} + \mathcal{J}_{(\mathrm{m})\mu\nu})\Big\}\,, \end{aligned}} \tag{5.99}$$

where again $\hat{\mathcal{D}}$ is the SO(2,3) (AdS) gauge covariant derivative

The symmetries of this action are essentially the same as in the ungauged case: GCTs, local Lorentz transformations,[11] U(1) gauge transformations, which now take the form

$$A'_\mu = A_\mu + \partial_\mu\chi\,, \qquad \psi'_\mu = e^{-ig\chi\sigma^2}\psi_\mu\,, \tag{5.100}$$

and local supersymmetry transformations, which take the same form as in the Poincaré case, but with the new supercovariant derivative

$$\tilde{\hat{\mathcal{D}}}_\mu = \hat{\mathcal{D}}_\mu + igA_\mu\sigma^2 + \tfrac{1}{4}\,\tilde{\not{F}}\gamma_\mu\sigma^2. \tag{5.101}$$

As mentioned before, the chiral–dual invariance of the ungauged theory is broken by the minimal coupling between gravitinos and vector field, which results in the new Maxwell equation with a new Noether current,

$$\partial_\nu(e\mathcal{F}^{\nu\mu}) - \frac{ig}{2}\epsilon^{\mu\nu\rho\sigma}\bar{\psi}_\nu\gamma_5\gamma_\rho\sigma^2\psi_\sigma. \tag{5.102}$$

For the sake of completeness, we give the remaining equations of motion

$$\begin{aligned} 0 &= G_a{}^\mu - 3g^2 e_a{}^\mu - 2T(\psi)_a{}^\mu - 2\tilde{T}(A)_a{}^\mu, \\ 0 &= e^{-1}\epsilon^{\mu\nu\rho\sigma}\gamma_5\gamma_\nu\Big(\hat{\mathcal{D}}_\rho + igA_\rho\sigma^2\Big)\psi_\sigma - i\Big(\tilde{F}^{\mu\nu} + i\,{}^\star\tilde{F}^{\mu\nu}\gamma_5\Big)\sigma^2\psi_\nu, \end{aligned} \tag{5.103}$$

where the equation of motion for $\omega_\mu{}^{ab}$ has been used and where

$$\begin{aligned} T(\psi)_a{}^\mu &= -\frac{1}{2e}\epsilon^{\mu\nu\rho\sigma}\bar{\psi}_\nu\gamma_5\gamma_a\Big(\hat{\mathcal{D}}_\rho + igA_\rho\sigma^2\Big)\psi_\sigma - \frac{ig}{2e}\epsilon^{\mu\nu\rho\sigma}\bar{\psi}_\nu\gamma_5\gamma_{\rho a}\psi_\sigma, \\ \tilde{T}(A)_a{}^\mu &= \tilde{F}_a{}^\rho\tilde{F}^\mu{}_\rho - \tfrac{1}{4}e_a{}^\mu\tilde{F}^2. \end{aligned} \tag{5.104}$$

[11] There is no invariance under the full SO(2,3).

To prove the invariance of the action under the local supersymmetry transformations, one has to check the $N = 2$, $d = 4$ AdS gauge identity

$$\tilde{\hat{\mathcal{D}}}_\mu \frac{\delta S}{\delta \bar{\psi}_\mu} = -i\left(\frac{\delta S}{\delta e^a{}_\mu}\gamma^a + \frac{\delta S}{\delta A_\mu}\sigma^2\right)\psi_\mu, \tag{5.105}$$

where, here,

$$\tilde{\hat{\mathcal{D}}}_\mu \frac{\delta S}{\delta \bar{\psi}_\mu} = \left[\hat{\mathcal{D}}_\mu + \tfrac{1}{4}\gamma_\mu\, \tilde{\not{F}}\sigma^2\right]\frac{\delta S}{\delta \bar{\psi}_\mu}. \tag{5.106}$$

To prove this identity we need only check the g-dependent terms (the g-independent ones work, as we checked in the previous section). To check the g-dependent terms, we need only the additional identities (see Section 5.7)

$$(\bar{\psi}_{[\nu|}\gamma_a\psi_{|\mu|})\gamma_5\gamma^a\sigma^2\psi_{|\rho]} = (\bar{\psi}_{[\nu|}\gamma_a\gamma_5\psi_{|\mu|})\gamma^a\sigma^2\psi_{|\rho]} \tag{5.107}$$

and

$$\begin{aligned}(\bar{\psi}_{[\nu|}\gamma_a\psi_{|\mu})\gamma_5\gamma^a\gamma_\rho\psi_{\sigma]} + (\bar{\psi}_{[\nu}\sigma^2\psi_\mu)\gamma_5\gamma_\rho\sigma^2\psi_{\sigma]} - (\bar{\psi}_{[\nu|}\gamma_5\sigma^2\psi_{|\mu})\gamma_\rho\sigma^2\psi_{\sigma]} \\ = -2(\bar{\psi}_{[\nu|}\gamma_5\gamma_{a|\rho}\psi_\mu)\gamma^a\psi_{\sigma]} - 2(\bar{\psi}_{[\nu|}\gamma_5\gamma_{|\rho}\sigma^2\psi_\mu)\sigma^2\psi_{\sigma]}.\end{aligned} \tag{5.108}$$

5.6.1 The local supersymmetry algebra

The commutator of two supersymmetry variations closes on-shell with the same parameters as in the ungauged case except for

$$\sigma^{ab} = \xi^\mu\omega_\mu{}^{ab} - g\bar{\epsilon}_2\gamma^{ab}\epsilon_1 - i\bar{\epsilon}_2\left(\tilde{F}^{ab} - i\gamma_5{}^\star\tilde{F}^{ab}\right)\sigma^2\epsilon_1. \tag{5.109}$$

From the point of view of the supersymmetry algebra, we are going from Poincaré supersymmetry to AdS supersymmetry in which the generator of SO(2) rotations has to appear in the anticommutator of two supersymmetry charges, for consistency. Although it appears in the same position as a central charge, it should be stressed that it is not a central charge because it does not commute with the supercharges.

5.7 Proofs of some identities

Using the $N = 2$ Fierz identities Eq. (B.57) we immediately find, for any spinor λ, the following two identities:

$$\begin{aligned}(\bar{\psi}_{[\nu|}\gamma_5\gamma_a\lambda)\gamma^a\psi_{|\mu]} = &-\tfrac{1}{2}(\bar{\psi}_{[\nu}\gamma_5\sigma^2\psi_{\mu]})\sigma^2\lambda - \tfrac{1}{4}(\bar{\psi}_{[\nu|}\gamma_a\gamma_5\sigma^2\psi_{|\mu]})\gamma^a\sigma^2\lambda \\ &+ \tfrac{1}{2}(\bar{\psi}_{[\nu}\sigma^2\psi_{\mu]})\gamma_5\sigma^2\lambda \\ &- \tfrac{1}{4}(\bar{\psi}_{[\nu|}\gamma_a\begin{pmatrix}\sigma^0\\ \sigma^1\\ \sigma^3\end{pmatrix}^{\mathrm{T}}\psi_{|\mu]})\gamma^a\gamma_5\begin{pmatrix}\sigma^0\\ \sigma^1\\ \sigma^3\end{pmatrix}\lambda\end{aligned} \tag{5.110}$$

and

$$(\bar{\psi}_{[\nu|}\gamma_a\chi)\gamma_5\gamma^a\psi_{|\mu]} = \tfrac{1}{2}(\bar{\psi}_{[\nu}\gamma_5\sigma^2\psi_{\mu]})\sigma^2\chi - \tfrac{1}{4}(\bar{\psi}_{[\nu|}\gamma_a\gamma_5\sigma^2\psi_{|\mu]})\gamma^a\sigma^2\chi$$
$$-\tfrac{1}{2}(\bar{\psi}_{[\nu}\sigma^2\psi_{\mu]})\gamma_5\sigma^2\chi$$
$$-\tfrac{1}{4}(\bar{\psi}_{[\nu|}\gamma_a\begin{pmatrix}\sigma^0\\ \sigma^1\\ \sigma^3\end{pmatrix}^{\mathrm{T}}\psi_{|\mu]})\gamma^a\gamma_5\begin{pmatrix}\sigma^0\\ \sigma^1\\ \sigma^3\end{pmatrix}\chi\,. \tag{5.111}$$

We can take $\lambda = \chi$ and subtract Eq. (5.111) from Eq. (5.110), giving

$$\begin{aligned}&(\bar{\psi}_{[\nu|}\gamma_5\gamma_a\lambda)\gamma^a\psi_{|\mu]} - (\bar{\psi}_{[\nu|}\gamma_a\lambda)\gamma_5\gamma^a\psi_{|\mu]}\\ &\quad = -(\bar{\psi}_{[\nu|}\gamma_5\sigma^2\lambda)\sigma^2\psi_{|\mu]}\sigma^2\lambda + (\bar{\psi}_{[\nu|}\sigma^2\lambda)\gamma_5\sigma^2\psi_{|\mu]}\gamma_5\sigma^2\lambda.\end{aligned} \tag{5.112}$$

We can take $\lambda = \psi_\rho$ and antisymmetrize in $\nu\rho\mu$, giving

$$(\bar{\psi}_{[\nu|}\gamma_a\psi_{|\mu|})\gamma_5\gamma^a\psi_{|\rho]} = -(\bar{\psi}_{[\nu|}\gamma_5\sigma^2\psi_{|\mu})\sigma^2\psi_{\rho]} + (\bar{\psi}_{[\nu}\sigma^2\psi_{\mu})\gamma_5\sigma^2\psi_{\rho]}, \tag{5.113}$$

from which Eq. (5.94) follows.

If we act with $\mathcal{D}_\mu$ on Eq. (5.94) and use Eq. (5.112) to $\lambda = \mathcal{D}_\mu\psi_\rho$ to relate $\mathcal{D}_\mu T_{\nu\rho}{}^a$ to $T(\psi)_a{}^\mu$, we obtain Eq. (5.95).

On substituting $\lambda = \sigma^2\psi_\rho$ into Eq. (5.110) and multiplying the result by an overall σ^2 and adding to it Eq. (5.111) with $\chi = \psi_\rho$, we obtain Eq. (5.107).

By combining Eqs. (5.110) and (5.111) with $\lambda = \gamma_\rho\psi_\sigma$ and $\chi = \sigma^2\gamma_\rho\psi_\sigma$ in several different ways, one obtains Eq. (5.108).

6
Conserved charges in general relativity

The definition of conserved charges in GR (and, in general, in non-Abelian gauge theories) is a very important and rather subtle subject, which is related to the definition of the energy–momentum tensor of the gravitational field. As we saw in the construction of the SRFT of gravity, perturbatively (that is, for asymptotically flat, well-behaved gravitational fields), GR gives a unique energy–momentum (Poincaré) tensor. It is natural to ask whether there is a fully general-covariant energy–momentum tensor for the gravitational field that would reduce to this in the weak-field limit. Many people (starting from Einstein himself) have unsuccessfully tried to find such a tensor, the current point of view being that it does not exist and that we have to content ourselves with energy–momentum *pseudotensors* for the gravitational field, which are covariant only under a restricted group of coordinate transformations (in most cases, Poincaré's). This, in fact, would be one of the characteristics of the gravitational field tied to the PEGI (see e.g. the discussion in Section 2.7 of [242]) that says that all the physical effects of the gravitational field (and one should include amongst them its energy density) can be locally eliminated by choosing a locally inertial coordinate system.[1]

The most important consequence of the absence of a fully general-covariant energy–momentum tensor for the gravitational field is the non-localizability of the gravitational energy: only the total energy of a spacetime is well defined (and conserved) because the integral of the energy–momentum pseudotensor over a finite volume would be dependent on the choice of coordinates. Some people find this unacceptable and, thus, the search for the general-covariant tensor goes on.[2]

[1] It must be mentioned that, in spite of all these considerations, the teleparallel approach to GR gives a local expression for the energy. This is, in fact, the reason why Møller [702] was led to the study of this class of theories.

[2] For instance, one can argue that the gravity field is actually characterized by the curvature tensor, not by the metric tensor. Even if we locally make the metric tensor flat by a coordinate transformation, we cannot do the same with the Riemann tensor. Thus one could look for energy–momentum tensors for the gravitational field constructed from the Riemann tensor. These are usually called "super-energy–momentum tensors" and an example of them is the Bel–Robinson tensor.

Apart from the problem of the gravitational energy–momentum tensor, the definition of conserved quantities in GR has many interesting points. Several approaches have been proposed and here we are going to study two: the construction of an energy–momentum pseudotensor for the gravitational field and the Noether approach. In both approaches there is a great deal of arbitrariness and in Section 6 we will study and compare several different results given in the literature in the weak field limit, finding complete agreement and a deep relation to the massless spin-2 relativistic field theory studied in Chapter 3.

6.1 The traditional approach

As we have stressed several times, the metric (or Rosenfeld) energy–momentum tensor of any general-covariant Lagrangian always satisfies (on-shell) the equation

$$\nabla_\mu T_{\text{matter}}{}^{\mu\nu} = 0, \tag{6.1}$$

as a direct consequence of general covariance. This equation is crucial for the consistency of the theory. Furthermore, it is the covariantization of the Minkowskian energy–momentum-conservation equation

$$\partial_\mu T_{\text{matter}}{}^{\mu\nu} = 0, \tag{6.2}$$

which is discussed at length in Chapter 2, and from which we can derive local conservation laws of the mass, momentum, and angular momentum and, in general, of those charges related to the invariance of a theory under certain coordinate transformations.

In curved spacetime, however, Eq. (6.1) is not equivalent to a continuity equation for the tensor density $\sqrt{|g|}\, T_{\text{matter}}{}^{\mu\nu}$ that holds in Minkowski spacetime. Actually, we can rewrite Eq. (6.1) in the form

$$\partial_\mu \left(\sqrt{|g|}\, T_{\text{matter}}{}^{\mu\nu} \right) = -\Gamma_{\rho\sigma}{}^{\nu} T_{\text{matter}}{}^{\rho\sigma}, \tag{6.3}$$

and, in general, the r.h.s. of this equation does not vanish. From this equation we cannot derive any local conservation law.

In a sense this was to be expected: only the *total* (matter plus gravity) energy and momentum should be conserved[3] and, therefore, we can only hope to be able to find local conservation laws for the total energy–momentum tensor. Now, how is the gravity energy–momentum tensor defined in GR? This is an old problem of GR.[4] It is clear that we cannot use the same definition (Rosenfeld's) as for the matter energy–momentum tensor because that leads to a total energy–momentum tensor that vanishes identically on-shell. On the other hand, if we found a covariantly divergenceless gravitational energy–momentum tensor, the total energy–momentum tensor would have the same problem as the matter one.

In fact, it can be argued, on the basis of the PEGI, that it is impossible to define a fully general-covariant gravitational energy–momentum tensor: according to the PEGI we can remove all the physical effects of a gravitational field locally, at any given point, by using an appropriate (free-falling) reference frame. This means that we could make the gravitational

[3] Actually, the coupling of gravity to the *total*, conserved, energy–momentum tensor was the main principle leading in Chapter 3 to GR.

[4] Some early references on the energy–momentum tensor of the gravitational field are [90, 355–627, 660, 839].

energy–momentum tensor vanish at any given point. However, that would mean that the energy–momentum tensor vanishes at any given point in any reference frame and, therefore, it is identically zero.[5]

Instead of being a problem, the lack of a gravitational energy–momentum (general-covariant) *tensor* really tells us that we should not be looking for such a tensor: after all, what we want is a total energy–momentum tensor satisfying the continuity equation $\partial_\mu T_{\text{total}}{}^{\mu\nu} = 0$, which is not a tensor equation. At most, it is a tensor equation w.r.t. the Poincaré group, if $T_{\text{total}}{}^{\mu\nu}$ behaves as a Lorentz tensor. Then we should simply be looking for a gravitational energy–momentum *pseudotensor* $t^{\mu\nu}$ transforming as a Lorentz tensor but not as a general-covariant tensor and such that

$$\partial_\mu \sqrt{|g|}\,(T_{\text{matter}}{}^{\mu\nu} + t^{\mu\nu}) = 0. \tag{6.4}$$

This should remind the reader of the self-consistency problem of the SRFT of gravitation that we studied in Chapter 3 in which we wanted to find the energy–momentum tensor of the gravitational field with respect to the vacuum which was Minkowski spacetime.

Another point to be stressed is that it looks as if we are forced to abandon general covariance to define conserved quantities. This is not so surprising: conserved quantities are in general naturally associated with the symmetries of the vacuum, not with the full symmetry of the theory. The vacuum is generically invariant under a finite-dimensional global symmetry group, in this case the Poincaré group. The conserved quantities we are after (momentum and angular momentum) are associated with that group.[6] In asymptotically flat spacetimes, only the infinity will have the invariances of the energy–momentum pseudotensor and, thus, only integrals over the boundary of (timelike hypersurfaces of) the whole spacetime will give well-defined conserved quantities. This implies the non-localizability of the energy mentioned in the introduction.

Our task now will be to find the gravitational energy–momentum pseudotensor and use it to define conserved quantities. Many candidates for a gravitational energy–momentum pseudotensor have been proposed in the literature. We are going to review just two of them that are physically very appealing: the Landau–Lifshitz pseudotensor [644], for asymptotically flat spacetimes, and the Abbott–Deser pseudotensor [1], for spacetimes with general asymptotics.

[5] This argument is not completely correct, though. In GR we can make the metric flat and its first derivatives vanishing at any given point, but not the second derivatives of the metric (i.e. the curvature). Although one can argue that, at one point (or any small enough neighborhood of a point), these non-vanishing derivatives will produce no observable physical effect (for instance, we need spatially separated test particles in order to measure tidal forces), this is not enough to say that the gravitational energy–momentum tensor will vanish identically at that point. In fact, the Landau–Lifshitz energy–momentum pseudotensor that we are going to study is precisely identified with the non-vanishing piece of the Einstein tensor at a point in a free-falling reference frame in which the metric is Minkowski's. The real problem, which is at the very foundations of GR, is that we do not have a good description of gravity in free-falling reference frames. That description could be covariantized (as happens with most other fields whose Lagrangians are well known in free-falling frames) and an energy–momentum tensor of the gravitational field and its coupling to itself could be found.

[6] We will later see what can be done for vacua different from Minkowski spacetime and for conserved quantities that are not necessarily associated with symmetries of the vacuum.

6.1.1 The Landau–Lifshitz pseudotensor

The main physical idea behind the definition of the Landau–Lifshitz energy–momentum pseudotensor is precisely that gravity can be locally eliminated at the point P by using a free-falling coordinate system at P. Then, the starting point is to choose, for instance, Riemann normal coordinates Eq. (3.268) at the given point P where we want to define the energy–momentum pseudotensor. In this coordinate system, at the point P the equation satisfied by the matter energy–momentum tensor takes the form

$$\nabla_\mu T_{\text{matter}}{}^{\mu\nu} = \partial_\mu T_{\text{matter}}{}^{\mu\nu} = 0, \tag{6.5}$$

and the matter energy–momentum tensor is conserved in the usual sense there because we have eliminated the gravitational field, its interaction with matter, and its own energy–momentum pseudotensor through the choice of coordinates. Thus, in this coordinate system, the gravitational energy–momentum pseudotensor vanishes.

Technically, this equation is satisfied identically due to the Bianchi identity of the r.h.s. of Einstein's equation. This means that, in this coordinate system, at the point P in question, Einstein's equation must be of the form (taking into account that the determinant of the metric can go through partial derivatives taken at P in this coordinate system)

$$\frac{1}{|g|}\partial_\rho \eta^{\mu\nu\rho} = T_{\text{matter}}{}^{\mu\nu}, \qquad \eta^{\mu\nu\rho} = -\eta^{\mu\rho\nu}. \tag{6.6}$$

Actually, it can be checked that, in this coordinate system, at the point in question, Einstein's equations take precisely the above form with

$$\eta^{\mu\nu\rho} = -\frac{2}{\chi^2}\partial_\sigma \mathfrak{g}^{\mu\sigma,\nu\rho}, \qquad \mathfrak{g}^{\mu\nu} = \sqrt{|g|}g^{\mu\nu}, \qquad \mathfrak{g}^{\mu\sigma,\nu\rho} = \tfrac{1}{2}(\mathfrak{g}^{\mu\nu}\mathfrak{g}^{\sigma\rho} - \mathfrak{g}^{\mu\rho}\mathfrak{g}^{\sigma\nu}). \tag{6.7}$$

Now, in any coordinate system we can define the *Landau–Lifshitz energy–momentum pseudotensor* by

$$t_{\text{LL}}{}^{\mu\nu} = \frac{1}{|g|}\partial_\rho \eta^{\mu\nu\rho} - T_{\text{matter}}{}^{\mu\nu}, \tag{6.8}$$

so, due to the symmetries of $\eta^{\mu\nu\rho}$,

$$\partial_\mu\{|g|(T_{\text{matter}}{}^{\mu\nu} + t_{\text{LL}}{}^{\mu\nu})\} = 0, \tag{6.9}$$

which is essentially what we wanted.

To determine the explicit form of $t_{\text{LL}}{}^{\mu\nu}$ we use Einstein's equation

$$t_{\text{LL}}{}^{\mu\nu} = -\frac{2}{\chi^2}\big(\partial_\rho\partial_\sigma \mathfrak{g}^{\mu\sigma,\nu\rho} + G^{\mu\nu}\big), \tag{6.10}$$

and by expanding both terms one obtains a very complicated and not very illuminating expression that is quadratic in the metric and quadratic in connections that can be found in most standard gravity textbooks [242, 644, 707]. Since it depends on connections, it does not transform as a world tensor, but it does transform as a Lorentz tensor (affine connections transform as Lorentz tensors), just as expected.

Having the total conserved energy–momentum pseudotensor obeying the local continuity equation, we go on to define the conserved charges (momentum and angular momentum) by the volume integrals[7]

$$\begin{aligned} P^{\mu} &= \int_{\Sigma} d^{d-1}\Sigma_{\nu}\sqrt{|g|}\left(T_{\text{matter}}{}^{\mu\nu} + t_{\text{LL}}{}^{\mu\nu}\right), \\ M^{\mu\alpha} &= \int_{\Sigma} d^{d-1}\Sigma_{\nu}\sqrt{|g|}\left[2x^{[\alpha}\left(T_{\text{matter}}{}^{\mu]\nu} + t_{\text{LL}}{}^{\mu]\nu}\right)\right], \end{aligned} \tag{6.11}$$

where it is assumed that one integrates over a timelike hypersurface Σ.

One of the shortcomings of this approach is that it is not clear why these are the (only) conserved charges and how we can generalize it to other spacetimes in which these are not necessarily the conserved charges. The Abbott–Deser approach will solve this problem.

A second shortcoming is the large number of terms that have to be calculated in order to find the conserved quantities. In practice, though, one uses Eq. (6.8) to rewrite, using Stokes' theorem,

$$P^{\mu} = \tfrac{1}{2}\int_{\partial\Sigma} d^{d-2}\Sigma_{\nu\rho}\frac{\eta^{\mu\nu\rho}}{\sqrt{|g|}}, \tag{6.12}$$

and similarly for $M^{\mu\alpha}$. This is an interesting expression that has to be evaluated at the boundary of the hypersurface Σ, which is, typically, for asymptotically flat spacetimes a $(d-2)$-sphere at spatial infinity $\mathrm{S}^{d-2}_{\infty}$. We could integrate over the boundary of smaller regions of the spacetime. However, the integrand is not a general-covariant tensor and the result of the integral would be coordinate-dependent and the momentum would not be well defined. Only when we integrate over $\mathrm{S}^{d-2}_{\infty}$ in asymptotically flat spacetimes does the integral transform as a Poincaré tensor. This is the common behavior of most superpotentials used to defined conserved quantities in GR, except for Møller's [702], which is a true tensor.

For asymptotically flat spaces, we can use the weak-field expansion[8] $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$. In this limit, we see that

$$\mathfrak{g}^{\mu\sigma,\nu\rho} = K^{\mu\sigma\nu\rho} + \mathcal{O}(h^2), \tag{6.13}$$

and

$$\eta^{\mu\nu\rho} = 2\eta^{\mu\nu\rho}_{\text{LL}} + \mathcal{O}(h^2), \qquad 2\partial_{\rho}\eta^{\mu\nu\rho}_{\text{LL}} = \mathcal{D}^{\mu\nu}(h), \tag{6.14}$$

where $\eta^{\mu\nu\rho}_{\text{LL}}$ was defined in Eq. (3.90) and $\mathcal{D}^{\mu\nu}(h)$ is the Fierz–Pauli wave operator.

Thus, in practice, all we have to do is to integrate the Fierz–Pauli wave operator over the volume Σ or η_{LL} over the boundary $\partial\Sigma$, if the asymptotic weak-field expansion of the metric is well defined. Many different gravity energy–momentum pseudotensors have been proposed in the literature but, in the end, one never uses them directly. Instead one integrates over Σ, using the equations of motion, an expression that, in the weak-field limit, is equivalent to the Fierz–Pauli wave operator. Usually this expression is rewritten as an integral over the boundary using Stokes' theorem. This can be done in many different ways, as we discussed in Chapter 3, and here is where the differences arise.[9]

---

[7] Observe the "extra" factors of $\sqrt{|g|}$.

[8] Observe that the $h_{\mu\nu}$ that we are using in this chapter is $\chi h_{\mu\nu}$ of Chapter 3.

[9] Some expressions may be better suited for certain boundary conditions. When we compare the weak-field limits of the various expressions proposed in the literature, we have to bear in mind that the expansions used are valid only under certain asymptotic conditions.

With the zeroth component of Eq. (6.12) in $d = 4$ for a stationary asymptotically flat metric in Cartesian coordinates, we obtain the Arnowitt–Deser–Misner (ADM) mass formula which was first derived by canonical methods in [54]:

$$\boxed{M_{\text{ADM}} = \frac{2}{\chi^2}\int_{S^2_\infty} d^2 S_k(\partial_k h_{ll} - \partial_l h_{lk}),} \tag{6.15}$$

where

$$dS_k \equiv \tfrac{1}{2}\epsilon_{ijk}dx^i \wedge dx^j. \tag{6.16}$$

We can immediately apply this formula to the simplest spacetime: Schwarzschild's spacetime. The four-dimensional Schwarzschild solution in Schwarzschild coordinates is

$$ds^2 = \left(1 - \frac{k}{r}\right)dt^2 - \left(1 - \frac{k}{r}\right)^{-1} dr^2 - r^2 d\Omega^2_{(2)}. \tag{6.17}$$

k is the integration constant. To apply the ADM mass formula Eq. (6.15), we first rewrite the metric in *isotropic coordinates*,

$$r = (\rho + k/4)^2/\rho, \tag{6.18}$$

obtaining

$$ds^2 = \left(\frac{\rho - k/4}{\rho + k/4}\right)^2 dt^2 - \left(1 + \frac{k/4}{\rho}\right)^4 d\vec{x}_3^{\,2}, \qquad \rho = |\vec{x}_3|. \tag{6.19}$$

With $\chi^2 = 16\pi G_{\text{N}}^{(4)}$, the ADM mass formula gives, in agreement with our results of Chapter 3,

$$k = 2G_{\text{N}}^{(4)} M. \tag{6.20}$$

6.1.2 The Abbott–Deser approach

In [1] Abbott and Deser proposed a general definition for spacetimes of arbitrary asymptotic behavior associating conserved charges with isometries of the asymptotic geometry which is supposed to be the vacuum (so we are physically calculating the conserved charges of an isolated system). This definition is very useful and can be extended to more complicated cases in which some dimensions are compactified [309] (see also [164]), other contexts such as supercharges in supersymmetric theories [1] (associated with Killing spinors of the vacuum, that will be studied in Chapter 13), and charges in non-Abelian gauge theories (associated with gauge Killing vectors) [2].

In this section we are essentially going to repeat and extend the calculations of Abbott and Deser [1] in our conventions, comparing the result with the one in the previous section. We will also use it to calculate the mass of a spacetime that is not asymptotically flat, as an example of its usefulness.

The first step in this approach is the expansion of the gravitational field around an arbitrary background metric $\bar{g}_{\mu\nu}$ that solves the vacuum cosmological Einstein equations, and the derivation of the linearized Einstein equations in that background. We already did this in Section 3.4.1, where we also gave a definition of the gravitational energy–momentum pseudotensor that was different from Landau and Lifshitz's. Here we use the notation and definitions of that section.

The second step consists in the construction of a conserved quantity. First, we observe that the linearized cosmological Einstein tensor satisfies the Bianchi identity with respect to the background metric:

$$\bar{\nabla}_\mu G_{c\,\mathrm{L}}{}^{\mu\nu} = 0. \tag{6.21}$$

This can be proven either by direct calculation or by taking the divergence of the cosmological Einstein tensor:

$$\nabla_\mu G_c{}^{\mu\nu} = \bar{\nabla}_\mu \bar{G}_c{}^{\mu\nu} + \gamma_{\mu\rho}{}^{\mu} \bar{G}_c{}^{\rho\nu} + \gamma_{\mu\rho}{}^{\nu} \bar{G}_c{}^{\mu\rho} + \bar{\nabla}_\mu G_{c\,\mathrm{L}}{}^{\mu\nu} + \mathcal{O}(h^2), \tag{6.22}$$

and observing that, by hypothesis, $\bar{G}_c{}^{\mu\nu} = 0$, and also that the Bianchi identity has to be satisfied order by order in h.

Using now Eq. (3.292) and Eq. (6.21), we find

$$\bar{\nabla}_\mu T^{\mu\nu}_{\text{total}} = 0, \qquad T^{\mu\nu}_{\text{total}} = T^{\mu\nu}_{\text{matter}} + t^{\mu\nu}_{\text{AD}}. \tag{6.23}$$

Finally, if $\bar{\xi}_\mu$ is a Killing vector field of the background metric, the above equation implies

$$\bar{\nabla}_\mu \left(T^{\mu\nu}_{\text{total}} \bar{\xi}_\nu\right) = 0, \tag{6.24}$$

and, from this, we find that the quantity

$$E(\bar{\xi}) \equiv \int_\Sigma d^{d-1}x \sqrt{|\bar{g}|}\, T^{0\nu}_{\text{total}} \bar{\xi}_\nu, \tag{6.25}$$

where the integral is performed over a constant time slice Σ, is a conserved quantity, namely the conserved quantity associated with the background Killing vector $\bar{\xi}^\mu$. If the Killing vector generates translations in time in the background, the conserved quantity is the energy (or mass). (In general, if $\bar{\xi}$ is a timelike Killing vector, $E(\bar{\xi})$ is called the *Killing energy*.) For Killing vectors that generate rotations we obtain components of the angular momentum etc.

The covariant form of the above expression is

$$E(\bar{\xi}) \equiv \int_\Sigma d^{d-1}\Sigma_\mu T^{\mu\nu}_{\text{total}} \bar{\xi}_\nu, \tag{6.26}$$

where

$$d^{d-1}\Sigma_\mu = \frac{1}{(d-1)!\sqrt{|\bar{g}|}} \epsilon_{\mu\rho_1\cdots\rho_{d-1}} dx^{\rho_1} \wedge \cdots \wedge dx^{\rho_{d-1}}. \tag{6.27}$$

This equation can be seen as a generalization of Landau and Lifshitz's Eq. (6.11), which is valid for any background metric $\bar{g}_{\mu\nu}$ and any of its Killing vectors $\bar{\xi}_\mu$, with a different definition of the gravitational energy–momentum pseudotensor. Indeed, Landau and Lifshitz's Eq. (6.11) can be written in the form

$$E(\xi) = \int_\Sigma d^{d-1}\Sigma_\mu \sqrt{|g|} (T_{\text{matter}}{}^{\mu\nu} + t_{\text{LL}}{}^{\mu\nu}) \xi_\nu, \tag{6.28}$$

where the Minkowski spacetime Killing vectors $\xi^{(\mu)\,\nu} = \eta^{\mu\nu}$ that generate constant translations are used to obtain the P^{μ}s and those which generate Lorentz transformations $\xi^{(\mu\alpha)\,\nu} = -2x^{[\mu}\eta^{\alpha]\nu}$ are used to obtain the $M^{\mu\alpha}$s. The different definition of $t^{\mu\nu}$ is responsible for the extra factor of $\sqrt{|g|}$ in this formula compared with Abbott and Deser's. On the other hand, in the Landau–Lifshitz approach we are forced not only to work with asymptotically flat spacetimes, but also to use Cartesian coordinates. The Abbott–Deser approach can be used for any spacetime in any coordinate system.

The main problem with Eq. (6.26) is also that the expression for $t^{\mu\nu}$ is very complicated; it is in fact, an infinite series in h. The solution is, again, to use the equation of motion Eq. (3.292) to rewrite it. The new integrand is, as we argued it would in general be, just the covariantized Fierz–Pauli wave operator $\bar{\mathcal{D}}^{\mu\nu}(h)$ contracted with a background Killing vector, that is,

$$E(\bar{\xi}) = \frac{2}{\chi^2}\int_{\Sigma} d^{d-1}\Sigma_{\mu}\bar{\mathcal{D}}^{\mu\nu}(h)\bar{\xi}_{\nu}. \tag{6.29}$$

At this point we notice that the integrand of this expression is nothing but the conserved Noether current $j^{\mu}_{\mathrm{N}}(\bar{\xi})$ in Eq. (3.314) and we can use the results of Section 3.4.1 to rewrite it as a total derivative and then use Stokes' theorem to rewrite it as a $(d-2)$-surface integral,

$$\boxed{E(\bar{\xi}) = -\frac{2}{\chi^2}\int_{\partial\Sigma=S^{d-2}_{\infty}} d^{d-2}\Sigma_{\mu\alpha}\left[\left(\bar{\nabla}_{\beta}K^{\mu\alpha\nu\beta}\right)\bar{\xi}_{\nu} - K^{\mu\beta\nu\alpha}\bar{\nabla}_{\beta}\bar{\xi}_{\nu}\right],} \tag{6.30}$$

where

$$d^{d-2}\Sigma_{\mu\alpha} = \frac{1}{(d-2)!\sqrt{|\bar{g}|}}\epsilon_{\mu\alpha\rho_1\cdots\rho_{d-2}}dx^{\rho_1}\wedge\cdots\wedge dx^{\rho_{d-2}}. \tag{6.31}$$

This is essentially Abbott and Deser's final result, although one can massage the above expression further to make it useful in specific situations. For instance, the following alternative expression is noteworthy. We first observe the identity

$$\bar{\nabla}_{\beta}K^{\mu\alpha\nu\beta} = 3\bar{g}^{\lambda\mu\alpha,\,\nu}{}_{\rho\sigma}\gamma_{\lambda}{}^{\rho\sigma}. \tag{6.32}$$

We can replace $\gamma_{\mu\nu}{}^{\rho}$ by $\Delta\Gamma_{\mu\nu}{}^{\rho} = \Gamma_{\mu\nu}{}^{\rho} - \bar{\Gamma}_{\mu\nu}{}^{\rho}$ because the difference is quadratic and higher in $h_{\mu\nu}$, which is assumed to go to zero at infinity fast enough. Then

$$E(\bar{\xi}) = -\frac{2}{\chi^2}\int_{S^{d-2}_{\infty}} d^{d-2}\Sigma_{\mu\alpha}\left[3\bar{g}^{\lambda\mu\alpha,\,\nu}{}_{\rho\sigma}\,\Delta\Gamma_{\lambda}{}^{\rho\sigma}\bar{\xi}_{\nu} - K^{\mu\beta\nu\alpha}\bar{\nabla}_{\beta}\bar{\xi}_{\nu}\right]. \tag{6.33}$$

Furthermore, in Minkowski spacetime in Cartesian coordinates the generators of translations are covariantly constant and the second term can be dropped, so we obtain for any component of the momentum (and, in particular, for the energy) of asymptotically flat spacetimes the expression

$$\boxed{E(\bar{\xi}) = -\frac{2}{\chi^2}\int_{\partial\Sigma} d^{d-2}\Sigma_{\mu\alpha}3\bar{g}^{\lambda\mu\alpha,\,\nu}{}_{\rho\sigma}\,\Delta\Gamma_{\lambda}{}^{\rho\sigma}\bar{\xi}_{\nu},} \tag{6.34}$$

which was first used in [719] and used afterwards in all proofs of the positivity of the mass or Bogomol'nyi bounds based on Nester's construction ([442, 596, 600] etc.).

It is also interesting to compare Eq. (6.30) with Landau and Lifshitz's result. In flat spacetime, with Cartesian coordinates, for translational Killing vectors (which are covariantly constant) Eq. (6.30) simplifies to

$$P^{\mu} = E(\bar{\xi}^{(\mu)}) = -\frac{2}{\chi^2}\int_{S^{d-2}_{\infty}} d^{d-2}\Sigma_{\nu\alpha}\partial_{\beta}K^{\nu\alpha\mu\beta}, \tag{6.35}$$

and we see that the integrand is nothing but $\eta_{\text{AD}}^{\nu\mu\alpha}$ defined in Eqs. (3.90). The difference from $\eta_{\text{LL}}^{\nu\mu\alpha}$ is just

$$\eta_{\text{LL}}^{\nu\mu\alpha} - \eta_{\text{AD}}^{\nu\mu\alpha} = \partial_{\beta}K^{\nu\mu\alpha\beta}, \tag{6.36}$$

so

$$\partial_{\alpha}\left(\eta_{\text{LL}}^{\nu\mu\alpha} - \eta_{\text{AD}}^{\nu\mu\alpha}\right) = 0, \tag{6.37}$$

and the difference should not contribute to the conserved charges.

To end this section, let us apply these results to a simple example: the four-dimensional Reissner–Nordström–de Sitter spacetime. First, for any static, spherically symmetric metric

$$ds^2 = g_{tt}(r)dt^2 + g_{rr}(r)dr^2 - r^2 d\Omega^2_{(2)} \tag{6.38}$$

and backgrounds

$$d\bar{s}^2 = \bar{g}_{tt}(r)dt^2 + \bar{g}_{rr}(r)dr^2 - r^2 d\Omega^2_{(2)} \tag{6.39}$$

and for the obvious timelike Killing vector $\bar{\xi}_{\nu} = \delta_{0\nu}\bar{g}_{tt}$ we obtain the mass formula

$$M = -\frac{1}{2G_{\text{N}}^{(4)}}\frac{|\bar{g}_{tt}|^{\frac{1}{2}}}{|\bar{g}_{rr}|^{\frac{3}{2}}} r(g_{rr} - \bar{g}_{rr}). \tag{6.40}$$

This formula can be directly applied to the Schwarzschild metric given in Eq. (6.17) and it gives the correct result. It can also be applied to asymptotically (anti-)de Sitter spacetimes. We can apply it, for instance to the Reissner–Nordström–(anti-)de Sitter metric in static coordinates

$$\begin{aligned} ds^2 &= V dt^2 - V^{-1}dr^2 - r^2 d\Omega^2_{(2)}, \\ V &= 1 - \frac{k}{r} + \frac{Z^2}{4r^2} - \frac{1}{3}\Lambda r^2. \end{aligned} \tag{6.41}$$

We obtain again $M = k/(2G_N^{(4)})$.

6.2 The Noether approach

The standard method used to obtain the conserved charge is through the Noether current. We have seen that, in fact, the Abbott–Deser formula for conserved charges can be seen from this point of view as the integral of the conserved Noether current associated with a background Killing vector of linearized gravity. Here we are going to investigate this point of view further, since there is a Noether current $j^{\mu}_{\text{N2}}(\xi)$ associated with *any* vector field ξ^{μ}, Killing or otherwise, as we proved in Section 3.4.1 and we could simply calculate the superpotential $j^{\mu\nu}_{\mathrm{N}2}(\xi)$ associated with an arbitrary vector field, by generalizing Abbott and Deser's result.[10] We will, however, content ourselves with reviewing some known results.

For GR we found the Noether current $j^{\mu}_{\mathrm{N}}(\xi)$ for any vector ξ^{μ} in Section 4.1.2 and we saw how it could be rewritten as the divergence of the antisymmetric superpotential tensor $j^{\mu\nu}_{\mathrm{N}}(\xi) = 2\nabla^{[\mu}\xi^{\nu]}$. Using it, we can define a conserved charge for each vector ξ^{μ}:

$$\boxed{E(\xi) = -\frac{2}{\chi^2}\int_{\partial\Sigma} d^{d-2}\Sigma_{\mu\alpha}\nabla^{\mu}\xi^{\alpha}.} \tag{6.42}$$

If ξ^{μ} is timelike, then $E(\xi)$ is the energy and the above formula is *Komar's formula* [631]. This formula can also be obtained from physical principles as in section 11.2 of [932].

We can compare Komar's formula with Abbott and Deser's Eq. (6.30). On rewriting it as

$$E(\xi) = -\frac{2}{\chi^2}\int_{\partial\Sigma} d^{d-2}\Sigma_{\mu\alpha}g^{\mu\alpha,\nu\beta}\nabla_{\nu}\xi_{\beta} \tag{6.43}$$

and using the weak-field expansion $g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu}$, we find that it reproduces exactly the first term in Eq. (6.30) (after integrating by parts) but not the second one. This difference is probably responsible for one of the known drawbacks of Komar's formula: it gives a wrong value for the angular momentum of the Kerr solution.

Komar's formula can be modified by adding to the Einstein–Hilbert action total derivative terms that modify the Noether current as explained in Section 4.1.2. The problem now is determining which total derivative should be added. In [770] some examples of total derivative terms that have been added in the literature can be found and a new one is proposed. Using it and using also, basically, Eq. (4.125) in the absence of torsion, the authors propose a new superpotential whose integral (if it is convergent) gives a conserved charge for any vector field ξ. In the weak-field limit, it can be written in the form

$$E(\bar{\xi}) = -\frac{2}{\chi^2}\int_{\partial\Sigma} d^{d-2}\Sigma_{\mu\alpha}\big[\big(\bar{\nabla}_{\beta}K^{\mu\alpha\nu\beta}\big)\xi_{\nu} - \bar{h}^{\sigma[\mu}\bar{\nabla}_{\sigma}\xi^{\alpha]}\big]. \tag{6.44}$$

The first term is identical to the first term in Eq. (6.30) and the second is identical to the second if $\xi^{\mu} = \bar{\xi}^{\mu}$, a background Killing vector, but the formula can be applied to more general cases. In fact, the complete formula in [770] gives correct results in the presence of radiation, whereas Abbott and Deser's does not, probably because the weak-field expansion is not consistent in those spacetimes.

6.3 The positive-energy theorem

Now that we know how to define conserved quantities in GR and, in particular, the mass (total energy), we are going to prove that the mass of an asymptotically flat spacetime that solves the Einstein equations

$$G_{\mu\nu} = \frac{\chi^2}{2}T_{\mathrm{matter}\,\mu\nu} \tag{6.45}$$

[10] Of course, not for every vector will the integral defining the corresponding conserved charge converge, but we will not deal with this problem here.

with a matter energy–momentum tensor satisfying the *dominant energy condition*

$$T_{\text{matter}}{}^{\mu\nu} k_\mu n_\nu \geq 0, \qquad \forall\, n_\mu,\, k_\mu \;\; \text{non-spacelike}, \tag{6.46}$$

is always non-negative, vanishing only for flat spacetime. This result was first obtained by Schoen and Yau in [830]. A new proof based on spinor techniques inspired by SUGRA was afterwards presented by Witten in [958] and subsequently by Nester in [719] and Israel and Nester in [596]. Previously, the positivity of mass in SUGRA and GR (as the bosonic part of $N = 1$ SUGRA) had been established in [310, 480]. Here we are going to use this Witten–Nester–Israel (WNI) technique because it can be generalized to more complicated cases and because it has a strong relation to supergravity that we will also use later on in Chapter 13.

The positive-energy theorem is a very important result associated with the *cosmic-censorship conjecture*: in the gravitational collapse of a star, the gravitational binding energy, which is negative, grows in absolute value. If the process continued indefinitely, the total energy of the collapsing star would become negative. However, according to the positive-mass theorem, this cannot happen and we expect a black-hole horizon to appear before the mass becomes negative.

The WNI technique starts with the construction of the Nester 2-form $E^{\mu\nu}$. In this case (pure $d = 4$ gravity; the extension to higher dimensions is straightforward) it is simply

$$E^{\mu\nu}(\epsilon) = +\frac{i}{2}\bar{\epsilon}\gamma^{\mu\nu\rho}\nabla_\rho\epsilon + \text{c.c.}, \tag{6.47}$$

where ϵ is a *commuting* Dirac spinor. The Nester form is manifestly real. Then, we define the integral I,

$$I(\epsilon) = \int_{\partial\Sigma} {}^\star E(\epsilon), \tag{6.48}$$

where Σ is a three-dimensional spacelike hypersurface (for instance, a constant time slice) whose boundary $\partial\Sigma$ is a 2-sphere at infinity S^2_∞. Observe that the Nester form can be rewritten in the form

$$E^{\mu\nu}(\epsilon) = +i\bar{\epsilon}\gamma^{\mu\nu\rho}\nabla_\rho\epsilon + \nabla_\rho\Big(-\frac{i}{2}\bar{\epsilon}\gamma^{\mu\nu\rho}\epsilon\Big), \tag{6.49}$$

and only the first term contributes to I.

The proof has two parts.

1. Prove that, for suitably chosen spinors ϵ and $T_{\text{matter}}{}^{\mu\nu}$ satisfying the dominant energy condition, $I(\epsilon) \geq 0$.

2. Relate $I(\epsilon)$ to conserved charges.

1. We use Stokes' theorem

$$I(\epsilon) = \int_{\partial\Sigma} {}^\star E(\epsilon) = \int_\Sigma d^\star E(\epsilon) = -\tfrac{1}{2}\int_\Sigma d^3\Sigma_\nu \big\{-i\nabla_\mu\bar{\epsilon}\gamma^{\mu\nu\rho}\nabla_\rho\epsilon - i\bar{\epsilon}\gamma^{\mu\nu\rho}\nabla_\mu\nabla_\rho\epsilon\big\}\,, \tag{6.50}$$

where

$$d^3\Sigma_\mu = \frac{1}{3!\sqrt{|g|}} dx^\rho \wedge dx^\sigma \wedge dx^\lambda \epsilon_{\rho\sigma\lambda\mu}. \tag{6.51}$$

The second term in the integral is proportional to the Lorentz curvature tensor due to the Ricci identities Eqs. (1.94). Expanding the product of the $\gamma^{\mu\nu\rho}$ and the γ_{ab} gives

$$-i\bar{\epsilon}\gamma^{\mu\nu\rho}\nabla_\mu\nabla_\rho\epsilon = \frac{i}{2}\bar{\epsilon}G_\mu{}^\nu\gamma^\mu\epsilon, \tag{6.52}$$

and, since the spacetime we are considering satisfies the Einstein equations,

$$-i\bar{\epsilon}\gamma^{\mu\nu\rho}\nabla_\mu\nabla_\rho\epsilon = \frac{\chi^2}{4} T_{\text{matter}\,\mu}{}^\nu k^\mu, \tag{6.53}$$

where we have defined the vector k^a as the following real bilinear of the spinor ϵ:

$$k^a = i\bar{\epsilon}\gamma^a\epsilon. \tag{6.54}$$

Now we want to show that k^μ is a non-spacelike vector, by calculating k^2 directly. Using the $d = 4$ Fierz identities *for commuting spinors*, we obtain

$$k^2 = 2(i\bar{\epsilon}\epsilon)^2 + 2(\bar{\epsilon}\gamma_5\epsilon)^2 + \ell^2, \tag{6.55}$$

where we have also defined the real pseudovector ℓ^a,

$$\ell^a = \bar{\epsilon}\gamma^a\gamma_5\epsilon. \tag{6.56}$$

Calculating now ℓ^2 using again the Fierz identities, we obtain

$$\ell^2 = -\tfrac{2}{3}(i\bar{\epsilon}\epsilon)^2 + \tfrac{1}{3}k^2 + \tfrac{2}{3}(\bar{\epsilon}\gamma_5\epsilon)^2, \tag{6.57}$$

from which we obtain

$$k^2 = 2(i\bar{\epsilon}\epsilon)^2 + 4(\bar{\epsilon}\gamma_5\epsilon)^2, \tag{6.58}$$

which is manifestly non-negative because the bilinears $i\bar{\epsilon}\epsilon$ and $\bar{\epsilon}\gamma_5\epsilon$ are real.[11]

On collecting these results and writing $d^3\Sigma_\mu = d^3\Sigma n_\mu$, where n_μ is the non-spacelike unit vector normal to the hypersurface Σ, we find that the integral of the second term in $I(\epsilon)$ is

$$\int_\Sigma d^3\Sigma_\nu\left\{-i\bar{\epsilon}\gamma^{\mu\nu\rho}\nabla_\mu\nabla_\rho\epsilon\right\} = \frac{\chi^2}{4}\int_\Sigma d^3\Sigma\, T_{\text{matter}}{}^{\mu\nu}k_\mu n_\nu. \tag{6.59}$$

The *dominant energy condition*, Eq. (6.46), implies that the second term in $I(\epsilon)$ is non-negative.

As for the second term, let us use a coordinate system in which $n_\mu = \delta_{\mu 0}$ ($\mu = 0, i$). Then, it can be rewritten in the form

$$\int_\Sigma d^3\Sigma(\nabla_i\epsilon)^\dagger_\alpha(\nabla_i\epsilon)^\alpha - \int_\Sigma d^3\Sigma\left(i\gamma^i\nabla_i\epsilon\right)^\dagger_\alpha\left(i\gamma^j\nabla_j\epsilon\right)^\alpha. \tag{6.60}$$

[11] If ϵ were a Majorana spinor, we would have $k^2 = 0$.

These two terms are manifestly positive. The second one vanishes if we use spinors satisfying the *Witten condition*

$$\gamma^i e_i{}^\mu \nabla_\mu \epsilon = 0. \tag{6.61}$$

Thus, we have proven that, if the dominant energy condition is satisfied and we use spinors satisfying the Witten condition, $I(\epsilon)$ is non-negative.

2. We rewrite $I(\epsilon)$ as follows:

$$I(\epsilon) = \tfrac{1}{2} \int_{\partial\Sigma} d^2\Sigma_{\mu\nu} \epsilon^{\mu\nu\rho\sigma} \bar{\epsilon} \gamma_5 \gamma_\sigma \nabla_\rho \epsilon, \tag{6.62}$$

and expand the integrand around the vacuum $\bar{g}_{\mu\nu}$ (Minkowski spacetime) to which the solution asymptotically tends. We also impose on the chosen spinors that they admit the expansion

$$\epsilon = \epsilon_0 + \mathcal{O}\left(\frac{1}{r}\right), \tag{6.63}$$

where $r \to \infty$ at spatial infinity and

$$\bar{\nabla}_\mu \epsilon_0 = 0. \tag{6.64}$$

A spinor satisfying this condition in $N = 1$ SUGRA is a *Killing spinor* of the solution $\bar{g}_{\mu\nu}$. Since $\nabla_\mu = \bar{\nabla}_\mu - \frac{1}{4}\Delta\omega_\mu{}^{ab}\gamma_{ab}$ and the integral is taken at spatial infinity,

$$\begin{aligned} I(\epsilon) &= -\tfrac{1}{8}\int d^2\Sigma_{\mu\nu}\epsilon^{\mu\nu\rho\sigma}\bar{\epsilon}_0\gamma_5\gamma_\sigma\nabla_\rho\epsilon_0 = \tfrac{1}{4}\int d^2\Sigma_{\mu\nu}\left[-3g^{\rho\mu\nu,\gamma}{}_{\alpha\beta}\Delta\omega_\rho{}^{\alpha\beta}k_{0\,\gamma}\right] \\ &= \tfrac{1}{4}E(k_0), \end{aligned} \tag{6.65}$$

where we have used Eq. (6.34) and the fact that

$$k_0^a = i\bar{\epsilon}_0\gamma^a\epsilon_0 \tag{6.66}$$

is, trivially, a Killing vector of the vacuum $\bar{g}_{\mu\nu}$. When it is timelike, k_0 is the generator of translations in time and $E(k_0)$ is just the mass.

This proves that $M \geq 0$ and $M = 0$ for Minkowski spacetime.

The above relation between Killing spinors and Killing vectors is quite generic and, with minor adaptations, is true in most SUGRAs. Just as the Killing vectors of a metric constitute a Lie algebra that generates its isometry group, the Killing spinors *and* Killing vectors of a solution of a SUGRA theory, which may involve other fields apart from the metric, constitute a superalgebra that generates a supergroup that leaves the solution invariant. The simplest case is Minkowski spacetime, whose invariance supergroup is the super-Poincaré one.

Part II
Gravitating point-particles

[The Universe] cannot be read until we have learnt the language and become familiar with the characters in which it is written. It is written in mathematical language, and the letters are triangles, circles and other geometrical figures, without which means it is humanly impossible to comprehend a single word.

Galileo Galilei

7
The Schwarzschild black hole

With this chapter we start the study of a number of important classical solutions of GR.[1] There is no doubt that the most important solution is Schwarzschild's, that describes the static, spherically symmetric gravitational field in the absence of matter that one finds outside any static, spherically symmetric object (star, planet . . .). It is this, the simplest non-trivial solution that leads to the concept of a black hole (BH), which affords a privileged theoretical laboratory for *Gedankenexperimente* in classical and quantum gravity.

It is, in fact, a firmly established belief in our scientific community that macroscopic BHs (of the size studied by astrophysicists) are the endpoints of *gravitational collapse* of stars, which, after a long time, gives rise to Schwarzschild BHs if the stars do not rotate. There should be many macroscopic Schwarzschild BHs in our Universe, since many stars have enough mass to undergo gravitational collapse and there is evidence of supermassive BHs in the centers of galaxies.[2] It has been suggested that smaller BHs could have been produced in the Big Bang. Here we are going to be interested in BHs of all sizes, independently of their origin (primordial, quantum-mechanical, astrophysical . . .).

We begin by deriving the Schwarzschild solution and studying its classical properties in order to find its physical interpretation. The physical interpretation of vacuum solutions of the Einstein equations is a most important and complicated point (see [168, 169]) since the source, located by definition in the region in which the vacuum Einstein equations are not solved, is unknown. In the case of the Schwarzschild solution, we will be led to the new concepts of the event horizon and BHs. Some of the classical properties of BHs can be formulated as laws of thermodynamics but, classically, the analogy cannot be complete. It is the existence of Hawking radiation, a quantum phenomenon, that makes the analogy complete and allows us to take it seriously, raising at the same time the problem of the statistical interpretation of the BH (Bekenstein–Hawking) entropy and the BH information problem. Finally, we are going to rederive the expression for the BH entropy in the Euclidean quantum-gravity approach and we are going to generalize our previous results and Schwarzschild's solution to higher dimensions.

[1] Some of these solutions (the Schwarzschild, Reissner–Nordström, Taub–NUT, etc.) are also reviewed in [149] from a different perspective and emphasizing different properties.

[2] For general references on astrophysical evidence for the existence of BHs see, for instance, [214, 693, 801].

There are many excellent books and reviews on these subjects. We would like to mention Frolov and Novikov's book [737], which is the most complete reference on BH physics, Townsend's lectures on BHs [903], the books on quantum-field theory (QFT) on curved spacetimes [157, 936], and the review articles [431, 938].

7.1 Schwarzschild's solution

To solve the vacuum Einstein equations

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R = 0, \quad \Rightarrow R_{\mu\nu} = 0. \tag{7.1}$$

is necessary to make a simplifying *Ansatz* for the metric. The *Ansatz* must, at the same time, reflect the physical properties that we want the solution to enjoy. In this case we want to obtain the metric in the spacetime outside a massive spherically symmetric body that is at rest in a given coordinate system. The latter property is contained in the assumption of *staticity*[3] of the metric and the first in the assumption of spherical symmetry.[4] Under these assumptions, the most general metric can always be cast in the form

$$ds^2 = W(r)(dct)^2 - W^{-1}(r)dr^2 - R^2(r)d\Omega^2_{(2)}, \tag{7.2}$$

where $W(r)$ and $R(r)$ are two undetermined functions of the coordinate r and $d\Omega^2_{(2)}$ is the metric on the unit 2-sphere S^2 (see Appendix C). On substituting this *Ansatz* into the equations of motion one finds (see for instance [932]) a general solution for W and R,

$$W = 1 + \omega/r, \qquad R^2 = r^2, \tag{7.3}$$

with one integration constant ω. We see that the solution is asymptotically flat; i.e. that, as the coordinate r, approaches infinity, the metric approaches Minkowski's. Physically, the requirement of asymptotic flatness means that we are dealing with an isolated system, with a source of gravitational field confined in a finite volume. The constant ω has dimensions of length and we will study its meaning in a moment.

The result is Schwarzschild's solution [840] in *Schwarzschild coordinates* $\{t, r, \theta, \varphi\}$:

$$\boxed{ds^2 = W(dct)^2 - W^{-1}dr^2 - r^2 d\Omega^2_{(2)}, \qquad W = 1 + \omega/r.} \tag{7.4}$$

Let us now review the properties of this solution.

[3] That is, the metric admits a timelike Killing vector with the property of *hypersurface-orthogonality*: the space can be foliated by a family of spacelike hypersurfaces that are orthogonal to the orbits of the timelike Killing vector, and can be labeled by the parameter of these orbits, which takes the same value at any point of each of these hypersurfaces. If the space does not have this property, the explicit dependence of the metric on the associated time coordinate can always be avoided, but there will always be non-vanishing off-diagonal terms in the metric mixing time components with space components, breaking at the same time spherical symmetry: all stationary, spherically symmetric spacetimes are also static.

[4] Invariance under the group SO(3) of spatial rotations in $d = 4$.

7.1.1 General properties

1. *Schwarzschild's is the only spherically symmetric solution of $R_{\mu\nu} = 0$ (static or not).* This is Birkhoff's theorem [155]. A simple proof can be found in [242, 707].

2. Schwarzschild's solution is stable under small perturbations, gravitational or associated with external fields [232]: the perturbations in the geometry grow small with time, being carried by waves to either $r \to \infty$ or $r \to 0$.

3. The integration constant ω is, in principle, arbitrary. It has the following meaning: for large values of r, where the gravitational field is weak, the trajectories of massive test particles (geodesics) approach the Keplerian orbits that they would describe if they were subject to the Newtonian gravitational field produced by a spherically symmetric object of total mass

$$M = -\frac{\omega c^2}{2G_{\mathrm{N}}^{(4)}}, \tag{7.5}$$

centered at $r = 0$. Then we can identify M with the mass of the object we are describing in GR and $-\omega$ is the Schwarzschild radius associated with such an object, defined in Eq. (4.7)

$$\omega = -R_{\mathrm{S}}. \tag{7.6}$$

We can arrive at the same conclusion by using the ADM mass formula Eq. (6.15) that we rewrite here for convenience,

$$M = \frac{c^2}{8\pi G_{\mathrm{N}}^{(4)}} \int d^2 S_i \left(\partial_j h_{ij} - \partial_i h_{jj}\right). \tag{7.7}$$

Therefore, M is the (ADM) mass of the Schwarzschild solution and it is taken to be positive for two reasons: first, nobody has seen an object with negative gravitational mass; and second, the Schwarzschild solution with negative mass has a *naked singularity*, as we will explain later, which is thought to be unacceptable on physical grounds.

4. We conclude that, as we wanted, the Schwarzschild metric describes the gravitational field created by a spherically symmetric, massive object as seen from far away (in the vacuum region) by a static observer to whom the above (Schwarzschild) coordinates $\{t, r, \theta, \varphi\}$ are adapted.[5]

5. Usually, the Schwarzschild solution is used from $r = \infty$ to some finite value $r = r_{\mathrm{E}} > R_{\mathrm{S}}$ (we will see why we have to have $r_{\mathrm{E}} > R_{\mathrm{S}}$) where it is glued to another

[5] Actually, the coordinate r does not have, *a priori*, the meaning of a radius, even though we have been referring to it as the *radial coordinate*. There are smooth, topologically non-trivial solutions with spherical symmetry but no center [643]. In the Schwarzschild solution r has the meaning of a radius only asymptotically, as we are going to see. Sometimes it is called the *area radius* because its meaning, anywhere, is that surfaces of constant t and r are spheres with area $4\pi r^2$.

static, spherically symmetric metric that is a solution of the Einstein equations for some matter energy–momentum tensor appropriate to describe a static, spherically symmetric star[6] or any other body whose surface is at $r = r_{\mathrm{E}}$. These metrics, called *Schwarzschild interior solutions*,[7] describe the spacetime in the interiors of stars and Schwarzschild's describes all their exteriors (by virtue of Birkhoff's theorem).

6. The Schwarzschild metric is singular (i.e. det $g_{\mu\nu} = 0$ or certain components of the metric blow up) at $r = 0$, R_{S}. We know that the Schwarzschild metric is physically sensible for large values of r, but we cannot take it seriously for $r \leq R_{\mathrm{S}}$ because we have to go through a singularity.

   The singularity at $r = R_{\mathrm{S}}$ can be physical or merely the result of a bad choice of coordinates (just like the singularity at the origin in the Euclidean plane in polar coordinates). If the singularity is physical, then the region $r \leq R_{\mathrm{S}}$ has nothing to do with the region $r > R_{\mathrm{S}}$ that describes the exterior of massive bodies. However, if the singularity at $r = R_{\mathrm{S}}$ is just a coordinate singularity, we can use another coordinate system that is related to Schwarzschild's in the region $r > R_{\mathrm{S}}$ by a standard coordinate change but such that the metric is regular at $r = R_{\mathrm{S}}$. The *analytic extension* of the Schwarzschild metric obtained in this way will also cover the region $r < R_{\mathrm{S}}$.

   To find the nature of the singularities it is necessary to perform an analysis of the curvature invariants and of the geodesics.[8]

   - Obviously $R = 0$ and $R_{\mu\nu}R^{\mu\nu} = 0$ because the Schwarzschild metric solves the equations of motion $R_{\mu\nu} = 0$. However, other higher-order curvature invariants do not vanish, for instance, the *Kretschmann invariant*

     $$R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma} = \frac{48M^2\cos^2\theta}{r^6} + \cdots. \tag{7.8}$$

     By examining all of them, one concludes that there is a curvature singularity at $r = 0$ but not at $r = R_{\mathrm{S}}$. This means that the singularity at $r = R_{\mathrm{S}}$ could be a coordinate singularity, but the singularity at $r = 0$ is certainly a physical singularity that will be present in any coordinate system.
   - If an observer[9] with rest mass m moves in the Schwarzschild field, its equation of motion obeys the general mass-shell constraint Eq. (3.259),

     $$g_{\alpha\beta}p^\alpha p^\beta = m^2c^2, \tag{7.9}$$

[6] It is possible to prove that all stellar models describing isolated stars in equilibrium have spherically symmetric metrics [94, 656].

[7] For more details see, for instance, [932].

[8] A general reference on the analysis of singularities is [243].

[9] Traditionally, in this *Gedankenexperiment* the observer sent to probe the Schwarzschild gravitational field at $r = R_{\mathrm{S}}$ is a graduate student who periodically sends reports to his/her advisor, who sits comfortably away from that point. If the gravitational field at the advisor's position is weak enough, the proper time will be well approximated by Schwarzschild's time t. We will break this cruel custom by referring to the former as a free-falling observer and to the latter as the Schwarzschild observer.

where $p^\alpha = -m dx^\alpha/d\tau$ is the observer's four-momentum, τ is the observer's proper time, and we have set $\xi = c\tau$. On the other hand, since the Schwarzschild metric admits a timelike Killing vector $k^\mu = \delta^{\mu 0}$, the observer's motion has an associated conserved momentum $p(k) \equiv p^0$ given by Eq. (3.266) that we can identify with the observer's total energy

$$E = -p^0 c. \tag{7.10}$$

To simplify the calculations, we assume that the observer has only radial motion (i.e. zero angular momentum) so $p^\theta = p^\varphi = 0$. Then, using the conservation of the energy, the mass-shell constraint becomes a simple equation for p^r, which is a differential equation for r,

$$\left(\frac{dr}{cd\tau}\right)^2 = \left(\frac{E}{mc^2}\right)^2 - W, \tag{7.11}$$

that can be integrated to give the total proper time:

$$\tau = \frac{1}{c}\int_{r_1}^{r_2} dr \left(\frac{R_\mathrm{S}}{r} - \frac{R_\mathrm{S}}{R_0}\right)^{-\frac{1}{2}}, \tag{7.12}$$

where

$$R_0 = \frac{R_\mathrm{S}}{1 - [E/(mc^2)]^2} \tag{7.13}$$

is the value of the radial coordinate r for which the speed of the observer is zero and $E = mc^2$.

We can use the above expression to calculate how long it takes for the free-falling observer to go from $r = R_0 > R_\mathrm{S}$ to the curvature singularity at $r = 0$, going through the surface $r = R_\mathrm{S}$. The answer is, surprisingly, finite:

$$\Delta\tau = \frac{\pi}{2c} R_0 \left(\frac{R_0}{R_\mathrm{S}}\right)^{\frac{1}{2}}. \tag{7.14}$$

This confirms that nothing unphysical happens at $r = R_\mathrm{S}$ and that the singularity is only a problem of Schwarzschild's coordinates. It should, then, be possible to find a coordinate system which is not singular there.[10]

This is essentially the idea on which the Eddington–Finkelstein coordinates $\{v, r, \theta, \varphi\}$ are based [345, 394]. In these coordinates the Schwarzschild solution takes the form

$$ds^2 = W dv^2 - 2dvdr - r^2 d\Omega^2_{(2)}, \tag{7.15}$$

where the coordinate v is related to t and r in the region $r > R_\mathrm{S}$ by

$$v = ct + r + R_\mathrm{S} \ln |W|, \tag{7.16}$$

[10] There is, of course, another issue: whether the tidal forces at the horizon are big or small. For big enough Schwarzschild BHs they are small, but this might not be a universal behavior of BHs [555, 556].

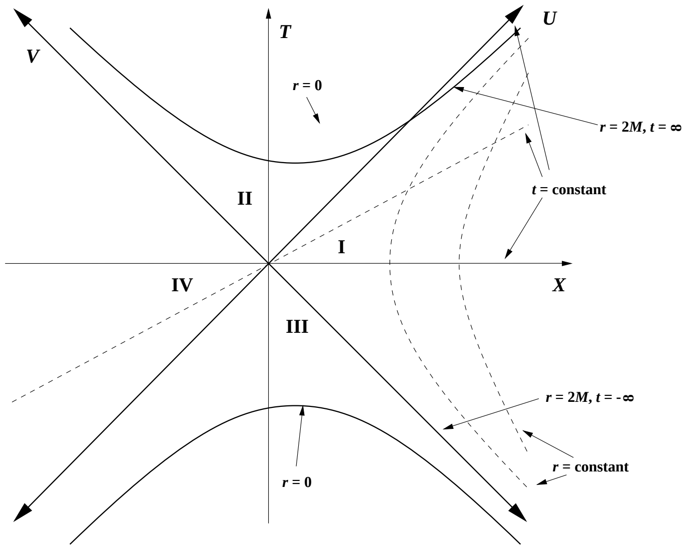

Fig. 7.1. Kruskal coordinates for Schwarzschild's solution. We have set $c = G_{\rm N}^{(4)} = 1$ and each point corresponds to a 2-sphere of radius $r(T, X)$.

and is constant for light like radial geodesics (the worldlines of free-falling photons). This metric is regular everywhere except at $r = 0$, as expected, and analytically extends the original solution to the region $r < R_{\rm S}$, allowing its study. Observe that, for the Schwarzschild observer, things look quite different, though. The proper time of the Schwarzschild observer (equal to the Schwarzschild time t) is related to the proper time of the free-falling observer τ by

$$\frac{dt}{d\tau} = \frac{E/mc^2}{1 - R_{\rm S}/r}, \tag{7.17}$$

and will approach infinity when r approaches $r = R_{\rm S}$. This infinite redshift factor is related to the singularity of the Schwarzschild metric in Schwarzschild coordinates. This seemingly paradoxical disagreement between the two observers is, however, not inconsistent, because, as we are going to see, the two observers cannot compare their observations.

7. To study the region $r < R_{\rm S}$ it is more convenient to use the Kruskal–Szekeres [641, 876] coordinates $\{T, X, \theta, \varphi\}$ that provide the maximal analytic extension of the Schwarzschild metric, describing regions not covered by the Eddington–Finkelstein coordinates (see Figure 6). The region covered by the original Schwarzschild coordinates is just the first quadrant in the figure, whereas the Eddington–Finkelstein coordinates cover the first two quadrants, separated by the $r = R_{\rm S}$ line. There are two additional regions in the quadrants III and IV. Of course, the curvature singularity at $r = 0$ is also present in these new coordinates. The Schwarzschild

metric in Kruskal–Szekeres coordinates takes the form

$$ds^2 = \frac{4R_S^3 e^{\frac{-r}{R_S}}}{r}\left[(dcT)^2 - dX^2\right] - r^2 d\Omega_{(2)}^2, \tag{7.18}$$

where r is a function of T and X that is implicitly given by the coordinate transformations between the pairs t and r and T and X:

$$\begin{aligned}\left(\frac{r}{R_S} - 1\right) e^{\frac{r}{R_S}} &= X^2 - c^2T^2, \\ \frac{ct}{R_S} = \ln\left(\frac{X + cT}{X - cT}\right) &= 2 \operatorname{arctanh}(cT/X),\end{aligned} \tag{7.19}$$

so the Schwarzschild time t is an angular coordinate and the constant-r lines are similar to hyperbolas that asymptotically approach the $X = \pm cT$ lines.

A convenient feature of the Kruskal–Szekeres coordinates is that the T, X part is conformally flat and at each point in the T, X plane the light cones have the same form as in Minkowski spacetime and no particle can have a worldline forming an angle smaller than $\pi/4$ with the X axis. The $r = R_S$ lightlike hypersurface that separates these two quadrants (I, the *exterior*, and II, the *interior*) is called the *event horizon*. It is then clear that particles or signals can go from the exterior to the interior but no signal or particle (including light signals) can go from the interior to the exterior. For this reason, the object described by the full Schwarzschild metric (with no star at $r_E > R_S$) is called a *black hole* (BH).

The existence of an event horizon has very important consequences. First, the free-falling observer can never come back from the BH and cannot send any information that contradicts the Schwarzschild observer's experiences. In this way, the two different observations are made compatible, completely against our classical intuition. Second, it is impossible for the Schwarzschild observer to have any experience of the physical singularity at $r = 0$. This is pictorially expressed by saying that "the singularity is covered by the event horizon."

8. There is another kind of diagram that can be useful for studying the causal structure of the spacetime: Penrose diagrams (see, for instance, [508]). They are obtained by performing a conformal transformation of the metric (that preserves the light-cone structure) such that the infinity is brought to a finite distance in the new metric. A Penrose diagram of Schwarzschild's spacetime is drawn in Figure 7.2.

   Apart from the existence of an event horizon, we also see clearly in this diagram that the fate of the free-falling observer will always be to reach the singularity $r = 0$, which is now a spacelike hypersurface in which he/she will be crushed by infinite tidal forces.

9. We know that there are many objects in the Universe whose gravitational fields are very well described by a region $r > r_E > R_S$ of the Schwarzschild spacetime, but what kind of object gives rise to the $r < R_S$ region, that is, to the BH metric?

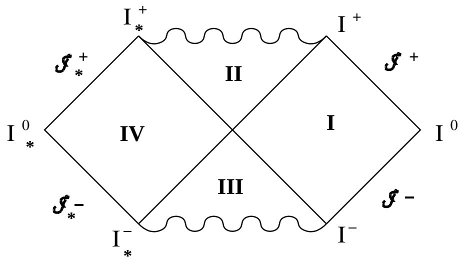

Fig. 7.2. A Penrose diagram of Schwarzschild's spacetime.

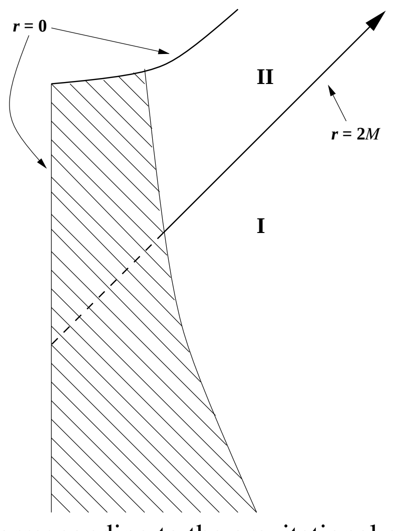

Fig. 7.3. The spacetime corresponding to the gravitational collapse of a star.

This question can be answered only by inventing a new kind of object, the BH, which is, by definition, an object giving rise to a spacetime with an event horizon.

How are BHs created in the Universe? In Thorne's book [885] the story is told of how, in a process that took almost 50 years, the scientific community arrived at the conclusion that BHs could originate from the *gravitational collapse* of very massive stars and, furthermore, that the gravitational collapse would be unavoidable if the star had a mass a few times the Sun's.

It is evident that the spacetime described by the maximally extended Schwarzschild solution cannot originate from a gravitational collapse (there is no star in the past). Instead it describes an *eternal BH*. In Figure 9 the spacetime corresponding to the spherically symmetric gravitational collapse of a star has been represented in Kruskal–Szekeres-like coordinates. The dashed region represents the star's interior and the exterior is just Schwarzschild's spacetime in Kruskal coordinates. The BH appears when the collapsing star has a radius smaller than R_S.

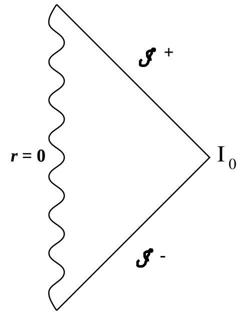

Fig. 7.4. The Penrose diagram of the typical spacetime of a naked singularity.

At this point it may seem exaggerated to assume that the collapse of any star, in any initial state, is going to give rise to a Schwarzschild-like BH. We will elaborate on this crucial point in a moment.

10. When M is negative, there is no horizon covering the singularity at $r = 0$ and it could be "seen" by all observers (see the Penrose diagram in Figure 7.4) that could be causally affected by it.

    This can be the source of many problems and, to avoid them, one can then argue that such a metric, with a singularity that can be seen from infinity, will never be the endpoint of the gravitational collapse of an ordinary star (or any kind of matter with a physically acceptable energy–momentum tensor). This is the essence of Penrose's *cosmic-censorship hypothesis*, which he first suggested in [762] (see also [765, 767] and the reviews [244, 855, 931]), in its weak form.[11] There is a strong relation between cosmic censorship and the positivity of energy. Since the gravitational binding energy is negative, when a cloud of self-gravitating matter starts compressing itself, the total energy diminishes more and more and eventually it would become negative. Before that happens an event horizon should appear.

11. There must be other BHs apart from Schwarzschild's: those corresponding to intermediate states of the gravitational collapse of a star, those that result from perturbing a Schwarzschild BH, or those that describe the gravitational collapse of an electrically charged star. Furthermore, a star can be in many possible states and it is reasonable to think that they will give rise to many different BHs (or to the same BH in many different states).

    On the contrary, the analysis of the perturbations of the Schwarzschild BH [791, 792] shows that all perturbations decay and, after a time of the order of the Schwarzschild radius, the perturbed BH will be Schwarzschild's,[12] determined solely

[11] In its strong form, the cosmic-censorship hypothesis states that, in physically acceptable spacetimes, no singularity, except for initial (Big-Bang) singularities is ever visible to any observer. Rigorous formulations can be found in [932].

[12] Or, in general, the Kerr–Newman BH, which is entirely determined by the mass M, the electric charge Q, and the angular momentum J. For simplicity we are going to ignore angular momentum.

by M, independently of the initial state of the gravitational collapse that originated it and of how it was perturbed. All the higher multipole momenta of the gravitational field (quadrupolar and higher[13]) and of the electromagnetic field (dipole and higher[14]) [96, 880, 929] and all momenta of any scalar field are radiated away to infinity so the resulting BH is always a Schwarzschild BH ($M \neq 0,\ Q,\ J = 0$), a Kerr BH ($M,\ J \neq 0,\ Q = 0$), a Reissner–Nordström BH ($M,\ Q \neq 0,\ J = 0$), or a Kerr-Newman BH ($M,\ J,\ Q \neq 0$).

Nevertheless, it is conceivable that there might be BH solutions with higher momenta of the gravitational and electromagnetic fields or with a non-trivial scalar field that is not created by perturbations or by gravitational collapse. However, it can be shown (*uniqueness theorems*[15]) that the only BH in the absence of angular momentum and other fields is Schwarzschild's [593], that with electric charge is the Reissner–Nordström BH [594], and that with mass and angular momentum is Kerr's [216, 928]. Furthermore, there are no BHs with a non-constant scalar field[16] [100, 234, 689, 872].

This does not mean that there are no solutions with the forbidden momenta: they actually exist but they are not BHs, they do not have an event horizon, and they have naked singularities. A simple example is the family of static, spherically symmetric solutions with a non-trivial scalar field discussed in Section 8.1. A more complex example is provided by Bonnor's magnetic-dipole solution[17] [166].

We conclude that there cannot be BHs with other characteristics (*hairs*) different from M, J, and Q (and, in general, other locally conserved charges). Although this has not been fully proven in all cases [101, 268, 532], this suggests that stationary BHs "have no hair" [814] (the *no-hair conjecture*, which has a somewhat imprecise formulation).

We would like to make two comments about this conjecture.

(a) Given that the presence of hair is associated with the absence of an event horizon, the no-hair conjecture is intimately related to cosmic censorship: for an event horizon to form in gravitational collapse, all the higher momenta of the fields have to be radiated away. Cosmic censorship is related to the positivity of energy[18] and, thus, the no-hair conjecture also is. Non-stationary BHs with scalar hair and positive energy are also known to exist [744], but the cosmic censorship and "baldness" conjectures tell us that the hair must disappear in the evolution of the BH toward a stationary state. This is possible because the "scalar charge" is not a locally conserved charge.

[13] The monopole momentum is the mass M and the dipole momentum is determined by the angular momentum J.
[14] The monopole momentum of the electromagnetic field is the electric charge.
[15] Two reviews on uniqueness theorems containing many references are [531, 533].
[16] This statement will be made more precise in coming chapters.
[17] For a physical interpretation see [371] and, for generalizations, see [295].
[18] If negative kinetic energies are allowed, BHs with non-trivial scalar fields are possible.

(b) Since the gravitational collapse of many different systems always gives rise to the same BHs, characterized by a very small number of parameters, it is natural to wonder what has happened to all the information about the original state. This is essentially the *BH information problem*, which can be stated more precisely in quantum-mechanical language. Furthermore, it is also natural to attribute to the BHs a very big entropy that we should be able to compute by standard statistical methods if we knew all the BH microstates that a BH characterized by M, Q, and J can be in. This is the essence of the *BH entropy problem*. To solve these two problems, we need a theory of quantum gravity.

12. The event horizons of stationary BHs are usually *Killing horizons*, hypersurfaces that are invariant under one isometry wherein the modulus of the corresponding Killing vector k^μ of the metric vanishes, $k^2 = 0$. In the Schwarzschild case, $k^\mu = \delta^{\mu t}$ and generates translations in time: $k^2\big|_{r=R_S} = g_{tt}|_{r=R_S} = 0$. Furthermore, the horizon hypersurface $r = R_S$ is, as a whole, time-translation-invariant.

    Killing horizons (and, hence, event horizons) are *null hypersurfaces*.[19] Furthermore, for each value of t, the Killing horizon is a two-sphere of radius R_S. This is the only topology allowed according to the *topological-censorship* theorems [407, 507]. Like many other important results in GR, these theorems depend heavily on energy-positivity conditions and, thus, it is not surprising that they break down in the presence of a negative cosmological constant and then it is possible to find *topological black holes* [42, 156, 186, 202, 203, 572, 628, 629, 648, 649, 650, 683, 684, 859, 921] whose event horizons can have the topology of any compact Riemann surface. In particular, generalizations of the asymptotically anti-de Sitter Schwarzschild BH with horizons with the topology of Riemann surfaces of arbitrary genus were given in [921].

13. The area of the event horizon is

$$A = \int_{r=R_S} d\Omega^2\, r^2 = 4\pi R_S^2. \tag{7.20}$$

    Hawking proved in [512] that the Einstein equations imply that the area A of the event horizon of a BH never decreases with time. On top of this, if two BHs coalesce to form a new BH, the area of the horizon of this final BH is larger than the sum of the areas of the horizons of the initial BHs. (This result holds for more general kinds of BHs having electric charge and angular momentum.)

    There is a clear analogy between the area A of a BH event horizon and the entropy of a thermodynamical system as never decreasing quantities [95, 97, 98, 241], which deserves to be investigated further.

[19] That is, the vector field normal to a Killing horizon is a null vector field. This vector field, due to the Lorentzian signature, always belongs to the tangent space of the null hypersurface.

14. For Killing horizons one can define, following Boyer [177], the quantity known as *surface gravity* κ, given by the formula

$$\kappa^2 = -\tfrac{1}{2}(\nabla^\mu k^\nu)(\nabla_\mu k_\nu)\big|_{\text{horizon}} . \tag{7.21}$$

If $\kappa \neq 0$ the Killing horizon is part of a *bifurcate horizon*, whereas if $\kappa = 0$ it is a *degenerate Killing horizon.*

In the particular case of static spherically symmetric metrics, which can always be written like this,

$$ds^2 = g_{tt}(r)dt^2 + g_{rr}(r)dr^2 - r^2 d\Omega^2_{(2)}, \tag{7.22}$$

the Killing vector k^μ is just $\delta^{\mu t}$ and the surface gravity takes the value

$$\kappa = \tfrac{1}{2}\frac{\partial_r g_{tt}}{\sqrt{-g_{tt}g_{rr}}}c, \tag{7.23}$$

which for the Schwarzschild BH is the non-vanishing constant

$$\kappa = \frac{c^4}{4G_{\rm N}^{(4)}M}. \tag{7.24}$$

It can be shown that the surface gravity is also constant over the horizon in more general cases [85, 217, 509]. This is analogous to the fact that the temperature is the same at any point of a system in thermodynamical equilibrium and it constitutes the first analogy between the surface gravity and the BH temperature (and the second between a BH and a thermodynamical system). Physically, the surface gravity is the force that must be exerted at ∞ to hold a unit mass in place when $r \to R_{\rm S}$ and has dimensions of acceleration, LT^{-2}.

15. Another set of coordinates that is useful in some problems is *isotropic coordinates* $\{t, \vec{x}_3\}$ with $\vec{x}_3 = (x^1, x^2, x^3)$ in which the three-dimensional constant-time slices are conformally flat and isotropic. The change of coordinates is given by

$$r = \left(\rho - \frac{\omega}{4}\right)^2 \Big/ \rho, \tag{7.25}$$

and the metric takes the form

$$\boxed{ds^2 = \left(1 + \frac{\omega/4}{\rho}\right)^2 \left(1 - \frac{\omega/4}{\rho}\right)^{-2} dt^2 - \left(1 - \frac{\omega/4}{\rho}\right)^4 d\vec{x}_3^{\,2},} \tag{7.26}$$

where $d\rho^2 + \rho^2 d\Omega^2_{(2)} \equiv d\vec{x}_3^{\,2}$ and $\rho = |\vec{x}_3|$.

16. Yet another system of coordinates: let us consider some arbitrary coordinate system $\{y^\alpha\}$ and let us take four scalar functions labeled by $\mu = 0, 1, 2, 3$ of the coordinates y^α and $H^\mu(y)$, which we require to be harmonic

$$\nabla^2 H^\mu = \frac{1}{\sqrt{|g|}} \partial_\alpha \left(\sqrt{|g|}\, g^{\alpha\beta} \partial_\beta H^\mu \right) = 0. \tag{7.27}$$

Now we can define new coordinates $x^\mu \equiv H^\mu(y)$, which are called *harmonic coordinates*. In the system of harmonic coordinates, the above equation takes the form of a condition on the metric:

$$\partial_\alpha \left(\sqrt{|g|}\, g^{\alpha\mu} \right) = 0. \tag{7.28}$$

If we expand the metric in a perturbation series around flat spacetime,

$$\begin{aligned}
g_{\mu\nu} &= \eta_{\mu\nu} + \chi h^{(0)}{}_{\mu\nu} + \chi^2 h^{(1)}{}_{\mu\nu} + \cdots, \\
g^{\mu\nu} &= \eta^{\mu\nu} - \chi h^{(0)\,\mu\nu} + \chi^2 \big(h^{(0)\,\mu\rho} h^{(0)}{}_\rho{}^\nu - h^{(1)\,\mu\nu} \big), \\
g &= 1 + \chi h^{(0)} + \chi^2 \big[h^{(1)} + \tfrac{1}{2} \big(h^{(0)\,2} - h^{(0)\,\mu\nu} h^{(0)}{}_{\mu\nu} \big) \big], \\
\sqrt{|g|} &= 1 + \tfrac{1}{2} \chi h^{(0)} + \tfrac{1}{4} \chi^2 \big[2h^{(1)} + h^{(0)\,2} - 2h^{(0)\,\mu\nu} h^{(0)}{}_{\mu\nu} \big], \\
\sqrt{|g|}\, g^{\mu\nu} &= \eta^{\mu\nu} - \chi \bar{h}^{(0)\,\mu\nu} + \chi^2 \big[-h^{(1)}{}_{\mu\nu} h^{(0)\,\mu\rho} h^{(0)}{}_\rho{}^\nu - h^{(0)} h^{(0)\,\mu\nu} \\
&\qquad + \tfrac{1}{4} \big(2h^{(1)} + h^{(0)\,2} - 2h^{(0)\,\alpha\beta} h^{(0)}{}_{\alpha\beta} \big)\, \eta^{\mu\nu} \big],
\end{aligned} \tag{7.29}$$

where, as usual, $h \equiv h^\rho{}_\rho$ and $\bar{h}^{(0)}{}_{\mu\nu} \equiv h^{(0)}{}_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} h^{(0)}$. On substituting these into the above equation, we find that the linear perturbation $h^{(0)}{}_{\mu\nu}$ of the metric in harmonic coordinates is in the *harmonic gauge*, Eq. (3.57), but the next order is not.

To set the Schwarzschild solution in a harmonic coordinate system it turns out that we just have to shift the Schwarzschild radial coordinate $r \equiv r_{\rm h} - \omega/2$ to obtain

$$ds^2 = \left(\frac{r_{\rm h} + \omega/2}{r_{\rm h} - \omega/2} \right) (dct)^2 - \left(\frac{r_{\rm h} - \omega/2}{r_{\rm h} + \omega/2} \right) dr_{\rm h}^2 + (r_{\rm h} - \omega/2)^2\, d\Omega^2_{(2)}, \tag{7.30}$$

and reexpress the metric in terms of coordinates $\vec{x}_3$ (having nothing to do with the isotropic coordinates introduced before) such that $r_{\rm h} = |\vec{x}_3|$ using $r_{\rm h} dr_{\rm h} = \vec{x}_3 \cdot d\vec{x}_3$

and $d\vec{x}_3^{\,2} = dr_{\rm h}^2 + r_{\rm h}^2 d\Omega_{(2)}^2$:

$$\boxed{\begin{aligned} ds^2 = \left(\frac{r_{\rm h} + \omega/2}{r_{\rm h} - \omega/2}\right)(dct)^2 - \left(1 - \frac{\omega/2}{r_{\rm h}}\right)^2 d\vec{x}_3^{\,2} \\ - \left(\frac{r_{\rm h} - \omega/2}{r_{\rm h} + \omega/2}\right)\frac{(\omega/2)^2}{r_{\rm h}^4}(\vec{x}_3 \cdot d\vec{x}_3)^2. \end{aligned}} \tag{7.31}$$

This is the metric whose first two non-trivial terms in a perturbative series expansion, Eq. (3.217), we obtained in Chapter 3 by imposing self-consistency of the SRFT of a spin-2 particle. Observe that the metric Eq. (3.217) has no event horizon. Only if we calculated all the higher-order corrections and summed them to obtain an exact solution of Einstein's equations could we obtain an event horizon. In this sense, BHs are a highly non-perturbative phenomenon.

The differences between GR and the SRFT of the spin-2 particle also become manifest when one compares the causal structures and the asymptotic behaviors. It seems that it is possible to make compatible either of them but not both for Schwarzschild's spacetime and the Minkowski spacetime in which the SRFT of gravity is defined [766]. This may be a serious problem for any SRFT of gravity.

The fact that we obtained the first approximation to the Schwarzschild solution by solving the Fierz–Pauli equation in the presence of a massive point-like source may lead us to think that the full solution also corresponds to a point-like source. This is an interesting point that we are going to discuss in the next section.

7.2 Sources for Schwarzschild's solution

We would like to identify the object which is the source of the full Schwarzschild gravitational field (with no interior solution). Although we have found it by solving the vacuum Einstein equations, it has a singularity ($r = 0$) where the sourceless Einstein equations are not solved and we can proceed by analogy with the Maxwell case: if we solve Maxwell's equations in vacuum imposing spherical symmetry and staticity, we find the Coulomb solution $A_\mu = \delta_{t\mu} q/(4\pi r)$, which is singular at $r = 0$ and there the equations are not solved. However, one can add at $r = 0$ a singular source corresponding to a point-like electric charge. The Maxwell equations are then solved everywhere by the Coulomb solution and one can say that the source of the Coulomb field is a point-like electric charge. The solution, however, is not completely consistent since the equations of motion of the charged particle in its own electric field are not solved because this diverges at the position of the particle. This is a well-known problem of the classical model of the electron[20] that the quantum theory solves.

[20] For a review see e.g. [884].

There are several reasons why we can expect a negative result: first of all, if the source for the Schwarzschild field were a massive point-particle, it would give rise to a timelike singularity along its worldline, but we know that the Schwarzschild singularity is spacelike. Second, the source for the gravitational field is not just mass, but any kind of energy, including the gravitational field itself. Thus, even if we have a mass distribution confined to a finite region of space (in an idealized case, a point), the gravitational field that it generates will fill the whole space and the source (mass and field) will not be confined to that region. In a sense this is already taken care of by Einstein's equations: in our construction of the self-consistent spin-2 theory we saw that the Einstein tensor contains the "gravitational energy–momentum (pseudo)tensor" and only the matter sources are on the r.h.s. of Einstein's equations.

Anyway, we are going to check explicitly that the massive point-particle cannot be the source for the Schwarzschild metric. This calculation will prepare us for future calculations of the same kind, which, in contrast, will be successful and will help us to understand the reason why.

We consider the action for a massive particle coupled to gravity (we ignore boundary terms):

$$S[g_{\mu\nu}, X^{\mu}(\xi)] = \frac{c^3}{16\pi G_{\rm N}^{(4)}} \int d^4x \sqrt{|g|}\, R - Mc \int d\xi \sqrt{|g_{\mu\nu}(X)\dot{X}^{\mu}\dot{X}^{\nu}|}. \tag{7.32}$$

The equations of motion of $g_{\mu\nu}(x)$ and $X^{\mu}(\xi)$ are, respectively,

$$G_{\mu\nu}(x) + \frac{8\pi M G_{\rm N}^{(4)} c^{-2}}{\sqrt{|g|}} \int d\xi \frac{g_{\mu\rho}(X) g_{\nu\sigma}(X)\dot{X}^{\rho}\dot{X}^{\sigma}}{\sqrt{|g_{\lambda\tau}(X)\dot{X}^{\lambda}\dot{X}^{\tau}|}} \delta^{(4)}[X(\xi) - x] = 0, \tag{7.33}$$

$$\gamma^{\frac{1}{2}} M \nabla^2(\gamma) X^{\lambda} + M\gamma^{-\frac{1}{2}} \Gamma_{\rho\sigma}{}^{\lambda}\dot{X}^{\rho}\dot{X}^{\sigma} = 0, \tag{7.34}$$

where

$$\gamma = g_{\mu\nu}(X)\dot{X}^{\mu}\dot{X}^{\nu}. \tag{7.35}$$

In the physical system that we are considering, the Schwarzschild gravitational field is produced by a point-particle that is at rest in the frame that we are going to use (Schwarzschild coordinates). Then, we expect the solution for $X^{\mu}(\xi)$ to be

$$X^{\mu}(\xi) = \delta^{\mu}{}_{0}\xi. \tag{7.36}$$

However, the X^{μ} equations of motion are not satisfied because the component $\Gamma_{00}{}^{r}$ does not vanish at the origin. Actually, it diverges, and we face here the problem of the infinite force that the gravitational field exerts over the source itself, which is similar to the infinite-self-energy problem of the classical electron mentioned at the beginning of this section. We will see that, in certain situations (in the presence of unbroken supersymmetry), this problem does not occur because the divergent gravitational field is canceled out by another divergent field (electromagnetic, scalar...) and the equation of motion of the particle (or *brane*) can be solved exactly.

The above solution for $X^{\mu}(\xi)$ leads to an energy–momentum tensor whose only non-vanishing component is $T_{00} \sim \delta^{(3)}(\vec{x})$. However, on recalculating carefully[21] the components of the Einstein tensor for the Schwarzschild metric, we find that all the diagonal components, not only G_{00}, are different from zero at the origin:

$$\begin{aligned} G_{00} &= -\frac{W}{\sin^2\theta} 4\pi R_{\mathrm{S}} \delta^{(3)}(r), & G_{rr} &= \frac{W^{-1}}{\sin^2\theta} 4\pi R_{\mathrm{S}} \delta^{(3)}(r), \\ G_{\theta\theta} &= -\frac{r^2}{\sin^2\theta} 2\pi R_{\mathrm{S}} \delta^{(3)}(r), & G_{\varphi\varphi} &= \sin^2\theta \; G_{\theta\theta}. \end{aligned} \tag{7.40}$$

This is related to the spacelike nature of the Schwarzschild singularity, as expected. In the cases in which we will be able to identify the source of a solution with a particle (or a *brane*) the singularity of the metric will be non-spacelike.

7.3 Thermodynamics

We have seen in previous sections that, classically, according to the Einstein equations, there are two magnitudes in a Schwarzschild BH, the area A and the surface gravity κ, that behave in some respects like the entropy S and the temperature T of a thermodynamical system. From this point of view the constancy of κ over the event horizon would be the "*zeroth law of BH thermodynamics*" and the never-decreasing nature of A would be the "*second law of BH thermodynamics*." In a thermodynamical system S, T, and the energy E are related by the *first law of thermodynamics*:

$$dE = TdS. \tag{7.41}$$

To take the thermodynamical analogy any further, it is necessary to prove that κ and A are also related to the analog of the energy E by a similar equation. The natural analog for the energy is the BH mass M (times c^2), and, thus, it is necessary to have (the factor of $G_{\mathrm{N}}^{(4)}$ appears for dimensional reasons)

$$dM \sim \frac{1}{G_{\mathrm{N}}^{(4)}} \kappa dA. \tag{7.42}$$

[21] In this calculation one has to be careful to keep singular (δ-like) contributions that are non-zero only at a certain point. These contributions come in two forms. One is the standard four-dimensional identity

$$\partial_i \partial_i \frac{1}{|\vec{x}|} = -4\pi \delta^{(3)}(\vec{x}), \qquad i = 1, 2, 3, \tag{7.37}$$

adapted to spherical coordinates

$$\partial_r \left[r^2 \partial_r \frac{1}{r} \right] = -\frac{4\pi}{\sin\theta} \delta^{(3)}(r), \tag{7.38}$$

and the other one is

$$\partial_r \left(r \frac{1}{r} \right) = \frac{4\pi}{\sin\theta} \delta^{(3)}(r), \tag{7.39}$$

both of which can be checked by partial integration. Here $\delta^{(3)}(\vec{x}) \neq \delta^{(3)}(r)$. The latter is defined by $\int dr d\varphi d\theta \delta^{(3)}(r) = 1$. The result obtained coincides with the one obtained by more rigorous methods in [71, 72].

This relation turns out to be true. The coefficient of proportionality can be determined [95, 241, 857] and the *first law of BH thermodynamics* takes the form

$$\boxed{dM = \frac{1}{8\pi G_N^{(4)}} \kappa dA.} \tag{7.43}$$

There is an integral version of this relation that can be checked immediately (the *Smarr formula* [857]) by simple substitution of the values of κ and A for the Schwarzschild BH:

$$M = \frac{1}{4\pi G_N^{(4)}} \kappa A. \tag{7.44}$$

The above two relations (conveniently generalized to include other conserved quantities such as the electric charge and the angular momentum) seem to hold under very general conditions [85] (see also [446, 534, 935]).

This surprising set of analogies suggests the identification between the area of the BH horizon A and the BH entropy and between the surface gravity κ and the BH temperature. Stimulated by these ideas, the authors of [85] conjectured, giving some plausibility arguments, a "*third law of BH thermodynamics*," namely that "it is impossible by any procedure, no matter how idealized, to reduce κ to zero by a finite sequence of operations." Several specific examples were studied by Wald in [930]. We will comment more on this in the case of the Reissner–Nordström BH.

The analogy is, though, not sufficient to make a full identification. Indeed, as the authors of [85] say,

> It can be seen that $\kappa/(8\pi)$ is analogous to the temperature in the same way that A is analogous to the entropy. It should, however, be emphasized that $\kappa/(8\pi)$ and A are distinct from the temperature and entropy of the BH.
>
> In fact the effective temperature of a BH is absolute zero. One way of seeing this is to note that a BH cannot be in equilibrium with black-body radiation at any non-zero temperature, because no radiation could be emitted from the hole whereas some radiation would always cross the horizon into the BH.

On the other hand, in the identification $A \sim S,\ \kappa \sim T$ it is not clear what the proportionality constants should be (apart from what the dimensional analysis dictates).

Hawking's discovery [510, 511] that, when the quantum effects produced by the existence of an event horizon are taken into account,[22] BHs radiate as if they were black bodies

[22] This was originally done in a semiclassical calculation in which the background geometry is classical and fixed and there are quantum fields around the BH. The existence of an event horizon gives rise to the Hawking radiation but the effect of the Hawking radiation on the BH horizon (*backreaction*) is not taken into account. A pedagogical review of this calculation can be found in [907].

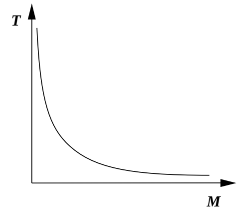

Fig. 7.5. The temperature T versus the mass M of a Schwarzschild black hole.

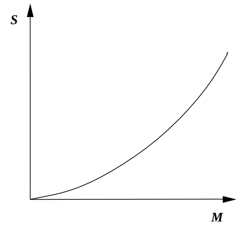

Fig. 7.6. The entropy S versus the mass M of a Schwarzschild black hole.

with temperature[23]

$$\boxed{T = \frac{\hbar\kappa}{2\pi c}} \tag{7.45}$$

dramatically changed this situation. On the one hand, it removed the last obstruction to a complete identification of BHs as thermodynamical systems. On the other, the coefficient of proportionality between κ and T was completely determined, and determined, in turn, that between A and S:

$$\boxed{S = \frac{Ac^3}{4\hbar G_{\mathrm{N}}^{(4)}}.} \tag{7.46}$$

Observe that this relation can be rewritten in this way:

$$S = \frac{1}{32\pi^2}\frac{A}{\ell_{\mathrm{Planck}}^2}, \tag{7.47}$$

that is, essentially the area of the horizon measured in Planckian units, a huge number for astrophysical-size BHs, in agreement with our discussions about the no-hair conjecture. Observe also that the appearance of $\hbar$ in T makes manifest its quantum-mechanical origin.

In particular, for Schwarzschild's BH we have (see Figures 7.5 and 7.6)

$$T = \frac{\hbar c^3}{8\pi G_{\mathrm{N}}^{(4)} M}, \qquad S = \frac{4\pi G_{\mathrm{N}}^{(4)} M^2}{\hbar c}, \tag{7.48}$$

[23] In our units Boltzmann's constant is 1 and dimensionless so T has dimensions of energy, ML^2T^{-2} or L^{-1} in natural units, and the entropy is dimensionless.

and so the first law of BH thermodynamics and Smarr's formula take the forms

$$dMc^2 = TdS, \qquad Mc^2 = 2TS. \tag{7.49}$$

How can a BH from which nothing can ever escape (classically) radiate? The physical mechanism behind the Hawking radiation seems to be the process of Schwinger-pair creation in strong background fields [195, 737], which was originally discovered for electric fields [824], rather than quantum tunneling across the horizon, which would violate causality. In the electric-field case, the background field gives energy to the particles of a virtual pair, separating them. In the BH case, one of the particles in the pair is produced inside the event horizon and the other outside the event horizon. The net effect is a loss of BH mass and the "emission" of radiation by the BH.

The same effect causes the spontaneous discharge of charged bodies (such as a positively charged sphere, say) left in vacuum: if the electric field is strong enough, the electron and positron of a virtual pair can be separated. The electron will move toward the sphere being captured by it, while the positron will be accelerated to infinity. From far away, one would observe a radiation of positrons coming from the charged sphere, whose charge would diminish little by little. In fact, this process is believed to cause the discharge of Reissner–Nordström BHs [218, 289, 429, 686, 716, 737, 750] and was discovered before the publication of Hawking's results.[24] The energy spectrum of the charged pairs produced in an electric field is also thermal [866], but only charged particles are produced and the temperature is different depending on the kind of charged particles considered (electron–positron, proton–antiproton, etc.), whereas in the gravitational case, due to the universal coupling of gravity to all forms of energy, all kinds of particles are produced with thermal spectra with a common Hawking temperature.

The thermodynamics of BHs has several problems or peculiarities.

1. The temperature of a Schwarzschild BH (and of all known BHs far from the *extreme* limit which we will define and discuss later) decreases as the mass (the energy) increases (see Figure 7.5) and therefore a Schwarzschild BH has a negative specific heat (Figure 7.7)
$$C^{-1} = \frac{\partial T}{\partial M} = -\frac{\hbar c^3}{8\pi G_{\mathrm{N}}^{(4)} M^2} < 0, \tag{7.50}$$
and becomes colder when it absorbs matter instead of when it radiates (as ordinary thermodynamical systems do). Thus, a BH cannot be put into equilibrium with an infinite heat reservoir because it would absorb the energy and grow without bounds.

2. The temperature grows when the mass decreases (in the evaporation, for instance) and diverges near zero mass.[25] At the same time the specific heat becomes bigger

[24] It should also be pointed out that the production of particles in the gravitational field of a rotating BH was also discovered before [698, 861, 917, 968], but this is not a purely quantum-mechanical effect, but the quantum translation of the well-known classical *super-radiance* effect.

[25] Precisely when the metric becomes (apparently, smoothly) Minkowski's. The temperature of the Minkowski spacetime is zero, rather than being infinite like the $M \to 0$ limit of the BH temperature. This result is, at first sight, paradoxical, but similar results are, though, very frequent and we will soon meet another one (see the

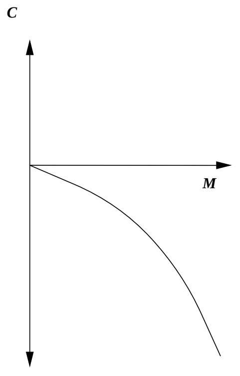

Fig. 7.7. The specific heat C versus the mass M of a Schwarzschild black hole.

in absolute value and stays negative. If these formulae remained valid all the way to $M = 0$, the final stage of the Hawking evaporation of a BH would be a violent explosion in which the BH would disappear. However, when R_S becomes of the order of the BH's Compton wavelength (this happens when $M \sim M_{\text{Planck}}$ and implies that $R_S \sim \ell_{\text{Planck}}$), quantum-gravity effects should become important and should determine (we do not know how) the BH's fate.

3. If a BH can radiate, its entropy can diminish. This is against the second law of BH thermodynamics (which is purely classical). However, the analogy with the second law of thermodynamics can still be preserved because it can be proven that the total entropy (BH plus radiation) never decreases. This is sometimes called the *generalized second law of BH thermodynamics* [99].
4. Returning to the BH information problem, the Hawking radiation seems to carry no more information about the BH than M, J, and Q (just like the metric itself, so it is not so surprising), but we can ask ourselves whether, in the real world, beyond the approximations made, it would carry more information and we may be able to see it in a full quantum computation of the gravitational collapse of matter in a well-defined quantum state and the subsequent evaporation of the resulting BH. For 't Hooft, Susskind, and many others the answer is a definite "yes," namely a BH is just another (peculiar) quantum system and all the information that comes in should unitarily come out: the theory of quantum gravity is unitary. From this point of view, the absorption and radiation of matter by a BH is similar to any standard scattering experiment.

footnote on page 216). *We will very often find that physical properties of a family of metrics parametrized by a number of continuous parameters are not themselves continuous functions of those parameters*. There is no paradox, though, because metrics in that family given by infinitesimally different values of the parameters are not always infinitely close in the space of metrics. Thus, the distance between the Minkowski metric and the Schwarzschild metric with an infinitesimal mass is not infinitesimal. Physically, this is easy to see: no matter how small the mass is, the Schwarzschild spacetime has an event horizon and does not look at all like the Minkowski spacetime.

If no information is carried by Hawking's radiation and the BH evaporates indefinitely, the information about the initial state from which the BH originated is completely lost forever and the theory of quantum gravity governing all these processes is non-unitary, in contrast to all the other physical theories. This is, for instance, Hawking's own viewpoint.

There is a third group that proposes that the information is not carried out of the BH by Hawking radiation but the evaporation process stops at some point, leaving a *BH remnant* containing that information.

There is a little-explored fourth possibility, which is consistent with the classical results on stability of BHs and the no-hair conjecture; namely that the information never enters BHs.

There is, however, no conclusive solution for the BH information problem. In the models based on string theory that we will explain here, BHs are standard quantum-mechanical systems and information is always recovered (even if after a long time).

5. Concerning the BH entropy problem, the statistical-mechanical entropy of systems of fixed energy E is given by

$$S(E) = \ln \rho(E), \tag{7.51}$$

where $\rho(E)$ is the density of states of the system whose energy is E. If a BH is just another quantum-mechanical system with $E = M$, a good theory of quantum gravity should allow us to calculate the Bekenstein–Hawking entropy S from knowledge of the density of BH microstates $\rho(M)$. Also, if that theory exists and the above relation is justified, our knowledge of the Bekenstein–Hawking entropy can be used to find $\rho(M)$ for large values of M (when the quantum corrections are small),

$$\rho(M) \sim \exp M^2. \tag{7.52}$$

We see that the number of BH states with a given mass must grow extremely fast if it is to explain the BH's huge entropy (for a solar-mass BH, $\rho \sim 10^{10^{76}}$). The thermodynamical description of systems whose densities of states grow so fast with the energy is, however, very complicated: the canonical partition function

$$\mathcal{Z}(T) \sim \int dE \rho(E) e^{-\frac{E}{T}} \tag{7.53}$$

diverges whenever $\rho(E)$ grows like e^E or faster. For instance, the density of states of any string theory grows exponentially with the mass and the partition function diverges above *Hagedorn's temperature* (see e.g. [31]). For p-branes [38]

$$\rho(M) \sim \exp\left(\lambda M^{\frac{2p}{p+1}}\right), \tag{7.54}$$

and for $p > 1$ the partition function diverges already at zero temperature. The density of states of BHs must grow faster than that of any of these theories.

As we are going to see, string theory allows us to calculate the entropy and temperature of certain BHs for which this theory provides quantum-mechanical models, from the density of the associated microstates. In this way string theory seems to solve (at least to some extent) the BH entropy and information problems by treating BHs as ordinary quantum-mechanical systems.

7.4 The Euclidean path-integral approach

It is desirable to have an independent and more direct calculation of the BH entropy and temperature. This can be achieved by using the Euclidean path integral as suggested by Gibbons and Hawking [436, 514].

The thermodynamical study of a statistical-mechanical system starts with the calculation of a thermodynamical potential. If there are certain conserved charges C_i (their related potentials being μ_i), it is convenient to work in the grand canonical ensemble, where the fundamental object is the grand partition function

$$\mathcal{Z} = \mathrm{Tr}\, e^{-\beta(H-\mu_i C_i)}, \tag{7.55}$$

and the thermodynamic potential

$$W = E - TS - \mu_i C_i \tag{7.56}$$

is related to the grand partition function by

$$e^{-\frac{W}{T}} = \mathcal{Z}. \tag{7.57}$$

All thermodynamic properties of the system can be obtained from knowledge of $\mathcal{Z}$. In particular, the entropy is given by

$$S = (E - \mu_i C_i)/T + \ln \mathcal{Z}. \tag{7.58}$$

The idea is to calculate the *thermal* grand partition function of quantum gravity through the path integral of a Euclidean version of the Einstein–Hilbert action Eq. (4.26), $\tilde{S}_{\mathrm{EH}}$,

$$\mathcal{Z} = \int Dg\, e^{-\frac{\tilde{S}_{\mathrm{EH}}}{\hbar}}, \tag{7.59}$$

where one has to sum over all metrics with period[26] $\beta = \hbar c/T$. The only modification that has to be made to the Einstein–Hilbert action is the addition of a surface term to normalize the action so that the on-shell Euclidean action vanishes for flat Euclidean spacetime (the vacuum). The Einstein–Hilbert action becomes [436]

$$\boxed{S_{\mathrm{EH}}[g] = \frac{c^3}{16\pi G_{\mathrm{N}}^{(4)}} \int_{\mathcal{M}} d^4x \sqrt{|g|}\, R + \frac{c^3}{8\pi G_{\mathrm{N}}^{(4)}} \int_{\partial\mathcal{M}} d^3\Sigma\, (\mathcal{K} - \mathcal{K}_0),} \tag{7.60}$$

[26] β has dimensions of length if T has dimensions of energy.

where $\mathcal{K}_0$ is calculated by substituting the vacuum metric into the expression for $\mathcal{K}$.

The path integral is now to be calculated in the (semiclassical) saddle-point approximation (from now on we set $\hbar = c = G_{\rm N}^{(4)} = 1$ for simplicity)

$$\mathcal{Z} = e^{-\tilde{S}_{\rm EH}(\text{on-shell})}. \tag{7.61}$$

The classical solution used to calculate the on-shell Euclidean action above is the Euclidean Schwarzschild solution that we now discuss.

7.4.1 The Euclidean Schwarzschild solution

The Euclidean Schwarzschild solution solves the Einstein equations with Euclidean metric (in our case $(-,-,-,-)$). It can be obtained by performing a Wick rotation $\tau = it$ of the Lorentzian Schwarzschild solution. If we use Kruskal–Szekeres (KS) coordinates $\{T, X, \theta, \varphi\}$, we have to define the Euclidean KS time $\mathcal{T} = iT$. This Wick rotation has important effects. The relation between the Schwarzschild coordinate r and T, X coordinates was

$$(r/R_{\rm S} - 1)e^{\frac{r}{R_{\rm S}}} = X^2 - T^2. \tag{7.62}$$

The l.h.s. is bigger than -1 and that is why the X, T coordinates also cover the BH interior. However, in terms of $\mathcal{T}$,

$$(r/R_{\rm S} - 1)e^{\frac{r}{R_{\rm S}}} = X^2 + \mathcal{T}^2 > 0, \tag{7.63}$$

and the interior $r < R_{\rm S}$ of the BH is not covered by the Euclidean KS coordinates. On the other hand, the relation between the Schwarzschild time t and $X, \mathcal{T}$,

$$\frac{X+T}{X-T} = e^{\frac{t}{R_{\rm S}}}, \tag{7.64}$$

becomes

$$\frac{X - i\mathcal{T}}{X + \mathcal{T}} = e^{-2i\ \mathrm{Arg}(X+i\mathcal{T})} = e^{-\frac{i\tau}{R_{\rm S}}}. \tag{7.65}$$

Since $\mathrm{Arg}(X + i\mathcal{T})$ takes values between 0 and 2π (which should be identified), for consistency (to avoid conical singularities) τ must take values in a circle of length $8\pi M$ [436, 969]. The period of the Euclidean time can be interpreted as the inverse temperature β which coincides with the known Hawking temperature. This is the reason why we can use this metric to calculate the thermal partition function.

The result is a Euclidean metric with periodic time that covers only the exterior of the BH (region I of the KS diagram). The $X, \mathcal{T}$ part of the metric describes a semi-infinite "cigar" (times a 2-sphere) that goes from the horizon to infinity with topology $\mathbb{R}^2 \times \mathrm{S}^2$.

Knowing the result beforehand, we could just as well have used Schwarzschild coordinates, which cover smoothly the BH exterior, and proceeded in this much more economical way [546]: given a static, spherically symmetric BH with regular horizon at $r = 0$, the r–τ part of its Euclidean metric can always be put in the form

$$-d\sigma^2 = f(r)d\tau^2 + f^{-1}(r)dr^2 \sim f'(0)rd\tau^2 + \frac{1}{f'(0)r}dr^2, \tag{7.66}$$

near the horizon. Defining another radial coordinate ρ such that $g_{\rho\rho} = 1$, we obtain

$$-d\sigma^2 \sim \big(f'(0)/2\big)^2 d\tau^2 + d\rho^2 \equiv \rho^2 d\tau^2 + d\rho^2. \tag{7.67}$$

Now this metric is just the 2-plane metric in polar coordinates if $\tau' \in [0, 2\pi]$. Otherwise it is the metric of a cone and has a conical singularity at $\rho = 0$ (the horizon). Then, $\tau \in [0, \beta = 4\pi/f'(0)]$.

In practice we do not even need the Euclidean Schwarzschild metric. We need only the information about the period of the Euclidean time (temperature) and the fact that the BH interior disappears (the integration region) and we can simply replace $-\tilde{S}_{\rm EH}$(on-shell) by $+i S_{\rm EH}$(on-shell) because it gives the same result once we take into account the above two points. Thus, in our calculation we will use the Lorentzian Schwarzschild metric in Schwarzschild coordinates using these observations.

7.4.2 The boundary terms

The Euclidean Schwarzschild solution, being a solution of the vacuum Einstein equations, has $R = 0$ everywhere (the singularity $r = 0$ together with the whole BH interior is not included) and only the boundary term contributes to the on-shell action. We are going to calculate its value in this section.

The only boundary of the Euclidean Schwarzschild metric, with the time compactified on a circle of length β, is $r \to \infty$. (If we gave the Euclidean time a different periodicity, there would be another boundary at the horizon, but there is no reason to do this.) This boundary is then the hypersurface $r = r_{\rm c}$ when the constant $r_{\rm c}$ goes to infinity. A vector normal to the hypersurfaces $r - r_{\rm c} = 0$ is $n_\mu \sim \partial_\mu(r - r_{\rm c}) = \delta_{\mu r}$, and, normalized to unity ($n_\mu n^\mu = -1$ because it is spacelike) with the right sign to make it outward-pointing, is, for a generic spherically symmetric metric Eq. (7.22),

$$n_\mu = -\frac{\delta_{\mu r}}{\sqrt{-n^2}} = -\sqrt{-g_{rr}}\,\delta_{\mu r}. \tag{7.68}$$

The four-dimensional metric $g_{\mu\nu}$ induces the following metric $h_{\mu\nu}$ on the hypersurface $r - r_{\rm c} = 0$:

$$ds^2_{(3)} = h_{\mu\nu}dx^\mu dx^\nu = g_{tt}dt^2 - r^2 d\Omega^2_{(2)}\Big|_{r=r_{\rm c}}. \tag{7.69}$$

The covariant derivative of n_μ is

$$\nabla_\mu n_\nu = -\sqrt{-g_{rr}}\left\{\delta_{\mu r}\delta_{\nu r}\partial_r \ln\sqrt{-g_{rr}} - \Gamma_{\mu\nu}{}^r\right\}, \tag{7.70}$$

and the trace of the extrinsic curvature of the $r - r_{\rm c} = 0$ hypersurfaces is (the Christoffel symbols can be found in Appendix F.1)

$$\mathcal{K} = h^{\mu\nu}\nabla_\mu n_\nu = \frac{1}{\sqrt{-g_{rr}}}\left\{\tfrac{1}{2}\partial_r \ln g_{tt} + 2/r\right\}\Bigg|_{r=r_{\rm c}}. \tag{7.71}$$

The regulator $\mathcal{K}_0$ can be found form this expression to be

$$\mathcal{K}_0 = (2/r)|_{r=r_0}. \tag{7.72}$$

On the other hand, for any static, spherically symmetric, asymptotically flat metric we must have for large r

$$g_{tt} \sim 1 - \frac{2M}{r}, \qquad g_{rr} \sim -\left(1 + \frac{2M}{r}\right), \quad \Rightarrow (\mathcal{K} - \mathcal{K}_0)|_{r=r_\mathrm{c}} \sim -M/r_\mathrm{c}^2. \tag{7.73}$$

Finally, we have

$$\begin{aligned} \frac{i}{8\pi}\int_{r_0\to\infty} d^3x\sqrt{|h|}\,(\mathcal{K} - \mathcal{K}_0) &= \lim_{r_\mathrm{c}\to\infty}\frac{i}{8\pi}\int_0^{-i\beta} dt\int_{\mathrm{S}^2} d\Omega^2 r_\mathrm{c}^2\sqrt{g_{tt}(r_\mathrm{c})}\,(\mathcal{K} - \mathcal{K}_0) \\ &= \lim_{r_0\to\infty}\frac{\beta}{2}r_\mathrm{c}^2(\mathcal{K} - \mathcal{K}_0) = -\frac{\beta M}{2}. \end{aligned} \tag{7.74}$$

For Schwarzschild $\beta = 8\pi M$ and Eqs. (7.58) and (7.61) lead to the expected result

$$S = \beta M + \ln\mathcal{Z} = \beta M/2 = 4\pi M^2. \tag{7.75}$$

7.5 Higher-dimensional Schwarzschild metrics

If we consider the d-dimensional vacuum Einstein equations, it is natural to look for the generalization of Schwarzschild's solution: static, spherically symmetric metrics. Here, spherical symmetry means invariance under global $\mathrm{SO}(d-1)$ transformations. The appropriate *Ansatz* that generalizes Eq. (7.2) is

$$ds^2 = W(r)(dct)^2 - W^{-1}(r)dr^2 - R^2(r)d\Omega^2_{(d-2)}, \tag{7.76}$$

where $d\Omega^2_{(d-2)}$ is the metric element on the $(d-2)$-sphere S^{d-2}(see Appendix C).

One finds the following generalization of Schwarzschild's solution [706, 877]:

$$\boxed{ds^2 = W(dct)^2 - W^{-1}dr^2 - r^2 d\Omega^2_{(d-2)}, \qquad W = 1 + \omega/r^{d-3},} \tag{7.77}$$

where $d \geq 4$: there are no Schwarzschild BHs in fewer than four dimensions.[27]

The integration constant ω is related to the d-dimensional analog of the Schwarzschild radius. To establish the above relation between the Schwarzschild radius and the mass, one can use for instance Komar's formula Eq. (6.42) correctly normalized [706]:

$$Mc^2 = -\frac{1}{16\pi G_\mathrm{N}^{(d)}}\frac{d-2}{d-3}\int_{\mathrm{S}^{d-2}_\infty} d^{d-2}\Sigma_{\mu\nu}\nabla^\mu k^\nu. \tag{7.78}$$

The result of the integral is $(d-3)\omega_{(d-2)}\omega c$, with $\omega_{(d-2)}$ given in Eq. (C.11), and, thus

$$\omega = -R_\mathrm{S}^{d-3} = -\frac{16\pi G_\mathrm{N}^{(d)} Mc^{-2}}{(d-2)\omega_{(d-2)}}. \tag{7.79}$$

[27] In the presence of a negative cosmological constant there is, though, an asymptotically AdS_3 three-dimensional solution that can be identified with a BH: the BH of Bañados, Teitelboim, and Zanelli ("BTZ") [82].

The solutions Eq. (7.77) are almost straightforward generalizations of the four-dimensional Schwarzschild solution in every sense. Their most interesting property is the existence of event horizons at $r = R_S$ in all of them, with properties that generalize those of the $d = 4$ ones and lead us to the study of their thermodynamics. The uniqueness of these (static BH) solutions was proved in [444, 589]. There is no uniqueness for stationary BHs in higher dimensions, as the existence of the rotating black ring of [373] shows.

7.5.1 Thermodynamics

In d dimensions, the first law of BH thermodynamics and Smarr's formula are [706]

$$dMc^2 = \frac{d-2}{2(d-3)} T dS, \qquad Mc^2 = \frac{d-2}{d-3} TS, \tag{7.80}$$

where the temperature T is now given in terms of the surface gravity κ by the same expression as in four dimensions Eq. (7.45) while κ is defined by the same formula Eq. (7.21) in any dimension. The entropy is given in terms of the volume of the $(d-2)$-dimensional constant-time slices of the event horizon $V^{(d-2)}$ by

$$\boxed{S = \frac{V^{(d)}}{4\hbar G_N^{(d)}}.} \tag{7.81}$$

The volume and surface gravity of the event horizon are

$$V^{(d-2)} = R_S^{d-2} \omega_{(d-2)}, \qquad \kappa = \frac{(d-3)c^2}{2R_S}, \tag{7.82}$$

and, therefore

$$T = \frac{(d-3)\hbar c}{4\pi R_S}, \qquad S = \frac{R_S^{d-2} \omega_{(d-2)} c^3}{4\hbar G_N^{(d)}}. \tag{7.83}$$

Smarr's formula can be easily checked using these results.

The temperature of the higher d-dimensional BHs can also be calculated in the Euclidean formalism with the criterion of avoiding conical singularities of the τ–r part of the metric on the event horizon. A Euclidean calculation of the entropy may also be done.

8
The Reissner–Nordström black hole

In the previous chapter we obtained and studied the Schwarzschild solution of the vacuum Einstein equations and arrived at the BH concept. However, many of the general features of BHs that we discussed, such as the no-hair conjecture, make reference to BHs in the presence of matter fields. In this chapter we are going to initiate the study and construction of BH solutions of the Einstein equations in the presence of matter fields, starting with the simplest ones: massless scalar and vector fields.

The (unsuccessful) search for BH solutions of gravity coupled to a scalar field will allow us to deepen our understanding of the no-hair conjecture.

The (successful) search for BH solutions of gravity coupled to a vector field will allow us to find the simplest BH solution different from Schwarzschild's: the Reissner–Nordström (RN) solution. Simple as it is, it has very interesting features, in particular, the existence of an extreme limit with a regular horizon and zero Hawking temperature that will be approached with positive specific heat, as in standard thermodynamical systems. Later on we will relate some of these properties to the unbroken supersymmetry of the extreme RN (ERN) solution, which will allow us to reinterpret it as a self-gravitating supersymmetric soliton interpolating between two vacua of the theory.

The ERN BH is the archetype of the more complicated self-gravitating supersymmetric solitons that we are going to encounter later on in the context of superstring low-energy effective actions (actually, one of our goals will be to recover it as a superstring solution) and many of its properties will be shared by them. Furthermore, the four-dimensional Einstein–Maxwell system exhibits *electric–magnetic duality* in its simplest form. Electric-magnetic duality will play a crucial role in many of the subsequent developments either as a classical solution-generating tool or as a tool that relates the weak- and strong-coupling regimes of QFTs.

It is, therefore, very important to study all these properties in this simple system.

In this chapter we are first going to study the coupling of a free massless real scalar to gravity, discussing the (non-)existence of BH solutions and its relation to the no-hair conjecture. Then, we will study the coupling of a massless vector field to gravity (the Einstein–Maxwell system), its gauge symmetry, and the notion and definition of electric charge and its conservation law. Immediately afterwards we will introduce and study the electrically charged RN BH, and its sources, thermodynamics, and Euclidean action. Once we are done with the electrically charged RN BH, we will introduce electric–magnetic duality, the notion and definition of magnetic charge, and the Dirac–Schwinger–Zwanziger quantization condition. Using electric–magnetic duality, we will construct magnetically charged and dyonic RN BHs. Finally, we will consider higher-dimensional RN BH solutions.

8.1 Coupling a scalar field to gravity and no-hair theorems

The simplest field to which we can couple gravity is a free (vanishing potential) massless real scalar field φ. The action of this system is (choosing the simplest normalization)

$$\boxed{S[g_{\mu\nu}, \varphi] = S_{\text{EH}} + \frac{c^3}{8\pi G_{\text{N}}^{(4)}} \int d^4x \sqrt{|g|}\, \partial_\mu\varphi\partial^\mu\varphi.} \tag{8.1}$$

The equations of motion for the metric and the scalar are

$$G_{\mu\nu} + 2\left[\partial_\mu\varphi\partial_\nu\varphi - \tfrac{1}{2}g_{\mu\nu}(\partial\varphi)^2\right] = 0, \qquad \nabla^2\varphi = 0. \tag{8.2}$$

If we take the divergence of the Einstein equation above and use the contracted Bianchi identity $\nabla^\mu G_{\mu\nu} = 0$, one obtains the equation

$$\nabla^2\varphi\nabla_\nu\varphi = 0, \tag{8.3}$$

which implies the equation of motion for the scalar field φ if $\nabla_\nu\varphi \neq 0$. If $\nabla_\nu\varphi = 0$ the scalar equation of motion is automatically solved and, thus, we can say that the Einstein equations imply the scalar field equation of motion and we only have to solve the former. If we subtract its trace, we are left with

$$R_{\mu\nu} + 2\partial_\mu\varphi\partial_\nu\varphi = 0 \tag{8.4}$$

as the only set of equations that we really need to solve.

One can then proceed by trying to find a BH-type solution (i.e. one with a metric similar to that of the Schwarzschild solution, possessing an event horizon) of the equation of motion of this system. It is clear that any solution of the vacuum Einstein equations (in particular, Schwarzschild's) will be a solution of these equations with a constant scalar $\varphi = \varphi_0$, but we are really interested only in solutions with a non-trivial φ. How could we characterize the non-triviality of φ? By analogy with other fields, we could consider multipole expansions of φ. The monopole momentum of φ (the coefficient of the $1/r$ term), which is the only one that respects spherical symmetry, could be understood as the "scalar charge" and we could characterize the simplest BH-type solutions (the static and spherically symmetric ones) by the mass (the monopole momentum of the gravitational field) and the "scalar charge."

We would like to have, though, a more physical definition of the "scalar charge." The first definition of "scalar charge" one could try is suggested by the form of a possible source for φ: it would have to be a scalar ρ satisfying $\nabla^2\varphi = \rho$, corresponding to a coupling of the form $\varphi\rho$ in the action. Then, the integral over some spatial volume (let us say a constant-time slice of the whole spacetime) of the source would give the charge, and, using the equation of motion, one could define

$$\int_\Sigma d^3\Sigma_\mu n^\mu \nabla^2\varphi, \tag{8.5}$$

where n^μ is the unit vector normal to the spacelike hypersurface Σ. This integral is indeed proportional to the coefficient of $1/r$ in the multipole expansion of φ. However, there is no way to show that this "charge' is conserved using the scalar equation of motion. Nothing prevents this kind of "charge" from disappearing and, in fact, according to the results on gravitational collapse and perturbations[1] of the Schwarzschild solution upon which the no-hair conjecture is based, this is actually what happens in the gravitational collapse, although no complete proof is available.

Still, one could conceive of a situation in which not all the "scalar charge" disappears and after a long time the system settles into a static, spherically symmetric state with non-vanishing scalar charge. The no-hair conjecture asserts that the solution describing this state will not be a BH, which in general means that it will have naked singularities. The cosmic-censorship conjecture then tells us that this state could not have been produced in the gravitational collapse of well-behaved matter with physically admissible initial conditions, in complete agreement with the no-hair conjecture.

Now we can put to the test the no-hair and cosmic-censorship conjectures either by trying to find static, spherically symmetric solutions with non-trivial scalar fields or by evolving initial data sets describing one or several regular BHs with mass and scalar charge that are not in equilibrium, such as those in [744]. This has not yet been done and, therefore, we will concentrate on finding scalar BH solutions. It is worth mentioning that some exceptions to the cosmic-censorship conjecture are known, especially in Einstein–Yang–Mills systems, and only by evolving the initial data can one really find out whether the same will happen here.

To find static, spherically symmetric solutions we make the *Ansatz* ($c = G_{\rm N}^{(4)} = 1$)

$$ds^2 = \lambda(r)dt^2 - \lambda^{-1}(r)dr^2 - R^2(r)d\Omega^2_{(2)}, \qquad \varphi = \varphi(r), \tag{8.6}$$

and, using the formulae in Appendix F.1.2, we find the Janis–Newman–Winicour (JNW) solutions [18, 607]

$$\boxed{\begin{aligned} ds^2 &= W^{\frac{2M}{\omega}-1}W dt^2 - W^{1-\frac{2M}{\omega}}\left[W^{-1}dr^2 + r^2 d\Omega^2_{(2)}\right],\\ \varphi &= \varphi_0 + \frac{\Sigma}{\omega}\ln W,\\ W &= 1+\tfrac{\omega}{r}, \qquad \omega = \pm 2\sqrt{M^2+\Sigma^2}. \end{aligned}} \tag{8.7}$$

[1] See e.g. [249], in which the wave equation for a scalar field on a Schwarzschild BH background is analyzed and it is shown that it has no physically acceptable solutions, the conclusion being that a BH cannot act as a source for the scalar field and that there will be no BH solutions with non-trivial scalar hair.

The three fully independent parameters that characterize each solution are the mass M, the "scalar charge" Σ, and the value of the scalar at infinity φ_0. As expected, only when the "scalar charge" vanishes ($\Sigma = 0$) does one have a regular solution (Schwarzschild's).[2] In all other cases there is a singularity at $r = r_0$, when $r_0 > 0$, or at $r = 0$.[3]

Although a regular BH cannot act as a source for scalar charge, other fields can. This is what happens in the "a-model" (also known as Einstein–Maxwell-dilaton (EMD) gravity, see Section 12.1) in which the scalar ("dilaton") equation of motion is roughly of the form

$$\nabla^2\varphi = \tfrac{1}{8} a e^{-2a\varphi} F^2. \tag{8.11}$$

In this theory we can expect BHs with non-trivial scalar fields. However, the scalar charge will be completely determined by the mass and electric and magnetic charges of the electromagnetic field, according to a certain formula. This kind of hair, which does depend on the mass, angular momentum, and conserved charges is called *secondary hair* [249]. If the scalar charge does not have the value dictated by the formula then there is another source for the scalar field apart from the electromagnetic field as in the solutions of [19], so the BH would also have *primary hair*. This is the only kind of hair that the solutions Eqs. (8.8) have and is the kind forbidden by the no-hair conjecture.

At this point it is worth mentioning that there are other kinds of scalar charges that are locally conserved. This discussion anticipates concepts that we will encounter in Part III. First, the equation of motion $\nabla^2\varphi = 0$ can be rewritten in the form $\partial_\mu\left(\sqrt{|g|}\, F^\mu\right) = 0$, where

---

[2] This is another example (see footnote on page 205) of a family of metrics parametrized by a continuous parameter whose physical properties are not continuous functions of those parameters.

[3] Observe that the above family of solutions includes a non-trivial *massless* solution. On setting $M = 0$ above, we find

$$\boxed{\begin{aligned} ds^2 &= dt^2 - dr^2 - W r^2 d\Omega^2_{(2)}, \\ \varphi &= \varphi_0 + \tfrac{1}{2}\ln W, \\ W &= 1 + \tfrac{\omega}{r}. \end{aligned}} \tag{8.8}$$

This solution is related to Schwarzschild's (with positive or negative mass) by a Buscher "T-duality" (to be explained later on) transformation on the time direction. It is still singular for any value of ω different from zero. This is perhaps best seen after the coordinate change

$$r = \frac{1}{\rho}\left(\rho - \frac{\omega}{4}\right)^2, \tag{8.9}$$

which allows us to rewrite the metric in the isotropic form

$$\begin{aligned} ds^2 &= dt^2 - \left(1 + \frac{\omega/4}{\rho}\right)^2\left(1 - \frac{\omega/4}{\rho}\right)^2 d\vec{x}_3^{\,2}, \\ \varphi &= \varphi_0 - \tfrac{1}{2}\ln\left[\left(1 - \frac{\omega/4}{\rho}\right)^2\left(1 + \frac{\omega/4}{\rho}\right)^{-2}\right], \qquad \rho = |\vec{x}_3|. \end{aligned} \tag{8.10}$$

The interpretation of these static, massless solutions is not easy. Since the mass of a spacetime is its total energy and the scalar field must contribute a positive amount to the total energy, we have to admit that the gravitational field contributes a negative amount to it. Here we see again the relation among the no-hair conjecture, cosmic censorship, and positivity of the energy.

$F^{\mu} = \nabla^{\mu}\varphi$. As will be explained later for the electric charge, this is just the continuity equation for the current F^{μ} and suggests the definition of scalar charge

$$\int_V d^3\Sigma_{\mu}\nabla^{\mu}\varphi, \tag{8.12}$$

which will be locally conserved. The conservation of this current is associated via Noether's theorem with the invariance of the action under constant shifts of the scalar.

Second, the Bianchi-type identity $\partial_{[\mu}\partial_{\nu]}\varphi = 0$ can be rewritten in the form $\nabla_{\mu}F^{\mu\nu\rho} = 0$, where we have defined the completely antisymmetric tensor $F^{\mu\nu\rho} = (1/\sqrt{|g|})\epsilon^{\mu\nu\rho\sigma}\partial_{\sigma}\varphi$. With this definition it is possible to show that the line integral

$$\frac{1}{3!}\oint_{\gamma} d^1\Sigma_{\mu\nu\rho}F^{\mu\nu\rho} = \oint_{\gamma} d\varphi, \tag{8.13}$$

along the curve γ is conserved. Observe that, if γ is closed, the integral will only be different from zero if φ is multivalued, for instance if φ is an axion (a pseudoscalar) that takes values in a circle.

How should we interpret these charges? We will see later in this chapter that the electromagnetic field A_{μ} has a natural coupling to the worldline of a particle with electric charge q given by Eq. (8.53). The particle's electric charge is given by the surface integral over a sphere S^2 of the Hodge dual of the electromagnetic-field-strength 2-form $F_{\mu\nu}$. The particle's magnetic charge is given by the surface integral over a sphere S^2 of the electromagnetic-field-strength 2-form. The electric charge is conserved due to the equation of motion and the magnetic charge is conserved due to the Bianchi identity. A topologically nontrivial configuration of the field is needed in order to have magnetic charge.

Potentials that are differential forms of higher rank couple to the worldvolumes of extended objects: a $(p+1)$-form potential $A_{(p+1)}$ naturally couples to p-dimensional objects with a $(p+1)$-dimensional worldvolume (we will explain how this comes about in Chapter 18). The electric charge is the integral over the sphere $\mathrm{S}^{d-(p+2)}$ transverse to the object's worldvolume of the Hodge dual of the $(p+2)$-form field strength $F_{(p+2)} = dA_{(p+1)}$. The magnetic charge would be the electric charge of the dual $(d-p-4)$-dimensional object, charged under the dual potential whose field strength is the Hodge dual of $F_{(p+2)}$.

Looking now at the above charges, we immediately realize that the charge defined in Eq. (8.13) is the charge of a one-dimensional object (string) and the former Eq. (8.12) is the charge of a "-1-dimensional object." Such an object would be an instanton, defined in Euclidean space and with zero-dimensional worldvolume. Then "charge conservation" is not a concept to be applied to it. In both cases φ has to be a pseudoscalar.

Observe that, indeed, a line integral as Eq. (8.13) cannot measure a point-like charge because we could continuously contract the loop γ to a point without meeting the singularity at which the charge rests. The line integral has to have a non-vanishing linking number with the one-dimensional object, which has to have either infinite length or the topology of S^1; otherwise the integral would be zero by the same argument. The behavior of the scalar field has to be $\varphi \sim \ln\rho$, where ρ measures the distance to the one-dimensional object in the two-dimensional plane orthogonal to it.

Similar arguments apply to the definition Eq. (8.12) and $\varphi \sim 1/\rho^2$, where now ρ measures the distance to the instanton in the four-dimensional Euclidean space.

From this point of view, if BHs can be understood as particle-like objects, looking for BHs with a well-defined scalar charge is utterly hopeless. One should look instead for "black strings" and instantons and in due time we will do so and find them.[4]

There is another point of view concerning scalar fields: in some cases they should be interpreted not as matter fields but as "local coupling constants" (as in the case of the string-theory dilaton) or, more generally, as *moduli fields*, which we will define in Chapter 11, in which case they should be treated as backgrounds and there would be no room for the notion of scalar charge.

In conclusion, if we want to find new BH solutions, we need to couple the Einstein–Hilbert action to matter fields that have associated conserved charges. The charges must be those of point-particles or we will naturally obtain solutions describing extended objects instead of black holes. Thus, we have to consider vector fields and the simplest one is an Abelian vector field A_μ. We are going to study in some detail the resulting system because later we will find generalizations of all the concepts and formulae developed here.

8.2 The Einstein–Maxwell system

The action for gravity coupled to an Abelian vector field A_μ is the so-called Einstein–Maxwell action[5] obtained by adding the Einstein–Hilbert and the Maxwell action with $\eta_{\mu\nu}$, ∂_μ, and d^4x replaced by $g_{\mu\nu}$, ∇_μ, and d^4x:

$$\boxed{S_{\text{EM}}[g_{\mu\nu}, A_\mu] = S_{\text{EH}}[g] + \frac{1}{c}\int d^4x\sqrt{|g|}\,\left[-\tfrac{1}{4}F^2\right].} \tag{8.15}$$

$F_{\mu\nu}$ is the field strength of the electromagnetic vector field A_μ and is again given by

$$F_{\mu\nu} = 2\partial_{[\mu}A_{\nu]}, \qquad F^2 = F_{\mu\nu}F^{\mu\nu}, \tag{8.16}$$

since, in the absence of torsion, $\nabla_{[\mu}A_{\nu]} = \partial_{[\mu}A_{\nu]}$. The components of A_μ and $F_{\mu\nu}$ in a given coordinate system are customarily split in this way,

$$(A_\mu) = (\phi, -\vec{A}), \qquad (F_{\mu\nu}) = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & -B_3 & B_2 \\ -E_2 & B_3 & 0 & -B_1 \\ -E_3 & -B_2 & B_1 & 0 \end{pmatrix}, \tag{8.17}$$

---

[4] This argument really applies to pseudoscalar fields.

[5] In this section we work in the Heaviside system of units, so the Coulomb force between two charges is

$$\frac{1}{4\pi}\frac{q_1q_2}{r_{12}^2}. \tag{8.14}$$

In the Gaussian system we should replace $1/(4c)$ by $1/(16\pi c)$ and the factor of 4π disappears from the Coulomb force. The dimensions of the vector field A_μ are $M^{1/2}L^{1/2}T^{-1}$ (that is, L^{-1} in natural units $\hbar = c = 1$) and the electric charge's units are $M^{1/2}L^{3/2}T^{-1}$, so it is dimensionless in natural units. At the end we will introduce another system of units, which will be the one we more often will work with, taking $c = 1$ and replacing the factor of $1/(4c)$ in front of F^2 by $1/\left(64G_{\text{N}}^{(4)}\right)$.

where $\vec{E} = (E_1, E_2, E_3)$ and $\vec{B} = (B_1, B_2, B_3)$ are the electric and magnetic 3-vector fields in that coordinate system, and, thus, with $\vec{\nabla} = (\partial_{\underline{1}}, \partial_{\underline{2}}, \partial_{\underline{3}})$

$$\begin{cases} E_i = F_{\underline{0i}}, \\ B_i = -\frac{1}{2}\epsilon_{ijk}F_{\underline{kl}}, \end{cases} \quad \Leftrightarrow \quad \begin{cases} \vec{E} = -\vec{\nabla}\phi - \dfrac{1}{c}\dfrac{\partial}{\partial t}\vec{A}, \\ \vec{B} = \vec{\nabla} \times \vec{A}. \end{cases} \tag{8.18}$$

The field strength (and the action) is invariant under the Abelian gauge transformations

$$A'_\mu = A_\mu + \partial_\mu \Lambda \tag{8.19}$$

with smooth, gauge parameter Λ. Depending on which gauge group we consider ($\mathbb{R}$ or U(1)), Λ must be a single-valued or multivalued function.[6] In differential-forms language

$$A = A_\mu dx^\mu, \qquad A' = A + d\Lambda, \qquad F = \tfrac{1}{2}F_{\mu\nu}dx^\mu \wedge dx^\nu = dA, \tag{8.20}$$

and the gauge invariance of F is a consequence of $d^2 = 0$. Using these differential forms, the Maxwell action can be rewritten as follows:

$$S_{\mathrm{M}}[A] = \frac{1}{8c}\int F \wedge {}^\star F. \tag{8.21}$$

Observe that there is no matter charged with respect to A_μ in this system. This is analogous to the presence of no matter fields in the Einstein–Hilbert action. However, the Einstein–Hilbert action contains the self-coupling of gravity and therefore the presence of a coupling constant in it makes sense, whereas in the Maxwell theory there are no *direct* interactions between photons and, in principle, there is neither an electromagnetic coupling constant nor a unit of electric charge. We will see that things are a bit more complicated in the presence of gravity, through which photons do interact.

The equations of motion of $g_{\mu\nu}$ and A_μ are

$$G_{\mu\nu} - \frac{8\pi G_{\mathrm{N}}^{(4)}}{c^4}T_{\mu\nu} = 0, \tag{8.22}$$

$$\nabla_\mu F^{\mu\nu} = 0 \qquad \text{(Maxwell's equation)}, \tag{8.23}$$

where

$$T_{\mu\nu} = \frac{-2c}{\sqrt{|g|}}\frac{\delta S_{\mathrm{M}}[A]}{\delta g^{\mu\nu}} = F_{\mu\rho}F_\nu{}^\rho - \tfrac{1}{4}g_{\mu\nu}F^2 \tag{8.24}$$

is the energy–momentum tensor of the vector field, which is traceless[7] in $d = 4$. The tracelessness of the electromagnetic energy–momentum tensor implies that $R = 0$ and the

[6] If the gauge group is $\mathbb{R}$, the elements of the group will be $e^{\Lambda/L}$, whereas, if it is U(1), they will be $e^{i\Lambda/L}$, where L is a constant introduced to make the exponent dimensionless because Λ is dimensionful. In the second case Λ will have to be identified with $\Lambda + 2\pi L$. When there is a unit of charge, L is related to it.

[7] This property is associated with the invariance of the Maxwell Lagrangian in curved spacetime under Weyl rescalings of the metric,

$$g'_{\mu\nu} = \Omega^2(x)g_{\mu\nu}. \tag{8.25}$$

In fact, if $\Omega = e^\sigma$, then for infinitesimal transformations $\delta_\sigma g_{\mu\nu} = 2\sigma(x)g_{\mu\nu}$ we have

$$\delta_\sigma S_{\mathrm{M}} = \frac{\delta S_{\mathrm{M}}}{\delta g_{\mu\nu}}\delta_\sigma g_{\mu\nu} \sim \sigma T^{\mu\nu}g_{\mu\nu} = 0. \tag{8.26}$$

Einstein equation takes the simpler form

$$R_{\mu\nu} = \frac{8\pi G_N^{(4)}}{c^4} T_{\mu\nu}. \tag{8.27}$$

On taking the divergence of the Einstein equation and using the contracted Bianchi identity for the Einstein tensor $\nabla_\mu G^{\mu\nu} = 0$, we find

$$F_{\nu\rho}\nabla_\mu F^{\mu\rho} - \tfrac{3}{2}F^{\mu\rho}\nabla_{[\mu}F_{\rho\nu]} = 0. \tag{8.28}$$

Since the Levi-Cività connection is symmetric,

$$\nabla_{[\mu}F_{\rho\nu]} = \partial_{[\mu}F_{\rho\nu]} = 0 \qquad \text{(the Bianchi identity)} \tag{8.29}$$

identically, using the definition of $F_{\mu\nu}$, and then we see that the Einstein equation implies generically the Maxwell equation. Using Eq. (1.62), the Maxwell equation can also be written in a simpler, equivalent, form:

$$\partial_\mu\left(\sqrt{|g|}F^{\mu\nu}\right) = 0. \tag{8.30}$$

The equations are written in terms of the field strength F and usually they are solved in terms of it. However, we are ultimately interested in the vector field A itself and we have to make sure that the F we obtain is such that it is related to some vector field by Eq. (8.16) or Eq. (8.20). It turns out that, locally, A exists if the electromagnetic Bianchi identity Eq. (8.29) is satisfied.[8]

The Bianchi identity can also be written in this form (by contracting Eq. (8.29) with $\epsilon^{\mu\nu\rho\sigma}$, introducing it into the partial derivative (because it is constant), and using the

[8] In fact, if we are given F and the Bianchi identity is satisfied, we can always find the corresponding vector potential by using, e.g., the formula

$$\boxed{A_\mu(x) = -\int_0^1 d\lambda \lambda x^\nu F_{\mu\nu}(\lambda x).} \tag{8.31}$$

To check this formula it is necessary to use the Bianchi identity: taking the curl of the l.h.s.,

$$\partial_{[\rho}A_{\mu]}(x) = -\int_0^1 d\lambda\lambda\partial_{[\rho}[x^\nu F_{\mu]\nu}(\lambda x)], \tag{8.32}$$

and operating,

$$\partial_{[\rho}\big(F_{\mu]\nu}(\lambda x)\big) = \lambda\partial_{[\rho}F_{\mu]\nu}(\lambda x) - \tfrac{1}{2}\lambda\big(\partial_\nu F_{\rho\mu}\big)(\lambda x), \tag{8.33}$$

where the Bianchi identity Eq. (8.29) has been used in the last identity, one obtains

$$\begin{aligned}\partial_{[\rho}A_{\mu]}(x) &= \tfrac{1}{2}\int_0^1 d\lambda\left\{\lambda^2 x^\nu\partial_\nu F_{\rho\mu}(\lambda x) - \lambda F_{\mu\rho}(\lambda x)\right\} = \tfrac{1}{2}\int_0^1 d\lambda\frac{d}{d\lambda}\left[\lambda^2 F_{\rho\mu}(\lambda x)\right]\\ &= \frac{1}{2}F_{\rho\mu}(x).\end{aligned} \tag{8.34}$$

definition of the Hodge dual and Eq. (8.30) for the divergence):

$$\nabla_\mu {}^\star F^{\mu\sigma} = 0. \tag{8.35}$$

In the language of differential forms, the Maxwell equation and Bianchi identity are

$$d^\star F = 0, \tag{8.36}$$

$$dF = 0, \tag{8.37}$$

and the Bianchi identity is just a consequence of the definition Eq. (8.20) and $d^2 = 0$.

Then, if we work with the field strength, we find that there are two pairs of equations, (8.23) and (8.35) and (8.36) and (8.23), which are (as pairs) invariant if one replaces F by ${}^\star F$ (by virtue of ${}^{\star\star}F = -F$). This is an *electric–magnetic-duality* transformation. The name is due to the fact that this transformation interchanges the electric and magnetic fields in any given coordinate system according to

$$\vec{E}\,' = \vec{B}, \qquad \vec{B}\,' = -\vec{E}. \tag{8.38}$$

Actually, this pair of homogeneous equations (the Maxwell equation and the Bianchi identity) would be invariant under the (invertible) substitution for F of any linear combination of F and ${}^\star F$. We would have a symmetry of all the equations of motion if the Einstein equation were also invariant under this replacement. We will later see in Section 8.7 that this is the case and that the Einstein–Maxwell theory is invariant under electric–magnetic duality.

The four Maxwell equations in Minkowski spacetime can be deduced from the Maxwell equation and the Bianchi identity (two of them imply the existence of the potential A_μ and are equivalent to the latter). We have

$$\begin{aligned}
\partial_\mu F^{\mu\nu} = 0 \quad &\Leftrightarrow \quad \begin{cases} \vec{\nabla}\cdot\vec{E} = 0, \\ \vec{\nabla}\times\vec{B} - \dfrac{1}{c}\dfrac{\partial}{\partial t}\vec{E} = 0, \end{cases} \\
\partial_\mu {}^\star F^{\mu\nu} = 0 \quad &\Leftrightarrow \quad \begin{cases} \vec{\nabla}\cdot\vec{B} = 0, \\ \vec{\nabla}\times\vec{E} + \dfrac{1}{c}\dfrac{\partial}{\partial t}\vec{B} = 0. \end{cases}
\end{aligned} \tag{8.39}$$

8.2.1 Electric charge

The electric charge can be defined in terms of a source coupled to the electromagnetic field (this is analogous to the energy–momentum-pseudotensor approach for the gravitational field) or in terms of the Noether current associated with the gauge invariance (the approach that leads to Komar's formula and its generalizations for the gravitational field). The two definitions are equivalent and are very closely related to each other because the gauge invariance of the free theory imposes strong constraints on the possible couplings.

Let us first introduce the electric charge using sources. A source for the Maxwell field is described by a current j^{μ}, which naturally couples to the vector field through a term in the action of the form

$$\frac{1}{c^2}\int d^4x\sqrt{|g|}\,\left[-A_{\mu}j^{\mu}\right]. \tag{8.40}$$

This additional interaction term spoils the action's gauge invariance unless the source j^{μ} is divergence-free,

$$\nabla_{\mu}j^{\mu}=0 \quad\Leftrightarrow\quad d\,{}^{\star}j=0 \qquad (j\equiv j_{\mu}dx^{\mu}), \tag{8.41}$$

which implies the *continuity equation* for the vector density $\mathrm{j}^{\mu}\equiv\sqrt{|g|}\,j^{\mu}$,

$$\partial_{\mu}\mathrm{j}^{\mu}=0. \tag{8.42}$$

The continuity equation can be used to establish the local conservation of the electric charge, as explained in Section 2.3, if the electric charge contained in a three-dimensional volume at a given time t, V^3_t, is defined by[9]

$$q(t)=-\frac{1}{c}\int_{\mathrm{V}^3_t}d^3x\,\mathrm{j}^0, \tag{8.43}$$

or, in a more covariant form,

$$q(t)=\frac{1}{c}\int_{\mathrm{V}^3_t}{}^{\star}j. \tag{8.44}$$

As explained in Section 2.3, this quantity is not constant: its variation is related to the flux of charge through the boundary of V^3_t. If V^3_t is a constant-time slice of the whole spacetime with no boundary, then the above integrals give the total charge, which will be constant in time. If we can foliate our spacetime with constant-time hypersurfaces, then we take the four-dimensional spacetime V^4 contained in between two constant-time slices $\mathrm{V}^3_{t_1}$ and $\mathrm{V}^3_{t_2}$, integrate the continuity equation over it, and use Stokes' theorem. The boundary of the four-dimensional region we have proposed is made up of the two constant-time slices with opposite orientations, so

$$0=\int_{\mathrm{V}^4}d\,{}^{\star}j=\int_{\mathrm{V}^3_{t_1}}{}^{\star}j-\int_{\mathrm{V}^3_{t_2}}{}^{\star}j, \tag{8.45}$$

and the total electric charge is constant in time.

Thus, gauge invariance of the action implies that the source is divergence-free and from this the local conservation of the electric charge (and the global conservation of the total electric charge) follows.

On the other hand, in the presence of the source, the Maxwell equation is modified into

$$\nabla_{\mu}F^{\mu\nu}=\frac{1}{c}j^{\nu}, \tag{8.46}$$

or, equivalently,

$$d^{\star}F=\frac{1}{c}{}^{\star}j, \tag{8.47}$$

[9] The sign is conventional.

and, using the antisymmetry of $F_{\mu\nu}$ or $d^2 = 0$, it is trivial to see that, since the l.h.s. of the equation is divergence-free, the r.h.s. of the equation is also, for consistency, divergence-free, as we knew it had to be in order to preserve the gauge invariance of the action. This is no coincidence: the fact that the r.h.s. of the Maxwell equation is divergence-free is in fact the gauge identity associated with the invariance under $\delta A_\mu = \partial_\mu \Lambda$, as we are going to see.

Finally, using the Maxwell equation (8.47), we can rewrite the definition of the total electric charge Eq. (8.44) in terms of the field strength and again use Stokes' theorem. If the boundary of a constant-time slice has the topology of a S^2 at infinity, we obtain

$$\boxed{q = \int_{S^2_\infty} {}^\star F.} \tag{8.48}$$

which is a useful definition of the total electric charge of a spacetime in terms of the field strength (the electric flux) and which we will generalize further in Part III.

This is the kind of formula that we will use because in the Einstein–Maxwell system there are no fields explicitly written that act as sources for A_μ. Just as in the case of the Maxwell equations in vacuum, we can obtain solutions describing the field of charges. These solutions are singular near the place where the charge ought to be and the solution is not a solution there (there are no charges explicitly included in the system). However, the above expression allows us to calculate the charge that ought to be placed there to produce the flux of electromagnetic field that we observe.[10] We have introduced sources as a device for understanding the definition.

We could also have used the invariance of the Einstein–Maxwell action to find the conserved Noether current and define the electric charge through it.

We studied the invariance of the Maxwell action and found the corresponding Noether current in Minkowski spacetime in Section 3.2.1. The coupling to gravity introduces only minor changes and the conclusion is, again, that the electric charge can be defined by Eq. (8.48).

It is useful to consider a simple example of a source: the current associated with a particle of electric charge q and worldline γ parametrized by $X^\mu(\xi)$. In a manifestly covariant form it is given by

$$j^\mu(x) = qc \int_\gamma dX^\mu \frac{1}{\sqrt{|g|}} \delta^{(4)}[x - X(\xi)], \tag{8.49}$$

where $dX^\mu = d\xi dX^\mu/d\xi$. On making the choice $\xi = X^0$ and integrating over X^0, we obtain

$$\begin{aligned} j^\mu(x^0, \vec{x}) &= qc \int dX^0 \frac{dX^\mu}{dX^0} \frac{1}{\sqrt{|g|}} \delta^{(3)}(\vec{x} - \vec{X}) \delta(x^0 - X^0) \\ &= -qc \frac{dX^\mu}{dx^0} \frac{\delta^{(3)}[\vec{x} - \vec{X}(x^0)]}{\sqrt{|g|}}. \end{aligned} \tag{8.50}$$

[10] Of course, this is just a covariant generalization of the Gauss theorem that relates the flux of electric field through a closed surface to the charge enclosed by it.

If the particle is at rest at the origin in the chosen coordinate system, the current is

$$j^{\mu}(x^0,\vec{x}) = -qc\delta^{\mu 0}\frac{\delta^{(3)}(\vec{x})}{\sqrt{|g|}}, \tag{8.51}$$

and it is easy to see that q is indeed the electric charge according to the above definitions.

j^{μ} is a conserved current:

$$\begin{aligned}
\nabla_{\mu} j^{\mu} &\sim \frac{\partial}{\partial x^{\mu}}\left\{\sqrt{|g(x)|}\int d\xi\,\dot{X}^{\mu}\frac{1}{\sqrt{|g(X)|}}\delta^{(4)}[x - X(\xi)]\right\}\\
&= \int d\xi\,\dot{X}^{\mu}\frac{\partial}{\partial x^{\mu}}\delta^{(4)}[x - X(\xi)] = -\int d\xi\,\dot{X}^{\mu}\frac{\partial}{\partial X^{\mu}}\delta^{(4)}[x - X(\xi)]\\
&= -\int d\xi\,\frac{d}{d\xi}\delta^{(4)}[x - X(\xi)] = -\delta^{(4)}[x - X(\xi)]\Big|_{\xi_1}^{\xi_2} = 0,
\end{aligned} \tag{8.52}$$

generically, except for the initial and final positions of the particle $X^{\mu}(\xi_1)$ and $X^{\mu}(\xi_2)$, which look like a 1-particle source and a sink and can be taken to infinity.

Observe that, for the current (8.49), the interaction term (8.40) becomes the integral of the 1-form A over the worldline γ:

$$-\frac{q}{c}\int_{\gamma(\xi)} A_{\mu}\dot{x}^{\mu}d\xi = -\frac{q}{c}\int_{\gamma} A. \tag{8.53}$$

This term has to be added to the action of the particle, Eq. (3.255), (3.257) or (3.258), in order to obtain the worldline action of a massive electrically charged particle,

$$\boxed{S[X^{\mu}(\xi)] = -Mc\int d\xi\,\sqrt{g_{\mu\nu}(X)\dot{X}^{\mu}\dot{X}^{\nu}} - \frac{q}{c}\int d\xi\,A_{\mu}\dot{X}^{\mu},} \tag{8.54}$$

or that of a massless one. That kind of term is known as a *Wess–Zumino* (WZ) term. In this form it is easy to see that, under a gauge transformation, the action changes by a total derivative. The integral of the total derivative vanishes exactly only for special boundary conditions, though.

This action can be used as a source, but it also describes the motion of a charged particle in a gravitational/electromagnetic background. In the special-relativistic limit, taking $\xi = X^0 = ct$, the action takes the standard form

$$S \sim \int dt\left\{-Mc^2 + \tfrac{1}{2}Mv^2 - q\phi + \frac{q}{c}\vec{A}\cdot\vec{v}\right\}. \tag{8.55}$$

If there is a point-like charge q at rest at the origin the only non-vanishing components of F are $F_{\underline{0}r}$ and they should depend only on r because of the spherical symmetry of the problem. Using the above definition of charge and working in general static spherical coordinates Eq. (7.22), we find

$$q = \int_{S^2_{\infty}} \frac{\epsilon_{\mu\nu\rho\sigma}}{4\sqrt{|g|}}F^{\rho\sigma}dx^{\mu}\wedge dx^{\nu} = \int_{S^2_{\infty}} d\Omega^2 r^2 F_{\underline{0}r} = \omega_{(2)}\lim_{r\to\infty}\left(r^2 F_{\underline{0}r}\right), \tag{8.56}$$

where $\omega_{(2)}$ is the volume of the 2-sphere 4π. Then, the electromagnetic field of a point-like charge must behave for large r as follows:

$$E_r = F_{\underline{0}r} \sim +\frac{1}{4\pi}\frac{q}{r^2}, \qquad \phi \sim +\frac{1}{4\pi}\frac{q}{r}. \tag{8.57}$$

(Of course, this result is exact in the absence of gravity, in Minkowski spacetime.) On plugging this result into Eq. (8.55), we find that the electrostatic force between two particles is, in this unit system, $q_1 q_2/(4\pi r^2)$, as we said.

In the units that we are using, M appears multiplied by $G_{\mathrm{N}}^{(4)}$ in the metric (as in the Schwarzschild solution) and q does not. Some simplification is achieved by using the following normalization and units that are standard in this field: we set $c = 1$ and rewrite the Einstein–Maxwell action as follows:

$$\boxed{S_{\mathrm{EM}}[g, A] = \frac{1}{16\pi G_{\mathrm{N}}^{(4)}} \int d^4x \sqrt{|g|}\, \left[R - \tfrac{1}{4}F^2\right].} \tag{8.58}$$

In these units both A_μ and $g_{\mu\nu}$ are dimensionless. The factor $16\pi G_{\mathrm{N}}^{(4)}$ disappears from the equations of motion. Furthermore, if we keep (by definition) the WZ term as in Eq. (8.54) without any additional normalization factor, the electric charge is now

$$\boxed{q = \frac{1}{16\pi G_{\mathrm{N}}^{(4)}} \int_{S^2_\infty} {}^\star F,} \tag{8.59}$$

and has dimensions of mass (energy). Finally, for a point-like charge we expect, for large r,

$$E_r = F_{0r} \sim \frac{4G_{\mathrm{N}}^{(4)} q}{r^2}, \tag{8.60}$$

which implies that the force between two charges is

$$F_{12} = 4G_{\mathrm{N}}^{(4)} \frac{q_1 q_2}{r_{12}^2}. \tag{8.61}$$

8.2.2 Massive electrodynamics

Before concluding this section it is worth considering which facts would be modified if the vector field were massive. A massive vector field in Minkowski spacetime is described by the Proca Lagrangian Eq. (3.67) and its generalization to curved spacetime is straightforward. The equation of motion is

$$\nabla_\nu F^{\nu\mu} + m^2 A^\mu = 0. \tag{8.62}$$

We immediately see that this equation is completely different from the Bianchi identity Eq. (8.35), which is also valid in the massive case, which implies that massive electrodynamics, apart from gauge invariance, has no electric–magnetic duality. This implies that, in principle, there will be no Dirac magnetic monopoles dual to the electric ones, which explains the results of [591].

If we take the divergence of this equation, we find the integrability condition

$$\nabla_\nu A^\nu = 0, \tag{8.63}$$

that removes one of the degrees of freedom described by the vector field, leaving only three that correspond to the three possible helicities of a massive spin-1 particle $(-1, 0, +1)$.

The quanta of the Proca field, being massive, will propagate at a speed smaller than 1 (c) and the interaction they mediate will be short-ranged. We can see this by finding the static, spherically symmetric solution that describes the field of an electric monopole in this theory in Minkowski spacetime. On substituting the *Ansatz*

$$A_\mu = \delta_{\mu 0}\frac{f(r)}{r} \tag{8.64}$$

into the equation of motion, we obtain the differential equation

$$f'' - m^2 f = 0, \tag{8.65}$$

whose solution is (with the boundary condition $A_\mu \to 0$ when $r \to \infty$)

$$A_\mu = Q\delta_{\mu 0}\frac{e^{-mr}}{r}. \tag{8.66}$$

Q is an integration constant that is somehow related to the "electric charge." However, the lack of gauge invariance suggests that the "electric charge" is not conserved in this system. In fact, it is not easy to define what is meant by electric charge here. It is then useful to consider a slightly more general system with the following classically equivalent action for A_μ and a scalar auxiliary field ϕ:

$$S[A_\mu, \phi] = \int d^4x\sqrt{|g|}\left[-\tfrac{1}{4}F^2 + \tfrac{1}{2}(\partial\phi + mA)^2\right]. \tag{8.67}$$

This action is invariant under the following *massive gauge transformations*:

$$\delta A_\mu = \partial_\mu \Lambda, \qquad \delta\phi = -m\Lambda. \tag{8.68}$$

Observe that, for consistency, the scalar ϕ has to live in the gauge group manifold: either $\mathbb{R}$ or S^1 (if the gauge group is U(1)). On fixing the gauge $\phi = 0$ we recover the Proca Lagrangian and any solution of the equations of motion of the original system is also a solution of this one in this gauge.

It is sometimes said that the scalar ϕ is "eaten" by the vector field, which acquires a mass in the process. ϕ is then referred to as a *Stückelberg field* [871]. Observe that the number of degrees of freedom before and after the gauge fixing are the same. Observe also that this procedure for obtaining a massive vector field is different from the standard spontaneous symmetry-breaking mechanism. There are two main differences: the scalar is real and carries no charge with respect to the vector field and there is no potential for the scalar. (Actually, there is no way to write a gauge-invariant potential with only one real scalar.)

The equations of motion corresponding to the new Lagrangian are

$$\nabla_\nu F^{\nu\mu} + m\big(\nabla^\mu \phi + mA^\mu\big) = 0, \qquad \nabla^2\phi + m\partial_\mu A^\mu = 0. \tag{8.69}$$

To define a conserved charge, we can either introduce a source j^μ into the first equation or use the conserved Noether current associated with *constant*[11] shifts of ϕ:

$$j_{\rm N}^\mu = \partial_\nu F^{\nu\mu} + m\big(\partial^\mu\phi + mA^\mu\big). \tag{8.70}$$

The source j^μ is conserved, but only on-shell (upon use of the ϕ equation of motion) and the same applies to the Noether current, which is associated with a global symmetry. In both cases we can define the electric charge in this system by

$$q = \int d^3x\sqrt{|g|}\, j_{\rm N}^0 = \int d^3x\sqrt{|g|}\,\big[\nabla_\nu F^{\nu 0} + m\big(\nabla^0\phi + mA^0\big)\big]. \tag{8.71}$$

On applying this definition to the electric monopole solution Eq. (8.66) and using

$$\vec{\nabla}^2\Big(\frac{e^{-mr}}{r}\Big) = -4\pi\delta^{(3)}(\vec{x}) + \frac{m^2 e^{-mr}}{r}, \tag{8.72}$$

we find $q = 4\pi Q$, as we naively expected. It should be stressed, though, that this charge is of a nature completely different from the usual electric charge since it is associated with a *global* symmetry of a different field. In principle, the no-hair conjecture should apply (negatively) to charges of this kind associated with short-range interactions and global (rather than local) symmetries.

Finally, let us notice that neither the original Proca action nor the new one with the Stückelberg field ϕ has any duality *symmetry*. However, the new action can be dualized (i.e. written in *dual variables*), as we will see later in Section 8.7.5.

8.3 The electric Reissner–Nordström solution

We are now ready to find BH-type solutions of the equations of motion derived from the Einstein–Maxwell action normalized as in Eqs. (8.58). Since the Maxwell equation is satisfied if the Einstein equation is, we only have to solve the latter with the trace subtracted,

$$R_{\mu\nu} = \tfrac{1}{2}\big[F_\mu{}^\rho F_{\nu\rho} - \tfrac{1}{4}g_{\mu\nu}F^2\big], \tag{8.73}$$

plus the Bianchi identity. We are looking for a static, spherically symmetric solution and, therefore, as usual, we make the *Ansatz* Eq. (7.2) for the metric. This time we also have to make an *Ansatz* for the electromagnetic field. If we are looking for a point-like electrically charged object at rest, taking into account Eq. (8.60), an appropriate *Ansatz* that is readily seen to satisfy the Maxwell equation and Bianchi identity for the metric Eq. (7.2) is

$$F_{tr} \sim \pm\frac{1}{R^2(r)}. \tag{8.74}$$

[11] If we use the full gauge invariance of the theory, we recover exactly the same Noether current and Bianchi identity as in the massless case. The definition Eq. (8.59) then gives zero charge because F goes to zero too fast at infinity.

The $\pm$ corresponds to the two possible signs of the electric charge. The metric cannot depend on this sign because the action is invariant under the (admittedly rather trivial) duality symmetry $F \to -F$, $g \to g$. The solution one obtains in this way is the Reissner–Nordström (RN) solution[12] [735, 802] and can be conveniently written as follows:

$$\boxed{\begin{aligned} ds^2 &= f(r)dt^2 - f^{-1}(r)dr^2 - r^2 d\Omega^2_{(2)}, \\ F_{tr} &= \frac{4G_{\rm N}^{(4)} q}{r^2}, \\ f(r) &= \frac{(r-r_+)(r-r_-)}{r^2}, \\ r_\pm &= G_{\rm N}^{(4)} M \pm r_0, \qquad r_0 = G_{\rm N}^{(4)} \left(M^2 - 4q^2\right)^{\frac{1}{2}}, \end{aligned}} \tag{8.75}$$

where q is the electric charge, normalized as in Eq. (8.59), and M is the ADM mass.

Some remarks are necessary.

1. This metric describes the gravitational and electromagnetic field created by a spherical (or point-like), electrically charged object of total mass M and electric charge q as seen from far away by a static observer to which the coordinates $\{t, r, \theta, \varphi\}$ (that we can keep calling "Schwarzschild coordinates") are adapted. Schwarzschild's solution is contained as the special case $q = 0$.

   Included in the (total) mass is the energy associated with the presence of an electromagnetic field. We cannot covariantly separate the energy associated with "matter" from the energy associated with the electromagnetic field and the gravitational field, but we must keep in mind that the mass of the spacetime contains all these contributions.

2. The vector field that gives the above field strength and whose local existence is guaranteed by the fact that F satisfies the Bianchi identity is

$$A_\mu = \delta_{\mu t} \frac{4G_{\rm N}^{(4)} q}{r}. \tag{8.76}$$

3. There is a generalization of Birkhoff's theorem for RN BHs (see exercise 32.1 of [707]): RN is the only spherically symmetric family of solutions (that includes Schwarzschild's) of the Einstein–Maxwell system.

4. The metric above is a solution for any values of the parameters M and q and, therefore, of $r_\pm$, including complex ones.

5. The metric is singular at $r = 0$ and also at r_- and r_+, if r_+ and r_- are real. At $r = r_\pm$ the signature changes and, in the region between r_+ and r_-, r is timelike and t is spacelike and in that region the metric is not static as in Schwarzschild's horizon

[12] The Reissner–Nordström solution is also a particular case (the spherically symmetric case) of the general static axisymmetric electrovacuum solutions discovered independently by Weyl in [949, 950] and should also bear his name.

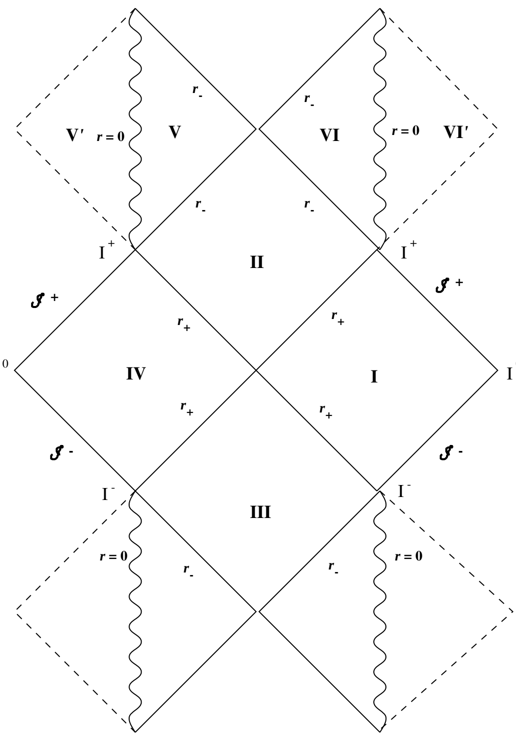

Fig. 8.1. Part of the Penrose diagram of a Reissner–Nordström black hole $M > 2|q|$. Only two “universes” are shown. The complete diagram repeats periodically the part shown.

interior. To find the nature of these singularities, we calculate curvature invariants and study the geodesics. $R = 0$ due to $T^{\mu}{}_{\mu} = 0$, but other curvature invariants (and F^2 as well) tell us that there is a curvature singularity at $r = 0$ but not at $r = r_{\pm}$. In fact, an analysis similar to the one made in the Schwarzschild case shows that, when it is real and positive, r_+ is an event horizon of area

$$A = 4\pi r_+^2, \tag{8.77}$$

surrounding the curvature singularity, in agreement with the weak form of the cosmic censorship conjecture, whereas r_- is a *Cauchy horizon*: in the RN spacetime there is no Cauchy hypersurface on which we can give initial data for arbitrary fields and predict their evolution in the whole spacetime. By definition, we can have a Cauchy hypersurface only for the region outside the Cauchy horizon. This horizon seems to be unstable under small perturbations [763] associated with the infinite blueshift that incoming radiation suffers in its neighborhood (opposite to the infinite redshift that incoming radiation suffers in the neighborhood of the event horizon), and it is conjectured that a spacelike singularity should appear in its place [197, 737].

Both horizons exist when $M > 2|q|$ and then the RN metric describes a BH. In Figure 8.1 we have represented part of its Penrose diagram, based on the maximal analytic extension of the RN metric Eq. (8.75) found in [465]. In this diagram there are two “universes” (quadrants I and IV, that have asymptotically flat regions), as in

Schwarzschild's, but the complete diagram consists of an infinite number of pairs of "universes" arranged periodically. The singularities are timelike, not spacelike like Schwarzschild's, and can be avoided by observers that enter the BH. In fact, there are timelike geodesics that, starting in a certain "universe," enter the BH crossing the event horizon r_+ and, after crossing two Cauchy horizons r_-, emerge in a different "universe." Analogous effects take place in the gravitational collapse of spherically symmetric shells of electrically charged matter [175]: depending on the characteristics of the shell, the gravitational collapse can end in a singularity or the shell can stop contracting, and start to expand in a different "universe."

Observe that, although the cosmic censorship conjecture is obeyed by the RN spacetime in its weak form, it is violated in its strong form: an observer that takes the inter-"universe" trip will see the singularity.[13] However, if the Cauchy horizon indeed became a spacelike singularity, such a problem would not arise.

6. When $M < -2|q|$ (negative) $r_\pm$ are real and negative and there is no horizon surrounding the curvature singularity at $r = 0$. The Penrose diagram of this spacetime again is the one in Figure 7.4. This case could be excluded by invoking cosmic censorship, which is violated in its weak form by this metric. It is reasonable to think (and the positive-energy theorem proves it) that, if we start with physically reasonable initial conditions, we will not end up with a negative mass.
7. When $-2|q| < M < 2|q|$, the constants $r_\pm$ are complex and there are no horizons and the only singularity left is the one at $r = 0$, and it is naked, the Penrose diagram being Figure 7.4. Again, cosmic censorship should exclude this range of values of M. This includes the special case $M = 0$. Observe that, otherwise, we would have a massless, charged object at rest, which is a rather strange object. The mass is the total energy of the spacetime. A non-trivial electromagnetic field such as the one produced by a point-like charge is a source of (positive) energy. Thus, our physical intuition tells us that, in order to have non-zero charge and at the same time zero mass, there must be some "negative energy density" present. It is thought that the same should happen in the other $-2|q| < M < 2|q|$ cases.

   Negative energies always seem to be at the heart of naked singularities and, in the spirit of cosmic censorship, if negative energies are not allowed initially, no naked singularities will appear in the evolution of the system.

   Thus, cosmic censorship restricts the possible values of M to the range $M \geq 2|q|$. What happens if we now throw into a regular RN BH charged matter with mass M' and charge q' such that $M + M' < 2|q + q'|$? In [930] it was proven that, if $M = 2|q|$ (an extreme RN BH), particles whose absorption by the BH would take it into the region of forbidden parameters are not captured by the BH. However, it seems that it is possible to "overcharge" a non-extremal ($M < 2|q|$) RN BH by sending into it a charged test particle (but not by using a charged collapsing shell of charged matter) [573], although the effects of the absorption of the particle on the BH geometry (which are assumed to be small) have not yet been worked out.

[13] An observer falling into a Schwarzschild BH cannot see the singularity, which always lies in its future, until he/she actually crashes onto it. This has to do with the spacelike nature of the Schwarzschild singularity.

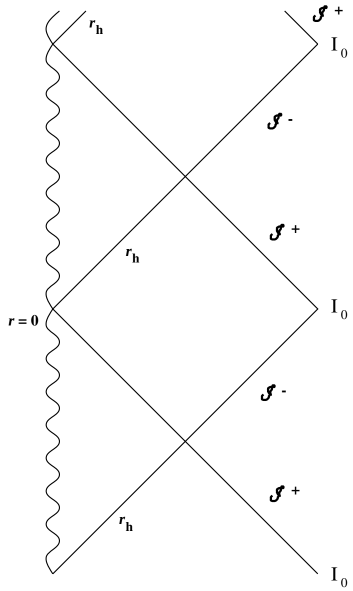

Fig. 8.2. Part of the Penrose diagram of an extreme Reissner–Nordström black hole. The complete diagram has an infinite number of "universes."

We see that the RN BH provides a very interesting playground on which to test cosmic censorship. We will see that the relation between cosmic censorship and positivity of the energy can be translated into supersymmetry (BPS) bounds.

8. The limiting case $M = 2|q|$ between the naked singularity and the regular BH is very special. When $M = 2|q|$ the two horizons coincide, $r_+ = r_- = G_N^{(4)} M$, and there is no change of signature across the resulting horizon (which is a degenerate Killing horizon), which still has a non-vanishing area

$$A_{\text{extreme}} = 4\pi r_+^2 = 4\pi \left(G_N^{(4)} M\right)^2. \tag{8.78}$$

This object is an *extreme* RN (ERN) BH and it will play a central role in much of what follows. Some of the properties of ERN BHs are the following.

(a) The proper distance to the horizon along radial directions at constant time,

$$\lim_{r_2 \to r_+} \int_{r_1}^{r_2} ds = \lim_{r_2 \to r_+} \int_{r_1}^{r_2} dr \left(1 - \frac{r_+}{r}\right)^{-1} = \infty, \tag{8.79}$$

diverges. This does not happen along timelike or null directions, though an observer can cross it in a finite proper time.

(b) The Penrose diagram is drawn in Figure 8.2. As we see, the causal structure is completely different from that of any regular RN BH no matter how close to the extreme limit it is. Thus, we can expect physical properties of the family of RN BHs to be discontinuous at the extreme limit.

(c) The relative values of their charge and mass are such that, if we have two of them, $M_1 = 2|q_1|$ and $M_2 = 2|q_2|$, it will always happen that

$$G_{\rm N}^{(4)} M_1 M_2 = 4 G_{\rm N}^{(4)} |q_1 q_2|, \tag{8.80}$$

and, *if both charges have the same sign* and we divide by the relative distance between them, we obtain

$$F_{12} = -G_{\rm N}^{(4)} \frac{M_1 M_2}{r_{12}^2} + 4 G_{\rm N}^{(4)} \frac{q_1 q_2}{r_{12}^2} = 0. \tag{8.81}$$

This is nothing but the force between two point-like, massive, charged, non-relativistic objects on account of Eqs. (3.140) and (8.61) and it vanishes, so they will be in equilibrium. Then, this suggests that it should be possible to find static solutions describing two (or many) ERN BHs in equilibrium.

(d) On shifting the radial coordinate $r = \rho + G_{\rm N}^{(4)} M$ of the ERN metric, it becomes

$$ds^2 = \left(1 + \frac{G_{\rm N}^{(4)} M}{\rho}\right)^{-2} dt^2 - \left(1 + \frac{G_{\rm N}^{(4)} M}{\rho}\right)^2 \left(d\rho^2 + \rho^2 d\Omega_{(2)}^2\right). \tag{8.82}$$

On defining new Cartesian coordinates $\vec{x}_3 = (x^1, x^2, x^3)$ such that $|\vec{x}_3| = \rho$ and $d\vec{x}_3^{\,2} = d\rho^2 + \rho^2 d\Omega_{(2)}^2$, we obtain a new form of the ERN solution:

$$\boxed{\begin{aligned} ds^2 &= H^{-2} dt^2 - H^2 d\vec{x}_3^{\,2}, \\ A_\mu &= -2\delta_{\mu t}\, \mathrm{sign}(q) \left(H^{-1} - 1\right), \\ H &= 1 + \frac{2 G_{\rm N}^{(4)} |q|}{|\vec{x}_3|} = 1 + \frac{G_{\rm N}^{(4)} M}{|\vec{x}_3|}. \end{aligned}} \tag{8.83}$$

Observe that, in this case, due to the shift in the radial coordinate, the event horizon is placed at $\vec{x}_3 = \vec{0}$, which in flat Minkowski spacetime is just a point. It is, though, easy to see that the surface labeled by $\vec{x}_3 = \vec{0}$ is not just a point but is a sphere of finite area because in the limit $\rho \to 0$ one has to take into account the ρ^2 factor of $d\Omega_{(2)}^2$ that cancels out the poles in H^2, so the induced metric in the $\rho = 0$, $t =$ constant hypersurface is, indeed,

$$ds^2 = -(G_{\rm N}^{(4)} M)^2 d\Omega_{(2)}^2. \tag{8.84}$$

H is a harmonic function in the three-dimensional Euclidean space spanned by the coordinates $\vec{x}_3$, i.e. it satisfies

$$\partial_{\underline{i}} \partial_{\underline{i}} H = 0. \tag{8.85}$$

This fact could just be a coincidence but, if we use Eq. (8.83) as an *Ansatz* in the equations of motion without imposing any particular form for H, we find that they are solved for *any harmonic function* H, not just for the one in Eq. (8.83). We have obtained in this way the *Majumdar–Papapetrou* (MP) family of solutions [676, 754]:

$$\boxed{\begin{aligned} ds^2 &= H^{-2}dt^2 - H^2 d\vec{x}_3^{\,2}, \\ A_\mu &= \delta_{\mu t}\alpha\big(H^{-1} - 1\big), \qquad \alpha = \pm 2, \\ \partial_{\underline{i}}\partial_{\underline{i}} H &= 0. \end{aligned}} \tag{8.86}$$

If we want to find solutions describing several ERN BHs in static equilibrium, it is, therefore, natural to search amongst this class of solutions.[14]

Maxwell's theory in Minkowski spacetime is a linear theory and obeys the superposition principle. It is possible to find a solution describing an arbitrary number of electric charges at rest in arbitrary positions by adding the corresponding Coulomb solutions. With our normalizations we would have

$$A_\mu = -\delta_{\mu t} \sum_{i=1}^{N} \frac{2G_{\rm N}^{(4)} q_i}{|\vec{x}_3 - \vec{x}_{3,i}|}, \tag{8.88}$$

in a certain gauge. As we have stressed before, Maxwell's theory in Minkowski spacetime does not know about interactions and this is why we can have a static solution, which we know would be possible in the real world only if there were another force holding the charges in place. If we introduce source terms for the charges (massive or massless point-like particles of electric charges q_i) then we will have to solve a (non-linear) coupled system of equations: the Maxwell field equations and the equations of motion for the particles. The solutions will be in general time-dependent (and realistic).

Newtonian gravity is another linear theory and, thus, there are static solutions corresponding to arbitrary mass distributions even if we know that external forces are needed to hold the masses in place. Again, on introducing sources, the solutions become realistic (and, in general, time-dependent).

Now, if we again introduce sources interacting both gravitationally and electrostatically, we can have static solutions describing particles with masses and charges $M_i = 2|q_i|$ in equilibrium. Newtonian gravity is insensitive to the electrostatic interaction energy and to the gravitational interaction energy.

---

[14] One can also try to look for solutions of this form in the scalar-coupled-to-gravity system. Since the force between two objects with "scalar charge" is always attractive, we do not expect on physical grounds to find any. In fact, it is possible to find such solutions if we pay the price of having purely imaginary "scalar charges" (which repel each other). The solutions have the following form:

$$\begin{cases} ds^2 = e^{2H}dt^2 - e^{-2H}d\vec{x}_3^{\,2}, \\ \quad \varphi = c \pm iH, \end{cases} \tag{8.87}$$

where c is any constant and H is any harmonic function $\partial_i\partial_i H = 0$.

In GR, a non-linear (non-Abelian, self-coupling) theory, things are quite different. There is no need to introduce sources: the theory *knows* that two Schwarzschild BHs, for instance, cannot be in static equilibrium and the corresponding solution does not exist. The coupling to gravity makes the electromagnetic interaction effectively non-Abelian, and it does not need the introduction of sources to *know* that only ERN BHs can be in static equilibrium[15] [185]. This coupling gives rise to many other interesting phenomena in RN backgrounds, such as the conversion of electromagnetic into gravitational waves [740].

Since the horizon of a single ERN BH looks like a point in isotropic coordinates, we can try harmonic functions with several point-like singularities:

$$H(\vec{x}_3) = 1 + \sum_{i=1}^{N} \frac{2G_{\rm N}^{(4)}|q_i|}{|\vec{x}_3 - \vec{x}_{3,i}|}. \tag{8.89}$$

The overall normalization is chosen so as to obtain an asymptotically flat solution and the coefficients of each pole are taken positive so that $H(\vec{x}_3)$ is nowhere vanishing and the metric is non-singular. Also this choice gives a potential like the one in Eq. (8.88) for large values of $|\vec{x}_3|$.

It can be seen [500] that each pole of H indeed corresponds to a BH horizon. In fact, to see that there is a surface of finite area at $\vec{x}_{3,i}$, one simply has to shift the origin of coordinates to that point and then examine the $\rho \to 0$ limit as in the single-BH case. The charge of each BH can be calculated most simply using Eq. (8.43), where the volume encloses only one singularity (the current is nothing but a collection of Dirac-delta terms). The charges turn out to be $\text{sign}(-\alpha)|\,q_i|$, i.e. all the charges have the same sign.

In GR it is, however, impossible to calculate the *mass* of each BH because there is no local conservation law for the mass and there is no such concept as the mass of some region of the spacetime. Only one mass can be defined, which is the total mass of the spacetime and this is $M = 2\sum_{i=1}^{N}|q_i|$. However, the equilibrium of forces existing between the black holes suggests that the electrostatic and gravitational interaction energies (to which GR gravity is sensitive) cancel out everywhere. If that were true, the masses and charges would be *localized* at the singularities and then we could assign a mass $M_i = 2|q_i|$ to each black hole [185]. It is, perhaps, this localization of the mass of ERN BHs which will allow us to find sources for them, something that turned out to be impossible for Schwarzschild BHs. This is physically a very appealing idea, but it is certainly not a rigorous proof.

If we do not care about singularities, we can also take some coefficients of the poles of the harmonic function to be negative. In this way it is possible to obtain solutions with vanishing total mass. Here, it is intuitively clear that

[15] As a matter of fact, the identity $M_1 M_2 = 4|q_1 q_2|$ does not imply that both objects are ERN BHs. It can be satisfied by a non-extremal RN BH with $M_1 > 2|q_1|$ and a naked singularity with $M_2 < 2|q_2|$, but the corresponding static solutions (if any) are not known.

a negative coefficient is associated with some "negative mass density" and cosmic censorship should eliminate these solutions.

(e) If we take the *near-horizon limit* $\rho \to 0$ in the ERN metric Eqs. (8.83), the constant 1 can be ignored and we find another MP solution with harmonic function $H = 2G_{\rm N}^{(4)}|q|/\rho$:

$$\boxed{\begin{aligned} ds^2 &= \left(\frac{\rho}{2G_{\rm N}^{(4)}|q|}\right)^2 dt^2 - \left(\frac{\rho}{2G_{\rm N}^{(4)}|q|}\right)^{-2} d\rho^2 - \left(2G_{\rm N}^{(4)}|q|\right)^2 d\Omega^2_{(2)}, \\ A_t &= -\frac{\rho}{G_{\rm N}^{(4)} q}, \qquad F_{t\rho} = \frac{1}{G_{\rm N}^{(4)} q}. \end{aligned}} \tag{8.90}$$

This exact solution is the Robinson–Bertotti (RB) solution [146, 812] and describes the ERN metric near the horizon. It is the only solution of the Einstein–Maxwell equations which is homogeneous and has a homogeneous non-null electromagnetic field (Theorem 10.3 in [640]). It is the direct product of two two-dimensional spaces of constant curvature: a two-dimensional anti-de Sitter (AdS_2) spacetime with "radius" $R_{\rm AdS} = 2G_{\rm N}^{(4)}|q|$ and therefore with two-dimensional scalar curvature $R^{(2)} = -1/[2(G_{\rm N}^{(4)}|q|)^2]$, in the $t-\rho$ part of the metric and a 2-sphere S^2 of radius $R_{\rm S} = 2G_{\rm N}^{(4)}|q|$ and curvature $R^{(2)} = +1/[2(G_{\rm N}^{(4)}|q|)^2]$ in the θ–φ part of the metric. The sum of the two-dimensional scalar curvatures vanishes, as it should, because all solutions of the Einstein–Maxwell system have $R = 0$. Evidently, it is not asymptotically flat.

AdS_2 is invariant under the isometry group $SO(1, 2)$ (which is also called AdS_2) and S^2 under $SO(3)$. If we compare the RB isometry group with the ERN isometry group ($SO(1, 1) \times SO(3)$ and $SO(1, 1) \sim \mathbb{R}^+ \times \mathbb{Z}_2$ are shifts in time and time inversions) we see that there is an enhancement of symmetry when we approach the horizon. As we will see in Chapter 13, there is also an enhancement of unbroken supersymmetry, which is maximal in this limit. This is enough to consider the RB solution as a vacuum of the theory alternative to Minkowski.

In turn, this allows us to view the ERN solution as interpolating between the Minkowski vacuum (which is at infinity) and the RB solution (which is at the horizon), and then we can interpret it as a *gravitational soliton* [433].

(f) There are many other solutions in the MP class. However, it has been argued in [500] that the only BH solutions in this class (and in a bigger class that we will study later, the IWP class) are the ones we have written above. One could look for solutions describing extended objects by allowing the harmonic function H to have one- or two-dimensional singularities. They are not asymptotically flat and they are not natural, so we will not consider them.

9. If we shift the radial coordinate by $r = \rho + r_\pm$ in the RN solution Eq. (8.75), it takes the following form:

$$
\begin{aligned}
ds^2 &= \left(1+\frac{r_\pm}{\rho}\right)^{-2}\left(1+\frac{\pm 2r_0}{\rho}\right)dt^2 \\
&\quad - \left(1+\frac{r_\pm}{\rho}\right)^{2}\left[\left(1+\frac{\pm 2r_0}{\rho}\right)^{-1}d\rho^2 + \rho^2 d\Omega^2_{(2)}\right], \\
A'_\mu &= -\delta_{\mu t}\frac{4G_N^{(4)}q}{r_\pm}\left[\left(1+\frac{r_\pm}{\rho}\right)^{-1} - 1\right].
\end{aligned}
\tag{8.91}
$$

The RN metric looks in this form (taking the minus sign) like a Schwarzschild metric with mass $r_0/G_N^{(4)}$ "dressed" with some factors related to the gauge potentials or, alternatively, as the ERN solution dressed with some Schwarzschild-like factors. The Schwarzschild component of this metric completely disappears in the extreme limit, leaving an ERN isotropic metric. This form of charged BH metric is quite common and occurs, as we will see, in various contexts, rewritten in this way:

$$
\boxed{
\begin{aligned}
&ds^2 = H^{-2}W dt^2 - H^2\left[W^{-1}d\rho^2 + \rho^2 d\Omega^2_{(2)}\right], \\
&A_\mu = \delta_{\mu t}\alpha\left(H^{-1} - 1\right), \\
&H = 1 + \tfrac{h}{\rho}, \qquad W = 1 + \tfrac{\omega}{\rho}, \qquad \omega = h\left[1 - (\alpha/2)^2\right].
\end{aligned}
}
\tag{8.92}
$$

We will obtain many solutions in this form. Afterwards, we will identify the integration constants that appear in them in terms of the physical constants:

$$
\alpha = -4G_N^{(4)}q/r_\pm, \qquad h = r_\pm, \qquad \omega = \pm 2r_0. \tag{8.93}
$$

10. Another useful coordinate system for charged BHs [446], which covers the BH exterior and in which the radial coordinate τ takes values between $-\infty$ on the horizon and 0 at spatial infinity, can be obtained by the transformation of the coordinate ρ above,

$$
\rho = -\frac{r_0 e^{-r_0\tau}}{\sinh(r_0\tau)}, \tag{8.94}
$$

so the metric takes the form

$$
\begin{aligned}
ds^2 &= e^{2U}dt^2 - e^{-2U}\left[\frac{r_0^4}{\sinh^4(r_0\tau)}d\tau^2 + \frac{r_0^2}{\sinh^2(r_0\tau)}d\Omega^2_{(2)}\right], \\
e^{2U} &= \left(1 + \frac{r_-}{2r_0} - \frac{r_-}{2r_0}e^{2r_0\tau}\right)^{-2} e^{2r_0\tau}.
\end{aligned}
\tag{8.95}
$$

11. Finally, BH solutions for an action containing several different vector fields A^I_μ, $I = 1, \ldots, N$, can easily be found. Let us consider the action

$$\boxed{S[g_{\mu\nu}, A^I{}_\mu] = \frac{1}{16\pi G_N^{(4)}} \int d^4x \sqrt{|g|} \left[R - \tfrac{1}{4} \sum_{I=1}^{I=N} \left(F^I\right)^2 \right].} \tag{8.96}$$

This action is invariant under global $O(N)$ rotations of the N vector field strengths. This is a simple example of *duality symmetry*. Now, any solution of the Einstein–Maxwell theory (one vector field) is a solution of this theory with the remaining $N-1$ vector fields equal to zero, and, by performing general $O(N)$ rotations, one can generate new solutions in which the N vector fields are non-trivial. It is clear that, if the original solution had the electric charge q_1, the electric charges of the new solution q_i will satisfy $\sum_{i=1}^N q_i'^2 = q_1^2$. This duality symmetry does not act on the metric and, therefore, all one has to do is to replace q_1^2 by $\sum_{i=1}^N q_i'^2$ in it.

For example, had we started from the RN solution (8.92), we would have obtained by this procedure a RN solution with many Abelian electric charges:

$$\boxed{\begin{aligned} ds^2 &= H^{-2} W dt^2 - H^2 \left[W^{-1} d\rho^2 + \rho^2 d\Omega^2_{(2)}\right], \\ A^I{}_\mu &= \delta_{\mu t} \alpha^I \left(H^{-1} - 1\right), \\ H &= 1 + h/\rho, \qquad W = 1 + \tfrac{\omega}{\rho}, \\ \omega &= h \left[1 - \sum_{I=1}^{I=N} \left(\frac{\alpha^I}{2}\right)^2 \right], \end{aligned}} \tag{8.97}$$

and

$$\alpha^I = -4 G_N^{(4)} q^I / r_\pm, \qquad h = r_\pm, \qquad \omega = \pm 2 r_0, \tag{8.98}$$

where now

$$r_\pm = G_N^{(4)} M \pm r_0, \qquad r_0 = G_N^{(4)} \left(M^2 - 4 \sum_{I=1}^{I=N} q_I^2 \right)^{\frac{1}{2}}. \tag{8.99}$$

This is the first and simplest example of the use of duality symmetries as solution-generating symmetries. We will find more-complex examples later on, but the main ideas are the same.

Observe that, in this procedure of generating new solutions out of known ones, the new solutions are expressed at the beginning in terms of the old physical parameters and the parameters of the duality transformation (in this case, $O(N)$ and sines and cosines of angles). Then one has to identify those constants in terms of the physical parameters of the new solution. This is usually quite a painful calculation (sometimes, in cases more complicated than this one, it is impossible to do) unless one uses invariance properties such as the invariance of $\sum_{i=1}^{i=N} q_i^2$ under $\mathrm{O}(N)$ transformations.

In the end, one should obtain a general duality-invariant family of solutions such that a further duality transformation takes us to another member of the family but the form of the general solution no longer changes. These families of solutions reflect many of the symmetries of the theory and depend only on duality-invariant combinations of charges and moduli.

The family we have obtained is duality-invariant: the effect of a further duality transformation is just to replace all charges by primed charges, but the general form of the solution does not change.

8.4 The Sources of the electric RN black hole

Just as we did with the Schwarzschild solution, we want to try to find a source for the RN solution such that it becomes a solution everywhere, including at the singularity $r = 0$. Our candidate source will be a point-like particle at rest at $r = 0$ whose mass and electric charge match those of the RN BH. As in the Schwarzschild case, our expectations are not good because most of the reasons why we were unsuccessful (delocalization of the gravitational energy and the infinite self-force of the particle) are valid also in the general RN case. The only change is the causal nature of the singularity: spacelike in the Schwarzschild case, timelike in the RN case. However, one could argue that the gravitational and electromagnetic energy densities in the ERN BH cancel each other out everywhere so they are somehow *localized* at the origin $r = 0$ and, thus, in this particular case we have some hope.

Our starting point is, therefore, the action of the Einstein–Maxwell system Eq. (8.58) coupled to the action of a massive, charged particle ($c = 1$):

$$S = \frac{1}{16\pi G_{\mathrm{N}}^{(4)}} \int d^4x \sqrt{|g|}\, \left[R - \tfrac{1}{4}F^2\right] - M \int d\xi\, \sqrt{g_{\mu\nu}(X)\dot{X}^\mu \dot{X}^\nu} - q \int d\xi\, A_\mu \dot{X}^\mu. \quad (8.100)$$

The equations of motion of the dynamical fields $g_{\mu\nu}$, A_μ, and X^μ are, respectively,

$$G_{\mu\nu} - 8\pi G_{\mathrm{N}}^{(4)} T_{\mu\nu}^{(A)} + \frac{8\pi G_{\mathrm{N}}^{(4)} M}{\sqrt{|g|}} \int d\xi\, \frac{g_{\mu\rho} g_{\nu\sigma} \dot{X}^\rho \dot{X}^\sigma}{\sqrt{|g_{\lambda\tau}\dot{X}^\lambda \dot{X}^\tau|}} \delta^{(4)}[X(\xi) - x] = 0, \quad (8.101)$$

$$\partial_\mu\left(\sqrt{|g|}F^{\mu\nu}\right) - 16\pi G_{\mathrm{N}}^{(4)} q \int d\xi\, \dot{X}^\nu \delta^{(4)}[X(\xi) - x] = 0, \quad (8.102)$$

$$\gamma^{1/2} M \nabla^2(\gamma) X^\lambda + M\gamma^{-1/2} \Gamma_{\rho\sigma}{}^{\lambda} \dot{X}^\rho \dot{X}^\sigma - qF^\lambda{}_\rho \dot{X}^\rho = 0, \quad (8.103)$$

where

$$\gamma = g_{\mu\nu}(X)\dot{X}^\mu \dot{X}^\nu. \quad (8.104)$$

Let us first consider the Einstein equation. We use the RN solution in the coordinates Eqs. (8.92) with the upper sign and, following the same steps as in the Schwarzschild case, we obtain the following non-vanishing components of the Einstein tensor:[16]

$$
\begin{aligned}
G_{00} &= g_{00}H^{-2}\{-\delta(W) + 2(WH^{-1})\delta(H) + (WH^{-1})'H'\},\\
G_{\rho\rho} &= g_{\rho\rho}H^{-2}\{-\delta(W) + (WH^{-1})'H'\},\\
G_{\theta\theta} &= g_{\theta\theta}H^{-2}\{\tfrac{1}{2}\delta(W) - (WH^{-1})'H'\},\\
G_{\varphi\varphi} &= \sin^2\theta G_{\theta\theta},
\end{aligned}
\tag{8.105}
$$

where we are using the notation

$$
\delta(W) = -\frac{4\pi}{\sin\theta}\omega\delta^{(3)}(\rho), \qquad \delta(H) = -\frac{4\pi}{\sin\theta}h\delta^{(3)}(\rho). \tag{8.106}
$$

The electromagnetic energy–momentum tensor does not need to be calculated explicitly. It does not have any distributional term (δ function) and we know that it cancels out exactly the finite terms of the Einstein tensor. Thus, in the Einstein equation, we need only focus on the distributional terms coming from the Einstein tensor and the particle's energy–momentum tensor, which depends on our *Ansatz* for X^μ. For a particle at rest at $\vec{x}_3 = \vec{0}$, we must set[17]

$$
X^\mu = \delta^{\mu 0}\xi, \tag{8.107}
$$

but with this choice only the 00 component of the particle's energy–momentum tensor is non-vanishing, as in the Schwarzschild case. However, in the extreme case $\omega = 0$ only the 00 component of the Einstein has a distributional term that matches exactly the particle's energy–momentum tensor

$$
8\pi G_N^{(4)} M H^{-5}\delta^{(3)}(\vec{x}_3) \tag{8.108}
$$

(after integration over ξ).

The Maxwell equation is also satisfied (even in the non-extreme case). Let us turn to the particle's equation of motion. The time component is just $dg_{00}^{-1/2}/d\xi = dH/d\xi$ in the extreme case. H diverges on the particle's path and, even though it is independent of ξ, we cannot say that this equation is truly solved. The radial component can be put in the form

$$
-M\partial_r g_{00}^{\frac{1}{2}} - q\partial_r A_0 = 0, \tag{8.109}
$$

and it is satisfied identically by the ERN solution.[18] If we considered the motion of any other particle with $M' = 2|q'|$, we would see that it can be at rest anywhere in the ERN solution.

These kinds of cancelations are indications of supersymmetry, which, as we will see, is present in the ERN solution.

---

[16] Needless to say, the mathematical rigor in all these manipulations is scarce. For instance, we feel free to multiply delta functions by functions that may diverge or be zero if we integrated the product. Some of these manipulations could possibly be justified by working with tensor densities instead of tensors, etc. The ultimate justification for presenting these calculations is the result, which allows us to match physical parameters such as mass and electric charge with integration constants of solutions.

[17] Observe that, in this coordinate system, $\vec{x}_3 = \vec{0}$ is the event horizon! The δ functions that we obtain have support only there and we are forced to make this *Ansatz* if we want the particle's energy–momentum tensor and electric current to reproduce the singularities of the Einstein tensor and the Maxwell equation.

[18] Again, we manipulate g_{00} etc., not taking into account that they are zero or diverge along the particle's path.

8.5 Thermodynamics of RN black holes

As we said in the discussion of Schwarzschild BH thermodynamics, most of the results can be generalized to BHs containing charges or angular momentum. In particular, the zeroth and second laws of BH thermodynamics take exactly the same form and so do the identifications between the surface gravity and temperature and horizon, Eq. (7.45), and between area and entropy, Eq. (7.46). The first law requires the addition of a new term that takes into account the possible changes in the BH mass due to changes in the charge ($\hbar = c = 1$),

$$\boxed{dM = \frac{1}{8\pi G_{\rm N}^{(4)}}\kappa dA + \phi^{\rm h} dq,} \tag{8.110}$$

where $\phi^{\rm h}$ is the electrostatic potential on the horizon. In this case

$$\begin{aligned} T &= \frac{r_0}{2\pi r_+^2} = \frac{1}{2\pi G_{\rm N}^{(4)}} \frac{\sqrt{M^2 - 4q^2}}{\left(M + \sqrt{M^2 - 4q^2}\right)^2}, \\ S &= \frac{\pi r_+^2}{G_{\rm N}^{(4)}} = \pi G_{\rm N}^{(4)} \left(M + \sqrt{M^2 - 4q^2}\right)^2, \\ \phi^h &= \phi(r_+) = \frac{4G_{\rm N}^{(4)} q}{r_+}. \end{aligned} \tag{8.111}$$

and the Smarr formula takes the form

$$M = 2TS + q\phi^{\rm h}. \tag{8.112}$$

It is worth stressing that the above formulae have been obtained using a generic RN metric (i.e. non-extremal). However, we know that the limit in which we approach the ERN solution with $M = 2|q|$ is not continuous: the topology of the ERN, its causal structure, is different from that of any non-extreme RN BH, no matter how close to the extreme limit it is. Furthermore, it seems that the extreme limit cannot be approached by a finite series of physical processes (the third law of BH thermodynamics) and it has also been argued that the thermodynamical description of the RN BH breaks down when we approach the extreme limit [790] (see also [653]): close enough to the extreme limit, the emission of a single quantum with energy equal to the Hawking temperature would take the mass of the RN BH *beyond* the extreme limit. Then, the change in the spacetime metric caused by Hawking radiation would be very big and Hawking's calculation in which backreaction of the metric to the radiation is ignored becomes inconsistent.

For all these reasons we may expect surprises if we naively take the limit $M \to |q|$ in the above formulae, but this seems not to happen: in that limit the temperature vanishes and the entropy remains finite and, if we calculate both directly on the ERN solution, we find the same result. In any case, this is a very important issue because essentially these are the only BHs for which a statistical computation of the entropy based on string theory has been performed, and we should try to compute both by other methods, for instance using

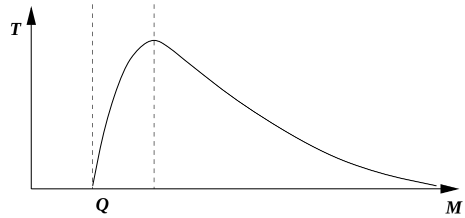

Fig. 8.3. The temperature T versus the mass M of a Reissner–Nordström black hole of charge $Q = 2q$.

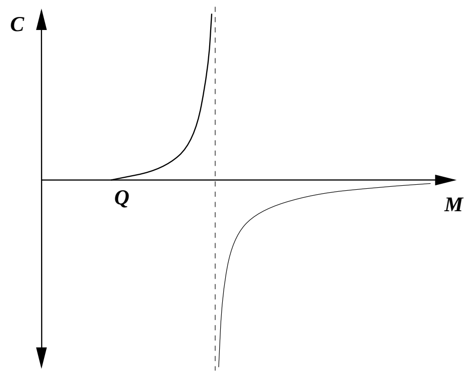

Fig. 8.4. The specific heat at constant charge C versus the mass M of a Reissner–Nordström black hole of charge $Q = 2q$.

the Euclidean path integral formalism. Before we do so, let us mention other remarkable aspects of the RN BH thermodynamics.

We have drawn the behavior of the RN BH temperature for fixed charge in Figure 8.3. In it we see that, for large values of the mass, the temperature diminishes when the mass grows, just as in the Schwarzschild BH, but, for values of the mass comparable to the charge, close to the extreme limit, the temperature grows with the mass, as in any ordinary thermodynamical system. There is a maximum temperature for RN BHs (for constant charge), which is reached for $M^\star = 4|q|/\sqrt{3}$. The maximum value for the temperature is given by $T^\star$:

$$T^\star = T(M^\star, q) = \frac{1}{12\sqrt{3}\pi G_{\rm N}^{(4)}|q|}. \tag{8.113}$$

In the plot of the specific heat at constant charge C of Figure 8.4 we clearly see the two regions in which the thermodynamical behavior is "standard" (positive specific heat) and "Schwarzschild-like" (negative specific heat). At the point $M^\star$ at which the temperature reaches its maximum value, $\partial T/\partial M = 0$ and the specific heat diverges. It is tempting to associate that divergence with a phase transition between the two kinds of behavior. It is also tempting to associate the success of the statistical calculation of the ERN BH entropy with its standard thermodynamical behavior in its neighborhood.

What would be the endpoint of the Hawking evaporation of a RN BH? As we mentioned before, the electric charge is lost faster than the mass and, before the extreme limit is reached, we should have an uncharged Schwarzschild BH, whose fate we have already discussed. We can, however, speculate what would happen if the charge of the RN were of a kind not carried by any elementary particle[19] so that it could not be lost by Hawking radiation, or if the carriers of that kind of charge were extremely heavy particles[20] (unlike electrons) so that the BH would discharge much more slowly than it would lose mass. In these cases, assuming that nothing special happens when the mass is such that $\partial T/\partial M = 0$, one would expect the RN BH to approach the extreme limit in a very long-lasting (perhaps eternal) process in which the BH losses mass and temperature at lower and lower rates. It has been conjectured that the ERN BH could be a BH remnant storing all the information contained in the original BH that is not radiated away.

8.6 The Euclidean electric RN solution and its action

The Euclidean (non-extreme) RN solution has a structure identical to that of the Euclidean Schwarzschild solution and, in particular, it covers only the BH exterior and it also requires the compactification of the Euclidean time in order to eliminate a conical singularity. This allows us to calculate the temperature again by finding the period of the Euclidean time that makes the metric on the horizon regular. If we use spherical coordinates with origin on the horizon (like those of Eq. (8.92) with the upper sign) then we see that $T = g'_{\tau\tau}(0)/(4\pi) = \omega/(4\pi h^2)$. With this period of the Euclidean time, the topology is $\mathbb{R}^2 \times S^2$.

Let us now study the Euclidean ERN solution directly. The interesting region is the neighborhood of the horizon and to study it we expand as usual the metric components of Eq. (8.92), (with $\omega = 0$) in a power series in the inverse of the radial coordinate around the origin and keep the lowest-order terms; instead of an approximate solution, we have the Euclidean continuation of the RB metric Eq. (8.90). This metric is completely regular for any periodicity of the Euclidean time. It is convenient to use the coordinate $r = R\ln(\rho/R)$ with $R = 2G_{\rm N}^{(4)}|q|$ in which the Lorentzian RB solution becomes

$$\boxed{\begin{aligned} & ds^2 = e^{2r/R}dt^2 - dr^2 - R^2 d\Omega^2_{(2)}, \\ & A_t = -2e^{r/R}, \qquad F_{tr} = \frac{2}{R}e^{\frac{r}{R}}, \qquad R = 2G_{\rm N}^{(4)}|q|. \end{aligned}} \tag{8.114}$$

[19] This BH could not have been created by standard gravitational collapse. Instead, a process like quantum pair creation has to be invoked to justify its existence.

[20] For instance, the carriers of Kaluza–Klein charges and the massive modes in string theory are usually assigned very large masses (of the order of the Planck mass) in order to explain why they have not yet been observed.

In these coordinates it is evident that the horizon $r = -\infty$ is at an infinite distance in the r direction and a constant-time slice of this spacetime looks like an infinite tube whose $r =$ constant sections are 2-spheres of constant radius R. It does not make much sense to talk about the period of τ that makes the Wick-rotated metric on the horizon regular because it is regular for any period. The same applies to flat Euclidean spacetime. The temperature cannot be uniquely assigned in this formalism. The reason could be the fact that both Minkowski spacetime and the RB solution can be considered vacua of the theory.

As a conclusion of this discussion, then, if we compactify the Euclidean time with some arbitrary period β, the topology is not $\mathbb{R}^2 \times \mathrm{S}^2$ as in the non-extreme case, but $\mathbb{R} \times \mathrm{S}^1 \times \mathrm{S}^2$. The factor $\mathbb{R} \times \mathrm{S}^1$, with the topology of a cylinder, corresponds to $\mathbb{R}^2 - \{0\}$, the τ–r plane with the point at the origin (the event horizon, which is at an infinite distance) removed. The Euclidean RN solution has, therefore, two boundaries: at infinity (as in the non-extreme case) and at the horizon. One way to check this fact is to calculate the Euler characteristic of the Euclidean ERN solution using the Gauss–Bonnet theorem adapted to manifolds with boundaries. The Euler characteristic χ is a topological invariant whose value is an integer and the Gauss–Bonnet theorem states that the integral of the 4-form

$$\frac{1}{32\pi^2}\epsilon_{abcd}R^{ab} \wedge R^{cd} \tag{8.115}$$

over a four-dimensional compact manifold M is precisely χ. If the manifold has a boundary ∂M, then χ(M) is given by the integral over M of the above 4-form plus the integral over the boundary of a 3-form [236, 347],

$$\chi(\mathrm{M}) = \frac{1}{32\pi^2}\int_{\mathrm{M}}\epsilon_{abcd}R^{ab} \wedge R^{cd} - \frac{1}{32\pi^2}\int_{\partial\mathrm{M}}\epsilon_{abcd}\left[2\theta^{ab} \wedge R^{cd} - \tfrac{4}{3}\theta^{ab} \wedge \theta^{c}{}_{e} \wedge \theta^{ed}\right], \tag{8.116}$$

where θ^{ab} is the *second fundamental 1-form* on ∂M, that can be constructed as explained in [347]. The contribution of the boundary integral is crucial in order to have $\chi = 2$ in the non-extreme case, corresponding to the topology $\mathbb{R}^2 \times \mathrm{S}^2$. In the extreme case, only by taking into account the boundary at the horizon does one obtain $\chi = 0$, the correct value for the topology $\mathbb{R} \times \mathrm{S}^1 \times \mathrm{S}^2$ [445].

This is going to have important consequences in what follows.

Once we have determined the period, we are ready to calculate the partition function using the Euclidean path-integral formalism in the saddle-point approximation. We are going to do it as in the Schwarzschild case, using the Lorentzian action and solution but taking into account the periodicity of the Euclidean time and the fact that the Euclidean solution covers only the exterior of the horizon.

In $\hbar = c = G_{\mathrm{N}}^{(4)} = 1$ the Einstein–Maxwell system with boundary terms is

$$S_{\mathrm{EM}}[g_{\mu\nu}, A_\mu] = \frac{1}{16\pi}\int d^4x\,\sqrt{|g|}\left[R - \tfrac{1}{4}F^2\right] + \frac{1}{8\pi}\int d^3\Sigma(\mathcal{K} - \mathcal{K}_0), \tag{8.117}$$

and, using the definition of $F_{\mu\nu}$ and integrating by parts, we rewrite it in the form

$$\begin{aligned} S_{\mathrm{EM}}[g_{\mu\nu}, A_\mu] = {} & \frac{1}{16\pi}\int d^4x\,\sqrt{|g|}[R + \tfrac{1}{2}A_\nu\nabla_\mu F^{\mu\nu}] \\ & + \frac{1}{8\pi}\int d^3\Sigma[(\mathcal{K} - \mathcal{K}_0) + \tfrac{1}{4}n_\mu F^{\mu\nu}A_\nu]. \end{aligned} \tag{8.118}$$

Only the boundary term contributes to the action because the volume term vanishes on-shell. Furthermore, for a generic non-extreme RN BH there is only one boundary at infinity. The contribution of the extrinsic curvature terms for large values of $r_{\rm c}$ is always given by Eq. (7.73) for spherically symmetric, static, asymptotically flat metrics such as the RN metric. The electromagnetic boundary term has to be computed in the gauge in which A_μ vanishes on the horizon, i.e. using $A'_\mu \equiv A_\mu - A_\mu(r_+)$, because the Killing vector $\partial/\partial\tau$ is singular on the event horizon (which is a Killing horizon). We find

$$\tfrac{1}{4} n_\mu F'^{\mu\nu} A'_\nu = -\frac{q}{r^2\sqrt{-g_{rr}g_{tt}}}\left(\frac{4q}{r} - \frac{4q}{r_+}\right)\Bigg|_{r=r_{\rm c}\rightarrow\infty} \sim \frac{4q^2}{r_{\rm c}^2 r_+} + \mathcal{O}(r_{\rm c}^{-3}). \quad (8.119)$$

Finally, taking into account that

$$d^3\Sigma = dt d\Omega^2 r^2 \sqrt{g_{tt}}\big|_{r=r_{\rm c}\rightarrow\infty} \sim dt d\Omega^2 r_{\rm c}^2, \quad (8.120)$$

we find in the limit $r_{\rm c} \rightarrow \infty$ for the Euclidean action

$$-\tilde{S}_{\rm EM} = -\frac{\beta}{2}[M - q\phi(r_+)] = -\frac{\beta}{2} r_0, \quad (8.121)$$

where $\phi(r_+) = A_t(r_+)$ is the electrostatic potential on the horizon. The entropy is

$$S = \beta[M - q\phi(r_+)] + \ln\mathcal{Z} = +\frac{\beta}{2}[M - q\phi(r_+)] = \frac{\beta}{2} r_0 = \pi r_+^2, \quad (8.122)$$

that is, one quarter of the area of the horizon.

This calculation is valid for generic non-extreme RN BHs. We should now repeat the calculation directly for ERN BHs. There are two important differences.

1. The period β of the Euclidean time is not determined.

2. The Euclidean ERN solution has another boundary at the horizon, and the action contains the contribution of the boundary at infinity, given in Eq. (8.121), and the contribution from the new boundary that we can calculate straightaway:

$$-\frac{i}{8\pi}\int_0^{-i\beta} dt \int_{S^2} d\Omega^2 r^2 \sqrt{g_{tt}}\left\{\frac{1}{\sqrt{-g_{rr}}}\left[\tfrac{1}{2}\partial_r \ln g_{tt} + \frac{2}{r}\right] - \frac{2}{r}\right.$$
$$\left. - \frac{q}{r^2\sqrt{-g_{rr}g_{tt}}}\left(\frac{4q}{r} - \frac{4q}{r_+}\right)\right\}, \quad (8.123)$$

where we have to substitute $r = r_+$. The result is $-\beta r_0/2$ and, thus, we have

$$-\tilde{S}_{\rm EM} = -\beta r_0, \quad (8.124)$$

which gives[21] identically [445, 516, 883]

$$S = \beta[M - q\phi(r_+)] + \ln\mathcal{Z} = 0. \quad (8.125)$$

[21] Of course, βr_0 would be identically zero for ERN BHs ($r_0 = 0$) if β were taken finite.

It has been suggested that the same is true for any extreme charged BH, not just ERN BHs, and also that the Bekenstein–Hawking entropy formula Eq. (7.46) should be [652]

$$S = \frac{\chi A c^3}{8 G_{\mathrm{N}}^{(4)} \hbar}. \tag{8.126}$$

Since one of the main successes of string theory has been the calculation of the (finite!) entropy of the ERN BH, this result is a bit disturbing. Actually it implies that string theory and the Euclidean path-integral approach to quantum gravity give different predictions for the entropy of the ERN BH. It has been argued in [547] that the near-horizon ERN geometry suffers important corrections in string theory. The reason would be that, although the topology is that of a cylinder, the geometry is rather that of a *pipette*, with a radius that tends to zero at infinity when we asymptotically approach the horizon. String theory compactified on a circle undergoes a phase transition when the radius reaches the *self-dual* value. Thus, beyond the point of the pipette at which the radius has that value, the geometry may indeed change,[22] although no precise calculations have been done so far.

8.7 Electric-magnetic duality

As we explained in Section 8.2, the full set of sourceless Maxwell equations (the Maxwell equation plus the Bianchi identity) is invariant (up to signs) under the replacement of the field strength F by its dual $\tilde{F} = {}^\star F$

$$F \to \tilde{F} = {}^\star F. \tag{8.127}$$

This is true in flat as well as in curved spacetime. In a given frame, this transformation corresponds to the interchange of electric and magnetic fields according to Eq. (8.38), hence the name *electric–magnetic duality*. This transformation squares to (minus) the identity and, therefore, it generates a $\mathbb{Z}_2$ electric–magnetic-duality group.

The $\mathbb{Z}_2$ can easily be extended to a continuous symmetry group:[23]

$$\tilde{F} = aF + b{}^\star F, \qquad \Rightarrow {}^\star\tilde{F} = -bF + a{}^\star F, \qquad a^2 + b^2 \neq 0, \tag{8.128}$$

is an invertible transformation that leaves the set of the two equations invariant (up to factors). It is convenient to define the *duality vector*

$$\vec{F} \equiv \begin{pmatrix} F \\ {}^\star F \end{pmatrix}. \tag{8.129}$$

It is subject to the constraint

$${}^\star\vec{F} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \vec{F}, \tag{8.130}$$

with which the Maxwell equations can be written as

$$\nabla_\mu \vec{F}^{\mu\nu} = 0, \tag{8.131}$$

[22] See analogous discussions on page 577 about the *correspondence principle*.

[23] For the moment, all these are classical considerations. We will see that quantum effects (in particular, charge quantization) break the continuous symmetry to a discrete subgroup.

and it transforms in the vector representation of the duality group, a subgroup of $\mathrm{GL}(2, \mathbb{R})$:

$$\tilde{\vec{F}} = M\vec{F}, \qquad M = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} = \pm\lambda \begin{pmatrix} \cos\xi & \sin\xi \\ -\sin\xi & \cos\xi \end{pmatrix}. \tag{8.132}$$

In this form we see that the duality group consists of rescalings and O(2) rotations of $\vec{F}$.

Observe that, if we integrate the Hodge dual of the duality vector ${}^\star\vec{F}$ over a 2-sphere at infinity we obtain a *charge vector* whose first component is $16\pi G_{\rm N}^{(4)}q$, in our conventions. The second component will be, by definition, the *magnetic charge* p:

$$\int_{S^2_\infty} {}^\star\vec{F} = \begin{pmatrix} 16\pi G_{\rm N}^{(4)} q \\ p \end{pmatrix} \equiv 16\pi G_{\rm N}^{(4)} \vec{q}, \qquad \vec{q} = \begin{pmatrix} q \\ p/(16\pi G_{\rm N}^{(4)}) \end{pmatrix}. \tag{8.133}$$

Although this transformation looks very simple written in terms of the electromagnetic field strength $F_{\mu\nu}$, it is very non-local in terms of the true field variable A_μ. To see this, we simply have to use Eq. (8.31) to obtain an explicit relation between A_μ and the dual vector field $\tilde{A}_\mu$:

$$\tilde{A}_\mu(x) = -\int_0^1 d\lambda \lambda x^\nu \frac{\epsilon_{\mu\nu}{}^{\rho\sigma}}{\sqrt{|g|}} \partial_\rho A_\sigma(\lambda x). \tag{8.134}$$

This non-locality is, at the same time, what makes this duality transformation interesting and the source of problems. To start with, the replacement of F by ${}^\star F$ is not a symmetry of the Maxwell action because $({}^\star F)^2 = -F^2$. The reason for this is that the transformation should be done on the right variable, namely the vector field, but this is difficult to do. Another possibility is to write an action that really is a functional of the field strength. On this action, the above replacement can be performed and gives the right results. This procedure is called *Poincaré duality* and we explain it in detail in Section 8.7.1.

Let us now see what modifications the coupling to gravity Eq. (8.58) produces. The main difference is that we now have one more equation (Einstein's). For our purposes, it is useful to rewrite it in this form (see Section 1.6 and Eq. (1.126)):

$$G_{\mu\nu} - \left(F_\mu{}^\rho F_{\nu\rho} + {}^\star F_\mu{}^{\rho\star} F_{\nu\rho}\right) = 0, \tag{8.135}$$

or, using the duality vector,

$$G_{\mu\nu} - \left(\vec{F}_\mu{}^\rho\right)^{\rm T} \vec{F}_{\nu\rho} = 0, \tag{8.136}$$

which makes it clear that only the O(2) subgroup leaves the Einstein equation invariant. Out of this O(2) group, the parity transformation clearly belongs to a different class (if we had N vector fields, it would belong to the $\mathrm{O}(N)$ group that rotates the vectors amongst themselves). Thus, the classical electric–magnetic-duality group of the Einstein–Maxwell theory is actually SO(2).

We are studying an Abelian theory without matter and therefore it has no coupling constant. However, we could think of this U(1) gauge symmetry as part of a bigger, non-Abelian, broken symmetry group and introduce a (dimensionless in natural units in $d = 4$) coupling constant g that appears as a g^{-2} factor in front of F^2 in the action and that we will not reabsorb into a rescaling of the vector field. The appropriate duality vector the integral of whose dual over S^2_∞ is $16\pi G_N^{(4)}\vec{q}$ is now

$$\vec{F} \equiv \begin{pmatrix} g^{-2}F \\ {}^\star F \end{pmatrix}. \tag{8.137}$$

In terms of this duality vector, the Einstein equation can be rewritten as follows:

$$G_{\mu\nu} + \left(\vec{F}_\mu{}^\rho\right)^{\mathrm{T}} \eta \, {}^\star\vec{F}_{\nu\rho} = 0, \qquad \eta \equiv \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{8.138}$$

and it is invariant under $\mathrm{Sp}(2,\mathbb{R}) \sim \mathrm{SL}(2,\mathbb{R})$. Now, it can be checked that, out of the full group, and allowing for transformations of g, only the following transformations (rescalings and $\mathbb{Z}_2$ duality rotations and their products) are consistent with the duality-vector constraint:

$$\begin{aligned} M &= \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix}, \qquad g' = a^{-1} g, \\ M &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad g' = 1/g. \end{aligned} \tag{8.139}$$

Now we see the main reason why this duality is interesting: if the coupling constant g of the original theory is large so perturbation theory cannot be used and non-perturbative states become light, then the coupling constant of the dual theory $g' = 1/g$ is small and can be used to do perturbative expansions and the dual theory gives a better description of the same phenomena and states. In particular, magnetic monopoles are typical non-perturbative states of gauge theories with masses proportional to $1/g^2$ and become perturbative, electrically charged states of the dual theory.

Although, originally, electric–magnetic duality arose as a symmetry of the theory, a better point of view is that it is a relation, a mapping, between two theories that describe the same degrees of freedom in different ways. One of them can describe better one region of the *moduli space*[24] than can the other. Dualities in which the coupling constant is inverted and perturbative (weak-coupling) and non-perturbative (strong-coupling) regimes are related go by the name of *S dualities*. Electric–magnetic duality in the Maxwell theory is the simplest example. Perturbative dualities such as the $\mathrm{O}(N)$ rotation between the N vector fields that we considered in Section 8.3 go by the name of *T dualities*, at least in the string-theory context. In some string theories (type II) the two kinds of dualities are part of a bigger duality group (which is not just the direct product of the S and T duality groups) which is called the *U duality* group [583].

A last comment on semantics: when talking about duality, there are always certain ambiguities in the use of the word "theory." Two theories that are dual are two different descriptions of the same physical system and many physicists would say that they are, therefore, the same "theory" written in different variables. We would like to call them different "theories" describing the same reality. Both points of view are legitimate and are similar to the active and passive points of view in symmetry transformations.

---

[24] The coupling constant g and other parameters necessary to describe completely a theory are usually called *moduli*. The space in which they take values is the *moduli space* of the theory.

8.7.1 Poincaré duality

One of the peculiarities of the electric–magnetic-duality transformation is that it does not leave the Einstein–Maxwell action invariant: the direct replacement of F by its dual in the action changes the sign of the kinetic terms F^2. The reason is that the action Eq. (8.58) is actually a functional of the vector potential. To be able to replace F by ${}^\star F$ we need an action that is a functional of F. The so-called Poincaré-dualization procedure provides a systematic way of finding actions that are functionals of the field strengths and on which we can perform electric-duality transformations, obtaining the correct dual action. Furthermore, this procedure can be generalized to other k-form potentials and dimensions.

Since the metric does not play a role, we consider only the vector-field kinetic term in Eq. (8.58). From that action one obtains only half of the Maxwell equations: the Bianchi identity has been solved and it is assumed that $F = dA$. Thus, if we want to have a functional of F that produces all the Maxwell equations (sometimes called a first-order action), it has to give also the Bianchi identity $dF = 0$. This action can be constructed simply by adding to the standard Einstein–Maxwell action a Lagrange multiplier term enforcing the Bianchi identity. dF is a 3-form and so the Lagrange multiplier has to be a 1-form $\tilde{A} = \tilde{A}_\mu dx^\mu$ (which will become the dual potential) and then the term to be added to the action is $\sim\int \tilde{A} \wedge dF$. Integrating by parts, this term is rewritten as $\sim\int d\tilde{A} \wedge F$. More explicitly, in component language, the action with the Lagrange-multiplier term is

$$S[F_{\mu\nu}, \tilde{A}_\mu] = \frac{1}{16\pi G_{\rm N}^{(4)}} \int d^4x \sqrt{|g|}\big[-\tfrac{1}{4}F^2\big] - \frac{1}{16\pi G_{\rm N}^{(4)}} \int d^4x\ \tfrac{1}{2}\epsilon^{\mu\nu\rho\sigma}\partial_\mu \tilde{A}_\nu F_{\rho\sigma}. \tag{8.140}$$

This action gives rise to the same equations of motion as the original action $S[A]$: the equation of motion of F is

$$F = {}^\star\tilde{F}, \tag{8.141}$$

where we have defined

$$\tilde{F} = d\tilde{A}. \tag{8.142}$$

The Bianchi identity $d\tilde{F} = 0$, a consequence of its definition, becomes the Maxwell equation $d{}^\star F = 0$ by virtue of the F equation of motion above. Furthermore, by construction, the equation of motion of $\tilde{A}$ is nothing but the Bianchi identity $dF = 0$ that implies the existence of the original vector field A_μ.

Since the equation of motion of F is purely algebraic, we can use it in the above action to eliminate it. The result is an action that is a functional of the dual potential $\tilde{A}$ and is identical to the original Einstein–Maxwell action (with the right sign):

$$S[\tilde{A}_\mu] = \frac{1}{16\pi G_{\rm N}^{(4)}} \int d^4x \sqrt{|g|}\,\big[-\tfrac{1}{4}\tilde{F}^2\big]. \tag{8.143}$$

8.7.2 Magnetic charge: the Dirac monopole and the Dirac quantization condition

The electric–magnetic-duality invariance of the vacuum Maxwell equations is automatically broken when one adds sources j^μ. This is not surprising since j^μ describes static or dynamical electric (only) charges. It is necessary to introduce magnetic sources that can be rotated into the electric ones in order to maintain duality invariance of the Maxwell equations. We have already seen in Eq. (8.133) that electric–magnetic duality needs the introduction of magnetic charges into which electric charges can transform. By definition, then, the magnetic charge is given by[25]:

$$p \equiv \tilde{q} = \int_{S^2_\infty} d^2\vec{S} \cdot \tilde{\vec{E}} = \int_{S^2_\infty} d^2\vec{S} \cdot \vec{B} = -\int_{S^2_\infty} F. \tag{8.144}$$

The simplest electric-charge distribution is a point-like electric charge and its dual is a magnetic point-like charge, which should be given by a magnetic field obeying

$$\vec{\nabla} \cdot \vec{B} = p\, \delta^{(3)}(\vec{x}_3), \tag{8.145}$$

which is the *Dirac monopole equation* for the vector potential.

Introducing magnetic sources to preserve electric–magnetic duality is, however, a very dangerous move: the Bianchi identity is not satisfied at the locations of the magnetic sources and there the vector potential, the true dynamical field, cannot be defined or, more precisely, it cannot be defined everywhere: it will have singularities. This may not be as bad as it looks at first sight, because, after all, the electrostatic potential is not defined at the location of an electric point-like charge, either. It depends on how bad the singularities of the vector field are. In the electric case, it is quite benign, since the singularity affects only the particle that gives rise to the field. Let us see what happens with the vector potential of a point-like magnetic monopole. First, we have to find it.

Knowing that

$$\nabla^2 \frac{1}{|\vec{x}_3|} = -4\pi\delta^{(3)}(\vec{x}_3), \tag{8.146}$$

we find that the magnetic field is given by

$$\vec{B} = -\frac{p}{4\pi}\vec{\nabla}\frac{1}{|\vec{x}_3|}, \tag{8.147}$$

which implies, due to $\vec{B} = \vec{\nabla} \times \vec{A}$, for the Dirac monopole equation

$$\vec{\nabla} \times \vec{A} = -\frac{p}{4\pi}\vec{\nabla}\frac{1}{|\vec{x}_3|}, \tag{8.148}$$

or, defining, to simplify matters $\vec{f} = -(4\pi/p)\vec{A}$, the following, standard form:

$$\boxed{\partial_m f_n - \partial_n f_m = \epsilon_{mnp}\partial_p \frac{1}{|\vec{x}_3|}.} \tag{8.149}$$

[25] We work again in the standard units of the beginning of Section 8.2.1 and in flat spacetime. At the end of this section we will say which changes have to be made when using our normalization Eq. (8.58).

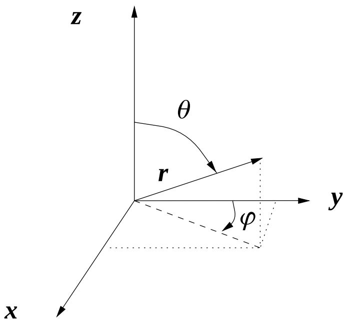

Fig. 8.5. Spherical versus Cartesian coordinates.

The integrability condition for this system of coupled partial differential equations is found by rewriting it in the equivalent form

$$\epsilon_{mnp}\partial_m f_n = \partial_p \frac{1}{|\vec{x}_3|}, \tag{8.150}$$

and acting with ∂_p,

$$\partial_p \partial_p \frac{1}{|\vec{x}_3|} = 0, \tag{8.151}$$

which is true everywhere except at the origin. Instead of $1/|\vec{x}_3|$ we could have used any other harmonic function on the three-dimensional Euclidean space.

A solution of the Dirac monopole equation is provided by (see e.g. [246, 459, 715])

$$\vec{f}^{\,+} = -\frac{(0,0,1) \times (x,y,z)}{|\vec{x}_3|(|\vec{x}_3| + z)}. \tag{8.152}$$

This solution is singular at $|\vec{x}_3| = -z$, i.e. the whole negative z axis, not just at the location of the magnetic monopole.

In spherical coordinates (Figure 8.5)

$$\begin{aligned} x^1 &= x = r\sin\theta\sin\varphi, \\ x^2 &= y = r\sin\theta\cos\varphi, \\ x^3 &= z = r\cos\theta, \end{aligned} \tag{8.153}$$

the above solution has as its only non-vanishing component

$$f^+_\varphi = 1 - \cos\theta. \tag{8.154}$$

In these coordinates the solution looks regular. However, one has to take into account that the unit vector orthogonal to constant φ surfaces is singular over the z axis. Over the positive z axis, $\vec{f}^{\,+}$ is regular because f^+_φ vanishes.

Owing to this singularity, $\vec{f}^{\,+}$ is not a solution of the Dirac monopole equation (8.149) everywhere. This can be seen just as one sees that $\nabla^2|\vec{x}|^{-1} \sim \delta^{(3)}(\vec{x})$ by integrating and applying Stokes' theorem. Let us consider the integral of $\vec{\nabla} \times \vec{f}^{\,+} + \vec{x}/|\vec{x}|^3$ over a surface

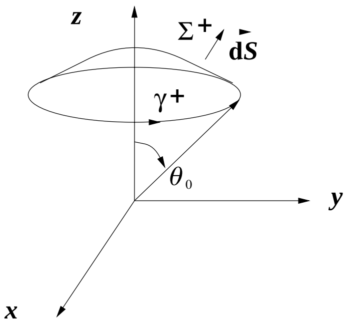

Fig. 8.6. The surface Σ^+ and its boundary γ^+.

Σ. If $\vec{f}^{\,+}$ were a solution of Eq. (8.149) everywhere, this integral would be zero for any surface Σ. Now let us apply Stokes' theorem to the first term. We find

$$\Phi_\Sigma = \int_\Sigma d^2\vec{S} \cdot \left(\vec{\nabla} \times \vec{f}^{\,+} + \frac{\vec{x}}{|\vec{x}|^3}\right) = \oint_{\gamma=\partial\Sigma} d\vec{x} \cdot \vec{f}^{\,+} + \int_\Sigma d^2\vec{S} \bullet \frac{\vec{x}}{|\vec{x}|^3}, \quad (8.155)$$

where γ is the one-dimensional boundary of the two-dimensional surface Σ. Let us consider the particular surface Σ_+ (a sector of the unit sphere whose boundary is $\gamma^+ : \theta = \theta_0$ oriented in the sense of negative φ) shown in Figure 8.6. We find

$$\left.\begin{aligned} \int_{\Sigma^+} d^2\vec{S} \cdot \frac{\vec{x}}{|\vec{x}|^3} &= 2\pi(1 - \cos\theta_0), \\ \int_{\gamma^+} d\vec{x} \cdot \vec{f}^{\,+} &= -2\pi(1 - \cos\theta_0), \end{aligned}\right\} \Rightarrow \Phi_{\Sigma^+} = 0. \quad (8.156)$$

Let us now consider a different surface Σ_- (a sector of the unit sphere whose boundary is $\gamma^- : \theta = \theta_0$ oriented in the sense of positive φ), shown in Figure 8.7:

$$\left.\begin{aligned} \int_{\Sigma^-} d^2\vec{S} \cdot \frac{\vec{x}}{|\vec{x}|^3} &= 2\pi(1 + \cos\theta_0), \\ \int_{\gamma^-} d\vec{x} \cdot \vec{f}^{\,+} &= 2\pi(1 - \cos\theta_0), \end{aligned}\right\} \Rightarrow \Phi_{\Sigma^-} = 4\pi. \quad (8.157)$$

The above results are valid for any value of θ_0 and we conclude that $\vec{f}^{\,+}$ does indeed solve the Dirac-monopole equation only away from the negative z axis $\theta = \pi$. More precisely

$$\vec{\nabla} \times \vec{f}^{\,+} = -\frac{\vec{x}_3}{|\vec{x}_3|^3} - 4\pi\delta(x)\delta(y)\theta(-z)\vec{u}_z, \quad (8.158)$$

where $\vec{u}_z$ is a unit vector along the z axis.

The singularity along $\theta = \pi$ is known as the *Dirac string*. Physically, the Dirac string can be visualized as the zero-section limit of a semi-infinite tube of magnetic flux. Thus,

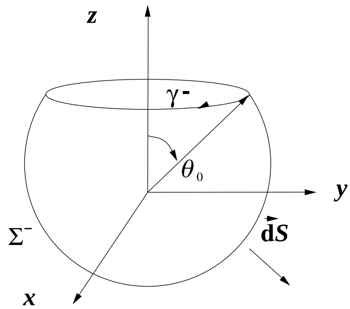

Fig. 8.7. The surface Σ^- and its boundary γ^-.

the flux of

$$\vec{B}^{\,+} = \vec{\nabla} \times \vec{A}^{\,+} = \frac{p}{4\pi}\frac{\vec{x}_3}{|\vec{x}_3|^3} + p\delta(x)\delta(y)\theta(-z)\vec{u}_z \tag{8.159}$$

across any closed 2-surface is zero. The Dirac string appears as a singularity of the magnetic field and hence, in principle, it should be considered a physical singularity.

Thus, at first sight we have not succeeded in finding a solution of the Dirac monopole equation. Still, we can ask ourselves whether the Dirac string is observable and has any physical effect. First of all, observe that the Dirac string can be moved (but not removed) by gauge transformations. For instance, the transformed gauge potential $\vec{A}^{\,-}$,

$$A^-_\varphi = A^+_\varphi + \partial_\varphi\left(\frac{p}{2\pi}\varphi\right) = \frac{p}{4\pi}(1+\cos\theta), \tag{8.160}$$

is now singular over the positive z axis only. We have changed the position of the Dirac string from the negative to the positive axis. From this one could naively conclude that the Dirac string is just a gauge artifact and, as such, unphysical. This is not strictly correct, though. First, $\vec{A}^{\,+}$ and $\vec{A}^{\,-}$ are related by a gauge transformation that is multivalued. Two configurations related by a multivalued gauge transformation are definitely not physically equivalent if the gauge group is $\mathbb{R}$. Second, and more important, no matter what the gauge group is, classically, $\vec{A}^{\,+}$ and $\vec{A}^{\,-}$ can be distinguished by a classical charged particle crossing the string singularity ($\vec{B}^{\,+}$ and $\vec{B}^{\,-}$ are indeed different).

However, quantum-mechanically they *may* be completely equivalent if the gauge group is U(1), provided that the gauge function has the right periodicity. To analyze this problem we have to consider the quantum-mechanical coupling of the U(1) vector field to charged matter. Thus, let us consider the Schrödinger equation for a particle of mass M and electric charge q in an electromagnetic field:

$$H\Psi = i\hbar\frac{\partial}{\partial t}\Psi. \tag{8.161}$$

To obtain the Hamiltonian H we start from the action for a massive relativistic particle in an electromagnetic background field, Eq. (8.54), which in the non-relativistic limit gives Eq. (8.55), from which, after subtracting the zero-point energy Mc^2, we can identify the non-relativistic Lagrangian L and construct the classical Hamiltonian

$$H = \vec{P} \cdot \dot{\vec{X}} - L = \frac{1}{2M}\left[\vec{P} - \frac{q}{c}\vec{A}\right]^2 + q\phi, \qquad \vec{P} \equiv M\dot{\vec{X}} + \frac{q}{c}\vec{A}. \tag{8.162}$$

In the quantization of this system the momentum $\vec{P}$ is replaced by the operator $-i\hbar\vec{\nabla}$. We obtain the Hamiltonian

$$H = -\frac{\hbar^2}{2M}\vec{D}^2 + q\phi, \tag{8.163}$$

where $\vec{D}$ is the *covariant derivative*

$$\vec{D} = \vec{\nabla} - ie\vec{A}, \tag{8.164}$$

and the gauge coupling constant is

$$e = \frac{q}{\hbar c}. \tag{8.165}$$

Under a gauge transformation of the vector field $A'_\mu = A_\mu + \partial_\mu \Lambda$ the Hamiltonian Eq. (8.163) is not invariant, but its transformation can be compensated by the following gauge transformation of the wave function:

$$\Psi' = e^{-ie\Lambda}\Psi. \tag{8.166}$$

So the Schrödinger equation is gauge-covariant (it changes by the above overall phase).

If the gauge group is $\mathbb{R}$ (i.e. $\Lambda \in \mathbb{R}$), Λ has to be single-valued and the same must be true for the wave function. If the gauge group is U(1), though, Λ lives in a circle, or equivalently in a lattice, and we have to identify two different values of Λ differing by the period T,

$$\Lambda \sim \Lambda + T. \tag{8.167}$$

In a topologically trivial spacetime any closed path is contractible to a point. This implies that the wave function has to be single-valued around any closed path. This implies in turn that only gauge transformations such that the gauge phase $e^{-ie\Lambda}$ is single-valued around any closed path are allowed. Since we just admitted that Λ can be multivalued with period T, we conclude that the only Ts allowed are those satisfying

$$Te = 2\pi n, \qquad n \in \mathbb{Z}. \tag{8.168}$$

For application to the Dirac-monopole case in which the space is topologically non-trivial ($\mathbb{R}^3$ minus the positive or negative z axis) but has no non-contractible closed paths, we conclude that the gauge transformation that moves the Dirac string relates two quantum-mechanically equivalent configurations in which the wave function is single-valued if the gauge parameter has the right periodicity. If the two configurations are equivalent in spite of the fact that they have Dirac strings in different places, then the Dirac strings have no physical effect. Going around the z axis once gives

$$\Lambda(\varphi + 2\pi) = \Lambda + p, \tag{8.169}$$

so we find that we can do consistent quantum mechanics ignoring the Dirac string if the magnetic charge is related to the electric charge by the *Dirac quantization condition*[26] [323],

$$\boxed{qp = n2\pi\hbar c.} \tag{8.170}$$

It is worth remarking that this formula is invariant (up to a global sign) under electric–magnetic-duality transformations $q \to p$, $p \to -q$.

Using the normalization of Eq. (8.58), the definitions of electric and magnetic charge that satisfy the Dirac quantization condition in the above form (without any extra factors) are

$$\boxed{q \equiv \frac{1}{16\pi G_{\mathrm{N}}^{(4)}} \int_{\mathrm{S}^2_\infty} {}^\star F, \qquad p \equiv -\int_{\mathrm{S}^2_\infty} F, \qquad pq = 2\pi n.} \tag{8.171}$$

In a non-simply connected spacetime there will be closed paths that are not contractible to a point (that is, it will have a non-trivial π_1). The wave function will not in general be single-valued around those closed paths but will pick up a phase, the *Aharonov–Bohm phase* [20, 21], which can be detected by interference experiments. The Dirac quantization condition can be considered as the condition of cancelation of a would-be Aharonov–Bohm phase around the Dirac string, which physically is unacceptable. The concept of the Aharonov–Bohm phase is, however, much more general and deals with the non-triviality of the topology of the gauge-field itself when it is seen as a section of a fiber bundle. To study the Aharonov–Bohm phase, thus, we first reformulate the Dirac monopole in this language.

8.7.3 The Wu–Yang monopole

Wu and Yang [964] were the first to reformulate the Dirac monopole in the modern language. The basic idea is to generalize the basic concepts of tensors in manifolds to gauge fields:[27] a manifold is a topological space that in general is not isomorphic to $\mathbb{R}^n$. Thus it needs to be covered by patches that are isomorphic to parts of $\mathbb{R}^n$. Each patch provides a local coordinate system. Neighboring patches must overlap and the two different coordinates of points in the overlaps are related by diffeomorphisms. Now one can define tensor fields on a manifold. A given well-defined tensor field will have different components in the overlaps, corresponding to the different coordinate systems that are defined there, but they will be related by the tensor-transformation laws corresponding to the diffeomorphisms that relate the different coordinate systems.

---

[26] There are other ways of finding this condition, such as studying the quantization of the angular momentum of the electromagnetic field created by the electric and magnetic particles. See, for instance, [459].

[27] For a less-pedestrian explanation, there are many reviews and textbooks that the interested reader can consult: for instance [240, 347, 630, 715, 717].

Now, a gauge field defined on a manifold is a 1-form field and its definitions in different patches will be related by the standard transformation rules of 1-forms under diffeomorphisms. The new freedom that we have in fiber bundles is that these different definitions can also be related by gauge transformations. The most basic example is precisely the Dirac monopole. The space manifold is just $\mathbb{R}^3$ and, in principle, we need only one patch to cover it. However, we are going to use two, because the topology of the gauge-field configuration requires it. The two patches will be the two halves of $\mathbb{R}^3$ with $z \geq 0$ and $z \leq 0$, which overlap over the plane $z = 0$. The two coordinate systems that we are going to use are trivially related and we will not distinguish them. The U(1) gauge field in the first patch $z \geq 0$ will be A^+, which is completely regular there (except at the origin) because its Dirac string lies in the second patch. In the second patch, $z \leq 0$, the gauge field will be A^-, which is also regular there for analogous reasons. In the overlap $z = 0$, we have two different values of the gauge field, but they are related (by construction) by the gauge transformation Eq. (8.160). The discussion of which gauge transformations are allowed is still valid here and we arrive at the same Dirac quantization condition.

One of the advantages of this formulation is that, at the expense of introducing non-trivial topology for the gauge field, we have eliminated completely the Dirac string and have a completely regular gauge field (except at the origin). The magnetic field $\vec{B}$ is only singular at the origin, too. A calculation of the magnetic charge through the magnetic flux should now give the right result. First we rewrite the flux in differential-forms language:

$$\int_{S^2} d\vec{S} \cdot \vec{B} = \int_{S^2} d\vec{S} \cdot \vec{\nabla} \times \vec{A} = \int_{S^2} dA. \tag{8.172}$$

We cannot use Stokes' theorem here because A is multivalued. We divide the 2-sphere into two halves $\Sigma^\pm$ overlapping at the equator $z = 0$. In each of these two halves, A is single-valued and Stokes' theorem can be applied:

$$\int_{S^2} d\vec{S} \cdot \vec{B} = \int_{S^2} dA = \int_{\Sigma^+} dA^+ + \int_{\Sigma^-} dA^- = \int_{\gamma^+} A^+ + \int_{\gamma^-} A^-, \tag{8.173}$$

where $\gamma^\pm$ are the boundaries of $\Sigma^\pm$: equatorial circumferences are oriented in the negative and positive φ directions, so $\gamma^+ = -\gamma^-$. Using the relation between A^+ and A^-, we find

$$\int_{S^2} d\vec{S} \cdot \vec{B} = -\int_{\gamma^+} \partial_\varphi \left(\frac{p}{2\pi} \varphi \right) = p. \tag{8.174}$$

The magnetic charge is given by the non-trivial monodromy of the gauge parameter.

The topology of gauge fields (fiber bundles) such as the monopole can be characterized by the values of topological invariants. In the case of the Abelian monopole it is the *first Chern class*,

$$c_1 = -\frac{1}{2\pi} \int_{S^2} F, \tag{8.175}$$

which is nothing but the magnetic charge $p/(2\pi)$ and should be an integer n, according to general arguments. This result is stated in units in which $q = \hbar = c = 1$ and then we see that this is nothing but the Dirac quantization condition.

8.7.4 Dyons and the DSZ charge-quantization condition

If objects with electric charge and objects with magnetic charge exist, then objects with both kinds of charges, called *dyons*, may exist. The electromagnetic field they produce is just a linear superposition of those produced by electric and magnetic monopoles.

Considering the quantum evolution of one dyon in the field of another dyon, it is found that consistency requires the four charges of these objects to obey the *Dirac–Schwinger–Zwanziger (DSZ) quantization condition* [825, 826, 970, 971]

$$\boxed{q_1 p_2 - q_2 p_1 = n 2\pi \hbar c.} \tag{8.176}$$

With the normalization and units of Eqs. (8.58) and (8.171) the condition takes the same form but with no $\hbar c$ constants.

Now, this condition is completely invariant under $\mathbb{Z}_2$ electric–magnetic-duality transformations. This can be more easily seen if we rewrite it in this very suggestive form using the charge vectors we introduced before:

$$\vec{q}_1{}^{\mathrm{T}} \eta \, \vec{q}_2 = \frac{n}{8G_{\mathrm{N}}^{(4)}}, \qquad \vec{q} = \begin{pmatrix} q \\ p/(16\pi G_{\mathrm{N}}^{(4)}) \end{pmatrix}. \tag{8.177}$$

We saw that the Einstein equation could also be written using duality vectors and the matrix $\eta = i\sigma^2$ (Eq. (8.138)). The presence of that matrix implied that the duality group was a subgroup of $\mathrm{SL}(2,\mathbb{R}) \sim \mathrm{Sp}(2,\mathbb{R})$. Now we obtain the same result from the DSZ quantization condition. This condition does not take into account all the quantum effects, such as the quantization of electric charge (independently of any magnetic-monopole charge). These effects will break the classical duality group to some discrete subgroup, but will not change the DSZ quantization condition.

Inclusion of a theta angle and the Witten effect. The Einstein–Maxwell action can be modified by the addition of a *topological term* of the form

$$-\frac{\theta}{8\pi c} \int d^4x \sqrt{|g|}\, F\, {}^\star F, \tag{8.178}$$

or, in differential-forms language,

$$-\frac{\theta}{4\pi c} \int F \wedge F, \tag{8.179}$$

where we see that the metric does not appear in it, which is the reason why it is called topological.

On the other hand, using the Bianchi identity, the integrand of this term can be shown to be the total derivative of the *Chern–Simons 3-form* $F \wedge A$

$$F \wedge F = d(A \wedge F), \tag{8.180}$$

and therefore it does not contribute to the classical equations of motion. However, a change in the Lagrangian produces a change in the Noether current and in the definition of the corresponding conserved charge. To make this more concrete, let us consider the Einstein–Maxwell Lagrangian Eq. (8.58) with θ-term and with coupling constant g with our conventions and units:

$$S[g_{\mu\nu}, A_\mu] = \frac{1}{16\pi G_N^{(4)}} \int d^4x \sqrt{|g|} \left[R - \frac{1}{4g^2} F^2 - \frac{\theta}{8\pi} F\, {}^\star F \right]. \tag{8.181}$$

The Noether current associated with the gauge transformations of the vector field is now

$$j_N^\mu = \frac{1}{16\pi G_N^{(4)}} \nabla_\nu \left[\frac{1}{g^2} F^{\nu\mu} + \frac{\theta}{2\pi} {}^\star F^{\nu\mu} \right]. \tag{8.182}$$

In this simple Abelian case that we are considering the second term vanishes by virtue of the Bianchi identity. Still we will keep it and, after using Stokes' theorem in the definition of electric charge, the second term gives a net contribution

$$q \equiv \frac{1}{16\pi G_N^{(4)}} \int_{S^2_\infty} \left[\frac{1}{g^2} {}^\star F - \frac{\theta}{2\pi} F \right]. \tag{8.183}$$

The magnetic charge is still given by Eq. (8.171).

If we start with a magnetic monopole in a vacuum with $\theta = 0$ and then "switch on" θ, we see in the above formulae that the magnetic monopole acquires an electric charge proportional to θ and becomes a dyon. This is the *Witten effect* [956].

We studied the classical duality group when we introduced the coupling constant g and allowed it to transform under it. It is interesting to see what happens after we introduce θ and we allow it to transform as well. This can be seen more easily if we redefine the duality vector

$$\vec{F} \equiv \begin{pmatrix} \frac{1}{g^2} F + \frac{\theta}{2\pi} {}^\star F \\ {}^\star F \end{pmatrix}, \tag{8.184}$$

whose two components F^1 and F^2 are subject to the constraint

$$F^1 = \frac{\theta}{2\pi} F^2 - \frac{1}{g^2} {}^\star F^2, \tag{8.185}$$

and define the complexified coupling constant

$$\tau = \frac{\theta}{2\pi} + \frac{i}{g^2}. \tag{8.186}$$

In terms of the new duality vector, the Einstein equation still has the form (8.138) and we see that any $SL(2, \mathbb{R})$ transformation leaves it invariant. Furthermore, we can see that the $SL(2, \mathbb{R})$-transformed duality vector has the same form as the original one but with τ transformed as follows: if the $SL(2, \mathbb{R})$ transformation is

$$\tilde{\vec{F}} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \vec{F}, \qquad \alpha\delta - \beta\gamma = 1, \tag{8.187}$$

then the complexified coupling constant transforms simultaneously as follows:

$$\tilde{\tau} = \frac{\alpha\tau + \beta}{\gamma\tau + \delta}. \tag{8.188}$$

So, the full set of equations of motion is invariant under $\mathrm{SL}(2,\mathbb{R})$-duality transformations.

Furthermore, in the presence of a θ-term, the electric and magnetic charges naturally fit into a duality vector, which is the integral of the Hodge dual of the duality vector of the 2-form field strengths defined above, with a $1/(16\pi G_{\mathrm{N}}^{(4)})$ normalization factor. The DSZ quantization condition still takes the form Eq. (8.177) (the θ-dependent terms drop out from it) and we see that it is fully form-invariant under $\mathrm{SL}(2,\mathbb{R})$ transformations. If the electric charge were quantized $q \in \mathbb{Z}$, it is clear that only the discrete subgroup $\mathrm{SL}(2,\mathbb{Z})$ would preserve its quantization and the spectrum of charged particles (generically dyons characterized by their charge vectors). This is the general S duality group and will appear in different forms in many places in what follows.

8.7.5 *Duality in massive electrodynamics*

To acquire some training in the use of the Poincaré-duality procedure explained in Section 8.7.1 in more general settings than that of Maxwell's theory, it is interesting to consider the dualization of the Proca Lagrangian rewritten using the Stückelberg scalar in Eq. (8.67), which seems to have no electric–magnetic-duality *symmetry*. We first rewrite it in this form:

$$S[A_\mu, \phi] = \int d^4x \sqrt{|g|}\left[\tfrac{1}{2}G^2 - \tfrac{1}{4}F^2\right], \tag{8.189}$$

where G and F are the scalar and vector gauge-invariant field strengths

$$G_\mu = \partial_\mu \phi + m A_\mu, \qquad F_{\mu\nu} = 2\partial_{[\mu}A_{\nu]}. \tag{8.190}$$

To the equations of motion of this system one can now add a Bianchi identity for G_μ:

$$\partial_{[\mu}\big(G_{\nu]} - mA_{\nu]}\big) = 0. \tag{8.191}$$

However, there is no duality *symmetry* because the dual of a 1-form field strength is a 3-form field strength. Nevertheless, we can perform a duality *transformation* to an equivalent system with a 3-form field strength. In other words, we can apply the Poincaré-duality procedure to the scalar (only its derivatives appear in the action), replacing it by a 2-form potential. Following the general dualization procedure, we want to find an equivalent action that is a functional of the field strength G_μ instead of the scalar ϕ. Thus, we add to the above action a Lagrange-multiplier term enforcing the Bianchi identity for G,

$$\tfrac{1}{2}\int dx^4 \epsilon^{\mu\nu\rho\sigma}\partial_\mu B_{\nu\rho}(G_\sigma - mA_\sigma), \tag{8.192}$$

where we have already integrated by parts. The new action is a functional of G_μ, A_μ, and $B_{\mu\nu}$. The equation of motion for G_μ is

$$G = {}^\star H, \qquad H_{\mu\nu\rho} = 3\partial_{[\mu}B_{\nu\rho]}, \tag{8.193}$$

where H is the field strength of the 2-form B, which is invariant under the gauge transformations $\delta B_{\mu\nu} = \partial_{[\mu}\Lambda_{\nu]}$. On substituting this back into the action and integrating again by parts, we obtain an action that is a functional of the fields A_ν and $B_{\mu\nu}$:

$$S[A_\mu, B_{\mu\nu}] = \int dx^4 \sqrt{|g|}\left[\frac{1}{2\cdot 3!}H^2 - \tfrac{1}{4}F^2 + \frac{m}{4}\frac{\epsilon}{\sqrt{|g|}}FB\right]. \tag{8.194}$$

We have completely dualized ϕ into $B_{\mu\nu}$. Now, A_μ does not occur explicitly any longer in this action, but only through its field strength, and thus we can now Poincaré-dualize with respect to it. By adding a term

$$\int d^4x\, \tfrac{1}{2}\epsilon\, \partial\tilde{A}F, \tag{8.195}$$

and eliminating F through its equation of motion

$$F = {}^\star\tilde{F}, \qquad \tilde{F} = 2\Big(\partial\tilde{A} + \frac{m}{2}B\Big), \tag{8.196}$$

we obtain the action dual to the original:

$$S[\tilde{A}_\mu, B_{\mu\nu}] = \int d^4x \sqrt{|g|}\left[\frac{1}{2\cdot 3!}H^2 - \tfrac{1}{4}\tilde{F}^2\right]. \tag{8.197}$$

The dual vector-field strength is now invariant under *dual massive gauge transformations*

$$\delta B_{\mu\nu} = \partial_{[\mu}\Lambda_{\nu]}, \qquad \delta\tilde{A}_\mu = -\frac{m}{2}\Lambda_\mu, \tag{8.198}$$

which allow us to eliminate $\tilde{A}$ completely, leaving us with a massive 2-form. $\tilde{A}$ now plays the role of the Stückelberg field for B.

The relation between the dual and original variables is

$$\begin{aligned} H &= -{}^\star G, \\ \tilde{F} &= -\left({}^\star F + mB\right). \end{aligned} \tag{8.199}$$

8.8 Magnetic and dyonic RN black holes

We have seen that the full set of equations of motion of the Einstein–Maxwell system without a θ-term and without the introduction of any coupling constant is invariant under the SO(2) group of electric–magnetic duality,

$$\tilde{F} = \cos(\xi)F + \sin(\xi)\,{}^\star F, \qquad {}^\star\tilde{F} = -\sin(\xi)F + \cos(\xi)\,{}^\star F. \tag{8.200}$$

Duality symmetries can be used as solution-generating transformations. For instance, we generated new solutions for a theory with N vector fields from the 1-vector RN solution using the O(N) duality symmetry that rotates the vector fields. We can now do the same and generate new solutions with both electric and magnetic charges out of the purely electric

RN or MP solutions. Let us take the single electric RN BH solution as given in Eq. (8.75). Trivially we obtain a new solution with the same metric and with

$$\tilde{F}_{tr} = \frac{4G_{\rm N}^{(4)}\cos(\xi)q}{r^2}, \qquad \tilde{F}_{\theta\varphi} = 4G_{\rm N}^{(4)}\sin(\xi)q\ \sin\theta. \tag{8.201}$$

After the new solution has been found, it has to be expressed in terms of the new physical parameters $\tilde{q}$ and $\tilde{p}$, which turn out to be related to the old ones by

$$\tilde{q} = \cos(\xi)q, \quad \tilde{p} = -16\pi G_{\rm N}^{(4)}\sin(\xi)q, \quad \Rightarrow |\tilde{\vec{q}}|^2 = \tilde{q}^2 + \tilde{p}/\big(16\pi G_{\rm N}^{(4)}\big) = q^2. \tag{8.202}$$

The last equation is due to the fact that SO(2) leaves the norm of the charge vector $\vec{q}$ invariant. In the metric, we need only replace q^2 everywhere by $|\tilde{\vec{q}}|^2$. In the vector-field strength the other two equations have to be used to replace q and ξ by $\tilde{q}$ and $\tilde{p}$. The result is (now suppressing tildes)

$$\boxed{\begin{aligned} ds^2 &= f(r)dt^2 - f^{-1}(r)dr^2 - r^2 d\Omega^2_{(2)},\\ F_{tr} &= \frac{4G_{\rm N}^{(4)}q}{r^2}, \qquad F_{\theta\varphi} = -\frac{1}{4\pi}p\sin\theta,\\ f(r) &= \frac{(r-r_+)(r-r_-)}{r^2},\\ r_\pm &= G_{\rm N}^{(4)}M \pm r_0, \qquad r_0 = G_{\rm N}^{(4)}\sqrt{M^2 - 4|\vec{q}|^2}. \end{aligned}} \tag{8.203}$$

This is a dyonic RN black hole. The metric is essentially the same as that of the purely electric one with the replacement $q^2 \to |\vec{q}|^2$ and most of its properties are also essentially identical.

Starting with the MP solutions, we find the dyonic MP solutions

$$\boxed{\begin{aligned} ds^2 &= H^{-2}dt^2 - H^2 d\vec{x}_3^{\,2},\\ F_{t\underline{i}} &= -2\cos\alpha\ \partial_{\underline{i}}H^{-1}, \qquad F_{\underline{i}\underline{j}} = 2\sin\alpha\ \epsilon_{ijk}\partial_{\underline{k}}H,\\ \partial_{\underline{i}}\partial_{\underline{i}}H &= 0. \end{aligned}} \tag{8.204}$$

From the point of view of finding new solutions, the important lesson to be learned is that we have generated a new solution with one more physical parameter (the magnetic charge) using a one-parameter solution-generating transformation group. Observe that, in the MP case, the family of solutions depends on only one arbitrary real harmonic function.

We could view these solutions and, in particular, the dyonic RN solution, as solutions of the more general theory with $g = 1$ and $\theta = 0$ and we can try to generate solutions of the more general theory using general SL(2, $\mathbb{R}$) transformations. These have three independent parameters, but we have already used the one corresponding to SO(2). The other two parameters would precisely generate non-trivial values of g and θ. Let us obtain these RN solutions.

First we use on the dyonic RN solution Eq. (8.203) SL(2, $\mathbb{R}$) rescalings, corresponding to matrices of the form

$$\begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix}. \tag{8.205}$$

They generate a nontrivial $\tilde{g} = a$ and rescale the electric and magnetic charges and field strength. The transformed solution, written in terms of the transformed parameters q, p, and g (without the tildes), takes the form

$$\begin{aligned} ds^2 &= f(r)dt^2 - f^{-1}(r)dr^2 - r^2 d\Omega^2_{(2)}, \\ F_{tr} &= \frac{4G_N^{(4)} g q}{r^2}, \qquad F_{\theta\varphi} = -\frac{1}{4\pi g} p \sin\theta, \\ f(r) &= \frac{(r - r_+)(r - r_-)}{r^2}, \\ r_\pm &= G_N^{(4)} M \pm r_0, \qquad r_0 = G_N^{(4)} \sqrt{M^2 - 4\vec{q}^{\,\mathrm{T}} \mathcal{M}^{-1} \vec{q}}, \end{aligned} \tag{8.206}$$

where $\mathcal{M}$ is the matrix

$$\mathcal{M} = \begin{pmatrix} 1/g^2 & 0 \\ 0 & g^2 \end{pmatrix}, \qquad \mathcal{M}^{-1} = \begin{pmatrix} g^2 & 0 \\ 0 & 1/g^2 \end{pmatrix}. \tag{8.207}$$

Now we use the SL(2, $\mathbb{R}$) transformations that shift the θ-parameter from its zero value, corresponding to matrices of the form

$$\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}. \tag{8.208}$$

They generate a non-trivial $\theta/(2\pi) = b$ and mix different components of the field strength and the electric and magnetic charges (the Witten effect). The transformed solution, written in terms of the transformed parameters q, p, g and θ (without the tildes), takes the form[28]

$$\begin{aligned} ds^2 &= f(r)dt^2 - f^{-1}(r)dr^2 - r^2 d\Omega^2_{(2)}, \\ \vec{F}_{tr} &= \frac{4G_N^{(4)} \vec{q}}{g r^2}, \\ f(r) &= \frac{(r - r_+)(r - r_-)}{r^2}, \\ r_\pm &= G_N^{(4)} M \pm r_0, \qquad r_0 = G_N^{(4)} \sqrt{M^2 - 4\vec{q}^{\,\mathrm{T}} \mathcal{M}^{-1} \vec{q}}, \end{aligned} \tag{8.209}$$

where $\mathcal{M}$ is now the matrix

$$\mathcal{M} = g^2 \begin{pmatrix} |\tau|^2 & \theta/(2\pi) \\ \theta/(2\pi) & 1 \end{pmatrix}, \qquad \mathcal{M}^{-1} = g^2 \begin{pmatrix} 1 & -\theta/(2\pi) \\ -\theta/(2\pi) & |\tau|^2 \end{pmatrix}, \tag{8.210}$$

[28] Giving the tr components of the two components of the duality vector is equivalent to, but much simpler than, giving the tr and $\theta\varphi$ components of F.

which has interesting properties: it belongs to SL(2, $\mathbb{R}$) but it is symmetric. It can be seen that it parametrizes the SL(2, $\mathbb{R}$)/SO(2) cosets. Furthermore, under an SL(2, $\mathbb{R}$) transformation Λ, it transforms (due to the transformation of g and θ) according to

$$\mathcal{M} = \Lambda\mathcal{M}\Lambda^{\mathrm{T}}, \tag{8.211}$$

so $\vec{q}^{\,\mathrm{T}}\mathcal{M}^{-1}\vec{q}$ is form-invariant under SL(2, $\mathbb{R}$) transformations. Thus, using SL(2, $\mathbb{R}$) duality, we cannot generate any new solutions not yet contained in the above family.

The result is that we have generated a family of solutions that contains three parameters more than the initial one by using a three-dimensional duality group. The solutions are expressed in the simplest form when one uses objects that have good transformation properties under the duality group: duality vectors and matrices. On the other hand, the family covers the most general BH-type solution of the Einstein–Maxwell theory that one can have according to the no-hair conjecture: the BH solution depends on only two conserved charges (electric and magnetic) and two moduli parameters, which are not really characteristic of the BH but rather of the vacuum of the theory.

This example may look quite simple, but it has the same features as some more complicated and juicy cases.

To end this section, let us comment on a couple of subtle points.

- Electric–magnetic-duality rotations and the Wick rotation do not commute. Although we did not stress it, the Euclidean electric RN solution has a purely imaginary electromagnetic field. Electric–magnetic-duality rotations of the Euclidean purely electric RN solution generate a Euclidean solution with imaginary magnetic charge that remains imaginary when we Wick-rotate back to the Lorentzian signature. If we Wick-rotate the dyonic RN solution, we obtain a Euclidean solution with real magnetic charge. This gives rise to problems in the calculation of the entropy in the Euclidean-path-integral formalism,[29] but they can be dealt with, as shown in [196, 302, 307, 520].
- In the extreme magnetic RN BH case, we could also try to look for a source. However, the only thing that works is to view the magnetic charge as the electric charge of the dual vector field.

8.9 Higher-dimensional RN solutions

Just as there are higher-dimensional analogs of the Schwarzschild BH, there are also higher-dimensional analogs of the electric RN BH, which are solutions of the equations of motion that one obtains from considering the Einstein–Maxwell action in d dimensions.

Let us first consider the higher-dimensional generalization of the Einstein-scalar system that we considered at the beginning of this chapter:

$$\boxed{S[g_{\mu\nu}, \varphi] = \frac{c^3}{16\pi G_{\mathrm{N}}^{(d)}}\int d^dx\sqrt{|g|}\,\left[R + 2\partial_\mu\varphi\partial^\mu\varphi\right].} \tag{8.212}$$

[29] The thermodynamical quantities that one derives from the Lorentzian metric of the dyonic RN solution are clearly S-duality-invariant.

It is natural to ask whether the no-hair conjecture that says that there are no regular BH-type solutions of that system in four dimensions holds also in more than four dimensions. Thus, we can try to find static, spherically symmetric BH solutions of this system. For the metric, we will use the straightforward generalization of the *Ansatz* of the "dressed Schwarzschild metric" form Eq. (8.91) that we found for the four-dimensional RN solution (and that is also valid for the four-dimensional solutions of this system, Eqs. (8.7)):

$$ds^2 = H^{2x} W dt^2 - H^{-2y}\left[W^{-1}dr^2 + r^2 d\Omega^2_{(d-2)}\right], \tag{8.213}$$

where W will be a function of the form

$$W = 1 + \frac{\omega}{r^{d-3}}, \tag{8.214}$$

and where H is related to the scalar by

$$\varphi = \varphi_0 + z \ln H, \tag{8.215}$$

z being a constant, so, when the scalar becomes constant, the above metric is just the higher-dimensional Schwarzschild metric. It is easy to see that we are forced to set $H = W$ and $y \neq 0$ in order to have a solution. This implies that the would-be "horizon" is always singular, except when the scalar is constant. For the sake of completeness we give below the form of these solutions, which generalize those obtained in [18, 607],

$$\boxed{\begin{aligned} ds^2 &= W^{\frac{M}{\omega}-1} W dt^2 - W^{\frac{1}{d-3}\left(1-\frac{M}{\omega}\right)}\left[W^{-1}dr^2 + r^2 d\Omega^2_{(d-2)}\right], \\ \varphi &= \varphi_0 \pm \frac{\Sigma}{\omega} \ln W, \\ W &= 1 + \frac{\omega}{r^{d-3}}, \qquad \omega = \pm 2\sqrt{M^2 + 2\left(\frac{d-3}{d-2}\right)\Sigma^2}. \end{aligned}} \tag{8.216}$$

For $\Sigma = 0$ we recover the d-dimensional Schwarzschild solution. In all other cases we have metrics with naked singularities at $r = 0$ or at $r^{d-3} = -\omega$ (if possible).

Now, let us return to the higher-dimensional Einstein–Maxwell system, normalized as in Eq. (8.58) ($c = 1$),

$$\boxed{S[g_{\mu\nu}, A_\mu] = \frac{1}{16\pi G_N^{(d)}} \int d^d x \sqrt{|g|}\,\left[R - \tfrac{1}{4}F^2\right].} \tag{8.217}$$

The Einstein and Maxwell equations are

$$G_{\mu\nu} - \tfrac{1}{2}T_{\mu\nu} = 0, \qquad \nabla_\mu F^{\mu\nu} = 0, \tag{8.218}$$

where the electromagnetic energy–momentum tensor $T_{\mu\nu}$ is again given by Eq. (8.24).

There are a few differences from the four-dimensional case. First we observe that, in more than four dimensions, the energy–momentum tensor of the Maxwell field is no longer traceless because the Maxwell action is not invariant under Weyl rescalings of the metric. This implies that, in general, the curvature scalar is not zero on solutions, but, instead

$$R = \frac{d-4}{4(d-2)} F^2, \tag{8.219}$$

and, thus, on subtracting the trace in the Einstein equation we are now left with the equation

$$R_{\mu\nu} - \tfrac{1}{2}\left[F_\mu{}^\rho F_{\nu\rho} - \frac{1}{2(d-2)} g_{\mu\nu} F^2\right] = 0 \tag{8.220}$$

plus the Maxwell equation to solve.

The second difference is the definition of the electric charge (we treated the definition of the mass in higher-dimensional spaces in the Schwarzschild case). If we follow exactly the same steps as in the four-dimensional case, we arrive at

$$\boxed{q = (-1)^d \frac{1}{16\pi G_{\rm N}^{(d)}} \int_{{\rm S}^{(d-2)}_\infty} \star F,} \tag{8.221}$$

where $\star F$ is now a $(d-2)$-form and ${\rm S}^{(d-2)}_\infty$ is a $(d-2)$-sphere at spatial infinity (constant t, $r \to \infty$). This means that, if there is a charge q at the origin in an asymptotically flat spacetime, the asymptotic behavior of F and the vector A is

$$F_{tr} \sim \frac{16\pi G_{\rm N}^{(d)} q}{\omega_{(d-2)}} \frac{1}{r^{d-2}}, \qquad A_\mu \sim -\delta_{\mu t} \frac{16\pi G_{\rm N}^{(d)} q}{(d-3)\omega_{(d-2)}} \frac{1}{r^{d-3}}, \tag{8.222}$$

where $\omega_{(d-2)}$ is the volume of the unit $(d-2)$-sphere (see Appendix C). In $d \neq 4$ one can perform an electric–magnetic-duality transformation, replacing F by its Hodge dual $\tilde{F} = {}^\star F$, which is a $(d-2)$-form field strength for a $(d-3)$-form potential $\tilde{F} = d\tilde{A}$. This transformation is not a symmetry. Now, we can define the electric charge associated with the dual $(d-3)$-form potential, which is what we would define as magnetic charge, by analogy with the four-dimensional case. However, the carrier of the electric charge of the dual $(d-3)$-form potential cannot be a point-like particle, but has to be a $(d-4)$-dimensional extended object (*brane*). Thus, a standard BH of the kind we are interested in now cannot carry that kind of charge and we will not consider it here, although we will in Part III.

Our immediate goal is, then, to find d-dimensional analogs of the RN BH. Again, we use an *Ansatz* of the "dressed Schwarzschild metric" form:

$$\begin{aligned} ds^2 &= H^{2a} W dt^2 - H^{-2b}\left[W^{-1} dr^2 + r^2 d\Omega^2_{(d-2)}\right], \\ A_\mu &= \alpha \delta_{\mu t}(H^{-1} - 1), \qquad H = 1 + \tfrac{h}{r^{d-3}}, \end{aligned} \tag{8.223}$$

where a, b, h, and α are constants to be found. The electric charge is proportional to h and, thus, we expect that, when it vanishes, h becomes zero ($H = 1$) and we recover the higher-dimensional Schwarzschild metric Eq. (7.77), so we can guess that

$$W = 1 + \frac{\omega}{r^{d-3}}. \tag{8.224}$$

On the other hand, with this *Ansatz*, the metric will have two horizons at $r = -h, -\omega$ when both h and ω are non-vanishing. When ω vanishes ($W = 1$) there is only one horizon and this should correspond to the extreme limit. In this case, the above metric becomes isotropic and we should be able to find whether H becomes a harmonic function and multi-BH solutions exist.

On substituting into the equations of motion, we find the d-dimensional RN solutions

$$\boxed{\begin{aligned} &ds^2 = H^{-2} W dt^2 - H^{\frac{2}{d-3}}\Big[W^{-1}dr^2 + r^2 d\Omega^2_{(d-2)}\Big],\\ &A_\mu = \delta_{\mu t}\alpha(H^{-1} - 1),\\ &H = 1 + \tfrac{h}{r^{d-3}}, \qquad W = 1 + \tfrac{\omega}{r^{d-3}},\\ &\omega = h\Big[1 - \tfrac{d-3}{2(d-2)}\alpha^2\Big]. \end{aligned}} \tag{8.225}$$

By examining the asymptotic behavior of the metric and vector field, we can relate the integration constants h, ω and α to the mass M and the electric charge q as follows:

$$\alpha = \frac{16\pi G_{\rm N}^{(d)}}{(d-3)\omega_{(d-2)}}\frac{q}{r_\pm^{d-3}}, \qquad h = r_\pm^{d-3}, \qquad \omega = \pm r_0^{d-3}, \tag{8.226}$$

where now

$$r_\pm^{d-3} = \frac{8\pi G_{\rm N}^{(d)}}{(d-2)\omega_{(d-2)}}M \pm r_0^{d-3}, \qquad r_0^{d-3} = \frac{8\pi G_{\rm N}^{(d)}}{(d-2)\omega_{(d-2)}}\sqrt{M^2 - \frac{2(d-2)}{d-3}q^2}. \tag{8.227}$$

If we take the lower signs, we obtain a BH solution very similar to the four-dimensional RN solution: if r_0^{d-3} is real and finite (we take it positive) and M positive, there is an event horizon at $r = r_0$ and a Cauchy horizon at $r = 0$. The reality of r_0^{d-3} implies a lower bound for the mass,

$$M \geq \sqrt{\frac{2(d-2)}{d-3}}|q|. \tag{8.228}$$

When this bound is saturated $r_0 = 0$ ($\omega = 0$) and there is only one horizon, which is regular, and we are in the extreme limit. Furthermore, if we set $W = 1$ in the *Ansatz*, the equations of motion are solved by any arbitrary harmonic function in $(d-1)$-dimensional Euclidean space H:

$$\boxed{\begin{aligned} &ds^2 = H^{-2}dt^2 - H^{\frac{2}{d-3}}d\vec{x}^{\,2}_{d-1},\\ &A_\mu = \delta_{\mu t}\alpha(H^{-1} - 1), \qquad \alpha = \pm 2,\\ &\partial_{\underline{i}}\partial_{\underline{i}}H = 0. \end{aligned}} \tag{8.229}$$

These solutions are the generalization of the MP solutions [712]. The d-dimensional ERN solution is a particular case of the d-dimensional MP family, which, evidently, contains also multi-BH solutions. If we take the near-horizon limit of the d-dimensional ERN solution, we find, after the coordinate change $r^{d-3} = (d-3)h\rho,\ \ t \to t/h^{\frac{1}{d-3}}$, a generalization of the RB solution,

$$\boxed{\begin{aligned} ds^2 &= \left(\frac{\rho}{R}\right)^2 dt^2 - \left(\frac{R}{\rho}\right)^2 d\rho^2 - h^{\frac{2}{d-3}} d\Omega^2_{(d-2)}, \\ A_\mu &= \delta_{\mu t}\alpha\rho, \end{aligned}} \tag{8.230}$$

where

$$R = \frac{h^{\frac{1}{d-3}}}{d-3}. \tag{8.231}$$

The metric is that of the direct product $AdS_2 \times S^{d-2}$.

9
The Taub–NUT solution

The asymptotically flat, static, spherically symmetric Schwarzschild and RN BH solutions that we have studied in the two previous chapters were the only solutions of the Einstein and Einstein–Maxwell equations with those properties. To find more solutions, we have to relax these conditions or couple to gravity more general types of matter, as we will do later on. If we stay with the Einstein(–Maxwell) theory, one possibility is to look for static, axially symmetric solutions and another possibility is to relax the condition of staticity and only ask that the solution be stationary, which implies that we have to relax the condition of spherical symmetry as well and look for stationary, axisymmetric spacetimes. In the first case one finds solutions like those in Weyl's family [949, 950] which can be interpreted as describing the gravitational fields of axisymmetric sources with arbitrary multipole momenta[1] or Melvin's solution [692] (which has cylindrical symmetry and was constructed earlier by Bonnor [165] via a Harrison transformation [499] of the vacuum), among many others. In the second case, we find the Kerr–Newman BHs [617, 723] with angular momentum and electric or magnetic charge and also the Taub–Newman–Unti–Tambourino (Taub–NUT) solution [724, 879], which may but need not include charges. The Taub–NUT metric does not describe a BH because it is not asymptotically flat. In fact, the only stationary axially symmetric BHs of the Einstein–Maxwell theory belong to the Kerr–Newman family of solutions (see e.g. [532, 533]).

The Taub–NUT solution has a number of features that are particularly interesting for us, which we are going to discuss in this chapter. In particular, it carries a new type of charge (*NUT charge*), which is of topological nature and can be viewed as "gravitational magnetic charge," so the solution is a sort of *gravitational dyon* and its Euclidean continuation (for certain values of the mass and NUT charge) is the solution known in other contexts as a *Kaluza–Klein monopole*. This is a very important solution with interesting properties such as the self-duality of its curvature and its relation to the Belavin–Polyakov–Schwarz–Tyupkin (BPST) SU(2) instanton and the 't Hooft–Polyakov monopole. In Chapter 11 we will study how it arises in KK theory. Here we will describe it as a self-dual gravitational instanton and we will take the opportunity to mention other gravitational instantons.

[1] For a review see [793].

The charged Taub–NUT solutions will help us to introduce a very large and interesting family of solutions; the Israel–Wilson–Perjés (IWP) solutions, which have very important properties from the point of view of supersymmetry and duality.

9.1 The Taub–NUT solution

General stationary, axially symmetric metrics have only two Killing vectors, $k = \partial_t$ and $m = \partial_\varphi$, that generate time translations and rotations around the symmetry axis (z). These two Killing vectors are not mutually orthogonal, which implies that the off-diagonal component of the metric $g_{t\varphi} = k^\mu m_\mu$ does not vanish (otherwise we would have a static spacetime). Furthermore, the components of the metric can depend on the other coordinates, which we call r and θ in $d = 4$. A general *Ansatz* for these spacetimes has the form

$$ds^2 = g_{tt}dt^2 + 2g_{t\varphi}dtd\varphi + g_{rr}dr^2 + g_{\theta\theta}d\theta^2 + g_{\varphi\varphi}d\varphi^2, \tag{9.1}$$

where all the components may depend on r and θ. The new interesting ingredient is the component $g_{t\varphi}(r,\theta)$. If the metric is asymptotically flat for $r \to \infty$ and $g_{t\varphi}(r,\theta)$ has the asymptotic behavior

$$g_{t\varphi} \sim 2J\frac{\sin^2\theta}{r}, \tag{9.2}$$

then the solution describes a spacetime with angular momentum J in the direction of the z axis. The only vacuum solution of this kind is Kerr's [617], which in Boyer–Lindquist coordinates takes the form

$$\boxed{\begin{aligned} ds^2 &= \left(1 - \frac{2Mr}{\Sigma}\right)dt^2 + 2\frac{2aMr\sin^2\theta}{\Sigma}dtd\varphi - \frac{\Sigma}{\Delta}dr^2 - \Sigma d\theta^2 - \frac{\mathcal{A}}{\Sigma}\sin^2\theta\, d\varphi^2, \\ \mathcal{A} &= \Sigma(r^2+a^2) + 2Mra^2\sin^2\theta, \\ \Sigma &= r^2 + a^2\cos^2\theta, \qquad \Delta = r^2 - 2Mr + a^2, \end{aligned}} \tag{9.3}$$

where $a = J/M$. If $M^2 \geq a^2$ this solution describes rotating BHs with mass M and angular momentum $J = Ma$. The event horizon is placed at $r = r_+ = M + \sqrt{M^2 - a^2}$ (the larger value of r for which $\Delta = 0$). When $a = 0$ we recover the Schwarzschild solution. Observe that, if we take $M \to 0$ keeping a finite, we also obtain Minkowski spacetime, as opposed to the limit $M \to 0$ with finite q in the RN case. If $M^2 < a^2$ the solution describes naked singularities. This resembles what happens in the RN case. Here we can think that a star with a large enough angular momentum cannot undergo spontaneous gravitational collapse and so the Kerr solutions with $M^2 < a^2$ and naked singularities never arise, according to the cosmic-censorship conjecture.

The Kerr solution for $r > r_+$ is not the metric of any known rotating body: there is no known "interior Kerr solution" as in the Schwarzschild case. Instead, such spacetimes are produced by certain rotating-disk sources (see Section 6.2 of [149] for a short review with references). However, the Kerr solution describes *all* isolated, rotating, uncharged BHs.

More details on the Kerr solutions can be found in most standard textbooks on GR and in the monograph [741]. Our subject now is the Taub–NUT solution.

If, asymptotically,

$$g_{t\varphi} \sim 2N\cos\theta, \tag{9.4}$$

the solution describes an object with *NUT charge N*. We will discuss soon the meaning of this new charge, for which there is no Newtonian analog. The simplest vacuum solution with this kind of charge is the Taub–NUT solution [724, 879]

$$\boxed{\begin{aligned} ds^2 &= f(r)(dt + 2N\cos\theta\, d\varphi)^2 - f^{-1}(r)dr^2 - \left(r^2 + N^2\right) d\Omega^2_{(2)}, \\ f(r) &= \frac{(r - r_+)(r - r_-)}{r^2 + N^2}, \\ r_\pm &= M \pm r_0, \qquad r_0^2 = M^2 + N^2, \end{aligned}} \tag{9.5}$$

which is a generalization of the Schwarzschild solution with NUT charge, and reduces to it when $N = 0$.

Let us list some immediate properties of this spacetime.

1. The solution is non-trivial in the $M \to 0$ limit, in which it may be interpreted as the gravitational field of a pure "spike" of spin [167, 329].

2. The mass of the solution can be found by standard methods and it is M. In particular, we know that we can determine the mass by studying the weak-field expansion and making contact with the Newtonian limit. The Newtonian gravitational potential is given in this approximation by $\phi \sim (g_{tt} - 1)/2 = -M/r$. The Taub–NUT solution has other non-vanishing components of the metric. The diagonal components are still related to the gravitostatic Newtonian potential ϕ, but the off-diagonal ones $g_{t\underline{i}}$ are related to a gravitomagnetic potential $\vec{A}$ according to Eq. (3.141). In the coordinates that we are using, we see that the Taub–NUT gravitational field has, as non-vanishing component of the gravitomagnetic potential,

$$A_\varphi = g_{t\varphi} = 2N\cos\theta. \tag{9.6}$$

   This is essentially the electromagnetic field of a magnetic monopole of charge proportional to N. Thus, the NUT charge N can be considered as a sort of "magnetic mass" [297] and so the Taub–NUT solution can be interpreted as a gravitational dyon [328].

3. This metric is not asymptotically flat but defines its own class of asymptotic behavior (*asymptotically Taub–NUT spacetimes*) labeled by N, which is associated with the non-vanishing at infinity of the off-diagonal $g_{t\varphi}$ component of the metric and, as we are going to see, with the periodicity of the time coordinate. The reason for this periodicity is the desire to avoid certain singularities and to have a spherically symmetric solution. Thus, let us first study the singularities.

4. This metric does not have curvature singularities and is perfectly regular at $r = 0$. However, it has the so-called "wire singularities" at $\theta = 0$ and $\theta = \pi$ where the metric fails to be invertible. These coordinate singularities cannot be cured simultaneously. Misner [697] found a way to make the metric regular everywhere by introducing two coordinate patches.

   (a) One patch covers the region $\theta \geq \pi/2$ around the north pole. In this region we change the time coordinate from t to $t^{(+)}$ defined by

   $$t = t^{(+)} - 2N\varphi, \tag{9.7}$$

   so

   $$ds^2_{(+)} = f(r)\left[dt^{(+)} - 2N(1-\cos\theta)d\varphi\right]^2 - f^{-1}(r)dr^2 - \left(r^2 + N^2\right) d\Omega^2_{(2)}. \tag{9.8}$$

   (b) The second patch covers the region $\theta \leq \pi/2$ around the south pole. In this region we change the time coordinate from t to $t^{(-)}$ defined by

   $$t = t^{(-)} + 2N\varphi, \tag{9.9}$$

   so

   $$ds^2_{(-)} = f(r)\left[dt^{(-)} + 2N(1+\cos\theta)d\varphi\right]^2 - f^{-1}(r)dr^2 - \left(r^2 + N^2\right) d\Omega^2_{(2)}. \tag{9.10}$$

   In the overlap region $t^{(+)} = t^{(-)} + 4N\varphi$ and, since φ is compact with period 2π, then both of $t^{(\pm)}$ have to be compact with period $8\pi N$.

5. The metric admits three Killing vectors whose Lie brackets are those of the so (3) Lie algebra. When the period of the time coordinates is precisely $8\pi N$ this local symmetry can be integrated to give a global SO (3) symmetry and the metric is indeed spherically symmetric [587]. Furthermore, the Taub–NUT spacetime now has a very different topology: the hypersurfaces of constant r are 3-spheres S^3 constructed as a Hopf fibration of S^2, the fiber being the time S^1. Thus, Taub–NUT has the topology of $\mathbb{R}^4$.

6. This way of eliminating the wire singularities is identical to the way in which we eliminated the string singularity in the vector field of the Dirac monopole because the mathematical problem is identical. The Dirac quantization condition translates into a relation between the periodicity of the time coordinate and the NUT charge.

   This relation is more than just a coincidence: in Chapter 11 we will generate by compactification of the Euclidean time of the Euclidean version of the Taub–NUT solution a magnetically charged black hole. For this reason, the Euclidean Taub–NUT solution, which we will study later, is also known as the *Kaluza–Klein monopole*.

7. The metric function $f(r)$ has two zeros at $r = r_\pm$ and the metric has coordinate singularities there. For $r > r_+$ and $r < r_-$ (where t is timelike and r spacelike) the metric has closed timelike curves. Thus, although the form of the metric is similar to the Reissner–Nordström metric, no black-hole interpretation is possible. Furthermore, the "extremality parameter" r_0 vanishes only for $M = N = 0$.
8. In the region $r_- < r < r_+$, the coordinate t is spacelike and r is timelike. This region describes a non-singular, anisotropic, closed cosmological model. It can be thought of as a closed universe containing gravitational radiation having the longest possible wavelength [184].
9. There is no known generalization to higher dimensions. It can be embedded in higher-dimensional spacetimes but always as a product metric. The NUT charge seems to be an intrinsically four-dimensional charge (see, however, [578])
10. There are interior Taub–NUT solutions [178].

9.2 The Euclidean Taub–NUT solution

The Euclidean Taub–NUT metric is interesting in itself, as we are going to see. We obtain it by Wick-rotating the time, which also has to be accompanied by a Wick rotation of the NUT charge N in order to keep the metric real. We denote the Euclidean time by τ. The result is (taking into account the two patches)

$$\boxed{\begin{aligned} -d\sigma_\pm^2 &= f(r)\big[d\tau^{(\pm)} \mp 2N\,(1 \mp \cos\theta)\,d\varphi\big]^2 + f^{-1}(r)dr^2 + \left(r^2 - N^2\right) d\Omega_{(2)}^2, \\ f(r) &= \frac{(r-r_+)(r-r_-)}{r^2 - N^2}, \\ r_\pm &= M \pm r_0, \qquad r_0^2 = M^2 - N^2. \end{aligned}} \tag{9.11}$$

We see that, in the Euclidean case, there is an extreme limit[2] $r_0 = 0$, which corresponds to $M = |N|$. In this case, after shifting the radial coordinate by M, we find that the solution can be written in isotropic coordinates in the following way (we suppress the $^\pm$ and it is understood that τ is a compact coordinate with period $8\pi N$ and the 1-form A is defined by patches so it is regular everywhere):

$$\boxed{\begin{aligned} -d\sigma^2 &= H^{-1}(d\tau + A)^2 + H d\vec{x}_3^{\,2}, \\ H &= 1 + \frac{2|N|}{|\vec{x}_3|}, \\ A &= A_i dx^i, \qquad \epsilon_{ijk}\partial_i A_j = \mathrm{sign}(N)\,\partial_k H. \end{aligned}} \tag{9.12}$$

[2] In the literature it is the extreme limit that usually receives the name of Euclidean Taub–NUT solution.

This solution is known as the (Sorkin–Gross–Perry) *Kaluza–Klein (KK) monopole* [483, 860]. The 1-form A satisfies the Dirac-monopole equation (8.149), which we know has to be solved in two different patches.

9.2.1 Self-dual gravitational instantons

If we use the above form of the solution as an *Ansatz* in the vacuum Einstein equations, we find that we have a solution for every function H that is harmonic in three-dimensional space:

$$\boxed{\begin{array}{l} -d\sigma^2 = H^{-1}(d\tau + A)^2 + H d\vec{x}_3^{\,2}, \\ \quad A = A_i dx^i, \qquad \epsilon_{ijk}\partial_i A_j = \pm\partial_k H, \\ \partial_{\underline{i}}\partial_{\underline{i}} H = 0. \end{array}} \tag{9.13}$$

In fact, we know that the Laplace equation is the integrability condition of the Dirac-monopole equation, ensuring that it can be (locally) solved. Now it is possible to have solutions with several KK monopoles in equilibrium by taking a harmonic function H with several point-like singularities (*Gibbons–Hawking multicenter metrics* [437]):

$$H = \epsilon + \sum_{I=1}^{k} \frac{2|N_I|}{|\vec{x}_3 - \vec{x}_{3\,I}|}. \tag{9.14}$$

If we choose $\epsilon = 1$, we have the multi-Taub–NUT metric. If all the NUT charges N_I are equal to N, then all the wire singularities associated with each pole can be removed simultaneously by taking the period of τ equal to $8\pi N$. Asymptotically the topology is that of a *lens space*: an S^3 in which k points have been identified, and so they are not asymptotically flat in general.

If we choose $\epsilon = 0$, the wire singularities can be eliminated by the same procedure, but the N_Is can all be made equal by a rescaling of the coordinates. The topology is the same as in the $\epsilon = 1$ case, but the metrics are *asymptotically locally Euclidean* (ALE), i.e. they are asymptotic to the quotient of Euclidean space by a discrete subgroup of SO(4). The $k = 1$ solution is just flat space. The $k = 2$ solution is equivalent [787] to the Eguchi–Hanson solution [348], which is usually written in the form

$$\boxed{-d\sigma^2 = \left(1 - \frac{a^4}{\rho^4}\right)\frac{\rho^2}{4}(d\tau + \cos\theta d\varphi)^2 + \left(1 - \frac{a^4}{\rho^4}\right)^{-1} d\rho^2 + \frac{\rho^2}{4} d\Omega^2_{(2)}.} \tag{9.15}$$

This solution has an apparent singularity at $\rho = a$ that can be removed by identifying $\tau \sim \tau + 2\pi$. With this identification, all the $\rho > a$ constant hypersurfaces are $\mathbb{R}\mathrm{P}^3$ (S^3 with antipodal points identified).

All these solutions are *gravitational instantons*, the gravitational analog of the SU(2) BPST Yang–Mills (YM) instantons discovered in [102], i.e. non-singular solutions of the Euclidean Einstein equations with finite action, i.e. local minima of the Euclidean Einstein action that can be used to compute the partition function in the saddle-point approximation[3] [513]. This definition also applies to the Euclidean Schwarzschild and RN solutions, of course. It also applies to the general Euclidean Taub–NUT Eq. (9.11) which, for the particular value $M = \frac{5}{4}|N|$, is known [347] as the *Taub-bolt solution* [751]. However, the gravitational instantons with Gibbons–Hawking metric Eq. (9.13) have a very special property that brings them closer to their YM counterparts: the SU(2) YM instantons have an (anti-)self-dual field strength[4]

$$\mathcal{F}_{\mu\nu} = {}^{\star}\mathcal{F}_{\mu\nu}, \tag{9.16}$$

and the above gravitational instantons have an (anti-)self-dual Lorentz (SO(4)) curvature

$$R_{\mu\nu}{}^{ab}(\omega) = \pm{}^{\star}R_{\mu\nu}{}^{ab}(\omega). \tag{9.17}$$

The (anti-)self-duality of the YM field strength implies, upon use of the Bianchi identity Eq. (A.43), the YM equations of motion Eq. (A.45). The (anti-)self-duality of the Lorentz curvature[5] implies, via the Bianchi identity $R_{[\mu\nu\rho]}{}^{\sigma} = 0$, the vanishing of the Ricci tensor and the Einstein equations. Both in the YM case and in the gravitational case, (anti-)self-duality is also related to special supersymmetry properties (see Chapter 13).

Four-dimensional SU(2) YM instantons can be characterized by topological invariants such as the *second Chern class*,

$$c_2 = \frac{1}{16\pi^2}\int d^4x \,\mathrm{Tr}(\mathcal{F}{}^{\star}\mathcal{F}). \tag{9.18}$$

Then, the manifestly positive integrals

$$\int d^4x \,\left(\mathcal{F} \pm {}^{\star}\mathcal{F}\right)^2 = 2\int d^4x \,\left(\mathcal{F}^2 \pm \mathcal{F}\,{}^{\star}\mathcal{F}\right) = 8S_{\mathrm{E\,YM}} \pm 16\pi^2 c_2 \geq 0, \tag{9.19}$$

can be used to obtain a bound for the Euclidean YM action $S_{\mathrm{E\,YM}}$:

$$S_{\mathrm{E\,YM}} \geq 2\pi^2|c_2|. \tag{9.20}$$

(Anti-)self-dual YM field configurations are the solutions that minimize the Euclidean action in a sector characterized by the given topological number c_2.

---

3 A table with the properties of these and other gravitational instantons can be found in Appendix D of [347]. A calculation of the Euclidean actions based on the isometries of the instantons was done in [438, 468] (for more recent references see [515, 517–9, 584]).

4 (Anti-)self-duality can be consistently imposed only in even dimensions and depending on the signature: with Lorentzian signature, only for $d = 4n + 2$; and with Euclidean signature, only in $d = 4n$.

5 Observe that, in Riemannian spaces, the symmetry property (Bianchi identity) $R_{\mu\nu\rho\sigma} = R_{\rho\sigma\mu\nu}$ implies that the Lorentz curvature 2-form $R_{\mu\nu}{}^{ab}$is also (anti-)self-dual in the Lorentz indices ab. Furthermore, if the SO(4) curvature is (anti-)self-dual, there is always a gauge (a frame $e_a{}^{\mu}$) in which the connection $\omega_{\mu}{}^{ab}$ is also (anti-)self-dual in the Lorentz indices ab [348]. The Gibbons–Hawking multicenter metric has an (anti-)self-dual connection in the frame Eq. (9.43), but not in the frame Eq. (9.50). This property of (anti-)self-dual curvatures is a particular case of a more general property: as we are going to see in the next section, an object with (anti-)self-dual SO(4) indices is in fact an object with SU(2) indices embedded in SO(4) and therefore (anti-)self-dual SO(4) curvatures are SU(2) curvatures or curvatures of *special* SU(2) *holonomy*. The "reduction theorem" (Section II.7 of Vol. 1 of [630]) states that there is always a frame in which the spin connection has the same holonomy as the curvature.

Four-dimensional gravitational instantons are characterized by two topological invariants: the *Hirzebruch signature* $\tau(\mathrm{M})$, which is a third of the integral of the *first Pontrjagin class* p_1,

$$\tau(\mathrm{M}) = \tfrac{1}{3}\int_{\mathrm{M}} p_1 = -\frac{1}{24\pi^2}\int_{\mathrm{M}} \mathrm{Tr}_{\mathrm{v}}(R \wedge R) = \frac{1}{96\pi^2}\int_{\mathrm{M}} d^4x\sqrt{|g|}\epsilon^{\mu\nu\rho\sigma} R_{\mu\nu}{}^{ab} R_{\rho\sigma\, ab}, \quad (9.21)$$

where Tr_{v} denotes the trace in the vector representation and $\chi(\mathrm{M})$ the *Euler characteristic*, given in Eq. (8.116), but there is no obvious direct relation between these invariants and the Einstein–Hilbert action (nevertheless, see the discussion in Section 8.6).

The relation between these YM and gravitational configurations is really worth investigating a little bit further. Let us first review the BPST SU(2) instanton in the form known as the *'t Hooft Ansatz* [601].

9.2.2 The BPST instanton

The so called *'t Hooft Ansatz* [601] for the SU(2) instanton connection 1-form $\mathcal{A}_m$ has the form[6]

$$\mathcal{A}_m^{(\pm)} = -\mathbf{M}^{\mp}_{mn} V_n, \quad (9.22)$$

where the 2×2 matrices $\mathbf{M}^{\mp}_{mn}$ are (anti-)self-dual generators of so(4) constructed from the Pauli matrices, and V_n is a vector field to be determined by the requirement that the field strength $\mathcal{F}^{(\pm)} = d\mathcal{A}^{(\pm)} - \mathcal{A}^{(\pm)}\mathcal{A}^{(\pm)}$ be (anti-)self-dual,

$$^\star\mathcal{F}^{(\pm)}_{mn} = \pm\mathcal{F}^{(\pm)}_{mn}. \quad (9.23)$$

This condition is satisfied if

$$\begin{aligned} \partial_m V_m + V_m V_m &= 0, \\ ^\star f_{mn} \pm f_{mn} &= 0, \qquad f_{mn} = 2\partial_{[m} V_{n]}. \end{aligned} \quad (9.24)$$

The second condition is usually satisfied by choosing a V_m that is the gradient of some scalar function, $V_m = \partial_m \ln V$. Then, the first condition becomes the equation

$$V^{-1}\partial_m\partial_m V = 0. \quad (9.25)$$

We have an instanton solution for each harmonic function V on four-dimensional Euclidean space. Not all of them have finite action, though. The most interesting choice is

$$V = \epsilon + \sum_{I=1}^{k} \frac{\lambda_I^2}{|\vec{x}_4 - \vec{x}_{4\,I}|^2}, \qquad \epsilon = 1, 0, \quad (9.26)$$

for k instantons. For $k = \epsilon = 1$ we recover the BPST instanton solution in the second gauge [102], which can be written in a suggestive form that resembles the electromagnetic vector field of the ERN BH,

$$\mathcal{A}_m^{(+)} = (V^{-1} - 1)\mathbf{g}^{-1}\,\partial_m\mathbf{g}, \qquad \mathcal{A}_m^{(-)} = -\left(V^{-1} - 1\right)\,\partial_m\mathbf{g}\,\mathbf{g}^{-1}, \quad (9.27)$$

[6] Here we are in flat four-dimensional Euclidean space and we use non-underlined Latin indices $m, n, p, q = 0, 1, 2, 3$ for convenience and calligraphic $\mathcal{A}$ for the YM connection to distinguish it from the 1-form A appearing in the Taub–NUT metric.

where $\mathbf{g}(x)$ is the SU(2)-valued function

$$\mathbf{g} = (x^0 - i x^i \sigma^i)/|\vec{x}_4|, \tag{9.28}$$

and the σ^i are the Pauli matrices Eq. (B.9).

The 't Hooft *Ansatz* makes the embedding of the SU(2) gauge connection in SO(4) easy (only the self-duality properties of the generators and their commutation relations play a role): we simply have to take the generators of so(4) in the fundamental representation

$$(\mathbf{M}_{mn})^{\ pq} = -2\delta_{mn}{}^{pq}, \tag{9.29}$$

and then take their (anti-)self-dual part,[7]

$$(\mathbf{M}^{(\pm)}_{mn})^{pq} = \tfrac{1}{2}\left(\delta_{mn}{}^{rs} \pm \tfrac{1}{2}\epsilon_{mn}{}^{rs}\right)(\mathbf{M}_{rs})^{\ pq} = -\left(\delta_{mn}{}^{pq} \pm \tfrac{1}{2}\epsilon_{mn}{}^{pq}\right). \tag{9.30}$$

If we split the four dimensions into $m = 0, i$, with $i = 1, 2, 3$, the components are

$$\begin{aligned} \mathcal{A}^{(\pm)}_{0\,i0} &= -\tfrac{1}{2}\partial_i \ln V, & \mathcal{A}^{(\pm)}_{i\,0j} &= -\tfrac{1}{2}\partial_0 \ln V\,\delta_{ij} \pm \epsilon_{ijk}\partial_k \ln V, \\ \mathcal{A}^{(\pm)}_{0\,ij} &= \mp\epsilon_{ijk}\partial_k \ln V, & \mathcal{A}^{(\pm)}_{i\,jk} &= +\delta_{i[j}\partial_{k]} \ln V \pm \epsilon_{ijk}\partial_0 \ln V. \end{aligned} \tag{9.31}$$

For reasons that will become clear, we are also interested in a slightly different choice of self-dual so(4) generators $\tilde{\mathbf{M}}^{\pm}_{ab}$ defined as follows:

$$\tilde{\mathbf{M}}^{\pm}_{ij} = -\mathbf{M}^{\mp}_{ij}, \qquad \tilde{\mathbf{M}}^{\pm}_{0i} = +\mathbf{M}^{\mp}_{0i}, \tag{9.32}$$

so the non-vanishing components of

$$\tilde{\mathcal{A}}^{(\pm)}_m = -\tilde{\mathbf{M}}^{\mp}_{mn}\,\partial_n V \tag{9.33}$$

are

$$\begin{aligned} \tilde{\mathcal{A}}^{(\pm)}_{0\,i0} &= -\tfrac{1}{2}\partial_i \ln V, & \tilde{\mathcal{A}}^{(\pm)}_{i\,0j} &= -\tfrac{1}{2}\partial_0 \ln V\,\delta_{ij} \mp \epsilon_{ijk}\partial_k \ln V, \\ \tilde{\mathcal{A}}^{(\pm)}_{0\,ij} &= \mp\epsilon_{ijk}\partial_k \ln V, & \tilde{\mathcal{A}}^{(\pm)}_{i\,jk} &= -\delta_{i[j}\partial_{k]} \ln V \pm \epsilon_{ijk}\partial_0 \ln V. \end{aligned} \tag{9.34}$$

9.2.3 Instantons and monopoles

There is an interesting relation between instantons and certain monopoles in spite of their different (Euclidean, Lorentzian) natures. Let us restrict ourselves to YM field configurations that do not depend on the coordinate $x^0 = \tau$. The restricted theory is, thus, effectively three-dimensional. The component $\mathcal{A}_0$ now has the interpretation of a three-dimensional scalar in the adjoint representation that we denote by Φ, while the other three components

[7] These matrices have the same duality properties in the Lie algebra indices ab and in the representation indices cd because they have the interchange property $(\mathbf{M}^{(\pm)}_{mn})_{pq} = (\mathbf{M}^{(\pm)}_{pq})_{mn}$. This property implies that the (SO(4)) connection is also (anti-)self-dual in the group indices and so will be the curvature. On the other hand, observe that we are basically using the fact that the algebra so(4) = su(2) ⊕ su(2). The self-dual part of the so(4) generators generates one of the su(2) subspaces and the anti-self-dual part generates the other.

become the components of the three-dimensional YM vector field. The $\mathcal{F}_{i0}$ components of the field strength are

$$\mathcal{F}_{i0} = \partial_i \Phi - [\mathcal{A}_i, \Phi] = \mathcal{D}_i \Phi, \tag{9.35}$$

i.e. the three-dimensional YM covariant derivative of the scalar Φ. After integrating over the redundant coordinate τ (which we take to be periodic with period 2π), the Euclidean YM action becomes

$$S_{\mathrm{E\,YM}} = 2\pi \int d^3x \,\mathrm{Tr}\left[\tfrac{1}{4}\mathcal{F}_{ij}\mathcal{F}_{ij} + \tfrac{1}{2}\mathcal{D}_i\Phi\mathcal{D}_i\Phi\right], \tag{9.36}$$

and the (anti-)self-duality equation for $\mathcal{F}$ becomes the *Bogomol'nyi equation* [163, 248]

$$\mathcal{F}_{ij} = \mp\epsilon_{ijk}\mathcal{D}_k\Phi. \tag{9.37}$$

Let us now consider the (four-dimensional, Lorentzian) Georgi–Glashow model [425] which consists of an SU(2) gauge field $\mathcal{A}$ coupled to a triplet of Higgs fields Φ with a potential $V(\Phi) = \frac{1}{2}\lambda[\mathrm{Tr}(\Phi^2) - 1]^2$

$$S_{\mathrm{GG}} = \int d^4x \,\left\{-\tfrac{1}{4}\mathrm{Tr}\,\mathcal{F}^2 + \tfrac{1}{2}\mathrm{Tr}(\mathcal{D}\Phi)^2 - \tfrac{1}{2}\lambda[\mathrm{Tr}(\Phi^2) - 1]^2\right\}. \tag{9.38}$$

't Hooft [890] and Polyakov [783] found a magnetic-monopole solution of this model that generalizes Dirac's. In the $\lambda = 0$ limit (the *Bogomol'nyi–Prasad–Sommerfield (BPS) limit*), the solution takes an especially simple form [788] and has special properties that can also be related to supersymmetry (see Chapter 13).

Let us focus on purely magnetic (i.e. $\mathcal{A}_0 = 0$) and static ($\partial_0\mathcal{A}_\mu = \partial_0\Phi = 0$) field configurations. Their energy (taking $\lambda = 0$) is given precisely by $[1/(2\pi)]S_{\mathrm{E\,YM}}$ in Eq. (9.36). It is not surprising that, therefore, the energy of these configurations is bounded: the manifestly positive integral

$$\int d^3x \,\mathrm{Tr}\left(\mathcal{F}_{ij} \pm \epsilon_{ijk}\mathcal{D}_k\Phi\right)^2 = 8E \pm \int d^3x \epsilon_{ijk}\mathrm{Tr}(\mathcal{F}_{ij}\mathcal{D}_k\Phi) \geq 0. \tag{9.39}$$

On integrating by parts and using the three-dimensional Bianchi identity, we find that

$$\int d^3x \epsilon_{ijk}\mathrm{Tr}(\mathcal{F}_{ij}\mathcal{D}_k\Phi) = \int d^3x \partial_i(\epsilon_{ijk}\Phi\mathcal{F}_{ij}) = 4\int_{S^2_\infty}\mathrm{Tr}(\Phi\mathcal{F}) = -4p, \tag{9.40}$$

where we have used Stokes' theorem and where p is the SU(2) magnetic charge. Thus,

$$E \geq \tfrac{1}{2}|p|, \tag{9.41}$$

which is the *Bogomol'nyi or BPS bound*. We know that p is quantized (for $g = 1$), $p = 2\pi n$. Using this fact and the relation $E = [1/(2\pi)]S_{\mathrm{E\,YM}}$, this relation is completely equivalent to Eq. (9.20). On the other hand, the configurations that minimize the energy $E = \frac{1}{2}|p|$ (saturate the BPS bound) are those satisfying the first-order Bogomol'nyi equation and it is easy to prove that these configurations also solve all the (second-order) equations of motion of the $\lambda = 0$ Georgi–Glashow model.

The immediate conclusion of this discussion is that, if we take SU(2) (anti-)self-dual instantons that do not depend on the τ coordinate, we have automatically a magnetic monopole solution of the Georgi–Glashow model with $\lambda = 0$ satisfying the Bogomol'nyi bound. In particular, the BPS limit of the 't Hooft–Polyakov SU(2) is obtained using the 't Hooft *Ansatz* with the harmonic function

$$V = 1 + \frac{\lambda}{|\vec{x}_3|}. \tag{9.42}$$

9.2.4 The BPST instanton and the KK monopole

We are now ready to establish a relation between the Euclidean Taub–NUT solution (KK monopole) and the BPST instanton. We are going to see that the spin-connection frame components $\omega_{m\,np}$ of the KK monopole are identical to the SO(4)-embedded components of the BPST instanton connection $\tilde{\mathcal{A}}_{m\,np}$ with the harmonic function V identical to the harmonic function H of the KK monopole, depending on just three coordinates $\vec{x}_3$.

In the simplest frame,

$$\begin{aligned} e^0 &= H^{-\frac{1}{2}}[d\tau + A_{\underline{i}}dx^i], & e_0 &= H^{\frac{1}{2}}\partial_\tau, \\ e^i &= H^{\frac{1}{2}}dx^i, & e_i &= H^{-\frac{1}{2}}[\partial_{\underline{i}} - A_{\underline{i}}\partial_\tau], \end{aligned} \tag{9.43}$$

the frame components of the spin connection (which is just an SO(4) connection) are

$$\begin{aligned} \omega_{0\,i0}(e) &= -\tfrac{1}{2}\partial_i \ln H, & \omega_{i\,0j}(e) &= H^{-1}\partial_{[i}A_{\underline{j}]}, \\ \omega_{0\,ij}(e) &= H^{-1}\partial_{[i}A_{\underline{j}]}, & \omega_{i\,jk}(e) &= -\delta_{i[j}\partial_{k]} \ln H. \end{aligned} \tag{9.44}$$

Here it is important to observe that all partial derivatives in this expression have frame indices. Using the Dirac-monopole equation for the 1-form A,

$$\epsilon_{ijk}\partial_{[\underline{i}}A_{\underline{j}]} = \pm\partial_{\underline{k}}H, \tag{9.45}$$

the KK-monopole spin connection becomes

$$\begin{aligned} \omega^{(\pm)}_{0\,i0}(e) &= -\tfrac{1}{2}\partial_i \ln H, & \omega^{(\pm)}_{i\,0j}(e) &= \pm\epsilon_{ijk}\partial_k \ln H, \\ \omega^{(\pm)}_{0\,ij}(e) &= \pm\epsilon_{ijk}\partial_k \ln H, & \omega^{(\pm)}_{i\,jk}(e) &= -\delta_{i[j}\partial_{k]} \ln H, \end{aligned} \tag{9.46}$$

which is identical to the connection $\tilde{A}$ in Eq. (9.34). It is, therefore, (anti-)self-dual and has SU(2) holonomy.

9.2.5 Bianchi IX gravitational instantons

In [449] the class of gravitational instantons with an SU(2) or SO(3) isometry group acting transitively (Bianchi IX metrics) was studied, with special emphasis on those with self-dual curvature. This class includes some of the gravitational instantons that we have studied, namely Taub–NUT, Taub-bolt, and Eguchi–Hanson instantons, and its discussion will provide us with some further interesting examples.

All Ricci-flat ($R_{\mu\nu}=0$) Bianchi IX metrics can locally be written in the form

$$d\sigma^2 = (a_1a_2a_3)d\eta^2 + \sum_{i=1,2,3}(a_ie^i)^2, \tag{9.47}$$

where the a_is depend only on η and the σ^is are the η-independent SU(2) Maurer–Cartan 1-forms denoted by e^i in Appendix A.3.1.

A simple solution of the Einstein equations with $a_1^2 = a_2^2$ is given by the Euclidean Taub–NUT solution ($M \neq N$),

$$\begin{aligned} a_1^1 = a_2^2 &= \tfrac{1}{4}q\sinh[q(\eta-\eta_2)]\,\mathrm{cosech}^2[q(\eta-\eta_1)], \\ q(\eta-\eta_2)a_3^2 &= \mathrm{cosech}[q(\eta-\eta_2)], \end{aligned} \tag{9.48}$$

where q, η_1, and η_2 are integration constants. The relation to the standard integration constants and coordinates is

$$\begin{aligned} N^2 &= -\tfrac{1}{4}q\,\mathrm{cosech}[q(\eta_2-\eta_1)], \\ M &= N\cosh[q(\eta_2-\eta_1)], \\ r &= \frac{q}{4N}\{\coth[\tfrac{1}{2}q(\eta-\eta_1)] - \coth[q(\eta_2-\eta_1)]\}, \\ \tau &= 4N\psi. \end{aligned} \tag{9.49}$$

On taking the limit $q\to 0$ we obtain the $M=|N|$ Taub–NUT metric with self-dual curvature. With the obvious frame choice

$$e^0 = a_1a_2a_3d\eta, \qquad e^i = a_i\sigma^i, \tag{9.50}$$

its connection is *not* (anti-)self-dual. With $\eta_1=\eta_2$ we obtain the Eguchi–Hanson metric Eq. (9.15) with

$$M = N + \frac{a^4}{128N^3}, \qquad r = M + \frac{\rho^2}{8N}, \tag{9.51}$$

after taking the $N\to\infty$ limit. This metric has self-dual curvature and connection (using the above frame). On setting $M=\frac{5}{4}|N|$ we obtain the Taub-bolt metric.

If we impose the condition that the Lorentz curvature is self-dual in the above frame, one obtains, after one integration, the equations

$$\begin{aligned} 2\frac{d}{d\eta}\ln a_1 &= \sum_{i=1,2,3} a_i^2 - 2a_1^2 - 2\lambda_1a_2a_3, \\ \lambda_1 &= \lambda_2\lambda_3, \end{aligned} \tag{9.52}$$

and the equations one obtains from these by cyclic permutations of the indices $i=1,2,3$. The algebraic equations for the constants λ_i admit three possible solutions:

$$(\lambda_1,\lambda_2,\lambda_3) = (0,0,0),\ (1,1,1),\ (-1,-1,1). \tag{9.53}$$

The first solution corresponds to metrics whose connection is self-dual and can be completely integrated. The general solution is [104]

$$
\begin{aligned}
d\sigma^2 &= (f_1 f_2 f_3)^{-\frac{1}{2}} d\eta^2 + (f_1 f_2 f_3)^{\frac{1}{2}} \frac{\rho^2}{4} \sum_{i=1,2,3} (f_i^{-\frac{1}{2}} \sigma^i)^2, \\
f_i &= 1 - \frac{b_i^4}{\rho^4}.
\end{aligned}
\tag{9.54}
$$

$b_1 = b_2 = a,\ b_3 = 0$ is the Eguchi–Hanson metric Eq. (9.15). Solutions of the second class have not been obtained except for the special case $a_1 = a_2$, that gives the self-dual Taub–NUT metric. The third case is not equivalent to the second and corresponds to the Atiyah–Hitchin metric [56] which governs the interaction of two slowly moving BPS SU(2) monopoles.

9.3 Charged Taub–NUT solutions and IWP solutions

Let us consider stationary, axially symmetric solutions of the Einstein–Maxwell system. Some of them are the result of adding electric or magnetic charges to vacuum solutions.

The charged version of the Kerr solution was found in [723] and is known as the Kerr–Newman solution, which takes the form

$$
\boxed{
\begin{aligned}
ds^2 &= \left(1 - \frac{2Mr - 4q^2}{\Sigma}\right) dt^2 + 2\frac{a(2Mr - 4q^2)\sin^2\theta}{\Sigma} dt d\varphi \\
&\quad - \frac{\Sigma}{\Delta} dr^2 - \Sigma d\theta^2 - \frac{\mathcal{A}}{\Sigma} \sin^2\theta\, d\varphi^2, \\
\Sigma &= r^2 + a^2\cos^2\theta, \qquad \Delta = r^2 - 2Mr + 4q^2 + a^2 \\
\mathcal{A} &= \Sigma(r^2 + a^2) + (2Mr - 4q^2)a^2\sin^2\theta, \\
A_\mu &= \frac{4qr}{\Sigma}[\delta_{\mu t} - \delta_{\mu\varphi} a \sin^2\theta].
\end{aligned}
}
\tag{9.55}
$$

Again, if $M^2 \geq 4q^2 + a^2$, this solution describes BHs with mass M, angular momentum $J = Ma$, and electric charge q, with the event horizon at $r = r_+ = M + \sqrt{M^2 - 4q^2 - a^2}$ (the larger value of r for which $\Delta = 0$).

Observe that, although the solution is only electrically charged, the rotation induces a magnetic dipole moment and the A_φ component of the vector field is non-zero.

The electrically charged Taub–NUT solution was found by Brill in [184] and is

$$
\boxed{
\begin{aligned}
& ds^2 = f(r)(dt + 2N\cos\theta\, d\varphi)^2 - f^{-1}(r)dr^2 - \left(r^2 + N^2\right) d\Omega^2_{(2)}, \\
& F_{tr} = \frac{4q(r^2 - N^2)}{r^2 + N^2}, \qquad ({}^\star F)_{tr} = \frac{8qNr}{(r^2 + N^2)^2}, \\
& f(r) = \frac{(r - r_+)(r - r_-)}{r^2 + N^2}, \\
& r_\pm = M \pm r_0, \qquad r_0^2 = M^2 + N^2 - 4q^2.
\end{aligned}
}
\tag{9.56}
$$

It reduces to the RN solution when we set the NUT charge to zero. It is trivial to generalize these solutions to the magnetic and dyonic cases.

In contrast to the Taub–NUT solution, the charged Taub–NUT solution does have an extremal limit $M^2 + N^2 = 4q^2$ in which the extremality parameter r_0 vanishes and the two zeros of the metric function $f(r)$ coincide. In this case, by shifting the radial coordinate to $\rho = r - M$ and defining Cartesian coordinates such that $\rho = |\vec{x}_3|$, we find a simple form of the solution,[8]

$$
\boxed{
\begin{aligned}
& ds^2 = |\mathcal{H}|^{-2}\,(dt + A)^2 - |\mathcal{H}|^2 d\vec{x}_3^{\,2}, \\
& A_t = 2\mathrm{Re}(e^{i\alpha}\mathcal{H}), \qquad \tilde{A}_t = 2\mathrm{Im}(e^{i\alpha}\mathcal{H}), \\
& H = 1 + \frac{M + iN}{|\vec{x}_3|}, \\
& A = A_{\underline{i}} dx^i, \qquad \epsilon_{ijk}\partial_i A_j = \pm\mathrm{Im}(\overline{\mathcal{H}}\partial_k\mathcal{H}).
\end{aligned}
}
\tag{9.57}
$$

As in some of the other "extreme" solutions that we have found so far,[9] it turns out that we obtain a solution for *any complex harmonic function* $\mathcal{H}(\vec{x}_3)$. By absorbing the complex phase $e^{i\alpha}$ into $\mathcal{H}$, we can write the general solution in this form:

$$
\boxed{
\begin{aligned}
& ds^2 = |\mathcal{H}|^{-2}(dt + A)^2 - |\mathcal{H}|^2 d\vec{x}_3^{\,2}, \\
& A_t = 2\mathrm{Re}\,\mathcal{H}, \qquad \tilde{A}_t = -2\mathrm{Re}(i\mathcal{H}), \\
& A = A_{\underline{i}} dx^i, \qquad \epsilon_{ijk}\partial_i A_j = \pm\mathrm{Im}(\overline{\mathcal{H}}\partial_k\mathcal{H}), \\
& \partial_{\underline{i}}\partial_{\underline{i}}\mathcal{H} = 0.
\end{aligned}
}
\tag{9.58}
$$

Metrics of the above form are known as *conformastationary metrics* [640]. Observe that the integrability condition of the equation for the 1-form A is the Laplace equation for $\mathcal{H}$.

[8] Here we are actually taking the extreme limit of the dyonic solution, which indeed has a simpler form. The information on the electric and magnetic charges is contained in the SO(2) electric–magnetic-duality phase $e^{i\alpha}$.

[9] But not in all of them. In particular, not in the Kerr BH.

This big family of solutions is known as the *Israel–Wilson–Perjés (IWP) solutions* [597, 769], although they were first discovered by Neugebauer [721]. This family contains all the "extreme" solutions (RN, charged Taub–NUT, and their multicenter generalizations) that we have found so far, plus many others that may have mass, electric and magnetic charges, NUT charge, and also angular momentum. In particular, the $M^2 = 4q^2$ Kerr–Newman solutions, for arbitrary angular momentum, belong to this family: their complex harmonic function is

$$\mathcal{H} = 1 + \frac{M}{\sqrt{x^2 + y^2 + (z - ia)^2}}. \tag{9.59}$$

In terms of more suitable oblate spheroidal coordinates,

$$\begin{aligned} x + iy &= [(r - M)^2 + a^2]^{\frac{1}{2}} \sin\theta \, e^{i\varphi}, \\ z &= (r - M)\cos\theta, \end{aligned} \tag{9.60}$$

the function $\mathcal{H}$ takes the form

$$\mathcal{H} = 1 + \frac{M}{r - M - ia\cos\theta}, \tag{9.61}$$

and the Euclidean three-dimensional metric becomes

$$d\vec{x}_3^{\,2} = \left[(r - M)^2 + a^2\cos^2\theta\right]\left[\frac{dr^2}{(r - M)^2 + a^2} + d\theta^2\right] + \left[(r - M)^2 + a^2\right]\sin^2\theta \, d\varphi^2. \tag{9.62}$$

Furthermore, the 1-form A is given by

$$A = \frac{(2Mr - M^2)a\sin^2\theta}{(r - M)^2 + a^2\cos^2\theta} d\varphi, \tag{9.63}$$

and

$$|\mathcal{H}|^2 = \frac{(r - m)^2 - a^2\cos^2\theta}{r^2 + a^2\cos^2\theta}, \tag{9.64}$$

and we recover the Kerr–Newman solutions with $M^2 = 4q^2$. These solutions are not BHs because they violate the bound $M^2 - 4q^2 - a^2 \geq 0$. In fact, it has been argued by Hartle and Hawking that the only BH-type solutions in the IWP family of metrics are the multi-ERN solutions.

For us, one of the main interests of this family is that it is electric–magnetic-duality-invariant and it is the most general family that we can have with the above charges always satisfying the identity $M^2 = 4|\vec{q}|^2$. An electric–magnetic-duality transformation is nothing but a change in the phase of $\mathcal{H}$. Non-extreme solutions can be constructed from the IWP class, by adding a "non-extremality function" W, as in the RN case [665]. We will study them as a subfamily of the most general BH-type solutions of pure $N = 4$, $d = 4$ SUEGRA.

10
Gravitational pp-waves

As we saw in Part I, the weak-field limit of GR is just a relativistic field theory of a massless spin-2 particle propagating in Minkowski spacetime. In the absence of sources, by choosing the De Donder gauge Eq. (3.100), it can be shown that the gravitational field $h_{\mu\nu}$ satisfies the wave equation (3.101) and, correspondingly, there are wave-like solutions of the weak-field equations like the one we found in Section 3.2.3 associated with a massless point-particle moving at the speed of light.

GR is, however, a highly non-linear theory and it is natural to wonder whether there are exact wave-like solutions of the full Einstein equations. The answer is definitely yes and in this chapter we are going to study some of them, the so-called pp-waves, which are especially interesting for us. In particular we are going to see that the linear solution we found in Section 3.2.3 is an exact solution of the full Einstein equations that has the same interpretation. We will use this solution many times in what follows to describe the gravitational field of Kaluza–Klein momentum modes, for instance.

10.1 pp-Waves

pp-waves (shorthand for *plane-fronted waves with parallel rays*) are metrics that, by definition, admit a covariantly constant null Killing vector field ℓ_μ:

$$\nabla_\mu \ell_\nu = 0, \qquad \ell^2 = \ell_\mu \ell^\mu = 0. \tag{10.1}$$

The first spacetimes with this property were discovered by Brinkmann in [193]. To describe pp-wave metrics, we define light-cone coordinates u and v in terms of the usual Cartesian coordinates

$$u = \frac{1}{\sqrt{2}}(t - z), \qquad v = \frac{1}{\sqrt{2}}(t + z), \tag{10.2}$$

which are related to the null Killing vector by

$$\ell_\mu = \partial_\mu u, \qquad \ell^\mu \partial_\mu v = 1, \tag{10.3}$$

i.e. v is the coordinate we can make the metric independent of, the only non-vanishing components of ℓ are $\ell_u = \ell^v = 1$, and the metric describes a gravitational wave propagating in the positive direction of the z axis. The most general metric admitting a covariantly constant null Killing vector in d dimensions [194] takes the form

$$ds^2 = 2Wu(dv + Kdu + A_{\underline{i}}dx^i) + \tilde{g}_{\underline{i}\,\underline{j}}dx^i dx^j, \tag{10.4}$$

where $i, j = 1, 2, \ldots, d-2$ and the vector (*Sagnac connection* [440]) $A_{\underline{i}}$ and the metric $\tilde{g}_{\underline{i}\,\underline{j}}$ in the transverse space do not depend on v. The connection and curvature for this metric are given in Appendix F.2.5. It is possible to eliminate either K or the $A_{\underline{i}}$s by performing a GCT $(u, v, x^i) \to (u, v', x^{i\,\prime})$ that preserves the above form of the metric. Under

$$x^i = x^i(u, x'), \qquad v = v' + f(u, x'), \tag{10.5}$$

we obtain a metric of the same form but with

$$\begin{aligned}
A'_{\underline{i}} &= A_{\underline{j}} M^{\underline{j}}{}_{\underline{i}} + \tilde{g}_{\underline{k}\,\underline{j}}\partial_{\underline{u}}x^k M^{\underline{j}}{}_{\underline{i}} + \frac{\partial f}{\partial x^{i\,\prime}},\\
K' &= K + A_{\underline{i}}\partial_{\underline{u}}x^i + \tfrac{1}{2}\tilde{g}_{\underline{i}\,\underline{j}}\partial_{\underline{u}}x^i\partial_{\underline{u}}x^j + \partial_{\underline{u}} f,\\
\tilde{g}'_{\underline{i}\,\underline{j}} &= \tilde{g}_{\underline{k}\underline{l}} M^{\underline{k}}{}_{\underline{i}} M^{\underline{l}}{}_{\underline{j}},\\
M^{\underline{i}}{}_{\underline{j}} &\equiv \frac{\partial x^j}{\partial x^{i\,\prime}}.
\end{aligned} \tag{10.6}$$

It is now possible to solve the equation $A'_{\underline{i}} = 0$ with $f = 0$ and the $x^{i\,\prime}$ given by the solutions of the first-order differential equation

$$\partial_{\underline{u}}x^i = -\tilde{g}^{\underline{i}\,\underline{j}}A_{\underline{j}}, \qquad \tilde{g}^{\underline{i}\,\underline{j}}\tilde{g}_{\underline{j}\,\underline{k}} = \delta^{\underline{i}}{}_{\underline{k}}, \tag{10.7}$$

if the matrix $M^{\underline{i}}{}_{\underline{j}}$ can be inverted. The equation $K' = 0$ can also be solved with

$$x^{i\,\prime} = x^i, \qquad \partial_{\underline{u}} f = -K. \tag{10.8}$$

10.1.1 Hpp-waves

A family of pp-waves known as *homogeneous pp-waves* or Hpp-waves was constructed by Cahen and Wallach as *symmetric* (not just homogeneous) Lorentzian spacetimes [201]. Some of these spacetimes (in $d = 4$ [637], $d = 6$ [690], $d = 10$ [159], and $d = 11$ [392, 636]) are *maximally supersymmetric*, as we will explain in Chapter 13, and are, therefore, vacua of the corresponding supersymmetric theory, just as the RB solution is another vacuum of $N = 2, d = 4$ SUGRA. In fact, the maximally supersymmetric Hpp-waves are the *Penrose limits* [495, 764] of RB-type ($\mathrm{AdS}_n \times \mathrm{S}^{d-n}$) vacua, which also occur in $d = 4, 6, 10$, and 11 [158, 160]. This makes them particularly interesting. Here we review their construction following [392] and using Appendix A.

First, we need some definitions: the *Heisenberg algebra* $H(2n+1)$ is the Lie algebra generated by $\{q_i, p_j, V\}$ $i, j = 1, \ldots, n$ with the only non-vanishing Lie brackets

$$[q_i, p_i] = \delta_{ij} V. \tag{10.9}$$

The *Heisenberg algebra* $H(2n+2)$ is the semidirect sum of $H(2n+1)$ and the Lie algebra generated by the automorphism U whose action is determined by the new non-vanishing Lie brackets

$$[U, q_i] = p_i, \qquad [U, p_i] = -q_i. \tag{10.10}$$

In the complex basis

$$\alpha_i = \frac{1}{\sqrt{2}}(q_i + ip_i), \qquad I = iV, \qquad N = -iU, \tag{10.11}$$

the Lie brackets take the form

$$[\alpha_i, \alpha_j^\dagger] = \delta_{ij} I, \qquad [N, \alpha_i] = -\alpha_i, \qquad [N, \alpha_i^\dagger] = +\alpha_i^\dagger, \tag{10.12}$$

in which we recognize N as the *number operator*.

All the Heisenberg algebras are solvable and have a singular Killing metric.[1] V (I) is always central.

The Heisenberg algebras can be *deformed* as follows: let us denote by x_r, $r = 1, \ldots, 2n$ the column vector formed by the q_is and p_is. The Lie brackets can be written in this form:

$$[x_r, x_s] = \eta_{rs} V, \qquad [U, x_r] = \eta_{rs} x_s, \qquad (\eta_{rs}) = \begin{pmatrix} 0 & \mathbb{I}_{n\times n} \\ -\mathbb{I}_{n\times n} & 0 \end{pmatrix}. \tag{10.13}$$

Now, we can define a new (solvable) Lie algebra with brackets

$$[x_r, x_s] = M_{rs} V, \qquad [U, x_r] = N_{rs} x_s, \qquad MN^{\mathrm{T}} - NM^{\mathrm{T}} = 0. \tag{10.14}$$

In some cases, but not always, this algebra is equivalent to the original Heisenberg algebra up to a $\mathrm{GL}(2n)$ transformation.

The $(n+2)$-dimensional Hpp-wave spacetimes are constructed starting from a $(2n+2)$-dimensional algebra of the above form with

$$(M_{rs}) = \begin{pmatrix} 0 & -2A \\ 2A & 0 \end{pmatrix}, \qquad (N_{rs}) = \begin{pmatrix} 0 & \mathbb{I}_{n\times n} \\ 2A & 0 \end{pmatrix}, \qquad A_{ij} = A_{ji}, \tag{10.15}$$

which is inequivalent to the original Heisenberg algebra $H(2n+2)$. In the coset construction $\mathfrak{h}$ will be the Abelian subalgebra generated by the $p_i \equiv M_i$s and its orthogonal complement $\mathfrak{k}$ is generated by $q_i \equiv P_i$, $V \equiv P_v$, and $U \equiv P_u$. $\mathfrak{h}$ and $\mathfrak{k}$ are a symmetric pair.

Using the coset representative

$$u = e^{vP_v} e^{uP_u} e^{x^i P_i}, \tag{10.16}$$

we obtain the 1-forms

$$\begin{aligned} e^u &= -du, & e^i &= -dx^i, \\ e^v &= -(dv + A_{ij} x^i x^j du), & \vartheta^i &= -x^i du. \end{aligned} \tag{10.17}$$

[1] Actually the algebras $H(2n+1)$ are nilpotent, which implies an identically vanishing Killing metric.

To construct an invariant Riemannian metric, we use the H-invariant metric[2] $B_{uv} = +1$, $B_{ij} = +\delta_{ij}$ on $\mathfrak{k}$, and the result is a pp-wave of the form

$$ds^2 = 2du(dv + A_{ij}x^i x^j du) + d\vec{x}_n^{\,2}. \tag{10.18}$$

These Hpp-waves are characterized by the eigenvalues of A. They are invariant under the $(2n+2)$-dimensional Heisenberg group but also under the rotations of the wavefront coordinates that preserve the eigenspaces.

10.2 Four-dimensional pp-wave solutions

In four dimensions it is useful to define complex coordinates on the (plane) wavefront $\xi, \overline{\xi}$,

$$\xi = \frac{1}{\sqrt{2}}(x + iy), \tag{10.19}$$

so, using the fact that any two-dimensional metric is conformally equivalent to flat space, the four-dimensional metric can always be written in the form

$$ds^2 = 2du[dv + K(u, \xi, \bar{\xi})du] - 2P(u, \xi, \bar{\xi})d\xi d\overline{\xi}. \tag{10.20}$$

The Einstein vacuum equations are solved if K is a harmonic function on the wavefront,

$$\partial_\xi \partial_{\bar{\xi}} K = 0, \tag{10.21}$$

and P is a function of u alone, and then we can absorb it into a redefinition of ξ that does not change the form of the metric. The only non-trivial element of the metric in this adapted coordinate system is, therefore, $g_{\underline{uu}} = K(u, \xi, \overline{\xi})$. Observe that this function K has exactly the form of a perturbation of the gravitational field about the vacuum (flat Minkowski space with metric $\eta_{\mu\nu}$) since

$$2dudv - 2d\xi d\overline{\xi} = \eta_{\mu\nu}dx^\mu dx^\nu, \tag{10.22}$$

and the metric Eq. (10.20) can also be written in the form

$$ds^2 = \eta_{\mu\nu}dx^\mu dx^\nu + 2K(u, \xi, \overline{\xi})du^2, \qquad h_{uu} = 2K. \tag{10.23}$$

The most general pp-wave solutions of the four-dimensional Einstein–Maxwell theory Eq. (8.58) are also known (see [640]), and take the form

$$\boxed{\begin{aligned} &ds^2 = 2du(dv + Kdu) - 2d\xi d\overline{\xi}, \\ &F_{\xi u} = \partial_\xi C, \\ &\;K = \mathrm{Re}\, f + \tfrac{1}{4}|C|^2, \qquad \partial_{\bar{\xi}} f = \partial_{\bar{\xi}} C = 0. \end{aligned}} \tag{10.24}$$

[2] This metric has mostly plus signature, because $B_{uv} = +1$, $B_{ij} = -\delta_{ij}$ is not H-invariant. We have to perform Wick rotations to obtain a mostly minus metric.

The specific properties of each pp-wave solution depend on the form of the function K.[3] K has two different terms. The first is independent of the electromagnetic field; only the second depends on it. The first term (the real part of the analytic $f(u,\xi)$) is just any harmonic function $H(u,\vec{x}_2)$ in the wavefront Euclidean two-dimensional space and it provides a purely gravitational solution. It represents a sort of perturbation of the electromagnetic and gravitational background described by the second term of K.

A particularly interesting type of pp-waves is *shock* or *impulse waves* with the first term of K given by

$$K(u,\xi,\bar{\xi}) = \delta(u)K(\xi,\bar{\xi}). \tag{10.25}$$

An example of a gravitational shock wave is provided by the purely gravitational *Aichelburg–Sexl solution* [24]

$$K = H(u,\vec{x}_2) = \delta(u)\ln|\xi|, \tag{10.26}$$

which describes the gravitational field of a massive point-like particle boosted to the speed of light. In [24] this metric was obtained by performing an infinite boost in the direction z to a Schwarzschild black hole. This method for generating impulsive waves also works in (anti-)de Sitter spacetimes [565] using the Schwarzschild–(anti-)de Sitter solution and has also been applied to the Kerr–Newman solution [73, 74, 385, 661] and to Weyl's axisymmetric vacuum solutions.[775].[4] However, in Section 10.3 we will identify d-dimensional Aichelburg–Sexl-type (AS) shock waves as the gravitational field produced by a massless particle moving at the speed of light, checking explicitly that (AS) shock waves satisfy the equations of motion of Einstein's action coupled to a massless particle.

This interpretation will later turn out to be very useful. In Chapter 11 we will be interested in the gravitational field produced by massless particles moving at the speed of light in compact dimensions. These particles appear as massive and charged in the non-compact dimensions and their gravitational field (a charged extreme black hole) can be derived from the massless-particle gravitational field. Then, we will simply have to adapt the AS shock-wave solution to a spacetime with compact dimensions.

Another example, this time with the first term of K vanishing, is provided by a solution with Hpp-wave-type metrics (10.18). A particular case is the *four-dimensional Kowalski–Glikman solution* KG4 [637],

$$\boxed{\begin{aligned} ds^2 &= 2du(dv + \tfrac{1}{8}\lambda^2|\vec{x}_2|^2 du) - d\vec{x}_2^{\,2}, \\ F_{\underline{u1}} &= \lambda, \end{aligned}} \tag{10.27}$$

which is a maximally supersymmetric solution of the $d = 4$ Einstein–Maxwell theory that is the Penrose limit of the RB solution. We will study the (super)symmetries of these vacua in Chapter 13.

Before studying shock wave sources, we consider the higher-dimensional generalization of the pp-wave solutions Eq. (10.24).

[3] A detailed classification and description of metrics of this kind that are solutions of the Einstein–Maxwell equations can be found in [640].

[4] For further results and references on impulse waves see e.g. [774, 865].

10.2.1 Higher-dimensional pp-waves

A general pp-wave solution of the d-dimensional Einstein–Maxwell theory Eq. (8.217) is given by [664]

$$\boxed{\begin{aligned} & ds^2 = 2du(dv + K du) - \tilde{g}_{\underline{i}\,\underline{j}}(\vec{x}_{d-2}) dx^i dx^j, \\ & F_{\underline{ui}} = C_{\underline{i}}, \\ & \tilde{\nabla}^2 K = \tfrac{1}{4}\tilde{C}_{\underline{i}}\tilde{C}^{\underline{i}}, \qquad \tilde{d}C = \tilde{d}^{\star}C = 0, \qquad \tilde{R}_{\underline{i}\,\underline{j}} = 0, \end{aligned}} \tag{10.28}$$

i.e. $C(u, \vec{x}_{d-2})_{\underline{i}} dx^i$ is a harmonic 1-form in the Ricci-flat wavefront space and K satisfies the above differential equation that can be integrated if the Green function for the Laplacian on the wavefront space is known. If the wavefront space is flat, $\tilde{g}_{\underline{i}\,\underline{j}} = -\delta_{\underline{i}\,\underline{j}}$, and we take the $\vec{x}_{d-2}$-independent harmonic 1-form $C_{\underline{i}}(u)$, K is given by

$$K = H(u, \vec{x}_{d-2}) + \tfrac{1}{4} C_{\underline{i}} C^{\underline{i}}(u) M_{ij}(u) x^i x^j, \qquad \mathrm{Tr}(M) = 1, \qquad \partial_{\underline{i}}\partial_{\underline{i}} H = 0. \tag{10.29}$$

Again, K consists of two terms: the first is a harmonic function on the Euclidean wavefront space $H(u, \vec{x}_{d-2})$. This is the part of K that can be related to singular sources (massless particles), as we are going to see in the next section. The second term in K describes the gravitational and electromagnetic background. The solutions with $H = 0$ and $C_{\underline{i}}$ and M_{ij} constant have, again, Hpp-wave metrics:

$$\boxed{\begin{aligned} & ds^2 = 2du(dv + A_{ij} x^i x^j du) - d\vec{x}^{\,2}_{d-2}, \\ & F_{\underline{ui}} = C_{\underline{i}}, \qquad \mathrm{Tr}(A) = \tfrac{1}{4} C_{\underline{i}} C^{\underline{i}}. \end{aligned}} \tag{10.30}$$

One particular case is the KG4 solution Eq. (10.27). Another interesting case is the five-dimensional Kowalski–Glikman solution KG5 [690], which is also maximally supersymmetric in $N = 1, d = 5$ SUGRA [261]:

$$\boxed{\begin{aligned} & ds^2 = 2du\left[dv + \frac{\lambda_5^2}{24}(4z^2 + x^2 + y^2)du\right] - dx^2 - dy^2 - dz^2, \\ & \mathcal{F} = \lambda_5 du \wedge dz. \end{aligned}} \tag{10.31}$$

10.3 Sources: the AS shock wave

We consider a massless particle moving in d-dimensional curved space coupled to the Einstein action for the gravitational field. This coupled system is described by the following action (see Section 7.2, where, in particular, the action for a massless particle Eq. (3.258) was derived) with $c = 1$:

$$S = \frac{1}{16\pi G_{\mathrm{N}}^{(d)}} \int d^d x \sqrt{|g|}\, R - \frac{p}{2} \int d\xi \sqrt{\gamma}\gamma^{-1} g_{\mu\nu}(X) \dot{X}^{\mu} \dot{X}^{\nu}. \tag{10.32}$$

The equations of motion for $g_{\mu\nu}(x)$, $X^{\mu}(\xi)$, and $\gamma(\xi)$ are, respectively,

$$
\begin{aligned}
\frac{16\pi G_{\mathrm{N}}^{(d)}}{\sqrt{|g|}}\frac{\delta S}{\delta g^{\mu\nu}} &= G_{\mu\nu} + \frac{8\pi G_{\mathrm{N}}^{(d)} p}{\sqrt{|g|}}\int d\xi \sqrt{\gamma}\gamma^{-1} g_{\mu\rho}g_{\nu\sigma}\dot{X}^{\rho}\dot{X}^{\sigma}\delta^{(d)}(x-X) = 0, \\
\frac{\gamma^{\frac{1}{2}}}{p} g^{\sigma\rho}\frac{\delta S}{\delta X^{\rho}} &= \ddot{X}^{\sigma} + \Gamma_{\rho\nu}{}^{\sigma}\dot{X}^{\rho}\dot{X}^{\nu} - \frac{d}{d\xi}(\ln\gamma)^{\frac{1}{2}}\dot{X}^{\sigma} = 0, \\
\frac{4\gamma^{\frac{3}{2}}}{p}\frac{\delta S}{\delta\gamma} &= g_{\mu\nu}\dot{X}^{\mu}\dot{X}^{\nu} = 0.
\end{aligned}
\tag{10.33}
$$

Since the particle is massless, it must move at the speed of light (this is the content of the equation of motion of γ). If it moves in the direction of the $x^{d-1} \equiv z$ axis, one can use the light-cone coordinates u and v defined above.

If it moves in the sense of increasing z at the speed of light, its equation of motion is $U(\xi) = 0$. We can set $V(\xi) = \sqrt{2}\,\xi$. Thus, our *Ansatz* for the $X^{\mu}(\xi)$ is

$$
U(\xi) = 0, \qquad V(\xi) = \sqrt{2}\xi, \qquad \vec{X} \equiv (X^{1}, \ldots, X^{d-2}) = \vec{0}. \tag{10.34}
$$

A gravitational wave moves at the speed of light, and thus our *Ansatz* for the spacetime metric is that of a gravitational pp-wave moving in the same direction (i.e. with null Killing vector $\ell_{\mu} = \delta_{\mu u}$ so, in particular, nothing depends on v):

$$
ds^{2} = 2du dv + 2K(u, \vec{x}_{d-2})du^{2} - d\vec{x}_{d-2}^{\,2}, \qquad \vec{x}_{d-2} = \left(x^{1}, \ldots, x^{d-2}\right). \tag{10.35}
$$

Now we plug our *Ansatz* into the equation of motion above. First, we immediately see that the equation for γ is satisfied because $\dot{X}^{\mu} = \sqrt{2}\delta^{\mu}{}_{v}$ and $g_{vv} = 0$. The equation of motion for X^{μ} is also satisfied by taking a constant worldline metric $\gamma = 1$ because $\Gamma_{vv}{}^{\sigma} = 0$.

Only one equation remains to be solved. On substituting our *Ansatz* for the coordinates and γ plus $|g| = 1$ (which holds for the above pp-waves), we find

$$
G_{\mu\nu} + 8\pi G_{\mathrm{N}}^{(d)} p \int d\xi\, \delta_{\mu u}\delta_{\nu u}\delta(u)\delta(v - \sqrt{2}\xi)\delta^{(d-2)}(\vec{x}_{d-2}) = 0. \tag{10.36}
$$

For the pp-wave metric Eq. (10.35) we also have *exactly* (that is, without using any property of the metric apart from the light-like character of ℓ^{μ})

$$
G_{\mu\nu} = -\delta_{\mu u}\delta_{\nu u}\vec{\partial}_{d-2}^{\,2} K(u, \vec{x}_{d-2}). \tag{10.37}
$$

Then, on integrating over ξ and substituting the above result, the Einstein equation reduces to the following equation for $K(u, \vec{x}_{d-2})$:

$$
\vec{\partial}_{d-2}^{\,2} K(u, \vec{x}_{d-2}) = -\sqrt{2}\, 8\pi G_{\mathrm{N}}^{(d)} p \delta(u)\delta^{(d-2)}(\vec{x}_{d-2}). \tag{10.38}
$$

In Chapter 3, Section 3.2.3, we found precisely the same equation and it has the same solution, Eqs. (3.133) and (3.134). Thus, we have found the solution

$$\boxed{\begin{aligned} ds^2 &= 2dudv + 2K(u,\vec{x}_{d-2})du^2 - d\vec{x}^{\,2}_{d-2}, \\ K(u,\vec{x}_{d-2}) &= \frac{\sqrt{2}\,p8\pi G_{\rm N}^{(d)}}{(d-4)\omega_{(d-3)}}\frac{1}{|\vec{x}_{d-2}|^{d-4}}\,\delta(u), \qquad d\geq 5, \\ K(u,\vec{x}_2) &= -\sqrt{2}\,p4G_{\rm N}^{(4)}\ln|\vec{x}_2|\,\delta(u), \qquad d=4. \end{aligned}} \tag{10.39}$$

The $d=4$ solution is the AS shock wave found in [24]. Observe that this solution is exactly the same as that which we obtained in Section 3.2.3 by solving the linear-order theory. There are no higher-order corrections to the first-order solution which is not *renormalized*. This is due to the special structure of the linear solution and can be related to supersymmetry as well.

There is another useful way to rewrite the pp-wave metrics that we have found. Defining the function

$$H \equiv 1 - K, \tag{10.40}$$

the solution takes the form

$$\boxed{\begin{aligned} ds^2 &= H^{-1}dt^2 - H\big[dz - \alpha(H^{-1}-1)dt\big]^2 - d\vec{x}^{\,2}_{d-2}, \qquad \alpha = \pm 1, \\ H &= 1 - \frac{\sqrt{2}\,p8\pi G_{\rm N}^{(d)}}{(d-4)\omega_{(d-3)}}\frac{1}{|\vec{x}_{d-2}|^{d-4}}\,\delta\left[\frac{1}{\sqrt{2}}(t-\alpha z)\right], \qquad d\geq 5, \\ H &= 1 + \sqrt{2}\,p4G_{\rm N}^{(4)}\ln|\vec{x}_2|\,\delta\left[\frac{1}{\sqrt{2}}(t-\alpha z)\right], \qquad d=4, \end{aligned}} \tag{10.41}$$

where we have introduced the constant $\alpha = \pm 1$ to take care of the two possible directions of propagation toward $z = \alpha\infty$.

Had we tried to solve the vacuum Einstein equations with the *Ansatz* Eq. (10.35), we would have arrived at the conclusion that any function K (or H) harmonic in $(d-2)$-dimensional Euclidean space transverse to z provides a solution. Thus, we obtain a family of pp-wave solutions of the form

$$\boxed{\begin{aligned} ds^2 &= H^{-1}dt^2 - H[dz - \alpha(H^{-1}-1)dt]^2 - d\vec{x}^{\,2}_{d-2}, \\ \vec{\partial}^{\,2}_{(d-2)}H &= 0, \qquad \alpha = \pm 1. \end{aligned}} \tag{10.42}$$

11

The Kaluza–Klein black hole

Kaluza [615] and Nordström's [728] original idea/observation that electromagnetism could be seen as part of five-dimensional gravity, combined with Klein's curling up of the fifth dimension in a tiny circle [626], constitutes one of the most fascinating and recurring themes of modern physics. Kaluza–Klein theories[1] are interesting both in their own right (in spite of their failure to produce realistic four-dimensional theories [960], at least when the internal space is a manifold) and because of the usefulness of the techniques of dimensional reduction for treating problems in which the dynamics in one or several directions is irrelevant. We saw an example in Chapter 9, when we related four-dimensional instantons to monopoles.

On the other hand, the effective-field theories of some superstring theories (which are supergravity theories) can be obtained by dimensional reduction of 11-dimensional supergravity, which is the low-energy effective-field theory of (there is no real consensus on this point) M theory or one of its dual versions. In turn, string theory needs to be "compactified" to take a four-dimensional form and, to obtain the four-dimensional low-energy effective actions, one can apply the dimensional-reduction techniques.

Here we want to give a simple overview of the physics of compact dimensions and the techniques used to deal with them (dimensional reduction etc.) in a non-stringy context. We will deal only with the compactification of pure gravity and vector fields, leaving aside compactification in the presence of more general matter fields (including fermions) until Part III. We will also leave aside many subjects such as spontaneous compactification and the issue of constructing realistic Kaluza–Klein theories, which are covered elsewhere [342, 957]. In addition to establishing the basic results, we want to study classical solutions of the original and dimensionally reduced theories and how Kaluza–Klein techniques can be used to generate new solutions of both of them.

This chapter is organized as follows. We first study in Section 11.1 the classical and quantum mechanics of a massless particle in flat spacetime with a compact spacelike dimension.

[1] Reference [45] contains many reprints of the most influential papers on the subject. Two old textbooks that describe the classical Kaluza–Klein theory are [109, 654]. More recent accounts can be found in [162, 799, 887]. Even more recent reviews are [331, 342]. A book that describes the geometrical foundations is [252].

We find that the spectrum consists of an infinite tower of massive states and explain the full spectrum of the compactified theory. Next, we perform the simplest dimensional reduction of pure gravity in $\hat{d}$ dimensions to $d = \hat{d} - 1$ dimensions using Scherk and Schwarz's formalism in Section 11.2. We find the action, equations of motion, and symmetries for the massless fields and study various choices of conformal frame. In Section 11.2.3 we study the ("direct") dimensional reduction of the effective action of a massless particle moving in curved spacetime with one compact dimension using the Scherk–Schwarz formalism. We recover the known results about the spectrum of the KK theory in the following form: the massless $\hat{d}$-dimensional particle effective action reduces to the action of a massive, charged, particle moving in $(\hat{d} - 1)$-dimensional space, with mass and charge proportional to the momentum in the compact direction. In Section 11.2.4 we obtain the S dual of the reduced KK theory by the procedure of Poincaré duality explained in Section 8.7.1. In Section 11.2.5 we reduce the $\hat{d}$-dimensional Einstein–Maxwell action and in Section 11.2.5 the bosonic sector of the $N = 1, d = 5$ SUGRA action Eq. (11.98) (which is a modification of the Einstein–Maxwell action). This will allow us to reduce the solutions of that theory studied earlier.

Once the reduction of theories has been established, in Section 11.3 we study the reduction of particular solutions of the Einstein–Maxwell theory and the "oxidation" of particular solutions of the dimensionally reduced Einstein–Maxwell theory. We will reduce ERN BHs in Section 11.3.1 and the AS shock-wave solution (obtaining in this way the electrically charged KK black hole) in Section 11.3.2, and study the possible reduction of Schwarzschild and non-extreme RN BHs in Section 11.3.3. Finally we will see some examples of the use of KK reduction and oxidation combined with dualities to generate new solutions in Section 11.3.4. In particular, exploiting the four-dimensional S-duality symmetry studied in Section 11.2.4, we will obtain the magnetically charged KK BH that becomes, after oxidation to five dimensions, the (Sorkin–Gross–Perry) KK monopole [483, 860] studied in Chapter 9.

In the remaining sections we give an overview of more general dimensional-reduction techniques: toroidal in Section 11.4, the Scherk–Schwarz generalized dimensional reduction in Section 11.5, and orbifold compactification in Section 11.6.

11.1 Classical and quantum mechanics on $\mathbb{R}^{1,3} \times S^1$

The main idea of all KK theories can be stated as follows.

> *KK principle:* our spacetime may have extra dimensions and spacetime symmetries in those dimensions are seen as internal (gauge) symmetries from the four-dimensional point of view. All symmetries could then be unified.

There are several versions of the extra dimensions (brane-worlds etc.) and here we will consider only the "standard" extra dimensions which are curled up in a very small compact manifold, the simplest case which we are going to study (and the one originally considered by Kaluza and Klein) being a circle. The motion of particles in this dimension should not be observable in the usual sense by (empirically well-established) assumption and that is why it is considered compact and small.

Spacetime symmetries are associated with the graviton. It is, thus, natural to start by studying the classical and quantum kinematics of a free massless particle representing a graviton in flat five-dimensional spacetime with a compact fifth dimension of length equal to $2\pi R_z$ and parametrized by the *periodic coordinate* $x^4 = z$ which takes values in $[0, 2\pi\ell]$,

$$z \sim z + 2\pi\ell, \tag{11.1}$$

that can be seen as the vacuum of the full KK theory just as Minkowski spacetime is the vacuum of GR. ℓ is some fundamental length unit (the Planck length ℓ_{Planck}, in string theory the string length $\ell_s = \sqrt{\alpha'}$ etc.) R_z is a fundamental *datum* defining our KK vacuum spacetime and is the simplest example of a *modulus*. The choice of vacuum in KK theory is, however, arbitrary and one of the main objections to KK theories is that no dynamical mechanisms explaining why one dimension is compact and has the size indicated by the modulus are provided. This is generically known as the *moduli problem*.

The five-dimensional metric of this spacetime is, then, in these coordinates[2]

$$d\hat{s}^2 = \eta_{\mu\nu}dx^\mu dx^\nu - (R_z/\ell)^2\, dz^2. \tag{11.2}$$

We can already see that the assumption that the fifth dimension is compact has an immediate and important consequence: *five-dimensional Poincaré invariance of the KK vacuum is spontaneously broken*,

$$\mathrm{ISO}(1,4) \rightarrow \mathrm{ISO}(1,3) \times \mathrm{U}(1).$$

The five-dimensional Lorentz transformations that mix the compact and non-compact dimensions are not symmetries of the metric (they leave it formally invariant if we set $\ell = R_z$ but they change the periodicity properties of the coordinates). Amongst the five-dimensional Poincaré transformations that do not mix compact and non-compact coordinates, clearly Poincaré transformations in the four non-compact dimensions are a symmetry of the theory and constant shifts in the internal coordinate z are also a U(1) symmetry of the theory. These are the symmetries of the KK vacuum.

The rescalings of the compact coordinate rescale ℓ, but not R_z, unless we choose to ignore the rescaling of the period of z, which is the point of view that is usually adopted. In this case, the rescalings are not a symmetry of the theory because they change the modulus R_z which is part of our definition of the (vacuum of the) theory. This is a *duality transformation* that takes us from one theory to another one (albeit of the same class).

We assume that the kinematics in the fifth dimension are the most straightforward generalization of the four-dimensional ones.[3] Thus, we assume that a free, massless particle

[2] Usually in the literature $\ell = R_z$. We prefer this parametrization which emphasizes the distinction between R_z, which is a physical parameter, and ℓ, the range of z which is unphysical. One could also normalize $\ell = 1/(2\pi)$ but coordinates have dimensions of length and it is useful to keep their dependence on ℓ. In some cases it is easier to take $\ell = R_z$ and we will do so by indicating it explicitly.

[3] It is always implicitly assumed that fundamental constants such as the speed of light c and Planck constant h have the same value in the five-dimensional world and the extra dimension is always taken to be space-like. These assumptions are completely *ad hoc* and should be taken as *minimal* assumptions, although it is known that extra timelike dimensions give fields with kinetic terms with the wrong sign in lower dimensions and this justifies the assumption.

moving in a flat five-dimensional spacetime always satisfies[4]

$$\hat{p}^{\hat{\mu}}\hat{p}_{\hat{\mu}} = 0. \tag{11.3}$$

If we separate the four- and five-dimensional pieces of the above equation, it takes the form of a four-dimensional mass-shell condition:

$$p^{\mu}p_{\mu} = \left(p^z R_z/\ell\right)^2, \tag{11.4}$$

and we see that the momentum in the fifth dimension is "seen" as a four-dimensional mass,

$$M = |\hat{p}^z| R_z/\ell. \tag{11.5}$$

We can now consider the quantum-mechanical side of the problem. A free-particle wave function is a momentum eigenmode

$$\hat{P}_{\hat{\mu}}\hat{\Psi} \equiv -i\hbar\partial_{\hat{\mu}}\hat{\Psi} = \hat{p}_{\hat{\mu}}\hat{\Psi}, \quad \Rightarrow \hat{\Psi} = e^{\frac{i}{\hbar}\hat{p}_{\hat{\mu}}\hat{x}^{\hat{\mu}}} \tag{11.6}$$

with $\hat{p}^2 = 0$. The wave function is supposed to be single-valued (periodic) in the compact dimension. For the above wave function, however, we have

$$\hat{\Psi}(x^{\mu}, z + 2\pi\ell) = e^{-\frac{i}{\hbar}\left(\frac{R_z}{\ell}\right)^2 2\pi\ell\hat{p}^z}\hat{\Psi}(x^{\mu}, z), \tag{11.7}$$

and therefore the momentum in the internal dimension can only take the values

$$\hat{p}^z = n\ell\hbar/R_z^2, \quad \hat{p}_z = n\hbar/\ell, \quad n \in \mathbb{Z}, \tag{11.8}$$

and, on account of Eq. (11.5), the spectrum of four-dimensional masses is given by

$$\boxed{M = \frac{|n|\hbar}{R_z}, \quad n \in \mathbb{Z}} \tag{11.9}$$

This is the first prediction of the KK theory: the five-dimensional graviton momentum modes give rise to a discrete spectrum of massive four-dimensional particles plus some massless ones related to $n = 0$. The mass of these *KK modes* is inversely proportional to the size of the internal dimension. If the size of the internal dimension is of the order of the Planck length, these particles will have masses that are multiples of the Planck mass, which would account for the fact that they are not observed.

Observe that M does not depend on ℓ, but only on the modulus R_z.

Let us now move to field theory and consider a five-dimensional, massless, complex scalar field $\hat{\varphi}$ satisfying the five-dimensional sourceless Klein–Gordon equation

$$\hat{\Box}\hat{\varphi} = 0. \tag{11.10}$$

It is natural to Fourier-expand the field:

$$\hat{\varphi}(\hat{x}) = \sum_{n\in\mathbb{Z}} e^{\frac{inz}{\ell}}\varphi^{(n)}(x). \tag{11.11}$$

[4] As we will always do in this and other chapters, we denote five- or, in general, higher-dimensional objects and indices with a hat. Therefore $(\hat{p}^{\hat{\mu}}) = (p^{\mu}, \hat{p}^z)$ and $(p^{\mu}) = (p^0, p^1, p^2, p^3)$.

Table 11.1. In this table the decomposition of the five-dimensional graviton in four-dimensional fields and the physical spectrum are displayed. As explained in the main text, the three four-dimensional fields $g^{(n)}_{\mu\nu}$, $A^{(n)}_{\mu}$, and $k^{(n)}$ for each n combine via the Higgs mechanism and represent a massive spin-2 particle (massive graviton) with mass $m = |n|/R_z$ which has five degrees of freedom (DOF). There are no massive scalars or vectors in the spectrum.

n	$\hat{d} = 5$	DOF	$d = 4$ fields	DOF	Physical spectrum
0	$\hat{g}^{(0)}_{\hat{\mu}\hat{\nu}}$	5	$g_{\mu\nu}$	2	Graviton $m = 0$
			A_{μ}	2	Vector $m = 0$
			k	1	Scalar $m = 0$
$n \neq 0$	$\hat{g}^{(n)}_{\hat{\mu}\hat{\nu}}$	5	$g^{(n)}_{\mu\nu}$	2	
			$A^{(n)}_{\mu}$	2	Graviton $m = \|n\|/R_z$
			$k^{(n)}$	1	

On substituting into the above equation, we see that each Fourier mode satisfies the Klein–Gordon equation for massive fields ($\hbar = 1$),

$$\left[\Box - (n/R_z)^2\right]\varphi^{(n)}(x) = 0, \tag{11.12}$$

and, therefore, each Fourier mode corresponds to a scalar KK mode. Dimensional reduction amounts to taking the zero mode alone. If $\hat{\varphi}$ is to be interpreted as a "relativistic wave function," this is all we need to know. However, if we want to do field theory, we are interested in the Green function for the Klein–Gordon equation. For instance, for time-independent sources we are interested in the Laplace equation

$$\Delta_{(4)}\hat{\varphi} = \delta^{(4)}(\vec{x}_4), \qquad \vec{x}_4 = (x^1, \ldots, x^4), \tag{11.13}$$

and we want to know which kind of equations it implies for each KK mode and what its solution is. That is, we want to know the harmonic function $H_{\mathbb{R}\times S^1}$ in $\mathbb{R}^3 \times S^1$ and its relation to harmonic functions in $\mathbb{R}^3$. We will deal with this problem in Appendix G.

The same analysis cannot be naively applied to the five-dimensional metric field $\hat{g}_{\hat{\mu}\hat{\nu}}$. The Fourier modes of a five-dimensional scalar field can be interpreted as scalar fields in four dimensions, but the Fourier modes of the five-dimensional metric cannot be interpreted as four-dimensional metrics because they are 5×5 matrices. The same applies to vector or spinor fields. We have to decompose the fields with respect to the four-dimensional Poincaré group.

For the graviton, the result is represented schematically in Table 11.1. Let us first focus on the Fourier zero mode, which is a 5×5 symmetric matrix. It can be decomposed (in several ways) into a 4×4 symmetric matrix that can be interpreted as the four-dimensional metric (graviton), a four-dimensional vector, and a scalar. We will see in detail in Section 11.2 how this four-dimensional massless mode of the five-dimensional graviton $\hat{g}^{(0)}_{\hat{\mu}\hat{\nu}}$ (five helicity states) can be decomposed into one massless graviton $g_{\mu\nu}$ (two helicity states), one massless vector A_{μ} (two helicity states), which we will call a *KK vector*, and one massless scalar k

(one helicity state), which we will call a *KK scalar*. The number of helicity states (degrees of freedom) is conserved in this decomposition. The massless spectrum is, thus,

$$\{g_{\mu\nu}, A_\mu, k\}, \tag{11.14}$$

and its symmetries are the local version of symmetries of the KK vacuum determined by the metric Eq. (11.2) plus a vanishing vacuum expectation value for the vector field $\langle A_\mu \rangle = 0$, i.e. four-dimensional GCTs times local U(1) whose gauge field is A_μ.

The infinite tower of four-dimensional massive modes is constituted by spin-2 particles (*massive gravitons*) [829]. They appear as interacting massless[5] gravitons, vectors, and scalars labeled by an integer:

$$\{g_{\mu\nu}^{(n)}, A_\mu^{(n)}, k^{(n)}\}. \tag{11.15}$$

As for the $n \neq 0$ modes, we will see that these fields are related by an infinite symmetry group that contains the Virasoro group [324]. These symmetries are spontaneously broken in the above KK vacuum, and the fields $A_\mu^{(n)}$ and $k^{(n)}$ are the corresponding Goldstone bosons. Owing to the Higgs mechanism, a massless vector and scalar are "eaten" by each massless graviton, giving rise to the massive gravitons [238, 239, 324]. Observe that the number of helicity states is also preserved.[6]

A brief and approximate description of how the Higgs mechanism works in this case is worth giving. Some of the symmetries acting on the $n \neq 0$ sector are *massive gauge transformations*, which include shifts of the scalars $k^{(n)}$ by arbitrary functions that are also standard gauge parameters for the vectors ${A_\mu}^{(n)}$ and shifts of the vectors ${A_\mu}^{(n)}$ by arbitrary vectors that are standard gauge transformations for the $g_{\mu\nu}^{(n)}$s. This means that the gauge-invariant field strengths of the scalars and vectors have, very roughly, the structure

$$\partial_\mu k^{(n)} + nA_\mu^{(n)}, \qquad \partial_\mu A_\nu^{(n)} + ng_{\mu\nu}^{(n)}. \tag{11.16}$$

---

[5] Strictly speaking, one cannot speak about the mass of these fields since, due to the interactions, neither of them is a mass eigenstate [238, 239]. By massless here we simply mean that they enjoy gauge invariances analogous to those of the massless fields.

[6] More generally, in $\hat{d}$ dimensions the graviton (spin 2) has $\hat{d}(\hat{d}-3)/2$ helicity states and a massless $(p+1)$-form potential has $(\hat{d}-2)!/[(p+1)!(\hat{d}-p-3)!]$ helicity states. In particular, a spin-1 particle (vector, $p=0$) has $\hat{d}-2$ and a spin-0 particle (scalar $p=-1$) always has one. A *massive graviton* (spin-2 particle) has $\hat{d}(\hat{d}-1)/2-1$ helicity states and a massive $(p+1)$-form potential has $(\hat{d}-1)!/[(p+1)!(\hat{d}-p-2)!]$ helicity states. In particular, a massive spin-1 particle (a massive vector, $p=0$) has $\hat{d}-1$ helicity states and a massive spin-0 particle (a massive scalar, $p=-1$) has just one. Thus, just on the basis of counting helicity states, the $\hat{d}$-dimensional graviton can always be decomposed into a $(\hat{d}-1)$-dimensional massless graviton, vector, and scalar, and, if the interactions allow it, via the Higgs mechanism, these massless particles can combine into a $(\hat{d}-1)$-dimensional *massive graviton*, which has the same number of helicity states as the massless $\hat{d}$-dimensional one. Analogously, a massless $\hat{d}$-dimensional $(p+1)$-form potential gives rise to massless $(\hat{d}-1)$-dimensional $(p+1)$- and p-form potentials. If the interactions allow it, these two potentials can combine via the Higgs mechanism into a $(\hat{d}-1)$-dimensional massive $(p+1)$-form potential that has the same number of helicity states as the massless $\hat{d}$-dimensional one. Since invariance under GCTs is (see Appendix 3.2) nothing but the gauge symmetry of the massless spin-2 particle, the theory of the *massive graviton* cannot have it. However, in the description of the *massive graviton* as a coupled system of massless graviton, vector, and scalar field, it is possible to have invariance under GCTs that is spontaneously broken by the Higgs mechanism.

These field strengths appear squared in the action. Using the massive gauge transformations, $k^{(n)}$ and $A_\mu{}^{(n)}$ can be gauged away, leaving mass terms for the $g^{(n)}_{\mu\nu}$s. Thus, $k^{(n)}$ and $A_\mu{}^{(n)}$ play the role of Stückelberg fields, like the scalar that one can introduce in massive electrodynamics to preserve a (formal) gauge invariance (see Section 8.2.2). More examples of massive gauge transformations can be found in Section 11.5.

In the above vacuum, the masslessness of the KK scalar is associated with this being the Goldstone boson of dilatations of the compact coordinate (under which it scales).

In the full KK theory ("compactification") all modes should be taken into account. More often, though, all massive modes (all KK modes) are ignored and only the massless spectrum is kept. This is equivalent to ignoring all dynamics in the internal dimensions and it is called *dimensional reduction*. This is the only consistent truncation of the full theory. It is, on the other hand, the effective theory which describes the low-energy behavior of the full theory and contains a good deal of information about the full theory. In particular, the massive modes reappear in it as solitonic solutions: extreme electrically charged KK BHs. This non-trivial fact makes the truncated action even more interesting.

In the "decompactification" limit $R_z \to \infty$ the difference between the masses of the nth and $(n+1)$th modes goes to zero and the spectrum becomes continuous, just like the usual momentum spectrum in a non-compact direction.

To complete our description of the KK spectrum, we should mention that, as we will see later, the KK modes also carry electric charge with respect to the massless KK vector field A_μ. However (as with the details on the spectrum that we have just given), this cannot be seen in flat spacetime. In fact, now we see only that they have a certain rest mass. We know that the gravitational field will couple to it, and we know this even if we do not introduce the gravitational field. However, we can see the electric charge only in the presence of an electromagnetic field. Both the gravitational field and the electric field originate from the five-dimensional gravitational field, which we have not included so far. We will show this in Section 11.2 and we will show that KK modes carry electric charge with respect to this field in Section 11.2.3.

To end this section, let us mention that it has been argued that the KK vacuum is quantum-mechanically unstable [959].

11.2 KK dimensional reduction on a circle S^1

In this section we are going to perform the dimensional reduction of $\hat{d}$-dimensional gravity to $d \equiv \hat{d} - 1$ dimensions in the formalism developed by Scherk and Schwarz in [836]. Thus, here we are going to consider only the massless modes of the graviton field, which by definition do not depend on the compact ("internal") spacelike coordinate $\hat{x}^{\hat{d}-1}$ which we denote by z and which is periodically identified with period $2\pi\ell$, where ℓ is some fundamental length in the theory, and we are going to see how the graviton field splits into d-dimensional fields.

At this point we would like to stress that, in KK theory, the use of distinguished coordinates is unavoidable: up to constant shifts, there is only one coordinate z that is periodic with period $2\pi\ell$ and the Fourier mode expansion has to be done with respect to that coordinate. The metric zero mode is defined by the fact that it does not depend on that coordinate.

Furthermore, technically, the dimensional-reduction procedure requires that we use the coordinate z.

Our starting point, therefore, is a $\hat{d}$-dimensional[7] metric $\hat{g}_{\hat{\mu}\hat{\nu}}$ independent of z.

It is sometimes convenient to give a coordinate-independent characterization of the metrics we are going to deal with. These are metrics admitting a spacelike Killing vector $\hat{k}^{\hat{\mu}}$. If the metric admits the Killing vector $\hat{k}^{\hat{\mu}}$ then its Lie derivative with respect to it vanishes:

$$\mathcal{L}_{\hat{k}}\hat{g}_{\hat{\mu}\hat{\nu}} = 2\hat{\nabla}_{(\hat{\mu}}\hat{k}_{\hat{\nu})} = 0 \tag{11.17}$$

(this is just the Killing equation, see Section 1.5) and this is the condition we would impose on other fields, if we had them.

To this local condition we have to add a global condition: that the integral curves of the Killing vector are closed. z will be the coordinate parametrizing those integral curves (the "adapted coordinate") and it can be rescaled to make it have period $2\pi\ell$. This global condition will not be explicitly used in most of what follows, but only it guarantees consistency. In adapted coordinates $\hat{k}^{\hat{\mu}} = \delta_{\underline{z}}{}^{\hat{\mu}}$.

It is reasonable to think of the hypersurfaces orthogonal to the Killing vector as the d-dimensional spacetime of the lower-dimensional theory. Then, the first object of interest is the metric induced on them. This is

$$\hat{\Pi}_{\hat{\mu}\hat{\nu}} \equiv \hat{g}_{\hat{\mu}\hat{\nu}} + k^{-2}\hat{k}_{\hat{\mu}}\hat{k}_{\hat{\nu}}, \qquad k^2 \equiv -\hat{k}^{\hat{\mu}}\hat{k}_{\hat{\mu}}. \tag{11.18}$$

$\hat{\Pi}^{\hat{\mu}}{}_{\hat{\nu}} = \hat{g}^{\hat{\mu}\hat{\rho}}\hat{\Pi}_{\hat{\rho}\hat{\nu}}$ can be used to project onto directions orthogonal to the Killing vector and $-k^{-2}\hat{k}^{\hat{\mu}}\hat{k}_{\hat{\nu}}$ to project onto directions parallel to it. In adapted coordinates, due to the orthogonality of $\hat{\Pi}$ and $\hat{k}$, we have

$$k = |\hat{k}^{\hat{\mu}}\hat{k}_{\mu}|^{\frac{1}{2}} = |\hat{g}_{\underline{zz}}|, \qquad \hat{\Pi}_{\hat{\mu}\underline{z}} = 0. \tag{11.19}$$

The remaining components define the $(\hat{d}-1)$-dimensional metric

$$g_{\mu\nu} \equiv \hat{\Pi}_{\mu\nu}. \tag{11.20}$$

To understand why this is the right definition of the $(\hat{d}-1)$-dimensional metric instead of just $\hat{g}_{\mu\nu}$ (apart from the reason to do with orthogonality to the Killing vector), we need to examine the effect of $\hat{d}$-dimensional GCTs on it. Under the infinitesimal GCTs $\delta_{\hat{\epsilon}}\,\hat{x}^{\hat{\mu}} = \hat{\epsilon}^{\hat{\mu}}(\hat{x})$, the $\hat{d}$-dimensional metric transforms as follows:

$$\delta_{\hat{\epsilon}}\,\hat{g}_{\hat{\mu}\hat{\nu}} = -\hat{\epsilon}^{\hat{\lambda}}\partial_{\hat{\lambda}}\hat{g}_{\hat{\mu}\hat{\nu}} - 2\hat{g}_{\hat{\lambda}(\hat{\mu}}\partial_{\hat{\nu})}\hat{\epsilon}^{\hat{\lambda}}. \tag{11.21}$$

For the moment, we are interested only in $\hat{d}$-dimensional GCTs that respect the KK *Ansatz*, i.e. that do not introduce any dependence on the internal coordinate z. These fall into two classes: those with infinitesimal generator $\hat{\epsilon}^{\hat{\mu}}$ independent of z and those generated by a z-dependent $\hat{\epsilon}^{\hat{\mu}}$. The latter act only on z and they are found to be only

$$\delta z = az, \qquad a \in \mathbb{R}, \tag{11.22}$$

[7] All $\hat{d}$-dimensional objects carry a hat, whereas $d = (\hat{d}-1)$-dimensional ones do not. The $\hat{d}$-dimensional indices split as follows: $\hat{\mu} = (\mu, \underline{z})$ (curved) and $\hat{a} = (a, z)$ (tangent-space indices).

which can be integrated to give global rescalings plus shifts of the coordinate z:

$$z' = az + b, \qquad a, b \in \mathbb{R}. \tag{11.23}$$

The former can be projected onto the directions orthogonal or parallel to the Killing vector. In orthogonal directions they are just $(\hat{d}-1)$-dimensional GCTs,

$$\delta_\epsilon x^\mu = \epsilon^\mu, \qquad \epsilon^\mu = \hat{\Pi}^\mu{}_{\hat{\nu}} \hat{\epsilon}^{\hat{\nu}} = \hat{\epsilon}^\mu. \tag{11.24}$$

In parallel directions they act only on z,

$$\delta_\Lambda z = -\Lambda, \qquad \Lambda = k^{-2} \hat{k}_{\hat{\nu}} \hat{\epsilon}^{\hat{\nu}} = \hat{\epsilon}^{\underline{z}}, \tag{11.25}$$

which must correspond to some local internal symmetry of the lower-dimensional theory. As we argued before, the $\hat{d}$-dimensional metric is going to give rise to the massless $(\hat{d}-1)$-dimensional fields (11.14). These fields should have good transformation properties under this internal symmetry. In particular, the metric must be invariant under it and the vector must transform under it in the standard way (because it is massless):

$$\delta_\Lambda A_\mu = \partial_\mu \Lambda. \tag{11.26}$$

Observe that the periodicity of Λ has to be the same as the periodicity of z, in order for it to be a well-defined coordinate transformation. We know that the period of the U(1) gauge parameters is related to the unit of electric charge, and we will see that this is also the case in KK theories.

Using the above transformation law for the various components of the $\hat{d}$-dimensional metric, we arrive at the conclusion that the lower-dimensional fields are the following natural combinations of them:

$$g_{\mu\nu} = \hat{g}_{\mu\nu} - \hat{g}_{\underline{z}\mu} \hat{g}_{\underline{z}\nu} / \hat{g}_{\underline{zz}}, \qquad A_\mu = \hat{g}_{\mu\underline{z}} / \hat{g}_{\underline{zz}}, \qquad k = |\hat{g}_{\underline{zz}}|^{\frac{1}{2}} = |\hat{k}^{\hat{\mu}} \hat{k}_{\hat{\mu}}|^{\frac{1}{2}}. \tag{11.27}$$

Equivalently, we can say that the higher-dimensional metric decomposes as follows:

$$\hat{g}_{\mu\nu} = g_{\mu\nu} - k^2 A_\mu A_\nu, \qquad \hat{g}_{\mu\underline{z}} = -k^2 A_\mu, \qquad \hat{g}_{\underline{zz}} = -k^2. \tag{11.28}$$

Furthermore, under the global transformations of the internal space Eq. (11.23), the metric is invariant and only A_μ and k transform. The shifts of z have no effect on them and we are left with a multiplicative $\mathbb{R}$ duality group that can be split according to $\mathbb{R} = \mathbb{R}^+ \times \mathbb{Z}_2$. Only $\mathbb{R}^+$ acts on k,

$$A'_\mu = a A_\mu, \qquad k' = a^{-1} k, \qquad a \in \mathbb{R}^+, \tag{11.29}$$

and only A_μ transforms under the $\mathbb{Z}_2$ factor,

$$A'_\mu = -A_\mu. \tag{11.30}$$

It is a general rule that, in dimensional reductions, global internal transformations give rise to non-compact global symmetries of the lower-dimensional-theory action which generally rescale and/or rotate the fields among themselves. In particular, they act on scalars, and thus scalars naturally parametrize a σ-model. In this case k parametrizes a σ-model with target space $\mathbb{R}^+$. As we explained before, these transformations should not be understood as symmetries but as *dualities* relating different theories.[8]

Observe that, in Section 11.1, the radius of the compact dimension R_z appeared explicitly in the metric. In curved spacetime and at each point of the lower-dimensional spacetime we can define a local radius of the compact dimension $R_z(x)$,

$$2\pi R_z(x^\mu) = \int_0^{2\pi\ell} dz|\hat{g}_{\underline{zz}}|^{\frac{1}{2}} = \int_0^{2\pi\ell} kdz. \tag{11.31}$$

Thus, we see that the KK scalar measures the local size of the internal dimension. We should require that, asymptotically, our five-dimensional metric approaches that of the vacuum Eq. (11.2). Then, we find the following relation among the modulus R_z, the fundamental scale length ℓ, and the asymptotic value of the KK scalar k_0:

$$\boxed{R_z = \ell k_0, \qquad k_0 = \lim_{r\to\infty} k.} \tag{11.32}$$

Sometimes the word modulus is used for the full scalar k. However, only its value at infinity, which we will see is not determined by the equations of motion and thus has to be set by hand as a datum defining the theory, really deserves that name.

Since masses are measured at infinity and, in KK theory, we know that these depend on the radius of the compact dimension through Eq. (11.9), we expect that the masses will depend on the value at infinity of the radius of the compact dimension R_z (which is why we have used the same symbol to denote them).

11.2.1 The Scherk–Schwarz formalism

Having determined the relations Eqs. (11.28) and (11.27) between the lower- and higher-dimensional fields, one can simply plug them into the equations of motion of the higher-dimensional fields (here just Einstein's equations) and obtain equations for the lower-dimensional ones. This procedure automatically ensures that any field configuration that solves the lower-dimensional equations of motion also solves (when it is translated to higher-dimensional fields) the higher-dimensional equations of motion.

In this way one can see that it is not correct to set the KK scalar to a constant as was usually done in the very early KK literature. As was first realized in [886], the KK scalar has a non-trivial equation of motion, which we will find later, and, if one sets it to a constant, this equation of motion transforms into a constraint for the vector-field strength. This constraint is not generically satisfied and, therefore, solutions with $k = k_0$ that do not satisfy this constraint are not solutions of the original theory.

[8] Observe that ℓ is fixed. Dualities change R_z only.

As a general rule, one cannot naively truncate actions by setting some fields to specific values. Doing this in the equations of motion (the correct procedure) would leave us with constraints that must be satisfied and cannot be obtained from the truncated actions. In other words, one cannot reproduce all the truncated equations of motion from a truncated action.

When will a truncation in the action be consistent? Also as a general rule, if there is a discrete symmetry in the action, eliminating only the fields which are not invariant under it will always be consistent. From this point if view, since there is no discrete symmetry acting on k, the inconsistency of its elimination is not surprising. On the other hand, there is a $\mathbb{Z}_2$ symmetry that acts only on A_μ and it is easy to see that it is consistent to eliminate only this field. For instance, this truncation is used to obtain $N = 1, d = 10$ supergravity from $N = 1, \hat{d} = 11$ supergravity (or the heterotic string from M theory) and can be related to dimensional reduction over the orbifold $\mathrm{S}^1/\mathbb{Z}_2$ (a segment of a line, with two boundaries) instead of on the circle S^1.

Performing the dimensional reduction on the equations of motion is in general a quite lengthy calculation (which we will nevertheless perform in Section 11.5). Furthermore, the above decomposition of higher-dimensional fields into lower-dimensional ones cannot be used in the presence of fermions.

In [836] Scherk and Schwarz described a systematic procedure for performing the dimensional reduction in the action and using the Vielbein formalism so it can also be applied to fermions. Another advantage of using Vielbeins is that we can work with objects that have only Lorentz indices and are, therefore, scalars under GCTs. Since some of the GCTs become internal gauge transformations, those objects are automatically GCT-scalars and gauge-invariant.

The first thing to do is to reexpress the relations Eqs. (11.27) and (11.28) in terms of Vielbeins. Using local Lorentz rotations, one can always choose an upper-triangular Vielbein basis of the form

$$\left(\hat{e}_{\hat{\mu}}{}^{\hat{a}}\right) = \begin{pmatrix} e_\mu{}^a & kA_\mu \\ 0 & k \end{pmatrix}, \qquad \left(\hat{e}_{\hat{a}}{}^{\hat{\mu}}\right) = \begin{pmatrix} e_a{}^\mu & -A_a \\ 0 & k^{-1} \end{pmatrix}, \tag{11.33}$$

where $A_a = e_a{}^\mu A_\mu$ and we will assume that all d-dimensional fields with Lorentz indices have been contracted with the d-dimensional Vielbeins.

This choice of Vielbein basis breaks the $\hat{d}$-dimensional local Lorentz invariance to the $d = (\hat{d} - 1)$-dimensional one, which is the subgroup that preserves our choice. If there were other symmetries (such as supersymmetry) acting on the Vielbeins, we would have to add to them compensating Lorentz transformations in order to preserve the choice of Vielbeins.

Next, we find the non-vanishing components of $\hat{\Omega}_{\hat{a}\hat{b}\hat{c}}$,

$$\hat{\Omega}_{abc} = \Omega_{abc}, \qquad \hat{\Omega}_{abz} = -\tfrac{1}{2}kF_{ab}, \qquad \hat{\Omega}_{azz} = -\tfrac{1}{2}\partial_a \ln k, \tag{11.34}$$

where

$$F_{ab} = e_a{}^\mu e_b{}^\nu F_{\mu\nu}, \qquad F_{\mu\nu} = 2\partial_{[\mu}A_{\nu]}, \tag{11.35}$$

is the vector-field strength. With these we find the non-vanishing components of the spin connection $\hat{\omega}_{\hat{a}\hat{b}\hat{c}}$:

$$\begin{aligned} \hat{\omega}_{abc} &= \omega_{abc}, & \hat{\omega}_{abz} &= \tfrac{1}{2} k F_{ab}, \\ \hat{\omega}_{zbc} &= -\tfrac{1}{2} k F_{bc}, & \hat{\omega}_{zbz} &= -\partial_b \ln k. \end{aligned} \tag{11.36}$$

Now, instead of calculating the Ricci scalar, which involves derivatives of the spin connection, we use the following simplifying trick: we first eliminate the derivatives of the spin connection from the action by integration by parts. The result is known as the *Palatini identity* and it is derived in Appendix D for a more general case. On plugging then the above results plus

$$\sqrt{|\hat{g}|} = \sqrt{|g|}\, k \tag{11.37}$$

into the $\hat{d}$-dimensional Palatini identity Eq. (D.4) with $K = 1$, we immediately find that the $\hat{d}$-dimensional Einstein–Hilbert action can be reexpressed, up to total derivatives, in $(\hat{d} - 1)$-dimensional language as follows:

$$\begin{aligned} \int d^{\hat{d}}\hat{x}\, \sqrt{|\hat{g}|}\, \hat{R} = \int dz \int d^{\hat{d}-1}x\, \sqrt{|g|}\, k\, \big\{ & -\omega_b{}^{ba}\omega_c{}^c{}_a - \omega_a{}^{bc}\omega_{bc}{}^a \\ & + 2\omega_b{}^{ba}\partial_a \ln k - \tfrac{1}{4} k^2 F^2 \big\}. \end{aligned} \tag{11.38}$$

Nothing depends on the internal coordinate z and we can integrate over it, obtaining a factor of $2\pi\ell$. Using now "backwards" the $(\hat{d} - 1)$-dimensional Palatini identity with $K = k$, we find at last

$$\boxed{\hat{S} = \frac{1}{16\pi G_N^{(\hat{d})}} \int d^{\hat{d}}\hat{x}\, \sqrt{|\hat{g}|}\, \hat{R} = \frac{2\pi\ell}{16\pi G_N^{(\hat{d})}} \int d^{\hat{d}-1}x \sqrt{|g|}\, k \big[R - \tfrac{1}{4} k^2 F^2 \big].} \tag{11.39}$$

This result is correct up to total derivatives (the ones ignored in applying the Palatini identity). In particular, let us stress that there was not a scalar $\hat{K}$ as in Eq. (4.43) in the original action, because objects that were total derivatives in the previous case would not be so in this case, and in the various integrations by parts factors of $\partial \hat{K}$ would be picked up. These factors are taken into account in the generalized Palatini identity Eq. (D.4). We will often deal with this kind of Lagrangian in Part III.

Another important point is to realize that this action rescales under the global rescalings Eq. (11.29). This happens, though, only because we have chosen to ignore the effect of the rescalings on the period of z. On taking that effect into account, the action would be a scalar, as is the original action.

The KK scalar appears in a strange way because it does not seem to have a kinetic term, so one would say that it has no dynamics. However, one has to remember that, in deriving the Einstein equations of motion, one has to integrate several times by parts. In these integrations, derivatives of k are picked up and one can see that k has standard equations of motion that are implicit in Einstein's.

The equations of motion are[9]

$$\frac{16\pi G_{\rm N}^{(\hat{d})}}{2\pi\ell}\frac{1}{k\sqrt{|g|}}\frac{\delta S}{\delta g^{\alpha\beta}} = G_{\alpha\beta} + \left[\partial_\alpha \ln k\, \partial_\beta \ln k - g_{\alpha\beta}(\partial \ln k)^2\right] \\ + \left[\nabla_\alpha \partial_\beta \ln k - g_{\alpha\beta}\nabla^2 \ln k\right] - \tfrac{1}{2}k^2\left[F_\alpha{}^\mu F_{\beta\mu} - \tfrac{1}{4}g_{\alpha\beta}F^2\right] = 0, \tag{11.40}$$

$$\frac{16\pi G_{\rm N}^{(\hat{d})}}{2\pi\ell}\frac{1}{\sqrt{|g|}}\frac{\delta S}{\delta k} = R - \tfrac{3}{4}k^2 F^2 = 0, \tag{11.41}$$

$$\frac{16\pi G_{\rm N}^{(\hat{d})}}{2\pi\ell}\frac{1}{\sqrt{|g|}}\frac{\delta S}{\delta A_\alpha} = \nabla_\beta\left(k^3 F^{\beta\alpha}\right) = 0. \tag{11.42}$$

On combining the KK scalar equation with the trace of the Einstein equation, we find a standard equation of motion for k,

$$\nabla^2 k = -\frac{\hat{d}-2}{4}k^3 F^2. \tag{11.43}$$

Setting $k = k_0$ is consistent only if $F^2 = 0$, which is not true in general. As we explained before, the KK scalar cannot be simply ignored, as was first realized in [886]. The truncation $A_\mu = 0$ is, nevertheless, consistent.

Another way to see that the KK scalar is dynamical is to rescale the metric to the so-called *Einstein conformal frame*. By definition, this frame is the one in which the Einstein–Hilbert action has the standard form (without the factor of k). The rescaled metric is the *Einstein metric* $g_{{\rm E}\,\mu\nu}$. In the context of Jordan–Brans–Dicke theories, the metric $g_{\mu\nu}$ is sometimes called the *Jordan metric*, but we will call it the *KK metric* and we will refer to the corresponding conformal frame as the *KK conformal frame*.

Using the formulae of Appendix E, we find that the conformal factor is[10] (for $\hat{d} \neq 3$)

$$\Omega = k^{\frac{-1}{d-2}}, \qquad g_{\mu\nu} = k^{\frac{-2}{d-2}} g_{{\rm E}\,\mu\nu}, \tag{11.44}$$

and with it we obtain

$$\boxed{S_{\rm E} = \frac{2\pi\ell}{16\pi G_{\rm N}^{(d+1)}}\int d^d x\sqrt{|g_{\rm E}|}\left[R_{\rm E} + \frac{d-1}{d-2}k^{-2}(\partial k)^2 - \tfrac{1}{4}k^{2\frac{d-1}{d-2}}F^2\right].} \tag{11.45}$$

This action is *not* invariant under the global rescalings Eq. (11.29) because the Einstein metric also rescales under them. Rather, it rescales by a global factor that could be absorbed into the rescaling of ℓ (which we have chosen not to do).

However, we can combine these rescalings with a rescaling of the $\hat{d}$-dimensional metric that rescales the $\hat{d}$- and d-dimensional actions in such a way that the Einstein metric is

[9] See Section 4.2 for a detailed derivation of Einstein's equations in the presence of an overall scalar factor.

[10] We replace $\hat{d} - 1$ by d to avoid confusion, since we are going to use these actions very often.

invariant and only the KK scalar and vector field rescale:

$$k' = ck, \qquad A'_\mu = c^{-\frac{d-1}{d-2}} A_\mu, \qquad c \in \mathbb{R}^+. \tag{11.46}$$

The KK action in the Einstein frame exhibits manifest invariance under these global rescalings which are, together with the $\mathbb{Z}_2$ transformations, the duality group of the theory. This is a standard feature of KK and supergravity theories in the Einstein frame: they are manifestly invariant under duality symmetries. In particular, the scalars that appear in these theories parametrize some σ-model. In this case, the kinetic term for k is the $\mathbb{R}^+$ σ-model. It is sometimes convenient to use a scalar with a standard kinetic term φ,

$$k = e^{\pm\sqrt{2\frac{d-2}{d-1}}\varphi}, \tag{11.47}$$

in terms of which the action takes the form

$$\boxed{S_{\rm E} = \frac{2\pi\ell}{16\pi G_{\rm N}^{(d+1)}} \int d^dx \sqrt{|g_{\rm E}|} \left[R_{\rm E} + 2(\partial\varphi)^2 - \tfrac{1}{4} e^{\pm 2\sqrt{2\frac{d-1}{d-2}}\varphi} F^2 \right].} \tag{11.48}$$

φ transforms under the global rescalings Eq. (11.46) by constant shifts,

$$\varphi' = \varphi \pm \sqrt{\frac{d-1}{2(d-2)}} \ln c. \tag{11.49}$$

The redefinition of the field above is just a change of variables. φ parametrizes $\mathbb{R}$. The two group manifolds are isomorphic, one as a multiplicative group and the other as an additive group.

Owing to its behavior under dilatations, the KK scalar is sometimes called the *dilaton*. We reserve this name for the string-theory dilaton. However, in Section 16.1 we will see that the KK scalar one obtains in the reduction of $\hat{d} = 11$ supergravity to $N = 2A, d = 10$ supergravity can be interpreted as the type-IIA string-theory dilaton.

In fact, the action Eq. (11.48) is an example of the general class of actions described by the "a-model" whose action Eq. (12.1) depends on a continuous parameter a. In this case

$$a = \pm\sqrt{\frac{2d-1}{d-2}}. \tag{11.50}$$

In Chapter 12 we will find BH-type solutions of the a-model for any value a and here we will simply use those results for the specific value of a given above.

11.2.2 Newton's constant and masses

In the presence of gravity, masses are measured at infinity in asymptotically flat spacetimes. When one dimension is compact, one can speak only about asymptotic flatness in the non-compact directions.[11] In particular, the diagonal component of the metric in the compact

[11] The definition of mass in spacetimes with compact dimensions has also been discussed in [164, 309].

dimension does not have to go to -1 at infinity but can be any real negative number. If the metric is asymptotically flat in the non-compact directions then the dimensionally reduced metric (assuming that the compact dimension is isometric) will be asymptotically flat in the KK conformal frame and the value of k at infinity will be some positive real number k_0.

When we rescale the metric to go to the Einstein conformal frame, the metric does not look asymptotically flat any longer, but

$$\lim_{r\to\infty} g_{\mathrm{E}\ \mu\nu} = k_0^{\frac{2}{d-2}} \eta_{\mu\nu}, \tag{11.51}$$

and a change of coordinates is necessary:

$$x^\mu \to x'^\mu = k_0^{\frac{1}{d-2}} x^\mu, \qquad \Rightarrow g_{\mathrm{E}\ \mu\nu} \quad \to g'_{\mathrm{E}\ \mu\nu} = k_0^{-\frac{2}{d-2}} g_{\mathrm{E}\ \mu\nu} \quad \stackrel{r'\to\infty}{\longrightarrow} \eta_{\mu\nu}. \tag{11.52}$$

Thus, if we start with $\hat{d}$-dimensional metrics that are asymptotically flat in the non-compact dimensions, we are forced to perform a rescaling of the coordinates, which is, at the very least, quite unusual. Of course, this change of coordinates, does not modify the action Eq. (11.45).

We could have decided to start with $\hat{d}$-dimensional metrics, which naturally lead to asymptotically flat Einstein metrics with no need for changes of coordinates, but this looks rather artificial.

As we pointed out before, a very interesting aspect of the massless sector of the KK theory is that the truncated massive modes reappear as solitonic solutions. A further problem of the standard Einstein conformal frame is that the masses one finds for solitons are not the ones expected in the spectrum of Kaluza–Klein theories. We are going to check this explicitly in Section 11.2.3.

The prescription we have used to go to the Einstein frame is not canonical, though. We just wanted to eliminate the unconventional (local) factor of k in front of the curvature scalar and the conformal factor that does the job is unique only up to an overall constant factor. In particular, we could have rescaled the KK metric by the factor $\tilde{\Omega} = (k/k_0)^{-\frac{1}{d-2}}$ which defines the *modified Einstein conformal frame*

$$g_{\mu\nu} = (k/k_0)^{-\frac{2}{d-2}}\, \tilde{g}_{\mathrm{E}\ \mu\nu}. \tag{11.53}$$

One of the main characteristics of this metric is that it is invariant under the scale transformations Eq. (11.29). It is appropriate to use with it fields that are also invariant under those rescalings:

$$\tilde{A}_\mu = k_0 A_\mu, \qquad \tilde{k} = k/k_0. \tag{11.54}$$

In terms of these scale-invariant fields, the action takes the form

$$\boxed{\tilde{S}_{\mathrm{E}} = \frac{2\pi \ell k_0}{16\pi G_{\mathrm{N}}^{(\hat{d})}} \int d^d x \sqrt{|\tilde{g}_{\mathrm{E}}|} \left[\tilde{R}_{\mathrm{E}} + \frac{d-1}{d-2} \tilde{k}^{-2} \left(\partial \tilde{k}\right)^2 - \tfrac{1}{4} \tilde{k}^{2\frac{d-1}{d-2}} \tilde{F}^2 \right],} \tag{11.55}$$

which is identical to the action in the original "Einstein frame" Eq. (11.45) except for the overall factor.

This is the frame that leads to correct results.[12] The main difference is the overall factor k_0 which modifies the effective value of the d-dimensional Newton constant which is given by (recall Eq. (11.32))

$$\boxed{G_N^{(d)} = \frac{G_N^{(\hat{d})}}{2\pi R_z} = \frac{G_N^{(\hat{d})}}{V_z}.} \tag{11.56}$$

Here V_z stands for the volume of the compact dimension.

Now, in d dimensions, in the Einstein frame, with the action normalized,

$$S = \frac{1}{16\pi G_N^{(d)}} \int d^d x \sqrt{|g_E|}\, R_E, \tag{11.57}$$

the mass M_E of a given asymptotically flat solution can be read off from $g_{E\,tt}$:

$$g_{E\,tt} \sim 1 - \frac{16\pi G_N^{(d)} M_E}{(d-2)\omega_{(d-2)}} \frac{1}{r^{d-3}}. \tag{11.58}$$

This definition can be used to find the mass in the modified Einstein frame, which we denote by M, or in the Einstein frame (after rescaling the coordinates so the metric is asymptotically flat), which we denote by M_E. The relation between these two masses for the same spacetime can easily be computed:

$$\begin{aligned} g'_{E\,t't'} &\sim 1 - \frac{16\pi G_N^{(\hat{d})} M_E}{2\pi\ell(d-2)\omega_{(d-2)}} \frac{1}{r'^{\,d-3}}, \\ \tilde{g}_{E\,tt} &\sim 1 - \frac{16\pi G_N^{(\hat{d})} M}{2\pi\ell k_0(d-2)\omega_{(d-2)}} \frac{1}{r^{d-3}}, \end{aligned} \tag{11.59}$$

and, using the relation between primed and unprimed coordinates, we find

$$\boxed{M_E = k_0^{-\frac{1}{d-2}} M.} \tag{11.60}$$

It is also handy to have the definition of the electric and magnetic charges $\tilde{q}$ and $\tilde{p}$ to which the scale-invariant KK vector $\tilde{A}_\mu$ couples. To define the charge, we first find the Noether current associated with U(1) gauge transformations in the modified Einstein-frame action:

$$\tilde{j}_N^\mu = \frac{1}{16\pi G_N^{(d)}} \nabla_\nu \left[\tilde{k}^{2\frac{d-1}{d-2}} \tilde{F}^{\nu\mu}\right]. \tag{11.61}$$

---

[12] The names "Einstein frame" and "modified Einstein frame" are a bit confusing and we keep them just because they are standard names in the literature. Both are Einstein frames in the sense that there is no scalar factor in the action in front of the Ricci scalar. However, there is an infinite number of conformal frames with that property, related by constant rescalings. Among that infinite number there is only one in which we recover what we knew about the spectrum: the "modified Einstein frame" which is related to the asymptotically flat $\hat{d}$-dimensional metric by a conformal factor that goes to 1 at infinity. The "Einstein frame" is just the simplest rescaling.

We define

$$\tilde{q} = \int_{\mathrm{B}_{(d-1)}} d\Sigma_\mu \tilde{j}^\mu_{\mathrm{N}}, \tag{11.62}$$

where B_{d-1} is a $(d-1)$-dimensional $t = $ constant hypersurface with boundary $\partial\mathrm{B}_{d-1} = \mathrm{S}^{d-2}$ at infinity. Using Stokes' theorem, we end up with the following definition of electric charge, which we write together with the definition of magnetic charge:

$$\boxed{\tilde{q} = \frac{1}{16\pi G_{\mathrm{N}}^{(d)}} \int_{\mathrm{S}^{d-2}_\infty} \tilde{k}^{2\frac{d-1}{d-2}} {}^\star\tilde{F}, \qquad \tilde{p} = -\int_{\mathrm{S}^2_\infty} \tilde{F}.} \tag{11.63}$$

These charges have the right normalization and so the Dirac quantization condition can be written in terms of them:

$$\tilde{q}\,\tilde{p} = 2\pi n. \tag{11.64}$$

Observe that the period of the gauge parameter of the rescaled vector field $\tilde{A}$,

$$\delta_{\tilde{\Lambda}} z = k_0^{-1}\tilde{\Lambda}, \qquad \delta_{\tilde{\Lambda}} \tilde{A}_\mu = \partial_\mu \tilde{\Lambda}, \tag{11.65}$$

has to be $2\pi R_z$, in agreement with the unit of electric charge $1/R_z$ and Eqs. (8.167) and (8.168).

11.2.3 KK reduction of sources: the massless particle

One of the most interesting things we have learned so far, in several different ways, is that gravitons (or any other massless particles) traveling at the speed of light in the compact dimension look like massive, electrically charged particles in one dimension fewer.

In this section we are going to recover this result in yet another, particularly useful, way. We are going to see that the action for a massless particle moving in a $\hat{d}$-dimensional spacetime, given in Eq. (3.258), becomes that of a massive, charged "K particle" moving in $d = (\hat{d} - 1)$-dimensional spacetime when the $\hat{d}$-dimensional spacetime has an isometry. Furthermore, the mass and electric charge are both proportional to the momentum in the isometric direction and, if we assume that this dimension is compact, we recover exactly the results about the KK spectrum of Section 11.1.

By a "K particle" we mean a slight generalization of the standard massive particle with an extra coupling to a scalar, which we denote generically by K. The Nambu–Goto-type action takes the form

$$\boxed{S = -MK_0^{-1} \int d\xi\, K(X) \sqrt{\left|g_{\mu\nu}\dot{X}^\mu \dot{X}^\nu\right|}.} \tag{11.66}$$

The scalar cannot appear anywhere else. In particular it cannot appear in the Wess–Zumino term which describes the coupling of the particle to an electromagnetic field,

$$WZ = -q \int d\xi\, A_\mu \dot{X}^\mu, \tag{11.67}$$

because that would spoil U(1) gauge invariance. The scalar acts as a sort of local coupling constant. In particular, its presence modifies the mass of the particle, which is no longer the coefficient in front of the action: if the metric is asymptotically flat and K_0 is the constant value at infinity of K, then the mass is the coefficient in front of the action times K_0. We have already taken this into account in writing Eq. (11.66).

KK modes and also string-theory objects called "winding modes" and "D0-branes" that we will study are examples of "K particles." The former couple to the inverse of the KK scalar, i.e. $K = k^{-1}$, as we are immediately going to see. Winding modes couple to k directly, $K = k$, and D0-branes couple to the dilaton $e^{-\phi}$ in string theory.

Although we are going to explain this procedure (called *direct dimensional reduction*) in full detail, it is worth stressing that we are not going to prove that the two actions are completely equivalent. Rather, what we are going to prove is that all the solutions of the first action are of the form of those of the second one for some value of the mass and charge. If we take only one specific value of the mass and charge, we are reducing the system to some sector with a given, fixed, momentum in the internal direction.

Our starting point is the action of a point-like massless particle given in Eq. (3.258), which we rewrite here for convenience:

$$\hat{S}[\hat{X}^{\hat{\mu}}(\xi), \gamma(\xi)] = -\frac{p}{2}\int d\xi\ \gamma^{-\frac{1}{2}}\hat{g}_{\hat{\mu}\hat{\nu}}(\hat{X})\dot{\hat{X}}^{\hat{\mu}}\dot{\hat{X}}^{\hat{\nu}}. \tag{11.68}$$

This action is usually said to be *invariant* under GCTs. In fact it is just *covariant*, since one goes from one metric to a different (even if physically equivalent) one. This happens typically when the action depends on potentials instead of field strengths. The infinitesimal transformations giving $\tilde{\delta}S = 0$ are

$$\begin{aligned}
\tilde{\delta}\hat{X}^{\hat{\mu}} &= \hat{X}'^{\hat{\mu}} - \hat{X}^{\hat{\mu}} &&= \hat{\epsilon}^{\hat{\mu}}(\hat{X}),\\
\tilde{\delta}\hat{g}_{\hat{\mu}\hat{\nu}} &= \hat{g}'_{\hat{\mu}\hat{\nu}}(\hat{X}') - \hat{g}_{\hat{\mu}\hat{\nu}}(\hat{X}) &&= -2\hat{g}_{\hat{\lambda}(\hat{\mu}}\partial_{\hat{\nu})}\hat{\epsilon}^{\hat{\lambda}}.
\end{aligned} \tag{11.69}$$

Let us now consider infinitesimal displacements in the direction $\hat{\epsilon}^{\hat{\mu}}$,

$$\begin{aligned}
\delta_{\hat{\epsilon}}\hat{X}^{\hat{\mu}} &= \hat{\epsilon}^{\hat{\mu}},\\
\delta_{\hat{\epsilon}}\hat{g}_{\hat{\mu}\hat{\nu}} &= \hat{g}_{\hat{\mu}\hat{\nu}}(\hat{X}') - \hat{g}_{\hat{\mu}\hat{\nu}}(\hat{X}) = \hat{\epsilon}^{\hat{\lambda}}\partial_{\hat{\lambda}}\hat{g}_{\hat{\mu}\hat{\nu}}.
\end{aligned} \tag{11.70}$$

Using the formulae in Chapter 1, we find that the change of the action is now

$$\delta_{\hat{\epsilon}}\hat{S} = -\frac{p}{2}\int d\xi\gamma^{-\frac{1}{2}}\left[\mathcal{L}_{\hat{\epsilon}}\,\hat{g}_{\hat{\mu}\hat{\nu}}\right]\dot{\hat{X}}^{\hat{\mu}}\dot{\hat{X}}^{\hat{\nu}}. \tag{11.71}$$

Thus, the action is invariant if and only if $\hat{\epsilon}^{\hat{\mu}} = \hat{\epsilon}\hat{k}^{\hat{\mu}}$, $\hat{\epsilon}$ being an infinitesimal constant parameter and $\hat{k}^{\hat{\mu}}$ being a Killing vector. In other words, if the metric admits an isometry, the above action is invariant under the above symmetry and there is a conserved quantity, namely the momentum in the $\hat{k}^{\hat{\mu}}$ direction:

$$\hat{P} = -p\gamma^{-\frac{1}{2}}\hat{k}_{\hat{\mu}}\dot{\hat{X}}^{\hat{\mu}}, \qquad \dot{\hat{P}} = 0. \tag{11.72}$$

What one would like to do now is to fix the value of this momentum, which completely determines the dynamics in the isometry direction, and find the effective dynamics in the remaining directions. Doing this in a general coordinate system is very complicated (if it is possible at all) and hence we have to work in adapted coordinates as before. We will use, then, all the machinery and notation developed in this section.

In adapted coordinates the fact that there is a conserved momentum becomes evident since the action no longer depends on the isometric coordinate z.

To simplify the problem further, we split the $\hat{d}$-dimensional fields and coordinates in terms of the d-dimensional ones according to Eq. (11.28), obtaining

$$\hat{S}[X^{\mu}(\xi), Z(\xi), \gamma(\xi)] = -\frac{p}{2}\int d\xi\, \gamma^{-\frac{1}{2}}\left[g_{\mu\nu}\dot{X}^{\mu}\dot{X}^{\nu} - k^2 F^2(Z)\right], \tag{11.73}$$

where the combination

$$F(Z) = \dot{Z} + A_{\mu}\dot{X}^{\mu} \tag{11.74}$$

that naturally appears in the action is the "field strength" of the extra worldline scalar Z, which now does not have a coordinate interpretation.

As we explained, the original action (11.68) above is covariant under target-space diffeomorphisms and so must the action (11.73) be, since it is a simple rewriting of the former. In particular, it must be covariant under X^{μ}-dependent shifts of the redundant coordinate Z,

$$\delta_{\Lambda} Z = -\Lambda(X^{\mu}), \tag{11.75}$$

which do not take us out of our choice of coordinates (i.e. coordinates adapted to the isometry) either. As discussed before, these transformations generate gauge transformations of the U(1) gauge potential,

$$\delta_{\Lambda} A_{\mu} = \partial_{\mu}\Lambda. \tag{11.76}$$

The field strength of Z is covariant under this transformation, which justifies its definition.

Related to the constant shifts of Z (which is an invariance) is the conservation of the momentum conjugate to Z,

$$P_z \equiv \frac{\partial\mathcal{L}}{\partial\dot{Z}} = p\gamma^{-\frac{1}{2}}F(Z), \qquad \dot{P}_z = 0. \tag{11.77}$$

Now we want to eliminate Z from the action completely, using its equation of motion ($\dot{P}_z = 0$), and thus obtain the action that governs the effective d-dimensional dynamics. However, we cannot simply substitute into the action Eq. (11.73) $P_z = p\gamma^{\frac{1}{2}}F(Z) =$ constant because from the resulting action one does not obtain the same equations of motion as one would from making the substitution into the equations of motion. The reason for this is that the equation of motion of Z is not algebraic because $\dot{Z}$ occurs in the action.

A consistent procedure by which to eliminate Z is to perform first the Legendre transformation of the Lagrangian with respect to the redundant coordinate Z, just as one would do to find the Hamiltonian if the Lagrangian depended only on Z. We express $\dot{Z}$ in terms of X, $\dot{X}$, and P_z by using the definition of the latter and then define the Legendre transform

$$\mathcal{H}_z(X, \dot{X}, P_z) \equiv -P_z \dot{Z}(X, \dot{X}, P_z) + \mathcal{L}[X, \dot{X}, \dot{Z}(X, \dot{X}, P_z)]. \tag{11.78}$$

After the Legendre transform has been performed, the action that gives the corresponding equations of motion is

$$\begin{aligned}\hat{S}_z[X, \dot{X}, Z, P_z, \gamma] &= \int d\xi \left(-\dot{P}_z Z + \mathcal{H}_z\right) \\ &= -\frac{p}{2}\int d\xi \gamma^{-\frac{1}{2}}\left[g_{\mu\nu}\dot{X}^\mu \dot{X}^\nu + \gamma k^{-2}(P_z/p)^2\right] + \int d\xi\, P_z F(Z).\end{aligned} \tag{11.79}$$

By explicit calculation one can now see that the equation for Z (which now appears explicitly in the first term of the action) is just $\dot{P}_z = 0$ and the equation for P_z is trivially satisfied. Nothing wrong happens, then, on using the equation of motion of Z in the action and replacing the variable P_z by the constant $-p_z$, giving

$$S[X, \gamma] = -\frac{p}{2}\int d\xi \gamma^{-\frac{1}{2}}\left[g_{\mu\nu}\dot{X}^\mu \dot{X}^\nu + \gamma k^{-2}(p_z/p)^2\right] - p_z \int d\xi \left(\dot{Z} + A_\mu \dot{X}^\mu\right). \tag{11.80}$$

Here $\dot{Z}$ still occurs, but in a total derivative term that we can eliminate. Otherwise, we can keep it as an auxiliary scalar, which maintains explicit covariance under gauge transformations. Eliminating this term may give rise to boundary terms under gauge transformations, and thus we prefer to keep it, although it is, admittedly, unusual.

For $p_z \neq 0$, this is the action of a massive charged “K particle” in $(\hat{d} - 1)$-dimensional spacetime. For $p_z = 0$ this is, again, the action of a massless particle moving in a $(\hat{d} - 1)$-dimensional spacetime. To rewrite the $p_z \neq 0$ action in the usual Nambu–Goto form we eliminate γ directly from the action (no derivatives of γ occur in it) by using its equation of motion:

$$\gamma = (p/p_z)^2 k^2 g_{\mu\nu}\dot{X}^\mu \dot{X}^\nu, \tag{11.81}$$

obtaining

$$\boxed{S = -|p_z| \int d\xi k^{-1}\sqrt{\left|g_{\mu\nu}\dot{X}^\mu \dot{X}^\nu\right|} - p_z \int d\xi \left(\dot{Z} + A_\mu \dot{X}^\mu\right),} \tag{11.82}$$

or, ignoring the total derivative and using the scale-invariant (tilded) fields,

$$\boxed{S = -|p_z| k_0^{-1}\int d\xi \tilde{k}^{-1}\sqrt{\left|g_{\mu\nu}\dot{X}^\mu \dot{X}^\nu\right|} - p_z k_0^{-1}\int d\xi\, \tilde{A}_\mu \dot{X}^\mu.} \tag{11.83}$$

This is a remarkable result. For a given momentum in the internal dimension, the massless particle looks like a “K particle” (in fact, a KK mode) with $K = k^{-1}$, mass

$$M = |p_z| k_0^{-1}, \tag{11.84}$$

and charge[13]

$$\tilde{q} = p_z k_0^{-1}. \tag{11.85}$$

The following identity, known as a *Bogomol'nyi identity*, is satisfied:[14]

$$\boxed{M = |\tilde{q}|.} \tag{11.86}$$

This is similar to the identity satisfied by the electric charge and mass of an ERN BH, between the mass and the NUT charge of an extreme Taub–NUT solution, or between the action and the second Chern class of instantons. We will see that this is not a coincidence. In Chapter 13 we will see that all of them are Bogomol'nyi identities (saturated *Bogomol'nyi bounds*) signaling the presence of residual supersymmetries in the background.

If Z is a compact coordinate with period $2\pi\ell$ then the single-valuedness of the wave function implies that the momentum p_z would be quantized,

$$p_z = n/\ell \tag{11.87}$$

(in natural units), and so would the mass and charge of the corresponding KK mode be, as we know. Actually, since $k_0 = R_z/\ell$, we find

$$M = |n|/R_z, \qquad \tilde{q} = n/R_z. \tag{11.88}$$

To finish this section we can try to see how far one can go without assuming that there is an isometry in the direction of the compact coordinate z. Using the split Eq. (11.28), we can equally well arrive at the action (11.73) but now with the fields having periodic dependences on z. Now we should proceed to Fourier-expand all of them. This is not trivial, though, since we do not know how to expand Z because it is not a periodic function of Z (although $\dot{Z}$ is).

11.2.4 Electric–magnetic duality and the KK action

As in the case of the four-dimensional Einstein–Maxwell theory, the four-dimensional KK theory has an electric–magnetic symmetry, but, instead of being a continuous symmetry (at the classical level), it is a discrete $\mathbb{Z}_2$ symmetry. The duality transformation has to be defined very carefully in order to give consistent results. When this is done, the duality can be used to construct new solutions of the *same* theory. In general the duality transformation is not a symmetry, but relates two different theories or different degrees of freedom of the same theory.

We start by performing a Poincaré-duality transformation on the (modified-Einstein-frame) KK action. We remind the reader that the replacement of $\tilde{F}$ by its dual in the action leads in general to an action with the wrong sign for the kinetic term, which does not give rise to the dual equations of motion. This is why one has to follow the Poincaré-duality procedure explained in Section 8.7.1. Only the term involving the vector field in Eq. (11.55)

[13] $q = p_z$ for the untilded A_μ field.
[14] $M = |q|k_0^{-1}$ for the untilded A_μ field.

is important here. We want to replace the vector (1-form) potential $\tilde{A}$ by its dual $(d-3)$-form potential $\tilde{A}_{(d-3)}$ and for this we have to rewrite the action in terms of the 2-form field strength $\tilde{F}$. We need to add a Lagrange-multiplier term to enforce the Bianchi identity and in order to be able to recover the equation $\tilde{F} = d\tilde{A}$. The Lagrange multiplier is the dual potential. The action is, therefore

$$S[\tilde{F}, \tilde{A}_{(d-3)}] = \frac{1}{16\pi G_{\mathrm{N}}^{(d)}} \int d^d x \sqrt{|\tilde{g}_{\mathrm{E}}|} \left[-\tfrac{1}{4} \tilde{k}^{2\frac{d-1}{d-2}} \tilde{F}^2 \right] - \frac{1}{16\pi G_{\mathrm{N}}^{(d)}} \int d^d x \frac{1}{2 \cdot (d-3)!} \epsilon^{\mu_1 \cdots \mu_{d-2} \nu_1 \nu_2} \partial_{\mu_1} \tilde{A}_{(d-3)\ \mu_2 \cdots \mu_{d-2}} \tilde{F}_{\nu_1 \nu_2}. \quad (11.89)$$

This action gives rise to the same equations of motion as does the original action $S[\tilde{A}]$. The equation of motion of $\tilde{F}$ is

$$\tilde{F} = \tilde{k}^{-2\frac{d-1}{d-2}} {}^\star \tilde{F}_{(d-2)}, \qquad \tilde{F}_{(d-2)\mu_1 \cdots \mu_{d-2}} = (d-2) \partial_{[\mu_1} \tilde{A}_{(d-3)\ \mu_2 \cdots \mu_{d-2}]}, \quad (11.90)$$

and, on substituting this into the action, we obtain the dual action, which we rewrite here in full:

$$\boxed{\tilde{S}_{\mathrm{dualE}} = \frac{1}{16\pi G_{\mathrm{N}}^{(d)}} \int d^d x \sqrt{|\tilde{g}_{\mathrm{E}}|} \left[\tilde{R}_{\mathrm{E}} + \frac{d-1}{d-2} \tilde{k}^{-2} (\partial \tilde{k})^2 + \frac{(-1)^{d-3}}{2 \cdot (d-2)!} \tilde{k}^{-2\frac{d-1}{d-2}} \tilde{F}_{(d-2)}^2 \right].} \quad (11.91)$$

This transformation has a chance of being a symmetry of the same theory only in four dimensions. However, even in four dimensions it is not a symmetry because the prefactor of the $\tilde{F}^2$ term was inverted in the transformation.[15] If we interpret the KK scalar as a sort of local coupling constant then we can say that the electric–magnetic-duality transformation relates two different regimes (strong and weak coupling) of the same theory. This can be made explicit by supplementing the electric–magnetic-duality transformation with an inversion of the "coupling constant." This does give us a transformation that leaves invariant the action (via the Poincaré-duality procedure) and the full set of equations of motion (including Bianchi identities),

$$\tilde{F}' = \tilde{k}^{+2\frac{d-1}{d-2}} {}^\star \tilde{F}, \qquad \tilde{k}' = \tilde{k}^{-1}. \quad (11.93)$$

Observe that this transformation does not involve any transformation of the modulus k_0 which defines our theory. (We have stressed several times that a theory is defined also by the expectation values of the moduli, in this asymptotically flat gravitational context by

[15] This is another particular example of the "a-model" action Eq. (12.1) with the opposite value of a,

$$a = \mp \sqrt{\frac{2(d-1)}{d-2}}. \quad (11.92)$$

their constant values at infinity). Thus, we can truly say that the above transformation is a *symmetry* of the theory.

We could have considered similar transformations for the untilded fields. For instance, the following transformation leaves the equations of motion invariant:

$$F' = k_0^{-2}(k/k_0)^{+2\frac{d-1}{d-2}\star}F, \qquad k' = k^{-1}. \tag{11.94}$$

However, this is *not* a symmetry of the theory. The above transformation inverts k_0. If we went back to the $\hat{d}$ theory, we would find that the radius of the compact dimension is inverted and that the $\hat{d}$-dimensional Newton constant does not have the same value.

The transformation Eqs. (11.93) is going to relate electric and magnetic objects in the same theory. If a quantum theory with electrically and magnetically charged states is going to make sense, all the possible pairs of electric and magnetic charges must satisfy the Dirac quantization condition Eq. (8.170). The electric–magnetic-duality symmetry allows us to generate magnetic charges from electric charges and we want the magnetic charges created to be compatible with the original electric charges that we have shown the KK theory to have. We defined the electric and magnetic charges of a solution in Eq. (11.63).

If we start with a field $\tilde{F}$ with electric charge $\tilde{q}$ and perform the electric–magnetic-duality transformation above, we generate the following magnetic charge:

$$\tilde{p}' = -\int_{S^2_\infty} \tilde{F}' = -\int_{S^2_\infty} \tilde{k}^{-2\frac{d-1}{d-2}\star}\tilde{F} = -16\pi G_N^{(4)}\tilde{q}, \tag{11.95}$$

where we have used the definition of $\tilde{q}$ in Eqs. (11.63). Then (ignoring the sign)

$$\tilde{p}'\tilde{q} = 16\pi G_N^{(4)}\tilde{q}^2 = 16\pi G_N^{(4)} n^2/R_z^2, \tag{11.96}$$

on account of Eqs. (11.88) and (11.32). This quantity will be an integer multiple of 2π if

$$\boxed{R_z = \sqrt{8G_N^{(4)}/|m|}, \qquad m \in \mathbb{Z}.} \tag{11.97}$$

The existence of electric–magnetic-duality symmetry (so that each object and its dual can coexist) requires the radius of the internal dimension to be of the order of the Planck length.

Similar constraints on the sizes of the internal dimensions or the values of other moduli can be found in string theory, requiring that each object and its U dual can coexist. A non-trivial check of U duality is that the constraints on moduli obtained from different dual object-pairs are consistent. We will see in Section 19.3, for instance, that the coexistence of all ten-dimensional D-p-branes and their electric–magnetic duals implies the same condition on the value of the ten-dimensional Newton constant.

We can say that, for values of the compactification radius, the theory can undergo a duality transformation into another theory, but, for the "self-dual compactification radius," the theory enjoys an additional symmetry. U duality will become a symmetry for the "self-dual values of the moduli." In this language, there is an enhancement of symmetry at the self-dual point in moduli space, a well-known phenomenon in the context of T duality, in which there is an enhancement of gauge symmetry at the self-dual points.

We should also stress that the electric–magnetic-duality transformation acts on the KK frame and $\hat{d}$ metric in a highly non-trivial way. Also, since it is only a discrete $\mathbb{Z}_2$ transformation even at the classical level, we cannot use it to construct dyonic solutions, although some dyonic solutions can be found.

A final remark: the dual KK action Eq. (11.91) in $d=5$ is identical to the five-dimensional string effective action up to k_0 factors, Eq. (15.13), with the identification $\tilde{k}=e^{\phi}$. Evidently, in the Einstein frame the two actions would be absolutely identical with the identification $k=e^{\phi}$. Then, if we are careful enough with factors of k_0, we can identify any solution of the five-dimensional string effective action involving only the dilaton, the Kalb–Ramond 2-form (these fields are introduced in Part III), and the metric in six-dimensional pure gravity.

11.2.5 Reduction of the Einstein–Maxwell action and $N=1, d=5$ SUGRA

Although the beauty of Kaluza–Klein theories is that they geometrize other interactions, unifying all of them in gravity, it is possible, and sometimes necessary, to introduce other fields in $\hat{d}$ dimensions. For instance, in the compactification of supergravity theories we have to include at the very least all the fields that enter into the supermultiplet in which the graviton lies. In higher dimensions, apart from gravitinos, the minimal supergravity multiplet necessarily contains other fermions plus scalars and k-form fields. In Part III we are going to reduce several of these supergravity theories but now we want to see in a simple example ($N=1, d=5$ SUGRA) how the Scherk–Schwarz formalism works in the presence of matter fields.

In $\hat{d}=5$ the minimal SUGRA [261] has a metric, a vector field, and a pair of symplectic Majorana gravitinos that are associated with eight real supercharges. The action of the bosonic sector is essentially the Einstein–Maxwell action with an extra *topological* (in the sense of metric-independent) cubic *Chern–Simons* term:

$$\boxed{\hat{S}=\int d^5\hat{x}\sqrt{|\hat{g}|}\left[\hat{R}-\tfrac{1}{4}\hat{G}^2+\frac{1}{12\sqrt{3}}\frac{\hat{\epsilon}}{\sqrt{|\hat{g}|}}\hat{G}\hat{G}\hat{V}\right],} \tag{11.98}$$

where $\hat{G}=2\partial\hat{V}$ is the 2-form field strength of the vector $\hat{V}$. The field strength and the action (up to a total derivative) are invariant under the gauge transformations $\delta_{\hat{\chi}}\hat{V}=\partial\hat{\chi}$.

We want to reduce this theory on a circle, but with the same effort we can first perform the reduction of the $\hat{d}$-dimensional Einstein–Maxwell theory (without any topological term) on a circle and then apply the results to our case.

Before we dimensionally reduce the action of the $\hat{d}$-dimensional Einstein–Maxwell theory, it is convenient to know the spectrum of new states that appear when we consider a massless spin-1 particle on a circle. According to general arguments, we expect an infinite tower of states with masses proportional to the inverse of the compactification radius. Furthermore, we know that these massive states will be electrically charged under the massless

KK vector that arises from the metric. On the other hand, we have to take into account that the $\hat{d}$-dimensional vector representation of $\mathrm{SO}(1, \hat{d}-1)$ gives rise to a vector and a scalar of $\mathrm{SO}(1, d-1)$ at each mass level:

$$\hat{V}^{(n)}_{\hat{\mu}} \rightarrow V^{(n)}_{\mu},\ l^{(n)}. \tag{11.99}$$

Our previous experience tells us that, in the $n \neq 0$ levels, the scalars $l^{(n)}$ will act as Stückelberg fields for the vectors $V^{(n)}_{\mu}$, giving rise to the mass terms for them that we expect according to the general KK arguments. For $n = 0$ we obtain a massless vector and a massless scalar, V_{μ} and l. These are the only ones we keep in the dimensional reduction of the theory. The massless scalar is associated with the spontaneous breaking of the $\hat{d}$-dimensional gauge transformations $\delta_{\hat{\chi}} \hat{V}_{\hat{\mu}} = \partial_{\hat{\mu}} \hat{\chi}$ that depend on the coordinate z. In fact, the only z-dependent gauge transformations that preserve the KK *Ansatz* are those linear in z that shift the component $\hat{V}_{\underline{z}}$ and they give rise to a global, non-compact symmetry (duality) of the reduced theory.

Of course, we need to identify the lower-dimensional fields that transform correctly under all the gauge symmetries in order to see all these arguments working. The action is

$$\hat{S}[\hat{g}_{\hat{\mu}\hat{\nu}}, \hat{V}_{\hat{\mu}}] = \frac{1}{16\pi G_{\mathrm{N}}^{(\hat{d})}} \int d^{\hat{d}}\hat{x}\sqrt{|\hat{g}|}\left[\hat{R} - \tfrac{1}{4}\hat{G}^2\right], \tag{11.100}$$

The reduction of the Einstein–Hilbert term goes exactly as before. We need only take care of the Maxwell term. In accord with the Scherk–Schwarz formalism, we use flat indices to identify fields that are invariant under the KK U(1) gauge transformations. Thus, the massless d-dimensional vector field V_{μ} is, using the Vielbein *Ansatz* Eq. (11.33),

$$e_a{}^{\mu} V_{\mu} \equiv \hat{e}_a{}^{\hat{\mu}} \hat{V}_{\hat{\mu}} = \left(\hat{V}_{\mu} - \hat{V}_{\underline{z}} A_{\mu}\right) e_a{}^{\mu} \quad \Rightarrow V_{\mu} = \hat{V}_{\mu} - \hat{V}_{\underline{z}} A_{\mu}. \tag{11.101}$$

The $\hat{V}_{\underline{z}}$ component becomes automatically the d-dimensional massless scalar l, and, thus, we have the decomposition

$$\begin{aligned} \hat{V}_{\underline{z}} &= l, & l &= \hat{V}_{\underline{z}}, \\ \hat{V}_{\mu} &= V_{\mu} + l A_{\mu}. & V_{\mu} &= \hat{V}_{\mu} - \hat{V}_{\underline{z}} \hat{g}_{\mu \underline{z}} / \hat{g}_{\underline{zz}}. \end{aligned} \tag{11.102}$$

It is easy to check that the d-dimensional scalar and vector fields obtained in this way are invariant under the KK U(1) δ_{Λ} transformations. Under the z-independent $\hat{d}$-dimensional transformations, only V_{μ} transforms,

$$\delta_{\chi} V_{\mu} = \partial_{\mu} \chi, \qquad \chi = \hat{\chi}(x), \tag{11.103}$$

and, under the linear gauge transformations $\hat{\chi} = mz$,

$$\delta_m l = m, \qquad \delta_m V_{\mu} = -m A_{\mu}. \tag{11.104}$$

Finally, under the rescalings of the z coordinate that rescale k and A_{μ}, only l transforms:

$$l' = a^{-1} l. \tag{11.105}$$

Now we need to identify the d-dimensional field strength. This is going to be related to $\hat{G}_{ab}$, which is invariant under δ_Λ and $\delta_{\hat{\chi}}$ transformations (including the linear ones δ_m):

$$\hat{G}_{ab} = e_a{}^\mu e_b{}^\nu (2\partial_{[\mu} V_{\nu]} + 2V \partial_{[\mu} A_{\nu]}), \tag{11.106}$$

and we define the gauge-invariant $G_{\mu\nu}$ and the gauge-plus-global-invariant $\mathcal{G}_{ab} = \hat{G}_{ab}$:

$$G_{\mu\nu} = 2\partial_{[\mu} V_{\nu]}, \qquad \mathcal{G} = G + lF. \tag{11.107}$$

On the other hand,

$$\hat{G}_{az} = k^{-1}\partial_a l \tag{11.108}$$

is also invariant under δ_m, and, therefore,

$$\hat{G}^2 = \hat{G}_{ab}\hat{G}^{ab} - 2\hat{G}_{az}\hat{G}^a{}_z = \mathcal{G}^2 - 2k^{-2}(\partial l)^2, \tag{11.109}$$

and the full dimensionally reduced Einstein–Maxwell action is

$$\boxed{\hat{S} = \frac{2\pi\ell}{16\pi G_{\mathrm{N}}^{(\hat{d})}} \int d^{\hat{d}-1}x\sqrt{|g|}\, k\left[R + \tfrac{1}{2}k^{-2}(\partial l)^2 - \tfrac{1}{4}k^2F^2 - \tfrac{1}{4}\mathcal{G}^2\right].} \tag{11.110}$$

Let us now go back to the $\hat{d} = 5$ and let us reduce the Chern–Simons term. First, we convert the Chern–Simons term into an expression with only Lorentz indices,

$$\hat{\epsilon}^{\hat{\mu}_1\cdots\hat{\mu}_5}\hat{G}_{\hat{\mu}_1\hat{\mu}_2}\hat{G}_{\hat{\mu}_3\hat{\mu}_4}\hat{\mathcal{V}}_{\hat{\mu}_5} = \sqrt{|\hat{g}|}\hat{\epsilon}^{\hat{a}_1\cdots\hat{a}_5}\hat{G}_{\hat{a}_1\hat{a}_2}\hat{G}_{\hat{a}_3\hat{a}_4}\hat{\mathcal{V}}_{\hat{a}_5}, \tag{11.111}$$

and then we use the relation

$$\hat{\epsilon}^{abcdz} = \epsilon^{abcd} \tag{11.112}$$

between the five- and four-dimensional Levi-Cività symbols:

$$\sqrt{|\hat{g}|}\hat{\epsilon}\hat{G}\hat{G}\hat{\hat{V}} = k\sqrt{|g|}\epsilon(\hat{G}\hat{G}\hat{\mathcal{V}}_z - 4\hat{G}\hat{G}_z\hat{\mathcal{V}}) = \sqrt{|g|}\epsilon(\mathcal{G}\mathcal{G}l - 4\mathcal{G}\partial l V). \tag{11.113}$$

On turning back to curved indices and integrating by parts, the action takes the form

$$\boxed{\begin{aligned} S = \frac{2\pi\ell}{16\pi G_{\mathrm{N}}^{(\hat{d})}} \int d^4x\sqrt{|g|}\, k\Big\{&R + \tfrac{1}{2}k^{-2}(\partial l)^2 - \tfrac{1}{4}k^2F^2(A) - \tfrac{1}{4}\mathcal{G}^2 \\ &+ \frac{k^{-1}l}{4\sqrt{3}\sqrt{|g|}}\epsilon[\mathcal{G} - 2A\partial l]^2\Big\}. \end{aligned}} \tag{11.114}$$

This theory is a four-dimensional SUGRA theory that is invariant under eight independent local z-independent supersymmetry transformations. Thus, it is an $N = 2, d = 4$

SUGRA theory. Pure $N = 2, d = 4$ SUGRA was described in Section 5.5 and its only bosonic fields are the metric and a vector. Therefore, the extra vector and two scalars that we obtain must be matter fields, actually the bosonic fields of an $N = 2, d = 4$ vector supermultiplet [230]. This reducibility of the gravity supermultiplet after dimensional reduction is a general characteristic of non-maximal SUGRAs. The matter and supergravity vector fields are combinations of the two vectors A and V. To identify them, we can use the fact that eliminating a matter supermultiplet is always a consistent truncation of the theory. The equations of motion of k and l after setting $k = 1$ and $l = 0$ (their truncation values) give the constraints

$$\begin{aligned} 3F^2(A) + F^2(V) &= 0, \\ \sqrt{3}F(A) - {}^\star F(V) &= 0. \end{aligned} \tag{11.115}$$

The second constraint implies the first and tells us that, with $k = 1$ and $l = 0$, the matter vector's field strength is, precisely, the combination[16] $(\sqrt{3}/2)F(A) - \frac{1}{2}{}^\star F(V)$ that has to be set to zero for the truncation of the (matter) scalars to be consistent. The orthogonal combination $\frac{1}{2}{}^\star F(A) - (\sqrt{3}/2)F(V)$ is the supergravity vector field. On setting the matter scalars and vector field to zero, we obtain the action of pure $N = 2, d = 4$ SUGRA (Einstein–Maxwell) with the normalization of Eq. (11.100).

We find the following relations between four-dimensional Einstein–Maxwell fields $g_{\mu\nu}$ and A_μ and five-dimensional fields satisfying the truncation condition:

$$\begin{aligned} \hat{g}_{\underline{z}\underline{z}} &= -1, & \hat{\mathcal{V}}_{\underline{z}} &= 0, \\ 2\partial_{[\mu}\hat{g}_{\nu]\underline{z}} &= -\frac{1}{4\sqrt{|g|}}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}(A), & \hat{\mathcal{V}}_\mu &= -\frac{\sqrt{3}}{2}A_\mu, \\ \hat{g}_{\mu\nu} &= g_{\mu\nu} - \hat{g}_{\mu\underline{z}}\hat{g}_{\nu\underline{z}}, \end{aligned} \tag{11.116}$$

which can be used to uplift *any* $N = 2, d = 4$ SUGRA (Einstein–Maxwell) solution to a $N = 1, d = 5$ SUGRA solution preserving the supersymmetry properties. Similar results can be found in the reduction of the minimal $N = (1, 0), d = 6$ SUGRA (which also has eight supercharges) to $d = 5$ [664] and we will make use of them in Section 13.4 to relate maximally supersymmetric solutions of these three theories by dimensional reduction.

11.3 KK reduction and oxidation of solutions

We have learned how to perform dimensional reduction for the action of pure gravity and the Einstein–Maxwell theory. In particular, we have learned to relate the fields in lower and higher dimensions and the main property of these reductions is that any solution of the lower-dimensional theory is automatically a solution of the higher-dimensional theory that does not depend on one coordinate and the other way around: any solution of the higher-dimensional theory that does not depend on a certain coordinate is a solution of the dimensionally reduced theory, *even if the coordinate is not periodic*.

[16] Actually, its Hodge dual. The supersymmetry transformation rules have to be examined in order to determine these ambiguities [664].

This puts in our hands an incredibly powerful tool for generating new solutions both of the higher- and of the lower-dimensional theories.

The simplest use consists in taking a solution of the higher-dimensional theory that does not depend on one coordinate, which we identify as the compact one, and reducing it using the relations between higher- and lower-dimensional fields; or taking a solution of the lower-dimensional theory and uplifting (*oxidizing*) it to a solution of the higher-dimensional theory.

A more sophisticated use combines reduction and oxidation with a duality transformation of the lower-dimensional solution or a GCT of the higher-dimensional solution.

In this section we are going to see the most important examples of these techniques.

11.3.1 ERN black holes

Periodic arrays and reduction. Let us consider the Einstein–Maxwell theory in $\mathbb{R}^d \times \mathrm{S}^1$. The action is given in Eq. (11.100) and is no different from the action in $\mathbb{R}^{d+1}$ and, thus, the equations of motion admit the same solutions, but now we have to impose different boundary conditions, namely periodicity in the coordinate z. Obviously, solutions that do not depend on the coordinate z are trivially periodic, but we are interested primarily in solutions that do depend on z.

The Einstein–Maxwell theory has MP-type solutions, Eq. (8.229), in any dimension, which depend on a completely arbitrary harmonic function H. Harmonic functions with a point-like singularity that tend to 1 at infinity give asymptotically flat ERN BHs. We can also require the harmonic function to be periodic in the coordinate z in order to obtain an ERN solution in $\mathbb{R}^d \times \mathrm{S}^1$. There is a systematic way to construct a harmonic function periodic in z with a point-like singularity [712] that makes use of the fact that we can construct solutions with an arbitrary number of ERN BHs by taking harmonic functions with that many point-like singularities. The idea is to place an infinite number of ERN BHs with identical masses at regular intervals along the z axis. The corresponding solution is physically equivalent to one with a single ERN BH and a periodic z coordinate. The harmonic function is given by the series

$$H = 1 + h \sum_{n=-\infty}^{n=+\infty} \frac{1}{(|\vec{x}_{\hat{d}-2}|^2 + (z + 2\pi n R_z)^2|)^{\frac{\hat{d}-3}{2}}}, \tag{11.117}$$

where we have assumed for simplicity that $z \in [0, 2\pi R_z]$ and it is (if it converges[17]) a periodic function of z with a pole in $\vec{x}_{\hat{d}-2} = z = 0$ in the interval $[0, 2\pi R_z]$, as we wanted.

Now that we have a solution of the Einstein–Maxwell theory in $\mathbb{R}^d \times \mathrm{S}^1$, we can follow the standard procedure: expand in Fourier series, take the z-independent zero mode, and use the relation between higher- and lower-dimensional fields to obtain a d-dimensional solution of the action Eq. (11.110). For $\hat{d} = 5$, this is done in Appendix G, but for general

[17] It certainly does converge for $\hat{d} = 5$. In fact, this procedure was first developed in [498] in order to obtain harmonic functions on $\mathbb{R}^3 \times \mathrm{S}^1$ and periodic SU(2) instanton solutions using the 't Hooft *Ansatz* Eq. (9.22). Some related calculations can be found in Appendix G.

$\hat{d}$ it is unnecessary to sum the infinite series and then calculate the zero mode [712]: it is possible to approximate the infinite sum by an integral. First we change variables,

$$u_n = \frac{z - 2\pi n R_z}{|\vec{x}_{\hat{d}-2}|}, \qquad u_n \in \left[\frac{2\pi n R_z}{|\vec{x}_{\hat{d}-2}|}, \frac{2\pi(n+1)R_z}{|\vec{x}_{\hat{d}-2}|}\right], \tag{11.118}$$

and we have

$$\begin{aligned} H &= 1 + \frac{h}{|\vec{x}_{\hat{d}-2}|^{\hat{d}-3}} \sum_{n=-\infty}^{n=+\infty} \frac{1}{(1+u_n^2)^{\frac{\hat{d}-3}{2}}} \\ &\sim 1 + \frac{h}{|\vec{x}_{\hat{d}-2}|^{\hat{d}-3}} \frac{1}{\frac{2\pi R_z}{|\vec{x}_{\hat{d}-2}|}} \int_{-\infty}^{+\infty} \frac{du}{(1+u^2)^{\frac{\hat{d}-3}{2}}} = 1 + \frac{h'}{|\vec{x}_{\hat{d}-2}|^{\hat{d}-4}}, \end{aligned} \tag{11.119}$$

with

$$h' = \frac{h\omega_{(\hat{d}-4)}}{2\pi R_z \omega_{(\hat{d}-5)}}. \tag{11.120}$$

It is clear that this approximation is valid if $|\vec{x}_{\hat{d}-2}| \gg R_z$ and for $\hat{d} \geq 5$. For $\hat{d} = 4$ the series does not converge. In fact, defining now

$$u_n = (z - 2\pi n R_z) \in [2\pi n R_z, 2\pi(n+1)R_z], \tag{11.121}$$

we have

$$\begin{aligned} H &= 1 + \frac{h}{|\vec{x}_2|} \sum_{n=-\infty}^{n=+\infty} \frac{1}{(|\vec{x}_2|^2 + u_n^2)^{\frac{1}{2}}} \sim 1 + \frac{h}{2\pi R_z} \int_{-\infty}^{+\infty} \frac{du}{(|\vec{x}_2|^2 + u^2)^{\frac{1}{2}}} \\ &= 1 + \lim_{v \to +\infty} \frac{h}{\pi R_z} \ln\left\{ \frac{v}{|\vec{x}_2|} + \sqrt{1 + \left(\frac{v}{|\vec{x}_2|}\right)^2} \right\} \\ &\sim -\frac{h}{\pi R_z} \ln|\vec{x}_2| + D, \end{aligned} \tag{11.122}$$

where D is a divergent constant. The solution to this problem [712] is to redefine each term in the H series with a constant chosen so as to cancel D out:

$$H = h \sum_{n=-\infty}^{n=+\infty} \frac{1}{(|\vec{x}_2|^2 + (z + 2\pi n R_z)^2|)^{\frac{1}{2}}} - 2h \sum_{n=1}^{n=+\infty} \frac{1}{2\pi n R_z}. \tag{11.123}$$

The solution is not asymptotically flat, but this is to be expected on physical grounds.

These are very useful formulae that we are going to use many times and they deserve to be rewritten and framed. For $n > 1$ and $n = 1$, respectively,

$$\boxed{\begin{aligned} H &= 1 + h \sum_{m=-\infty}^{m=+\infty} \frac{1}{[|\vec{x}_{n+1}|^2 + (z + 2\pi m R_z)^2]^{\frac{n}{2}}} \\ &\sim 1 + \frac{h\omega_{(n-1)}}{2\pi R_z \omega_{(n-2)}} \frac{1}{|\vec{x}_{n+1}|^{n-1}}, \\ H &= h \sum_{m=-\infty}^{m=+\infty} \frac{1}{[|\vec{x}_2|^2 + (z + 2\pi m R_z)^2]^{\frac{1}{2}}} - 2h \sum_{m=1}^{m=+\infty} \frac{1}{2\pi m R_z} \\ &\sim -\frac{h}{\pi R_z} \ln |\vec{x}_2|. \end{aligned}} \tag{11.124}$$

Using this approximated H (the zero mode of the periodic one) in the $\hat{d}$-dimensional MP solution, we obtain a solution that does not depend on the periodic coordinate z and now we can rewrite the solution in terms of the $d = (\hat{d} - 1)$-dimensional fields:[18]

$$\boxed{\begin{aligned} ds^2_{\text{KK}} &= H^{-2} dt^2 - H^{\frac{2}{d-2}} d\vec{x}^{\,2}_{d-1}, \\ ds^2_{\text{E}} &= H^{-\frac{5}{2}} dt^2 - H^{-\frac{d-6}{d-2}} d\vec{x}^{\,2}_{d-1}, \\ V_\mu &= \delta_{\mu t} \alpha (H^{-1} - 1), \qquad \alpha = \pm 2, \\ k &= H^{\frac{1}{d-2}}, \qquad V = V_0. \qquad \partial_{\underline{i}} \partial_{\underline{i}} H = 0, \end{aligned}} \tag{11.125}$$

where we have included a possible constant value for $\hat{V}_{\underline{z}}$. This form is valid for any $\hat{d}$-dimensional MP solution with a z-independent harmonic function, and, in particular, for the above H that corresponds to the zero mode of the $\hat{d}$-dimensional periodic ERN BH.

Why have we gone through the long procedure of finding periodic ERN BH solutions and finding their zero modes when we could simply reduce the whole MP family assuming independence of z? The reason is that, in the cases that will interest us, we will have a well-defined $\hat{d}$-dimensional source that will determine the coefficient h of the $\hat{d}$-dimensional harmonic function and only by going through all this procedure can we relate it to the coefficient of the d-dimensional harmonic function.

The dimensionally reduced ERN solution does not have a regular horizon: near the origin (the only place where the horizon could be placed), using spherical coordinates $r = |\vec{x}_{d-1}|$,

$$\left(1 + h'/r^{d-3}\right)^{-\frac{d-6}{d-2}} r^2 d\Omega^2_{(d-2)} \sim h' r^{\frac{(d-3)(d-6)}{d-2}} d\Omega^2_{(d-2)}, \tag{11.126}$$

[18] Since we have absorbed the asymptotic value of the KK scalar into the period of the coordinate z, $k = \tilde{k}$ and there is no difference between the Einstein and modified Einstein frames.

which never goes to a $(d-2)$-sphere with finite radius. However, we *know* that this solution corresponds to a solution with a regular horizon in $d+1$ dimensions! One possible way to explain what is happening here is the following: the results of the dimensional–reduction procedure are meaningful within certain approximations. In particular, we assume that the massive modes can be ignored because their masses are very large, which means that the compactification radius is small. In this geometry, the compactification radius, measured by the modulus k, is not constant over the space but depends on r, blowing up when $r \to 0$ (the locus of the putative horizon). Thus, near this point, there are KK modes whose masses become small enough to be taken into account, but we have not done this and the solution cannot be considered valid near $r=0$. Near $r=0$ the solution is indeed $\hat{d}$-dimensional and regular. Similar mechanisms have been proposed in other cases and in the context of string theory to show how some singularities disappear when we take into account the higher-dimensional origin of the solution [441].

Oxidation. Dimensional oxidation is in general a much simpler operation than reduction: we simply take a solution of the lower-dimensional theory and rewrite it in terms of the $\hat{d}$-dimensional fields, obtaining a solution of the higher-dimensional theory that does not depend on the compact coordinate. However, this solution may (but need not) be the zero mode of a solution that does depend periodically on the compact coordinate and in general we cannot know which of these possibilities is true.

In any case, the first step consists in having a solution of the lower-dimensional theory and our problem is that the d-dimensional ERN solution (in general, the MP solutions) is not a solution of the dimensionally reduced $\hat{d}=(d+1)$-dimensional Einstein–Maxwell theory. Let us examine the KK scalar equation of motion in the Einstein frame. It takes the form

$$\nabla^2 \ln k \sim c_1 k^{a_1} F^2 + c_2 k^{a_2} \mathcal{G}^2, \tag{11.127}$$

and requires a non-trivial k if $F^2 \neq 0$ or $\mathcal{G} \neq 0$, as is the case here. Thus, the MP solutions cannot, in general, be considered solutions of the reduced Einstein–Maxwell equations and, thus, cannot be dimensionally oxidized.

There are, however, exceptions. For instance

1. Solutions with $F^2 = 0$ satisfy the KK scalar equation of motion and thus can be oxidized to a purely gravitational solution. One example is the dyonic ERN BH with electric and magnetic charges related by $p = \pm 16\pi G_{\mathrm{N}}^{(4)} q$ (see page 330). Another example is provided by electromagnetic pp-waves.

2. We have seen in Section 11.2.5 that any solution of the four-dimensional Einstein–Maxwell theory ($N=2, d=4$ SUGRA) can be oxidized to a solution of $N=2, d=5$ SUGRA using Eqs. (11.116) and we have mentioned that solutions of the latter can be further oxidized to $N=(1,0), d=6$ SUGRA.

Observe that we can oxidize the four-dimensional Einstein–Maxwell solutions with $F^2=0$ in two different ways to $d=5$. The second form makes use of the supersymmetric structure of the theory and ensures that the supersymmetry properties will be preserved in the oxidation, whereas in the first case they will not.

11.3.2 Dimensional reduction of the AS shock wave: the extreme electric KK black hole

Now we are going to consider the dimensional reduction of the AS shock-wave solution Eq. (10.41). We must distinguish between two cases: when the wave propagates in the compact coordinate and when it propagates in an orthogonal direction. The second case is simpler, and we study it first.

To avoid confusion, we are going to call y the direction in which the wave propagates and z the compact direction. The AS solution depends on a harmonic function of the transverse coordinates $H(\vec{x}_{\hat{d}-2})$ and on a delta function $\delta[(1/\sqrt{2})(t-\alpha y)]$. If the compact coordinate is $x^{\hat{d}-2} \equiv z$, we split the transverse-coordinates vector into $\vec{x}_{\hat{d}-2} = (\vec{x}_{d-2}, z)$. We know that any harmonic function H provides a solution and hence we can repeat the construction of a harmonic function of $(\vec{x}_{d-2}, z)$ that has a single point-like pole and is periodic in the coordinate z by constructing a periodic array and taking the average. For $\hat{d} \geq 5$ the reduced solution is another AS shock wave with the same metric in one dimension fewer and with the coefficients of the harmonic functions related as above.

The case in which the wave propagates in the compact direction is far more interesting. We should be able to guess the result, since we have reduced the source of the AS shock wave (the massless point-particle action) in Section 11.2.3 and found the action of a massive KK mode that is electrically charged with respect to the KK vector field and with charge and mass equal to the momentum in the compact direction. We expect, then, that the reduction in the direction in which the wave propagates should give a metric describing a massive, electrically charged object which will be "extreme" in some sense, corresponding to the special relation between its mass and charge.

First, we adapt the solution Eq. (10.41) to the compactness of z, rescaling it to $k_0 z$, and at the same time rescaling ℓ so the periodicity of z is always $2\pi\ell$. These rescalings introduce k_0 factors in several places. The solution we are going to start with is

$$
\begin{aligned}
d\hat{s}^2 &= H^{-1}dt^2 - H\left[k_0 dz - \alpha(H^{-1}-1)dt\right]^2 - d\vec{x}^{\,2}_{\hat{d}-2}, && \alpha = \pm 1,\\
H &= 1 - \frac{\sqrt{2}\, p 8\pi G_{\rm N}^{(\hat{d})}}{(\hat{d}-4)\omega_{(\hat{d}-3)}} \frac{1}{|\vec{x}_{\hat{d}-2}|^{\hat{d}-4}}\, \delta\!\left[\frac{1}{\sqrt{2}}(t-\alpha k_0 z)\right], && \hat{d} \geq 5, \qquad (11.128)\\
H &= 1 + \sqrt{2}\, p 4 G_{\rm N}^{(4)} \ln|\vec{x}_2|\, \delta\!\left[\frac{1}{\sqrt{2}}(t-\alpha k_0 z)\right], && \hat{d} = 4.
\end{aligned}
$$

Before we proceed, it is necessary to identify the constant p. In asymptotically flat cases, p is just the absolute value of the momentum carried by the massless particle. In the present case, the momentum of the massless particle in the z direction is given by (just take the KK vacuum spacetime limit)

$$p_z = \alpha p k_0, \qquad (11.129)$$

and we should replace p by $|p_z|/k_0$ accordingly in the above harmonic functions.

Now, we should Fourier-expand all the components of the metric, but we are going to content ourselves with taking the zero mode, which will be a solution of the KK-theory action Eq. (11.39). We expand

$$\delta\!\left[\frac{1}{\sqrt{2}}(t-\alpha k_0 z)\right] = -\frac{\sqrt{2}}{k_0}\sum_n \frac{1}{2\pi\ell} e^{in\left(z-\frac{t}{k_0}\right)/\ell}, \qquad (11.130)$$

and keep only the zero mode $-1/(\sqrt{2}\pi\ell k_0)$. The replacement of the δ function by its constant zero mode gives us the z-independent harmonic functions and metric, which can be immediately rewritten in terms of d-dimensional fields that we express both in the KK frame and in the modified Einstein frame for the interesting, asymptotically flat $d > 4$ cases:

$$\boxed{\begin{aligned} ds^2_{\mathrm{KK}} &= H^{-1}dt^2 - d\vec{x}^{\,2}_{d-1},\\ d\tilde{s}^2_{\mathrm{E}} &= H^{-\frac{d-3}{d-2}}dt^2 - H^{\frac{1}{d-2}}d\vec{x}^{\,2}_{d-1},\\ \tilde{A}_t &= \alpha(H^{-1}-1), \qquad \tilde{k} = H^{\frac{1}{2}}, \qquad \alpha = \pm 1,\\ H &= 1 + \frac{h}{|\vec{x}_{d-1}|^{d-3}}, \qquad h = \frac{16\pi G_{\mathrm{N}}^{(\hat{d})} p_z}{2\pi\ell k_0^2(d-3)\omega_{(d-2)}}. \end{aligned}} \tag{11.131}$$

This is the d-dimensional *extreme electric KK BH* solution. As expected, it describes a massive, electrically charged object that should be a KK mode. It does not have a regular horizon. It is clear that, had we started from the general family of pp-wave solutions Eqs. (10.42), we would have obtained a family of solutions of the same form but with arbitrary harmonic functions. Thus, we can construct solutions of the KK action Eq. (11.39) with several of these objects with charges of the same sign in static equilibrium by the standard procedure. Now, the equilibrium is more difficult to describe because a third interaction, mediated by the KK scalar k, comes into play. On the other hand, in the reduction of the ERN solution we also found a solution describing a massive object charged with respect to a vector field and with a non-trivial scalar, but different from this one. The reason is that they obey different equations of motion, the difference being the strength with which the KK scalar couples to the vector field. We will study these *dilaton "BHs"* in more detail in Section 12.1.

We can calculate the mass and charge of the above solutions to check that they do indeed correspond to those of a KK mode. From

$$\tilde{g}_{\mathrm{E}\,tt} = H^{-\frac{d-3}{d-2}} \sim 1 - \frac{d-3}{d-2}\frac{h}{|\vec{x}_{d-1}|^{d-3}}, \tag{11.132}$$

and the definition of the mass M,

$$\tilde{g}_{\mathrm{E}\,tt} \sim 1 - \frac{16\pi G_{\mathrm{N}}^{(d)} M}{(d-2)\omega_{(d-2)}}\frac{1}{|\vec{x}_{d-1}|^{d-3}}, \tag{11.133}$$

we find $M = p_z k_0^{-1}$, as expected.

The electric charge can be calculated using the definition in Eqs. (11.63), finding first

$$\tilde{k}^{2\frac{d-1}{d-2}\star}\tilde{F} = \pm(d-3)h\, d\Omega^{d-2}, \tag{11.134}$$

where $d\Omega^{d-2}$ is the unit $(d-2)$-sphere volume form, whose integral over the sphere just gives $\omega_{(d-2)}$ (see Appendix C). The final result is $\tilde{q} = \pm p_z k_0^{-1}$ (p_z was taken to be positive), also as expected.

We conclude that the extreme electric KK BH solution does indeed describe the long-range fields of a KK mode.

The name "extreme BH" for a solution that does not have a regular event horizon needs some justification: the reason is that this solution belongs to a larger family of BH solutions with regular event horizons and also with Cauchy horizons, which we will construct in Section 11.3.4. When the mass and electric charge are equal (the "extreme limit"), the event and Cauchy horizons coincide and become singular. The general families of non-extreme dilaton BHs will be studied in Section 12.1. Those with the right dilaton coupling can be oxidized to one dimension more.

Finally, observe that purely gravitational pp-waves can always be oxidized to one dimension more by taking the product with the metric of a flat line. We know that the dependence of the harmonic functions can be extended to that coordinate. The first observation is also true for any purely gravitational solution, which is always a solution of the KK action Eq. (11.39). However, the dependences of the functions in the metric cannot always be extended to the new compact coordinate. This is the case for the Schwarzschild BH solution, as we are going to see.

11.3.3 Non-extreme Schwarzschild and RN black holes

Dimensional reduction. Paradoxically, the simplest and most fundamental BH solutions are also the most difficult to reduce because it is also more difficult to generalize them to the case in which one coordinate is compact. We certainly cannot construct, in a simple and naive way, infinite periodic arrays of Schwarzschild and non-extreme RN BHs because it is not at all clear how to construct solutions for more than one non-extreme BH, and, on physical grounds, one does not expect them even to exist because the interaction between non-extreme BHs is not balanced and they cannot be in static equilibrium.

Nevertheless, there are solutions describing an arbitrary number of aligned Schwarzschild BHs: the Israel–Khan solutions [595]. They belong to Weyl's family of axisymmetric vacuum solutions [640, 949, 950] and, thus, they have a metric that, in Weyl's canonical coordinates $\{t, \rho, z, \varphi\}$, takes the form

$$ds^2 = e^{2U}dt^2 - e^{-2U}\left[e^{2k}(d\rho^2 + dz^2) + \rho^2 d\varphi^2\right], \tag{11.135}$$

where U is a harmonic function in three-dimensional Euclidean space that is independent of φ (because of axisymmetry) and k depends on U through two first-order differential equations that can be integrated straightaway:

$$\begin{aligned} \partial_{\underline{i}}\partial_{\underline{i}}U &= 0, \\ \partial_\rho k &= \rho[(\partial_\rho U)^2 - (\partial_z U)^2], \\ \partial_z k &= 2\rho\,\partial_\rho U\,\partial_z U. \end{aligned} \tag{11.136}$$

The simplest choice of U is, in spherical coordinates $r^2 = \rho^2 + z^2$,

$$U = -\frac{G_N^{(4)}M}{r}, \qquad k = -\frac{(G_N^{(4)}M)^2\sin^2\theta}{2r^2}, \tag{11.137}$$

and gives the Chazy–Curzon solution [235, 269]. In spite of the spherically symmetric U, the solution is only axisymmetric. The Schwarzschild solution corresponds to a U equal to the Newtonian gravitational potential for an ideal homogeneous rod of finite length $2G_{\mathrm{N}}^{(4)}M$ and total mass M,

$$\begin{aligned} U &= \tfrac{1}{2}\ln\left(\frac{r_+ + z_+}{r_- + z_-}\right) = \tfrac{1}{2}\ln\left(\frac{r_+ + r_- + (z_+ - z_-)}{r_+ + r_- - (z_+ - z_-)}\right), \\ k &= \tfrac{1}{2}\ln\left(\frac{r_+ r_- + z_+ z_- + \rho^2}{2r_+ r_-}\right), \end{aligned} \tag{11.138}$$

where

$$r_\pm \equiv \sqrt{\rho^2 + z_\pm^2}, \qquad z_\pm \equiv z - \left(z_0 \pm G_{\mathrm{N}}^{(4)}M\right), \tag{11.139}$$

and z_0 is the value of z at the center of the rod. The two very different-looking forms of the function U are completely equivalent. The coordinate transformation

$$\rho = \sqrt{r\left(r - 2G_{\mathrm{N}}^{(4)}M\right)}, \qquad z - z_0 = \left(r - 2G_{\mathrm{N}}^{(4)}M\right)\cos\theta, \tag{11.140}$$

gives back the Schwarzschild metric in Schwarzschild coordinates.

This metric is singular at the position of the rod over the z axis ($M > 0$): $\rho = 0$, $z_0 - G_{\mathrm{N}}^{(4)}M < z < z_0 + G_{\mathrm{N}}^{(4)}M$, where U diverges, but the singularity can be removed by a coordinate transformation and indicates only the presence of the event horizon [595]. On the other hand, $k = 0$ on the axis, so there are no conical singularities there, as we are going to explain.

Since U satisfies a linear equation, we can linearly superpose the potentials of N separated rods with masses M_i and lengths $2G_{\mathrm{N}}^{(4)}M_i$ to give a solution that, in principle, can describe several Schwarzschild BHs in static equilibrium. We just have to calculate k:

$$\begin{aligned} U &= \sum_{i=1}^{N} U_i, \quad U_i = \tfrac{1}{2}\ln\left(\frac{r_{+i} + r_{-i} + (z_{+i} - z_{-i})}{r_{+i} + r_{-i} - (z_{+i} - z_{-i})}\right), \\ k &= \sum_{i,j=1}^{N} k_{ij}, \quad k_{ij} = \tfrac{1}{4}\ln\left(\frac{r_{+i} r_{-j} + z_{+i} z_{-j} + \rho^2}{r_{+i} r_{+j} + z_{+i} z_{+j} + \rho^2} + (+ \leftrightarrow -)\right), \end{aligned} \tag{11.141}$$

where now

$$r_{\pm i} \equiv \sqrt{\rho^2 + z_{\pm i}^2}, \qquad z_{\pm i} \equiv z - (z_{0i} \pm G_{\mathrm{N}}^{(4)}M_i), \tag{11.142}$$

and the centers of the rods are at z_{0i}. These are Israel–Khan solutions [595]. Since, physically, we did not expect these solutions to exist, where is the catch? These solutions have additional conical singularities over the z in between the rods (BHs): $e^{\pm 2U}$ is completely regular in between the axes because the U_is are. The k_{ij}s vanish when z is not in between the rods i and j, and, in between the rods i and j (that is, assuming that $z_{0i} < z_{0j}$, $z_{\pm i} > 0$ and $z_{\pm j} < 0$), on taking the $\rho \to 0$ limit carefully, we obtain

$$k_{ij}^0 \equiv \lim_{\rho \to 0} k_{ij} = \tfrac{1}{2}\ln\left|\frac{(z_{+i} - z_{-j})((z_{-i} - z_{+j})}{((z_{+i} - z_{+j})((z_{-i} - z_{-j})}\right|, \tag{11.143}$$

which is constant and proportional to the Newtonian gravitational force between the rods i and j. Thus, in general k will be a constant $k^0 = \sum_{i.j} k^0_{ij}$ that differs from zero over the z axis. This implies the existence of conical singularities over the axis: when $\rho \to 0$ the spatial part of the metric Eq. (11.135) takes the form

$$-ds^2_{(3)} \sim e^{-2(U-k)}\Big[d\rho^2 + dz^2 + \rho^2 e^{-2k^0} d\varphi^2\Big]. \tag{11.144}$$

The metric in brackets is the Euclidean metric in cylindrical coordinates if $e^{-k^0}\varphi$ has period $\Delta e^{-k^0}\varphi = 2\pi$; otherwise, there is a conical singularity with a deficit angle $\delta = 2\pi - \Delta e^{-k^0}\varphi$. However, for analogous reasons, the period of φ has to be 2π if the metric is to be asymptotically flat rather than asymptotically conical and, in general, there is a defect angle $\delta = 2\pi(1 - e^{-k^0})$. For instance, for two rods separated by a coordinate distance Δz (so, $z_{01} - z_{02} = \Delta z + G^{(4)}_{\mathrm{N}}(M_1 + M_2)$) the deficit angle is (see e.g. [257])

$$\delta = -2\pi \frac{4G^{(4)\,2}_{\mathrm{N}} M_1 M_2}{\Delta z[\Delta z + 2G^{(4)}_{\mathrm{N}}(M_1 + M_2)]}. \tag{11.145}$$

This conical singularity can be considered as a *strut* that holds the two BHs in place in spite of their gravitational attraction. The BH horizons are deformed by all these interactions [257]. The conical singularities are unavoidable: it can be shown that the only non-singular solution is the one with a single BH [428, 705]. In fact, in [430] it is shown that the only static, axisymmetric, asymptotically flat solutions with many BHs are the MP solutions. Nevertheless, the Euclidean action is well defined even in the presence of conical singularities [448].

It is clear that the Israel–Khan solution can be used to construct an infinite periodic array of identical Schwarzschild BHs of mass M whose rod centers are separated by a coordinate distance $2\pi R_z$. This construction was made in [712]. The series $U = \sum_{n=-\infty}^{n=\infty} U_n$ diverges (asymptotically, it is similar to the second series H in Eq. (11.124)) and we need to redefine it:

$$U = \sum_{n=-\infty}^{n=\infty} U_n - \sum_{n=1}^{+\infty} \ln\left(\frac{1 - G^{(4)}_{\mathrm{N}} M/(n2\pi R_z)}{1 + G^{(4)}_{\mathrm{N}} M/(n2\pi R_z)}\right). \tag{11.146}$$

The same is true for the k_{nm}s:

$$k = \sum_{n,m=-\infty}^{n,m=+\infty} k_{nm} - \sum_{n,m=0}^{n,m=+\infty} \ln\left[1 - \frac{4G^{(4)\,2}_{\mathrm{N}} M^2}{(n+m+1)^2(2\pi R_z)^2}\right]. \tag{11.147}$$

When the number of BHs is infinite, we expect the total force exerted on each BH by all the others (an infinite number to its left and to its right) to vanish and, indeed, one finds $k^0 = 0$, a total absence of conical singularities in between the BHs.

This solution can now be considered as a Schwarzschild BH with a compact dimension, asymptotically $\mathbb{R}^3 \times \mathrm{S}^1$. We could extract its Fourier zero mode and then dimensionally reduce it to three dimensions using the standard procedure. This construction can be generalized to other non-extreme BHs such as the RN BH with like [257] or opposite charges [374], and the generalization to $d = 5$ dimensions can be performed using the higher-dimensional generalization of the Weyl class recently found in [372], although for higher dimensions there are still problems.

Oxidation: black branes. We have already mentioned that any purely gravitational solution is automatically a solution of the KK action given in Eq. (11.39) with constant KK scalar $k = k_0$ and, therefore, it is also a solution of the higher-dimensional purely gravitational theory. The procedure can be repeated as many times as we want (p, say) and the result is a purely gravitational solution with a metric that is the direct product of the original metric and the metric of p circles (a p-torus T^p).

This remark applies in particular to $d = (\hat{d} - p)$-dimensional Schwarzschild BHs. On oxidizing them to $\hat{d}$ dimensions and adding p coordinates $\vec{y}_p = (y^1, y^2, \ldots, y^p)$ with $y^i \in [0, 2\pi R_i]$, we obtain the following metric [557]:

$$\boxed{\begin{aligned} d\hat{s}^2 &= W dt^2 - d\vec{y}_p^{\,2} - W^{-1} dr^2 - r^2 d\Omega^2_{[\hat{d}-(p+2)]}, \\ W &= 1 - \frac{16\pi G_{\mathrm{N}}^{(\hat{d}-p)} M}{[\hat{d} - (p+2)]\omega_{[\hat{d}-(p+2)]}} \frac{1}{r^{\hat{d}-p-3}}. \end{aligned}} \tag{11.148}$$

These solutions are known as *(Schwarzschild) black p-brane* solutions and represent the gravitational field of massive, extended objects of p spatial dimensions (*p-branes*), which are parametrized by the coordinates $\vec{y}_p$. They are asymptotically flat only in the directions orthogonal (or transverse) to the *worldvolume* directions t and $\vec{y}_p$ (even in the $R_i \to \infty$ limit). Since the mass is the same even for infinite compactification radii, it is clear that these objects are really characterized by a mass density per unit p-volume (in the $\vec{y}_p$ directions), which is called the *brane tension* $T_{(p)}$, rather than by M, which is the mass of the point-like object they give rise to after compactification on T^p. To calculate the tension $T_{(p)}$ we use p times the relation between the Newton constants in different dimensions,

$$G_{\mathrm{N}}^{(\hat{d}-p)} = G_{\mathrm{N}}^{(\hat{d})} / V_p, \qquad V_p = (2\pi)^p R_1 \cdots R_p, \tag{11.149}$$

and define

$$G_{\mathrm{N}}^{(\hat{d}-p)} M = G_{\mathrm{N}}^{(\hat{d})} T_{(p)}, \quad \Rightarrow M = V_p T_{(p)}. \tag{11.150}$$

Solutions like this are going to be studied in detail in Part III.

It is also possible to oxidize to purely gravitational solutions the solutions Eqs. (8.216) of the Einstein-scalar theory, but the resulting solutions, a sort of generalized black p-branes, do not have a clear interpretation.

11.3.4 Simple KK solution-generating techniques

KK oxidation and reduction can be used to generate new solutions. In general, the procedure consists in using a well-defined symmetry of the higher- or lower-dimensional theory as an intermediate step between oxidation and reduction or reduction and oxidation. Let us study some examples.

Generation of charged solutions by higher-dimensional boosts. The first example consists of three steps.

1. Oxidation of the Schwarzschild solution to $\hat{d} = d + 1$ dimensions.
2. Lorentz-boosting the Schwarzschild black 1-brane solution in the compact direction.
3. Reduction in the same direction.

We have already performed the first operation in the previous section. The $\hat{d}$-dimensional solution is

$$d\hat{s}^2 = W dt^2 - dz^2 - W^{-1} dr^2 - r^2 d\Omega^2_{[\hat{d}-3]}, \qquad W = 1 + \frac{\omega}{r^{\hat{d}-4}}, \tag{11.151}$$

and we are ready to perform a Lorentz boost in the positive- or negative-z direction, which evidently transforms a solution of the $\hat{d}$-dimensional Einstein equations into another one:

$$\begin{pmatrix} t \\ z \end{pmatrix} \to \begin{pmatrix} \cosh\gamma & \pm\sinh\gamma \\ \pm\sinh\gamma & \cosh\gamma \end{pmatrix} \begin{pmatrix} t \\ z \end{pmatrix}, \qquad \gamma > 0. \tag{11.152}$$

The new solution can be rewritten in the form

$$\boxed{\begin{array}{l} d\hat{s}^2 = H^{-1} W dt^2 - H\big[dz - \alpha(H^{-1} - 1)dt\big]^2 - W^{-1} dr^2 - r^2 d\Omega^2_{[\hat{d}-3]}, \\ W = 1 + \dfrac{\omega}{r^{\hat{d}-4}}, \qquad H = 1 + \dfrac{h}{r^{\hat{d}-4}}, \qquad \omega = h(1-\alpha^2), \end{array}} \tag{11.153}$$

if we parametrize $\alpha = \pm\coth\gamma$, which is a sort of "black pp-wave" metric. The non-extremality function W disappears when we boost at the speed of light $\alpha = \pm 1$ and then we recover exactly the pp-wave solutions Eq. (10.42), for which H can be any general harmonic function in $(\hat{d} - 2)$-dimensional Euclidean space.

Now, the third step gives a new d-dimensional class of solutions whose existence we announced:

$$\boxed{\begin{array}{l} ds^2_{\text{KK}} = H^{-1} W dt^2 - W^{-1} dr^2 - r^2 d\Omega^2_{(d-2)}, \\ d\tilde{s}^2_{\text{E}} = H^{-\frac{d-3}{d-2}} W dt^2 - H^{\frac{1}{(d-2)}} \Big(W^{-1} dr^2 - r^2 d\Omega^2_{(d-2)}\Big), \\ \tilde{A}_t = \alpha(H^{-1} - 1), \qquad \tilde{k} = H^{\frac{1}{2}}, \\ W = 1 + \dfrac{\omega}{r^{d-3}}, \qquad H = 1 + \dfrac{h}{r^{d-3}}, \qquad \omega = h(1-\alpha^2). \end{array}} \tag{11.154}$$

These are the *non-extreme electric KK BHs*. They have regular event horizons and Cauchy horizons (for negative ω) and, in the extreme limit $\omega = 0$, they become the extreme electric KK BHs, Eq. (11.131).

The same procedure can be used with higher-p branes and also with "charged p-branes."

Lower-dimensional S dualities and generation of KK branes. In this example, we are also going to study a three-step mechanism for generating new solutions that exploits the existence of an S-duality symmetry in the four-dimensional KK theory, as discussed in Section 11.2.4.

1. Reduction of a purely gravitational five-dimensional solution to a four-dimensional KK-theory solution.
2. S dualization of the four-dimensional KK-theory solution.
3. Oxidation of the S-dualized KK-theory solution to a new purely gravitational five-dimensional solution.

In particular, we are going to apply this recipe to the "black pp-wave" solutions given in Eqs. (11.153). Using the transformation Eq. (11.93) on the four-dimensional solution Eq. (11.154), we immediately obtain

$$\boxed{\begin{aligned} ds^2_{\rm KK} &= W dt^2 - H\Big[W^{-1}dr^2 + r^2 d\Omega^2_{(2)}\Big],\\ d\tilde{s}^2_{\rm E} &= H^{-\frac{1}{2}} W dt^2 - H^{\frac{1}{2}}\Big[W^{-1}dr^2 + r^2 d\Omega^2_{(2)}\Big],\\ \tilde{F} &= \alpha h d\Omega^2, \qquad \tilde{k} = H^{-\frac{1}{2}},\\ H &= 1 + \frac{h}{r}, \qquad W = 1 + \frac{\omega}{r}, \qquad \omega = h(1-\alpha^2). \end{aligned}} \tag{11.155}$$

The solution is naturally given in terms of the field strength $\tilde{F}$. Finding the potential is equivalent to solving the Dirac-monopole problem, which we already solved in Section 8.7.2. Here we simply quote the result: in spherical coordinates the non-vanishing components of $\tilde{F}$ are

$$\tilde{F}_{\theta\varphi} = \alpha h \sin\theta = \partial_\theta \tilde{A}_\varphi, \ \ \Rightarrow \tilde{A}_\varphi = -\alpha h \cos\theta, \tag{11.156}$$

up to gauge transformations. This potential is singular at $\theta = 0$ and $\theta = \pi$ and the solution to this problem is to define the potential in two different patches $\tilde{A}^{(\pm)}_\varphi$ related by a gauge transformation:

$$\tilde{A}^{(\pm)}_\varphi = \pm\alpha h(1 \mp \cos\theta). \tag{11.157}$$

It is useful to rewrite the equation that the untilded A has to satisfy (the Dirac-monopole equation) in a coordinate-independent way that will allow us to generalize the solutions,

$$\partial_{[\underline{i}} A_{\underline{j}]} = \alpha k_0^{-1} \tfrac{1}{2} \epsilon_{ijk} \partial_{\underline{k}} H. \tag{11.158}$$

All the properties that depend only on the modified Einstein-frame metric (singularities, horizons, causality, extremality, thermodynamics etc.) are the same as in the electric case. The characteristic features of the magnetic BH appear in the KK frame and in $\hat{d}$ dimensions.

Using the relations Eqs. (11.28), we can immediately find the $\hat{d}$-dimensional metric which gives rise to the fields of the magnetic solution given in Eqs. (11.155):

$$\begin{aligned} d\hat{s}^2 &= W dt^2 - H\Big[W^{-1}dr^2 + r^2 d\Omega^2_{(2)}\Big] - k_0^2 H^{-1} W^{-1}[dz + A]^2, \\ A &= A_{\underline{i}} dx^i, \qquad \partial_{[\underline{i}} A_{\underline{j}]} = \alpha k_0^{-1} \tfrac{1}{2}\epsilon_{ijk}\partial_{\underline{k}} H, \\ H &= 1 + \tfrac{h}{r}, \qquad W = 1 + \tfrac{\omega}{r}, \qquad \omega = h(1-\alpha^2). \end{aligned} \tag{11.159}$$

This solution has no simple interpretation. The extreme $\omega = 0$ case is particularly interesting because the metric becomes a product of time and a non-trivial four-dimensional Euclidean manifold:

$$\boxed{\begin{aligned} d\hat{s}^2 &= dt^2 - H d\vec{x}_3^{\,2} - k_0^2 H^{-1}[dz + A]^2, \\ A &= A_{\underline{i}} dx^i, \qquad \partial_{[\underline{i}} A_{\underline{j}]} = \alpha k_0^{-1} \tfrac{1}{2}\epsilon_{ijk}\partial_{\underline{k}} H, \\ H &= 1 + \frac{|\tilde{p}|}{4\pi}\frac{1}{|\vec{x}_3|}, \qquad \alpha = \pm 1, \end{aligned}} \tag{11.160}$$

where we have identified the constant h in H in terms of the four-dimensional magnetic charge $\tilde{p}$. As usual, in the extreme case the function H can be any harmonic function in flat three-dimensional space. The non-trivial four-dimensional manifold is nothing but the Euclidean Taub–NUT solution[19] Eq. (9.12) up to a rescaling

$$z = k_0^{-1}\tau, \qquad A_{\mathrm{KK}} = k_0^{-1} A_{\mathrm{TN}}. \tag{11.161}$$

This is the reason why the latter is called the (Sorkin–Gross–Perry) *Kaluza–Klein monopole* [483, 860]. We have to identify the magnetic charge and the Taub–NUT charge,

$$|\tilde{p}|/(4\pi) = 2|N|. \tag{11.162}$$

N is related to the period of τ, $8\pi|N|$, which, upon making the identification $\tau = k_0 z$, implies $4|N| = R_z$ and $|\tilde{p}| = 2\pi R_z$, which is consistent with the known quantization of the KK modes' electric charge $\tilde{q} = n/R_z$ and the Dirac quantization condition Eq. (11.64).

Summarizing, we have performed a purely gravitational five-dimensional duality transformation that interchanges momentum in the direction z with (Euclidean) NUT charge. These two purely gravitational charges are seen in four dimensions as electric and magnetic U(1) charges.[20] This mechanism can be used in more general contexts, whenever the dimensionally reduced theory has an S-duality symmetry (see, for instance, [666]).

The KK S-duality symmetry is just a discrete $\mathbb{Z}_2$ transformation and it is natural to wonder whether there are dyonic solutions, even if they cannot be generated by continuous S-duality transformations. There is, to the best of our knowledge, only one dyonic KK BH solution that is also a dyonic ERN BH.

[19] But the solution Eq. (11.159) is not the Euclidean continuation of the non-extreme Taub–NUT solution, which is four-dimensional.

[20] For a discussion of the geometrical NUT charge and its representation as a $(d-3)$-form potential in d dimensions see [578].

The RN–KK dyon. Let us consider the dyonic MP solutions Eqs. (8.204). A quick calculation gives

$$F^2 = 8(\cos^2\alpha - \sin^2\alpha)\partial_{\underline{i}} H^{-1} \partial_{\underline{i}} H^{-1}, \tag{11.163}$$

which vanishes for $\alpha = \pm\pi/4$. For this value of the charges (whose signs we can still change, preserving $F^2 = 0$) the dyonic MP solutions are also solutions of the KK action with constant KK scalar $k = k_0$ (=1 for simplicity) and can be uplifted to a purely gravitational five-dimensional solution [621]:

$$\boxed{\begin{aligned} d\hat{s}^2 &= H^{-2}dt^2 - H^2 d\vec{x}_3^{\,2} - \left[dz + \sqrt{2}\left(\alpha_q H^{-1}dt + \alpha_p A_{\underline{i}}dx^i\right)\right]^2, \\ \partial_{[\underline{i}}A_{\underline{j}]} &= \alpha_p \tfrac{1}{2}\epsilon_{ijk}\partial_{\underline{k}}H, \qquad \partial_{\underline{k}}\partial_{\underline{k}}H = 0, \qquad \alpha_q^2 = \alpha_p^2 = 1, \end{aligned}} \tag{11.164}$$

where α_q and α_p are the possible signs of the electric and magnetic charges.

Skew KK reduction and generation of fluxbranes. Our last example, "skew KK reduction" [326, 327] shows the power of the KK techniques to generate new solutions from "almost nothing." The general setup is the following.[21] Let us consider a metric that admits two isometries, one compact (a U(1)), associated with the coordinate θ, and one non-compact (an $\mathbb{R}$), associated with the coordinate z with a metric of the product form

$$d\hat{s}^2 = -dz^2 - f^2\left[d\theta + f_m dx^m\right]^2 + f_{mn}dx^m dx^n, \tag{11.165}$$

where we have normalized the period of $\theta \in [0, 2\pi]$. We want now to construct a new spacetime by identifying points in the above spacetime according to

$$\begin{pmatrix} z + 2\pi R_z \\ \theta \end{pmatrix} \sim \begin{pmatrix} z \\ \theta - 2\pi B \end{pmatrix}. \tag{11.166}$$

To apply the standard Scherk–Schwarz formalism, we need to use a coordinate independent of z and thus we define a new coordinate θ' adapted to the above identifications,

$$\theta' = \theta - \frac{B}{R_z}z, \tag{11.167}$$

adapted to the Killing vector $R_z\partial_{\underline{z}} - B\partial_{\underline{\theta}}$ and rewrite the metric, adapting it to KK reduction in the direction z. The lower-dimensional fields are

$$\begin{aligned} ds^2_{\mathrm{KK}} &= -\frac{R_z}{B}k^{-2}\left(d\theta' + f_m dx^m\right)^2 + f_{mn}dx^m dx^n, \\ A_{\theta'} &= \frac{B}{R_z}k^{-2}, \qquad A_{\underline{m}} = \frac{B}{R_z}k^{-2}f_m, \\ k^2 &= 1 + \frac{B^2}{R_z^2}f^2. \end{aligned} \tag{11.168}$$

[21] Here we follow [325], where more uses of this technique to construct new solutions can be found.

If we start from flat spacetime in polar coordinates [325],

$$d\hat{s}^2 = -dz^2 - (d\rho^2 + \rho^2 d\theta^2) + dt^2 - d\vec{y}^{\,2}_{(\hat{d}-4)}, \tag{11.169}$$

$f = \rho, \;\; f_m = 0$, and we obtain

$$\boxed{\begin{aligned} &ds^2_{\text{KK}} = dt^2 - d\vec{y}^{\,2}_{(d-3)} - \frac{R_z}{B} k^{-2} d\theta'^2, \\ &A_{\theta'} = \frac{B}{R_z} k^{-2}, \qquad\qquad k^2 = 1 + \frac{B^2}{R_z^2}\rho^2. \end{aligned}} \tag{11.170}$$

This is the *Kaluza–Klein Melvin solution* [325]. It generalizes the original Melvin solution that *describes a parallel bundle of magnetic flux held together by its own gravitational pull* [692]. These solutions are also known as $(d-3)$-fluxbranes since they have a magnetic flux orthogonal to a $(d-3)$-dimensional spacelike submanifold that is invariant under all possible translations, just like the Schwarzschild black p-branes which we saw have a p-dimensional spacelike translation-invariant submanifold.

11.4 Toroidal (Abelian) dimensional reduction

The next simplest case we can consider is a $\hat{d}$-dimensional spacetime that locally (and asymptotically) is the product of d-dimensional Minkowski spacetime and n circles ($\hat{d} = d + n$). The product of n circles is *topologically* an n-torus T^n and this case, which is a trivial generalization of the single-circle case, is called *toroidal compactification*. *Metrically*, the relative sizes and angles of the circles define the torus.

A useful way to characterize tori is the following: a circle of length $2\pi R$ can be considered as a coset manifold, namely the quotient of the group of continuous translations $\mathbb{R}$ by the subgroup of discrete translations of size $2\pi R$, which we can denote by $2\pi R\mathbb{Z}$. Thus, $\mathrm{S}^1 = \mathbb{R}/(2\pi R\mathbb{Z})$.

A torus T^n can be similarly considered as the quotient of the group of n-dimensional translations $\mathbb{R}^n$ by a discrete n-dimensional subgroup called an n-dimensional *lattice*[22] Γ^n, $\mathrm{T}^n = \mathbb{R}^n/\,\Gamma^n$. The information about sizes and angles is evidently contained in Γ^n.

The quotient affects only the global properties of the torus, which locally is just $\mathbb{R}^n$, and therefore it has n independent translational isometries. We can choose n independent mutually commuting Killing vectors in the directions of the lattice generators. The n adapted coordinates z^m that parametrize their integral curves will then be periodic coordinates that can be used simultaneously.

The analysis of the theory in these spaces proceeds along the same lines as in the case of a single compact dimension. First, to find the spectrum, one performs an n-dimensional Fourier-mode expansion in the vacuum. The single zero mode will be the only massless mode and all the other modes will be massive. The massless $\hat{d}$-dimensional graviton mode

[22] A lattice Γ^n is generated by linear combinations with integer coefficients of n linearly independent vectors of $\mathbb{R}^n$, $\{\vec{u}_i\}$, $i = 1, \ldots, n$. Thus, a generic element $\vec{u} \in \Gamma^n$ can be written as $\vec{u} = n^i \vec{u}_i$, $n^i \in \mathbb{Z}$.

has to be decomposed into d-dimensional fields. It takes little effort to see that one obtains a d-dimensional graviton, n d-dimensional vectors, and $(n+1)n/2$ scalars. The graviton and the n vectors gauge the unbroken symmetries of the vacuum: $\mathrm{ISO}(1, d-1) \times \mathrm{U}(1)^n$ (d-dimensional Poincaré times the n periodic isometries of the torus T^n). The (asymptotic values of the) scalars are *moduli*: they appear naturally arranged in an n-dimensional metric, which is the metric of the internal space T^n and they carry the information about circle sizes and relative angles. Evidently they generalize k, which contains only information about the size of the single internal circle.

On the moduli will act the global symmetries of the torus: the affine group $\mathrm{IGL}(n, \mathbb{R})$,

$$z^{m\,\prime} = (R^{-1\,\mathrm{T}})^m{}_n z^n + a^m, \qquad R \in \mathrm{GL}(n, \mathbb{R}), \quad a^m \in \mathbb{R}^n, \tag{11.171}$$

which will give rise to the duality symmetries of the lower-dimensional theory.

It makes sense again to perform dimensional reduction of the theory, keeping only the massless mode. Our goal in this section will therefore be to perform the dimensional reduction of the $\tilde{d}$-dimensional Einstein–Hilbert action to $d = \hat{d} - n$ dimensions.

The setup is the following: since we keep only the zero mode of the $\hat{d}$-dimensional metric, in practice we will be considering a metric that does not depend on the n coordinates z^m which parametrize the torus.[23] This is equivalent to saying that our metric does admit n mutually commuting, translational, and periodic spacelike Killing vectors $\hat{k}_{(m)}^{\hat{\mu}}$, which we identify with those of the internal torus. We assume that all the internal coordinates have the same period $2\pi\ell$.

We can find the right definitions of the d-dimensional fields as in the $n=1$ case. There is not much new to be learned there, so we start by performing the following decomposition of the $\hat{d}$-dimensional Vielbeins $\hat{e}_{\hat{\mu}}{}^{\hat{a}}$ (KK *Ansatz*) into d-dimensional Vielbeins $e_\mu{}^a$, vector fields $A^m{}_\mu$, and the n-dimensional internal metric Vielbeins $e_m{}^i$, which become scalars of the $(\hat{d}-n)$-dimensional theory:

$$\left(\hat{e}_{\hat{\mu}}{}^{\hat{a}}\right) = \begin{pmatrix} e_\mu{}^a & A^m{}_\mu e_m{}^i \\ 0 & e_m{}^i \end{pmatrix}, \qquad \left(\hat{e}_{\hat{a}}{}^{\hat{\mu}}\right) = \begin{pmatrix} e_a{}^\mu & -A^m{}_a \\ 0 & e_i{}^m \end{pmatrix}. \tag{11.172}$$

This *Ansatz* is always possible because there always is a Lorentz rotation of the Vielbeins that brings them into this upper-triangular form. As usual, the d-dimensional metric is built out of the Vielbeins in this way,

$$g_{\mu\nu} = e_\mu{}^a e_\nu{}^b \eta_{ab}, \tag{11.173}$$

and we use them to trade curved and flat lower-dimensional indices, so, for instance,

$$A^m{}_a = A^m{}_\mu e_a{}^\mu. \tag{11.174}$$

We also have for the internal metric scalars (recall our mostly minus signature)

$$G_{mn} = -e_m{}^i e_n{}^j \delta_{ij}. \tag{11.175}$$

[23] We split coordinates and indices as follows: $(\hat{x}^{\hat{\mu}}) = (x^\mu, z^m)$ and, for Lorentz indices, $(\hat{a}) = (a, i)$.

The relation between $\hat{d}$- and d-dimensional fields is

$$\begin{aligned}
\hat{g}_{\mu\nu} &= g_{\mu\nu} + A^m{}_\mu A^n{}_\nu G_{mn}, \\
\hat{g}_{\mu n} &= A^m{}_\mu G_{mn} = \hat{k}_{(n)\,\mu}, \\
\hat{g}_{mn} &= G_{mn} = \hat{k}_{(m)}{}^{\hat{\mu}} \hat{k}_{(n)\,\hat{\mu}}.
\end{aligned} \tag{11.176}$$

These fields transform correctly as tensors, vectors, and scalars under $\hat{d}$-dimensional GCTs in the non-compact dimensions ($\hat{\epsilon}^\mu \equiv \epsilon^\mu$). Furthermore, under $\hat{d}$-dimensional GCTs in the internal dimensions ($\hat{\epsilon}^m \equiv -\Lambda^m$), the vectors undergo standard U(1) transformations,

$$\delta_{\Lambda^m} A^n{}_\mu = \delta_m{}^n \partial_\mu \Lambda^m. \tag{11.177}$$

The constant shifts of the internal coordinates have no effect whatsoever on the d-dimensional fields. Furthermore, under the GL$(n, \mathbb{R})$ transformations only objects with internal indices transform. Thus, the d-dimensional metric is invariant and, in matrix notation, the internal metric and vectors transform according to

$$G' = R\, G\, R^{\mathrm{T}}, \qquad \vec{A}'_\mu = R^{-1\,\mathrm{T}} \vec{A}_\mu. \tag{11.178}$$

The group GL$(n, \mathbb{R})$ can be decomposed into SL$(n, \mathbb{R}) \times \mathbb{R}^+ \times \mathbb{Z}_2$, the $\mathbb{R}^+$ factor corresponding to rescalings analogous to those of the $n = 1$ case, that change the determinant of the internal metric, and later we will want to redefine the fields so they transform well under those factors.

To calculate now the components of the spin connection in the above Vielbein basis, we first calculate the Ricci rotation coefficients $\hat{\Omega}_{\hat{a}\hat{b}\hat{c}}$ and the non-vanishing ones are

$$\hat{\Omega}_{abc} = \Omega_{abc}, \qquad \hat{\Omega}_{abi} = \tfrac{1}{2} e_{mi} F^m{}_{ab}, \qquad \hat{\Omega}_{ibj} = -\tfrac{1}{2} e_i{}^m \partial_b e_{mj}. \tag{11.179}$$

They give

$$\begin{aligned}
\hat{\omega}_{abc} &= \omega_{abc}, & \hat{\omega}_{abi} &= -\tfrac{1}{2} e_{im} F^m{}_{ab}, \\
\hat{\omega}_{ibc} &= -\hat{\omega}_{bci}, & \hat{\omega}_{aij} &= -e_{[i|}{}^m \partial_a e_{m|j]}, \\
\hat{\omega}_{ibj} &= \tfrac{1}{2} e_i{}^m e_j{}^n \partial_b G_{mn}, & &
\end{aligned} \tag{11.180}$$

where we have used

$$e_{(i|}{}^m \partial_a e_{|m|j)} = \tfrac{1}{2} e_i{}^m e_j{}^n \partial_a G_{mn}, \tag{11.181}$$

and we have defined

$$F^m{}_{\mu\nu} \equiv 2\partial_{[\mu} A^m{}_{\nu]}, \qquad F^m{}_{ab} = e_a{}^\mu e_b{}^\nu F^m{}_{\mu\nu}. \tag{11.182}$$

Next, we plug this result into the Ricci scalar term in the action expressed in terms of the spin-connection coefficients with the help of Palatini's identity Eq. (D.4) and obtain

$$\begin{aligned}
\int d^{\hat{d}}\hat{x} \sqrt{|\hat{g}|}\, \hat{R} = \int d^n z \int d^d x \sqrt{|g|}\, K \big\{ &-\omega_b{}^{ba} \omega_c{}^c{}_a - \omega_a{}^{bc} \omega_{bc}{}^a + 2\omega_b{}^{ba} \partial_a \ln K \\
&- (\partial \ln K)^2 + \tfrac{1}{4} F^2 - \tfrac{1}{4} \partial_a G_{mn} \partial^a G^{mn} \big\},
\end{aligned} \tag{11.183}$$

where

$$K^2 \equiv |\det G_{mn}|, \qquad F^2 \equiv F^{m\,\mu\nu} F^{n}{}_{\mu\nu} G_{mn}, \tag{11.184}$$

so

$$G^{mn}\partial G_{mn} = 2\partial \ln K, \qquad \sqrt{|\hat{g}|} = \sqrt{|g|}\, K. \tag{11.185}$$

The sign of F^2 looks wrong, but one has to take into account the internal metric G_{mn} which is negative-definite.

Using again the Palatini identity but now in d dimensions and integrating over the internal coordinates, we find

$$\boxed{S = \frac{(2\pi\ell)^n}{16\pi G_{\rm N}^{(\hat{d})}} \int d^d x \sqrt{|g|}\, K \left\{ R - (\partial \ln K)^2 + \tfrac{1}{4}F^2 - \tfrac{1}{4}\partial_a G_{mn}\partial^a G^{mn} \right\}.} \tag{11.186}$$

We now want to use variables that are invariant under the $\mathbb{R}^+$ subgroup of rescalings, just as in the $n = 1$ case (the tilded variables). First, we observe that any transformation $R \in \mathrm{GL}(n, \mathbb{R})$ can be written as follows:

$$R = a^{\frac{1}{n}} S X, \qquad a = |\det R| \in \mathbb{R}^+, \qquad S \in \mathrm{SL}(n, \mathbb{R}), \qquad X^2 = \mathbb{I}_{n\times n}. \tag{11.187}$$

Second, we define the modulus K_0 as the value of the scalar K at infinity. According to its definition, K is nothing but the volume element of the internal torus and it generalizes the scalar k of the $n = 1$ case. The volume of the internal torus at a point x of the d-dimensional space is

$$V_n(x) = \int_{\mathrm{T}^n} d^n z \sqrt{|\det G_{mn}(x)|} = K(x) \int_{\mathrm{T}^n} d^n z = (2\pi\ell)^n K(x), \tag{11.188}$$

and its value at infinity V_n is measured in terms of the modulus K_0:

$$V_n = \lim_{r\to\infty} V_n(x) = (2\pi\ell)^n K_0. \tag{11.189}$$

If the torus were made up of orthogonal circles of local radii $R_m(x)$, then the internal metric would be diagonal

$$G_{mn} = -\delta_{(m)n} K_{(m)}, \qquad K_m = (R_m(x)/\ell)\,, \tag{11.190}$$

and the volume would factorize into the product of the volumes of the circles. We would have

$$V_n = \prod_{m=1}^{m=n} (2\pi\ell K_{m\,0}) = \prod_{m=1}^{m=n} (2\pi R_m), \tag{11.191}$$

but it is worth stressing that this is not the case in general.

Under the transformation $R \in \mathrm{GL}(n, \mathbb{R})$ decomposed as above, the scalar K and the modulus K_0 transform only under the $\mathbb{R}^+$ factor,

$$K' = a^{-1} K, \qquad K_0' = a^{-1} K_0, \tag{11.192}$$

and thus we can use them to define fields that are invariant under this factor:

$$
\begin{aligned}
\tilde{K} &= K/K_0, & \tilde{g}_{E\,\mu\nu} &= \tilde{K}^{\frac{2}{d-2}} g_{\mu\nu}, \\
\tilde{A}^m{}_\mu &= K_0^{\frac{1}{n}} A^m{}_\mu, & \mathcal{M}_{mn} &= -K^{-\frac{2}{n}} G_{mn}.
\end{aligned}
\tag{11.193}
$$

$\mathcal{M}$ and $\tilde{A}_\mu$ transform only under the $SX \in \mathrm{SL}(n, \mathbb{R}) \times \mathbb{Z}_2$ factor as expected:

$$
\mathcal{M}' = S\mathcal{M}S^{\mathrm{T}}, \qquad \tilde{\vec{A}}'_\mu = S^{-1\,\mathrm{T}} \tilde{\vec{A}}_\mu. \tag{11.194}
$$

The metric is the "modified Einstein-frame metric" and the action takes the form

$$
\boxed{
\begin{aligned}
S = \frac{V_n}{16\pi G_{\mathrm{N}}^{\hat{d}}} \int d^d x \sqrt{|\tilde{g}_{\mathrm{E}}|} \Bigg\{ & \tilde{R}_{\mathrm{E}} + \frac{\hat{d}-2}{n(d-2)} (\partial \ln \tilde{K})^2 - \tfrac{1}{4} \partial_\mu \mathcal{M}_{mn} \partial^\mu \mathcal{M}^{mn} \\
& - \tfrac{1}{4} \tilde{K}^{2\frac{\hat{d}-2}{n(d-2)}} \mathcal{M}_{mn} \tilde{F}^{m\,\mu\nu} \tilde{F}^n{}_{\mu\nu} \Bigg\}.
\end{aligned}
}
\tag{11.195}
$$

In this action, $\tilde{K}$ parametrizes an $\mathbb{R}^+$ σ-model, but what about $\mathcal{M}_{mn}$? This is a unimodular $n \times n$ matrix and, therefore, it belongs to $\mathrm{SL}(n, \mathbb{R})$ itself. Furthermore, it is symmetric and, therefore, it is not the most general $\mathrm{SL}(n, \mathbb{R})$ matrix we can find and it does not parametrize $\mathrm{SL}(n, \mathbb{R})$. In fact, with its $n(n+1)/2 - 1$ degrees of freedom, it parametrizes the coset space $\mathrm{SL}(n, \mathbb{R})/\mathrm{SO}(n, \mathbb{R})$. This can be seen as follows: we can view $\mathcal{M}$ as the product of two unimodular n-beins $\mathcal{V}_m{}^i$,

$$
\mathcal{M}_{mn} = \mathcal{V}_m{}^i \mathcal{V}_n{}^j \delta_{ij}, \qquad \mathcal{V}_m{}^i = K^{-\frac{1}{n}}\, \hat{e}_m{}^i. \tag{11.196}
$$

These unimodular n-beins transform under global $S \in \mathrm{SL}(n, \mathbb{R})$ transformations and local $\Lambda(x) \in \mathrm{SO}(n, \mathbb{R})$ transformations according to

$$
\mathcal{V}' = S\mathcal{V}\Lambda^{\mathrm{T}}(x). \tag{11.197}
$$

We can now choose $\mathcal{V}$ to be upper triangular. This can always be achieved by a suitable local $\mathrm{SO}(n, \mathbb{R})$ rotation. That matrix contains $n(n+1)/2 - 1$ degrees of freedom and parametrizes the coset space $\mathrm{SL}(n, \mathbb{R})/\mathrm{SO}(n, \mathbb{R})$ because it is an $\mathrm{SL}(n, \mathbb{R})$ matrix generated by the exponentiation of all the generators of that group except for those of the $\mathrm{SO}(n, \mathbb{R})$ subgroup which necessarily generate non-upper-triangular matrices.[24] We can see our choice of upper-triangular matrices as a coset-representative or gauge choice. S transformations take us out of our gauge choice but we can implement an S-dependent compensating Λ transformation to restore the upper-triangular form.

The constant value of $\mathcal{M}$ at infinity, $\mathcal{M}_0$, contains the modular parameters of the torus (relative sizes and angles of the circles).

[24] The transpose of an upper-triangular matrix with all terms above and on the diagonal non-vanishing can never be the inverse of that matrix.

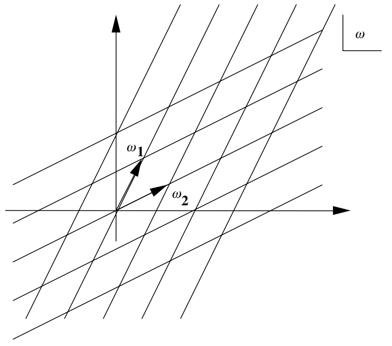

Fig. 11.1. The lattice generated in the ω plane by ω_1 and ω_2.

11.4.1 The 2-torus and the modular group

In our study of the global transformations of the internal torus we have not yet taken into account the periodic boundary conditions of the coordinates, which have to be preserved by the diffeomorphisms in the KK setting. Clearly the rescalings R do not respect the torus boundary conditions, but they rescale ℓ. The rotations S respect the boundary conditions only if $S^{-1}\vec{n} \in \mathbb{Z}^n$; the matrix entries are integers, i.e. $S \in \mathrm{SL}(n, \mathbb{Z})$.

The case $n = 2$ is particularly interesting because it occurs in many instances,[25] some (but not all of them) associated with S dualities. In the case $n = 2$, up to a reflection $S = -\mathbb{I}_{2\times 2}$, these diffeomorphisms are known as *Dehn twists* and are not connected to the identity (in fact, they constitute the mapping class group of torus diffeomorphisms) and they constitute the *modular group* $\mathrm{PSL}(2, \mathbb{Z}) = \mathrm{SL}(2, \mathbb{Z})/\{\pm\mathbb{I}_{2\times 2}\}$. This is the group that acts on $\mathcal{M}$.

It is convenient to relate $\mathcal{M}$ to the complex modular parameter τ of the torus. We start by defining a complex modular-invariant coordinate ω on T^2 by

$$\omega = \frac{1}{2\pi\ell}\,\vec{\omega}^{\mathrm{T}} \cdot \vec{z}, \qquad \vec{\omega} = \mathbb{C}^2, \tag{11.198}$$

where, under $\mathrm{PSL}(2, \mathbb{Z})$ modular transformations, we assume that the complex vector $\vec{\omega}$ transforms according to

$$\vec{\omega}' = S\vec{\omega}. \tag{11.199}$$

The periodicity of ω is

$$\omega \sim \omega + \vec{\omega}^{\mathrm{T}} \cdot \vec{n}, \qquad \vec{n} \in \mathbb{Z}^2. \tag{11.200}$$

The lattice generated in the ω plane by $\vec{\omega}$ is represented in Figure 11.1. In terms of the modular-invariant complex coordinate, the torus metric element

$$ds^2_{\mathrm{Int}} = d\vec{z}^{\,\mathrm{T}} G d\vec{z} \tag{11.201}$$

takes the form

$$ds^2_{\mathrm{Int}} = K^{\frac{1}{2}} \frac{1}{\mathrm{Im}(\omega_1\bar{\omega}_2)} d\omega d\bar{\omega}. \tag{11.202}$$

(Observe that $\mathrm{Im}(\omega_1\bar{\omega}_2)$ is a modular-invariant term, and a quite important one.)

---

[25] Owing to the isomorphisms $\mathrm{SL}(2, \mathbb{R}) \sim \mathrm{Sp}(2, \mathbb{R}) \sim \mathrm{SU}(1, 1)$ it takes several different, but equivalent, forms.

What we have just done is to transfer the information contained in the metric (more precisely, in $\mathcal{M}$) into the complex periods $\vec{\omega}$. The relation between these two is

$$\mathcal{M} = \frac{1}{\mathrm{Im}(\omega_1\bar{\omega_2})}\begin{pmatrix} |\omega_1|^2 & \mathrm{Re}(\omega_1\bar{\omega_2}) \\ \mathrm{Re}(\omega_1\bar{\omega_2}) & |\omega_2|^2 \end{pmatrix}. \tag{11.203}$$

We can check that the transformation rules for the complex periods Eq. (11.199) and for the matrix $\mathcal{M}$ Eq. (11.194) are perfectly compatible.

It should be clear that not all pairs of complex periods characterize different tori. Recall that $\mathcal{M}$ has only two independent entries whereas $\vec{\omega}$ contains four real independent quantities. In particular, we can see that multiplying $\vec{\omega}$ by any complex number leaves the matrix $\mathcal{M}$ invariant. It is customary to define the complex modulus parameter τ,

$$\tau = \omega_1/\omega_2, \tag{11.204}$$

that can always be chosen to belong to the upper half of the complex plane $\mathbb{H}$, $\mathrm{Im}(\tau) \geq 0$ ($-\omega_1$ defines the same torus as ω_1).

Under a modular transformation with S parametrized by

$$S = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \tag{11.205}$$

with $\alpha\delta - \beta\gamma = 1$, the modular parameter τ undergoes a fractional-linear transformation:

$$\tau' = \frac{\alpha\tau + \beta}{\gamma\tau + \delta}. \tag{11.206}$$

Finally, in terms of τ, the matrix $\mathcal{M}$ reads

$$\mathcal{M} = \frac{1}{\mathrm{Im}(\tau)}\begin{pmatrix} |\tau|^2 & \mathrm{Re}(\tau) \\ \mathrm{Re}(\tau) & 1 \end{pmatrix}. \tag{11.207}$$

The linear transformation of the matrix $\mathcal{M}$ Eq. (11.194) and the (non-linear) fractional-linear transformation Eq. (11.206) are completely equivalent.

The parametrization of the unimodular $\mathcal{V}$, in terms of τ, is

$$\mathcal{V} = \begin{pmatrix} [\mathrm{Im}(\tau)]^{\frac{1}{2}} & [\mathrm{Im}(\tau)]^{-\frac{1}{2}}\,\mathrm{Re}(\tau) \\ 0 & [\mathrm{Im}(\tau)]^{-\frac{1}{2}} \end{pmatrix}, \tag{11.208}$$

and the $\mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$ σ-model action takes the form

$$\int d^dx\sqrt{|\tilde{g}_\mathrm{E}|}\left[-\tfrac{1}{4}\partial_\mu\mathcal{M}_{mn}\partial^\mu\mathcal{M}^{mn}\right] = \int d^dx\sqrt{|\tilde{g}_\mathrm{E}|}\left[\tfrac{1}{2}\frac{\partial_\mu\tau\partial^\mu\bar{\tau}}{(\mathrm{Im}(\tau))^2}\right]. \tag{11.209}$$

As said, this σ-model and the global symmetry group $\mathrm{SL}(2,\mathbb{R})$ (broken by boundary conditions or quantum effects to $\mathrm{SL}(2,\mathbb{Z})$) appear in many instances, apart from T^2

compactifications. To start with, $\mathrm{SL}(2,\mathbb{Z})$ is the S-duality group and, under it, the complexified coupling constant that we also called τ transforms in the same way as the modular parameter of a torus (see Eq. (8.188)), but not every $\mathrm{SL}(2,\mathbb{Z})$ is an S duality.

Another important example is provided by $N=2B, d=10$ SUGRA (the effective-field theory of the type IIB superstring), which we will review in Chapter 17, in which there is an $\mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$ σ-model[26] with $\mathrm{Re}(\tau)=\hat{C}^{(0)}$, the RR (pseudo)scalar, and $\mathrm{Im}(\tau)=e^{-\hat{\varphi}}$, $\hat{\varphi}$ being the dilaton, and invariance under global $\mathrm{SL}(2,\mathbb{R})$ duality transformations that are interpreted as an S duality that rotates perturbative into non-perturbative states of the theory. In this case, the scalar σ-model does not arise from compactification. However, as we will explain in detail there, in $d=9$ dimensions it can be identified with the σ-model that arises in the compactification of $N=1, d=11$ supergravity to $d=9$ on T^2.

In the compactification of $N=1, d=11$ supergravity to $d=8$ dimensions on T^3 there is an $\mathrm{SL}(3,\mathbb{R})/\mathrm{SO}(3)$ σ-model that naturally contains the $\mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$ σ-model we just mentioned, but there is another $\mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$ σ-model that arises because the eleven-dimensional 3-form gives rise to an eight-dimensional pseudoscalar [28]. Similar effects give rise to many $\mathrm{SL}(2,\mathbb{R})$ subgroups of the total (U) duality group in various compactifications of eleven- and ten-dimensional SUGRAs [666].

In $N=4, d=4$ SUGRA (the theory which results from the compactification on T^6 of $N=1, d=10$ supergravity, the effective field theory of the heterotic and type-I strings), which we will review in Section 12.2, $\mathrm{Re}(\tau)=a$ is a pseudoscalar that is the Hodge dual of the dimensionally reduced Neveu–Schwarz–Neveu–Schwarz (for the heterotic) Ramond–Ramond (for the type-I) 2-form and $\mathrm{Im}(\tau)=e^{-2\phi}$, where ϕ is the four-dimensional dilaton.[27] In this case, the scalar $\mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$ σ-model does not arise from compactification on T^2 either.

11.4.2 Masses, charges and Newton's constant

In the tilded, scale-invariant variables that we have defined we can immediately see that the d-dimensional Newton constant is given by

$$\boxed{G_{\mathrm{N}}^{(d)}=G_{\mathrm{N}}^{(\hat{d})}/V_n.} \tag{11.210}$$

[26] In [820] this coset was described in the form SU(1,1)/U(1). This is natural if one wants to construct the supergravity theory from scratch, using complex fields, but, from the point of view of string theory, the natural parametrization is the real one $\mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$. The relation between the SU(1,1)/U(1) variable S and the $\mathrm{SL}(2,\mathbb{R})$ parameter τ is

$$\tau=i\frac{1-S}{1+S},$$

and the relation between the kinetic terms is

$$\tfrac{1}{2}\frac{\partial_\mu\tau\partial^\mu\overline{\tau}}{(\mathrm{Im}(\tau))^2}=2\frac{\partial_\mu S\partial^\mu\overline{S}}{(1-S\overline{S})^2}.$$

[27] In [266] this coset space was also described in the form SU(1,1)/U(1).

To find the right definitions for the n electric charges, we need the Noether currents. These are

$$j_n^{\nu} = \frac{1}{16\pi G_{\rm N}^{(d)}} \tilde{\nabla}_{\mu}\left(\tilde{K}^{2\frac{\hat{d}-2}{n(d-2)}} \mathcal{M}_{nm} \tilde{F}^{m\,\mu\nu}\right), \tag{11.211}$$

and then the electric and magnetic charges of the vector fields are defined by

$$\tilde{q}_n = \frac{1}{16\pi G_{\rm N}^{(d)}} \int_{S_{\infty}^{d-2}} \tilde{K}^{2\frac{(\hat{d}-2)}{n(d-2)}} \mathcal{M}_{nm}{}^{\star}\tilde{F}^{m\,\mu\nu}, \qquad \tilde{p}^n = -\int_{S_{\infty}^2} \tilde{F}^n. \tag{11.212}$$

With these definitions, the electric and magnetic charges of the vector field $\tilde{A}_{\mu}^{n}$ satisfy the Dirac quantization condition

$$\tilde{q}_n \tilde{p}^n = 2\pi m, \qquad m \in \mathbb{Z}. \tag{11.213}$$

11.5 Generalized dimensional reduction

In [835, 836] Scherk and Schwarz introduced the idea of *generalized dimensional reduction* (GDR) and developed a general formalism. Here we want to explain the principle underlying the idea of GDR.

We can understand GDR[28] as the answer to the question "How do we dimensionally reduce multivalued fields?" There at least two types of multivalued fields: fields that take values in some topologically non-trivial space (e.g. a circle) and fields that are defined up to some kind of local transformation (e.g. gauge vector fields, spinors (defined up to local Lorentz transformations) etc.). Let us take the simplest: a real scalar field $\hat{\varphi}$ taking values on a circle of radius m (like an axion, which is, as a matter of fact, a pseudoscalar). In practice, to represent a multivalued field one takes a field living on the real line and then identifies

$$\hat{\varphi} \sim \hat{\varphi} + 2\pi m. \tag{11.214}$$

A single-valued field has to be a strictly periodic function of the compact coordinate: on going once around the compact dimension, we return to the same point and there the field has to have the same value. However, a multivalued field such as $\hat{\varphi}$ is allowed to take a different value as long as it is a multiple of $2\pi m$ because the two values of the field are assumed to be physically equivalent. Thus, in general, we can have

$$\hat{\varphi}(x, z + 2\pi\ell) = \hat{\varphi}(x, z) + 2\pi N m \sim \hat{\varphi}(x, z). \tag{11.215}$$

The Fourier expansion of such a multivalued field in z is now

$$\hat{\varphi}^{(N)}(x, z) = \frac{mN}{\ell} z + \sum_{n\in\mathbb{Z}} e^{\frac{2\pi i n z}{\ell}} \hat{\varphi}^{(n)}(x). \tag{11.216}$$

[28] Originally, GDR was introduced as just a generalized KK *Ansatz* in which the $\hat{d}$-dimensional fields were allowed to depend on the internal coordinates z^m in such a way that the lower-dimensional fields did not depend on them and, at the same time, some symmetries were broken. Here, we prefer to take the view that GDR is *the* KK *Ansatz* for multivalued fields and it is not an option or just a clever trick.

The extra term linear in z is responsible for the multivaluedness. This term is clearly non-dynamical, unlike the KK modes $\hat{\varphi}^{(n)}(x)$ which are dynamical, which means that the value of N cannot change (at least, classically). N is chosen once and for all and its value defines the vacuum. Therefore, it is a (discrete) modulus of the theory.

It should be obvious that the above field configurations are topologically non-trivial: the field is "wound" N times around the compact dimension. The topological number that characterizes these configurations is the winding number N,

$$N = \frac{1}{2\pi \ell m} \oint \hat{\varphi}. \tag{11.217}$$

The choice of vacuum is also a choice of topological sector in the space of configurations.

It should be stressed that all this makes sense if there are solutions of the form

$$\hat{\varphi} = \frac{mN}{\ell} z \tag{11.218}$$

compatible with the vacuum configurations of the other fields. Otherwise, one cannot talk about those new vacua labeled by N.

How do we perform the dimensional reduction of this field in the vacuum N? The logic is always the same: we simply ignore the massive modes and keep the massless ones. This means that, to carry out dimensional reduction of the above field, we should consider the KK *Ansatz*

$$\hat{\varphi} = \frac{mN}{\ell} z + \hat{\varphi}^{(0)}(x) = \frac{mN}{\ell} z + \varphi(x). \tag{11.219}$$

Now the question of how we are supposed to obtain a truly $d = (\hat{d} - 1)$-dimensional theory if we start with a field that depends on the internal coordinate z arises. We can argue that the dependence on z will always disappear in the lower-dimensional theory: a field that lives on a circle necessarily appears in the action in a form that is invariant under *arbitrary constant shifts*. This means that the action can always be rewritten in terms of derivatives of $\hat{\varphi}$. Then, the linear term will either completely disappear (if it is hit by the derivative with respect to x^μ) or remain without the z (if it is hit by the derivative with respect to z). The surviving term will play the role of a mass term in general, as we will see.

This argument leads us to three observations.

1. The rule of thumb for how to perform GDR in this context is to implement a z-dependent shift in the scalar field's standard KK *Ansatz*. If we consider more general multivalued fields Φ, which are identified by

$$\hat{\Phi} \sim e^{i\omega \mathcal{Q}} \hat{\Phi}, \tag{11.220}$$

where $\mathcal{Q}$ is some symmetry of the theory, then the generalized KK *Ansatz* is, ignoring higher KK modes,

$$\hat{\Phi}(\hat{x}) \sim e^{\frac{i\omega \mathcal{Q} z}{2\pi \ell}} \Phi(x). \tag{11.221}$$

The symmetry generated by $\mathcal{Q}$ is generically broken.

2. The converse is not true: the invariance of the action under constant shifts of $\hat{\varphi}$ does not mean that the field lives on a circle and GDR makes sense. Formally, the GDR procedure can be performed, but the result could be meaningless since no vacuum solution associated with the GDR *Ansatz* is guaranteed to exist. We are going to see an example of this fact in Section 11.5.1.

3. Under U(1) gauge transformations

$$\delta_\Lambda z = -\Lambda(x), \qquad \delta_\Lambda A_\mu = \partial_\mu \Lambda(x), \qquad \delta_\Lambda \varphi = \frac{mN}{\ell}\Lambda(x), \tag{11.222}$$

i.e. the lower-dimensional scalar field transforms by shifts of the gauge parameter! This kind of gauge transformation is called a *massive gauge transformation* and allows us to eliminate φ completely by fixing the gauge. φ plays the role of Stückelberg field for A_μ [871]. KK gauge invariance is broken after this gauge fixing and this is reflected, as we will see, in a new mass term for the vector field. It is usually said that the vector has "eaten" the scalar, becoming massive. This is a sort of Higgs phenomenon, the difference being that there is no scalar potential. Observe that φ can be removed consistently by a gauge transformation if both Λ and φ live in circles, as we have assumed.

In the next sections we are going to see some examples of GDR that illustrate these ideas. In the first example we perform the complete GDR of the real scalar field that we have discussed above and give an alternative interpretation.

11.5.1 Example 1: a real scalar

Let us consider the simple model

$$\hat{S} = \int d^{\hat{d}}\hat{x}\sqrt{|\hat{g}|}\left[\hat{R} + \tfrac{1}{2}(\partial\hat{\varphi})^2\right], \tag{11.223}$$

where $\hat{\varphi}$ is a real scalar field. This action is invariant under constant shifts of the scalar and therefore it is possible to use the standard recipe for GDR: we perform now a z-dependent shift of the usual z-independent *Ansatz* $\hat{\varphi}(x,z) = \varphi(x) + mNz/\ell$, which will lead us to a d-dimensional theory with no dependence on z.

However, as we have stressed repeatedly, this recipe makes real sense only if the scalar field lives in a circle and is identified periodically, $\hat{\varphi} \sim \hat{\varphi} + 2\pi m$. Although it looks as if we can simply decree that identification, the above action does not contain enough structure to enforce it and we will see that, in particular, there is no vacuum solution with $\hat{\varphi}(x,z) = mNz/\ell$. This example is therefore just an academic exercise.

Using the standard *Ansatz* for the Vielbein Eq. (11.33) but adding a subscript (1) to the KK scalar field, we find

$$S = \int d^d x\sqrt{|g|}\,k\left[R - \tfrac{1}{4}k^2 F_{(2)}^2 + \tfrac{1}{2}(\mathcal{D}\phi)^2 - \tfrac{1}{2}\left(\frac{mN}{\ell}\right)^2 k^{-2}\right], \tag{11.224}$$

where the field strengths are defined by

$$F_{(2)\ \mu\nu} = 2\partial_{[\mu} A_{(1)\ \nu]}, \qquad \mathcal{D}_\mu \varphi = \partial_\mu \varphi - \frac{mN}{\ell} A_{(1)\ \mu}. \tag{11.225}$$

and are invariant under the massive gauge transformations

$$\delta_\Lambda z = -\Lambda, \qquad \delta_\Lambda \varphi = \frac{mN}{\ell}\Lambda, \qquad \delta A_{(1)\ \mu} = \partial_\mu \Lambda. \tag{11.226}$$

As we expected from our general discussion, φ is a Stückelberg field for $A_{(1)\ \mu}$, which becomes massive by "eating" it, and the KK U(1) symmetry is broken by our choice of vacuum if $N \neq 0$.

Now, let us try to find a vacuum solution of the reduced theory. It will correspond to the gauge-breaking vacuum of the $\hat{d}$-dimensional theory. We can assume that the vacuum solutions will have $A_\mu = 0$ and $\varphi = \varphi_0$, a constant. Solutions of this kind can be derived consistently from the above action by setting those fields to zero. On going to the Einstein frame and redefining k as in Eq. (11.47) but now calling χ the new scalar, we find the action of a real scalar with an unbounded potential coupled to gravity:

$$S = \int d^d x \sqrt{|g|} \left[R + 2(\partial\chi)^2 - \tfrac{1}{2}\left(\frac{mN}{\ell}\right)^2 e^{-2\sqrt{2\frac{d-1}{d-2}}\chi} \right]. \tag{11.227}$$

Since the potential has no minima, there are no vacuum solutions with constant χ equal to some minimum of the potential and a Minkowski metric. The vacuum has to have a non-trivial metric.

Typical solutions of actions of this kind, with generic potentials, are *domain-wall* solutions that interpolate between two asymptotic regions in which the scalar field takes the value of a different minimum of the potential, i.e. two vacua in which the scalar has a constant value equal to the minimum of the potential.[29] The region in which the value of the scalar switches from one vacuum value to another one is the domain wall. It is a $(d-2)$-dimensional region (plus the time) that is orthogonal to the coordinate on which the scalar typically depends. In fact, it can have zero thickness or some finite thickness in the direction of the transverse coordinate.

Although this potential has no minima, there might be some domain-wall-type solution since potentials like this one admit them: $(d-2)$-branes. However, precisely for the above potential, the generic solution given in [668] breaks down. Although this is far from a proof, it seems plausible that no such solution exists, confirming our suspicion that the GDR that we have performed is not consistent because it is based on a non-existent vacuum.

GDR and $(\hat{d}-3)$*-branes.* We have mentioned that actions such as Eq. (11.224) generically admit $(d-2)$-brane solutions. However, we have said that p-branes couple to a $(p+1)$-form potential with a $(p+2)$-form field strength and there is no d-form field strength in that action but only a potential proportional to the square of the mass parameter. However, terms of this kind, which are typical of massive supergravities, should not naively be interpreted

[29] A general reference for domain-wall solutions in $d = 4$ dimensions is [273].

as potentials. Instead, we should compare such a term with the kinetic term for the 1-form, which is also multiplied by a power of k. The analogy (and the fact that the sign is the correct one) suggests that we should interpret that term as a sort of "kinetic" term for a 0-form field strength (the mass constant), which happens to be the Hodge dual of the d-form field strength associated with the $(d-2)$-brane solutions.

There is another way, different from GDR, to see how the $(d-2)$-brane solutions arise: before performing the reduction of the scalar, we could have Hodge-dualized the $\hat{d}$-dimensional scalar into a $(\hat{d}-2)$-form potential by the Poincaré-duality procedure explained in Section 8.7.1,

$$^{\star}d\hat{\varphi} = d\hat{A}_{(\hat{d}-2)} \equiv F^{2}_{(\hat{d}-1)}, \tag{11.228}$$

obtaining the (on-shell) equivalent model

$$\tilde{\hat{S}} = \int d^{\hat{d}}\hat{x}\sqrt{|\hat{g}|}\left[\hat{R} + \frac{(-1)^{\hat{d}-2}}{2\cdot(\hat{d}-1)!}\hat{F}^{2}_{(\hat{d}-1)}\right]. \tag{11.229}$$

The KK dimensional reduction of p-forms follows the pattern of the reduction of the Maxwell vector field performed in Section 11.2.5: a p-form in $\hat{d}$ dimensions gives rise to a p-form and a $(p-1)$-form in d dimensions. The potentials and gauge-invariant field strengths are identified using tangent-space indices. In this case, we obtain a $(d-2)$-form potential and a $(d-1)$-form potential and the action

$$\tilde{S} = \int d^{\hat{d}-1}x\sqrt{|g|}\,k\left[R - \tfrac{1}{4}k^{2}F^{2}_{(2)} + \frac{(-1)^{d-1}}{2\cdot d!}F^{2}_{(d)} + \frac{(-1)^{d-2}}{2\cdot(d-1)!}k^{-2}F^{2}_{(d-1)}\right], \tag{11.230}$$

where[30]

$$F_{(d)} = d\partial A_{(d-1)} + (-1)^{d}A_{(1)}F_{(d-1)}, \qquad F_{(d-1)} = (d-1)\partial A_{(d-2)}, \tag{11.231}$$

are the field strengths. We can now dualize the potentials. A $(d-1)$-form potential in d dimensions has a d-form field strength whose Hodge dual is some function $f = {}^{\star}F_{(d)}$. The equation of motion of the $(d-1)$-form potential $d^{\star}F_{(d)} = 0$ becomes the Bianchi identity for the dual $df = 0$, which implies that f is a constant that we call Nm/ℓ. On adding the term

$$-\frac{1}{d!}\int d^{d}x\,\frac{mN}{\ell}\epsilon\Big[F_{(d)} + (-1)^{d+1}dA_{(1)}F_{(d-1)}\Big], \tag{11.232}$$

to the action and eliminating $F_{(d)}$ using its equation of motion,

$$mN/\ell = k^{\star}F_{(d)}, \tag{11.233}$$

in the action, we obtain

$$\begin{aligned}\tilde{S} = \int d^{d}x\sqrt{|g|}\Bigg\{k\Bigg[R - \tfrac{1}{4}k^{2}F^{2}_{(2)} + \frac{(-1)^{d-2}}{2\cdot(d-1)!}k^{-2}F^{2}_{(d-1)} - \tfrac{1}{2}\left(\frac{mN}{\ell}\right)^{2}k^{-2}\Bigg] \\ - \frac{1}{(d-1)!}\left(\frac{mN}{\ell}\right)\frac{\epsilon}{\sqrt{|g|}}F_{(d-1)}A_{(1)}\Bigg\}.\end{aligned} \tag{11.234}$$

[30] When indices are not explicitly shown, we assume all indices to be antisymmetrized with weight unity.

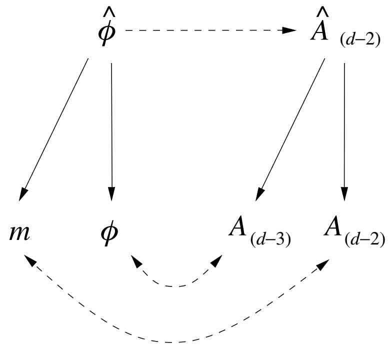

Fig. 11.2. This diagram represents two different ways of obtaining the same result: generalized dimensional reduction and "dual" standard dimensional reduction.

Now we dualize into a scalar field the $(d-2)$-form potential: we add the term

$$\frac{1}{(d-1)(d-1)!}\int d^d x\, \epsilon F_{(d-1)}\partial\varphi, \tag{11.235}$$

and eliminate $F_{(d-1)}$ by substituting into the action its equation of motion

$$F_{(d-1)} = (-1)^{(d-1)} k\, {}^\star\mathcal{D}\varphi, \tag{11.236}$$

obtaining the same result as with GDR. The two possible routes by which to arrive at the same d-dimensional theory are represented in Figure 11.2.

Thus, the standard recipe for GDR is just a way to take into account all the fields and degrees of freedom that can arise in the dimensional reduction. The new degrees of freedom are discrete degrees of freedom described by a $(d-1)$-form potential or by the dual variable that can take the values Nm/ℓ, $N \in \mathbb{Z}$ and are associated with a choice of vacuum.

Now, with the form $\hat{A}_{(\hat{d}-2)}$ we can associate a $(\hat{d}-3)$-brane. If one dimension is compact, there are two possibilities: either one of the dimensions of the brane is wrapped around the compact dimension or none is. From the d-dimensional point of view, the first configuration looks like a $(\hat{d}-4) = (d-3)$-brane and the second like a $(\hat{d}-3) = (d-2)$-brane. The $(\hat{d}-3) = (d-2)$-brane has no dynamics and has only one degree of freedom: its charge (or mass, which is usually proportional), which is the mass parameter that appears in the d-dimensional action. The mass parameters are to be considered fields, although one can equally consider them as expectation values of those fields. In this language we can say that our vacuum contains a $(d-2)$-brane.[31]

The charge of the $(\hat{d}-3)$-brane can be associated with the monodromy of $\hat{\varphi}$:

$$q \sim \oint {}^\star\hat{F}_{(\hat{d}-1)} \sim \oint d\hat{\varphi} \sim mN. \tag{11.237}$$

[31] We have said that it actually does not in this academic example, although it will in more general cases.

11.5.2 Example 2: a complex scalar

The simplest example just considered failed GDR because we did not really have a multi-valued scalar field. Let us consider a more interesting model with a complex scalar $\hat{\Phi}$:

$$\hat{S} = \int d^{\hat{d}}\hat{x}\sqrt{|\hat{g}|}\Big[\hat{R} + \tfrac{1}{2}\partial\hat{\Phi}\partial\hat{\Phi}^*\Big]. \tag{11.238}$$

It is invariant under phase shifts of the scalar. Actually, we must consider as equivalent $\hat{\Phi}$ and $e^{2\pi i}\hat{\Phi}$. If we split it into its modulus $\hat{\rho}$ and phase $\hat{\sigma}$, $\hat{\Phi} = \hat{\rho}e^{\frac{i\hat{\sigma}}{m}}$, $\hat{\sigma}$ must be identified with $\hat{\sigma} + 2\pi m$ and we can say that it lives in a circle of radius m. Following the rule of thumb of GDR, our *Ansatz* now has to be

$$\hat{\sigma}(x, z) = \hat{\sigma}(x) + \frac{Nm}{\ell}z, \quad \Rightarrow \hat{\Phi}(\hat{x}) = e^{\frac{iNz}{\ell}}\Phi(x). \tag{11.239}$$

We obtain the action

$$S = \int d^d x\sqrt{|g|}\, k\Big[R - \tfrac{1}{4}k^2F^2 + \tfrac{1}{2}\mathcal{D}\Phi\mathcal{D}\Phi^* - \tfrac{1}{2}\left(\frac{N}{\ell}\right)^2 k^{-2}|\Phi|^2\Big], \tag{11.240}$$

where the field strengths are now given by

$$F_{(2)\ \mu\nu} = 2\partial_{[\mu}A_{\nu]}, \qquad \mathcal{D}_\mu\Phi = \partial_\mu\Phi + i\frac{N}{\ell}A_\mu\Phi, \tag{11.241}$$

and are invariant under the massive U(1) gauge transformations

$$\delta_\Lambda A_\mu = \partial_\mu\Lambda, \qquad \delta_\Lambda\Phi = e^{\frac{iN\Lambda}{\ell}}\Phi. \tag{11.242}$$

This is (ignoring k) the Lagrangian for a complex massive scalar field with U(1) charge N/ℓ. In this case, the massive gauge transformations are simply the standard gauge transformations for a charged scalar field.

In terms of the real fields $\Phi = \rho e^{\frac{i\sigma}{m}}$ we find

$$\mathcal{D}\Phi\mathcal{D}\Phi^* = (\partial\rho)^2 + \frac{1}{m^2}\rho^2(\mathcal{D}\sigma)^2, \qquad \mathcal{D}_\mu\sigma = \partial_\mu\sigma + \frac{mN}{\ell}A_\mu. \tag{11.243}$$

ρ is invariant, but σ transforms under massive gauge transformations,

$$\delta_\Lambda\sigma = \frac{mN}{\ell}\Lambda, \tag{11.244}$$

and is a Stückelberg field for A_μ and can be gauged away, leaving a mass term for A_μ. U(1) can be spontaneously broken. A $|\hat{\Phi}|^4$ potential would produce a gravity-coupled version of the *Ginzburg–Landau Lagrangian*.

This model has an obvious solution $\rho = A_\mu = 0$ with the Minkowski metric. For $\rho = 0$ the σ-model defined by the scalars' kinetic terms is singular and we cannot distinguish between the different vacua labeled by N. Thus, this is also a failed example of GDR.

11.5.3 Example 3: an $\mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$ σ-model

It is given by the action

$$\hat{S} = \int d^{\hat{d}}\hat{x}\sqrt{|\hat{g}|}\left[\hat{R} + \tfrac{1}{2}\frac{\partial\hat{\tau}\partial\hat{\tau}^*}{(\mathrm{Im}(\tau))^2}\right]. \tag{11.245}$$

This action is invariant under global $\mathrm{SL}(2,\mathbb{R})$ fractional-linear transformations of $\hat{\tau} = \hat{a} + ie^{-\hat{\varphi}}$ and, in particular, under constant shifts $\hat{\tau} \to \hat{\tau} + b$, which act only on the real part. Furthermore, we have argued that, in many cases, we should consider as equivalent two values of τ related by $\mathrm{SL}(2,\mathbb{Z})$ transformations, which in particular means that a lives in a circle of unit length. In this case, the σ-model metric is regular for finite values of $\hat{\varphi}$.

The general recipe of GDR tells us to use the *Ansatz*

$$\hat{\tau}(\hat{x}) = \tau(x) + \frac{N}{2\pi\ell}z, \tag{11.246}$$

and we obtain the action

$$S = \int d^dx\sqrt{|g|}\,k\left[R - \tfrac{1}{4}k^2F^2 + \tfrac{1}{2}(\partial\varphi)^2 + \tfrac{1}{2}(\mathcal{D}a)^2 - \tfrac{1}{2}\left(\frac{N}{2\pi\ell}\right)^2 k^{-2}e^{-2\varphi}\right], \tag{11.247}$$

where

$$\mathcal{D}_\mu a = \partial_\mu a + \frac{N}{2\pi\ell}A_\mu, \tag{11.248}$$

and there is invariance under the massive U(1) gauge transformations

$$\delta_\Lambda A_\mu = \partial_\mu\Lambda, \qquad \delta_\Lambda a = \frac{N}{2\pi\ell}\Lambda. \tag{11.249}$$

Global $\mathrm{SL}(2,\mathbb{R})$ invariance is now clearly broken and the KK U(1) gauge invariance is also broken by the standard Stückelberg mechanism.

Let us look for a vacuum solution that will have $A_\mu = 0$ and $a = a_0$, a constant. The action for the remaining fields, in the Einstein frame, is

$$S = \int d^dx\sqrt{|g|}\left[R + 2(\partial\chi)^2 + \tfrac{1}{2}(\partial\varphi)^2 - \tfrac{1}{2}\left(\frac{N}{2\pi\ell}\right)^2 e^{-2\sqrt{2\frac{d-1}{d-2}}\chi - 2\varphi}\right]. \tag{11.250}$$

The potential for the remaining scalars has no lower bound, but we can still look for $(d-2)$-brane solutions. First, we diagonalize the potential by redefining the scalars:

$$\begin{aligned}
\chi' &= \sqrt{\frac{d-1}{3d-5}}\chi + \frac{2}{\sqrt{\frac{2(3d-5)}{d-2}}}\frac{\varphi}{2},\\
\frac{\varphi'}{2} &= -\frac{2}{\sqrt{\frac{2(3d-5)}{d-2}}}\chi + \sqrt{\frac{d-1}{3d-5}}\frac{\varphi}{2},
\end{aligned} \tag{11.251}$$

leaving the action in the form

$$S = \int d^d x \sqrt{|g|} \left[R + 2(\partial \chi')^2 + \tfrac{1}{2}(\partial \varphi')^2 - \tfrac{1}{2}\left(\frac{N}{2\pi \ell} \right)^2 e^{-2\sqrt{\frac{2(3d-5)}{d-2}}\chi'} \right]. \tag{11.252}$$

There are $(d-2)$-brane solutions with $\varphi' = 0$ in any dimension $d > 2$ [668]. These solutions are associated with $(\hat{d}-3)$-brane solutions in $\hat{d}$ dimensions of the $\mathrm{SL}(2,\mathbb{R})$ σ-model that, in the context of the ten-dimensional type-IIB superstring theory, are known as D-7-branes [118, 435]. We can say that the above action is the result of compactifying in a vacuum that contains N $(\hat{d}-3)$-branes.

At last we have a successful realization of GDR and its relation to $(\hat{d}-3)$-branes.

11.5.4 Example 4: Wilson lines and GDR

Another simple and interesting example is provided by a Dirac spinor $\hat{\psi}$ coupled to a U(1) gauge field $\hat{A}_{\hat{\mu}}$ in flat (for simplicity) $\hat{d} = 5$ spacetime,

$$\hat{S} = \int d^5 \hat{x} \left\{ \frac{i}{2} \bar{\hat{\psi}} \left(\hat{\partial\!\!\!/} - ig\hat{A\!\!\!/} \right) \hat{\psi} + \text{c.c.} \right\}. \tag{11.253}$$

This action is invariant under local U(1) transformations,

$$\hat{A}'_{\hat{\mu}} = \hat{A}_{\hat{\mu}} + \partial_{\hat{\mu}} \hat{\chi}, \qquad \hat{\psi}' = e^{ig\hat{\chi}} \hat{\psi}, \tag{11.254}$$

where the period of $\hat{\chi}$ has to be $2\pi/g$, and also under global phase shifts of the Dirac spinor (gauge transformations with constant $\hat{\chi}$). Thus, we can use the GDR *Ansatz*

$$\hat{\psi}(\hat{x}) = e^{\frac{iNgz}{2\pi\ell}} \psi(x). \tag{11.255}$$

The dependence on the coordinate z can be eliminated by a gauge transformation[32] with $\hat{\chi} = Nz/(2\pi\ell)$, but then the z component of the gauge vector acquires a constant value (a non-vanishing vacuum expectation value (VEV)

$$\hat{A}'_{\hat{\mu}} = \hat{A}_{\hat{\mu}} + \frac{N}{2\pi\ell} \delta_{\hat{\mu} z}. \tag{11.256}$$

The line integral of the vector field $\hat{A}_{\hat{\mu}}$ around the compact dimension is finite:

$$\oint_{\gamma} \hat{A} = N. \tag{11.257}$$

This configuration is said to have a U(1) *Wilson line*. The effect of the Wilson line (or the non-trivial dependence of the spinor on z) is to give a mass to the Dirac fermion. This is known as the *Hosotani or Wilson-line mechanism* [562–4] and we see that it can be transformed into Scherk–Schwarz GDR.

[32] Observe that the gauge parameter does not have the right periodicity.

11.6 Orbifold compactification

Sometimes it is possible to compactify on spaces that are not manifolds. The prototypes of these spaces are *orbifolds*. These can be constructed as the quotients of manifolds by discrete symmetries. The simplest case is the segment, which can be constructed as the quotient $S^1/\mathbb{Z}_2$. To describe the quotient we need to define the action of $\mathbb{Z}_2$, and for this it is convenient to describe the circle itself as the quotient of the real line parametrized by z by the group $\mathbb{Z}$ of discrete translations $z \to z + 2\pi n\ell$. There are no fixed points of the real line under this group and, therefore, we obtain a non-singular manifold.

Now, in terms of this coordinate z, $\mathbb{Z}_2$ acts by $z \to -z$. The result is the segment of line that goes from $z = 0$ to $z = 2\pi$. There are two fixed points under this group $z = 0$, for obvious reasons, and $z = \pi\ell$, since $-\pi\ell \sim -\pi\ell + 2\pi\ell = \pi\ell$, and they are the singular endpoints of the segment, which is not a manifold.[33]

The description of orbifolds as quotients is very convenient because in general the discrete symmetries have a well-defined action on the fields of the theory. In standard KK theory there are only tensor fields and their behavior under z reflections depends on the number of z indices they have: they acquire a minus sign for each index z. Only the KK vector has an odd number of z indices, $A_\mu = \hat{g}_{\mu\underline{z}}/\hat{g}_{\underline{zz}}$, and thus it reverses its sign while the metric and KK scalar remain invariant.

The rule is that the spectrum of the KK theory on an orbifold can contain only fields that are invariant under the discrete symmetry. The reason is that odd fields will be given in solutions by odd functions of z on the circle and they would be double-valued (i.e. not well defined) on the orbifold. Thus, in the standard KK theory the KK vector is projected out of it. It is precisely this mechanism that was used by Hořava and Witten in [543, 544] to eliminate the RR 1-form $\hat{C}^{(1)}$ in the reduction of 1one-dimensional supergravity (the effective-field theory of "M theory" in some corner of moduli space) to obtain chiral $N = 1, d = 10$ supergravity (the effective-field theory of the heterotic string) instead of unchiral $N = 2A, d = 10$ supergravity (see Section 16.4).

In supersymmetric KK theory one has to define the action of the $\mathbb{Z}_2$ group on fermions. In odd dimensions one typically defines

$$\hat{\psi}' = \pm\hat{\Gamma}_z\hat{\psi}, \tag{11.258}$$

where Γ_z is the gamma matrix associated with the direction z and is proportional to the chirality matrix in one dimension fewer. Then, in the orbifold compactification only one chiral half of the spinors survives the projection.

[33] The corresponding spacetime, taking into account the metric would have a size of πR_z.

12
Dilaton and dilaton/axion black holes

In the previous chapter we have seen how scalar fields coupled to gravity arise naturally in KK compactification. In Part III we are also going to see that scalar fields are also present, even before compactification, in some higher-dimensional supergravity theories that are the low-energy effective-field theories of certain superstring theories. In all these examples the scalar fields couple in a characteristic way to vector (or p-form in higher dimensions) field strengths. In this chapter we are going to study first, in Section 12.1, a simple model that synthesizes the main features of those theories.

The a-model describes a real scalar coupled to gravity and to a vector-field strength. The coupling is exponential and depends on a parameter a (hence the name "a-model" that we are giving it here). Since the scalar can be identified in some cases with the string dilaton (or with the KK scalar, which is called also the dilaton sometimes), these models are also generically referred to as *dilaton gravity*. We will be able to obtain BH-type solutions for general values of a and in any dimension $d \geq 4$; however, only a handful of values of a actually occur in the theories of interest, although they occur in many different ways (*embeddings* [620]).

After studying the main properties of these *dilaton BHs*, we are going to study in Section 20.1 a more complex (four-dimensional) model that involves several scalar and vector fields. We are going to obtain extreme BH solutions that can be understood as *composite* BHs. This interpretation will open the door to the construction of four-dimensional extreme BH solutions in string theory as composite objects, the building blocks being p-branes and other extended objects that we will study in Chapter 20.

In Section 12.2 we add to the a-model with $a = 1$ and $d = 4$ a second scalar that couples not to the vector-field kinetic term F^2 but to $F^\star F$ and also couples to the dilaton. This kind of scalar (actually, a pseudoscalar, to preserve invariance of the action under parity) is called an *axion*. The model obtained has equations of motion that are invariant under global $\mathrm{SL}(2, \mathbb{R})$ (S) duality transformations (the dilaton and the axion parametrize an $\mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$ coset space) and it is sometimes called *axion–dilaton* gravity. The S-duality transformations can be used to obtain new solutions from known solutions.

As a matter of fact, the axion–dilaton-gravity model is a truncation of the bosonic sector of pure, ungauged, $N = 4$, $d = 4$ SUGRA which has five additional vector fields. This theory is a consistent truncation of the effective-field theory of the heterotic superstring compactified on T^6, as we will see in Chapter 16, and we will therefore study its BH solutions. The most general solution (compatible with the no-hair conjecture) takes a very interesting duality-invariant form. The most general BH solution of the heterotic superstring compactified on T^6 should have a similarly duality-invariant form but it is, unfortunately, unknown.

12.1 Dilaton black holes: the a-model

The d-dimensional "a-model" action is[1]

$$\boxed{S = \frac{1}{16\pi G_N^{(d)}} \int d^d x \sqrt{|g|} \left[R + 2(\partial\varphi)^2 - \tfrac{1}{4} e^{-2a\varphi} F^2\right],} \tag{12.1}$$

where, as usual $F_{\mu\nu} = 2\partial_{[\mu}A_{\nu]}$ (or the equations of motion have to be supplemented by the Bianchi identity for F).

Our goal in this section is to find and study BH-type solutions of this model. Since there is a scalar, we might think that the only BH solutions are those with a trivial (constant) scalar field, because those are the only ones with no scalar hair. However, as we discussed in Section 8.1, we should distinguish between primary and secondary scalar hair. Secondary scalar hair is related to other conserved charges by a certain, fixed, formula, and is compatible with the existence of event horizons. We have already met in the previous chapter some examples of dilaton BHs with non-trivial scalar fields and regular horizons. We do not know *a priori* the formula that relates the allowed, secondary, scalar "charge" to the conserved charges (mass and electric or magnetic charges), but we can deduce it from explicit BH solutions, if we find them.

The equations of motion are

$$G_{\mu\nu} + 2T^{\varphi}_{\mu\nu} - \tfrac{1}{2} e^{-2a\varphi} T^{A}_{\mu\nu} = 0, \qquad \nabla^2\varphi - \tfrac{1}{8} a e^{-2a\varphi} F^2 = 0, \qquad \nabla_\mu\left(e^{-2a\varphi} F^{\mu\nu}\right) = 0, \tag{12.2}$$

where

$$T^{\varphi}_{\mu\nu} = \partial_\mu\varphi\partial_\nu\varphi - \tfrac{1}{2} g_{\mu\nu}(\partial\varphi)^2, \qquad T^{A}_{\mu\nu} = F_\mu{}^\rho F_{\nu\rho} - \tfrac{1}{4} g_{\mu\nu} F^2. \tag{12.3}$$

Observe that, when $a = 0$, this is the Einstein–Maxwell system with an uncoupled scalar, which we can take to be constant. For $a \neq 0$ the only solutions that have a trivial dilaton (and are, therefore, solutions of the Einstein–Maxwell system) are those with $F^2 = 0$. We have already made use of this observation to embed solutions of the Einstein–Maxwell theory (dyonic RN BHs) into the KK theory (page 330) to obtain the RN–KK dyon.

In Section 11.2 we showed that, in the KK reduction of pure $(d+1)$-dimensional gravity on a circle, we always obtain an a-model with a given by

$$a_{KK} = \pm\sqrt{\frac{2(d-1)}{d-2}}. \tag{12.4}$$

---

[1] Generalizations with a massive dilaton have been studied in [475, 545].

(The two signs are related by the transformation $\varphi \to -\varphi$.) We will see that, in the reduction of the heterotic string on T^6, we naturally obtain $a = 1$.

If we take the divergence of the Einstein equation above and use both the Bianchi identity for the metric and the Bianchi identity for F, we obtain

$$\left[\nabla^2\varphi - \tfrac{1}{8}ae^{-2a\varphi}F^2\right]\partial_\nu\varphi = 0, \tag{12.5}$$

which implies the scalar equation of motion *provided that the scalar is not constant*. For a non-constant scalar the only equations that we have to solve are, then, the Maxwell equation and the Einstein equation (with the trace subtracted for convenience):

$$\begin{aligned} R_{\mu\nu} + 2\partial_\mu\varphi\partial_\nu\varphi - \tfrac{1}{2}e^{-2a\varphi}\left[F_\mu{}^\rho F_{\nu\rho} - \frac{1}{2(d-2)}g_{\mu\nu}F^2\right] &= 0, \\ \nabla_\mu\left(e^{-2a\varphi}F^{\mu\nu}\right) &= 0. \end{aligned} \tag{12.6}$$

We want to find solutions of these equations describing electrically charged BHs that have to have a non-trivial scalar field. The solutions will reduce to the RN BH when $a = 0$, and when $F = 0$ we expect to recover the solutions of [18, 607] and the higher-dimensional analogs presented in Eq. (8.216). On the basis of our previous experience and discussions, it is natural to make an *Ansatz* for the (static, spherically symmetric) metric that generalizes the "dressed Schwarzschild" metric in such a way that we can call it the "dressed RN" metric:

$$\begin{aligned} ds^2 &= f^{2x}H^{-2}Wdt^2 - f^{-2y}H^{\frac{2}{d-3}}\left[W^{-1}dr^2 + r^2d\Omega^2_{(d-2)}\right], \\ A_\mu &= \alpha\delta_{\mu t}(H^{-1} - 1), \qquad e^{-2a\varphi} = f^z, \end{aligned} \tag{12.7}$$

where $f = HW^b$ and

$$H = 1 + \frac{h}{r^{d-3}}, \qquad W = 1 + \frac{\omega}{r^{d-3}}, \tag{12.8}$$

and x, y, z, b, h, ω, and α are constants to be found. Observe that the action is invariant under constant shifts of φ accompanied by rescalings of the vector field. We can use this symmetry later to add a constant value at infinity to φ and have

$$A_\mu = e^{a\varphi_0}\alpha\delta_{\mu t}(H^{-1} - 1), \qquad e^{-2a\varphi} = e^{-2a\varphi_0}f^z, \tag{12.9}$$

On substituting into the above equations of motion, one finds that there are two families of solutions, one with $b = 0$ and another one with $b = -1$. All the constants are identical in the two families, so all the fields (except for the metric) are identical. The first family contains regular BHs, but the second doesn't. The difference between the two families is the relation between the scalar charge and the mass and electric charge. We can view this difference as the presence of secondary scalar hair in the $b = 0$ family and of primary scalar hair in the $b = -1$ family. We are primarily interested in regular BHs and, therefore, we only write the $b = 0$ family of *dilaton-BH* solutions in its final form:

$$\boxed{\begin{array}{l} ds^2 = e^{-2a(\varphi-\varphi_0)}H^{-2}W dt^2 - \left(e^{-2a(\varphi-\varphi_0)}H^{-2}\right)^{-\frac{1}{d-3}}\left[W^{-1}dr^2 + r^2 d\Omega^2_{(d-2)}\right], \\ A_\mu = \alpha e^{a\varphi_0}\delta_{\mu t}(H^{-1} - 1), \qquad e^{-2a\varphi} = e^{-2a\varphi_0}H^{2x}, \\ H = 1 + \dfrac{h}{r^{d-3}}, \qquad W = 1 + \dfrac{\omega}{r^{d-3}}, \qquad \omega = h\left[1 - \dfrac{a^2}{4x}\alpha^2\right], \\ x = \dfrac{(a^2/2)c}{1 + (a^2/2)c}, \qquad c = \dfrac{d-2}{d-3}. \end{array}} \tag{12.10}$$

On adding the corresponding factors of W to this solution, one obtains the $b = -1$ family. Here a and d are given parameters that determine our theory and α, φ_0 (the value of the dilaton at infinity), and h (the coefficient of $r^{-(d-3)}$ in H) are the independent parameters. The relation between ω and h is valid only for $h \neq 0$. If $h = 0$, then ω is an arbitrary constant, there is no electromagnetic field, and we recover the solutions (8.216) which have primary scalar hair (the scalar charge is unrelated to the conserved charges) and are singular except for $b = 0$ or for $a = 0$ (which implies that $x = 0$), which is the Schwarzschild solution. For $a = 0$ we recover the RN solution, as we wanted.

Furthermore, for all values of d, a, and b, when $\omega = 0$ (*extreme dilaton BHs*) H can be any arbitrary harmonic function in the transverse $(d-2)$-dimensional Euclidean space. This allows us to construct multi-BH (in general multicenter) solutions, as in the MP family (which is included in this one with $a = 0$).

The $b = 0$ solutions were first obtained and studied in [432, 447]. The $d = 4$ solutions were rediscovered from a string-theory point of view in [416] and those with arbitrary a were also studied in [539]. The multicenter solutions were found in [853] (see also [743]) and the solutions with $b = -1$ are presented for the first time here.

Let us now study the properties and the geometry of the $b = 0$ family. First, we want to relate the integration constants to the physical parameters: mass, electric charge, and "scalar charge." Only the first two are independent. For the sake of clarity we omit most numerical factors and define these charges by the asymptotic expansions of the fields:

$$g_{tt} \sim 1 - \frac{\mathcal{M}}{r^{d-3}}, \qquad A_t \sim -\frac{\mathcal{Q}}{r^{d-3}}, \qquad \varphi \sim \varphi_0 - \frac{\mathcal{S}}{r^{d-3}}. \tag{12.11}$$

These charges are related to the integration constants by

$$\mathcal{M} = 2(1-x)h - \omega, \qquad \mathcal{Q} = \alpha e^{a\varphi_0}h, \qquad \mathcal{S} = xh. \tag{12.12}$$

The inverse relations are, for $x \neq \frac{1}{2}$,

$$\begin{aligned}
h &= \frac{\mathcal{M} \pm \sqrt{\mathcal{M}^2 - \frac{1-2x}{x} a^2 e^{-2a\varphi_0} \mathcal{Q}^2}}{2(1-2x)}, \\
\alpha &= \frac{2(1-2x) e^{-a\varphi_0} \mathcal{Q}}{\mathcal{M} \pm \sqrt{\mathcal{M}^2 - \frac{1-2x}{x} a^2 e^{-2a\varphi_0} \mathcal{Q}^2}}, \\
\omega &= \frac{x}{1-2x} \mathcal{M} \pm \frac{1-x}{1-2x} \sqrt{\mathcal{M}^2 - \frac{1-2x}{x} a^2 e^{-2a\varphi_0} \mathcal{Q}^2},
\end{aligned} \tag{12.13}$$

and they give us the expression of the "scalar charge" $\mathcal{S}$ in terms of $\mathcal{M}$ and $\mathcal{Q}$:

$$\mathcal{S} = \frac{a^2 e^{-2a\varphi_0} \mathcal{Q}^2}{2\left(\mathcal{M} \mp \sqrt{\mathcal{M}^2 - \frac{1-2x}{x} a^2 e^{-2a\varphi_0} \mathcal{Q}^2}\right)}. \tag{12.14}$$

We see that it vanishes for vanishing a (the RN solution) or vanishing $\mathcal{Q}$. This does not happen in the $b = -1$ family. If $\mathcal{S}$ has a different value (as in the $b = -1$ family) then there is primary scalar hair and we have solutions without regular event horizons.

The integration constants h, ω, and α are real only when

$$\mathcal{M}^2 \geq \frac{1-2x}{x} a^2 e^{-2a\varphi_0} \mathcal{Q}^2. \tag{12.15}$$

This is a constraint on $\mathcal{M}$ and $\mathcal{Q}$ only for $x < \frac{1}{2}$, that is, for $(a^2/2)c < 1$. This includes, as we know, the RN case.

When $x = \frac{1}{2}$, that is, $a = \pm\sqrt{2(d-3)/(d-2)}$ we find

$$h = \frac{d-3}{d-2} \frac{e^{-2a\varphi_0} \mathcal{Q}^2}{\mathcal{M}}, \qquad \alpha = \frac{d-2}{d-3} \frac{e^{a\varphi_0} \mathcal{M}}{\mathcal{Q}}, \qquad \omega = -\frac{\mathcal{M}^2 - \frac{d-3}{d-2} e^{-2a\varphi_0} \mathcal{Q}^2}{\mathcal{M}}, \tag{12.16}$$

and the "scalar charge" is given by

$$\mathcal{S} = \pm \frac{d-3}{2(d-2)} \frac{e^{-2a\varphi_0} \mathcal{Q}^2}{\mathcal{M}}. \tag{12.17}$$

These metrics are a generalization of the RN metric and they have two horizons, at $r = 0$ and $r = -\omega$. If we take the lower sign, $r = -\omega$ is the (regular in all cases with $\omega \neq 0$) event horizon but the "horizon" at $r = 0$ is generically singular, except for $a = 0$. When $\omega = 0$ (the extremal limit) the two horizons coincide. This happens when

$$\mathcal{M} = \frac{2(1-x)}{x} \mathcal{S} = \frac{1-x}{\sqrt{x}} a e^{-a\varphi_0} |\mathcal{Q}|. \tag{12.18}$$

However, in this limit we have a regular BH only for $a = 0$ (the ERN BH). All the other extreme dilaton BHs have a singular "horizon," i.e. a naked singularity.

There is a better way to express the extremality condition, using the scalar charge:

$$\omega^2 = \mathcal{M}^2 + 4\left(\frac{1}{x} - 1\right)\mathcal{S}^2 - a^2\left(\frac{1}{x} - 1\right)e^{-2a\varphi_0}\mathcal{Q}^2 = 0. \tag{12.19}$$

This form suggests that, as in the ERN case, the extremality condition can be viewed as a no-force condition. The difference here is that the dilaton field carries an additional interaction proportional to the "scalar charge" $\mathcal{S}$. This gives a physical explanation for the existence of regular multi-dilaton BH solutions. A pair of dilaton BHs with charges $(\mathcal{M}_i, \mathcal{Q}_i)$, satisfying separately the extremality condition, will also satisfy the no-force condition,

$$\mathcal{M}_1\mathcal{M}_2 + 4\left(\frac{1}{x} - 1\right)\mathcal{S}_1\mathcal{S}_2 - a^2\left(\frac{1}{x} - 1\right)e^{-2a\varphi_0}\mathcal{Q}_1\mathcal{Q}_2 = 0. \tag{12.20}$$

12.1.1 The *a*-model solutions in four dimensions

The general solution for the four-dimensional a-model is

$$\boxed{\begin{array}{l} ds^2 = H^{-\frac{2}{1+a^2}} W dt^2 - H^{-\frac{2}{1+a^2}}\left[W^{-1}dr^2 + r^2 d\Omega^2_{(2)}\right], \\ A_\mu = \alpha e^{a\varphi_0}\delta_{\mu t}(H^{-1} - 1), \qquad e^{-2\varphi} = e^{-2\varphi_0} H^{\frac{2a}{1+a^2}}, \\ H = 1 + \dfrac{h}{r}, \qquad W = 1 + \dfrac{\omega}{r}, \qquad \omega = h\left[1 - (1 + a^2)(\alpha/2)^2\right]. \end{array}} \tag{12.21}$$

Among all the possible values of a, only a few are relevant, at least in SUGRA theories (and hence for string theory). The most important value is $a = \sqrt{3}$. This is the value that we obtain in KK compactification from five to four dimensions, but it also appears in many other ways. The metric and dilaton field are[2]

$$\boxed{\begin{array}{l} ds^2 = H^{-\frac{1}{2}} W dt^2 - H^{\frac{1}{2}}\left[W^{-1}dr^2 + r^2 d\Omega^2_{(2)}\right], \\ e^{-2\varphi} = e^{-2\varphi_0} H^{\frac{\sqrt{3}}{2}}, \qquad \omega = h\left[1 - \alpha^2\right]. \end{array}} \tag{12.22}$$

All the extended objects of type-II string theory compactified on tori give rise precisely to this Einstein metric (see, for instance, Eqs. (20.8) and (20.10)). This illustrates the comment we made in the introduction about the many possible embeddings of the a-model solutions into SUGRA theories.[3]

[2] The vector field and the functions H and W always have the same form as in Eqs. (12.21).

[3] A systematic study of embeddings of these four-dimensional dilaton BHs in the effective-field theory of the heterotic string ($N = 1$, $d = 10$ SUGRA plus 16 vector multiplets) and their unbroken supersymmetries was presented in [620].

The next values of interest from the string-theory-supergravity point of view are $a = 1$;

$$\boxed{\begin{aligned} ds^2 &= H^{-1}Wdt^2 - H\left[W^{-1}dr^2 + r^2 d\Omega^2_{(2)}\right],\\ e^{-2\varphi} &= e^{-2\varphi_0}H, \qquad \omega = h\Big[1 - (\alpha/\sqrt{2})^2\Big], \end{aligned}} \tag{12.23}$$

and $a = 1/\sqrt{3}$;

$$\boxed{\begin{aligned} ds^2 &= H^{-\frac{3}{2}}Wdt^2 - H^{\frac{3}{2}}\big[W^{-1}dr^2 + r^2 d\Omega^2_{(2)}\big],\\ e^{-2\varphi} &= e^{-2\varphi_0}H^{\frac{\sqrt{3}}{2}}, \qquad \omega = h\big[1 - \alpha^2/3\big]. \end{aligned}} \tag{12.24}$$

$d = 4$ stringy solutions with these metrics will appear in the compactification of solutions that describe the intersection of two and three extended objects, respectively, instead of just one as in the previous case. We are going to see how this comes about in Section 20.1.

Finally, we have $a = 0$, the RN BH:

$$\boxed{\begin{aligned} ds^2 &= H^{-2}Wdt^2 - H^{2}\big[W^{-1}dr^2 + r^2 d\Omega^2_{(2)}\big],\\ e^{-2\varphi} &= e^{-2\varphi_0}, \qquad \omega = h\big[1 - (\alpha/2)^2\big]. \end{aligned}} \tag{12.25}$$

This case will be seen to arise from the intersection of four extended objects in higher dimensions.

In four dimensions, we can define the mass M, electric charge q, and “scalar charge” Σ more precisely by the asymptotic expansions

$$g_{tt} \sim 1 - \frac{2G_N^{(4)}M}{r}, \qquad A_t \sim \frac{4G_N^{(4)}e^{2a\varphi_0}q}{r}, \qquad \varphi \sim \varphi_0 + \frac{G_N^{(4)}\Sigma}{r}. \tag{12.26}$$

The dilaton-dependent factor $e^{2a\varphi_0}$ in the definition of the electric charge is related to the integral definition

$$q = \frac{1}{16\pi G_N^{(4)}}\int_{S^2_\infty} e^{-2a\varphi\star}F, \tag{12.27}$$

which is in turn related to the modification of the Gauss law introduced by the dilaton.

For $a \neq 1$ the integration constants in the solutions are given by

$$\begin{aligned} h &= \frac{1+a^2}{1-a^2}G_N^{(4)}\Big(M \pm \sqrt{M^2 - 4(1-a^2)e^{2a\varphi_0}q^2}\Big),\\ \alpha &= \frac{1-a^2}{1+a^2}\,\frac{4e^{a\varphi_0}q}{M \pm \sqrt{M^2 - 4(1-a^2)e^{2a\varphi_0}q^2}},\\ \omega &= \frac{2a^2}{1-a^2}G_N^{(4)}M \pm \frac{2}{1-a^2}G_N^{(4)}\sqrt{M^2 - 4(1-a^2)e^{2a\varphi_0}q^2}, \end{aligned} \tag{12.28}$$

and Σ is related to the conserved charges by

$$\Sigma = \frac{4a^2 e^{2a\varphi_0} q^2}{M \pm \sqrt{M^2 - 4(1-a^2)e^{2a\varphi_0} q^2}}. \tag{12.29}$$

For $a = 1$

$$h = \frac{4G_N^{(4)} e^{2\varphi_0} q^2}{M}, \qquad \alpha = -\frac{M}{e^{\varphi_0} q}, \qquad \omega = -2G_N^{(4)} \frac{M^2 - 2e^{2\varphi_0} q^2}{M}, \tag{12.30}$$

and Σ is related to the conserved charges by

$$\Sigma = -\frac{2G_N^{(4)} e^{2\varphi_0} q^2}{M}. \tag{12.31}$$

The extremality condition always takes the form

$$\left(\frac{\omega}{2G_N^{(4)}}\right)^2 = M^2 + \frac{1}{a^2}\Sigma^2 - 4e^{2\varphi_0} q^2 = 0. \tag{12.32}$$

Thermodynamics of four-dimensional dilaton BHs. The coupling to scalar fields requires a modification of the first law of BH thermodynamics, which has to include a new term [259, 260, 446] in order to take into account possible variations of the energy due to variations of the scalar fields. This term is proportional to the "scalar charges" and to the variations of the values of the scalar fields at infinity (*moduli*) that characterize the vacuum of the theory,

$$\Sigma_a d\varphi^a. \tag{12.33}$$

Apart from this new term, the temperature and the entropy of dilaton BHs are related to the area and surface gravity of the event horizon by the standard formulae. When $\omega \leq 0$ (as we will assume) the event horizon is placed at $r = -\omega$. Its area is given by

$$A = 4\pi H^{\frac{2}{1+a^2}} r^2 \Big|_{r=-\omega}, \tag{12.34}$$

and so the entropy is given by

$$S = \pi (h + |\omega|)^{\frac{2}{1+a^2}} |\omega|^{\frac{2a^2}{1+a^2}}. \tag{12.35}$$

The temperature is given by

$$T = \frac{1}{4\pi} (h + |\omega|)^{-\frac{2}{1+a^2}} |\omega|^{\frac{1-a^2}{1+a^2}}. \tag{12.36}$$

These expressions can be compared with those in [539], where the thermodynamics of four-dimensional dilaton BHs was studied, with $|\omega| = r_+ - r_-$ and $h + |\omega| = r_+$.

The behavior of T and S in the extreme limit depends on the value of the parameter a:

$$\lim_{\omega \to 0} \begin{cases} T \to 0, & S \to \pi h^{\frac{2}{1+a^2}}, \quad 0 \leq a < 1, \\ T \to h, & S \to 0, \quad a = 1, \\ T \to \infty, & S \to 0, \quad a > 1, \end{cases} \tag{12.37}$$

Below $a = 1$ the behavior is similar to that of the RN BH, which also means that near the extreme limit the specific heat is positive. Above $a = 1$ the behavior is similar to that of the Schwarzschild BH in the zero-mass limit and the specific heat is negative.

Electric–magnetic duality in the four-dimensional a-model. In four dimensions the a-model has electric–magnetic duality: the equations of motion are invariant under the discrete transformation

$$F' = \tilde{F} \equiv e^{-2a\varphi\star}F, \qquad \varphi' = \tilde{\varphi} \equiv -\varphi. \tag{12.38}$$

Tilded fields are, by definition, the S-dual fields. This symmetry allows us to transform the above electrically charged solutions into magnetic solutions that have the same (Einstein-frame) metric. All the properties that depend on the Einstein metric (for instance, thermodynamical properties) are not affected by this transformation. However, in some cases we are interested in properties that depend on the metric given in a different frame (such as the string frame, that we will study, or the KK frame that we studied in Chapter 11) that is related to Einstein's by a conformal rescaling by a function of the dilaton. Since the dilaton changes in this electric–magnetic transformation, so does the (KK or stringy) metric. A good example is provided by the electric–magnetic-duality rotation of the electrically charged KK BH studied on page 328.

For special values of the parameter a there are also dyonic dilaton BH solutions, carrying both electric and magnetic charges.[4] This trivially happens for $a = 0$, the Einstein–Maxwell plus uncoupled scalar case, because in this case (as we have already seen) the electric–magnetic-duality symmetry is a continuous symmetry and one can continuously rotate the purely electric solution into the purely magnetic one. In the case $a = 1$ there is no obvious reason for this to happen. However, the $a = 1$ model is a truncation of the $N = 4, d = 4$ SUEGRA action that we are going to see next, which does have a continuous electric–magnetic-duality symmetry.

The dyonic solutions take the form[5] [432, 447, 612]

$$\boxed{\begin{aligned}
ds^2 &= (H_1H_2)^{-1}Wdt^2 - H_1H_2\left[W^{-1}dr^2 + r^2 d\Omega^2_{(2)}\right],\\
A_t &= \frac{-4G_N^{(4)}e^{\varphi_0}q}{r_- - G_N^{(4)}\Sigma}(H_1^{-1}-1), \qquad \tilde{A}_t = \frac{1}{4\pi}\frac{e^{-\varphi_0}p}{r_- + G_N^{(4)}\Sigma}(H_2^{-1}-1),\\
e^{-2\varphi} &= e^{-2\varphi_0}H_1/H_2,\\
H_1 &= 1 + \frac{r_- - G_N^{(4)}\Sigma}{r}, \qquad H_2 = 1 + \frac{r_- + G_N^{(4)}\Sigma}{r},\\
W &= 1 - \frac{2r_0}{r}, \qquad r_\pm = M \pm r_0,\\
r_0^2 &= M^2 + \Sigma^2 - 4\left[e^{2\varphi_0}q^2 + e^{-2\varphi_0}\left(\frac{p}{16\pi G_N^{(4)}}\right)^2\right],\\
\Sigma &= \frac{2}{M}\left[e^{2\varphi_0}q^2 - e^{-2\varphi_0}\left(\frac{p}{16\pi G_N^{(4)}}\right)^2\right].
\end{aligned}} \tag{12.39}$$

[4] There is no theorem ensuring this, but all attempts to build dyonic solutions for other values of a have been unsuccessful.

[5] Solutions with additional scalar hair are also possible, but we will not deal with them any further.

Here we have used the S-dual potential $\tilde{A}_\mu$ which is the potential related to the S-dual field strength,

$$\tilde{F}_{\mu\nu} = 2\partial_{[\mu}\tilde{A}_{\nu]}, \tag{12.40}$$

whose existence is ensured by the equation of motion of A_μ, which is just the Bianchi identity for $\tilde{F}_{\mu\nu}$. Knowledge of the electric components F_{tr} and $\tilde{F}_{tr}$ and of the metric and dilaton is enough to find all the components $F_{\mu\nu}$, but this form of presenting the result is more elegant and convenient since it exhibits the symmetries of the theory acting on the solution. In particular, we see that S duality interchanges A_μ and $\tilde{A}_\mu$ and q and $p/(16\pi G_N{}^{(4)})$, and takes φ_0 to $-\varphi_0$, which also takes Σ to $-\Sigma$.

The purely electric dilaton BH solutions with $a = 1$ are recovered when $H_2 = 1$ and the purely magnetic ones when $H_1 = 1$. When $H_1 = H_2 = H$ the scalar becomes trivial and we recover the RN solutions. Thus, these solutions are the most general from the point of view of electric–magnetic duality.

As usual, when $W = 1$, H_1 and H_2 can be arbitrary harmonic functions in three-dimensional Euclidean space. They may but need not have coincident poles and, thus, the solutions describe electric and magnetic monopoles and dyons in static equilibrium.

Solutions of the four-dimensional ($a = 1$)-model with primary scalar hair and electric charge have been presented in [19] and probably can be generalized to all values of a and to higher dimensions. We will not pursue this issue any further.

12.2 Dilaton/axion black holes

The a-model is a good starting point from which to study BH solutions of supergravity/superstring theories, but it is clearly too simple. It is natural to introduce successive generalizations to this model that make it closer to the real thing. In higher dimensions we can introduce differential-form potentials of higher rank, but these are associated with extended objects. In four dimensions we can introduce, as a first step, additional vector fields, all of them coupled in the same way to the scalar field. Then, we can introduce new scalars or different couplings of the scalar(s) to the vector fields. We would have an action of the form

$$S = \frac{1}{16\pi G_{\rm N}^{(4)}} \int d^4x \sqrt{|g|} \left[R + \tfrac{1}{2} g_{ij}\partial_\mu\varphi^i\partial^\mu\varphi^j - \tfrac{1}{4} M_{ij} F^i{}_{\mu\nu} F^{j\,\mu\nu}\right], \tag{12.41}$$

where $g^{ij}(\varphi)$ and $M_{ij}(\varphi)$ are some square matrices depending on the scalars. g^{ij} can be interpreted as the inverse metric of some space of which the scalars φ_i are the coordinates. The scalar kinetic term is a *σ-model*.

A good example of an action of this kind is provided by the four-dimensional KK action that one obtains from $\hat{d} = 4 + N$ dimensions by compactification on T^N, Eq. (11.195). The scalars parametrize an $\mathbb{R}^+ \times \mathrm{SL}(N, \mathbb{R})/\mathrm{SO}(2)$ coset space.

There is another kind of couplings of scalars to vectors that we can introduce in four dimensions: couplings of the form

$$-\tfrac{1}{4} N_{ij}(\varphi) F^i{}_{\mu\nu} {}^\star F^{j\,\mu\nu}. \tag{12.42}$$

As a matter of fact, the bosonic sectors of all four-dimensional SUEGRAs can be written in this form. Each of them is characterized by the number of vectors and scalars, by the σ-model metric $g_{ij}(\varphi)$, and by the matrices of couplings $M_{ij}(\varphi)$ and $N_{ij}(\varphi)$. The general case is studied in [445].

The simplest model with a coupling of the above kind is the so-called *axion/dilaton-gravity* model

$$\boxed{S = \frac{1}{16\pi G_{\mathrm{N}}^{(4)}}\int d^4x\sqrt{|g|}\left[R + 2(\partial\varphi)^2 + \tfrac{1}{2}e^{4\varphi}(\partial a)^2 - e^{-2\varphi}F^2 + aF^{\star}F\right].} \tag{12.43}$$

The scalar field that couples to $F^{\star}F$ is called the *axion* and should be a pseudoscalar for the above action to be parity-invariant. It plays the role of a local θ-parameter (see Eq. (8.178)) just as the dilaton plays the role of local coupling constant. In fact, this model is a version of the one studied in Section 8.7.4 with local coupling constants (moduli) and, as we are going to see, it exhibits the same S-duality symmetry [821].

1. The factor $e^{4\varphi}$ of the axion kinetic term allows us to combine the axion and the dilaton into a complex scalar field, the *axidilaton* τ;

$$\tau = a + ie^{-2\varphi}, \tag{12.44}$$

   and its kinetic term takes the form of an $\mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$ σ-model, Eq. (11.209). We can also use the symmetric $\mathrm{SL}(2,\mathbb{R})$ matrix $\mathcal{M}$ defined in Eq. (11.207). As discussed in Sections 11.4.1, the σ-model is invariant under global $\mathrm{SL}(2,\mathbb{R})$ transformations that are fractional-linear transformations of τ given by Eqs. (11.205) and (11.206). This group contains three different kinds of transformations.

   (a) Rescalings of τ:

$$S = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{pmatrix}, \qquad \tau' = \alpha^2\tau. \tag{12.45}$$

   These transformations rescale the axion and shift the value of the dilaton at infinity, $\varphi_0' = \varphi_0 - \ln\alpha$.

   (b) Constant shifts:

$$S = \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix}, \qquad \tau' = \tau + \beta. \tag{12.46}$$

   These transformations only shift the value of the axion at infinity, $a_0' = a_0 + \beta$.

   (c) SO(2) rotations:

$$S = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \qquad \tau' = \frac{\cos\theta\ \tau + \sin\theta}{-\sin\theta\ \tau + \cos\theta}. \tag{12.47}$$

   The rotation with $\theta = \pi/2$ inverts τ:

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad \tau' = -1/\tau. \tag{12.48}$$

When $a = 0$ this transformation is just the electric–magnetic-duality transformation of the dilaton model $\varphi' = -\varphi$.

2. The action is invariant under the first two kinds of transformations: the rescalings of τ can be compensated by opposite rescalings of F:

$$F' = \frac{1}{\alpha} F, \tag{12.49}$$

and the shifts of a simply change the action by a total derivative $\beta\sqrt{|g|}F\,{}^\star F$. This is just an Abelian version of the Peccei–Quinn symmetry. In the Euclidean non-Abelian SU(2) case the total derivative is proportional to a topological invariant; namely the second Chern class defined in Eq. (9.18) that takes integer values. If the Euclidean action is properly normalized, the Peccei–Quinn transformation simply shifts it by β times an integer, which results in a phase change in the integrand of the path integral. Thus, the classical continuous Peccei–Quinn symmetry is broken to $\mathbb{Z}$ since the only transformations that leave the path integral invariant are those with $\beta = 2\pi n$, $n \in \mathbb{Z}$. This is one of the quantum effects[6] that breaks SL(2, $\mathbb{R}$) to SL(2, $\mathbb{Z}$), the group of S duality.

3. The equations of motion (but not the action) of the whole theory are also invariant under SO(2) rotations. To see this (to check invariance under the whole SL(2, $\mathbb{R}$)), it is convenient to define the $SL(2, \mathbb{R})$-dual $\tilde{F}$ of the vector-field strength F:

$$\tilde{F}_{\mu\nu} \equiv e^{-2\varphi}\,{}^\star F_{\mu\nu} + aF_{\mu\nu}. \tag{12.50}$$

The Maxwell equation is now the Bianchi identity of the S-dual field strength:

$$\nabla_\mu{}^\star\tilde{F}^{\mu\nu} = 0. \tag{12.51}$$

It is convenient to define two S-duality vectors $\vec{F}$ and $\vec{\mathcal{F}}$,

$$\vec{F} \equiv \begin{pmatrix} {}^\star F \\ F \end{pmatrix}, \qquad \vec{\mathcal{F}} \equiv e^{-\varphi}\mathcal{V}\vec{F} = \begin{pmatrix} \tilde{F} \\ F \end{pmatrix}, \tag{12.52}$$

where $\mathcal{V}$ is the upper-triangular unimodular matrix that we defined in Eq. (11.208) that satisfies $\mathcal{V}\mathcal{V}^{\mathrm{T}} = \mathcal{M}$. $\vec{\mathcal{F}}$ transforms covariantly under $S \in$ SL(2, $\mathbb{R}$):

$$\vec{\mathcal{F}}' = S\vec{\mathcal{F}}. \tag{12.53}$$

The two components of this vector are not independent, but are related by a constraint that involves τ. This constraint must be preserved by S and one can check that this happens if, and only if, τ transforms according to Eqs. (11.205) and (11.206). The transformation $\tau' = -1/\tau$ interchanges the two components of the duality vector. For a vanishing axion field, this is the discrete electric–magnetic-duality transformation of the dilaton-gravity model.

[6] The other one is charge quantization.

In terms of the duality vector $\vec{\mathcal{F}}$ the Maxwell equation and the Bianchi identity take the SL(2, $\mathbb{R}$)-invariant form

$$\nabla_\mu \vec{\mathcal{F}}^{\mu\nu} = 0, \tag{12.54}$$

and the Einstein equation can also be written in invariant form (see Eq. (8.138)):

$$R_{\mu\nu} + \frac{\partial_\mu \tau \partial_\nu \bar{\tau}}{(\mathrm{Im}(\tau))^2} + \vec{\mathcal{F}}^{\mathrm{T}} \eta^\star \vec{\mathcal{F}} = 0, \tag{12.55}$$

with $\eta = i\sigma^2$, due to the property $S\eta S^{\mathrm{T}} = \eta$ of $\mathrm{Sp}(2) \sim \mathrm{SL}(2,\mathbb{R})$ matrices S. The remaining two equations of motion,

$$\begin{aligned} &\nabla^2\varphi - \tfrac{1}{2}e^{4\varphi}(\partial a)^2 - \tfrac{1}{2}e^{-2\varphi}F^2 = 0,\\ &\nabla^2 a + 4\partial_\mu\varphi\, \partial^\mu a - e^{-4\varphi}F\,{}^\star F = 0, \end{aligned} \tag{12.56}$$

can also be rewritten in a manifestly duality-invariant form:

$$\nabla_\mu\left(\partial^\mu \mathcal{M}\mathcal{M}^{-1}\right) + \vec{\mathcal{F}}\vec{\mathcal{F}}^{\mathrm{T}}\eta = 0. \tag{12.57}$$

The action Eq. (12.43) is a truncation of the bosonic sector of ungauged $N = 4, d = 4$ SUEGRA [266], that contains the metric $g_{\mu\nu}$, complex scalar τ, and six Abelian vector fields $A^{(n)}{}_\mu$, $n = 1, \ldots, 6$. On setting $G_{\mathrm{N}}^{(4)} = 1$, it takes the form

$$\boxed{S = \frac{1}{16\pi}\int d^4x\sqrt{|g|}\left[R + 2(\partial\varphi)^2 + \tfrac{1}{2}e^{4\varphi}(\partial a)^2 - e^{-2\varphi}F^{(n)}F^{(n)} + aF^{(n)}\,{}^\star F^{(n)}\right].} \tag{12.58}$$

This theory, in turn, can be obtained by dimensional reduction and consistent truncation from $N = 1, d = 10$ SUGRA, the effective theory of the heterotic string, as we will see in Chapter 16. In this context φ coincides with the four-dimensional string dilaton and there are many things about the general stringy case that we can learn by studying this simpler case.

Apart from S duality, this action has a trivial invariance under SO(6) (T-duality) rotations of the vector fields. This may seem to suggest that considering just one vector field would be enough to obtain the most general BH solution (up to SO(6) rotations), but we are going to see that this is not the case: at least two vectors are needed if one wants to obtain a BH solution from which we can generate the most general one[7] by more or less trivial SO(6) rotations (a *generating solution*). To explain why this is the case, we need to discuss how the conserved charges enter in the metric and scalar fields.

First we use Eq. (12.54) to define the conserved electric and magnetic charges of the six Abelian vector fields $\vec{q}^{(n)}$,

$$\vec{q}^{(n)} \equiv \frac{1}{4\pi}\int_{S^2_\infty} \vec{\mathcal{F}}^{(n)}, \qquad \vec{q}^{(n)} = \begin{pmatrix} q^{(n)} \\ p^{(n)} \end{pmatrix}, \tag{12.59}$$

[7] By definition, the one with the highest possible number of charges (mass, angular momentum, and electric and magnetic charges) and moduli (the asymptotic value of τ) allowed by the no-hair conjecture.

that we can arrange into a twelve-dimensional vector $\mathbf{q}$. $\mathbf{q}$ transforms linearly under S and T-duality transformations S and R:

$$\mathbf{q}' = S \otimes R\mathbf{q}. \tag{12.60}$$

The charges[8] must enter into the metric in duality-invariant combinations because the metric is duality-invariant. There are only two such invariants that are quadratic and quartic in the charges:

$$I_2 \equiv \mathbf{q}^{\mathrm{T}} \mathcal{M}_0^{-1} \otimes \mathbb{I}_{6\times 6}\mathbf{q}, \qquad I_4 \equiv \det\left[\sum_{n=1}^{n=6} \vec{q}^{\,(n)} \vec{q}^{\,(n)\,\mathrm{T}}\right]. \tag{12.62}$$

Here $\mathcal{M}_0$ is the asymptotic value of the scalar matrix $\mathcal{M}$. Thus, I_2 is moduli-dependent and I_4 is moduli-independent. On the other hand, I_4 vanishes when only one vector field is non-trivial and, therefore, starting from the most general charge configuration with only one vector field and $I_4 = 0$, we cannot generate the most general charge configuration with $I_4 \neq 0$ by S- and T-duality transformations. The generating solution has to have both I_2 and I_4 generically non-vanishing.

To attain a better understanding, we can try to construct the most general solution starting from the $d = 4$, $a = 1$ dilaton BH solutions we studied in the previous section. We simply have to observe that the equations of motion of the axion/dilaton model coincide[9] with those of the four-dimensional $a = 1$ model if the axion $a = 0$ and $F^{\star}F = 0$. Then, the purely electric BH Eq. (12.23) provides a solution of the axion/dilaton model with one independent charge and one non-trivial modulus (φ_0). By performing one SO(2) S-duality transformations Eq. (12.47), we can generate a solution that has electric and magnetic charge. As in the Einstein–Maxwell case, the SO(2) parameter becomes a new independent charge. A non-trivial axion is generated. Further $\mathrm{SL}(2, \mathbb{R})$ transformations only shift φ_0 and add an asymptotic value to the axion a_0. In this way we have obtained the most general axion/dilaton BH solution with one vector field [850], but it has the same metric as the purely dilatonic BH.

This solution is also a solution of $N = 4, d = 4$ SUEGRA with five vanishing vector fields. We could excite them by performing $\mathrm{SO}(6)/\mathrm{SO}(5)$ T-duality rotations that do not leave the charge vector invariant. However, in this way we can obtain only solutions in which all the magnetic charges are proportional to all the electric charges with the same proportionality factor. We would have added only five new independent parameters to the solution and the metric would still be the same (because $I_4 = 0$).

A more general solution with two non-vanishing charges in different vectors $q^{(1)}$ and $p^{(2)}$ was found in [432] and, later on, studied in [612]. It has a different metric (and

[8] Observe that, the axion being a local θ-parameter, it induces a Witten effect on the charges, as explained in Section 8.7.4. Furthermore, the DSZ quantization condition takes the manifestly $\mathrm{SL}(2, \mathbb{R})$-invariant form

$$\vec{q}_1^{\,(n)\,\mathrm{T}} \eta\, \vec{q}_2^{\,(n)} = m/2, \qquad m \in \mathbb{Z}. \tag{12.61}$$

($q^{(n)}$ is canonically normalized, but $p^{(n)}$ is $1/(4\pi)$ times the canonical magnetic charge. The product of the canonical charges is quantized in integer multiples of 2π.)

[9] The vector fields have a different normalization.

non-vanishing I_4). It has a non-vanishing dilaton, a vanishing axion, and trivial moduli, which, however, could be generated by S-duality transformations. In fact, it is clear that S and T dualities suffice to generate the four possible independent charges of the two vector fields and, actually, the $2N$ independent charges of N vector fields and, thus, it is the (static) generating solution of this theory.

This static generating solution is essentially the $d = 4$, $a = 1$ dyonic dilaton BH solution given in Eq. (12.39) but where the electric and magnetic components of the vector field belong to two different vector fields.[10] In the conventions that we are using in this section, it takes the form

$$\boxed{\begin{aligned}
ds^2 &= (H_1 H_2)^{-1} W dt^2 - H_1 H_2 \big[W^{-1} dr^2 + r^2 d\Omega^2_{(2)}\big], \\
A^{(1)}{}_t &= \frac{-q}{r_- - \Sigma}(H_1^{-1} - 1), \qquad \tilde{A}^{(2)}{}_t = \frac{p}{r_- + \Sigma}(H_2^{-1} - 1), \\
e^{-2\varphi} &= H_1/H_2, \\
H_1 &= 1 + \frac{r_- - \Sigma}{r}, \qquad H_2 = 1 + \frac{r_- + \Sigma}{r}, \qquad W = 1 - \frac{2r_0}{r}, \\
r_\pm &= M \pm r_0, \qquad r_0^2 = M^2 + \Sigma^2 - (q^2 + p^2), \qquad \Sigma = 2(q^2 - p^2)/M.
\end{aligned}}$$

(12.63)

It is, however, very convenient to have the most general solution written explicitly in terms of the physical charges. Moreover, the most general static solution can be immediately generalized in a natural way by adding angular momentum and NUT charge, becoming the truly most general stationary BH-type solution that we will call the *SWIP solution*[11] [665]. It will be S- and T-duality-invariant by definition, and its physical properties will be given in terms of duality-invariant combinations of charges. Ungauged $N = 4$, $d = 4$ SUEGRA is the most complicated case in which the most general solution is explicitly known and the attempts to write the most general solution of more complicated theories are inspired by it. For these reasons, it is worth studying.

12.2.1 The general SWIP solution

The general solution is determined by two complex harmonic functions, $\mathcal{H}_{1,2}$, the non-extremality function, W, the spatial background metric, ${}^{(3)}\gamma_{ij}$, and N complex constants

[10] Sometimes these solutions are called $U(1)^2$ BHs.

[11] The construction of the most general BH-type solution was initiated in [743] and the most general static solution was obtained in [613]. There, the solution was written in terms of two complex functions $\mathcal{H}_{1,2}$ (harmonic in the extreme limit) that obeyed a constraint. It was realized in [132] that removing the constraint in the extreme case immediately resulted in the natural inclusion of NUT charge and angular momentum. The new solutions had been obtained independently in [894]. Finally, the general, non-extreme solution was constructed in [665]. Related work was done in [67, 245, 409–13, 415, 610, 806–9].

$k^{(n)}$:

$$\boxed{\begin{array}{l} ds^2 = e^{2U} W\big(dt + A_\varphi d\varphi\big)^2 - e^{-2U} W^{-1}\, {}^{(3)}\gamma_{ij} dx^i dx^j, \\ A^{(n)}{}_t = 2e^{2U}\,\mathrm{Re}\big(k^{(n)}\mathcal{H}_2\big), \qquad \tilde{A}^{(n)}{}_t = 2e^{2U}\,\mathrm{Re}\big(k^{(n)}\mathcal{H}_1\big), \qquad \tau = \mathcal{H}_1/\mathcal{H}_2. \end{array}} \tag{12.64}$$

where

$$\begin{aligned} e^{-2U} &= 2\,\mathrm{Im}\big(\mathcal{H}_1\bar{\mathcal{H}}_2\big), \\ A_\varphi &= 2N\cos\theta + \alpha\sin^2\theta\left(e^{-2U}W^{-1} - 1\right). \end{aligned} \tag{12.65}$$

The functions $\mathcal{H}_{1,2}$ take the form

$$\mathcal{H}_1 = \frac{1}{\sqrt{2}} e^{\varphi_0} e^{i\beta}\left(\tau_0 + \frac{\tau_0\mathfrak{M} + \bar{\tau}_0\Upsilon}{r + i\alpha\cos\theta}\right), \qquad \mathcal{H}_2 = \frac{1}{\sqrt{2}} e^{\varphi_0} e^{i\beta}\left(1 + \frac{\mathfrak{M} + \Upsilon}{r + i\alpha\cos\theta}\right), \tag{12.66}$$

and W and the background metric ${}^{(3)}\gamma_{ij}$ take the forms

$$\begin{aligned} W &= 1 - \frac{r_0{}^2}{r^2 + \alpha^2\cos^2\theta}, \\ {}^{(3)}\gamma_{ij} dx^i dx^j &= \frac{r^2 + \alpha^2\cos^2\theta - r_0^2}{r^2 + \alpha^2 - r_0^2}\, dr^2 + \left(r^2 + \alpha^2\cos^2\theta - r_0{}^2\right) d\theta^2 \\ &\quad + \left(r^2 + \alpha^2 - r_0{}^2\right)\sin^2\theta\, d\varphi^2. \end{aligned} \tag{12.67}$$

The complex constants are given by

$$k^{(n)} = -\frac{1}{\sqrt{2}} e^{-i\beta}\frac{\mathfrak{M}\Gamma^{(n)} + \overline{\Upsilon\Gamma^{(n)}}}{|\mathfrak{M}|^2 - |\Upsilon|^2}. \tag{12.68}$$

The metric can also be written in a more standard form:

$$\begin{aligned} ds^2 &= \frac{\Delta - \alpha^2\sin^2\theta}{\Sigma}\, dt^2 + 2\alpha\sin^2\theta\,\frac{\Sigma + \alpha^2\sin^2\theta - \Delta}{\Sigma}\, dt d\varphi \\ &\quad - \frac{\Sigma}{\Delta}\, dr^2 - \Sigma\, d\theta^2 - \frac{\left(\Sigma + \alpha^2\sin^2\theta\right)^2 - \Delta\alpha^2\sin^2\theta}{\Sigma}\sin^2\theta\, d\varphi^2, \\ \Delta &= r^2 - R_0{}^2 = r^2 + \alpha^2 - r_0{}^2, \\ \Sigma &= (r + M)^2 + (n + \alpha\cos\theta)^2 - |\Upsilon|^2. \end{aligned} \tag{12.69}$$

We have expressed the functions that enter the solution in terms of physical constants (charges and moduli). $\alpha = J/M$ is the angular momentum (J) per unit mass (M), and we have combined the mass and NUT charge (N) into the complex "mass"

$$\mathfrak{M} \equiv M + iN, \tag{12.70}$$

and the electric and magnetic charges into

$$\Gamma^{(n)} \equiv Q^{(n)} + iP^{(n)}, \qquad \vec{Q}^{(n)} \equiv \mathcal{V}_0^{-1}\vec{q}^{(n)}. \tag{12.71}$$

Υ, the (complex) axion/dilaton charge, and τ_0, its asymptotic value, are defined by

$$\tau \sim \tau_0 - ie^{-2\varphi_0}\frac{2\Upsilon}{r}. \tag{12.72}$$

In these solutions Υ depends on the conserved charges in this fixed way:

$$\Upsilon = -\tfrac{1}{2}\sum_n \frac{(\bar{\Gamma}^{(n)})^2}{\mathfrak{M}}. \tag{12.73}$$

Finally, the "non-extremality" parameter r_0 is given by

$$r_0^2 = |\mathfrak{M}|^2 + |\Upsilon|^2 - \sum_n |\Gamma^{(n)}|^2. \tag{12.74}$$

In non-static cases when $r_0 = 0$ the solution is *supersymmetric*, but for $\alpha \neq 0$ it is not an extreme BH. A more appropriate name is *supersymmetry parameter*. The *extremality parameter* will be $R_0^2 = r_0^2 - \alpha^2$. When it is positive, we have two horizons placed at $r_\pm = M \pm R_0$. The area of the event horizon (the one at r_+) is given, for BH solutions with zero NUT charge, by

$$A = 4\pi\left(r_+{}^2 + \alpha^2 - |\Upsilon|^2\right). \tag{12.75}$$

12.2.2 Supersymmetric SWIP solutions

When $r_0 = 0$ $W = 1$ the general SWIP solution has special properties. First, the background metric ${}^{(3)}\gamma_{ij}$ is nothing but the metric of Euclidean three-dimensional space in oblate spheroidal coordinates, which are related to the ordinary Cartesian ones by

$$\begin{aligned} x &= \sqrt{r^2+\alpha^2}\ \sin\theta\cos\varphi, \\ y &= \sqrt{r^2+\alpha^2}\ \sin\theta\sin\varphi, \\ z &= r\cos\theta. \end{aligned} \tag{12.76}$$

On rewriting the solution Eqs. (12.64) in Cartesian coordinates, we find the solutions

$$\boxed{\begin{aligned} ds^2 &= 2\,\mathrm{Im}(\mathcal{H}_1\bar{\mathcal{H}}_2)\,(dt + A)^2 - [2\,\mathrm{Im}(\mathcal{H}_1\bar{\mathcal{H}}_2)]^{-1}d\vec{x}_3^{\,2}, \\ A^{(n)}{}_t &= 2e^{2U}\,\mathrm{Re}\big(k^{(n)}\mathcal{H}_2\big), \qquad \tilde{A}^{(n)}{}_t = 2e^{2U}\,\mathrm{Re}\big(k^{(n)}\mathcal{H}_1\big), \qquad \tau = \mathcal{H}_1/\mathcal{H}_2\,. \\ A &= A_{\underline{i}}dx^i, \qquad \epsilon_{ijk}\partial_i A_j = \pm\mathrm{Re}\big(\mathcal{H}_1\partial_k\bar{\mathcal{H}}_2 - \bar{\mathcal{H}}_2\partial_k\mathcal{H}_1\big), \\ \partial_{\underline{i}}\partial_{\underline{i}}\mathcal{H}_{1,2} &= 0, \qquad \sum_{n=1}^{N}(k^{(n)})^2 = 0, \qquad \sum_{n=1}^{N}|k^{(n)}|^2 = \frac{1}{2}. \end{aligned}} \tag{12.77}$$

That is, for any arbitrary pair of complex harmonic functions $\mathcal{H}_{1,2}(\vec{x}_3)$ in the three-dimensional Euclidean space, it is clear that we can construct multi-BH solutions and that $r_0 = 0$ can be reinterpreted as a no-force condition between the BHs.

These solutions include the IWP metrics Eqs. (9.58) when

$$\mathcal{H}_1 = i\mathcal{H}_2 = \frac{1}{\sqrt{2}}\mathcal{H} \tag{12.78}$$

(which trivializes the axidilaton τ) that in turn include the MP solutions Eqs. (8.86). These are the only BH-type solutions in the IWP family: the addition of angular momentum eliminates the event horizon and the addition of NUT charge eliminates the asymptotic flatness. Something similar is true for the supersymmetric SWIP solutions above: the only supersymmetric BHs in this family are the static ones with $A_i = 0$, which imposes a very non-trivial constraint on the harmonic functions, which was implicit in [613].

The general solutions include, for vanishing axidilaton charge $\Upsilon = 0$ (which corresponds to special choices of the electric and magnetic charges), the Kerr–Newman solution in Boyer–Lindquist coordinates Eq. (9.55).

12.2.3 Duality properties of the SWIP solutions

Solutions of the general and supersymmetric SWIP families are the most general BH-type solutions of $N = 4$, $d = 4$ SUEGRA and, therefore, an S- or T-duality transformation takes one member of the family into another member of the family. Thus, the effect of duality transformations is just to replace all the constants and functions that enter the solutions with primed constants and functions. The structure of the solutions thus reflects the $\mathrm{SL}(2,\mathbb{R})\times \mathrm{SO}(6)$ duality invariance of the equations of motion.

Let us see in a bit more detail how the charges and functions transform under duality. $\mathfrak{M}$ is obviously invariant. The complex combinations of electric and magnetic charges $\Gamma^{(n)}$ are SO(6) vectors and change by a phase under $\mathrm{SL}(2,\mathbb{R})$,

$$\Gamma^{(n)\,\prime} = e^{i\,\arg(\gamma\tau_0+\delta)}\Gamma^{(n)}, \tag{12.79}$$

while the axidilaton charge also changes by a phase but is an SO(6) scalar,

$$\Upsilon' = e^{-2i\,\arg(\gamma\tau_0+\delta)}\Upsilon, \tag{12.80}$$

and, therefore, its absolute value is duality-invariant and can be expressed in terms of the two invariants I_2 and I_4:

$$|\Upsilon|^2 = \frac{1}{4|\mathfrak{M}|^2}(I_2^2 - 4I_4). \tag{12.81}$$

It is also easy to show that

$$\sum_n |\Gamma^{(n)}|^2 = I_2. \tag{12.82}$$

Since $\mathfrak{M}$ is trivially duality-invariant, the last two equations imply the duality invariance of the supersymmetry parameter r_0, given in Eq. (12.74), and of the *supersymmetry bound* (to be defined in Chapter 13) $r_0^2 \geq 0$.

It is useful to define the two combinations of charges [130]

$$|Z_{1,2}|^2 \equiv \tfrac{1}{2}I_2 \pm I_4^{\frac{1}{2}}, \tag{12.83}$$

which are at most interchanged by duality transformations. In terms of them, the supersymmetry parameter and the BH entropy take the suggestive forms

$$r_0^2 = \frac{1}{|\mathfrak{M}|^2}\big(|\mathfrak{M}|^2 - |Z_1|^2\big)\big(|\mathfrak{M}|^2 - |Z_2|^2\big).$$

$$S = \pi\left\{(M^2 - |Z_1|^2) + (M^2 - |Z_2|^2) + 2\sqrt{(M^2 - |Z_1|^2)(M^2 - |Z_2|^2) - J^2}\right\}, \tag{12.84}$$

which we will discuss in Chapter 13. Observe that $|\Upsilon|$ is given in general by $|Z_1 Z_2|^2\mathfrak{M}^{-2}$.

The functions $\mathcal{H}_{1,2}$ transform as a doublet under SL(2, $\mathbb{R}$), whereas the $k^{(n)}$s are invariant because, although they transform with the same phase as $\Gamma^{(n)}$, they can be absorbed into the arbitrary phase β that appears in the solution. The $k^{(n)}$s are clearly SO(6) vectors, as are the corresponding vector potentials.

12.2.4 $N = 2, d = 4$ SUGRA solutions

The form of the supersymmetric SWIP solutions strongly reflects the structure of the dualities of the theory and suggested to the authors of [382] a relation to the special-geometry formalism of $N = 2, d = 4$ SUGRA theories[12] [222, 225, 265, 317, 868] that describes the geometry of the scalar manifold (the space in which the scalars φ_i of the theory take values and, hence, the σ-model metric $g_{ij}(\varphi)$ in the action Eq. (12.41)), the couplings of the scalars to the vector fields (the functions $M_{ij}(\varphi)$ and $N_{ij}(\varphi)$), and, for gauged SUGRAS, the gauge groups and the scalar potential. Pure $N = 4, d = 4$ SUGRA with only two vector fields (which still supports the most general SWIP solution) can be seen as $N = 2, d = 4$ SUGRA coupled to an $N = 2$ vector multiplet with two new "accidental" supersymmetries just as pure $N = 2$ SUGRA (Section 5.5) can be seen as $N = 1$ coupled to a vector multiplet [380] with one "accidental" supersymmetry and, therefore, that formalism can be applied to it. There are many other theories arising from compactifications of ten-dimensional superstring effective actions that can be described with this formalism.

The coupling of n $N = 2$ vector multiplets to $N = 2$ SUGRA can, in some cases, be completely described by a *prepotential* function F of the complex projective coordinates X^Λ, $\Lambda = 0, 1, \ldots, n$, that parametrize the scalar manifold. From F one can derive the *Kähler potential* K,

$$K = -\ln\big(N_{\Lambda\Sigma}X^\Lambda X^\Sigma\big), \qquad N_{\Lambda\Sigma} = \tfrac{1}{2}\,\mathrm{Re}(\partial_\Lambda\partial_\Sigma F), \tag{12.85}$$

from which the Kähler metric of the scalar σ-model,

$$g_{i\bar{j}} = \frac{\partial^2 K}{\partial\varphi^i\partial\bar{\varphi}^j}, \qquad \varphi^i \equiv X^i/X^0, \qquad i = 1, \ldots, n, \tag{12.86}$$

the *chiral connection* $\mathcal{A}_\mu$,

$$\mathcal{A}_\mu = \frac{i}{2}N_{\Lambda\Sigma}\Big[\overline{X}^\Lambda\partial_\mu X^\Sigma - (\partial_\mu\overline{X}^\Lambda)X^\Sigma\Big], \tag{12.87}$$

and also the couplings of the scalars to the vector fields can be derived.

[12] For an introduction to the special-geometry formalism of $N = 2$ supergravity, see e.g. [400, 402].

The most general BH-type solution of an $N = 2$ theory has to be duality-invariant and thus has to be built out of the only invariants that the special-geometry formalism contains: the Kähler potential and the chiral connection. In [382] it was realized that the metric for extreme BHs in $N = 2$ theories can always be written in the form

$$ds^2 = e^K dt^2 - e^{-K} d\vec{x}^{\,2}, \tag{12.88}$$

where the projective coordinates X^Λ are identified with real harmonic functions H^Λ that are also related to the $n + 1$ U(1) vector potentials of the theory. In [132] it was realized that one could also use complex harmonic functions, and then the 1-form A_i that appears in non-static SWIP BH-solutions

$$ds^2 = e^K (dt^2 + A_i dx^i)^2 - e^{-K} d\vec{x}^{\,2}, \tag{12.89}$$

is related to the chiral 1-form of the $N = 2$ SUGRA theory by

$$\epsilon_{ijk} \partial_{\underline{j}} A_{\underline{k}} = \mathcal{A}_{\underline{i}}. \tag{12.90}$$

More precisely, $N = 4, d = 4$ SUGRA with only two vector fields corresponds to an $N = 2, d = 4$ SUGRA with prepotential $F = 2X^0 X^1$. The axidilaton is just $\tau = X^1/X^0$. It is a simple exercise to check that the above recipe, with

$$X^0 = i\mathcal{H}_2, \qquad X^1 = \mathcal{H}_1, \tag{12.91}$$

gives the SWIP solutions.

It is natural to conjecture that the same (or a similar) recipe should work in more general cases since the basic principle of correspondence between components of the metric and special-geometry invariants should be valid.[13] However, in practice, the SWIP solutions remain the only solutions whose complete explicit form is known. Also, from our experience with the general (non-supersymmetric) SWIP solutions, it is to be expected that general (non-supersymmetric) BH-type solutions of $N = 2$ SUEGRA can also be constructed by introducing non-extremality functions and a background metric.

[13] The construction of extreme BH solutions of $N = 2$ SUEGRAs is reviewed in [43, 57, 708, 709].

13
Unbroken supersymmetry

In our study of several solutions in the previous chapters we have mentioned that some special properties that arise for special values of the parameters (mass, charges) are related to supersymmetry; more precisely, to the existence of (*unbroken*) supersymmetry. Those statements were a bit surprising because we were dealing with solutions of purely bosonic theories (Einstein–Maxwell, Kaluza–Klein . . .).

The goal of this chapter is to explain the concept and implications of unbroken supersymmetry and how it can be applied in purely bosonic contexts, including pure GR. Supersymmetry will be shown to have a very deep meaning, underlying more familiar symmetries that can be constructed as squares of supersymmetries. At the very least, supersymmetry can be considered as an extremely useful tool that simplifies many calculations and demonstrations of very important results in GR that are related directly or indirectly to the positivity of energy (a manifest property of supersymmetric theories).

As a further reason to devote a full chapter to this topic, unbroken supersymmetry is a crucial ingredient in the stringy calculation of the BH entropy by the counting of microstates. It ensures the stability of the solution and the calculation under classical and quantum perturbations.

To place this subject in a wider context, we will start by giving in Section 13.1 a general definition of residual (unbroken) symmetry and we will relate it to the definition of a vacuum. Vacua are characterized by their symmetries, which determine the conserved charges of point-particles moving in them and, ultimately, the spectra of quantum-field theories (QFTs) defined on them. These definitions will be applied in Section 13.2 to supersymmetry as a particular case. In this section we will have to develop a new tool, the *covariant Lie derivative*, which will be used to find the unbroken-supersymmetry algebra of any given solution according to Figueroa-O'Farrill's prescription in [390]. In Section 13.3 we will apply this prescription and the geometrical methods of [25] to the vacua of the simplest four-dimensional supergravity theories and we will try to recover the supersymmetry algebras that we gauged to construct them in Chapter 5. These vacuum superalgebras will then be used in Section 13.5 to understand the properties of other solutions (with or without unbroken supersymmetry) with the same asymptotic behavior. In particular, they can be used to derive *supersymmetry* or *BPS bounds*. We will also discuss the results known for minimal $d = 5, 6$ supergravities, but we will leave higher-dimensional supergravities and theories with more supercharges for Part III because these theories can be derived from ten-dimensional superstring effective theories, but we will say what can be expected from general arguments based on the structures of the respective superalgebras.

In Section 13.5 we will study the properties of solutions with partially broken supersymmetry that cannot be considered vacua but instead can be considered as excitations of some vacuum to which they tend asymptotically. By associating states in a quantum theory with these solutions and using the vacuum superalgebra, general supersymmetry bounds for the mass can be derived. These bounds are saturated by (*supersymmetric or "BPS"*) states with partially unbroken supersymmetry. The bounds can be extended to solutions of the theory, even in the absence of supersymmetry, if certain conditions on the energy–momentum tensor are imposed. These are very powerful techniques.

In Section 13.5.2 we will review important examples of solutions with unbroken supersymmetries in $N = 1, 2, 4$, $d = 4$ Poincaré supergravity, including the general families of supersymmetric solutions which are known only for these cases. In particular, we will discuss the relations among BH thermodynamics, cosmic censorship, and unbroken supersymmetry in these theories.

13.1 Vacuum and residual symmetries

The solutions of the equations of motion of a given theory usually break most (or all) of its symmetries. Sometimes a solution has (*preserves*) some of them, which receive the name of *residual (or unbroken) symmetries*, and, being symmetries, they form a symmetry group. The solution is said to be *symmetric*. The symmetries of the theory which are broken by the symmetric solution can be used to generate new solutions of the theory. Let us see two examples.

Classical mechanics. The Lagrangian of a free relativistic particle moving in Minkowski spacetime is invariant under the whole Poincaré group ISO(1,3). However, every solution is a straight line, invariant only under translations parallel to it and rotations with it as the axis. These are the residual symmetries of every solution and form a two-dimensional group $\mathbb{R} \times \mathrm{SO}(2)$. The remaining Poincaré transformations move the line and generate other solutions.

Field theory. Einstein's equations are invariant under the infinite-dimensional group of GCTs. However, a given solution (metric) is invariant only under a finite-dimensional group of *isometries*. By definition, an infinitesimal isometry is an infinitesimal GCT that leaves the metric invariant, that is

$$\delta_\xi g_{\mu\nu} = -\mathcal{L}_\xi g_{\mu\nu} = -2\nabla_{(\mu}\xi_{\nu)} = 0, \tag{13.1}$$

which is known as the *Killing equation*. The solutions $\xi^\mu = \xi k^\mu$ are each the product of an infinitesimal constant ξ times a *Killing vector* k^μ, the generator of the isometry.

The isometries of a metric form an isometry group. This is a finite-dimensional Lie group, whose generators are Killing vectors. The finite-dimensional Lie algebra of isometries coincides with the Lie algebra of the Killing vectors with the Killing bracket by virtue of the property of the Lie derivative

$$[\mathcal{L}_{k_1}, \mathcal{L}_{k_2}] = \mathcal{L}_{[k_1,k_2]}. \tag{13.2}$$

(This structure is induced from the infinite-dimensional group of all GCTs, of which the isometry group is a subgroup.)

Formally we can associate a generator of the abstract symmetry algebra of the solution $P_{(I)}$ with each of its Killing vectors $k_{(I)}$. This abstract generator is represented on the metric by an operator, which is just minus the Lie derivative with respect to the corresponding Killing vector $P_{(I)} \sim -\mathcal{L}_{k_{(I)}}$. Then, if the Lie algebra of the isometries is $[k_{(I)}, k_{(J)}] = -f_{IJ}{}^{K} k_{(K)}$, the abstract symmetry algebra takes the form

$$[P_{(I)}, P_{(J)}] = f_{IJ}{}^{K} P_{(K)}. \tag{13.3}$$

What happens if there are matter fields in the theory? If they are standard tensor fields[1] T, infinitesimal GCTs act on them through (minus) the Lie derivative:

$$\delta_\xi T = -\mathcal{L}_\xi T. \tag{13.4}$$

Only those GCTs that leave invariant all fields of a solution will be (unbroken) symmetries of that solution. Thus, only those isometries that leave invariant the matter fields,

$$-\mathcal{L}_{k_{(I)}} T = 0, \tag{13.5}$$

generate the symmetry algebra of the solution.

Finally, GCTs that are not symmetries transform the solution into another solution, which may be physically equivalent if the boundary conditions are invariant, but will be inequivalent otherwise.

The second example is evidently richer and more interesting. In it the presence of residual symmetries has far-reaching consequences. For instance, we have proven in Section 3.3 that point-particles moving in a curved spacetime with isometries have a conserved quantity associated with every isometry. If we construct QFTs in such a spacetime, the quanta of the fields will appear in unitary representations of the symmetry (isometry) group, according to Wigner's theorem. The spectrum and the kinematics of the QFT are thus determined by the symmetry group.

The simplest and best-known example is Minkowski spacetime, whose isometry group is Poincaré's $\mathrm{ISO}(1, d-1)$: a particle moving in Minkowski spacetime has $d(d+1)/2$

[1] If the matter fields are not standard tensor fields, i.e. if they are spinors or fields transforming covariantly under some other local symmetry of the theory (local Lorentz transformations for spinors and Vielbeins, gauge transformations for charged fields ...), then the standard Lie derivative does not give a good representation of the infinitesimal GCTs because it is not covariant under those local symmetries and the results would depend on the frame or gauge chosen. Instead we have to use a generalized *covariant Lie derivative*, as we will see in the next section, since this problem is relevant in supergravity theories.

conserved quantities (the d components of the momentum and the $d(d-1)/2$ components of the angular momentum). QFTs in Minkowski spacetime are constructed preserving Poincaré symmetry and the quanta of the fields will be particles defined by the values of the invariants that can be constructed with the conserved quantities (mass and spin).

It is natural to associate solutions with a maximal number of unbroken symmetries with possible vacuum states of the QFT. These states will be annihilated by the operators associated with these symmetries in the quantum theory. In GR with no cosmological constant, the only maximally symmetric solution is the Minkowski spacetime (ten isometries in $d = 4$ dimensions). With a (negative) positive cosmological constant, the Minkowski metric is not a solution and the only maximally symmetric solutions are the (anti-)de Sitter spacetimes whose isometry group (SO(2,3)) SO(1,4) is also ten-dimensional. These are the only maximally symmetric solutions of GR.

It is possible to define field theories in (anti-)de Sitter spacetime, but it is also possible (albeit unusual) to do it in spacetimes with fewer isometries, except in higher dimensions: for instance, we have studied in Chapter 11 Kaluza–Klein vacua that are the products of d-dimensional Minkowski spacetime and a circle whose isometry group is considerably smaller than that of $(d+1)$-dimensional Minkowski spacetime, which is spontaneously broken by the choice of vacuum. The spectrum of the KK theory is determined by the unbroken symmetry group, and it is the spectrum of a d-dimensional theory with gravity. The name *spontaneous compactification* could be applied to this and other cases in which there is a classical solution that we associated with a vacuum in which the spacetime is a product of a lower-dimensional spacetime and a compact space.

We can also consider other solutions of GR that asymptotically approach one of the three vacua we just mentioned. As we have stressed repeatedly, solutions of this kind represent isolated systems in GR. We can use the Abbott–Deser formalism of Section 6.1.2 to find the values of the $d(d+1)/2$ conserved quantities of those spacetimes which are associated with the isometries of the vacuum (even if the solutions themselves do not have any isometry). If we associate with the systems described by the asymptotically vacuum solutions states of a QFT built over the associated vacuum state, then the generators of the symmetry algebra have a well-defined action on them.[2] On the other hand, only the vacuum state is annihilated by all those generators, corresponding to its invariance under all the isometries. In particular, the vacuum state will be annihilated by the energy operator, and thus (if we restrict ourselves to states with non-negative energy) it will be the state with minimal energy. This point is problematic in de Sitter spacetimes, which compromises their stability.

This association of solutions that approach asymptotically a vacuum and states of a quantum theory on which the generators of the vacuum isometries act is a very fruitful point of view that we will use extensively. It can be extended to less-symmetric vacua, defining its own class of asymptotic behavior.

We are now ready to extend this concept to the supersymmetry context.

---

[2] It should be stressed that this can be done for all the states corresponding to spacetimes with the same asymptotic behavior. We cannot compare the energies of, say, asymptotically flat and asymptotically anti-de Sitter spacetimes.

13.2 Supersymmetric vacua and residual (unbroken) supersymmetries

In general, the solutions of a supergravity theory are not invariant under any of the (infinite) supersymmetry transformations that leave the theory invariant. Those which are invariant under some (always a finite number of *residual or unbroken supersymmetries*) are said to be *supersymmetric*, *BPS*, or *BPS-saturated*.[3]

Schematically, the local supersymmetry transformations take the form

$$\begin{aligned} \delta_\epsilon B &\sim \epsilon F, \\ \delta_\epsilon F &\sim \partial\epsilon + B\epsilon, \end{aligned} \tag{13.6}$$

for boson (B) and fermion (F) fields. We are interested in purely bosonic solutions since these are the ones that correspond to classical solutions.[4] They are also solutions of the bosonic action that one obtains by setting to zero all the fermion fields of the supergravity theory, because this is always a consistent truncation. These bosonic actions are just well-known actions of GR coupled to matter fields (for instance, the Einstein–Maxwell theory in the $N=2$, $d=4$ supergravity case).

According to the general definition, a bosonic solution will be supersymmetric if the above transformations vanish for some infinitesimal supersymmetry parameter $\epsilon(x)$. In the absence of fermion fields, the bosonic fields are always invariant, and it is necessary only that the supersymmetry transformations of the fermion fields vanish:

$$\delta_\kappa F \sim \partial\epsilon + B\epsilon = 0. \tag{13.7}$$

From the superspace point of view, this can be seen as invariance under an infinitesimal super-reparametrization. Thus, by analogy with GR, this is called the *Killing spinor equation* and its solutions can be seen as the product of an infinitesimal anticommuting number ϵ and a finite commuting spinor κ called a *Killing spinor* that also satisfies the above equation. There is a different Killing spinor equation for each supergravity theory but, since we have defined it for purely bosonic configurations, it can be used without any reference to supergravity or fermion fields.

What is the symmetry group generated by the Killing spinors? Clearly, it has to be a finite-dimensional *supergroup* of which the Killing spinors are the fermionic generators. The supergroup is part of the infinite-dimensional supergroup of superspace super-reparametrizations that includes all the local supersymmetry transformations, GCTs, etc. However, where are the bosonic generators?

In the case of the isometry group of a metric, the structures of the finite-dimensional group and of the algebra of its generators are inherited from those of the

[3] We focus on local supersymmetries, although it is evidently possible to define unbroken supersymmetry in theories that are invariant only under global supersymmetry. For instance, in the context of super-Yang–Mills theory, the Bogomol'nyi–Prasad–Sommerfield (BPS) limit of the 't Hooft–Polyakov monopole discussed in Section 9.2.3 has some unbroken supersymmetries. This is why supersymmetric solutions are sometimes called BPS solutions. The reason why they are called BPS-saturated will be explained when we discuss *supersymmetry, Bogomol'nyi*, or *BPS bounds*.

[4] We observe only macroscopic bosonic fields in nature. However, technically, we could equally well consider non-vanishing fermionic fields. Also, we can *generate* fermionic fields by performing supersymmetry transformations on purely bosonic solutions.

infinite-dimensional group of all GCTs. In this case, the structure of the finite-dimensional supersymmetry group of a solution is inherited from that of the infinite-dimensional supergroup of all local supersymmetry transformations, GCTs, etc. of the supergravity theory. The commutator of two local supersymmetry transformations is a combination of all the symmetries of the theory: for instance, in $N = 2, d = 4$ Poincaré supergravity, given by Eq. (5.96), a GCT, a local Lorentz rotation, a U(1) gauge transformation, and a local supersymmetry transformation with parameters that depend on $\epsilon_{1,2}$ and the fields of the theory. Now, if $\kappa_{1,2}$ are Killing spinors of a bosonic solution, the commutator will give bosonic symmetries of the same solution. In particular, we find that the solution will be invariant[5] under GCTs generated by bilinears of the form[6]

$$k^{\mu} = -i\bar{\kappa}_1\gamma^{\mu}\kappa_2. \tag{13.8}$$

Other Killing spinor bilinears will be associated with generators of other (non-geometrical) symmetries of the solution. This is how the bosonic generators of the supersymmetry group of a bosonic solution arise.

Following our previous discussion of isometries in general-covariant theories, we can associate solutions admitting a maximal number of Killing spinors (maximally supersymmetric solutions) with vacua of the supergravity theory. Now, a given supergravity can have more than one maximally supersymmetric solution (vacuum). Usually, one of the vacua is also a maximally symmetric solution (Minkowski or AdS), but the other vacua are not and have non-vanishing matter fields. Each of these vacua defines a class of solutions with the same asymptotic behavior, which can be associated with states of the QFT that one would construct on the corresponding vacuum. The vacuum supersymmetry algebras can be used to define conserved quantities for those spacetimes/states. Thus, we can study the supersymmetries of these spacetimes using knowledge of their conserved charges and the superalgebra of the asymptotic vacuum spacetime or by solving the Killing spinor equation directly. We will do this in Section 13.5.

Our immediate task is to develop a method by which to find the supersymmetry algebras of the vacuum (or any other) solutions. Let us proceed by analogy with the non-supersymmetric-gravity case discussed in the previous section. There will be a bosonic generator $P_{(I)}$ of the abstract supersymmetry algebra for each Killing vector $k_{(I)}{}^{\mu}$ that generates a GCT that leaves invariant all the fields of the solution, there will be other "internal" bosonic generators $B_{(M)}$ associated with each invariance of the matter fields, and there will be a fermionic generator $Q_{(A)}$ of the abstract supersymmetry algebra for each Killing spinor $\kappa^{\alpha}_{(A)}$.

Now, we have to identify all the generators of the abstract supersymmetry algebra with operators acting on the supergravity fields. The (anti)commutators of these operators will give the corresponding (anti)commutators of the superalgebra generators.

Let us start with the bosonic generators $P_{(I)}$. On world tensors, each $P_{(I)}$ is represented by (minus) the standard Lie derivative with respect to the corresponding Killing vector $k_{(I)}$, which transforms world tensors into world tensors of the same rank. However, most of the

[5] This statement will be made more precise shortly.

[6] If κ_1 and κ_2 are identical commuting Killing spinors, the bilinear does not vanish. Furthermore, it can be shown that $k^{\mu} = -i\bar{\kappa}\gamma^{\mu}\kappa$ is always timelike or null in $d = 4$, null in $N = 1, d = 10$ supergravity, etc.

fields in supergravity theories are Lorentz tensors (with vector or spinor indices) and the standard Lie derivative is not covariant under local Lorentz transformations and its action on Lorentz tensors is frame-dependent.

This has annoying consequences: for instance the Lie derivatives of Vielbeins with respect to a Killing vector will not be zero in general, even though the same Lie derivative of the metric always will.

On the other hand, Lorentz tensors (and, in particular, spinors) in curved spaces are treated as scalars under GCTs in (Weyl's) standard formalism explained in Section 1.4. Then, if we work in Minkowski spacetime in curvilinear coordinates using Weyl's formalism and perform a Lorentz transformation, all Lorentz tensors and spinors will be invariant. This looks strange, but is not unphysical: in practice one always makes a choice of frame based on some simplicity criterion. For instance, we could always set the Vielbein matrix in an upper-triangular form using local Lorentz transformations. This choice can be seen as a gauge-fixing condition that uses up all the Lorentz gauge symmetry. If we now perform a GCT (for instance, the Lorentz transformation we were discussing), it will be necessary to implement a compensating local Lorentz transformation in order to keep the Vielbein matrix upper-triangular. This local Lorentz transformation will act on all Lorentz tensors and can be understood as the effect of the GCT on them.

It is necessary for our purposes to find an operator acting on Lorentz tensors that implements the adequate compensating local Lorentz transformation for each GCT. This operator is the *Lie–Lorentz derivative* [748], which was first introduced for spinors by Lichnerowicz and Kosmann in [632, 633, 655] and used in supergravity by Figueroa-O'Farrill in [390] (see also [586, 919, 920]). In simple terms, it is just a Lorentz-covariant Lie derivative.

Analogous problems arise whenever there are additional local symmetries. For instance, in $N = 2, d = 4$ supergravity there is a local U(1) symmetry. In the Poincaré case only the gauge potential A_μ transforms under it, but in the AdS case ("gauged $N = 2, d = 4$ supergravity") the gravitinos and infinitesimal supersymmetry parameters transform as doublets (they are charged). A U(1)-covariant derivative (*Lie–Maxwell derivative*) is needed in order to represent infinitesimal GCTs on these fields.

Covariant Lie derivatives can be found also in the context of the geometry of reductive coset spaces G/H (see Appendix A.4) on which there is a well-defined action of H. In fact, the Lie–Lorentz derivative coincides with it in coset spaces in which spinors can be defined and H is a subgroup of the Lorentz group [25].

More generally, they can be defined in principal bundles with a reductive G-structure[7] [460], but here we will not make use of this formalism.

13.2.1 Covariant Lie derivatives

The Lie–Lorentz derivative The spinorial Lie–Lorentz derivative with respect to any vector v of a Lorentz tensor T transforming in the representation r is given by

$$\mathbb{L}_v T \equiv v^\rho \nabla_\rho T + \tfrac{1}{2} \nabla_{[a} v_{b]}\, \Gamma_r(M^{ab}) T, \tag{13.9}$$

[7] Recall that G/H is a principal bundle over G/H with structure group H, so this is a special case.

and, on mixed world–Lorentz tensors $T_{\mu_1\cdots\mu_m}{}^{\nu_1\cdots\nu_n}$,

$$\begin{aligned}\mathbb{L}_v T_{\mu_1\cdots\mu_m}{}^{\nu_1\cdots\nu_n} \equiv\; & v^\rho \nabla_\rho T_{\mu_1\cdots\mu_m}{}^{\nu_1\cdots\nu_n} - \nabla_\rho v^{\nu_1} T_{\mu_1\cdots\mu_m}{}^{\rho\nu_2\cdots\nu_n} - \cdots \\ & + \nabla_{\mu_1} v^\rho T_{\rho\mu_2\cdots\mu_m}{}^{\nu_1\cdots\nu_n} + \cdots + \tfrac{1}{2}\nabla_{[a} v_{b]}\, \Gamma_r(M^{ab}) T_{\mu_1\cdots\mu_m}{}^{\nu_1\cdots\nu_n},\end{aligned} \tag{13.10}$$

where ∇_μ is the full (affine plus Lorentz) torsionless covariant derivative satisfying the first Vielbein postulate and $\Gamma_r(M^{ab})$ are the generators of the Lorentz algebra in the representation r. $\nabla_{[a} v_{b]}$ is the parameter of the compensating Lorentz transformation.

This derivative enjoys certain properties only when it is taken with respect to a Killing vector or a conformal Killing vector. In particular, the property Eq. (13.12) which allows us to define a Lie algebra structure holds only for conformal Killing vectors and we are going to restrict our study to that case.

For any two mixed tensors T_1 and T_2 and any two conformal Killing vectors k_1 and k_2 and constants a^1 and a^2 we have the following.

1. $\mathbb{L}_k$ satisfies the Leibniz rule:

$$\mathbb{L}_k(T_1 T_2) = \mathbb{L}_k(T_1) T_2 + T_1 \mathbb{L}_k T_2. \tag{13.11}$$

2. The commutator of two Lie–Lorentz derivatives

$$[\mathbb{L}_{k_1}, \mathbb{L}_{k_2}]\, T = \mathbb{L}_{[k_1,k_2]}\, T, \tag{13.12}$$

   where $[k_1, k_2]$ is the Lie bracket.

3. $\mathbb{L}_k$ is linear in the vector fields

$$\mathbb{L}_{a^1 k_1 + a^2 k_2}\, T = a^1\, \mathbb{L}_{k_1} T + a^2\, \mathbb{L}_{k_2} T. \tag{13.13}$$

Thus, $\mathbb{L}_k$ is a derivative and provides a representation of the Lie algebra of conformal isometries of the manifold.

Some further properties are the following.

1. The Lie–Lorentz derivative of the Vielbein is

$$\mathbb{L}_k e^a{}_\mu = \frac{1}{d} \nabla_\rho k^\rho e^a{}_\mu, \tag{13.14}$$

   and vanishes when k is a Killing vector (not just conformal). In this case, we have

$$\mathbb{L}_k \xi^a = e^a{}_\mu \mathcal{L}_k \xi^\mu. \tag{13.15}$$

2. If $k^\mu = \sigma^\mu{}_\nu x^\nu$ with $\sigma^{\mu\nu} = -\sigma^{\nu\mu}$ and constant is an infinitesimal global Lorentz transformation in Minkowski spacetime with Cartesian coordinates, then, on a spinor ψ, as we wanted

$$\mathbb{L}_k = k^\mu \mathcal{D}_\mu \psi + \tfrac{1}{4}\mathcal{D}_{[a} k_{b]} \gamma^{ab} \psi = k^\mu \partial_\mu \psi + \tfrac{1}{4}\sigma_{ab}\gamma^{ab}\psi. \tag{13.16}$$

3.

$$\mathbb{L}_k \gamma^a = 0. \tag{13.17}$$

4. Owing to Eqs. (13.11), (13.14), and (13.17), the Lie–Lorentz derivative with respect to Killing vectors preserves the Clifford action of vectors v on spinors ψ, $v \cdot \psi \equiv v_a \Gamma^a \psi = \not{v}\psi$:

$$[\mathbb{L}_k, \not{v}]\, \psi = [k, v] \cdot \psi. \tag{13.18}$$

5. Also, for Killing vectors k only, it preserves the covariant derivative

$$[\mathbb{L}_k, \nabla_v]\, T = \nabla_{[k,v]}\, T. \tag{13.19}$$

6. All this implies that the Lie–Lorentz derivative with respect to Killing vectors preserves the supercovariant derivative of $N = 1, 2, d = 4$ Poincaré and $N = 1, d = 4$ AdS supergravity theories,

$$[\mathbb{L}_k, \tilde{\mathcal{D}}_v]\, \psi = \tilde{\mathcal{D}}_{[k,v]}\, \psi, \tag{13.20}$$

if

$$\mathcal{L}_k F_{\mu\nu} = 0. \tag{13.21}$$

It should be clear that (minus) the Lie–Lorentz derivative with respect to the Killing vectors of the theory $-\mathbb{L}_{k_{(I)}}$ should be the operator that represents the bosonic generators $P_{(I)}$ on the Vielbein $e^a{}_\mu$ and on the infinitesimal supersymmetry parameters ϵ in $N = 1, 2, d = 4$ Poincaré and $N = 1, d = 4$ AdS theories.

The Lie–Maxwell derivative. How are the $P_{(I)}$s represented on the other fields A_μ and ψ_μ? A_μ is defined in any solution up to U(1) gauge transformations[8] and, even though it transforms under GCTs as a vector field, the action of the standard Lie derivative is also gauge-dependent. This is similar to our problem with Lorentz tensors, but not quite the same, because A_μ is a connection and does not transform as a U(1) tensor. Thus, we do not expect to find a U(1)-covariant non-trivial generalization of the Lie derivative for it. It is easy to construct a gauge-invariant generalization of the Lie derivative by adding a compensating U(1) gauge transformation:

$$\mathcal{L}_k A_\mu - \partial_\mu (k^\nu A_\nu), \tag{13.22}$$

but it does not have the crucial Lie-algebra property. We could try to add another gauge transformation with parameter Ω,

$$\mathcal{L}_k A_\mu - \partial_\mu (k^\nu A_\nu + \Omega), \tag{13.23}$$

but it works only if

$$\partial_\mu \Omega = k^\lambda F_{\mu\lambda}, \tag{13.24}$$

and then Eq. (13.23) vanishes for any A_μ. This is in fact how the $P_{(I)}$s are represented on A_μ: on looking into the commutator Eq. (5.96), we find on the r.h.s. a GCT and a gauge

[8] Of course, the $P_{(I)}$s are represented on the field strength $F_{\mu\nu}$ by the standard Lie derivative. The condition Eq. (13.21) is a necessary condition for the corresponding $P_{(I)}$ to be a symmetry of the complete solution.

transformation with parameter $\chi = k^\nu A_\nu + \Omega$ with $\Omega = -i\bar{\epsilon}_2\sigma^2\epsilon_1$ and it can be checked that, for Killing spinors, we have, precisely,

$$\partial_\mu\big(-i\bar{\kappa}_{(B)}\sigma^2\kappa_{(A)}\big) = k_{(I)}{}^\lambda F_{\mu\lambda} \qquad k_{(I)}{}^\mu \equiv -i\bar{\kappa}_{(A)}\gamma^\mu\kappa_{(B)}. \tag{13.25}$$

This exercise is useful because there can be other gauge-dependent fields in the supergravity theory: in $N = 2, d = 4$ AdS (gauged) supergravity the gravitinos ψ_μ and the supersymmetry parameters ϵ (and, therefore, the Killing spinors, if any) are electrically charged and transform according to

$$A'_\mu = A_\mu + \partial_\mu\chi, \qquad \psi'_\mu = e^{-ig\chi\sigma^2}\psi_\mu, \qquad \epsilon' = e^{-ig\chi\sigma^2}\epsilon, \tag{13.26}$$

and we need to define a U(1) and Lorentz-covariant Lie derivative for them. For the supersymmetry parameters ϵ and Killing vectors k, we define

$$\mathsf{L}_k\epsilon \equiv \mathbb{L}_k\epsilon + ig(k^\mu A_\mu + \Omega)\sigma^2\epsilon, \tag{13.27}$$

where Ω has been defined above and exists if k is Killing and Eq. (13.21) is satisfied. This derivative has the Lie algebra property and also preserves the $N = 2, d = 4$ AdS supercovariant derivative

$$[\mathsf{L}_k, \hat{\tilde{\mathcal{D}}}_v]\,\epsilon = \hat{\tilde{\mathcal{D}}}_{[k,v]}\,\epsilon, \tag{13.28}$$

under the condition Eq. (13.21), which is necessary anyway in order for the associated $P_{(I)}$ to be a symmetry of the whole solution.

A non-Abelian generalization of all these formulae can be found in Appendix A.4.1.

For the gravitinos ψ_μ we expect problems similar to those we found for A_μ since they can be considered (super) gauge fields and transform inhomogeneously under supersymmetry. The role of the supersymmetry transformation that appears in the commutators Eqs. (5.45), (5.58) and (5.96) will clearly be that of compensating the effect of the GCT.

13.2.2 Calculation of supersymmetry algebras

We have developed all the tools we need to calculate the symmetry superalgebra of any supergravity solution. Now we just have to follow this six-step recipe [390].

1. First we have to solve the Killing and Killing-spinor equations. We keep only the Killing vectors that leave invariant all the fields of the solution. Furthermore, we have to find any other "internal" invariance of the fields.

2. With each Killing vector $k_{(I)}{}^\mu$ we associate a bosonic generator of the superalgebra $P_{(I)}$, with any internal symmetry of the fields another bosonic generator $B_{(M)}$, and with each Killing spinor $\kappa_{(A)}{}^\alpha$ we associate a fermionic generator (supercharge) $Q_{(A)}$.

   The bosonic subalgebra is in general the sum of two subalgebras generated by the $P_{(I)}$s and the $B_{(M)}$s with structure constants $f_{IJ}{}^K$ and $f_{MN}{}^P$. The fermionic generators are in representations of these bosonic subalgebras. These representations are determined by the structure constants $f_{AI}{}^B$ and $f_{AM}{}^B$ that appear in

$$[Q_{(A)}, P_{(I)}] = f_{AI}{}^B Q_{(B)}, \qquad [Q_{(A)}, B_{(M)}] = f_{AM}{}^B Q_{(B)}. \tag{13.29}$$

The superalgebra is determined by these four sets of structure constants plus the structure constants $f_{AB}{}^{I}$ that appear in the anticommutators

$$\{Q_{(A)}, Q_{(B)}\} = f_{AB}{}^{I} P_{(I)}. \tag{13.30}$$

3. The structure constants $f_{IJ}{}^{K}$ of the bosonic subalgebra are simply those of the isometry Lie algebra
$$[k_{(I)}, k_{(J)}] = -f_{IJ}{}^{K} k_{(K)}. \tag{13.31}$$

4. The commutators $[Q_{(A)}, P_{(I)}]$ can be interpreted as the action of the bosonic generators on the fermionic generators, which transform under some (spinorial) representation of the bosonic subalgebra with matrices $\Gamma_{\mathrm{s}}(P_{(I)})^{B}{}_{A} = f_{AI}{}^{B}$. Since the covariant Lie derivative has been defined to represent the action of infinitesimal GCTs on any kind of Lorentz tensors or spinors, and, according to Eqs. (13.19) and (13.20), transforms Killing spinors into Killing spinors, which, therefore, furnish a representation of the bosonic subalgebra, it is natural to expect that the structure constants $f_{AI}{}^{B}$ are given by the covariant Lie derivatives
$$\mathsf{L}_{k_{(I)}} \kappa_{(A)} \equiv f_{AI}{}^{B} \kappa_{(B)}. \tag{13.32}$$

5. We have mentioned that the bilinears of Killing spinors $-i\bar{\kappa}_{(A)}\gamma^{\mu}\kappa_{(B)}$ are Killing vectors. In fact, in the commutator of two local $N = 1, 2, d = 4$ supersymmetry transformations with parameters $\epsilon_{1,2}$ given in Eqs. (5.45), (5.58), and (5.96) we found a GCT (more precisely, (minus) a standard Lie derivative) with parameter $-i\bar{\epsilon}_1\gamma^{\mu}\epsilon_2$, a local Lorentz transformation, and a gauge transformation. When we use two Killing spinors $\kappa_{(A),(B)}$ instead, on the r.h.s. we always find $-\mathsf{L}_k$, where k^{μ} is the Killing vector $-i\bar{\kappa}_{(A)}\gamma^{\mu}\kappa_{(B)}$, which must be a linear combination of the Killing vectors $k_{(I)}$. The structure constants $f_{AB}{}^{I}$ are thus given by the decomposition of the bilinears
$$-i\bar{\kappa}_{(A)}\gamma^{\mu}\kappa_{(B)} \equiv f_{AB}{}^{I} k_{(I)}{}^{\mu}. \tag{13.33}$$

6. The structure constants involving the internal generators $B_{(M)}$ have to be determined case by case. They appear in extended supergravities and in general they are constant gauge transformations of vector fields that also act on the spinors.

We are now ready to apply these prescriptions to some basic examples, but it is useful to present some general considerations first.

13.3 $N = 1, 2, d = 4$ vacuum supersymmetry algebras

We have defined supergravity vacua as the classical solutions that admit a maximal number of Killing spinors, i.e. four in $N = 1, d = 4$ theories and eight in $N = 2, d = 4$ theories. A necessary (and locally sufficient) condition for a solution to be maximally supersymmetric is that the integrability condition of the Killing spinor equation admit the maximal number of possible solutions.

The Killing spinor equation takes the generic form $\tilde{D}_\mu \kappa = 0$, where $\tilde{D}_\mu = \partial_\mu - \Omega_\mu$ (the supercovariant derivative) can be understood as a standard covariant derivative with a connection Ω_μ that is the combination of the spin connection and other supergravity fields contracted with gamma matrices:

$$\Omega_\mu = \Omega_\mu{}^I \Gamma_{\mathrm{s}}(T_I), \tag{13.34}$$

where $\Gamma_{\mathrm{s}}(T_I)$ stands for different antisymmetrized products of gamma matrices that constitute a (spinorial) representation of some of the generators of some algebra. Thus, the Killing spinor equation can be understood as an equation of parallelism. This is why Killing spinors are sometimes called *parallel spinors*.

The integrability condition says that the commutator of the supercovariant derivative on the Killing spinor has to be zero, that is

$$[\tilde{D}_\mu, \tilde{D}_\nu]\kappa = 0, \quad \Rightarrow R_{\mu\nu}(\Omega)\kappa = 0, \tag{13.35}$$

where $R_{\mu\nu}(\Omega)$ is the curvature associated with the connection Ω. This is a homogeneous equation. The space of non-trivial solutions is determined by the rank of the matrix $R_{\mu\nu}(\Omega)$, which is a linear combination of $\Gamma_{\mathrm{s}}(T_I)$s with coefficients that depend on the values of the supergravity fields in the solution. In particular, we can have maximal supersymmetry only if $R_{\mu\nu}(\Omega) = 0$ identically (the connection is *flat*), which means that all the coefficients in the linear combination have to vanish.

All the maximally supersymmetric solutions known have homogeneous reductive spacetimes with invariant metrics and the connection 1-form Ω turns out to be the Maurer–Cartan 1-form V defined in Eq. (A.106) in a spinorial representation [25, 26]. In symmetric spaces, the spin connection contributes with the vertical components of V:

$$-\tfrac{1}{4}\omega_{ab}\gamma^{ab} = -\vartheta^i \Gamma_{\mathrm{s}}(M_i), \qquad \Gamma_{\mathrm{s}}(M_i) \equiv \tfrac{1}{4} f_{ia}{}^b \gamma_b{}^a, \tag{13.36}$$

due to Eq. (A.117) and the fact that the structure constants $f_{ia}{}^b$ are a representation of $\mathfrak{h}$ on $\mathfrak{k}$, which makes the above $\Gamma_{\mathrm{s}}(M_i)$ a spinorial representation of $\mathfrak{h}$.

All the horizontal components of V must come from the contribution of the supergravity fields. In the non-symmetric case [26] a combination of the two contributions gives V.

The curvature of the 1-form V is identically zero: in the language of differential forms

$$\tilde{D} = d - V, \quad \Rightarrow R(V) = dV - V \wedge V = 0, \tag{13.37}$$

which are precisely the Maurer–Cartan equations. The Killing spinor equations admit a maximal number of solutions and, actually, since $V = -\Gamma_{\mathrm{s}}(u^{-1})d\Gamma_{\mathrm{s}}(u)$ where $\Gamma_{\mathrm{s}}(u)$ is the coset representative defined in Eq. (A.104) using the spinorial representation $\Gamma_{\mathrm{s}}(P_{(a)})$ dictated by the supergravity theory, the Killing spinors take the form

$$\kappa = \Gamma_{\mathrm{s}}(u^{-1})\kappa_0, \tag{13.38}$$

where κ_0 is any constant spinor. Choosing independent constant spinors we find the following basis of Killing spinors:

$$\kappa_{(\alpha)}{}^\beta = \Gamma_{\mathrm{s}}(u^{-1})^\beta{}_\alpha. \tag{13.39}$$

This result reproduces the construction of Killing spinors on spheres and AdS space made in [667] but the calculations are dramatically simplified and the geometrical meaning of the construction is clearer.

On the other hand, this form of the Killing spinors is extremely useful for computing the supersymmetry algebra. First, it can be shown [25] that the Lie–Lorentz derivatives associated with the Killing vectors coincide with the H-covariant Lie derivatives defined in Eq. (A.4.1). Then, using the property Eq. (A.128), we immediately find

$$\mathbb{L}_{k_{(I)}}\kappa_{(\alpha)}{}^{\beta} = -\kappa_{(\gamma)}^{\beta}\Gamma_{\rm s}(T_I)^{\gamma}{}_{\alpha}, \quad \Rightarrow f_{\alpha I}{}^{\gamma} = -\Gamma_{\rm s}(T_I)^{\gamma}{}_{\alpha}. \tag{13.40}$$

Let us now consider the Killing-spinor bilinears. We consider only Majorana spinors. Then the bilinears take the form

$$-i\bar{\kappa}_{(\alpha)}\gamma^{\mu}\kappa_{(\beta)}\partial_{\mu} = -i\Gamma_{\rm s}(u^{-1})_{\alpha}{}^{\gamma}\mathcal{C}_{\gamma\delta}(\gamma^{a})^{\delta}{}_{\epsilon}\Gamma_{\rm s}(u^{-1})^{\epsilon}{}_{\beta}, \tag{13.41}$$

where $\mathcal{C}$ is the charge-conjugation matrix.

Usually, one finds that the spinorial representation is proportional to the gamma matrices,

$$\Gamma_{\rm s}(P_a) = \mathcal{S}\gamma_a. \tag{13.42}$$

When $\mathcal{S}$ is invertible and the Killing metric is also invertible[9] we can write

$$\gamma^{a} = \mathcal{S}\Gamma_{\rm s}(P^{a}), \tag{13.43}$$

where $\Gamma_{\rm s}(P^a)$ is a dual representation. The combination $\tilde{\mathcal{C}} \equiv \mathcal{C}\mathcal{S}$ acts as a charge-conjugation matrix in the subspace spanned by the horizontal generators in the spinorial representation

$$\tilde{\mathcal{C}}^{-1}\Gamma_{\rm s}(P^{a})^{\rm T}\tilde{\mathcal{C}} = -\Gamma_{\rm s}(P^{a}), \tag{13.44}$$

so

$$\Gamma_{\rm s}(u^{-1})^{\rm T}\mathcal{C}\gamma^{a} = \Gamma_{\rm s}(u^{-1})^{\rm T}\tilde{\mathcal{C}}\Gamma_{\rm s}(P^{a}) = \tilde{\mathcal{C}}\Gamma_{\rm s}(u)\Gamma_{\rm s}(P^{a}). \tag{13.45}$$

and, thus,

$$-i\bar{\kappa}_{(\alpha)}\gamma^{\mu}\kappa_{(\beta)}\partial_{\mu} = -i\tilde{\mathcal{C}}_{\alpha\gamma}\Gamma_{\rm s}(u)^{\gamma}{}_{\delta}\Gamma_{\rm s}(P^{a})^{\delta}{}_{\epsilon}\Gamma_{\rm s}(u^{-1})^{\epsilon}{}_{\beta}e_{a}. \tag{13.46}$$

In this expression we can recognize $uP^{a}u^{-1}$ in the spinorial representation, which is the coadjoint action of the coset element u on P^a,

$$\begin{aligned} -i\bar{\kappa}_{(\alpha)}\gamma^{\mu}\kappa_{(\beta)}\partial_{\mu} &= -i\tilde{\mathcal{C}}_{\alpha\gamma}\Gamma_{\rm s}(T^{I})^{\gamma}{}_{\beta}\Gamma_{\rm Adj}(u^{-1})^{a}{}_{I}e_{a} = -i\tilde{\mathcal{C}}_{\alpha\gamma}\Gamma_{\rm s}(T^{I})^{\gamma}{}_{\beta}k_{(I)}, \\ &\Rightarrow f_{\alpha\beta}{}^{I} = -i\tilde{\mathcal{C}}_{\alpha\gamma}\Gamma_{\rm s}(T^{I})^{\gamma}{}_{\beta}, \end{aligned} \tag{13.47}$$

where we have used Eq. (A.114).

In extended supergravities, one may also have to compute other Killing-spinor bilinears $-i\bar{\kappa}_{(\alpha)}\Sigma^{(M)}\kappa_{(\beta)}$, where $\Sigma^{(M)}$ acts on internal spinor indices that we are not showing. In some cases, the internal symmetries are related to the vertical generators M_i,

$$\Sigma^{(i)} = \mathcal{S}\Gamma_{\rm s}(M^{i}), \tag{13.48}$$

[9] This is usually the case. The exceptions are the Kowalski–Glikman Hpp-waves and the five-dimensional Gödel-like solution of [420].

and, using the property of the matrix $\mathcal{S}$, one finds

$$-i\bar{\kappa}_{(\alpha)}\Sigma^{(i)}\kappa_{(\beta)} = -i\tilde{C}_{\alpha\gamma}\Gamma_{\rm s}(T^I)^\gamma{}_\beta\Gamma_{\rm Adj}(u^{-1})^i{}_I = i\tilde{C}_{\alpha\gamma}\Gamma_{\rm s}(T^I)^\gamma{}_\beta(k_{(I)}^\mu\vartheta^i{}_\mu + W^i{}_I). \quad (13.49)$$

In the light of Appendix A.4.1 this means that $-i\bar{\kappa}_{(\alpha)}\Sigma^{(i)}\kappa_{(\beta)}$ gives the infinitesimal parameter of the gauge transformation of the vertical Maurer–Cartan 1-forms ϑ^i that can be compensated by a diffeomorphism.

This result can be understood only after the realization that, in all these cases, the vertical Maurer–Cartan 1-forms ϑ^i enter the solution in a non-trivial form. The simplest example is the Robinson–Bertotti solution of $N=2, d=4$ Poincaré supergravity, with $\mathrm{AdS}_2\times \mathrm{S}^2$ geometry in which the Maxwell vector field is identical to a linear combination of the two vertical Maurer–Cartan 1-forms. The same is true for the KG4 solution. In the higher-dimensional cases that we will study in Part III (Section 19.5.1), the matter fields (differential forms of higher rank) are also given in terms of the ϑ^is, which are non-Abelian.

We are now going to focus again on the simplest $N=1,2, d=4$ cases. First, we are going to derive the Killing-spinor integrability condition for $N=2, d=4$ gauged supergravity because, in the $g\to 0$ limit, we can recover the integrability condition for the ungauged case, and, on setting $A_\mu=0$ and ignoring one fermion, we obtain the $N=1$ cases.

13.3.1 The Killing-spinor integrability condition

The integrability condition in $N=2, d=4$ gauged supergravity is

$$[\tilde{\hat{\mathcal{D}}}_\mu, \tilde{\hat{\mathcal{D}}}_\nu]\kappa = 0. \quad (13.50)$$

The supercovariant derivative $\tilde{\hat{D}}_\mu$ is given in Eq. (5.101). We immediately find

$$[\tilde{\hat{\mathcal{D}}}_\mu, \tilde{\hat{\mathcal{D}}}_\nu] = [\hat{\mathcal{D}}_\mu, \hat{\mathcal{D}}_\nu] + gF_{\mu\nu}i\sigma^2 - \frac{i}{2}\hat{\nabla}_{[\mu}\not{F}\gamma_{\nu]}i\sigma^2 + \frac{1}{8}\not{F}\gamma_{[\mu}\not{F}\gamma_{\nu]}. \quad (13.51)$$

The first term gives the SO(2,3) curvature that can be put into the form

$$[\hat{\mathcal{D}}_\mu, \hat{\mathcal{D}}_\nu] = -\tfrac{1}{4}\left(R_{\mu\nu}{}^{ab} + 2g^2 e_\mu{}^{[a}e_\nu{}^{b]}\right)\gamma_{ab}. \quad (13.52)$$

The third term can also be put into the form

$$-\frac{i}{2}\hat{\mathcal{D}}_{[\mu}\not{F}\gamma_{\nu]}i\sigma^2 = -\frac{i}{2}\nabla_{[\mu}\not{F}\gamma_{\nu]}i\sigma^2 - \frac{g}{4}\left(\gamma_{[\mu}\not{F}\gamma_{\nu]} + \not{F}\gamma_{\mu\nu}\right)i\sigma^2. \quad (13.53)$$

The g-dependent terms together with the second term in Eq. (13.51) combine into

$$-\frac{g}{8}F_{ab}\left(3\gamma^{ab}\gamma_{\mu\nu} + \gamma_{\mu\nu}\gamma^{ab}\right)i\sigma^2, \quad (13.54)$$

and, with some gamma gymnastics, it is possible to rewrite it in the form

$$-\frac{i}{2}\nabla_{[\mu}\not{F}\gamma_{\nu]}i\sigma^2 = -\frac{i}{2}\not{\nabla}\left(F_{\mu\nu} + i{}^\star F_{\mu\nu}\gamma_5\right)i\sigma^2 + \frac{i}{2}\gamma_{\mu\nu\rho}\nabla_\sigma\left(F^{\sigma\rho} + i{}^\star F^{\sigma\rho}\gamma_5\right)i\sigma^2. \quad (13.55)$$

This term is manifestly electric–magnetic-duality invariant. Since it is not proportional to g, this agrees with the chiral–dual invariance of the ungauged theory. Now

$$\not{F}\gamma_{[\mu}\not{F}\gamma_{\nu]} = 8i F_{[\mu}{}^{\rho}\,{}^{\star}F_{\nu]\rho}\gamma_5 + 8T(A)_{[\mu|\rho}\gamma^{\rho}{}_{|\nu]}. \tag{13.56}$$

The first term here can be shown to be identically zero[10] and $T(A)_{\mu\nu}$ is the bosonic part of the vector-field energy–momentum tensor Eq. (5.79). These two terms are also electric–magnetic-duality invariant and independent of g.

Putting everything together, we find

$$\begin{aligned}[\tilde{\tilde{\mathcal{D}}}_\mu, \tilde{\tilde{\mathcal{D}}}_\nu] = &-\tfrac{1}{4}R_{\mu\nu}{}^{ab}\gamma_{ab} - \tfrac{1}{2}g^2\gamma_{\mu\nu} + T(A)_{[\mu|\alpha}\gamma^{\alpha}{}_{|\nu]} - \frac{i}{2}\not{\nabla}\big(F_{\mu\nu} + i{}^{\star}F_{\mu\nu}\gamma_5\big)i\sigma^2 \\ &+ \frac{i}{2}\gamma_{\mu\nu\rho}\nabla_\sigma(F^{\sigma\rho} + i{}^{\star}F^{\sigma\rho}\gamma_5)i\sigma^2 - \frac{g}{8}F_{ab}\big(3\gamma^{ab}\gamma_{\mu\nu} + \gamma_{\mu\nu}\gamma^{ab}\big)i\sigma^2.\end{aligned} \tag{13.57}$$

We can now use the bosonic part of the Einstein equation rewritten in this form (substituting the value of $R = -12g^2$)

$$T(A)_{\mu\nu} = \tfrac{1}{2}R_{\mu\nu} + \tfrac{3}{2}g^2 g_{\mu\nu}, \tag{13.58}$$

plus the bosonic part of the Maxwell equation and the Bianchi identity. We obtain

$$-\tfrac{1}{4}\Big\{C_{\mu\nu}{}^{ab}\gamma_{ab} + 2i\,\not{\nabla}\big(F_{\mu\nu} + i{}^{\star}F_{\mu\nu}\gamma_5\big)i\sigma^2 + \frac{g}{2}F_{ab}\big(3\gamma^{ab}\gamma_{\mu\nu} + \gamma_{\mu\nu}\gamma^{ab}\big)i\sigma^2\Big\}\kappa = 0. \tag{13.59}$$

This is a homogeneous linear equation for κ. The 8×8 matrix is a linear combination of tensor products of gamma matrices and Pauli matrices, all of them linearly independent. There are terms with two gammas (and $\otimes\mathbb{I}_{2\times 2}$), whose coefficients are the components of the Weyl tensor, there are terms proportional to one gamma and γ_5 ($\otimes\sigma^2$), whose coefficients are the components of the covariant derivative of the electromagnetic tensor and its dual, and, finally, there are terms with zero, two, and four gammas ($\otimes\sigma^2$), whose coefficients are the components of the electromagnetic tensor. In order to have maximally supersymmetric solutions, each of these terms has to vanish. This imposes severe constraints on the vacuum candidates. Let us now study each case separately, and let us calculate the symmetry superalgebra using the recipe of Section 13.2.2.

13.3.2 The vacua of $N = 1, d = 4$ Poincaré supergravity

On setting $g = F_{\mu\nu} = 0$ in Eq. (13.59), we find the integrability condition for the Killing spinors of $N = 1, d = 4$ Poincaré supergravity. The maximally supersymmetric solutions are those with vanishing Weyl tensor, which (since the equations of motion are $R_{\mu\nu} = 0$) implies a vanishing Riemann curvature tensor. Thus, Minkowski spacetime is the only maximally supersymmetric vacuum of $N = 1, d = 4$ Poincaré supergravity.

[10] One has to use the self-evident four-dimensional identity

$$\eta_{a[b}\epsilon_{a_1\ldots a_4]}F^{a_1a_2}F^{a_3a_4} = 0.$$

In this simple case it is unnecessary to use the coset description of Minkowski spacetime. We can compute Killing spinors and vectors directly. In Cartesian coordinates and with the trivial frame $e^a{}_\mu = \delta^a{}_\mu$ the spin and Levi-Cività connections vanish, the Killing spinors are just constant, and the Killing vectors are the ten known generators of the Poincaré group.

Solution:

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu, \qquad e^a{}_\mu = \delta^a{}_\mu, \qquad e_a{}^\mu = \delta_a{}^\mu. \tag{13.60}$$

Killing spinors:

$$\{\kappa_{(A)}{}^\beta\} = \{\kappa_{(\alpha)}{}^\beta = \delta_{(\alpha)}{}^\beta\}. \tag{13.61}$$

Killing vectors:

$$\{k_{(I)}{}^\mu\} = \{k_{(a)}{}^\mu = e_a{}^\mu,\ k_{(ab)}{}^\mu = 2e_{[a}{}^\mu e_{b]\sigma} x^\sigma\}. \tag{13.62}$$

The commutators of the bosonic generators are given by the Lie algebra of the Killing vectors which gives the Poincaré algebra, that we do not need to write explicitly. To find the anticommutator of two supercharges, we need to decompose the Killing vectors $-i\bar{\kappa}_{(\alpha)}\gamma^\mu\kappa_{(\beta)}$ as a combination of the $k_{(I)}{}^\mu$s for each pair of indices (α) and (β). This is easy:

$$-i\bar{\kappa}_{(\alpha)}\gamma^\mu\kappa_{(\beta)} = -i(\mathcal{C}\gamma^a)_{(\alpha)(\beta)} k_{(a)}{}^\mu, \quad \Rightarrow \{Q_{(\alpha)}, Q_{(\beta)}\} = -i(\mathcal{C}\gamma^a)_{(\alpha)(\beta)} P_{(a)}, \tag{13.63}$$

which we can convert into the standard form by raising the indices (α) and (β) with $\mathcal{C}^{-1}$. The fact that only the translational symmetries occur could have been predicted since all the $N = 1, d = 4$ Killing vectors $-i\bar{\kappa}_{(\alpha)}\gamma^\mu\kappa_{(\beta)}$ are covariantly constant. These are also the ones with vanishing compensating Lorentz transformation in the Lie–Lorentz derivative. Since the Killing spinors are constant, we find that only $\mathbb{L}_{k_{(ab)}}\kappa_{(\alpha)}{}^\beta$ is different from zero and, using $\nabla_{[a|}k_{(bc)|d]} = -2\eta_{ad,bc}$, we find, as expected,

$$[Q_{(\alpha)}, P_{(ab)}] = -Q_{(\beta)}\tfrac{1}{2}(\gamma_{ab})^\beta{}_\alpha. \tag{13.64}$$

13.3.3 *The vacua of* $N = 1, d = 4$ AdS$_4$ *supergravity*

The integrability condition in $N = 1, d = 4$ AdS supergravity again implies the vanishing of the Weyl tensor, but now this implies

$$R_{\mu\nu}{}^{ab} + 2g^2 e_{[\mu}{}^a e_{\nu]}{}^b = 0, \tag{13.65}$$

i.e. the space is a maximally symmetric space of constant curvature $-2g^2$, so it is locally AdS$_4$ with AdS radius $R = 1/g$.

To construct the Killing spinors and find the symmetry superalgebra using the coset method, we have to construct the metric and Vierbeins using the method explained in Appendix A.4.[11]

[11] Actually, we do not even need to implement this construction explicitly. It is enough to assume that the Vierbeins and metric (and hence the spin connection) have been constructed by that procedure. We are going to construct the AdS$_4$-invariant metric just to illustrate the procedure and fix the notation.

AdS$_4$ can be identified with the coset space $\mathrm{SO}(2, 3)/\mathrm{SO}(1, 3)$. We use the index conventions of Section 4.5. $\mathfrak{g} = \mathrm{so}(2, 3)$, the Lie algebra of $G = \mathrm{SO}(2, 3)$, is written in Eq. (4.152). It is convenient to rescale and rename the generators as in Eq. (4.154), and the commutation relations take the form Eqs. (4.155).

The M_{ab}s generate the subalgebra $\mathfrak{h} = \mathrm{so}(1, 3)$ of the Lorentz subgroup. The orthogonal complement $\mathfrak{k}$ is generated by the P_as and, on looking at the commutation relations Eqs. (4.155), we see that we have a symmetric pair.

Now we construct the coset representative u,

$$u(x) = e^{x^3 P_3} e^{x^2 P_2} e^{x^1 P_1} e^{x^0 P_0}, \tag{13.66}$$

and the Maurer–Cartan 1-form V,

$$\begin{aligned} V &= -u^{-1} du \\ &= -P_0 dx^0 - e^{-x^0 P_0} P_1 e^{x^0 P_0} dx^1 - e^{-x^1 P_1} e^{-x^0 P_0} P_2 e^{x^0 P_0} e^{x^1 P_1} dx^2 \\ &\quad - e^{-x^2 P_2} e^{-x^1 P_1} e^{-x^0 P_0} P_3 e^{x^0 P_0} e^{x^1 P_1} e^{x^2 P_2} dx^3. \end{aligned} \tag{13.67}$$

Using the definition of the adjoint action of the group on the algebra, we see that

$$e^{-x^0 P_0} P_1 e^{x^0 P_0} = T_I \Gamma_{\mathrm{Adj}} (e^{-x^0 P_0})^I{}_1, \tag{13.68}$$

etc. and, by projecting onto the horizontal subspace, we find the Vierbeins,

$$\begin{aligned} &e^0 = -dx^0, \;\; e^1 = -\cos x^0\, dx^1, \;\; e^2 = -\cos x^0 \cosh x^1\, dx^2, \\ &e^3 = -\cos x^0 \cosh x^1 \cosh x^2\, dx^2, \end{aligned} \tag{13.69}$$

which,with the Killing metric $(+ - --)$, give the following form of the AdS$_4$ metric:

$$ds^2 = (dx^0)^2 - \cos^2 x^0 \{(dx^1)^2 + \cosh^2 x^1 [(dx^2)^2 + \cosh^2 x^2 (dx^3)^2]\}. \tag{13.70}$$

The explicit form of the vertical 1-forms ϑ^{ab} is not necessary, but we need to know how they enter the spin connection. According to Eq. (A.117)

$$\omega^a{}_b = \tfrac{1}{2} \vartheta^{cd} f_{cd\,-1b}{}^{-1a} = \vartheta^{ac} \eta_{cb}. \tag{13.71}$$

The Killing spinor equation is

$$dx^\mu \hat{\mathcal{D}}_\mu \kappa = \left(d - \tfrac{1}{4} \omega_{ab} \gamma^{ab} - \frac{ig}{2} \gamma_a e^a \right) \kappa = 0, \tag{13.72}$$

and can be written in the form $(d - V)\kappa = 0$ with[12]

$$\Gamma_{\mathrm{s}}(P_a) = \frac{ig}{2} \gamma_a, \qquad \Gamma_{\mathrm{s}}(M_{ab}) = \tfrac{1}{2} \gamma_{ab}. \tag{13.73}$$

The Killing spinors are thus of the form Eqs. (13.38) and (13.39).

[12] Compare this with Eq. (B.132).

The dual generators $\Gamma_{\mathrm{s}}(P^a)$ can be defined by

$$\begin{aligned}\mathrm{Tr}\,[\Gamma_{\mathrm{s}}(P^a)\Gamma_{\mathrm{s}}(P_b)] &= \delta^a{}_b, \;\; \Rightarrow \Gamma_{\mathrm{s}}(P^a) = -\frac{i}{2g}\gamma^a,\\ \mathrm{Tr}\,[\Gamma_{\mathrm{s}}(M^{ab})\Gamma_{\mathrm{s}}(P_{cd})] &= \delta^{ab}{}_{cd}, \;\; \Rightarrow \Gamma_{\mathrm{s}}(M^{ab}) = -\tfrac{1}{2}\gamma^{ab}.\end{aligned} \tag{13.74}$$

We see that the matrix $\mathcal{S}$ is just the identity in this case. The bilinears give

$$\begin{aligned}-i\bar{\kappa}\gamma^a\kappa e_a &= 2g\Gamma_{\mathrm{s}}(u^{-1})^{\mathrm{T}}\mathcal{C}\Gamma_{\mathrm{s}}(P^a)\Gamma_{\mathrm{s}}(u^{-1})e_a\\ &= g\mathcal{C}\Gamma_{\mathrm{s}}(\hat{M}^{\hat{b}\hat{c}})\Gamma_{\mathrm{Adj}}(u^{-1})^a{}_{\hat{b}\hat{c}}e_a\\ &= g\mathcal{C}\Gamma_{\mathrm{s}}(\hat{M}^{\hat{b}\hat{c}})k_{(\hat{b}\hat{c})},\end{aligned} \tag{13.75}$$

and we recover the standard anticommutator of the supercharges,

$$\{Q_{(\alpha)}, Q_{(\beta)}\} = g[\mathcal{C}\Gamma_{\mathrm{s}}(\hat{M}^{\hat{a}\hat{b}})]_{\alpha\beta}\hat{M}_{\hat{a}\hat{b}} = -i(\mathcal{C}\gamma^a)_{\alpha\beta}P_a - \frac{g}{2}(\mathcal{C}\gamma^{ab})_{\alpha\beta}M_{ab}. \tag{13.76}$$

The commutators $[Q_{(\alpha)}, \hat{M}_{\hat{a}\hat{b}}]$ are found using Eq. (13.40):

$$[Q_{(\alpha)}, \hat{M}_{\hat{a}\hat{b}}] = -Q_{(\beta)}\Gamma_{\mathrm{s}}(\hat{M}_{\hat{a}\hat{b}})^{\beta}{}_{\alpha}. \tag{13.77}$$

13.3.4 The vacua of $N = 2$, $d = 4$ Poincaré supergravity

The integrability condition Eq. (13.59) now gives two independent conditions for maximal supersymmetry: a vanishing Weyl tensor and a covariantly constant Maxwell field strength. Only three solutions satisfy them: Minkowski spacetime, the Robinson–Bertotti (RB) solution [146, 812] given in Eq. (8.90), whose metric is that of the $\mathrm{AdS}_2 \times \mathrm{S}^2$ symmetric space, and the $d = 4$ Kowalski–Glikman solution (KG4) [637] given in Eq. (10.27) with a Hpp-wave metric (again, a symmetric space, that we have constructed as a coset space in Section 10.1.1). The symmetry superalgebra of the Minkowski spacetime is identical to that of the $N = 1$ case, with additional indices $i, j = 1, 2$ and, thus, we will focus on the RB and KG4 solutions since they are the simplest of a series of maximally supersymmetric solutions with metrics of the same form: $\mathrm{AdS}_m \times \mathrm{S}^n$ and Hpp whose symmetry superalgebras can be calculated in a very similar fashion [25]. The five- and six-dimensional cases will be discussed in Section 13.4 and the ten- and eleven-dimensional cases in Section 19.5.1, but these four-dimensional examples already exhibit all the interesting features.

The Robinson–Bertotti superalgebra. The solution is given in Eq. (8.90), but we rewrite it in a more convenient form, adapted to the normalization we used for the Maxwell field of $N = 2$, $d = 4$ supergravity (a factor of two difference):

$$ds^2 = R_2^2\, d\Pi^2_{(2)} - R_2^2\, d\Omega^2_{(2)}, \qquad F = -R_2\, \omega_{\mathrm{AdS}_2}. \tag{13.78}$$

Here $d\Pi^2_{(2)}$ is the metric of the AdS_2 spacetime of unit radius, $d\Omega^2_{(2)}$ the metric of the

2-sphere of unit radius. That is why there are factors of R_2^2 in the metric. On the other hand, ω_{AdS_2} is the volume 2-form of the AdS_2 spacetime of unit radius.

Both AdS_2 and S^2 are symmetric spacetimes, $\mathrm{SO}(2,1)/\mathrm{SO}(2)$ and $\mathrm{SO}(3)/\mathrm{SO}(2)$, and their product is a symmetric space as well. We call the $\mathrm{SO}(2,1)$ generators $\{T_I\}$, $I = 1, 2, 3$, with commutators given by Eq. (A.84) with $\mathsf{Q} = \mathrm{diag}(++-)$, and the $\mathrm{SO}(3)$ generators $\{\tilde{T}_I\}$, $I = 1, 2, 3$, with commutators given by Eq. (A.84) with $\mathsf{Q} = \mathrm{diag}(+++)$.

The subalgebra $\mathfrak{h}$ is generated by T_1 and $\tilde{T}_3$ and $\mathfrak{k}$ is generated by T_2, T_3, $\tilde{T}_1$, and $\tilde{T}_2$. We perform the following redefinitions:

$$T_1 = M_1,\;\; T_2 = R_2 P_1,\;\; T_3 = R_2 P_0,\;\; \tilde{T}_1 = R_2 P_3,\;\; \tilde{T}_2 = R_2 P_2,\;\; \tilde{T}_3 = M_2. \tag{13.79}$$

The coset representative is the product of the coset representatives u and $\tilde{u}$,

$$u = e^{R_2\phi P_0} e^{R_2\chi P_1}, \qquad \tilde{u} = e^{R_2\varphi P_3} e^{R_2(\theta-\frac{\pi}{2})P_2}. \tag{13.80}$$

We obtain the Maurer–Cartan 1-forms

$$\begin{aligned} e^0 &= -R_2\cosh\chi\, d\phi, & e^2 &= -R_2 d\theta, & \vartheta^1 &= -\sinh\chi\, d\phi, \\ e^1 &= -R_2\, d\chi, & e^3 &= -R_2\sin\theta d\varphi, & \vartheta^2 &= -\cos\theta\, d\varphi, \end{aligned} \tag{13.81}$$

and, with the Minkowski metric, we obtain the metric of Eq. (13.78) with

$$d\Pi^2_{(2)} \equiv \cosh^2\chi\, d\phi^2 - d\chi^2, \qquad d\Omega^2_{(2)} \equiv d\theta^2 + \sin^2\theta\, d\varphi^2. \tag{13.82}$$

Observe that we can construct not just the metric, but also the vector field of the solution, using the geometry. The RB solution Eq. (13.78) is purely electric. The gauge field is just

$$A = R_2\vartheta^1. \tag{13.83}$$

We could also use the magnetic RB solution. The gauge field of that solution is

$$A = R_2\vartheta^2. \tag{13.84}$$

We will work with the electric RB solution. The $N = 2, d = 4$ Killing-spinor equation takes the form

$$dx^\mu\tilde{\mathcal{D}}_\mu\kappa = (d - \tfrac{1}{4}\omega_{ab}\gamma^{ab} + \tfrac{1}{4}\sigma^2\not{F}\gamma_a e^a)\kappa = 0. \tag{13.85}$$

Equation (13.36) identifies the spin-connection term with the vertical components of the Maurer–Cartan 1-form V. The coefficients of the Vierbeins e^a are the horizontal generators

$$\Gamma_{\mathrm{s}}(P_a) = -\tfrac{1}{4}\sigma^2\not{F}\gamma_a. \tag{13.86}$$

It can be checked that the representation, which is explicitly given by

$$\begin{aligned} \Gamma_{\mathrm{s}}(P_0) &= \frac{1}{2R_2}\gamma^1\sigma^2, & \Gamma_{\mathrm{s}}(P_2) &= \frac{1}{2R_2}\gamma^0\gamma^1\gamma^3\sigma^2, & \Gamma_{\mathrm{s}}(M_1) &= \tfrac{1}{2}\gamma^0\gamma^1, \\ \Gamma_{\mathrm{s}}(P_1) &= -\frac{1}{2R_2}\gamma^0\sigma^2, & \Gamma_{\mathrm{s}}(P_3) &= \frac{1}{2R_2}\gamma^0\gamma^1\gamma^2\sigma^2, & \Gamma_{\mathrm{s}}(M_2) &= \tfrac{1}{2}\gamma^2\gamma^3, \end{aligned} \tag{13.87}$$

satisfies the algebra and thus the Killing-spinor equation takes the form $(d - V)\kappa = 0$ and the Killing spinors have the standard form (here $\kappa = \Gamma_{\mathrm{s}}(u^{-1}\tilde{u}^{-1})\kappa_0$).

To calculate the bilinears $-i\bar{\kappa}\gamma^{\mu}\kappa$ we need the duals $\Gamma_{\rm s}(P^a)$:

$$\gamma^a = -\frac{4}{R_2}\mathcal{S}\Gamma_{\rm s}(P^a), \qquad \mathcal{S} = \gamma^0\gamma^1\sigma^2, \qquad \mathrm{Tr}[\Gamma_{\rm s}(P^a)\Gamma_{\rm s}(P_b)] = \delta^a{}_b. \tag{13.88}$$

The modified charge-conjugation matrix $\tilde{\mathcal{C}} = \mathcal{C}\mathcal{S}$ has the required property

$$\tilde{\mathcal{C}}^{-1}\Gamma_{\rm s}(P^a)^{\rm T}\tilde{\mathcal{C}} = -\Gamma_{\rm s}(P^a), \quad \Rightarrow (u\tilde{u})^{-1\,\rm T}\tilde{\mathcal{C}} = \tilde{\mathcal{C}}u\tilde{u}, \tag{13.89}$$

that allows us to express the bilinears in the form

$$-i\bar{\kappa}_{(\alpha i)}\gamma^a\kappa_{(\beta j)}e_a = \frac{4i}{R_2}\left\{\tilde{\mathcal{C}}[\Gamma_{\rm s}(T^I)k_{(I)} + \Gamma_{\rm s}(\tilde{T}^I)\tilde{k}_{(I)}]\right\}_{(\alpha i\,\beta j)}, \tag{13.90}$$

where the $k_{(I)}$s are the Killing vectors of AdS_2 and the $\tilde{k}_{(I)}$s are those of S^2. This translates into the anticommutator

$$\{Q_{(\alpha i)}, Q_{(\beta j)}\} = -i\delta_{ij}(\mathcal{C}\gamma^a)_{\alpha\beta}P_a + \frac{i}{R_2}\mathcal{C}_{\alpha\beta}\epsilon_{ij}M_1 + \frac{1}{R_2}(\mathcal{C}\gamma_5)_{\alpha\beta}\epsilon_{ij}M_2, \tag{13.91}$$

that looks exactly like the $N = 2, d = 4$ Poincaré anticommutator with the two possible central charges. The difference is that now the P_as commute neither with each other nor with the M_is, which are no longer central.

The appearance of the two "central-charge" terms is a bit surprising. Actually, they are related to the invariance of the solution under the combined action of a gauge transformation with parameter $W_I{}^i$ and a reparametrization generated by $k_{(I)}$, due to the identification of the gauge field and the vertical 1-forms ϑ^i (see the discussion in Appendix A.4.1). This invariance does not commute with other reparametrizations and thus they no longer lead to central charges.

The "central-charge" terms can also be found by calculating the bilinears $-i\bar{\kappa}_{(\alpha i)}\sigma^2\kappa_{(\beta j)}$ and $-i\bar{\kappa}_{(\alpha i)}\sigma^2\gamma_5\kappa_{(\beta j)}$, using

$$\sigma^2 = \mathcal{S}\Gamma_{\rm s}(M_1), \qquad \sigma^2\gamma_5 = \mathcal{S}\Gamma_{\rm s}(M_2), \tag{13.92}$$

the definition of the adjoint action, and the expression of $W_I{}^i$, Eq. (A.114).

The commutators of the fermionic and bosonic generators are given by Eq. (13.40).

The RB solution induces spontaneous compactification on S^2. The bosonic generators of the so(2, 1) subalgebra will have a spacetime interpretation, and the so(3) ones will have the interpretation of internal symmetries, which is typical of extended AdS superalgebras. Since these act non-trivially on the supercharges, we would obtain a gauged supergravity with so(3)-charged gravitinos.

Let us consider now the contraction limit $R_2 \to \infty$. This contraction takes AdS_2 into the two-dimensional Poincaré algebra ISO(1, 1) (M_1 being the single Lorentz generator) and SO(3) into ISO(2) (M_2 being the generator of SO(2) rotations of P_2 and P_3, which now have the interpretation of two-dimensional central charges).

There is another contraction limit one can take: $R_2 \to \infty$ after redefining $M_1 \equiv R_2 Q$ and $M_2 \equiv R_2 P$. In this limit, all the bosonic generators again commute with each other and one recovers exactly the $N = 2, d = 4$ Poincaré superalgebra (without Lorentz generators, which can be understood here as R-parity generators) with Q and P as electric and magnetic central charges.

Generalizations of these facts will take place in other $\mathrm{AdS}_n \times \mathrm{S}^m$ solutions.

The KG4 superalgebra The solution is given in Eq. (10.27), but we have to take into account the factor of two due to the normalizations being different, which changes $F_{\underline{u1}} = \frac{1}{2}\lambda$. The metric is that of a Cahen–Wallach symmetric space, whose coset construction is reviewed in Section 10.1.1, with $A_{ij} = \frac{1}{8}\lambda^2\delta_{ij}$. There is one subtlety: in order to use the coset method, we are forced to use mostly plus signature, in which the Killing-spinor equation takes the form

$$\left(d - \tfrac{1}{4}\omega^a{}_b\gamma_a{}^b + \frac{i}{4}\sigma^2 \not{F}\gamma_a e^a\right)\kappa = 0. \tag{13.93}$$

As in the RB case, not just the metric, but also the Maxwell field, can be constructed using the vertical Maurer–Cartan 1-forms. In fact

$$A = \tfrac{1}{2}\lambda\vartheta^1. \tag{13.94}$$

An electric–magnetic-duality rotation is equivalent, in this solution, to a rotation in the wavefront plane. The new Maxwell field would be

$$A = \tfrac{1}{2}\lambda\vartheta^2. \tag{13.95}$$

On substituting the background fields into the above Killing-spinor equation, we find that it takes the form $(d - V)\kappa = 0$, with the spinorial representation of the Heisenberg algebra:

$$\Gamma_s(P_a) = -\frac{i}{4}\sigma^2\gamma^u\gamma^1\gamma_a, \qquad \Gamma_s(M_i) = -\tfrac{1}{8}\lambda^2\gamma^u\gamma^i, \qquad \Gamma_s(V) = 0. \tag{13.96}$$

The Killing spinors can be constructed immediately and the action of the Heisenberg-algebra generators on them is trivial to find, using Eq. (13.40). However, note the following.

1. H(6) is not the whole isometry algebra of the KG4 metric: SO(2) rotations in the wavefront plane leave the metric invariant and the full isometry group is their semidirect product SO(2) $\ltimes$ H(6) (because SO(2) acts on the H(6) generators). We would have to include SO(2) in the coset construction in order to obtain the commutator between the generator of SO(2) and the supercharges. However, SO(2) is not a symmetry of the full solution because it does not leave the field strength invariant.

2. Owing to the non-semisimplicity of the Heisenberg algebra, it is not possible to find a relation between γ^a and the dual representation $\Gamma_s(P^a)$. Thus, the Killing-spinor bilinears have to be calculated by brute force, but we will not do it here.

13.3.5 The vacua of $N = 2, d = 4$ AdS supergravity

The integrability condition now imposes a third constraint in order for the terms with zero, two, and four gammas to vanish: the Maxwell field-strength tensor has to vanish. Then, the only maximally supersymmetric solutions are those of $N = 1, d = 4$ AdS supergravity, i.e. AdS_4 spacetime, and a basis of Killing spinors will be provided by

$$\kappa_{(\alpha i)}{}^{\beta j} = \Gamma_s(u^{-1})^\beta{}_\alpha\delta^j{}_j. \tag{13.97}$$

The superalgebra will now include a bosonic generator associated with the constant-gauge transformations that rotate the spinor doublets and leave the vector field invariant. This generator must appear in the anticommutator of two supercharges and the corresponding structure constants are given by the bilinears $-i\bar{\kappa}_{(\alpha i)}\sigma^2\kappa_{(\beta j)} = \mathcal{C}_{\alpha\beta}\epsilon_{ij}$.

13.4 The vacua of $d = 5, 6$ supergravities with eight supercharges

To complete our overview of supergravity vacua in lower dimensions, we are going to study the vacua of the minimal supergravities in $d = 5$ and 6 dimensions that are related by dimensional reduction with $N = 2$, $d = 4$ Poincaré supergravity and have the same number of supercharges. Almost all the maximally supersymmetric vacua of these theories also happen to be related by dimensional reduction [664].

13.4.1 $N = (1, 0), d = 6$ supergravity

The fields of this theory are the metric $\hat{\hat{e}}^{\hat{\hat{a}}}{}_{\hat{\hat{\mu}}}$, 2-form field $\hat{\hat{B}}^{-}_{\hat{\hat{\mu}}\hat{\hat{\nu}}}$ with anti-self-dual field strength $\hat{\hat{H}}^{-} = 3\partial\hat{\hat{B}}^{-}$ and positive-chirality symplectic Majorana–Weyl gravitino $\hat{\hat{\psi}}^{+}_{\hat{\hat{\mu}}}$ [727] (we use the gamma matrices of Appendix B).

To write an action for an anti-self-dual 3-form, one has to introduce auxiliary fields. Alternatively, one can write an action for a generic 3-form,

$$\boxed{\hat{\hat{S}} = \int d^6\hat{\hat{x}}\sqrt{|\hat{\hat{g}}|}\left[\hat{\hat{R}} + \frac{1}{12}\hat{\hat{H}}^2\right],} \tag{13.98}$$

and impose the anti-self-duality constraint ${}^\star\hat{H}^{-} = -\hat{H}^{-}$ on the equations of motion. The Killing-spinor equation is

$$\left(\hat{\hat{\nabla}}_{\hat{\hat{a}}} - \frac{1}{48}\hat{\hat{\not{H}}}^{-}\hat{\hat{\gamma}}_{\hat{\hat{a}}}\right)\hat{\hat{\kappa}}^{+} = 0, \tag{13.99}$$

where $\hat{\hat{\kappa}}^{+}$ is a spinor of positive chirality.

Three maximally supersymmetric vacua of this theory are known:[13] Minkowski spacetime, the near-horizon limit of the extreme anti-self-dual string solution [441] that has an $\mathrm{AdS}_3 \times \mathrm{S}^3$ geometry, and the KG6 Hpp-wave solution [690]. The latter can be obtained by taking the Penrose limit of the former [158, 495, 764].

The $\mathrm{AdS}_3 \times \mathrm{S}^3$ solution can be written as follows:

$$\boxed{d\hat{\hat{s}}^2 = R_3^2\, d\Pi^2_{(3)} - R_3^2\, d\Omega^2_{(3)}, \qquad \hat{\hat{H}}^{-} = 4R_3(\omega_{\mathrm{AdS}_3} + \omega_{\mathrm{S}^3}),} \tag{13.100}$$

[13] Actually, it has been shown in [229] that the maximally supersymmetric vacua of this theory and of $N = (2, 0)$, $d = 6$ supergravity are in one-to-one correspondence, and their metrics are locally isometric to bi-invariant Lorentzian metrics of six-dimensional Lie groups with anti-self-dual parallelizing torsion. The three solutions that we present have this property and exhaust all the possibilities.

where $d\Pi^2_{(3)}$ is the metric of an AdS_3 spacetime of unit radius and ω_{AdS_3} is its volume 3-form. Remarkably, the electric and magnetic components of the 2-form potential can be written in terms of the vertical Maurer–Cartan 1-forms in AdS_3 and S^3, which are SO(2, 1) and SO(3) Yang–Mills solutions (the c_a are constants):

$$\hat{\hat{B}}^{-a} = c_{(a)}\epsilon^{(a)bc}\vartheta^i f_{ibc}. \tag{13.101}$$

The KG6 solution can be written as follows:

$$\boxed{\begin{aligned} d\hat{s}^2 &= 2du\left[dv + \frac{\lambda_6^2}{8}\vec{x}^{\,2}_{(4)}du\right] - d\vec{x}^{\,2}_{(4)}, \\ \hat{B}^- &= -\lambda_6 du \wedge (x^1dx^2 - x^3dx^4). \end{aligned}} \tag{13.102}$$

The potential can be also written as a Yang–Mills field:

$$\hat{\hat{B}}^{-i} \sim \vartheta^j f_j{}^{iu}. \tag{13.103}$$

The calculation of the Killing spinors and superalgebras can be done following the general method. It is, however, more interesting to see how the dimensional reduction of these solutions gives all the maximally supersymmetric vacua of $N = 1, d = 5$ supergravity.

13.4.2 $N = 1, d = 5$ supergravity

The dimensional reduction of $N = (1, 0), d = 6$ supergravity gives $N = 1, d = 5$ supergravity (Section 11.2.5) coupled to a vector multiplet. We are interested in reducing maximally supersymmetric six-dimensional solutions preserving all their unbroken supersymmetries. When is this possible?

Let us consider the component of the gravitino in the compact direction w (which gives rise to a five-dimensional spin $-\frac{1}{2}$ "gaugino") and its supersymmetry variation,

$$\delta_{\hat{\epsilon}}\hat{\hat{\psi}}^+_{\underline{w}} \sim (\partial_{\underline{w}} + M)\hat{\epsilon}^+, \tag{13.104}$$

where M is a combination of gamma matrices (different from the unit matrix). This equation has to vanish identically for a w-independent Killing spinor $\hat{\kappa}^+$ in order to have five-dimensional unbroken supersymmetry, which implies that M has to vanish identically. If we reduce a maximally supersymmetric six-dimensional solution and M does not vanish identically, then, since the above equation vanishes for some Killing vectors, they must depend on w and five-dimensional supersymmetry is broken. The amount of supersymmetry broken depends on the rank of M, which tells us how many non-trivial solutions of $M\hat{\hat{\kappa}}^+ = 0$ exist, and how many six-dimensional Killing spinors are independent of w.

Up to this point, this discussion carries over to any other case without modification. However, there are two different possibilities concerning the vanishing of M: in the present case, we obtain a five-dimensional reducible theory of which we can always truncate consistently the matter multiplet. Matter fields in a given multiplet are identified in the supersymmetry transformation rules because they transform among themselves. Now, $\hat{\hat{\psi}}{}^{+}_{\underline{w}}$ corresponds to a matter spinor and therefore M consists basically of matter fields whose truncation sets $M = 0$. This is indeed possible because all the terms in M contain two gammas and a combination of them can always be set to zero. This truncation gives the pure $N = 1, d = 5$ supergravity action Eq. (11.98) and the supersymmetry transformation rule of the five-dimensional gravitino [664]:

$$\delta_{\hat{\epsilon}}\hat{\psi}_{\hat{a}} = \left\{\hat{\nabla}_{\hat{a}} - \frac{1}{8\sqrt{3}}(\hat{\gamma}^{\hat{b}\hat{c}}\hat{\gamma}_{\hat{a}} + 2\hat{\gamma}^{\hat{b}}\hat{g}^{\hat{c}}{}_{\hat{a}})\hat{G}_{\hat{b}\hat{c}}\right\}\hat{\epsilon}. \tag{13.105}$$

The relation between the six- and five-dimensional fields is

$$\hat{\hat{g}}_{\underline{ww}} = -1, \quad \hat{\hat{B}}{}^{-}_{\hat{\mu}\underline{w}} = \frac{1}{\sqrt{3}}\hat{V}_{\hat{\mu}}, \quad \hat{\hat{g}}_{\hat{\mu}\underline{w}} = \frac{1}{\sqrt{3}}\hat{V}_{\hat{\mu}}, \quad \hat{\hat{g}}_{\hat{\mu}\hat{\nu}} = \hat{g}_{\hat{\mu}\hat{\nu}} - \tfrac{1}{3}\hat{V}_{\hat{\mu}}\hat{V}_{\hat{\nu}}, \tag{13.106}$$

with the $\hat{B}^{-}_{\mu\nu}$ components determined by anti-self-duality.

The $AdS_3 \times S^3$ solution can be reduced preserving all the supersymmetry in two directions: the direction of the Hopf fiber of the S^3 when we see it as a fibration over S^2 (i.e. the coordinate ψ in the Euler-angle parametrization Eq. (A.97)) and the analog in AdS^3 when we see it as a fibration over AdS_2, i.e. the coordinate η in

$$d\Pi^2_{(3)} \equiv \tfrac{1}{4}\left[d\Pi^2_{(2)} - (d\eta + \sinh(\chi/2)d\phi)^2\right]. \tag{13.107}$$

Actually it is also possible to perform a dimensional reduction in a combination of the two directions: on rotating by an angle ξ,

$$w = \frac{R_3}{2}(\cos\xi\,\eta + \sin\xi\,\psi), \qquad y = -\sin\xi\,\eta + \cos\xi\,\psi, \tag{13.108}$$

and reducing in the direction w, we obtain the two-parameter family [272]

$$\boxed{\begin{aligned} d\hat{s}^2 &= R_2^2 d\Pi^2_{(2)} - R_2^2 d\Omega^2_{(2)} - R_2^2(dy + \cos\xi\cos\theta\,d\varphi - \sin\xi\sinh\chi\,d\phi)^2, \\ \hat{G} &= \sqrt{3}R_2(\cos\xi\,\omega_{AdS_2} - \sin\xi\,\omega_{S^2}), \qquad R_2 = R_3/2. \end{aligned}} \tag{13.109}$$

It is maximally supersymmetric for any ξ and can be obtained as the near-horizon limit of the $d = 5$ rotating extreme BH [422] which, as usual, has only half of the maximal supersymmetry [614]. $\sin\xi$ plays the role of the rotation parameter $\jmath < 1$ of [906] and it is no surprise that, when $\xi = 0$, we recover the near-horizon limit of the static $d = 5$ extreme BH, which has the geometry of $AdS_2 \times S^3$ $(y = \psi)$ [228]. It is a bit more surprising that for $\xi = \pi/2$, we recover the near-horizon limit of the $d = 5$ extreme string solution [441]. Although many computations can be performed for arbitrary ξ and then we can take the limits $\xi \to 0, \pi/2$, it is clear that this is yet another example of a family of solutions that seems to depend continuously on a parameter, while the physical properties (the symmetry superalgebras, for instance) do not.

The above solution admits a description as a homogeneous, reductive but non-symmetric spacetime that can be used to compute its symmetry superalgebra [26].

Before the KG6 solution can be reduced, preserving all its supersymmetries, we need to identify isometric directions satisfying the truncation condition. One of the directions is found after performing a u-dependent rotation,

$$x^3 = \cos\left(\frac{\lambda_6}{2}u\right) x^{3\prime} + \sin\left(\frac{\lambda_6}{2}u\right) x^{4\prime}, \;\; x^4 = -\sin\left(\frac{\lambda_6}{2}u\right) x^{3\prime} + \cos\left(\frac{\lambda_6}{2}u\right) x^{4\prime},$$
$$v = v' + \frac{\lambda_6}{2} x^{3\prime} x^{4\prime},$$

that leaves the solution in the form

$$\boxed{\begin{aligned} d\hat{\hat{s}}^2 &= 2du\left[dv' + \frac{\lambda_6^2}{8}(x^2 + y^2)du + \lambda_6 x^{3\prime} dx^{4\prime}\right] - d\vec{x}^{\prime\,2}_{(4)}, \\ \hat{\hat{B}}^- &= -\lambda_6 du \wedge (x^1 dx^2 - x^{3\prime} dx^{4\prime}). \end{aligned}} \tag{13.110}$$

On reducing now in the isometric coordinate $w \equiv x^{4\prime}$, we obtain the KG5 solution Eq. (10.31) with $\lambda_5 = -\sqrt{3}\lambda_6$. A similar rotation in x^1 and x^2 produces the same result. The isometric direction $w = (1/\sqrt{2})(u + v)$ can also be used. If we perform the two rotations and reduce in the direction w, we obtain a $d = 5$ maximally supersymmetric Gödel-like solution [420]:

$$\boxed{d\hat{s}^2 = (dt + \omega)^2 - d\vec{x}^{\,2}_{(4)}, \qquad \hat{V} = \sqrt{3}\,\omega, \qquad \omega = \lambda_6(x^1 dx^2 - x^3 dx^4).} \tag{13.111}$$

It is not known whether there are more maximally supersymmetric solutions in $N = 1$, $d = 5$ supergravity, although in principle it is known how to construct all the solutions of this theory that preserve some supersymmetry [420]. The relations among all the known vacua are represented in Figure 13.1.

13.4.3 *Relation to the $N = 2, d = 4$ vacua*

The dimensional reduction of $N = 1, d = 5$ supergravity gives $N = 2, d = 4$ supergravity coupled to a vector multiplet that can be consistently truncated as shown in Section 11.2.5. The discussion at the beginning of the previous section applies to this situation: maximal supersymmetry survives the dimensional reduction only if no matter fields are generated. This condition cannot be satisfied for the Gödel-like solution Eq. (13.111) and only the KG5 Eq. (10.31) and the near-horizon limit of the rotating BH Eq. (13.109) and string give maximally supersymmetric four-dimensional solutions: the KG4 solution Eq. (10.27) with $\lambda_4 = (2/\sqrt{3})\lambda_5$ and the dyonic RB solution in which the rotation parameter $\sin\xi$ now plays

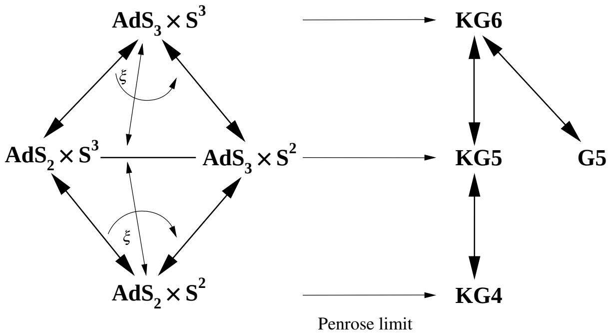

Fig. 13.1. Relations between the non-trivial $d = 4$, 5, 6 vacua with eight supercharges.

the role of the electric–magnetic-duality rotation parameter:

$$\boxed{ds^2 = R_2^2\, d\Pi^2_{(2)} - R_2^2\, d\Omega^2_{(2)}, \qquad F = -R_2\big(\cos\xi\, \omega_{\mathrm{AdS}_2} - \sin\xi\, \omega_{\mathrm{S}^2}\big).} \tag{13.112}$$

It is amusing to observe that a spatial rotation in $d = 6$ has become an electric–magnetic-duality rotation in $d = 4$ [338]. This is one of the phenomena we will study in Part III.

13.5 Partially supersymmetric solutions

Now that we have studied maximally supersymmetric vacua, we are ready to study solutions with less unbroken supersymmetry that tend asymptotically to one of them.

As explained in Section 6.1.2, spacetimes (supersymmetric or not) with these asymptotics have well-defined conserved bosonic charges associated with the isometries of the vacua they approach asymptotically, but it is also possible to assign to them supercharges using the same formalism [1]. Then, it is possible (at least formally) to associate with these solutions states in the quantum theory transforming under the vacuum supersymmetry algebra, with well-defined quantum numbers associated with all the generators. If the solution is invariant under the supersymmetry transformation generated by a spinor ϵ^α that asymptotically approaches a combination of vacuum Killing spinors $\epsilon^\alpha \to \epsilon^A \kappa_{(A)}{}^\alpha$, where the ϵ^A are (commuting) constants, then the (BPS) state will be annihilated by the supercharge $\bar{\epsilon}^A Q_{(A)}$:

$$\delta_\epsilon |s\rangle \sim \epsilon^A Q_{(A)} |s\rangle = 0. \tag{13.113}$$

The game now will be to use results about the states obtained through the study of the superalgebra to predict properties of the associated solutions.

The superalgebra is a very powerful tool. It is a fact that, in all cases, one can construct a manifestly positive quadratic combination of the supercharges that equals a combination of the bosonic charges including the energy,[14] which turns out to bound it below.

[14] This has been studied mostly in Poincaré superalgebras that can be applied to asymptotically flat spacetimes. There are less general results for AdS superalgebras and none for KG Hpp-wave superalgebras.

These *supersymmetry* or *Bogomol'nyi (B) bounds*[15] have very important consequences, for instance the positivity of the energy itself and the stability of the theory and the BPS states that saturate the bound. When the bound is saturated, the state is annihilated by some combination of supercharges and it is supersymmetric. In the supergravity theory general bounds on the energy and other charges of solutions can be established under mild assumptions using *Witten–Nester–Israel constructions* [442, 596, 719, 937] (see also [600]), the simplest of which was used in Section 6.3 to prove the positive-energy theorem. The solutions associated with supersymmetric states (with the same charges/quantum numbers) are also supersymmetric (they admit Killing spinors) and also have special stability properties.

Our goal in this section is to study these B bounds in Minkowski spacetimes and their consequences for supergravity solutions.

13.5.1 Partially unbroken supersymmetry, supersymmetry bounds, and the superalgebra

Let us consider a supersymmetric state $|s\rangle$ in the quantum theory associated with the N-extended $d = 4$ Poincaré superalgebra, including central charges. The state describes a quantum object in Minkowski spacetime and will be annihilated by a combination of the Minkowski (Poincaré) supercharges (the Minkowski vacuum is annihilated by all of them):

$$\delta_\epsilon |s\rangle \sim \bar{\epsilon}^i_\alpha Q^{i\,\alpha} |s\rangle = 0. \tag{13.114}$$

The anticommutator of this supercharge with itself gives (assuming that $|s\rangle$ is normalizable)

$$\bar{\epsilon}\mathfrak{M}\epsilon = 0, \tag{13.115}$$

where, using Eq. (5.70), the matrix $\mathfrak{M}$ is given by the expressionc

$$\mathfrak{M} \equiv i\delta^{ij}\gamma^a P_a + iZ^{[ij]} + \gamma_5 \tilde{Z}^{[ij]} + \gamma^a Z_a^{(ij)} + i\gamma_5\gamma^a Z_a^{[ij]} + i\gamma^{ab} Z_{ab}^{(ij)} + \gamma_5\gamma^{ab}\tilde{Z}_{ab}^{(ij)}, \tag{13.116}$$

in which we have replaced the operators by their values on $|s\rangle$. This equation has $4N - \mathrm{rank}(\mathfrak{M})$ independent solutions (preserved supersymmetries) if $\det\mathfrak{M} = 0$. Finding all the values of P_a, Z^{ij}, $\tilde{Z}^{ij}, \ldots$ for which $\mathfrak{M}$ is singular would be exceedingly difficult. However, there are simple solutions with a simple physical interpretation.

Let us consider first a state describing a point-like, massless, uncharged (all $Z = 0$) particle. In the reference system in which it moves in direction 1 its null momentum takes the form $(P^\mu) = (p, \pm p, 0, \ldots, 0)$ and $\mathfrak{M}$ is

$$\mathfrak{M} = ip\delta^{ij}\gamma^0\big(1 \pm \gamma^0\gamma^1\big). \tag{13.117}$$

$\mathfrak{M}$ is singular: $(\gamma^0\gamma^1)^2 = 1$ and $\mathrm{Tr}(\gamma^0\gamma^1) = 0$ imply that $\frac{1}{2}(1 \pm \gamma^0\gamma^1)$ is idempotent (and, therefore, a projector) with trace equal to 2, and, thus, with eigenvalues $+1$ with multiplicity 2 and 0 with multiplicity 2. Half of the supersymmetries of this state will be preserved,

[15] These bounds were first discovered by Bogomol'nyi [163] in a purely bosonic context and then by Witten and Olive [963] as a consequence of supersymmetry.

precisely those generated by $\bar{\epsilon}^{i}_{\alpha} Q^{i\,\alpha}$, where $\gamma^{\pm}\epsilon = \frac{1}{2}(\gamma^0 \pm \gamma^1)\epsilon = 0$. This is precisely the constraint satisfied by the Killing spinors of generic pp-wave solutions, some of which can be associated with the gravitational field of massless point-particles (see Section 10.3). All supergravity theories admit these supersymmetric states and solutions.

The next simplest case is that of a massive point-particle with mass M and electric charge Q (massive neutral particle states cannot be supersymmetric) in $N = 2, d = 4$ Poincaré SUEGRA ($Z^{[ij]} = Q\epsilon^{ij}$). In the particle's rest frame $(P^{\mu}) = (M, 0, \ldots, 0)$ and

$$\mathfrak{M} = i\gamma^0 M\left[\delta^{ij} + Q/\left(M\gamma^0\epsilon^{ij}\right)\right]. \tag{13.118}$$

The 8×8 matrix $\mathfrak{M}$ is singular if and only if $Q = \pm M$, in which case it is proportional to another projector: $\frac{1}{2}(\delta^{ij} \pm \gamma^0\epsilon^{ij})$. The unbroken supersymmetries are generated by the supercharges $\bar{\epsilon}^i Q^i$, where ϵ^i satisfies the constraint

$$\tfrac{1}{2}(\delta^{ij} \pm \gamma^0\epsilon^{ij})\epsilon^j = 0. \tag{13.119}$$

The rank of this projector (and $\mathfrak{M}$) is 4: half of the supersymmetries are unbroken.

Which solution of $N = 2, d = 4$ SUEGRA could we associate this supersymmetric state with? It should be a static, spherically symmetric solution (the closest to a point-like object we can have) with charge $M = |Q| = 2|q|$ with the conventions of Chapter 8: an extreme, electrically charged Reissner–Nordström BH.[16] We will see that indeed the ERN BH admits Killing spinors that satisfy the above constraint.

What happens in the general case? Let us focus first on the $d = 4$ N-extended case with strictly central charges. Following [384], to study this case, it is convenient to use a (complex) Weyl representation in which the electric and magnetic central-charge matrices are combined into a single, complex, antisymmetric matrix $\mathbf{Z}^{ij}$, which has $[N/2]$ skew eigenvalues Z_i. For massive states, in the rest frame, $\mathfrak{M}$ is singular whenever the absolute value of one of these skew eigenvalues is equal to the mass, $|Z_i| = M$. The amount of supersymmetry preserved depends on the number of skew eigenvalues whose absolute values are equal to M. When all of them are equal to $M \neq 0$, half of the supersymmetries will be unbroken. Let us see what happens in the most interesting cases $N = 1, 2, 4, 8, d = 4$.

N = 1 There are no massive supersymmetric states.

N = 2 Half of the supersymmetries are unbroken when[17]

$$M = |Z|.$$

N = 4 Half of the supersymmetries are preserved when

$$M = |Z_1| = |Z_2|,$$

and a quarter when

$$M = |Z_1| \neq |Z_2|.$$

---

[16] With magnetic charge $\tilde{Z}^{[ij]} = P\epsilon^{ij}$ and $P = \pm M$, the projector would be $\frac{1}{2}(\delta^{ij} \pm \gamma_5\gamma^0\epsilon^{ij})$ and, in the dyonic case $P^2 + Q^2 = M^2$, the projector would be $[\delta^{ij} \pm (\cos\xi + i\sin\xi\,\gamma_5)\gamma^0\epsilon^{ij}] = [\delta^{ij} \pm e^{i\gamma_5}\gamma^0\epsilon^{ij}]$.

[17] In $N = 2$, $\mathbf{Z}^{ij} = Z\epsilon^{ij}$ and $Z = Q + iP$ is the only skew eigenvalue.

N = 8 Half of the supersymmetries are unbroken when

$$M = |Z_1| = |Z_2| = |Z_3| = |Z_4|,$$

a quarter when

$$M = |Z_1| = |Z_2| \neq |Z_{3,4}|,$$

and an eighth when

$$M = |Z_1| \neq |Z_{2,3,4}|.$$

In all cases, the projectors on the spinors are of the form $\delta^{ij} + \gamma^0 \alpha^{ij}$, where α^{ij} depends on the SUEGRA theory and the specific solution we are considering: different combinations of charges can lead to the same $|Z_i|$s. These combinations are usually related by dualities of the theory (e.g. electric–magnetic duality in $N = 2$, $d = 4$ SUEGRA, which multiplies Z by a complex phase). There is one projector for each $\frac{1}{2}$ factor of broken supersymmetry and all the projectors must commute among themselves in order for them to be compatible. In all cases, the associated solutions are generalizations of the ERN BH like those of Section 12.2.

Let us now consider the case in which there is one (real) quasi-central charge $Z^{(p)}_{a_1\cdots a_p}$. In the appropriate reference frame it is always possible to write

$$\mathfrak{M} = i\gamma^0 M\big(\delta^{ij} + (Z^{(p)}/M)\gamma^0\gamma^1 \cdots \gamma^p \alpha^{ij}\big), \tag{13.120}$$

which is singular when $M = Z^{(p)}$. The projector is not one of those associated with point-particles. Actually, quasi-central charges with p Lorentz indices are associated in the corresponding SUEGRA theory with $(p + 1)$-form potentials[18] $A^{(p+1)}{}_{\mu_1\cdots\mu_{p+1}}$. As we will see in Chapter 18, these potentials couple naturally to the $(p + 1)$-dimensional worldvolumes of p-dimensional extended objects (*p-branes*) just as vector fields couple to the worldlines of point-particles. Thus, the projector and the supersymmetric state correspond to a half-supersymmetric p-brane. We will study the SUEGRA solutions in Chapter 19.

In a more general case, there can be several non-vanishing quasi-central charges. The corresponding state is then interpreted as composed of several supersymmetric p-branes that intersect (see Section 19.6), generically one for each projector. This interpretation, which can also be applied to supersymmetric point-particle states (Chapter 20), is based on the observed property that supersymmetric objects can be in equilibrium: there is a cancelation between gravitational attraction and other interactions and this allows the existence of solutions that describe several of these supersymmetric objects in equilibrium. The simplest example is provided by the Majumdar–Papapetrou solutions that describe several ERN BHs in equilibrium.

$\mathfrak{M}$ has another important property: it is roughly the square of the supercharges and, thus, we can expect its eigenvalues to have some positivity properties. Indeed, it can be shown that the Hamiltonian of supersymmetric QFTs is non-negative [310] or, equivalently, that $M \geq 0$. In four-dimensional theories with extended supersymmetry one can go even further

[18] One can view them as the components of the gauge superpotential associated with the quasi-central charges: $A^{(p+1))}{}_\mu{}^{a_1\cdots a_p} Z^{(p)}_{a_1\cdots a_p}$ if we think of SUGRA theories as gauge theories of given superalgebras, as we did in Chapter 5.

and prove that the mass is bounded from below by the skew eigenvalues of the central charges matrix $\mathbf{Z}^{ij}$ [384, 963],

$$M \geq |Z_i|, \qquad \forall i = 1, \ldots, [N/2], \tag{13.121}$$

and the results can be generalized to other dimensions. These bounds are known as *Bogomol'nyi*, *BPS*, or *supersymmetry bounds* and play a crucial role in the stability of states and theories. Broken supersymmetry is restored when one of these bounds is saturated, as we have seen. Supersymmetric ("BPS") states then have the minimal masses allowed for given values of the central charges. The central charges cannot change, because there are no (perturbative) states in the theory that carry them and, then, the masses of BPS states cannot diminish and the states cannot decay and are stable.

As for the associated BPS solutions and SUEGRA theories, some stability properties can also be proven, such as the positive-energy theorem of GR whose proof, inspired by $N = 1, d = 4$ SUGRA, we gave in Section 6.3. The relation between positivity of the energy and the supersymmetry algebra was studied in [575]. There are generalizations based on the WNI construction (see e.g. [442] for $N = 2, d = 4$). As for the stability of solutions, it manifests itself, most remarkably, in the absence of Hawking radiation ($T = 0$) from ERN and other supersymmetric BHs. This establishes an interesting link between BH thermodynamics and supersymmetry (see e.g. [745]), which we will use in Chapter 20.

In the next section we are going to review the (not maximally) supersymmetric solutions of the simplest theories, $N = 1, 2, 4, d = 4$ SUGRA, and their relation to the supersymmetry bounds one can derive from the superalgebras.

13.5.2 Examples

$N = 1, d = 4$ *Poincaré supergravity.* The only solutions with partially unbroken supersymmetry in this theory are the purely gravitational pp-waves given by Eqs. (10.24) with $C = 0$. The Killing-spinor equation $\nabla_\mu \kappa = 0$ is solved for any constant spinor κ satisfying the constraint $\gamma^u \kappa = 0$, which is precisely what we expected from the superalgebra. The absence of massive supersymmetric solutions can be understood as a consequence of the $N = 1$ supersymmetry bound $M \geq 0$. This is in agreement with the finite temperature and entropy of all Schwarzschild BHs. The bound is also in agreement with the cosmic-censorship conjecture.

Things are different in Euclidean $N = 1, d = 4$ supergravity: the integrability condition of the Killing-spinor equation,

$$R_{\mu\nu}{}^{ab} \gamma_{ab} \kappa = 0, \tag{13.122}$$

also admits solutions when the curvature is (anti-)self-dual because then it is proportional to a projector $\frac{1}{2}(1 \pm \gamma^1 \gamma^2 \gamma^3 \gamma^4)$,

$$R_{\mu\nu}{}^{ab} \gamma_{ab} = R_{\mu\nu}{}^{ab} \gamma_{ab} \tfrac{1}{2}(1 \pm \gamma^1 \gamma^2 \gamma^3 \gamma^4) \kappa, \tag{13.123}$$

and half of the supersymmetries are preserved. Then all metrics (in any dimension!) with (anti-)self-dual curvature (special SU(2) holonomy, see footnote 5) preserve half of the supersymmetries. These are the metrics of (anti-)self-dual gravitational instantons, which will be discussed in Section 9.2.1. In the frame in which the spin connection is also (anti-)self-dual, the Killing-spinor equation takes the form

$$\nabla_\mu \kappa = \left[\partial_\mu - \tfrac{1}{4}\omega_\mu{}^{ab}\gamma_{ab}\tfrac{1}{2}(1 \pm \gamma^1\gamma^2\gamma^3\gamma^4)\right]\kappa = \partial_\mu \kappa = 0, \tag{13.124}$$

where we have used $\frac{1}{2}(1 \pm \gamma^1\gamma^2\gamma^3\gamma^4)\kappa = 0$. Then, in this frame, the Killing spinors are the constant spinors that satisfy that constraint.

It is, actually, possible to reformulate the Killing-spinor-equation problem in terms of holonomy, which turns out to be an extremely powerful tool. See, for instance, [484].

An interesting case of a metric of special SU(2) holonomy is the Euclidean Taub–NUT solution because its supersymmetry can be understood from the point of view of BPS saturation: it is in principle possible to include the NUT charge in the BPS bound in asymptotically Taub–NUT Lorentzian spaces: $M + |N| \geq 0$. This explains why the massless Lorentzian Taub–NUT solution is not supersymmetric. In the rotation to Euclidean time, the bound becomes $M - |N| \geq 0$, although M no longer has the interpretation of mass,[19] and can be saturated for $M = |N|$, which corresponds to the self-dual Euclidean Taub–NUT solution Eq. (9.12).

$N = 2, d = 4$ *Poincaré supergravity.* All the field configurations (metric and U(1) vector field) admitting Killing spinors in $N = 2, d = 4$ SUEGRA were constructed by Tod in [893]. His construction includes field configurations that do not solve the (Einstein–Maxwell) equations of motion. The supersymmetric solutions fall into two classes: pp-waves such as those of Eqs. (10.24) (including the maximally supersymmetric KG4 solution Eq. (10.27)) and the IWP solutions of Eqs. (9.58) (including the maximally supersymmetric RB solution Eq. (8.90)). Generically, the solutions of both families preserve half of the supersymmetries.

The IWP family contains many different solutions. We can use the superalgebra to study the asymptotically flat ones which are the (multi-)Kerr–Newman solutions Eq. (9.55) with the saturated BPS bound $M = |Q + iP| = |Z|$ for any value of the angular momentum, which does not appear in the bound, basically because the Lorentz generators do not appear in the supercharge anticommutator. All these solutions are singular for non-vanishing angular momentum. The only regular ones are the ERN solutions. These can be seen as interpolating between the Minkowski vacuum at asymptotic infinity and the RB vacuum at the near-horizon limit, which supports their interpretation as solitons [433]. This vacuum interpolation, which takes place in more general cases [334, 452], plays a very important role in many instances, such as the AdS/CFT correspondence [679] (for a review see [23]) and non-conformal generalizations [170].

Let us now consider the thermodynamics of generic dyonic RN BHs. Their temperature and entropy are given by

$$T = \frac{1}{2\pi}\frac{\sqrt{M^2 - |Q + iP|^2}}{M + \sqrt{M + \sqrt{M^2 - |Q + iP|^2}}} \overset{M \to |Q+iP|}{\longrightarrow} 0, \tag{13.125}$$

$$S = \pi\left[M + \sqrt{M^2 - |Q + iP|^2}\right]^2 \overset{M \to |Q+iP|}{\longrightarrow} \pi M^2. \tag{13.126}$$

[19] Compare this bound with Eq. (9.41).

Observe that, like any other physical quantity that depends only on the metric, which is invariant under the electric–magnetic duality of the theory, both T and S are also duality-invariant: they can be written in terms of the single mass matrix skew eigenvalue $Z = |Q + iP|$. On the other hand, in the extreme limit, which is also here (in the absence of angular momentum) the supersymmetric limit, $T \to 0$ and there is no Hawking radiation, in agreement with our previous arguments about stability, while $S \to \pi M^2 = \pi |Z|^2$ is completely determined by the central charges. We discussed at length the arguments in favor of and against there being a finite value of the entropy for ERN BHs in Section 8.6. In any case, we can always say that the area of the ERN BH is finite and given by a duality-invariant combination of the central charges. This observation is relevant because we will find that, in general, charged supersymmetric spherically symmetric solutions have a regular horizon with finite area only when they preserve the same amount of supersymmetry as the ERN BH (i.e. four supersymmetries).

If we consider asymptotically Taub–NUT IWP solutions, we find that all of them satisfy the saturated bound $M + |N| = |Q + iP| = |Z|$ (with or without angular momentum).

The supersymmetry of some solutions of $N = 2, d = 4$ AdS supergravity have also been studied in [203, 634, 811]. The supersymmetry of the largest class of solutions, given by Plebański and Demiański in [772, 773], was studied in [27].

$N = 4, d = 4$ *supergravity.* The bosonic part of this theory is described by the action Eq. (12.58). All the field configurations admitting Killing spinors were found by Tod in [894] and they include pp-waves and the SWIP solutions of [130, 613] described in Section 12.2.1. These have been studied more thoroughly. There are SWIP solutions that preserve half of the supersymmetries (i.e. eight) and SWIP solutions that preserve a quarter (i.e. four) (these include the IWP solutions).

Actually, to study them, it is convenient to focus on the asymptotically flat ones and use the supersymmetry bounds of the theory. There are two BPS bounds: $M^2 - |Z_{1,2}|^2 \geq 0$, neither of which is invariant under the dualities of the theory. However, we can construct a duality-invariant *generalized BPS bound* by taking their product and dividing by M^2:

$$M^2 + \frac{|Z_1 Z_2|^2}{M^2} - |Z_1|^2 - |Z_2|^2 \geq 0. \tag{13.127}$$

This bound is satisfied by all the regular static non-extreme SWIP solutions of [665] and it is saturated by all the supersymmetric SWIP solutions of [130]: the skew eigenvalues of the central-charge matrix correspond precisely to the combinations of electric and magnetic charges of Eq. (12.83) and the product of the two supersymmetry bounds gives precisely the supersymmetry parameter r_0^2 of the SWIP solutions, Eq. (12.84).

The bound can be saturated in two different ways: when $M = |Z_{1,2}| \neq |Z_{2,1}|$, a quarter of the supersymmetries are unbroken, and when $M = |Z_1| = |Z_2|$, half of the supersymmetries are unbroken. The only supersymmetric solutions with regular horizons are the static ones with only a quarter of the supersymmetries preserved. All the supersymmetric solutions have zero temperature.

The entropy (see the second of Eqs. (12.84)) and the temperature of the BHs of $N = 4, d = 4$ can be expressed in a form that is manifestly invariant under the two duality groups of the theory: SL(2, $\mathbb{R}$) (S duality) and SO(6) (T duality). The entropy of the static supersymmetric BHs that preserve a quarter of ths supersymmetries is given by

$$S = 4\pi \left| |Z_1|^2 - |Z_2|^2 \right| = 4\pi \sqrt{I_4}, \tag{13.128}$$

where I_4 is the quartic invariant defined in Eq. (12.62), which depends on the charges but not on the asymptotic value of the axidilaton. In the extreme limit, thus, when the entropy is finite, its expression in terms of the central charges is not only duality-invariant but also *moduli-independent*. This is a very important property that suggests that the BH entropy does indeed count microstates, since this counting would not be changed by small variations of the moduli.

We will find generalizations of these facts when we study BH solutions of $N = 8, d = 4$ SUEGRA in Chapter 20.

Part III

Gravitating extended objects of string theory

There is geometry in the humming of the strings.

Pythagoras

14
String theory

In this chapter we start the study of the extended objects that appear in the non-perturbative spectrum of string theory, the subject of the third part of this book. In this part we will make use of all the techniques we have developed in the first and second parts, whose main goal was to serve as a preparation for the third.

In a certain sense, this third part also presents the synthesis and (it is hoped) culmination of the ideas presented in the previous two in the framework of string theory: on the one hand, string theory includes a presumably consistent theory of quantum gravity that contains the gravitons described at lowest order by the Fierz–Pauli theory we studied in Chapter 3 [833, 834]. There are two main differences from the non-renormalizable theory of GR: the presence of a dimensionless coupling constant different from the Planck length and the presence of terms of higher order in derivatives. Furthermore, consistent string theories have spacetime supersymmetry and, therefore, supergravity, which we studied in Chapters 5 and 13. On the other hand, string theory incorporates naturally extra dimensions that have to be compactified. Thus, the ideas of Kaluza and Klein studied in Chapter 11 are also integrated into the picture.

Finally, the Schwarzschild, Reissner–Nordström, pp-wave, etc. solutions studied in other chapters are also solutions of string theory and it is natural to try to use them to solve the puzzles that arise when one tries to do quantum mechanics in those backgrounds: the information and entropy problems. If string theory is really a good theory of quantum gravity, then it should help us to solve them and we will see to what extent it succeeds in the last chapter (Chapter 20) of this part.

The attempts to solve these long-standing problems have been made possible by recent developments in string theory (essentially dualities and D-branes) and also by a change of perspective that we could call the “spacetime approach,” which is based on the effective-field theories, when further advance with the “worldsheet approach” was becoming increasingly difficult and slow. Of course, the two approaches are complementary and there has been a considerable amount of feedback between them. In fact, some of the most interesting things that we have learned in this period are the relations between the two of them. The logic of these relations is represented schematically in Figure 14.1 and it is worth pausing for a moment to describe it in detail.

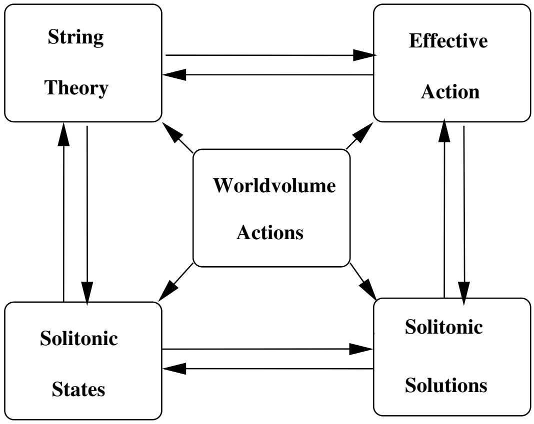

Fig. 14.1. Most of the recent progress in string theory has been based on the relations represented in this diagram between different aspects of the theory. The implications of duality symmetries in each box must be related to similar implications in other boxes.

The box in the center represents the worldvolume theories of all the extended objects of string theory, including fundamental strings. They give, to start with, the perturbative definition of string theory (the worldsheet formulation), represented by the box in the upper-left-hand corner. This worldsheet formulation allows us to calculate the perturbative spectrum and also find some non-perturbative states (the D-branes), which are represented by the lower-left-hand corner of the diagram (their worldvolume actions are also represented by the box in the center). The low-energy effective-field theories that describe the dynamics of the massless states of the perturbative spectrum are represented in the upper-right-hand corner. These theories are usually supergravity theories that have certain (on-shell) global symmetries that transform and mix the fields that represent the string massless modes. The equations of motion of these effective-field theories can be derived (a most important point) from consistency conditions (conformal invariance or κ-symmetry) of worldvolume theories (the box in the center) in general backgrounds. The effective field theories also admit solitonic solutions, which are represented in the lower-right-hand corner. The solitonic solutions can be excited and deformed and their effective dynamics is, yet again, given by worldvolume actions.

The meaning of the $\downarrow, \rightarrow,$ and $\downarrow$ arrows and the arrows that come from the box in the center of this diagram is clear. The progress made in this field comes from the realization of the existence and use of the remaining arrows of the figure.

The main idea is that, generically, the global symmetries of the effective-field theories correspond to *dualities* of the string theories.[1]

Some of these dualities are essentially perturbative (in the string coupling constant, that we will define later) and can be found and studied using the worldsheet approach. They are

[1] This is the point of view proposed in [583], but a more precise statement that includes more cases would be that the relations between different effective field theories correspond to dualities of the corresponding string theories. Some of the relations can be described as global symmetries of a single effective action, but in other cases there are relations between very different effective actions.

generically called *T dualities*, and the archetype of them (known long before as T duality, T from "target space," see e.g. [456]) relates string theories compactified on circles of dual radii. This is an exact symmetry at all orders in string perturbation theory [33].

On top of these, there are the so-called *S dualities* which are non-perturbative in the string coupling constant and cannot be studied using the standard worldsheet approach. In fact, their existence in four-dimensional heterotic string theory was suggested by the existence of the corresponding global symmetry in the effective action [397, 803, 842] which is nothing but that of $N = 4, d = 4$ SUGRA. As we saw in Section 12.2, the equations of motion of this theory are invariant under global $\mathrm{SL}(2, \mathbb{R})$ transformations, some of which invert the dilaton field which can be interpreted as the string coupling constant. In general S dualities are associated with this group[2] [583] and involve the inversion of the dilaton (string coupling constant) and the interchange or electric and magnetic fields. Another interesting example that we will study in Chapter 17 is $N = 2B, d = 10$ SUGRA, the effective-field theory of the type-IIB superstring: the equations of motion of the supergravity theory are invariant under global $\mathrm{SL}(2, \mathbb{R})$ rotations that invert the dilaton, which suggests the existence of an S duality between type-IIB superstring theories.

The T- and S-duality groups are sometimes (in type-II theories) part of a bigger duality group called in [583] the U *duality group*.

The interpretation as string dualities of the supergravity symmetries (the upper $\leftarrow$ arrow) is possible if the string spectra reflect the same duality properties. For T dualities, this is easy to see in the perturbative spectrum, but S dualities necessarily imply the existence of new non-perturbative states. Thus, using the upper $\leftarrow$ relation, we start learning new things about string theory and its spectrum. Needless to say, the worldvolume actions for the corresponding states must also be related by the same dualities.

The symmetries of the supergravity theories also relate different solitonic solutions. In some cases this has been used to generate new solitonic solutions out of old ones. In the end one should be able to obtain complete duality-invariant families of solutions. Now, the relation between effective-field theories and string theories can be used to relate solitonic supergravity solutions to perturbative and non-perturbative string states (the lower $\leftarrow$ arrow). Whereas in the supergravity theories the duality groups are continuous, in string theory quantum effects generically break them to the discrete subgroups that result from the restriction from real numbers to integers. In particular, the S-duality group is broken by charge quantization of the string states to $\mathrm{SL}(2, \mathbb{Z})$ [843]. The full spectrum is invariant under that group [844].

To develop these relations, we must find a dictionary to interpret, in terms of string theory, supergravity field configurations and symmetries, because only the fields associated with massless string-theory modes appear in the effective field theories, but string dualities often involve massive modes. However, these massive modes are usually charged and couple to the massless long-range potentials that appear in supergravity,[3] and we know their transformations under dualities, which help us to know how the massive and charged string

[2] But not the other way around: the T-duality group can contain $\mathrm{SL}(2, \mathbb{R})$ subgroups.

[3] The only bosonic fields apart from the potentials and the metric are the scalars, whose interpretation is different: they represent coupling constants or geometrical data of the compactification space, *moduli*.

modes behave under those dualities. These couplings can be read off from the worldvolume actions of those states.

We encountered a similar phenomenon in KK theories in Chapter 11: the massive KK modes of the spectrum are charged with respect to the massless KK vector field that appears in the KK (effective) action. The massive KK modes reappear in the theory as massive, electrically charged solitonic solutions (extreme, electrically charged BHs). The S duality of the four-dimensional KK theory (in this case, just a $\mathbb{Z}_2$ group) could be used to generate extreme, magnetic BH solutions, which should represent non-perturbative states of the KK theory.

A final ingredient in all these relations is supersymmetry and supersymmetric ("BPS") states, which are associated with supersymmetric supergravity solutions. BPS states enter into shortened supermultiplets, which have fewer states. This and other properties (the saturation of the Bogomol'nyi bounds, which are duality-invariant [743, 844]) do not depend on any continuous parameters and remain valid in the limits of strong and weak coupling. Therefore, these states are particularly well suited for studying, for instance, S dualities.

In this part we are going to investigate all these relations using as our main tool the effective actions, supergravity actions, in order finally to apply them to the calculation of entropies of extreme BHs. This chapter is devoted to a very brief introduction to string theory.[4] We will study in it, from the worldsheet point of view, the simplest, and most characteristic, string duality: T duality.

In Chapter 15 we study the origin and meaning of the string effective field theories which we will use to investigate string dualities in the following chapters. In this chapter we will start studying T duality in curved spaces from the effective-action point of view.

The ten-dimensional supergravities which are effective field theories of ten-dimensional superstring theories will be studied in the next two chapters (Chapters 16 and 17). We will be able to interpret some of their symmetries in stringy terms, but many results will have to wait until Chapter 18, in which we will introduce extended objects and we will realize that the $(p + 1)$-form potentials that appear in the supergravities are naturally the fields to which extended, charged p-dimensional objects (*p-branes*) couple. Then we will see the implications of the supergravity symmetries for the non-perturbative spectrum of string theory.

In Chapter 18 we are going to study generically extended objects and in Chapter 19 we will focus on the extended objects of string theory. We will first find which these extended objects, implied by duality, are. Then, we will identify the corresponding classical solutions and, to strengthen the connection between classical solutions and string states, we will calculate the masses that duality predicts and we will use the results to identify the integration constants of the classical solutions. Here we can interpret these solutions as the long-range fields produced by one of these extended objects. We will check that they preserve the right amount of unbroken supersymmetry according to the supersymmetry algebra. Next, we will study the worldvolume effective actions of these objects. Finally, we will see solutions that describe several of these objects intersecting each other, still preserving some amount of supersymmetry. Related states must exist in string theory.

[4] This chapter is no substitute for the study of the many reviews and books on string theory: [37, 39, 468, 473, 609, 624, 625, 673, 749, 779, 831, 832] and D-branes [604, 605, 780]. We will simply present the concepts and results that we will use, establishing the notation and conventions.

As advertised, in Chapter 20 we will use many of these results to perform a microscopical computation of the entropies of some four- and five-dimensional BHs (generalizations of the extreme Reissner–Nordström BH). This is one of the main successes of string theory and has stimulated the research in this field from many different points of view.

14.1 Strings

Even though duality relates all the extended objects that appear in string theory, leading to the idea of *p-brane democracy* [898], we know how to quantize only particles and strings, zero- and one-dimensional objects that have one-dimensional worldlines and two-dimensional *worldsheets*, respectively. The latter are the fundamental objects of all string theories, although some of them contain also particles (D0-*branes*). For the moment, there is no fully satisfactory formulation of a *string field theory*, so we must content ourselves with a "first quantization" and with a recipe for how to compute Feynman diagrams.

The action for a single string is the generalization, invariant under general worldsheet and target-space coordinate transformations, of the action for a point-particle Eq. (3.8) proposed by Nambu [714] and Goto [463], and measures essentially the area of the worldsheet swept out by the string when it moves in a d-dimensional ambient (*target*) space with a metric $g_{\mu\nu}$:

$$\boxed{S_{\text{NG}}[X^{\mu}(\xi)] = -T\int_{\Sigma} d^2\xi\ \sqrt{|g_{ij}|}.} \tag{14.1}$$

Here ξ^i, $i = 0, 1$, are coordinates on the two-dimensional worldvolume, $X^{\mu}(\xi)$, $\mu = 0, \ldots, d-1$ are the spacetime coordinates of the string, which give the embedding of the worldvolume into the d-dimensional spacetime, and $|g_{ij}|$ stands for the determinant of the induced metric g_{ij} on the worldvolume (the pull-back of the spacetime metric $g_{\mu\nu}$),

$$g_{ij} \equiv \partial_i X^{\mu}\partial_j X^{\nu} g_{\mu\nu}(X). \tag{14.2}$$

T is the *string tension*, a positive constant with (natural) dimensions of L^{-2} or M^2, which is equivalent to mass per unit length: when the string is wrapped on a compact dimension of radius R, it is seen from the uncompactified dimensions as a particle with mass $2\pi RT$. It is related to *the Regge slope* α' by

$$T = 1/(2\pi\alpha'). \tag{14.3}$$

The Regge slope sets the fundamental length and mass of the theory, the *string length* ℓ_{s} and the *string mass* m_{s}:

$$\ell_{\text{s}} = \sqrt{\alpha'}, \qquad m_{\text{s}} = 1/\sqrt{\alpha'}. \tag{14.4}$$

The Nambu–Goto action is highly non-linear and very difficult to quantize [832], even if we do it in Minkowski spacetime $g_{\mu\nu} = \eta_{\mu\nu}$, avoiding the non-linearities associated with the X-dependence of the spacetime metric. An action that is quadratic in derivatives of the worldsheet fields $X^{\mu}(\xi)$ can be constructed by introducing an auxiliary worldsheet metric

γ_{ij}. The so-called *Polyakov action*[5] reads

$$\boxed{S_{\text{P}}[X^\mu(\xi), \gamma_{ij}(\xi)] = -\frac{T}{2}\int_\Sigma d^2\xi \sqrt{|\gamma|}\, \gamma^{ij}\partial_i X^\mu \partial_j X^\nu g_{\mu\nu}(X).} \tag{14.5}$$

The general variation of this action is given by

$$\begin{aligned}\delta S_{\text{P}} = {} & T\int_\Sigma d^2\xi \sqrt{|\gamma|}\, \delta X^\mu g_{\mu\nu}\big[\nabla^2 X^\nu + \gamma^{ij}\partial_i X^\rho \partial_j X^\sigma \Gamma_{\rho\sigma}{}^\nu(g)\big] \\ & - \frac{T}{2}\int_\Sigma d^2\xi \sqrt{|\gamma|}\, \delta\gamma^{ij}\big[\partial_i X^\mu \partial_j X^\nu - \tfrac{1}{2}\gamma_{ij}\gamma^{kl}\partial_k X^\mu \partial_l X^\nu\big] g_{\mu\nu} \\ & - T\int_{\partial\Sigma} d\Sigma^i \delta X^\mu \partial_i X^\nu g_{\mu\nu}.\end{aligned} \tag{14.6}$$

Since there is no kinetic term for the worldsheet metric, its equation of motion just gives the Rosenfeld energy–momentum tensor of the worldsheet fields X^μ and tells us that it must be zero. This equation is just a primary constraint that we can use to eliminate γ_{ij} in the Polyakov action: on writing it in the form

$$\gamma_{kl} = 2g_{kl}/g_i{}^i, \tag{14.7}$$

and substituting this into the Polyakov action we recover the Nambu–Goto action. Observe that the equation of motion of the worldsheet metric γ_{ij} tells us only that it is proportional to the induced metric g_{ij}, but, in just two worldsheet dimensions, it is impossible to determine the proportionality coefficient $g_i{}^i = \gamma^{ij}g_{ij}$ because this equation of motion (and, hence, the energy–momentum tensor) is (off-shell) traceless. This property is related to an additional symmetry of the Polyakov action for strings: invariance under Weyl rescalings of the worldsheet metric,

$$\gamma_{ij} \to \Omega^2(\xi)\gamma_{ij}. \tag{14.8}$$

This symmetry plays a crucial role in the quantization of the Polyakov action, allowing one to gauge away the worldsheet metric completely and consistently: in two dimensions it is always possible to put the metric in the *conformal gauge* $\gamma_{ij} \propto \eta_{ij}$ by using reparametrizations. Then, with a Weyl rescaling, we can always obtain $\gamma_{ij} = \eta_{ij}$. This symmetry is, however, potentially broken by anomalies that impose further restrictions on the spacetime dimensions, metric, etc., in order to have consistent string theories (see e.g. [832]).

Let us now consider the equation of motion of the X^μs. The boundary term can be nonvanishing if we consider the propagation of *open strings*, with free endpoints and the topology of a line segment (the alternative is to consider *closed strings*, with the topology of a circle). In order to eliminate it, we have to impose special boundary conditions (BCs) for open strings. There are two main possibilities (for each disconnected piece of the boundary $\partial\Sigma^{(n)}$ and for each embedding coordinate X^μ)

[5] It was Polyakov who first quantized it and its supersymmetric version (Eq. 14.23) using the path-integral formalism in [784, 785].

Neumann (N) boundary conditions:

$$n^i \partial_i X^\mu \big|_{\partial\Sigma^{(n)}} = 0, \tag{14.9}$$

where n^i is a unit vector normal to the boundary. These conditions respect target-space Poincaré invariance when $g_{\mu\nu} = \eta_{\mu\nu}$. In the simplest situation, the propagation of a free open string, in which the worldsheet is a strip swept out with length and width parametrized by ξ^0 and ξ^1, the N boundary conditions are

$$\partial_{\underline{1}} X^\mu \big|_{\partial\Sigma^{(n)}} = 0. \tag{14.10}$$

Dirichlet (D) boundary conditions:

$$t^i \partial_i X^\mu \big|_{\partial\Sigma^{(n)}} = 0, \tag{14.11}$$

where t^i is a unit vector tangential to the boundary. For the free open string it is just

$$\partial_{\underline{0}} X^\mu \big|_{\partial\Sigma^{(n)}} = 0. \tag{14.12}$$

This condition is equivalent to requiring that

$$X^\mu|_{\partial\Sigma^{(n)}} = c^{(n)\,\mu}, \tag{14.13}$$

where the $c^{(n)\,\mu}$s are constants, which restrict the endpoints of the open strings (one or both, depending on how many pieces of the boundary we impose this kind of condition on) to moving on $(p+1)$-dimensional hypersurfaces if these conditions are imposed on $(d-p-1)$ embedding coordinates. This explicitly breaks translation invariance in those $(d-p-1)$ coordinates. The hypersurfaces will later be interpreted as the worldvolumes of dynamical p-dimensional objects: D*p-branes* [777].

In Minkowski spacetime and in the conformal gauge the equation of motion of the X^μs is just the two-dimensional wave equation for free bosonic fields $\partial_i \partial^i X^\mu = 0$, whose solutions are the sum of a function $X^\mu_+(\xi^+)$ of $\xi^+ = \xi^1 + \xi^0$ (a *left-moving* wave) and a function $X^\mu_-(\xi^-)$ of $\xi^- = \xi^1 - \xi^0$ (a *right-moving* wave). In open-string worldsheets, left- and right-moving waves are related by the boundary conditions $\partial_+ X_+ = -\partial_- X_-$ for N boundary conditions and $\partial_+ X_+ = \partial_- X_-$ for D boundary conditions.

On the other hand, even if we eliminate the worldsheet metric by gauge transformations, we still have to take into account its equation of motion (the vanishing of the worldsheet energy–momentum tensor).

Apart from being open or closed, strings can be *oriented* or *unoriented*, and in each case only oriented or unoriented worldsheet surfaces are considered.

It is possible to add another Weyl-invariant term to the Polyakov action without any additional background field: a worldsheet Einstein–Hilbert term[6]

$$-\frac{\phi_0}{4\pi} \int d^2\xi \sqrt{|\gamma|}\, R(\gamma). \tag{14.14}$$

[6] In the presence of boundaries, it must be supplemented by a boundary term as in Eq. (4.26).

This term does not change the classical equations of motion because the two-dimensional Einstein–Hilbert Lagrangian density is just the curvature 2-form, which is locally a total derivative: the spin-connection 1-form is $\omega^{ab} = \epsilon^{ab}\omega$ and $R^{ab} = \epsilon^{ab}d\omega$. Actually, this term is the constant ϕ_0 times a topological invariant: the *Euler characteristic* $\chi = 2-2g-b-c$, where g is the *genus* (number of handles) of the two-dimensional worldsheet, b the number of boundaries, and c the number of *crosscaps* (which are present only when the worldsheet is non-orientable). As we are going to see, ϕ_0 is the vacuum expectation value of the *dilaton*, a massless scalar field present in all string theories, and $g = e^{\phi_0}$ can be interpreted as the *string coupling constant* that counts loops in string amplitudes.

14.1.1 Superstrings

The string theories we have studied so far are called *bosonic* because they include only bosonic worldsheet fields and, furthermore, their spectra only contain spacetime bosons, as we will see. To construct a string theory whose spectrum includes fermions, we can generalize the action of a *spinning* point-particle (although the historical order did not follow this logic). The action for a spinning particle contains, in addition to the commuting variables $X^\mu(\xi)$ that describe the position of the particle, anticommuting variables $\psi^\mu(\xi)$ that describe the spin degrees of freedom and were first proposed in [105, 219] and studied and generalized in [70, 86–8, 106, 188, 190, 251]. The simplest action one can write has global *worldline supersymmetry* transformations that relate X^μ and ψ^μ, which form a scalar supermultiplet [188, 190], and the natural generalization, which is invariant under worldline reparametrizations, is also naturally invariant under local worldline supersymmetry transformations. This generalization requires the introduction of auxiliary fields: an Einbein $e(\xi)$, as in the bosonic case ($e^2 = \gamma$), and a gravitino χ, which form a one-dimensional supergravity multiplet that has no dynamics. The action for the massless case is

$$\boxed{S = -\tfrac{1}{2}\int d\xi\, e\left[e^{-2}\dot{X}^\mu \dot{X}_\mu + e^{-1}\psi^\mu \dot{\psi}_\mu - e^{-2}\chi\psi^\mu \dot{X}_\mu\right].} \tag{14.15}$$

and the supersymmetry transformations that leave it invariant are

$$\begin{aligned} \delta_\epsilon X^\mu &= \epsilon\psi^\mu, & \delta_\epsilon e &= \epsilon\chi, \\ \delta_\epsilon \psi^\mu &= -\epsilon(\dot{X}^\mu - \tfrac{1}{2}\chi\psi^\mu)e^{-1}, & \delta_\epsilon \chi &= 2\dot{\epsilon}. \end{aligned} \tag{14.16}$$

Invariance under worldline reparametrizations and supersymmetry transformations leads to constraints (gauge identities) that are necessary for consistency. Furthermore, it can be shown that the quantization of this model leads to the Dirac equation for spin-$\frac{1}{2}$ particles [86, 106, 190] (for more recent references, see [424, 529]) and this action can be used to obtain path-integral representations of propagators and Feynman diagrams (see e.g. [537] and references therein) for spin-$\frac{1}{2}$ particles. It is also remarkable that, when it is coupled to gravity [87], the action leads to the Dirac equation in curved spacetime as given in [187] and the equations of motion for a pole–dipole singularity derived by Papapetrou in [755] using the method of Einstein, Infeld, and Hoffmann [366].[7]

The action for the spinning particle can be generalized to include other couplings to vector fields and odd-rank forms preserving worldline local supersymmetry, which can be made manifest by using *superworldsheet* coordinates and geometry (see, for instance, [536] and references therein). In particular, the coupling to a 3-form can be understood as the coupling to a completely antisymmetric torsion field, and leads to the Dirac equation of the CSK theory and a modification of the Papapetrou equations [816].

On the other hand, the spectrum also contains spin-0 and-1 particles and, thus, in spite of the worldline supersymmetry, there is no spacetime supersymmetry.

The action for a massive Dirac particle can be obtained by adding the term[8]

$$-\tfrac{1}{2}\int d\xi\, e[m^2 + e^{-1}\psi_5\dot{\psi}_5 + me^{-1}\chi\psi_5], \tag{14.17}$$

where ψ_5 is another anticommuting variable that transforms as follows:

$$\delta_\epsilon\psi_5 = m\epsilon + \frac{\epsilon}{me}\psi_5(\dot{\psi}_5 - \tfrac{1}{2}\chi). \tag{14.18}$$

By analogy, the worldsheet action for a fermionic string that can describe spacetime fermions (but, also, as we will see, spacetime bosons, depending on boundary conditions) must incorporate two-dimensional spinors. Worldsheet supersymmetry can be used as a guiding principle to introduce worldsheet fermions. Then, the theory that generalizes the Polyakov action in Minkowski spacetime is the theory of d two-dimensional scalar superfields (X^μ, ψ^μ) coupled to the two-dimensional supergravity multiplet $(e^a{}_i, \chi_i)$ with the action [189, 316]

$$\boxed{\begin{aligned} S = -\frac{T}{2}\int_\Sigma d^2\xi\, e\big[&\gamma^{ij}\partial_i X^\mu\partial_j X_\mu - i\bar{\psi}^\mu \not{\mathcal{D}}\psi_\mu \\ &+ 2\bar{\chi}_i\rho^j\rho^i\psi^\mu\partial_j X_\mu + \tfrac{1}{2}(\bar{\chi}_i\rho^j\rho^i\chi_j)(\bar{\psi}^\mu\psi_\mu)\big].\end{aligned}} \tag{14.19}$$

Both ψ^μ and χ_i are real two-dimensional spinors with hidden spinorial indices and $\rho^i = \rho^a e_a{}^i$, where ρ^a are the two-dimensional gamma matrices (see Appendix B.1.7). This action is invariant under worldsheet reparametrizations and local Lorentz transformations, as usual in supergravity (see Chapter 5), and also under global Poincaré transformations of the embedding coordinates $X^{\mu\prime} = \Lambda^\mu{}_\nu X^\nu + a^\mu$. It is also invariant under local worldsheet supersymmetry transformations:

$$\begin{aligned} \delta_\epsilon X^\mu &= \bar{\epsilon}\psi^\mu, & \delta_\epsilon e^a{}_i &= -2i\bar{\epsilon}\rho^a\chi_i, \\ \delta_\epsilon\psi^\mu &= i(\partial_i X^\mu + \tfrac{1}{4}\bar{\chi}_i\psi^\mu)\rho^i\epsilon, & \delta_\epsilon\chi_i &= \tilde{\mathcal{D}}_i\epsilon, \end{aligned} \tag{14.20}$$

[7] This method was used in [396] to re-obtain the equations of motion of a point-particle and in [490] to recover the equations of motion of the Nambu–Goto string.

[8] This action can be derived by dimensional reduction of the action of the massless spinning particle.

where

$$\tilde{D}_i\epsilon = (\partial_i + \tilde{\omega}_i\rho_3)\epsilon, \qquad \tilde{\omega}_i = \omega_i + i\bar{\chi}_i\rho_3\rho^j\chi_j. \tag{14.21}$$

Furthermore, and most importantly, it is also invariant under super-Weyl transformations:

$$\begin{aligned} X^{\mu\,\prime} &= X^\mu, & e^{a\,\prime}{}_i &= \Omega e^a{}_i, \\ \psi^{\mu\,\prime} &= \Omega^{-\frac{1}{2}}\psi^\mu & \chi_i' &= \Omega^{\frac{1}{2}}\chi_i + i\rho_i\lambda, \end{aligned} \tag{14.22}$$

where λ is any real spinor. This symmetry allows the complete decoupling of the Zweibein e^a and the worldsheet gravitino χ_i, whose elimination gives the so-called *Ramond–Neveu–Schwarz (RNS)* model [722, 796]

$$\boxed{S = -\frac{T}{2}\int_\Sigma d^2\xi[\eta^{ij}\partial_i X^\mu\partial_j X_\mu - i\bar{\psi}^\mu\partial\!\!\!/\psi_\mu].} \tag{14.23}$$

This model is only globally supersymmetric and has to be supplemented with the equations of motion of the worldsheet metric and gravitino (even after we have eliminated them).

One can consider open, closed, oriented and unoriented superstrings. In the open-string case, at the endpoints $\xi^1 = 0, 2\pi\ell$ N and D boundary conditions can be chosen for the X^μs, and for the ψ^μs one can choose between

Ramond (R) boundary conditions,

$$\psi_+^\mu(\xi^1 = 0) = \psi_-^\mu(\xi^1 = 0), \qquad \psi_+^\mu(\xi^1 = 2\pi\ell) = \psi_-^\mu(\xi^1 = 2\pi\ell), \tag{14.24}$$

and Neveu–Schwarz (NS) boundary conditions,

$$\psi_+^\mu(\xi^1 = 0) = \psi_-^\mu(\xi^1 = 0), \qquad \psi_+^\mu(\xi^1 = 2\pi\ell) = -\psi_-^\mu(\xi^1 = 2\pi\ell), \tag{14.25}$$

where $\psi_\pm^\mu$ are the left- and right-moving components of ψ^μ.

For closed superstrings, $\xi \sim \xi^1 + 2\pi\ell$ and we can have, for each component ψ_+^μ and ψ_-^μ, independently,

Ramond (R) (periodic) boundary conditions,

$$\psi_\pm^\mu(\xi^1 = 0) = \psi_\pm^\mu(\xi^1 = 2\pi\ell), \tag{14.26}$$

and Neveu–Schwarz (NS) (antiperiodic) boundary conditions,

$$\psi_\pm^\mu(\xi^1 = 0) = -\psi_\pm^\mu(\xi^1 = 2\pi\ell). \tag{14.27}$$

So we can have RR, RNS, NSR, and NSNS boundary conditions.

Just as in the spinning-particle case, worldsheet supersymmetry leads to spacetime supersymmetry of the quantized theory only under quite restrictive conditions [457] and, in any case, spacetime supersymmetry is never explicit. In the alternative *Green–Schwarz (GS)* formulation [471] spacetime supersymmetry is explicit from the outset.

14.1.2 Green–Schwarz Actions

Green–Schwarz-type actions can be constructed for particles, strings, and other extended objects, as we will see. The simplest example describes a massless particle moving in flat (target) superspace with supercoordinates $Z^M(\xi) = (X^\mu(\xi), \theta^{\alpha\, I}(\xi))$ where $I = 1, \ldots, N$ numbers the supersymmetries, α is a spacetime spinorial index, and the θ^I are anticommuting spacetime spinors (but worldsheet scalars) [191]:

$$\boxed{S = -\frac{p}{2}\int d\xi\, e^{-1}(\dot{X}^\mu \delta^a{}_\mu - i\bar{\theta}^I \gamma^a \dot{\theta}^I)\eta_{ab}(\dot{X}^\nu \delta^b{}_\nu - i\bar{\theta}^I \gamma^b \dot{\theta}^I).} \tag{14.28}$$

This action is invariant under worldline reparametrizations, under target-space Poincaré transformations, and also under the standard global superspace transformations

$$\delta_\epsilon \theta^I = \epsilon^I, \qquad \delta_\epsilon X^\mu = i\bar{\epsilon}^I \gamma^a \theta^I \delta_a{}^\mu. \tag{14.29}$$

In principle, there is no worldline supersymmetry, a very desirable property. A necessary condition for having linearly realized worldline supersymmetry is that the numbers of on-shell bosonic and fermionic degrees of freedom should be equal.[9] To find the numbers of on-shell degrees of freedom, it is necessary to know all the local symmetries of the action. The above action turns out to have worldline-reparametrization invariance, which can be used to gauge away one of the X^μs, and a new local symmetry generated by a fermionic infinitesimal parameter κ (*κ-symmetry*), which halves the number of fermionic degrees of freedom [60, 61, 854], which has already been halved by the Dirac equation. Under κ-symmetry

$$\delta_\kappa \theta^I = -i\Pi_\mu \gamma^a \delta_a{}^\mu \kappa^I, \quad \delta_\kappa X^\mu = i\bar{\theta}^I \gamma^a \delta_a{}^\mu \delta_\kappa \theta^I, \quad \delta_\kappa e = 4e\dot{\bar{\theta}}^I \kappa^I, \tag{14.30}$$

where $\Pi_\mu = \delta S/\delta \dot{X}^\mu$ is the momentum conjugate to X^μ.

Taking into account this new symmetry, if we denote by M the number of real components of the minimal spinor in the spacetime dimension d considered, then, the necessary condition for having worldsheet supersymmetry reads

$$NM = 4(d-1), \tag{14.31}$$

and, taking into account the values of M in Table B.1, it can be satisfied for $d = 2, 3, 4, 5,$ and 9 with $N = 4, 4, 3, 2,$ and 2, respectively. Thus, the dimensions in which these actions can be consistent are restricted.

The condition can be generalized to objects with p extended dimensions. Using worldvolume reparametrizations, we can always gauge away $p+1$ X^μs. The condition of worldvolume supersymmetry becomes now

$$NM = 4(d-p-1), \tag{14.32}$$

[9] In just one dimension, talking about degrees of freedom does not make much sense. However, the same reasoning carries over to higher-dimensional cases, which allows a classification of all the possible supersymmetric extended objects [13].

Table 14.1. A brane scan taking into account only scalar supermultiplets. N is given by the quotient between the numbers given and M.

		$p=-1$	$p=0$	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
d	M	d	$(d-1)$	$(d-2)$	$(d-3)$	$(d-4)$	$(d-5)$	$(d-6)$
2	1	**8**	**4**					
3	2	**12**	**8**	**4**				
4	4	**16**	**12**	**8**	**4**			
5	8		**16**		**8**			
6	8			**16**		**8**		
7	16				**16**			
8	16	**32**				**16**		
9	16		**32**				**16**	
10	16			**32**				**16**
11	32				**32**			

and it can be solved in the cases represented in Table 14.1 [13]. There are five series of solutions. Four of them (with $NM = 32$, 16, 8, and 4) are associated with the four division algebras $\mathbb{O}$, $\mathbb{Q}$, $\mathbb{C}$, and $\mathbb{R}$, respectively. These series correspond to objects related by *double dimensional reduction*, as we will see.

There is another way to understand this result: if there is linearly realized worldvolume supersymmetry, the worldvolume fields must fit in $(p+1)$-dimensional scalar supermultiplets. Each solution in the table corresponds to a scalar multiplet. There is agreement with the fact that scalar multiplets exist only in up to six dimensions.

For us, it is interesting that spacetime supersymmetric string actions could in principle be constructed in $d = 3, 4, 6$, and 10, provided that the action has κ-symmetry. Green and Schwarz showed in [472] that the ten-dimensional action previously constructed by them in [471] has κ-symmetry. The Green–Schwarz (GS) action is not just a straightforward generalization of the superparticle action Eq. (14.28), because the kinetic term would not be κ-symmetric by itself. The key to κ-symmetry is the addition of a (super-)Wess–Zumino (WZ) term: the integral of a 2-form Ω_2 such that $\Omega_3 = d\Omega_2$ is (target) Poincaré- and supersymmetry invariant [528]:[10]

$$\Omega_2 = -idX^{\mu}\delta^{a}{}_{\mu} \wedge (\bar{\theta}^1\Gamma_a\theta^1 - \bar{\theta}^2\Gamma_a d\theta^2) + (\bar{\theta}^1\Gamma_a d\theta^1) \wedge (\bar{\theta}^2\Gamma^a d\theta^2). \tag{14.33}$$

The GS action is then given by

$$\boxed{S = -\frac{T}{2}\int_{\Sigma} d^2\xi\sqrt{|\gamma|}\gamma^{ij}(\partial_i X^{\mu}\delta^{a}{}_{\mu} - i\bar{\theta}^I\gamma^a\partial_i\theta^I)(\partial_j X^{\nu}\delta^{b}{}_{\nu} - i\bar{\theta}^J\gamma^b\partial_j\theta^J)\eta_{ab} + T\int_{\Sigma}\Omega_2.} \tag{14.34}$$

[10] According to Table 14.1, this is an $N = 2$ theory, with two minimal (Majorana–Weyl spinors with 16 real components) ten-dimensional spinors, θ^1 and θ^2, with equal or opposite chiralities: type-IIB and type-IIA strings, respectively. These theories can also be described with the RNS theory, but only after quantization.

This action, whose covariant quantization poses many problems, has very interesting features. First, it can be generalized to the other string theories with $N = 2$ spacetime supersymmetry [137, 481] and other higher-dimensional objects of Table 14.1, always with a WZ term (see also [78, 79]). This was done for the super 3-brane in $d = 6$ in [574] and for the super 2-brane (*supermembrane*, now *M2-brane*) of $d = 11$ in [138, 139]. The M2-brane is related to the ten-dimensional string by double dimensional reduction [14, 335] (the last two members of the $NM = 32$ series).

These generalizations include the coupling of the supersymmetric extended objects to supergravity fields: since GS actions are manifestly spacetime-supersymmetric, the coupling to gravity implies the coupling to all the background fields of a supergravity theory.[11] The fact that it can and must be coupled to supergravity explains in part the necessity of the WZ term: supergravity theories contain potentials that are $(p+1)$-forms in superspace that can naturally couple to p-dimensional objects through a WZ term (the integral of the pull-back of the $(p+1)$-form over the $(p+1)$-dimensional worldvolume), just like particles couple to the Maxwell 1-form Eq. (8.53) (see also Eq. (15.4)). The expansion of the $(p+1)$-form fields in components contains terms that do not vanish in flat spacetime and Ω_2 above is one such term. The full WZ term for the superstring coupled to supergravity also contains the purely bosonic term Eq. (15.4).

κ-symmetry imposes constraints on the supergravity fields [137–9, 335, 481, 961]. In particular [138, 139], the action for the 11-dimensional supermembrane coupled to the fields of 11-dimensional supergravity can be κ-invariant only if certain constraints are solved by the equations of motion of that theory. These constraints coincide with the superspace constraints of $d = 11$ supergravity [262]. Thus worldvolume κ-invariance implies the equations of motion of the spacetime supergravity fields, a highly non-trivial fact. Something similar happens in string theory coupled to background fields: by requiring invariance under worldsheet Weyl transformations in the quantum theory one obtains the equations of motion of the spacetime fields (see Section 15.1).

Although there is no clear motivation at this point, we could also include other supermultiplets in the super-p-brane actions [339]. The new supersymmetric extended objects include the *Dp-branes* we will study in more detail later [16, 17, 81, 140, 223, 569, 899].

14.2 Quantum theories of strings

In this section we are going to overview the quantization in Minkowski spacetime of the bosonic and fermionic string actions that we have introduced in the previous section. We will focus on the definition of quantum string theory (in particular on string interactions) and on the results: the critical dimensions, mode expansions, and massless spectra of the simplest consistent string theories.

14.2.1 *Quantization of free-bosonic-string theories*

Free strings can translate, rotate, and vibrate. The various allowed vibrational modes are seen as different particle states in spacetime. These particle states must fit into Poincaré

[11] Only recently has it been learned how to couple superstrings to RR backgrounds (to be defined later) [695, 694, 696].

multiplets characterized by mass and spin (or helicity). Failure to do so implies breaking of spacetime Poincaré invariance.

The simplest way to quantize the Polyakov or the superstring action, Eqs. (14.5) and (14.1.1), and obtain their spectra is to use the physical light-cone gauge in which all the gauge invariances of the action are used to eliminate unphysical degrees of freedom. This was first done for the bosonic string in [458], where it was found that Poincaré invariance is recovered only in the critical dimension $d = 26$.

Here we will follow the careful treatment of [779] for the bosonic string, where it is shown how, using worldsheet reparametrizations and Weyl rescalings, one can always set

$$X^+ \equiv \frac{1}{\sqrt{2}}(X^0 + X^1) = \xi^0, \qquad \partial_{\underline{1}}\gamma_{\underline{11}} = 0, \qquad \det(\gamma_{ij}) = -1, \tag{14.35}$$

which allows the elimination of X^+, X^-, and γ_{ij} and leads to the Hamiltonian

$$H = -\frac{c}{2}\int_0^{2\pi\ell} d\xi^1\big[T^{-1}\Pi^i\Pi^i + T\partial_{\underline{1}}X^i\partial_{\underline{1}}X^i\big], \qquad c = 2\pi\ell T/p^+, \tag{14.36}$$

where $\xi^1 \in [0, 2\pi\ell]$, the Π^i are the momenta conjugate to the X^i, and p^+ is the momentum conjugate to x^-, a cyclic variable, so p^+ is a constant of motion. The equations of motion for the X^i that follow from this Lagrangian are

$$(\partial_{\underline{0}}^2 - c^2\partial_{\underline{1}}^2)X^i = 0 \quad \Rightarrow X^i = X^i_+ + X^i_-, \qquad X^i_\pm = X^i_\pm(\xi^1 \pm c\xi^0). \tag{14.37}$$

The left- X^i_+ and right-moving X^i_- components are related by boundary conditions in the open-string case and we consider only X^i. The boundary conditions play their role in the mode expansions of the X^is: for open strings with N boundary conditions

$$X^i = x^i + \frac{p^i}{p^+}\xi^0 - i\sqrt{2\alpha'}\sum_{n\neq 0}\frac{\alpha^i_n}{n}e^{\frac{icn\xi^0}{2\ell}}\cos\left(\frac{n\xi^1}{2\ell}\right). \tag{14.38}$$

If X has D boundary conditions at both ends, $X(\xi^1 = 0) = x_1$ and $X(\xi^1 = 2\pi\ell) = x_2$, then

$$X = x_1 - \frac{x_2 - x_1}{2\pi\ell}\xi^1 + \sqrt{2\alpha'}\sum_{n\neq 0}\frac{\alpha_n}{n}e^{\frac{icn\xi^0}{2\ell}}\sin\left(\frac{n\xi^1}{2\ell}\right). \tag{14.39}$$

For closed strings

$$X^i = x^i + \frac{p^i}{p^+}\xi^0 - i\sqrt{\frac{\alpha'}{2}}\sum_{n\neq 0}\left[\frac{\alpha^i_n}{n}e^{\frac{in}{\ell}(\xi^1 + c\xi^0)} + \frac{\tilde{\alpha}^i_n}{n}e^{-\frac{in}{\ell}(\xi^1 - c\xi^0)}\right]. \tag{14.40}$$

Reality implies in all cases ${\alpha^i_n}^\dagger = \alpha^i_{-n}$ and ${\tilde{\alpha}^i_n}^\dagger = \tilde{\alpha}^i_{-n}$. On substituting the mode expansions into the equal time commutators (using $\Pi^i = p^+\partial_{\underline{0}}X^i/(2\pi\ell)$),

$$[x^-, p^+] = i, \qquad [X^i(\xi^1), \Pi^j(\xi^{1\prime})] = i\delta^{ij}\delta(\xi^1 - \xi^{1\prime}), \tag{14.41}$$

we find the commutation relations

$$[x^i, p^j] = i\delta^{ij}, \qquad [\alpha^i_m, \alpha^j_n] = m\delta^{ij}\delta_{m,-n}, \qquad [\tilde{\alpha}^i_m, \tilde{\alpha}^j_n] = m\delta^{ij}\delta_{m,-n}. \tag{14.42}$$

The vacuum is defined to be annihilated by all the oscillators α_n^i and $\tilde{\alpha}_n^i$ with $n > 0$ and states are created by acting on it with creation operators α_{-n}^i and $\tilde{\alpha}_{-n}^i$, with $n > 0$, on the momentum eigenstates $|0, k\rangle$

$$p^i |0, k\rangle = k^i |0, k\rangle, \qquad p^+ |0, k\rangle = k^+ |0, k\rangle. \tag{14.43}$$

Relative to this vacuum, the mass operator $M^2 = -2p^+ H - p^i p^i$ takes the form, for open strings with only N boundary conditions,

$$M^2 = \frac{1}{\alpha'}(N + A). \tag{14.44}$$

For open strings with DD boundary conditions in one coordinate

$$M^2 = \left(\frac{x_2 - x_1}{2\pi\alpha'}\right)^2 + \frac{1}{\alpha'}(N + A), \tag{14.45}$$

and for closed strings

$$M^2 = \frac{2}{\alpha'}(N + \tilde{N} + A + \tilde{A}). \tag{14.46}$$

In all these cases

$$N = \sum_{n>0} n\alpha_{-n}^i \alpha_n^i, \qquad \tilde{N} = \sum_{n>0} n\tilde{\alpha}_{-n}^i \tilde{\alpha}_n^i, \tag{14.47}$$

are the *level* operators that take only positive integer values and A and $\tilde{A}$ are constants that arise in the normal ordering of the Hamiltonian and take the value $A = \tilde{A} = (2 - d)/24$.

In the closed-string case there is still one constraint that has not been eliminated, which is associated with the ξ^1 shift symmetry:

$$N = \tilde{N}. \tag{14.48}$$

Let us now consider the lightest states of these three theories. The lightest states of the open string with N boundary conditions are the $|0, k\rangle$, whose mass is, according to Eq. (14.44), $M^2 = (2 - d)/(24\alpha')$, which is negative for $d > 2$, corresponding to a spacetime scalar tachyon and indicating the instability of the open-bosonic-string vacuum. The next lightest states are obtained by acting with the α_{-1}^i operators on $|0, k\rangle$ and have masses $M^2 = (26 - d)/(24\alpha')$. They fill a vector representation of $SO(d - 2)$, just like a d-dimensional massless spacetime vector particle. Poincaré invariance then requires the mass of these states to be zero and $d = 26$ and the spectrum contains a scalar tachyon, a massless vector, and massive and higher-spin states.

The lightest state of the DD open string is again $|0, k\rangle$, with $(d = 26)$ $M^2 = [(x_2 - x_1)/(2\pi\alpha')]^2 - 1/\alpha'$, whose value and sign depend on the distance between the hypersurfaces $x = x_1, x_2$ to which the string endpoints are attached. When $x_2 = x_1$ there are massless states $\alpha_{-1}^i|0, k\rangle$, $i \neq x$ and $\alpha_{-1}^x|0, k\rangle$ (x being the DD direction), a massless vector, and a scalar in $d - 1$ dimensions. When $x_2 \neq x_1$ we have the $d - 2$ states of a massive $(d - 1)$-dimensional vector. The scalar plays the role of a Higgs scalar.

Observe that the α_ns would represent oscillations of the strings in directions perpendicular to the $x = x_1, x_2$ hypersurfaces, which is not possible. Hence they must represent oscillations of the $x = x_1, x_2$ hypersurfaces which, thus, are the 25-dimensional worldvolumes of dynamical 24-dimensional objects: D*24-branes*. These objects can be understood as new physical string states that must be non-perturbative since they do not appear in the perturbative spectrum. Indeed, as we will see, their tension is $\sim g^{-1}$, where g is the string coupling constant that we will define later.

In the presence of two D24-branes parallel at $x = x_1, x_2$, open strings can have both ends attached to either one of them or have each end attached to each of them.[12] Since we are considering oriented strings, we must distinguish between the strings that connect 1 to 2 and those that connect 2 to 1. There are, then, four *sectors* and, according to the preceding paragraphs, two of them include two massless vectors and scalars and two massive vectors, and scalars, all of them living in the directions parallel to the D-branes. When the two 24-branes coincide, we have two extra massless vector and scalar fields. The four massless vector fields turn out to be U(2) gauge fields. In general, when n D-branes coincide, there are massless U(n) vector fields labeled by two indices $I, J = 1, 2$ that indicate to which D-brane each string endpoint is attached and the gauge symmetry is spontaneously broken to smaller groups when some of the D-branes become separated: the Higgs scalars give mass to the gauge fields.

It is possible to introduce labels (*Chan–Paton factors*) for open-string endpoints even if all coordinates have N boundary conditions, and the theory will have U(n) gauge vector fields. In this case we can think that the spacetime is filled with n D25-branes.

Symmetry enhancements at particular values of moduli are some of the most interesting features of string theories.

Let us consider now closed strings. The lightest states obeying the constraint (14.48) after the tachyon $|0, 0; k\rangle$ are of the form $\alpha^i_{-1}\tilde{\alpha}^j_{-1}|0, 0; k\rangle$ and, just in $d = 26$, they fit into Poincaré representations: the part symmetric and traceless in ij corresponds to a massless graviton, the trace part to a massless scalar, the *dilaton*, and the antisymmetric part to a massless 2-form field: the Kalb–Ramond (KR) field.

After introducing interactions and following the reasoning of Chapter 3, closed-string theories must contain gravity, which in a certain limit must coincide with Einstein's GR.

Finally, let us consider unoriented strings. They can be obtained from oriented strings by taking the quotient by the *worldsheet parity operator* $\Omega\colon \xi^1 \to 2\pi\ell - \xi^1$, which interchanges right- and left-moving sectors,

$$\Omega X^\mu_\pm = X^\mu_\mp, \tag{14.49}$$

and is an involution, $\Omega^2 = 1$, and generates a $\mathbb{Z}_2$ symmetry group. Taking the quotient means keeping only Ω-invariant states. On level-N states

$$\Omega|N; k\rangle = (-1)N|N; k\rangle, \qquad \Omega|N, \tilde{N}; k\rangle = |\tilde{N}, N; k\rangle. \tag{14.50}$$

Thus, the KR tensor is removed from the closed-string spectrum and the massless vector is removed from the open-string spectrum unless the endpoints have Chan–Paton factors

[12] This system is not stable in bosonic-string theory, as the presence of tachyons indicates, but it will be in type-II superstring theories.

attached. In that case, only the U(n) vectors V_μ^{IJ} antisymmetric in IJ survive, i.e. the SO(n) or Sp(n) subgroups.

In general, the quotient of a theory by a discrete symmetry of the theory (such as Ω here) is called an *orbifold* by analogy with the spacetime orbifolds discussed in Section 11.6. If in a string theory worldsheet parity is combined with a discrete symmetry of the spacetime, then the quotient is called an *orientifold* [147, 148, 283, 464, 541, 542, 768, 827],[13] but since orbifolds and orientifolds are often related by dualities, it is customary to call all of them orientifolds. In this language, closed-unoriented-string theory is an orientifold of closed-oriented-string theory. The hypersurfaces left invariant by the discrete spacetime symmetries are called *orientifold planes* and, although they are similar in other respects to Dp-branes, they are not dynamical objects in the sense that they are attached to the fixed points of the spacetime orbifold and cannot translate or oscillate. However, there are dynamical fields on them. In the above case, there is no spacetime symmetry, the whole spacetime is invariant and we can say that there is an orientifold plane of 25 spatial dimensions (O25-plane) that fills the entire spacetime, as a D25-brane does.

There is a crucial difference between orientifolds of point-particle theories and closed-string theories: in the latter one must include, for consistency, *twisted sectors*: strings that are closed up to a symmetry operation associated with the orientifold. In general, the inclusion of these twisted sectors makes the string theory non-singular at the orbifold fixed points, as distinct from point-particle theories. Twisted sectors also appear in other contexts: for instance, the *winding modes* that appear in toroidal compactifications (see Section 14.3) can be seen as strings in $\mathbb{R}^n$ closed up to an element of Γ^n, where Γ^n is the discrete group used to define the torus: $\mathrm{T}^n \equiv \mathbb{R}^n / \Gamma^n$ ($\Gamma^n = \mathbb{Z}^n$ in the simplest case).

In bosonic-string theory we can add D-branes and O-planes more or less at will, because the theory is already inconsistent due to the tachyon. In the consistent superstring theories, D-branes and O-planes have to be introduced, paying attention to anomaly and tadpole cancelations. These conditions are, in turn, related to the possibility of solving the equations of motion of the effective string theory for a background that contains those objects. In particular one has to be able to solve the harmonic equation for $(p+1)$-form potentials in compact spaces, which is possible only if the total charge associated with the potentials vanishes. (Super-)Dp-branes and Op-planes of superstring theories are charged with respect to (so-called) RR $(p+1)$-form potentials, which we will define later, and a consistent background will be one in which the sum of those charges vanishes. A very interesting example, as we will see, is the construction of the type-I SO(32) superstring theory by adding D9-branes and O9-planes to the type-IIB superstring theory [827].

There is one last consideration we must make: as we are going to see, open-string interactions can produce closed strings. Thus, open strings are not fully consistent by themselves and have to be combined with a closed-string sector with the same orientability. The fields of the massless spectra of the resulting theories (without D-branes) are represented in Table 14.2. The consistency of the interacting theory also requires the addition of twisted sectors in orbifolds and orientifolds.

[13] For a review on orientifolds, see, for instance, [280] and also [44].

Table 14.2. In this table we describe the massless fields of the various bosonic-string theories.

Theory	Massless fields
Closed oriented	$g_{\mu\nu},\ B_{\mu\nu},\ \phi$
Closed unoriented	$g_{\mu\nu},\ \phi$
Open (and closed) oriented	$g_{\mu\nu},\ B_{\mu\nu},\ V_\mu^{IJ},\ \phi$
Open (and closed) unoriented	$g_{\mu\nu},\ V_\mu^{[IJ]},\ \phi$

14.2.2 Quantization of free-fermionic-string theories

Again, the simplest way to arrive to the physical spectrum is to go to the light-cone gauge,

$$X^+ = \xi^0, \qquad e^0{}_{\underline{1}} = \partial_{\underline{1}} e^1{}_{\underline{1}} = \psi^+ = \rho^i \chi_i = 0, \qquad \det(e^a{}_i) = +1, \tag{14.51}$$

which allows the elimination of $e^a{}_i$, χ_i, $X^\pm$, and $\psi^\pm$ and leads to the Hamiltonian

$$H = -\frac{c}{2}\int_0^{2\pi\ell} d\xi^1 \left[T^{-1}\Pi_X^i \Pi_X^i + T\partial_{\underline{1}} X^i \partial_{\underline{1}} X^i + 2\Pi_\psi^i \rho_3 \partial_{\underline{1}} \psi^i\right], \tag{14.52}$$

and to the equations of motion (14.37) for the X^i and, for the upper (ψ_+^i) and lower (ψ_-^i) components of the spinors ψ^i,

$$(\partial_{\underline{0}} \mp \partial_{\underline{1}})\psi_\pm^i = 0, \tag{14.53}$$

they are left- and right-moving, respectively.

The quantization proceeds essentially along the same lines as in the bosonic string, paying attention to the second-class constraints that this theory has. Here we describe just the general structure of the consistent superstring theories, leaving all details aside.

Superstring theories are Poincaré-invariant only in the critical dimension $d = 10$. It is necessary to introduce the *worldvolume fermion number F*, defined modulo 2. The R and NS sectors are separated into $\mathrm{R}_\pm$ and $\mathrm{NS}_\pm$ subsectors with respect to the operator $e^{i\pi F}$. Then, consistency and the absence of tachyons require the combination of these subsectors (*GSO projection* [457]) in very precise ways.

Closed strings There are several possibilities.

Type-IIB$_+$ superstring $\mathrm{R}_+\mathrm{R}_+ \oplus \mathrm{R}_+\mathrm{NS}_+ \oplus \mathrm{NS}_+\mathrm{R}_+ \oplus \mathrm{NS}_+\mathrm{NS}_+$, whose massless spectrum corresponds to an $N = 2B_+$, $d = 10$ supergravity multiplet with self-dual 4-form potential and chiral fermions.

Type-IIB$_-$ superstring $\mathrm{R}_-\mathrm{R}_- \oplus \mathrm{R}_-\mathrm{NS}_+ \oplus \mathrm{NS}_+\mathrm{R}_- \oplus \mathrm{NS}_+\mathrm{NS}_+$, whose massless spectrum corresponds to an $N = 2B_-$, $d = 10$ supergravity multiplet with anti-self-dual 4-form potential and chiral fermions with the opposite chirality to the previous case.

Type-IIA$_{+-}$ superstring $\mathrm{R}_+\mathrm{R}_- \oplus \mathrm{R}_+\mathrm{NS}_+ \oplus \mathrm{NS}_+\mathrm{R}_- \oplus \mathrm{NS}_+\mathrm{NS}_+$, whose massless spectrum corresponds to an $N = 2A_{+-}$, $d = 10$ supergravity multiplet.

Table 14.3. In this table we describe the massless fields of the various ten-dimensional superstring (supergravity) theories.

Theory	NSNS bosonic	RR bosonic	Chiral fermionic	Non-chiral fermionic	Vector supermultiplets
Type IIA	$\hat{g}_{\hat{\mu}\hat{\nu}}, \hat{B}_{\hat{\mu}\hat{\nu}}, \hat{\phi}$	$\hat{C}^{(1)}{}_{\hat{\mu}}, \hat{C}^{(3)}{}_{\hat{\mu}\hat{\nu}\hat{\rho}}$		$\hat{\psi}_{\hat{\mu}}, \hat{\lambda}$	
Type IIB	$\hat{\jmath}_{\hat{\mu}\hat{\nu}}, \hat{\mathcal{B}}_{\hat{\mu}\hat{\nu}}, \hat{\varphi}$	$\hat{C}^{(0)}, \hat{C}^{(2)}{}_{\hat{\mu}\hat{\nu}}, \hat{C}^{(4\pm)}{}_{\hat{\mu}_1\cdots\hat{\mu}_4}$	$\hat{\zeta}^{i\,(\mp)}_{\hat{\mu}}, \hat{\chi}^{i\,(\pm)}$		
Type I	$\hat{\jmath}_{\hat{\mu}\hat{\nu}}, \hat{\varphi}$	$\hat{C}^{(2)}{}_{\hat{\mu}\hat{\nu}}$	$\hat{\zeta}^{(\pm)}_{\hat{\mu}}, \hat{\chi}^{(\mp)}$		$(V^I{}_{\mu}, \eta^I)$
Heterotic	$\hat{g}_{\hat{\mu}\hat{\nu}}, \hat{B}_{\hat{\mu}\hat{\nu}}, \hat{\phi}$		$\hat{\psi}^{(\pm)}_{\hat{\mu}}, \hat{\lambda}^{(\mp)}$		$(V^I{}_{\mu}, \eta^I)$

Type-IIA$_{-+}$ superstring $\mathrm{R}_-\mathrm{R}_+ \oplus \mathrm{R}_-\mathrm{NS}_+ \oplus \mathrm{NS}_+\mathrm{R}_+ \oplus \mathrm{NS}_+\mathrm{NS}_+$, whose massless spectrum corresponds to an $N = 2A_{-+}, d = 10$ supergravity multiplet. The sign of the Chern–Simons term of the action is different from that of the $N = 2A_{+-}$ theory.

There is a fifth possibility that gives the so-called type-0A and -0B strings, which are bosonic and have a tachyon but have sometimes been considered.

Heterotic strings They are constructed by combining the right-moving fields of the closed type-II superstring with the left-moving fields of the closed bosonic string. The 16 extra spacetime dimensions of the bosonic string must be compactified, which gives rise to gauge symmetry. A generic toroidal compactification gives the gauge group[14] $\mathrm{U}(1)^{16}$, but, as explained in Section 14.3, for special values of the radii of the circles and of the angles between them, there the gauge group can be bigger and non-Abelian. In particular, one can show that the anomaly-free groups are SO(32) and $\mathrm{E}_8 \times \mathrm{E}_8$. The massless modes are those of $N = 1_{\pm}, d = 10$ supergravity coupled to vector supermultiplets with those gauge groups. The couplings between the massless fields are, however, different from those of the type-I SO(32) theory.

Open superstrings They result from the combination of two subsectors: $\mathrm{R}_+ \oplus \mathrm{NS}_+$ or $\mathrm{R}_- \oplus \mathrm{NS}_+$. In both cases, the massless spectrum corresponds to an $N = 1, d = 10$ vector supermultiplet V^{IJ}_{μ}, χ, where χ is a Majorana–Weyl gaugino with positive chirality in one case and negative in the other and where we have added gauge indices associated with $\mathrm{U}(n)$ Chan–Paton factors.

Open superstrings also need a closed superstring sector, but with the same spacetime supersymmetry ($N = 1$), which is constructed by taking the quotient of one type IIB by Ω. Then, we have to take also open unoriented superstrings, with gauge group $\mathrm{SO}(n)$ or $\mathrm{Sp}(n)$. The only anomaly-free group is SO(32). The result is the

type-I$_{\pm}$ SO(32) superstring The massless modes correspond to $N = 1_{\pm}, d = 10$ supergravity coupled to SO(32) vector supermultiplets.

The fields corresponding to the massless modes of all these theories are given in Table 14.3.

[14] If we had to compactify both right- and left-moving fields we would obtain $\mathrm{U}(1)^{32}$. The explanation for this phenomenon can be found in Section 14.3.

14.2.3 D-Branes and O-planes in superstring theories

In the bosonic-string case we have seen how theories (oriented or unoriented) with Dirichlet boundary conditions could be constructed by adding to the simplest oriented theories D-branes or O-planes. The presence of these objects changes the boundary conditions and the orientability of the theory. In the bosonic case consistency of the construction was not so important since the original theory was already sick, but in the construction of superstring theories we have chosen only those which are completely (self-)consistent, free of anomalies and tachyons, and we have found all of them in ten dimensions. We can, however, try to construct new theories by compactification of these or by the addition of D-branes and O-planes, breaking in general ten-dimensional Lorentz invariance and supersymmetry.

The rules for how to add D-branes and O-planes consistently to superstring theories are much more restrictive than in the bosonic case. To start with, the following facts have to be taken into account.

1. The type-IIA (B) theory admits only Dp-branes and Op-planes with p even (odd).

2. Dp-branes and Op-planes are charged with respect to the RR $(p+1)$-form potentials of each theory [777] (see Table 14.3) and their duals. This agrees with the fact that type IIA (B) has odd (even)-rank RR potentials only. However, one has to introduce a 9 (10)-form potential $\hat{C}^{(9)}$ ($\hat{C}^{(10)}$) for the D8 (9)-brane and O8 (9)-plane of the Type IIA (B) theory. These fields carry no local degrees of freedom and, therefore, they are not associated with states of the spectrum.

3. Dp-branes carry a unit of positive or negative RR charge that equals its tension $q_{\mathrm{D}p} = \pm T_{\mathrm{D}p}$. O$p$-planes have charges and tensions that depend on the symmetries involved. The prototype is the type-IIB$_\pm$ O9-plane associated with worldsheet parity Ω, which is a symmetry of both theories (but interchanges the two type IIAs) and has $\hat{C}^{(10)}$ charge $q_{\mathrm{O}9} = \pm T_{\mathrm{O}9}$ and tension $T_{\mathrm{O}9} = -32 T_{\mathrm{D}9}$ (negative). The Op-planes related to it by T duality have
$$T_{\mathrm{O}p} = -2^{p-5} T_{\mathrm{D}p}. \tag{14.54}$$

4. The presence of a single Dp-brane or Op-plane halves the supersymmetry of the theory. This is the amount preserved by those objects considered as superstring states. From this point of view they are BPS states and they must saturate some Bogomol'nyi bound, which implies a relation between their tensions and charges (identity in proper units). Then, equilibrium of forces between objects of the same kind (including spatial orientation) is to be expected, as discussed in Chapter 13, which means that we can have parallel Dp-branes with the same charge in equilibrium and enhancement of gauge symmetry when they coincide. We will see that classical solutions that describe these systems can be found in the effective supergravity theories. The stability of these systems is reflected in the absence of tadpoles in the string theory.

5. It is possible to introduce Dp-branes that intersect (at any angle) if the resulting system preserves supersymmetry. There are simple rules for allowed intersections at right angles, which will be studied later.

6. As we have mentioned, the absence of anomalies and tadpoles is related to the stability of the system of Dp-branes and Op-planes. In particular, the system should be able to solve the equations of motion of the string effective theory which we are going to study in the next few chapters. The equations of motion for $(p+1)$-forms are generalizations of the harmonic equation which, in compact spaces, can be solved only if the total charge is zero.[15]

These rules have been used extensively for building new string theories. The simplest construction leads to the type-I SO(32) theory starting from type IIB: on introducing an O9-plane (i.e. taking the quotient of the type-IIB theory by Ω), consistency requires the addition of 32 D9-branes in order to obtain zero total RR charge, which results in the introduction of an open, unoriented-string sector with gauge group SO(32).

14.2.4 String interactions

Strings interact by joining and splitting. It is then easy to understand that open strings can interact to give closed strings and that consistency (unitarity) requires a closed-string sector in open-string theories.

String amplitudes are defined as path integrals over all embeddings X^μ and all worldsheet metrics γ_{ij} with given boundaries and boundary data that determine the string states that are scattered. The boundary data are included as vertex operators in the path integral. Without vertex operators, we have vacuum amplitudes, given by the path integral

$$Z = \int \mathcal{D}X\mathcal{D}\gamma\, e^{-S_{\rm P}-S_{\rm Euler}}, \tag{14.55}$$

where $S_{\rm P}$ is the Euclidean Polyakov integral Eq. (14.5) and $S_{\rm Euler}$ is the topological term Eq. (14.14). For closed strings we will restrict ourselves to (just oriented or oriented plus unoriented) compact surfaces. For open strings we will add surfaces with boundaries.

The sum over metrics can be decomposed into a sum of path integrals over worldsheets with given topologies. The topology of two-dimensional surfaces can be characterized completely by the numbers $g, b,$ and c, combined into the Euler characteristic χ, which is given by the topological term Eq. (14.14) as explained on page 412. The result takes the form

$$Z = \sum_t (e^{\phi_0})^{-\chi(t)} \int_{\{\Sigma_t\}} \mathcal{D}X\mathcal{D}\gamma\, e^{-S_{\rm P\Sigma_t}}, \tag{14.56}$$

where t stands for given topologies and $\{\Sigma_t\}$ is the space of surfaces with topology t. Now, each topology can be associated with a loop order, given precisely by $-\chi(t)$, and the above sum can be understood as a perturbative series expansion in which e^{ϕ_0} plays the role of the *string coupling constant* g:

$$\boxed{g \equiv e^{\phi_0}.} \tag{14.57}$$

In Section 15.1 we will see that ϕ_0 is the vacuum expectation value of the *dilaton field*.

[15] The lines of force of the field can only go to sources or to infinity. In compact spacetimes, they have to start and end on sources and the total charge has to be zero.

14.3 Compactification on S^1: T duality and D-branes

We can obtain four-dimensional string theories by compactification. The simplest compactification would be on a circle. Already in this case we are going to start seeing *stringy* effects (T duality, first discovered in [622, 819], and enhancement of gauge symmetry) that we did not see in the field-theory (KK) case, which are a manifestation of the extended-object nature of strings and a suggestion that there is a minimal length in string theory. General references on T duality are [35, 456].

We are going to study first the compactification of closed bosonic strings on a circle.

14.3.1 Closed bosonic strings on S^1

If $Z \equiv X^{\hat{d}-1}$ is the compact coordinate, it is convenient to identify $Z \sim Z + 2\pi R_z$, where R_z is the compactification radius, and keep using the Minkowski metric. Now, in the mode expansion Eq. (14.40) of Z the following applies.

1. There is another zero mode compatible with the periodicities of ξ^1 and Z:

$$\frac{R_z w}{\ell}\xi^1, \qquad w \in \mathbb{Z}. \tag{14.58}$$

   When we go around the closed string once, $\xi^1 \to \xi^1 + 2\pi\ell$, we go w times around the compact dimension: $Z \to Z + 2\pi R_z w$. This is a *winding mode*, a purely stringy animal that corresponds to the capacity of closed strings to be wrapped w times around compact dimensions.

2. There are also string KK modes as in Chapter 11,

$$\frac{n}{R_z p^+}\xi^0, \qquad n \in \mathbb{Z}. \tag{14.59}$$

Quantization leads to the mass formula and constraint

$$M^2 = \frac{n^2}{R_z^2} + \frac{R_z^2 w^2}{\alpha'^2} + \frac{2}{\alpha'}(N + \tilde{N} - 2), \qquad N = \tilde{N} + nw. \tag{14.60}$$

Observe that the mass of the $w = 1$ mode agrees with the definition of the string tension on page 409: it is the product of the length of the compact dimension and the string tension.

The spectrum is now that of the uncompactified theory (the $n = w = 0$ sector) plus new sectors with non-vanishing KK momentum or winding number. The spectrum of the uncompactified theory has to be interpreted now in $\hat{d} - 1$ dimensions: the $\hat{d}$-dimensional graviton gives rise to a graviton, a (KK) vector and a (KK) scalar in $\hat{d} - 1$ dimensions, while the KR 2-form gives rise to another 2-form and another (*winding*) vector and the dilaton gives another dilaton. For generic values of the compactification radius R_z there are no more massless states and the vector gauge symmetry group is $\mathrm{U}(1)^2$. As discussed in Chapter 11, KK modes are charged with respect to the KK U(1) vector field. We will see that winding modes are charged with respect to the winding U(1) vector field.

The spectrum is invariant under the *T-duality* transformation

$$\boxed{n' = w, \qquad w' = n, \qquad R'_z = \alpha'/R_z.} \tag{14.61}$$

Two bosonic-string theories with one dimension compactified on a circle of radius R_z and α'/R_z have the same spectra, with the winding modes of one of them having the same masses as the KK modes of the other and vice-versa. Not only do they have the same spectra, but also they have the same interactions and scattering amplitudes, but one has to take into account that the string coupling constants of the two theories are related by [33]

$$\boxed{g' = g\ell_{\mathrm{s}}/R_z.} \tag{14.62}$$

This has very important consequences: if we diminish the size of the compactification radius beyond the *self-dual radius* $R_z = \ell_{\mathrm{s}} = \sqrt{\alpha'}$, there is another completely equivalent bosonic-string theory defined on a circle of radius bigger than ℓ_{s}. The self-dual radius can be interpreted as the minimal radius on which a bosonic-string theory can be compactified.

On the other hand, at the self-dual radius there are four additional massless vectors: $N = \pm n = \pm w = 1$ and $\tilde{N} = \pm n = \mp w = 1$. These, plus the KK and the winding vector turn out to be the gauge vectors of $\mathrm{SU}(2) \times \mathrm{SU}(2)$ and we find a new enhancement of symmetry at a special point (a T-duality fixed point) of the moduli space, which is one of the most striking properties of string theory.

Is T duality related to some property of the Polyakov action that we have missed? Actually, yes: the Polyakov action is invariant under Poincaré-dualization of one of the embedding coordinates, just as Maxwell's action is invariant under the dualization of the vector field (Section 8.7.1): we put $\partial_i Z \equiv F_i$ and add to the Polyakov action in Minkowski spacetime a Lagrange multiplier term enforcing the Bianchi identity $\partial_{[i} F_{j]} = 0$:

$$S_P[X^\mu, \gamma_{ij}, Z] = -\frac{T}{2}\int d^2\xi \sqrt{|\gamma|}\, \gamma^{ij}(\partial_i X^\mu \partial_j X_\mu - F_i F_j) - T\int d^2\xi \epsilon^{ij}\partial_i Z' F_j. \tag{14.63}$$

The equations of motion of the Lagrange multiplier Z' and of F_i are

$$\partial_{[i} F_{j]} = 0, \qquad F^i = \epsilon^{ij}\partial_j Z', \tag{14.64}$$

and we can use the latter to eliminate F_i. The result is the Polyakov action with Z replaced by the dual embedding coordinate Z'. This procedure can be used in the presence of non-trivial spacetime metrics that are independent of z, as we will see in Section 15.2.2.

14.3.2 Open bosonic strings on S^1 and D-branes

Equations (14.64) imply $\partial^i Z = \epsilon^{ij}\partial_j Z'$, which implies that, under Poincaré duality, left-moving objects transform into left-moving objects and right-moving ones into minus right-moving objects. Then T duality can be described as the equivalent transformation

$$Z = Z_+ + Z_- \to Z' = Z_+ - Z_-. \tag{14.65}$$

This description is useful for open strings. Open strings with N boundary conditions have KK modes but no winding modes on a circle. Clearly, a T-duality transformation will not take us into another similar open-string theory. Yet, we can perform the transformation and try to identify the resulting theory. The main observation is that N boundary conditions $\partial_+ Z_+ = \partial_- Z_-$ and D boundary conditions $\partial_+ Z_+ = -\partial_- Z_-$ are interchanged by T duality. Indeed, on applying the above transformation to the open string with N boundary conditions mode expansion Eq. (14.38), which we rewrite here in the form

$$Z_\pm = \frac{z}{2} \pm \frac{p^z}{2cp^+}\xi^\pm - i\sqrt{\frac{\alpha'}{2}}\sum_{n\neq 0} e^{\pm\frac{in}{2\ell}\xi^\pm}, \tag{14.66}$$

we find

$$Z' = Z_+ - Z_- = \frac{p^z}{cp^+}\xi^1 + \sqrt{2\alpha'}\sum_{n\neq 0}\frac{\alpha_n}{n}e^{\frac{icn\xi^0}{2\ell}}\sin\left(\frac{n\xi^1}{2\ell}\right), \tag{14.67}$$

which coincides with the expansion with D boundary conditions Eq. (14.39) with $z^1 = 0$ and $z^2 = p^z/T$. If we take into account that $p^z = n/R_z$, we see that $z^2 = n2\pi R'_z$, where $R'_z = \alpha'/R_z$ is the T-dual radius. The coordinate Z with N boundary conditions has been transformed into the compact coordinate Z' with dual compactification radius and D boundary conditions at both ends, $z^1 = 0$ and $z^2 = 2\pi R'_z \sim 0$. The momentum mode n has become a winding mode with winding number n: the string has one endpoint attached to the D24-brane, winds around the compact dimension n times, and ends again on the same D24-brane.

We could repeat the procedure in another compact coordinate, giving a D23-brane. Thus, T duality in a direction parallel to a Dp-brane transforms it into a D$(p-1)$-brane with a compact transverse dimension and vice-versa. Furthermore, with Chan–Paton factors, we would have found endpoints with different labels on the same hypersurface. Thus, we would have overlapping D-branes and the gauge group would be preserved by T duality.

What happens when we perform T duality on two parallel non-overlapping D24-branes whose transverse dimension is compact? There are, as we discussed, four sectors labeled by pairs ij with $ij = 1, 2$ indicating on which of the two D24-branes the first and second endpoints are. The spectrum is consistent with spontaneously broken U(2) gauge symmetry. The 11 and 22 sectors are T dual to open strings with N boundary conditions. If the D24-branes are placed at angles θ_1, and θ_2, the 12 and 21 sectors with winding number w have expansions

$$Z^{ij}_\pm = \frac{R_z\theta_i}{2} + \frac{[2\pi w - (\theta_j - \theta_i)]R_z}{4\pi\ell}\xi^\pm \mp i\sqrt{\frac{\alpha'}{2}}\sum_{n\neq 0} e^{\pm\frac{in}{2\ell}\xi^\pm}, \tag{14.68}$$

and the T-dual expansion is that of an N open string with momentum given by $(p^{z\,ij})' = [w - (\theta_i - \theta - j)]/R'_z$. The shift with respect to the standard KK momentum w/R'_z is caused by the appearance of *Wilson lines* of the U(2) gauge field V^{ij} of the N open strings with two Chan–Paton factors. As explained on page 347, in a background with a compact dimension we can have topologically non-trivial configurations of the gauge fields characterized by the line integral of the gauge field around the compact direction (known as a

Wilson line):

$$W[C] = \exp\left\{ i \oint_C A \right\}. \tag{14.69}$$

In this case, in the dual-string theory, we have the gauge-field configuration

$$\left(V_{\underline{z}}^{ij} \right) = -\frac{1}{2\pi R'_z} \begin{pmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{pmatrix}. \tag{14.70}$$

This non-trivial background breaks U(2) down to $\mathrm{U}(1)^2$, as can be seen in the spectrum[16] [779]. Thus, the symmetry in the original configuration is equal to that in the dual configuration. In one case, the symmetry breaking is associated with the separation of the D24-branes and in the T dual it is associated with the presence of Wilson lines (D25-branes, being spacetime-filling branes, cannot be spatially separated), but gauge symmetry is preserved by T duality. The generalization to n D-branes and $\mathrm{U}(n)$ gauge symmetry is straightforward.

14.3.3 *Superstrings on* S^1

Closed and heterotic superstring theories compactified on circles also have T duals: the type-IIA and type-IIB theories compactified on circles of T-dual radii are each other's dual [283, 322] and the SO(32) and $E_8 \times E_8$ heterotic strings compactified on circles of T-dual radii are each other's T-dual theory. Type-I SO(32) strings also have a T-dual theory, the so-called type-I$'$, which has a nine-dimensional interpretation.

To make further progress in the study of string dualities, we need to study string effective actions. They will allow us to extend these results easily to more general backgrounds and they will also allow us to find possible non-perturbative dualities that the perturbative formulation of string theory that we have just sketched in this chapter is not powerful enough to exhibit.

---

[16] To find the spectrum, one has to take into account the coupling of the U(2) gauge vector to the Polyakov action through the boundary term Eq. (15.6).

15

The string effective action and T duality

After the previous chapter's brief introduction to perturbative string theory and T duality we are going to discuss how one arrives at the low-energy string effective (field theory) actions, what their meaning is, and what their limits of validity are. We are going to start exploiting them to study T duality and find *Buscher's T-duality rules*, that relate different curved backgrounds that are equivalent from the string-theory point of view. These rules are some of the most powerful tools of string theory.

15.1 Effective actions and background fields

The low-energy string effective action describes the low-energy dynamics of a given string theory. Here low energy means energies lower than the relevant energy scale: the string mass $m_{\rm s}$. Thus, the low-energy limit is the $\alpha' \to 0$ limit, heuristically, the limit in which the string length can be ignored and a theory of particles (a field theory) is recovered.

On the other hand, at low energies only the massless modes are relevant and their dynamics is described by a theory of the corresponding massless fields. The obvious way to find this field theory is to compute string amplitudes for the massless modes, take the $\alpha' \to 0$ limit and then construct a field theory that reproduces these amplitudes. In principle, the effective field theory has an expansion in powers of α', although usually only the lowest-order terms are considered. The terms of higher order in α' are also of higher order in derivatives. Also, string amplitudes are calculated order by order in string perturbation theory and the effective action can also be expanded in powers of the string coupling constant, which here is the exponential of the dilaton field. Again, only the lowest orders are usually considered.

Actually, for some superstring theories, it is possible to arrive at the effective theory using (super)symmetry arguments. In particular, the massless modes of the type-II superstrings fill the supergravity multiplets of the (two only) maximal ten-dimensional supergravity theories: those of the type-IIA (non-chiral) theory fill the supergravity multiplet of $N = 2A, d = 10$ SUEGRA [211, 427, 571, 585] whose action is given in Eq. (16.38) and those of the type-IIB (chiral) theory fill the supergravity multiplet $N = 2B, d = 10$ SUEGRA [571, 820, 823] whose action is given Eq. (17.4). Similarly, since there is only one $N = 1$ supergravity in $d = 10$ dimensions [114, 233, 457], coupled to vector fields with the right gauge group it must be the effective action of the heterotic and type-I strings (but already here the couplings of the 2-form are different in the two theories).

The fields of these effective theories are given in Table 14.3. The NSNS fields are sometimes called the *common sector* since they occur in all of them, including the bosonic (oriented)-string theories. The fields in the common sector are the metric $g_{\mu\nu}$, associated with the graviton, the KR 2-form $B_{\mu\nu}$, and the *dilaton* ϕ whose vacuum expectation value ϕ_0 gives the string coupling constant $g = e^{\phi_0}$ (see Eqs. (15.8) and (14.14)). The action for the common sector in the *string frame* to be defined below is given (in d dimensions with $d = 10$ and 26) by

$$\boxed{S = \frac{g^2}{16\pi G_{\rm N}^{(d)}} \int d^d x \sqrt{|g|}\, e^{-2\phi}\left[R - 4(\partial\phi)^2 + \frac{1}{2\cdot 3!}H^2\right],} \tag{15.1}$$

where, in our notation in which indices not shown are all antisymmetrized,

$$H = 3\partial B \tag{15.2}$$

is the KR field strength, which is invariant under the gauge transformations necessary for the consistent quantization of a massless 2-form field,

$$\delta B = 2\partial\Sigma, \tag{15.3}$$

where Σ_μ is an arbitrary vector field. The overall factor $e^{-2\phi}$ is associated with the genus-0 (tree-level) origin of these terms and the normalization is conventional. In particular, the factor g^2 (g is defined in Eq. (14.57)) compensates the factor $e^{-2\phi_0}$ that appears in the weak field expansion of the action around the vacuum $g_{\mu\nu} = \eta_{\mu\nu}$, $\phi = \phi_0$, so $G_{\rm N}^{(d)}$ can be interpreted as the d-dimensional Newton constant. Its value Eq. (19.26) can be determined by using duality arguments that relate it to the string coupling constant and the string length just as we determined the Newton constant in terms of the compactification radius of KK theory in Eq. (11.97), as we will see in Section 19.3.

Observe that, up to normalization, the above action is also invariant under shifts of the dilaton field that change its vacuum expectation value and thus g, which is a free parameter. A potential for the dilaton could give it a mass and fix g, solving two problems simultaneously: the determination of g and the existence of a massless scalar that couples as the Jordan–Brans–Dicke field to all kinds of matter, inducing violations of the equivalence principle [287, 288]. The massless KR 2-form that also couples to matter can also be a source of violations of the equivalence principle. The KR 3-form field strength can be understood as a completely antisymmetric dynamical torsion field, as we discussed on page 132 and, in the same spirit, the dilaton can be understood as part of a non-metricity tensor of the type considered by Weyl [834]. The above string effective action can then be written as an Einstein–Hilbert action for a torsionful and non-metric-compatible connection using Eq. (1.55) (see also Eq. (1.58)).

The RR fields are differential forms ${C^{(n)}}_{\mu_1\cdots\mu_n}$ of even (odd) rank in the $N = 2B$ (A) theory and appear in the respective actions Eqs. (17.4) and Eq. (16.38) with no couplings to the dilaton in the string frame. They couple to the KR 2-form due to the definitions of the field strengths and also to the presence of Chern–Simons (CS) topological terms in the supergravity actions. These CS terms contain a great deal of information on the possible intersections of extended objects of the theory [900].

Although the identification of the field theories on the basis of symmetry arguments is correct, the identification of the fields with the string modes is ambiguous, since the supergravity theories are unique up to field redefinitions. To establish completely the relation between supergravity fields and string modes, it is necessary to have more information. For instance, making use of the relations in Figure 14.1, the supergravity fields must be related by dualities in the same way as the string modes are.

String effective actions also arise in a different way: string theories are usually quantized in flat spacetime, but the string worldsheet action can be written in a curved background as a non-linear σ-model, Eq. (14.5), and, furthermore, can be generalized to describe the coupling to all background fields associated with the string massless modes.[1] The coupling of the string to the Kalb–Ramond 2-form $B_{\mu\nu}$ is represented by a WZ term that generalizes the coupling of the Maxwell vector field to a charged point-particle, Eq. (8.53), i.e. it is the integral of the pull-back of the 2-form over the two-dimensional worldsheet:

$$\frac{T}{2}\int_{\Sigma} B, \tag{15.4}$$

where B is given by

$$B = \tfrac{1}{2} B_{ij} d\xi^i \wedge d\xi^j = d^2\xi \epsilon^{ij} B_{ij} = d^2\xi \epsilon^{ij} \partial_i X^\mu \partial_j X^\nu B_{\mu\nu}. \tag{15.5}$$

Observe that the role of the electric charge is played here by the string tension. This coefficient can be changed but we will take it as above, defining the normalization of $B_{\mu\nu}$ that we will use. This is the normalization that leads to the effective action Eq. (15.1). Observe also that the WZ term, being topological (metric-independent), is automatically Weyl-invariant and does not contribute to the γ_{ij} equation of motion. Furthermore, the WZ term is invariant, up to total derivatives, under gauge transformations of the 2-form, Eq. (15.3), which means that strings are charged with respect to the KR 2-form and the charge is conserved. In open-string worldsheets the non-vanishing boundary term is canceled out by the variation of the term that represents the coupling of the open string to the 1-form V_μ:

$$\int_{\partial\Sigma} V, \tag{15.6}$$

provided that the vector transforms under the KR 2-form gauge symmetry Eq. (15.3),

$$\delta V_\mu = T\Sigma_\mu. \tag{15.7}$$

Finally, this term is not parity-invariant and occurs only in oriented-string theories.

[1] This is always true for the fields in the common sector for the bosonic and fermionic strings, although worldsheet supersymmetry has to be studied case by case. The inclusion of RR massless superstring fields in the σ-model is more complicated and how to do it is known only in certain cases.

The coupling to the dilaton is related to the topological term Eq. (14.14):

$$-\frac{1}{4\pi}\int d^2\xi\sqrt{|\gamma|}\,\phi(X)R(\gamma). \tag{15.8}$$

This term, which is of higher order in α' for dimensional reasons, gives no contribution to the γ_{ij} equation of motion but breaks Weyl invariance for generic dilaton fields.

Since Weyl invariance is absolutely necessary for the consistency of string theory, the natural question to be asked is that of in which backgrounds $g_{\mu\nu}$, $B_{\mu\nu}$, and ϕ Weyl invariance is quantum-mechanically preserved. The background fields can be understood as coupling functions and then the question can be reformulated in terms of the vanishing of the beta *functionals* associated with them.

To lowest order, these beta functionals are given by [205]

$$\begin{aligned}
\beta^g_{\mu\nu} &= \alpha'\left[R_{\mu\nu} - 2\nabla_\mu\nabla_\nu\phi + \tfrac{1}{4}H_\mu{}^{\alpha\beta}H_{\nu\alpha\beta}\right] + \mathcal{O}(\alpha'^2),\\
\beta^B_{\mu\nu} &= \frac{\alpha'}{2}e^{2\phi}\nabla^\rho(e^{-2\phi}H_{\rho\mu\nu}) + \mathcal{O}(\alpha'^2),\\
\beta^\phi &= \frac{d-26}{6} - \frac{\alpha'}{2}\left[\nabla^2\phi - (\partial\phi)^2 - \tfrac{1}{4}R - \tfrac{1}{48}H^2\right] + \mathcal{O}(\alpha'^2),
\end{aligned} \tag{15.9}$$

and it turns out that the vanishing of these beta functionals is equivalent to the equations of motion derived from the action Eq. (15.1) plus a term

$$\frac{g^2}{16\pi G_{\rm N}^{(d)}}\int d^dx\sqrt{|g|}\,e^{-2\phi}[-2(d-2)\Lambda], \qquad \Lambda = \frac{2(d-26)}{3\alpha'(d-2)}, \tag{15.10}$$

for the bosonic string, which vanishes in the critical dimension $d = 26$ (the same happens for the fermionic string for $d = 10$). Indeed, the equations of motion are (see Section 4.2)

$$\begin{aligned}
\frac{16\pi G_{\rm N}^{(d)}e^{2(\phi-\phi_0)}}{\sqrt{|g|}}\frac{\delta S}{\delta g^{\mu\nu}} &\sim \frac{1}{\alpha'}\left(\beta^g_{\mu\nu} - 4g_{\mu\nu}\beta^\phi\right) + \mathcal{O}(\alpha'),\\
\frac{16\pi G_{\rm N}^{(d)}e^{2(\phi-\phi_0)}}{\sqrt{|g|}}\frac{\delta S}{\delta\phi} &\sim -\frac{16}{\alpha'}\beta^\phi + \mathcal{O}(\alpha'),\\
\frac{16\pi G_{\rm N}^{(d)}e^{2(\phi-\phi_0)}}{\sqrt{|g|}}\frac{\delta S}{\delta B_{\mu\nu}} &\sim -\frac{1}{\alpha'}\beta^{B\,\mu\nu} + \mathcal{O}(\alpha').
\end{aligned} \tag{15.11}$$

Thus, we see that quantum conformal invariance leads (in the critical dimension) to the same effective action for the string common sector Eq. (15.1). The metric that appears in that action is the same metric as that to which the string couples and therefore appears in the σ-model and is called the *string-frame metric*. A conformal rescaling,

$$g_{\mu\nu} = e^{\frac{4}{d-2}\phi}g_{{\rm E}\,\mu\nu}, \tag{15.12}$$

can eliminate the factor $e^{-2\phi}$ in front of the Einstein–Hilbert term (see Appendix E). The rescaled metric $g_{{\rm E}\,\mu\nu}$ is called the *Einstein-frame metric*. In the Einstein frame, the string action is given by

$$S = \frac{1}{16\pi G_{\mathrm{N}}^{(d)}} \int d^d x \sqrt{|g_{\mathrm{E}}|} \Big[R_{\mathrm{E}} + \frac{4}{d-2} (\partial \phi)^2 + \frac{1}{2 \cdot 3!} e^{\frac{-8}{d-2}\phi} H^2 - (d-2) \Lambda e^{\frac{4}{d-2}\phi} \Big]. \tag{15.13}$$

The solutions to these equations describe backgrounds (vacua) for bosonic-string theory in which strings can be consistently quantized, to lowest order in α' and the string coupling constant. The simplest is evidently ten-dimensional Minkowski spacetime, which should remain a good vacuum to all orders[2] because all fields are trivial. Other vacua can be argued to be exact and not to receive higher α' corrections due to their unbroken supersymmetry and/or the vanishing of their curvature invariants as is the case with pp-wave solutions [41, 129, 493, 558–60, 910] and the four-dimensional solutions of [250] which are based on the classification of the four-dimensional metrics that have all the curvature invariants vanishing [789]. The next step is to try to quantize string theory on these vacua (for instance in the KG10 solution [694, 696]).

It is amusing to see that quantizing string theory in non-trivial backgrounds amounts to finding the generalization of Pythagoras' law for vibrating strings (arguably the first law in the history of physics) and that the generalization is possible only for backgrounds that satisfy the above generalization of the Einstein equations.

The effective action has been obtained perturbatively in both α' and the string coupling constant g and, furthermore, in the low-energy (long-distance) approximation. As a general rule, the results obtained working with it can be trusted as long as $e^{\phi} \ll 1$ and the curvature scalar $R \ll \ell_{\mathrm{s}}^{-2}$ or $R\alpha' \ll 1$. Sometimes there are lengths in the system under consideration that do not appear in the curvature, such as the radius of a circle. These distances should be bigger than the string scale ℓ_{s}. At distances of the order of ℓ_{s}, or curvatures of the order of ℓ_{s}^2, there are stringy effects that invalidate the effective action and its solutions, unless higher order corrections to both are taken into account.

For instance, we have seen in the previous chapter that, when the distance between two D-branes becomes of the order of ℓ_{s} or when the compactification radius takes the self-dual value of ℓ_{s}, new massless states appear in the string spectrum that were not taken into account in the calculation of the string effective action and should be included by hand at that point. Beyond that point, the T-dual description of the effective-field theory should be used. In more complex geometries one can find radii that are functions of other coordinates, so there are regions in which they are smaller than ℓ_{s}. The solution cannot be trusted in those regions. This is the basis of Sen's argument, which we mentioned on page 245.

15.1.1 The D-brane effective action

The effective action of a Dp-brane (or, rather, the effective action of open-plus-closed-string theory in the presence of a Dp-brane so the open-string sector has D boundary conditions. in $25 - p$ coordinates) in a background with metric, KR 2-form, dilaton, and $(p+1)$-dimensional U(1) vector is given by the following generalization of the

[2] In superstring theories. In bosonic-string theories, the 26-dimensional Minkowski spacetime is not a stable vacuum, as the presence of the tachyon suggests, and is also α'-corrected [395], although it is not known to which solution the corrections converge.

Nambu–Goto action [647]:

$$\boxed{S = -T_{\mathrm{D}p} g \int d^{p+1}\xi \, e^{-\phi} \sqrt{|g_{ij} + B_{ij} + 2\pi\alpha' F_{ij}|},} \tag{15.14}$$

where g_{ij} and B_{ij} are the pull-backs over the $(p+1)$-dimensional worldvolume of the spacetime fields and $F_{ij} = 2\partial_{[i} V_{j]}$ is the standard field strength of a gauge vector that in this context is called the *Born–Infeld* (BI) vector field since the above action, with flat spacetime metric and zero dilaton and KR field, was proposed by Born and Infeld in [171–3] as a nonlinear model for electrodynamics: the power expansion of the action contains a Maxwell term F^2 and an infinite series of higher-order terms. The main property of this theory is that the spherically symmetric field was singularity-free and had a *core* with characteristic size $\sim\ell_{\mathrm{s}}$ with the above normalization.[3] $T_{\mathrm{D}p}$ is the Dp-brane tension.[4]

This action is invariant under spacetime and worldvolume reparametrizations, and also under the KR and BI vector-field gauge transformations Eqs. (15.3) and (15.7) and has to be added to the *bulk* closed-string effective action Eq. (15.1). If the strings are unoriented, then there is no KR 2-form in either of them.

The fact that, in the presence of D boundary conditions, the string effective action includes the worldvolume action of an extended object, the Dp-brane, is a final argument in favor of the interpretation of the latter as a dynamical object. On the other hand, the Dp-brane action is a generalization of the string σ-model action and it may constitute the starting point for quantization. Actually, the Dp-brane actions of superstring theories can (and must) couple to all the fields of the (bulk) closed-superstring effective action, which is a supergravity action. The consistent coupling to all these fields requires, first of all, a WZ term for the action to be invariant under κ-symmetry transformations. The bosonic part of the WZ term describes the coupling to a RR $(p+1)$-form potential and other RR potentials of lower rank. Thus, as first shown by Polchinski in [777], superstring Dp-branes carry RR charges and are sources of the RR fields of those actions. We will discuss the effective actions of these super-D-branes in Chapter 19.

15.2 T duality and background fields: Buscher's rules

In the preceding section we have introduced string effective-field-theory actions and in this section we want to show how to use them to study string dualities in the simplest case: T duality. We will consider the effective action for the string common sector and the results will be valid only for closed bosonic strings but will later be generalized to the heterotic and type-II cases. We essentially follow [131], where the heterotic case was studied, with a few notational changes.

---

[3] A recent review actions of on this kind is [909].

[4] In the literature, the same quantity as that which we define as tension is called *effective tension*, and the coefficient in front of the Dp-brane effective action is called *tension*. We use only the physical tension parameter defined above, which should avoid confusion.

15.2.1 T duality in the bosonic-string effective action

T duality relates closed $\hat{d}$-dimensional string theories compactified on circles of relatively dual radii. The effective-field theories will be $\hat{d}-1=d$-dimensional field theories for the massless modes and the KK formalism that was developed in Chapter 11 is perfectly suited to obtaining them from the effective actions of the uncompactified $\hat{d}$-dimensional effective theories.[5]

Our starting point is the action Eq. (15.1) with hats on every object, following the notation of Chapter 11. We denote the compact coordinate by $x^{\hat{d}-1} \equiv z \in [0, 2\pi\ell_{\rm s}]$, and assume that all fields are independent of it. We can use the standard KK *Ansatz* Eq. (11.33) and the results concerning the spin connection, Eqs. (11.36) and (11.35), and volume element, Eq. (11.37). Before substituting in the Einstein–Hilbert part of the action, we use the $\hat{d}$-dimensional Palatini identity Eq. (D.4) with $K=e^{-2\hat{\phi}}$ and immediately obtain

$$\begin{aligned}\int d^{\hat{d}}\hat{x}\,\sqrt{|\hat{g}|}\,e^{-2\hat{\phi}}\hat{R} = \int dz \int d^{\hat{d}-1}x\,\sqrt{|g|}\,e^{-2\hat{\phi}}k\Big\{ &-\omega_b{}^{ba}\omega_c{}^c{}_a - \omega_a{}^{bc}\omega_{bc}{}^a \\ &+2\omega_b{}^{ba}\partial_a \ln(e^{-2\hat{\phi}}k) - 2\partial_a \ln k\,\partial^a \ln e^{-2\hat{\phi}} - \tfrac{1}{4}k^2F^2(A)\Big\}.\end{aligned} \tag{15.15}$$

It is evident that the combination $e^{-2\hat{\phi}}k$ now plays the role of a d-dimensional dilaton, and thus we define

$$\phi \equiv \hat{\phi} - \tfrac{1}{2}\ln k. \tag{15.16}$$

The kinetic term for the dilaton gives

$$\int d^{\hat{d}}\hat{x}\,\sqrt{|\hat{g}|}\,e^{-2\hat{\phi}}\Big[-4(\partial\hat{\phi})^2\Big] = \int dz \int d^{\hat{d}-1}x\,\sqrt{|g|}\,e^{-2\hat{\phi}}k\Big[-4(\partial\hat{\phi})^2\Big]. \tag{15.17}$$

On combining these two terms and using now the d-dimensional Palatini identity with $K=e^{-2\phi}$, we obtain, straightforwardly,

$$\begin{aligned}\int d^{\hat{d}}\hat{x}\,\sqrt{|\hat{g}|}\,e^{-2\hat{\phi}}[\hat{R}-4(\partial\hat{\phi})^2] = \int dz\int d^{\hat{d}-1}x\,\sqrt{|g|}\,e^{-2\phi}\big[&R-4(\partial\phi)^2 \\ &+(\partial\ln k)^2 - \tfrac{1}{4}k^2F^2(A)\big].\end{aligned} \tag{15.18}$$

The reduction of the KR 2-form is a bit trickier. First, we reduce the field strength by identifying, in tangent-space indices,

$$\hat{H}_{abc} \equiv H_{abc}, \qquad \hat{H}_{abz} = k^{-1}F_{ab}(B), \tag{15.19}$$

where

$$F(B) = 2\partial B, \qquad B_\mu \equiv \hat{B}_{\mu\underline{z}}. \tag{15.20}$$

[5] Observe that the KK formalism may describe the massive KK modes of the string but not the massive winding modes. The massless modes have zero momentum and winding number except at the self-dual radius, at which there are additional massless modes with non-trivial KK momentum and winding number. The effective action that we are going to write cannot describe the enhancement of symmetry that takes place at the self-dual radius. A direct calculation of the string effective action for that radius is necessary.

The above identification of H does not completely determine the reduction of the KR 2-form. It simply gives

$$H_{\mu\nu\rho} = 3(\partial_{[\mu}\hat{B}_{\nu\rho]} - A_{[\mu}F(B)_{\nu\rho]}). \tag{15.21}$$

Now, we could define $B_{\mu\nu} = \hat{B}_{\mu\nu}$, but we are free to implement field redefinitions, for the sake of convenience. Here, it is convenient to define

$$\hat{B}_{\mu\nu} = B_{\mu\nu} - A_{[\mu}B_{\nu]}, \quad \Rightarrow H = 3[\partial B - \tfrac{1}{2}AF(B) - \tfrac{1}{2}BF(A)], \tag{15.22}$$

so $B_{\mu\nu}$ is invariant under the interchange of the two vector fields A_μ and B_μ.

The presence of two additional terms in the field strength H (apart from $3\partial B$), generically known as *Chern–Simons terms*, is a new feature that is quite common in higher-dimensional supergravities. H is invariant under the gauge transformations of the vector fields and the 2-form, but the 2-form must also transform under gauge transformations of the vector fields (the so-called *Nicolai–Townsend transformations*):

$$\begin{aligned} \delta_\Lambda A_\mu &= \partial_\mu \Lambda, \qquad \delta_\Sigma B_\mu = \partial_\mu \Sigma, \\ \delta B_{\mu\nu} &= 2\partial_{[\mu}\Sigma_{\nu]} + B_{[\mu}\partial_{\nu]}\Lambda + A_{[\mu}\partial_{\nu]}\Sigma. \end{aligned} \tag{15.23}$$

The origin of the gauge transformation of the KK vector field A is the GCTs of the compact coordinate $\delta z = -\Lambda$, while the gauge transformations of the vector field B and the 2-form are the gauge transformations of the $\hat{d}$-dimensional 2-form Eq. (15.3):

$$\Sigma_\mu = \hat{\Sigma}_\mu, \qquad \Sigma = \hat{\Sigma}_{\underline{z}}. \tag{15.24}$$

After integration over the compact coordinate, the dimensionally reduced effective action takes the form

$$S \sim \int d^dx\sqrt{|g|}\, e^{-2\phi}\Big[R - 4(\partial\phi)^2 + \frac{1}{2\cdot 3!}H^2 + (\partial \log k)^2 - \tfrac{1}{4}k^2F^2(A) \\ - \tfrac{1}{4}k^{-2}F^2(B)\Big]. \tag{15.25}$$

We could now rescale the d-dimensional fields as we did in Eq. (11.54) and rewrite the action in terms of the scale-invariant fields:

$$\tilde{k} = k/k_0, \qquad \tilde{A}_\mu = k_0 A_\mu, \qquad \tilde{B}_\mu = k_0^{-1}B_\mu. \tag{15.26}$$

The action Eq. (15.25) is manifestly invariant under the transformations

$$A_\mu \to B_\mu, \qquad B_\mu \to A_\mu, \qquad k \to k^{-1}, \tag{15.27}$$

which invert the KK scalar (and therefore the radius of compactification) and interchange the KK vector field that couples to KK modes with the vector field that, as we are going to see, couples to the winding modes. These are, evidently, part of the string T-duality transformations.

There are two ways to understand these transformations: first, we compactify a string background, T-dualize it, and decompactify it into a different background. Another way to think about them is to think of two compactifications of T-dual backgrounds: in one of them we call the KK scalar and vector field k and A and, in the other, we call them k^{-1} and B. These two dual compactifications give the same d-dimensional string background. This is a better way to think about T-duality transformations because it is the one that generalizes to type-II effective actions.

In both cases, it is easy to relate the $\hat{d}$-dimensional metric, KR 2-form, and dilaton of the two dual string backgrounds (primed and unprimed):[6]

$$\boxed{\begin{array}{ll} \hat{g}'_{\underline{zz}} = 1/\hat{g}_{\underline{zz}}, & \hat{B}'_{\mu\underline{z}} = \hat{g}_{\mu\underline{z}}/\hat{g}_{\underline{zz}}, \\ \hat{g}'_{\mu\underline{z}} = \hat{B}_{\mu\underline{z}}/\hat{g}_{\underline{zz}}, & \hat{B}'_{\mu\nu} = \hat{B}_{\mu\nu} + 2\hat{g}_{[\mu|\underline{z}|}\hat{B}_{\nu]\underline{z}}/\hat{g}_{\underline{zz}}, \\ \hat{g}'_{\mu\nu} = \hat{g}_{\mu\nu} - (\hat{g}_{\mu\underline{z}}\hat{g}_{\nu\underline{z}} - \hat{B}_{\mu\underline{z}}\hat{B}_{\nu\underline{z}})/\hat{g}_{\underline{zz}}, & \hat{\phi}' = \hat{\phi} - \frac{1}{2}\ln|\hat{g}_{\underline{zz}}|. \end{array}} \tag{15.28}$$

These relations are known as *Buscher's rules* [198–200] and relate two backgrounds with one isometry that are completely equivalent[7] from the string-theory point of view and, in particular, are classical solutions of the string effective-action Eq. (15.1). If we set $\ell_z = \ell_s$, we immediately obtain the relations Eqs. (14.61) and (14.62) between the moduli of the two dual theories.

The rules were originally derived using the string σ-model, as we are going to do in the next section (although at the classical level), but the effective-action method [121, 125, 130, 675] turns out to give the correct rules in a much simpler way. In Section 15.3 we will study some simple examples of string solutions and T dualization using Buscher's rules, although string solutions and their duality relations are the main theme of Part III and we will see many more examples in later chapters.

Buscher's rules refer only to solutions with an isometry.[8] However, from the string point of view, it seems that it should be possible to define T duality whenever strings can be wrapped around non-contractible cycles. However, the only (partial) realization of this more general duality has been achieved in [476].

To end this discussion on Buscher's T-duality rules, let us make some important remarks.

1. These rules are valid only to lowest order in α'.
2. T duality does not commute with gauge transformations (reparametrizations or gauge transformations of the KR 2-form).
3. In the presence of fermions, Buscher's rules have to be formulated in terms of the Vielbein instead of the metric. We have used the Scherk–Schwarz recipe, which employs the Vielbein formalism, to derive the rules and one could draw the conclusion

[6] These rules are valid only for the heterotic-string background fields (all in the NSNS sector) at lowest order in α'. At higher orders in α' one has to take into account the Yang–Mills fields and also corrections to these rules [34, 126].

[7] If the isometric direction is not compact or corresponds to an isometry with fixed points (a rotation instead of a translation) so that strings cannot wrap around it, the stringy equivalence between the two solutions related by Buscher's rules need not be true. Still, the new configuration solves the string equations of motion and it *is* another string solution [805].

[8] It is clear, though, that they can be extended to the case of several mutually commuting symmetries (toroidal compactifications). The rules follow from the results of Section 16.5.

that our results automatically imply a transformation rule for the Vielbein. However, the rules involve only world tensors and they determine the transformation rules for the Vielbeins up to (z-independent) local Lorentz transformations, and only by considering T duality with fermions is the indeterminacy eliminated and one finds just two possible transformation rules for the Vielbein [121]:

$$\boxed{\hat{e}^{\hat{a}\,\prime}{}_{\underline{z}} = \mp \hat{e}^{\hat{a}}{}_{\underline{z}}/\hat{g}_{\underline{zz}}, \qquad \hat{e}^{\hat{a}\,\prime}{}_{\mu} = \hat{e}^{\hat{a}}{}_{\mu} - (\hat{g}_{\mu\underline{z}} \pm \hat{B}_{\mu\underline{z}})\hat{e}^{\hat{a}}{}_{\underline{z}}/\hat{g}_{\underline{zz}}.} \tag{15.29}$$

Both signs lead to the same Buscher rules for world tensors Eqs. (15.28). Now, if we start with the standard gauge choice for the Vielbein Eq. (11.33), the two possible T-dual Vielbeins are given by

$$\left(\hat{e}_{\hat{\mu}}{}^{\hat{a}\,\prime}\right) = \begin{pmatrix} e_{\mu}{}^{a} & \pm k^{-1} B_{\mu} \\ 0 & \pm k^{-1} \end{pmatrix}, \qquad \left(\hat{e}_{\hat{a}}{}^{\hat{\mu}\,\prime}\right) = \begin{pmatrix} e_{a}{}^{\mu} & -B_{a} \\ 0 & \pm k \end{pmatrix}. \tag{15.30}$$

We will see in Section 17.4 that T duality in type-II theories requires the use of the lower ("non-standard") sign for it to work in the fermionic sector.

15.2.2 T duality in the bosonic-string worldsheet action

We can also gain some insight by studying T duality from the point of view of the two-dimensional σ-model that describes the motion of a string in a $\hat{d}$-dimensional spacetime with a metric $\hat{g}_{\hat{\mu}\hat{\nu}}$ and a KR 2-form $\hat{B}_{\hat{\mu}\hat{\nu}}$:

$$\hat{S} = -\frac{T}{2}\int d^2\xi \sqrt{|\gamma|}\gamma^{ij}\hat{g}_{\hat{\mu}\hat{\nu}}(\hat{X})\partial_i \hat{X}^{\hat{\mu}}\partial_j \hat{X}^{\hat{\nu}} + \frac{T}{2}\int d^2\xi \; \epsilon^{ij}\hat{B}_{\hat{\mu}\hat{\nu}}(\hat{X})\partial_i \hat{X}^{\hat{\mu}}\partial_j \hat{X}^{\hat{\nu}}. \tag{15.31}$$

We do not include the dilaton term Eq. (15.8) since, in our purely classical approach, it is not going to play any role at all.[9] As in the effective action, we assume that the spacetime fields are independent of $z = x^{\hat{d}-1}$ and, thus, the embedding coordinate Z appears only through its derivatives. We then decompose the $\hat{d}$-dimensional fields into $(\hat{d}-1)$-dimensional fields using Eqs. (11.28), (15.20), and (15.22) and, on substituting into the above, we obtain

$$\hat{S} = -\frac{T}{2}\int d^2\xi \sqrt{|\gamma|}\left[\gamma^{ij}g_{ij} - k^2 F^2\right] + \frac{T}{2}\int d^2\xi \; \epsilon^{ij}\left[B_{ij} + A_i B_j - 2F_i B_j\right], \tag{15.32}$$

where g_{ij}, B_{ij}, A_i, and B_i are the pull-backs of the d-dimensional metric, KR 2-form, KK vector, and winding vector and where

$$F_i = \partial_i Z + A_i, \tag{15.33}$$

is the field strength of Z, which is invariant under the shifts

$$\delta_\Lambda Z = -\Lambda(X), \qquad \delta_\Lambda A_\mu = \partial_\mu \Lambda. \tag{15.34}$$

[9] The T-duality transformation rule of the dilaton is a quantum effect. We found it in the string effective action because this action contains information about the quantum theory.

There is a conserved current associated with this invariance that coincides with the momentum canonically conjugate to the cyclic coordinate Z:

$$P_z{}^i = \frac{1}{\sqrt{|\gamma|}}\frac{\delta\hat{S}}{\delta\partial_i Z} = \frac{1}{\sqrt{|\gamma|}}\frac{\delta\hat{S}}{\delta F_i} = T(k^2F^i - {}^\star B^i), \qquad \nabla_i P_z{}^i = 0, \tag{15.35}$$

and there is, as usual, an associated magnetic-like (i.e. topologically) conserved current:

$$W_z{}^i = T{}^\star F^i - {}^\star A^i, \qquad \nabla_i W_z{}^i \sim \epsilon^{ij}\partial_i\partial_j Z = 0. \tag{15.36}$$

The charge associated with this current is the string winding number in the compact dimension: up to normalization

$$\int d\xi^1\sqrt{|\gamma|}W_{\underline{z}}^{\underline{0}} \sim \int d\xi^1\partial_{\underline{1}}Z. \tag{15.37}$$

We will see that the charges associated with these currents are the momentum of the string in the compact dimension and the winding number, respectively.

As in the Minkowski-spacetime case, we want to perform a Poincaré-duality transformation of the action (15.32) with respect to the scalar Z. The Lagrange-multiplier term that we have to add to enforce the Bianchi identity is now

$$+T\int d^2\xi\ \epsilon^{ij}\partial_i Z'\ (F_j - A_j), \tag{15.38}$$

Now we want to eliminate F_i by using its equation of motion, which is the constraint

$$F_i = k^{-2}\,{}^\star F_i', \qquad F_i' \equiv \partial_i Z' + B_i. \tag{15.39}$$

Since B_μ transforms under δ_Σ in Eq. (15.23), for $\tilde{F}_i$ to be gauge-invariant (as the l.h.s. of the above equation is) Z' has to transform simultaneously as follows:

$$\delta_\Sigma Z' = -\Sigma. \tag{15.40}$$

On substituting Eq. (15.39) into the the action modified with the Lagrange-multiplier term, we obtain the dual action

$$\hat{S}' = -\frac{T}{2}\int d^2\xi\sqrt{|\gamma|}\left[\gamma^{ij}g_{ij} - k^{-2}F'^2\right] + \frac{T}{2}\int d^2\xi\epsilon^{ij}\left(B_{ij} + B_iA_j - 2F_i'A_j\right), \tag{15.41}$$

which has exactly the same form as the original action (15.32) with the replacements Eq. (15.27), which imply the Buscher T-duality rules Eqs. (15.28) for all fields except for the dilaton, as we have explained.

The dual action has conserved currents $P^i_{z'}$, and $W^i_{z'}$, which are related to the conserved currents of the original theory by

$$P^i_{z'} = W^i_z, \qquad W^i_{z'} = P^i_z, \tag{15.42}$$

also as expected.

To gain more insight into these transformations, we are going to find the worldline actions of the string momentum modes and winding modes and see that they have the expected dependences of the masses on the compactification radius and are interchanged by T duality, as expected. The relation between these two worldline actions is similar to the relation between T-dual D-brane worldvolume actions, as we will see in Section 15.2.3.

Winding and momentum modes. We want to find the action of winding modes as seen from the $(\hat{d}-1)$-dimensional point of view. We will perform a *double dimensional reduction* of the spacetime fields (as we did to obtain Eq. (15.32)) and also of the worldvolume fields, since we assume that the worldsheet coordinate ξ^1 is compact and that we can use the KK formalism also in the worldsheet. The result will describe a particle moving in $\hat{d}-1$ dimensions and coupled to the $(\hat{d}-1)$-dimensional fields in a specific way. Its mass will identify it as a bosonic-string winding mode.

Our starting point is thus Eq. (15.32). The next step consists in using part of the gauge freedom to set

$$Z = w\ell_z/(\ell\xi^1), \tag{15.43}$$

where we take $\xi^1 \in [0, 2\pi\ell]$ and $Z \in [0, 2\pi\ell_z]$. This configuration has winding number w and, indeed, on computing the conserved charge associated with $W_z^{\,i}$, we obtain a number proportional to w. ℓ and ℓ_z can be changed at will by worldsheet and spacetime reparametrizations and the final physical results will not depend on either of them. With this normalization, the asymptotic value of the KK scalar is $k_0 = R_z/\ell_z$. All the other worldsheet fields are taken to be independent of ξ^1. We split the worldsheet metric as follows:

$$\begin{aligned}
\hat{\gamma}_{\tau\tau} &= l^2(\gamma - a^2), & \hat{\gamma}^{\tau\tau} &= l^{-2}\gamma^{-1},\\
\hat{\gamma}_{\tau\sigma} &= -l^2 a_i, & \hat{\gamma}^{\tau\sigma} &= -l^{-2}\gamma^{-1}a,\\
\hat{\gamma}_{\sigma\sigma} &= -l^2, & \hat{\gamma}^{\sigma\sigma} &= -l^{-2}(1-\gamma^{-1}a^2),
\end{aligned} \tag{15.44}$$

where γ is going to be the worldline metric after reduction.

Using this *Ansatz* in the action Eq. (15.32) and integrating ξ^1, we obtain

$$\begin{aligned}
S = -\pi\ell T \int d\xi^0 \Bigg\{ \gamma^{\frac{-1}{2}} \left[(g_{\mu\nu} - k^2 A_\mu A_\nu)\dot{X}^\mu \dot{X}^\nu + 2\frac{w\ell_z}{\ell} a k^2 A_\mu \dot{X}^\mu \right] \\
+ \left(\frac{w\ell_z}{\ell}\right)^2 \gamma^{\frac{1}{2}} (1-\gamma^{-1}a^2) k^2 \Bigg\} + 2\pi\ell_z w T \int d\xi^0 B_\mu \dot{X}^\mu,
\end{aligned} \tag{15.45}$$

and, on solving the equation for the metric component a and substituting back into the action, we obtain the worldline action of a point-particle charged with respect to the winding vector:

$$S = -\pi\ell T \int d\xi^0 \left[\gamma^{\frac{-1}{2}} g_{\mu\nu}\dot{X}^\mu \dot{X}^\nu + \left(\frac{w\ell_z}{\ell}\right)^2 \gamma^{\frac{1}{2}} k^2 \right] + 2\pi\ell_z w T \int d\xi^0 B_\mu \dot{X}^\mu. \tag{15.46}$$

To read off the mass, we eliminate the worldline metric γ using its equation of motion. The final result is the worldline action for a winding mode with winding number w:

$$\boxed{S = -2\pi |w| \ell_z k_0 T \int d\xi^0 (k/k_0) \sqrt{|g_{\mu\nu} \dot{X}^\mu \dot{X}^\nu|} + 2\pi \ell_z w T \int d\xi^0 B_\mu \dot{X}^\mu.} \tag{15.47}$$

The mass is given by

$$M = 2\pi |w| \ell_z k_0 T = 2\pi |w| \ell_z \frac{R_z}{\ell_z} \frac{1}{2\pi\alpha'} = |w| \frac{R_z}{\alpha'}, \tag{15.48}$$

and it is equal (up to the sign) to the charge with respect to the scale-invariant winding vector $\tilde{B}_\mu$. This action thus describes a winding state w, $n = 0$, $N = \tilde{N} = 1$ in Eq. (14.60).

Let us now find the action for a string momentum mode. Our starting point is again Eq. (15.32). We want now to eliminate Z using the conservation of the current P_z^i. The situation is similar to the one we found in Section 11.2.3 when we reduced the action of a massless particle moving in a space with an isometry in the direction Z using the conservation of momentum in that direction. The first step is then to replace $\partial_i Z$ by P_z^i in the action by performing a Legendre transformation of the Lagrangian $\hat{L}$ in Eq. (15.32), $\hat{S} = \int d^2\xi \sqrt{|\gamma|} \hat{L}$, with respect to Z and only then use the equation of motion.

Therefore, we take the transformed action

$$\tilde{\hat{S}}[X^\mu, P_z^i] = \int d^2\xi \sqrt{|\gamma|} \left(-P_z{}^i \partial_i Z + L\right), \tag{15.49}$$

and eliminate $\partial_i X$ (and F_i) completely by using (15.35), obtaining

$$\begin{aligned} \hat{S}'[X^\mu, P_z^i] = &-\frac{T}{2} \int d^2\xi \sqrt{|\gamma|} \gamma^{ij} \left[g_{ij} + k^{-2} T^{-2} \mathcal{F}_i \mathcal{F}_j - \frac{2}{T} A_i \mathcal{F}_j \right] \\ &+ \frac{T}{2} \int d^2\xi \, \epsilon^{ij} [B_{ij} - A_i B_j], \end{aligned} \tag{15.50}$$

where

$$\mathcal{F}_i \equiv [P_{z\,i} + T^\star B_i]. \tag{15.51}$$

Now we can use consistently the equations of motion and replace P_z^i in Eq. (15.50) by

$$P_z{}^i = \frac{1}{\sqrt{|\gamma|}} \delta^{i0} C, \tag{15.52}$$

which is automatically conserved. The constant C is fixed by noting that the conserved charge is the momentum in the direction z and is quantized in units n/ℓ_z:

$$\int d\xi^1 \sqrt{|\gamma|} P_z^0 = \frac{n}{\ell_z}, \quad \Rightarrow C = \frac{n}{2\pi \ell \ell_z}. \tag{15.53}$$

After elimination of all the components of the worldsheet metric, we arrive immediately at the worldline action of a KK mode with momentum number n Eq. (11.83), which correspond to the states n, $w = 0$, $N = \tilde{N} = 1$ in Eq. (14.60). These momentum modes are related to the winding modes whose worldline action we found before by T-duality and their worldline actions are related by Buscher's T-duality rules Eqs. (15.28).

15.2.3 T duality in the bosonic Dp-brane effective action

We have found in Section 14.3.2 how Dp-branes transform under T duality into D$(p+1)$- or D$(p-1)$-branes depending on whether the compact coordinate is transverse or parallel to the Dp-brane worldvolume. These relations between Dp-branes in spacetimes with compact directions are also realized in their worldvolume effective actions Eq. (15.14), as shown in [36, 113]. Not only do the spacetime fields transform following Buscher's rules, but also the relation between components of the BI vector and embedding coordinates is realized in them. In the type-II superstring case there are WZ terms that describe the coupling to RR forms and we need the type-II Buscher rules for them. We will study them in Chapter 17 and the T-duality relations between the worldvolume actions of super-Dp-branes in Chapter 19.

The T duality between the effective action of a D$\hat{p}$-brane ($\hat{p}=p+1$) wrapped on a compact spacetime dimension and the effective action of a Dp-brane can be established by performing the double dimensional reduction of the former and the direct dimensional reduction of the latter and showing that the resulting actions are identical. We follow [113].

Double dimensional reduction of the D$\hat{p}$*-brane effective action.* We use hats both for spacetime and for worldvolume objects, since each of them will be reduced in one dimension, parametrized, respectively, by $\hat{x}^d \equiv z \in [0, 2\pi\ell_z]$ and $\hat{\xi}^{\hat{p}} \equiv \zeta \in [0, 2\pi\ell]$. We proceed in two steps: First, we rewrite the $\hat{d}$-dimensional spacetime fields in terms of the d-dimensional ones, as in Section 15.2.2, giving

$$\hat{H}_{\hat{\imath}\hat{\jmath}} \equiv \hat{g}_{\hat{\imath}\hat{\jmath}} + \hat{B}_{\hat{\imath}\hat{\jmath}} + 2\pi\alpha'\hat{F}_{\hat{\imath}\hat{\jmath}} = g_{\hat{\imath}\hat{\jmath}} - k^2 F_{\hat{\imath}} F_{\hat{\jmath}} + B_{\hat{\imath}\hat{\jmath}} + A_{[\hat{\imath}} B_{\hat{\jmath}]} - 2F_{[\hat{\imath}} B_{\hat{\jmath}]} + 2\pi\alpha'\hat{F}_{\hat{\imath}\hat{\jmath}}, \tag{15.54}$$

where the pull-backs of $\hat{d}$-dimensional fields over the $(\hat{p}+1)$-dimensional worldvolume are computed with $\partial_{\hat{\imath}}\hat{X}^{\hat{\mu}}$, the pull-backs of d-dimensional fields with $\partial_{\hat{\imath}}X^{\mu}$, and where

$$F_{\hat{\imath}} = \partial_{\hat{\imath}} Z + A_{\hat{\imath}}. \tag{15.55}$$

Next, we make the gauge choice $Z = c\zeta$, with the remaining embedding coordinates independent of ζ. c is the constant ℓ_z/ℓ. The $(\hat{p}+1)\times(\hat{p}+1)$ matrix $\hat{H}_{\hat{\imath}\hat{\jmath}}$ defined above splits as follows:

$$\left(\hat{H}_{\hat{\imath}\hat{\jmath}}\right) = \begin{pmatrix} M_{ij} & -U_i^- \\ -U_j^+ & -c^2k^2 \end{pmatrix} = \begin{pmatrix} M_{ij} & 0 \\ 0 & -c^2k^2 \end{pmatrix} \begin{pmatrix} \mathbb{I} & -(M^{-1})_{ik}U_k^- \\ c^{-2}k^{-2}U_j^+ & 1 \end{pmatrix}, \tag{15.56}$$

where

$$M_{ij} \equiv H_{ij} - k^2 A_i A_j - A_{[i}B_{j]}, \qquad U_i^{\pm} \equiv c(k^{-1}\mathcal{F}_i \pm kA_i), \qquad \mathcal{F}_i \equiv \partial_i \hat{V}_{\underline{\zeta}} + B_i. \tag{15.57}$$

The determinant of $\hat{H}_{\hat{\imath}\hat{\jmath}}$ is the product of the determinants of the two matrices above:

$$\det(\hat{H}_{\hat{\imath}\hat{\jmath}}) = -c^2k^2 \det(M_{ij})\,[1 - U_i^+(M^{-1})_{ik}U_k^-], \tag{15.58}$$

and, on substituting this into the D$\hat{p}$-brane effective action and integrating over ζ, we obtain the effective action of a Dp-brane moving in d spacetime dimensions:

$$S = -T_{\mathrm{D}\hat{p}} 2\pi \ell |c| k_0 \int d^{p+1}\xi \, e^{-(\phi-\phi_0)} (k/k_0)^{\frac{1}{2}} [1 - U_i^{+} (M^{-1})_{ik} U_k^{-}]^{\frac{1}{2}} \sqrt{|M|}. \qquad (15.59)$$

Direct dimensional reduction of the Dp*-brane effective action.* Now we use hats and primes (indicating that we are in the T-dual situation) for the spacetime fields which we are going to reduce in the direction parametrized by z', and primes but no hats for the $(p+1)$-dimensional worldvolume fields. We split the spacetime fields as in Eq. (15.54),

$$\hat{H}'_{ij} \equiv \hat{g}'_{ij} + \hat{B}'_{ij} + 2\pi\alpha' F'_{ij} = M'_{ij} - U_i^{-\prime} U_j^{+\prime}, \qquad (15.60)$$

where

$$M'_{ij} \equiv H_{ij} - k'^{-2} B'_i B'_j - B'_{[i} A'_{j]}, \qquad U_i^{\pm\prime} \equiv (k' \mathcal{F}'_i \pm k'^{-1} B'_i), \qquad \mathcal{F}'_i \equiv \partial_i Z' + A'_i. \qquad (15.61)$$

M'_{ij}, $U_i^{\pm\prime}$, and $\mathcal{F}_i^{i}$ are exactly the Buscher T duals of the unprimed ones plus the relation

$$Z' = \hat{V}_{\underline{\zeta}}. \qquad (15.62)$$

Now, it only remains to see that the action we obtain after this reduction is equivalent to Eq. (15.59). First, we rewrite $\hat{H}'_{ij}$ as the product of two matrices,

$$\hat{H}'_{ij} = M'_{ik} [\delta_{kj} - M'^{-1}_{ik} U_k^{-\prime} U_j^{+\prime}], \qquad (15.63)$$

the second of which has p times the eigenvalue $+1$ (for each of the p vectors orthogonal to $U^{+\prime}$) and one time the eigenvalue $1 - U_i^{+\prime} M'^{-1}_{ij} U_j^{-\prime}$ for the eigenvector $M'^{-1} U^{-\prime}$, and

$$\det(\hat{H}'_{ij}) = \det(M'_{ij}) \, [1 - U_i^{+\prime} (M'^{-1})_{ik} U_k^{-\prime}]. \qquad (15.64)$$

Writing $e^{-(\hat{\phi}'-\hat{\phi}'_0)} = e^{-(\phi'-\phi'_0)} (k'/k'_0)^{-\frac{1}{2}}$, the action takes the form

$$S = -T'_{\mathrm{D}p} \int d^{p+1}\xi \, e^{-(\phi-\phi_0)} (k'/k'_0)^{-\frac{1}{2}} [1 - U_i^{+\prime} (M'^{-1})_{ik} U_k^{-\prime}]^{\frac{1}{2}} \sqrt{|M'|}. \qquad (15.65)$$

This action is identical to Eq. (15.59) after application of Buscher's rules if we identify

$$T'_{\mathrm{D}p} = T_{\mathrm{D}(p+1)} 2\pi \ell |c| k_0^{\frac{1}{2}} = 2\pi R_z T_{p+1}. \qquad (15.66)$$

The Dp-brane tension T_p should be essentially independent of the compactification radius (since it takes the same value in uncompactified spacetimes). It can depend solely on ℓ_{s} and g. Since the string coupling constants of T-dual theories are related by Eqs. (14.61) and (14.62) and since it has units of mass by p-volume, for all p

$$T_{\mathrm{D}p} \sim \frac{1}{\ell_{\mathrm{s}}^{p+1} g}. \qquad (15.67)$$

We will see later on other methods by which to find the same result and the proportionality constant. The g^{-1} dependence of the tension of these objects gives them unique status, intermediate between standard solitons, whose mass is proportional to g^{-2}, and the *fundamental* objects that appear in the perturbative spectrum, with masses (tensions) independent of g, such as the *fundamental string* whose action is the string σ-model action and whose tension is $T = 1/(2\pi\ell_s^2)$.

15.3 Example: the fundamental string (F1)

Many string solutions (i.e. solutions of the string effective action Eq. (15.1)) are known. For instance, all the vacuum Einstein solutions are string solutions with constant dilaton and pure gauge KR 2-form and, for each of them that admits an isometry, we can find a T-dual string solution (possibly with non-trivial dilaton and KR 2-form). To investigate T duality, however, we must choose a convenient solution to which we can give a physical and stringy interpretation, in the same spirit as that in which we chose a gravitational wave in Section 11.2.3 to illustrate KK reduction, four-dimensional electric–magnetic duality and KK oxidation. The so-called *fundamental string* solution [281, 282] (see also [337]) denoted by F1 represents a string at rest and can play this role.

We can understand the F1 solution as a solution of the ("bulk-plus-brane") action that results from the addition of the string effective action Eq. (15.1) to the string σ-model action Eq. (15.31) which, denoted from now on by S_{F1}, acts as a singular one-dimensional source for the former. The equations of motion of the spacetime fields[10] are Eqs. (15.11) plus the source terms

$$\begin{aligned}
\frac{16\pi G_{\text{N}}^{(\hat{d})} e^{2(\hat{\phi}-\hat{\phi}_0)}}{\sqrt{|\hat{g}|}} \frac{\delta \hat{S}_{\text{F1}}}{\delta \hat{g}^{\hat{\mu}\hat{\nu}}} &= +\frac{8\pi G_{\text{N}}^{(\hat{d})} e^{2(\hat{\phi}-\hat{\phi}_0)} T}{\sqrt{|\hat{g}|}} \int d^2\xi \sqrt{|\gamma|}\gamma^{ij} \hat{g}_{i\hat{\mu}} \hat{g}_{j\hat{\nu}} \delta^{(\hat{d})}(\hat{x} - \hat{X}), \\
\frac{16\pi G_{\text{N}}^{(\hat{d})} e^{2(\hat{\phi}-\hat{\phi}_0)}}{\sqrt{|\hat{g}|}} \frac{\delta \hat{S}_{\text{F1}}}{\delta \hat{\phi}} &= 0, \\
\frac{16\pi G_{\text{N}}^{(\hat{d})} e^{2(\hat{\phi}-\hat{\phi}_0)}}{\sqrt{|\hat{g}|}} \frac{\delta \hat{S}_{\text{F1}}}{\delta \hat{B}_{\hat{\mu}\hat{\nu}}} &= +\frac{8\pi G_{\text{N}}^{(\hat{d})} e^{2(\hat{\phi}-\hat{\phi}_0)} T}{\sqrt{|\hat{g}|}} \int d^2\xi \epsilon^{ij} \partial_i \hat{X}^{\hat{\mu}} \partial_j \hat{X}^{\hat{\nu}} \delta^{(\hat{d})}(\hat{x} - \hat{X}).
\end{aligned} \tag{15.68}$$

We also have to solve the equations of motion of the worldvolume fields,

$$\begin{aligned}
-\frac{2}{T\sqrt{|\gamma|}} \frac{\delta \hat{S}_{\text{F1}}}{\delta \gamma^{ij}} &= \hat{g}_{ij} - \tfrac{1}{2}\gamma_{ij}\hat{g}^{k}{}_{k} = 0, \\
\frac{1}{T\sqrt{|\gamma|}} \frac{\delta \hat{S}_{\text{F1}}}{\delta \hat{X}^{\hat{\mu}}} &= \hat{g}_{\hat{\mu}\hat{\nu}} \left[\nabla^2 \hat{X}^{\hat{\nu}} + \gamma^{ij} \hat{\Gamma}_{ij}{}^{\hat{\nu}}\right] + \frac{\epsilon^{ij}}{2\sqrt{|\gamma|}} \hat{H}_{\hat{\mu}ij} = 0.
\end{aligned} \tag{15.69}$$

We work in the *static gauge*, identifying the worldvolume coordinates with the first spacetime coordinates $\hat{X}^i = \xi^i \equiv (T, Y)$, $i = 0, 1$. The remaining spacetime coordinates are transverse to the string worldvolume and we make for them the *Ansatz* $X^m(\xi) = 0$, $m = 1, \ldots, \hat{d} - 2$. If the solution is to describe a fundamental string at rest, it is natural to make an *Ansatz* for the metric with Poincaré symmetry in the worldvolume directions. On the other hand, our experience with ERN BHs tells us that a full solution of the equations with sources can be expected only when there is supersymmetry/extremality and the solution depends solely on a reduced number of functions that are harmonic in transverse

[10] We add hats to all $\hat{d}$-dimensional fields.

space, whose singularities are associated with the sources. In this case, the solution should depend on just one function, $H_{\mathrm{F1}}(x^m)$.

A solution of all the equations that satisfy the above criteria and *Ansatz* is the *fundamental-string solution* given, for $\hat{d} \geq 5$, by

$$\boxed{\begin{aligned} d\hat{s}^2 &= H_{\mathrm{F1}}^{-1}[dt^2 - dy^2] - d\vec{x}^{\,2}_{(\hat{d}-2)}, \\ \hat{B}_{t\underline{y}} &= -\big(H_{\mathrm{F1}}^{-1} - 1\big), \\ e^{-2(\hat{\phi}-\hat{\phi}_0)} &= H_{\mathrm{F1}}, \\ H_{\mathrm{F1}} &= \varepsilon + \frac{h_{\mathrm{F1}}}{|\vec{x}_{(\hat{d}-2)}|^{\hat{d}-4}}, \qquad h_{\mathrm{F1}} = \frac{16\pi G_{\mathrm{N}}^{(\hat{d})} T}{(\hat{d}-4)\omega_{(\hat{d}-3)}}. \end{aligned}} \tag{15.70}$$

The integration constant h_{F1} is completely determined by the source and the value of the Newton constant, and ε has to be taken equal to 1 in order to have asymptotic flatness.

It is reasonable to assume that this solution describes a string lying at rest in the direction of the isometric coordinate y. Let us now take y to be compact, rescaling it by $y \to k_0 y$, where $k_0 = R_y/\ell_{\mathrm{s}}$, R_y being the compactification radius as usual and $y \in [0, 2\pi\ell_{\mathrm{s}}]$. The dimensionally reduced solution has the non-vanishing fields

$$\boxed{\begin{aligned} ds^2 &= H_{\mathrm{F1}}^{-1} dt^2 - d\vec{x}^{\,2}_{(d-1)}, \\ B_t &= -k_0\big(H_{\mathrm{F1}}^{-1} - 1\big), \\ e^{-2(\phi-\phi_0)} &= H_{\mathrm{F1}}^{\frac{1}{2}}, \qquad k = k_0 H_{\mathrm{F1}}^{-\frac{1}{2}}, \end{aligned}} \tag{15.71}$$

with the same H_{F1}. This is the metric of a point-like object with mass and charged with respect to the winding vector B_μ, just as corresponds to a string winding mode.[11] The T dual is charged with respect to the KK vector A_μ and the $\hat{d}$-dimensional T dual is the purely gravitational solution

$$\boxed{\begin{aligned} d\hat{s}'^2 &= H_{\mathrm{F1}}^{-1} dt^2 - k_0^{-2}[dy' + k_0(H_{\mathrm{F1}}^{-1} - 1)dt] - d\vec{x}^{\,2}_{(\hat{d}-2)}, \\ e^{-2\hat{\phi}'} &= e^{-2\hat{\phi}'} = e^{-2\hat{\phi}_0} k_0^2, \end{aligned}} \tag{15.72}$$

which is the zero mode of the shock wave solution Eq. (10.41) that can be shown to represent not just a point-particle, but also a fundamental string moving in the compact coordinate y', as expected.

[11] To compare this with the masses predicted, we need the string value of $G_{\mathrm{N}}^{(\hat{d})}$.

16
From eleven to four dimensions

In the previous chapter we started our study of string dualities in the effective action by treating the simplest case: T duality in the string common sector. Now we are ready to handle more complicated cases: type-II T duality, type-IIB S duality, heterotic/type-I string duality and the strong-coupling limit of the type-IIA superstring. Actually, only the first of these dualities (the only one which is perturbative) was known from the worldsheet point of view; the rest were conjectured after they had been observed in the corresponding effective actions and were interpreted in string language.

For instance, it was well known that the $N = 2A, d = 10$ supergravity theory can be obtained from $N = 1, d = 11$ supergravity [264] by dimensional reduction (i.e. compactifying on a circle and ignoring all the massive Kaluza–Klein modes). This reduction was first performed in the Einstein frame. The reduction in the string frame [125, 962] gave new and useful information. To go to the string frame, it is necessary to identify the dilaton, which turns out to be essentially the moduli field that measures the radius of the circle in the 11th dimension, namely the KK scalar $\hat{\hat{g}}_{\underline{xx}}$. Since the dilaton is the string coupling constant, the strong-coupling limit of the type-IIA string theory corresponds to the limit of decompactification (large radius) of the 11th dimension. The surprising fact is that this statement is true including the massive Kaluza–Klein modes and string modes if one also includes the *solitonic modes* which appear in the non-perturbative spectrum of the string theory. These non-perturbative states at strong coupling should be identified with the ordinary Kaluza–Klein modes of 11-dimensional supergravity.

This relation between field theories is in agreement with the relation between the worldvolume action of the M2-brane wrapped on a compact dimension and that of the GS type-IIA superstring theory [14, 335] which give, in a certain sense, the 11-dimensional and $N = 2A, d = 10$ supergravity theories.

The other dualities form a chain that relates all string theories (at least under certain circumstances) to 11-dimensional supergravity. There has been conjectured the existence of the so-called *M theory* whose low-energy limit would be described by 11-dimensional supergravity which, in different limits, would give all the ten-dimensional string theories which would be different, dual, manifestations of the same unique theory.[1]

Our interest is mainly in four-dimensional string effective-field theories and classical solutions and their connection to higher-dimensional theories and solutions. It is then natural to start by introducing 11-dimensional supergravity and then performing the reduction on a circle to find the type-IIA superstring effective action. We will do this in Section 16.1. Since we are interested in classical solutions, we will study only the bosonic sectors of these theories. However, we will also need the supersymmetry transformation rules for the fermions in order to study their unbroken supersymmetries.

To study T duality between the effective actions of the type-IIA and -IIB theories, following the philosophy of Section 15.2, we will have to perform dimensional reduction of both theories to nine dimensions. The reduction of the IIA theory will be done in Section 16.3 whereas the reduction of the IIB theory will be postponed to the next chapter, in which we will find the type-II Buscher rules.

The $E_8 \times E_8$ heterotic string theory can be obtained by compactification of M theory on a segment (the simplest orbifold), each E_8 factor group living on one of the ten-dimensional boundaries. From the point of view of effective actions, we can easily obtain the heterotic string effective action (without the gauge fields) by compactifying 11-dimensional supergravity on an orbifold, which amounts to the S^1 compactification which we carry out in Section 16.1 followed by a truncation that we study in Section 16.4. A similar truncation of the type-IIB theory that gives the type-I theory (again without the gauge fields) will be studied in the next chapter and corresponds to the O9 plus 32 D9 construction of the type-I SO(32) theory.

Further compactification increases the number of dualities: on the one hand, one can perform T dualities in more directions that can also be rotated into each other. Also, in even dimensions, new dualities that involve the Hodge-dualization of differential-form potentials appear: in $d = 4$, vectors can be dualized, in $d = 6$ 2-forms, and in $d = 8$ 3-forms. Furthermore, in odd dimensions, Hodge-dualization of differential forms can increase the number of vector fields that can be rotated into other vector fields, enhancing the duality group. Usually these dualities are manifest only in the Einstein frame and were known as *hidden symmetries* of supergravity theories. In Section 16.5 we are going to study the toroidal compactification of $N = 1, d = 10$ supergravity, the effective theory of the heterotic string down to $d = 4$, as an example and we will find that, generically, the classical duality group[2] is $O(n, n + 16)$ for compactification on an n-torus, all of it due to T duality, but, in $d = 4$, vectors can be dualized into vectors and the symmetry is increased by the S-duality group $SL(2, \mathbb{R})$. (The duality groups that appear in toroidal compactifications of $N = 2, d = 10$ theories are given in Table 16.1.)

Finally, we are going to study the preservation of unbroken supersymmetry under duality transformations in Section 16.6.

[1] A 12-dimensional origin for M theory and type-IIB superstring theory by the name of F theory has also been suggested.

[2] Quantum effects such as charge quantization break the classical supergravity duality groups to discrete subgroups, typically the ones obtained by restricting the matrix entries to taking integer values [583]. On the other hand, if G is the classical duality group, the scalars parametrize a coset space G/H, where H is the maximal compact subgroup of H. For the heterotic-string case $H = O(n) \times O(n + 16)$.

Table 16.1. Hidden symmetries of toroidally compactified $N = 2, d = 10$ supergravities [261, 608]. $E_{3(+3)} = SL(3, \mathbb{R}) \times SL(2, \mathbb{R})$, $E_{4(+4)} = SL(5, \mathbb{R})$, and $E_{5(+5)} = SO(5, 5)$. The tilded numbers indicate that the corresponding fields are dualized by some of the duality transformations which will merely be symmetries of the equations of motion. The discrete subgroups are the U duality groups.

d	G	H	$e^a{}_\mu$	$C_{\mu\nu\rho}$	$B_{\mu\nu}$	A_μ	φ	ψ_μ	λ
9	$GL(2, \mathbb{R})$	SO(2)	1	1	2	3	3	2	4
8	$E_{3(+3)}$	SO(3) × SO(2)	1	$\tilde{1}$	3	6	7	2	6
7	$E_{4(+4)}$	SO(5)	1	0	5	10	14	4	16
6	$E_{5(+5)}$	SO(5) × SO(5)	1	0	$\tilde{5}$	16	25	4	20
5	$E_{6(+6)}$	USp(8)	1	0	0	27	42	8	48
4	$E_{7(+7)}$	SU(8)	1	0	0	$\widetilde{28}$	70	8	56
3	$E_{8(+8)}$	SO(16)	1	0	0	0	128	16	128

A scheme of the dimensional reductions and truncations that we are going to study in this chapter and the next is given in Figure 16.1. It is not a diagram of the web of dualities that relates string theories, although it is closely related to it. A general reference for this and the next chapter is [828], where much information and most original references to supergravity theories can be found.

16.1 Dimensional reduction from $d = 11$ to $d = 10$

Here we are going to obtain the bosonic sector of $N = 2A, d = 10$ (also known as type-IIA) supergravity and the supersymmetry transformation rules by straightforward dimensional reduction of $N = 1, d = 11$ supergravity using the techniques developed in [836]. This dimensional reduction has been performed in [585], but now we will obtain the type-IIA theory directly in the string frame.

We are going to describe the procedure used to perform the dimensional reduction in some detail since our goal is to relate the various supergravity theories in different dimensions. Throughout this and the next few sections we will use double hats for 11-dimensional objects, single hats for ten-dimensional objects, and no hats for nine-dimensional objects. We first introduce the theory of 11-dimensional supergravity.

16.1.1 11-dimensional supergravity

The fields of $N = 1, d = 11$ supergravity [264] are the Elfbein, a three-form potential, and a 32-component Majorana gravitino,[3]

$$\left\{\hat{\hat{e}}^{\hat{\hat{a}}}{}_{\hat{\hat{\mu}}},\ \hat{\hat{C}}_{\hat{\hat{\mu}}\hat{\hat{\nu}}\hat{\hat{\rho}}},\ \hat{\hat{\psi}}_{\hat{\hat{\mu}}}\right\}. \tag{16.1}$$

[3] Our conventions for 11-dimensional gamma matrices and spinors are given in Appendix B.1.3 and are essentially identical to those of the original reference [264] except for the relation between $\hat{\hat{\Gamma}}^{10}$, the totally antisymmetric tensor, and the remaining ten gamma matrices, which is not explicitly given in that reference. In our case, that relation is given in Eq. (B.71), and in the supersymmetry transformation rules below the sign of the topological term in the action is the opposite to that in [264]. Observe also that our spin connection and contorsion have opposite signs, though, and that we have set the constant $K = \frac{1}{2}$.

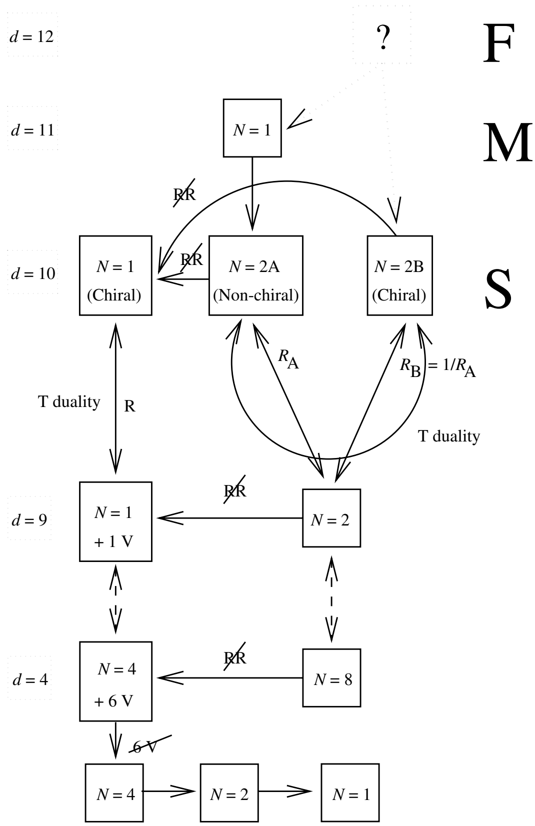

Fig. 16.1. A directory of supergravities. The relations between various supergravity theories upon compactification on circles (lines with two arrowheads) and truncation (lines with a single arrowhead) are schematically represented here. The letters F, M, and S indicate that the corresponding supergravity theory is the low-energy limit of F theory, M theory, or a superstring theory.

The action for these fields is

$$\boxed{\begin{aligned}\hat{\hat{S}} = \frac{1}{16\pi G_{\mathrm{N}}^{(11)}} \int d^{11}\hat{\hat{x}}\, \sqrt{|\hat{\hat{g}}|}\, \Bigg[& \hat{\hat{R}}(\hat{\omega}) - \frac{i}{2}\bar{\hat{\hat{\psi}}}_{\hat{\hat{\mu}}}\hat{\hat{\Gamma}}^{\hat{\hat{\mu}}\hat{\hat{\nu}}\hat{\hat{\rho}}}\nabla_{\hat{\hat{\nu}}}\left(\frac{\hat{\hat{\omega}}+\tilde{\hat{\hat{\omega}}}}{2}\right)\hat{\hat{\psi}}_{\hat{\hat{\rho}}} - \frac{1}{2\cdot 4!}\hat{\hat{G}}^{2} \\ & + \frac{1}{192}\bar{\hat{\hat{\psi}}}_{\hat{\hat{\mu}}}\,\hat{\hat{\Gamma}}^{[\hat{\hat{\mu}}}\hat{\hat{\Gamma}}_{\hat{\hat{\alpha}}\hat{\hat{\beta}}\hat{\hat{\gamma}}\hat{\hat{\delta}}}\hat{\hat{\Gamma}}^{\hat{\hat{\nu}}]}\,\hat{\hat{\psi}}_{\hat{\hat{\nu}}}\left(\hat{\hat{G}}^{\hat{\hat{\alpha}}\hat{\hat{\beta}}\hat{\hat{\gamma}}\hat{\hat{\delta}}} + \tilde{\hat{\hat{G}}}^{\hat{\hat{\alpha}}\hat{\hat{\beta}}\hat{\hat{\gamma}}\hat{\hat{\delta}}}\right) - \frac{1}{(144)^{2}}\frac{1}{\sqrt{|\hat{\hat{g}}|}}\hat{\hat{\epsilon}}\hat{\hat{G}}\hat{\hat{G}}\hat{\hat{C}}\Bigg].\end{aligned}} \tag{16.2}$$

Let us now describe each object in this action:

$$\hat{\hat{G}} = 4\partial\hat{\hat{C}} \tag{16.3}$$

is the field strength of the 3-form and is obviously invariant under the gauge transformations

$$\delta_{\hat{\hat{\chi}}} \hat{\hat{C}} = 3\partial \hat{\hat{\chi}}, \tag{16.4}$$

where $\hat{\hat{\chi}}$ is a 2-form;

$$\tilde{\hat{\hat{G}}}_{\hat{\hat{\mu}}\hat{\hat{\nu}}\hat{\hat{\rho}}\hat{\hat{\sigma}}} = \hat{\hat{G}}_{\hat{\hat{\mu}}\hat{\hat{\nu}}\hat{\hat{\rho}}\hat{\hat{\sigma}}} - \tfrac{3}{2}\bar{\hat{\hat{\psi}}}_{[\hat{\hat{\mu}}}\hat{\hat{\Gamma}}_{\hat{\hat{\nu}}\hat{\hat{\rho}}}\hat{\hat{\psi}}_{\hat{\hat{\sigma}}]} \tag{16.5}$$

is the supercovariant field strength; and

$$\nabla_{\hat{\hat{\mu}}}\left(\hat{\hat{\omega}}\right)\hat{\hat{\psi}}_{\hat{\hat{\nu}}} = \partial_{\hat{\hat{\mu}}}\hat{\hat{\psi}}_{\hat{\hat{\nu}}} - \tfrac{1}{4}\hat{\hat{\omega}}_{\hat{\hat{\mu}}}{}^{\hat{\hat{a}}\hat{\hat{b}}}\hat{\hat{\Gamma}}_{\hat{\hat{a}}\hat{\hat{b}}}\hat{\hat{\psi}}_{\hat{\hat{\nu}}} \tag{16.6}$$

is the covariant derivative with

$$\begin{aligned}
\tilde{\hat{\hat{\omega}}}_{\hat{\hat{\mu}}}{}^{\hat{\hat{a}}\hat{\hat{b}}} &= \hat{\hat{\omega}}_{\hat{\hat{\mu}}}{}^{\hat{\hat{a}}\hat{\hat{b}}} - \frac{i}{16}\bar{\hat{\hat{\psi}}}_{\hat{\hat{\alpha}}}\,\hat{\hat{\Gamma}}_{\hat{\hat{\mu}}}{}^{\hat{\hat{a}}\hat{\hat{b}}\hat{\hat{\alpha}}\hat{\hat{\beta}}}\hat{\hat{\psi}}_{\hat{\hat{\beta}}}, \\
\hat{\hat{\omega}}_{\hat{\hat{\mu}}}{}^{\hat{\hat{a}}\hat{\hat{b}}} &= \hat{\hat{\omega}}_{\hat{\hat{\mu}}}{}^{\hat{\hat{a}}\hat{\hat{b}}}\left(\hat{\hat{e}}\right) + \hat{\hat{K}}_{\hat{\hat{\mu}}}{}^{\hat{\hat{a}}\hat{\hat{b}}}, \\
\hat{\hat{K}}_{\hat{\hat{\mu}}}{}^{\hat{\hat{a}}\hat{\hat{b}}} &= -\frac{i}{16}\left[-\bar{\hat{\hat{\psi}}}_{\hat{\hat{\alpha}}}\,\hat{\hat{\Gamma}}_{\hat{\hat{\mu}}}{}^{\hat{\hat{a}}\hat{\hat{b}}\hat{\hat{\alpha}}\hat{\hat{\beta}}}\,\hat{\hat{\psi}}_{\hat{\hat{\beta}}} + 2\left(\bar{\hat{\hat{\psi}}}_{\hat{\hat{\mu}}}\,\hat{\hat{\Gamma}}^{\hat{\hat{b}}}\,\hat{\hat{\psi}}^{\hat{\hat{a}}} - \bar{\hat{\hat{\psi}}}_{\hat{\hat{\mu}}}\,\hat{\hat{\Gamma}}^{\hat{\hat{a}}}\,\hat{\hat{\psi}}^{\hat{\hat{b}}} + \bar{\hat{\hat{\psi}}}_{\hat{\hat{b}}}\,\hat{\hat{\Gamma}}_{\hat{\hat{\mu}}}\,\hat{\hat{\psi}}^{\hat{\hat{a}}}\right)\right].
\end{aligned} \tag{16.7}$$

This action is invariant under the local supersymmetry transformations with parameter $\hat{\hat{\epsilon}}$ (a Majorana spinor):

$$\boxed{\begin{aligned}
\delta_{\hat{\hat{\epsilon}}}\hat{\hat{e}}^{\hat{\hat{a}}}{}_{\hat{\hat{\mu}}} &= -\frac{i}{2}\bar{\hat{\hat{\epsilon}}}\,\hat{\hat{\Gamma}}^{\hat{\hat{a}}}\,\hat{\hat{\psi}}_{\hat{\hat{\mu}}}, \\
\delta_{\hat{\hat{\epsilon}}}\hat{\hat{\psi}}_{\hat{\hat{\mu}}} &= 2\tilde{\nabla}_{\hat{\hat{\mu}}}\hat{\hat{\epsilon}} = 2\nabla_{\hat{\hat{\mu}}}\left(\tilde{\hat{\hat{\omega}}}\right)\hat{\hat{\epsilon}} + \frac{i}{144}\left(\hat{\hat{\Gamma}}^{\hat{\hat{\alpha}}\hat{\hat{\beta}}\hat{\hat{\gamma}}\hat{\hat{\delta}}}{}_{\hat{\hat{\mu}}} - 8\hat{\hat{\Gamma}}^{\hat{\hat{\beta}}\hat{\hat{\gamma}}\hat{\hat{\delta}}}\hat{\hat{\eta}}_{\hat{\hat{\mu}}}{}^{\hat{\hat{\alpha}}}\right)\hat{\hat{\epsilon}}\,\tilde{\hat{\hat{G}}}_{\hat{\hat{\alpha}}\hat{\hat{\beta}}\hat{\hat{\gamma}}\hat{\hat{\delta}}}, \\
\delta_{\hat{\hat{\epsilon}}}\hat{\hat{C}}_{\hat{\hat{\mu}}\hat{\hat{\nu}}\hat{\hat{\rho}}} &= \tfrac{3}{2}\bar{\hat{\hat{\epsilon}}}\,\hat{\hat{\Gamma}}_{[\hat{\hat{\mu}}\hat{\hat{\nu}}}\,\hat{\hat{\psi}}_{\hat{\hat{\rho}}]}.
\end{aligned}} \tag{16.8}$$

Observe that the "topological" *Chern–Simons* term in the action $\hat{\hat{\epsilon}}\hat{\hat{G}}\hat{\hat{G}}\hat{\hat{C}}$ seems to break parity ($\hat{\hat{\epsilon}}$ is a tensor density, or pseudotensor). Parity is, however, preserved because the 3-form $\hat{\hat{C}}$ is also a pseudotensor and so transforms with an extra sign under reflections.[4]

[4] The sign of the topological term can be changed by a field redefinition $\hat{\hat{C}} \to -\hat{\hat{C}}$. However, as we will see in Chapter 17, with our conventions, to establish T duality between the type-IIA theory that we will obtain by dimensional reduction to ten dimensions and the conventional type-IIB theory with *self-dual* 5-form and positive-chirality gravitinos, taking into account everything (i.e. not just the bosonic sector), we are forced to take the negative sign. The type-IIA theory that one would obtain with positive sign seems to be related to the unconventional type-IIB theory with *anti-self-dual* 5-form and negative-chirality gravitinos. In other words; there are *two* "different" 11-dimensional supergravity theories, which differ in the sign of the topological term and are related by the (rather trivial) duality transformation $\hat{\hat{C}} \to -\hat{\hat{C}}$. Each of these theories gives rise to a different type-IIA theory (type IIA$_{+-}$ and type IIA$_{-+}$), which in turn are related by T duality to the two "different" $N = 2B_+$ and $N = 2B_-$ supergravity theories. We mentioned all these theories in Section 14.2.2.

Now we are going to perform the dimensional reduction. For many purposes it is enough to reduce the bosonic fields by setting to zero all fermions in the action and also reducing the supersymmetry transformation laws, and this is what we are going to do.

16.1.2 Reduction of the bosonic sector

The action for the bosonic fields is

$$\boxed{\hat{\hat{S}} = \frac{1}{16\pi G_{\mathrm{N}}^{(11)}} \int d^{11}\hat{\hat{x}}\, \sqrt{|\hat{\hat{g}}|}\left[\hat{\hat{R}} - \frac{1}{2\cdot 4!}\hat{\hat{G}}^{2} - \frac{1}{(144)^2}\frac{1}{\sqrt{|\hat{\hat{g}}|}}\hat{\hat{\epsilon}}\hat{\hat{G}}\hat{\hat{G}}\hat{\hat{C}}\right],} \tag{16.9}$$

and the equations of motion are

$$\begin{aligned} &\hat{\hat{R}}_{\hat{\hat{\mu}}\hat{\hat{\nu}}} - \frac{1}{12}\left[\hat{\hat{G}}_{\hat{\hat{\mu}}}{}^{\hat{\hat{\alpha}}_1\hat{\hat{\alpha}}_2\hat{\hat{\alpha}}_3}\hat{\hat{G}}_{\hat{\hat{\nu}}\hat{\hat{\alpha}}_1\hat{\hat{\alpha}}_2\hat{\hat{\alpha}}_3} - \frac{1}{12}\hat{\hat{g}}_{\hat{\hat{\mu}}\hat{\hat{\nu}}}\hat{\hat{G}}^{2}\right] = 0, \\ &\nabla_{\hat{\hat{\mu}}}\hat{\hat{G}}^{\hat{\hat{\mu}}\hat{\hat{\nu}}\hat{\hat{\rho}}\hat{\hat{\sigma}}} - \frac{1}{3^2\cdot 2^7}\frac{1}{\sqrt{|\hat{\hat{g}}|}}\hat{\hat{\epsilon}}^{\hat{\hat{\nu}}\hat{\hat{\rho}}\hat{\hat{\sigma}}\hat{\hat{\mu}}_1\cdots\hat{\hat{\mu}}_4\hat{\hat{\nu}}_1\cdots\hat{\hat{\nu}}_4}\hat{\hat{G}}_{\hat{\hat{\mu}}_1\cdots\hat{\hat{\mu}}_4}\hat{\hat{G}}_{\hat{\hat{\nu}}_1\cdots\hat{\hat{\nu}}_4} = 0. \end{aligned} \tag{16.10}$$

The equation of motion of the 3-form potential can be rewritten in this way:

$$\partial\left({}^{\star}\hat{\hat{G}} + \frac{35}{2}\hat{\hat{C}}\hat{\hat{G}}\right) = 0. \tag{16.11}$$

This equation has the form of a Bianchi identity and we could identify the expression in parentheses with $7\partial\hat{\hat{\tilde{C}}}$ where $\hat{\hat{\tilde{C}}}$ is a 6-form potential that is the dual of the 3-form potential. This implies that the field strength of the dual 6-form is [22, 119, 136]

$${}^{\star}\hat{\hat{G}} = 7(\partial\hat{\hat{\tilde{C}}} - 10\hat{\hat{C}}\partial\hat{\hat{C}}) \equiv \hat{\hat{\tilde{G}}}. \tag{16.12}$$

This field strength is obviously invariant under the gauge transformations

$$\delta_{\hat{\hat{\tilde{\chi}}}}\hat{\hat{\tilde{C}}} = 6\partial\hat{\hat{\tilde{\chi}}}, \tag{16.13}$$

where $\hat{\hat{\tilde{\chi}}}$ is a 5-form. However, in its definition the 3-form appears explicitly and, to make it invariant under the 3-form gauge transformations (16.4), $\hat{\hat{\tilde{C}}}$ has to transform as follows:

$$\delta_{\hat{\hat{\chi}}}\hat{\hat{\tilde{C}}} = -30\partial\hat{\hat{\chi}}\hat{\hat{C}}. \tag{16.14}$$

This procedure for defining the dual of some field in which the original field has not been completely eliminated, by making use of the equations of motion, is usually referred to as "on-shell" dualization. The fact that the 3-form potential appears in the field strength of its dual 6-form makes it very difficult (but not impossible, see [136]) to find a formulation of 11-dimensional supergravity in terms of the 6-form alone. On-shell dualization is enough for determining the gauge-transformation laws and a field strength in which the 3-form should be interpreted as a complicated function of the dual 6-form.

We now focus on the dimensional reduction of the theory in its 3-form formulation. We assume that all fields are independent of the coordinate $z = x^{\underline{10}}$ which we choose to correspond to a spacelike direction ($\hat{\hat{\eta}}_{\underline{zz}} = -1$) and we rewrite the fields and action in a ten-dimensional form. The dimensional reduction of the metric gives rise to the ten-dimensional metric, a vector field, and a scalar (the dilaton), whereas the dimensional reduction of the 3-form potential gives rise to a ten-dimensional 3-form and a (KR) 2-form, which are the fields of the ten-dimensional $N = 2A$, $d = 10$ supergravity theory

$$\left\{\hat{g}_{\hat{\mu}\hat{\nu}},\, \hat{B}_{\hat{\mu}\hat{\nu}},\, \hat{\phi},\, \hat{C}^{(3)}{}_{\hat{\mu}\hat{\nu}\hat{\rho}},\, \hat{C}^{(1)}{}_{\hat{\mu}},\right\}. \tag{16.15}$$

The metric, the KR two-form, and the dilaton are NSNS fields and the 3-form and the vector are RR fields. We are going to use for RR forms the conventions proposed and used in [112, 136, 470, 691], which we will explain later. Furthermore, we are going to use the string metric, to which the NSNS sector couples as in the bosonic-string effective action Eq. (15.1). The coupling to the RR sector is then completely determined just by defining RR fields with "natural" gauge-transformation rules (with no scalars involved in them). On the other hand, we have chosen for the RR fields the simplest normalization, which is also the most common: they appear at the same order in α' as the NSNS ones.[5]

Using the Scherk–Schwarz procedure as explained in Chapter 11 and rescaling the metric so as to obtain the action in the string frame, we find that the fields of the 11-dimensional theory have to be expressed in terms of the ten-dimensional ones as follows:

$$\begin{aligned}
\hat{\hat{g}}_{\hat{\mu}\hat{\nu}} &= e^{-\frac{2}{3}\hat{\phi}}\hat{g}_{\hat{\mu}\hat{\nu}} - e^{\frac{4}{3}\hat{\phi}}\hat{C}^{(1)}{}_{\hat{\mu}}\hat{C}^{(1)}{}_{\hat{\nu}}, \qquad & \hat{\hat{C}}_{\hat{\mu}\hat{\nu}\hat{\rho}} &= \hat{C}^{(3)}{}_{\hat{\mu}\hat{\nu}\hat{\rho}},\\
\hat{\hat{g}}_{\hat{\mu}\underline{z}} &= -e^{\frac{4}{3}\hat{\phi}}\hat{C}^{(1)}{}_{\hat{\mu}}, & \hat{\hat{C}}_{\hat{\mu}\hat{\nu}\underline{z}} &= \hat{B}_{\hat{\mu}\hat{\nu}},\\
\hat{\hat{g}}_{\underline{zz}} &= -e^{\frac{4}{3}\hat{\phi}}. & &
\end{aligned} \tag{16.16}$$

For the Elfbeins we have

$$\left(\hat{\hat{e}}^{\hat{\hat{a}}}{}_{\hat{\hat{\mu}}}\right) = \begin{pmatrix} e^{-\frac{1}{3}\hat{\phi}}\hat{e}^{\hat{a}}{}_{\hat{\mu}} & e^{\frac{2}{3}\hat{\phi}}\hat{C}^{(1)}{}_{\hat{\mu}} \\ 0 & e^{\frac{2}{3}\hat{\phi}} \end{pmatrix}, \qquad \left(\hat{\hat{e}}_{\hat{\hat{a}}}{}^{\hat{\hat{\mu}}}\right) = \begin{pmatrix} e^{\frac{1}{3}\hat{\phi}}\hat{e}_{\hat{a}}{}^{\hat{\mu}} & -e^{\frac{1}{3}\hat{\phi}}\hat{C}^{(1)}{}_{\hat{a}} \\ 0 & e^{-\frac{2}{3}\hat{\phi}} \end{pmatrix}. \tag{16.17}$$

[5] This is not the only normalization used in the literature. For instance, in [501], all the kinetic terms of the RR fields in the action have an extra factor of α' with respect to ours, so their RR fields have units of L^{-1} (ours are dimensionless). In [780, 778] the RR fields have dimensions L^{-4} and their kinetic terms appear in the action with a purely numerical factor instead of the factor $(16\pi G_{\mathrm{N}}^{(10)})^{-1}$ that they carry in our case. Our conventions are also those of [40], where the issue of normalizations has been studied exhaustively.

The inverse relations are

$$\begin{aligned}
\hat{g}_{\hat{\mu}\hat{\nu}} &= \left(-\hat{\hat{g}}_{\underline{zz}}\right)^{\frac{1}{2}} \left(\hat{\hat{g}}_{\hat{\mu}\hat{\nu}} - \hat{\hat{g}}_{\hat{\mu}\underline{z}}\hat{\hat{g}}_{\hat{\nu}\underline{z}}/\hat{\hat{g}}_{\underline{zz}}\right), & \hat{C}^{(3)}{}_{\hat{\mu}\hat{\nu}\hat{\rho}} &= \hat{\hat{C}}_{\hat{\mu}\hat{\nu}\hat{\rho}}, \\
\hat{C}^{(1)}{}_{\hat{\mu}} &= \hat{\hat{g}}_{\hat{\mu}\underline{z}}/\hat{\hat{g}}_{\underline{zz}}, & \hat{B}_{\hat{\mu}\hat{\nu}} &= \hat{\hat{C}}_{\hat{\mu}\hat{\nu}\underline{z}}, \\
\hat{\phi} &= \tfrac{3}{4}\ln\left(-\hat{\hat{g}}_{\underline{zz}}\right).
\end{aligned} \tag{16.18}$$

Now we can perform the reduction of the action Eq. (16.9). We first consider the Ricci scalar term. To reduce this term, we use a slight generalization of Palatini's identity Eq. (D.4). With the above *Ansatz* for the Elfbeins the non-vanishing components of the 11-dimensional spin connection are

$$\begin{aligned}
\hat{\hat{\omega}}_{z\hat{a}z} &= -\tfrac{2}{3}e^{\frac{1}{3}\hat{\phi}}\partial_{\hat{a}}\hat{\phi}, & \hat{\hat{\omega}}_{z\hat{a}\hat{b}} &= -\tfrac{1}{2}e^{\frac{4}{3}\hat{\phi}}\hat{G}^{(2)}{}_{\hat{a}\hat{b}}, \\
\hat{\hat{\omega}}_{\hat{a}\hat{b}z} &= \tfrac{1}{2}e^{\frac{4}{3}\hat{\phi}}\hat{G}^{(2)}{}_{\hat{a}\hat{b}}, & \hat{\hat{\omega}}_{\hat{a}\hat{b}\hat{c}} &= e^{\frac{1}{3}\hat{\phi}}\left(\hat{\omega}_{\hat{a}\hat{b}\hat{c}} + \tfrac{2}{3}\delta_{\hat{a}[\hat{b}}\partial_{\hat{c}]}\hat{\phi}\right),
\end{aligned} \tag{16.19}$$

where

$$\hat{G}^{(2)} = 2\partial\hat{C}^{(1)} \tag{16.20}$$

is the field strength of the ten-dimensional RR 1-form $\hat{C}^{(1)}{}_{\hat{\mu}}$. Using

$$\sqrt{\hat{\hat{g}}} = \sqrt{|\hat{g}|}\, e^{-\frac{8}{3}\hat{\phi}}, \tag{16.21}$$

plus Palatini's identity Eq. (D.4) for $d = 11$ and $\hat{\hat{\phi}} = 0$, plus the fact that the coordinate z conventionally lives in a circle of radius equal to the reduced 11-dimensional Planck length $\ell\!\!\!/^{(11)}_{\text{Planck}} \equiv \ell^{(11)}_{\text{Planck}}/(2\pi)$, we find

$$\int d^{11}\hat{\hat{x}}\,\sqrt{|\hat{\hat{g}}|}\left[\,\hat{\hat{R}}\,\right] = 2\pi\ell\!\!\!/^{(11)}_{\text{Planck}}\int d^{10}\hat{x}\,\sqrt{|\hat{g}|}\left\{-e^{-2\hat{\phi}}\left[\left(\hat{\omega}_{\hat{b}}{}^{\hat{b}\hat{a}} + 2\partial^{\hat{a}}\hat{\phi}\right)^2 + \hat{\omega}_{\hat{a}}{}^{\hat{b}\hat{c}}\hat{\omega}_{\hat{b}\hat{c}}{}^{\hat{a}}\right] - \tfrac{1}{4}\left(\hat{G}^{(2)}\right)^2\right\}.$$

Finally, using Palatini's identity Eq. (D.4) again, but now for $d = 10$ and $\phi = \hat{\phi}$, we obtain for the Ricci scalar term

$$\int d^{11}\hat{\hat{x}}\,\sqrt{|\hat{\hat{g}}|}\left[\,\hat{\hat{R}}\,\right] = 2\pi\ell\!\!\!/^{(11)}_{\text{Planck}}\int d^{10}\hat{x}\,\sqrt{|\hat{g}|}\left\{e^{-2\hat{\phi}}\left[\hat{R} - 4\left(\partial\hat{\phi}\right)^2\right] - \tfrac{1}{4}\left(\hat{G}^{(2)}\right)^2\right\}. \tag{16.22}$$

Now we have to reduce the $\hat{\hat{G}}$-term in Eq. (16.9). We identify field strengths in 11 and ten dimensions with flat indices (this automatically ensures gauge invariance), taking into account the scaling of the ten-dimensional metric

$$\hat{G}^{(4)}{}_{\hat{a}\hat{b}\hat{c}\hat{d}} = e^{-\frac{4}{3}\hat{\phi}}\;\hat{\hat{G}}_{\hat{a}\hat{b}\hat{c}\hat{d}}, \tag{16.23}$$

which leads to

$$\hat{G}^{(4)} = 4\left(\partial\hat{C}^{(3)} - \hat{H}\hat{C}^{(1)}\right), \tag{16.24}$$

where $\hat{H}$ is the field strength of the NSNS two-form $\hat{B}$,

$$\hat{H} = 3\partial\hat{B}. \tag{16.25}$$

The remaining components of $\hat{\hat{G}}$ are given by

$$\hat{\hat{G}}_{\hat{a}\hat{b}\hat{c}z} = e^{\frac{1}{3}\hat{\phi}}\hat{H}_{\hat{a}\hat{b}\hat{c}}, \tag{16.26}$$

and the contribution of the $\hat{\hat{G}}$-term to the ten-dimensional action becomes

$$\int d^{11}\hat{\hat{x}}\,\sqrt{|\hat{\hat{g}}|}\left[-\frac{1}{2\cdot 4!}\left(\hat{\hat{G}}\right)^2\right] = 2\pi\ell_{\text{Planck}}^{(11)}\int d^{10}\hat{x}\,\sqrt{|\hat{g}|}\left[\frac{1}{2\cdot 3!}e^{-2\hat{\phi}}\hat{H}^2 - \frac{1}{2\cdot 4!}\left(\hat{G}^{(4)}\right)^2\right]. \tag{16.27}$$

Finally, taking into account

$$\hat{\hat{\epsilon}}^{\;\hat{\mu}_0\cdots\hat{\mu}_9\underline{z}} = \hat{\epsilon}^{\;\hat{\mu}_0\cdots\hat{\mu}_9}, \tag{16.28}$$

the third term in the $d = 11$ action Eq. (16.9) (all terms with curved indices) gives

$$\hat{\hat{\epsilon}}\hat{\hat{G}}\hat{\hat{G}}\hat{\hat{C}} = 48\hat{\epsilon}\partial\hat{C}^{(3)}\partial\hat{C}^{(3)}\hat{B} - 96\hat{\epsilon}\partial\hat{C}^{(3)}\partial\hat{B}\hat{C}^{(3)}, \tag{16.29}$$

and, on integrating by parts, we obtain

$$\int d^{11}\hat{\hat{x}}\left[-\frac{1}{(144)^2}\hat{\hat{\epsilon}}\hat{\hat{G}}\hat{\hat{G}}\hat{\hat{C}}\right] = 2\pi\ell_{\text{Planck}}^{(11)}\int d^{10}\hat{x}\left[-\frac{1}{144}\hat{\epsilon}\partial\hat{C}^{(3)}\partial\hat{C}^{(3)}\hat{B}\right]. \tag{16.30}$$

On putting all these results together, we find what is described in the literature as the bosonic part of the $N = 2A, d = 10$ supergravity action in ten dimensions in the string frame

$$\begin{aligned}\hat{S} = \frac{2\pi\ell_{\text{Planck}}^{(11)}}{16\pi G_{\text{N}}^{(11)}}\int d^{10}\hat{x}\,\sqrt{|\hat{g}|}\Bigg\{&e^{-2\hat{\phi}}\left[\hat{R} - 4\left(\partial\hat{\phi}\right)^2 + \frac{1}{2\cdot 3!}\hat{H}^2\right]\\ &-\left[\frac{1}{4}\left(\hat{G}^{(2)}\right)^2 + \frac{1}{2\cdot 4!}\left(\hat{G}^{(4)}\right)^2\right] - \frac{1}{144}\frac{1}{\sqrt{|\hat{g}|}}\,\hat{\epsilon}\partial\hat{C}^{(3)}\partial\hat{C}^{(3)}\hat{B}\Bigg\}.\end{aligned} \tag{16.31}$$

In the first line of Eq. (16.38) we can see the known action for the bosonic NSNS fields in the string frame (characterized by the overall factor $e^{-2\hat{\phi}}$). The second line has no dilaton factor and describes the RR sector whose truncation leaves us with the action of $N = 1, d = 10$ supergravity.

However, in string theory we want the string metric to be asymptotically flat and we wanted, in 11-dimensional supergravity, to have asymptotically flat metrics. Both things cannot be true at the same time with the above relations between eleven- and ten-dimensional fields since, with z compact, $\hat{\hat{g}}_{\underline{zz}}$ does not have to go to -1 at infinity in the non-compact directions and the ten-dimensional metric is essentially the 11-dimensional one rescaled by powers of $\hat{\hat{g}}_{\underline{zz}}$ (the dilaton). If the 11-dimensional metric is asymptotically flat and we denote by $\hat{\phi}_0$ the asymptotic value of the dilaton, then the ten-dimensional metric that we have defined behaves at infinity as follows:

$$\hat{g}_{\hat{\mu}\hat{\nu}} \to e^{\frac{2}{3}\hat{\phi}_0}\hat{\eta}_{\hat{\mu}\hat{\nu}}. \tag{16.32}$$

The exponential of the asymptotic value of the dilaton is the (type-IIA) *string coupling constant* $\hat{g}_{\mathrm{A}}$ that counts loops in string amplitudes,

$$\hat{g}_{\mathrm{A}} = e^{\hat{\phi}_0}. \tag{16.33}$$

We could repeat here the discussion of Section 11.2.2 with some minor modifications, but it should be clear that, to obtain the "right" string metric, we have to rescale the metric. To eliminate the additional factors of $e^{\hat{\phi}_0}$ that would appear, we have to rescale all the other fields, although this does not seem strictly necessary. Thus, we make the following substitutions:

$$\begin{aligned} \hat{g}_{\hat{\mu}\hat{\nu}} &\to e^{\frac{2}{3}\hat{\phi}_0}\hat{g}_{\hat{\mu}\hat{\nu}}, & \hat{C}^{(1)}{}_{\hat{\mu}} &\to e^{\frac{1}{3}\hat{\phi}_0}\hat{C}^{(1)}{}_{\hat{\mu}}, \\ \hat{B}_{\hat{\mu}\hat{\nu}} &\to e^{\frac{2}{3}\hat{\phi}_0}\hat{B}_{\hat{\mu}\hat{\nu}}, & \hat{C}^{(3)}{}_{\hat{\mu}\hat{\nu}\hat{\rho}} &\to e^{\hat{\phi}_0}\hat{C}^{(3)}{}_{\hat{\mu}\hat{\nu}\hat{\rho}}, \end{aligned} \tag{16.34}$$

after which the relation between 11- and 10-dimensional fields is

$$\begin{aligned} \hat{\hat{g}}_{\hat{\mu}\hat{\nu}} &= e^{-\frac{2}{3}(\hat{\phi}-\hat{\phi}_0)}\hat{g}_{\hat{\mu}\hat{\nu}} - e^{\frac{4}{3}\hat{\phi}}e^{\frac{2}{3}\hat{\phi}_0}\hat{C}^{(1)}{}_{\hat{\mu}}\hat{C}^{(1)}{}_{\hat{\nu}}, & \hat{\hat{C}}_{\hat{\mu}\hat{\nu}\hat{\rho}} &= e^{\hat{\phi}_0}\hat{C}^{(3)}{}_{\hat{\mu}\hat{\nu}\hat{\rho}}, \\ \hat{\hat{g}}_{\hat{\mu}\underline{z}} &= -e^{\frac{4}{3}\hat{\phi}}e^{\frac{1}{3}\hat{\phi}_0}\hat{C}^{(1)}{}_{\hat{\mu}}, & \hat{\hat{C}}_{\hat{\mu}\hat{\nu}\underline{z}} &= e^{\frac{2}{3}\hat{\phi}_0}\hat{B}_{\hat{\mu}\hat{\nu}}, \\ \hat{\hat{g}}_{\underline{zz}} &= -e^{\frac{4}{3}\hat{\phi}}. \end{aligned} \tag{16.35}$$

For the Elfbein we have

$$\begin{aligned} \left(\hat{\hat{e}}^{\hat{\hat{a}}}{}_{\hat{\hat{\mu}}}\right) &= \begin{pmatrix} e^{-\frac{1}{3}(\hat{\phi}-\hat{\phi}_0)}\hat{e}^{\hat{a}}{}_{\hat{\mu}} & e^{\frac{2}{3}\hat{\phi}}e^{\frac{1}{3}\hat{\phi}_0}\hat{C}^{(1)}{}_{\hat{\mu}} \\ 0 & e^{\frac{2}{3}\hat{\phi}} \end{pmatrix}, \\ \left(\hat{\hat{e}}_{\hat{\hat{a}}}{}^{\hat{\hat{\mu}}}\right) &= \begin{pmatrix} e^{\frac{1}{3}(\hat{\phi}-\hat{\phi}_0)}\hat{e}_{\hat{a}}{}^{\hat{\mu}} & -e^{\frac{1}{3}\hat{\phi}}\hat{C}^{(1)}{}_{\hat{a}} \\ 0 & e^{-\frac{2}{3}\hat{\phi}} \end{pmatrix}. \end{aligned} \tag{16.36}$$

The inverse relations are

$$\begin{aligned} \hat{g}_{\hat{\mu}\hat{\nu}} &= e^{-\frac{2}{3}\hat{\phi}_0}\left(-\hat{\hat{g}}_{\underline{zz}}\right)^{\frac{1}{2}}\left(\hat{\hat{g}}_{\hat{\mu}\hat{\nu}} - \hat{\hat{g}}_{\hat{\mu}\underline{z}}\hat{\hat{g}}_{\hat{\nu}\underline{z}}/\hat{\hat{g}}_{\underline{zz}}\right), & \hat{C}^{(3)}{}_{\hat{\mu}\hat{\nu}\hat{\rho}} &= e^{-\hat{\phi}_0}\hat{\hat{C}}_{\hat{\mu}\hat{\nu}\hat{\rho}}, \\ \hat{C}^{(1)}{}_{\hat{\mu}} &= e^{-\frac{1}{3}\hat{\phi}_0}\hat{\hat{g}}_{\hat{\mu}\underline{z}}/\hat{\hat{g}}_{\underline{zz}}, & \hat{B}_{\hat{\mu}\hat{\nu}} &= e^{-\frac{2}{3}\hat{\phi}_0}\hat{\hat{C}}_{\hat{\mu}\hat{\nu}\underline{z}}, \\ \hat{\phi} &= \tfrac{3}{4}\ln\left(-\hat{\hat{g}}_{\underline{zz}}\right), \end{aligned} \tag{16.37}$$

and, finally, the action becomes

$$\boxed{\begin{aligned}\hat{S} \;=\; \frac{\hat{g}_{\mathrm{A}}^{2}}{16\pi G_{\mathrm{NA}}^{(10)}}\int d^{10}\hat{x}\,\sqrt{|\hat{g}|}\Big\{& e^{-2\hat{\phi}}\left[\hat{R} - 4\left(\partial\hat{\phi}\right)^{2} + \frac{1}{2\cdot 3!}\hat{H}^{2}\right] \\ & - \left[\tfrac{1}{4}\left(\hat{G}^{(2)}\right)^{2} + \frac{1}{2\cdot 4!}\left(\hat{G}^{(4)}\right)^{2}\right] \\ & - \frac{1}{144}\frac{1}{\sqrt{|\hat{g}|}}\,\hat{\epsilon}\partial\hat{C}^{(3)}\partial\hat{C}^{(3)}\hat{B}\Big\},\end{aligned}} \tag{16.38}$$

where we have made the identification of the prefactor of the action

$$\frac{2\pi\ell_{\mathrm{Planck}}^{(11)}e^{\frac{8}{3}\hat{\phi}_{0}}}{16\pi G_{\mathrm{N}}^{(11)}} = \frac{\hat{g}_{\mathrm{A}}^{2}}{16\pi G_{\mathrm{NA}}^{(10)}}. \tag{16.39}$$

The factor g_{A}^{2} absorbs the asymptotic value of the dilaton in the action.

Then, we find the following relation:

$$G_{\mathrm{NA}}^{(10)} = \frac{G_{\mathrm{N}}^{(11)}}{2\pi\ell_{\mathrm{Planck}}^{(11)}\hat{g}_{\mathrm{A}}^{\frac{2}{3}}}. \tag{16.40}$$

Observe that we have taken the 11th coordinate z to live in a circle of radius equal to the reduced 11-dimensional Planck length $\ell_{\mathrm{Planck}}^{(11)}$, which, up to numerical factors, is the only scale available in 11-dimensional supergravity. Actually, we have to distinguish between the interval in which z takes values ($z \in [0, 2\pi\ell_{\mathrm{Planck}}^{(11)}]$) and the actual radius of the 11th dimension (measured at infinity) which we denote here by R_{11} and which is naturally measured with the 11-dimensional metric:

$$R_{11} = \frac{1}{2\pi}\lim_{r\to\infty}\int\sqrt{|\hat{\hat{g}}_{\underline{zz}}|}dz = \ell_{\mathrm{Planck}}^{(11)}e^{\frac{2}{3}\hat{\phi}_{0}} = \ell_{\mathrm{Planck}}^{(11)}\hat{g}_{\mathrm{A}}^{\frac{2}{3}}. \tag{16.41}$$

By using this relation in (16.40), we find

$$G_{\mathrm{NA}}^{(10)} = \frac{G_{\mathrm{N}}^{(11)}}{2\pi R_{11}} = \frac{G_{\mathrm{N}}^{(11)}}{V_{11}}, \tag{16.42}$$

as usual in KK theory (V_{11} is the volume of the internal space).

Now, in this case, we define the 11-dimensional Planck length[6] by

$$\boxed{16\pi G_{\mathrm{N}}^{(11)} \equiv \frac{(\ell_{\mathrm{Planck}}^{(11)})^{9}}{2\pi},} \tag{16.43}$$

[6] The reason for this somewhat unusual definition is explained in Section 19.1.1.

and, thus, in terms of it and $\hat{g}_{\rm A}$ only, the ten-dimensional Newton constant is given by

$$G_{\rm N\,A}^{(10)} = \frac{(\ell_{\rm Planck}^{(11)})^8}{32\pi^2 \hat{g}_{\rm A}^{\frac{2}{3}}}. \tag{16.44}$$

This result has to be compared with the value we will obtain in Section 19.1.1 for $G_{\rm N}^{(10)}$ in terms of the type-IIA stringy variables $\ell_{\rm s}$ and $\hat{g}_{\rm A}$, Eq. (19.26):

$$G_{\rm N\,A}^{(10)} = 8\pi^6 \hat{g}_{\rm A}^2 (\alpha')^4. \tag{16.45}$$

They are consistent if

$$\ell_{\rm Planck}^{(11)} = 2\pi \ell_{\rm s} \hat{g}_{\rm A}^{\frac{1}{3}}, \tag{16.46}$$

and, thus,

$$R_{11} = \ell_{\rm s} \hat{g}_{\rm A}. \tag{16.47}$$

These are the two main relations between 11- and 10-dimensional type-IIA constants and they deserve to be framed together:

$$\boxed{\begin{aligned} \ell_{\rm Planck}^{(11)} &= 2\pi \ell_{\rm s} \hat{g}_{\rm A}^{\frac{1}{3}}, \\ R_{11} &= \ell_{\rm s} \hat{g}_{\rm A}. \end{aligned}} \tag{16.48}$$

Once we have expressed R_{11} in these stringy variables, it is easy to see that the strong-coupling limit of the type-IIA theory ($\hat{g}_{\rm A} \to \infty$) coincides with the decompactification limit R_{11} in which a new dimension becomes macroscopic [962].

Just as the KK scalar of the 11-dimensional theory gives the string-theory dilaton, the KK vector gives the RR 1-form and the 3-form gives the RR 3-form and the NSNS 2-form. Since we know that there are D0- and D2-branes associated with the RR 1-form and 3-form, we find that they originate, respectively, from the 11-dimensional graviton moving in the compact direction and from a two-dimensional object that couples to the 11-dimensional 3-form: the M2-brane. An M2-brane wrapped around the compact dimension gives the type-IIA string ($\hat{B}_{\hat{\mu}\hat{\nu}} = e^{-\frac{2}{3}\hat{\phi}_0} \hat{\hat{C}}_{\hat{\mu}\hat{\nu}\underline{z}}$) and, unwrapped, gives the D2-brane ($\hat{C}^{(3)}{}_{\hat{\mu}\hat{\nu}\hat{\rho}} = e^{-\hat{\phi}_0} \hat{\hat{C}}_{\hat{\mu}\hat{\nu}\hat{\rho}}$). These and more relations between extended objects are represented in Figure 19.5 on page 552.

16.1.3 *Magnetic potentials*

For each of the differential-form potentials $\hat{C}^{(1)}$, $\hat{C}^{(3)}$, and $\hat{B}$ a magnetic dual $\hat{C}^{(7)}$, $\hat{C}^{(5)}$, and $\hat{B}^{(6)}$, respectively, whose equation of motion is equivalent to the Bianchi identity of the original (electric) potential and vice-versa, can be introduced. (In general we can dualize only on-shell since the electric potentials occur without derivatives in the action.) These potentials will be useful in studying D-brane solutions and D-brane effective actions.

Dual field strengths are defined through the Hodge dual of the original field strengths, but potentials can be defined in many ways. A way to define the dual (magnetic) potentials is to fix the form of the corresponding field strength. For RR potentials, a convenient general form for the field strengths was proposed in [112, 470] and used explicitly in [136, 691]. In differential-forms and component language, it is

$$\begin{aligned}\hat{G}^{(2n)} &= d\hat{C}^{(2n-1)} - \hat{H}\hat{C}^{(2n-3)},\\ \hat{G}^{(2n)} &= 2n\Big[\partial\hat{C}^{(2n-1)} - \tfrac{1}{2}(2n-1)(2n-2)\partial\hat{B}\hat{C}^{(2n-3)}\Big].\end{aligned} \tag{16.49}$$

The differential-forms language considerably simplifies the expressions. The gauge transformations of the RR form potentials are

$$\delta\hat{C}^{(2n-1)} = d\hat{\Lambda}^{(2n-2)} - d\hat{B}\hat{\Lambda}^{(2n-4)}. \tag{16.50}$$

This normalization is extremely useful because it can be generalized to the type-IIB and massive type-IIA fields. For massless type-II supergravities, we can also introduce the notation

$$\begin{aligned}\hat{C} &= \hat{C}^{(0)} + \hat{C}^{(1)} + \hat{C}^{(2)} + \cdots,\\ \hat{G} &= \hat{G}^{(1)} + \hat{G}^{(2)} + \cdots,\\ \hat{\Lambda}^{(\cdot)} &= \hat{\Lambda}^{(0)} + \hat{\Lambda}^{(1)} + \cdots,\end{aligned} \tag{16.51}$$

with which we can write[7] (as we are going to show)

$$\begin{aligned}\hat{H} &= d\hat{B}, & \delta\hat{B} &= d\hat{\Lambda},\\ \hat{G} &= d\hat{C} - \hat{H}\wedge\hat{C}, & \delta\hat{C} &= d\hat{\Lambda}^{(\cdot)} - d\hat{B}\wedge\hat{\Lambda}^{(\cdot)}.\end{aligned} \tag{16.52}$$

Now we have to prove that it is indeed possible to have magnetic RR potentials with field strengths of that kind. First of all, the field strengths $\hat{G}^{(2)}$ and $\hat{G}^{(4)}$ we are already using conform to this normalization. Their Bianchi identities are

$$d\hat{G}^{(2)} = 0, \qquad d\hat{G}^{(4)} - \hat{H}\wedge\hat{G}^{(2)} = 0, \tag{16.53}$$

and, from the action, the equations of motion are found to be

$$d^{\star}\hat{G}^{(2)} + H\wedge{}^{\star}\hat{G}^{(4)} = 0, \qquad d^{\star}\hat{G}^{(4)} - \hat{H}\wedge\hat{G}^{(4)} = 0. \tag{16.54}$$

The Bianchi identities for the dual field strengths $\hat{G}^{(8)}$ and $\hat{G}^{(6)}$ are, according to the general normalization,

$$d\hat{G}^{(8)} - \hat{H}\wedge\hat{G}^{(6)} = 0, \qquad d\hat{G}^{(6)} - \hat{H}\wedge\hat{G}^{(4)} = 0. \tag{16.55}$$

By comparison with the equations of motion for the electric potentials, we find the relations

$$\hat{G}^{(8)} = -{}^{\star}\hat{G}^{(2)}, \qquad \hat{G}^{(6)} = +{}^{\star}\hat{G}^{(4)}. \tag{16.56}$$

[7] The RR gauge transformations are also written, after a redefinition of the gauge parameters, in the form

$$\delta\hat{C} = d\hat{\Lambda}^{(\cdot)}e^{\hat{B}}.$$

These relations define the dual field strengths, which, together with the general form for the field strengths, defines the magnetic RR potentials.

The above relations between electric- and magnetic-field strengths and their Bianchi identities can be written in the general form

$$\boxed{\begin{aligned} \hat{G}^{(10-k)} &= (-1)^{[k/2]}\,{}^\star\hat{G}^{(k)}, \\ d\hat{G} - \hat{H}\wedge\hat{G} &= 0, \end{aligned}} \tag{16.57}$$

which we will see apply in this form to the massive type-IIA theory in Section 16.2 and to the type-IIB theory in Chapter 17.

Observe that the set of all the Bianchi identities plus the duality relations between electric and magnetic potentials determine all the equations of motion of the RR fields. Observe also that all the "electric" RR potentials are pseudotensors and that their duals are, by definition, tensors.

Now let us dualize the NSNS 2-form potential. Using the above general definitions for RR potentials, field strengths etc., we can write

$$\boxed{\begin{aligned} \hat{H}^{(7)} &= e^{-2\hat{\phi}}\,{}^\star\hat{H}, \\ dH &= 0, \\ d\hat{H}^{(7)} + \tfrac{1}{2}\,{}^\star\hat{G}\wedge\hat{G} &= 0, \\ d\left(e^{-2\phi}\,{}^\star\hat{H}\right) + \tfrac{1}{2}\,{}^\star\hat{G}\wedge\hat{G} &= 0, \\ d\left(e^{2\phi}\,{}^\star\hat{H}^{(7)}\right) &= 0, \end{aligned}} \tag{16.58}$$

which, again, apply in precisely this form to massive type-IIA and type-IIB theories. In the type-IIA case, one can take [603]

$$\hat{H}^{(7)} = d\hat{B}^{(6)} - \tfrac{1}{2}\sum_{n=1}^{n=4}{}^\star\hat{G}^{(2n+2)}\wedge\hat{C}^{(2n-1)}. \tag{16.59}$$

It is more difficult to relate the magnetic potentials to 11-dimensional potentials since some of them can be defined only in ten dimensions. For instance, the $\hat{C}^{(7)}$ associated with the D6-brane is the dual of the $\hat{C}^{(1)}$, which in $d = 11$ is part of the metric. In fact, the D6-brane can be obtained by compactification of the 11-dimensional KK monopole (KKM), which is a purely gravitational solution that, by definition, always has a compact direction. The potential to which the 11-dimensional KKM couples has been studied in [578]. The potentials $\hat{B}^{(6)}$ and $\hat{C}^{(5)}$ can be related to $\hat{\hat{\tilde{C}}}$, and the associated extended objects (the so-called *solitonic 5-brane* S5A which, by definition, couples to $\hat{B}^{(6)}$ and the D4-brane) originate from the 11-dimensional 5-brane M5 which, by definition, is the object that couples to $\hat{\hat{\tilde{C}}}$. These relations are represented in Figure 19.5 on page 552.

16.1.4 Reduction of fermions and the supersymmetry rules

Here we want to reduce to ten dimensions the fermions and the supersymmetry transformation laws in Eq. (16.8). We will keep only terms up to second order in fermions.

First we need to decompose the 11-dimensional gamma matrices in terms of the ten-dimensional ones. This is done in Appendix B.1.4. Next, we have to decompose the 11-dimensional spinors into ten-dimensional spinors. Eleven-dimensional Majorana spinors are also ten-dimensional Majorana spinors. However, in ten-dimensional supergravity, the elementary spinor is a Majorana–Weyl spinor. Thus, each 11-dimensional spinor can be considered as a pair of Majorana–Weyl spinors with opposite chiralities (this is why this theory is non-chiral and it is $N = 2$). In principle we could split all the spinors into their chiral halves, but is is not worth doing it for the moment. Later on, we will have to do it in order to relate the spinors to those of the type-IIB theory, which are of the same chirality and cannot be considered the two halves of any Majorana spinor. As we will see, there are two options in the type-IIB case: either we use indices $i = 1, 2$ for the spinors or we combine them into a chiral complex spinor. We will use the first possibility.

We express the 11-dimensional spinors in terms of the ten-dimensional spinors (gravitino $\hat{\psi}_{\hat{\mu}}$, dilatino $\hat{\lambda}$, and the supersymmetry transformation parameter $\hat{\epsilon}$) as follows:[8]

$$\hat{\hat{\epsilon}} = e^{-\frac{1}{6}\hat{\phi}}\,\hat{\epsilon}, \quad \hat{\hat{\psi}}_{\hat{a}} = e^{\frac{1}{6}\hat{\phi}}\Big(2\hat{\psi}_{\hat{a}} - \tfrac{1}{3}\hat{\Gamma}_{\hat{a}}\hat{\lambda}\Big), \quad \hat{\hat{\psi}}_{z} = \frac{2i}{3}e^{\frac{1}{6}\hat{\phi}}\hat{\Gamma}_{11}\hat{\lambda}. \tag{16.60}$$

Observe that, with these definitions, the gravitino $\hat{\psi}_{\hat{\mu}}$ is real but the dilatino $\hat{\lambda}$ is purely imaginary. We could use a purely real dilatino just by multiplying by i, but then its supersymmetry rule would look unconventional.

We now want to use the relation between the 11- and ten-dimensional bosonic fields that we have already obtained. However, we have performed the dimensional reduction working in a special Lorentz gauge $\hat{\hat{e}}^{\hat{a}}{}_{\underline{z}} = 0$ and supersymmetry transformations do not preserve this gauge. In fact,

$$\delta_{\hat{\hat{\epsilon}}}\hat{\hat{e}}^{\hat{a}}{}_{\underline{z}} = \tfrac{1}{3}e^{\frac{1}{3}\hat{\phi}}\,\bar{\hat{\epsilon}}\,\hat{\Gamma}^{\hat{a}}\hat{\Gamma}_{11}\hat{\lambda}. \tag{16.61}$$

We have to introduce a compensating local Lorentz transformation in order to preserve our gauge choice. Then, the ten-dimensional supersymmetry transformation $\delta_{\hat{\epsilon}}$ will be a combination of an 11-dimensional supersymmetry transformation $\delta_{\hat{\hat{\epsilon}}}$ and an 11-dimensional compensating local Lorentz transformation $\delta_{\hat{\hat{\sigma}}}$ such that

$$\delta_{\hat{\epsilon}}\hat{\hat{e}}^{\hat{a}}{}_{\underline{z}} \equiv \big(\delta_{\hat{\hat{\epsilon}}} + \delta_{\hat{\hat{\sigma}}}\big)\hat{\hat{e}}^{\hat{a}}{}_{\underline{z}} = \tfrac{1}{3}e^{\frac{1}{3}\hat{\phi}}\,\bar{\hat{\epsilon}}\,\hat{\Gamma}^{\hat{a}}\hat{\Gamma}_{11}\hat{\lambda} + \tfrac{1}{2}\hat{\hat{\sigma}}^{\hat{b}\hat{c}}\Gamma_{\mathrm{v}}\Big(\hat{\hat{M}}_{\hat{b}\hat{c}}\Big)^{\hat{a}}{}_{\hat{d}}\,\hat{\hat{e}}^{\hat{d}}{}_{\underline{z}} = 0. \tag{16.62}$$

Since the generators of the Lorentz group in the vector representation are given by Eq. (A.60), the parameter of the compensating Lorentz transformation is given by

$$\hat{\hat{\sigma}}^{\hat{a}}{}_{z} = -\tfrac{1}{3}e^{\frac{1}{3}\hat{\phi}}\,\bar{\hat{\epsilon}}\,\hat{\Gamma}^{\hat{a}}\hat{\Gamma}_{11}\hat{\lambda}. \tag{16.63}$$

[8] As a first step, we simply identify, up to factors involving the dilaton, the dilatino $\hat{\lambda}$ with the (flat) component $\hat{\hat{\psi}}_{z}$ and the gravitino $\hat{\psi}_{\hat{a}}$ with the (flat) components $\hat{\hat{\psi}}_{\hat{a}}$. Then we see that it is natural to combine this dilatino and this gravitino into a new one whose supersymmetry transformation rules are much simpler. The final combinations are the ones we write.

We now have to apply this definition of a ten-dimensional supersymmetry transformation to all fields, performing a compensating Lorentz transformation on all of them with the above parameter and using the explicit form of the Lorentz transformations on each individual field.

After some calculations we find the following $N = 2A$, $d = 10$ supersymmetry transformation laws (only for the "electric" NSNS and RR potentials):

$$\boxed{\begin{aligned}
\delta_{\hat\epsilon}\hat{e}^{\hat a}{}_{\hat\mu} &= -i\bar{\hat\epsilon}\hat\Gamma^{\hat a}\hat\psi_{\hat\mu},\\
\delta_{\hat\epsilon}\hat\psi_{\hat\mu} &= \left\{\partial_{\hat\mu} - \tfrac{1}{4}\left(\hat{\not\omega}_{\hat\mu} + \tfrac{1}{2}\Gamma_{11}\hat{\not H}_{\mu}\right)\right\}\hat\epsilon\\
&\quad + \frac{i}{8}e^{\hat\phi}\sum_{n=1,2}\frac{1}{(2n)!}\,\hat{\not G}^{(2n)}\hat\Gamma_{\hat\mu}\left(-\hat\Gamma_{11}\right)^{n}\hat\epsilon,\\
\delta_{\hat\epsilon}\hat{B}_{\hat\mu\hat\nu} &= -2i\bar{\hat\epsilon}\hat\Gamma_{[\hat\mu}\hat\Gamma_{11}\hat\psi_{\hat\nu]},\\
\delta_{\hat\epsilon}\hat{C}^{(1)}{}_{\hat\mu} &= -e^{\hat\phi}\bar{\hat\epsilon}\hat\Gamma_{11}\left(\hat\psi_{\hat\mu} - \tfrac{1}{2}\hat\Gamma_{\hat\mu}\hat\lambda\right),\\
\delta_{\hat\epsilon}\hat{C}^{(3)}{}_{\hat\mu\hat\nu\hat\rho} &= 3e^{\hat\phi}\bar{\hat\epsilon}\hat\Gamma_{\hat\mu\hat\nu}\left(\hat\psi_{\hat\rho]} - \frac{1}{3!}\hat\Gamma_{\hat\rho]}\hat\lambda\right) + 3\hat{C}^{(1)}{}_{[\hat\mu}\delta_{\hat\epsilon}\hat{B}_{\hat\mu\hat\nu]},\\
\delta_{\hat\epsilon}\hat\lambda &= \left(\hat{\not\partial}\hat\phi + \frac{1}{12}\hat\Gamma_{11}\hat{\not H}\right)\hat\epsilon + \frac{i}{4}e^{\hat\phi}\sum_{n=1,2}\frac{5-2n}{(2n)!}\,\hat{\not G}^{(2n)}\left(-\hat\Gamma_{11}\right)^{n}\hat\epsilon,\\
\delta_{\hat\epsilon}\hat\phi &= -\frac{i}{2}\bar{\hat\epsilon}\hat\lambda.
\end{aligned}} \tag{16.64}$$

Observe that, in principle, one obtains two additional terms in $\delta_{\hat\epsilon}\hat{e}_{\hat\mu}{}^{\hat a}$:

$$\frac{i}{3}\bar{\hat\epsilon}\hat\Gamma^{\hat a}\hat\Gamma_{\hat\mu}\hat\lambda + \tfrac{1}{3}\delta_{\hat\epsilon}\hat\phi\,\hat{e}_{\hat\mu}{}^{\hat a}, \tag{16.65}$$

which combine into an infinitesimal Lorentz transformation of the Zehnbein with parameter

$$\hat\sigma^{\hat a\hat b} = -\frac{i}{6}\bar{\hat\epsilon}\hat\Gamma^{\hat a\hat b}\hat\lambda. \tag{16.66}$$

The same transformation also appears in the other fields with tangent-space indices (i.e. just the fermions) but at higher orders in fermions. Thus, it can be absorbed into a redefinition of the ten-dimensional local Lorentz transformations and that is why we have ignored it.

Another point is that we have obtained these expressions without taking into account the final rescaling of the bosonic fields by powers of $e^{\hat\phi_0}$. It can be checked that, if we also rescale the fermions according to

$$\hat\psi_{\hat\mu} \to e^{\frac{1}{6}\hat\phi_0}\hat\psi_{\hat\mu}, \quad \hat\lambda \to e^{-\frac{1}{6}\hat\phi_0}\hat\lambda, \quad \hat\epsilon \to e^{\frac{1}{6}\hat\phi_0}\hat\epsilon, \tag{16.67}$$

the supersymmetry transformation rules remain invariant.

In order to study which solutions of this theory preserve some supersymmetry, it is desirable to include the magnetic potentials in the fermionic supersymmetry transformation rules, since many solutions are naturally expressed in terms of the magnetic variables.[9] Using the general definitions Eqs. (16.57) and (16.58) plus the identity Eq. (B.77), it is easy to prove that

$$\begin{aligned} \frac{1}{k!}\,\hat{\not{G}}^{(k)} &= \frac{(-1)^k}{(10-k)!}\,\hat{\not{G}}^{(10-k)}\hat{\Gamma}_{11} = \tfrac{1}{2}\left\{\frac{1}{k!}\,\hat{\not{G}}^{(k)} + \frac{(-1)^k}{(10-k)!}\,\hat{\not{G}}^{(10-k)}\hat{\Gamma}_{11}\right\}, \\ \frac{1}{3!}\hat{\Gamma}_{11}\,\hat{\not{H}} &= -\frac{1}{7!}e^{2\hat{\phi}}\,\hat{\not{H}}^{(7)} \qquad\qquad = \tfrac{1}{2}\left\{\frac{1}{3!}\hat{\Gamma}_{11}\,\hat{\not{H}} - \frac{1}{7!}e^{2\hat{\phi}}\,\hat{\not{H}}^{(7)}\right\}. \end{aligned} \tag{16.68}$$

Furthermore,

$$\hat{\not{H}}_{\hat{\mu}} = \frac{2}{7!}e^{2\hat{\phi}}\hat{\Gamma}_{11}\hat{\Gamma}_{\hat{\mu}\hat{\nu}_1\cdots\hat{\nu}_7}\hat{H}^{(7)\,\hat{\nu}_1\cdots\hat{\nu}_7} = \tfrac{1}{2}\left\{\hat{\not{H}}_{\hat{\mu}} + \frac{2}{7!}e^{2\hat{\phi}}\hat{\Gamma}_{11}\hat{\Gamma}_{\hat{\mu}\hat{\nu}_1\cdots\hat{\nu}_7}\hat{H}^{(7)\,\hat{\nu}_1\cdots\hat{\nu}_7}\right\}. \tag{16.69}$$

We simply have to substitute these identities into Eqs. (16.64) to obtain

$$\boxed{\begin{aligned} \delta_{\hat{\epsilon}}\hat{\psi}_{\hat{\mu}} &= \left\{\partial_{\hat{\mu}} - \tfrac{1}{4}\left(\hat{\not{\omega}}_{\hat{\mu}} + \tfrac{1}{4}\Gamma_{11}\,\hat{\not{H}}_{\mu} + \frac{1}{2\cdot 7!}e^{2\hat{\phi}}\hat{\Gamma}_{\hat{\mu}\hat{\nu}_1\cdots\hat{\nu}_7}\hat{H}^{(7)\,\hat{\nu}_1\cdots\hat{\nu}_7}\right)\right\}\hat{\epsilon} \\ &\quad + \frac{i}{16}e^{\hat{\phi}}\sum_{n=1}^{n=4}\frac{1}{(2n)!}\,\hat{\not{G}}^{(2n)}\hat{\Gamma}_{\hat{\mu}}\left(-\hat{\Gamma}_{11}\right)^n\hat{\epsilon}, \\ \delta_{\hat{\epsilon}}\hat{\lambda} &= \left[\not{\partial}\hat{\phi} + \tfrac{1}{4}\left(\frac{1}{3!}\hat{\Gamma}_{11}\,\hat{\not{H}} - \frac{1}{7!}e^{2\hat{\phi}}\hat{\not{H}}^{(7)}\right)\right]\hat{\epsilon} \\ &\quad + \frac{i}{8}e^{\hat{\phi}}\sum_{n=1}^{n=4}\frac{5-2n}{(2n)!}\,\hat{\not{G}}^{(2n)}\left(-\hat{\Gamma}_{11}\right)^n\hat{\epsilon}. \end{aligned}} \tag{16.70}$$

When using these expressions one should always keep in mind that the magnetic potentials are *not* independent and that, if the magnetic potentials do not vanish, the electric ones will not and vice-versa and their contributions have to be added.

Finally, it is possible to assign on-shell supersymmetry transformation rules to all the RR potentials (electric and magnetic) [117]. It should also be possible to assign on-shell supersymmetry transformation rules to the dual KR 6-form.

16.2 Romans' massive $N = 2A, d = 10$ supergravity

Romans showed in [810] that (in contrast to 11-dimensional or $N = 2B, d = 10$ supergravity) $N = 2A, d = 10$ supergravity can be deformed by introducing a mass parameter

9 It is possible to find supersymmetry transformation rules for the magnetic RR potentials for which the supersymmetry algebra is satisfied on-shell (which is not a problem since these potentials are, anyway, defined only on-shell). See e.g. [117]. The same is probably true for the NSNS magnetic potentials, but the transformation rules have not been given in the literature.

m while keeping, formally, all the supersymmetry and gauge symmetry, although, actually, both are broken, as we will see. It is not known how to derive this theory from the standard 11-dimensional SUGRA and it is not possible to deform 11-dimensional SUGRA in any way to include a cosmological constant, while preserving at the same time 11-dimensional Poincaré invariance [91, 303, 304].[10]

The deformation of the $N = 2A$, $d = 10$ theory consists in a deformation of the RR field strengths and in a deformation of the Lagrangian. Both are consistent with a generalization of the systematic definition of electric and magnetic RR field strengths in Section 16.1.3 in which we replace the second and fourth of Eqs. (16.52) by[11]

$$\begin{aligned}\hat{G} &= d\hat{C} - \hat{H} \wedge \hat{C} + me^{\hat{B}},\\ \delta\hat{C} &= d\hat{\Lambda}^{(\cdot)} - d\hat{B} \wedge \hat{\Lambda}^{(\cdot)} - m\hat{\Lambda}e^{\hat{B}},\end{aligned} \tag{16.71}$$

while the equations of motion, Bianchi identities, and duality relations Eqs. (16.57) remain valid as they are. In particular, the 2- and 4-form field strengths are

$$\hat{G}^{(2)} = 2\partial\hat{C}^{(1)} + m\hat{B}. \qquad \hat{G}^{(4)} = 4\partial\hat{C}^{(3)} - 12\partial\hat{B}\hat{C}^{(1)} + 3m\hat{B}\hat{B}, \tag{16.72}$$

and the gauge transformations of the RR 1- and 3-forms are

$$\delta\hat{C}^{(1)} = \partial\hat{\Lambda}^{(0)} - m\hat{\Lambda}, \qquad \delta\hat{C}^{(3)} = 3\partial\hat{\Lambda}^{(2)} - \hat{H}\hat{\Lambda}^{(0)} - 3m\hat{\Lambda}\hat{B}. \tag{16.73}$$

The Lagrangian is deformed by replacing the RR field strengths by the deformed ones, by adding a "cosmological-constant" term proportional to m^2, and by adding new terms to the Chern–Simons piece of the action. The action of the bosonic sector is

$$\boxed{\begin{aligned}\hat{S} = \frac{g_A^2}{16\pi G_{\mathrm{NA}}^{(10)}}\int d^{10}\hat{x}\,\sqrt{|\hat{g}|}\Bigg\{&e^{-2\hat{\phi}}\left[\hat{R} - 4\left(\partial\hat{\phi}\right)^2 + \frac{1}{2\cdot 3!}\hat{H}^2\right]\\ &-\left[\tfrac{1}{2}m^2 + \frac{1}{2\cdot 2!}\left(\hat{G}^{(2)}\right)^2 + \frac{1}{2\cdot 4!}\left(\hat{G}^{(4)}\right)^2\right]\\ &-\frac{1}{144}\frac{1}{\sqrt{|\hat{g}|}}\,\hat{\epsilon}\Big[\partial\hat{C}^{(3)}\partial\hat{C}^{(3)}\hat{B} + \tfrac{1}{2}m\partial\hat{C}^{(3)}\hat{B}\hat{B}\hat{B}\\ &\qquad + \frac{9}{80}m^2\hat{B}\hat{B}\hat{B}\hat{B}\hat{B}\Big]\Bigg\}.\end{aligned}} \tag{16.74}$$

Actually, there is no cosmological-constant term: in the Einstein frame m^2 carries a dilaton factor and so it is a potential for the dilaton. On the other hand, the $\hat{\Lambda}$ gauge transformations can be used to gauge away $\hat{C}^{(1)}$ completely, which is nothing but a Stückelberg

[10] See, however, [28, 29, 136, 691]. Also, it is possible to derive other ten-dimensional massive theories from $d = 11$ [566].

[11] On redefining $\hat{\Lambda}^{(\cdot)}$, the RR gauge transformations can also be rewritten in the form

$$\delta\hat{C} = (d\hat{\Lambda}^{(\cdot)} - m\hat{\Lambda})e^{\hat{B}}.$$

field for the NSNS 2-form $\hat{B}$ which transforms its kinetic term in the action into a conventional mass term $m^2\hat{B}^2$, and is the reason why this theory is called *massive*.

One may wonder how some of the fields of the supergravity multiplet can become massive while the theory is formally invariant under $N = 2$ local supersymmetry since linearly realized supersymmetry implies that all states in the same supermultiplet have the same mass. The reason is not just gauge symmetry but also that supersymmetry is broken in this theory. There are two ways to see this: on the one hand, a local supersymmetry transformation can be used to gauge away one dilatino and give mass to one gravitino; on the other hand, the most (super)symmetric vacuum of this theory, which is not Minkowski spacetime but the D8-brane, breaks half of the supersymmetries[12] as well as some of the isometries of Minkowski spacetime. Romans' massive $N = 2A, d = 10$ supergravity can be interpreted as the effective-field theory of type-IIA superstrings with an open-string sector associated with a D8-brane that breaks translation invariance and supersymmetry [782].

There are good reasons to interpret the mass parameter m as another (0-form) RR field strength,

$$\hat{G}^{(0)} \equiv m. \tag{16.75}$$

A 0-form field strength has to be constant due to the Bianchi identity $d\hat{G}^{(0)} = 0$. It can be dualized (on-shell) into a 10-form field strength $\hat{G}^{(10)}$ whose equation of motion is this Bianchi identity and whose Bianchi identity also follows the general rule Eqs. (16.57) and implies the existence of a RR 9-form $\hat{C}^{(9)}$, which must be non-trivial in order to have ${}^\star\hat{G}^{(10)} = -\hat{G}^{(0)} = -m$. A RR 9-form potential was required by string theory since the type-IIA theory admits all even p Dp-branes, which couple to RR $(p+1)$-form potentials and in massless $N = 2A, d = 10$ supergravity there is no potential for the D8-brane. Romans' theory describes the effective type-IIA string theory in the presence of D8-branes. The trouble with the 9-form potential is that it does not have dynamical degrees of freedom and, if we include it in the form of a mass parameter, there is a D8-brane and if we do not include it, there is not a D8-brane, whereas, for lower-rank potentials, the *same* theory admits solutions with and without branes.[13]

The 11-dimensional origin of the D8-brane and its associated mass parameter are unknown, although there are arguments based on the superalgebras of $N = 2A, d = 10$ and $d = 11$ supergravity that support the idea that there is a nine-dimensional extended object in $d = 11$ (the M9-brane discussed in [123, 142], also known as the KK9M-brane [666]), which could also be associated with the $(9+1)$-dimensional boundaries of the Hořava–Witten construction discussed on page 16.4.

The supersymmetry transformation rules are given by Eqs. (16.70), where the sums have to be extended up to $n = 5$ to include $\hat{G}^{(10)}$ and the new field strengths have to be used. The same is true for the expression for $\hat{H}^{(7)}$, Eq. (16.70).

[12] Actually, the mass of the KR 2-form should be determined in the D8 background. The Stückelberg mechanism for higher-rank form potentials underlies many gauged higher-dimensional supergravities, but there are cases in which the theory admits a maximally supersymmetric AdS vacuum with respect to which the forms are massless in spite of the explicit "mass terms" they have in the action.

[13] Actually, the theory can be generalized slightly, admitting the possibility that $\hat{G}^{(0)}$ is only piecewise constant, which is equivalent to the introduction of sources for the dual $\hat{C}^{(9)}$ potential which are D8-branes placed at the discontinuities of $\hat{G}^{(0)}$ [118, 133], but we will not consider this generalization here.

16.3 Further reduction of $N = 2A, d = 10$ SUEGRA to nine dimensions

Now we consider the dimensional reduction of the action of the massless $N = 2A, d = 10$ supergravity Eq. (16.38) to nine dimensions. This reduction should give us the effective field theory of the type-IIA superstring compactified on a circle, which we will later compare with the effective theory of the type-IIB theory on another circle in order to find their relations under T duality.[14]

We could have reduced the 11-dimensional theory directly on a 2-torus, obtaining an equivalent result with manifest invariance under global GL(2, $\mathbb{R}$) transformations (in the Einstein frame), according to the general arguments of Section 11.4. This would facilitate the comparison with the reduction of the $N = 2B, d = 10$ theory in its manifestly SL(2, $\mathbb{R}$)-invariant form [691] since these two symmetries coincide in nine dimensions [125], although they have very different (geometrical and non-geometrical) origins. However, the compactification in two steps is necessary in order to obtain the T-duality relations between the ten-dimensional fields, since these and their physical interpretation are much simpler in the string frame, with string variables.

We start by reducing the bosonic NSNS sector of the action Eq. (16.38). Apart from the fact that we are going to call x the compact coordinate, $A^{(1)}$ the KK vector, and $A^{(2)}$ the winding vector, the result of this reduction was given in Eq. (15.25) and we can use it directly. The only subtlety has to do with the normalization factor: after integration of the compact coordinate $x \in [0, 2\pi\ell_s]$, we obtain

$$\frac{2\pi\ell_s\hat{g}_A^2}{16\pi G_{NA}^{(10)}} = \frac{2\pi\ell_s g_A^2 k_0}{16\pi G_{NA}^{(10)}} = \frac{g_A^2}{16\pi G_{NA}^{(9)}}, \tag{16.76}$$

where we have used

$$g_A = \hat{g}_A k_0^{-\frac{1}{2}}, \qquad G_{NA}^{(9)} = G_{NA}^{(10)}/(2\pi R_x). \tag{16.77}$$

Next, we perform the dimensional reduction of the bosonic RR sector.

16.3.1 Dimensional reduction of the bosonic RR sector

The task of reducing the RR field strengths is simplified very much by using the normalization Eqs. (16.52). We find that the ten-dimensional odd-rank RR potentials split into the following nine-dimensional RR potentials of odd and even rank,

$$\begin{aligned}
\hat{C}^{(2n-1)}{}_{\mu_1\cdots\mu_{2n-1}} &= C^{(2n-1)}{}_{\mu_1\cdots\mu_{2n-1}} + (2n-1)A^{(1)}{}_{[\mu_1}C^{(2n-2)}{}_{\mu_2\cdots\mu_{2n-1}]}, \\
\hat{C}^{(2n-1)}{}_{\mu_1\cdots\mu_{2n-2}\underline{x}} &= C^{(2n-2)}{}_{\mu_1\cdots\mu_{2n-2}},
\end{aligned} \tag{16.78}$$

and the even-rank RR field strengths reduce to nine dimensions according to

$$\hat{G}^{(2n)}{}_{a_1\cdots a_{2n}} = G^{(2n)}{}_{a_1\cdots a_{2n}}, \qquad \hat{G}^{(2n)}{}_{a_1\cdots a_{2n-1}x} = k^{-1}G^{(2n-1)}{}_{a_1\cdots a_{2n-1}}, \tag{16.79}$$

[14] T duality can also be established between the massive $N = 2A, d = 10$ supergravity and $N = 2B, d = 10$ supergravity [118, 691], but, due to lack of space, we will restrict ourselves to the simplest case.

where the nine-dimensional RR field strengths are defined as follows,

$$\begin{aligned} G^{(2n+1)} &= dC^{(2n)} - HC^{(2n-2)} + F^{(2)}C^{(2n-1)}, \\ G^{(2n)} &= dC^{(2n-1)} - HC^{(2n-3)} + F^{(1)}C^{(2n-2)}, \end{aligned} \tag{16.80}$$

and the nine-dimensional gauge transformations that leave them invariant are

$$\begin{aligned} \delta A^{(i)} &= d\Sigma^{(i)}, \\ \delta B &= d\Lambda - d\Sigma^{(1)}A^{(2)} - d\Sigma^{(2)}A^{(1)}, \\ \delta C^{(2n)} &= d\Lambda^{(2n-1)} - H\Lambda^{(2n-3)} - F^{(2)}\Lambda^{(2n-2)}, \\ \delta C^{(2n+1)} &= d\Lambda^{(2n)} - H\Lambda^{(2n-2)} - F^{(1)}\Lambda^{(2n-1)}. \end{aligned} \tag{16.81}$$

Using the notation introduced in Section 16.1.3, we can write the nine-dimensional RR field strengths and gauge transformations in this way:

$$\begin{aligned} G &= dC - HC + F^{(2)}\Pi_{\text{odd}}C + F^{(1)}\Pi_{\text{even}}C. \\ \delta C &= d\Lambda^{(\cdot)} - H\Lambda^{(\cdot)} - F^{(2)}\Pi_{\text{even}}\Lambda^{(\cdot)} - F^{(1)}\Pi_{\text{odd}}\Lambda^{(\cdot)}. \end{aligned} \tag{16.82}$$

The RR kinetic terms in the action reduce as follows:

$$-\frac{\sqrt{|\hat{g}|}}{2\cdot(2n+2)!}\left(\hat{G}^{(2n+2)}\right)^2 = -\frac{\sqrt{|g|}}{2\cdot(2n+2)!}k\left(G^{(2n+2)}\right)^2 + \frac{\sqrt{|g|}}{2\cdot(2n+2)!}k^{-1}\left(G^{(2n+1)}\right)^2. \tag{16.83}$$

The reduction of the Chern–Simons term is straightforward. On putting everything together, after some integrations by parts, we obtain

$$\boxed{\begin{aligned} S = \frac{g_{\text{A}}^2}{16\pi G_{\text{NA}}^{(9)}}\int d^9x\,\sqrt{|g|}\Bigg\{&e^{-2\phi}\Bigg[R - 4(\partial\phi)^2 + \frac{1}{2\cdot 3!}H^2 + (\partial\ \ln\,k)^2 \\ &-\tfrac{1}{4}k^2\left(F^{(1)}\right)^2 - \tfrac{1}{4}k^{-2}\left(F^{(2)}\right)^2\Bigg] - \tfrac{1}{2}\sum_{n=1,\ldots,4}\frac{(-1)^n k^{(-1)^n}}{n!}\left(G^{(n)}\right)^2 \\ &-\frac{1}{2^3\cdot 3^2}\frac{\epsilon}{\sqrt{|g|}}\Big[\partial C^{(3)}\partial C^{(3)}A^{(2)} - 3\partial C^{(3)}\partial C^{(2)}\left(B + A^{(1)}A^{(2)}\right) \\ &+ 6\partial C^{(3)}\partial A^{(1)}C^{(2)}A^{(2)} + 9\partial C^{(2)}\partial A^{(1)}C^{(2)}\left(B + A^{(1)}A^{(2)}\right) \\ &+ 9\partial A^{(1)}\partial A^{(1)}C^{(2)}C^{(2)}A^{(2)}\Big]\Bigg\}. \end{aligned}} \tag{16.84}$$

16.3.2 Dimensional reduction of fermions and supersymmetry rules

We can use the decomposition of ten- into nine-dimensional gamma matrices explained in Appendix B.1.5, which corresponds to the decomposition of ten-dimensional

32-component fermions,

$$\hat{f} = \begin{pmatrix} f^1 \\ f^2 \end{pmatrix}, \tag{16.85}$$

into their two chiral 16-dimensional halves f^1 and f^2. Thus, from each ten-dimensional spinor we obtain a pair of nine-dimensional spinors with internal indices $i = 1, 2$, which we do not write explicitly in general and on which the Pauli matrices act. Now, from the two ten-dimensional spinors $\hat{\lambda}$ and $\hat{\psi}_{\hat{\mu}}$ we obtain three pairs of nine-dimensional spinors, ρ, λ, and ψ_μ. The ten-dimensional supersymmetry parameter $\hat{\epsilon}$ gives a pair of nine-dimensional supersymmetry parameters ϵ. The explicit relations are

$$\begin{aligned} \hat{\psi}_\mu &= \psi_\mu + kA^{(1)}{}_\mu \sigma^3 \rho, & \hat{\lambda} &= \sigma^2(\lambda + \rho), \\ \hat{\psi}_{\underline{x}} &= k\sigma^3\rho, & \hat{\epsilon} &= \epsilon. \end{aligned} \tag{16.86}$$

Observe that all these nine-dimensional spinors are real (we remind the reader that $\hat{\lambda}$ was defined to be imaginary).

Observe also that, in the dimensional reduction, the Dirac conjugates acquire an extra σ^2. For instance

$$\bar{\hat{\epsilon}} = \bar{\epsilon}\sigma^2. \tag{16.87}$$

The dimensional reduction of the supersymmetry rules is a repetition of what we did in the reduction from 11 to ten dimensions and we quote only the final results.

For the NSNS bosons:

$$\boxed{\begin{aligned} &\delta_\epsilon e^a{}_\mu = -i\bar{\epsilon}\Gamma^a\psi_\mu, \qquad \delta_\epsilon k = -ik\bar{\epsilon}\rho, \qquad \delta_\epsilon \phi = -\frac{i}{2}\bar{\epsilon}\lambda, \\ &\delta_\epsilon A^{(1)}{}_\mu = -ik^{-1}\bar{\epsilon}\sigma^3\left(\psi_\mu - \Gamma_\mu\rho\right), \qquad \delta_\epsilon A^{(2)}{}_\mu = -ik\bar{\epsilon}\left(\psi_\mu + \Gamma_\mu\rho\right), \\ &\delta_\epsilon B_{\mu\nu} = -2i\bar{\epsilon}\sigma^3\Gamma_{[\mu}\psi_{\nu]} + \delta_\epsilon A^{(1)}{}_{[\mu}A^{(2)}{}_{\nu]} + \delta_\epsilon A^{(2)}{}_{[\mu}A^{(1)}{}_{\nu]}. \end{aligned}} \tag{16.88}$$

For the RR bosons:

$$\boxed{\begin{aligned} \delta_\epsilon C^{(2n)}{}_{\mu_1\cdots\mu_{2n}} = {}& 2nie^{-\phi}k^{1/2}\bar{\epsilon}\mathcal{P}_{n-1}\Gamma_{[\mu_1\cdots\mu_{2n-1}}\left[\psi_{\mu_{2n}]} - \frac{1}{4n}\Gamma_{\mu_{2n}]}(\lambda - \rho)\right] \\ & - i2n(2n-1)\bar{\epsilon}\sigma^3 C^{(2n-2)}{}_{[\mu_1\cdots\mu_{2n-2}}\Gamma_{\mu_{2n-1}}\psi_{\mu_{2n}]} \\ & - 2n\delta_\epsilon A^{(2)}{}_{[\mu_1}C^{(2n-1)}{}_{\mu_2\cdots\mu_{2n}]}, \\ \delta_\epsilon C^{(2n-1)}{}_{\mu_1\cdots\mu_{2n-1}} & \\ = - i(2n-1)&e^{-\phi}k^{-\frac{1}{2}}\bar{\epsilon}\mathcal{P}_{n-1}\Gamma_{[\mu_1\cdots\mu_{2n-2}}\left[\psi_{\mu_{2n-1}]} - \frac{1}{2(2n-1)}\Gamma_{\mu_{2n-1}]}(\lambda + \rho)\right] \\ - i(2n-1)&(2n-2)\bar{\epsilon}\sigma^3 C^{(2n-3)}{}_{[\mu_1\cdots\mu_{2n-3}}\Gamma_{\mu_{2n-2}}\psi_{\mu_{2n-1}]} \\ - (2n-1)&\delta_\epsilon A^{(1)}{}_{[\mu_1}C^{(2n-2)}{}_{\mu_2\cdots\mu_{2n-1}]}. \end{aligned}} \tag{16.89}$$

For the fermions:

$$
\boxed{
\begin{aligned}
\delta_\epsilon \psi_\mu &= \left\{\partial_\mu - \tfrac{1}{4}\left[\not{\omega}_\mu + \tfrac{1}{2}\not{H}_\mu \sigma^3 + \left(k \not{F}^{(1)}{}_\mu \sigma^3 + k^{-1} \not{F}^{(2)}{}_\mu\right)\right]\right\}\epsilon \\
&\quad + \frac{1}{16} e^{\phi} \sum_{i=1}^{8} \frac{1}{n!} k^{\frac{(-1)^n}{2}} \not{G}^{(n)} \Gamma_\mu \mathcal{P}_{[\frac{n}{2}]+1}\epsilon. \\
\delta_\epsilon \rho &= \left[\tfrac{1}{2} \not{\partial} \ln k + \tfrac{1}{8}\left(k \not{F}^{(1)}\sigma^3 - k^{-1} \not{F}^{(2)}\right)\right]\epsilon \\
&\quad + \frac{1}{16} e^{\phi} \sum_{i=1}^{8} \frac{1}{n!} k^{\frac{(-1)^n}{2}} \not{G}^{(n)} \mathcal{P}_{[\frac{n}{2}]+1}\epsilon, \\
\delta_\epsilon \lambda &= \left\{\not{\partial}\phi - \frac{1}{12}\not{H}\sigma^3 - \tfrac{1}{8}\left(k \not{F}^{(1)}\sigma^3 + k^{-1} \not{F}^{(2)}\right)\right\}\epsilon \\
&\quad + \frac{1}{16} e^{\phi} \sum_{i=1}^{8} (-1)^n \frac{(9-2n)}{n!} k^{\frac{(-1)^n}{2}} \not{G}^{(n)} \mathcal{P}_{[\frac{n}{2}]+1}\epsilon.
\end{aligned}
}
\tag{16.90}
$$

16.4 The effective-field theory of the heterotic string

The full action and supersymmetry transformation rules of the $N = 2A, d = 10$ supergravity theory are invariant under the two $\mathbb{Z}_2$ transformations:

$$
\hat{C}^{(2n+1)} \to -\hat{C}^{(2n+1)}, \qquad \hat{\psi}_{\hat{\mu}} \to \pm\hat{\Gamma}_{11}\hat{\psi}_{\hat{\mu}}, \qquad \hat{\lambda} \to \mp\hat{\Gamma}_{11}\hat{\lambda}, \qquad \hat{\epsilon} \to \pm\hat{\Gamma}_{11}\hat{\epsilon}. \tag{16.91}
$$

These transformations correspond to the 11-dimensional transformation of the compact coordinate[15] $z \to -z$ combined with the transformation $f \to \pm f$ for all the fermions of the theory, which is always a symmetry. Eliminating all the fields which are odd under these transformations is always a consistent truncation of $N = 2A, d = 10$ supergravity that is equivalent, according to the discussion in Section 11.6, to the compactification of 11-dimensional supergravity on the orbifold $S^1/\mathbb{Z}_2$. In the two possible truncations all the RR fields and half of the fermions (a chiral half) are eliminated. The result is a chiral theory that is invariant under supersymmetry transformations generated by a single Majorana–Weyl fermion, i.e. $N = 1, d = 10$ supergravity. The action for the bosonic sector of this theory is that of the common sector Eq. (15.1). As for the supersymmetry transformation rules, defining for all fermions $\hat{f}$

$$
\hat{f} = \hat{f}^{(+)} + \hat{f}^{(-)}, \qquad \hat{\Gamma}_{11}\hat{f}^{(\pm)} \equiv \pm\hat{f}^{(\pm)}, \tag{16.92}
$$

[15] One should remember that the 11-dimensional $\hat{\hat{C}}_{\hat{\hat{\mu}}\hat{\hat{\nu}}\hat{\hat{\rho}}}$ is a pseudotensor.

we obtain, for the two possible truncations,

$$\boxed{\begin{aligned} \delta_{\hat{\epsilon}}\hat{\psi}^{(\pm)}_{\hat{a}} &= \hat{\nabla}^{(\pm)}_{\hat{a}}\hat{\epsilon}^{(\pm)}, \\ \delta_{\hat{\epsilon}}\hat{\lambda}^{(\mp)} &= \left(\not{\partial}\hat{\phi} \pm \frac{1}{12}\,\not{\hat{H}}\right)\hat{\epsilon}^{(\pm)}, \end{aligned}} \tag{16.93}$$

where $\hat{\nabla}^{(\pm)}_{\hat{a}}$ are the covariant derivatives associated with the two torsional spin connections

$$\hat{\Omega}^{(\pm)}_{\hat{a}\hat{b}\hat{c}} = \hat{\omega}_{\hat{a}\hat{b}\hat{c}} \pm \tfrac{1}{2}\hat{H}_{\hat{a}\hat{b}\hat{c}}. \tag{16.94}$$

From the string/M-theoretical point of view, though, this is not the whole story: first of all one expects to obtain the effective-field theory of some $N = 1, d = 10$ superstring theory. There are three of these: type-I SO(32), heterotic with gauge group $E_8 \times E_8$, and heterotic with gauge group SO(32). Hořava and Witten showed in [543, 544] that, on each of the $(1 + 9)$-dimensional boundaries of the compactified spacetime that correspond to the endpoints of the segment $S^1/\mathbb{Z}_2$, there is an E_8 vector supermultiplet, so the orbifold compactification of M theory gives the $E_8 \times E_8$ heterotic-string theory.

This gives an entirely new way to view this string theory which, as is type-IIA, is also related to M theory. For instance, the heterotic-string dilaton measures the distance between the $(1 + 9)$-dimensional boundaries. Also, the dimensional reduction of the heterotic string on a circle can be related to toroidal compactifications of M theory, which are related to type-II string theory. In the end, this will give a relation between heterotic and type-I string theories, as we will discuss in the next chapter.

As for the effective action of the heterotic-string theory, it is obtained by coupling the action of pure $N = 1, d = 10$ supergravity [114] whose bosonic sector is given by Eq. (15.1) to the corresponding vector supermultiplets. In the bosonic sector, this requires the addition of the Yang–Mills kinetic term to the action,

$$\boxed{\hat{S} = \frac{\hat{g}_{\rm h}^{2}}{16\pi G^{(10)}_{\rm N\,h}}\int d^{10}\hat{x}\,\sqrt{|\hat{g}|}\,e^{-2\hat{\phi}}\left[\hat{R} - 4\left(\partial\hat{\phi}\right)^{2} + \frac{1}{2\cdot 3!}\hat{H}^{2} - \tfrac{1}{4}\alpha'\hat{F}^{I}{}_{\hat{\mu}\hat{\nu}}\hat{F}^{I\,\hat{\mu}\hat{\nu}}\right],} \tag{16.95}$$

together with a modification of the KR field strength by the addition of the Yang–Mills Chern–Simons 3-form term defined in Eq. (A.50) [233]:

$$\hat{H} = 3\partial\hat{B} - \tfrac{1}{2}\alpha'\hat{\omega}_3, \qquad \hat{\omega}_3 = -3\hat{A}^{I}\hat{F}^{I} + 2f_{IJK}A^{I}A^{J}A^{K}. \tag{16.96}$$

The supersymmetry transformation rules of the gravitino and dilatino fields are still given by Eqs. (16.93), with $\hat{H}$ as defined above, but we also have to consider that of the gauginos,

$$\boxed{\delta_{\hat{\epsilon}}\hat{\chi}^{I} = -\frac{\sqrt{2\alpha'}}{8}\,\not{\hat{F}}^{I}.} \tag{16.97}$$

16.5 Toroidal compactification of the heterotic string

In this section we are going to study toroidal compactifications of the heterotic-string effective-field theory from $\hat{d} = 10$ to $d = 10 - n$ dimensions. Our goal is to find the T and S dualities that arise in the compactification, especially in $d = 4$ dimensions. In contrast to maximal supergravities (32 supercharges, $N = 2, d = 10$ theories and their toroidal reductions, for example), after dimensional reduction on an n-torus the supergravity multiplet becomes reducible into a lower-dimensional supergravity multiplet and n vector multiplets. We will study how to separate the two kinds of fields. This is important, since matter fields can always be consistently truncated, but they have to be correctly identified in order to preserve supersymmetry.

This dimensional reduction was first done in [226] in the Einstein frame and in [675] in the string frame, in which a stringy interpretation was given to the dualities they found. Here we repeat what they did, first for pure $N = 1, d = 10$ supergravity using our own conventions and emphasizing the relations between ten-dimensional and lower-dimensional fields that will allow us to relate solutions in different dimensions. Later we will add Yang–Mills fields in order to have the complete heterotic-string effective-field theory.

16.5.1 Reduction of the action of pure $N = 1, d = 10$ supergravity

The Ricci scalar and dilaton terms. We can use the notation and *Ansatz* we made for the metric in Section 11.4 and apply immediately Eqs. (11.183)–(11.185), although we have to insert into the first of them the dilaton prefactor $e^{-2\hat{\phi}}$. On defining the $d = (\hat{d} - n)$-dimensional dilaton field by

$$e^{-2\phi} \equiv e^{-2\hat{\phi}} K, \tag{16.98}$$

integrating over the n redundant coordinates, and applying again Palatini's identity to re-express the d-dimensional spin-connection coefficients in terms of the Ricci scalar, we obtain

$$\begin{aligned}&\frac{\hat{g}^2}{16\pi G_{\mathrm{N}}^{(10)}} \int d^{10}\hat{x} \sqrt{|\hat{g}|}\, e^{-2\hat{\phi}} \Big[\hat{R} - 4(\partial\hat{\phi})^2\Big] \\ &\quad = \frac{g^2}{16\pi G_{\mathrm{N}}^{(d)}} \int d^d x \sqrt{|g|}\, e^{-2\phi} \big[R - 4(\partial\phi)^2 + \tfrac{1}{4}F^2 - \tfrac{1}{4}\partial_a G_{mn}\partial^a G^{mn}\big],\end{aligned} \tag{16.99}$$

where the d-dimensional string coupling constant is

$$g = e^{\phi_0} = e^{\hat{\phi}_0} \frac{1}{\sqrt{K_0}} = \hat{g} \left(\frac{V_n}{(2\pi \ell_{\mathrm{s}})^n}\right)^{\frac{1}{2}}, \qquad V_n = (2\pi)^n R_9 \cdots R_{(10-n)}, \tag{16.100}$$

and the d-dimensional Newton constant $G_{\mathrm{N}}^{(d)}$ is related to $G_{\mathrm{N}}^{(10)}$ by

$$G_{\mathrm{N}}^{(d)} = G_{\mathrm{N}}^{(10)} / V_n. \tag{16.101}$$

The KR term. As usual, we define the lower-dimensional KR field strength as identical to the higher-dimensional one in flat indices so the gauge invariance is automatically inherited:

$$H_{abc} \equiv \hat{H}_{abc} = e_a{}^{\mu} e_b{}^{\nu} e_c{}^{\rho} \Big[\hat{H}_{\mu\nu\rho} - 3A^m{}_{[\mu} \hat{H}_{\nu\rho]m} + 3A^m{}_{[\mu} A^n{}_{\nu} \hat{H}_{\rho]mn} \Big]. \tag{16.102}$$

On the other hand, we have another two gauge-invariant combinations,

$$\hat{H}_{abi} = e_a{}^{\mu} e_b{}^{\nu} e_i{}^{m} \Big[\hat{H}_{\mu\nu m} - 2A^n{}_{[\mu} \hat{H}_{\nu]mn} \Big], \tag{16.103}$$

$$\hat{H}_{aij} = e_a{}^{\mu} e_i{}^{m} e_j{}^{n} \hat{H}_{\mu mn}, \tag{16.104}$$

where

$$\hat{H}_{\mu\nu m} = 2\partial_{[\mu} \hat{B}_{\nu]m}, \qquad \hat{H}_{\mu mn} = \partial_{\mu} \hat{B}_{mn}. \tag{16.105}$$

$\hat{H}_{\mu mn}$ is just the field strength for the d-dimensional scalars,

$$B_{mn} = \hat{B}_{mn}. \tag{16.106}$$

Although $\hat{H}_{\mu\nu n}$ looks like a good vector-field strength, it is not gauge-invariant. It enters into the gauge-invariant combination $\hat{H}_{abi}$, which can be rewritten in this way:

$$\hat{H}_{abi} = e_a{}^{\mu} e_b{}^{\nu} e_i{}^{m} \Big[2\partial_{[\mu} \Big(\hat{B}_{\nu]m} + A^n{}_{\nu]} \hat{B}_{nm} \Big) - 2\partial_{[\mu} A^n{}_{\nu]} \hat{B}_{nm} \Big]. \tag{16.107}$$

$2\partial_{[\mu} A^n{}_{\nu]} = F^n{}_{\mu\nu}$ is gauge-invariant by itself, and so the other piece on the r.h.s. must also be gauge-invariant. This suggests the following (re)definition of the d-dimensional vector fields and their strengths

$$\begin{aligned} A^{(1)m}{}_{\mu} &= A^m{}_{\mu}, & F^{(1)m}{}_{\mu\nu} &= 2\partial_{[\mu} A^{(1)m}{}_{\nu]}, \\ A^{(2)}{}_{m\ \mu} &= \hat{B}_{\mu m} - A^n{}_{\mu} \hat{B}_{nm}, & F^{(2)}{}_{m\ \mu\nu} &= 2\partial_{[\mu} A^{(2)}{}_{|m|\nu]}, \end{aligned} \tag{16.108}$$

which leads to

$$\hat{H}_{abi} = e_i{}^{m} \big(F^{(2)}{}_{m\ ab} + F^{(1)n}{}_{ab} B_{nm} \big). \tag{16.109}$$

Using these results in the KR field strength Eq. (16.102), we arrive at the following natural definition for the d-dimensional axion field:

$$B = \hat{B} - A^{(1)m} \hat{B}_{mn} A^{(1)n} + A^{(1)m} A^{(2)}{}_{m}, \tag{16.110}$$

which implies

$$H = 3\partial B - \tfrac{3}{2} A^{(1)m} F^{(2)}{}_{m} - \tfrac{3}{2} A^{(2)}{}_{m} F^{(1)m}. \tag{16.111}$$

Finally, we have

$$\hat{H}^2 = H^2 + 3\hat{H}_{abi} \hat{H}^{abi} + 3\hat{H}_{aij} \hat{H}^{aij} = H^2 + 3\mathcal{F}^2 + 3G^{mn} G^{pq} \partial_{\mu} B_{mp} \partial^{\mu} B_{nq}, \tag{16.112}$$

where we use the shorthand notation

$$\mathcal{F}_m = F^{(2)}{}_m + F^{(1)p} B_{pm}, \qquad \mathcal{F}^2 = G^{mn} \mathcal{F}_m \mathcal{F}_n. \tag{16.113}$$

On putting together the results of this and the previous section, we obtain

$$\boxed{\begin{aligned} S = \frac{g_{\mathrm{h}}^2}{16\pi G_{\mathrm{N}}^{(d)}} \int d^d x \sqrt{|g|}\, e^{-2\phi} \Big\{ & R - 4(\partial\phi)^2 + \frac{1}{2\cdot 3!} H^2 \\ & -\tfrac{1}{4}\left[\partial_\mu G_{mn} \partial^\mu G^{mn} - G^{mn} G^{pq} \partial_\mu B_{mp} \partial^\mu B_{nq}\right] + \tfrac{1}{4}\left(F^{(1)}\right)^2 + \tfrac{1}{4}\mathcal{F}^2 \Big\}. \end{aligned}} \tag{16.114}$$

This is essentially the result we were after, although we will have to massage it a bit more and we will also have to add the fermions to the picture. We also wanted the relation between the ten- and d-dimensional fields in order to be able to reconstruct in ten-dimensional language four-dimensional solutions. If we have such a solution in terms of the fields $g_{\mu\nu}$, $B_{\mu\nu}$, $A^{(1)m}_{\mu}$, $A^{(2)}{}_{m\mu}$, G_{mn}, B_{mn}, and ϕ, the ten-dimensional fields of the corresponding ten-dimensional solutions are given by

$$\begin{aligned} \hat{g}_{\mu\nu} &= g_{\mu\nu} + A^{(1)m}{}_\mu A^{(1)n}{}_\nu G_{mn}, & \hat{g}_{mn} &= G_{mn}, \\ \hat{B}_{\mu\nu} &= B_{\mu\nu} + A^{(1)m}{}_\mu A^{(1)n}{}_\nu B_{mn} - A^{(1)m}{}_{[\mu} A^{(2)}{}_{|m|\nu]}, & \hat{B}_{mn} &= B_{mn} \\ \hat{B}_{\mu m} &= A^{(2)}{}_{m\mu} + A^{(1)n}{}_\mu B_{nm}, & \hat{g}_{\mu m} &= A^{(1)n}{}_\mu G_{nm}, \\ \hat{\phi} &= \phi + \tfrac{1}{4}\ln(\det G). \end{aligned} \tag{16.115}$$

Manifestly O(n,n)*-symmetric action.* The d-dimensional action that we have just obtained has a global GL$(n,\mathbb{R})$ invariance that acts on the indices associated with the compact dimensions. In particular, the O(n) subgroup acts irreducibly on the vectors $A^{(1)m}$ and $A^{(2)}{}_m$ without mixing them. The same happens for the scalars G_{mn} and B_{mn}. We know, though, that, in the $n=1$ case, Eq. (15.25), the action is invariant under the $\mathbb{Z}_2$ T-duality transformations Eqs. (15.27) that interchange these fields and which, combined with the SO(1,1) rescalings of the fields, form an O(1, 1) duality group. We expect now that all the KK and winding vectors $A^{(1)m}$ and $A^{(2)}{}_m$ can be interchanged independently and also rotated. The resulting T-duality group will be O(n,n) but this cannot be seen with the action written as above.

To make O(n,n) manifest, following [675] define the matrices $G \equiv (G_{mn})$ and $B \equiv (B_{mn})$, construct the $2n \times 2n$ symmetric matrix

$$M = \begin{pmatrix} -G^{-1} & G^{-1}B \\ -BG^{-1} & -G + BG^{-1}B \end{pmatrix}, \qquad M^{-1} = \begin{pmatrix} -G + BG^{-1}B & -BG^{-1} \\ G^{-1}B & -G^{-1} \end{pmatrix}, \tag{16.116}$$

and introduce the $2n \times 2n$ matrix L,

$$L = \begin{pmatrix} 0 & \mathbb{I}_{n\times n} \\ \\ \mathbb{I}_{n\times n} & 0 \end{pmatrix} = L^{-1}, \tag{16.117}$$

which is nothing but the $\mathrm{O}(n,n)$ metric $(\mathrm{diag}(+,\cdots,+,-,\cdots,-))$ in a non-diagonal form.[16] The essential property of M is that, as we can see in Eq. (16.116),

$$LML = M^{-1}, \quad \Rightarrow M^{\mathrm{T}}LM = L, \tag{16.118}$$

which is the definition of an $\mathrm{O}(n,n)$ matrix in this non-diagonal basis, i.e. $M \in \mathrm{O}(n,n)$. Actually, M parametrizes the coset space $\mathrm{O}(n,n)/(\mathrm{O}(n)\times\mathrm{O}(n))$, as can be seen by counting the number of independent scalars and comparing it with the dimension of the coset space and also in the construction of Section 16.5.2.

Using the cyclic property of the trace and Eq. (16.118), we can rewrite the kinetic term of the scalars in the action Eq. (16.114) in this manifestly $\mathrm{O}(n,n)$-invariant way:

$$\tfrac{1}{4}\left[\partial_\mu G_{mn}\partial^\mu G^{mn} - G^{mn}G^{pq}\partial_\mu B_{mp}\partial^\mu B_{nq}\right] = \tfrac{1}{8}\mathrm{Tr}(\partial ML\partial ML). \tag{16.119}$$

The scalars are coupled to the vectors and we also need to rewrite their kinetic terms. Defining the $\mathrm{O}(n,n)$ column vectors

$$A^\Sigma = \begin{pmatrix} A^{(1)m}{}_\mu \\ A^{(2)}{}_{m\ \mu} \end{pmatrix}, \qquad F^\Sigma = 2\partial A^\Sigma, \tag{16.120}$$

we can rewrite the kinetic term with them as follows:

$$\tfrac{1}{4}\left(F^{(1)}\right)^2 + \tfrac{1}{4}\mathcal{F}^2 = -\tfrac{1}{4}\left(M^{-1}\right)_{\Sigma\Lambda} F^\Sigma F^\Lambda = -\tfrac{1}{4}(LML)_{\Sigma\Lambda}F^\Sigma F^\Lambda, \tag{16.121}$$

and the KR field strength in the form

$$H = 3\partial B - \tfrac{3}{2}L_{\Sigma\Lambda}A^\Sigma F^\Lambda, \tag{16.122}$$

and we arrive at the following $\mathrm{O}(n,n)$-invariant action:

$$\boxed{\begin{aligned} S = \frac{g_{\mathrm{h}}^2}{16\pi G_{\mathrm{N}}^{(d)}}\int d^dx\sqrt{|g|}\,e^{-2\phi}\Bigg\{ & R - 4(\partial\phi)^2 + \frac{1}{2\cdot 3!}H^2 \\ & - \tfrac{1}{8}\mathrm{Tr}\,(\partial ML\partial ML) - \tfrac{1}{4}(LML)_{\Sigma\Lambda}F^\Sigma F^\Lambda\Bigg\}. \end{aligned}} \tag{16.123}$$

The scalars M and vectors transform under $\Omega \in \mathrm{O}(n,n)$ according to

$$M' = \Omega M \Omega^{\mathrm{T}}, \qquad F' = \Omega F. \tag{16.124}$$

---

[16] We reserve the symbol η for the diagonal form of the $\mathrm{O}(n,n)$ metric.

In the supergravity literature [134, 241, 298] η, the diagonal metric of $\mathrm{O}(n,n)$, is used instead of L. The diagonalization is important since, as we are going to see, it distinguishes vector fields in the supergravity multiplet (*graviphotons*) from vectors in the matter supermultiplets. To do this, it suffices to perform a change of basis:

$$\mathcal{A} = RA, \qquad \mathcal{F} = RF, \qquad \mathcal{M} = (R^{-1})^{\mathrm{T}} M R^{-1}, \tag{16.125}$$

$$\eta = (R^{-1})^{\mathrm{T}} L R^{-1} = \begin{pmatrix} \mathbb{I}_{n\times n} & 0 \\ 0 & -\mathbb{I}_{n\times n} \end{pmatrix}, \tag{16.126}$$

$$R = (R^{-1})^{\mathrm{T}} = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbb{I}_{n\times n} & \mathbb{I}_{n\times n} \\ -\mathbb{I}_{n\times n} & \mathbb{I}_{n\times n} \end{pmatrix}. \tag{16.127}$$

The action and the KR field strength take the same form with L replaced by η and the vector and scalar fields A and M replaced by $\mathcal{A}$ and $\mathcal{M}$.

Manifestly $\mathrm{O}(p, p)$*-covariant equations of motion.* It is clear that the equations of motion for the metric, KR, dilaton, and vector fields are automatically $\mathrm{O}(p, p)$-covariant. However, $\mathcal{M}$ is a constrained matrix and its equations of motion have to be calculated with care. It can be shown that they can be put into the manifestly $\mathrm{O}(p, p)$-covariant form

$$\nabla_\mu \big(e^{-2\phi} \mathcal{M}^{-1} \partial^\mu \mathcal{M}\big) = \tfrac{1}{2} e^{-2\phi} \big(\eta \mathcal{M} \eta \mathcal{F} \mathcal{F}^{\mathrm{T}} - \eta \mathcal{F} \mathcal{F}^{\mathrm{T}} \eta \mathcal{M}\big). \tag{16.128}$$

This equation transforms in the adjoint representation of $\mathrm{O}(p, p)$.

16.5.2 Reduction of the fermions and supersymmetry rules of $N = 1, d = 10$ SUGRA

To reduce the fermions and the supersymmetry rules, we need to construct Vielbeins for the σ-model scalars, i.e. Vielbeins in the coset space $\mathrm{O}(n, n)/(\mathrm{O}(n) \times \mathrm{O}(n))$. We define the matrix $E \equiv (e^i{}_m)$ so $E^{-1} = (e^m{}_i)$ and $E^{\mathrm{T}} E = -G$ and construct the $2n \times 2n$ matrix

$$V \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} -E + (E^{-1})^{\mathrm{T}} B & -(E^{-1})^{\mathrm{T}} \\ -E - (E^{-1})^{\mathrm{T}} B & (E^{-1})^{\mathrm{T}} \end{pmatrix}, \tag{16.129}$$

with inverse

$$V^{-1} = -\frac{1}{\sqrt{2}} \begin{pmatrix} E^{-1} & E^{-1} \\ E^{\mathrm{T}} + B E^{-1} & -E^{\mathrm{T}} + B E^{-1} \end{pmatrix}. \tag{16.130}$$

This matrix satisfies

$$V^{\mathrm{T}} V = M^{-1}, \qquad V^{-1} (V^{-1})^{\mathrm{T}} = M, \tag{16.131}$$

and transforms under global $\Omega \in \mathrm{O}(n, n)$ transformations on the right and local $N \in$

$O(n) \times O(n)$ on the left,

$$V' = NV\Omega. \tag{16.132}$$

It is natural to use indices $V = \left(V^A{}_\Sigma\right)$ and $V^{-1} = \left(V^\Sigma{}_A\right)$ where $A, B = 1, \ldots, 2n$. On the other hand, the change of basis from L to the diagonal metric η is

$$\mathcal{V} \equiv VR^{-1}. \tag{16.133}$$

The combinations $\mathcal{V}\mathcal{F}$ and $\mathcal{V}\partial\mathcal{V}^{-1}$ which are invariant under global $O(n, n)$ transformations appear naturally in the reduction of the fermionic supersymmetry transformation rules. We are interested only in the purely bosonic transformation rules (with all fermions set to zero). The reduction is made in two steps: first one reduces all the tensor fields, then one decomposes ten-dimensional 32-component spinors into d-dimensional spinors with extra internal indices and ten-dimensional gamma matrices into tensor products of d-dimensional gamma matrices and matrices associated with the internal symmetries. The second step depends strongly on n and, thus, we are going to perform only the first step, with the ultimate goal of finding the right truncation of the matter vector fields.

It is convenient to split the $2n$-dimensional indices A, B into A_1, B_1 running from 1 to n and A_2, B_2 that take values between $n+1$ and $2n$. Furthermore, in order to indicate the correct contractions with the gamma matrices, we have defined the $2n$ "vector"

$$\hat{\Gamma}^A = (\hat{\Gamma}^{d+1}, \ldots, \hat{\Gamma}^{d+p}, \hat{\Gamma}^{d+1}, \ldots, \hat{\Gamma}^{d+p}). \tag{16.134}$$

The result of the reduction is, then,

$$\begin{aligned}
\delta_{\hat{\epsilon}}\hat{\psi}_a^{(+)} &= \nabla_a^{(+)}\,\hat{\epsilon}^{(+)} + \frac{\sqrt{2}}{4}\hat{\Gamma}_{A_1}\mathcal{V}^{A_1}{}_\Sigma\,\mathcal{F}^\Sigma{}_a\hat{\epsilon}^{(+)} \\
&\quad - \tfrac{1}{4}\hat{\Gamma}_{A_1}\mathcal{V}^{A_1}{}_\Sigma\partial_a\mathcal{V}^\Sigma{}_{B_1}\hat{\Gamma}^{B_1}\hat{\epsilon}^{(+)}, \\
\delta_{\hat{\epsilon}}\hat{\psi}^{(+)\,A_2} &= -\frac{\sqrt{2}}{8}\mathcal{V}^{A_2}{}_\Sigma\,\not{\mathcal{F}}^\Sigma\,\hat{\epsilon}^{(+)} - \tfrac{1}{2}\mathcal{V}^{A_2}{}_\Sigma\,\not{\partial}\mathcal{V}^\Sigma{}_{A_1}\hat{\Gamma}^{A_1}\,\hat{\epsilon}^{(+)}, \\
\delta_{\hat{\epsilon}}(\hat{\lambda}^{(-)} - \hat{\Gamma}^i\hat{\psi}_i^{(+)}) &= \left(\not{\partial}\phi - \frac{1}{12}\not{H}\right)\hat{\epsilon}^{(+)} + \frac{\sqrt{2}}{8}\hat{\Gamma}_{A_1}\mathcal{V}^{A_1}{}_\Sigma\,\not{\mathcal{F}}^\Sigma\,\hat{\epsilon}^{(+)}.
\end{aligned} \tag{16.135}$$

It is clear that $\hat{\psi}_a^{(+)}$ will split into several d-dimensional gravitinos (four in $d = 4$, since the ten-dimensional chiral spinors $\hat{\psi}_a^{(+)}$ have 16 real components for each a, which corresponds to $N = 4, d = 4$ SUEGRA) and the combination $\hat{\lambda}^{(-)} - \hat{\Gamma}^i\hat{\psi}_i^{(+)}$ will split into several d-dimensional dilatinos (again four in $d = 4$) since they transform into the dilaton under supersymmetry. The n spinors $\hat{\psi}^{(+)\,A_2}$ transform into vectors and so they will split into d-dimensional gauginos ($4n$ in $d = 4$) of the n vector supermultiplets. The supersymmetry parameter splits into as many d-dimensional supersymmetry parameters as the gravitino, giving the number N of independent supersymmetry transformations that can be performed ($N = 4$ in $d = 4$).

These supersymmetry transformation rules are clearly covariant under $\mathrm{O}(n, n)$ T-duality transformations. This means that any d-dimensional solution will have the same number of unbroken supersymmetries after an $\mathrm{O}(n, n)$ rotation. The corresponding ten-dimensional solutions may but need not have the same amount of supersymmetry. This is due to the fact that unbroken supersymmetry can be broken in dimensional reduction. We are going to discuss this subtle point in Section 16.6, but obviously it applies to many other situations: for instance the relation between the unbroken supersymmetries of $N = 1, d = 11$ and $N = 2A, d = 10$ supergravity solutions.

16.5.3 The truncation to pure supergravity

The fields of the reduced theory correspond to pure supergravity (16 supercharges in d dimensions) coupled to n vector supermultiplets. The fields in the supergravity multiplet are the gravitinos $\hat{\psi}_a^{(+)}$, the dilatino $\hat{\lambda}^{(-)} - \hat{\Gamma}^i \hat{\psi}_i^{(+)}$, the graviton $e^a{}_\mu$, the dilaton ϕ, the KR 2-form $B_{\mu\nu}$, and n of the $2n$ KK and winding vectors A^Σ. The n vector supermultiplets are made out of the n^2 scalars contained in $\mathcal{V}^A{}_\Sigma$, n of the $2n$ KK and winding vectors A^Σ, and the gauginos $\hat{\psi}^{(+)\ A_2}$. Thus, we know to which supermultiplet each field belongs, except for the vector fields. These, however, can be identified by studying the truncation of the vector multiplets, which consists in

$$E = \mathbb{I}_{n\times n}, \qquad B = 0, \qquad \hat{\psi}^{(+)\ A_2} = 0, \tag{16.136}$$

plus the vanishing of the matter vector fields. Since the truncation has to be consistent at the level of the equations of motion, if we substitute the above values of the fields into the equations of motion of the theory, we will be forced to set to zero n combinations of the $2n$ vector fields A^Σ, which are then identified with the matter vector fields. The n orthogonal combinations that remain are the supergravity vector fields.

Substituting $\mathcal{M} = \mathbb{I}_{2n\times 2n}$ into Eq. (16.128) tells us only that $\mathcal{F}^{\Sigma_1}\mathcal{F}^{\Sigma_2} = 0$, though, and we also have to impose consistency of the truncation of the supersymmetry transformation rules Eqs. (16.135). On substituting $\mathcal{V} = 0$ and $\delta_\epsilon \hat{\psi}^{(+)\ A_2} = 0$, we find

$$\mathcal{F}^{\Sigma_2} = 0 \quad \Rightarrow -\frac{1}{\sqrt{2}}(F^{\Sigma_1} - F^{\Sigma_2}) = 0, \tag{16.137}$$

which implies that the combinations $\mathcal{F}^{\Sigma_2}$ are the matter vector fields and the $\mathcal{F}^{\Sigma^1}$ are the supergravity ones.

The action of the truncated pure supergravity d-dimensional theory is

$$\boxed{S = \frac{g_{\mathrm{h}}^2}{16\pi G_{\mathrm{N}}^{(d)}} \int d^d x \sqrt{|g|}\, e^{-2\phi}\left[R - 4(\partial\phi)^2 + \frac{1}{2\cdot 3!}H^2 - \tfrac{1}{4}\mathcal{F}^{\Sigma_1}\mathcal{F}^{\Sigma_2}\right],} \tag{16.138}$$

where

$$H = 3\partial B - \tfrac{3}{2}\mathcal{A}^{\Sigma_1}\mathcal{F}^{\Sigma_1}. \tag{16.139}$$

In $d = 4$ (the $n = 6$ case), as we are going to see in Section 16.5.5, the KR 2-form can be dualized into a pseudoscalar axion field a and the action is exactly as we wrote it in Eq. (12.58) (up to notational details and the rescaling to the Einstein frame). Then, any solution of $N = 4, d = 4$ SUEGRA like those we studied in Chapter 12 can be oxidized to a solution of $N = 1, d = 10$ supergravity and understood as a string solution. Furthermore, if we take into account that the four-dimensional supergravity vectors are

$$\mathcal{A}^m = \frac{1}{\sqrt{2}}(A^{(1)\,m} + A^{(2)}{}_m), \tag{16.140}$$

and the truncation condition implies $A^{(1)\,m} = A^{(2)}{}_m$, we will find that all supersymmetric four-dimensional solutions are also supersymmetric using the ten-dimensional supersymmetry transformation rules.

16.5.4 Reduction with additional U(1) *vector fields*

Let us now consider the reduction of the the full action Eq. (16.95). We are going to do this for generic points of the moduli space of toroidal compactifications at which, as discussed on page 423, the symmetry group is the Abelian $U(1)^{16}$, but here we will keep the number of vector fields generic, p. Observe that the Yang–Mills terms appear in the heterotic-string effective action at first order in α' and we will explicitly exhibit this constant.

Before we work out this generic case, we should mention a particular but most interesting case: the $E_8 \times E_8$ and SO(32) are related by T duality after compactification on a circle.

All we need to reduce now is the KR and vector fields. We just quote the results, since the procedure followed is the same as before. The p vector fields give p vector fields $A^I{}_\mu$ and $p \times n$ scalars $a^I{}_m$:

$$\hat{A}^I{}_\mu = \frac{1}{\sqrt{\alpha'}}A^I{}_\mu + \hat{A}^I{}_m A^{(1)\,m}{}_\mu, \qquad \hat{A}^I{}_m = \frac{1}{\sqrt{\alpha'}}a^I{}_m. \tag{16.141}$$

The KR 2-form gives the same fields in d dimensions but with different definitions:

$$\begin{aligned}
\hat{B}_{\mu\nu} &= B_{\mu\nu} + \hat{B}_{mn}A^{(1)\,m}{}_\mu A^{(1)\,n}{}_\nu - A^{(1)\,m}{}_{[\mu|}A^{(2)}{}_{m|\nu]} - a^I{}_m A^{(1)\,m}{}_{[\mu}A^I{}_{\nu]},\\
\hat{B}_{\mu\,m} &= A^{(2)}{}_{m\,\mu} + A^{(1)\,n}{}_\mu B_{nm} + \tfrac{1}{2}a^I{}_m A^I{}_\mu,\\
\hat{B}_{mn} &= B_{mn}.
\end{aligned} \tag{16.142}$$

The vector-field strength decomposes as

$$\begin{aligned}
\hat{F}^I{}_{ab} &= \frac{1}{\sqrt{\alpha'}}\big(F^I{}_{ab} + a^I{}_m F^{(1)\,m}{}_{ab}\big),\\
\hat{F}^I{}_{ai} &= \frac{1}{\sqrt{\alpha'}}e_i{}^m \partial_a a^I{}_m,
\end{aligned} \tag{16.143}$$

where ${F^I}_{\mu\nu} = 2\partial_{[\mu}{A^I}_{\nu]}$, whereas the KR field strength decomposes as

$$\begin{aligned}
\hat{H}_{aij} &= e_i{}^m e_j{}^n \left(\partial_a B_{mn} - \tfrac{1}{2}a^I{}_m \partial_a a^I{}_n + \tfrac{1}{2}a^I{}_n \partial_a a^I{}_m\right),\\
\hat{H}_{abi} &= e_i{}^m \left(F^{(2)}{}_{m\,ab} - C_{mn} F^{(1)\,n}{}_{ab} + a^I{}_m F^I{}_{ab}\right),\\
\hat{H}_{abc} &= H_{abc},
\end{aligned} \tag{16.144}$$

where the d-dimensional KR field strength and the scalars C_{mn} are given by

$$H_{\mu\nu\rho} = 3\partial_{[\mu} B_{\nu\rho]} - \tfrac{3}{2} L_{\Sigma\Lambda} A^{\Sigma}_{[\mu} F^{\Lambda}{}_{\nu\rho]}, \qquad C_{mn} = B_{mn} - \tfrac{1}{2} a^I{}_m a^I{}_n, \tag{16.145}$$

we have defined the $(2n + p)$-dimensional vector

$$\left(A^{\Sigma}{}_{\mu}\right) = \begin{pmatrix} A^{(1)m}{}_{\mu} \\ A^{(2)}{}_{m\ \mu} \\ A^I{}_{\mu} \end{pmatrix}, \qquad F^{\Sigma}{}_{\mu\nu} = 2\partial_{[\mu} A^{\Sigma}{}_{\nu]}, \qquad \Sigma = 1, \ldots, 2n + p, \tag{16.146}$$

and $L_{\Sigma\Lambda}$ is the $\mathrm{O}(p, p + n)$ metric in a non-diagonal basis:

$$(L_{\Sigma\Lambda}) = \begin{pmatrix} 0 & \mathbb{I}_{p\times p} & 0 \\ \mathbb{I}_{p\times p} & 0 & 0 \\ 0 & 0 & -\mathbb{I}_{n\times n} \end{pmatrix}. \tag{16.147}$$

On putting everything together, we obtain an action of the form Eq. (16.123) but with the fields and $L_{\Sigma\Lambda}$ defined above and the matrix M now of dimension $(2n + p) \times (2n + p)$ parametrizing an $\mathrm{O}(n, n + p)/(\mathrm{O}(n) \times \mathrm{O}(n + p))$ coset space and given by

$$M = \begin{pmatrix} -G^{-1} & G^{-1}C & G^{-1}a^{\mathrm{T}} \\ C^{\mathrm{T}}G^{-1} & -G - C^{\mathrm{T}}G^{-1}C + a^{\mathrm{T}}a & -C^{\mathrm{T}}G^{-1}a^{\mathrm{T}} + a^{\mathrm{T}} \\ aG^{-1} & -aG^{-1}C + a & \mathbb{I}_{p\times p} - aG^{-1}a^{\mathrm{T}} \end{pmatrix}. \tag{16.148}$$

It can be constructed with the Vielbein

$$V \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} -E + (E^{-1})^{\mathrm{T}}C & -(E^{-1})^{\mathrm{T}} & -(E^{-1})^{\mathrm{T}}a^{\mathrm{T}} \\ -E - (E^{-1})^{\mathrm{T}}C & (E^{-1})^{\mathrm{T}} & (E^{-1})^{\mathrm{T}}a^{\mathrm{T}} \\ \sqrt{2}\,a & 0 & \sqrt{2}\,\mathbb{I}_{p\times p} \end{pmatrix}, \qquad V^{\mathrm{T}}V = M^{-1}. \tag{16.149}$$

All the properties enjoyed by the old M as an $\mathrm{O}(p, p)$ matrix are now enjoyed by the new M as an $\mathrm{O}(p, p + n)$ matrix with the new metric L. This metric can also be diagonalized to $\eta = \mathrm{diag}(\mathbb{I}_{n\times n}, -\mathbb{I}_{(n+p)\times(n+p)})$ by the same matrix R given in Eq. (16.127) acting only on the first $2n$ indices. Clearly, all the vector fields associated with the $n + p$ negative eigenvalues of η are matter vector fields. The truncation to pure supergravity now includes the conditions

$$a^I{}_m = A^I{}_\mu = 0. \tag{16.150}$$

The T-duality group is now $O(n, n + p)$ and includes the interchange of KK and winding vectors with the U(1) gauge vectors. This is not too surprising if we take into account that the gauge fields of the heterotic string originate from the compactification of the right- or left-moving part of 16 worldsheet scalars.

16.5.5 *Trading the KR 2-form for its dual*

As we mentioned in the introduction, in certain dimensions, the symmetry of the compactified theory can be bigger than $O(n, n + p)$, for instance due to the possibility of dualizing fields that can be rotated into other already existing fields. Here we are going to see an important example in $d = 4$ dimensions, in which the heterotic-string KR 2-form can be dualized into a pseudoscalar axion field, which, together with the dilaton, parametrizes an $SL(2, \mathbb{R})/SO(2)$ coset space, Eq. (11.209) (the one present in $N = 4, d = 4$ supergravity, studied in Section 12.2). It turns out that the equations of motion (but not the action) of the full theory are $SL(2, \mathbb{R})$-covariant because the $SL(2, \mathbb{R})$ transformations involve the dualization of the vector fields (which are dual to vector fields precisely in $d = 4$). This new hidden symmetry of the supergravity theory has been conjectured to be a non-perturbative S duality of the full heterotic-string theory [397, 803, 842].

We are also interested in the dualization of the KR 2-form in $d = 6$, in which one obtains another 2-form potential. A transformation of the dilaton and metric brings the theory into the form of the theory that one obtains by compactifying $N = 2A, d = 10$ supergravity on K3, which is evidence of a strong–weak-coupling duality between the full heterotic-string theory compactified on T^4 and the full type-IIA string theory compactified on K3.

The SO(32) heterotic string is also related by another strong–weak-coupling duality to the type-I SO(32) superstring but the relation does not involve the dualization of the (NSNS) KR 2-form, but rather its interchange by a RR 2-form, as we will see in Section 17.5.

Then we are going to perform the dualization of the KR 2-form in arbitrary dimension d. The general procedure for Poincaré dualizations is explained in Section 8.7.1: we consider the action as a functional of H and add a Lagrange multiplier term to enforce the Bianchi identity

$$E^{\mu_1\cdots\mu_{6-n}} \equiv \nabla_\mu {}^\star H^{\mu\mu_1\cdots\mu_{6-n}} + \frac{(-1)^{6-n}}{4}\eta_{\Sigma\Lambda}\mathcal{F}^\Sigma{}_{\nu\rho}{}^\star\mathcal{F}^{\Lambda\nu\rho\mu_1\cdots\mu_{6-n}} = 0. \tag{16.151}$$

The Lagrange-multiplier term that has to be added to Eq. (16.123) (diagonalized, so L is replaced by η, F by $\mathcal{F}$, etc.) is

$$\frac{g_{\rm h}^2}{16\pi G_{\rm N}^{(d)}}\int d^dx\sqrt{|g|}\,\frac{1}{(6-n)!}\tilde{B}_{\mu_1\cdots\mu_{6-n}}E^{\mu_1\cdots\mu_{6-n}}. \tag{16.152}$$

Now we want to eliminate H completely from the action by using its equation of motion:

$$H^{\nu_1\nu_2\nu_3} = \frac{1}{(7-n)!} e^{2\phi} \frac{1}{\sqrt{|g|}} \epsilon^{\mu_1\cdots\mu_{7-n}\nu_1\nu_2\nu_3} \tilde{H}_{\mu_1\cdots\mu_{7-n}}, \tag{16.153}$$

where

$$\tilde{H} = (7-n)\partial\tilde{B} \tag{16.154}$$

is the $(7-n)$-form dual to H.

Since B appears in the action only through H, B is completely eliminated from this action and replaced by $\tilde{B}$, and we obtain the dual action

$$\boxed{\begin{aligned} S = \frac{g_{\rm h}^2}{16\pi G_{\rm N}^{d)}} \int d^dx \sqrt{|g|}\, e^{-2\phi} \Big\{ & R - 4(\partial\phi)^2 + \frac{(-1)^d}{2\cdot(7-n)!} e^{4\phi}\tilde{H}^2 - \tfrac{1}{8}\,\mathrm{Tr}\,(\partial\mathcal{M}\eta\partial\mathcal{M}\eta) \\ & - \tfrac{1}{4}(\eta\mathcal{M}\eta)_{\Sigma\Lambda}\,\mathcal{F}^\Sigma\mathcal{F}^\Lambda \\ & + \frac{(-1)^{6-p}}{8\cdot(6-p)!}\eta_{\Sigma\Lambda}\frac{1}{\sqrt{|g|}}\epsilon\tilde{B}\mathcal{F}^\Sigma\mathcal{F}^\Lambda \Big\}. \end{aligned}} \tag{16.155}$$

Now we study two special cases.

The $n=4$, $d=6$ case. In this case $\tilde{H}$ is just another 3-form field strength. The field content of the effective action coincides with the massless sector of the type-IIA string compactified on K3 and, in fact, there is a field redefinition that takes us from the massless fields of the heterotic string compactified on T^4 to the massless fields of the type-IIA string compactified on K3, which supports the duality between the two theories (including the massive modes) [332, 336, 583, 847, 962]:

$$\tilde{H} = H_{\rm IIA}, \qquad \phi = -\phi_{\rm IIA}, \qquad g_{\mu\nu} = e^{-2\phi_{\rm IIA}} g_{\rm IIA\,\mu\nu}. \tag{16.156}$$

On performing this change of variables, we obtain

$$\boxed{\begin{aligned} S = \frac{g_{\rm h}^2}{16\pi G_{\rm N}^{(6)}} \int d^6x \sqrt{|g_{\rm IIA}|}\, \Big\{ e^{-2\phi_{\rm IIA}} \Big[& R_{\rm IIA} - 4(\partial\phi_{\rm IIA})^2 + \frac{1}{2\cdot 3!} H_{\rm IIA}^2 \\ & - \tfrac{1}{8}\,\mathrm{Tr}(\partial\mathcal{M}\eta\partial\mathcal{M}\eta)\Big] - \tfrac{1}{4}\,(\eta\mathcal{M}\eta)_{\Sigma\Lambda}\,\mathcal{F}^\Sigma\mathcal{F}^\Lambda + \frac{1}{16}\eta_{\Sigma\Lambda}\frac{1}{\sqrt{|g_{IIA}|}}\epsilon B_{IIA}\mathcal{F}^\Sigma\mathcal{F}^\Lambda \Big\}. \end{aligned}} \tag{16.157}$$

Observe that, in these variables, the vector fields do not carry the factor $e^{-2\phi_{\rm IIA}}$. This is due to the fact that all of them are of RR origin.

The $n = 6$, $d = 4$ case. To exhibit the new symmetry, we first go to the Einstein frame by rescaling the metric,

$$g_{\mu\nu} = e^{2\phi} g_{E\mu\nu}. \tag{16.158}$$

Using the formulae in Appendix E to rescale the Ricci scalar and defining the complex scalar field τ that parametrizes the coset space $\mathrm{SL}(2, \mathbb{R})/\mathrm{SO}(2)$,

$$\tau \equiv a + ie^{-2\phi}, \qquad a \equiv \tilde{B}, \tag{16.159}$$

we obtain the action of $N = 4$, $d = 4$ SUEGRA coupled to p vector multiplets:

$$\boxed{\begin{aligned} S = \frac{1}{16\pi G_{\mathrm{N}}^{(4)}} \int d^4x \sqrt{|g_{\mathrm{E}}|} \, \Big\{ & R_{\mathrm{E}} - \tfrac{1}{2} \frac{\partial_\mu \tau \partial^\mu \overline{\tau}}{(\mathrm{Im}(\tau))^2} - \tfrac{1}{8} \mathrm{Tr}(\partial \mathcal{M} \eta \partial \mathcal{M} \eta) \\ & - \tfrac{1}{4} e^{-2\phi} \, (\eta \mathcal{M} \eta)_{\Sigma\Lambda} \, \mathcal{F}^\Sigma \mathcal{F}^\Lambda + \tfrac{1}{8} \eta_{\Sigma\Lambda} \frac{1}{\sqrt{|g|}} a \epsilon \mathcal{F}^\Sigma \mathcal{F}^\Lambda \Big\}. \end{aligned}} \tag{16.160}$$

The truncation of the vector fields $\mathcal{A}^\Sigma = \Sigma = 7, \ldots, 12 + p$ and the scalar fields $\mathcal{M} = \mathbb{I}_{(12+p)\times(12+p)}$ takes us to the pure supergravity action Eq. (12.58). The truncated action is invariant under SO(6) rotations of the vector fields, which are associated with rotations[17] in the internal T^6 and, as discussed in Section 12.2, the equations of motion are covariant under $\mathrm{SL}(2, \mathbb{R})$ transformations of τ, a non-perturbative symmetry that will not have a simple interpretation until we introduce the solitonic 5-brane.

16.6 T duality, compactification, and supersymmetry

The hidden symmetries of supergravity theories that we have studied can be used to transform 11- or ten-dimensional solutions with the appropriate number of isometries using one of the mechanisms we described in Chapter 11 to generate new solutions: reduce, use the d-dimensional hidden symmetry transformation, and oxidize again. The T-duality Buscher rules of the string common sector can be interpreted as the simplest application of this mechanism, using the $\mathbb{Z}_2$ symmetry of the theory reduced on a circle.

On the other hand, these hidden symmetries of supergravity theories are evidently symmetries of the supersymmetry transformation rules, which means that they preserve the unbroken supersymmetries of the d-dimensional solutions, acting covariantly on their Killing spinors.

This may lead us to think that the whole procedure of reduction–dualization–oxidation preserves the unbroken supersymmetry of the 11- or 10-dimensional solutions. There is, however, one loophole in all these arguments: unbroken supersymmetry has to be preserved by dimensional reduction and this requires that the 11- or 10-dimensional Killing spinors

[17] This SO(6) is part of the original T-duality group $\mathrm{O}(6, 6 + p)$, but does not contain any interchanges of winding and KK vectors, which are constrained to be equal by the truncation conditions.

are independent of the internal coordinates, but this need not be true *even if all the bosonic fields of the solution are independent of them*. We have indeed assumed that the supersymmetry parameters are independent of the internal coordinates to obtain the supersymmetry transformation rules of the d-dimensional theories and a Killing spinor depending on the internal coordinates simply would not appear as a d-dimensional Killing spinor, the d-dimensional solution would not be supersymmetric and, after dualization and oxidation, in general, the new 11- or ten-dimensional solution will not be supersymmetric either. In those cases, duality does not respect supersymmetry [32, 68, 131, 340, 504].

This is a very interesting phenomenon with potentially important implications and it is worth studying a concrete example: the T-dualization of Minkowski spacetime [131]. We have found T duality by studying string theory in Minkowski spacetime with a compact dimension parametrized by a Cartesian coordinate with radius R_z, realizing that the spectrum was identical to that of another string theory in Minkowski spacetime with a compact dimension parametrized by a Cartesian coordinate with radius α'/R_z, which is the string solution T dual to the former. Both spacetimes preserve all supersymmetries of $N = 1, d = 10$ supergravity. The Buscher rules allow us, however, to find T duals to Minkowski spacetime associated with any other of its isometries, not only with the translational ones.

Let us consider, then, the maximally supersymmetric ten-dimensional Minkowski spacetime with a two-dimensional subspace written in polar coordinates ρ and θ:

$$d\hat{s}^2 = dt^2 - d\vec{x}_7^{\,2} - d\rho^2 - \rho^2 d\theta^2. \tag{16.161}$$

The T dual with respect to the isometry associated with shifts in the angular coordinate $\theta \in [0, 2\pi]$ is the solution

$$d\hat{s}'^2 = dt^2 - d\vec{x}_7^{\,2} - d\rho^2 - \rho^{-2} d\theta^2, \qquad e^{-2\hat{\phi}'} = \rho^2. \tag{16.162}$$

On looking into the supersymmetry transformation rules Eqs. (16.93), it takes no time to see that this solution has no unbroken supersymmetries at all because the dilatino variation $\delta_{\hat{\epsilon}}\hat{\lambda} = \rho^{-1}\hat{\Gamma}^9\hat{\epsilon}$ will never vanish for non-trivial $\hat{\epsilon}$.

Apparently, the T-duality transformation has broken completely the supersymmetry of the original background. The technical reason can be traced to the reduction to nine dimensions: the nine-dimensional solution

$$ds^2 = dt^2 - d\vec{x}_7^{\,2} - d\rho^2, \qquad e^{-2\phi} = k = \rho, \tag{16.163}$$

is not supersymmetric according to Eqs. (16.135). The ten-dimensional Killing spinor equations are satisfied and the nine-dimensional ones are not, and the reason why is that the ten-dimensional Killing spinor depends on the internal coordinate θ. This seems to happen whenever the isometry is not translational, but rotational, acting with fixed points.

Physically, the dimensional reduction in the direction θ is not a compactification in the standard KK sense. On the one hand, there are no non-contractible loops associated with θ and there are no associated winding modes. On the other hand, the "radius" of the θ direction, ρ, goes to zero in one limit and to infinity in another limit, and this space is certainly not asymptotically the KK vacuum. There seems to be no clear reason to expect these two backgrounds, which are related by Buscher's rules, to be equivalent for strings in the standard sense of interchange of KK and winding modes. Nevertheless, the two backgrounds are related by Buscher's rules and there seems to be no reason why supersymmetry should disappear [32]. It has been suggested that supersymmetry is still present but realized non-locally [504]. For the moment, there is no definite physical explanation for this phenomenon.

17

The type-IIB superstring and type-II T duality

In the previous chapter we initiated the study of the 11- and ten-dimensional supergravity theories which arise in the low-energy limits of the various string theories and M theory. Our goal was to study the dualities that relate the various string theories and M theory using effective-field-theory actions as we did in Section 15.2 with T duality in the effective action of the common string sector. In the coming chapters we will study these dualities from the point of view of their effect on classical solutions of the effective actions that represent the classical long-range fields of perturbative and non-perturbative states of these theories, as we did in Section 15.3 with the solutions associated with string and winding modes.

In this chapter we are going to study the $N = 2B$, $d = 10$ (chiral) supergravity theory, the effective field theory of the type-IIB superstring, and how it is related to the $N = 2A$ (non-chiral) theory after compactification on a circle (type-II T duality). Furthermore, we are also going to study the truncations to $N = 1$ theories, which are the effective-field theories of the type-I and heterotic superstrings, finding the field-theory version of the type-I/heterotic-string duality.

First, in Section 17.1, we will study the bosonic sector of the theory, giving a non-self-dual action from which one can derive equations of motion that have to be supplemented by the self-duality constraint of the 5-form field strength [111]. We will also give the supersymmetry transformation rules to lowest order in fermions. Then, in Section 17.2 we will study the S-duality symmetry of this theory, which becomes manifest only after several redefinitions.

Next, in Section 17.3 we will perform the dimensional reduction to nine dimensions of $N = 2B, d = 10$ supergravity compactified on a circle. As we will see, the nine-dimensional theory thus obtained is *identical* to the theory obtained by dimensional reduction of the $N = 2A$, $d = 10$ theory, Eq. (16.84). This will allow us to establish a correspondence between fields of the ten-dimensional $N = 2A$ and $N = 2B$ theories compactified on circles. This correspondence is part of the T duality existing between the two corresponding superstring theories and with our procedure we will have obtained the generalization of the T-duality Buscher rules to type-II theories found in [125] and generalized in [691].

Finally, in Section 17.5 we will study various consistent truncations of the theory and their relations to $N = 1$ theories and the corresponding full string theories. In particular, we will find a non-perturbative strong–weak-coupling relation between the type-I SO(32) and the heterotic SO(32) superstring theories.

17.1 $N = 2B, d = 10$ supergravity in the string frame

The fields of $N = 2B, d = 10$ SUEGRA [571, 820, 823] associated with the massless modes of the type-IIB superstring are given in Table 14.2. Actually, as we have mentioned a few times, there are two $N = 2B, d = 10$ theories, with all fermionic chiralities reversed and opposite self-duality properties of the RR 5-form field strength. Here we are going to describe the theory with a RR 4-form $\hat{C}^{(4)+}$ with self-dual 5-form field strength (suppressing the upper index $+$ for convenience),

$$\hat{G}^{(5)} = +{}^{\star}\hat{G}^{(5)} \tag{17.1}$$

(all the RR field strengths are normalized as explained in Section 16.1.3 according to Eqs. (16.51) and (16.52), the only difference being the use of calligraphic letters for the NSNS fields $\hat{\jmath}_{\hat{\mu}\hat{\nu}}$, $\hat{\mathcal{B}}_{\hat{\mu}\hat{\nu}}$, and $\hat{\varphi}$) and negative-chirality pairs of gravitinos and supersymmetry transformation parameters $\hat{\zeta}^{i\,(-)}_{\hat{\mu}}$ and $\hat{\varepsilon}^{i\,(-)}$ and positive-chirality dilatinos $\hat{\chi}^{i\,(+)}$ (although we also suppress these $\pm$ and the $i = 1, 2$ indices of the SO(2) global symmetry that rotates the fermions for convenience):

$$\hat{\Gamma}_{11}\hat{\zeta}_{\hat{\mu}} = -\hat{\zeta}_{\hat{\mu}}, \qquad \hat{\Gamma}_{11}\hat{\varepsilon} = -\hat{\varepsilon}, \qquad \hat{\Gamma}_{11}\hat{\chi} = +\hat{\chi}. \tag{17.2}$$

Although the self-duality equation for $\hat{G}^{(5)}$ looks like a constraint, it is indeed one of the equations of motion of the theory [820], and it arises as such in the superspace formalism [571]. It is not hard to see that, combined with the Bianchi identity, it gives a conventional-looking equation of motion:

$$\partial\hat{G}^{(5)} - \frac{10}{3}\hat{\mathcal{H}}\hat{G}^{(3)} = 0 \;\Rightarrow \partial^{\star}\hat{G}^{(5)} - \frac{10}{3}\hat{\mathcal{H}}\hat{G}^{(3)} = 0. \tag{17.3}$$

It is known that it is not possible to write a covariant action for the above self-duality equation of motion [685]. Nevertheless, having an action is very useful (for instance, to perform dimensional reductions) and we would like to write one. The main observation is that, if we do away with the self-duality of the 5-form, we can find an action that gives the conventional equation of motion Eq. (17.3) [111]. This equation of motion does not imply self-duality when it is combined with the Bianchi identity, but only $\partial\hat{G}^{(5)} = \partial^{\star}\hat{G}^{(5)}$. However, it is consistent with self-duality. This should be reflected in the following property of the action: if we dualize the 4-form, we must end up with an identical action written in terms of the dual 4-form. In other words, the action of the theory of the non-self-dual (NSD) 5-form must itself be "Poincaré self-dual." In our conventions the action of such a Poincaré self-dual NSD theory is

$$\boxed{\begin{aligned} S_{\text{NSD}} = \frac{\hat{g}_{\text{B}}^{2}}{16\pi G_{\text{NB}}^{(10)}}\int d^{10}\hat{x}\,\sqrt{|\hat{\jmath}|}\,\Big\{ & e^{-2\hat{\varphi}}\Big[\hat{R}(\hat{\jmath}) - 4\left(\partial\hat{\varphi}\right)^{2} + \frac{1}{2\cdot 3!}\hat{\mathcal{H}}^{2}\Big] \\ & + \tfrac{1}{2}\Big(\hat{G}^{(0)}\Big)^{2} + \frac{1}{2\cdot 3!}\Big(\hat{G}^{(3)}\Big)^{2} + \frac{1}{4\cdot 5!}\Big(\hat{G}^{(5)}\Big)^{2} - \frac{1}{192}\frac{1}{\sqrt{|\hat{\jmath}|}}\epsilon\;\partial\hat{C}^{(4)}\partial\hat{C}^{(2)}\hat{\mathcal{B}}\Big\}, \end{aligned}} \tag{17.4}$$

It is worth remarking that the sign of the last term is linked to the self-duality of the 5-form. If we wanted to impose consistently instead anti-self-duality of $\hat{G}^{(5)}$ and have an action for the opposite chirality $N = 2B, d = 10$ theory, the sign would have been exactly the opposite. This sign will ultimately be related, via T duality, to the sign of the Chern–Simons term of the $N = 2A, d = 10$ theory and, via dimensional oxidation, to the analogous sign of 11-dimensional supergravity.

It is also important to observe that the kinetic term for the 4-form has an extra factor of $\frac{1}{2}$ with respect to the standard normalization. In a sense, this factor takes into account that we have twice the right number of degrees of freedom. When, after dimensional reduction, we eliminate the self-duality constraint, we will obtain fields with the correct normalization thanks to this extra $\frac{1}{2}$.

Observe also that we have introduced, as in the $N = 2A$ case, a prefactor g_{B}^2 to absorb in it the asymptotic value of the dilaton, using the definition

$$g_{\mathrm{B}} \equiv e^{\hat{\varphi}_0}. \tag{17.5}$$

Assuming, as we do here, that the metric in the string frame is asymptotically flat, then it will also be asymptotically flat in the "modified-Einstein-frame" metric, and we conclude that the Newton constant of this theory is $G_{\mathrm{N\,B}}^{(10)}$. Its value is different from that of the IIA theory, so the quotient $\hat{g}_{\mathrm{B}}^2/G_{\mathrm{N\,B}}{}^{(10)}$ does not depend on $\hat{g}_{\mathrm{B}}$ (see Eq. (19.26)). This will be important in finding a frame in which the action is invariant under S duality.

17.1.1 Magnetic potentials

On comparing the equations of motion of the RR 0- and 2-form potentials,

$$d^{\star}\hat{G}^{(1)} - \hat{\mathcal{H}}^{\star}\hat{G}^{(3)} = 0, \qquad d^{\star}\hat{G}^{(3)} + \hat{\mathcal{H}}\hat{G}^{(5)} = 0, \tag{17.6}$$

with the Bianchi identities of the corresponding dual potentials,

$$d\hat{G}^{(9)} - \hat{\mathcal{H}}\hat{G}^{(7)} = 0, \qquad d\hat{G}^{(7)} - \hat{\mathcal{H}}\hat{G}^{(5)} = 0, \tag{17.7}$$

we find the relation between the original and the dual field strengths:

$$\hat{G}^{(9)} = +{}^{\star}\hat{G}^{(1)}, \qquad \hat{G}^{(7)} = -\hat{G}^{(3)}, \tag{17.8}$$

which defines them, in complete agreement with the first of Eqs. (16.57). Observe that $\hat{C}^{(0)}$ and $\hat{C}^{(2)}$ are pseudotensors and acquire an extra minus sign under parity transformations. $\hat{C}^{(5)}$ has no definite parity properties: in fact, parity transforms the self-duality constraint into an anti-self-duality constraint and chiral spinors into anti-chiral spinors and vice-versa. Thus the $N = 2B, d = 10$ supergravity is not invariant under parity, which, in fact, relates the two possible $N = 2B, d = 10$ supergravities. The magnetic potentials $\hat{C}^{(8)}$ and $\hat{C}^{(6)}$ are tensors.

There is a RR potential, electric or magnetic, for all the associated p odd Dp-branes of the type-IIB string theory, except for the D9-brane that requires a $\hat{C}^{(10)}$ potential, which can be added to the theory at no cost.

Finally, the dual of the NSNS 2-form is a 6-form whose field strength can be written in the form

$$\hat{\mathcal{H}}^{(7)} = d\hat{\mathcal{B}}^{(6)} + \tfrac{1}{2}\sum_{n=0}^{n=3} {}^{\star}\hat{G}^{(2n+3)} \wedge \hat{C}^{(2n)}. \tag{17.9}$$

17.1.2 The type-IIB supersymmetry rules

The supersymmetry transformation rules of $N = 2B, d = 10$ supergravity, generalized to include the magnetic RR potentials and field strengths plus $\hat{C}^{(10)}$, are, suppressing the $i, j = 1, 2$ SO(2) indices in fermions and Pauli matrices [117],

$$\boxed{\begin{aligned}
\delta_{\hat\varepsilon}\hat{e}_{\hat\mu}{}^{\hat a} &= -i\bar{\hat\varepsilon}\hat\Gamma^{\hat a}\hat\zeta_{\hat\mu},\\
\delta_{\hat\varepsilon}\hat\zeta_{\hat\mu} &= \nabla_{\hat\mu}\hat\varepsilon - \tfrac{1}{8}\,\hat{\not{\mathcal{H}}}_{\hat\mu}\sigma_3\hat\varepsilon + \frac{1}{16}e^{\hat\varphi}\sum_{n=1,\cdots,5}\frac{1}{(2n-1)!}\,\hat{\not{G}}^{(2n-1)}\hat\Gamma_{\hat\mu}\mathcal{P}_n\hat\varepsilon,\\
\delta_{\hat\varepsilon}\hat{\mathcal{B}}_{\hat\mu\hat\nu} &= -2i\bar{\hat\varepsilon}\sigma^3\hat\Gamma_{[\hat\mu}\hat\zeta_{\hat\nu]},\\
\delta_{\hat\varepsilon}\hat{C}^{(2n-2)}{}_{\hat\mu_1\cdots\hat\mu_{2n-2}} &= i(2n-2)e^{-\hat\varphi}\bar{\hat\varepsilon}\mathcal{P}_n\hat\Gamma_{[\hat\mu_1\cdots\hat\mu_{2n-3}}\left(\hat\zeta_{\hat\mu_{2n-2}]} - \frac{1}{2(2n-2)}\hat\Gamma_{\hat\mu_{2n-2}]}\hat\chi\right)\\
&\quad + \tfrac{1}{2}(2n-2)(2n-3)\hat{C}^{(2n-4)}{}_{[\hat\mu_1\cdots\hat\mu_{2n-4}}\delta_{\hat\varepsilon}\hat{\mathcal{B}}_{\hat\mu_{2n-3}\hat\mu_{2n-4}]},\\
\delta_{\hat\varepsilon}\hat\chi &= \left(\not{\partial}\hat\varphi - \frac{1}{12}\,\hat{\not{\mathcal{H}}}\sigma^3\right)\hat\varepsilon + \tfrac{1}{4}e^{\hat\varphi}\sum_{n=1,\cdots,5}\frac{(n-3)}{(2n-1)!}\,\hat{\not{G}}^{(2n-1)}\mathcal{P}_n\hat\varepsilon,\\
\delta_{\hat\varepsilon}\hat\varphi &= -\frac{i}{2}\bar{\hat\varepsilon}\hat\chi,
\end{aligned}} \tag{17.10}$$

where

$$\mathcal{P}_n = \begin{cases} \sigma^1, & n \text{ even},\\ i\sigma^2, & n \text{ odd}. \end{cases} \tag{17.11}$$

Observe that the consistency of these supersymmetry transformations demands that the gravitinos and supersymmetry transformation parameters have the same chirality, opposite to that of the dilatinos. Observe also that, due to the self-duality of $\hat{G}^{(5)}$,

$$\hat{\not{G}}^{(5)} = \hat{\not{G}}^{(5)}\tfrac{1}{2}(1 + \hat\Gamma_{11}). \tag{17.12}$$

The $\hat{\not{G}}^{(5)}$ term in $\delta_{\hat\varepsilon}\hat\zeta_{\hat\mu}$ survives due to the negative chirality of $\hat\varepsilon$ and does not survive in $\delta_{\hat\varepsilon}\hat\chi$ for the same reason. This fact plays an important role in the existence of maximally supersymmetric solutions of this theory.

17.2 Type-IIB S duality

In the original version of the ten-dimensional, chiral $N = 2$ supergravity [820] the theory has a classical SU(1,1) global symmetry. The two scalars parametrize the coset space

SU(1,1)/U(1), U(1) being the maximal compact subgroup of SU(1,1), and transform under a combination of a global SU(1,1) transformation and a local U(1) transformation that depends on the global SU(1,1) transformation. They are combinations of the dilaton and the RR scalar. The group SU(1,1) is isomorphic to $SL(2,\mathbb{R})$, the conjectured classical S-duality symmetry group for the type-IIB string theory [583]. A simple field redefinition [125] is enough to rewrite the action in terms of two real scalars parametrizing the coset space $SL(2,\mathbb{R})/SO(2)$, which can now be identified with the dilaton and the RR scalar. Now the S-duality symmetry becomes manifest only when we rescale the metric to work in the Einstein frame:

$$\hat{\jmath}_{\mathrm{E}\,\mu\nu} = e^{-\frac{\varphi}{2}} \jmath_{\mu\nu}. \tag{17.13}$$

However, the RR potentials we are working with here are not the most appropriate to exhibit manifest $SL(2,\mathbb{R})$ symmetry. In fact, they have been chosen because they are the most appropriate to study T duality and the worldvolume effective actions of D-branes. In particular, while the NSNS and RR 2-forms we are using form an $SL(2,\mathbb{R})$ doublet (as we are going to see), their field strengths do not. Furthermore, our self-dual RR 4-form potential $\hat{C}^{(4)}$ is not $SL(2,\mathbb{R})$-invariant. Thus, for the purpose of exhibiting the $SL(2,\mathbb{R})$ symmetry it is convenient to perform the following field redefinitions:[1,2]

$$\hat{\vec{\mathcal{B}}} = \begin{pmatrix} \hat{C}^{(2)} \\ \hat{\mathcal{B}} \end{pmatrix}, \qquad \hat{D} = \hat{C}^{(4)} - 3\hat{\mathcal{B}}\hat{C}^{(2)}. \tag{17.14}$$

These new fields undergo the following gauge transformations:

$$\delta\hat{\vec{\mathcal{B}}} = 2\hat{\vec{\Sigma}}, \qquad \delta\hat{D} = 4\partial\hat{\Delta} + 2\hat{\vec{\Sigma}}^{\mathrm{T}}\eta\,\hat{\vec{\mathcal{H}}}, \tag{17.15}$$

and have field strengths

$$\begin{aligned} \hat{\vec{\mathcal{H}}} &= 3\partial\hat{\vec{\mathcal{B}}}, \\ \hat{F} &= \hat{G}^{(5)} = +{}^{\star}\hat{F} = 5\Big(\partial\hat{D} - \hat{\vec{\mathcal{B}}}^{\,T}\eta\,\hat{\vec{\mathcal{H}}}\Big), \end{aligned} \tag{17.16}$$

where η is the 2×2 matrix

$$\eta = i\sigma^2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = -\eta^{-1} = -\eta^{\mathrm{T}}. \tag{17.17}$$

Given the isomorphism $SL(2,\mathbb{R}) \sim Sp(2,\mathbb{R})$, it plays the role of an invariant metric:

$$S\eta S^{\mathrm{T}} = \eta, \;\Rightarrow\; \eta S\eta^{\mathrm{T}} = (S^{-1})^{\mathrm{T}}, \qquad S \in SL(2,\mathbb{R}). \tag{17.18}$$

Next, we define the complex scalar $\hat{\tau}$ that parametrizes the coset space $SL(2,\mathbb{R})/SO(2)$,

$$\hat{\tau} = \hat{C}^{(0)} + ie^{-\hat{\varphi}}, \tag{17.19}$$

---

[1] In our conventions all fields are either invariant or transform *covariantly* as opposed to *contravariantly*.

[2] The complete relations (including fermions) between the formulation of the $N = 2B$ theory in "stringy variables" that we have introduced in the previous section and the manifestly S-duality-covariant formulation that we are going to introduce in this section can be found in [426].

and the 2×2 symmetric $\mathrm{SL}(2,\mathbb{R})$ matrix $\hat{\mathcal{M}}_{ij}$, given in Eq. (11.207) in terms of $\hat{\tau}$, that satisfies the property

$$\hat{\mathcal{M}}^{-1} = \eta \hat{\mathcal{M}} \eta^{\mathrm{T}}. \tag{17.20}$$

Under $S \in \mathrm{SL}(2,\mathbb{R})$ given by Eq. (11.205) the new variables that we have defined transform as follows:

$$\hat{\mathcal{M}}' = S \hat{\mathcal{M}} S^{\mathrm{T}}, \qquad \hat{\tau}' = \frac{\alpha \hat{\tau} + \beta}{\gamma \hat{\tau} + \delta}, \qquad \hat{\vec{\mathcal{B}}}' = S \hat{\vec{\mathcal{B}}}, \tag{17.21}$$

and the 4-form $\hat{D}$ and the Einstein metric $\hat{\jmath}_{\mathrm{E}}$ are inert.

Now, it is a simple exercise to rewrite the NSD $N = 2B$ action in the following manifestly S-duality-invariant form:

$$\boxed{\begin{aligned}\hat{S}_{\mathrm{NSD}} = \frac{\hat{g}_{\mathrm{B}}^2}{16\pi G_{\mathrm{NB}}^{(10)}} \int d^{10}\hat{x}\, \sqrt{|\hat{\jmath}_{\mathrm{E}}|} \Big\{ & \hat{R}(\hat{\jmath}_{\mathrm{E}}) + \tfrac{1}{4} \mathrm{Tr}\Big(\partial \hat{\mathcal{M}} \hat{\mathcal{M}}^{-1}\Big)^2 \\ & + \frac{1}{2 \cdot 3!} \hat{\vec{\mathcal{H}}}^{\mathrm{T}} \hat{\mathcal{M}}^{-1} \hat{\vec{\mathcal{H}}} + \frac{1}{4 \cdot 5!} \hat{F}^2 - \frac{1}{2^7 \cdot 3^3} \frac{1}{\sqrt{|\hat{\jmath}_{\mathrm{E}}|}} \epsilon \hat{D} \hat{\vec{\mathcal{H}}}^{\mathrm{T}} \eta\, \hat{\vec{\mathcal{H}}} \Big\}.\end{aligned}} \tag{17.22}$$

Observe that the factor $\hat{g}_{\mathrm{B}}^2/(16\pi G_{\mathrm{NB}}^{(10)})$ is S-duality-invariant because it does not depend on $\hat{g}_{\mathrm{B}}$ (see Eq. (19.26)). Thus it is in the Einstein frame that the full action is invariant under S duality and masses measured in this frame are S-duality-invariant. As usual, this is not the metric in which we should measure masses (at least masses that we want to compare with the string spectrum) because, if the string metric is asymptotically flat, the Einstein metric is not. We should use the modified Einstein frame. This metric will also be S-duality-invariant, but the action will have the prefactor $1/(16\pi G_{\mathrm{NB}}^{(10)})$ which is not invariant and, thus, masses measured in it will not be invariant.

It is easy to find how the stringy fields $\hat{\mathcal{H}}$, $\hat{G}^{(3)}$, and $\hat{C}^{(4)}$ transform under $\mathrm{SL}(2,\mathbb{R})$:

$$\begin{aligned}\hat{\mathcal{H}}' &= \Big(\delta + \gamma \hat{C}^{(0)}\Big)\hat{\mathcal{H}} + \gamma \hat{G}^{(3)}, \\ \hat{G}^{(3)\,\prime} &= \frac{1}{|\gamma\hat{\tau} + \delta|^2} \Big[\Big(\delta + \gamma \hat{C}^{(0)}\Big)\hat{G}^{(3)} - \gamma e^{-2\hat{\varphi}} \hat{\mathcal{H}}\Big], \\ \hat{C}^{(4)\,\prime} &= \hat{C}^{(4)} - 3\big(\hat{C}^{(2)}\ \hat{\mathcal{B}}\big)\begin{pmatrix} \alpha\gamma & \beta\gamma \\ \beta\gamma & \delta\beta \end{pmatrix}\begin{pmatrix} \hat{C}^{(2)} \\ \hat{\mathcal{B}} \end{pmatrix}.\end{aligned} \tag{17.23}$$

$\hat{\tau}$ transforms as above and we stress that the string metric does transform under $\mathrm{SL}(2,\mathbb{R})$:

$$\hat{\jmath}' = |\gamma\hat{\tau} + \delta|\, \hat{\jmath}. \tag{17.24}$$

Some of the $\mathrm{SL}(2,\mathbb{R})$ transformations of $N = 2B$, $d = 10$ SUEGRA involve an inversion of the dilaton, and, hence, of the string coupling constant $\hat{g}_{\mathrm{B}}$, just as we discussed in the case of $N = 4$, $d = 4$ SUEGRA, the effective theory of the heterotic string (Sections 12.2 and 16.5.5). These are, therefore, non-perturbative transformations from the string-theory point of view, and the perturbative description that we have of it gives little information about them. We can take the point of view that the existence of this symmetry of the supergravity theory indicates the existence of a similar string S duality (with $SL(2, \mathbb{R})$ broken to $SL(2, \mathbb{Z})$ by quantum effects such as charge quantization) that relates strongly coupled type-IIB string theory to another weakly coupled type-IIB theory (because the supergravity is invariant under these transformations), and try to check the implications.

As we learnt in Section 8.7 and applied later in the case of S duality in four-dimensional KK theory, one of the main characteristics of S duality is that it interchanges fundamental, perturbative states with solitonic, non-perturbative states. If the S dual of a theory is another theory of the same kind (as is the case here), then the full spectrum of the theory, including perturbative and non-perturbative states, must be S-duality-invariant. Thus, if there is S duality in type-IIB superstring theory, there must be (or we must add) non-perturbative states in it that are interchanged with the fundamental string states that we already know. These non-perturbative states turn out to be D1-branes, viz. *D-strings*, and dyonic states known as *pq-strings*. Their presence implies the possible addition of open-string sectors to the IIB theory. Since the type-IIA and -IIB theories are T dual to each other, namely D-strings are T dual to D0- and D2-branes, we will have to admit the possibility of having the corresponding open-string sectors in the IIA theory and, again, due to T duality, D3-branes in the IIB and so on and so forth.

The very first string solution that we studied, the F1 solution Eq. (15.70), represents the long-range fields associated with a kind of fundamental, non-perturbative string state. S duality requires the existence of S-dual solutions: D-string and pq-string solutions. We will study all these issues related to the stringy interpretation of the supergravity symmetry that we have uncovered in this section in more detail later on.

17.3 Dimensional reduction of $N = 2B$, $d = 10$ SUEGRA and type-II T duality

In this section we are going to study from the point of view of the type-II string effective actions the T duality between the IIA and IIB superstring theories compactified on circles discovered in [283, 322]. As in the string common sector, it is possible to find T-duality relations ("type-II Buscher rules") between fields and solutions of both theories by performing the dimensional reduction of the effective-field theories ($N = 2A$ and $N = 2B$, $d = 10$ SUEGRAs) to nine dimensions on a circle. The dimensional reduction of the $N = 2A$ theory was performed in Section 16.3 and here we are going to do the same to the action of the $N = 2B$ theory Eq. (17.4), assuming now that all the fields are independent of the dual compact coordinate that we call y in this case. Since the NSD action has to be complemented by the self-duality constraint, at the end we will also have to reduce the constraint and then use it in the reduced nine-dimensional action to obtain exactly Eq. (16.84).

We start with the reduction of the NSNS sector, which we have already performed. The result is given in Eq. (15.25) but now we call y the compact coordinate, $A^{(2)}$ the KK vector, $A^{(1)}$ the winding vector, and k^{-1} the KK scalar (a summary of the relations between the ten- and nine-dimensional fields can be found in Section 17.3.1). The asymptotic value of this KK scalar is R_y/ℓ_s but, since k is to be identified with the KK scalar coming from the reduction of the IIA theory, it is also ℓ_s/R_x, and thus we have the T-duality relation between compactification radii

$$\frac{R_y}{\ell_s} = \frac{\ell_s}{R_x}. \tag{17.25}$$

Furthermore, the normalization factor in front of the action is now

$$\frac{2\pi \ell_s \hat{g}_B^2}{16\pi G_{NB}^{(10)}} = \frac{2\pi \ell_s g_B^2 k_0^{-1}}{16\pi G_{NB}^{(10)}} = \frac{g_B^2}{16\pi G_{NB}^{(9)}}, \tag{17.26}$$

where we have used

$$g_B = \hat{g}_B k_0^{\frac{1}{2}}, \qquad G_{NB}^{(9)} = G_{NB}^{(10)}/(2\pi R_y). \tag{17.27}$$

The normalization factor is independent both of the $d = 10$ or $d = 9$ string coupling constant and of the compactification radius, although the $d = 9$ string coupling constant squared and the $d = 9$ Newton constant do depend on them in precisely the same way.

This is the T-dual reduction to the one we performed in Section 16.3 for the IIA theory. The ten-dimensional IIB NSNS fields $\hat{\jmath}_{\hat{\mu}\hat{\nu}}$, $\hat{\mathcal{B}}_{\hat{\mu}\hat{\nu}}$, and $\hat{\varphi}$ decompose in terms of the same nine-dimensional IIA/B NSNS fields in the T-dual way.

As for the dimensional reduction of the RR sector, when we reduced the $N = 2A$ theory to nine dimensions, we defined the field strengths and gauge transformations for the nine-dimensional RR potentials in Eqs. (16.79) and (16.80). Since we have completely determined the reduction of the NSNS fields, the reduction of the $N = 2B$ RR potentials is completely determined by the requirement that they have the same nine-dimensional field strengths as in the $N = 2A$ case. This is achieved by the following identifications:

$$\begin{aligned} \hat{C}^{(2n)}{}_{\mu_1\cdots\mu_{2n}} &= C^{(2n)}{}_{\mu_1\cdots\mu_{2n}} - (2n) A^{(2)}{}_{[\mu_1} C^{(2n-1)}{}_{\mu_2\cdots\mu_{2n}]}, \\ \hat{C}^{(2n)}{}_{\mu_1\cdots\mu_{2n-1}\underline{x}} &= -C^{(2n-1)}{}_{\mu_1\cdots\mu_{2n-1}}. \end{aligned} \tag{17.28}$$

The field strengths reduce as follows:

$$\hat{G}^{(2n+1)}{}_{a_1\cdots a_{2n+1}} = G^{(2n+1)}{}_{a_1\cdots a_{2n+1}}, \qquad \hat{G}^{(2n+1)}{}_{a_1\cdots a_{2n} y} = -k G^{(2n)}{}_{a_1\cdots a_{2n}}, \tag{17.29}$$

and the corresponding kinetic terms in the action reduce as follows:

$$\frac{\sqrt{|\hat{\jmath}|}}{2\cdot(2n+1)!}\left(\hat{G}^{(2n+1)}\right)^2 = \frac{\sqrt{|g|}}{2\cdot(2n+1)!} k^{-1}\left(G^{(2n+1)}\right)^2 - \frac{\sqrt{|g|}}{2\cdot(2n)!} k\left(G^{(2n)}\right)^2. \tag{17.30}$$

Although only the electric RR forms appear in the action, the reduction formulae work for the magnetic ones as well. The final check for the RR fields is that the field strengths satisfy the same duality relations as one obtains in the $N = 2A$ case because, since we have used the same definitions for the field strengths as in the $N = 2A$ case, they satisfy the same Bianchi identities and, then, if they satisfy the same duality relations, they satisfy the same equations of motion. Indeed, the RR field strengths obtained from both theories satisfy

$$G^{(9-k)} = -{}^\star G^{(k)}, \tag{17.31}$$

which is always consistent in $d = 9$.

This is clearly enough to conclude that the reduction of the $N = 2A$ and $N = 2B$ theories to $d = 9$ gives the same nine-dimensional theory. However, just as a check, we can complete the reduction of the action of the $N = 2B$ theory and see that it coincides with Eq. (16.84). We are only going to outline how this is done: the NSD action can be reduced to $d = 9$ straightforwardly using the above formulae, but we obtain a theory with 5- and 4-form RR field strengths that originate from the ten-dimensional self-dual 5-form field strength. The nine-dimensional 5- and 4-forms are related by Eq. (17.31), which is a constraint of the action that we have to eliminate in order to arrive at the action Eq. (16.84). To eliminate consistently the constraint and the 5-form, we first Poincaré-dualize it into a second 4-form, following the standard procedure. Finally, we identify the two 4-forms and the result is Eq. (16.84) with a single 4-form and with the correct sign in the Chern–Simons term.

This result allows us to map fields of one ten-dimensional theory onto fields of the other ten-dimensional theory (which is always independent of one coordinate). This mapping is the generalization of Buscher's T-duality rules to type-II theories [125, 691] that we describe in the next section, but it is worth making some preliminary remarks.

1. The rules reflect the T-duality rules for D-branes that we discussed on page 428.
2. We could have reduced the manifestly S-duality-invariant action Eq. (17.22) and we would have obtained a manifestly SL(2, $\mathbb{R}$)-invariant action in $d = 9$. As usual in KK compactification, the action would also be invariant under a group $\mathbb{R}^+$ of rescalings of the internal dimension and other $\mathbb{Z}_2$ factors which combine into GL(2, $\mathbb{R}$), which is the invariance group that one obtains in the reduction from $d = 11$ to $d = 9$. The IIB S duality now has a geometrical interpretation in the IIA theory.
3. It is possible to use the full SL(2, $\mathbb{R}$) invariance of the action to perform a GDR of the theory [426, 691]. The result is a family of *massive* supergravity theories that depends on three mass parameters transforming in the adjoint of SL(2, $\mathbb{R}$) that fit into a symmetric *mass matrix*. One of the theories, which depends on a single parameter, is precisely the theory one would obtain by reducing Romans' theory to $d = 9$ [118][3], but there are other theories that cannot be obtained from known 11- and ten-dimensional supergravities. Most of the $d = 9$ theories obtained in this way are gauged supergravities [258, 726] and the gauge group is determined by the conjugacy class of the chosen mass matrix [120]. While there is a simple string interpretation for Romans' theory based on the D8-brane, the remaining massive/gauged theories have a less-conventional interpretation that is based on non-conventional extended branes.
4. Although only low-rank RR potentials appear in the action, we have extended the relation to the high-rank magnetic RR potentials. This implies the existence of new high-rank RR potentials unrelated to the electric ones: the IIB $\hat{C}^{(8)}$ can be related to the IIA $\hat{C}^{(7)}$, but also to a $\hat{C}^{(9)}$ that exists in Romans' theory only since it is the

[3] This is the theory that one obtains with the GDR *Ansatz* studied in Section 11.5.3. Observe that the mass parameter is naturally quantized since it is a winding number. This implies that the mass parameter of Romans' theory must also be quantized, if we insist on identifying these theories as string theory indicates. This GDR *Ansatz* is related to the RR 9-form potential we are going to discuss next.

magnetic dual of the mass parameter.[4] In turn, the IIA $\hat{C}^{(9)}$ implies the existence of a IIB $\hat{C}^{(10)}$ associated with the D9-brane. On combining these results with S duality, we arrive at the possible existence of the S dual of $\hat{C}^{(10)}$, $\hat{\mathcal{B}}^{(10)}$, whose on-shell supersymmetry transformation is given in Eq. (17.41). A new T duality implies the existence of another $\hat{B}^{(10)}$ in the IIA theory. These potentials play an interesting role, as we are going to see in Section 17.5 [123, 580]. T-duality rules for $\mathcal{B}^{(10)}$ and $\mathcal{B}^{(6)}$ have been given in [375].

17.3.1 The type-II T-duality Buscher rules

We are now ready to relate the $N = 2A$, B, $d = 10$ fields. For the sake of completeness we summarize the relation between ten- and nine-dimensional fields for each theory first.

Summary of the type-IIA reduction

NSNS fields:

$$
\begin{aligned}
\hat{g}_{\mu\nu} &= g_{\mu\nu} - k^2 A^{(1)}{}_{\mu} A^{(1)}{}_{\nu}, & g_{\mu\nu} &= \hat{g}_{\mu\nu} - \hat{g}_{\mu\underline{x}}\hat{g}_{\nu\underline{x}}/\hat{g}_{\underline{xx}},\\
\hat{B}_{\mu\nu} &= B_{\mu\nu} - A^{(1)}{}_{[\mu} A^{(2)}{}_{\nu]}, & B_{\mu\nu} &= \hat{B}_{\mu\nu} + \hat{g}_{[\mu|\underline{x}|}\hat{B}_{\nu]\underline{x}}/\hat{g}_{\underline{xx}},\\
\hat{\phi} &= \phi + \tfrac{1}{2}\ln k, & \phi &= \hat{\phi} - \tfrac{1}{4}\ln|\hat{g}_{\underline{xx}}|,\\
\hat{g}_{\mu\underline{x}} &= -k^2 A^{(1)}{}_{\mu}, & A^{(1)}{}_{\mu} &= \hat{g}_{\mu\underline{x}}/\hat{g}_{\underline{xx}},\\
\hat{B}_{\mu\underline{x}} &= A^{(2)}{}_{\mu}, & A^{(2)}{}_{\mu} &= \hat{B}_{\mu\underline{x}},\\
\hat{g}_{\underline{xx}} &= -k^2, & k &= |\hat{g}_{\underline{xx}}|^{\frac{1}{2}}.
\end{aligned}
\tag{17.32}
$$

RR fields:

$$
\begin{aligned}
\hat{C}^{(2n-1)}{}_{\mu_1\cdots\mu_{2n-1}} &= C^{(2n-1)}{}_{\mu_1\cdots\mu_{2n-1}} + (2n-1)A^{(1)}{}_{[\mu_1}C^{(2n-2)}{}_{\mu_2\cdots\mu_{2n-1}]},\\
\hat{C}^{(2n+1)}{}_{\mu_1\cdots\mu_{2n}\underline{x}} &= C^{(2n)}{}_{\mu_1\cdots\mu_{2n}},\\
C^{(2n-1)}{}_{\mu_1\cdots\mu_{2n-1}} &= \hat{C}^{(2n-1)}{}_{\mu_1\cdots\mu_{2n-1}} - (2n-1)\hat{g}_{[\mu_1|\underline{x}|}\hat{C}^{(2n-1)}{}_{\mu_2\cdots\mu_{2n-1}]\underline{x}}/\hat{g}_{\underline{xx}},\\
C^{(2n)}{}_{\mu_1\cdots\mu_{2n}} &= \hat{C}^{(2n+1)}{}_{\mu_1\cdots\mu_{2n}\underline{x}}.
\end{aligned}
\tag{17.33}
$$

Summary of the type-IIB reduction

NSNS fields:

$$
\begin{aligned}
\hat{\jmath}_{\mu\nu} &= g_{\mu\nu} - k^{-2} A^{(2)}{}_{\mu} A^{(2)}{}_{\nu}, & g_{\mu\nu} &= \hat{\jmath}_{\mu\nu} - \hat{\jmath}_{\mu\underline{y}}\hat{\jmath}_{\nu\underline{y}}/\hat{\jmath}_{\underline{yy}},\\
\hat{\mathcal{B}}_{\mu\nu} &= B_{\mu\nu} + A^{(1)}{}_{[\mu} A^{(2)}{}_{\nu]}, & B_{\mu\nu} &= \hat{\mathcal{B}}_{\mu\nu} + \hat{\jmath}_{[\mu|\underline{y}|}\hat{\mathcal{B}}_{\nu]\underline{y}}/\hat{\jmath}_{\underline{yy}},\\
\hat{\varphi} &= \phi - \tfrac{1}{2}\ln k, & \phi &= \hat{\varphi} - \tfrac{1}{4}\ln|\hat{\jmath}_{\underline{yy}}|,\\
\hat{\jmath}_{\mu\underline{y}} &= -k^{-2} A^{(2)}{}_{\mu}, & A^{(1)}{}_{\mu} &= \hat{\mathcal{B}}_{\mu\underline{y}},\\
\hat{\mathcal{B}}_{\mu\underline{y}} &= A^{(1)}{}_{\mu}, & A^{(2)}{}_{\mu} &= \hat{\jmath}_{\mu\underline{y}}/\hat{\jmath}_{\underline{yy}},\\
\hat{\jmath}_{\underline{yy}} &= -k^{-2}, & k &= |\hat{\jmath}_{\underline{yy}}|^{-\frac{1}{2}}.
\end{aligned}
\tag{17.34}
$$

[4] As explained on page 342, a non-vanishing value of the potential $\hat{C}^{(9)}$ is related to the GDR associated with the shifts of $\hat{C}^{(0)}$ that we discussed before, which are in turn related to the mass parameter of Romans' theory. Clearly, the whole picture is consistent.

RR fields:

$$
\begin{aligned}
\hat{C}^{(2n)}{}_{\mu_1\cdots\mu_{2n}} &= C^{(2n)}{}_{\mu_1\cdots\mu_{2n}} - (2n)A^{(2)}{}_{[\mu_1}C^{(2n-1)}{}_{\mu_2\cdots\mu_{2n}]}, \\
\hat{C}^{(2n)}{}_{\mu_1\cdots\mu_{2n-1}\underline{y}} &= -C^{(2n-1)}{}_{\mu_1\cdots\mu_{2n-1}}, \\
C^{(2n)}{}_{\mu_1\cdots\mu_{2n}} &= \hat{C}^{(2n)}{}_{\mu_1\cdots\mu_{2n}} + (2n)\hat{\jmath}_{[\mu_1|\underline{y}|}\hat{C}^{(2n)}{}_{\mu_2\cdots\mu_{2n}]\underline{y}}/\hat{\jmath}_{\underline{yy}}, \\
C^{(2n-1)}{}_{\mu_1\cdots\mu_{2n-1}} &= -\hat{C}^{(2n)}{}_{\mu_1\cdots\mu_{2n-1}\underline{y}}.
\end{aligned}
\tag{17.35}
$$

We obtain the following generalization of Buscher's rules [125, 691].

From IIA to IIB:

$$
\boxed{
\begin{aligned}
\hat{\jmath}_{\mu\nu} &= \hat{g}_{\mu\nu} - \left(\hat{g}_{\mu\underline{x}}\hat{g}_{\nu\underline{x}} - \hat{B}_{\mu\underline{x}}\hat{B}_{\nu\underline{x}}\right)\Big/\hat{g}_{\underline{xx}}, & \hat{\jmath}_{\mu\underline{y}} &= \hat{B}_{\mu\underline{x}}/\hat{g}_{\underline{xx}}, \\
\hat{\mathcal{B}}_{\mu\nu} &= \hat{B}_{\mu\nu} + 2\hat{g}_{[\mu|\underline{x}}\hat{B}_{\nu]\underline{x}}/\hat{g}_{\underline{xx}}, & \hat{\mathcal{B}}_{\mu\underline{y}} &= \hat{g}_{\mu\underline{x}}/\hat{g}_{\underline{xx}}, \\
\hat{\varphi} &= \hat{\phi} - \tfrac{1}{2}\ln|\hat{g}_{\underline{xx}}|, & \hat{\jmath}_{\underline{yy}} &= 1/\hat{g}_{\underline{xx}}, \\
\hat{C}^{(2n)}{}_{\mu_1\cdots\mu_{2n}} &= \hat{C}^{(2n+1)}{}_{\mu_1\cdots\mu_{2n}\underline{x}} + 2n\hat{B}_{[\mu_1|\underline{x}|}\hat{C}^{(2n-1)}{}_{\mu_2\cdots\mu_{2n}]} \\
&\quad - 2n(2n-1)\hat{B}_{[\mu_1|\underline{x}|}\hat{g}_{\mu_2|\underline{x}|}\hat{C}^{(2n-1)}{}_{\mu_3\cdots\mu_{2n}]\underline{x}}/\hat{g}_{\underline{xx}}, \\
\hat{C}^{(2n)}{}_{\mu_1\cdots\mu_{2n-1}\underline{y}} &= -\hat{C}^{(2n-1)}{}_{\mu_1\cdots\mu_{2n-1}} \\
&\quad + (2n-1)\hat{g}_{[\mu_1|\underline{x}|}\hat{C}^{(2n-1)}{}_{\mu_2\cdots\mu_{2n-1}]\underline{x}}/\hat{g}_{\underline{xx}}.
\end{aligned}
}
\tag{17.36}
$$

From IIB to IIA:

$$
\boxed{
\begin{aligned}
\hat{g}_{\mu\nu} &= \hat{\jmath}_{\mu\nu} - \left(\hat{\jmath}_{\mu\underline{y}}\hat{\jmath}_{\nu\underline{y}} - \hat{\mathcal{B}}_{\mu\underline{y}}\hat{\mathcal{B}}_{\nu\underline{y}}\right)\Big/\hat{\jmath}_{\underline{yy}}, & \hat{g}_{\mu\underline{x}} &= \hat{\mathcal{B}}_{\mu\underline{y}}/\hat{\jmath}_{\underline{yy}}, \\
\hat{B}_{\mu\nu} &= \hat{\mathcal{B}}_{\mu\nu} + 2\hat{\jmath}_{[\mu|\underline{y}}\hat{\mathcal{B}}_{\nu]\underline{y}}/\hat{\jmath}_{\underline{yy}}, & \hat{B}_{\mu\underline{x}} &= \hat{\jmath}_{\mu\underline{y}}/\hat{\jmath}_{\underline{yy}}, \\
\hat{\phi} &= \hat{\varphi} - \tfrac{1}{2}\ln|\hat{\jmath}_{\underline{yy}}|, & \hat{g}_{\underline{xx}} &= 1/\hat{\jmath}_{\underline{yy}}, \\
\hat{C}^{(2n+1)}{}_{\mu_1\cdots\mu_{2n+1}} &= -\hat{C}^{(2n+2)}{}_{\mu_1\cdots\mu_{2n+1}\underline{y}} + (2n+1)\hat{\mathcal{B}}_{[\mu_1|\underline{y}|}\hat{C}^{(2n)}{}_{\mu_2\cdots\mu_{2n+1}]} \\
&\quad + 2n(2n+1)\hat{\mathcal{B}}_{[\mu_1|\underline{y}|}\hat{\jmath}_{\mu_2|\underline{y}|}\hat{C}^{(2n)}{}_{\mu_3\cdots\mu_{2n+1}]\underline{y}}/\hat{\jmath}_{\underline{yy}}, \\
\hat{C}^{(2n+1)}{}_{\mu_1\cdots\mu_{2n}\underline{x}} &= \hat{C}^{(2n)}{}_{\mu_1\cdots\mu_{2n}} + 2n\hat{\jmath}_{[\mu_1|\underline{y}|}\hat{C}^{(2n)}{}_{\mu_2\cdots\mu_{2n}]\underline{y}}/\hat{\jmath}_{\underline{yy}}.
\end{aligned}
}
\tag{17.37}
$$

17.4 Dimensional reduction of fermions and supersymmetry rules

As we discussed on page 439, we have to take into account carefully the T-duality transformation of the Vielbeins when dealing with fermions. There are two possible rules compatible with fermions, which, combined with the standard KK *Ansatz* Eqs. (11.33), give Eqs. (15.30). In our case we have already reduced the $N = 2A$ fermions and supersymmetry transformation rules using the standard KK *Ansatz* (with A_μ renamed $A^{(1)}{}_\mu$) and it just turns out that we can obtain agreement with those results only by using the lower sign in Eqs. (15.30) (with B_μ renamed $A^{(2)}{}_\mu$) [505, 506]. To be more explicit, the *Ansatz* that we must use in the reduction of the $N = 2B$, $d = 10$ theory is

$$\left(\hat{e}_{\hat{\mu}}{}^{\hat{a}}\right) = \begin{pmatrix} e_\mu{}^a & -k^{-1}A^{(2)}{}_\mu \\ 0 & -k^{-1} \end{pmatrix}, \qquad \left(\hat{e}_{\hat{a}}{}^{\hat{\mu}}\right) = \begin{pmatrix} e_a{}^\mu & -A^{(2)}{}_a \\ 0 & -k \end{pmatrix}. \tag{17.38}$$

The sign is irrelevant in the reduction of the bosonic sector, as we stressed, and, thus, it does not change the type-II Buscher rules we just derived.

The $N = 2B$, $d = 10$ spinors are pairs of Majorana–Weyl spinors with only 16 real non-vanishing components out of the 32 in the chiral basis that we are using with $\hat{\Gamma}_{11} = \mathbb{I}_{16\times 16} \otimes \sigma^3$. The indices that label each pair of fermions are usually not explicitly shown. The Pauli matrices that appear in the supersymmetry rules Eqs. (17.10) act only on those indices and both survive in the nine-dimensional theory. In the decomposition of the ten-dimensional gamma matrices that we have used in the type-IIA case new Pauli matrices appear but they do not act on those indices; rather, they act on the chiral (upper and lower) components of the 32-component spinors. These Pauli matrices do not survive the reduction.

Taking all these facts into account, the ten-dimensional 32-component fermions, which, including the supersymmetry parameter, are $\hat{\zeta}^i_{\hat{\mu}}$, $\hat{\chi}^i$, and $\hat{\varepsilon}^i$ and the nine-dimensional, 16-component fermions ψ^i_μ, λ^i, ρ^i, and ϵ^i, are related by

$$\begin{aligned} \hat{\zeta}^i_{\underline{y}} &= \begin{pmatrix} 0 \\ -k^{-1}\rho^i \end{pmatrix}, & \hat{\zeta}^i_\mu &= \begin{pmatrix} 0 \\ \psi^i_\mu - k^{-1}A^{(2)}{}_\mu\rho^i \end{pmatrix}, \\ \hat{\chi}^i &= \sigma^2\begin{pmatrix} 0 \\ \lambda^i - \rho^i \end{pmatrix}, & \hat{\varepsilon}^i &= \begin{pmatrix} 0 \\ \epsilon^i \end{pmatrix}, \end{aligned} \tag{17.39}$$

and, using these relations, we obtain complete agreement with the nine-dimensional supersymmetry transformation rules that we derived from the $N = 2A$ theory, Eqs. (16.88)–(16.90).

Using Eqs. (16.86) and (17.39), we can derive Buscher's rules for the fermions. They are not too interesting except for the supersymmetry parameters, since they can be used for Killing spinors, if they are independent of the compact coordinates, which is not always the case, as we discussed in Section 16.6. Recalling that the gamma matrix which points into the direction into which we T dualize is $\hat{\Gamma}^9 = \mathbb{I}_{16\times 16} \otimes i\sigma^1$, it is immediately possible to derive the two T-duality rules:

$$\boxed{\begin{aligned} &\hat{\epsilon} = \hat{\varepsilon}^2 - i\hat{\Gamma}^9\hat{\varepsilon}^1, \\ &\hat{\varepsilon}^1 = -\frac{i}{2}\hat{\Gamma}^9\left(1 + \hat{\Gamma}_{11}\right)\hat{\epsilon}, \qquad \hat{\varepsilon}^2 = \tfrac{1}{2}\left(1 - \hat{\Gamma}_{11}\right)\hat{\epsilon}. \end{aligned}} \tag{17.40}$$

17.5 Consistent truncations and heterotic/type-I duality

In Section 16.4 we saw how the $N = 2A, d = 10$ theory could be truncated to $N = 1, d = 10$ SUGRA to which vector supermultiplets could be coupled. We also studied how this truncation and the addition of $\mathrm{E}_8 \times \mathrm{E}_8$ vector supermultiplets could be justified from the string/M-theory point of view as arising from an orbifold compactification of 11-dimensional supergravity (M theory) with ten-dimensional E_8 vector supermultiplets living on the two boundaries.

In this section we are going to study the possible consistent truncations of the $N = 2B, d = 10$ theory and their relations to other string effective-field theories. We will follow [117, 804] and, in particular, we will include in our discussion the RR 10-form potential and its S dual $\mathcal{B}^{(10)}$, to which one can give the on-shell supersymmetry transformation rule

$$\delta_{\hat{\varepsilon}}\hat{\mathcal{B}}^{(10)} = -ie^{-2\hat{\varphi}}\bar{\hat{\varepsilon}}\sigma^3\left(10\hat{\Gamma}_{[\hat{\mu}_1\cdots\hat{\mu}_9}\hat{\psi}_{\hat{\mu}_{10}]} - \hat{\Gamma}_{\hat{\mu}_1\cdots\hat{\mu}_{10}}\hat{\lambda}\right). \tag{17.41}$$

Several $\mathbb{Z}_2$ symmetries of the $N = 2B, d = 10$ theory are known.

1. $(-1)^F$, which changes the signs of all fermions and leaves bosons invariant. It is present in all theories and it is related to the truncation to $N = 0$ eliminating all fermions.

2. The symmetry associated with the worldsheet parity symmetry Ω,

$$\hat{f} \to \sigma^1\hat{f}, \quad \hat{C}^{(2n-2)} \to (-1)^n\hat{C}^{(2n-2)}, \quad \hat{\mathcal{B}} \to -\hat{\mathcal{B}}, \quad \hat{\mathcal{B}}^{(10)} \to -\hat{\mathcal{B}}^{(10)}, \tag{17.42}$$

where $\hat{f}$ stands for any fermion doublet. The associated truncations are

$$\hat{C}^{(2n-2)} = 0, \quad n = 1, 3, 5, \qquad \hat{\mathcal{B}} = 0, \qquad \hat{\mathcal{B}}^{(10)} = 0, \qquad (1 + \sigma^1)\hat{f} = 0. \tag{17.43}$$

The remaining fields are those of the $N = 1, d = 10$ supergravity multiplet (plus $\hat{C}^{(10)}$) but they now appear as in the type-I string effective action:

$$\boxed{S = \frac{\hat{g}_{\mathrm{I}}^2}{16\pi G_{\mathrm{NI}}^{(10)}}\int d^{10}\hat{x}\,\sqrt{|\hat{\jmath}|}\left\{e^{-2\hat{\varphi}}\left[\hat{R}(\hat{\jmath}) - 4\left(\partial\hat{\varphi}\right)^2\right] + \frac{1}{2\cdot 3!}\left(\hat{G}^{(3)}\right)^2\right\}.} \tag{17.44}$$

We know that the quotient of the type-IIB string theory by Ω is equivalent to the introduction of an O9-plane and that consistency requires the introduction of 32 D9-branes whose RR charges and tensions will cancel out exactly those of the O9-plane, introducing at the same time open strings whose massless states fill an SO(32) gauge supermultiplet. At lowest order in α' these vector supermultiplets will contribute to the above action with a term [64]

$$\boxed{\frac{\hat{g}_{\mathrm{I}}^2}{16\pi G_{\mathrm{NI}}^{(10)}}\int d^{10}\hat{x}\,\sqrt{|\hat{\jmath}|}\left\{\frac{\alpha'}{4}e^{-\hat{\varphi}}\,\mathrm{Tr}_{\mathrm{Adj}}\left(F^2\right)\right\}} \tag{17.45}$$

(with the Killing metric normalized to $K_{IJ} = -\delta_{IJ}$) plus the addition of a Chern–Simons term Eq. (A.50) to $\hat{G}^{(3)}$,

$$\hat{G}^{(3)} \to 3\partial\hat{C}^{(2)} - \tfrac{1}{2}\alpha'\hat{\omega}_3, \tag{17.46}$$

which is needed for the supersymmetric coupling. The sum of Eqs. (17.44) and (17.45) is the effective action of the type-I SO(32) superstring theory. Observe that the vector fields kinetic term carries a dilaton factor $e^{-\hat{\varphi}}$, which is associated with the fact that these terms come from different string diagrams (worldsheet topologies).

3. The product $(-1)^F\Omega$ induces a truncation of the $N = 2B, d = 10$ SUEGRA that consists in keeping the same bosonic fields and the combination of fermions orthogonal to that of the previous case. The result is again an $N = 1, d = 10$ SUGRA with bosonic action Eq. (17.44). From the string-theory point of view this truncation has been associated in [804] with one in which the tensions of the O9-plane and the D9-branes cancel out, but the charges do not. The corresponding string theory, which has gauge group USp(32) and was constructed in [873], is not supersymmetric, even though it is tachyon-free. The supersymmetry of the supergravity theory is broken by the coupling to matter, which fails to be supersymmetric because the vector fields are not in the adjoint representation (a necessary condition for supersymmetry). The consistency of the coupling is due to the fact that supersymmetry is spontaneously broken [330, 703], which makes it a fascinating theory.

4. The S-duality transformation $S = \eta$. It does not lead to any supersymmetric truncation, but it allows us to discuss the S duals of other truncations.

5. Those that correspond to the worldsheet transformations $(-1)^{F_{\mathrm{L}}}$ and $(-1)^{F_{\mathrm{R}}}$, where $F_{\mathrm{L(R)}}$ is the spacetime fermion number coming from the left- (right-)movers. These two transformations are related by

$$\Omega(-1)^{F_{\mathrm{L}}}\Omega = (-1)^{F_{\mathrm{R}}}, \qquad (-1)^F(-1)^{F_{\mathrm{L(R)}}} = (-1)^{F_{\mathrm{R(L)}}}. \tag{17.47}$$

Their action on the supergravity fields is

$$\hat{f} \to \pm\sigma^3\hat{f}, \qquad \hat{C}^{(2n-2)} \to -\hat{C}^{(2n-2)}. \tag{17.48}$$

The truncation is

$$\hat{C}^{(2n-2)} = 0, \quad n = 1, \ldots, 6, \qquad (1 \mp \sigma^3)\hat{f} = 0. \tag{17.49}$$

The remaining fields are those of pure $N = 1, d = 10$ supergravity just as they appear in the heterotic-string effective action Eq. (15.1) (plus $\hat{\mathcal{B}}^{(10)}$). In string theory one has to take into account the twisted sectors that arise and which have been argued to give the type-IIA superstring theory [280]. On the other hand, in [123, 580] it has been argued that $(-1)^{F_{\mathrm{L}}}$ is actually the S dual of Ω,

$$S\Omega S^{-1} = (-1)^{F_{\mathrm{L}}}, \tag{17.50}$$

so one can also consider the S dual of the construction that leads to the type-I theory: an O9-plane associated with $(-1)^{F_\mathrm{L}}$ and 32 S duals of the D9-branes (S9-branes). The result is the heterotic SO(32) superstring theory which arises, then, as the S dual of the type-I SO(32) superstring. These theories are each other's strong-coupling limit [278, 577, 782]. The fields of the effective theories are related by the strong–weak-coupling transformation

$$\boxed{\hat{\jmath}_{\hat{\mu}\hat{\nu}} = e^{-\hat{\phi}}\hat{g}_{\hat{\mu}\hat{\nu}}, \qquad \hat{\varphi} = -\hat{\phi}_h, \qquad \hat{C}^{(2)}{}_{\hat{\mu}\hat{\nu}} = \hat{B}_{\hat{\mu}\hat{\nu}}.} \tag{17.51}$$

Since $S(-1)^F\Omega S^{-1} = (-1)^{F_\mathrm{R}}$, one may also expect to find a (non-supersymmetric) heterotic dual of the USp(32) superstring.

We can combine the construction of the type-I theory and this heterotic/type-I duality with our knowledge of type-II T duality. Since we can consider the type-I theory as simply type IIB with one O9-plane and 32 D9-branes and we know what the T dual of each of them is (type IIA with an O8-plane and 32 D8-branes), we can immediately say that the T dual of the type-I theory (called type I′ [778]) is essentially a nine-dimensional theory with $N = 1$ supersymmetry (16 supercharges) and gauge group SO(32)× U(1)^2. Since

$$T\Omega T^{-1} = I_x\Omega, \qquad I_x x = -x, \tag{17.52}$$

the presence of the O8-plane implies that, instead of $\mathbb{R}^9 \times \mathrm{S}^1$, we have $\mathbb{R}^9 \times \mathrm{S}^1/\mathbb{Z}_2$ and we actually have two O8-planes with RR charge -16 in the two nine-dimensional boundaries. Introduction of Wilson lines into the compactification separates the D8-branes and one can obtain different gauge groups [66, 710].

The S-dual version of this T duality is well known to lead from the heterotic SO(32) theory to the heterotic $\mathrm{E}_8 \times \mathrm{E}_8$ theory (up to the possible introduction of Wilson lines) with one dimension compactified, which is associated with the Hořava–Witten scenario with one extra dimension compactified, that is, M theory on $\mathrm{S}^1 \times \mathrm{S}^1/\mathbb{Z}_2$. We have learned that type-IIB S duality is a rotation of the 2-torus on which we compactify M theory and here we are seeing precisely that the type I′ theory is a rotated version of the heterotic $\mathrm{E}_8 \times \mathrm{E}_8$ theory compactified on a circle and both are related to M theory. Furthermore, the mysterious objects at the boundaries of the Hořava–Witten scenario, compactified on a circle, are related to the O8-planes and D8-branes. More consequences of these chains of dualities were studied in [123].

18
Extended objects

Introduction

In the previous chapters we have studied the upper-left- and upper-right-hand boxes of Figure 14.1 that concern the standard perturbative formulation of string theory and the effective actions of the ten-dimensional string theories (and M theory). We have also learned a bit about the existence of some non-perturbative states in the string spectrum, in particular D-branes and KK and winding modes in compactified theories (the lower-left-hand box of Figure 14.1). We have studied in the three cases the existence of dualities that related various theories and how these dualities are realized in the worldsheet action (when this is possible, i.e. for T duality) and in the effective actions. We have also mentioned that S dualities and T dualities imply the existence of new solitonic states in the string spectrum.

In this chapter and the next we are going to study systematically the lower-right-hand and central boxes of Figure 14.1, that is, the solitonic solutions of the string effective-field theories and their worldvolume actions. We will study the implications that the various dualities have for them (which are evidently related to the effects of dualities on the effective actions) and for the non-perturbative string spectrum. This chapter will be devoted to a general introduction to extended objects and in the next chapter we will deal specifically with those that occur in string/M theory.

These are subjects with many facets that are related in many ways to each other and to the subjects of the previous chapters. Therefore, it is hopeless to try to give a complete, or even half-complete, account of them in the space that we have at our disposal. Our aim will be to cover the basic material and the essential results and solutions in a unified system of conventions (like the rest of the book), giving pointers to the literature for further developments.

We start in Section 18.1 with a general introduction to the kinematics and dynamics of generic extended objects in which we will discuss various forms of the actions for these objects, their coupling to background fields (Section 18.1.1), and the generalization of the Dirac quantization condition for extended objects (Section 18.1.2).

In Section 18.2 we treat the simplest generic black and extreme solutions of the "p-brane a-model," which is itself a generalization of the "a-model" studied in Section 12.1. The string-theory solutions that we will study later are in general special cases of these general families of solutions.

18.1 Generalities

The basic extended objects are known as *p-branes*, objects with p spatial dimensions that sweep out $(p+1)$-dimensional worldvolumes as they evolve in time in a d-dimensional ambient (or target) spacetime. Strings, which we have already studied, are the simplest examples of p-branes ($p=1$), but there are many other examples that differ by their worldvolume dimensions, their worldvolume fields, their couplings to background fields, and other characteristics such as being associated with a compact dimension (such as the KK monopole and more general KK*p-branes* [666]). We will discuss these variants of the basic p-brane in order of increasing complexity and in the next chapter we will see which of them occur in string/M theory.

The dynamics of all these objects is governed by their $(p+1)$-dimensional worldvolume actions and in what follows we are going to study them and use them as tools to classify the objects. In this chapter we consider only bosonic actions, but in the next chapter we will briefly discuss the κ-symmetric addition of fermions which is required by the coupling to supergravity.

18.1.1 Worldvolume actions

The basic dynamical variables of a p-brane are the spacetime coordinates of the object $X^\mu(\xi)$, $\mu = 0, \dots, d-1$ (ξ^i , $i = 0, \dots, p$ are the worldvolume coordinates), which give the embedding of the worldvolume in the d-dimensional target spacetime and are worldvolume scalar fields. Some p-branes may have additional worldvolume fields (scalars, vectors (such as the BI vector of Dp-branes), and tensors), whose physical meanings will be discussed in Section 19.6.

The simplest worldvolume and spacetime reparametrization-invariant action for a p-brane is the generalization of the Nambu–Goto action Eq. (14.1) (in the same notation),

$$\boxed{S^{(p)}_{\text{NG}}[X^\mu(\xi)] = -T_{(p)} \int d^{p+1}\xi \, \sqrt{|g_{ij}|}\,,} \tag{18.1}$$

which is proportional to the volume swept out by the p-brane. The proportionality constant $T_{(p)}$ is the p-brane tension and has natural dimensions of $L^{-(p+1)}$ or, equivalently, of mass per unit of spatial p-dimensional volume. In fact, let us consider a spacetime that is the direct product of a non-compact $(d-p)$-dimensional spacetime and a p-dimensional compact space so the metric of the original spacetime $\hat{g}_{\hat{\mu}\hat{\nu}} = \text{diag}(g_{\mu\nu}, g_{mn})$ and a configuration in which the p-brane[1] wraps the p-dimensional space so $\hat{X}^m = \xi^m$ $m = 1, \dots, p$ and the remaining embedding coordinates are independent of the $\hat{X}^m = \xi^m$. The Nambu–Goto (NG) action becomes the action of a massive particle moving in the $(d-p)$-dimensional

[1] This is the straightforward generalization of a string winding-mode configuration.

spacetime,

$$S^{(p)}_{\rm NG}[X^{\mu}(\xi)] = -T_{(p)} V_{(p)} \int d\xi^{0} \sqrt{|g_{\mu\nu}\dot{X}^{\mu}\dot{X}^{\nu}|}\,, \qquad V_{(p)} = \int d\xi^{1}\cdots d\xi^{p}\sqrt{|g_{mn}|}, \tag{18.2}$$

where $V_{(p)}$ is the volume of the internal manifold and $T_{(p)}V_{(p)}$ is the mass of the particle if the spacetime metric is asymptotically flat in the $d-p$ directions orthogonal ("transverse") to the worldvolume. Transverse space, the space in which the wrapped p-brane moves as a particle, plays a very important role in the definition of mass (as we have just seen) and charge.

As we have discussed several times, on the one hand, the NG action is highly non-linear and, on the other hand, it cannot describe massless (tensionless) objects (also known as *null branes*). These problems are solved by introducing auxiliary fields. Several possibilities have been proposed in the literature. For instance, we can introduce a scalar density field v and write the action

$$\boxed{S^{(p)}[X^{\mu}(\xi), v] = \int d^{p+1}\xi\, \frac{1}{2v}\left[\, |g| + v^{2}T^{2}_{(p)}\right],} \tag{18.3}$$

which is equivalent to the NG action upon elimination of v using its equation of motion. In this action we can take consistently the tensionless limit to obtain a null brane action [657]. Although this action is still highly non-linear, it is useful for certain purposes: we can replace the tension (a constant) by a worldvolume p-form potential[2] $c_{(p)\,i_1\cdots i_p}$ whose equation of motion tells us that the dual of its field strength $\mathcal{G}_{(p+1)} = (p+1)\partial c_{(p)}$ is just a constant. The action is [896]

$$\boxed{S^{(p)}[X^{\mu}(\xi), v, c_{(p)}] = \int d^{p+1}\xi\, \frac{1}{2v}\left[\, |g| + \left({}^{\star}\mathcal{G}_{(p+1)}\right)^{2}\right],} \tag{18.4}$$

where here ${}^{\star}\mathcal{G}_{(p+1)} = [1/(p+1)!]\epsilon^{i_1\cdots i_{p+1}}\mathcal{G}_{(p+1)\,i_1\cdots i_{p+1}}$ and the equation of motion of $c_{(p)}$ has the solution

$$\mathcal{G}_{(p+1)\,i_1\cdots i_p} = \frac{T_{(p)}}{v}\epsilon_{i_1\cdots i_p}, \tag{18.5}$$

where $T_{(p)}$ arises as just an integration constant. On substituting this solution into the action, we recover exactly the action Eq. (18.3) and then $T_{(p)}$ is identified as the p-brane tension. One can also consider solutions in which ${}^{\star}\mathcal{G}_{(p+1)}/v$ is only piecewise constant, the discontinuities being associated with brane intersections (see, for instance, [902] for an application involving string and D-string junctions). On the other hand, these actions are also suitable for supersymmetric objects and can be made κ-symmetric [135, 141].

---

[2] This is similar to the replacement of the mass parameter by the RR 9-form potential in Romans' massive supergravity, Section 16.2.

Another, more common, possibility is to introduce an independent metric on the worldvolume $\gamma_{ij}(\xi)$ and write the following Polyakov-type action:

$$\boxed{S_{\mathrm{P}}^{(p)}[X^{\mu},\gamma_{ij}] = -\frac{T_{(p)}}{2}\int d^{p+1}\xi\sqrt{|\gamma|}\left[\gamma^{ij}g_{ij} + (1-p)\right].} \tag{18.6}$$

The "cosmological-constant" term $(p-1)$ has been chosen to vanish in the string case $p = 1$ (otherwise conformal invariance would be broken) and for $p \neq 1$ to give $\gamma_{ij} = g_{ij}$ identically as the solution of the equation of motion for γ_{ij}. On substituting this solution into the above action to eliminate the auxiliary worldvolume metric, one recovers the NG action with the above normalization. (The $p = 1$ case was discussed in Section 14.1.)

In the $p \neq 1$ case, the "cosmological-constant" term can be considered as a mass (tension) term by following the same procedure as in the particle case. First we rescale the worldvolume metric,

$$\gamma_{ij} = T_{(p)}^{-\frac{2}{p-1}}\gamma'_{ij}, \tag{18.7}$$

giving

$$S_{\mathrm{P}}^{(p)}[X^{\mu},\gamma_{ij}] = -\tfrac{1}{2}\int d^{p+1}\xi\sqrt{|\gamma'|}\Big[\gamma'^{\,ij}g_{ij} + T_{(p)}^{\frac{p+1}{p-1}}(1-p)\Big]. \tag{18.8}$$

Now we can take the tensionless limit $T_{(p)} \to 0$ and then again rescale the worldvolume metric to a dimensionless metric, again obtaining an action suitable for describing null branes:

$$\boxed{S_{\mathrm{P}}^{(p)}[X^{\mu},\gamma_{ij}] = -\frac{P_{(p)}}{2}\int d^{p+1}\xi\sqrt{|\gamma''|}\,\gamma''^{\,ij}g_{ij}.} \tag{18.9}$$

The procedure of introducing auxiliary fields can also be used to linearize the Born–Infeld action of D-branes [3], as we will see.

KK-brane actions. Certain objects necessarily live in spacetimes with compact dimensions, with some of their worldvolume dimensions wrapped around them and no dynamics in those directions. Their worldvolume actions can be formulated as *gauged σ-models*. The main example is the KK monopole [122, 127, 375], whose interpretation as an extended object of string/M theory is implied by duality, as we will see. The effective actions of the M-theory objects which reduce to those of the objects of the massive type-IIA theory are also given by gauged σ-models [136, 376, 663, 746]. What follows is a short introduction to gauged σ-models (with no Wess–Zumino (WZ) term).

Our starting point is the $(p+1)$-dimensional σ-model (p-brane Polyakov-type action)

$$S = -\frac{T_{(p)}}{2}\int d^{p+1}\xi\sqrt{|\gamma|}\left[\gamma^{ij}g_{ij} - (p-1)\right], \tag{18.10}$$

which is invariant under the GCTs

$$\begin{aligned} X^{\mu} &\to X^{\mu\,\prime} = X^{\mu} + \epsilon^{\mu}(X),\\ g_{\mu\nu}(X) &\to g'_{\mu\nu}(X') = g_{\mu\nu}(X) - 2\big(\partial_{(\mu|}\epsilon^{\rho}\big)g_{\rho|\nu)}. \end{aligned} \tag{18.11}$$

Let us assume now that the metric admits an isometry group generated by the Killing vectors $k_{(I)}{}^{\mu}(X), \quad I = 1, \ldots, r$,

$$[k_{(I)}, k_{(J)}] = f_{IJ}{}^{K} k_{(K)}, \tag{18.12}$$

and let us consider the following infinitesimal transformations (which are *not* infinitesimal GCTs):

$$\begin{aligned} X^{\mu} &\rightarrow X^{\mu\,\prime} = X^{\mu} + \eta^{I} k_{(I)}{}^{\mu}, \\ g_{\mu\nu}(X) &\rightarrow g_{\mu\nu}(X') = g_{\mu\nu}(X) + \eta^{I} k_{(I)}{}^{\lambda} \partial_{\lambda} g_{\mu\nu}, \end{aligned} \tag{18.13}$$

where the η^{I}s are constant infinitesimal parameters. The variation of the action is

$$\delta_{\eta} S = -T_{(p)} \int d^{p+1}\xi \sqrt{|\gamma|} \eta^{I} \gamma^{ij} \partial_{i} X^{\mu} \partial_{j} X^{\nu}\ \nabla_{(\mu|} k_{(I)|\nu)} \tag{18.14}$$

and vanishes if (as is assumed) the $k_{(I)}$ satisfy the Killing equation $\nabla_{(\mu|} k_{(I)|\nu)} = 0$.

Now we want to gauge this symmetry. The infinitesimal transformations will be

$$\begin{aligned} \delta_{\eta} X^{\mu} &= \eta^{I}(\xi) k_{(I)}{}^{\mu}(X), \\ \delta_{\eta} g_{\mu\nu}(X) &= \eta^{I}(\xi) k_{(I)}{}^{\lambda} \partial_{\lambda} g_{\mu\nu}\,. \end{aligned} \tag{18.15}$$

Observe that the Killing vectors transform as follows:

$$\begin{aligned} \delta_{\eta} k_{(I)} &= \eta^{(J)} k_{(J)}{}^{\nu} \partial_{\nu} k_{(I)}{}^{\mu} = \eta^{(J)} \left(\left[k_{(J)}, k_{(I)} \right]^{\mu} + k_{(I)}{}^{\nu} \partial_{\nu} k_{(J)}{}^{\mu} \right) \\ &= \eta^{(J)} f_{JI}{}^{K} k_{(K)}{}^{\mu} + \left(\eta^{(J)} \partial_{\nu} k_{(J)}{}^{\mu} \right) k_{(I)}{}^{\nu}. \end{aligned} \tag{18.16}$$

To make the σ-model invariant under the above local transformations, it suffices to replace the partial derivative of the worldvolume scalars X^{μ} by the covariant derivative

$$\mathcal{D}_{i} X^{\mu} = \partial_{i} X^{\mu} + C^{I}{}_{i} k_{(I)}{}^{\mu}, \tag{18.17}$$

where we have introduced the non-dynamical worldvolume vector fields $C^{(I)}{}_{i}$ which transform as standard gauge potentials:

$$\delta_{\eta} C^{I}{}_{i} = -\left(\partial_{i} \eta^{I} + f_{JK}{}^{I} C^{J}{}_{i} \eta^{K} \right) = -\mathcal{D}_{i} \eta^{I}. \tag{18.18}$$

The covariant derivative defined above transforms covariantly, that is, with no derivatives of the gauge parameter:

$$\delta_{\eta} \mathcal{D}_{i} X^{\mu} = \left(\eta^{J} \partial_{\nu} k_{(J)}{}^{\mu} \right) \mathcal{D}_{i} X^{\nu}. \tag{18.19}$$

The gauged σ-model action (without WZ term) then takes the form

$$\boxed{S = -\frac{T_{(p)}}{2} \int d^{p+1}\xi \sqrt{|\gamma|} \left[\gamma^{ij} \mathcal{D}_{i} X^{\mu} \mathcal{D}_{j} X^{\nu} g_{\mu\nu} - (p-1) \right].} \tag{18.20}$$

Since the worldvolume vector fields $C^{a}{}_{i}$ are not dynamical (their derivatives do not occur in the action), they play the role of Lagrange multipliers and may be eliminated by using their equations of motion directly in the action (at least in this simple σ-model with no WZ term). These are

$$k_{(I)\ \mu}\mathcal{D}_i X^\mu = 0, \tag{18.21}$$

and their solution is

$$C^I{}_i = -h^{IJ} k_{(J)}{}^\mu \partial_i X^\nu g_{\mu\nu}, \tag{18.22}$$

where we have defined the metric h_{IJ} (*assumed* to be invertible, which is not true in general, but is in many cases of interest),

$$h_{IJ} = k_I{}^\mu k_J{}^\nu g_{\mu\nu}, \qquad h^{IJ} h_{JK} = \delta^I{}_K. \tag{18.23}$$

In this case, we have on-shell

$$\mathcal{D}_i X^\mu = \left(g^\mu{}_\nu - h^{IJ} k_{(I)}{}^\mu k_{(J)\ \nu}\right) \partial_i X^\nu. \tag{18.24}$$

Observe that the matrix

$$\Pi^\mu{}_\nu \equiv \left(g^\mu{}_\nu - h^{IJ} k_{(I)}{}^\mu k_{(J)\ \nu}\right), \qquad \Pi^\mu{}_\nu \Pi^\nu{}_\rho = \Pi^\mu{}_\rho, \tag{18.25}$$

projects onto the space orthogonal to the orbits of the isometry group:

$$\Pi^\mu{}_\nu k_{(I)}{}^\nu = 0, \qquad \forall\, I = 1, \ldots, r, \tag{18.26}$$

so r directions simply disappear from the σ-model action. We will see this more clearly when we perform the target-space (*direct*) dimensional reduction of the gauged σ-model.

After eliminating the auxiliary vector fields (assuming that this was possible), the gauged σ-model takes the form

$$S = -\frac{T_{(p)}}{2} \int d^{p+1}\xi \sqrt{|\gamma|}\left[\gamma^{ij}\partial_i X^\mu \partial_j X^\nu \Pi_{\mu\nu} - (p-1)\right], \tag{18.27}$$

since $g_{\rho\sigma}\Pi^\rho{}_\mu \Pi^\sigma{}_\nu = \Pi_{\mu\nu}$. In the case of just one isometry, we have seen in Section 11.2 that, in adapted coordinates, $\hat{\Pi}_{\hat{\mu}\hat{\nu}}$ is zero except for the $(\hat{d}-1)\times(\hat{d}-1)$ submatrix $\hat{\Pi}_{\mu\nu}$ which is the metric in $\hat{d}-1$ dimensions. The above σ-model (adding hats everywhere) does not depend on the isometric coordinate Z and reduces simply to a σ-model with $(\hat{d}-1)$-dimensional target space.

In this simple case, then, the gauged σ-model with d-dimensional target space is actually a σ-model with $(d-1)$-dimensional target space in disguise, written in d-dimensional covariant language. In more general cases it is not possible to eliminate completely the non-physical degrees of freedom (such as Z), but the σ-model still has $d-r$ degrees of freedom.

Let us now consider the coupling of p-branes to other spacetime fields. The simplest and more natural ones are the couplings to scalar fields (which act as local coupling "constants") and to $(p+1)$-form potentials, the fields p-branes can be charged with respect to. Let us start with $(p+1)$-form potentials.

18.1.2 Charged branes and Dirac charge quantization for extended objects

The Lagrangian of a $(p+1)$-form potential $A_{(p+1)\,\mu_1\cdots\mu_{p+1}}$ is usually constructed in terms of its $(p+2)$-form field strength,

$$F_{(p+2)\,\mu_1\cdots\mu_{p+2}} = (p+2)\partial_{[\mu_1}A_{(p+1)\,\mu_2\cdots\mu_{p+2}]} \tag{18.28}$$

($F_{(p+2)} = dA_{(p+1)}$ in differential-forms language), which is invariant under the gauge transformations generated by a p-form $\Lambda_{(p)}$,

$$\delta A_{(p+1)\,\mu_1\cdots\mu_{p+1}} = (p+1)\partial_{[\mu_1}\Lambda_{(p)\,\mu_2\cdots\mu_{p+1}]} \tag{18.29}$$

($\delta A_{(p+a)} = d\Lambda_{(p)}$ in differential forms language). The Lagrangian is[3]

$$S[A_{(p+1)}] = \int d^dx\sqrt{|g|}\left[\frac{(-1)^{(p+1)}}{2\cdot(p+2)!}F^2_{(p+2)}\right], \tag{18.30}$$

and the equation of motion is just

$$\nabla_\mu F_{(p+2)}{}^{\mu\mu_1\cdots\mu_{p+1}} = 0 \tag{18.31}$$

($d^\star F_{(p+2)} = 0$ in differential-forms language). As usual, we can work directly with the field strength, provided that we impose on it the Bianchi identity

$$\nabla_\mu{}^\star F_{(p+2)}{}^{\mu\mu_1\cdots\mu_{d-p-3}} = 0 \tag{18.32}$$

($dF_{(p+2)} = 0$) to ensure the local existence[4] of the $A_{(p+1)}$.

$A_{(p+1)}$ couples naturally to a current $j^{\mu_1\cdots\mu_{p+1}}$, the coupling being represented by the Lagrangian term

$$\int d^dx\sqrt{|g|}\frac{(-1)^{(p+1)}}{(p+1)!}A_{(p+1)\mu_1\cdots\mu_{p+1}}j^{\mu_1\cdots\mu_{p+1}}. \tag{18.34}$$

This term is gauge-invariant if the current is divergence-free ("conserved"),

$$\nabla_\mu j^{\mu\mu_1\cdots\mu_p} = 0 \tag{18.35}$$

($d^\star j = 0$ in differential-forms language). This condition follows from the gauge identity of the free theory $\nabla_\mu\nabla_\nu F_{(p+2)}{}^{\mu\nu\mu_1\cdots\mu_p} = 0$ (off-shell) plus the equation of motion

$$\nabla_\mu F_{(p+2)}{}^{\mu\mu_1\cdots\mu_{p+1}} = j^{\mu_1\cdots\mu_{p+1}}, \tag{18.36}$$

or, in differential-forms language,

$$\delta F = (-1)^d j, \qquad d\,{}^\star F_{(p+2)} = (-1)^{d+p}\,{}^\star j. \tag{18.37}$$

---

[3] We are ignoring here possible scalar couplings, Chern–Simons terms, etc.

[4] Given a field strength $F_{(p+2)}$ satisfying the Bianchi identity, then $A_{(p+1)}$ is given, up to gauge transformations, by the formula that generalizes Eq. (8.31):

$$A_{\mu_1\cdots\mu_{p+1}} = (-1)^{(p+1)}\int_0^1 d\lambda\lambda^{(p+1)}F_{(p+2)\,\mu_1\cdots\mu_{p+2}}(\lambda x)x^{\mu_{p+2}}. \tag{18.33}$$

The conservation law for the current $d^\star j = 0$ suggests the following definition for the charge q_p associated with $A_{(p+1)}$:

$$q_p \equiv \int_{\mathrm{B}^{(d-p-1)}} {}^\star j, \tag{18.38}$$

where, by definition, $\mathrm{B}^{(d-p-1)}$ is a *capping surface* whose boundary is (topologically) $\partial\mathrm{B}^{(d-p-1)} = \mathrm{S}^{(d-p-2)}$. More precisely, this is the charge contained in the capping surface. The total charge would be calculated by integrating over a capping surface whose boundary is the $(d-p-2)$-sphere at infinity $\mathrm{S}^{(d-p-2)}_\infty$. As usual this definition is invariant under smooth deformations of the capping surface in source-free ($j = 0$) regions.

Using the generalization of the Gauss law Eq. (18.37) and Stokes' theorem, we obtain the usual definition

$$\boxed{q_p = (-1)^{d+p} \int_{\mathrm{S}^{(d-p-2)}} {}^\star F_{(p+2)},} \tag{18.39}$$

which is also invariant under smooth deformations of the $(d-p-2)$-dimensional surface in source-free regions.

Given a p-brane sweeping out a $(p+1)$-dimensional worldvolume $\mathrm{W}^{(p+1)}$, we can immediately construct a conserved current for it that generalizes Eq. (8.49) for the current of a point-like charged object:

$$\begin{aligned} j^{\mu_1\cdots\mu_{p+1}}(x) &= q_p \int_{\mathrm{W}^{(p+1)}} dX^{\mu_1} \wedge \cdots \wedge dX^{\mu_{p+1}} \frac{\delta^{(d)}(x - X(\xi))}{\sqrt{|g(X)|}} \\ &= q_p \int_{\mathrm{W}^{(p+1)}} d^{p+1}\xi \frac{\partial(X^{\mu_1}\cdots X^{\mu_{p+1}})}{\partial(\xi^{\mu_1}\cdots\xi^{\mu_{p+1}})} \frac{\delta^{(d)}(x - X(\xi))}{\sqrt{|g(X)|}}. \end{aligned} \tag{18.40}$$

The charge associated with this current is q_p, except when the p-brane worldvolume has boundaries, since, in that case, we can continuously deform the surface of integration and contract it to a point without meeting the p-brane source, obtaining zero charge. This is easy to visualize for a string of finite length in a four-dimensional target space: the string charge with respect to a 2-form potential is calculated through a closed line integral around the string that can be slid off the string and contracted to a point as shown in Figure 18.1. In a sense this explains why the KR 2-form does not appear in the open-string spectrum. It is clear now that $q_p \neq 0$ only when the p-brane has compact topology or extends to infinity. A brane with boundaries can also carry charge if the boundaries are attached to another object just as open strings are attached to D-branes. This case, which is more complicated (and interesting), will be studied in Section 19.6 since it depends strongly on the theory we are considering.

On substituting the above p-brane current into the interaction term Eq. (18.34), we obtain the standard form of the WZ term for p-branes which appears in the p-brane worldvolume action,

$$(-1)^{(p+1)} q_p \int_{\mathrm{W}^{(p+1)}} A_{(p+1)}. \tag{18.41}$$

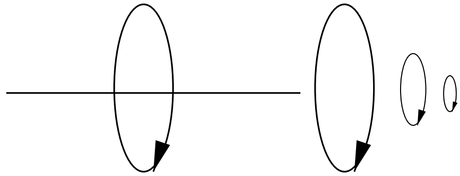

Fig. 18.1. Open strings cannot carry charge.

Electric–magnetic duality for extended objects. The electric–magnetic dual of a charged p-brane is, by definition, an object that couples to the electric-magnetic dual of the $(p+1)$-form potential $A_{(p+1)}$. We have seen several examples, for instance, in Section 16.5.5, in which we used Poincaré duality to replace the KR 2-form completely by its dual in various dimensions and also in Sections 16.1.3 and 17.1.1 in which we defined the on-shell duals of the RR potentials that appear in the $N = 2A, B, d = 10$ supergravity actions, not being able to replace the original potentials completely by their duals.

In all (massless) cases, the dual is a $(\tilde{p}+1)$-form $\tilde{A}_{(\tilde{p}+1)}$ with $\tilde{p} = d - p - 4$, whose field strength is, by definition, the Hodge dual of $\tilde{F}_{\tilde{p}+2} = {}^{\star}F_{(p+2)}$. The electric-magnetic dual of a p-brane is therefore a $\tilde{p}$-brane with $\tilde{p} \neq p$ in general, which is electrically charged with respect to $\tilde{A}_{(\tilde{p}+1)}$. Only in even dimensions do objects with $p = (d-4)/2$ have duals of the same dimension, and then we can have p-brane dyons.

A new feature with respect to point-particles is that there can be *self-dual p-branes*, charged with respect to a self-dual potential: point-particles couple only to vectors, which are dual to vectors in $d = 4$, but, in $d = 4$, self- or anti-self-duality is consistent only with a Euclidean signature. However, self-dual 2-forms in $d = 6$ and 4-forms in $d = 10$ can be consistently defined. They occur in $N = (1, 0), d = 6$ SUGRA (Section 13.4.1) and in $N = 2B, d = 10$ SUEGRA (Chapter 17), and are associated with chiral theories.

Dirac charge quantization. The charge of a p-brane moving in the background $A_{(p+1)}$ field sourced by a dual $\tilde{p}$-brane is quantized as in the case of point-particles [718, 881, 882].[5]

Let us consider the quantum propagation of a charged p-brane moving along a closed path[6] so its worldvolume is topologically a $(p+1)$-sphere $\mathrm{S}^{(p+1)}$. The interesting term in the path integral is the WZ term which, using Stokes' theorem, can be written in the form

$$(-1)^{(p+1)} q_p \int_{\mathrm{B}^{(p+2)}} dA_{(p+1)} = (-1)^{(p+1)} q_p \int_{\mathrm{B}^{(p+2)}} F_{(p+2)}, \qquad (18.42)$$

where $\mathrm{B}^{(p+2)}$ is one of the many possible $(p+2)$-dimensional capping surfaces with $\partial\mathrm{B}^{(p+2)} = \mathrm{S}^{(p+1)}$. To avoid having any ambiguities in the path integral, the difference between choosing two different capping surfaces $\mathrm{B}_1^{(p+2)}$ and $\mathrm{B}_2^{(p+2)}$ must be an integer

[5] The derivation of the Dirac quantization condition we are about to explain is different from those used in Section 8.7.2 for point-particles, which can also be generalized to charged p-branes.

[6] This will be the analog of a point-particle moving along a closed path encircling a Dirac string.

multiple of 2π:

$$(-1)^{(p+1)} q \int_{\mathrm{B}_1^{(p+2)}} F_{(p+2)} - (-1)^{(p+1)} q \int_{\mathrm{B}_2^{(p+2)}} F_{(p+2)} = 2\pi n. \tag{18.43}$$

Now, the first capping surface plus the second (with reversed orientation) form (topologically) a $(p+2)$-sphere $\mathrm{S}^{(p+2)}$, and we arrive at the condition

$$q_p \int_{\mathrm{S}^{(p+2)}} F_{(p+2)} = (-1)^{(p+1)} 2\pi n. \tag{18.44}$$

Defining now the "magnetic" charge, or, better, the electric charge of the dual $\tilde{p}$-brane, by

$$\boxed{p_p = q_{\tilde{p}} = (-1)^{d+\tilde{p}} \int_{\mathrm{S}^{p+2}} F_{(p+2)},} \tag{18.45}$$

the above result takes the form of the generalization of the Dirac quantization condition we were after:

$$\boxed{q_p q_{\tilde{p}} = 2\pi n, \qquad n \in \mathbb{Z}.} \tag{18.46}$$

The charge quantization condition for p-brane dyons was found in [305].

Observe that, in the string σ-model action Eq. (15.31), the constant that plays the role of the charge with respect to the KR 2-form is the string tension T. This identity, which has the form of a BPS bound, indicates that the object described by the action is a BPS object. All the extended objects we are going to deal with have the same property.

Observe also that, even though we have defined a scalar charge q_p because this is a useful quantity, this number does not give all the information. Actually, the charge should be the tensor

$$\boxed{Z^{\mu_1 \cdots \mu_p} \sim \int_\Sigma d^{d-1} \Sigma_\mu j^{\mu\mu_1 \cdots \mu_p}} \tag{18.47}$$

(where Σ is a spacelike hypersurface) and includes information about spatial orientation, etc. This explains why objects with the same tensions and q_ps can be in equilibrium if they are parallel (so their $Z^{\mu_1 \cdots \mu_p}$s are identical) but not when one of them is tilted. These tensorial charges appear naturally in supersymmetry algebras, as we are going to see in Section 19.5.

18.1.3 The coupling of p-branes to scalar fields

A spacetime scalar $K(X)$ can be introduced in only one place in the NG action if we want to preserve reparametrization invariance and the gauge invariance of the WZ term,

$$S_{\mathrm{NG}}^{(p)}[X^\mu(\xi)] = -T_{(p)} \int d^{p+1}\xi (K/K_0) \sqrt{|g_{ij}|}. \tag{18.48}$$

We have introduced K_0, the asymptotic value of K at infinity (assuming that the metric $g_{\mu\nu}$ is asymptotically flat, at least in the directions transverse to the p-brane), so $T_{(p)}$ is the physical p-brane tension[7] and usually is proportional to K_0.

The coupling to K can always be changed or eliminated by a Weyl rescaling of the spacetime metric. It is important to make clear in which Weyl conformal reference frame we are writing the p-brane action. There are two special frames that can always be defined. We use the fundamental string worldsheet action to illustrate the definitions.

(Fundamental) p-brane frames. These are defined as the Weyl conformal frames in which the p-brane action does not couple to the scalar K. At the same time, all the terms in the action for the spacetime fields should carry the same K-dependent factor [337].

For instance, by definition, the fundamental string worldsheet action Eq. (15.31) does not depend on the dilaton when it is written in the string conformal frame. At the same time, all the terms in the action for the spacetime fields that couple to the string Eq. (15.1) carry the same $e^{-2\phi}$ factor.

Dual p-brane frames. These are the frames in which the electric-magnetic dual $\tilde{p}$-brane would be fundamental. In the fundamental string case in d dimensions the KR 2-form is dual to a $(d-4)$-form potential C with field strength $K = (d-3)\partial C$ and the dual object has $\tilde{p} = d-5$ (a 5-brane in $d = 10$). On Poincaré dualizing the KR 2-form, we obtain

$$S = \frac{g^2}{16\pi G_{\mathrm{N}}^{(d)}} \int d^d x \sqrt{|g|} e^{-2\phi} \left[R - 4(\partial\phi)^2 + \frac{(-1)^{(d-4)}}{2\cdot(d-3)!} e^{4\phi} G^2 \right]. \tag{18.49}$$

Thus, $g_{\mu\nu}$, the string metric, is not the metric to which the dual $(d-5)$-brane naturally couples. On performing a conformal transformation,

$$g = \Omega_{(1)-(d-5)}\, g_{(d-5)}, \tag{18.50}$$

and imposing that the dilaton factor is the same for all terms in the action, one obtains

$$\Omega_{(1)-(d-5)} = e^{\frac{4}{d-4}\phi}, \tag{18.51}$$

and

$$S = \frac{g^2}{16\pi G_{\mathrm{N}}^{(d)}} \int d^d x \sqrt{|g_{(d-5)}|}\, e^{\frac{4}{d-4}\phi} \left[R + \frac{(-1)^{(d-4)}}{2\cdot(d-3)!} G^2 \right]. \tag{18.52}$$

In the frame $g_{(d-5)}$, by definition, the NG $(d-5)$-brane action has no dilaton factors. Then, going back to the string frame, we find

$$S_{\mathrm{NG}}^{(d-5)}[X^\mu(\xi)] = -T_{(d-5)} \int d^{d-4}\xi\; e^{-2\phi} \sqrt{|g_{ij}|}. \tag{18.53}$$

The dual of the fundamental string has, then, a tension proportional to g^{-2} (g being the string coupling constant), as a typical solitonic object. The dual object to the fundamental

[7] This is called in the literature *effective tension*. See footnote 4 on page 435.

string in $d = 10$ dimensions is known as the *solitonic 5-brane* or S5-brane (also called *NS5-brane* to distinguish it from the D5-brane that couples to the RR 6-form potential).

It is natural to use in all cases the string frame because it is the theory of strings that we know how to quantize (even if imperfectly). From the string point of view, we can classify all the extended objects that we are going to see in terms of the scalar couplings that appear in their NG actions

Fundamental p-branes They do not couple to the dilaton (to lowest order in α'), which does not occur in the NG action

$$S_{\rm NG}^{(\rm pf)}[X^\mu(\xi)] = -T_{(p)} \int d^{p+1}\xi \, \sqrt{|g_{ij}|}. \tag{18.54}$$

Their mass is independent of the string coupling constant $g = e^{\phi_0}$ (ϕ_0 being the constant value of the dilaton at infinity). In $d = 10$ there is only one: the fundamental string. In $d = 11$ the M2- and M5-branes can both be considered fundamental.

Solitonic p-branes They couple to the dilaton as follows:

$$S_{\rm NG}^{(ps)}[X^\mu(\xi)] = -T_{(p)} \int d^{p+1}\xi \; e^{-2\phi}\sqrt{|g_{ij}|}. \tag{18.55}$$

Their mass is proportional to g^{-2}, which is typical of standard solitons. In $d = 10$ there is only one: the S5-brane.

Dirichlet (D) p-branes They couple to the dilaton as follows:

$$S_{\rm NG}^{(\rm D)}[X^\mu(\xi)] = -T_{(p)} \int d^{p+1}\xi \; e^{-\phi}\sqrt{|g_{ij}|}. \tag{18.56}$$

Their mass is proportional to g^{-1}. They are a new type of purely stringy solitons. They occur only in $d = 10$ and lower dimensions and couple to RR potentials.

Momentum modes They are charged point-like objects that couple to the KK scalar k as follows:

$$S_{\rm NG}[X^\mu(\xi)] = -T_{(0)} \int d\xi \; k^{-1}\sqrt{|g_{\xi\xi}|}. \tag{18.57}$$

Their mass is proportional to k_0^{-1}, that is, to the inverse of the radius of the compact dimension, and they couple to the KK vector. We have met them in several places. We are going to see that $d = 11$ momentum modes can be seen as $d = 10$ D0-branes, given the relation between the KK scalar k and the dilaton and between the KK vector and the RR 1-form, Eqs. (16.35).

Winding modes[8] They are charged point-like objects that couple to the KK scalar k as follows:

$$S_{\rm NG}[X^\mu(\xi)] = -T_{(0)} \int d\xi \; k\sqrt{|g_{\xi\xi}|}. \tag{18.58}$$

[8] There are also winding modes associated with other branes wrapped on compact spaces. Here we refer only to the string winding modes.

Their mass is proportional to k_0, that is, to the radius of the compact dimension, and they couple to the winding vector.

Kaluza–Klein (KK) branes They are described by gauged σ-models and couple to a positive power of the volume of the compact space associated with the gauging. In the decompactification limit the tension becomes infinite. Thus, these objects exist only when there are compact dimensions. The archetype of these objects is the KK monopole, which is described by a U(1)-gauged σ-model and coupled both to the dilaton and to the KK scalar k as follows:

$$S_{\rm NG}^{\rm (KK)}[X^{\mu}(\xi)] = -T_{(p)} \int d^{p+1}\xi \; e^{-2\phi} k^2 \sqrt{|\Pi_{ij}|}. \tag{18.59}$$

18.2 General p-brane solutions

In this section we are going to construct the simplest classical solutions that describe uncharged (*Schwarzschild*) and charged p-branes in a d-dimensional spacetime. They are solutions of a generalization proposed in [557] of the a-model discussed in Section 12.1 in which we will replace the 1-form potential adequate for BHs (which, in a sense, are point-like objects, 0-branes) by a $(p+1)$-form potential that is adequate for p-branes. This p-brane a-model is, for specific as and ps, a simplified version of most of the supergravity actions we are dealing with and the solutions we obtain will be supergravity (superstring) solutions.

We are also going to see how, according to [337], it is possible to find a p-brane worldvolume source for the "extreme" ones, generalizing the results we obtained for ERN and KK BHs (Sections 8.4 and 11.2.3, respectively), and for the fundamental string solution (Section 15.3), which will be particular cases of our general solution.

18.2.1 *Schwarzschild black p-branes*

The Schwarzschild solution describes the gravitational field of a massive, point-like object in vacuum. Is there an analogous solution of the d-dimensional Einstein–Hilbert action describing the gravitational field of a massive p-brane in vacuum?

Let us consider the simplest p-brane configuration (a "flat" p-brane with trivial topology). This configuration should give rise to an asymptotically flat spacetime characterized by $p+1$ translational isometries associated with the p-brane worldvolume. The requirement that the solution be asymptotically flat is essential, if we want to describe an *isolated* p-brane. However, we cannot impose asymptotic flatness in the direction along which the isometries act, but only in the $d-(p+1)$ spacelike transverse dimensions. In what follows, we will use the concept of asymptotic flatness in this restricted sense.

Solutions with these properties, *Schwarzschild p-branes*, were constructed using KK techniques in Section 11.3.3, and the metric is given by Eqs. (11.148), although here we are going to ignore all the hats. The first thing we can do is compute the p-brane tension[9] using

[9] The restricted asymptotic flatness of the metric does not allow us to define a finite d-dimensional mass.

the definition we gave in Section 18.1.1. This was done in Section 11.3.3, Eqs. (11.149) and (11.150), and the result can be expressed in this way: we can identify the p-brane tension as the constant $T_{(p)}$ in the expansion of the g_{tt} component of the metric

$$g_{tt} \sim 1 - \frac{16\pi G_{\mathrm{N}}^{(d)} T_{(p)}}{(\tilde{p}+2)\omega_{(\tilde{p}+2)}} \frac{1}{r^{\tilde{p}+1}}. \tag{18.60}$$

The coefficient of $r^{-(\tilde{p}+1)}$ is, by definition, the p-brane Schwarzschild radius to the power $\tilde{p}+1$.

It is clear from the definition that infinite (uncompactified) p-branes have infinite energy.

Is this Schwarzschild p-brane solution analogous to the Schwarzschild BH solution in the sense that it exhibits an event horizon? In other words: is it a *black* p-brane solution? In the presence of the p translational isometries the only sensible definition of an event horizon is equivalent to the standard definition of an event horizon in the transverse space. The event horizon thus defined becomes a $(p+2)$-dimensional extended object: the product of a BH horizon and the p-dimensional Euclidean space spanned by the p-brane. Clearly, for positive tension, the Schwarzschild p-brane solution has an event horizon of that form, whereas, for negative tension, the curvature singularity at $r=0$ will be naked.

When the spacetime has n compact dimensions, it is possible to have black p-branes, $p \leq n$, wrapping the compact dimensions with the same (finite) mass and event horizons of different topologies. Therefore, the uniqueness of four-dimensional BHs is not true in higher dimensions if some of the dimensions are compact. We know now that it is not true even in the absence of compact dimensions, as the existence of the asymptotically flat rotating black ring of [373] shows.

Black p-brane solutions (the Schwarzschild ones and the charged non-extreme ones that we are going to see next) are classically unstable [477, 478] (the charged, extreme ones are stable [479], as is usual in supersymmetric solutions) under linear perturbations along the worldvolume dimensions with wavelengths larger than the Schwarzschild radius. On the other hand, they are also quantum-mechanically unstable, because there is Hawking radiation associated with their event horizons, but also because the area of the event horizon (entropy) of several BHs with the same total mass is in general larger than the area of the event horizon of a Schwarzschild p-brane. For a Schwarzschild string compactified on a circle this happens whenever the length of the circle is larger than the Schwarzschild p-brane radius. The two instabilities seem to be related in the sense that the classical one is present whenever the thermodynamical one is present [485, 486, 800].

Although the thermodynamical argument seems to indicate that a black p-brane will break up into several BHs that will eventually merge into one, it has been argued in [551] that, for the black string, this process cannot take place in a finite time and that, instead, the black string decays into a new non-translationally invariant ("inhomogeneous") black string. Initial data sets for inhomogeneous p-brane solutions have subsequently been proposed in [552].

As we did in the BH case, we are going to use the Schwarzschild p-brane solution Eqs. (11.148) as our basic black p-brane solution and we are going to see that the charged

p-brane solutions of the p-brane a-model we are about to study can be built by "dressing" it with appropriate factors.

18.2.2 The p-brane a-model

As we have already said, this is the simplest model that embodies the main characteristics of the string effective action (and supergravity actions): gravity coupled to one scalar and a $(p+1)$-form to which the scalar couples non-minimally. It generalizes the a-model for BHs that we used before, which appears here as the $p=0$ case, but here we canonically normalize the $(p+1)$-form field strengths. The action is[10]

$$\boxed{S = \frac{1}{16\pi G_{\rm N}^{(d)}} \int d^d x \sqrt{|g|} \left[R + 2(\partial\varphi)^2 + \frac{(-1)^{p+1}}{2\cdot(p+2)!} e^{-2a\varphi} F_{(p+2)}^2 \right],} \tag{18.61}$$

where $F_{(p+2)} = dA_{(p+1)}$ is the field strength of the $(p+1)$-form potential $A_{(p+1)}$.

The equations of motion are

$$\begin{aligned} G_{\mu\nu} + 2T^{\varphi}_{\mu\nu} - \tfrac{1}{2} e^{-2a\varphi} T^{A_{(p+1)}}_{\mu\nu} &= 0, \\ \nabla^2\varphi + \frac{(-1)^{p+1}}{4\cdot(p+2)!} a e^{-2a\varphi} F_{(p+2)}^2 &= 0, \\ \nabla_\mu \left(e^{-2a\varphi} F_{(p+2)}{}^{\mu\nu_1\cdots\nu_{p+1}} \right) &= 0, \end{aligned} \tag{18.62}$$

where $T^{A_{(p+1)}}$ is the $(p+1)$-form energy–momentum tensor, given in Eq. (1.122). As usual, not all of them are independent in general (they can be derived from the Bianchi identities). On the other hand, the solutions we will find will be defined up to gauge transformations, including *large* gauge transformations that change the asymptotic behavior of the fields so the physics could be inequivalent. The classical equations of motion are insensitive to these subtleties.

We want to find single-charged-black-p-brane solutions of an analogous nature to the black Schwarzschild p-brane of the previous section. The method we used to construct them in Section 11.3.3 cannot be used in the presence of p-forms: we cannot simply add extra dimensions to a lower-dimensional charged BH metric. For instance, if we took a four-dimensional RN BH and added several extra dimensions, we would obtain a higher-dimensional metric with vanishing scalar curvature (as in four dimensions) but now the trace of the Maxwell energy-momentum tensor would not be zero in more than four dimensions. Another way of expressing the same fact is that, if we dimensionally reduce the above action to $d-p$ dimensions, one does not simply obtain the above action written directly in $d-p$ dimensions, but one finds extra fields. In particular, one finds extra

[10] This action is equivalent to the original one written in [557] (whose main results we reobtain in a different fashion here) which was given in the string frame. Here we do not want to assume that the scalar is the dilaton and thus we prefer to use an Einstein-frame action. The constant a is, then, not the same as in [557] but we obtain simpler, more-symmetric expressions.

scalars that couple to the form. The approach taken originally by Horowitz and Strominger in [557] was to perform this dimensional reduction, simultaneously reducing the $(p+2)$-form field strength to a 2-form. The resulting problem is a BH problem. In fact, we obtain the a-model for BHs! (The two parameters a are related but different.) Then, they used the solutions found in [432, 447] and rewrote them back into d-dimensional language.

Our approach will be to solve the problem directly in d dimensions with the *Ansatz*

$$ds^2 = f[Wdt^2 - d\vec{y}_p^{\,2}] - g^{-1}[W^{-1}d\rho^2 - \rho^2 d\Omega^2_{(\tilde{p}+2)}], \qquad W = 1 + \frac{\omega}{\rho^{\tilde{p}+1}}, \quad (18.63)$$

which can be seen as a dressing of the Schwarzschild p-brane metric Eqs. (11.148) by the functions f and g, which are related to the presence of the dilaton and the $(p+1)$-form. The natural *Ansatz* for $A_{(p+1)}$, if the p-brane is electrically charged and we want $A_{(p+1)}$ to vanish at infinity, is, by analogy with the BH case,

$$A_{t\underline{y}^1\cdots\underline{y}^p} = \alpha\big(H^{-1} - 1\big), \qquad H = 1 + \frac{h}{\rho^{\tilde{p}+1}}. \quad (18.64)$$

The dilaton and the functions f and g are just functions of H, if we do not want to consider primary scalar hair. We are going to assume that the value of the scalar field at infinity vanishes, $\varphi_0 = 0$. We can always generate a non-vanishing value by rescalings of $A_{(p+1)}$ and shifts of φ, but, since these rescalings will be different in different conformal frames, we postpone them until we specify those frames.

On substituting this *Ansatz* into the equations of motion and using Appendix F.2.2, we arrive at the final form for the desired solutions, which is valid for $\tilde{p} \leq 0$:

$$\boxed{\begin{aligned} ds^2 &= \big(e^{-2a\varphi}H^{-2}\big)^{\frac{1}{p+1}}\big[Wdt^2 - d\vec{y}_p^{\,2}\big] \\ &\quad - \big(e^{-2a\varphi}H^{-2}\big)^{-\frac{1}{\tilde{p}+1}}\big[W^{-1}d\rho^2 + \rho^2 d\Omega^2_{(\tilde{p}+2)}\big], \\ e^{-2a\varphi} &= H^{2x}, \quad A_{t\underline{y}^1\cdots\underline{y}^p} = \alpha(H^{-1} - 1), \quad H = 1 + \frac{h}{\rho^{\tilde{p}+1}}, \quad W = 1 + \frac{\omega}{\rho^{\tilde{p}+1}}, \\ \omega &= h\Big[1 - \frac{a^2}{4x}\alpha^2\Big], \qquad x = \frac{(a^2/2)c}{1 + (a^2/2)c}, \qquad c = \frac{(p+1) + (\tilde{p}+1)}{(p+1)(\tilde{p}+1)}. \end{aligned}}$$

(18.65)

For $p = 0$ these solutions reduce to those of the dilaton a-model Eqs. (12.10). As has happened in all the cases we have previously studied, when the extremality parameter $\omega = 0$ so the "Schwarzschild factor" W becomes 1 (and W disappears), H becomes an arbitrary harmonic function in the $(\tilde{p}+3)$-dimensional Euclidean space transverse to the object. These will be solutions describing extreme p-branes. When $h = 0$ so $H = 1$ and disappears, we recover the Schwarzschild p-brane solutions. Finally, this general solution is written in such a way that it is valid for the case $a = 0$ as well. In that case, the scalar decouples from $A_{(p+1)}$ and the only solutions with no primary scalar hair have a completely trivial scalar field.

In general, these solutions have a regular event horizon at $\rho = \omega^{\frac{1}{\tilde{p}+1}}$ when $h \geq 0$ and $\omega < 0$.

The general solution Eq. (18.65) can be generalized further by assuming that there are q additional translational isometries in the directions $z_1, \ldots, z_q$. Using as *Ansatz* the metric and the results of Section F.2.3 we, quite straightforwardly, obtain

$$\boxed{\begin{aligned}
ds^2 &= \left(e^{-2a\varphi} H^{-2}\right)^{\frac{1}{p+1}} \left[W dt^2 - d\vec{y}_p^{\,2}\right] \\
&\quad - \left(e^{-2a\varphi} H^{-2}\right)^{-\frac{1}{\tilde{p}+1}} \left[d\vec{z}_q^{\,2} + W^{-1} d\rho^2 + \rho^2 d\Omega^2_{(\delta-2)}\right], \\
e^{-2a\varphi} &= H^{2x}, \quad A_{\underline{t}\underline{y}^1\cdots\underline{y}^p} = \alpha(H^{-1} - 1), \quad H = 1 + \frac{h}{\rho^{\delta-3}}, \quad W = 1 + \frac{\omega}{\rho^{\delta-3}}, \\
\omega &= h\left[1 - \frac{a^2}{4x}\alpha^2\right], \qquad x = \frac{(a^2/2)c}{1 + (a^2/2)c}, \qquad c = \frac{(p+1) + (\tilde{p}+1)}{(p+1)(\tilde{p}+1)},
\end{aligned}}
\tag{18.66}$$

where $\delta = d - (p+q) > 3$. Observe that q is essentially arbitrary. This solution can be considered as the zero mode of the original solution when the qs of the transverse dimensions are compact. Equivalently, we can say that it is the original solution "smeared" over q dimensions.

Electric–magnetic duality in the p-brane a-model. Except in a few particular cases, the p-brane a-model does not have any electric–magnetic-duality symmetry. Instead, in general, there is an electric–magnetic duality relating *pairs* of these models: the equations of motion (not the actions, as usual) of the (a, p) model and the $(a, \tilde{p})$ models ($\tilde{p} = d - p - 4$) are related by the transformation (see the formulae in Section 1.6)

$$F_{(p+2)} = e^{-2\varphi_{\tilde{p}}} F_{(\tilde{p}+2)}, \qquad \varphi_p = -\varphi_{\tilde{p}}. \tag{18.67}$$

We have a symmetry of the same theory only when $\tilde{p} = p$ (BHs in $d = 4$, strings in $d = 6$, membranes in $d = 8$, 3-branes in $d = 10$, etc.).

At the level of solutions, these transformations allow us to rewrite an electric solution of the (a, p) model as a magnetic solution of the $(a, \tilde{p})$ model and vice-versa. Thus, we do not obtain, strictly speaking, more solutions by this procedure: the magnetic dual of an electric p-brane is the electric $\tilde{p}$-brane and all these solutions are already contained in the general one, Eqs. (18.66). In many cases, though, the theory is written in terms of $(p+2)$-forms directly and it is useful to write these "magnetic" solutions directly in terms of them. On the other hand, since the dilaton is inverted in the magnetic solution, if we use it to reexpress the Einstein metric in a different conformal frame (the string frame, say) the metrics of the electric and magnetic solutions will be different. We straightforwardly find

$$
\boxed{
\begin{aligned}
ds^2 &= \left(e^{+2a\varphi}H^{-2}\right)^{\frac{1}{\tilde{p}+1}}\left[Wdt^2 - d\vec{y}_{\tilde{p}}^{\,2}\right] \\
&\quad - \left(e^{+2a\varphi}H^{-2}\right)^{-\frac{1}{p+1}}\left[d\vec{z}_q^{\,2} + W^{-1}d\rho^2 + \rho^2 d\Omega^2_{(\tilde{\delta}-2)}\right], \\
e^{-2a\varphi} &= H^{-2x}, \qquad F_{(p+2)z_1\cdots z_q\psi_1\cdots\psi_{(\tilde{\delta}-2)}} = (\tilde{\delta}-3)\alpha h \Omega^{(\tilde{\delta}-2)}_{\psi_1\cdots\psi_{(\tilde{\delta}-2)}}, \\
H &= 1 + \frac{h}{\rho^{\tilde{\delta}-3}}, \qquad W = 1 + \frac{\omega}{\rho^{\tilde{\delta}-3}}, \\
\omega &= h\left[1 - \frac{a^2}{4x}\alpha^2\right], \qquad x = \frac{(a^2/2)c}{1+(a^2/2)c}, \qquad c = \frac{(p+1)+(\tilde{p}+1)}{(p+1)(\tilde{p}+1)},
\end{aligned}
}
\tag{18.68}
$$

where $\tilde{\delta} = d - (\tilde{p} + q)$ and $\Omega^{(n)}$ is the volume form of an n-sphere.

Most of the single-p-brane solutions we are going to deal with are included in these general solutions, except for the self-dual ones. It is easy to deal with them using solutions describing two p-branes and therefore we will study them after we study intersecting p-brane solutions in Section 19.6.

18.2.3 Sources for solutions of the p-brane a-model

Our experience tells us that we may find charged-p-brane sources for the extreme, and only for the extreme ($\omega = 0$), charged-p-brane solutions of the a-model that we have just found. On finding these sources, we will be able to relate the integration constants h and α of the solution to the brane tension $T_{(p)}$ and charge parameter μ_p.

We consider the following generic coupled system:

$$S = S_a + S_p, \tag{18.69}$$

where S_a is the bulk a-model action (18.61) and S_p is the charged p-brane action:

$$
\begin{aligned}
S_p[X^\mu, \gamma_{ij}] = &-\frac{T_{(p)}}{2}\int d^{p+1}\xi\sqrt{|\gamma|}\left[e^{-2b\varphi}\gamma^{ij}\partial_i X^\mu \partial_j X^\nu g_{\mu\nu} - (p-1)\right] \\
&+ \frac{(-1)^{p+1}\mu_p}{(p+1)!}\int d^{p+1}\xi\, \epsilon^{i_1\cdots i_{p+1}} A_{(p+1)\,\mu_1\cdots\mu_{p+1}}\partial_{i_1}X^{\mu_1}\cdots\partial_{i_{p+1}}X^{\mu_{p+1}}.
\end{aligned}
\tag{18.70}
$$

The coupling of the scalar to the p-brane is, in principle, arbitrary. However, in all the relevant cases the parameters a and b turn out to be related by

$$a = -(p+1)b, \tag{18.71}$$

and, actually, only then do we have a solution of the coupled system, as we are going to see.

The equations of motion of the spacetime fields are

$$\begin{aligned}
&G_{\mu\nu} + 2T^{\varphi}_{\mu\nu} + \frac{(-1)^{p+1}}{2\cdot(p+1)!} e^{-2a\varphi} T^{A_{(p+1)}}_{\mu\nu} \\
&\qquad - \frac{8\pi G_{\rm N}^{(d)} T_{(p)}}{\sqrt{|g|}} \int d^{p+1}\xi \sqrt{|\gamma|}\, e^{-2b\varphi} \gamma^{ij} \partial_i X^{\mu} \partial_j X^{\nu} g_{\rho\mu} g_{\sigma\nu}\, \delta^{(d)}(x - X(\xi)) = 0, \\
&\nabla^2\varphi + \frac{(-1)^{p+1} a}{4\cdot(p+2)!} e^{-2a\varphi} F^2_{(p+2)} \\
&\qquad - \frac{4\pi G_{\rm N}^{(d)} T_{(p)} b}{\sqrt{|g|}} \int d^{p+1}\xi \sqrt{|\gamma|}\, e^{-2b\varphi}\, \gamma^{ij} \partial_i X^{\mu} \partial_j X^{\nu} g_{\mu\nu}\, \delta^{(d)}(x - X(\xi)) = 0, \\
&\nabla_{\mu}\left(e^{-2a\varphi} F_{(p+2)}{}^{\mu\nu_1\cdots\nu_{p+1}}\right) \\
&\qquad - \frac{16\pi G_{\rm N}^{(d)} \mu_p}{\sqrt{|g|}} \int d^{p+1}\xi\, \epsilon^{i_1\cdots i_{p+1}} \partial_{i_1} X^{\nu_1} \cdots \partial_{i_{p+1}} X^{\nu_{p+1}} \delta^{(d)}(x - X(\xi)) = 0,
\end{aligned} \tag{18.72}$$

and those of the worldvolume fields are

$$\begin{aligned}
&\gamma_{ij} - e^{-2b\varphi} g_{\mu\nu} \partial_i X^{\mu} \partial_j X^{\nu} = 0, \\
&\nabla^2(\gamma) X^{\mu} + \gamma^{ij} \partial_i X^{\rho} \partial_j X^{\sigma} \left(\Gamma_{\rho\sigma}{}^{\mu} - 2b\partial_{(\rho}\varphi g_{\sigma)}{}^{\mu}\right) \\
&\qquad + \frac{(-1)^{p+1}\mu_p/T_{(p)}}{(p+1)!\sqrt{|\gamma|}} e^{2b\varphi} F_{(p+2)}{}^{\mu}{}_{\mu_1\cdots\mu_{p+1}} \partial_{i_1} X^{\mu_1} \cdots \partial_{i_{p+1}} X^{\mu_{p+1}} \epsilon^{i_1\cdots i_{p+1}} = 0.
\end{aligned} \tag{18.73}$$

The first of the worldvolume equations can be used immediately in all the other equations to eliminate the worldvolume metric. Furthermore, using the static gauge for the first $(p+1)$ p-brane embedding coordinates that we denote by Y^i,

$$Y^i(\xi) = \xi^i, \tag{18.74}$$

and the following *Ansatz* for the transverse embedding coordinates;

$$X^m(\xi) = 0, \tag{18.75}$$

which corresponds to a p-brane at rest at $x^m = 0$, it is possible to perform the worldvolume integrals in the equations of motion of the spacetime fields and only $(\tilde{p}+3)$-dimensional Dirac δ functions remain as sources.

Our *Ansatz* for the spacetime fields is given by the extreme ($\omega = 0$) p-brane solutions Eqs. (18.65), H being now a function of the transverse coordinates x^m to be determined.

In the absence of sources (or outside of them) H can be any harmonic function of those $\tilde{p}+3$ transverse coordinates, satisfying

$$\partial_m \partial_m H(\vec{x}_{(\tilde{p}+3)}) = 0. \tag{18.76}$$

In general, H, and therefore the solution, has singularities that can be understood as originated by sources that are not included explicitly in the action. When we include source terms in the action, the singularities of H have to match them. In this case, the sources are

$(\tilde{p}+3)$-dimensional Dirac δ functions placed at the origin in transverse space, and H has to have a single pole there, with the coefficient h necessary to match that of the Dirac δ functions.

We find that all the equations are solved everywhere (including at the Dirac-δ-function singularity) if and only if a and b are related by Eq. (18.71) and the tension $T_{(p)}$ and charge parameter $\mu_{(p)}$ are related by

$$\boxed{\mu_p = (-1)^p T_{(p)}/\alpha.} \tag{18.77}$$

Then, H is given by

$$\boxed{H = \begin{cases} \epsilon + \dfrac{h}{|\vec{x}_{(\tilde{p}+3)}|}, & h = \dfrac{16\pi G_{\mathrm{N}}^{(d)} T_{(p)}}{(\tilde{p}+1)\omega_{(\tilde{p}+2)}\alpha^2}, \qquad \tilde{p} \geq 0, \\ \epsilon + h \ln |\vec{x}_2|, & h = -\dfrac{16\pi G_{\mathrm{N}}^{(d)} T_{(p)}}{2\pi\alpha^2}, \qquad \tilde{p} = -1, \\ \epsilon + h|x|, & h = -\dfrac{16\pi G_{\mathrm{N}}^{(d)} T_{(p)}}{2\alpha^2}, \qquad \tilde{p} = -2, \end{cases}} \tag{18.78}$$

where ϵ is not determined by the equations of motion alone. Asymptotic flatness in transverse space ($\tilde{p} \geq 0$) requires that $\epsilon = +1$. The solutions with $\epsilon = 0$ can sometimes be understood as the $\vec{x}_{\tilde{p}+3} \to 0$ limit of the asymptotically flat ones. Typically, the limit $\vec{x}_{\tilde{p}+3} \to 0$ is a near-horizon limit. We will see some very important examples in Section 19.5.1, but we have already studied the simplest example: the near-horizon limit of the ERN BH which corresponds to the RB solution Eq. (8.90).

Observe that, for $\tilde{p} \geq 0$ ($\tilde{p} < 0$), h is naturally positive (negative) ($T_{(p)}$ being naturally positive). In general, the metrics of solutions with a single p-brane with $\tilde{p} < 0$ will unavoidably have singularities that are unrelated to the p-brane sources because H will become zero or negative at some point. To obtain consistent solutions, one must combine several branes. An example of this kind of construction of a regular solution for branes with $\tilde{p} = -1$, i.e. $(d-3)$-branes, can be found in [474]. For branes with $\tilde{p} = -2$, i.e. $(d-2)$-branes (*domain walls*), a popular way of obtaining a regular metric consists in "cutting" the space before the critical distance $x = 1/|h|$ at which $H = 0$ is reached, for instance by constructing a one-dimensional orbifold with the positive-tension $(d-2)$-brane at one of the fixed points.[11] Consistency requires us to place a brane with opposite tension and charge at the other fixed point so the total charge and tension are zero. For this system

$$H = 1 - |hx|, \qquad x \in [0, \pi R], \quad \pi R < 1/|h|. \tag{18.79}$$

In stringy constructions, the negative-tension brane is an orientifold plane associated with the symmetry $x \to -x$ that gives rise to the orbifold.

[11] This is the basis of the Hořava–Witten scenario, and also of the Randall–Sundrum scenarios [797, 798]. See also [133].

19

The extended objects of string theory

After the general introduction to extended objects of the preceding chapter, in this one we are going to study specifically the extended objects that appear in string theory. The existence of these objects is implied by our previous knowledge of existing objects (strings and Dp-branes) combined with duality. This path will be followed in Section 19.1, in which we will arrive at the diagrams in Figures 19.4.1 and 19.4.1 that represent, respectively, more- and less-conventional extended string/M-theory objects and their duality relations. The duality relations can be used to find the masses of all these objects compactified on tori (Tables 19.1–19.3) using as input the mass of a string wound once on a circle (i.e. the mass of a winding mode). To obtain consistent results (in particular for electric–magnetic dual branes to coexist satisfying the Dirac quantization condition), the ten-dimensional Newton constant has to have a specific value in terms of the string coupling constant and the string length that we will determine.

The next step (Section 19.2) will be to identify which are, among the general solutions of the p-brane a-model, those that represent the long-range fields of the basic extended objects of string and M theory that we found before. We will first identify families of solutions and then we will study one by one the most important solutions. In Section 19.3 we will check the values of the integration constants of those solutions against the masses and charges of the extended objects that we determined using duality arguments. Then, the duality relations between the solutions will be checked in Section 19.4.

Next, in Section 19.5 we will learn how a great deal of information about all these objects is encoded in the spacetime superalgebras of the effective (supergravity) theories. In particular, the superalgebras tell us (up to a point) which extended objects may exist and the amount of unbroken supersymmetry preserved by each of them (always half of the total), as we will check by solving explicitly the Killing-spinor equations (Section 19.5.1).

In Section 19.6 we will study the possible intersections between several of these objects. The worldvolume fields of the extended objects contain a large amount of information about these intersections and we will briefly review the worldvolume theories of the extended objects of string/M theory first. We will construct solutions describing the simplest intersections, which will be used in Chapter 20 to construct four-dimensional BH solutions.

Some general references with emphasis on p-brane solutions of the string/M-theory effective actions are [333, 337, 417, 858, 862, 863, 901, 966]. The standard general references on D-branes are the second volume of Polchinski's book [779] and [64, 604, 780] and Johnson's book [605].

19.1 String-theory extended objects from duality

We have already met some of the extended objects of string theory: (fundamental) strings (F1) and, in type-II and type-I theories, Dp-branes (Dp) with $p = 0, 2, 4, 6, 8$ for the type-IIA theory, $p = 1, 3, 5, 7, 9$ for the type-IIB theory, and $p = 1, 5$ for the type-I theory. Although Dp-branes have masses proportional to the inverse string coupling constant, their existence has been inferred from the perturbative formulation of string theory, which was reviewed in Chapter 14. It is not surprising that T duality, a perturbative string duality, does not require the existence of any new extended objects in the theory: it just relates Dp-branes to D$(p \pm 1)$-branes and fundamental strings to fundamental strings in different states.[1] These relations are represented from the viewpoint of the associated classical solutions in Figure 19.4.1, which also contains many other relations and new objects required by duality.[2]

We have stressed, however, that non-perturbative S dualities require in general the existence of new non-perturbative states dual to the ones present in the perturbative spectrum. $N = 2B, d = 10$ SUEGRA has a global SL$(2, \mathbb{R})$ symmetry and it was proposed in [583] that this symmetry of the effective action reflects an S duality between type-IIB superstring theories, which would be related by the discrete subgroup SL$(2, \mathbb{Z})$, as discussed in Section 17.2. Let us consider systematically what the implications of the existence of this S duality, fundamental strings, and Dp-branes (with p odd) in type-IIB superstring theory are.

Extended objects from type-IIB S duality. Fundamental strings couple to the KR 2-form potential $\hat{\mathcal{B}}_{\hat{\mu}\hat{\nu}}$, which is interchanged with the RR 2-form $\hat{C}^{(2)}_{\hat{\mu}\hat{\nu}}$ to which D1-branes (*D-strings*) couple, by the S-duality transformation $S = \eta$. Thus, the S dual of the fundamental string is the D-string and the two objects form an S-duality doublet, as represented in Figure 19.4.1. This is not new information, but it fits nicely in the conjectured S duality.

Let us now consider Dp-branes beyond the D-string. The D3-brane couples to the 4-form potential with self-dual field strength, which transforms into itself under S duality. Thus, the D3-brane is an S-duality singlet. The D5-brane couples to $\hat{C}^{(6)}$, which is the electric–magnetic dual of $\hat{C}^{(2)}$ (D5-branes are the electric–magnetic duals of D-strings, but not their S duals). Under S duality $\hat{C}^{(6)}$ must transform into $\hat{\mathcal{B}}^{(6)}$, the electric–magnetic dual of $\hat{\mathcal{B}}$. We have to add to the string spectrum the 5-brane (the *solitonic 5-brane*, *NS 5-brane*, or *S5-brane*) that couples to $\hat{\mathcal{B}}^{(6)}$ that we mentioned on page 510. The D5-brane and the S5-brane

[1] To the standard Dp-branes with $p \geq 0$ we can add a IIB D(-1)-brane, the *D-instanton*, with zero worldvolume directions and ten transverse Euclidean directions. It can be obtained by T dualizing the D0-brane in the Euclidean time direction.

[2] Less-conventional objects whose existence is also implied by string dualities are represented in Figure 19.4.1. On the other hand, the T duality between fundamental strings is represented in Figure 19.4.1 as T duality between gravitational-wave solutions and fundamental-string solutions, which we know represent momentum and winding string states (Section 15.3). Waves and KK6 monopoles can be viewed as electric–magnetic duals.

transform as an S-duality doublet. We will determine more properties of the S5-brane using S duality later on.

Next, there is the D7-brane, which couples to $\hat{C}^{(8)}$, the electric–magnetic dual of $\hat{C}^{(0)}$. This pseudoscalar field does not transform linearly under S duality and it is very difficult (if it is possible at all) to find how $\hat{C}^{(8)}$ transforms. Still, the existence of an object related to D7 by $S = \eta$ is required for the S-duality conjecture to be true. Such an object can be called an *S7-brane*. We will see that it is possible to find a classical solution that represents it and that it is related to other solutions by T duality, which makes its presence necessary from this point of view.

Finally, it has been conjectured that the D9-brane is dual to an S9-brane that couples to a 10-form potential $\hat{\mathcal{B}}^{(10)}$. The S9-brane plays an important role in the type-IIB construction of the heterotic SO(32) superstring theory that we have reviewed in Section 17.5.

Extended objects from type-II T duality. We have completed the string spectrum to make it consistent with type-IIB S duality. Now, the enhanced type-IIB spectrum has to be consistent with type-IIA/B T duality. First, what is the T dual of the S5-brane in a transverse direction? It must be an object that couples to the T dual of $\hat{\mathcal{B}}^{(6)}$, that is, to the electric–magnetic dual of the type-IIA metric components that give rise to the KK vector field in $d = 9$. We know only one object of this kind: the KK monopole, Eq. (11.160), which we generated via electric–magnetic duality of the KK vector on page 329. We will check that the classical solutions that represent the S5 and the KK monopole are indeed related by T duality. It is clear that we can add to the type-IIA theory an S5A that will be dual to the KK monopole of the type-IIB theory.

This is a remarkable relation with very important implications.

1. So far, we have seen the KK monopole just as a topologically non-trivial solution (a gravitational instanton). Now we are going to view it as a new kind of extended object, a KKp-brane with non-trivial worldvolume dynamics. The KK monopole of the type-IIA theory has a $(5 + 1)$-dimensional worldvolume with a vector and a scalar in addition to the ten embedding coordinates $\hat{X}^{\hat{\mu}}$ and we will denote it by KK6A. The worldvolume theory of the KK6B contains a self-dual 2-form potential plus two extra scalars [127].

2. The KK-monopole solution Eq. (11.160) is well defined only when the coordinate z (which is the one associated with the S5–KK6 T duality) is compact and has the right periodicity, which is related to its charge and tension. Thus, we cannot take the decompactification limit.[3] KK6-branes do not exist in uncompactified theories.

3. The above two points are consistent with the requirements for a κ-symmetric worldvolume action: the above bosonic-field contents of the KK6A and KK6B theories have one too many local degrees of freedom, but, as explained in Section 18.1.1, they can be eliminated by gauging one isometry, which would be associated with the coordinate z.

[3] In other words; the tension and charge are proportional to the compactification radius and would diverge in the decompactification limit.

The KK6B, being purely gravitational, is an S-duality singlet and does not require the introduction of any other object. On the other hand, T duality in the worldvolume directions of the KK6- and S5-branes takes us into another KK6-brane or S5-brane.

Now, what is the T dual of an S7-brane? The actual transformation of the S7 solution shows that the T dual, in a worldvolume direction, is a solution called the $D6_1$-brane in [666]. It shares with KK-branes the requirement that a transverse direction should be compact and it shares with the D6-brane the fact that it has a non-trivial $\hat{C}^{(7)}$.

We can go on tracing all the objects related to these by S and T dualities. In general, starting with the D7/S7 doublet,[4] one finds KK-branes. Their presence is, nevertheless, required by these dualities in ten and 11 dimensions and by U duality in lower dimensions [161, 578, 579, 738], but they have not been studied thoroughly and they are often considered exotic objects.

Extended objects from type-IIA/M Theory duality. We have shown in Section 16.1 that the fact that the dimensional reduction of 11-dimensional supergravity on a circle gives $N = 2A$, $d = 10$ SUEGRA can be interpreted as a suggestion that the strong-coupling limit of type-IIA superstring theory is 11-dimensional supergravity or a theory that reduces to it at low energies (*M theory*). In that limit, then, all the type-IIA extended objects must be related to some 11-dimensional M-theory object [897].

The identification of the $d = 11$ objects requires the oxidation of the $d = 10$ classical solutions associated with the type-IIA extended objects which will not be performed until Section 19.4, after we find such solutions in Section 19.2. Here we will only give the results, but we can check their internal consistency[5] and their consistency with $d = 10$ dualities.

It has long being known that the type-IIA fundamental string (F1A) can be seen as the 11-dimensional membrane (2-brane) (M2) wrapped on a circle [335]. However, if there was a membrane in $d = 11$, there should have been a membrane in $d = 10$ as well. Now we know that type-IIA superstring theory has a membrane: the D2-brane. We will see that the tensions, charges, and worldvolume effective actions of the fundamental string and the D2-brane can be derived from that of the M2-brane.

How about the other Dp-branes? The D0-brane turns out to correspond to a $d = 11$ graviton KK mode, and its electric–magnetic dual, the D6-brane, is a $d = 11$ KK monopole compactified on the U(1) fiber direction (z in Eq. (11.160)). The D4-brane has to originate on a $d = 11$ 4- or 5-brane, but there are no 3-branes in $d = 10$ and there is a 5-brane, the S5A. Both the D4 and the S5A come from a $d = 11$ 5-brane (M5) associated with the dual 6-form potential.

[4] The reduction to $d = 9$ of the D7 and S7 gives a pair of objects that can be constructed by reducing the KK7M on two directions in different orders. All S-duality doublets of the IIB theory have this property (see the next footnote). The situation is, actually, more complicated, since there is an infinite number of "pq 7-branes" of which the D7 and S7 are just two examples that do not give rise to all the possibilities.

[5] If we assume that one $d = 11$ object exists, all the different ways of reducing that object over a circle must give different $d = 10$ objects. On the other hand, if we compactify the same object down to $d = 9$ over the same directions but in different orders, then, since these compactifications are related by an $SL(2, \mathbb{R})$ transformation in the internal torus, we must obtain $d = 9$ S-duality doublets. Many examples of this fact can be found in Figure 19.4.1.

The D8-brane has with be associated with a $d = 11$ 8- or 9-brane, but none of these is explicitly known. As we explained in Section 16.2, the natural candidate would be the KK9M-brane, but no associated solution of $d = 11$ supergravity corresponding to it is known, and possibly no such solution exists. However, it may exist as a solution of a modified theory [123, 142, 666]. Its inclusion is necessary for the consistency of the diagrams in Figures 19.4.1 and 19.4.1 on pages 552 and 553.

19.1.1 *The masses of string- and M-theory extended objects from duality*

We can immediately use our knowledge of the duality relations between the extended objects of string and M theories to find their masses when they are compactified on tori. All we need to know is the duality transformation rules of compactification radii, the string coupling constant and the string masses, and the mass of one of the extended objects. Let us first recall the duality transformation rules.

In T duality, the relation between the compactification radii $R_{\mathrm{A(B)}}$ and coupling constants $g_{\mathrm{A(B)}}$ of the type-IIA(B) theories (or any other pair of string theories) is given by Eqs. (14.61) and (14.62), which we rewrite here for convenience:

$$\boxed{R_{\mathrm{A,B}} = \ell_{\mathrm{s}}^2/R_{\mathrm{B,A}}, \qquad g_{\mathrm{A,B}} = g_{\mathrm{B,A}}\ell_{\mathrm{s}}/R_{\mathrm{B,A}}.} \tag{19.1}$$

In type-IIB S duality (with $S = \eta$ and vanishing $\hat{C}^{(0)}$), the transformation rules for the string coupling constant and radii can be deduced froms Eqs. (17.21) and (17.24), and take the forms

$$\boxed{g'_{\mathrm{B}} = 1/g_{\mathrm{B}}, \qquad R'_i = R_i/g_{\mathrm{B}}^{\frac{1}{2}}.} \tag{19.2}$$

These rules have to be supplemented by the following transformation rule for the masses, which follows from Eq. (17.24) and the definition of mass, as we will explain in Section 19.3:

$$\boxed{M' = g_{\mathrm{B}}^{\frac{1}{2}} M.} \tag{19.3}$$

In type-IIA/M-theory duality we have to use the relations Eqs. (16.48) (rewritten below) between the string length and the 11-dimensional Planck length and between the string coupling constant and the compactification radius of the 11th coordinate that we call here R_{10} for convenience:

$$\boxed{\ell_{\mathrm{s}} = \ell_{\mathrm{Planck}}^{(11)\,2}/[(2\pi)^2 R_{10}], \qquad g_{\mathrm{A}} = (2\pi)^2 R_{10}^2/\ell_{\mathrm{Planck}}^{(11)\,2}.} \tag{19.4}$$

Masses from $d = 10$ string dualities. We are going to apply these rules to the transformation of the (almost) only object whose mass we know: the (fundamental) string F1 wound once around a compact coordinate ($x^9 \in [0, 2\pi R_9]$, say). Its mass is just the mass of a winding mode with $w = 1$ in the mass formula Eq. (14.60) (which is valid for superstrings if we include left- and right-moving fermionic oscillators),

$$M_{\mathrm{F1}w} = \frac{R_9}{\ell_{\mathrm{s}}^2}, \tag{19.5}$$

which is just the string tension times the volume of the compact space.

As a warm-up exercise, let us first perform a T duality in the x^9 direction, in which we know we should obtain an F1 with minimal momentum in the compact direction:

$$M_{\mathrm{F1}m} = M'_{\mathrm{F1}w} = \frac{R_9}{\ell_{\mathrm{s}}^2} = \frac{1}{R'_9}, \tag{19.6}$$

in agreement with Eq. (14.60) with $n = 1$. In this example, it did not matter whether we were dealing with the IIA or IIB fundamental string. Let us now assume that it is the IIB one F1B. An S-duality transformation should take us to the D-string wound once around x^9. Using Eqs. (19.2) and (19.3), we find

$$M_{\mathrm{D1}} = M'_{\mathrm{F1B}w} = g_{\mathrm{B}}^{\frac{1}{2}} M_{\mathrm{F1B}w} = g_{\mathrm{B}}^{\frac{1}{2}} \frac{R_9}{\ell_{\mathrm{s}}^2} = \frac{R'_9}{g'_{\mathrm{B}} \ell_{\mathrm{s}}^2}. \tag{19.7}$$

This mass should be equal to the D-string tension times the volume of the circle, so

$$T_{\mathrm{D1}} = \frac{M_{\mathrm{D1}}}{2\pi R'_9} = \frac{1}{(2\pi \ell_{\mathrm{s}}) g'_{\mathrm{B}} \ell_{\mathrm{s}}}. \tag{19.8}$$

We can now perform successive T-duality transformations to find the masses and tensions of all the Dp-branes. A T duality in the direction x^9 takes us to the D0-brane, whose mass and tension are

$$T_{\mathrm{D0}} = M_{\mathrm{D0}} = M'_{\mathrm{D1}} = \frac{R_9}{g_{\mathrm{B}} \ell_{\mathrm{s}}^2} = \frac{\ell_{\mathrm{s}}^2 / R'_9}{g'_{\mathrm{A}} \ell_{\mathrm{s}} / (R'_9 \ell_{\mathrm{s}}^2)} = \frac{1}{g'_{\mathrm{A}} \ell_{\mathrm{s}}}. \tag{19.9}$$

If we T-dualize the D-string in a transverse direction (x^8), we obtain instead the D2-brane:

$$M_{\mathrm{D2}} = M'_{\mathrm{D1}} = \frac{R_9}{g_{\mathrm{B}} \ell_{\mathrm{s}}^2} = \frac{R_9}{g'_{\mathrm{A}} \ell_{\mathrm{s}} / (R'_8 \ell_{\mathrm{s}}^2)} = \frac{R_8 R_9}{g'_{\mathrm{A}} \ell_{\mathrm{s}}^3}, \qquad \Rightarrow T_{\mathrm{D2}} = \frac{1}{(2\pi \ell_{\mathrm{s}})^2 g'_{\mathrm{A}} \ell_{\mathrm{s}}}. \tag{19.10}$$

By repeating this procedure, we obtain the mass of the Dp-brane wrapped around a p-torus and its tension (removing the primes):

$$M_{\mathrm{D}p} = \frac{R_{10-p} \cdots R_9}{g \ell_{\mathrm{s}}^{p+1}}, \qquad T_{\mathrm{D}p} = \frac{1}{(2\pi \ell_{\mathrm{s}})^p g \ell_{\mathrm{s}}}. \tag{19.11}$$

The S5B-brane is the S-dual of the D5-brane:

$$\begin{aligned} M_{\mathrm{S5B}} &= g^{\frac{1}{2}} M'_{\mathrm{D5}} = g'^{-\frac{1}{2}} \frac{R'_5 / g'^{\frac{1}{2}} \cdots R'_9 / g'^{\frac{1}{2}}}{g'^{-1} \ell_{\mathrm{s}}^6} = \frac{R_5 \cdots R_9}{g^2 \ell_{\mathrm{s}}^6}, \\ T_{\mathrm{S5B}} &= \frac{1}{(2\pi \ell_{\mathrm{s}})^5 g^2 \ell_{\mathrm{s}}}, \end{aligned} \tag{19.12}$$

which confirms its non-perturbative character ($M \sim g^{-2}$). On S-dualizing the D7- and D9-branes, we obtain the S7- and S9-branes,

$$M_{\mathrm{S7}} = \frac{R_9 \cdots R_3}{g_{\mathrm{B}}^3 \ell_{\mathrm{s}}^8}, \qquad T_{\mathrm{S7}} = \frac{1}{(2\pi \ell_{\mathrm{s}})^7 g_{\mathrm{B}}^3 \ell_{\mathrm{s}}}, \tag{19.13}$$

$$M_{\mathrm{S9}} = \frac{R_9 \cdots R_1}{g_{\mathrm{B}}^4 \ell_{\mathrm{s}}^{10}}, \qquad T_{\mathrm{S9}} = \frac{1}{(2\pi \ell_{\mathrm{s}})^9 g_{\mathrm{B}}^4 \ell_{\mathrm{s}}}, \tag{19.14}$$

which are even more non-perturbative ($M \sim g^{-3}, g^{-4}$).

On T-dualizing the S5B-brane in a transverse direction (x^9), we find the IIA KK monopole (KK6A) with the U(1) fiber in the dual x^9 direction. Its mass is

$$M_{\mathrm{KK6A}} = \frac{R_9^2 R_8 \cdots R_4}{g_{\mathrm{A}}^2 \ell_{\mathrm{s}}^8}, \qquad T_{\mathrm{KK6A}} = \frac{R_9}{(2\pi \ell_{\mathrm{s}})^6 g_{\mathrm{A}}^2 \ell_{\mathrm{s}}^2}. \tag{19.15}$$

This object is non-perturbative ($M \sim g^{-2}$). Furthermore, its tension is proportional to the radius of the U(1) fiber, and diverges in the decompactification limit, as we announced.

Since this is a purely gravitational object, there is an identical object in the IIB theory, KK6B, whose mass is identical with the replacement of g_{A} by g_{B}. Furthermore, by virtue of T duality, the S5A-brane also has the same mass as the S5B with the obvious replacements.

If we dualize the S7-brane in the transverse direction x^2, we find the KK8A,

$$M_{\mathrm{KK8A}} = \frac{R_9 \cdots R_3 R_2^3}{g_{\mathrm{A}}^3 \ell_{\mathrm{s}}^{11}}, \qquad T_{\mathrm{KK8A}} = \frac{R_2^3}{(2\pi \ell_{\mathrm{s}})^8 g_{\mathrm{A}}^3 \ell_{\mathrm{s}}^3}, \tag{19.16}$$

which is highly non-perturbative and tightly wrapped around x^2. Similar results are obtained when we T-dualize the S9-brane in any direction (say x^9): we obtain the the KK9A,

$$M_{\mathrm{KK9A}} = \frac{R_9^3 R_8 \cdots R_2}{g_{\mathrm{A}}^4 \ell_{\mathrm{s}}^{12}}, \qquad T_{\mathrm{KK9A}} = \frac{R_9^3}{(2\pi \ell_{\mathrm{s}})^9 g_{\mathrm{A}}^4 \ell_{\mathrm{s}}^3}. \tag{19.17}$$

All these results are collected in Tables 19.1 and 19.2.

Masses from type-IIA/M-theory duality. We said that the F1A is the M2 wrapped in the 11th dimension. Thus, the mass of the M2 wrapped on a torus must equal that of the F1A wrapped on a circle:

$$M_{\mathrm{M2}} = M_{\mathrm{F1A}} = \frac{R_9}{\ell_{\mathrm{s}}^2} = \frac{R_9 R_{10}}{(\ell_{\mathrm{Planck}}^{(11)})^3}, \quad \Rightarrow T_{\mathrm{M2}} = \frac{1}{(2\pi \ell_{\mathrm{Planck}}^{(11)})^2 \ell_{\mathrm{Planck}}^{(11)}}. \tag{19.18}$$

On the other hand, the mass of the M2 wrapped in two directions different from the one that we consider the 11th (say x^8 and x^9) must coincide with that of the D2. Indeed,

$$M_{\mathrm{M2}} = \frac{R_8 R_9}{(\ell_{\mathrm{Planck}}^{(11)})^3} = \frac{R_8 R_9}{g_{\mathrm{A}} \ell_{\mathrm{s}}^3} = M_{\mathrm{D2}}. \tag{19.19}$$

Table 19.1. In this table the masses of the various extended objects of type-IIA superstring theory are given in ten-dimensional language (the compactification radii R_i, the string coupling constant g_{A}, and the string length ℓ_{s}) and in 11-dimensional (M-theory) language (the compactification radii R_i and the reduced 11-dimensional Planck length $\tilde{\ell}^{(11)}_{\mathrm{Planck}} = \ell^{(11)}_{\mathrm{Planck}}/(2\pi)$). The coordinate which is compactified to go from the 11- to the ten-dimensional theory is assumed to be x^{10}, so the "11th-dimensional radius" is here denoted by $R_{10} = g_{\mathrm{A}}\ell_{\mathrm{s}} = g_{\mathrm{A}}^{2/3}\tilde{\ell}^{(11)}_{\mathrm{Planck}}$. Furthermore, the configurations of the various 11-dimensional objects that give rise to the ten-dimensional ones are also provided in a notation whose meaning is the following: the array corresponds to the 11 coordinates, starting from $\hat{\hat{x}}^0$ up to $\hat{\hat{x}}^{10}$. A plus means that one of the worldvolume directions occupies that spacetime direction. A star means that the object has a special isometry in the corresponding direction. The corresponding direction cannot be decompactified.

Type-IIA object	Mass in $d = 10$ constants	Mass in $d = 11$ constants	11-Dimensional object
F1m	R_9^{-1}		
D0	$g_{\mathrm{A}}^{-1}\ell_{\mathrm{s}}^{-1}$	R_{10}^{-1}	$\mathrm{WM}(+, -^{10})$
F1w	$R_9\ell_{\mathrm{s}}^{-2}$	$R_{10}R_9(\tilde{\ell}^{(11)}_{\mathrm{Planck}})^{-3}$	$\mathrm{M2}(+, -^8, +^2)$
D2	$R_9R_8g_{\mathrm{A}}^{-1}\ell_{\mathrm{s}}^{-3}$	$R_9R_8(\tilde{\ell}^{(11)}_{\mathrm{Planck}})^{-3}$	$\mathrm{M2}(+, -^7, +^2, -)$
D4	$R_9\cdots R_6g_{\mathrm{A}}^{-1}\ell_{\mathrm{s}}^{-5}$	$R_{10}R_9\cdots R_5(\tilde{\ell}^{(11)}_{\mathrm{Planck}})^{-6}$	$\mathrm{M5}(+, -^5, +^5)$
S5A	$R_9\cdots R_5g_{\mathrm{A}}^{-2}\ell_{\mathrm{s}}^{-6}$	$R_9\cdots R_5(\tilde{\ell}^{(11)}_{\mathrm{Planck}})^{-6}$	$\mathrm{M5}(+, -^4, +^5, -)$
D6	$R_9\cdots R_4g_{\mathrm{A}}^{-1}\ell_{\mathrm{s}}^{-7}$	$R_{10}^2R_9\cdots R_4(\tilde{\ell}^{(11)}_{\mathrm{Planck}})^{-9}$	$\mathrm{KK7M}(+, -^3, +^6, -^\star)$
KK6A	$R_9^2R_8\cdots R_4g_{\mathrm{A}}^{-2}\ell_{\mathrm{s}}^{-8}$	$R_{10}R_9^2\cdots R_4(\tilde{\ell}^{(11)}_{\mathrm{Planck}})^{-9}$	$\mathrm{KK7M}(+, -^3, +^5, +^\star, +)$
D8	$R_9\cdots R_2g_{\mathrm{A}}^{-1}\ell_{\mathrm{s}}^{-9}$	$R_{10}^3R_9\cdots R_4(\tilde{\ell}^{(11)}_{\mathrm{Planck}})^{-12}$	$\mathrm{KK9M}(+, -, +^8, +^\star)$
KK8A	$R_9^3R_8\cdots R_2g_{\mathrm{A}}^{-3}\ell_{\mathrm{s}}^{-11}$	$R_{10}R_9^3R_8\cdots R_2(\tilde{\ell}^{(11)}_{\mathrm{Planck}})^{-12}$	$\mathrm{KK9M}(+, -, +^7, +^\star, +)$
KK9A	$R_9^3R_8\cdots R_1g_{\mathrm{A}}^{-4}\ell_{\mathrm{s}}^{-12}$	$R_{10}R_9^3R_8\cdots R_1(\tilde{\ell}^{(11)}_{\mathrm{Planck}})^{-12}$	$\mathrm{KK9M}(+, +^8, +^\star, -)$

Table 19.2. In this table the masses of the various extended objects of type-IIB superstring theory are given in terms of the compactification radii R_i, the string coupling constant g_{B}, and the string length ℓ_{s}. When a radius appears with a power different from 1, it means that that is a special isometric direction of the object (a KK object).

Type-IIB object	Mass	Type-IIB object	Mass
F1m	R_9^{-1}	KK6A	$R_9^2R_8\cdots R_4g_{\mathrm{B}}^{-2}\ell_{\mathrm{s}}^{-8}$
F1w	$R_9\ell_{\mathrm{s}}^{-2}$	D7	$R_9\cdots R_3g_{\mathrm{B}}^{-1}\ell_{\mathrm{s}}^{-8}$
D1	$R_9g_{\mathrm{B}}^{-1}\ell_{\mathrm{s}}^{-2}$	S7	$R_9\cdots R_3g_{\mathrm{B}}^{-3}\ell_{\mathrm{s}}^{-8}$
D3	$R_9\cdots R_7g_{\mathrm{B}}^{-1}\ell_{\mathrm{s}}^{-4}$	D9	$R_9\cdots R_1g_{\mathrm{B}}^{-1}\ell_{\mathrm{s}}^{-10}$
D5	$R_9\cdots R_5g_{\mathrm{B}}^{-1}\ell_{\mathrm{s}}^{-6}$	S9	$R_9\cdots R_1g_{\mathrm{B}}^{-4}\ell_{\mathrm{s}}^{-10}$
S5B	$R_9\cdots R_5g_{\mathrm{B}}^{-2}\ell_{\mathrm{s}}^{-6}$		

Table 19.3. In this table the masses of the extended objects of M theory are given in terms of the the compactification radii R_i and the reduced 11-dimensional Planck length $\ell\!\!\!{}^{-}{}^{(11)}_{\text{Planck}} = \ell^{(11)}_{\text{Planck}}/(2\pi)$. When a radius appears with a power different from 1, it means that that is a special isometric direction of the object (a KK object).

M object	Mass
WM	0
M2	$R_{10}R_9(\ell\!\!\!{}^{-}{}^{(11)}_{\text{Planck}})^{-3}$
M5	$R_{10}\cdots R_6(\ell\!\!\!{}^{-}{}^{(11)}_{\text{Planck}})^{-6}$
KK7M	$R_{10}^2 R_9\cdots R_4(\ell\!\!\!{}^{-}{}^{(11)}_{\text{Planck}})^{-9}$
KK9M	$R_{10}^3 R_9\cdots R_4(\ell\!\!\!{}^{-}{}^{(11)}_{\text{Planck}})^{-12}$

The D4 is an M5 wrapped in the 11th dimension,

$$M_{\text{M5}} = M_{\text{D4}} = \frac{R_6\cdots R_9}{g_{\text{A}}\ell_{\text{s}}^5} = \frac{R_6\cdots R_{10}}{(\ell\!\!\!{}^{-}{}^{(11)}_{\text{Planck}})^6}, \qquad \Rightarrow T_{\text{M5}} = \frac{1}{(2\pi\ell\!\!\!{}^{-}{}^{(11)}_{\text{Planck}})^5\ell\!\!\!{}^{-}{}^{(11)}_{\text{Planck}}}, \tag{19.20}$$

and the S5A is an M5 that is not wrapped there,

$$M_{\text{M5}} = \frac{R_5\cdots R_9}{(\ell\!\!\!{}^{-}{}^{(11)}_{\text{Planck}})^6} = \frac{R_5\cdots R_9}{g_{\text{A}}^2\ell_{\text{s}}^6} = M_{\text{S5A}}. \tag{19.21}$$

To complete this section, we can see that the D0 is nothing but a KK mode moving in the 11th direction,

$$M_{\text{D0}} = \frac{1}{g_{\text{A}}\ell_{\text{s}}} = \frac{1}{R_{10}}. \tag{19.22}$$

These results are collected in Tables 19.1 and 19.3.

The Newton constant. If we want all the extended objects just found to be quantum-mechanically compatible, we know that the charges of those pairs of extended objects which are electric–magnetic duals must satisfy the Dirac quantization condition. We do not know the charges of all of these objects, except in the case of the F1, since we know the coefficient of the WZ term in the string σ-model action Eq. (15.31); namely the string tension T. This coefficient coincides with the F1 charge, canonically normalized taking into account the normalization of the spacetime effective action Eq. (15.1). Then

$$q_{\text{F1}} = T = \frac{1}{(2\pi\ell_{\text{s}})\ell_{\text{s}}}. \tag{19.23}$$

This identity between the canonically normalized charge and the coefficient of the WZ term in the extended objects' effective action is completely general. Furthermore, those coefficients are always identical to the coefficients of the kinetic (NG) terms given on page 511 for various kinds of objects. Let us, then, consider the electric–magnetic dual of the F1, the S5, whose tension is given in Eq. (19.12). Then

$$q_{\mathrm{S5}} = T_{\mathrm{S5}} g^2 = \frac{1}{(2\pi \ell_{\mathrm{s}})^5 \ell_{\mathrm{s}}}. \tag{19.24}$$

With the normalization Eq. (15.1), the Dirac quantization condition for $n = 1$ reads

$$q_{\mathrm{F1}} q_{\mathrm{S5}} = 2\pi \frac{g^2}{16\pi G_{\mathrm{N}}^{(10)}}, \tag{19.25}$$

which is possible only if

$$\boxed{G_{\mathrm{N}}^{(10)} = 8\pi^6 g^2 \ell_{\mathrm{s}}^8.} \tag{19.26}$$

We obtained a similar result in Chapter 11 in the context of $\hat{d} = 5$ KK theory.

There are more pairs of electric–magnetic duals: Dp- and D$\tilde{p}$-branes. According to the above observations, the Dp-brane charge is

$$q_{\mathrm{D}p} = T_{\mathrm{D}p} g = \frac{1}{(2\pi \ell_{\mathrm{s}})^p \ell_{\mathrm{s}}}, \tag{19.27}$$

and we find again ($d = 10$)

$$q_{\mathrm{D}p} q_{\mathrm{D}\tilde{p}} = \frac{1}{(2\pi \ell_{\mathrm{s}})^6 \ell_{\mathrm{s}}^2} = q_{\mathrm{F1}} q_{\mathrm{S5}}. \tag{19.28}$$

For the two M-branes, we have

$$q_{\mathrm{M}p} = \frac{1}{(2\pi \ell_{\mathrm{Planck}}^{(11)})^p \ell_{\mathrm{Planck}}^{(11)}}, \tag{19.29}$$

so, on account of the definition of $\ell_{\mathrm{Planck}}^{(11)}$, Eq. (16.43),

$$q_{\mathrm{M}p} q_{\mathrm{M}\tilde{p}} = \frac{1}{(2\pi \ell_{\mathrm{Planck}}^{(11)})^7 (\ell_{\mathrm{Planck}}^{(11)})^2} = 2\pi \frac{2\pi}{(\ell_{\mathrm{Planck}}^{(11)})^9} = 2\pi \frac{1}{16\pi G_{\mathrm{N}}^{(11)}}, \tag{19.30}$$

which is the correct form of Dirac's quantization condition with the standard normalization of $d = 11$ supergravity. This is the reason behind the unusual definition of $\ell_{\mathrm{Planck}}^{(11)}$.

19.2 String-theory extended objects from effective-theory solutions

The results discussed in the previous section and depicted in Figures 19.4.1 and 19.4.1 are supported by (and, in many cases, based on) the study of explicit classical solutions of the string effective action that can be interpreted as the long-range fields associated with the extended objects of string theory. That interpretation is based on a comparison between the charges of the sources and those of the solutions. p-branes can be charged with respect to $(p+1)$-form potentials and, therefore, it is natural to start looking for p-brane solutions that are charged with respect to each of the $(p+1)$-form potentials of the string effective theory. The existence of these solutions is a clear argument in support of the existence of the associated string-theory extended object.

The search for these solutions can be systematized using the p-brane a-model solutions as follows: for solutions that involve only one brane, the Chern–Simons terms can be ignored and the relevant part of the string effective action is just

$$S = \frac{g^2}{16\pi G_{\rm N}^{(d)}} \int d^d x \sqrt{|g|}\, \left\{ e^{-2\phi}\left[R - 4(\partial\phi)^2 + \frac{(-1)^{p_1+1}}{2\cdot(p_1+2)!}(H^{(p_1+2)})^2 \right] + \frac{(-1)^{p_2+1}}{2\cdot(p_2+2)!}(G^{(p_2+2)})^2 \right\}, \tag{19.31}$$

where

$$H^{(p_1+2)} = dB^{(p_1+1)}, \qquad G^{(p_2+2)} = dC^{(p_1+1)}. \tag{19.32}$$

By definition, since it is written in the string frame, it describes the potentials that couple to fundamental p_1-branes and Dp_2-branes. We have to rewrite it in the *modified Einstein-frame metric*[6] $\tilde{g}_{\mathrm{E}\,\mu\nu}$ in which the p-brane a-model is given.[7] Using the relation

$$g_{\mu\nu} = e^{\frac{4}{d-2}(\phi-\phi_0)}\, \tilde{g}_{\mathrm{E}\,\mu\nu}, \tag{19.34}$$

we obtain (ignoring tildes) (see Appendix E)

$$S = \frac{1}{16\pi G_{\rm N}^{(d)}} \int d^d x \sqrt{|g|}\Big[R + \frac{4}{d-2}(\partial\phi)^2 + \frac{(-1)^{p_1+1}}{2\cdot(p_1+2)!} e^{-4\frac{p_1+1}{d-2}(\phi-\phi_0)}(H^{(p_1+2)})^2 + \frac{(-1)^{p_2+1}}{2\cdot(p_2+2)!} e^{2\frac{\tilde{p}_2-p_1}{d-2}(\phi-\phi_0)} g^2 (G^{(p_2+2)})^2 \Big]. \tag{19.35}$$

On comparing this action with the a-model action Eq. (18.61), we find that the relations between the string dilaton ϕ and the a-model scalar φ, and between the string potentials $B^{(p_1+1)}$ and $C^{(p_2+1)}$ and the a-model potentials $A_{(p_{1,2}+1)}$, are

$$\phi - \phi_0 = \sqrt{\frac{d-2}{2}}\varphi, \qquad B^{(p_1+1)} = A_{(p_1+1)}, \qquad C^{(p_2+1)} = g^{-1} A_{(p_2+1)}, \tag{19.36}$$

and, furthermore,

$$a_1 = \frac{2(p_1+1)}{\sqrt{2(d-2)}}, \qquad a_2 = \frac{-(\tilde{p}_2 - p_2)}{\sqrt{2(d-2)}}. \tag{19.37}$$

[6] We always take the string-frame metric $g_{\mu\nu}$ to be asymptotically flat (at least in the non-compact directions). On rescaling that metric by a power of the dilaton,

$$g_{\mu\nu} = e^{\frac{4}{d-2}\phi} g_{\mathrm{E}\,\mu\nu}, \tag{19.33}$$

we obtain the *Einstein-frame metric* $g_{\mathrm{E}\,\mu\nu}$ in which the Einstein–Hilbert term has no dilaton factors. However, this metric is not asymptotically flat and a constant rescaling by the dilaton VEV is necessary. The result is the asymptotically flat *modified Einstein-frame metric* [677] $\tilde{g}_{\mathrm{E}\,\mu\nu}$ given in Eq. (19.34).

[7] Recall that we chose $\varphi_0 = 0$.

We could have Poincaré-dualized the field strengths. In that case, we would have obtained for the duals of fundamental p_1-branes ($p_3 = \tilde{p}_1$-branes) an a-model with

$$a_3 = -\frac{2(p_3+1)}{\sqrt{2(d-2)}}, \qquad p_3 = \tilde{p}_1, \qquad \tilde{B}^{(p_3+1)} = g^{-2} A_{(p_3+1)}, \tag{19.38}$$

whereas for the dual Dp_2-branes we would have the same a_2.

Finally, to find solutions of $d = 11$ supergravity that describe M-theory branes (Mp-branes), we just have to set $a = 0$, since there is no scalar in that theory.

The four families of solutions, which are valid for $\delta > 3$, are given by the following.

(Black) M-theory p-Branes (Mp).

$$\boxed{\begin{aligned} &d\tilde{s}_{\mathrm{E}}^2 = H_{\mathrm{M}p}^{-\frac{2}{p+1}}\left[W dt^2 - d\vec{y}_p^{\,2}\right] - H_{\mathrm{M}p}^{\frac{2}{\tilde{p}+1}}\left[d\vec{z}_q^{\,2} + W^{-1}d\rho^2 + \rho^2 d\Omega_{(\delta-2)}^2\right],\\ &C^{(p+1)}{}_{\underline{t}\underline{y}^1\cdots\underline{y}^p} = \alpha\left(H_{\mathrm{M}p}^{-1} - 1\right),\\ &H_{\mathrm{M}p} = 1 + \frac{h_{\mathrm{M}p}}{\rho^{\delta-3}}, \qquad W = 1 + \frac{\omega}{\rho^{\delta-3}}, \qquad \omega = h_{\mathrm{M}p}\left[1 - \frac{1}{2c}\alpha^2\right], \end{aligned}} \tag{19.39}$$

(Black) fundamental p-branes (Fp).

$$\boxed{\begin{aligned} &d\tilde{s}_{\mathrm{E}}^2 = H_{\mathrm{F}p}^{-\frac{2}{d-2}\frac{\tilde{p}+1}{p+1}}\left[W dt^2 - d\vec{y}_p^{\,2}\right] - H_{\mathrm{F}p}^{\frac{2}{d-2}}\left[d\vec{z}_q^{\,2} + W^{-1}d\rho^2 + \rho^2 d\Omega_{(\delta-2)}^2\right],\\ &ds_{\mathrm{s}}^2 = H_{\mathrm{F}p}^{-\frac{2}{p+1}}\left[W dt^2 - d\vec{y}_p^{\,2}\right] - \left[d\vec{z}_q^{\,2} + W^{-1}d\rho^2 + \rho^2 d\Omega_{(\delta-2)}^2\right],\\ &e^{-2\phi} = e^{-2\phi_0} H_{\mathrm{F}p}, \qquad B^{(p+1)}{}_{\underline{t}\underline{y}^1\cdots\underline{y}^p} = \alpha\left(H_{\mathrm{F}p}^{-1} - 1\right),\\ &H_{\mathrm{F}p} = 1 + \frac{h_{\mathrm{F}p}}{\rho^{\delta-3}}, \qquad W = 1 + \frac{\omega}{\rho^{\delta-3}}, \qquad \omega = h_{\mathrm{F}p}\left[1 - \frac{p+1}{2}\alpha^2\right], \end{aligned}} \tag{19.40}$$

(Black) solitonic p-branes (Sp).

$$\boxed{\begin{aligned} &d\tilde{s}_{\mathrm{E}}^2 = H_{\mathrm{S}p}^{-\frac{2}{d-2}}\left[W dt^2 - d\vec{y}_p^{\,2}\right] - H_{\mathrm{S}p}^{\frac{2}{d-2}\frac{p+1}{\tilde{p}+1}}\left[d\vec{z}_q^{\,2} + W^{-1}d\rho^2 + \rho^2 d\Omega_{(\delta-2)}^2\right],\\ &ds_{\mathrm{s}}^2 = W dt^2 - d\vec{y}_p^{\,2} - H_{\mathrm{S}p}^{\frac{2}{\tilde{p}+1}}\left[d\vec{z}_q^{\,2} + W^{-1}d\rho^2 + \rho^2 d\Omega_{(\delta-2)}^2\right],\\ &e^{-2\phi} = e^{-2\phi_0} H_{\mathrm{S}p}^{-1}, \qquad \tilde{B}^{(p+1)}{}_{\underline{t}\underline{y}^1\cdots\underline{y}^p} = \alpha e^{-2\phi_0}\left(H_{\mathrm{S}p}^{-1} - 1\right),\\ &H_{\mathrm{S}p} = 1 + \frac{h_{\mathrm{S}p}}{\rho^{\delta-3}}, \qquad W = 1 + \frac{\omega}{\rho^{\delta-3}}, \qquad \omega = h_{\mathrm{S}p}\left[1 - \frac{\tilde{p}+1}{2}\alpha^2\right]. \end{aligned}} \tag{19.41}$$

(Black) Dp-branes.

$$
\boxed{
\begin{aligned}
&d\tilde{s}_{\mathrm{E}}^2 = H_{\mathrm{D}p}^{-8\frac{\tilde{p}+1}{(d-2)^2}}\left[W dt^2 - d\vec{y}_p^{\,2}\right] - H_{\mathrm{D}p}^{8\frac{p+1}{(d-2)^2}}\left[d\vec{z}_q^{\,2} + W^{-1}d\rho^2 + \rho^2 d\Omega_{(\delta-2)}^2\right],\\
&ds_{\mathrm{s}}^2 = H_{\mathrm{D}p}^{-\frac{4}{d-2}}\left[W dt^2 - d\vec{y}_p^{\,2}\right] - H_{\mathrm{D}p}^{\frac{4}{d-2}}\left[d\vec{z}_q^{\,2} + W^{-1}d\rho^2 + \rho^2 d\Omega_{(\delta-2)}^2\right],\\
&e^{-2\phi} = e^{-2\phi_0} H_{\mathrm{D}p}^{-2\frac{\tilde{p}-p}{d-2}}, \qquad C^{(p+1)}{}_{\underline{t}\underline{y}^1\cdots\underline{y}^p} = \alpha e^{-\phi_0}\left(H_{\mathrm{D}p}^{-1} - 1\right),\\
&H_{\mathrm{D}p} = 1 + \frac{h_{\mathrm{D}p}}{\rho^{\delta-3}}, \qquad W = 1 + \frac{\omega}{\rho^{\delta-3}}, \qquad \omega = h_{\mathrm{D}p}\left[1 - \frac{d-2}{8}\alpha^2\right].
\end{aligned}
}
\tag{19.42}
$$

19.2.1 Extreme p-brane solutions of string and M-theories and sources

The four families of solutions above contain subfamilies of extreme solutions with $\omega = 0$ in which the H functions can be arbitrary functions of the transverse coordinates $\vec{x}_{(\delta-1)}$ ($\rho = |\vec{x}_{(\delta-1)}|$) for $\delta \geq 2$. These are isotropic coordinates in which the metric of the transverse space is conformally flat. As in the ERN BH case, in some cases they do not cover the whole spacetime which can be analytically extended beyond $\rho = 0$, which is only a coordinate singularity.

In Section 18.2.3 we saw that some of the extreme solutions of the p-brane a-model (with $q = 0$ and a single-pole H) could be matched against some charged-p-brane sources (obeying Eqs. (18.71) and (18.77)), which allowed us to determine h, the coefficient of the pole of H, in terms of the tension and charge of the source and in terms of the Newton constant.

It turns out that the four families of objects that we are considering always satisfy Eqs. (18.71) and (18.77) and we can use those results to determine h in the extreme solutions with $q = 0$ and a single pole in terms of the tensions and Newton constants. Then, for the ten- and 11-dimensional objects whose tensions we found in Section 19.1.1, we can determine h as a function of ℓ_{s}, g, and $\ell_{\mathrm{Planck}}^{(11)}$, using the values of $G_{\mathrm{N}}^{(10)}$ and $G_{\mathrm{N}}^{(11)}$ that we determined there. For these $d = 10, 11$ objects $\alpha = \pm 1$ is just the relative sign between the charge parameter μ_p and the tension $T_{(p)}$ which we consider positive.

All we have to do to find h for the families of extreme Mp-branes, Dp-branes etc. is to substitute into Eqs. (18.78) the values of $T_{(p)}$ (which are unknown in general, except in the cases studied in the previous sections) and α, determined by setting $\omega = 0$ in the solutions;

$$
\alpha_{\mathrm{M}p}^2 = \frac{2(d-2)}{(p+1)(\tilde{p}+1)}, \quad \alpha_{\mathrm{F}p} = \frac{2}{p+1}, \quad \alpha_{\mathrm{S}p} = \frac{2}{\tilde{p}+1}, \quad \alpha_{\mathrm{D}p} = \frac{8}{d-2}. \tag{19.43}
$$

Observe that, indeed, for the string/M-theory branes (M2 and M5 in $d = 11$ and F1, S5, and Dp in $d = 10$) $\alpha^2 = +1$.

Writing the four families of extreme solutions with the right values for h is straightforward, but not very interesting, except in the case of the $d = 11, 10$ string/M-theory branes, that we are going to write and study next.

19.2.2 The M2 solution

On substituting the values $d = 11$ and $p = 2$ into the general black-Mp-brane family of solutions Eqs. (19.39), we immediately find the black M2 solution [494],

$$\boxed{\begin{aligned} d\hat{\hat{s}}^2 &= H_{\rm M2}^{-\frac{2}{3}}\big[W dt^2 - d\vec{y}_2^{\,2}\big] - H_{\rm M2}^{\frac{1}{3}}\big[W^{-1}d\rho^2 + \rho^2 d\Omega_{(7)}^2\big],\\ \hat{\hat{C}}_{t\underline{y}^1\underline{y}^2} &= \alpha\big(H_{\rm M2}^{-1} - 1\big),\\ H_{\rm M2} &= 1 + \frac{h_{\rm M2}}{\rho^6}, \qquad W = 1 + \frac{\omega}{\rho^6}, \qquad \omega = h_{\rm M2}\big[1-\alpha^2\big], \end{aligned}} \tag{19.44}$$

and, in the extreme limit $\omega = 0$, $\alpha = \pm 1$, we obtain [344]

$$\boxed{\begin{aligned} d\hat{\hat{s}}^2 &= H_{\rm M2}^{-\frac{2}{3}}\big[dt^2 - d\vec{y}_2^{\,2}\big] - H_{\rm M2}^{\frac{1}{3}}\, d\vec{x}_8^{\,2},\\ \hat{\hat{C}}_{t\underline{y}^1\underline{y}^2} &= \pm\big(H_{\rm M2}^{-1} - 1\big), \qquad H_{\rm M2} = 1 + \frac{h_{\rm M2}}{|\vec{x}_8|^6}, \end{aligned}} \tag{19.45}$$

which is the solution commonly called an M2-brane in the literature written in isotropic coordinates. The integration constant $h_{\rm M2}$ can be determined by the first of Eqs. (18.78), the value of the M2 tension by Eq. (19.18) and the value of $G_{\rm N}^{(11)}$ by Eq. (16.43),

$$\boxed{h_{\rm M2} = \frac{(\ell_{\rm Planck}^{(11)})^6}{6\omega_{(7)}}.} \tag{19.46}$$

In Section 19.3 we will check by classical field-theory methods that this value of this integration constant and the values of other integration constants really correspond to the tension and charge of the M2-brane and the other objects.

It is clear that, if we want the solution to describe $N_{\rm M2}$ parallel M2-branes, we just have to replace $H_{\rm M2}$ by another harmonic function with as many poles as branes, each of them with the coefficient $h_{\rm M2}$. When $N_{\rm M2}$ M2s coincide, there is a single pole with coefficient $N_{\rm M2}h_{\rm M2}$. Similar observations are valid for all the extreme solutions that follow.

The black M2-brane has a regular non-degenerate event horizon at $\rho = (-\omega)^{\frac{1}{6}}$ whose constant-time sections have the topology of S^7 and a singular (would-be) inner horizon covered by the event horizon.

In the extreme limit the event horizon does not become singular but becomes degenerate (as in the ERN BH case) and the singularity covered by it becomes timelike. The Penrose diagram of the (extreme) M2-brane is similar to that of the ERN BH but with only one asymptotically flat region [334, 452] and is represented in Figure 19.1 (a more detailed diagram can be found in [863]).

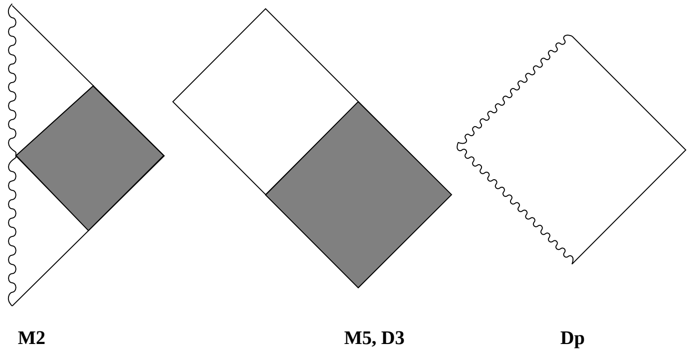

Fig. 19.1. Penrose diagrams of different (extreme) string/M-brane solutions: the M2-brane which has a timelike singularity covered by a horizon which does not allow it to be seen from the asymptotic region covered by the isotropic coordinates (shaded), the M5- and D3-brane that are regular everywhere and have two asymptotic regions separated by the horizon, and the Dp-branes with $p \neq 3$ which have singular horizons. In all cases the angular coordinates of the transverse spheres and the spacelike worldvolume coordinates have been ignored.

In fact, on taking the near-horizon limit $\rho = |\vec{x}_8| \to 0$ (which consists in the deletion of the constant 1 in H_{M2}) in spherical coordinates, we obtain a solution whose metric is the direct product of those of AdS_4 and S^7 with radii R_4 and $2R_4$ (after a rescaling of the worldvolume coordinates),

$$\boxed{d\hat{\hat{s}}^2 = R_4^2\, d\Pi^2_{(4)} - (2R_4)^2\, d\Omega^2_{(7)}, \qquad \hat{\hat{C}}_{t\underline{y}^1\underline{y}^2} = \left(\frac{r}{R_4}\right)^3, \qquad R_4 = \frac{h_{\mathrm{M2}}^{\frac{1}{2}}}{2},} \tag{19.47}$$

where we are using the following form of the metric of the AdS_n space with radius R_n:

$$R_n^2\, d\Pi^2_{(n)} \equiv \left(\frac{r}{R_n}\right)^2 \left(dt^2 - d\vec{y}^{\,2}_{n-2}\right) - \left(\frac{R_n}{r}\right)^2 dr^2. \tag{19.48}$$

The dual 7-form field strength is given by the S^7 volume form $\tilde{\tilde{\hat{\hat{G}}}}0 = 6(2R_4)^6 \omega_{(7)}$.

All the extreme string/M-theory solutions that we are going to study preserve half of the supersymmetries, but this near-horizon limit preserves all the supersymmetry and can be considered a vacuum of M theory. The M2-brane can then be seen as a soliton interpolating between two maximally supersymmetric vacua, Minkowski at infinity and $\mathrm{AdS}_4 \times \mathrm{S}^7$ at the horizon [452]. Since the S^7 is a compact space, this vacuum can be seen to induce spontaneous compactification to $d = 4$ [406]. The compactification of $\mathrm{D} = 11$ supergravity on S^7 gives rise to a gauged $N = 8$, $d = 4$ SUEGRA with gauge group SO(8) (the isometry group of the compact space) and an AdS_4 vacuum [343] (see also [342] and references therein).

19.2.3 The M5 solution

On substituting the values $d = 11$ and $p = 5$ into the general black-Mp-brane family of solutions Eqs. (19.39), we obtain the black-M5-brane solution [494]

$$\boxed{\begin{aligned} d\hat{\hat{s}}^2 &= H_{\mathrm{M5}}^{-\frac{1}{3}}\left[W dt^2 - d\vec{y}_5^{\,2}\right] - H_{\mathrm{M5}}^{\frac{2}{3}}\left[W^{-1}d\rho^2 + \rho^2 d\Omega_{(4)}^2\right], \\ \tilde{\hat{\hat{C}}}_{t\underline{y}^1\cdots\underline{y}^5} &= \alpha\left(H_{\mathrm{M5}s}^{-1} - 1\right), \\ H_{\mathrm{M5}} &= 1 + \frac{h_{\mathrm{M5}}}{\rho^3}, \qquad W = 1 + \frac{\omega}{\rho^3}, \qquad \omega = h_{\mathrm{M5}}[1-\alpha^2], \end{aligned}} \tag{19.49}$$

and, in the extreme limit $\omega = 0$, $\alpha = \pm 1$,

$$\boxed{\begin{aligned} d\hat{\hat{s}}^2 &= H_{\mathrm{M5}}^{-\frac{1}{3}}\left[dt^2 - d\vec{y}_5^{\,2}\right] - H_{\mathrm{M5}}^{\frac{2}{3}}\, d\vec{x}_5^{\,2}, \\ \tilde{\hat{\hat{C}}}_{t\underline{y}^1\cdots\underline{y}^5} &= \pm\left(H_{\mathrm{M5}}^{-1} - 1\right), \qquad H_{\mathrm{M5}} = 1 + \frac{h_{\mathrm{M5}}}{|\vec{x}_5|^3}, \end{aligned}} \tag{19.50}$$

which is what is usually called in the literature the M5-brane solution in isotropic coordinates and is regular everywhere.

Using the first of Eqs. (18.78), and Eqs. (19.20) and (16.43), we obtain

$$\boxed{h_{\mathrm{M5}} = \frac{(\ell_{\mathrm{Planck}}^{(11)})^3}{3\omega_{(4)}}.} \tag{19.51}$$

The event horizon of the black M5-brane, placed at $\rho = (-\omega)^{\frac{1}{3}}$, has (constant-time sections of) S^4 topology and a singular event horizon.[8]

As in the M2 case, the event horizon remains regular but degenerate in the extreme limit but now the singularity disappears and there is a new asymptotically flat region across the horizon (see its Penrose diagram in Figure 19.1). Using the coordinate r defined by

$$\rho = h_{\mathrm{S5}}^{\frac{1}{3}} r^2/(1-\rho^6)^{\frac{1}{3}}, \tag{19.52}$$

the metric takes the form [863]

$$d\hat{\hat{s}}^2 = r^2\left[dt^2 - d\vec{y}_5^{\,2}\right] - h_{\mathrm{S5}}^{\frac{2}{3}}\left[\frac{4}{r^2(1-r^6)^{\frac{8}{3}}}dr^2 + \frac{d\Omega_{(4)}^2}{(1-r^6)^{\frac{2}{3}}}\right] \tag{19.53}$$

and covers both sides of the event horizon $\rho = r = 0$. The spatial infinities are now at $r = \pm 1$. This metric is invariant under $r \to -r$, which allows us to identify the two asymptotic regions. There are more asymptotic regions, all of which may also be identified.

[8] This is a common feature of all extreme p-branes with $p \geq 1$ [863].

In the near-horizon limit $\rho = |\vec{x}_5| \to 0$ of the extreme M5 we obtain another maximally supersymmetric solution, whose metric is the direct product of those of AdS_7 and S^4 with radii R_7 and $\mathrm{R}_7/2$,

$$\boxed{d\hat{\hat{s}}^2 = R_7^2\, d\Pi^2_{(7)} - (R_7/2)^2\, d\Omega^2_{(4)}, \qquad \tilde{\hat{\hat{C}}}_{t\underline{y}^1\cdots\underline{y}^5} = \left(\frac{r}{R_7}\right)^6, \qquad R_7 = 2h^{\frac{1}{2}}_{\mathrm{M5}},} \tag{19.54}$$

where we are using again the notation of Eq. (19.48). The 4-form field strength is given by the S^4 volume form $\hat{\hat{G}} = 3(R_7/2)^3\omega_{(4)}$ and, again, the M5 can be seen as a vacuum-interpolating soliton [452]. This vacuum induces spontaneous compactification on S^4, which is described by a $d = 7$ gauged SUEGRA with SO(5) gauge group (the isometry group of S^4) and an AdS_7 vacuum [771].

These vacua play a crucial role in the AdS/CFT correspondence proposed by Maldacena in [679] (for a review see [23]).

19.2.4 The fundamental string F1

The ten-dimensional black-fundamental-string solution is given, in the modified Einstein frame and in the string frame, by

$$\boxed{\begin{aligned} d\tilde{\hat{s}}^2_{\mathrm{E}} &= H_{\mathrm{F1}}^{-\frac{3}{4}}[W dt^2 - dy^2] - H_{\mathrm{F1}}^{\frac{1}{4}}\big[W^{-1}d\rho^2 + \rho^2 d\Omega^2_{(7)}\big], \\ d\hat{s}^2_{\mathrm{s}} &= H_{\mathrm{F1}}^{-1}[W dt^2 - dy^2] - \big[W^{-1}d\rho^2 + \rho^2 d\Omega^2_{(7)}\big], \\ e^{-2\hat{\phi}} &= e^{-2\hat{\phi}_0} H_{\mathrm{F1}}, \qquad \hat{B}_{t\underline{y}} = \alpha\big(H_{\mathrm{F1}}^{-1} - 1\big), \\ H_{\mathrm{F1}} &= 1 + \frac{h_{\mathrm{F1}}}{\rho^6}, \qquad W = 1 + \frac{\omega}{\rho^6}, \qquad \omega = h_{\mathrm{F1}}[1-\alpha^2]. \end{aligned}} \tag{19.55}$$

The extreme limit $\omega = 0$, $\alpha = \pm 1$ is known as the *fundamental-string solution* [281, 282]:

$$\boxed{\begin{aligned} d\tilde{\hat{s}}^2_{\mathrm{E}} &= H_{\mathrm{F1}}^{-\frac{3}{4}}\big[dt^2 - dy^2\big] - H_{\mathrm{F1}}^{\frac{1}{4}} d\vec{x}_8^{\,2}, \\ d\hat{s}^2_{\mathrm{s}} &= H_{\mathrm{F1}}^{-1}[dt^2 - dy^2] - d\vec{x}_8^{\,2}, \\ e^{-2\hat{\phi}} &= e^{-2\hat{\phi}_0} H_{\mathrm{F1}}, \qquad \hat{B}_{t\underline{y}} = \pm\big(H_{\mathrm{F1}}^{-1} - 1\big), \qquad H_{\mathrm{F1}} = 1 + \frac{h_{\mathrm{F1}}}{|\vec{x}_8|^6}. \end{aligned}} \tag{19.56}$$

Using the first of Eqs. (18.78), and Eqs. (14.3), (14.4), and (19.26), we obtain

$$\boxed{h_{\mathrm{F1}} = \frac{(2\pi\ell_{\mathrm{s}})^6 g^2}{6\omega_{(7)}},} \tag{19.57}$$

which, in the weak-coupling limit $g \to 0$, with ℓ_s fixed, goes quickly to zero, giving a flat spacetime metric. The F1 solution can then be understood as the long-range fields produced by a fundamental string in the strong-coupling limit. In the weak-coupling limit, the string decouples from the supergravity fields.

The event horizon of the black solution becomes singular in the extreme limit both in the Einstein and in the string frame, and in that limit the dilaton also diverges at the horizon as $\hat{\phi} \sim \ln|\vec{x}_8|$. In the dual string frame (the frame in which the S5-brane is fundamental and there is no dilaton factor in its worldvolume action, which is related to the string frame by Eqs. (18.50) and (18.51)), ignoring the constant in $H_{\rm F1}$ leads to the solution with metric

$$d\hat{s}^2_{\rm S5} = \frac{\rho^4}{h_{\rm F1}^{\frac{2}{3}}}\left[dt^2 - dy^2\right] - h_{\rm F1}^{\frac{1}{3}}\frac{d\rho^2}{\rho^2} - h_{\rm F1}^{\frac{1}{3}} d\Omega^2_{(7)}, \tag{19.58}$$

which is the direct product of the round S^7 metric with radius $h_{\rm F1}^{\frac{1}{6}}$ and the metric of a 1-brane in three dimensions (i.e. a domain wall), which is singular. This near-horizon limit is well defined in this frame, even if it leads to a metric with singularities, but, unlike the M2 and M5 cases, it is not a maximally supersymmetric vacuum. These geometries play roles analogous to the $\mathrm{AdS}_n \times S^m$ geometries in non-conformal versions of the AdS/CFT correspondence [170].

19.2.5 The S5 solution

The ten-dimensional solitonic 5-brane solution is given by

$$\boxed{\begin{aligned}
d\tilde{\hat{s}}^2_{\rm E} &= H_{\rm S5}^{-\frac{1}{4}}\left[W dt^2 - d\vec{y}_5^{\,2}\right] - H_{\rm S5}^{\frac{3}{4}}\left[W^{-1}d\rho^2 + \rho^2 d\Omega^2_{(3)}\right],\\
d\hat{s}^2_{\rm s} &= W dt^2 - d\vec{y}_5^{\,2} - H_{\rm S5}\left[W^{-1}d\rho^2 + \rho^2 d\Omega^2_{(3)}\right],\\
e^{-2\hat{\phi}} &= e^{-2\hat{\phi}_0} H_{\rm S5}^{-1}, \qquad \hat{B}^{(6)}{}_{\underline{t}\underline{y}^1\cdots\underline{y}^5} = \alpha e^{-2\hat{\phi}_0}\left(H_{\rm S5}^{-1} - 1\right),\\
H_{\rm S5} &= 1 + \frac{h_{\rm S5}}{\rho^2}, \qquad W = 1 + \frac{\omega}{\rho^2}, \qquad \omega = h_{\rm S5}\left[1 - \alpha^2\right],
\end{aligned}} \tag{19.59}$$

and, in the extreme limit $\omega = 0, \alpha = \pm 1$ in which it is usually known as the *solitonic 5-brane solution* [206, 207] (also known as the *NS 5-brane*), it takes the form

$$\boxed{\begin{aligned}
d\tilde{\hat{s}}^2_{\rm E} &= H_{\rm S5}^{-\frac{1}{4}}\left[dt^2 - d\vec{y}_5^{\,2}\right] - H_{\rm S5}^{\frac{3}{4}} d\vec{x}_4^{\,2},\\
d\hat{s}^2_{\rm s} &= dt^2 - d\vec{y}_5^{\,2} - H_{\rm S5}\, d\vec{x}_4^{\,2},\\
e^{-2\hat{\phi}} &= e^{-2\hat{\phi}_0} H_{\rm S5}^{-1}, \qquad \tilde{\hat{B}}^{(6)}{}_{\underline{t}\underline{y}^1\cdots\underline{y}^5} = \pm e^{-2\hat{\phi}_0}\left(H_{\rm S5}^{-1} - 1\right),\\
H_{\rm S5} &= 1 + \frac{h_{\rm S5}}{|\vec{x}_4|^2},
\end{aligned}} \tag{19.60}$$

and is, like the M5-brane, regular everywhere. Using the first of Eqs. (18.78), the S5 tension, Eqs. (19.12), and $G_N^{(10)}$, Eq. (19.26), we obtain

$$\boxed{h_{S5} = \ell_s^2,} \tag{19.61}$$

which is independent of g and remains constant in the weak-coupling limit.

In the near-horizon limit $\rho = |\vec{x}_4| \to 0$ in the string frame (which is the dual S5-brane frame) we obtain a metric that is the product of Minkowski $6+1$ and that of a round S^3,

$$d\hat{s}^2 = dt^2 - d\vec{y}_5^{\,2} - dz^2 - h_{S5} d\Omega_{(3)}^2, \qquad z = h_{S5}^{\frac{1}{2}} \ln\left(\frac{\rho}{h_{S5}^{\frac{1}{2}}}\right). \tag{19.62}$$

The S5-brane metric interpolates, then, between Minkowski spacetime at infinity and the above regular metric at the horizon, which is at an infinite proper distance. There is no need to continue the metric analytically beyond the horizon and, actually, it would be more correct to say that, in the limit $\rho \to 0$, one finds another asymptotic region with the above metric.

19.2.6 The Dp-branes

The generic solution for black Dp-branes $N = 2A, B, d = 10$ SUEGRA with $p < 7$ is

$$\boxed{\begin{aligned}
d\tilde{\hat{s}}_E^2 &= H_{Dp}^{-\frac{7-p}{8}}\left[W dt^2 - d\vec{y}_p^{\,2}\right] - H_{Dp}^{\frac{p+1}{8}}\left[W^{-1} d\rho^2 + \rho^2 d\Omega_{(8-p)}^2\right],\\
d\hat{s}_s^2 &= H_{Dp}^{-\frac{1}{2}}\left[W dt^2 - d\vec{y}_p^{\,2}\right] - H_{Dp}^{\frac{1}{2}}\left[W^{-1} d\rho^2 + \rho^2 d\Omega_{(8-p)}^2\right],\\
e^{-2\hat{\phi}} &= e^{-2\hat{\phi}_0} H_{Dp}^{\frac{p-3}{2}}, \qquad \hat{C}^{(p+1)}{}_{\underline{t}\underline{y}^1\cdots\underline{y}^p} = \alpha e^{-\hat{\phi}_0}\left(H_{Dp}^{-1} - 1\right),\\
H_{Dp} &= 1 + \frac{h_{Dp}}{\rho^{7-p}}, \qquad W = 1 + \frac{\omega}{\rho^{7-p}}, \qquad \omega = h_{Dp}[1-\alpha^2].
\end{aligned}} \tag{19.63}$$

The above solution is not entirely correct for $p = 3$ since it does not take into account the self-duality of the 5-form field strength, but the only change that has to be made is in the 4-form potential: the metric and dilaton fields are correct, as we will see.

In the extreme limit $\omega = 0, \alpha = \pm 1$ the solutions are valid for all $p = 0, \ldots, 9$ (with the same caveats in the case $p = 3$) for harmonic functions with poles of the right order. These are the solutions usually known as *Dp-brane solutions* in the literature:

$$\boxed{\begin{aligned}
d\tilde{\hat{s}}_E^2 &= H_{Dp}^{\frac{p-7}{8}}\left[dt^2 - d\vec{y}_p^{\,2}\right] - H_{Dp}^{\frac{p+1}{8}} d\vec{x}_{9-p}^{\,2},\\
d\hat{s}_s^2 &= H_{Dp}^{-\frac{1}{2}}\left[dt^2 - d\vec{y}_p^{\,2}\right] - H_{Dp}^{\frac{1}{2}} d\vec{x}_{9-p}^{\,2},\\
e^{-2\hat{\phi}} &= e^{-2\hat{\phi}_0} H_{Dp}^{\frac{p-3}{2}}, \qquad \hat{C}^{(p+1)}{}_{\underline{t}\underline{y}^1\cdots\underline{y}^p} = \pm e^{-\hat{\phi}_0}\left(H_{Dp}^{-1} - 1\right),\\
H_{Dp} &= 1 + \frac{h_{Dp}}{|\vec{x}_{9-p}|^{7-p}}, \quad p < 7, \qquad H_{D7} = 1 + h_{D7} \ln|\vec{x}_2|, \quad H_{D8} = 1 + h_{D8}|x|.
\end{aligned}} \tag{19.64}$$

Using Eqs. (18.78), the Dp-brane tension formula, Eq. (19.11), and the value of $G_N^{(10)}$, Eq. (19.26), we find

$$\boxed{h_{\mathrm{D}p} = \frac{(2\pi\ell_s)^{7-p} g}{(7-p)\omega_{(8-p)}}, \;\; p<7, \qquad h_{\mathrm{D}7} = -\frac{g}{2\pi}, \qquad h_{\mathrm{D}8} = -\frac{g}{4\pi\ell_s}.} \tag{19.65}$$

Several remarks are in order here.

1. The D-instanton ($p = -1$) solution is not included in this general case. It will be dealt with in Section 19.2.7.

2. The D-string ($p = 1$) solution is related by IIB S duality (with $S = \eta$) to the F1B solution. More general S-duality transformations generate solutions that represent bound states of q F1Bs and p D1s called pq-strings [822]. The same can be said about the D5 and the S5B, which can be combined into pq 5-branes [670]. There are also pq 7-brane solutions, but they have a more complicated interpretation. We will study these solutions in Section 19.4.3.

3. The metric and dilaton of the $p = 3$ solution are those of the self-dual D3-brane solution, but the RR potential is different. The correct field strength is just the self-dual part of the field strength of the generic solution and its components are

$$\hat{G}^{(5)}{}_{\underline{m}t\underline{y}^1\underline{y}^3\underline{y}^3} = \mp\frac{e^{-\hat{\varphi}_0}}{2} H_{\mathrm{D}3}^{-2}\partial_{\underline{m}} H_{\mathrm{D}3}, \qquad \hat{G}^{(5)}{}_{\underline{m}_1\cdots\underline{m}_5} = \pm\frac{e^{-\hat{\varphi}_0}}{2}\epsilon_{\underline{m}_1\cdots\underline{m}_5\underline{m}_6}\partial_{\underline{m}_6} H_{\mathrm{D}3}. \tag{19.66}$$

4. The solutions for $p < 7$ are well defined for all values of $|\vec{x}_{9-p}| > 0$ ($H_{\mathrm{D}p} > 0$). The D7 and D8 solutions are well defined only in certain regions of the transverse space for which $H_{\mathrm{D}p} > 0$ due to the negative signs of $h_{\mathrm{D}7}$ and $h_{\mathrm{D}8}$. To obtain solutions that are well defined everywhere in the transverse space, one has to consider configurations with several branes and compact transverse spaces. The only singularities are then at the positions of the branes.[9] We are going to study the simplest of these combinations of D7-branes in Section 19.2.8. The simplest combination of D8-branes which leads to a regular metric is the orbifold construction discussed on page 519.

5. The Hodge dual of the 10-form field strength associated with the D8-brane solution is ${}^\star\hat{G}^{(10)} = \pm h_{\mathrm{D}8} g^{-1} = \mp 1/(4\pi\ell_s)$. This must, then, be the value of the mass parameter $m = \hat{G}^{(0)}$ of the Romans massive $N = 2A$, $d = 10$ supergravity that describes the effective string theory in the presence of one D8-brane [118, 782]. The parameter m which was completely arbitrary from the supergravity point of view must be quantized from the string-theory point of view.

6. The $p = 9$ solution is just flat spacetime.

[9] The positions of the branes can be identified in general with the poles of the harmonic functions, although we know that, in many cases, these poles are just coordinate singularities and correspond to regular event horizons.

7. In all the $p < 7$ cases, except for $p = 3$, the Dp-brane horizon is singular and its Penrose diagram is given in Figure 19.1. The near-horizon geometries (in the dual frame) also correspond to solutions with singularities.

8. In the $p = 3$ case the solution has a regular horizon and the analytic continuation across it is completely regular, as in the M5 case. There is again a discrete isometry that relates the old and new asymptotically flat regions. The near-horizon geometry of the D3-brane is a maximally supersymmetric solution with the metric of the completely regular space $\mathrm{AdS}_5 \times \mathrm{S}^5$:

$$\boxed{d\hat{s}_{\mathrm{s}}^2 = R_5^2\, d\Pi_{(5)}^2 - R_5^2\, d\Omega_{(5)}^2, \quad \hat{G}^{(5)} = \pm 2 e^{-\hat{\varphi}_0} R_5^4 \big(\omega_{\mathrm{AdS}_5} - \omega_{\mathrm{S}^5}\big)\,, \quad R_5 = h_{\mathrm{D3}}^{\frac{1}{4}}} \tag{19.67}$$

(compare with Eq. (13.100)), which has played a crucial role in the AdS/CFT-correspondence conjecture [23, 679]. This solution induces spontaneous compactification on S^5. The theory is described by gauged $N = 4, d = 5$ SUEGRA with gauge group SO(6) [487, 623].

19.2.7 The D-instanton

If we extrapolate the association of $(p+1)$-forms to objects with a $(p+1)$-dimensional worldvolume to a 0-form (scalar) potential such as the type-IIB RR 0-form $\hat{C}^{(0)}$, we conclude that the IIB theory admits a "-1-brane," an object with a zero-dimensional worldvolume: just a point in spacetime. Such an object must be an instanton, which is a Euclidean solution. Then, associated with the type-IIB RR 0-form $\hat{C}^{(0)}$, we expect to find a *D-instanton solution* to the Euclidean equations of motion of the $N = 2B, d = 10$ supergravity theory.[10] It is not clear how to define the complete Euclidean $N = 2B, d = 10$ supergravity because it is not possible to have a *real* self-dual 5-form (a formal definition has nevertheless been given in [435]), but this problem does not arise if one ignores the 5-form field strength, as we do here.

We start with the action of the truncated $N = 2B, d = 10$ theory in the Einstein frame and with Lorentzian signature in which we keep only the metric and the dilaton and the RR 0-form combined in the complex scalar $\hat{\tau} = \hat{C}^{(0)} + i e^{-\hat{\varphi}}$:

$$\hat{S} = \frac{\hat{g}_{\mathrm{B}}^2}{16\pi G_{\mathrm{N}}^{(10)}} \int d^{10}\hat{x} \sqrt{|\hat{\jmath}_{\mathrm{E}}|} \left\{ \hat{R}_{\mathrm{E}} + \tfrac{1}{2} \frac{\partial_\mu \hat{\tau}\, \partial^\mu \bar{\hat{\tau}}}{(\mathrm{Im}\,\hat{\tau})^2} \right\}, \tag{19.68}$$

To obtain the Euclidean action we have to perform a Wick rotation, which can be understood as a coordinate redefinition $t = x^0 = i\bar{x}^{10}$, where τ is treated as real afterwards. $\hat{\varphi}$ is a scalar and transforms as such under this reparametrization. However, $\hat{C}^{(0)}$ can be treated

[10] D-instantons of bosonic-string theories had been considered in [467, 776] before the relation between RR charges and D-branes was discovered in [777]. The IIB instanton BPS condition had been obtained in [466].

either as a scalar (and then its Hodge dual, the RR 7-form $\hat{C}^{(7)}$, has to be treated as a pseudotensor) or as a pseudoscalar (and the RR 7-form $\hat{C}^{(7)}$ has to be treated as a tensor).[11] The second option (which is also the one we adopted for consistency when we defined the magnetic RR potentials) was chosen by the authors of [435], who performed the Wick rotation using the RR 7-form and treating it as a tensor. If $\hat{C}^{(0)}$ is a pseudoscalar then it acquires an extra factor of i in the Wick rotation: $\hat{C}^{(0)} = i\bar{\hat{C}}^{(0)}$. The Euclidean action becomes

$$\bar{S} = \frac{\hat{g}_{\rm B}^2}{16\pi G_{\rm N}^{(10)}} \int d^{10}\bar{\hat{x}}\sqrt{|\bar{\hat{g}}_{\rm E}|}\left\{\bar{\hat{R}}_{\rm E} + \tfrac{1}{2}e^{2\hat{\varphi}}\left[(\partial e^{-\hat{\varphi}})^2 - (\partial\bar{\hat{C}}^{(0)})^2\right]\right\}, \tag{19.69}$$

Observe that the two scalars contribute with different signs to the action. Their "energy–momentum" tensors will appear with opposite signs in the Einstein equation. Thus, one can obtain a solution with flat spacetime by taking the derivatives of the two scalars to be equal, up to a global sign. Then one need only solve the scalar equations. The *D-instanton solution* takes the following form in the (unmodified) Einstein and string frames [435]:

$$\boxed{\begin{array}{ll} d\bar{\hat{s}}_{\rm E}^2 = e^{-\frac{\hat{\varphi}_0}{2}}d\bar{\vec{x}}_{10}^{\,2}, & d\bar{\hat{s}}_{\rm s}^2 = H_{{\rm D}i}^{\frac{1}{2}}d\bar{\vec{x}}_{10}^{\,2}, \\ e^{-2\hat{\varphi}} = e^{-2\hat{\varphi}_0}H_{{\rm D}i}^{-2}, & \bar{\hat{C}}^{(0)} = \pm e^{-\hat{\varphi}_0}\left(H_{{\rm D}i}^{-1} - 1\right), \\ H_{{\rm D}i} = 1 + \dfrac{h_{{\rm D}i}}{|\bar{\vec{x}}_{10}|^8}. & \end{array}} \tag{19.70}$$

The value of $h_{{\rm D}i}$ is the extrapolation to $p=-1$ of the value of $h_{{\rm D}p}$ Eq. (19.65). This value can also be obtained via a T-duality relation with the D0-brane.

At first sight there is a singularity at $\rho = |\bar{\vec{x}}_{10}| = 0$ (in the string frame). However, the string metric is invariant under the reparametrization

$$\rho = h_{{\rm D}i}^{\frac{1}{4}}/\tilde{\rho}, \tag{19.71}$$

which shows that, in the limit $\rho \to 0$, one finds another asymptotically flat region identical to the one at $\rho \to \infty$. The metric, therefore, describes a sort of Euclidean wormhole joining the two asymptotically flat regions and is regular everywhere.

The value of the Euclidean action of the D-instanton can be calculated. In the modified Einstein frame, and normalized in our conventions, it takes the value

$$I = 2\pi/g_{\rm B} = q_{{\rm D}i}/g_{\rm B}. \tag{19.72}$$

The action Eq. (19.68) appears in other contexts in $d \neq 10$ dimensions. We have met it, for instance, as a truncation of the $N=4, d=4$ SUEGRA theory that arises in the toroidal reduction of the heterotic-string effective action Eq. (16.160), but it appears in many other reductions, as shown in [666]. It is possible to find instanton solutions for all of them that are almost identical with the D-instanton solution, differing only in the harmonic function which, in $d>2$ dimensions, will be $H = 1 + h/|\vec{x}_d|^{d-2}$. The string-frame geometry will

[11] This difference between these two options can be seen as the reason why Wick rotations and Hodge duality do not commute.

be different in each case. The $d = 4$ solution associated with the heterotic string was found in [453] and also has a wormhole interpretation.

19.2.8 The D7-brane and holomorphic $(d-3)$-branes

The D7-brane solution Eqs. (19.64) is just the simplest of a very rich family of solutions of the action Eq. (19.68), which share many interesting properties and some pathologies that can be eliminated after a careful analysis. They are the subject of this section.

The $\mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$ σ-model of Eq. (19.68) is invariant under transformations of the whole group $\mathrm{SL}(2,\mathbb{R})$, Eqs. (11.205) and (11.206), but only the discrete subgroup $\mathrm{SL}(2,\mathbb{Z})$ is supposed to relate equivalent (dual) type-IIB theories. On the other hand, as discussed in Section 11.4.1, only the modular group $\mathrm{G} \equiv \mathrm{PSL}(2,\mathbb{Z}) = \mathrm{SL}(2,\mathbb{Z})/\{\pm\mathbb{I}_{2\times 2}\}$ acts on $\hat{\tau}$. In conclusion, type-IIB S duality tells us, then, that values of the $\hat{\tau}$ field that are related by modular transformations must be considered equivalent and should be *identified*. The same will be true in the cases in which τ can be viewed as the modular parameter of a torus.[12] Thus τ, which in principle takes values in the whole complex upper half plane $\mathbb{H}$, can be restricted to take values in the *fundamental domain* of the modular group in $\mathbb{H}$, which we are going to discuss now.

The modular group G is generated by the elements T and S

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad\qquad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \tag{19.73}$$

whose actions on τ are $T(\tau) = \tau + 1$ and $S(\tau) = -1/\tau$. Observe that $S^2 = -\mathbb{I}_{2\times 2} \sim \mathbb{I}_{2\times 2}$ in G and also $(ST)^3 = (T^{-1}S)^3 \sim \mathbb{I}$ in G. Thus S and ST generate two cyclic subgroups of orders 2 and 3, respectively.

The fundamental domain of G in $\mathbb{H}$ can be defined as the quotient $\mathbb{H}/\mathrm{G}$ and corresponds to the region $|\tau| \geq 1$ and $-\frac{1}{2} \leq \mathrm{Re}(\tau) \leq \frac{1}{2}$ with the lines $\mathrm{Re}(\tau) = -\frac{1}{2}$ and $\mathrm{Re}(\tau) = +\frac{1}{2}$ identified by a T transformation and with the arc of unit radius $e^{i\theta},\ \ \theta \in [\pi/3, 2\pi/3]$ joining the fundamental domain corners $e^{\frac{2\pi i}{3}}$ and $e^{\frac{\pi i}{3}}$ identified with itself ("orbifolded") according to $e^{i\theta} \sim e^{i(\pi-\theta)}$ (the S transformation) (see Figure 19.2).

$\mathbb{H}/\mathrm{G}$ has, therefore, two special points associated with the two cyclic subgroups generated by S and ST: $\tau = i$, which is invariant under S; and $\tau = \rho \equiv e^{\frac{2\pi i}{3}}$, which is invariant under ST. If we consider its compactification $\widehat{\mathbb{H}/\mathrm{G}}$ in which the point at infinity is added, then a third special point appears: ∞ itself, which is invariant under the infinite subgroup of integer powers of T and can be understood as an infinite-order orbifold point.

Since the fundamental domain in which τ takes values is topologically non-trivial, we expect $\tau(x)$, which maps the transverse space on the fundamental domain, to be a multivalued function of x whose monodromies are in G. On the other hand, the real part of τ has, therefore, the typical behavior of an axion field and takes values in a circle.

[12] We remove all the hats henceforth, since the results of this section will be valid for many cases and dimensions apart from the $N = 2B$, $d = 10$ case. The results for D7-branes in $d = 10$ dimensions will be valid for $(d-3)$-branes in d dimensions.

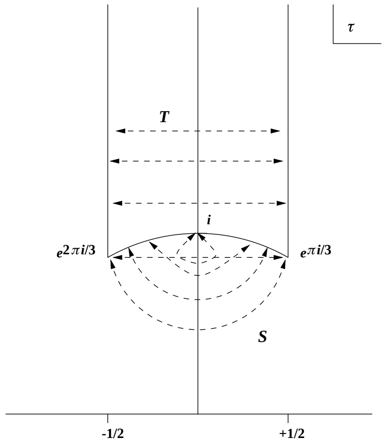

Fig. 19.2. The fundamental domain of the modular group.

We can now interpret the D7-brane solution in the light of the preceding discussion. First, it is useful to rewrite it using $\omega = x^1 + ix^2$ in the form

$$\boxed{d\tilde{\hat{s}}_{\rm E}^2 = dt^2 - d\vec{y}_7^{\,2} - \mathrm{Im}(\mathcal{H})\, d\omega d\bar{\omega}, \qquad \tau = \mathcal{H},} \tag{19.74}$$

where

$$\mathcal{H} = ie^{-\hat{\varphi}_0} h_{\rm D7} \ln\omega, \qquad \text{or} \qquad \mathcal{H} = ie^{-\hat{\varphi}_0} h_{\rm D7} \ln\bar{\omega}, \tag{19.75}$$

for D7- and anti-D7-branes (positive and negative charge w.r.t. $\hat{C}^{(0)}$), respectively, where we have eliminated the constant 1 in $H_{\rm D7}$ since the solution is not asymptotically flat anyway. If we go around the origin $\omega = 0$ at which the (anti-)D7 is placed, then, according to the source calculation,

$$\omega \to e^{2\pi i}\omega, \quad \Rightarrow \tau \to \tau \pm 1 = T^{\pm 1}(\tau). \tag{19.76}$$

The D7-brane solution has, as we expected from our general discussion, non-trivial monodromy. Furthermore, the monodromy around a D7-brane with charge n is T^n. For $(d-3)$-branes monodromy plays the role of charge (they are equivalent, when standard charge can be defined), which can be represented by a monodromy matrix.

The D7-brane solution is, however, defined only in the disk $|\omega| < 1$ due to the negative sign of $h_{\rm D7}$ and we may be interested in different transverse spaces. The corresponding solutions may be found thanks to the following observation: the *Ansatz* Eqs. (19.74) is a solution for any $\mathcal{H}$ that is a holomorphic or antiholomorphic function of ω [474]. D7-branes will be placed at points around which the monodromy of τ is T^n. The restriction to (anti)holomorphicity has to do with the impossibility of having objects with opposite charges in equilibrium.

Observe that what appears in the metric is $\mathrm{Im}(\mathcal{H})$, *not* $\mathrm{Im}(\tau)$, even if they coincide in this form of the solution. Then $g_{\omega\bar{\omega}}$ does not transform under G transformations of τ, but the relation $g_{\omega\bar{\omega}} = \mathrm{Im}(\tau)$ breaks down. In fact, the general 7-brane solution can be written in a more general form[13]

$$\boxed{d\tilde{s}_{\mathrm{E}}^2 = dt^2 - d\vec{y}_7^{\,2} - \mathrm{Im}(\mathcal{H})\,|f(\omega)|^2 d\omega d\bar{\omega}, \qquad \tau = \mathcal{H},} \tag{19.78}$$

where $f(\omega)$ is any holomorphic function of ω, but $f(\omega)$ can always be reabsorbed (locally!) into a change of coordinates $\omega' = F(\omega),\ dF/d\omega = f$, and $\tau(\omega') = \tau[F^{-1}(\omega')]$.

$\mathcal{H}$ is in general multivalued and hence $g_{\omega\bar{\omega}}$ could also be. When it is, we use a multivalued $f(\omega)$ in order to make $g_{\omega\bar{\omega}}$ single-valued. A general solution for $f(\omega)$ can be given on the basis of the observation that $\mathrm{Im}(\tau)$ transforms under G (or under monodromy) as the absolute value squared of a modular form of weight -1 would,[14] i.e.

$$\mathrm{Im}(\tau') = \frac{\mathrm{Im}(\tau)}{|\gamma\tau + \delta|^2}, \tag{19.79}$$

and, therefore, going around closed loops in transverse (ω) space $\mathrm{Im}(\mathcal{H})$ can transform in this way for some values of γ and δ. Then, we can build a single-valued function by multiplying $\mathrm{Im}(\mathcal{H})$ by the absolute value squared of a modular form of $\mathcal{H}$ of opposite weight 1, i.e. an $f[\mathcal{H}(\omega)]$ such that

$$f\left[\frac{\alpha\mathcal{H} + \beta}{\gamma\mathcal{H} + \delta}\right] = (\gamma\mathcal{H} + \delta)\, f(\mathcal{H}). \tag{19.80}$$

Then $g_{\omega\bar{\omega}} = \mathrm{Im}(\mathcal{H})\,|f(\omega)|^2$ will always be single-valued. The choice of modular form is not unique, though. The choice in [474] was

$$f = \eta^2(\mathcal{H}) \prod_{n=1}^{N} (\omega - \omega_n)^{-\frac{1}{12}}, \tag{19.81}$$

where η is Dedekind's function[15]

$$\eta(z) = q^{\frac{1}{24}} \prod_{n=1}^{\infty} (1 - q^n), \qquad q = e^{2\pi i z}, \tag{19.82}$$

---

[13] There is another form of the general solution, which is manifestly SL(2, $\mathbb{R}$)-invariant:

$$\boxed{d\tilde{s}_{\mathrm{E}}^2 = dt^2 - d\vec{y}_7^{\,2} - e^{-2U}\, d\omega d\bar{\omega}, \qquad \tau = \mathcal{H}_1/\mathcal{H}_2, \qquad e^{-2U} = \mathrm{Im}\big(\mathcal{H}_1\bar{\mathcal{H}}_2\big),} \tag{19.77}$$

where $\mathcal{H}_{1,2}$ are two arbitrary holomorphic functions ω transforming as a doublet under SL(2, $\mathbb{R}$), both in τ and in the metric (but e^{-2U} is invariant, as it must be). The structure of this family is similar to that of the SWIP solutions of $N = 4, d = 4$ SUGRA (Section 12.2.1). We can relate it either to the solution Eq. (19.74) as the particular case $\mathcal{H}_1 = \mathcal{H},\ \ \mathcal{H}_2 = 1$ or to the solution Eq. (19.78) as the particular case $\mathcal{H}_1/\mathcal{H}_2 = \mathcal{H},\ \ f = \mathcal{H}_2$ since $\mathrm{Im}(\mathcal{H}_1\bar{\mathcal{H}}_2) = |\mathcal{H}_2|^2\mathrm{Im}(\mathcal{H}_1/\mathcal{H}_2)$.

[14] There are no modular forms of negative weight, however; see [848].

[15] Dedekind's η function is not, strictly speaking, a modular form: only its 24th power is a weight-12 modular form (actually, a cusp form, since it vanishes at infinity where it admits the expansion $\eta^{24} = q - 24q^2 + \cdots$) because its transform contains a phase factor that is a 24th root of unity. However, since we are taking its absolute value, this phase is immaterial.

and the ω_ns are the *decompactification points*[16] at which $\mathcal{H} \sim i \ln(\omega - \omega_n)$ so $\mathrm{Im}(\tau)$ diverges when we approach them. Near the decompactification points $\eta^2 \sim (\omega - \omega_i)^{\frac{1}{12}}$ and the metric would become singular without the additional factor $\Pi_{n=1}^{N}(\omega - \omega_n)^{-\frac{1}{12}}$. As a result of the presence of this factor, the spacetime will be asymptotically conical: when $|\omega| \to \infty$, $|f(\omega)| \sim |\omega|^{-\frac{N}{12}}$ and the transverse metric takes the form

$$dr^2 + (1 - N/12)^{12} r^2 d\theta^2, \qquad r = |\omega|^{1-\frac{N}{12}}/(1 - N/12), \tag{19.83}$$

and it is asymptotically conical with deficit angle $N/12$ when $N < 12$. For $N = 12$ (i.e. 12 D7-branes) the space is asymptotically conical; for $N > 12$ it has finite volume, but it is singular except in the exceptional case $N = 24$ [435, 474].

Let us now go back to the problem of finding a globally well defined D7 ($(d-3)$-brane) solution. This problem was first considered and solved in [474] with the Riemann sphere as the transverse space. The crucial observation is that there is essentially a unique function that maps the fundamental region of the modular group bijectively onto the sphere [848]: the modular invariant $j(\tau)$ given in terms of the even Jacobi θ functions $\theta_2(\tau), \theta_3(\tau),$ and $\theta_4(\tau)$ by

$$j(\tau) = \frac{(\theta_2^8 + \theta_3^8 + \theta_4^8)^3}{\eta^{24}}. \tag{19.84}$$

The simplest solution is then implicitly given by $j(\tau) = \omega$ or

$$\tau(\omega) = j^{-1}(\omega). \tag{19.85}$$

More general solutions can be obtained by replacing ω by a holomorphic function $h(\omega)$, which is often a quotient of polynomials.

$j(\tau)$ is single-valued on $\widehat{\mathbb{H}/\mathrm{G}}$ and $j^{-1}(\omega)$ is multivalued on S^2 with monodromy in G. The only points around which there are non-trivial monodromies are the inverse images of the orbifold points $\widehat{\mathbb{H}/\mathrm{G}}$; namely $j(\infty) = \infty, \;\; j(i) \equiv \omega_i,$ and $j(\rho) \equiv \omega_\rho$, and the monodromies are related to the transformations that leave them invariant. In fact we can describe the monodromies of $\tau(\omega) = j^{-1}(\omega)$ by a sphere with two branch cuts joining the points ∞ and ω_ρ and ω_ρ and ω_i as in Figure 19.3. Crossing the first cut in the sense of the arrow, the function jumps from τ to $T(\tau)$; and crossing the second, it jumps from τ to $S(\tau)$. To check this, we can compute the monodromy along closed paths around those points. The paths are represented in Figure 19.3 and are the inverse images of the open paths in $\mathbb{H}$ (closed in $\mathbb{H}/\mathrm{G}$) represented in Figure 19.4.

Around ∞ the above solution admits the expansion $\tau(\omega) \sim -[1/(2\pi i)] \ln \omega + \cdots$. Evidently, on going once around infinity ($\omega \to \omega e^{-2\pi i}$), $\tau \to \tau + 1 = T(\tau)$ and we can say that this solution describes a $(d-3)$-brane (a D7-brane in the ten-dimensional string context) of unit charge at $\omega = \infty$ and, due also to some other welcome properties, this is why this solution Eq. (19.85) is generally known as *the* finite-energy $(d-3)$-brane solution.

Around ω_i the above solution admits the expansion $\tau(\omega) \sim i + \alpha(\omega - \omega_i)^{1/2} + \cdots$, and, on going once around it (i.e. $\omega - \omega_i \to (\omega - \omega_i)e^{+2\pi i}$), we see that, to leading order in $\omega - \omega_i$, $\tau \to -1/\tau = S(\tau)$. Finally, around ω_ρ, $\tau(\omega) \sim \rho + \beta(\omega - \omega_\rho)^{\frac{1}{3}} + \cdots$, and, on going once around it (i.e. $\omega - \omega_\rho \to (\omega - \omega_\rho)e^{+2\pi i}$), we see that $\tau \to -(1 + 1/\tau) = T^{-1}S(\tau)$.

[16] These are the points at which there are D7-branes, associated with T^n monodromy.

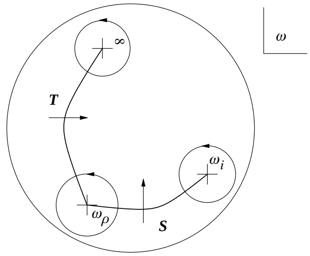

Fig. 19.3. The image of the fundamental domain of the modular group by $j(\tau)$.

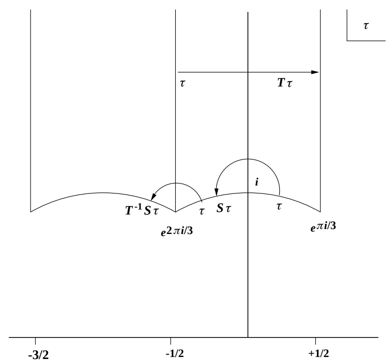

Fig. 19.4. The inverse image of the monodromy paths.

Apart from the fact that these singular points are given to us by the structure of the solution Eq. (19.85), it should be clear that the consistency of monodromy implies the existence of other singular points apart from ∞ such that the monodromy around all of them is T^{-1} [435]. The same conclusion could have been drawn from conservation of charge. Here the consistency of monodromy plays the same role as conservation of charge: we are dealing with a compact space and, just as the total charge has to be zero in such a space, the "total monodromy" has to be trivial.[17]

It is natural to associate with each singular point characterized by a monodromy matrix in G a $(d-3)$-brane. The standard IIB D7-branes are associated with T^n monodromies, but a consistent solution on a sphere requires, as we have seen, 7-branes with S and ST monodromy that may also have negative tensions. These 7-brane solutions are generically known as *pq 7-branes*, but, as distinct from pq-strings or 5-branes (Section 19.4.3), they are characterized by a PSL$(2\mathbb{Z})$ matrix, not by a pair of charges.

19.2.9 Some simple generalizations

The extreme p-brane solutions admit many generalizations. The most interesting ones describe intersections of branes and we will study them in Section 19.6. For a single p-brane,

[17] Just as in the D8-brane case, the $(d-3)$-branes that we have to add may have negative tensions.

the simplest generalizations involve either a modification of the worldvolume geometry, which we have taken so far to be flat $(p+1)$-dimensional spacetime, or a modification of the geometry of the transverse space.

Since the supergravity equations of motion are local, it is clear that global modifications of the worldvolume geometry such as imposing periodicity conditions on the n coordinates will give new solutions describing the p-branes *wrapped* on a rectangular n-torus. Another interesting possibility is to replace the flat worldvolume metric η_{ij} by $g_{ij}(t, \vec{y}_p)$ and the transverse metric δ_{mn} by $h_{mn}(x)$:

$$ds^2 = H^{\alpha}(x) g_{ij}(x) dy^i dy^j - H^{\beta}(x) h_{mn}(x) dx^m dx^n. \tag{19.86}$$

The equations of motion (with a minor modification of the $(p+1)$-form potential *Ansatz*) are still solved if g_{ij} and h_{mn} are Ricci-flat and H is harmonic in the new transverse space (see [179, 391, 418, 602]). This kind of solution can also be used to describe the wrapping of branes on cycles of more complicated spaces and also intersections.

A possible choice of transverse metric h_{mn} consists in the replacement of the round $\mathrm{S}^{(\tilde{p}+2)}$ metric $d\Omega^2_{(\tilde{p}+2)}$ by the metric of an Einstein space with the same curvature as the round $\mathrm{S}^{(\tilde{p}+2)}$ [341] (G/H coset spaces in the cases studied in [220]) in the flat metric written in spherical coordinates $d\rho^2 + \rho^2 d\Omega^2_{(\tilde{p}+2)}$. In the M2, M5, and D3 cases, the near-horizon limits are now the product of an AdS space and the Einstein space. This kind of solution induces spontaneous compactification in the Einstein space and the theory is described by a gauged supergravity with a gauge group related to the isometry group of the Einstein space (for a review, see e.g. [401]). Furthermore, since they do not preserve all the supersymmetries (unlike the $\mathrm{AdS}_n \times \mathrm{S}^m$ solutions), the supergravities will also have fewer supercharges.

It is also possible to look directly for metrics that can be understood as near-horizon limits (see, for instance, [389]).

19.3 The masses and charges of the p-brane solutions

In Section 19.1.1 we found the masses and charges of the extended objects of string/M theory using duality arguments. We matched these with the coefficients of the harmonic functions of the extreme solutions by studying the coupling of supergravity to the sources. This procedure is, however, difficult or impossible to follow for generic solutions that represent complex systems of extended objects or are "black" (non-extreme). In those cases we need a procedure by which to calculate the masses and charges of the objects described by the solutions using only the solution and the normalizations of the fields that appear in the action. This is the subject of this section. We will follow [40, 677].

19.3.1 Masses

We need to collect here several pieces of data that are scattered over several chapters.

1. The closed-superstring worldvolume action is given in Eq. (14.1.1). $T = 1/(2\pi\alpha')$ is the string tension, $\alpha' = \ell_s^2$ is the Regge slope, and ℓ_s is the string length. With that normalization of the worldvolume fields, the low-energy effective action of the

common sector of the different superstring theories is, in the string frame, given by Eq. (15.1). The complete effective actions of the type-IIA and -IIB superstring theories are given in Eqs. (16.38) and (17.4), respectively.

2. In the normalization factor of Eq. (15.1) $\hat{g}$ is the (dimensionless) $d = 10$ string coupling constant, which is related to the dilaton vacuum expectation value by Eq. (14.57). (Sometimes we use $\hat{g}_{\rm A}$ ($\hat{g}_{\rm B}$, $\hat{g}_{\rm I}$, $\hat{g}_{\rm h}$) for the coupling constant of the type-IIA (-IIB, -I, heterotic) superstring theory, but it should be clear from the context which coupling constant we are talking about.)

   The explicit factor of $\hat{g}^2$ is meant to absorb the asymptotic value of the dilaton, so that $G_{\rm N}^{(10)}$ is really the ten-dimensional Newton constant. We are assuming here that string-frame metrics are asymptotically flat.

3. When we are dimensionally reducing the above actions to d dimensions, we integrate over the compact, redundant, dimensions and obtain an overall factor that is the volume of these compact dimensions, V_{10-d}. One finds the following relations between the d-dimensional string coupling constant g and Newton constant $G_{\rm N}^{(d)}$ and the ten-dimensional ones (see Section 11.2.2):

$$G_{\rm N}^{(d)} = G_{\rm N}^{(10)}/V_{10-d}, \qquad g = \hat{g}/V_{10-d}. \tag{19.87}$$

   Here we are going to compactify on rectangular tori, with orthogonal circles, and so the volume of the compact space is

$$V_{10-d} = (2\pi)^{10-d} R_9 \cdots R_d. \tag{19.88}$$

4. The ten-dimensional Einstein metric $\hat{g}_{{\rm E}\,\hat{\mu}\hat{\nu}}$ is related to the string metric $\hat{g}_{\hat{\mu}\hat{\nu}}$ by

$$\hat{g}_{{\rm E}\,\hat{\mu}\hat{\nu}} = e^{-\frac{\hat{\phi}}{2}} \hat{g}_{\hat{\mu}\hat{\nu}}, \tag{19.89}$$

   i.e. by performing this rescaling in the above action we bring it into the canonical form because the factor $e^{-2\hat{\phi}}$ in front of the curvature disappears. However, the Einstein metric cannot be asymptotically flat if the string metric is and the factor in front of the action is $\hat{g}^2(16\pi G_{\rm N}^{(10)})^{-1}$.

   Following [677], we define a *modified Einstein metric* $\tilde{g}_{{\rm E}\,\hat{\mu}\hat{\nu}}$ that has the same value at infinity as the string metric and is the one to use to define masses. These masses are *the same as those masses that appear in the string spectrum* and, in particular, the mass of the fundamental string is independent of the string coupling constant:[18]

$$\tilde{\hat{g}}_{{\rm E}\,\hat{\mu}\hat{\nu}} \equiv e^{\frac{-\hat{\phi}-\hat{\phi}_0}{2}} \hat{g}_{\hat{\mu}\hat{\nu}} = \hat{g}^{\frac{1}{2}} \hat{g}_{{\rm E}\,\hat{\mu}\hat{\nu}}. \tag{19.91}$$

   If we rewrite the string effective action in terms of the modified Einstein metric, we obtain the correct normalization factor $(16\pi G_{\rm N}^{(10)})^{-1}$ and no dilaton factors.

[18] In d dimensions, the relation between the modified Einstein metric and the string metric is

$$\tilde{g}^{(d)}_{{\rm E}\,\mu\nu} = e^{-\frac{4}{d-2}(\phi-\phi_0)} g_{\mu\nu}. \tag{19.90}$$

5. In d dimensions, the mass M of any static, asymptotically flat metric describing a point-like object can be found from its asymptotic behavior at spatial infinity [706, 877]. In the string-theory context, the mass $M_{\rm E}$ associated with the "Einstein metric" (i.e. with the "wrong" normalization factor $g^2(16\pi G_{\rm N}^{(d)})^{-1}$ in the action) can be implicitly defined by

$$g_{{\rm E}\,tt} \sim 1 - \frac{16\pi G_{\rm N}^{(d)} M_{\rm E}}{g^2(d-2)\omega_{(d-2)}} \frac{1}{|\vec{x}_{d-1}|^{d-3}}, \qquad \vec{x}_{d-1} = (x^1, \ldots, x^{d-1}), \tag{19.92}$$

The mass M associated with the *modified Einstein metric* (i.e. with the "right" normalization factor $(16\pi G_{\rm N}^{(d)})^{-1}$) is defined analogously by,

$$\tilde{g}_{{\rm E}\,tt} \sim 1 - \frac{16\pi G_{\rm N}^{(d)} M}{(d-2)\omega_{(d-2)}} \frac{1}{|\vec{x}_{d-1}|^{d-3}}. \tag{19.93}$$

If both the string metric and the "modified string metric" are asymptotically flat (as we have assumed) then the Einstein metric is not, and it is necessary to rescale the coordinates with factors of $g^{\frac{1}{4}}$ in order to be able to use Eq. (19.92). Taking this into account, we find that the relation (in any dimension) between $M_{\rm E}$ and M is given by

$$M_{\rm E} = g^{\frac{1}{4}} M. \tag{19.94}$$

6. Under IIB S duality, it is the (unmodified) Einstein-frame metric that is invariant. Then $M_{\rm E}$ is S-duality-invariant, which implies the following S-duality transformation rule for M, which was already given in Eq. (19.3):

$$M'_{\rm E} = M_{\rm E}, \qquad \Rightarrow M' = g_{\rm B}'^{-\frac{1}{4}} g_{\rm B}^{\frac{1}{4}} M = g_{\rm B}^{\frac{1}{2}} M. \tag{19.95}$$

We can apply these formulae to find the masses of any of the solutions we have studied. They should coincide with the masses of the corresponding states in string/M theory. Let us take, for example, the F1 solution given in Eq. (19.56), assuming that the string is compactified on a circle and y is a compact dimension so we can dimensionally reduce the above solution, and calculate the modified Einstein mass of the resulting point-like object that lives in $d = 9$ by using Eq. (19.93).

First, we need the nine-dimensional dilaton (see Eq. (15.16)),

$$e^{-2(\phi-\phi_0)} = e^{-2(\hat{\phi}-\hat{\phi}_0)} \sqrt{|\hat{g}_{\underline{yy}}|} = H_{\rm F1}^{\frac{1}{2}}, \tag{19.96}$$

so, in this case, the relation between the nine-dimensional metrics is

$$\tilde{g}_{{\rm E}\mu\nu} = H_{\rm F1}^{\frac{1}{7}}\, g_{\mu\nu}, \qquad \Rightarrow \tilde{g}_{{\rm E}\,tt} = H_{\rm F1}^{-\frac{6}{7}} \sim 1 - \frac{6h_{\rm F1}}{7\rho^6}, \tag{19.97}$$

which, compared with Eq. (19.93), gives the right value,

$$M_{{\rm F1}w} = \frac{6h_{\rm F1}\omega_{(7)}}{16\pi G_{\rm N}^{(9)}} = \frac{12\pi R_9 h_{\rm F1}\omega_{(7)}}{16\pi G_{\rm N}^{(10)}} = \frac{R_9}{\ell_{\rm s}^9}, \tag{19.98}$$

on account of Eqs. (19.87), (19.26), and (19.57).

For the Dp-brane solutions ($p < 7$) Eqs. (19.64) compactified on p circles we calculate first the $(10 - p)$-dimensional dilaton, which is given by

$$e^{-2(\phi-\phi_0)} = e^{-2(\hat{\phi}-\hat{\phi}_0)}\sqrt{\hat{g}_{\underline{y}^1\underline{y}^1}\cdots\hat{g}_{\underline{y}^p\underline{y}^p}} = H_{\mathrm{D}p}^{\frac{p-6}{4}}. \tag{19.99}$$

Then, according to Eq. (19.90),

$$\tilde{g}_{\mathrm{E}\,\mu\nu} = H_{\mathrm{D}p}^{\frac{p-6}{2(8-p)}} g_{\mu\nu}, \quad \Rightarrow \quad \tilde{g}_{\mathrm{E}\,tt} = H_{\mathrm{D}p}^{-\frac{p-7}{p-8}} \sim 1 - \frac{p-7}{p-8} h_{\mathrm{D}p}\frac{1}{|\vec{x}_{p-9}|^{7-p}}, \tag{19.100}$$

which, compared with Eq. (19.93), gives again the right value,

$$M_{\mathrm{D}p} = \frac{(7-p)h_{\mathrm{D}p}\omega_{(8-p)}}{16\pi G_{\mathrm{N}}^{(10-p)}} = \frac{(7-p)(2\pi)^p R_9\cdots R_{10-p}h_{\mathrm{D}p}\omega_{(8-p)}}{16\pi G_{\mathrm{N}}^{(10)}} = \frac{R_9\cdots R_{10-p}}{\ell_{\mathrm{s}}^{p+1} g}. \tag{19.101}$$

19.3.2 Charges

With the normalization of the superstring effective actions Eqs. (16.38) and (17.4), the (electric) charges associated with the KR 2-form and the RR $(p+1)$-form potentials, which are carried, respectively, by fundamental strings and Dp-branes, can be defined by the integrals[19]

$$q_{\mathrm{F1}} = \frac{g^2}{16\pi G_{\mathrm{N}}^{(10)}}\int_{\mathrm{S}^7_\infty} e^{-2\hat{\phi}}\,{}^\star\hat{H}, \qquad q_{\mathrm{D}p} = \frac{g^2}{16\pi G_{\mathrm{N}}^{(10)}}\int_{\mathrm{S}^{8-p}_\infty} {}^\star\hat{G}^{(p+2)}, \tag{19.102}$$

whereas the charge associated with the NSNS 6-form potential, carried by the S5, is given by

$$q_{\mathrm{S5}} = \frac{g^2}{16\pi G_{\mathrm{N}}^{(10)}}\int_{\mathrm{S}^4_\infty} e^{2\hat{\phi}}\,{}^\star\hat{H}^{(7)} = \frac{g^2}{16\pi G_{\mathrm{N}}^{(10)}}\int_{\mathrm{S}^4_\infty}\hat{H}. \tag{19.103}$$

q_{F1} and q_{S5} are the electric–magnetic duals of each other, as are $q_{\mathrm{D}p}$ and $q_{\mathrm{D}\tilde{p}}$. With the above normalization, the generalization of the Dirac quantization condition for extended objects reads

$$q_{\mathrm{D}p}q_{\mathrm{D}\tilde{p}} = 2\pi n\frac{16\pi G_{\mathrm{N}}^{(10)}}{\hat{g}^2}, \qquad q_{\mathrm{F1}}q_{\mathrm{S5}} = 2\pi n\frac{16\pi G_{\mathrm{N}}^{(10)}}{\hat{g}^2}, \qquad n \in \mathbb{Z}. \tag{19.104}$$

It is easy to see that the values of the charges of the string/M-theory solutions coincide with the values we gave in Section 19.1.1. It should be stressed that both the masses and the charges of the extreme solutions are determined by the same h. This is due to the fact that the masses (tensions) and charges of these objects saturate BPS bounds. The solutions preserve half of the supersymmetries of the corresponding supergravity theory (see Section 19.5.1).

---

[19] These definitions are valid for field configurations in which only one potential is non-trivial. In general, the charge is obtained by integrating the form F such that $dF = 0$ is the equation of motion. (This is sometimes called the *Page charge* [687].) The presence of non-trivial Chern–Simons terms in the action implies that F consists in various terms, as we will discuss in Section 19.6.1.

19.4 Duality of string-theory solutions

In Section 19.1 we used string dualities to find and relate all the extended objects of string and M theories. In the subsequent sections we have established a relation between those objects and certain classical solutions of the string effective actions and $d = 11$ supergravity using arguments based on the symmetries of the solutions which determine the dimensions of the worldvolumes of the objects they describe, on the basis of the charges they carry and the matching with p-brane sources.

On the other hand, in Chapters 15–17 we learned how string dualities manifest themselves in string effective actions and, to close the loop, here we are going to see how the duality relations between string states are realized as relations between solutions of the effective actions. These relations are represented in Figures 19.4.1 and 19.4.1.

The three main types of duality relations that we are going to study are (i) those between the solutions of $d = 11$ supergravity and solutions of $N = 2A, d = 10$ supergravity, via the dimensional-reduction formulae Eqs. (16.35); (ii) those between solutions of $N = 2A, d = 10$ and $N = 2B, d = 10$ supergravity, via the type-II Buscher T-duality rules Eqs. (17.36) and (17.37); and (iii) those between solutions of $N = 2B$, $d = 10$ supergravity, via SL(2, $\mathbb{Z}$) transformations Eqs. (17.21) or (17.23) and (17.24). We are also going to need the results of Section 11.3.1 in order to perform reductions on transverse directions.

The supergravity duality transformations can also be used to construct new solutions. We will study two families of solutions constructed in this way: *pq* strings and *pq* 5-branes.

19.4.1 $N = 2A, d = 10$ *SUEGRA solutions from* $d = 11$ *SUGRA solutions*

There are two basic p-brane solutions of $d = 11$ SUGRA: the M2- and M5-brane solutions Eqs. (19.45) and (19.50). If they really describe the M2- and M5-brane states of M theory, their reduction must give rise to the F1A, D2, D4, and S5A solutions Eqs. (19.56), (19.64) and (19.60) of $N = 2A, d = 10$ SUEGRA [14, 335, 899] under *double* and *direct* dimensional reductions, i.e. in a worldvolume direction (corresponding to branes wrapped in the compact dimension) or in a transverse direction.

Double dimensional reductions are, by definition, made in a direction none of the fields depends on, and one just has to rewrite the solution in $d = 10$ variables using Eqs. (16.35) in a straightforward manner. The only subtlety is that, in order to have a non-trivial value for $\hat{g} = e^{\hat{\phi}_0}$, one must first rescale the compact worldvolume coordinate (that we call here z) $z \to e^{\frac{2}{3}\hat{\phi}_0} z$ in Eqs. (19.45) and (19.50).

Direct dimensional reductions are made precisely in one of the directions on which the p-brane metric depends. We could substitute the harmonic function for another one independent of the compact direction but, in that case, we would lose the relation to the quantum object it represents. The right procedure is, as we explained in Section 11.3.1, to construct first the correct solution that describes the p-brane in a transverse space with a compact coordinate, which amounts to solving the Laplace equation in such a space, and then Fourier-expand the solution, keeping only the zero mode. The solution is the same harmonic function as that which describes an infinite periodic array of parallel p-branes separated by a distance equal to the length of the compact direction. This harmonic function

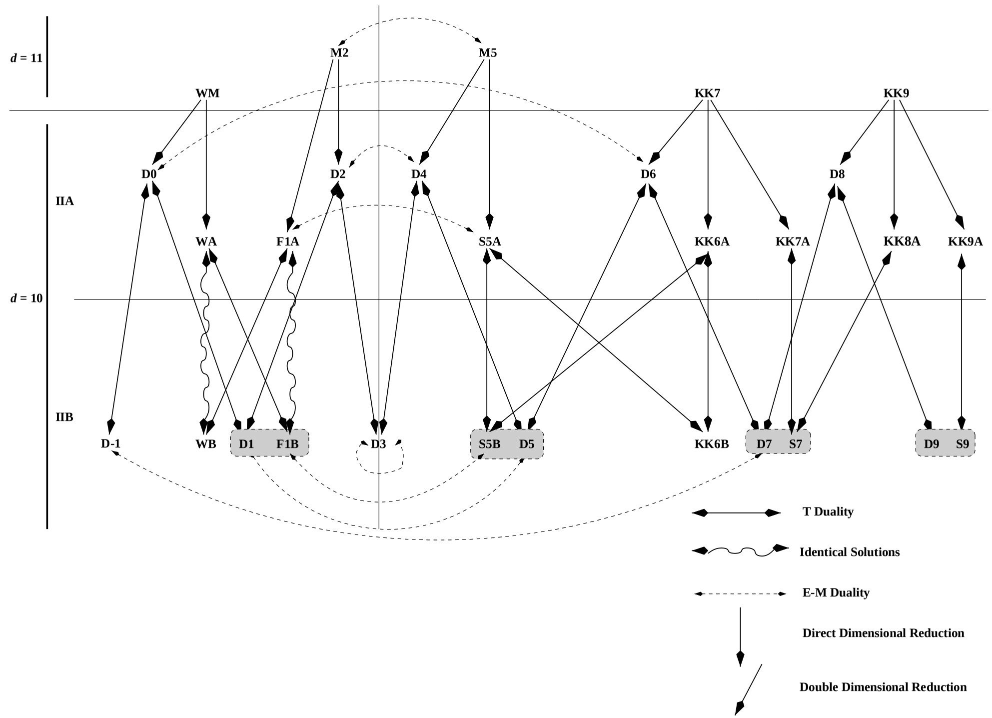

Fig. 19.5. Duality relations between classical solutions of ten- and 11-dimensional supergravity theories: p-branes, M-branes, D-branes, waves, and Kaluza–Klein monopoles. Lines with two arrows denote T-duality relations; dashed lines denote S-duality relations. Lines with a single arrow denote relations of dimensional reduction, either vertical (direct dimensional reduction) or diagonal (double dimensional reduction). Pairs of branes in boxes are type-IIB S-duality doublets.

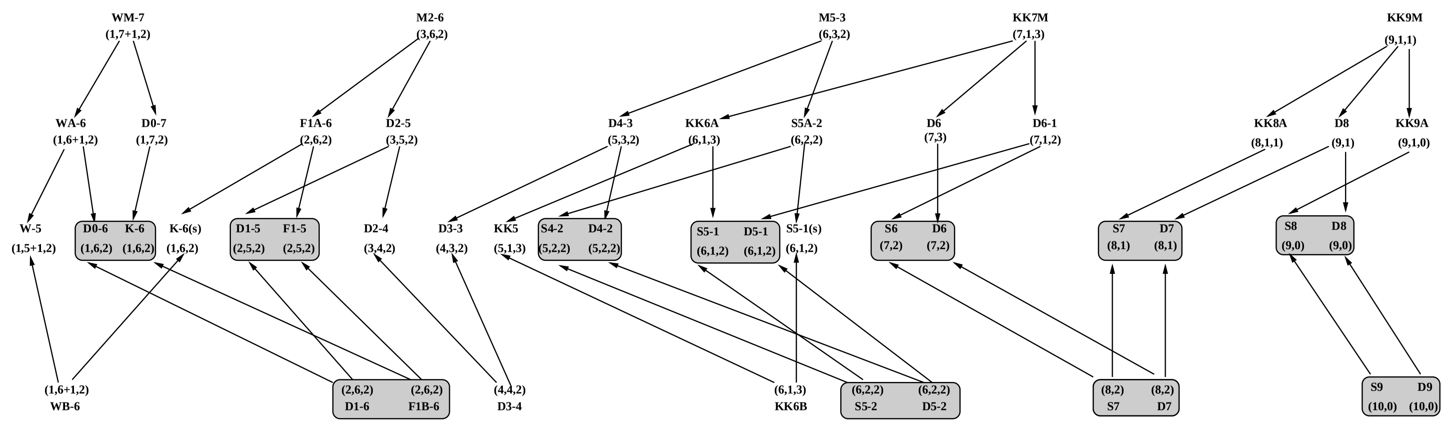

Fig. 19.6. Duality relations between KK branes. The numbers in parentheses represent the worldvolume dimension and isometric and transverse directions. The arrows indicate dimensional reduction in the corresponding kind of direction. In the upper row we represent M-theory KK branes, immediately below are ten-dimensional type-IIA branes, and below them nine-dimensional branes. Type-IIB KK branes are in the bottom row. Pairs of branes in boxes are S-duality doublets. Singlets are denoted with an (s).

is the linear superposition of those of each p-brane:

$$H_p = 1 + \frac{h_p}{|\vec{x}_{\tilde{p}+3}|^{\tilde{p}+1}} \longrightarrow H_p = 1 + \sum_{m\in\mathbb{Z}} \frac{h_p}{[|\vec{x}_{\tilde{p}+2}|^2 + (z + 2\pi m R_z)]^{\frac{\tilde{p}+1}{2}}}. \quad (19.105)$$

The zero mode can then be found using Eqs. (11.124). For M2- and M5-branes with a compact transverse coordinate ($\tilde{p} = 5, 2$ and $n = 6, 3$, respectively) we find

$$H_{\rm M2} \sim 1 + \frac{h_{\rm M2}\omega_{(5)}}{2\pi R_z \omega_{(4)}} \frac{1}{|\vec{x}_7|^5} = H_{\rm D2}, \qquad H_{\rm M5} \sim 1 + \frac{h_{\rm M5}\omega_{(2)}}{2\pi R_z \omega_{(1)}} \frac{1}{|\vec{x}_4|^2} = H_{\rm S5A}, \quad (19.106)$$

using the actual values of the integration constants h. The reduction is now straightforward.

This procedure is sometimes called *smearing*. The brane is said to be *delocalized* in one dimension. Sometimes duality gives delocalized solutions (for instance, oxidizing the F1A and D4 solutions using Eqs. (16.37)) and one has to show that one can indeed add the missing coordinate. This is a clear insufficiency of these methods.

There are more extended solutions in $N = 2A, d = 10$ SUEGRA that do not originate on M2- or M5-branes: the D0, D6, and D8 solutions. The $d = 11$ origin of the D8 is not known. To find the origins of the D0 and D6, we can simply apply the oxidation formulae Eqs. (16.37).

For the D0, we find that the only non-trivial field is the metric, which has the form of the pp-wave Eq. (10.42) with H replaced by $H_{\rm D0}$ and z by $e^{\frac{2}{3}\hat{\phi}_0}$. Actually, $H_{\rm D0}$ is the zero mode of the harmonic function H of an AS shock pp-wave moving in the compact 11th dimension. The reduction of such a shock wave was studied in Section 11.3.2 and, if we compare $h_{\rm D0}$ with h in Eq. (11.131), we find that it corresponds to a shock wave with the minimal momentum $p_z = 1/\ell_{\rm Planck}^{(11)}$. The D0-brane is, therefore, nothing but a KK mode.

The D6 also oxidizes to a purely gravitational solution: the KK monopole Eq. (11.160) (with six extra dimensions) with k_0 replaced by $e^{\frac{2}{3}\hat{\phi}_0}$ and H by $H_{\rm D6}$. Observe that $h_{\rm D6} = \ell_{\rm s} g/2 = R_z/2$ also has the right value, which is related to the periodicity of z and diverges in the decompactification limit. This means that the KK monopole is not a solution of standard $d = 11$ supergravity and can be included only when a dimension is compact.

This exercise shows the need to add purely gravitational solutions such as KK monopoles and waves in order to be able to explain the spectrum of objects of type-IIA superstring theory. Of course, once they are included in $d = 11$, if we reduce in a different dimension we find a gravitational wave and a KK monopole in (compactified) $N = 2A, d = 10$ supergravity.

19.4.2 $N = 2A/B, d = 10$ SUEGRA T-dual solutions

The type-II T-duality rules Eqs. (17.36) and (17.37) were derived using dimensional reduction and, therefore, if we want to T-dualize p-brane solutions in a transverse direction, we have to *delocalize* them previously in that direction. Conversely, T duality in a worldvolume direction typically takes us to a p-brane solution that is delocalized in a transverse direction. The main example is the whole chain of T dualities that relates the Dp-branes: if we start with the D0, fully localized in nine-dimensional transverse space, we have to delocalize it in one direction in order to find the D-string, which will be fully localized in eight dimensions and vice-versa.

One finds a precise correspondence through T duality between the coefficients $h_{\mathrm{D}p}$ that we have determined before: on T-dualizing a Dp in a transverse direction, we find the following coefficient after smearing:

$$\frac{h_{\mathrm{D}p}\omega_{(6-p)}}{2\pi R\omega_{(5-p)}} = \frac{(2\pi\ell_{\mathrm{s}})^{7-p}g}{2\pi R}\frac{\omega_{(6-p)}}{(7-p)\omega_{(8-p)}\omega_{(6-p)}}. \tag{19.107}$$

Using the T-duality rules for g and R Eqs. (19.1) and the identity

$$\frac{\omega_{(n-1)}}{n\omega_{(n+1)}\omega_{(n-2)}} = \frac{1}{(n-1)\omega_{(n)}},$$

we find $h_{\mathrm{D}(p+1)}(g')$, as given in Eq. (19.65).

Another example of T duality, namely that between the F1 and an AS shock wave that describes a string moving in a compact direction, can be found in Section 15.3. It is also a simple exercise to relate the S5, and KK monopole solutions by T duality in a transverse direction of the S5, which becomes the special isometric direction of the KK monopole.

19.4.3 S duality of $N = 2B, d = 10$ SUEGRA solutions: pq-branes

Performing S-duality transformations to relate the F1B to the D1 and the S5B to the D5 poses no problems but offers some opportunities: SL(2, $\mathbb{R}$) is a three-dimensional group and, after a general transformation, the new solution may have up to three new independent physical parameters. This procedure was used by Schwarz in [822] to construct a solution with four independent parameters. In the (unmodified) Einstein frame in which SL(2, $\mathbb{R}$) invariance is manifest, it takes the form

$$\boxed{\begin{array}{rcl} d\hat{s}_{\mathrm{E}}^2 &=& H_{pq1}^{-\frac{3}{4}}[dt^2 - dy^2] - H_{pq1}^{\frac{1}{4}}d\vec{x}_8^{\,2}, \\ \hat{\mathcal{B}}_{t\underline{y}} &=& \vec{a}(H_{pq1}^{-1} - 1), \qquad \hat{\mathcal{M}} = \vec{a}\vec{a}^{\mathrm{T}}H_{pq1}^{-\frac{1}{2}} + \vec{b}\vec{b}^{\mathrm{T}}H_{pq1}^{\frac{1}{2}}, \\ H_{pq1} &=& 1 + \dfrac{h_{pq1}\hat{g}^{-\frac{3}{2}}}{|\vec{x}_8|^6}, \qquad \vec{a}^{\mathrm{T}}\eta\vec{b} = 1. \end{array}} \tag{19.108}$$

Observe that h_{pq1} comes with a factor $\hat{g}^{-\frac{3}{2}}$ to take into account the rescaling of coordinates necessary to relate this metric to an asymptotically flat string of modified string metric with the usual formulae. The constant vectors $\vec{a}$ and $\vec{b}$ can be seen as the two column vectors of an SL(2, $\mathbb{R}$) matrix,

$$\left(\vec{a}\,\vec{b}\right) = \begin{pmatrix}\alpha & \beta \\ \gamma & \delta\end{pmatrix}, \qquad \vec{a}^{\mathrm{T}}\eta\vec{b} = \alpha\delta - \beta\gamma = 1, \tag{19.109}$$

and transform covariantly under SL(2, $\mathbb{R}$), which leaves this family invariant. The four independent parameters correspond to the asymptotic values of the two scalars, combined in the matrix $\hat{\mathcal{M}}_0$, and the charges $q_{\rm F1}$ and q_{D1} . The mass must be a function of those four parameters (again, a saturated BPS bound). To find the values of the physical parameters of this solution in terms of the constants $\vec{a}$, $\vec{b}$, and h_{pq1}, we can use an SL(2, $\mathbb{R}$) definition for the charges:

$$\begin{pmatrix} q_{\rm D1} \\ q_{\rm F1B} \end{pmatrix} = \vec{q} = \frac{g^2}{16\pi G_{\rm N}^{(10)}} \int_{{\rm S}^7_\infty} {}^\star\hat{\mathcal{M}}^{-1}\vec{\hat{\mathcal{H}}} = \frac{6\omega_{(7)} h_{pq1}\hat{g}^{-\frac{3}{2}}}{(2\pi\ell_{\rm s})^7\ell_{\rm s}} \hat{\mathcal{M}}_0^{-1}\vec{a}, \tag{19.110}$$

where $\hat{\mathcal{M}}_0 = \vec{a}\vec{a}^{\rm T} + \vec{b}\vec{b}^{\rm T}$. Using the property $\vec{a}^{\rm T}\hat{\mathcal{M}}_0^{-1}\vec{a} = 1$, we find the relation between $\vec{q}$ and h_{pq1}:

$$h_{pq1} = \frac{(2\pi\ell_{\rm s})^7\ell_{\rm s}\hat{g}^{\frac{3}{2}}}{6\omega_{(7)}} \sqrt{\vec{q}^{\rm T}\hat{\mathcal{M}}_0^{-1}\vec{q}}. \tag{19.111}$$

We can now express the full solution in terms of the physical parameters $\mathcal{M}_0$ and $\vec{q}$.

The object described by any of these solutions (usually called the *pq-string*) is a $p = 1$ object (a string) that has both $q_{\rm F1B}$ and $q_{\rm D1}$ charges in a IIB vacuum characterized by the moduli $\mathcal{M}_0$ ($e^{\hat{\varphi}_0} = \hat{g}$ and $\hat{C}_0^{(0)} = \hat{\theta}/(2\pi)$), and can be understood as the superposition of D1s and F1Bs. The values of the charges are therefore quantized: they can only be multiples of those of one (D, F) string: $n/(2\pi\ell_{\rm s}^2)$. The tension of this object is proportional to h_{pq1}, and, therefore, for trivial moduli, to

$$\sqrt{q_{\rm D1}^2 + q_{\rm F1B}^2} < |q_{\rm D1}| + |q_{\rm F1B}|.$$

The tension of two parallel (or coincident) strings of the same kind would be the sum of the tensions of each of them (*bound states at threshold*), which means that there is zero interaction energy and it costs zero energy to disintegrate the system. In this case, the tension is in general smaller, which means that this solution represents a bound state of F1Bs and D1s with non-zero binding energy (*non-threshold bound states*). However, the solution is stable with respect to disintegration only if the numbers of D1s and F1As are relatively prime: if they have a GCT different from 1, say N, the tension is N times that of a single pq-string with $n_{\rm D1}/N$ and $n_{\rm F1}/N$ strings, and it takes zero energy to disintegrate it. The pq-strings with coprime numbers of strings are the basic states of the theory.

A solution describing analogous bound states of D5 and S5Bs (*pq-5-branes*) was constructed in [670]:

$$\boxed{\begin{array}{rcl} d\hat{s}_{\rm E}^2 & = & H_{pq5}^{-\frac{1}{4}}[dt^2 - d\vec{y}_5^{\,2}] - H_{pq5}^{\frac{3}{4}} d\vec{x}_5^{\,2}, \\ \hat{\mathcal{B}}_{t\underline{y}^1\cdots\underline{y}^5} & = & \eta\vec{b}(H_{pq5}^{-1} - 1), \qquad \hat{\mathcal{M}} = \vec{a}\vec{a}^{\rm T}H_{pq5}^{\frac{1}{2}} + \vec{b}\vec{b}^{\rm T}H_{pq5}^{-\frac{1}{2}}, \\ H_{pq5} & = & 1 + h_{pq5}\dfrac{\hat{g}^{-\frac{1}{2}}}{|\vec{x}_4|^2}, \qquad \vec{a}^{\rm T}\eta\vec{b} = 1. \end{array}} \tag{19.112}$$

On T-dualizing these solutions, one obtains new bound states of F1s and Dps [671, 672]. These solutions describe intersecting branes with non-zero interaction energy. Other intersecting solutions with non-zero interaction energy are, for instance, the systems of Dp-branes and D$(p + 2)$-branes studied in [181, 256].

19.5 String-theory extended objects from superalgebras

In Chapter 5 we introduced supergravities as the gauge theories of the supersymmetry algebras. These contain a great deal of information about the global and local symmetries of each supergravity theory. When we studied four-dimensional Poincaré-extended supersymmetry algebras, we saw that, associated with each of the possible "electric" central charges Q^{ij}, there was an SO(2) gauge potential A^{ij}_{μ} whose gauge symmetry is generated by Q^{ij}. They contributed to the gauge superpotential with a term $\frac{1}{2}A^{ij}_{\mu}Q^{ij}$. The "magnetic" central charges could be associated with the electric–magnetic dual potentials, which are not independent. The central charges could be associated with electric and magnetic charges of supergravity solutions in Chapter 13 and the superalgebra could be used to see whether the solutions preserved any supersymmetries.

This correspondence between central charges and Abelian potentials holds for "quasi-central charges" with Lorentz indices as well, and with each charge $Z^{(p)}_{a_1\cdots a_p}$ we can associate in the supergravity theory a $(p+1)$-form potential $A^{(p+1)}$ that transforms under Abelian gauge transformations. They contribute to the gauge superpotential with a term $(1/p!)A^{(p+1)}{}_{\mu}{}^{a_1\cdots a_p}Z^{(p)}_{a_1\cdots a_p}$. The electric–magnetic dual $(\tilde{p}+1)$-form potential is associated with a $Z^{(\tilde{p})}_{a_1\cdots a_{\tilde{p}}}$ quasi-central charge that must also be present in the superalgebra. It is clear that these quasi-central charges must be associated with p-brane solutions of the supergravity theory [59] and that the superalgebra can be used to study their unbroken supersymmetries.

This is a very powerful tool that can be used to determine which objects/states may exist in a supergravity theory knowing just which quasi-central charges are algebraically allowed in the anticommutator of the supercharges of a given superalgebra.[20] Here we are going to write the superalgebras of the string/M-theory effective actions (supergravities) and we are going to study some examples, following in part [905] and starting with the algebra of $d=11$ supergravity.

The superalgebra of $d=11$ superalgebra (also known as *M superalgebra*) admits quasi-central charges of ranks $1, 2, 5, 6, 9,$ and 10. The last three values are just the duals of the first three. Therefore, the M superalgebra is usually written in the form

$$\left\{\hat{\bar{Q}}^{\alpha},\hat{\bar{Q}}^{\beta}\right\} = c(\hat{\hat{\Gamma}}^{\hat{a}}\hat{\hat{\mathcal{C}}}^{-1})^{\alpha\beta}\hat{\hat{P}}_{\hat{a}} + \frac{c_2}{2!}(\hat{\hat{\Gamma}}^{\hat{a}\hat{b}}\hat{\hat{\mathcal{C}}}^{-1})^{\alpha\beta}\hat{\hat{\mathcal{Z}}}^{(2)}_{\hat{a}\hat{b}} + \frac{c_5}{5!}(\hat{\hat{\Gamma}}^{\hat{a}_1\cdots\hat{a}_5}\hat{\hat{\mathcal{C}}}^{-1})^{\alpha\beta}\hat{\hat{\mathcal{Z}}}^{(5)}_{\hat{a}_1\cdots\hat{a}_5}, \quad (19.113)$$

with constants $c, c_2,$ and c_5 that are convention-dependent and immaterial for our discussion.[21] We immediately recognize the momentum and the charges associated with the M2- and M5-branes. The gravitational wave is associated with the momentum, but what is the charge associated with the KK monopole (KK7M)? Furthermore, is there a charge for the KK9M? As a matter of fact, as we have stressed repeatedly, these KK-branes are not states

[20] We found all the possibilities in $d=4$ in Section 5.4.1. In higher dimensions the analysis is almost identical.

[21] This formula can be interpreted as a decomposition of a symmetric bi-spinor into Lorentz tensors. A consistency check is provided by the counting of independent components on both sides of the equation: $33\times 32/2$. Physically, this formula should be understood as an inventory of possibilities: for instance, the superalgebra of $N=1, d=10$ supergravity admits quasi-central charges of ranks 1 and 5, but, physically, we expect on the r.h.s. one rank-5 and two rank-1 charges: momentum and the string charge. The counting on the two sides gives different results, but physically it is correct.

of the *uncompactified* $d = 11$ theory and appear only after compactification. Still, we can include them in the $d = 11$ superalgebra using dual charges *and* vectors k^a, and l^a that project the charges in the compact direction. The two terms that correspond to the KK7M and the KK9M and should be added are [666]

$$+\frac{c_6}{6!}(\hat{\hat{\Gamma}}^{\hat{\hat{a}}_1\cdots\hat{\hat{a}}_6}\hat{\hat{\mathcal{C}}}^{-1})^{\alpha\beta}\hat{\hat{\mathcal{Z}}}^{(7)}_{\hat{\hat{a}}_1\cdots\hat{\hat{a}}_7}k^{\hat{\hat{a}}_7}+\frac{c_9}{9!}(\hat{\hat{\Gamma}}^{\hat{\hat{a}}_1\cdots\hat{\hat{a}}_9}\hat{\hat{\mathcal{C}}}^{-1})^{\alpha\beta}\hat{\hat{\mathcal{Z}}}^{(8)}_{\hat{\hat{a}}_1\cdots\hat{\hat{a}}_8}l_{\hat{\hat{a}}_9}. \tag{19.114}$$

The dimensional reduction of the M algebra should give the $N = 2A, d = 10$ superalgebra with a charge for each of the known objects of this theory. We need only reduce the vector indices (as we did in the reduction of $d = 11$ supergravity). Each of the standard quasi-central charges gives two in $d = 10$: $\hat{\hat{P}}_{\hat{a}} = \hat{P}_{\hat{a}}$, $\hat{\hat{P}}_z = \hat{\mathcal{Z}}^{(0)}$, $\hat{\hat{\mathcal{Z}}}^{(2)}_{\hat{a}\hat{b}} = \hat{\mathcal{Z}}^{(2)}_{\hat{a}\hat{b}}$, $\hat{\hat{\mathcal{Z}}}^{(2)}_{\hat{a}z} = \hat{\mathcal{Z}}^{(1)}_{\hat{a}}$, etc. The non-standard ones give rise to three, for instance $\hat{\hat{\mathcal{Z}}}^{(7)}_{\hat{a}_1\cdots\hat{a}_7} = \hat{\mathcal{Z}}^{(7)}_{\hat{a}_1\cdots\hat{a}_7}$, $\hat{\hat{\mathcal{Z}}}^{(7)}_{\hat{a}_1\cdots\hat{a}_6 z} = \hat{\mathcal{Z}}^{(6)}_{\hat{a}_1\cdots\hat{a}_6}$, and $\hat{\hat{\mathcal{Z}}}^{(7)}_{\hat{a}_1\cdots z\hat{a}_6} = \hat{\mathcal{Z}}^{(6)}_{\hat{a}_1\cdots\hat{a}_6}$, corresponding, respectively, to the KK7A, the D6, and the KK6A. The KK6A is the standard KK monopole. The KK7A is the solution one obtains by reducing the KK7M (the M-theory KK monopole) in a genuine transverse direction (the harmonic function is smeared by the usual procedure and then one solves for the vector field in the metric [691]).

The result of the reduction is the $N = 2A, d = 10$ superalgebra generalized with the inclusion of KK-brane charges:

$$\begin{aligned}\left\{\hat{\mathcal{Q}}^\alpha, \hat{\mathcal{Q}}^\beta\right\} &= c(\hat{\Gamma}^{\hat{a}}\hat{\mathcal{C}}^{-1})^{\alpha\beta}\hat{P}_{\hat{a}} + \sum_{n=0,1,4,8}\frac{c_n}{n!}(\hat{\Gamma}^{\hat{a}_1\cdots\hat{a}_n}\hat{\Gamma}_{11}\hat{\mathcal{C}}^{-1})^{\alpha\beta}\mathcal{Z}^{(n)}_{\hat{a}_1\cdots\hat{a}_n}\\
&\quad+\sum_{n=2,5,6}\frac{c_n}{n!}(\hat{\Gamma}^{\hat{a}_1\cdots\hat{a}_n}\hat{\mathcal{C}}^{-1})^{\alpha\beta}\hat{\mathcal{Z}}^{(n)}_{\hat{a}_1\cdots\hat{a}_n}\\
&\quad+\frac{c_5}{5!}(\hat{\Gamma}^{\hat{a}_1\cdots\hat{a}_5}\hat{\Gamma}_{11}\hat{\mathcal{C}}^{-1})^{\alpha\beta}\hat{\mathcal{Z}}^{(6)}_{\hat{a}_1\cdots\hat{a}_5\hat{a}_6}\hat{k}^{\hat{a}_6}+\frac{c_6}{6!}(\hat{\Gamma}^{\hat{a}_1\cdots\hat{a}_6}\hat{\mathcal{C}}^{-1})^{\alpha\beta}\hat{\mathcal{Z}}^{(7)}_{\hat{a}_1\cdots\hat{a}_6\hat{a}_7}\hat{l}^{\hat{a}_7}\\
&\quad+\frac{c_8}{8!}(\hat{\Gamma}^{\hat{a}_1\cdots\hat{a}_8}\hat{\mathcal{C}}^{-1})^{\alpha\beta}\hat{\mathcal{Z}}^{(7)}_{\hat{a}_1\cdots\hat{a}_7}\hat{m}_{\hat{a}_8}+\frac{c_9}{9!}(\hat{\Gamma}^{\hat{a}_1\cdots\hat{a}_9}\hat{\mathcal{C}}^{-1})^{\alpha\beta}\hat{\mathcal{Z}}^{(8)}_{\hat{a}_1\cdots\hat{a}_8}\hat{n}_{\hat{a}_9}.\end{aligned} \tag{19.115}$$

Let us now turn to the $N = 2B, d = 10$ superalgebra. It contains an SO(2) pair of chiral supercharges labeled by $i, j = 1, 2$ and the charges that appear on the r.h.s. of their anti-commutator carry a pair symmetric or antisymmetric in these indices. The allowed ranks for antisymmetric indices are 3 and 7 and those for symmetric indices are 1,5, and 9. The charges with antisymmetric indices are proportional to σ^2 and those with symmetric indices can be decomposed into a basis of symmetric 2×2 matrices: $\mathbb{I}$, σ^1 and σ^3. The charges proportional to σ^2 and $\mathbb{I}$ are invariant under SO(2), and charges proportional to σ^1 and σ^3 form SO(2) doublets. Combining the latter into symmetric traceless charges denoted by (ij), the algebra is usually written in the form

$$\begin{aligned}\left\{\hat{\mathcal{Q}}^{i\,\alpha}, \hat{\mathcal{Q}}^{j\,\beta}\right\} &= c\delta^{ij}(\hat{\Gamma}^{\hat{a}}\hat{\mathcal{C}}^{-1})^{\alpha\beta}\hat{P}_{\hat{a}} + c_1(\hat{\Gamma}^{\hat{a}}\hat{\mathcal{C}}^{-1})^{\alpha\beta}\hat{\mathcal{Z}}^{(1)(ij)}_{\hat{a}}\\
&\quad+\frac{c_3}{3!}(\sigma^2)^{ij}(\hat{\Gamma}^{\hat{a}_1\hat{a}_2\hat{a}_3}\hat{\mathcal{C}}^{-1})^{\alpha\beta}\hat{\mathcal{Z}}^{(3)}_{\hat{a}_1\hat{a}_2\hat{a}_3}\\
&\quad+\frac{c_5}{5!}\delta^{ij}(\hat{\Gamma}^{\hat{a}_1\cdots\hat{a}_5}\hat{\mathcal{C}}^{-1})^{\alpha\beta}\hat{\mathcal{Z}}^{(5)}_{\hat{a}_1\cdots\hat{a}_5}+\frac{c_5}{5!}(\hat{\Gamma}^{\hat{a}_1\cdots\hat{a}_5}\hat{\mathcal{C}}^{-1})^{\alpha\beta}\hat{\mathcal{Z}}^{(5)(ij)}_{\hat{a}_1\cdots\hat{a}_5},\end{aligned} \tag{19.116}$$

where it is understood that the r.h.s. has to be projected over the positive-chirality subspace.

SL(2, $\mathbb{R}$) acts on the spinors through SO(2) rotations and, therefore, $\hat{\mathcal{Z}}^{(1)(ij)}$ and $\hat{\mathcal{Z}}^{(5)(ij)}$ correspond to the two doublets of strings (D1 and F1B) and 5-branes (D5 and S5B). $\hat{\mathcal{Z}}^{(3)}$ corresponds to the S-duality-invariant D3. There is no invariant 5-brane and $\hat{\mathcal{Z}}^{(5)}_{\hat{a}_1\cdots\hat{a}_5}$ should be replaced by $\hat{\mathcal{Z}}^{(6)}_{\hat{a}_1\cdots\hat{a}_5\hat{a}_6}\hat{k}^{\hat{a}_6}$ associated with the KK6B. To these charges one should add $\hat{\mathcal{Z}}^{(9)(ij)}$ for the D9–S9 doublet and two $\hat{\mathcal{Z}}^{(7)}$s for the D7–S7 doublet[22] if we are to relate this superalgebra to the $N = 2A, d = 10$ by T duality, as we should expect. As Figures 19.4.1 and 19.4.1 show, more charges with more auxiliary vectors need to be included if one really wants to have complete agreement and consistency with all the dualities conjectured.

19.5.1 Unbroken supersymmetries of string-theory solutions

If we follow now the reasoning of Section 13.5.1, we arrive at the conclusion that the quantum theory based on the above superalgebras admits states with momentum $P_0 = T_{(p)}$ (the tension) and the quasi-central charge $\mathcal{Z}^{(p)}_{1\cdots p} = T_{(p)}$ corresponding to extended objects that are invariant under the supersymmetry transformations generated by spinors that satisfy constraints of the form Eq. (13.120) that we can rewrite in the form

$$(\mathbb{I} \pm \Gamma^{01\cdots p}\mathcal{O})\epsilon = 0, \tag{19.117}$$

where $\mathcal{O}$ is an operator that depends on the theory and the state. In all cases $\Gamma^{01\cdots p}\mathcal{O}$ is a traceless operator that squares to the identity and half of its eigenvalues are $+1$ and the other half -1. Therefore $\frac{1}{2}(\mathbb{I} \pm \Gamma^{01\cdots p}\mathcal{O})$ is a projector that eliminates half of the components of ϵ. There is, therefore, a half-supersymmetric state for each quasi-central charge in the above superalgebras and we expect the associated solutions of the supergravity theories (which will be extreme solutions) to have unbroken supersymmetries generated by Killing spinors that satisfy the same constraints. Let us see some examples.

Unbroken supersymmetries of the M2-brane. We are going to work out in detail this example to illustrate how the Killing-spinor equations are usually solved. The rest of the examples follow the same pattern and we will give only the results.

The $d = 11$ SUGRA Killing-spinor equations are $\delta_{\hat{\hat{\kappa}}}\hat{\hat{\psi}}_{\hat{\hat{\mu}}} = 0$ with $\delta_{\hat{\hat{\kappa}}}\hat{\hat{\psi}}_{\hat{\hat{\mu}}}$ given by Eq. (16.8). We just have to substitute into it the spin connection and 4-form components of the M2 solution Eq. (19.45). Choosing the Elfbeins

$$\hat{\hat{e}}_{\underline{i}}{}^{j} = H_{\mathrm{M2}}^{-\frac{1}{3}}\delta_i{}^j, \qquad \hat{\hat{e}}_{\underline{m}}{}^{n} = H_{\mathrm{M2}}^{\frac{1}{6}}\delta_m{}^n, \tag{19.118}$$

and using the results of Appendix F.2.4, we find the non-vanishing components

$$\begin{aligned}
\hat{\hat{\omega}}_{\underline{m}}{}^{nl} &= -\tfrac{1}{3}H_{\mathrm{M2}}^{-1}\partial_{\underline{q}}H_{\mathrm{M2}}\eta_m{}^{[n}\eta^{l]q},\\
\hat{\hat{\omega}}_{\underline{i}}{}^{mj} &= \tfrac{2}{3}H_{\mathrm{M2}}^{-\frac{3}{2}}\partial_{\underline{q}}H_{\mathrm{M2}}\eta_i{}^{[m}\eta^{j]q},\\
\hat{\hat{G}}_{mijk} &= \mp\epsilon_{ijk}\,H_{\mathrm{M2}}^{-\frac{7}{6}}\partial_{\underline{m}}H_{\mathrm{M2}},
\end{aligned} \tag{19.119}$$

[22] Each of them is SO(2) invariant, but we may assume that they are interchanged by S duality. The situation is still not completely clear since, as we have stressed before, there is an infinite number of pq 7-branes, not just a doublet, and this is difficult to reflect in the superalgebra.

and substituting, and assuming that $\partial_{\underline{i}}\hat{\hat{\kappa}} = 0$, we find the equations

$$\begin{aligned}
\delta_{\hat{\hat{\kappa}}}\hat{\hat{\psi}}_{\underline{i}} &= \tfrac{1}{3}H_{\mathrm{M2}}^{-\frac{3}{2}}\partial_{\underline{n}}H_{\mathrm{M2}}\hat{\hat{\Gamma}}_{(i)}{}^{n}\left(1 \mp \frac{i}{2}\epsilon_{(i)jk}\hat{\hat{\Gamma}}^{(i)jk}\right)\hat{\hat{\kappa}} = 0,\\
\delta_{\hat{\hat{\kappa}}}\hat{\hat{\psi}}_{\underline{m}} &= 2\partial_{\underline{m}}\hat{\hat{\kappa}} - \tfrac{1}{6}H_{\mathrm{M2}}^{-1}\partial_{\underline{n}}H_{\mathrm{M2}}\left[\hat{\hat{\Gamma}}^{mn} \mp i\left(\hat{\hat{\Gamma}}^{mn} + 2\delta^{mn}\right)\hat{\hat{\Gamma}}^{012}\right]\hat{\hat{\kappa}} = 0.
\end{aligned} \tag{19.120}$$

The first equation is purely algebraic and can be solved only if $\hat{\hat{\kappa}}$ satisfies the constraint

$$\left(1 \mp i\hat{\hat{\Gamma}}^{012}\right)\hat{\hat{\kappa}} = 0. \tag{19.121}$$

Using this constraint in the second equation, it takes the form

$$2\left(\partial_{\underline{m}} + \tfrac{1}{6}H_{\mathrm{M2}}^{-1}\partial_{\underline{m}}H_{\mathrm{M2}}\right)\hat{\hat{\kappa}} = 0, \tag{19.122}$$

whose solution is $\hat{\hat{\kappa}} = H_{\mathrm{M2}}^{-\frac{1}{6}}\hat{\hat{\kappa}}_0$, where $\hat{\hat{\kappa}}_0$ is a constant spinor. The M2 Killing spinors are, therefore,

$$\boxed{\hat{\hat{\kappa}} = H_{\mathrm{M2}}^{-\frac{1}{6}}\,\hat{\hat{\kappa}}_0, \qquad \left(1 \mp i\hat{\hat{\Gamma}}^{012}\right)\hat{\hat{\kappa}}_0 = 0.} \tag{19.123}$$

The constraint has the form Eq. (19.117) predicted by the supersymmetry algebra and therefore only half of the components of $\hat{\hat{\kappa}}_0$ are independent and only half of the supersymmetries are unbroken. We have included the two possible signs of the M2 charge. They are irrelevant for a single brane but may be crucial in the presence of other branes.

Observe that the Killing spinor exists for *any* function H_{M2} (not necessarily harmonic!). The Killing-spinor equations do not imply the equations of motion that restrict H_{M2} to be harmonic. On the other hand, strictly speaking, the solution is supersymmetric only if the Killing spinors have the correct asymptotic behavior: if the solution is asymptotically the vacuum, the Killing spinors have to approach the vacuum Killing spinors asymptotically. Furthermore, they have to be normalizable. These conditions are satisfied if H_{M2} is the harmonic function that describes parallel M2-branes. If H_{M2} corresponds to the $\mathrm{AdS}_4 \times \mathrm{S}^7$ solution, which is a vacuum solution itself, the asymptotic behavior is right by definition. Furthermore, the Killing-spinor equation can be solved for spinors that do not satisfy the constraint by introducing dependence on the "worldvolume" coordinates and the solution is maximally supersymmetric. The group-theoretical methods explained in Chapter 13 are better suited for solving the equation.

Unbroken supersymmetries of the M5-brane. An entirely analogous calculation gives

$$\boxed{\hat{\hat{\kappa}} = H_{\mathrm{M5}}^{-\frac{1}{12}}\,\hat{\hat{\kappa}}_0, \qquad \left(1 \mp \hat{\hat{\Gamma}}^{012345}\right)\hat{\hat{\kappa}}_0 = 0.} \tag{19.124}$$

As in the M2 case, in the near-horizon limit (or, equivalently, on choosing the H_{M5} that describes the $\mathrm{AdS}_7 \times \mathrm{S}^4$ solution) the spinors that do not satisfy the constraint also solve the Killing-spinor equation and the solution is maximally supersymmetric.

Unbroken supersymmetries of KK monopoles. The KK-monopole solution, in all theories, is the direct product of the $d = 4$ Euclidean Taub–NUT solution and $(d-4)$-dimensional Minkowski spacetime and can be seen as a $(d-5)$-brane. In all cases, the Killing-spinor equations reduce to

$$\nabla_{\underline{m}}\kappa = 0, \qquad m = (d-4), (d-3), (d-2), (d-1), \tag{19.125}$$

since the solution is trivial in the worldvolume directions. As discussed on page 399, in the frame in which the spin connection is (anti-)self-dual, Eq. (9.43), the Killing spinors are constant spinors satisfying the constraint $(1 \pm \Gamma^{(d-4)(d-3)(d-2)(d-1)})\kappa = 0$, which can be rewritten in the form Eq. (19.117) with $p = d-5$ and $\mathcal{O}$ depending on the specific theory. For $N = 1, d = 11$, $N = 2A, d = 10$, and $N = 2B, d = 10$ (with self-dual RR 5-form), respectively, we have

$$\boxed{\left(1 \mp i\hat{\hat{\Gamma}}{}^{0123456}\right)\hat{\hat{\kappa}} = 0 \qquad \left(1 \mp \hat{\Gamma}^{012345}\hat{\Gamma}_{11}\right)\hat{\kappa} = 0, \qquad \left(1 \pm \hat{\Gamma}^{012345}\right)\hat{\kappa} = 0.} \tag{19.126}$$

Unbroken supersymmetries of the Dp-branes. Taking into account that the NSNS 3-form field strength is zero for these solutions, and that only the field strength $\hat{G}^{(p+2)}$ and its dual, whose combinations add up, is different from zero, the $N = 2A$ and $N = 2B$ Killing-spinor equations can be written in the following unified way:

$$\begin{aligned}
\delta_{\hat{\kappa}}\hat{\psi}_{\hat{\mu}} &= \left\{\partial_{\hat{\mu}} - \tfrac{1}{4}\,\hat{\not{\omega}}_{\hat{\mu}} + \tfrac{1}{8}e^{\hat{\phi}}\frac{1}{(p+2)!}\,\hat{\not{G}}^{(p+2)}\hat{\Gamma}_{\hat{\mu}}\mathcal{O}_{\mathrm{D}p}\right\}\hat{\kappa}, \\
\delta_{\hat{\kappa}}\hat{\lambda} &= \left\{\not{\partial}\hat{\phi} - \tfrac{1}{4}e^{\hat{\phi}}\frac{p-3}{(p+2)!}\,\hat{\not{G}}^{(p+2)}\mathcal{O}_{\mathrm{D}p}\right\}\hat{\kappa},
\end{aligned} \tag{19.127}$$

where

$$\mathcal{O}_{\mathrm{D}p} = i(-\hat{\Gamma}_{11})^{\frac{p+2}{2}}, \quad p \ \text{odd (IIA)}, \qquad \mathcal{O}_{\mathrm{D}p} = \mathcal{P}_{\frac{p+3}{2}}, \quad p \ \text{even (IIB)}, \tag{19.128}$$

where $\mathcal{P}_n$ is defined in Eq. (17.11).

With this notation, the Killing spinors are given by

$$\boxed{\hat{\kappa} = H_{\mathrm{D}p}^{-\frac{1}{8}}\hat{\kappa}_0, \qquad \left(1 \mp \hat{\Gamma}^{01\cdots p}\mathcal{O}_{\mathrm{D}p}\right)\hat{\kappa}_0 = 0.} \tag{19.129}$$

Unbroken supersymmetries of the fundamental string. Since, for this solution, the RR potentials vanish, we can write the Killing-spinor equations for the $N = 2$ and $N = 1$ theories in the unified form

$$\begin{aligned}
\delta_{\hat{\kappa}}\hat{\psi}_{\hat{\mu}} &= \left\{\partial_{\hat{\mu}} - \tfrac{1}{4}\left(\hat{\not{\omega}}_{\hat{\mu}} + \tfrac{1}{2}\,\hat{\not{H}}_{\hat{\mu}}\mathcal{O}\right)\right\}\hat{\kappa} = 0, \\
\delta_{\hat{\kappa}}\hat{\lambda} &= \left\{\not{\partial}\hat{\phi} - (1/12)\,\hat{\not{H}}\mathcal{O}\right\}\hat{\kappa} = 0,
\end{aligned} \tag{19.130}$$

where $\hat{\kappa}$ is a Majorana spinor, a pair of Majorana–Weyl spinors, or a Majorana–Weyl spinor and $\mathcal{O}_{\rm F1} = \hat{\Gamma}_{11}, \sigma^3,$ and $\mathbb{I}$, for the $N = 2A, 2B,$ and 1 theories, respectively. The solutions are given by

$$\boxed{\hat{\kappa} \;=\; H_{\rm F1}^{\frac{1}{4}}\kappa_0, \qquad \left(1 \pm \hat{\Gamma}^{01}\mathcal{O}_{\rm F1}\right)\hat{\kappa}_0 \;=\; 0.} \tag{19.131}$$

Unbroken supersymmetries of the solitonic 5-brane. The $N = 2A, 2B, 1, d = 10$ cases can also be treated in a unified way. The result is

$$\boxed{\hat{\kappa} \;=\; \hat{\kappa}_0, \qquad \left(1 \pm \hat{\Gamma}^{0\cdots 5}\mathcal{O}_{\rm S5}\right)\hat{\kappa}_0 \;=\; 0.} \tag{19.132}$$

where now $\mathcal{O}_{\rm S5} = \sigma^3$ for the $N = 2B$ theory and $\mathbb{I}$ in the other two cases.

Observe that, up to possible $\hat{\Gamma}_{11}$ factors, this is essentially the projector for a metric of SU(2) holonomy (see the discussion on page 399) in the frame in which the spin connection is (anti-)self-dual (because the Killing spinor is constant). Actually, the Killing-spinor equation for the S5 can be seen as the condition of covariant constancy of the spinor with respect to the torsion spin connection $\hat{\Omega}^{(\pm)}_{\hat{\mu}} = \hat{\omega}_{\hat{\mu}} \pm \hat{\not{H}}_{\hat{\mu}}$ in the proper subspaces of $\mathcal{O}$. The torsionful spin connection for the S5 is *identical* to that of the BPST instanton given in Section 9.2.2.

T duality of the Killing spinors. Several of the solutions whose Killing spinors we have calculated are related by T duality and so must be the Killing spinors themselves. The T-duality transformation rules for the Killing spinors are given in Eqs. (17.40) and here we are going to check them on Dp-brane Killing spinors. Clearly, the H-dependent factor plays no role and we are going to focus on the projectors.

The D0-brane Killing spinor satisfies

$$\left(1 \pm i\hat{\Gamma}^0\hat{\Gamma}_{11}\right)\hat{\kappa}_{\rm D0} = 0, \qquad \Rightarrow \hat{\Gamma}_{11}\hat{\kappa}_{\rm D0} = \pm i\hat{\Gamma}^0\hat{\kappa}_{\rm D0}. \tag{19.133}$$

A T-duality transformation in the ninth direction gives the following two Majorana–Weyl Killing spinors of the $N = 2B$ theory

$$\hat{\kappa}^1_{\rm D1} = -\frac{i}{2}\hat{\Gamma}^9(1 + \hat{\Gamma}_{11})\hat{\kappa}_{\rm D0}, \qquad \hat{\kappa}^2_{\rm D1} = \tfrac{1}{2}\hat{\Gamma}^9(1 - \hat{\Gamma}_{11})\hat{\kappa}_{\rm D0}. \tag{19.134}$$

Using the D0 constraint above, we can rewrite them in the form

$$\hat{\kappa}^1_{\rm D1} = \mp\hat{\Gamma}^{09}\tfrac{1}{2}(1 \mp i\hat{\Gamma}^0)\hat{\kappa}_{\rm D0}, \quad \hat{\kappa}^2_{\rm D1} = \tfrac{1}{2}(1 \mp i\hat{\Gamma}^0)\hat{\kappa}_{\rm D0}, \quad \Rightarrow (\mathbb{I} \pm \hat{\Gamma}^{09}\sigma^1)\hat{\kappa}_{\rm D1} = 0, \tag{19.135}$$

which is the constraint of the D1 Killing spinor. The dependence on H is clearly correct.

Another T-duality transformation in the eighth direction gives, using the constraint of $\hat{\kappa}_{\rm D1}$,

$$\hat{\kappa}_{\rm D2} = \hat{\kappa}^2_{\rm D1} - i\hat{\Gamma}^8\hat{\kappa}^1_{\rm D1} \pm i\hat{\Gamma}^{098}\left(\hat{\kappa}^2_{\rm D1} - i\hat{\Gamma}^8\hat{\kappa}^1_{\rm D1}\right), \qquad \Rightarrow \left(1 \mp i\hat{\Gamma}^{098}\right)\hat{\kappa}_{\rm D2} = 0, \tag{19.136}$$

which is the D2-brane Killing-spinor algebraic constraint.

Maximally supersymmetric vacua of string and M theories. We have mentioned that the $\mathrm{AdS}_4 \times \mathrm{S}^7$ and $\mathrm{AdS}_7 \times \mathrm{S}^4$ solutions of $d = 11$ supergravity Eqs. (19.54) and Eqs. (19.47) are maximally supersymmetric solutions and, therefore, vacua of the theory. The metrics of these spaces are products of those of symmetric spaces and the Killing spinors and symmetry superalgebras can be constructed and studied using the methods of Chapter 13 (see [25]). The superalgebras are extended AdS superalgebras of the kind we studied in Section 5.4, written in 11-dimensional notation. These are also the superalgebras of the gauged SUEGRAs one obtains by compactification on S^7 and S^4.

These are not the only maximally supersymmetric solutions of $d = 11$ supergravity since we can always take the Penrose limit of any solution while preserving (or increasing) the number of supersymmetries [158, 160, 495, 764]. The Penrose limits of the above two vacua give the same KG11 solution (first found in [636]) which has an Hpp-wave metric of the form Eq. (10.18) with

$$\hat{\hat{G}}_{\underline{u}\underline{x}^1\cdots\underline{x}^3} = \lambda, \qquad A_{ij} = \begin{cases} -\dfrac{1}{18}\lambda^2\delta_{ij} & i, j = 1, 2, 3, \\ -\dfrac{1}{72}\lambda^2\delta_{ij} & i, j = 4, \ldots, 9. \end{cases} \tag{19.137}$$

The symmetry superalgebra of this solution was studied in [392]. It does not seem to be associated with any known supergravity. The same happens for the other KG solutions.

There are no more maximally supersymmetric vacua in $d = 11$ [393, 636]. Let us turn now to the ten-dimensional theories. In the $N = 2A, 1$ cases the only maximally supersymmetric vacuum is Minkowski spacetime [393]. In the $N = 2B$ theory there is, as we have seen, a maximally supersymmetric solution with the metric of $\mathrm{AdS}_5 \times \mathrm{S}^5$ Eq. (19.67). The superalgebra is that of gauged $N = 4, d = 5$ SUEGRA with gauge group SO(6), but it is naturally written in ten-dimensional notation. The Penrose limit gives the maximally supersymmetric KG10 solution [159] which also has an Hpp-wave metric of the form Eq. (10.18) with

$$\hat{G}^{(5)}_{\underline{u}\underline{x}^1\cdots\underline{x}^4} = \hat{G}^{(5)}_{\underline{u}\underline{x}^5\cdots\underline{x}^8} = \lambda, \qquad A_{ij} = -\tfrac{1}{2}\delta_{ij}\lambda^2, \quad i, j, = 1, \ldots, 8, \tag{19.138}$$

in our conventions.

There are no more maximally supersymmetric vacua in $d = 10$ [393], but there are other vacua with fewer supersymmetries, which are perhaps more interesting from a phenomenological point of view. We have mentioned some of them (those that can be obtained by replacing the spheres by other Einstein spaces). A complete classification is still lacking, but there is currently intense work in this direction (see, for instance, [423, 582]).

19.6 Intersections

Although now it may seem natural to look for solutions that represent simultaneously branes of different kinds in equilibrium, the first solutions of that kind were only identified in [753] among general solutions found years before in [494]. After that, very many solutions were quickly constructed and the basic rules that govern their existence studied. Trying to review all these solutions and the various approaches in depth in the space available would be utterly hopeless. Thus, we will be pragmatic, focusing on the simplest families of solutions and the general rules. More information can be obtained from reviews such as [417, 588, 858].

The solutions we have studied so far describe p-branes at rest, in their lower energy states in which none of their worldvolume fields is excited and has a non-trivial configuration. We have checked this by matching the solutions with p-brane sources in Sections 18.2.3 and 19.2.1. Excited worldvolume fields describe, first of all, deformations of the p-brane in the target spacetime when they involve the embedding coordinates $X^{\mu}(\xi)$ (as happens in all the supersymmetric cases). These deformations can be seen, in certain cases, as other branes that end on or intersect the original brane. The converse relation is always true: a brane that ends on or intersects another brane always corresponds to an excitation of the worldvolume fields of the latter. The dimension of the intersection, which behaves as a dynamical solitonic object in the worldvolume of the *host* brane, determines the nature of the intersecting brane. That dimension is associated with the rank of the excited worldvolume differential-form fields: k-brane intersections to $(k+1)$-form worldvolume fields. There are three main examples of this correspondence.

1. By definition, open strings end on Dp-branes and their endpoints are seen as point-particles electrically charged w.r.t. the BI vector field (*BIons*). Worldvolume half-supersymmetric solutions of the Dp-brane action in flat spacetime describing a point-like electric charge were found in [209, 434] (see also [502, 646]. There is always an excited embedding scalar that corresponds to a *spike* sticking out of the Dp-brane. The energy of this solitonic worldvolume solution per unit length of the spike is precisely equal to the fundamental-string tension. Furthermore, perturbations along the spike have Dirichlet boundary conditions and one concludes that this solution represents a fundamental string attached to the Dp-brane.

   Since these are supersymmetric worldvolume solutions, it is not surprising that there are solutions describing several parallel (or antiparallel) spikes in equilibrium.

   The BI vector field can be dualized into a $(p-2)$-form potential, which is another BI vector for the D3-brane [450, 451, 469]. The D3 electric–magnetic self-duality is related to type-IIB S duality and the dual BIons turn out to describe D-strings ending on a D3-brane.[23] The dual $(p-3)$-BIons of other Dp-branes are related to this one by T duality: a D$(p-2)$ ending on a Dp with a $(p-3)$-dimensional intersection associated with the $(p-2)$-form dual of the BI vector field.

   These intersections can be written in the form F1 $\perp$ D$p(0)$ and D$(p-2) \perp$ D$p(p-3)$.

2. In [567] a supersymmetric worldvolume solution of the M5 equations of motion in flat spacetime [15, 80, 119, 570, 757] in which two embedding scalars and the self-dual 2-form[24] were excited was found. It corresponds to the intersection

[23] If the D-string ended on N coincident D3-branes, whose worldvolume field theory, contains a non-Abelian SU(N) BI vector field (in fact, it is a non-linear generalization of $N=4$, $d=4$ super-Yang–Mills theory), the intersection would be seen as an SU(N) magnetic monopole in the worldvolume [321, 469].

[24] The bosonic worldvolume fields of the string/M-theory extended objects can be found in Table 19.6.

Table 19.4. Elementary intersections of ten-dimensional extended objects.

F1 $\parallel$ S5, F1 $\perp$ D$p(0)$,
S5 $\perp$ S5(1), S5 $\perp$ S5(3), S5 $\perp$ D$p(p-1)$ $(p>1)$,
D$p \perp$ D$p'(m)$, $p+p'=4+2m$,
W $\parallel$ F1, W $\parallel$ S5, W $\parallel$ Dp,
KK6 $\perp$ D$p(p-2)$

Table 19.5. Elementary intersections of 11-dimensional extended objects.

M2 $\perp$ M2(0), M2 $\perp$ M5(1), M5 $\perp$ M5(1), M5 $\perp$ M5(3),
W $\parallel$ M2, W $\parallel$ M5,
KK7M $\parallel$ M2, KK7M $\perp$ M2(0), KK7M $\parallel$ M5, KK7M $\perp$ M5(1), KK7M $\perp$ M5(3),
W $\parallel$ KK, W $\perp$ KK7M(2), W $\perp$ KK7M(4)

M2 $\perp$ M5(1). The dimensional reduction along the intersection corresponds to F1A $\perp$ D4(0), which was discussed above.

3. A solution describing the supersymmetric M5 worldvolume soliton associated with the intersection M5 $\perp$ M5(3) was constructed in [568]. The worldvolume gauge field is here the dual of an embedding scalar. These are present in any p-brane with $p < d-1$ and are $(p-1)$-forms. They indicate the possibility of two p-branes intersecting over a $(p-2)$-brane.

   Indeed, on T-dualizing the D1 $\perp$ D3(0) in a direction parallel to the D3 and perpendicular to the D1, we find D2 $\perp$ D2(0). T duality in directions transverse to both branes generates another sequence of possible intersections, D$p \perp$ D$p(p-2)$.

   Had we dualized in a direction perpendicular to the D3 and parallel to the D1, we would have generated D0 $\perp$ D4(0) and then further T dualities would have generated the sequence D$p \perp$ D$p+4(p)$.

It is clear that we can go on generating new intersections via dualities. The results, in terms of supergravity solutions (including gravitational waves and KK monopoles [116, 669]), are summarized in Tables 19.4 and 19.5. Some of the intersections (named *overlaps* in [421]) cannot be associated with excited worldvolume fields. They arise, in fact, in degenerate limits of intersections involving more than two branes. For instance, the M5 $\perp$ M5(1) intersection corresponds to an M2 ending on two M5s in a limit in which these become infinitely close and the M2 disappears.

As we have mentioned, these intersections, seen as excited worldvolume configurations (branes within branes), always preserve some supersymmetry. Actually, in general

Table 19.6. Bosonic worldvolume fields of string/M-theory branes. S and T are worldvolume scalars, V_i is a worldvolume vector, and V^+_{ij} is a 2-form with self-dual field strength.

Object	Worldvolume dimension	Worldvolume fields
M2	$2+1$	X^μ
M5	$5+1$	X^μ, V^+_{ij}
KK7M	$6+1$	X^μ, V_i
S5A	$5+1$	X^μ, V^+_{ij}, S
KK6B	$5+1$	X^μ, V^+_{ij}, S, T
S5B	$5+1$	X^μ, V_i
KK6A	$5+1$	X^μ, V_i, S
Dp	$p+1$	X^μ, V_i

excitations,[25] no supersymmetry would be preserved, but we are interested in the cases in which some supersymmetry is preserved, in part because they are easier to deal with and we can also expect to find a classical (supersymmetric) solution associated with that brane configuration. Worldvolume supersymmetry and the worldvolume superalgebras have been used to study the possible intersections [124, 419].

What is the spacetime version of these worldvolume arguments? As we have seen, the low-energy effective actions are very powerful tools with which to study extended objects that arise as classical solutions. Supersymmetric solutions describing single branes can typically be related to elementary brane sources and we expect the same to be true for intersecting brane sources, although not many results have been obtained in this direction for specific solutions. However, one can use general arguments based on the field equations (charge conservation [53, 869, 900]) and spacetime supersymmetry to determine which intersections are allowed. Then, one can try to find the corresponding supergravity solutions.

Let us review these arguments.

19.6.1 Brane-charge conservation and brane surgery

Following [900], let us consider the charge carried by a F1B solution. In the absence of any other object (i.e. with only the $\hat{\mathcal{B}}$ potential excited), the charge is given by the first of Eqs. (19.102) and it is different from zero only if the string has no free endpoints at a finite distance since, otherwise, we could slide the S^7 on which we integrate along the string beyond its endpoint and contract it to a point[26] without encountering any singularity (source) because (this is just the $\hat{\mathcal{B}}$ equation of motion)

$$d\left(e^{-2\hat{\varphi}\star}\hat{\mathcal{H}}\right) = 0 \qquad (19.139)$$

[25] There are non-supersymmetric BIon solutions with no scalars excited, and, therefore, such solutions are not associated with deformations of the worldvolume [434].

[26] This can be visualized best in $d = 4$, in which S^7 is replaced by S^1.

outside the string in the presence of other fields, the equation of motion has additional terms and the homotopy-invariant definition of charge is

$$q_{\rm F1} \sim \int_{S^7} \left(e^{-2\hat{\varphi}\star}\hat{\mathcal{H}} - {}^\star\hat{G}^{(3)}\hat{C}^{(0)} - \hat{G}^{(5)}\hat{C}^{(2)}\right), \tag{19.140}$$

where S^7 surrounds the string. Let us consider a semi-infinite string. At a large enough distance L from the endpoint, boundary effects are not important and the charge is still approximately given by the first of Eqs. (19.102). The larger L is for a fixed value of the S^7 radius R_7, the better the approximation. Closer to the endpoint, the additional terms must contribute (otherwise, we are back in the previous case), but we can obtain the same value for the integral by making $R_7 \to 0$ keeping R_7/L constant until the only contribution to the integral comes from the endpoint. The degenerate S^7 can be decomposed, for convenience, into the product $S^5 \times S^2$, if we assume that the contribution to $q_{\rm F1B}$ comes from the last term in the above integral. The integral decomposes into a product of integrals,

$$\int_{S^5} \hat{G}^{(5)} \int_{S^2} \hat{C}^{(2)}. \tag{19.141}$$

The first integral gives the D3-brane charge (assuming, as we are doing here, that $\hat{\mathcal{H}}$ does not contribute), $\hat{G}^{(5)} = {}^\star\hat{G}^{(5)}$, and thus the string endpoint must be at a D3-brane. If there is no D1-brane present, then, $\hat{G}^{(3)} \sim d\hat{C}^{(2)} = 0$ inside the D3-brane and, *locally*, $\hat{C}^{(2)} = dV$, where V is a vector that lives in the D3-brane worldvolume. Then

$$q_{\rm F1B} \sim \int_{S^2} dV. \tag{19.142}$$

The interpretation is clear: an F1B can end on a D3-brane and at the intersection point there is an excited worldvolume vector field (the dual BI vector field) whose magnetic charge is proportional to the F1B charge. This is the same result as we obtained before.

A similar reasoning indicates that, if it is the second term that contributes to the charge integral, the F1B can also end on a D-string and at the intersection the BI vector field is excited so its dual field strength is a constant.[27]

These arguments that determine the opening of branes seem to depend on field redefinitions (the charge integrand is defined up to total derivatives). However, the different expressions for the charge are just choices that are more or less adequate to describe a given physical situation. The most symmetric expression for $q_{\rm F1B}$ can be obtained by using Eq. (17.9). Each of the four possible terms corresponds to the F1B ending on one of the four Dp-branes $p = 1, 3, 5,$ and 7 and exciting the dual BI field magnetically.

19.6.2 Marginally bound supersymmetric states and intersections

We are considering only supersymmetric brane intersections, in which the branes that intersect do not interact and are in supersymmetric equilibrium. These intersections can be

[27] This is similar to viewing the mass parameter of Romans' $N = 2A$, $d = 10$ SUEGRA as the dual of the RR 10-form field strength.

considered as bound states with zero binding energy (or *marginally bound* states) and their existence depends on whether it is possible to impose the simultaneous annihilation of that state by the supercharges that annihilate those associated with each individual brane.

As we saw, the annihilation of a p-brane state by a given set of supercharges is entirely equivalent to the action of a projector P_p of the generic form Eq. (19.117) on a spinor, $P_p\epsilon = 0$. Then, the existence of a supersymmetric state composed of a p- and a p'-brane depends on the compatibility of the respective projectors P_p and $P_{p'}$: it will exist if

$$[P_p, P_{p'}] = 0, \tag{19.143}$$

and the state will preserve a quarter of the supersymmetries.

This equation depends on p and p' but also on the spatial orientation of the branes. A general analysis is complicated because of the different $\mathcal{O}_p$ that occur in the projectors. Let us consider a simple example first: two p-branes of the same kind, S5A for simplicity, extended along five Cartesian coordinates (so they are either parallel or orthogonal). It is relatively easy to see that the two associated projectors commute if the number of relative transverse dimensions (those which are parallel to one brane and transverse to the other) is 0 mod 4, which leads to the allowed (supersymmetric) intersections S5 $\perp$ S5(3) and S5 $\perp$ S5(1), which are included in Table 19.4. For Dp-branes $\mathcal{O}_p = i\mathbb{I}, i\Gamma_{11}, \sigma^1, i\sigma^2$ depend on p mod 4 and the analysis of intersections between Dp-branes gives the allowed intersection D$p \perp$ D$(p+4)(p)$ and, with a little more effort, the other cases in the table [52, 346].

This analysis, which is essentially based on the spacetime supersymmetry algebra, allows the study of more complicated intersections involving more branes [115] or non-orthogonal intersections (*branes at angles* [128, 143, 182, 852, 904]). The inclusion of another brane is allowed if its associated projector commutes with the other ones. The amount of unbroken supersymmetry is generically halved each time a brane is included, except in the case in which the projector of the additional brane does not impose any new constraint on the spinor. The canonical example is that of a D5 in the directions 12345 and an S5B in the directions 12678, so they intersect in two directions. Their associated projectors $P_{\mathrm{D5}}(12345)$ and $P_{\mathrm{S5B}}(12679)$ (in the obvious notation) commute, and

$$P_{\mathrm{D5}}(12345)\,P_{\mathrm{S5B}}(12679)\epsilon \sim P_{\mathrm{D3}}(129)\epsilon = 0, \tag{19.144}$$

and, thus, including a D3 in the directions 129 does not break any additional supersymmetry (for any sign of the charge). This property gives rise to the phenomenon of D3-brane creation when a D5 and an S5B cross [65, 107, 290, 370, 497]. Adding a fourth brane may break all or no additional supersymmetry, depending on its charge's sign (Section 20.1).

Another case in which no additional constraint is imposed is when we add a brane of the same kind, but rotated by a supersymmetry-preserving angle ("branes at angles") [143]. We have no space to review this important case and we refer the reader to the literature.

19.6.3 Intersecting-brane solutions

The worldvolume and spacetime arguments that we have reviewed above suggest that classical solutions of the low-energy effective string/M theories describing intersecting-branes should exist. In this section we are going to study the simplest intersecting-brane solutions. These (like most of the solutions that have been found so far) are actually "imperfect" and represent intersecting branes (even in the cases in which we expect a brane ending on another brane) that are partially delocalized, smeared along the relative transverse directions. These intersecting-brane solutions have to be understood, then, as approximations to the true field configurations, but are, nevertheless, worth studying.

The construction of these solutions is surprisingly simple using the *harmonic-superposition rule* [421, 908]. This rule can be used for marginally bound systems of supersymmetric (extreme) branes when the number of overall transverse dimensions is finite (although it gives asymptotically flat solutions only when it is ≥ 3) and when the Chern–Simons terms in the field strengths do not contribute to the equations of motion. This rule gives an *Ansatz* for the metric, $(p+1)$-form potentials, and dilaton (in $d = 10$) that is based on the forms of the solutions that describe each of the extreme branes independently, namely Eqs. (19.45), (19.50), (19.56), (19.64), and (19.60). Of course, the solutions constructed using this rule have to be checked directly in the equations of motion. For all combinations of two branes this can be done using a generalization of the p-brane a-model that we will briefly study in the next section (see [49, 50] for further generalizations to more branes and non-extremal branes).

Basically, the harmonic-superposition rule says that the metric is diagonal and each component consists in the same factors as each of the individual solutions smeared over the relative transverse directions,[28] multiplied. The same is true for the dilaton. Finally, the differential-form potentials are the sum of those of each individual solution.

Let us use this rule to construct the solution describing the intersection of a F1 in the direction y with a Dp-brane extended in the orthogonal directions $\vec{z}_p \equiv (z^1, \ldots, z^p)$. By combining Eqs. (19.56) and (19.64), smearing H_{F1} along $\vec{z}_p$ and $H_{\mathrm{D}p}$ along y, we obtain

$$\boxed{\begin{aligned} d\hat{s}_{\mathrm{s}}^2 &= H_{\mathrm{D}p}^{-\frac{1}{2}} H_{\mathrm{F1}}^{-1} dt^2 - H_{\mathrm{D}p}^{+\frac{1}{2}} H_{\mathrm{F1}}^{-1} dy^2 - H_{\mathrm{D}p}^{-\frac{1}{2}} d\vec{z}_p^{\,2} - H_{\mathrm{D}p}^{+\frac{1}{2}} d\vec{x}_{8-p}^{\,2}, \\ e^{-2\hat{\phi}} &= e^{-2\hat{\phi}_0} H_{\mathrm{D}p}^{\frac{p-3}{2}} H_{\mathrm{F1}}, \qquad \hat{C}^{(p+1)}{}_{t\underline{z}^1\cdots\underline{z}^p} = \pm e^{-\hat{\phi}_0}\left(H_{\mathrm{D}p}^{-1} - 1\right), \\ \hat{B}_{t\underline{y}} &= \pm\left(H_{\mathrm{F1}}^{-1} - 1\right), \qquad H_{\mathrm{D}p,\mathrm{F1}} = 1 + \frac{h_{\mathrm{D}p,\mathrm{F1}}}{|\vec{x}_{8-p}|^{6-p}}. \end{aligned}} \tag{19.145}$$

It is straightforward to apply the rule to other cases and we will do this in Chapter 20, where more examples can be found. The metrics of some basic intersections can be found, for instance, in [417, 588, 858].

The main insufficiency of these solutions is the delocalization of the branes. For instance, the above solution does not tell us at which point along the coordinate y and in the hyperplane $\vec{z}_p$ the string intersects the Dp-brane. This is immaterial if at the end we want to wrap the solution in those directions and to perform dimensional reduction of the solution, as we will do in Chapter 20 to construct $d = 4, 5$ BHs. Nevertheless, since we have a clear worldvolume picture of supersymmetric intersections, it is natural to try to

[28] That is, all the harmonic functions depend only on the same overall transverse directions.

find the solutions that would have BIons and their generalizations as sources (this was the approach of [461]). Partially localized solutions and some special fully localized solutions have been found in [48, 379, 503, 561, 598, 658, 874, 967], but in [688, 760] it was argued, using AdS/CFT-correspondence arguments, that fully localized solutions might not exist in general, and these arguments seem to be confirmed by the results in [461]. If the string in the F1 $\perp$ D$p(0)$ intersection can be seen as "hair" on the Dp-brane, then the absence of a fully localized solution for that configuration can be seen as a sort of "no-hair theorem" for Dp-brane solutions.

Nevertheless, an *Ansatz* for fully localized intersections of this and other kinds has been given in [794]. The solutions depend on an unknown function that satisfies a highly non-linear differential equation, but it is not known whether this equation has solutions with the appropriate boundary conditions.

19.6.4 The $(a_1$–$a_2)$ *model for* p_1*- and* p_2*-branes and black intersecting branes*

This is a straightforward generalization of the p-brane a-model that includes (p_1+1)- and (p_2+1)-form potentials, coupled to a scalar with parameters a_1 and a_2:

$$\boxed{S = \frac{1}{16\pi G_{\rm N}^{(d)}}\int d^dx\sqrt{|g|}\left[R + 2(\partial\varphi)^2 + \sum_{i=1,2}\frac{(-1)^{p_i+1}}{2\cdot(p_i+2)!}e^{-2a_i\varphi}F^2_{(p_i+2)}\right],} \tag{19.146}$$

and it is just a convenient simplification of the higher-dimensional supergravity actions we are dealing with. Notice, in particular, the absence of Chern–Simons terms: the solutions we will obtain will be solutions of the full supergravity action only when those terms do not contribute to the equations of motion. This condition will be fulfilled in most cases.

The equations of motion corresponding to this action are

$$\begin{aligned} G_{\mu\nu} + 2T^{\varphi}_{\mu\nu} + \sum_{i=1,2}\frac{(-1)^{p_i+1}}{2\cdot(p_i+1)!}e^{-2a_i\varphi}T^{A_{(p_i+1)}}_{\mu\nu} &= 0,\\ \nabla^2\varphi + \sum_{i=1,2}\frac{(-1)^{p_i+1}}{4\cdot(p_i+2)!}a_ie^{-2a_i\varphi}F^2_{(p_i+2)} &= 0,\\ \nabla_\mu\left(e^{-2a_i\varphi}F_{(p_i+2)}{}^{\mu\nu_1\ldots\nu_{p_i+1}}\right) &= 0,\quad i=1,2. \end{aligned} \tag{19.147}$$

The harmonic-superposition rule and our experience indicate that an adequate *Ansatz* for a p_1- and a p_2-brane intersecting over r spatial directions may depend on three functions: H_i, $i=1,2$, which will be independent harmonic functions in the extreme limit and are associated with the potentials, and the "Schwarzschild factor" W, which becomes 1 in the extreme limit. The *Ansatz* must be such that, when a given H_i is set to 1 to recover a solution for a single p_j-brane, $i\neq j$. All these conditions are fulfilled by the metric *Ansatz*

$$\begin{aligned} ds^2 = {}& H_1^{2z_1}H_2^{2z_2}\left[Wdt^2 - d\vec{y}_r^{\,2}\right] - H_1^{2z_1}H_2^{-2y_2}d\vec{y}^{\,2}_{(p_1-r)} - H_1^{-2y_1}H_2^{2z_2}d\vec{y}^{\,2}_{(p_2-r)}\\ & - H_1^{-2y_1}H_2^{-2y_2}\left[d\vec{y}_q^{\,2} + W^{-1}d\rho^2 + \rho^2 d\Omega^2_{(\delta-2)}\right]. \end{aligned} \tag{19.148}$$

The coordinates $\vec{y}_r = (y_1^1, \ldots, y_1^r)$ (plus, of course, time) correspond to the common directions of the two branes relative to the worldvolume of the intersection and the solution is assumed to be independent of them. The coordinates $\vec{y}_{(p_1-r)} = (y_1^2, \ldots, y_2^{(p_1-r)})$ and $\vec{y}_{(p_2-r)} = (y_1^3, \ldots, y_3^{(p_2-r)})$ are relative transverse coordinates (to the p_1- and p_2-branes, respectively). For simplicity, we will also assume the solution to be independent of them (i.e. it will be delocalized). The solution may depend only on the overall transverse coordinates (the rest), but, as we did for single-brane solutions, we include, for completeness, q additional isometries and the solution will not depend on $\vec{y}_q$. Finally, the parameters $z_1, z_2, y_1,$ and y_2 are determined by the single-brane solutions of the a-model, Eq. (18.66).

The dilaton is assumed to be a certain product of powers of H_1 and H_2. Finally, the *Ansatz* for the potentials is the usual one and we need only take into account that the p_i-brane lies in the directions $\vec{y}_r$ and $\vec{y}_{(p_i-r)}$:

$$A_{(p_i+1)1_1\cdots r_1 1_2\cdots(p_i-r)_2} = \alpha_i(H_i^{-1} - 1), \qquad i = 1, 2. \tag{19.149}$$

If this *Ansatz* is to work, then, by insisting on the independence of the two would-be harmonic functions H_1 and H_2, we should simply acquire constraints on $\omega, r, a_1,$ and a_2. On plugging the *Ansatz* into the equations of motion, we find, after a long and boring calculation, that it does indeed lead to solutions under a few conditions on those constants:

$$\boxed{\begin{aligned}
ds^2 &= \left(e^{-2a_1\varphi_1}H_1^{-2}\right)^{\frac{1}{p_1+1}}\left(e^{-2a_2\varphi_2}H_2^{-2}\right)^{\frac{1}{p_2+1}}\left[W dt^2 - d\vec{y}_r^{\,2}\right] \\
&\quad - \left(e^{-2a_1\varphi_1}H_1^{-2}\right)^{\frac{1}{p_1+1}}\left(e^{-2a_2\varphi_2}H_2^{-2}\right)^{-\frac{1}{\tilde{p}_2+1}}d\vec{y}^{\,2}_{(p_1-r)} \\
&\quad - \left(e^{-2a_1\varphi_1}H_1^{-2}\right)^{-\frac{1}{\tilde{p}_1+1}}\left(e^{-2a_2\varphi_2}H_2^{-2}\right)^{\frac{1}{p_2+1}}d\vec{y}^{\,2}_{(p_2-r)} \\
&\quad - \left(e^{-2a_1\varphi_1}H_1^{-2}\right)^{-\frac{1}{\tilde{p}_1+1}}\left(e^{-2a_2\varphi_2}H_2^{-2}\right)^{-\frac{1}{\tilde{p}_2+1}} \\
&\qquad \left[d\vec{y}_q^{\,2} + W^{-1}d\rho^2 + \rho^2 d\Omega^2_{(\delta-2)}\right]. \\
&\quad A_{(p_i+1)1_1\cdots r_1 1_2\cdots(p_i-r)_2} = \alpha_1(H_1^{-1} - 1), \\
e^{-2a_i\varphi_i} &\equiv H_i^{2x_i}, \qquad e^{-2a_i\varphi} = e^{-2a_i\varphi_i}(e^{-2a_j\varphi_j})^{\frac{l_i}{x_j}}, \qquad i \neq j, \\
H_i &= 1 + \frac{h_i}{\rho^{\delta-3}}, \qquad W = 1 + \frac{\omega}{\rho^{\delta-3}}, \\
\omega &= h_i\left[1 - \frac{a_i^2}{4x_i}\alpha_i^2\right], \qquad l_i = (x_i - 1)\left[c_i(r+1) - \frac{p_j+1}{\tilde{p}_i+1}\right], \qquad i \neq j, \\
x_i &= \frac{(a_i^2/2)c_i}{1 + (a_i^2/2)c_i}, \qquad c_i = \frac{(p_i+1) + (\tilde{p}_i+1)}{(p_i+1)(\tilde{p}_i+1)}, \\
a_1 a_2 &= -2(r - r_0), \qquad r_0 = \frac{(p_1+1)(p_2+1)}{d-2} - 1.
\end{aligned}}$$

(19.150)

This solution generalizes to the non-extreme regime extreme intersecting solutions obtained in [51] in another (a_1–a_2) model (see also [50, 925] and references therein). Some of these generalizations had already been obtained in certain cases in [254, 255, 274]. As usual, in the extremal limit $W = 1$ the H_i are arbitrary independent harmonic functions of the overall transverse coordinates.

The most interesting relation that we obtain is the one in the last line, that among the a_is, the dimensionality of the branes, and r:

$$\boxed{r = \frac{(p_1+1)(p_2+1)}{d-2} - \frac{a_1 a_2}{d-2} - 1.} \tag{19.151}$$

This equation contains the intersection rules and we can apply it to some basic examples, using the values for the a_i constants appropriate for each kind of brane (see page 531).

As an example, let us consider the case of $d = 11$ SUGRA.[29] Since there is no scalar, $a_1 = a_2 = 0$, which implies $r = r_0$. Equation (19.151) immediately gives the three intersections M2 $\perp$ M2(0), M2 $\perp$ M5(1), and M5 $\perp$ M5(3) (but not the overlap M5 $\perp$ M5(1), which requires a different *Ansatz*, see e.g. [417]). For example, the solution corresponding to a black intersection M2 $\perp$ M2(0) in which brane 1 lies in the directions $\vec{y}_2 = (y^1, y^2)$ and brane 2 lies in the directions $\vec{z}_2 = (z^1, z^2)$ is given by

$$\boxed{\begin{aligned}
d\hat{\hat{s}}^2 &= H_1^{-\frac{2}{3}} H_2^{-\frac{2}{3}} W dt^2 - H_1^{-\frac{2}{3}} H_2^{\frac{1}{3}} d\vec{y}_2^{\,2} - H_1^{\frac{1}{3}} H_2^{-\frac{2}{3}} d\vec{z}_2^{\,2} \\
&\quad - H_1^{\frac{1}{3}} H_2^{\frac{1}{3}} \big[d\vec{w}_q^{\,2} + W^{-1} d\rho^2 + \rho^2 d\Omega^2_{(5-q)} \big]. \\
\hat{\hat{C}}_{t\underline{y}^1\underline{y}^2} &= \alpha_1 (H_1^{-1} - 1), \qquad \hat{\hat{C}}_{t\underline{z}^1\underline{z}^2} = \alpha_2 (H_2^{-1} - 1), \\
H_i &= 1 + \tfrac{h_i}{\rho^{4-q}}, \qquad W = 1 + \tfrac{\omega}{\rho^{4-q}}, \qquad \omega = h_i\big[1 - \alpha_i^2\big].
\end{aligned}} \tag{19.152}$$

On reducing in a relative transverse dimension, we obtain a black intersecting solution, F1A $\perp$ D2(0). On reducing in one of the extra isometric transverse directions $\vec{w}_q$, we obtain D2 $\perp$ D2(0) (this is the reason why the extra isometric $\vec{w}_q$s are introduced into the *Ansatz*).

The solution corresponding to a black intersection M2 $\perp$ M5(1) in which the M2 lies in the directions $\vec{y}_2 = (y^1, y^2)$ and the M5 in $y^1, \vec{z}_4 = (z^1, \ldots, z^4)$ is

$$\boxed{\begin{aligned}
d\hat{\hat{s}}^2 &= H_1^{-\frac{2}{3}} H_2^{-\frac{1}{3}} [W dt^2 - (dy^1)^2] - H_1^{-\frac{2}{3}} H_2^{\frac{2}{3}} (dy^2)^2 - H_1^{\frac{1}{3}} H_2^{-\frac{2}{3}} d\vec{z}_4^{\,2} \\
&\quad - H_1^{\frac{1}{3}} H_2^{\frac{2}{3}} \big[W^{-1} d\rho^2 + \rho^2 d\Omega^2_{(2)} \big]. \\
\hat{\hat{C}}_{t\underline{y}^1\underline{y}^2} &= \alpha_1 (H_1^{-1} - 1), \qquad \hat{\tilde{\hat{C}}}_{t\underline{y}^1\underline{z}^1\cdots\underline{z}^4} = \alpha_2 (H_2^{-1} - 1), \\
H_i &= 1 + \tfrac{h_i}{\rho}, \qquad W = 1 + \tfrac{\omega}{\rho}, \qquad \omega = h_i\big[1 - \alpha_i^2\big].
\end{aligned}} \tag{19.153}$$

[29] Examples in $d = 10$ can be found in the next chapter, where they are used to construct BH solutions: the black D1 $\parallel$ D5 is given in Eqs. (20.14) and the black D2 $\parallel$ S5 $\parallel$ D6 is given in Eqs. (20.37).

20

String black holes in four and five dimensions

Following our general plan, in the previous chapter we have started to see classical solutions that describe the long-range fields generated by configurations of extended objects in string/M theory. In general, the solutions do not reflect some of the characteristics of the brane configuration which may be understood as "hair," but in many cases of interest (in general, in the presence of unbroken supersymmetry), given a classical supergravity solution, we can tell which brane configurations give rise to it. This is in itself a very interesting development, but there is more, because, if the brane configurations only involve D-branes, they can be associated with two-dimensional CFTs (string theories) over which we have good control. Furthermore, each of the branes considered here (D- or not D-) has a world-volume supersymmetric field theory associated. All this allows us to relate supergravity configurations to QFTs whose degrees of freedom can be understood as the microscopical degrees of freedom of the quantum (super)gravity theory contained in string/M theory. This is, roughly speaking, the basis of the AdS/CFT correspondence and generalizations [170, 679] and also the basis for the microscopical computations of BH entropies [870], the subject of this final chapter.

In this chapter we are going to present $N = 2A/B, d = 10$ SUEGRA solutions associated with configurations of extended objects of type-II superstring theories that lead to BH solutions of maximal $d = 5, 4$ SUEGRAs ($N = 4, d = 5$ and $N = 8, d = 4$) (Section 20.2). The association can be understood as a strong–weak-coupling limit (see Figure 20.1). We will carefully relate the solutions' integration constants to the physical parameters of the stringy sources and then, using our knowledge of the QFTs associated with those sources in the extreme and supersymmetric cases, we will count the states of these QFTs at each energy level and the corresponding entropy will be shown to coincide with one quarter of the area of the BH's horizon (Section 20.3).

Although it is quite self-contained, this chapter is far from complete due to lack of space and also to the immense amount of literature on this subject. Fortunately, there are some good reviews such as [294, 759] and also Maldacena's Ph.D. Thesis, [677]. Other reviews that are interesting for their emphasis on particular aspects of the problem or as sources of bibliography are [43, 57, 62, 547, 548, 678, 708, 713, 758, 856, 927].

Before we study how to construct BHs as composed of intersecting systems of branes, it is interesting to review how the idea of compositeness of BHs came about by studying a

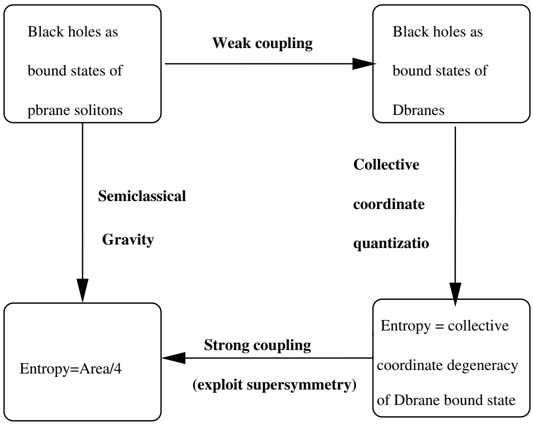

Fig. 20.1. The logic behind the string-theory calculation of extreme BH entropies is represented in this diagram [75].

$d = 4$ model whose BH-type solutions are related to the dilaton BHs studied in Section 12.1.1.

20.1 Composite dilaton black holes

Let us consider the following string-inspired model (which is actually an inconsistent truncation of the heterotic-string effective action compactified on T^6, so not all its solutions will be string solutions):

$$\boxed{\begin{aligned} S = \int d^4x\, \sqrt{|g|}\, \Big\{ R + 2\left[(\partial\phi)^2 + (\partial\sigma)^2 + (\partial\rho)^2\right] - \tfrac{1}{4}e^{-2\phi}\Big[e^{-2(\sigma+\rho)}(F^{(1)\,1})^2 \\ + e^{-2(\sigma-\rho)}(F^{(1)\,2})^2 + e^{2(\sigma+\rho)}(F^{(2)}{}_1)^2 + e^{2(\sigma-\rho)}(F^{(2)}{}_2)^2\Big]\Big\}\,. \end{aligned}} \tag{20.1}$$

Extreme BH-type solutions of this model (which are actually heterotic (or type-II) string solutions) were found in [277], further discussed in [275, 276], and later rediscovered and reinterpreted in [298]. They can be written as follows:

$$\boxed{\begin{aligned} ds^2 &= (H^{(1)\,1}H^{(1)\,2}H^{(2)}{}_1H^{(2)}{}_2)^{-\frac{1}{2}}dt^2 - (H^{(1)\,1}H^{(1)\,2}H^{(2)}{}_1H^{(2)}{}_2)^{-\frac{1}{2}}d\vec{x}_3^{\,2}, \\ A^{(1)\,m}{}_t &= \alpha^{(1)\,m}(H^{(1)\,m}-1)^{-1}, \quad (\alpha^{(1)\,m})^2 = 1, \quad m = 1, 2, \\ \tilde{A}^{(2)}{}_{m\,t} &= \alpha^{(2)}{}_m(H^{(2)}{}_m - 1)^{-1}, \quad (\alpha^{(2)}{}_m)^2 = 1, \quad m = 1, 2, \\ e^{-4\phi} &= \frac{H^{(1)\,1}H^{(2)}{}_1}{H^{(1)\,2}H^{(2)}{}_2}, \qquad e^{-4\sigma} = \frac{H^{(1)\,1}H^{(2)}{}_2}{H^{(1)\,2}H^{(2)}{}_1}, \qquad e^{-4\rho} = \frac{H^{(1)\,1}H^{(1)\,2}}{H^{(2)}{}_1H^{(2)}{}_2}, \end{aligned}} \tag{20.2}$$

where the $H^{(i)}{}_m$s are independent harmonic functions in three-dimensional Euclidean space and the potentials $\tilde{A}^{(2)}{}_m$ correspond to the dual field strengths

$$\tilde{F}^{(2)}{}_1 = e^{-2(\phi+\sigma-\rho)\star}F^{(2)}{}_1, \qquad \tilde{F}^{(2)}{}_2 = e^{-2(\phi-\sigma+\rho)\star}F^{(2)}{}_2. \tag{20.3}$$

The harmonic functions appropriate to describe a single BH are

$$H^{(i)}{}_m = 1 + \frac{|q^{(i)}{}_m|}{|\vec{x}_3|}. \tag{20.4}$$

The $q^{(1)}{}_m$s are electric charges and the $q^{(2)}{}_m$s are magnetic charges. Their signs are given by the $\alpha^{(i)}{}_m$ constants. The signs in the harmonic functions are chosen in order to have a regular metric. The ADM mass and horizon are ($G_{\rm N}^{(4)} = 1$)

$$M = \tfrac{1}{4}\sum_{i,m=1,2} |q^{(i)}{}_m|, \qquad A = 4\pi\sqrt{\prod_{i,m=1,2} |q^{(i)}{}_m|}. \tag{20.5}$$

The area is non-zero (and the horizon is regular) only when the four charges are finite.

The metrics can be related to those of the extreme a-model dilaton BHs of Section 12.1.1: when only one charge is different from zero, the metric is that of the extreme $a = \sqrt{3}$ BH, Eq. (12.22). If there are two non-vanishing charges that are equal, the metric is that of the extreme $a = 1$ BH, Eq. (12.23). Three identical non-vanishing charges give the metric of the extreme $a = 1/\sqrt{3}$ BH, Eq. (12.24); and four identical non-vanishing charges give the metric of the $a = 0$ BH (the ERN BH), Eq. (12.25). This fact suggests the interpretation of ERN BHs as objects composed of four extreme $a = \sqrt{3}$ "BHs" [298], each of which breaks/preserves separately half of the supersymmetries while the ERN preserves an eighth as a type-II (i.e. $N = 8$, $d = 4$ SUEGRA) solution.

It is interesting to study in a bit more detail the preservation of supersymmetries in terms of BPS bounds. As we discussed in Section 13.5.1, there are four central-charge skew eigenvalues Z_i in $N = 8$, $d = 4$ SUEGRA. Their absolute values are in this case [611]

$$\begin{aligned} |Z_1| &= \tfrac{1}{4}|q^1{}_1 + q^2{}_1 + q^1{}_2 + q^2{}_2|, \quad & |Z_2| &= \tfrac{1}{4}|q^1{}_1 - q^2{}_1 + q^1{}_2 - q^2{}_2|, \\ |Z_3| &= \tfrac{1}{4}|q^1{}_1 + q^2{}_1 - q^1{}_2 - q^2{}_2|, \quad & |Z_4| &= \tfrac{1}{4}|q^1{}_1 - q^2{}_1 - q^1{}_2 + q^2{}_2|. \end{aligned} \tag{20.6}$$

If only one of the charges q is different from zero (the extreme $a = \sqrt{3}$ dilaton BH), $M = |Z_i|,\ i = 1, \ldots, 4$, and half of the supersymmetries are preserved. If two are different from zero (the extreme $a = \sqrt{3}$ dilaton BH) (say $q^1{}_1$ and $q^2{}_1$, both positive), then $M = |Z_{1,2}| < |Z_{3,4}|$ and a quarter of the supersymmetries are preserved. For three (say $q^1{}_1$, $q^2{}_1$ and $q^1{}_2$, all positive), $M = |Z_1| < |Z_{2,3,4}|$ and an eighth of the supersymmetries are preserved. If we add a fourth charge $q^2{}_2$, then, if it is positive, no additional supersymmetries are broken, but all are broken if it is negative.

This discussion parallels the discussion of the addition of branes to a type-II configuration on page 568 and we need only establish the link between the $d = 4$ solutions and $d = 10$ brane solutions wrapped on T^6 to arrive at the conclusion that $d = 4$ BHs can be understood as composed of wrapped branes and that, in order to obtain $d = 4$ BHs with regular horizons, we need to include enough branes to break seven eighths of the supersymmetries. After reaching this conclusion, it is natural to try the construction of $d = 4$ and $d = 5$ BHs directly from $d = 10$ extended objects in order to identify precisely which elementary string-theory objects these BHs are made of. This information will later be used in the entropy calculation.

20.2 Black holes from branes

20.2.1 Black holes from single wrapped branes

To gain some insight, we are first going to investigate the construction of BHs in any dimension by wrapping completely p-brane solutions on T^p. The harmonic functions do not change and we need only reduce the metrics. These are diagonal and lead to a metric and the KK scalar modulus associated with the volume of the torus and we just have to rescale with it the reduced metric to the modified Einstein frame. The $(p+1)$-form gives directly a vector in $10-p$ dimensions and higher forms that vanish in these simple solutions.

The simplest example is provided by the F1, Eq. (19.56), wrapped on a circle. It gives rise to a $d=9$ charged extremal BH solution:

$$\begin{aligned} d\tilde{s}_{\mathrm{E}}^2 &= H_{\mathrm{F1}}^{-\frac{6}{7}} dt^2 - H_{\mathrm{F1}}^{\frac{1}{7}} d\vec{x}_8^{\,2}, & A_t &= \pm\big(H_{\mathrm{F1}}^{-1} - 1\big), \\ ds_{\mathrm{s}}^2 &= H_{\mathrm{F1}}^{-1} dt^2 - d\vec{x}_8^{\,2}, & e^{-2\phi} &= e^{-2\phi_0} H_{\mathrm{F1}}^{\frac{1}{2}}, \\ k &= k_0 H_{\mathrm{F1}}^{-\frac{1}{2}}. \end{aligned} \tag{20.7}$$

The horizon would be at the pole of H_{F1}, $\vec{x}_8 = 0$. However, its area (volume) is zero and so it is singular. Furthermore, the size of the compact coordinate, measured in terms of the modulus k, vanishes there and the dilaton diverges.

If we reduce further on T^5, smearing the harmonic function, we obtain the $d=4$ solution

$$\begin{aligned} d\tilde{s}_{\mathrm{E}}^2 &= H_{\mathrm{F1}}^{-\frac{1}{2}} dt^2 - H_{\mathrm{F1}}^{\frac{1}{2}} d\vec{x}_3^{\,2}, & A_t &= \pm\big(H_{\mathrm{F1}}^{-1} - 1\big), \\ ds_{\mathrm{s}}^2 &= H_{\mathrm{F1}}^{-1} dt^2 - d\vec{x}_3^{\,2}, & e^{-2\phi} &= e^{-2\phi_0} H_{\mathrm{F1}}^{\frac{1}{2}}, \\ k &= k_0\, H_{\mathrm{F1}}^{-\frac{1}{2}}, & V_5 &= V_{5\,0}, \end{aligned} \tag{20.8}$$

which has the metric of the $a=\sqrt{3}$ dilaton BH, one component of the solution Eq. (20.2). The metric and moduli of this solution are also singular at the horizon.

Let us try with Dp-brane solutions Eq. (19.64) wrapped on T^p $(p \leq 6)$. They give

$$\begin{aligned} d\tilde{s}_{\mathrm{E}}^2 &= H_{\mathrm{D}p}^{-\frac{7-p}{8-p}} dt^2 - H_{\mathrm{D}p}^{\frac{1}{8-p}} d\vec{x}_{9-p}^{\,2}, & A_t &= \pm\Big(H_{\mathrm{D}p}^{-1} - 1\Big), \\ ds_{\mathrm{s}}^2 &= H_{\mathrm{D}p}^{-\frac{1}{2}} dt^2 - H_{\mathrm{D}p}^{\frac{1}{2}} d\vec{x}_{9-p}^{\,2}, & e^{-2\phi} &= e^{-2\phi_0} H_{\mathrm{D}p}^{\frac{p-6}{4}}, \\ V_p &= V_{p,0}\, H_{\mathrm{D}p}^{-\frac{p}{4}}. \end{aligned} \tag{20.9}$$

In all cases the metric, the modulus that measures the volume of the torus, V_p, and the dilaton are singular on the horizon.[1]

[1] Except for the $p=3$ metric in the string reference frame and the $p=6$ dilaton, which is constant.

If we reduce further to $d = 4$ ($p \leq 6$) on a T^{6-p}, we obtain the solutions

$$
\begin{aligned}
d\tilde{s}^2_{\mathrm{E}} &= H_{\mathrm{D}p}^{-\frac{1}{2}} dt^2 - H_{\mathrm{D}p}^{\frac{1}{2}} d\vec{x}_3^{\,2}, & A_t &= \pm\left(H_{\mathrm{D}p}^{-1} - 1\right),\\
ds^2_{\mathrm{s}} &= H_{\mathrm{D}p}^{-\frac{1}{2}} dt^2 - H_{\mathrm{D}p}^{\frac{1}{2}} d\vec{x}_3^{\,2}, & e^{-2\phi} &= e^{-2\phi_0},\\
V_p &= V_{p\,0}\, H_{\mathrm{D}p}^{-\frac{p}{4}}, & V_{6-p} &= V_{6-p\,0}\, H_{\mathrm{D}p}^{\frac{6-p}{4}},
\end{aligned}
\tag{20.10}
$$

which also has the (singular) metric of the extreme $a = \sqrt{3}$ dilaton BH. The moduli are different but, clearly, Dp-branes and F1s can also be seen as the building blocks of the $d = 4$ BHs, Eqs. (20.2). An analogous calculation can be done for S5s, with analogous results, and, if we consider KK monopoles or gravitational waves, we are going to obtain the same result. All these objects can be used to construct the $d = 4$ BHs and we know that we need at least four of them in order to break more supersymmetries and obtain one extreme BH with a regular horizon.

There is a BH solution of the model in Eq. (20.1) that also describes non-extreme BHs with regular horizons and reduces to the solution Eqs. (20.2) in the extremal limit. It can also be seen as originating from a combination of simpler objects, which turn out to be the dimensional reductions of the non-extremal F1, Dp, S5, etc. solutions. This is the kind of solution we will try to obtain in Section 20.2.2 in $d = 4$ and also in $d = 5$.

Now, however, we are going to discuss briefly what happens near the singularity of the extreme $d =$ BHs we have obtained by reducing a single $d = 10$ object. These are solutions of the string effective action and, as we discussed in Section 15.1, in general they can be trusted only as long as their curvature, measured in string units ℓ_{s}^{-2}, is small. Furthermore, these are solutions of the *compactified* string effective action and we know that the standard KK dimensional reduction of the effective action is valid when the compactification radii are larger than the self-dual radius ℓ_{s} because close to it new stringy massless degrees of freedom that were not considered in the effective action appear. Similar effects are expected to take place when the curvature approaches ℓ_{s}^{-2}.

Thus, new massless stringy degrees of freedom must come into play near the singularities of the above $d = 4$ solutions since the moduli that measure the volumes of the tori in general vanish there, and the solution cannot be trusted beyond a surface of radius $\sim\ell_{\mathrm{s}}$ around the singularity that is sometimes called the *stretched horizon*. For the BH obtained by compactifying the F1 on T^6 Sen suggested in [846] that an entropy that would coincide with the entropy associated with the degeneracy of string states with the same mass could be associated with the stretched horizon. This suggestion goes in the direction of the idea that BHs could be identified with highly excited string states and vice-versa [875].

This correspondence between string states and BHs was made more precise in [553], taking into account the dependence of the radius of the stretched horizon on the string coupling constant $\sim g^2\ell_{\mathrm{s}}$: at strong coupling the size of the stretched horizon is bigger than the string length and the solution that describes the microscopical configuration should be a BH with a regular horizon. At weak coupling, the string picture is the right one. The transition between BHs and strings takes place at the value of g at which the BH's Schwarzschild radius $R_{\mathrm{S}} \sim \ell_{\mathrm{s}}$ and at this point the string and BH description of the BH/string object of a given mass M should agree.[2] The mass of a highly excited closed-string state is, according to Eqs. (14.46) and (14.48), $M \sim N^{\frac{1}{2}}/\ell_s$, while, at the correspondence point, the Schwarzschild BH's mass is $M \sim R_S/(g\ell_s)^2 \sim 1/(g^2\ell_s)$. If both are the mass of the same object then $g \sim N^{-\frac{1}{4}}$.

Let us now compute the entropy of this object in the string description and in the BH description at the correspondence point $R_S \sim \ell_s$, $g \sim N^{-\frac{1}{4}}$. The BH entropy is

$$S \sim R_S^2/G_N^{(4)} \sim g^{-2} \sim \sqrt{N}. \tag{20.11}$$

The string entropy is the logarithm of the degeneracy of states at the mass level M. String theories are a particular case of two-dimensional CFTs [455], which in general have an infinite spectrum of states. The degeneracy of states of a CFT characterized by a *central charge* c, for large values of the two-dimensional energy E, is given by *Cardy's formula*,

$$\rho(E) \sim e^{\sqrt{\pi(c-24E_0)EL/3}}, \tag{20.12}$$

where E_0 is the lowest energy and L the size of the spatial coordinate of the two-dimensional theory. For string theories $L \sim \ell_s$ and $E = M^2 \sim N/\ell_s^2$. Therefore, $\rho \sim e^{M/M_0}$ and $S = \ln\rho \sim \sqrt{N}$, which is in good qualitative agreement with the result in the BH picture.

The BH–string correspondence principle can be extended to higher-dimensional Schwarzschild BHs and also to charged BHs, generalizing at the same time the string picture to a string/brane picture characterized by the same conserved quantities.

Clearly, this principle underlies the logic of the calculation of entropies of stringy BHs depicted in Figure 20.1. It works best when there is unbroken supersymmetry (extreme BHs); it can be argued that the counting of states remains unmodified when we vary g. In the next few sections we are going to construct these $d = 4, 5$ stringy extreme supersymmetric BH solutions.

20.2.2 Black holes from wrapped intersecting branes

We have seen that, in order to construct $d = 4$ extreme BH solutions with regular horizons, we need at least four extended objects that break seven eighths of the supersymmetries. In $d = 5$, three are necessary and, since this case is a bit simpler and the counting of microstates for it clearer, we are going to start with it.

There are many possible configurations of three extended objects that give rise to a regular BH in $d = 5$ upon compactification on T^5. They are related by string (U) dualities in the five compact dimensions, which appear in $d = 5$ as hidden symmetries of the maximal $N = 4$, $d = 5$ SUEGRA. These symmetries do not act on the Einstein metric (although they do act on the moduli), thus any of these configurations is equally good for obtaining a BH metric (the issue of U duality will be studied in Section 20.2.3). Not all the corresponding $d = 10$ configurations are equally simple to treat, basically because we do not have good string descriptions of KK monopoles of S5-branes. D-brane configurations are clearly preferred. The simplest configurations of this kind are D5 $\parallel$ D1 $\parallel$W, as proposed in [208] as a

[2] The BH–fundamental-string transition has been studied further in [284, 554, 618, 619].

simpler alternative to the original configuration proposed by Strominger and Vafa in [870], and the T-dual configuration W‖ D2 ‖ D6 proposed in [680]. We are going to study the former.

In $d = 4$, a configuration that fits the requirements, W ‖ D2 ‖ S5A ‖ D6, was also proposed in [680] and two T-dual alternatives D0 ‖ D4 and D1 ‖ D5 plus an F1 and a KK monopole were proposed in [606]. We are going to study the configuration proposed in [680].

We will not study rotating BHs, although we have mentioned them in several places, remarking, in particular, on the non-existence of supersymmetric rotating BHs with a regular horizon in $d = 4$. It is interesting to mention their existence in $d = 5$, where they can also be modeled with string-theory extended objects [180, 183].

$d = 5$ *Black holes from intersecting branes.* To obtain regular extreme $d = 5$ BHs, it is necessary to construct them as intersections of at least three $d = 10$ extended objects. A possible configuration that leads to regular extreme $d = 5$ BHs is

	0	1	2	3	4	5	6	7	8	9
D1	$+$	$+$	$\sim$	$\sim$	$\sim$	$\sim$	$-$	$-$	$-$	$-$
D5	$+$	$+$	$+$	$+$	$+$	$+$	$-$	$-$	$-$	$-$
W	$+$	$\rightarrow$	$\sim$	$\sim$	$\sim$	$\sim$	$-$	$-$	$-$	$-$

(20.13)

where $+$ signs stand for worldvolume dimensions (isometric in the solutions), $-$ signs stand for overall transverse directions on which the solution depends, $\sim$ signs stand for transverse directions in which the solution has been smeared, and $\rightarrow$ indicates the direction in which the wave propagates. The direction with $+$, $\sim$, or $\rightarrow$ signs will be compactified on a $\mathrm{T}^5 = \mathrm{S}^1 \times \mathrm{T}^4$ with volume $V^5 = 2\pi R_1 V^4$, where R_1 is the radius of the coordinate y^1 and $V^4 = (2\pi)^4 R_2 \cdots R_5$ is the volume of the T^4 on which the coordinates $y^2, \ldots, y^5$ are compactified. Then the solution will depend only on the overall transverse coordinates $\vec{x}_4$.

Our procedure will be to construct first the $d = 10$ solution of $N = 2B$, $d = 10$ SUEGRA that describes this system and then reduce it to $d = 5$ in two steps (S^1 and T^4). Since we will also be interested in non-extremal BHs, we are going to construct the black intersecting solution first and then we will take the extreme limit.

The harmonic-superposition rule cannot be used directly in this black intersection. We start from the non-extreme $D1 \parallel D5$ intersection (contained in Eq. (19.150):

$$\boxed{\begin{aligned} d\hat{s}_{\mathrm{s}}^2 &= H_{\mathrm{D1}}^{-\frac{1}{2}} H_{\mathrm{D5}}^{-\frac{1}{2}}[W dt^2 - (dy^1)^2] - H_{\mathrm{D1}}^{\frac{1}{2}} H_{\mathrm{D5}}^{\frac{1}{2}}[W^{-1}dr^2 + r^2 d\Omega_{(3)}^2] \\ &\quad - H_{\mathrm{D1}}^{\frac{1}{2}} H_{\mathrm{D5}}^{-\frac{1}{2}}[(dy^2)^2 + (dy^3)^2 + (dy^4)^2 + (dy^5)^2], \\ \hat{C}^{(2)}{}_{t\underline{y}^1} &= \alpha_{\mathrm{D1}} e^{-\hat{\phi}_0}\big(H_{\mathrm{D1}}^{-1} - 1\big), \qquad \hat{C}^{(6)}{}_{t\underline{y}^1\cdots\underline{y}^5} = \alpha_{\mathrm{D5}} e^{-\hat{\phi}_0}\big(H_{\mathrm{D5}}^{-1} - 1\big), \\ e^{-2\hat{\phi}} &= e^{-2\hat{\phi}_0} H_{\mathrm{D5}}/H_{\mathrm{D1}}, \qquad H_i = 1 + \frac{r_i^2}{r^2}, \qquad W = 1 + \frac{\omega}{r^2}, \\ \omega &= r_i^2(1 - \alpha_i^2), \qquad i = \mathrm{D1}, \mathrm{D5}. \end{aligned}} \tag{20.14}$$

Now, to add a wave, we use the procedure studied on page 327 and boost the above solution in the direction y^1, obtaining [208, 550]

$$\boxed{\begin{aligned} d\hat{s}_s^2 &= H_{D1}^{-\frac{1}{2}} H_{D5}^{-\frac{1}{2}} \left\{ H_W^{-1} W dt^2 - H_W [dy^1 + \alpha_W (H_W^{-1} - 1) dt]^2 \right\} \\ &\quad - H_{D1}^{\frac{1}{2}} H_{D5}^{\frac{1}{2}} [W^{-1} dr^2 + r^2 d\Omega_{(3)}^2] \\ &\quad - H_{D1}^{\frac{1}{2}} H_{D5}^{-\frac{1}{2}} [(dy^2)^2 + (dy^3)^2 + (dy^4)^2 + (dy^5)^2], \\ \hat{C}^{(2)}{}_{\underline{t y^1}} &= \alpha_{D1} e^{-\hat{\phi}_0} \big(H_{D1}^{-1} - 1 \big), \qquad \hat{C}^{(6)}{}_{\underline{t y^1 \cdots y^5}} = \alpha_{D5} e^{-\hat{\phi}_0} \big(H_{D5}^{-1} - 1 \big), \\ e^{-2\hat{\phi}} &= e^{-2\hat{\phi}_0} H_{D5} / H_{D1}, \qquad H_i = 1 + \frac{r_i^2}{r^2}, \qquad W = 1 + \frac{\omega}{r^2}, \\ \omega &= r_i^2 (1 - \alpha_i^2), \qquad i = \mathrm{D1, D5, W}. \end{aligned}} \tag{20.15}$$

In the extreme limit $\omega = 0$ we recover a D1 $\parallel$ D5 $\parallel$ W that could have been constructed using the harmonic-superposition rule.

Let us now dimensionally reduce this solution in the direction y^1. In this reduction a modulus field that measures the size of that direction arises,[3]

$$\frac{k_1}{k_{10}} = |\hat{g}_{\underline{y^1 y^1}}|^{\frac{1}{2}} = \frac{H_W^{\frac{1}{2}}}{H_{D1}^{\frac{1}{4}} H_{D5}^{\frac{1}{4}}}, \tag{20.16}$$

and $k_{10} = R_1/\ell_s$. In the reduction on T^4 we obtain another modulus field associated with its volume,[4]

$$k_{V^4} = |\hat{g}_{\underline{y^2 y^2}} \hat{g}_{\underline{y^3 y^3}} \hat{g}_{\underline{y^4 y^4}} \hat{g}_{\underline{y^5 y^5}}|^{\frac{1}{2}} = H_{D1}/H_{D5}. \tag{20.17}$$

The $d = 5$ dilaton is given by

$$e^{-2\phi} = e^{-2\hat{\phi}} k_1 = e^{-2\phi_0} \frac{H_W^{\frac{1}{2}}}{H_{D1}^{\frac{1}{4}} H_{D5}^{\frac{1}{4}}}, \qquad e^{-2\phi_0} = e^{-2\hat{\phi}_0} k_{1,0}. \tag{20.18}$$

[3] Before reducing, we can rescale y^1 so that it takes values in $[0, 2\pi\ell_s]$. We will do the same systematically in common worldvolume directions, but not in relative transverse directions, since, in order to apply Eqs. (11.124), the coordinates have to take values in $[0, 2\pi R]$. The value at infinity of the corresponding modulus is 1.

[4] All we are doing here is a standard toroidal compactification of the kind we have performed in Section 16.5 and studied in general in Section 11.4. In general we would obtain a bunch of moduli fields coming from the internal metric. For this and the solutions that will follow, the internal metric is proportional to the identity and there is only one non-trivial modulus: its determinant. Its square root is k_V. We have not performed the toroidal reduction of RR fields, but it is clear that they give rise to a series of form potentials of equal and lower ranks. Since the potentials in this and the other solutions that we are going to study have components only in compact directions (plus time) only the time component of the vector fields that originate from the reduction will be non-trivial and have the obvious value.

The solution of maximal $d = 5$ SUEGRA in the modified Einstein frame is

$$\boxed{\begin{aligned}
&d\tilde{s}_{\rm E}^2 = (H_{\rm D1}H_{\rm D5}H_{\rm W})^{-\frac{2}{3}}Wdt^2 - (H_{\rm D1}H_{\rm D5}H_{\rm W})^{\frac{1}{3}}\,[W^{-1}dr^2 + r^2 d\Omega_{(3)}^2],\\
&A^{(i)}{}_t = \alpha_i\big(H_i{}^{-1} - 1\big), \qquad H_i = 1 + \frac{r_i^2}{|\vec{x}_4|^2}, \qquad \alpha_i^2 = 1, \qquad i = \rm D1, D5, W,\\
&k_{V^4} = \frac{H_{\rm D1}}{H_{\rm D5}}, \qquad e^{-2\phi} = e^{-2\phi_0}\frac{H_{\rm W}^{\frac{1}{2}}}{H_{\rm D1}^{\frac{1}{4}}H_{\rm D5}^{\frac{1}{4}}}, \qquad k_1 = k_{1\,0}\frac{H_{\rm W}^{\frac{1}{2}}}{H_{\rm D1}^{\frac{1}{4}}H_{\rm D5}^{\frac{1}{4}}},\\
&W = 1 + \frac{\omega}{r^2}, \qquad \omega = r_i^2(1-\alpha_i^2), \qquad i = \rm D1, D5, W.
\end{aligned}}$$

(20.19)

This is a BH solution with event horizon at $\rho = -\omega$ ($\omega < 0$). As usual, in the extreme limit $W = 1$ disappears and we can replace $dr^2 + r^2 d\Omega_{(3)}^2$ by $d\vec{x}_4^{\,2}$ and the H_i could be arbitrary harmonic functions. However, we will take them as above in that limit and, with that choice, it is easy to see that there is a regular horizon at $\vec{x}_4 = 0$, with finite area (and, hence, finite entropy) given by

$$A = \omega_{(3)}\left(\lim_{|\vec{x}_4|\to 0}|\vec{x}_4|^6\; H_{\rm D1}H_{\rm D5}H_{\rm W}\right)^{\frac{1}{2}} = 2\pi^2(r_{\rm D1}r_{\rm D5}r_{\rm W})^{\frac{1}{2}}\,. \tag{20.20}$$

Furthermore, the moduli fields are finite there, as we wanted. If $r_{\rm D1} = r_{\rm D5} = r_{\rm W}$ then all the moduli are constant[5] and the metric takes the form

$$ds^2 = H^{-2}Wdt^2 - H[W^{-1}dr^2 + r^2 d\Omega_{(3)}^2], \qquad H = H_{\rm D1} = H_{\rm D5} = H_{\rm W}, \tag{20.21}$$

which is just that of the $d = 5$ RN BH! (See Eq. (8.225).) The only difference is the number of vector fields of the total solution, which is dictated in our case by the requirement that we have an $N = 2B$, $d = 10$ SUEGRA solution that we can relate to a type-IIB superstring configuration.

Our next task is to relate the constants r_i to the physical parameters of the solution. This is very easy to do in the extreme case in which we can immediately associate the supergravity solution with a supersymmetric configuration with $N_{\rm D1}$ D-strings (all with the same kind of charge, either positive or negative, to preserve supersymmetry), $N_{\rm D5}$ D5s (again all of them with the same kind of charge), and $N_{\rm W}$ units of momentum in the direction y^1. Since the D5s are not smeared, using Eq. (19.65) we obtain

$$r_{\rm D5}^2 = N_{\rm D5}h_{\rm D5} = N_{\rm D5}\ell_{\rm s}^2\hat{g}. \tag{20.22}$$

For the D-strings, which are smeared in four directions, we have to use repeatedly Eqs. (11.124):

$$r_{\rm D1}^2 = N_{\rm D1}h_{\rm D1}\frac{\omega_{(5)}}{V^4\omega_{(1)}} = \frac{N_{\rm D1}\ell_{\rm s}^6\hat{g}}{R_2\cdots R_5}. \tag{20.23}$$

5 When we relate these constants to the numbers of branes of each kind, it will become clear that, in general, it is not possible to attain this equality except for special values of the moduli g, R, and V, although, for large numbers of branes, we can be arbitrarily close to the equality.

For the gravitational wave that propagates in a compact direction and has four compact transverse directions, we have to use first the coefficient in Eq. (11.131) which, for N_{W} units of momentum, takes the form

$$h_{\mathrm{W}} = \frac{8NG_{\mathrm{N}}^{(\hat{d})}}{R_z^2(\hat{d}-4)\omega_{(\hat{d}-3)}}. \tag{20.24}$$

Here $z = y^1$, $\hat{d} = 10$, and $G_{\mathrm{N}}^{(10)}$ is given by Eq. (19.26). Smearing in four directions, we find

$$r_{\mathrm{W}}^2 = h_{\mathrm{W}} \frac{\omega_{(5)}}{V^4 \omega_{(1)}} = \frac{N_{\mathrm{W}} \ell_{\mathrm{s}}^8 \hat{g}^2}{R_1^2 R_2 \cdots R_5}. \tag{20.25}$$

This complete identification of all the parameters of the solution in terms of stringy quantities and compactification moduli can be used in the expressions for the area of the horizon and the entropy of the extreme BH. In any dimension, the entropy is one quarter of the area of the horizon in Planck units [706], and we obtain

$$G_{\mathrm{N}}^{(5)} = \frac{G_{\mathrm{N}}^{(10)}}{V^5} = \frac{\pi}{4} \frac{\ell_{\mathrm{s}}^8 g^2}{R_1 \cdots R_5}, \qquad S = \frac{A}{4G_{\mathrm{N}}^{(5)}}, \;\Rightarrow \boxed{S = 2\pi \sqrt{N_{\mathrm{D1}} N_{\mathrm{D5}} N_{\mathrm{W}}},} \tag{20.26}$$

a beautiful formula that does not depend either on any moduli g and R_i or on the string length ℓ_{s}: it depends only on integers, which suggests that it can be explained in terms of a counting of possible string states on the background of the intersecting branes that are compatible with the same supergravity solution. Observe that the mass of the extreme BH (which is typical of a marginally bound configuration) does depend on the moduli:

$$\boxed{M = \frac{N_{\mathrm{D1}} R_1}{\hat{g}\ell_{\mathrm{s}}} + \frac{N_{\mathrm{D5}} R_1 \cdots R_5}{\hat{g}\ell_{\mathrm{s}}^6} + \frac{N_{\mathrm{W}}}{R_1}.} \tag{20.27}$$

The identification of the physical parameters in the non-extreme case is just a bit more complicated and the interpretation in terms of stringy objects is quite a bit more complicated. The physical parameters of the solution are the mass, D1 charge, D5 charge, and momentum, which are not proportional to the numbers of D1s, D5s, and momentum modes. The Chern-Simons terms are zero in this solution and the charges, measured in units of the fundamental D1 and D5 charges Eq. (19.27) (actually, charge densities), are simply

$$\begin{aligned}
\frac{Q_{\mathrm{D1}}}{2\pi\ell_{\mathrm{s}}^2} &= \frac{\hat{g}^2}{16\pi G_{\mathrm{N}}^{(10)}} \int_{\mathrm{S}^3 \times \mathrm{T}^4} {}^\star\hat{G}^{(3)} = \frac{\alpha_{\mathrm{D1}} r_{\mathrm{D1}}^2 R_2 \cdots R_5}{2\pi \ell_{\mathrm{s}}^8 \hat{g}}, && \Rightarrow \alpha_{\mathrm{D1}} r_{\mathrm{D1}}^2 = \frac{Q_{\mathrm{D1}} \ell_{\mathrm{s}}^6 \hat{g}}{R_2 \cdots R_5}, \\
\frac{Q_{\mathrm{D5}}}{(2\pi)^5 \ell_{\mathrm{s}}^6} &= \frac{\hat{g}^2}{16\pi G_{\mathrm{N}}^{(10)}} \int_{\mathrm{S}^3} {}^\star\hat{G}^{(7)} = \frac{\alpha_{\mathrm{D5}} r_{\mathrm{D5}}^2}{(2\pi)^5 \ell_{\mathrm{s}}^8 \hat{g}}, && \Rightarrow \alpha_{\mathrm{D5}} r_{\mathrm{D5}}^2 = Q_{\mathrm{D5}} \ell_{\mathrm{s}} \hat{g}, \\
Q_{\mathrm{W}} &= \frac{\alpha_{\mathrm{W}} r_{\mathrm{W}}^2 R_1^2 R_2 \cdots R_5}{\ell_{\mathrm{s}}^8 \hat{g}^2}, && \Rightarrow \alpha_{\mathrm{W}} r_{\mathrm{W}}^2 = \frac{Q_{\mathrm{W}} \ell_{\mathrm{s}}^8 \hat{g}^2}{R_1^2 R_2 \cdots R_5},
\end{aligned} \tag{20.28}$$

where the last result was obtained by comparison with Eq. (20.24). The Q_is are integers.

The mass of this brane configuration has to be measured in $d = 5$ in the modified Einstein frame. On expanding the g_{tt} component of the metric, we find

$$\omega - \tfrac{2}{3}(r_{\mathrm{D1}}^2 + r_{\mathrm{D5}}^2 + r_{\mathrm{W}}^2) = -\frac{8}{3\pi} G_{\mathrm{N}}^{(5)} M = -\frac{2M\ell_{\mathrm{s}}^8 \hat{g}^2}{3R_1 \cdots R_5}. \qquad (20.29)$$

These four relations plus the three relations among ω, α_i, and r_i in the last of Eqs. (20.19) allow us in principle to express all the integration constants in terms of the physical charges and moduli. In practice, however, one arrives at the equation

$$\sum_{i=\mathrm{D1,D5,W}} \sqrt{\omega^2 + 4\mathcal{Q}_i^2} = \frac{2M\ell_{\mathrm{s}}^8 \hat{g}^2}{3R_1 \cdots R_5} \equiv \mathcal{M}, \qquad \mathcal{Q}_i \equiv \alpha_i r_i^2, \qquad (20.30)$$

and it is difficult in general to write $\omega(\mathcal{M}, \mathcal{Q}_i)$ in a manageable way. So it is better to express all the constants in terms of ω and $\mathcal{Q}_i$: namely, M using the above equation, the $\alpha_i s$ using

$$\alpha_i = \frac{2\mathcal{Q}_i}{\omega + \sqrt{\omega^2 + 4\mathcal{Q}_i}}, \qquad (20.31)$$

and the r_is using Eqs. (20.28). The horizon area takes the value

$$A = 2\pi^2 \sqrt{\prod_i (r_i^2 - \omega)} \sim 2\pi^2 \sqrt{\mathcal{Q}_{\mathrm{D1}} \mathcal{Q}_{\mathrm{D5}} \mathcal{Q}_{\mathrm{W}}} \left(1 + \tfrac{1}{2}|\omega|\right)^{\frac{1}{2}}, \qquad \omega \ll |\mathcal{Q}_i|, \qquad (20.32)$$

from which we can immediately find the entropy. However, this formula is difficult to explain in terms of counting of states since the Q_is are not related to the numbers of stringy objects. In [550] it was proposed that one should interpret the Q_is as total charge densities associated with N_i branes *and* $\bar{N}_i$ antibranes; that is, $Q_i \equiv N_i - \bar{N}_i$. Furthermore, we also assume[6] that

$$\frac{\omega^2}{4\ell_{\mathrm{s}}^4 \hat{g}^2} \equiv 4N_{\mathrm{D5}}\bar{N}_{\mathrm{D5}}, \qquad \frac{\omega^2 R_2^2 \cdots R_5^2}{4\ell_{\mathrm{s}}^{12} \hat{g}^2} \equiv 4N_{\mathrm{D5}}\bar{N}_{\mathrm{D5}}, \qquad \frac{\omega^2 R_1^4 R_2^2 \cdots R_5^2}{4\ell_{\mathrm{s}}^{16} \hat{g}^4} \equiv 4N_{\mathrm{W}}\bar{N}_{\mathrm{W}}. \qquad (20.33)$$

The remaining physical parameters can be written in terms of the N_is and $\bar{N}_i$s:

$$\boxed{\begin{aligned} M &= (N_{\mathrm{D1}} + \bar{N}_{\mathrm{D1}}) \frac{R_1}{\hat{g}\ell_{\mathrm{s}}} + (N_{\mathrm{D5}} + \bar{N}_{\mathrm{D5}}) \frac{R_1 \cdots R_5}{\hat{g}\ell_{\mathrm{s}}^6} + (N_{\mathrm{W}} + \bar{N}_{\mathrm{W}}) \frac{1}{R_1}, \\ R_1^2 &= \hat{g}^{\frac{1}{2}} \ell_{\mathrm{s}}^2 \frac{N_{\mathrm{D1}}\bar{N}_{\mathrm{D1}}}{N_{\mathrm{W}}\bar{N}_{\mathrm{W}}}, \qquad R_2^2 \cdots R_5^2 = \ell_{\mathrm{s}}^8 \frac{N_{\mathrm{D1}}\bar{N}_{\mathrm{D1}}}{N_{\mathrm{D5}}\bar{N}_{\mathrm{D5}}}, \end{aligned}} \qquad (20.34)$$

[6] This could always be seen as a change of variables, although the resulting expressions have a very appealing physical interpretation. Observe that, when all the branes of one kind have charges of the same sign, $\omega = 0$ and we have an extremal BH as before.

The mass formula should be compared with that of the extreme BH Eq. (20.27): branes and antibranes contribute equally to it, and there seems to be no interaction/binding energy between them.

The entropy takes the form

$$\boxed{S = 2\pi\left(\sqrt{N_{\rm D5}} - \sqrt{\bar{N}_{\rm D5}}\right)\left(\sqrt{N_{\rm D1}} - \sqrt{\bar{N}_{\rm D1}}\right)\left(\sqrt{N_{\rm W}} - \sqrt{\bar{N}_{\rm W}}\right).} \tag{20.35}$$

$d = 4$ *Black holes from intersecting branes.* In [549] a stringy model based on a system of intersecting D6, D2, S5A, and W in the configuration

	0	1	2	3	4	5	6	7	8	9
D6	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$-$	$-$	$-$
S5	$+$	$+$	$+$	$+$	$+$	$+$	$\sim$	$-$	$-$	$-$
D2	$+$	$+$	$\sim$	$\sim$	$\sim$	$-$	$+$	$-$	$-$	$-$
W	$+$	$\rightarrow$	$\sim$	$\sim$	$\sim$	$\sim$	$\sim$	$-$	$-$	$-$

(20.36)

was proposed in order to describe $d = 4$ BHs. In the extreme limit (which had been constructed before in [606, 680]) it has a regular horizon and moduli that are regular there (if the D2 were placed in directions 012, then the moduli would be singular at the horizon). The construction of the $d = 10$ solution is similar to that of the solution associated with the $d = 5$ BH which we have discussed in detail and we will therefore omit unnecessary details here.

The black intersecting solution is

$$\boxed{\begin{aligned}
d\hat{s}_{\rm s}^2 &= H_{\rm D6}^{-\frac{1}{2}} H_{\rm D2}^{-\frac{1}{2}} \left\{ H_{\rm W}^{-1} W dt^2 - H_{\rm W}\left[dy^1 + \alpha_{\rm W}(H_{\rm W}^{-1} - 1)dt\right]^2 \right\} \\
&\quad - H_{\rm D6}^{-\frac{1}{2}} H_{\rm D2}^{\frac{1}{2}} [(dy^2)^2 + (dy^3)^2 + (dy^4)^2 + (dy^5)^2] \\
&\quad - H_{\rm D6}^{-\frac{1}{2}} H_{\rm D2}^{-\frac{1}{2}} H_{\rm S5} (dy^6)^2 - H_{\rm D6}^{\frac{1}{2}} H_{\rm D2}^{\frac{1}{2}} H_{\rm S5}[W^{-1}dr^2 + r^2 d\Omega_{(2)}^2], \\
e^{-2\hat{\phi}} &= e^{-2\hat{\phi}_0} H_{\rm D6}^{\frac{3}{2}} H_{\rm D2}^{-\frac{1}{2}} H_{\rm S5}^{-1}, \qquad \hat{B}^{(6)}{}_{t\underline{y}^1\cdots\underline{y}^5} = \alpha_{\rm S5} e^{-2\hat{\phi}_0}\left(H_{\rm S5}^{-1} - 1\right), \\
\hat{C}^{(3)}{}_{t\underline{y}^1\underline{y}^6} &= \alpha_{\rm D2} e^{-\hat{\phi}_0}\left(H_{\rm D2}^{-1} - 1\right), \qquad \hat{C}^{(7)}{}_{t\underline{y}^1\cdots\underline{y}^6} = \alpha_{\rm D6} e^{-\hat{\phi}_0}\left(H_{\rm D6}^{-1} - 1\right), \\
H_i &= 1 + \frac{r_i}{r}, \qquad \alpha_i^2 = +1, \qquad\qquad i = {\rm D6, D2, S5, W}, \\
W &= 1 + \frac{\omega}{r}, \qquad \omega = r_i(1 - \alpha_i^2), \qquad i = {\rm D6, D2, S5, W},
\end{aligned}} \tag{20.37}$$

We dimensionally reduce this solution in three steps: first on S^1 (y^1), then on T^4

$(y^2, \ldots, y^5)$, and then on S^1 (y^6). In the modified Einstein frame it takes the form

$$\boxed{\begin{aligned}
d\tilde{s}_{\mathrm{E}}^2 &= (H_{\mathrm{D6}}H_{\mathrm{D2}}H_{\mathrm{W}}H_{\mathrm{S5}})^{-\frac{1}{2}}\, W dt^2 \\
&\quad - (H_{\mathrm{D6}}H_{\mathrm{D2}}H_{\mathrm{W}}H_{\mathrm{S5}})^{\frac{1}{2}}[W^{-1}dr^2 + r^2 d\Omega_{(2)}^2], \\
A^{(i)}{}_t &= \alpha_i\big(H_i^{-1} - 1\big), \\
e^{-2\phi} &= e^{-2\phi_0}\frac{H_{\mathrm{W}}^{\frac{1}{2}}}{H_{\mathrm{S5}}^{\frac{1}{2}}}, \qquad k_1 = k_{1\,0}\frac{H_{\mathrm{W}}^{\frac{1}{2}}}{H_{\mathrm{D6}}^{\frac{1}{4}}H_{\mathrm{D2}}^{\frac{1}{4}}}, \\
K_6 &= \frac{H_{\mathrm{S5}}^{\frac{1}{2}}}{H_{\mathrm{D6}}^{\frac{1}{4}}H_{\mathrm{D2}}^{\frac{1}{4}}}, \qquad K_{V^4} = \frac{H_{\mathrm{D2}}}{H_{\mathrm{D6}}}. \\
H_i &= 1 + \frac{r_i}{r}, \qquad \alpha_i^2 = +1, \qquad i = \mathrm{D6, D2, S5, W}, \\
W &= 1 + \frac{\omega}{r}, \qquad \omega = r_i(1 - \alpha_i^2),\ i = \mathrm{D6, D2, S5, W}.
\end{aligned}} \tag{20.38}$$

This solution has a regular horizon at $r = -\omega$ $(\omega < 0)$. Actually, when all the constants r_i have the same value, all the $H_i = H$ and the metric is identical to that of the $d = 4$ RN BH. In all cases, the extreme-limit $W = 1$ solution (compare it with the solution Eq. (20.2)) has a regular horizon.

In this case we are going to find the physical parameters directly in the non-extreme case. Using definitions similar to those of the $d = 5$ case, we find

$$\begin{aligned}
\alpha_{\mathrm{D2}}r_{\mathrm{D2}} &= \frac{Q_{\mathrm{D2}}\ell_{\mathrm{s}}^5\hat{g}}{2R_2\cdots R_5}, \qquad & \alpha_{\mathrm{D6}}r_{\mathrm{D6}} &= \frac{Q_{\mathrm{D6}}\ell_{\mathrm{s}}\hat{g}}{2}, \\
\alpha_{\mathrm{S5}}r_{\mathrm{S5}} &= \frac{Q_{\mathrm{S5}}\ell_{\mathrm{s}}^2}{2R_6}, \qquad & \alpha_{\mathrm{W}}r_{\mathrm{W}} &= \frac{Q_{\mathrm{W}}\ell_{\mathrm{s}}^8\hat{g}^2}{2R_1^2R_2\cdots R_6},
\end{aligned} \tag{20.39}$$

and, for the mass,

$$\omega - \tfrac{1}{2}\sum_i r_i = -\frac{M\ell_{\mathrm{s}}^8\hat{g}^2}{4R_1\cdots R_6}. \tag{20.40}$$

We find the following equation relating charges, mass, and ω:

$$\sum_i \sqrt{\omega^2 + 4\mathcal{Q}_i} = \frac{M\ell_{\mathrm{s}}^8\hat{g}^2}{4R_1\cdots R_6} \equiv \mathcal{M}, \qquad \mathcal{Q}_i \equiv \alpha_i r_i, \tag{20.41}$$

which is, again, very difficult to solve. We therefore use ω as a parameter and find, for the α_is, again Eq. (20.31), etc. The area of the horizon is given by

$$A = 4\pi\sqrt{\prod_i (r_i - \omega)}. \tag{20.42}$$

Now, introducing a parametrization similar to the five-dimensional $Q_i \equiv N_i - \bar{N}_i$ and

$$
\begin{aligned}
\frac{\omega^2}{\ell_s^2 \hat{g}^2} &\equiv 4 N_{\mathrm{D6}} \bar{N}_{\mathrm{D6}}, & \frac{\omega^2 R_2^2 \cdots R_5^2}{\ell_s^{10} \hat{g}^2} &\equiv 4 N_{\mathrm{D2}} \bar{N}_{\mathrm{D2}}, \\
\frac{\omega^2 R_1^4 R_2^2 \cdots R_6^2}{\ell_s^{16} \hat{g}^4} &\equiv 4 N_{\mathrm{W}} \bar{N}_{\mathrm{W}}, & \frac{\omega^2 R_6^2}{\ell_s^2 \hat{g}^2} &\equiv 4 N_{\mathrm{D6}} \bar{N}_{\mathrm{D6}},
\end{aligned}
\tag{20.43}
$$

we obtain for the mass and entropy

$$
\boxed{
\begin{aligned}
M = {} & (N_{\mathrm{D2}} + \bar{N}_{\mathrm{D2}}) \frac{R_1 R_6}{\hat{g} \ell_s^3} + (N_{\mathrm{D6}} + \bar{N}_{\mathrm{D6}}) \frac{R_1 \cdots R_6}{\hat{g} \ell_s^7} \\
& + (N_{\mathrm{S5}} + \bar{N}_{\mathrm{S5}}) \frac{R_1 \cdots R_5}{\hat{g}^2 \ell_s^6} + (N_{\mathrm{W}} + \bar{N}_{\mathrm{W}}) \frac{1}{R_1}, \\
S = {} & 2\pi \prod_i \left(\sqrt{N_i} - \sqrt{\bar{N}_i} \right).
\end{aligned}
}
\tag{20.44}
$$

We can also express the moduli in terms of them, ℓ_s, and $\hat{g}$.

20.2.3 Duality and black-hole solutions

The solutions we have obtained are particular solutions that have only a few vectors and scalars excited of maximal $N = 4$, $d = 5$ and $N = 8$, $d = 4$ SUEGRA. These theories have, respectively, 27 and 56 U(1) vector fields, which are rotated among themselves by the U-duality groups $\mathrm{E}_{6(+6)}$ and $\mathrm{E}_{7(+7)}$ (see Table 16.1) and scalars that parametrize the coset spaces $\mathrm{E}_{6(+6)}/\mathrm{USp}(8)$ and $\mathrm{E}_{7(+7)}/\mathrm{SU}(8)$ [263] and are also rotated the same U-duality groups, but the Einstein metrics are U-duality-invariant and all unbroken supersymmetries are preserved.

Several questions arise immediately. How does U duality act on these solutions and on their $d = 10$ description? How does U duality act on physical parameters such as the masses and entropies? What is the more general BH-type solution of these theories?

Most U-duality rotations correspond to T and S dualities in higher dimensions, whose effects on the components are well known to us. We can use them to find configurations that are more convenient for our purposes. For instance, we can dualize the $d = 4$ BH we have obtained into one composed entirely of D-branes [76, 77], whose stringy description is much better known than that of the S5s we have used. If we denote by T^n a T-duality transformation in the nth coordinate, ignoring time and the three overall transverse coordinates, we find

$$
\begin{array}{|c||cccccc|}
\hline
 & 1 & 2 & 3 & 4 & 5 & 6 \\
\hline
\mathrm{D6} & + & + & + & + & + & + \\
\mathrm{S5} & + & + & + & + & + & \sim \\
\mathrm{D2} & + & \sim & \sim & \sim & \sim & + \\
\mathrm{W} & \rightarrow & \sim & \sim & \sim & \sim & \sim \\
\hline
\end{array}
\;\; \xrightarrow{\mathrm{ST}^2\mathrm{T}^1\mathrm{T}^5} \;\;
\begin{array}{|c||cccccc|}
\hline
 & 1 & 2 & 3 & 4 & 5 & 6 \\
\hline
\mathrm{D3} & \sim & \sim & + & + & \sim & + \\
\mathrm{D5} & + & + & + & + & + & \sim \\
\mathrm{D3} & \sim & + & \sim & \sim & + & + \\
\mathrm{D1} & + & \sim & \sim & \sim & \sim & \sim \\
\hline
\end{array}
\tag{20.45}
$$

which can be T-dualized further into a configuration involving only D3s etc.,

$$\overset{\longrightarrow}{\mathrm{T}^2\mathrm{T}^3}\ \begin{array}{|c||cccccc|}\hline & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline \mathrm{D3} & \sim & + & \sim & + & \sim & + \\ \mathrm{D3} & + & \sim & \sim & + & + & \sim \\ \mathrm{D3} & \sim & \sim & + & \sim & + & + \\ \mathrm{D3} & + & + & + & \sim & \sim & \sim \\ \hline \end{array} \tag{20.46}$$

U-duality transformations do not change the $d = 5, 4$ Einstein metric of these solutions; they amount simply to a relabeling of the vector fields and to a complicated non-linear transformation of the scalars. The physical parameters defined in terms of the Einstein metric (such as the mass and entropy) do not change, which means that there must be duality-invariant expressions for them. If the mass is an independent parameter (as it would be from the supergravity point of view), this is an empty statement, but, since the mass depends on the masses of the constituents and the moduli, which are transformed by duality, it is not, as a matter of fact.

The U-duality-invariant expressions for the masses and entropies are particularly simple in the extreme limits since they are completely determined by the moduli and the vector's electric and magnetic charges (27 electric charges in $d = 5$ and 28 electric and 28 magnetic charges in $d = 4$) encoded in the superalgebra's central-charge matrix. Actually, if the entropy counts microstates, it should not depend on any moduli, but only on charges. This is what happens in the explicit solutions that we have constructed. Since, at least in this case, there is only one U-duality invariant that can be constructed from the charges alone, the U-duality invariant expressions for the entropy of extreme BHs of $N = 4, d = 5$ and $N = 8, d = 4$ SUEGRA can easily be determined [381, 383, 611]. In the $N = 8, d = 4$ case, the entropy must be given by the beautiful formula

$$\boxed{S = 4\pi\sqrt{|\diamondsuit|},} \tag{20.47}$$

where $\diamondsuit$ is the unique quartic invariant of E_7. In the SO(8) form

$$-\diamondsuit \equiv x^{ij}y_{jk}x^{kl}y_{li} - \tfrac{1}{4}x^{ij}y_{ij}x^{kl}y_{kl} + \frac{1}{96}\epsilon_{ijklmnop}(x^{ij}x^{kl}x^{mn}x^{op} + y^{ij}y^{kl}y^{mn}y^{op}), \tag{20.48}$$

where x^{ij} and y^{ij} are, respectively, the real and imaginary parts of the 8×8 antisymmetric central-charge matrix $z_{ij} = (x^{ij} + iy^{ij})/\sqrt{2}$ that contains the 28 electric and 28 magnetic charges. This parametrization is very convenient because it can easily be related to the charges of extended objects of $N = 2A, d = 10$ SUEGRA compactified on T^6 [75]: if the indices $i, j, k, l, m, n = 1, \ldots, 6$ then

- $x^{ij} = \dfrac{1}{4!\sqrt{2}}\epsilon^{ijklmn}\mathcal{Z}^{(4)}_{klmn}$ correspond to D4s wrapped on the $klmn$ directions,
- $y^{ij} = \dfrac{1}{\sqrt{2}}\mathcal{Z}^{(2)}_{ij}$ correspond to D2s wrapped on the ij directions,

- $x^{78} = \frac{1}{\sqrt{2}}\mathcal{Z}^{(0)}$ corresponds to D0 charge,
- $y^{78} = -\frac{1}{\sqrt{2}}\mathcal{Z}^{(6)}_{123456}$ corresponds to a D6 wrapped on T^6,
- $x^{i7} = \frac{1}{5!\sqrt{2}}\epsilon^{ijklmn}\mathcal{Z}^{(5)}_{jklmn}$ correspond to S5As wrapped on $jklmn$,
- $x^{i8} = \frac{1}{\sqrt{2}}p^i$ correspond to KK momentum in the directions i,
- $y^{i7} = \frac{1}{5!\sqrt{2}}\epsilon^{(i)jklmn}\mathcal{Z}^{(6)}_{jklmn(i)}$ correspond to KK6As wrapped on $jklmn$ with isometric direction i, and
- $y^{i8} = \frac{1}{\sqrt{2}}\mathcal{Z}^{(1)\,i}$ correspond to F1s wrapped in the directions i.

The $d=4$ extremal BH that we have constructed corresponds to $x^{18} = (1/\sqrt{2})N_{\mathrm{W}}$, $y^{16} = (1/\sqrt{2})N_{\mathrm{D2}}$, $x^{67} = (1/\sqrt{2})N_{\mathrm{S5}}$, and $y^{78} = -(1/\sqrt{2})N_{\mathrm{D6}}$, and the diamond formula immediately gives the right value for the entropy. The dual configurations give exactly the same result. In fact, the above identification between entries of the central-charge matrix and charges of extended objects is clearly not unique, but is defined only up to U-duality rotations. We can also look at it from a different point of view: the objects we have considered are all wrapped on T^6 and are T dual to each other and, essentially, they cannot be distinguished from the $d=4$ central-matrix point of view. Their masses may be different if they are related by S dualities, though. On the other hand, since U duality acts on the moduli, too, we have to use the description in which compactification radii are bigger than the critical self-dual radius and the coupling constant is small.

It would be very interesting to have explicitly the most general U-duality-invariant BH-type solutions of these theories consistent with the no-hair theorems, which would be similar to the SWIP solutions of pure $N=4$, $d=4$ SUEGRA that we studied in Section 12.2.1 to check these formulae, but obtaining them turns out to be an extremely difficult problem. A simpler problem consists in finding a generating solution that would give the general solution when we act on it with a general U-duality transformation, which would preserve the metric. Simple arguments [75, 271] tell us that such solutions must have, respectively, four and five independent charge parameters in $d=4$ and 5 dimensions. It is not hard to find brane configurations with these parameters and generating solutions have been proposed in [144, 145].

20.3 Entropy from microstate counting

In the previous section we constructed BH solutions of maximal $d=4,5$ SUEGRA and identified the $N=2A/B, d=10$ SUEGRA solutions they originate from. In the extreme cases, we identified unambiguously their "components," which places us in the upper-left-hand box of Figure 20.1, and allows us to move clockwise in the figure and calculate the microscopical entropy and see that it coincides with $A/(4G_{\mathrm{N}}^{(d)})$, the commonly accepted semiclassical value.[7]

From the string-theory point of view, the BH solutions that we have obtained are just vacua on which strings should be quantized taking into account the boundary conditions imposed by the presence of D-branes and other extended objects. However, this may be difficult to do since, in general, the string coupling constant $\hat{g}$ is going to be large in order for the solutions to have a macroscopical Schwarzschild radius, and we only know how to quantize perturbatively, for small $\hat{g}$. We have argued above, however, that the entropy is independent of the moduli and, therefore, we can try to calculate it in the limit $\hat{g} \to 0$ in which the Schwarzschild radius goes to zero and the BH description has to be replaced by a system of branes in Minkowski spacetime, a background we do know how to quantize type-II superstrings on.[8] The microscopic entropy is calculated in this limit and its value is then extrapolated to the large-$\hat{g}$ (BH) regime. It is also important for the validity of this calculation that only BPS microstates contribute to the entropy,[9] and that the dimension of supermultiplets (and, hence, the counting of BPS states) is independent of $\hat{g}$.

All we have to do now is to identify the string theory defined by the vacuum of extended objects associated with the BH in the weak-coupling limit and find, in particular the central charge c, and the values of L and E_0 that characterize it as a CFT. Then Cardy's formula Eq. (20.12) gives the state degeneracy and its logarithm gives the entropy.

The first calculation along these lines was performed in [870] and in [208] a similar calculation with a simpler model (the extreme D1 $\parallel$ D5 $\parallel$W $d=5$ BH that we constructed above) was performed. The background of worldsheet string-theory that we have presented in this book is not sufficient to explain in detail the identification of the two-dimensional CFT associated with this BH (i.e. with N_{D1} parallel D1s intersecting in one dimension, N_{D5} parallel D5s, and momentum in the direction of the D1s), but its essential aspects are not too difficult to understand. First of all, since the theory will be supersymmetric, $E_0 = 0$. Then, one has to realize that the only string states that are going to contribute correspond to the open sector with one endpoint on one of the N_{D1} parallel D1s and the other on one of the N_{D5} parallel D5s, which have momentum in the direction y^1 that contributes to their mass[10] the amount N_{W}/R_1. The number of these string modes (states of the two-dimensional CFT) is clearly proportional to the product $N_{\mathrm{D1}}N_{\mathrm{D5}}$ and we expect c, which measures the number of local degrees of freedom of a CFT, to be proportional to it. A more precise calculation gives $c = 6N_{\mathrm{D1}}N_{\mathrm{D5}}$. Finally, in this case[11] $L = 2\pi R_1$ and, on substituting this into Cardy's formula, we obtain the entropy Eq. (20.35) in the extreme limit, a very interesting result whose validity has been reviewed more recently in [927]. Similar arguments explain why

[7] See, however, the discussion on page 245, which is clearly related to the correspondence principle.

[8] We are assuming implicitly that the r_i coefficients of our BH solutions are proportional to positive powers of $\hat{g}$, which implies that they are composed of D-branes, F1s, or Ws.

[9] There are subleading contributions from non-BPS states as well. They introduce small corrections to the entropy [212].

[10] There are no more contributions to the masses of these states, apart from the oscillators, but states with excited oscillator modes do not have the required properties of supersymmetry. The same happens to the states that start and end on branes of the same kind.

[11] Clearly, we are considering the theory that lives in the intersection of the D-branes.

the entropy of the $d = 4$ BH is also proportional to the square root of the product of the numbers of D-branes.

There is, unfortunately, no more space to study more general examples of these calculations (see, for instance, [75]); neither is there time to see how these string models explain qualitatively and quantitatively the Hawking radiation [292, 293, 681], nor extra dimensions to relate these calculations to the $d = 3$ BH of Bañados, Teitelboim, and Zanelli and non-stringy CFTs[12] [82, 213, 590, 849]. We hope that the reader will have found these introductory notes useful.

[12] To what extent are strings fundamental to this result?

Appendix A
Lie groups, symmetric spaces, and Yang–Mills fields

In this appendix we review some basic definitions and properties of Lie groups and algebras and their use in the construction of homogeneous and symmetric spaces and in field theory. Rigorous definitions and proofs can be found, for instance, in [527, 630], and in the physicist-oriented [89, 221 (Volume I), 252, 267 (Volume II), 454].

A.1 Generalities

A Lie group G of dimension n is both a group and a differential manifold of dimension n: the points of the manifold are the elements of the group and the maps

$$\begin{aligned} \mathrm{G} \times \mathrm{G} &\rightarrow \mathrm{G}, & \qquad \mathrm{G} &\rightarrow \mathrm{G}, \\ (g_1, g_2) &\rightarrow g_1 g_2, & \qquad g &\rightarrow g^{-1}, \end{aligned} \tag{A.1}$$

are differentiable. For each element $g \in \mathrm{G}$ there are also two natural diffeomorphisms: *left* and *right translations by* g, denoted by L_g and R_g respectively, and defined by

$$\begin{aligned} L_g : \mathrm{G} &\rightarrow \mathrm{G}, & \qquad R_g : \mathrm{G} &\rightarrow \mathrm{G}, \\ h &\rightarrow L_g(h) \equiv gh, & \qquad h &\rightarrow R_g(h) \equiv hg. \end{aligned} \tag{A.2}$$

The identity e is a naturally distinguished point. The tangent space at the identity $\mathrm{T}_e^{(1,0)}$ is the *Lie algebra* $\mathfrak{g}$ of G. This name will be justified later. Each element $v(e) \in \mathfrak{g}$ can be extended to a vector field $v(g)$ defined at all points $g \in \mathrm{G}$ by taking the push-forward of the left- or the right-translation diffeomorphisms

$$v_{\mathrm{L}}(g) \equiv L_{g\,*}\, v(e), \qquad v_{\mathrm{R}}(g) \equiv R_{g\,*}\, v(e). \tag{A.3}$$

Sometimes we use the following notation for them:

$$L_{g\,*}\, v \equiv gv, \qquad R_{g\,*}\, v \equiv vg. \tag{A.4}$$

The vector fields defined in this way have the property of being, respectively, *left-* and *right-invariant*, i.e. they satisfy

$$L_{g\,*}\, v_{\mathrm{L}}(h) = v_{\mathrm{L}}(gh), \qquad R_{g\,*}\, v_{\mathrm{R}}(h) = v_{\mathrm{R}}(hg). \tag{A.5}$$

Similarly, we can define left- and right-invariant differential forms ω and η of any rank using the pull-backs associated with the left- and right-translation diffeomorphisms:

$$L_g{}^* \omega(h) = \omega(gh), \qquad R_g{}^* \eta(h) = \eta(hg). \tag{A.6}$$

It is customary to work with left-invariant vector fields and 1-forms. We can always construct a basis of left-invariant vector fields $\{e_I(g)\}$ using the above procedure starting with a basis of vector fields at the identity $\{e_I(e)\}$ and a dual basis of left-invariant 1-forms $\{e^I(g)\}$. The Lie bracket of two left-invariant vector fields is another left-invariant vector field that we can write as a linear combination of elements of the $\{e_I(g)\}$ basis.[1] Thus,

$$[e_I, e_J] = -f_{IJ}{}^K e_K, \tag{A.7}$$

where $f_{IJ}{}^K = -f_{JI}{}^K = +(f_{IJ}{}^K)^*$ are the *structure constants*. The Jacobi identity Eq. (1.14) implies

$$f_{J[K}{}^I f_{LM]}{}^K = 0. \tag{A.8}$$

The vector fields at the identity are thus an n-dimensional vector space endowed with an antisymmetric, bilinear (but non-associative) product $[\cdot, \cdot]$ (the Lie bracket of the associated left-invariant vector fields) that satisfies the Jacobi identity; that is, by definition, a Lie algebra, which justifies our definition of $\mathfrak{g}$. We will denote a basis of $\mathfrak{g}$ by $\{T_I\}$ and, by convention,

$$e_I(e) \equiv -T_I, \quad \Rightarrow [T_I, T_J] = f_{IJ}{}^K T_K. \tag{A.9}$$

The dual left-invariant 1-forms satisfy the *Maurer–Cartan equations*

$$de^I = \tfrac{1}{2} f_{JK}{}^I e^J \wedge e^K, \tag{A.10}$$

and $d^2 e^I = 0$ is equivalent to the Jacobi identity.

The exponential map provides a local parametrization of G in a neighborhood of the identity with coordinates σ^I:

$$g(\sigma) = e^{\sigma^I T_I}. \tag{A.11}$$

If the group is a connected and compact manifold, any of its elements can be expressed in this way. With this parametrization it is easy to construct a basis of left-invariant[2] 1-forms by expanding the *Maurer–Cartan 1-form* V,

$$V = -g^{-1} dg = e^I T_I, \tag{A.12}$$

in terms of which the Maurer–Cartan equations are $dV - V \wedge V = 0$.

For matrix groups the generators T_I are just matrices and the Lie bracket is just the standard commutator. The left and right translations are just matrix multiplications from the left, gT, or from the right, Tg. We can take different sets of matrices of different dimensions or operators that satisfy the same commutation relations and provide different

[1] The same is true for right-invariant vector fields. On the other hand, the Lie bracket of any left-invariant vector field with any right-invariant vector field vanishes.

[2] Right-invariant 1-forms are provided by dgg^{-1}.

representations that we denote by the subscript r in $\Gamma_r(T_I)$. Their exponentiation generates a representation of the group,

$$\Gamma_r[g(\sigma)] = e^{\sigma^I \Gamma_r(T_I)}, \qquad [\Gamma_r(T_I), \Gamma_r(T_J)] = f_{IJ}{}^K \Gamma_r(T_K). \tag{A.13}$$

A Lie algebra can be complexified and we can then perform complex changes of basis that complexify the structure constants. If all the structure constants are real (as we will usually take them to be) and all the generators have the same definite Hermiticity properties, then all of them must be anti-Hermitian, $\Gamma_r(T_I)^\dagger = -\Gamma_r(T_I)$, and the elements of the group are represented by unitary operators $\Gamma_r[g(\sigma)]^\dagger = \Gamma_r[g(\sigma)]^{-1}$, i.e. we obtain a *unitary representation*.

Any representation of a *compact* Lie group is equivalent to a unitary representation. The unitary representations of compact Lie groups are also finite-dimensional and the operators can be represented by their matrices in a given basis. All matrix representations automatically satisfy the Jacobi identity.

Sometimes we will be interested in non-compact groups (for instance, the Lorentz group in d dimensions $\mathrm{SO}(1, d-1)$). Their unitary representations are infinite-dimensional and their finite-dimensional representations are necessarily non-unitary. Then not all the generators can be either Hermitian or anti-Hermitian at the same time (with real structure constants). The "non-compact generators" will be represented by Hermitian operators.

In the *adjoint representation* each element $T \in \mathfrak{g}$ is represented by an operator $\Gamma_{\mathrm{Adj}}(T)$ acting on $\mathfrak{g}$ itself according to

$$\Gamma_{\mathrm{Adj}}(T)T' \equiv [T, T'], \quad \forall\ T, T' \in \mathfrak{g}. \tag{A.14}$$

Thus, the generators themselves are represented by

$$\Gamma_{\mathrm{Adj}}(T_I)T_J \equiv [T_I, T_J] = f_{IJ}{}^K T_K \equiv T_K \Gamma_{\mathrm{Adj}}(T_I)^K{}_J, \tag{A.15}$$

so the components of the operator $\Gamma_{\mathrm{Adj}}(T_I)$ in the basis $\{T_I\}$ are the structure constants

$$\Gamma_{\mathrm{Adj}}(T_I)^K{}_J = f_{IJ}{}^K. \tag{A.16}$$

These matrices satisfy the Lie algebra (A.9) due to the Jacobi identity (A.8).

The adjoint representation of the Lie algebra allows us to define the adjoint representation of the group by exponentiation,

$$\Gamma_{\mathrm{Adj}}[g(\sigma)] \equiv \exp\{\sigma^I \Gamma_{\mathrm{Adj}}(T_I)\}, \tag{A.17}$$

and the adjoint action of the group on the algebra,

$$T'_J = T_L \big(\Gamma_{\mathrm{Adj}}[g(\sigma)]\big)^L{}_J. \tag{A.18}$$

An equivalent definition of the adjoint action of the group on the algebra is

$$\Gamma_{\mathrm{Adj}}(g)T \equiv L_{g\,*} R_{g^{-1}\,*} T = gTg^{-1}. \tag{A.19}$$

In any representation we have

$$\Gamma_r(g)\Gamma_r(T_I)\Gamma_r(g^{-1}) = \Gamma_r[\Gamma_{\mathrm{Adj}}(g)T_I] = \Gamma_r(T_J)(\Gamma_{\mathrm{Adj}}(g))^J{}_I. \tag{A.20}$$

We can now introduce the symmetric, bilinear *Killing form* (or *metric*) $K(\cdot,\cdot)$ into the Lie algebra:

$$K(T,T') \equiv \mathrm{Tr}\{\Gamma_{\mathrm{Adj}}(T)\Gamma_{\mathrm{Adj}}(T')\}, \qquad K_{IJ} = K(T_I,T_J) = f_{IK}{}^{L} f_{JL}{}^{K}. \tag{A.21}$$

The Killing metric is invariant under the adjoint action of the group on the algebra due to the cyclic property of the trace. Infinitesimally, we have

$$f_{I(J}{}^{K} K_{L)K} = 0, \tag{A.22}$$

and this is the condition that any other invariant metric should satisfy. We can define

$$f_{IJK} \equiv f_{IJ}{}^{L} K_{LK}, \tag{A.23}$$

which is fully antisymmetric on account of (A.22).

It is easy to see that

$$\mathrm{Tr}[\Gamma_{\mathrm{Adj}}(T_{[I})\Gamma_{\mathrm{Adj}}(T_J)\Gamma_{\mathrm{Adj}}(T_{K]})] = -\tfrac{1}{2} f_{IJK}. \tag{A.24}$$

The Killing metric contains a great deal of information. Let us first make some definitions.

A Lie algebra $\mathfrak{g}$ is Abelian $[T,T']=0,\ \forall\ T,T'\in\mathfrak{g}$. This is sometimes expressed as $[\mathfrak{g},\mathfrak{g}]=0$. Abelian Lie algebras generate by exponentiation Abelian Lie groups.

A subgroup $\mathrm{H}\subset\mathrm{G}$ is an *invariant subgroup* if $ghg^{-1}\in\mathrm{H}\ \forall\ h\in\mathrm{H},\ g\in\mathrm{G}$. A subalgebra $\mathfrak{h}\subset\mathfrak{g}$ is an *invariant subalgebra* or *ideal* if $[M,T]\in\mathfrak{h}\ \forall\ M\in\mathfrak{h},\ T\in\mathfrak{g}$. This is sometimes expressed as $[\mathfrak{h},\mathfrak{g}]\subset\mathfrak{h}$. The Lie algebra $\mathfrak{h}$ of an invariant subgroup $\mathrm{H}\subset\mathrm{G}$ is an invariant subalgebra of the Lie algebra $\mathfrak{g}$ of G.

A Lie group (algebra) is *simple* if it does not have any proper invariant subgroup (subalgebra). Simple Lie algebras generate simple Lie groups.

A Lie group (algebra) is *semisimple* if it does not have any non-trivial invariant Abelian subgroup (subalgebra). Semisimple Lie algebras generate semisimple Lie groups.

For any Lie algebra, the set of all possible Lie brackets of its elements $[\mathfrak{g},\mathfrak{g}]\equiv\mathfrak{g}^{(1)}\equiv\mathfrak{g}_{(1)}$ is an ideal called the *derived subalgebra*. We can define two sequences of ideals $\mathfrak{g}^{(n)}$ and $\mathfrak{g}_{(n)}$:

$$[\mathfrak{g}^{(n-1)},\mathfrak{g}^{(n-1)}]\equiv\mathfrak{g}^{(n)}, \qquad [\mathfrak{g}_{(n-1)},\mathfrak{g}]\equiv\mathfrak{g}_{(n)}. \tag{A.25}$$

A Lie algebra is *solvable* (*nilpotent*) if $\mathfrak{g}^{(n)}$ ($\mathfrak{g}_{(n)}$) is just the 0 element for some n. Every nilpotent algebra is solvable. It can be shown that a Lie algebra is semisimple iff it does not possess any invariant solvable subalgebra.

Now the following can be shown.

Cartan's first criterion $\mathfrak{g}$ is solvable if $g(T,T')=0,\ \forall\ T,T'\in\mathfrak{g}^{(1)}$.

Cartan's second criterion $\mathfrak{g}$ is semisimple iff its Killing form is non-degenerate.

(Weyl) A connected semisimple (linear) Lie group is compact iff its Lie-algebra Killing metric is definite negative.

If the Killing metric of a Lie algebra is 0, then the algebra is solvable.

The Killing metric of a nilpotent Lie algebra is 0.

We can diagonalize and normalize the Killing metric using $\mathrm{GL}(n,\mathbb{R})$ transformations so that it only has $\pm 1, 0$ in the diagonal. The zeros are associated with invariant Abelian subalgebras and the $+1$s with non-compact directions.

A.2 Yang–Mills fields

A.2.1 Fields and covariant derivatives

Fields always transform in finite-dimensional representations of the symmetry group, even if they are not unitary.[3] $\Gamma_r(g)^i{}_j$ $i, j = 1, \ldots, \dim(r)$ denotes the matrix corresponding to group element g in the representation labeled by r. The indices which are also carried by the fields will in general not be shown. In any representation, there are three different types of fields according to the way they transform: *contravariant fields*, represented by a column vector and transforming according to

$$\psi^{i\,\prime} = (\Gamma_r(g))^i{}_j \psi^j, \tag{A.26}$$

covariant fields, represented by a row vector and transforming according to

$$\xi_i' = \xi_j (\Gamma_r^{-1}(g))^j{}_i, \tag{A.27}$$

and Lie-algebra-valued fields that transform under the adjoint action of the group

$$\begin{aligned} \varphi &= \varphi^I \Gamma_r(T_I)\,, \\ \varphi' &= \Gamma_r(g)\varphi\Gamma_r^{-1}(g), \quad \Rightarrow \varphi'^I = \left[\Gamma_{\mathrm{Adj}}(g)\right]^I{}_J \varphi^J. \end{aligned} \tag{A.28}$$

The relation among the three kinds of fields depends on the group and representation we are considering. If the representation r is unitary and ψ is contravariant then $\psi^\dagger$ is covariant. If the group is defined by the property that it preserves the scalar product associated with a metric η $\langle u|v\rangle = u^\dagger \eta v$ so $u' = \Gamma_{\mathrm{v}}(g)u$ and $\Gamma_v^\dagger(g)\eta\Gamma_{\mathrm{v}}(g)$, where $\Gamma_{\mathrm{v}}(g)$ is the matrix associated with the group element g in the defining fundamental or vector representation (these are the groups $\mathrm{SO}(n_+, n_-)$, $\mathrm{SU}(n_+, n_-)$, and $\mathrm{Sp}(n)$) then, given a contravariant vector field ψ, the row vector $\psi^\dagger\eta$ transforms as a covariant vector field. It is also possible to relate contravariant and covariant fields in the spinor representations of $\mathrm{SO}(n_+, n_-)$ groups (see Appendix B).

Since

$$\Gamma_r(g) = \exp\{\sigma^I \Gamma_r(T_I)\} \equiv \exp\{\sigma_r\}, \tag{A.29}$$

for infinitesimal values of the parameters (group manifold coordinates) σ^I, i.e. for transformations near the identity or infinitesimal transformations, the various fields transform as follows:

$$\begin{aligned} \delta_\sigma \psi &= \sigma^I \Gamma_r(T_I)\,\psi = \sigma_r \psi, \\ \delta_\sigma \xi &= -\xi \sigma^I \Gamma_r(T_I) = -\xi\sigma_r, \\ \delta_\sigma \varphi &= [\sigma_r, \varphi], \quad \Rightarrow \delta_\sigma \varphi^I = \sigma^I_{\mathrm{Adj}\,J}\varphi^J. \end{aligned} \tag{A.30}$$

---

[3] Their *solutions* correspond to *states* in the quantum theory and therefore must fit into unitary representations, according to Wigner's theorem, however.

If we now consider *local* (i.e. *gauge*) transformations of the fields with $\sigma^I = \sigma^I(x)$, the derivatives[4] of the fields $\partial_\mu \psi$, $\partial_\mu \xi$, and $\partial_\mu \varphi$ do not transform as do the fields themselves, i.e. they do not transform covariantly under local transformations.[5] It is then necessary to introduce a compensating field A_μ transforming under the adjoint action of the group on the Lie algebra (like φ) to define a new *covariant derivative*. This compensating field is the *gauge field* and can be defined in any representation

$$A_{r\,\mu} = A^I{}_\mu \Gamma_r(T_I). \tag{A.31}$$

With it we define the covariant derivative D_μ by its action on the fields (here g is a coupling constant):

$$\begin{aligned} D_\mu \psi &= \partial_\mu \psi - g A_{r\,\mu} \psi, \\ D_\mu \xi &= \partial_\mu \xi + g \xi A_{r\,\mu}, \\ D_\mu \varphi &= \partial_\mu \varphi - g[A_{r\,\mu}, \varphi], \quad \Rightarrow D_\mu \varphi^I = \partial_\mu \varphi^I - g f_{JK}{}^I A^J{}_\mu \varphi^K. \end{aligned} \tag{A.32}$$

The covariant derivative transforms covariantly under gauge transformations, i.e.

$$\begin{aligned} \left(D_\mu \psi\right)' &= \Gamma_r[g(x)] D_\mu \psi, \\ \left(D_\mu \xi\right)' &= (D_\mu \xi) \Gamma_r^{-1}[g(x)], \\ \left(D_\mu \varphi\right)' &= \Gamma_r[g(x)](D_\mu \varphi)(\Gamma_r[g(x)])^{-1}, \end{aligned} \tag{A.33}$$

if the gauge field transforms as follows:

$$\begin{aligned} A'_{r\,\mu} &= \Gamma_r[g(x)] A_\mu (\Gamma_r[g(x)])^{-1} + \frac{1}{g}(\partial_\mu \Gamma_r[g(x)])(\Gamma_r[g(x)])^{-1}, \\ \delta_\sigma A^I{}_\mu &= \frac{1}{g}\left(\partial_\mu \sigma^I - g f_{JK}{}^I A^J{}_\mu \sigma^K\right), \\ \delta_\sigma A_{r\,\mu} &= \frac{1}{g} D_\mu \sigma_r. \end{aligned} \tag{A.34}$$

Observe that it transforms as φ or σ up to an inhomogeneous term typical of a connection (which is its geometrical meaning). The spin connection $\hat{\omega}_\mu$ defined in Chapter 1 is just the connection for the gauge group $SO(1, d-1)$.

Observe also that the covariant derivatives of contravariant and covariant fields are compatible: if the representation is unitary (and therefore all the generators anti-Hermitian so the real gauge field is anti-Hermitian as well)

$$(D\psi)^\dagger = \partial_\mu \psi^\dagger + g \psi^\dagger A_\mu, \tag{A.35}$$

etc. If we are dealing with a metric-preserving group, then in the vector representation

$$(D\psi)^\dagger \eta = \partial_\mu \psi^\dagger \eta + g \psi^\dagger \eta A_\mu, \tag{A.36}$$

on account of

$$\Gamma_{\rm v}(T_I^\dagger)\eta = -\eta \Gamma_{\rm v}(T_I)\,. \tag{A.37}$$

[4] Partial derivatives should be replaced by derivatives covariant with respect to GCTs.

[5] Sometimes in the literature covariant transformation means a transformation in which there are no derivatives of the local parameters σ^I, which is, obviously, a necessary condition.

A.2.2 Kinetic terms

Now we want to build gauge-invariant actions for the fields ψ, φ, and A_μ (ξ can simply be transformed into a contravariant field). Invariants can be built in several ways. The simplest is to take the product of covariant and contravariant objects that, by definition, transform oppositely. This works both for ψ and for φ (which is once covariant and once contravariant): With our convention for the signature $(+, -, \cdots, -)$, gauge-invariant kinetic terms for ψ and φ are

$$\begin{aligned} & g^{\mu\nu}\,(D_\mu\psi)^\dagger D_\nu\psi, \\ & -g^{\mu\nu}\,\mathrm{Tr}\big(D_\mu\varphi D_\nu\varphi\big) \sim K_{IJ}\,g^{\mu\nu}D_\mu\varphi^I D_\nu\varphi^J. \end{aligned} \tag{A.38}$$

(If the group is not compact, some kinetic terms will have a wrong sign.)

There is no covariant derivative for the gauge field because it does not transform covariantly (due to the inhomogeneous term). The closest to the covariant derivative of the gauge field that we can use to construct a kinetic term is its *gauge field strength*, which is just the curvature of the connection A_μ. We can define it as we defined the Riemann curvature tensor for the Levi-Cività connection in terms of the Ricci identity:

$$\begin{aligned} \big[D_\mu, D_\nu\big]\psi &= -gF_{r\,\mu\nu}\psi, \\ \big[D_\mu, D_\nu\big]\xi &= +g\xi F_{r\,\mu\nu}, , \\ \big[D_\mu, D_\nu\big]\varphi &= -g[F_{r\,\mu\nu}, \varphi]. \end{aligned} \tag{A.39}$$

From these relations we find

$$F_{r\,\mu\nu} = F^I{}_{\mu\nu}\Gamma_r(T_I) = 2\partial_{[\mu}A_{r\,\nu]} - g[A_{r\,\mu}, A_{r\,\nu}], \tag{A.40}$$

or, in components,

$$F^I{}_{\mu\nu} = 2\partial_{[\mu}A^I{}_{\nu]} - gf_{JK}{}^I A^J{}_\mu A^K{}_\nu. \tag{A.41}$$

The gauge field strength transforms under the adjoint action of the group on the algebra (like φ) and, thus,

$$D_\mu F_{r\,\nu\rho} = \partial_\mu F_{r\ \nu\rho} - g[A_{r\,\mu}, F_{r\,\nu\rho}]. \tag{A.42}$$

The gauge field strength always satisfies the *Bianchi identity*[6]

$$D_{[\mu}F_{\nu\rho]} = 0. \tag{A.43}$$

Finally, the kinetic term for A_μ which is invariant is

$$\mathrm{Tr}\big(F_{r\,\mu\nu}F^{r\,\mu\nu}\big) \sim K_{IJ}F^I{}_{\mu\nu}F^{J\mu\nu}, \tag{A.44}$$

where there is a proportionality coefficient that depends on conventions and on the representation r. The Yang–Mills equation of motion is, therefore

$$D_\mu F^{\mu\nu} = 0. \tag{A.45}$$

[6] This can be checked by expanding in powers of g. The first term (of zeroth order in g) vanishes in the absence of torsion, the second due to several cancelations, and the last term $\mathcal{O}(g^2)$ due to the Jacobi identity (A.8).

There is another invariant that we can build in four dimensions,

$$\mathrm{Tr}\,(F_{\mu\nu}{}^{\star}F^{\mu\nu}) \sim -\frac{1}{2\sqrt{|g|}}K_{IJ}\epsilon^{\mu\nu\rho\sigma}F^{I}{}_{\mu\nu}F^{J}{}_{\rho\sigma}, \tag{A.46}$$

but it is a total derivative and does not contribute to the A_μ equations of motion. In the Euclidean signature, the integral of the above term is (up to numerical factors) the *instanton number*, a topological invariant that does contribute to the Euclidean path integral.

Sometimes it is useful to work with differential forms. Thus, we define the Lie-algebra-valued 1- and 2-forms,

$$A \equiv A_\mu dx^\mu, \qquad F = \tfrac{1}{2}F_{\mu\nu}dx^\mu \wedge dx^\nu \equiv dA - A \wedge A, \tag{A.47}$$

where we have defined the exterior covariant derivative D. The kinetic term for A can now be written as the d-form (in d dimensions)

$$\int d^dx\sqrt{|g|}\,\mathrm{Tr}_{\mathrm{Adj}}F^2 \sim \int \mathrm{Tr}_{\mathrm{Adj}}(F \wedge {}^{\star}F), \tag{A.48}$$

and the four-dimensional topological term can be rewritten as

$$\int d^4x\sqrt{|g|}\,\mathrm{Tr}_{\mathrm{Adj}}(F^{\star}F) \sim \int \mathrm{Tr}_{\mathrm{Adj}}(F \wedge F). \tag{A.49}$$

Now we define the *Chern–Simons 3-form*

$$\omega_3 = \frac{1}{3!}\omega_{3\ \mu\nu\rho}dx^\mu \wedge dx^\nu \wedge dx^\rho \equiv \mathrm{Tr}_{\mathrm{Adj}}\big(A \wedge dA - \tfrac{2}{3}A \wedge A \wedge A\big), \tag{A.50}$$

or, in components, using the property Eq. (A.24) and the normalization $K_{IJ} = \delta_{IJ}$ for a compact group:

$$\omega_{3\ \mu\nu\rho} = -3!\big(A^{I}{}_{[\mu}\partial_\nu A^{I}{}_{\rho]} - \tfrac{1}{3}f_{IJK}A^{I}{}_{[\mu}A^{J}{}_{\nu}A^{K}{}_{\rho]}\big). \tag{A.51}$$

The Chern–Simons 3-form has the very important property[7]

$$d\omega_3 = \mathrm{Tr}_{\mathrm{Adj}}(F \wedge F), \tag{A.52}$$

which makes it evident that the topological term $F \wedge F$ is a total derivative.

A.2.3 SO(n_+, n_-) *gauge theory*

The group SO(n_+, n_-) is defined as the group of $n \times n$ (where $n = n_+ + n_-$) real matrices[8] $\hat{\Lambda}^{\hat{a}}{}_{\hat{b}}$ that act on (contravariant) n-dimensional vectors by

$$\hat{V}'^{\hat{a}} = \hat{\Lambda}^{\hat{a}}{}_{\hat{b}}\,\hat{V}^{\hat{b}}, \tag{A.53}$$

[7] To prove it one simply has to realize that the trace over the exterior product of four As vanishes because complete antisymmetry in four indices is the opposite to cyclic symmetry. (For three indices, complete antisymmetry and cyclic symmetry are the same thing and this is why ω_3 can be defined at all.)

[8] We use hats to avoid confusion with (Lorentzian) tangent-space indices. When $n_+ = 1$ and $n_- = d - 1$ they are, of course, identical.

have determinant $+1$, and preserve the metric $\hat{\eta}_{\hat{a}\hat{b}} = \mathrm{diag}(+\cdots+-\cdots-)$ ($n_\pm$ plus (minus) signs), so

$$\hat{V}'^{\hat{a}}\hat{\eta}_{\hat{a}\hat{b}}\,\hat{V}'^{\hat{b}} = \hat{V}^{\hat{a}}\hat{\eta}_{\hat{a}\hat{b}}\,\hat{V}^{\hat{b}}. \tag{A.54}$$

This implies that SO(n_+, n_-) matrices satisfy the defining property

$$\hat{\eta}_{\hat{a}\hat{b}}\,\hat{\Lambda}^{\hat{b}}{}_{\hat{d}}\,\hat{\eta}^{\hat{d}\hat{c}} = (\hat{\Lambda}^{-1})^{\hat{c}}{}_{\hat{a}}, \qquad \hat{\eta}^{\hat{a}\hat{b}}\hat{\eta}_{\hat{b}\hat{c}} = \delta^{\hat{a}}{}_{\hat{c}}. \tag{A.55}$$

This generalizes to arbitrary signature the $n_- = 0$ orthogonality condition $\hat{\Lambda}^{\mathrm{T}} = \hat{\Lambda}^{-1}$.

If we consider also translations in the n-dimensional vector space, we obtain the group ISO(n_+, n_-), which acts on contravariant vectors as follows:

$$\hat{V}'^{\hat{a}} = \hat{\Lambda}^{\hat{a}}{}_{\hat{b}}\,\hat{V}^{\hat{b}} + \hat{W}^{\hat{a}}. \tag{A.56}$$

The Poincaré group in d spacetime dimensions is ISO$(1, d-1)$ in this notation.

We can immediately define the action of SO(n_+, n_-) on covariant vectors $\hat{V}_{\hat{a}}$:

$$\hat{V}'_{\hat{a}} = \hat{V}_{\hat{b}}(\hat{\Lambda}^{-1})^{\hat{b}}{}_{\hat{a}}. \tag{A.57}$$

Using the defining property of SO(n_+, n_-) matrices Eq. (A.55), we can relate covariant and contravariant vectors in the standard way, raising and lowering indices with $\hat{\eta}$:

$$\hat{V}_{\hat{a}} = \hat{\eta}_{\hat{a}\hat{b}}\,\hat{V}^{\hat{b}}, \qquad \hat{V}^{\hat{a}} = \hat{\eta}^{\hat{a}\hat{b}}\,\hat{V}_{\hat{b}}. \tag{A.58}$$

Let us now consider infinitesimal SO(n_+, n_-) transformations $\hat{\Lambda}^{\hat{a}}{}_{\hat{b}} \sim \delta^{\hat{a}}{}_{\hat{b}} + \hat{\sigma}^{\hat{a}}{}_{\hat{b}}$. The defining property of SO(n_+, n_-) matrices in the vector representation Eq. (A.55) implies that the infinitesimal parameters of the transformation satisfy $\hat{\sigma}^{\hat{a}\hat{b}} = \hat{\sigma}^{[\hat{a}\hat{b}]}$ and thus the group has $n(n-1)/2$ independent generators $\hat{M}_{\hat{a}\hat{b}}$ (one for each independent parameter), which are conveniently labeled by an antisymmetric pair of indices $\hat{a}\hat{b}$ (this expresses the fact that the adjoint representation is just the antisymmetric product of two vector representations). We can write

$$\hat{\sigma}^{\hat{a}}{}_{\hat{b}} = \tfrac{1}{2}\hat{\sigma}^{\hat{c}\hat{d}}\,\Gamma_{\mathrm{v}}\Big(\hat{M}_{\hat{c}\hat{d}}\Big)^{\hat{a}}{}_{\hat{b}}, \tag{A.59}$$

where

$$\Gamma_{\mathrm{v}}\Big(\hat{M}_{\hat{c}\hat{d}}\Big)^{\hat{a}}{}_{\hat{b}} = +2\hat{\eta}_{[\hat{c}}{}^{\hat{a}}\hat{\eta}_{\hat{d}]\hat{b}} \tag{A.60}$$

are the SO(n_+, n_-) generators in the vector representation. Observe that we need to divide by two in order to avoid counting the same generator twice. These generators are normalized so that

$$\mathrm{Tr}\Big[\Gamma_{\mathrm{v}}\Big(\hat{M}_{\hat{a}\hat{b}}\Big)\Gamma_{\mathrm{v}}\Big(\hat{M}_{\hat{c}\hat{d}}\Big)\Big] = -4\hat{\eta}_{[\hat{a}\hat{b}][\hat{c}\hat{d}]}. \tag{A.61}$$

The infinitesimal transformations of contravariant and covariant vectors take the forms:

$$\begin{aligned}
\delta_{\hat{\sigma}}\,\hat{V}^{\hat{a}} &= \tfrac{1}{2}\hat{\sigma}^{\hat{c}\hat{d}}\,\Gamma_{\mathrm{v}}\Big(\hat{M}_{\hat{c}\hat{d}}\Big)^{\hat{a}}{}_{\hat{b}}\hat{V}^{\hat{b}},\\
\delta_{\hat{\sigma}}\,\hat{V}_{\hat{a}} &= \hat{V}_{\hat{b}}\Big[-\tfrac{1}{2}\hat{\sigma}^{\hat{c}\hat{d}}\,\Gamma_{\mathrm{v}}\Big(\hat{M}_{\hat{c}\hat{d}}\Big)^{\hat{b}}{}_{\hat{a}}\Big].
\end{aligned} \tag{A.62}$$

Using this representation of the generators, one can find the so(n_+, n_-) algebra

$$\left[\hat{M}_{\hat{a}\hat{b}}, \hat{M}_{\hat{c}\hat{d}}\right] = -\hat{\eta}_{\hat{a}\hat{c}}\hat{M}_{\hat{b}\hat{d}} - \hat{\eta}_{\hat{b}\hat{d}}\hat{M}_{\hat{a}\hat{c}} + \hat{\eta}_{\hat{a}\hat{d}}\hat{M}_{\hat{b}\hat{c}} + \hat{\eta}_{\hat{b}\hat{c}}\hat{M}_{\hat{a}\hat{d}}, \tag{A.63}$$

which can be also be written

$$\left[\hat{M}_{\hat{a}\hat{b}}, \hat{M}_{\hat{c}\hat{d}}\right] = -\hat{M}_{\hat{e}\hat{b}}\Gamma_{\mathrm{v}}\left(\hat{M}_{\hat{c}\hat{d}}\right)^{\hat{e}}{}_{\hat{a}} - \hat{M}_{\hat{a}\hat{e}}\Gamma_{\mathrm{v}}\left(\hat{M}_{\hat{c}\hat{d}}\right)^{\hat{e}}{}_{\hat{b}}. \tag{A.64}$$

These commutation relations can be interpreted as the action of $\hat{M}_{\hat{c}\hat{d}}$ on $\hat{M}_{\hat{a}\hat{b}}$, which transforms as the antisymmetric product of two covariant vectors, indicating that the adjoint representation is the antisymmetric product of two vector representations. This can be seen, for instance, by using the *Eckart–Schrödinger* representation of the iso(n_+, n_-) algebra[9]

$$\hat{M}_{\hat{a}\hat{b}} = \hat{x}_{\hat{a}}\partial_{\hat{b}} - \hat{x}_{\hat{b}}\partial_{\hat{a}} = -\hat{x}^{\hat{c}}\Gamma_{\mathrm{v}}\left(\hat{M}_{\hat{a}\hat{b}}\right)^{\hat{d}}{}_{\hat{c}}\partial_{\hat{d}}, \qquad \hat{P}_{\hat{a}} = -\partial_{\hat{a}}, \tag{A.67}$$

which gives the additional commutator

$$\left[\hat{P}_{\hat{a}}, \hat{M}_{\hat{b}\hat{c}}\right] = -\hat{P}_{\hat{d}}\,\Gamma_{\mathrm{v}}\left(\hat{M}_{\hat{b}\hat{c}}\right)^{\hat{d}}{}_{\hat{a}}, \tag{A.68}$$

indicating that the linear momentum $\hat{P}_{\hat{a}}$ is a covariant SO(n_+, n_-) vector.

The so(n_+, n_-) structure constants are defined by

$$\left[\hat{M}_{\hat{a}\hat{b}}, \hat{M}_{\hat{c}\hat{d}}\right] = \tfrac{1}{2}f_{\hat{a}\hat{b}\ \hat{c}\hat{d}}{}^{\hat{e}\hat{f}}\hat{M}_{\hat{e}\hat{f}}, \quad \Rightarrow \Gamma_{\mathrm{Adj}}\left(\hat{M}_{\hat{a}\hat{b}}\right)^{\hat{e}\hat{f}}{}_{\hat{c}\hat{d}} = f_{\hat{a}\hat{b}\ \hat{c}\hat{d}}{}^{\hat{e}\hat{f}} = 4\Gamma_{\mathrm{v}}\left(\hat{M}_{\hat{a}\hat{b}}\right)^{[\hat{e}}{}_{[\hat{c}}\,\hat{\eta}^{\hat{f}]}{}_{\hat{d}]}, \tag{A.69}$$

where, as we are going to do systematically, we have introduced an additional factor of $\frac{1}{2}$ in order to sum over each generator only once. The Killing metric is

$$\begin{aligned}\hat{K}_{\hat{a}\hat{b}\,\hat{c}\hat{d}} &= \mathrm{Tr}\left[\Gamma_{\mathrm{Adj}}\left(\hat{M}_{\hat{a}\hat{b}}\right)\Gamma_{\mathrm{Adj}}\left(\hat{M}_{\hat{c}\hat{d}}\right)\right] \\ &= (d-2)\,\mathrm{Tr}\left[\Gamma_{\mathrm{v}}\left(\hat{M}_{\hat{a}\hat{b}}\right)\Gamma_{\mathrm{v}}\left(\hat{M}_{\hat{c}\hat{d}}\right)\right] = 4(d-2)\hat{\eta}_{[\hat{a}\hat{b}]\,[\hat{c}\hat{d}]}.\end{aligned} \tag{A.70}$$

Apart from the vector and adjoint representations and other tensor representations (that can be built as tensor products of a number of covariant and contravariant vector representations), SO(n_+, n_-) groups also admit *spinorial representations* that are complex, in general. These are $2^{[n/2]}$-dimensional and can be constructed from a representation of a

---

[9] If we consider infinitesimal Poincaré transformations

$$\delta x^{\mu} = a^{\mu} + \sigma^{\mu}{}_{\nu}x^{\nu} \tag{A.65}$$

of vector fields in a space of signature (n_+, n_-), we find

$$\begin{aligned}\delta V^{\mu} &= \left[a^{\nu}P_{\nu} + \tfrac{1}{2}\sigma^{\alpha\beta}M_{\alpha\beta}\right]V^{\mu} + \tfrac{1}{2}\sigma^{\alpha\beta}\Gamma_{\mathrm{v}}(M_{\alpha\beta})^{\mu}{}_{\rho}V^{\rho}, \\ \delta V_{\mu} &= \left[a^{\nu}P_{\nu} + \tfrac{1}{2}\sigma^{\alpha\beta}M_{\alpha\beta}\right]V_{\mu} - \tfrac{1}{2}\sigma^{\alpha\beta}\Gamma_{\mathrm{v}}(M_{\alpha\beta})^{\rho}{}_{\mu}V_{\rho}.\end{aligned} \tag{A.66}$$

The generators in the Eckart–Schrödinger representation appear in the universal transport term.

Clifford algebra as explained in Appendix B. We denote the so(n_+, n_-) generators in a spinorial representation by

$$\Gamma_s\left(\hat{M}_{\hat{a}\hat{b}}\right)^{\alpha}{}_{\beta}, \qquad \alpha, \beta = 1, \ldots, 2^{[n/2]}. \tag{A.71}$$

Spinors are the elements of the representation space, and they are represented by $2^{[n/2]}$-component contravariant vectors $\hat{\psi}^{\alpha}$ and covariant vectors $\hat{\xi}_{\alpha}$. They transform infinitesimally under SO(n_+, n_-) according to

$$\begin{aligned} \delta_{\hat{\sigma}}\hat{\psi}^{\alpha} &= \tfrac{1}{2}\hat{\sigma}^{\hat{c}\hat{d}}\,\Gamma_s\left(\hat{M}_{\hat{c}\hat{d}}\right)^{\hat{\alpha}}{}_{\hat{\beta}}\hat{\psi}^{\beta}, \\ \delta_{\hat{\sigma}}\hat{\xi}_{\alpha} &= \hat{\xi}_{\beta}\left[-\tfrac{1}{2}\hat{\sigma}^{\hat{c}\hat{d}}\,\Gamma_s\left(\hat{M}_{\hat{c}\hat{d}}\right)^{\hat{\beta}}{}_{\hat{\alpha}}\right]. \end{aligned} \tag{A.72}$$

Covariant and contravariant spinors are related by the operation of *Dirac conjugation*: with each contravariant spinor $\hat{\psi}^{\alpha}$ one can associate a covariant spinor (its *Dirac conjugate*) denoted by $\bar{\hat{\psi}}_{\alpha}$ and related to it by

$$\bar{\hat{\psi}}_{\alpha} = \hat{\psi}^{\beta\,\star}\,\mathcal{D}_{\beta\alpha}, \tag{A.73}$$

where $\mathcal{D}$ is the *Dirac conjugation matrix* and satisfies

$$\mathcal{D}^{-1}\Gamma_s\left(\hat{M}_{\hat{a}\hat{b}}\right)^{\dagger}\mathcal{D} = -\Gamma_s\left(\hat{M}_{\hat{a}\hat{b}}\right). \tag{A.74}$$

There is another conjugation operation that transforms a contravariant spinor into a covariant one: *Majorana conjugation*. More details can be found in Appendix B.

The SO(n_+, n_-) connection, customarily denoted by $\hat{\omega}_{\mu}$ and called *spin connection*, is

$$\hat{\omega}_{\mu} = \tfrac{1}{2}\hat{\omega}_{\mu}{}^{\hat{a}\hat{b}}\Gamma\left(\hat{M}_{\hat{a}\hat{b}}\right), \tag{A.75}$$

and the SO(n_+, n_-)-covariant derivative acting on contravariant fields $\hat{\psi}$ is

$$\hat{\mathcal{D}}_{\mu}\hat{\psi} = \partial_{\mu}\hat{\psi} - \hat{\omega}_{\mu}\hat{\psi}, \tag{A.76}$$

whereas that acting on Lie-algebra-valued fields is

$$\hat{\varphi} = \tfrac{1}{2}\hat{\varphi}^{\hat{a}\hat{b}}\Gamma\left(\hat{M}_{\hat{a}\hat{b}}\right), \qquad \hat{\mathcal{D}}_{\mu}\hat{\varphi} = \partial_{\mu}\hat{\varphi} - \left[\hat{\omega}_{\mu}, \hat{\varphi}\right]. \tag{A.77}$$

$\hat{\psi}$, $\hat{\varphi}$, and the connection $\hat{\omega}_{\mu}$ undergo the following infinitesimal gauge transformations with local parameter $\hat{\sigma}^{\hat{a}\hat{b}}$:

$$\begin{aligned} \delta_{\hat{\sigma}}\hat{\psi} &= \hat{\sigma}\hat{\psi}, \\ \delta_{\hat{\sigma}}\hat{\varphi} &= \left[\hat{\sigma}, \hat{\varphi}\right], \\ \delta_{\hat{\sigma}}\hat{\omega}_{\mu} &= \hat{\mathcal{D}}_{\mu}\hat{\sigma} = \left(\partial_{\mu}\hat{\sigma} - \left[\hat{\omega}_{\mu}, \hat{\sigma}\right]\right), \qquad \hat{\sigma} \equiv \tfrac{1}{2}\hat{\sigma}^{\hat{a}\hat{b}}\Gamma\left(\hat{M}_{\hat{a}\hat{b}}\right), \end{aligned} \tag{A.78}$$

As usual, the curvature is another Lie-algebra-valued field defined through the commutator of two covariant derivatives in any representation:

$$\left[\hat{\mathcal{D}}_\mu, \hat{\mathcal{D}}_\nu\right]\hat{\psi} = -\hat{R}_{\mu\nu}\hat{\psi}, \tag{A.79}$$

where

$$\hat{R}_{\mu\nu} = \tfrac{1}{2}\hat{R}_{\mu\nu}{}^{\hat{a}\hat{b}}\Gamma\left(\hat{M}_{\hat{a}\hat{b}}\right), \qquad \hat{R}_{\mu\nu}{}^{\hat{a}\hat{b}} = 2\partial_{[\mu}\hat{\omega}_{\nu]}{}^{\hat{a}\hat{b}} - 2\hat{\omega}_{[\mu|}{}^{\hat{a}\hat{c}}\,\hat{\omega}_{|\nu]\hat{c}}{}^{\hat{b}}, \tag{A.80}$$

transforms as $\hat{\sigma}$ and satisfies the Bianchi identities

$$\hat{\mathcal{D}}_{[\mu}\hat{R}_{\nu\rho]} = 0. \tag{A.81}$$

SO(3) *and three-dimensional real Lie algebras.* For the particular case $n = 3$ the adjoint representation coincides with the vector representation. Then we have two different notations (with one and with two antisymmetric indices) for the same representation. The relation between the two of them is

$$T_i = \tfrac{1}{2}\epsilon_{ijk}T_{jk}, \qquad T_{ij} = \epsilon_{ijk}T_k, \tag{A.82}$$

and the T_is satisfy the algebra

$$[T_i, T_j] = -\epsilon_{ijk}\eta^{kl}T_l, \quad \Rightarrow f_{ij}{}^l = -\epsilon_{ijk}\eta^{kl}, \quad \Rightarrow K_{ij} = -2\eta_{ij}. \tag{A.83}$$

For $\eta^{kl} = \delta^{kl}$ (for the group SO(3)) an explicit representation is given in Eq. (A.94). The only other possibility, $\eta = \text{diag}\,(+ + -)$, is SO(2, 1). They are identical as complex algebras (it suffices to multiply T_1 and T_2 by i), but not as real algebras.

All the possible real three-dimensional Lie algebras can be written in terms of a matrix Q^{kl}.

$$[T_i, T_j] = -\epsilon_{ijk}\mathsf{Q}^{kl}T_l, \qquad \mathsf{Q}^{(lk)}\epsilon_{kij}\mathsf{Q}^{ij} = 0. \tag{A.84}$$

If we make the separation $\mathsf{Q}^{lk} = \mathsf{Q}^{(kl)} - \epsilon^{kli}a_i$, the constraint on Q is just $\mathsf{Q}^{(kl)}a_l = 0$. $\mathsf{Q}^{(kl)}$ can be diagonalized and its eigenvectors a_i can be found, and all the three-dimensional Lie algebras (nine in total) can be classified. This is the Bianchi classification (see e.g. [640]). The only semisimple ones are those of SO(3) and SO(2, 1).

A.3 Riemannian geometry of group manifolds

We can define Riemannian metrics on Lie groups. The most interesting ones are those invariant under the left- and right-translation diffeomorphisms (*bi-invariant metrics*). If B_{IJ} are the components of a non-singular metric in the basis of left-invariant vector fields $\{e_I\}$, then

$$ds^2 = B_{IJ}e^I \otimes e^J, \tag{A.85}$$

where the e^I are a basis of left-invariant 1-forms constructed for instance as in Eq. (A.12), is automatically a metric invariant under the left-translation diffeomorphisms and has an isometry group G. Under right translations $g \to gh$

$$e^I \to \Gamma_{\text{Adj}}(h^{-1})^I{}_J e^J, \quad \Rightarrow B_{IJ} \to \Gamma_{\text{Adj}}(h^{-1})^K{}_I\Gamma_{\text{Adj}}(h^{-1})^L{}_J B_{KL}. \tag{A.86}$$

Infinitesimally, B_{IJ} will be invariant if the analog of Eq. (A.22) holds. In that case the metric will have an isomorphism group of G × G (or smaller, if some left and right actions coincide).

For semisimple groups it is natural to use the Killing metric. If we diagonalize it and normalize it so that its diagonal contains only $+1$s and -1s, then the metric is just η_{IJ} and a metric-compatible connection will be an SO(n_+, n_-) connection, $\omega^{IJ} = -\omega^{JI}$. In the absence of torsion, the connection 1-form can be determined by comparing the Maurer–Cartan equations (A.10) with the structure equations (1.143),[10]

$$\omega^I{}_J = \tfrac{1}{2} e^K f_{KJ}{}^I = \tfrac{1}{2} e^K \Gamma_{\text{Adj}}(T_K)^I{}_J. \tag{A.87}$$

The curvature 2-form is given simply by

$$R^I{}_J = \tfrac{1}{8} e^K \wedge e^L f_{KL}{}^M f_{MJ}{}^I, \tag{A.88}$$

and all its components in this basis are constant and completely determined by the structure constants. The Ricci tensor is proportional to the Killing metric. Furthermore, repeated use of the Jacobi identity shows that the curvature is covariantly constant,

$$\nabla_I R_{JKLM} = 0. \tag{A.89}$$

Let us consider a simple but useful example:

A.3.1 Example: the SU(2) *group manifold*

SU(2) matrices U ($U^\dagger = U^{-1}$, $\det U = +1$) can be parametrized by $z_0, z_1 \in \mathbb{C}$,

$$U \equiv \begin{pmatrix} z_0 & z_1 \\ -\bar{z}_1 & \bar{z}_0 \end{pmatrix}, \qquad |z_0|^2 + |z_1|^2 = 1, \tag{A.90}$$

so SU(2) has the topology of S^3. We can parametrize both by the *Euler angles* $\{\theta, \varphi, \psi\}$,

$$z_0 = \cos\left(\frac{\theta}{2}\right) e^{i(\varphi+\psi)/2}, \qquad z_1 = \sin\left(\frac{\theta}{2}\right) e^{i(\varphi-\psi)/2}, \tag{A.91}$$

where[11] $\theta \in [0, \pi]$, $\varphi \in [0, 2\pi]$, and $\psi \in [0, 4\pi]$, whose main property is that we can construct any general SU(2) rotation as the product of three rotations:

$$U(\varphi, \theta, \psi) = U(\varphi, 0, 0) U(0, \theta, 0) U(0, 0, \psi). \tag{A.92}$$

The left-invariant Maurer–Cartan 1-forms e^i, $i = 1, 2, 3$, are

$$U^{-1} dU \equiv -e^i T_i, \tag{A.93}$$

where the T_is are the anti-Hermitian generators of the su(2) Lie algebra Eq. (A.83) (with $\eta^{kl} = \delta^{kl}$),

$$T_i = \frac{i}{2} \sigma^i. \tag{A.94}$$

[10] If the metric is not $\eta_{IJ} = (+\cdots, -\cdots)$ then $\omega^I{}_J = e^K f_{KJ}{}^I + e^K K_{KJ}{}^I$, where $K_{KJ}{}^I = K_{JK}{}^I$ has to be determined case by case.

[11] On imposing the restriction $\psi \in [0, 2\pi]$, we obtain $\mathbb{RP}^3$, which is homeomorphic to SO(3).

With the Euler-angles parametrization they are

$$\begin{aligned} e^1 &= \sin\psi\, d\theta - \sin\theta\cos\psi\, d\varphi, \\ e^2 &= -(\cos\psi\, d\theta + \sin\theta\sin\psi\, d\varphi), \\ e^3 &= -(d\psi + \cos\theta\, d\varphi), \end{aligned} \tag{A.95}$$

and it can easily be checked that they satisfy the Maurer–Cartan equation

$$de^i = -\tfrac{1}{2}\epsilon_{ijk}e^j \wedge e^k. \tag{A.96}$$

Using the Killing metric we can construct a bi-invariant (that is $\mathrm{SU}(2) \times \mathrm{SU}(2) \sim \mathrm{SO}(4)$ -invariant) metric on SU(2) (S^3). On normalizing to obtain the volume of S^3, we obtain[12]

$$d\Omega^2_{(3)} = \tfrac{1}{4}\big[(e^1)^2 + (e^2)^2 + (e^3)^2\big] = \tfrac{1}{4}\big[d\Omega^2_{(2)} + (e^3)^2\big]. \tag{A.97}$$

A.4 Riemannian geometry of homogeneous and symmetric spaces

We can define the action of a (*transformation*) group G on a space M as a continuous map

$$\begin{aligned} \mathrm{G} \times \mathrm{M} &\to \mathrm{M} \\ (g, x) &\to gx \end{aligned} \tag{A.98}$$

such that

$$g_1(g_2x) = (g_1g_2)x, \quad ex = x, \quad \forall\ g_1, g_2 \in G,\ x \in M. \tag{A.99}$$

Each $g \in \mathrm{G}$ induces a homeomorphism of M into M. G is said to act *transitively* on M if, given any two points of M, there is always a transformation of the group that relates them. M is then a *homogeneous space*. The subgroup $\mathrm{H} \subset \mathrm{G}$ that leaves a given point invariant is the *isotropy* group of that point. The isotropy groups of all points of M are isomorphic and we can talk about the isotropy group of M. Then, it can be shown that M and the coset space G/H defined by the equivalence classes under right multiplication by elements of H, sometimes denoted by $\{g\mathrm{H}\}$, are homeomorphic. Observe that G acts from the left on these equivalence classes.

It can also be shown that, if H is topologically closed, the coset space G/H can be given the structure of a manifold of dimension $\dim \mathrm{G} - \dim \mathrm{H}$ on which G acts transitively. A manifold M that is a homogeneous space is always diffeomorphic to the coset manifold G/H, which is what we will mean by homogeneous space henceforth.

A very important theorem states that G can be seen as a *principal bundle* with base space G/H, structure group H, and projection $\mathrm{G} \to \mathrm{G/H}$.

The Lie algebra of any homogeneous space G/H can be decomposed as the direct sum of vector spaces $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{p}$, where $\mathfrak{h}$ is the Lie subalgebra of H and $\mathfrak{k}$ is its orthogonal complement. Since $\mathfrak{h}$ is a subalgebra

$$[\mathfrak{h}, \mathfrak{h}] \subset \mathfrak{h}. \tag{A.100}$$

[12] See Appendix C for the definitions of spheres and their volumes.

G/H is *reductive* if

$$[\mathfrak{k}, \mathfrak{h}] \subset \mathfrak{k}, \tag{A.101}$$

which means that $\mathfrak{k}$ is a representation of H.

G/H is *symmetric* if it is reductive and

$$[\mathfrak{k}, \mathfrak{k}] \subset \mathfrak{h}. \tag{A.102}$$

The pair $(\mathfrak{k}, \mathfrak{h})$ is then a *symmetric pair*.[13] The two components of a symmetric pair are mutually orthogonal with respect to the Killing metric which is block-diagonal.

If G/H is reductive and $\mathfrak{k}$ is a subalgebra,

$$[\mathfrak{k}, \mathfrak{k}] \subset \mathfrak{k} \tag{A.103}$$

(so it is an ideal), then $\mathfrak{g}$ is the *semidirect sum* of $\mathfrak{h}$ and $\mathfrak{k}$, and G is the *semidirect product* of the corresponding subgroups $\mathrm{G} = \mathrm{H} \ltimes \mathrm{K}$.

By construction, there is a natural transitive action of G on homogeneous spaces and it is natural to define on them Riemannian metrics that are invariant under the left action of G. If the isotropy group H of a homogeneous (not necessarily symmetric or reductive) space G/H is compact, then there is always at least one G-invariant Riemannian metric. If G/H is symmetric, G is connected, and H is compact, and it is equipped with the G-invariant metric, then G/H is a *(Riemannian) globally symmetric* space.

Thus, these homogeneous spaces with a G-invariant metric have an isometry group G that acts transitively from the left. If they are reductive, their Riemann curvature tensor is covariantly constant[14] as in Eq. (A.89).

We have already seen a particular example of homogeneous reductive spaces with a G-invariant Riemannian metric: group manifolds equipped with a bi-invariant metric on $\mathfrak{g}$. They have a trivial isotropy subgroup $\mathrm{H} = e$ and are trivial coset manifolds. They are clearly reductive but not symmetric because $\mathfrak{k} = \mathfrak{g}$ and then $[\mathfrak{k}, \mathfrak{k}] = \mathfrak{k}$. The isometry group is the product of the left isometry group G and the right isometry group which is also G.[15]

Now, we are going to show a procedure by which to construct G-invariant Riemannian metrics on homogeneous spaces. Let us introduce some notation: we denote by $\{M_i\}$ $(i, j = 1, \ldots, \dim \mathrm{H})$ a basis of $\mathfrak{h}$, and by $\{P_a\}$ $(a, b = 1, \ldots, d = \dim \mathrm{G} - \dim \mathrm{H})$ a basis of $\mathfrak{k}$. By exponentiating the generators of $\mathfrak{k}$ we can construct a coset representative $u(x) = u(x^1, \ldots, x^d)$. We can construct the coset representative as a product of generic elements of the one-dimensional subgroups generated by the P_as:

$$u(x) = e^{x^1 P_1} \cdots e^{x^q P_q}. \tag{A.104}$$

Under a left transformation $g \in \mathrm{G}$, u transforms into another element of G, which becomes a coset representative $u(x')$ only after a right transformation with an element $h \in \mathrm{H}$,

---

[13] This decomposition can be characterized by an involutive $(\sigma^2 = 1)$ automorphism σ of $\mathfrak{g}$ such that $\sigma(T) = +T,\ \forall\ T \in \mathfrak{h}$ and $\sigma(T) = -T,\ \forall\ T \in \mathfrak{k}$.

[14] These spaces should not be confused with *spaces of constant curvature*, that have $R_{IJKL} = K\delta_{IJ\,KL}$. These are a particular type of symmetric space that has the maximal number of Killing vectors allowed $(\frac{1}{2}d(d+1))$. They are also known as *maximally symmetric spaces*.

[15] In general, the right isometry group of G/H will be $N(\mathrm{H})/\mathrm{H}$, where $N(\mathrm{H})$ is the normalizer of H.

which is a function of g and x:

$$gu(x) = u(x')h. \tag{A.105}$$

Now we construct the left-invariant Maurer–Cartan 1-form and expand it in *horizontal*, e^a, and *vertical*, ϑ^i, components:

$$V \equiv -u^{-1}du = e^a P_a + \vartheta^i M_i. \tag{A.106}$$

The horizontal components e^a can be used as Vielbeins for G/H and, given any metric B_{ab} on $\mathfrak{k}$, we can construct a Riemannian metric

$$ds^2 = B_{ab} e^a \otimes e^b. \tag{A.107}$$

To find under what conditions this metric will be (left-)G-invariant, we have to look at the transformation of the Maurer–Cartan 1-forms under left multiplication by a constant element $g \in \mathrm{G}$, $u(x') = gu(x)h^{-1}$:

$$\begin{aligned} e^a(x') &= (he(x)h^{-1})^a = \Gamma_{\mathrm{Adj}}(h)^a{}_b e^b(x), \\ \vartheta^i(x') &= (h\vartheta(x)h^{-1})^i + (h^{-1}dh)^i + (he(x)h^{-1})^i. \end{aligned} \tag{A.108}$$

The last term in the second equation is zero in the reductive case and the ϑ^is transform as a connection. Furthermore, the restriction of $\Gamma_{\mathrm{Adj}}(h)$ to $\mathfrak{k}$ is a representation of $\mathfrak{h}$. The Riemannian metric will be invariant under the left action of G if

$$f_{i(a}{}^c B_{b)c} = 0, \tag{A.109}$$

which is guaranteed if we can set $B_{ab} = K_{ab}$, the projection on $\mathfrak{k}$ of the (non-singular) Killing metric. There are many important cases in which G is not semisimple but there is a non-degenerate invariant metric. For instance, we can describe Minkowski space as the quotient of the Poincaré group (which is not semisimple because it contains the Abelian invariant subgroup of translations) by the Lorentz subgroup. The Minkowski metric is a non-degenerate invariant metric for this coset. Another example is provided by the Hpp-wave spacetimes constructed in Section 10.1.1.

The resulting Riemannian metric contains G in its isometry group (generically $\mathrm{G} \times N(\mathrm{H})/\mathrm{H}$) and must admit n Killing vector fields $k_{(I)}$. The Killing vectors $k_{(I)}$ and the *H-compensator* $W_I{}^i$ are defined through the infinitesimal version of $gu(x) = u(x')h$ with

$$\begin{aligned} g &= 1 + \sigma^I T_I, \\ h &= 1 - \sigma^I W_I{}^i M_i, \\ x^{\mu\,\prime} &= x^\mu + \sigma^I k_{(I)}{}^\mu. \end{aligned} \tag{A.110}$$

Using these equations in

$$u(x + \delta x) = u(x) + \sigma^I k_{(I)} u, \tag{A.111}$$

we obtain

$$T_I u = k_{(I)} u - u W_I{}^i M_i. \tag{A.112}$$

Acting with u^{-1} on the left and using the definitions of the adjoint action and the Maurer–Cartan 1-forms, we obtain

$$T_J \Gamma_{\mathrm{Adj}}(u^{-1})^J{}_I = -k_{(I)}{}^a P_a - (k_{(I)}{}^\mu \vartheta^i{}_\mu + W_I{}^i) M_i, \tag{A.113}$$

which, projected onto the horizontal and vertical subspaces, gives

$$k_{(I)}{}^a = -\Gamma_{\mathrm{Adj}}(u^{-1}(x))^a{}_I, \tag{A.114}$$
$$W_I{}^i = -k_{(I)}{}^\mu \vartheta^i{}_\mu - \Gamma_{\mathrm{Adj}}(u^{-1}(x))^i{}_I. \tag{A.115}$$

The Killing vectors associated with the right isometry group $N(\mathrm{H})/\mathrm{H}$ are just the vectors e_a dual to the horizontal Maurer–Cartan 1-forms in the directions of $N(\mathrm{H})/\mathrm{H}$.

These formulae simplify considerably the calculation of Killing vectors, if we construct the space with the above recipe. As in group manifolds, the spin connection[16] can easily be found: on comparing the Maurer–Cartan equations

$$de^a - \vartheta^i \wedge e^b f_{ib}{}^a - \tfrac{1}{2} e^b \wedge e^c f_{bc}{}^a = 0 \tag{A.116}$$

with the structure equation (1.143), we obtain

$$\omega^a{}_b = \vartheta^i f_{ib}{}^a + \tfrac{1}{2} e^c f_{cb}{}^a, \tag{A.117}$$

if we do not allow for torsion, or

$$\omega^a{}_b = \vartheta^i f_{ib}{}^a, \qquad T^a = -\tfrac{1}{2} e^c \wedge e^b f_{cb}{}^a. \tag{A.118}$$

It is straightforward to compute the curvature using the Maurer–Cartan equations:

$$d\vartheta^i - \tfrac{1}{2}\vartheta^j \wedge \vartheta^k f_{jk}{}^i - \tfrac{1}{2} e^a \wedge e^b f_{ab}{}^i = 0. \tag{A.119}$$

In the symmetric case ($f_{cb}{}^a = 0$) and in the reductive case with the torsionful connection Eq. (A.118)

$$R^a{}_b = [d\vartheta^i - \tfrac{1}{2}\vartheta^j \wedge \vartheta^k f_{jk}{}^i] f_{ib}{}^a = \tfrac{1}{2} e^c \wedge e^d f_{cd}{}^i f_{ib}{}^a \tag{A.120}$$

(using the Maurer–Cartan equations) and is covariantly constant. In the reductive (non-symmetric) case

$$R^a{}_b = \tfrac{1}{2} e^c \wedge e^d \big(f_{cd}{}^i f_{ib}{}^a + \tfrac{1}{2} f_{cd}{}^e f_{eb}{}^a - \tfrac{1}{2} f_{ce}{}^a f_{db}{}^e \big). \tag{A.121}$$

The reductive case (symmetric or not) is particularly interesting because, as we have said, according to Eqs. (A.108) the vertical 1-forms ϑ^i transform as a connection for the group H. The above formulae Eqs. (A.117) and (A.118) relate this gauge connection to the spin connection (and torsion). Sometimes this is expressed by saying that the gauge group has been embedded into the tangent-space group. These relations are used very often in the construction of solutions. This suggests the following definitions.

---

[16] We assume here that B_{ab} is diagonal with only $+1$s and -1s on the diagonal.

A.4.1 H-*covariant derivatives*

The H-*covariant derivative* of any object that transforms contravariantly, $\phi' = \Gamma_r(h)\phi$, or covariantly, $\psi' = \psi\Gamma_r(h^{-1})$ (for instance, $u(x)$ itself), in the representation r of H is

$$\mathcal{D}_\mu\phi \equiv \partial_\mu\phi - \vartheta_\mu{}^i\Gamma_r(M_i)\phi, \qquad \mathcal{D}_\mu\psi \equiv \partial_\mu\psi + \psi\vartheta_\mu{}^i\Gamma_r(M_i). \tag{A.122}$$

The curvature F^i is

$$F^i = d\vartheta^i - \tfrac{1}{2}\vartheta^j \wedge \vartheta^k f_{jk}{}^i \tag{A.123}$$

and it is covariantly constant with respect to the full (Lorentz-plus-gauge) covariant derivative if we use the torsional spin connection Eq. (A.118) and also in the symmetric case. This statement is clearly equivalent to the covariant constancy of the Lorentz curvature.

This implies that, in any reductive coset space G/H, there is a solution of the Yang–Mills equations of motion for the group H in the curved geometry associated with the (torsionful) spin connection defined above. This result is implicitly or explicitly used in many places. The simplest example is provided by the coset manifold SU(2)/U(1), which gives a round 2-sphere. The U(1) connection solves the Maxwell equations and corresponds to the Dirac monopole (see the calculation of the Robinson–Bertotti superalgebra on page 386 and Appendix C.1).

The gauge field ϑ^i is invariant under the combination of the diffeomorphisms generated by the Killing vectors $k_{(I)}$ and gauge transformations generated by $W_I{}^i$, i.e.

$$-\mathcal{L}_{k_{(I)}}\vartheta^i = \mathcal{D}_\mu W_I{}^i, \tag{A.124}$$

where $\mathcal{L}_{k_{(I)}}$ is the standard Lie derivative.

The H-*covariant Lie derivative with respect to the Killing vectors*[17] $k_{(I)}$ of contravariant (ϕ) or covariant (ψ) objects in the representation r of H is

$$\mathbb{L}_{k_{(I)}}\phi \equiv \mathcal{L}_{k_{(I)}}\phi + W_I{}^i\Gamma_r(M_i)\phi, \qquad \mathbb{L}_{k_{(I)}}\psi \equiv \mathcal{L}_{k_{(I)}}\psi - \psi W_I{}^i\Gamma_r(M_i). \tag{A.125}$$

This Lie derivative satisfies, among other properties

$$[\mathbb{L}_{k_{(I)}}, \mathbb{L}_{k_{(J)}}] = \mathbb{L}_{[k_{(I)},k_{(J)}]}, \tag{A.126}$$

$$\mathbb{L}_{k_{(I)}}e^a = 0, \tag{A.127}$$

$$\mathbb{L}_{k_{(I)}}u = \mathcal{L}_{k_{(I)}}u - uW_I{}^iM_i = T_Iu, \tag{A.128}$$

where the last property follows from Eqs. (A.115) and (A.112). The connection 1-forms ϑ^i are not covariant or contravariant objects and this definition does not apply to them. The best one can do for them is to combine the standard Lie derivative and a compensating gauge transformation. The resulting operator acting on ϑ^i is identically zero, due to Eq. (A.124).

[17] H-covariant Lie derivatives can be defined with respect to any vector, but the property Eq. (A.126) holds only for Killing vectors. The spinorial Lie derivative [632, 633, 655] and the more general Lie–Lorentz and Lie–Maxwell derivatives that appear in calculations of supersymmetry algebras [390, 748] discussed in Section 13.2.1 can actually be seen as particular examples of this more general operator (see e.g. [460]) and, actually, are identical objects when they are acting on Killing spinors of maximally supersymmetric spacetimes [25].

A.4.2 Example: round spheres

The n-dimensional sphere S^n (see Appendix C) is a homogeneous topological space on which the orthogonal group $SO(n+1)$ acts transitively.[18] Any point is invariant under rotations around the axis that crosses that point: the isotropy group is thus $SO(n)$, and S^n is therefore homeomorphic to $SO(n+1)/SO(n)$. If the $SO(n+1)$ generators are $\{\hat{M}_{\hat{a}\hat{b}}\}$, $\hat{a}, \hat{b} = 1, \ldots, n+1$, the generators of $SO(n)$ can be chosen as $\{M_{ab} \equiv \hat{M}_{ab}\}$ $a, b = 1, \ldots, n$ and those of the orthogonal complement as $P_a \equiv M_{n+1\,a}$. We immediately see that S^n is a symmetric space:

$$[P_a, M_{bc}] = 2\delta_{a[c}P_{b]} = -P_d\Gamma_{\mathrm{v}}(M_{bc})^d{}_a, \qquad [P_a, P_b] = M_{ab}. \tag{A.129}$$

To construct an $SO(n+1)$-symmetric Riemannian metric,[19] we first construct a coset representative u as above and then the Maurer–Cartan 1-form V:

$$\begin{aligned} -V &= P_n dx^n + e^{-x^n P_n} P_{n-1} e^{x^n P_n} dx^{n-1} \\ &\quad + e^{-x^n P_n} e^{x^{n-1} P_{n-1}} P_{n-2} e^{x^n P_{n-1}} e^{x^n P_n} dx^{n-2} \cdots. \end{aligned} \tag{A.130}$$

Here we can use repeatedly the formula

$$[X, Y] = Z, \quad [Y, Z] = X, \quad [Z, X] = Y, \quad \Rightarrow e^{aX} Y e^{-aX} = \cos(a)\, Y + \sin(a)\, Z, \tag{A.131}$$

for the triplets P_a, P_b, and M_{ab}, or, far better, the definition of the adjoint action:

$$e^{-x^n P_n} P_{n-1} e^{x^n P_n} = \tfrac{1}{2}\hat{M}_{\hat{a}\hat{b}} \Gamma_{\mathrm{Adj}}(e^{-x^n P_n})^{\hat{a}\hat{b}}{}_{n-1}. \tag{A.132}$$

In both cases, the result is

$$\begin{aligned} -e &= P_n dx^n + P_{n-1}\cos x^n\, dx^{n-1} + P_{n-2}\cos x^n \cos x^{n-1}\, dx^{n-2} + \cdots, \\ -\vartheta &= \sin x^n \sum_{a=1}^{a=n-1} M_{n\,a} dx^a + \cos x^n \sin x^{n-1} \sum_{a=1}^{a=n-2} M_{n-1\,a} dx^a \\ &\quad + \cos x^n \cos x^{n-1} \sin x^{n-2} \sum_{a=1}^{a=n-3} M_{n-2\,a} dx^a + \cdots. \end{aligned} \tag{A.133}$$

Using the $SO(n)$-invariant metric[20] δ_{ab}, we obtain the $SO(n+1)$-invariant metric

$$ds^2 = (dx^n)^2 + \cos^2 x^n \left[(dx^{n-1})^2 + \cos^2 x^{n-1}\left[(dx^{n-2})^2 + \cdots\right.\right. . \tag{A.134}$$

On comparing this with the metric Eq. (C.4) in standard spherical coordinates, we see that the coset coordinates that we have used are related to them by $x^n = \theta_{n-1} + \pi/2, \ldots, x^1 = \varphi$. The spin connection is given by

$$\omega^a{}_b = \tfrac{1}{2}\vartheta^{cd} f_{cd\,n+1b}{}^{n+1a} = \tfrac{1}{2}\vartheta^{cd}\Gamma_{\mathrm{v}}(M_{cd})^a{}_b = \vartheta^a{}_b, \tag{A.135}$$

---

[18] $SO(n+1)$ rotates in the standard form the coordinates of the ambient space $\mathbb{R}^{n+1}$, respecting the defining equation $(x^1)^2 + \cdots + (x^{n+1})^2 = 1$.

[19] With $\frac{1}{2}n(n+1)$ isometries, spheres with this metric (*round spheres*) are also maximally symmetric spaces.

[20] On rescaling some entries of δ_{ab}, one obtains *squashed spheres* (see Appendix C.2) that have less symmetry.

and the curvature, which corresponds to a maximally symmetric space, by

$$R_{cd}{}^{a}{}_{b} = \tfrac{1}{2} f_{n+1c\,n+1d}{}^{ef} f_{ef\,n+1b}{}^{n+1a} = \tfrac{1}{2}\delta_{cd}{}^{ef}\Gamma_{\mathrm{v}}(M_{ef})^{a}{}_{b} = \delta_{[c}{}^{a}\delta_{d]b}. \qquad \text{(A.136)}$$

We can construct AdS_d spacetimes in an almost identical way as the coset manifolds $\mathrm{SO}(2, d-1)/\mathrm{SO}(1, d-1)$ with $P_a \equiv \hat{M}_{-1\,a}$ $a = 0, 1, \ldots, d-1$ and $\{M_{ab} \equiv \hat{M}_{ab}\}$ using the metric η_{ab} on $\mathfrak{k}$. The $d = 4$ case is worked out in Section 13.3.3 and the $d = 2$ case in Section 13.3.4, page 386.

Appendix B
Gamma matrices and spinors

In this appendix we explain our conventions for gamma matrices in diverse dimensions and their relations (via dimensional reduction). We start by reviewing basic facts about spinors and gamma matrices in diverse dimensions. At the end we review spinors and gamma matrices in spaces of arbitrary dimensions and signatures.

B.1 Generalities

Let us first review some facts about gamma matrices.[1] Gamma matrices are the generators of the d-dimensional *Clifford algebra* associated with the metric $\eta_{ab} = \mathrm{diag}\,(+-\cdots-)$, $a, b = 0, \ldots, d-1$ and, therefore, satisfy the anticommutation relations

$$\{\Gamma_a, \Gamma_b\} = +2\eta_{ab}. \tag{B.1}$$

Any other element of the Clifford algebra can be constructed as a linear combination of the gamma matrices and their products.

Clifford algebras are relevant in physics due to the fact that a representation of the d-dimensional Clifford algebra for the above metric η_{ab} can be used to construct a representation of the d-dimensional Lorentz algebra so$(1, d-1)$,

$$[M_{ab}, M_{cd}] = -\eta_{ac}M_{bd} - \eta_{bd}M_{ac} + \eta_{ad}M_{bc} + \eta_{bc}M_{ad}, \tag{B.2}$$

that we denote by Γ_{s} by taking antisymmetric products of two gamma matrices:

$$\Gamma_{\mathrm{s}}(M_{ab}) = \tfrac{1}{2}\Gamma_{ab}, \qquad \Gamma_{ab} \equiv \Gamma_{[a}\Gamma_{b]}. \tag{B.3}$$

(We use the notation

$$\Gamma^{a_1\cdots a_n} = \Gamma^{[a_1}\Gamma^{a_2}\cdots\Gamma^{a_n]} \tag{B.4}$$

for the antisymmetrized (with weight unity) product of n gamma matrices.)

[1] We will follow [913] but using our mostly minus-signature metric. See also [221, 923, 947].

Lorentz transformations in this representation are constructed by exponentiation[2]

$$\Gamma_s(\Lambda) = \exp\left\{\tfrac{1}{2}\sigma^{ab}\Gamma_s(M_{ab})\right\}. \tag{B.5}$$

How many different representations can be built in this way? It can be shown that

> There is only one (physically[3]) inequivalent irreducible representation of the Clifford algebra in d dimensions and it is $2^{[d/2]}$-dimensional.

The corresponding representation of the Lorentz algebra Γ_s is a *spinorial representation* and the elements of the complex $2^{[d/2]}$-dimensional vector representation space are called (Dirac) *spinors*. We use the first Greek letters as indices in this vector space: $\alpha, \beta = 1, \ldots, 2^{[d/2]}$.

It is worth stressing that, even if we use an irreducible representation of the Clifford algebra, the corresponding representation of the Lorentz group is reducible for d even. In that case, the representation space is the direct sum of two subspaces of dimension $2^{[d/2]-1}$ whose elements are called *Weyl spinors* and will be discussed later.

We consider only unitary representations. The definition of the algebra means that, if the gamma matrices are unitary, Γ^0 is Hermitian and the rest of the Γ^i are anti-Hermitian:

$$\Gamma^{0\,\dagger} = +\Gamma^0, \qquad \Gamma^{i\,\dagger} = -\Gamma^i, \quad i = 1, \ldots, d-1. \tag{B.6}$$

This implies

$$\Gamma^0\Gamma^a\Gamma^0 = \Gamma^{a\,\dagger}. \tag{B.7}$$

In any representation, all gamma matrices are *traceless*. This can be seen by considering $\mathrm{tr}\left(\Gamma^a\Gamma^b\Gamma^a\right)$ with $a \neq b$ and using the anticommutators and the cyclic property of the trace.

We can prove the existence of a $2^{[d/2]}$-dimensional representation of the Clifford algebra by explicit construction as in [221]. This will provide us with gamma matrices in any dimension. Let us first consider the case d even. We can proceed by induction: if we assume that a representation of the $(d-2)$-dimensional gamma matrices $\{\Gamma^a_{(d-2)}\}$, $a = 0, \ldots, d-3$ exists and that it is $2^{[(d-1)/2]}$-dimensional, then

$$\Gamma^a_{(d)} = \Gamma^a_{(d-2)} \otimes \sigma^1, \qquad \Gamma^{d-2}_{(d)} = \mathbb{I} \otimes i\sigma^2, \qquad \Gamma^{d-1}_{(d)} = \mathbb{I} \otimes i\sigma^3, \tag{B.8}$$

---

[2] The Lorentz group O $(1, d-1)$ is neither connected nor simply connected for $d \geq 3$. Exponentiation of the generators gives only transformations in the component of the group manifold connected with the identity (the proper (determinant $+1$), orthochronous ($\Lambda^1{}_1 > 1$) Lorentz group). To obtain all the Lorentz transformations, one needs to multiply by discrete transformations such as time-reversal and parity (depending on the dimension d). More precisely, the exponentiation of the generators of the Lorentz algebra so$(1, d-1)$ in this representation Γ_s gives elements of the simply connected group which is locally isomorphic to SO$(1, d-1)$ (and, therefore, has the same Lie algebra) and which is called, by definition, Spin$(1, d-1)$. The Lorentz group is doubly connected and, thus, some representations of Spin$(1, d-1)$ (in particular the spinorial Γ_s) are double-valued as representations of the Lorentz group.

[3] By physically equivalent we mean representations of the Clifford algebra that give, with the above construction, equivalent representations (i.e. that are representations related by similarity transformations) of the Lorentz group. Two physically equivalent representations of the Clifford algebra may but need not be strictly equivalent.

where $\sigma^{1,2,3}$ are the (Hermitian, unitary, 2×2) Pauli matrices

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{B.9}$$

that satisfy

$$\sigma^i \sigma^j = \delta^{ij} + i\epsilon^{ijk}\sigma^k, \tag{B.10}$$

is a $2^{[d/2]}$-dimensional representation of the d-dimensional Clifford algebra. A $d = 2$ representation of the two-dimensional Clifford algebra is provided by (for instance) $\{\mathbb{I}_{2\times 2}, i\sigma^2\}$ and this completes the proof for d even.

Now, if d is even and $\{\Gamma^0, \ldots, \Gamma^{d-1}\}$ are $2^{d/2} \times 2^{d/2}$ gamma matrices satisfying the d-dimensional Clifford algebra, then the gamma matrices $\{\Gamma^0, \ldots, \Gamma^{d-1}, \Gamma^d\}$ with

$$\Gamma^d \equiv -i\varphi(d)\Gamma^0 \cdots \Gamma^{d-1}, \qquad \varphi(d) = (-1)^{\frac{1}{4}(d-2)+1}, \tag{B.11}$$

satisfy the $(d+1)$-dimensional Clifford algebra.[4] Thus, the even d irreducible representations determine the $d+1$ irreducible representations and this completes the proof. Observe that this matrix is different from the chirality matrix $\mathcal{Q} = \Gamma_{d+1}$ (γ_5 in $d = 4$):

$$\begin{aligned} &\Gamma_{d+1} = i\Gamma^d = \varphi(d)i\Gamma^0 \cdots \Gamma^{d-1}, \\ &\Gamma_{d+1}^2 = +1, \qquad \Gamma_{d+1}^\dagger = +\Gamma_{d+1}. \qquad \Gamma^0\Gamma_{d+1}\Gamma^0 = -\Gamma_{d+1}^\dagger. \end{aligned} \tag{B.13}$$

Observe also that, in odd dimensions, by construction, the product of all gamma matrices is proportional to a constant whose sign can be chosen at will (by changing the sign of Γ^d). The two possible signs give inequivalent representations of the Clifford algebra (which are, nevertheless, physically equivalent).

Let us now consider equivalent representations of the Clifford algebra, related by a similarity transformation

$$\Gamma^{a\,\prime} = S\Gamma^a S^{-1}. \tag{B.14}$$

If d is even, then, if we change the sign of all the gamma matrices, we obtain an equivalent representation with $S = \mathcal{Q}$, the chirality matrix. If d is odd, changing the signs of all the gamma matrices does not provide an equivalent representation because it changes the sign of the product of all the gamma matrices.

For both even and odd d the Hermitian conjugates of the gamma matrices constitute another representation related to the original one by $S = \mathcal{D}$, where $\mathcal{D}$ is the *Dirac conjugation matrix* and can be taken to be $\mathcal{D} = i\Gamma^0$.

In even d the transposed gamma matrices also provide another equivalent representation of the Clifford algebra. In that case, by definition, $S = \mathcal{C}$, the *charge-conjugation matrix*, which we will use later. For d odd, sometimes it is the transposed gamma matrices that

[4] By putting together

$$\begin{aligned} \Gamma^0 \cdot \Gamma^{d-1} &= (-1)^{[d/2]}\Gamma^{d-1} \cdot \Gamma^0, \\ \left(\Gamma^0 \cdot \Gamma^{d-1}\right)\left(\Gamma^{d-1} \cdot \Gamma^0\right) &= (-1)^{d-1}, \end{aligned} \tag{B.12}$$

it is easy to check that Γ^d anticommutes with all the other gammas, squares to -1, and is anti-Hermitian.

provide a representation and sometimes it is the transposed gamma matrices with the sign reversed that provide an equivalent representation.

Clearly, for even d (and for odd d, up to a sign) the complex conjugates of gamma matrices give also an equivalent representation and, by definition, $\mathcal{S} = \mathcal{B}$. The matrix $\mathcal{B}$ is related to $\mathcal{D}$ and $\mathcal{C}$ and we will not use it now.

Gamma matrices carry a vector Lorentz index that is raised and lowered with η. They are invariant under Lorentz transformations that act on their three (two spinorial, one vector) indices:

$$\left(\Gamma'^{a}\right)^{\alpha}{}_{\beta} = \Lambda^{a}{}_{b}\Lambda^{\alpha}{}_{\gamma}\left(\Gamma^{b}\right)^{\gamma}{}_{\delta}\left(\Lambda^{-1}\right)^{\delta}{}_{\beta} = \left(\Gamma^{a}\right)^{\alpha}{}_{\beta}, \qquad \Lambda^{\alpha}{}_{\gamma} = \Gamma_{\mathrm{s}}(\Lambda)^{\alpha}{}_{\gamma}. \tag{B.15}$$

If we consider infinitesimal transformations, we find that gamma matrices and generators of the Lorentz group in the spinorial representation must obey the commutation relations[5]

$$\left[\Gamma^{a}, \Gamma_{\mathrm{s}}(M_{bc})\right] = \Gamma_{\mathrm{v}}(M_{bc})^{a}{}_{d}\Gamma^{d}. \tag{B.16}$$

These commutation relations are identical to those of the momentum P^a and the Lorentz generators in the Poincaré algebra. Thus, they indicate that the gamma matrices transform as vectors in the spinorial representation.[6]

Let us now study spinors. Dirac spinors transform under the Lorentz group as expected (contravariantly),

$$\begin{aligned} \psi'^{\alpha} &= \Lambda^{\alpha}{}_{\beta}\psi^{\beta} = \Gamma_{\mathrm{s}}(\Lambda)^{\alpha}{}_{\beta}\psi^{\beta}, \\ \Gamma_{\mathrm{s}}(\Lambda) &= \exp\left\{\tfrac{1}{2}\sigma^{ab}\Gamma_{\mathrm{s}}(M_{ab})\right\} = \exp\left\{\tfrac{1}{4}\sigma^{ab}\Gamma_{ab}\right\}, \end{aligned} \tag{B.17}$$

or, infinitesimally,

$$\delta_{\sigma}\psi^{\alpha} = \tfrac{1}{2}\sigma^{ab}\Gamma_{\mathrm{s}}(M_{ab})^{\alpha}{}_{\beta}\psi^{\beta}. \tag{B.18}$$

In flat Minkowski spacetime, Lorentz transformations are global. In curved spacetime Lorentz transformations make sense only at one point in tangent space and therefore Lorentz transformations are naturally local ($\sigma^{ab} = \sigma^{ab}(x)$). Hence theories containing spinors are required to be invariant under local Lorentz transformations, and are naturally gauge theories of the Lorentz group. The Lorentz covariant derivative acting on a (contravariant) spinor is, according to the results of Appendix A.2.3 and Eq. (B.3),

$$\nabla_{\mu}\psi = \left(\partial_{\mu} - \tfrac{1}{4}\omega_{\mu}{}^{ab}\Gamma_{ab}\right)\psi. \tag{B.19}$$

In field theory, a Dirac spinor ψ^{α} is a field $\psi^{\alpha}(x)$ that satisfies the massive or massless ($m = 0$), charged or uncharged ($e = 0$) d-dimensional *Dirac equation*

$$(i\,\not{\nabla} - m + e\,\not{A})\psi = 0, \qquad \not{\nabla} \equiv \Gamma^{a}e_{a}{}^{\mu}\nabla_{\mu}, \qquad \not{A} \equiv \Gamma^{a}e_{a}{}^{\mu}A_{\mu}. \tag{B.20}$$

We can also define spinors ξ_{α} transforming covariantly,

$$\begin{aligned} \xi'_{\alpha} &= \xi_{\beta}\left(\Lambda^{-1}\right)^{\beta}{}_{\alpha} = \xi_{\beta}\Gamma_{\mathrm{s}}\left(\Lambda^{-1}\right)^{\beta}{}_{\alpha}, \\ \Gamma_{\mathrm{s}}\left(\Lambda^{-1}\right) &= \exp\left\{-\tfrac{1}{2}\sigma^{ab}\Gamma_{\mathrm{s}}(M_{ab})\right\} = \exp\left\{-\tfrac{1}{4}\sigma^{ab}\Gamma_{ab}\right\}, \end{aligned} \tag{B.21}$$

[5] This consistency check is satisfied in our conventions.

[6] Observe, though, that gamma matrices do not commute and therefore they do not provide a spinorial representation of the translation generators of the Poincaré algebra.

or, infinitesimally,

$$\delta_\sigma \xi_\alpha = \xi_\beta \left[-\tfrac{1}{2}\sigma^{ab}\Gamma_{\mathrm{s}}(M_{ab})^\beta{}_\alpha \right]. \tag{B.22}$$

The Lorentz covariant derivative acts on covariant spinors according to

$$\nabla_\mu \xi = \partial_\mu \xi + \xi \left(-\tfrac{1}{4}\omega_\mu{}^{ab}\Gamma_{ab} \right) \equiv \xi \overleftarrow{\nabla}_\mu. \tag{B.23}$$

Just as we can transform contravariant vectors into covariant vectors by "lowering the index" with the metric, we can transform contravariant spinors into covariant spinors by *conjugation*. Given a Dirac spinor transforming contravariantly, there are two kinds of conjugate spinors that transform covariantly.

The Dirac conjugate $\bar\psi$ of a spinor ψ is a new spinor that transforms covariantly and whose components $\bar\psi_\alpha$ are linear combinations of those of $(\psi^\alpha)^*$:

$$\bar\psi_\alpha = \left(\psi^\dagger \mathcal{D} \right)_\alpha = (\psi^\beta)^* \mathcal{D}_{\beta\alpha}, \tag{B.24}$$

where $\mathcal{D}$ is the *Dirac conjugation matrix*. According to the above definition

$$\mathcal{D}\,\Gamma_{ab}\,\mathcal{D}^{-1} = -\Gamma_{ab}^\dagger. \tag{B.25}$$

Taking the Hermitian conjugate of Eq. (B.17) and using $\Gamma^{ab\,\dagger} = -\Gamma^0\Gamma^{ab}\Gamma^0$, we find that, up to a phase, the Dirac conjugation matrix is given by Γ^0. Our convention is $\mathcal{D} = i\Gamma^0$, so

$$\bar\psi \equiv i\psi^\dagger\Gamma^0. \tag{B.26}$$

Obviously, by definition, the product $\bar\psi_\alpha\psi^\alpha$ is a Lorentz invariant. Also $\bar{\bar\psi} = \psi$.

The Dirac conjugate satisfies the conjugate Dirac equation

$$i\nabla_a\bar\psi\Gamma^a + m\bar\psi = \bar\psi(i\,\overleftarrow{\not\nabla} + m - e\not{A}) = 0. \tag{B.27}$$

The Majorana conjugate ψ^c of a spinor ψ is a new spinor that transforms covariantly and whose components ψ^c_α are linear combinations of ψ^α,

$$\psi^c_\alpha = \left(\psi^{\mathrm{T}}\mathcal{C} \right)_\alpha = \psi^\beta \mathcal{C}_{\beta\alpha} \tag{B.28}$$

(not $\psi^\star$ as in the Dirac conjugate), and that transform as the Dirac conjugate under Lorentz transformations. Here $\mathcal{C}$ is the *charge-conjugation matrix*. By transposing Eq. (B.17) and using the definition of ψ^c, we find that $\mathcal{C}$ must satisfy

$$\mathcal{C}\,\Gamma_{ab}\,\mathcal{C}^{-1} = -\Gamma_{ab}^{\mathrm{T}}. \tag{B.29}$$

The matrices $-\frac{1}{2}\Gamma_{ab}^{\mathrm{T}}$ also satisfy the Lorentz algebra. The charge-conjugation matrix $\mathcal{C}$ relates this representation and the standard one. It is natural to look for a charge-conjugation matrix that also relates a representation of the Clifford algebra and the representation obtained by transposing all gamma matrices. There are two possibilities: $\mathcal{C}_+$ and $\mathcal{C}_-$ defined by

$$\mathcal{C}_\pm\Gamma^a\mathcal{C}_\pm^{-1} = \pm\Gamma^{a\,\mathrm{T}}. \tag{B.30}$$

When d is even, both matrices exist because $+\Gamma^{a\,\mathrm{T}}$ and $-\Gamma^{a\,\mathrm{T}}$ generate the same finite group as Γ^a. When d is odd, only one of them exists because of the above definition of Γ^{d-1} in terms of $\Gamma^0, \ldots, \Gamma^{d-2}$. Furthermore, both matrices, when they exist, are either symmetric or antisymmetric.

When d is even, the charge-conjugation matrices act on Γ_{d+1} as follows:

$$\mathcal{C}_\pm \Gamma_{d+1} \mathcal{C}_\pm^{-1} = \varphi^2 \Gamma_{d+1}^{\mathrm{T}}. \tag{B.31}$$

Using the $\mathcal{C}_\pm$ charge-conjugation matrices and taking the Majorana conjugate of the Dirac equation, we find that the Majorana conjugate satisfies the following equation:

$$\psi^c (i \overleftarrow{\not{\nabla}} \mp m + e \not{A}) = 0, \tag{B.32}$$

which implies that ψ^c has charge opposite to that of $\bar{\psi}$ (hence the name "charge-conjugation matrix"). It is obviously desirable that both ψ^c and $\bar{\psi}$ have the same mass and, thus, in the massive case the only acceptable charge-conjugation matrix is $\mathcal{C}_-$.

By construction $\psi^c \psi$ is Lorentz-invariant and $(\psi^c)^c = \psi$.

We can now study various types of spinors that are in general associated with special representations of gamma matrices.

Weyl spinors (Also called *chiral* spinors.) For even d it is possible to define as before the *chirality matrix* Γ_{d+1} which anticommutes with all the gamma matrices and therefore commutes with the generators of the Lorentz group $\Gamma_s(M_{ab})$ and with their exponentials, which span the Spin$(1, d-1)$ group. Thus (Schur's lemma) this representation of the Spin$(1, d-1)$ group, and Dirac spinors, are reducible even if the gamma matrices provide an irreducible representation of the d-dimensional Clifford algebra of η_{ab}. The chirality matrix is traceless and squares to unity and therefore half of its eigenvalues are $+1$s and the other half are -1s. It is natural to split the space of Dirac spinors into the direct sum of the subspaces of spinors with eigenvalues $+1$ and -1. The elements of each of these subspaces are called *Weyl spinors* and, by definition, satisfy the Weyl or chirality condition

$$\tfrac{1}{2}(1 \pm \Gamma_{d+1})\psi = \psi. \tag{B.33}$$

For the positive sign, the spinors are called left-handed (negative chirality); and for the negative sign they are called right-handed (positive chirality). A Weyl spinor describes half the degrees of freedom of a Dirac spinor.

Observe that, while Weyl spinors are irreducible representations of the Spin$(1, d-1)$ group, they are not irreducible representations of the Lorentz group SO$(1, d-1)$ because this group contains discrete transformations that interchange the two subspaces of opposite chiralities. In particular, the parity transformation is implemented by $P = i\Gamma^0$, which does not commute but anticommutes with the chirality matrix, switching the chirality of the spinors.

Using the Dirac equation, it can be seen that the Weyl condition is preserved in time for $m = 0$.

Associated with Weyl spinors there are *Weyl (or chiral) representations* of gamma matrices. In a Weyl representation the generators of the Lorentz group are diagonal and the chirality matrix is $\Gamma_{(d+1)} = \mathbb{I} \otimes \sigma^3$. In a chiral basis of gamma matrices, half of the components of a Weyl spinor are zero and it is sometimes advantageous to use half-size spinors.

Majorana spinors are spinors whose Majorana conjugate is proportional to their Dirac conjugate (the Majorana condition):

$$\hat{\lambda} = \alpha\,\bar{\lambda}. \tag{B.34}$$

This is a reality condition because it relates the components of λ to those of its complex conjugate. Thus, it describes half the degrees of freedom of a Dirac spinor. Using our definitions of Majorana and Dirac conjugates in the definition of a Majorana spinor, we find that it implies

$$|\alpha|^2 \Gamma^{0\star} \big(\mathcal{C}^{-1}\big)^{\star}\, \Gamma^0 \mathcal{C}^{-1} = +1. \tag{B.35}$$

This condition cannot be fulfilled in all dimensions and this is the reason why Majorana spinors exist only in certain dimensions. This condition and the (anti)symmetry of $\mathcal{C}$ do not depend on the representation and the results found in any representation are valid in general.[7]

Associated with Majorana spinors there are *Majorana representations* in which all Γs are purely imaginary.[8] If a Majorana representation exists, then the condition for the existence of Majorana spinors is automatically satisfied by the choice

$$\mathcal{C} = i\alpha\Gamma^0. \tag{B.36}$$

With any other representation we have to check explicitly whether the above equation holds.

Below $d = 11$, there are Majorana spinors in all but $d = 5$, 6, and 7.

Majorana–Weyl spinors satisfy both Majorana and Weyl conditions. They exist in even dimensions if, in addition to the Majorana condition, one can satisfy a compatibility condition between Majorana and Weyl conditions:

$$\mathcal{C}^{-1}\, \Gamma^{\mathrm{T}}_{d+1}\, \mathcal{C} = \Gamma^0\, \Gamma^{\dagger}_{d+1}\, \Gamma^0. \tag{B.37}$$

This condition is representation-dependent and, again, can be satisfied only in certain dimensions. Using the definition of Γ_{d+1} we have chosen and its properties, we see that the above condition is satisfied whenever $\eta^2(d) = -1$ (d even); that is, when

$$d = 2(\mathrm{mod}\ 4) = 2, 6, 10, \ldots. \tag{B.38}$$

---

[7] This condition and similar conditions will be studied in detail, for arbitrary signature, in Section B.2. For the moment we will simply quote the results for Lorentzian signature and $d \leq 11$.

[8] With mostly plus signature all Γs are purely real (essentially the same matrices multiplied by i). If we used the Pauli metric, so that $\{\Gamma^a, \Gamma^b\} = +2\delta^{ab}$, all Γs would be real except for Γ^4, which would be imaginary.

Table B.1. Spinors that exist in various dimensions with signature $(1, d-1)$. M is the number of real independent components of the smaller spinorial representation in the given dimension. (W, Weyl; M, Majorana; M–W, Majorana–Weyl; and S–M, Symplectic–Majorana spinors.)

d	W	M	M–W	S-M	S-M and W	M
2	x	x	x			1
3		x				2
4	x	x		x		4
5				x		8
6	x			x	x	8
7				x		16
8	x	x		x		16
9		x				16
10	x	x	x			16
11		x				32

Majorana spinors do not exist in $d = 6$ and so Majorana–Weyl spinors exist only in $d = 2$ and 10 (at least in our representation and in fewer than 11 dimensions). It can be shown that these are also the only dimensions in which they exist.

Symplectic-Majorana spinors When defining Majorana spinors (i.e. spinors satisfying a reality condition) is not possible, one can take an even number of Dirac spinors labeled by $i = 1, \ldots, 2n$ and impose a reality condition on the whole set:

$$\bar{\psi}^i = \psi_i{}^c \equiv \Omega_{ij} \psi^{i\,c}, \tag{B.39}$$

where Ω is real and satisfies

$$\Omega_{ij} \Omega_{jk} = -\delta_{ik}. \tag{B.40}$$

Below $d = 11$ this can be done consistently in $d = 4$, 5, 6, 7, and 8. In $d = 6$ we can impose simultaneously the symplectic-Majorana and Weyl conditions.

In many cases, $n = 2$ symplectic-Majorana spinors appear combined in a single, unconstrained, Dirac spinor that contains the same number of degrees of freedom.

See Table B.1 for a summary.

B.1.1 Useful identities

Most of the gamma identities (for up to four gammas) that we need can be obtained from the following products by symmetrization, antisymmetrization, etc.:

$$\Gamma^a \Gamma^b = \Gamma^{ab} + \eta^{ab}, \tag{B.41}$$

$$\Gamma^a \Gamma^b \Gamma^c = \Gamma^{abc} + \eta^{ab} \Gamma^c - \eta^{ca} \Gamma^b + \eta^{bc} \Gamma^a, \tag{B.42}$$

$$\begin{aligned} \Gamma^a \Gamma^b \Gamma^c \Gamma^d = {} & \Gamma^{abcd} + \eta^{ab} \Gamma^{cd} - \eta^{cb} \Gamma^{da} + \eta^{cd} \Gamma^{ab} + \eta^{da} \Gamma^{bc} \\ & - \eta^{ac} \Gamma^{bd} - \eta^{bd} \Gamma^{ac} + \eta^{ab} \eta^{cd} - \eta^{ac} \eta^{bd} + \eta^{ad} \eta^{bc}. \end{aligned} \tag{B.43}$$

For instance, we can obtain

$$\Gamma^{ab}\Gamma_{cd} = \Gamma^{ab}{}_{cd} + 4\Gamma_{[d}{}^{[a}\eta_{c]}{}^{b]} + 2\eta_{[d}{}^{a}\eta_{c]}{}^{b}, \tag{B.44}$$

from which we can derive the Lorentz algebra. With more than four, we use repeatedly

$$\Gamma^{a}\Gamma^{b_1\cdots b_n} = \Gamma^{ab_1\cdots b_n} + n\eta^{a[b_1}\Gamma^{b_2\cdots b_n]}, \tag{B.45}$$

$$\Gamma^{b_1\cdots b_n}\Gamma^{a} = \Gamma^{b_1\cdots b_n a} + n\Gamma^{[b_1\cdots b_{n-1}}\eta^{b_n]a}. \tag{B.46}$$

Using them, we find, for instance, the (anti)commutator

$$\left[\Gamma^{a}, \Gamma^{b_1\cdots b_n}\right]_{\pm} = [1 \mp (-1)^n]\Gamma^{ab_1\cdots b_n} + n[1 \pm (-1)^n]\eta^{a[b_1}\Gamma^{\cdots b_n]} \tag{B.47}$$

and the general formula

$$\begin{aligned}\Gamma^{b_1\cdots b_n}\Gamma_{a_1\cdots a_m} &= \sum_{p=0}^{\min(n,m)} \frac{n!m!}{(n-p)!(m-p)!p!}\Gamma^{[b_1\cdots b_{n-p}}{}_{[a_{p+1}\cdots a_m}\eta^{b_{n-p+1}\cdots b_n]}{}_{a_1\cdots a_{m-p}]}\\ &= \Gamma^{b_1\cdots b_n}{}_{a_1\cdots a_m} + nm\Gamma^{[b_1\cdots b_{n-1}}{}_{[a_2\cdots a_m}\eta^{b_n]}{}_{a_1]} + \cdots .\end{aligned} \tag{B.48}$$

B.1.2 Fierz identities

These identities are used very often in supergravity theories. To derive these identities, we first need a basis $\{\mathcal{O}_I\}$ of the vector space of $2^{[d/2]} \times 2^{[d/2]}$ matrices. This basis can be built out of the gamma matrices

$$\left\{\mathcal{O}^I\right\} = \left\{\mathbb{I}, \Gamma^a, i\Gamma^{ab}, i\Gamma^{abc}, \Gamma^{abcd}, \ldots\right\}. \tag{B.49}$$

(Observe that there are $2^{2[d/2]}$ matrices in this basis. Furthermore, this is why a decomposition like Eq. (B.48) is always possible.) All these matrices are linearly independent except for the last one (the product of all gamma matrices) in odd dimensions. Now we construct a dual basis orthogonal to the one above:

$$\{\mathcal{O}_I\} = \{\mathbb{I}, \Gamma_a, i\Gamma_{ab}, i\Gamma_{abc}, \Gamma_{abcd}, \ldots\}, \qquad \mathcal{O}_I(\mathcal{O}^J) \equiv \mathrm{tr}(\mathcal{O}_I\mathcal{O}_J) = 2^{[d/2]}\delta_{IJ}, \tag{B.50}$$

as we can easily check:

$$\begin{aligned}\mathrm{tr}\,\mathbb{I}^2 &= 2^{[d/2]},\\ \mathrm{tr}\Big(\tilde{\Gamma}^a\tilde{\Gamma}^b\Big) &= \mathrm{tr}\Big(\tilde{\Gamma}^{(a}\tilde{\Gamma}^{b)}\Big) = \mathrm{tr}\big(\mathbb{I}\delta^{ab}\big) = 2^{[d/2]}\delta^{ab},\\ \mathrm{tr}\Big(\mathbb{I}\tilde{\Gamma}^a\Big) &= 0,\end{aligned} \tag{B.51}$$

etc. Any $2^{[d/2]} \times 2^{[d/2]}$ matrix P is a linear combination of the $\{\mathcal{O}^I\}$:

$$\begin{aligned}P &= p_I\mathcal{O}^I,\\ \mathrm{tr}(\mathcal{O}_I P) &= 2^{[d/2]}p_I,\\ P^{\alpha}{}_{\beta} &= p_I\mathcal{O}^{I\,\alpha}{}_{\beta} = 2^{-[d/2]}\sum_I P^{\gamma}{}_{\delta}\mathcal{O}^{\delta}_{I\,\gamma}\mathcal{O}^{I\,\alpha}_{\beta}.\end{aligned} \tag{B.52}$$

Let us now consider the product of bilinears of spinors,

$$\mathcal{Q} = \left(\bar{\lambda} M \chi\right)\left(\bar{\psi} N \varphi\right) = \bar{\lambda}_\alpha \chi^\beta \bar{\psi}_\gamma \varphi^\delta M^\alpha{}_\beta N^\gamma{}_\delta. \tag{B.53}$$

For fixed indices α and δ, $M^\alpha{}_\beta N^\gamma{}_\delta$ is just a matrix $P^\gamma{}_\beta(\alpha, \delta)$ to which we can apply the above formula, obtaining

$$M^\alpha{}_\beta N^\gamma{}_\delta = 2^{-[d/2]} \sum_I (M \mathcal{O}_I N)^\alpha{}_\delta \mathcal{O}^{I\,\gamma}_\beta. \tag{B.54}$$

On substituting this identity into $\mathcal{Q}$ and taking into account the necessary permutation of the spinors, we obtain the general *Fierz identities*

$$\boxed{\begin{aligned} &\mathcal{Q} = p\, 2^{-[d/2]} \sum_I \left(\bar{\lambda} M \mathcal{O}^I N \varphi\right) \left(\bar{\psi} \mathcal{O}_I \chi\right), \\ &p = \begin{cases} -1 & \text{anticommuting spinors,} \\ +1 & \text{commuting spinors.} \end{cases} \end{aligned}} \tag{B.55}$$

Now, depending on the dimensions and the particular properties of the spinors, this expression can be further simplified. For instance, if we are dealing with Majorana–Weyl spinors, then terms with n and $d - n$ gammas will be related.

In some cases ($N = 2$ SUGRA theories) spinors appear in SO(2) doublets ψ^i, $i = 1, 2$. It is then convenient to arrange them in a vector,

$$\psi \equiv \begin{pmatrix} \psi^1 \\ \psi^2 \end{pmatrix}, \tag{B.56}$$

and it is useful to have Fierz identities for these vectors.[9] In the space of 2×2 matrices a convenient basis is provided by the Pauli matrices and the identity. Is is easy to arrive at the following $N = 2$ Fierz identities:

$$\boxed{\begin{aligned} &\mathcal{Q} = p\, 2^{-[d/2]-1} \sum_{I,A} \left(\bar{\lambda} M \mathcal{O}^I N \sigma^A \varphi\right)\left(\bar{\psi} \mathcal{O}_I \sigma^A \chi\right), \\ &p = \begin{cases} -1 & \text{anticommuting spinors,} \\ +1 & \text{commuting spinors.} \end{cases} \end{aligned}} \tag{B.57}$$

with $A = 0, 1, 2, 3$ and $\sigma^0 = 1$.

B.1.3 Eleven dimensions

Our 11-dimensional gamma matrices satisfy the anticommutation relations

$$\{\hat{\hat{\Gamma}}^{\hat{\hat{a}}}, \hat{\hat{\Gamma}}^{\hat{\hat{b}}}\} = +2\hat{\hat{\eta}}^{\hat{\hat{a}}\hat{\hat{b}}}. \tag{B.58}$$

[9] In higher-N supergravity theories, spinors come in higher-dimensional multiplets and further generalizations exist.

It is possible to choose (in a way consistent with all the properties that we are going to enumerate) the 11th gamma matrix $\hat{\hat{\Gamma}}^{10}$ to be

$$\hat{\hat{\Gamma}}^{\hat{10}} = i\hat{\hat{\Gamma}}^{\hat{\hat{0}}} \cdots \hat{\hat{\Gamma}}^{\hat{\hat{9}}} \equiv -i\hat{\Gamma}_{11}, \tag{B.59}$$

where $\hat{\Gamma}_{11}$ will be the ten-dimensional chirality matrix $\{\hat{\Gamma}_{11}, \hat{\Gamma}^{\hat{a}}\} = 0$.
They are in a purely imaginary (i.e. Majorana) representation, i.e. $\hat{\hat{\Gamma}}^{\hat{\hat{a}}\,\star} = -\hat{\hat{\Gamma}}^{\hat{\hat{a}}}$. They are all anti-Hermitian, except for $\hat{\hat{\Gamma}}^{\hat{\hat{0}}}$, which is Hermitian:

$$\begin{aligned} \hat{\hat{\Gamma}}^{\hat{\hat{0}}\,\dagger} &= +\hat{\hat{\Gamma}}^{\hat{\hat{0}}}. \\ \hat{\hat{\Gamma}}^{\hat{\hat{i}}\,\dagger} &= -\hat{\hat{\Gamma}}^{\hat{\hat{i}}}, \qquad \hat{\hat{i}} = 1, \ldots, 10. \end{aligned} \tag{B.60}$$

We have the property

$$\hat{\hat{\Gamma}}^{\hat{\hat{0}}}\, \hat{\hat{\Gamma}}^{\hat{\hat{a}}}\, \hat{\hat{\Gamma}}^{\hat{\hat{0}}} = \hat{\hat{\Gamma}}^{\hat{\hat{a}}\,\dagger}. \tag{B.61}$$

The Dirac conjugation matrix $\hat{\hat{\mathcal{D}}}$ is the real antisymmetric matrix

$$\hat{\hat{\mathcal{D}}} = i\hat{\hat{\Gamma}}^{0}, \tag{B.62}$$

and thus we have

$$\hat{\hat{\mathcal{D}}}\, \hat{\hat{\Gamma}}^{\hat{\hat{a}}_1 \cdots \hat{\hat{a}}_n}\, \hat{\hat{\mathcal{D}}}^{-1} = (-1)^{[n/2]} \left(\hat{\hat{\Gamma}}^{\hat{\hat{a}}_1 \cdots \hat{\hat{a}}_n}\right)^{\dagger}. \tag{B.63}$$

Their Hermiticity properties combined with their imaginary nature mean that all are symmetric except for $\hat{\hat{\Gamma}}^{\hat{\hat{0}}}$, which is antisymmetric:

$$\hat{\hat{\Gamma}}^{\hat{\hat{0}}\,\mathrm{T}} = -\hat{\hat{\Gamma}}^{\hat{\hat{0}}}, \qquad \hat{\hat{\Gamma}}^{\hat{\hat{i}}\,\mathrm{T}} = +\hat{\hat{\Gamma}}^{\hat{\hat{i}}}, \qquad \hat{\hat{i}} = 1, \ldots, 10. \tag{B.64}$$

We choose a charge-conjugation matrix equal to the Dirac conjugation matrix,

$$\hat{\hat{\mathcal{C}}} = \hat{\hat{\mathcal{D}}} = i\hat{\hat{\Gamma}}^{0}, \tag{B.65}$$

which satisfies

$$\hat{\hat{\mathcal{C}}}^{\mathrm{T}} = \hat{\hat{\mathcal{C}}}^{\dagger} = \hat{\hat{\mathcal{C}}}^{-1} = -\hat{\hat{\mathcal{C}}}, \qquad \hat{\hat{\mathcal{C}}}\, \hat{\hat{\Gamma}}^{\hat{\hat{a}}}\, \hat{\hat{\mathcal{C}}}^{-1} = -\hat{\hat{\Gamma}}^{\hat{\hat{a}}\,\mathrm{T}}. \tag{B.66}$$

The last property implies

$$\hat{\hat{\mathcal{C}}}\, \hat{\hat{\Gamma}}^{\hat{\hat{a}}_1 \cdots \hat{\hat{a}}_n}\, \hat{\hat{\mathcal{C}}}^{-1} = (-1)^{n+[n/2]} \left(\hat{\hat{\Gamma}}^{\hat{\hat{a}}_1 \cdots \hat{\hat{a}}_n}\right)^{\mathrm{T}}. \tag{B.67}$$

The standard definitions of the Dirac conjugates and Majorana conjugates and our specific choice of Dirac and charge-conjugation matrices $\hat{\hat{\mathcal{C}}} = \hat{\hat{\mathcal{D}}}$ imply that the Majorana condition

$$\bar{\hat{\hat{\lambda}}} = \hat{\hat{\lambda}}^{c} \tag{B.68}$$

is equivalent to requiring that all components of a Majorana spinor are real. Using the property (B.67) and the definition of (*anticommuting*) Majorana spinors, one finds

$$\bar{\hat{\hat{\epsilon}}}\,\hat{\hat{\Gamma}}^{\hat{\hat{a}}_1\cdots\hat{\hat{a}}_n}\,\hat{\hat{\psi}} = (-1)^{n+[n/2]}\,\bar{\hat{\hat{\psi}}}\,\hat{\hat{\Gamma}}^{\hat{\hat{a}}_1\cdots\hat{\hat{a}}_n}\,\hat{\hat{\epsilon}}, \tag{B.69}$$

so the above bilinear is symmetric for $n = 0, 3, 4, 7,$ and 8 and antisymmetric for $n = 1, 2, 5, 6, 9,$ and 10.

On the other hand, taking the Hermitian conjugate[10] and using Eq. (B.63), we find

$$\left(\bar{\hat{\hat{\epsilon}}}\,\hat{\hat{\Gamma}}^{\hat{\hat{a}}_1\cdots\hat{\hat{a}}_n}\,\hat{\hat{\psi}}\right)^\dagger = (-1)^{[n/2]}\bar{\hat{\hat{\psi}}}\,\hat{\hat{\Gamma}}^{\hat{\hat{a}}_1\cdots\hat{\hat{a}}_n}\,\hat{\hat{\epsilon}}, \tag{B.70}$$

which implies, on comparison with Eq. (B.69), that the above bilinear is real for even n and imaginary for odd n.

Finally, we have the useful identity

$$\hat{\hat{\Gamma}}^{\hat{\hat{a}}_1\cdots\hat{\hat{a}}_n} = i\,\frac{(-1)^{[n/2]+1}}{(11-n)!}\hat{\hat{\epsilon}}^{\hat{\hat{a}}_1\cdots\hat{\hat{a}}_n\hat{\hat{b}}_1\cdots\hat{\hat{b}}_{11-n}}\,\hat{\hat{\Gamma}}_{\hat{\hat{b}}_1\cdots\hat{\hat{b}}_{11-n}}. \tag{B.71}$$

B.1.4 Ten dimensions

The 11-dimensional Majorana representation of gamma matrices can be constructed from the ten-dimensional Majorana (purely imaginary) representation, according to

$$\begin{aligned} \hat{\hat{\Gamma}}^{\hat{a}} &= \hat{\Gamma}^{\hat{a}}, \quad \hat{a} = 0, \ldots, 9, \\ \hat{\hat{\Gamma}}^{10} &= +i\,\hat{\Gamma}^0 \cdots \hat{\Gamma}^9. \end{aligned} \tag{B.72}$$

Ten-dimensional Majorana spinors are identical to 11-dimensional spinors and the same definitions and identities apply to them.

However, in ten dimensions we can also have Weyl spinors that satisfy many additional identities. They are defined in terms of the chirality matrix $\hat{\Gamma}_{11}$,

$$\hat{\Gamma}_{11} = -\hat{\Gamma}^0 \cdots \hat{\Gamma}^9 = i\,\hat{\hat{\Gamma}}^{10}, \tag{B.73}$$

so $\hat{\Gamma}_{11}$ is Hermitian and satisfies $(\hat{\Gamma}_{11})^2 = +1$. Spinors of positive, $\hat{\psi}^{(+)}$, and negative, $\hat{\psi}^{(-)}$, chiralities are defined as usual:

$$\hat{\Gamma}_{11}\hat{\psi}^{(\pm)} = \pm\hat{\psi}^{(\pm)}. \tag{B.74}$$

Furthermore, in $d = 10$ we can define Majorana–Weyl fermions. It is useful to work in a Majorana–Weyl representation of the gamma matrices in which, in addition to having imaginary gamma matrices, the chirality matrix $\hat{\Gamma}_{11}$ has the form

$$\hat{\Gamma}_{11} = \mathbb{I}_{16\times16} \otimes \sigma^3 = \begin{pmatrix} \mathbb{I}_{16\times16} & 0 \\ 0 & -\mathbb{I}_{16\times16} \end{pmatrix}. \tag{B.75}$$

[10] We use the convention $(ab)^\star = +a^\star b^\star$ for anticommuting numbers. This is the convention used in [264, 599, 795] etc. The opposite convention is used in [946, 948] etc.

We will see explicitly that it is possible to have $\hat{\Gamma}_{11}$ defined as above in terms of the ten gamma matrices and at the same time having precisely that form. The sign of $\hat{\Gamma}_{11}$ is chosen in order to have that relation with positive sign and to have as well

$$\hat{\Gamma}_{11} = \frac{1}{10!}\hat{\epsilon}_{\hat{a}_1\cdots\hat{a}_{10}}\hat{\Gamma}^{\hat{a}_1\cdots\hat{a}_{10}} = \frac{1}{10!\sqrt{|\hat{g}|}}\hat{\epsilon}_{\hat{\mu}_1\cdots\hat{\mu}_{10}}\hat{\Gamma}^{\hat{\mu}_1\cdots\hat{\mu}_{10}}, \tag{B.76}$$

which leads to

$$\Gamma_{11}\hat{\Gamma}^{\hat{a}_1\cdots\hat{a}_n} = \frac{(-1)^{[(10-n)/2]+1}}{(10-n)!}\hat{\epsilon}^{\hat{a}_1\cdots\hat{a}_n\hat{b}_1\cdots\hat{b}_{10-n}}\hat{\Gamma}_{\hat{b}_1\cdots\hat{b}_{10-n}}. \tag{B.77}$$

In the Majorana–Weyl representation each 32-component real Majorana spinor $\hat{\psi}$ can be constructed from one positive-chirality and one negative-chirality 16-component spinor:

$$\hat{\psi} = \begin{pmatrix} \hat{\psi}^{(+)} \\ \hat{\psi}^{(-)} \end{pmatrix}. \tag{B.78}$$

B.1.5 Nine dimensions

We have chosen a Majorana–Weyl representation for the ten-dimensional gamma matrices. They can be constructed from a purely real representation of the nine-dimensional ones:

$$\hat{\Gamma}^a = \Gamma^a \otimes \sigma^2, \quad a = 0, \ldots, 8, \qquad \hat{\Gamma}^9 = \mathbb{I}_{16\times 16} \otimes i\sigma^1, \tag{B.79}$$

where Γ^8 satisfies

$$\Gamma^8 = \Gamma^0 \cdots \Gamma^7. \tag{B.80}$$

As usual, it will be proportional to the eight-dimensional chiral matrix $\Gamma_{(8)\,9}$ (see below). One can explicitly check that, with these definitions, the ten-dimensional representation of the gamma matrices is indeed chiral and $\hat{\Gamma}_{11} = \mathbb{I}_{16\times 16} \otimes \sigma^3$.

B.1.6 Eight dimensions

The purely nine-dimensional gamma matrices we are using can be constructed in the standard way from a purely real eight-dimensional representation (which is not chiral):

$$\Gamma^a = \Gamma^a_{(8)}, \quad a = 0, \ldots, 7, \qquad \Gamma^8 = \Gamma^0 \cdots \Gamma^7. \tag{B.81}$$

The chirality matrix is defined by

$$\Gamma_{(8)\,9} = i\Gamma^8 = i\Gamma^0 \cdots \Gamma^7. \tag{B.82}$$

We will not be able to decompose this representation in terms of a seven-dimensional representation. There are no purely real or imaginary (in Lorentzian signature) representations of the gamma matrices in seven dimensions. Thus, we cannot decompose $\Gamma^a_{(8)} = \Gamma^a_{(7)} \otimes A_{2\times 2}$ with the same factor matrix A for all $a = 0, \ldots, 6$. Thus, it is impossible to use this representation to perform a dimensional reduction from $d = 8$ to $d = 7, 6, 5$ dimensions because we would break Lorentz invariance.

However, it is possible to reduce directly to four dimensions. If $\{\gamma^a\}$ is a Majorana (purely imaginary) representation of the four-dimensional gamma matrices, the purely real representation of the eight-dimensional ones can be constructed in this way:

$$\begin{aligned} \Gamma^a_{(8)} &= \gamma^a \otimes \sigma^2 \otimes \sigma^1, \quad a = 0, 1, 2, 3, & \Gamma^4_{(8)} &= \Gamma^4_{(5)} \otimes \sigma^3 \otimes \sigma^3, \\ \Gamma^5_{(8)} &= \Gamma^4_{(5)} \otimes \sigma^1 \otimes \sigma^3, & \Gamma^6_{(8)} &= \mathbb{I}_{4\times 4} \otimes \mathbb{I}_{2\times 2} \otimes i\sigma^2, \\ \Gamma^7_{(8)} &= \mathbb{I}_{4\times 4} \otimes i\sigma^2 \otimes \sigma^3, \end{aligned} \tag{B.83}$$

where

$$\Gamma^4_{(5)} = \gamma^0\gamma^1\gamma^2\gamma^3. \tag{B.84}$$

Since a four-dimensional Majorana representation exists, this proves the existence of the nine-, ten-, and 11-dimensional representations we are using. For the purpose of dimensional reduction of the gamma matrices from $d = 10$ to $d = 4$ this is not the best representation. It is more convenient to use those of [192, 457].

B.1.7 Two dimensions

A (purely imaginary) Majorana–Weyl representation of the two-dimensional Clifford algebra is given by

$$\gamma^0_{(2)} = \sigma^2, \qquad \gamma^1_{(2)} = i\sigma^1. \tag{B.85}$$

The chiral matrix is, as expected in a Weyl representation,

$$\gamma_{(2)\,3} = \gamma^0_{(2)}\gamma^1_{(2)} = \sigma^3. \tag{B.86}$$

The two-dimensional gamma matrices are sometimes denoted by ρ^a.

B.1.8 Three dimensions

We can build a purely imaginary Majorana representation from the two-dimensional (purely imaginary) Majorana–Weyl representation:

$$\gamma^a_{(3)} = \gamma^a_{(2)}, \quad a = 0, 1, \qquad \gamma^2_{(3)} = -i\gamma^0_{(2)}\gamma^1_{(2)} = -i\sigma^3. \tag{B.87}$$

B.1.9 Four dimensions

Given the above Majorana representation (purely imaginary) of the three-dimensional gamma matrices, we can build a Majorana representation (purely imaginary) of the four-dimensional gamma matrices:

$$\gamma^a = \gamma^a_{(3)} \otimes \sigma^3, \quad a = 0, 1, 2, \qquad \gamma^3 = \mathbb{I}_{2\times 2} \otimes i\sigma^1. \tag{B.88}$$

The chiral matrix is

$$\gamma_5 = -i\gamma^0\gamma^1\gamma^2\gamma^3 = \frac{i}{4!}\epsilon_{abcd}\gamma^{abcd}, \tag{B.89}$$

and, using the explicit form of the three-dimensional gamma matrices,

$$\gamma_5 = \mathbb{I}_{2\times 2} \otimes \sigma^2. \tag{B.90}$$

It is obviously Hermitian and imaginary (and therefore antisymmetric). In this representation we can use as charge-conjugation matrix $\mathcal{C}_- = i\gamma^0$, which is real and antisymmetric. The Majorana condition says that Majorana spinors are purely real spinors, $\psi = \psi^*$.

There is another possible choice; namely $\mathcal{C}_+ = \gamma_5\gamma^0$, which is also real and antisymmetric and would impose the following condition on Majorana spinors: $\psi = -i\gamma_5\psi^*$. This is inconsistent (just take the complex conjugate of this relation) and so with it we can define only symplectic-Majorana spinors.

We can also build a Weyl representation that is complex:

$$\gamma^a = \gamma^a_{(3)} \otimes \sigma^2, \quad a = 0, 1, 2, \qquad \gamma^3 = \mathbb{I}_{2\times 2} \otimes i\sigma^1. \tag{B.91}$$

With the above Majorana representation of three-dimensional gamma matrices, we find

$$\gamma_5 = \mathbb{I}_{2\times 2} \otimes \sigma^3. \tag{B.92}$$

There are no Majorana–Weyl fermions in four dimensions and there are no Majorana–Weyl representations of the gamma matrices. The Weyl and Majorana representations given here are related by the similarity transformation (which is valid also for γ_5)

$$\gamma^a_{\mathrm{M}} = S\gamma^a_{\mathrm{W}}S^{-1}, \qquad S = \mathbb{I}_{2\times 2} \otimes (\mathbb{I}_{2\times 2} - i\sigma^1), \tag{B.93}$$

We also have the identity

$$\gamma^{a_1\cdots a_n} = \frac{(-1)^{[n/2]}i}{(4-n)!}\epsilon^{a_1\cdots a_n b_1\cdots b_{4-n}}\gamma_{b_1\cdots b_{4-n}}\gamma_5. \tag{B.94}$$

Using this identity the $d = 4$ Fierz identities for anticommuting spinors take the form

$$\begin{aligned}(\bar{\lambda}M\chi)(\bar{\psi}N\varphi) = {}& -\tfrac{1}{4}(\bar{\lambda}MN\varphi)(\bar{\psi}\chi) - \tfrac{1}{4}(\bar{\lambda}M\gamma^a N\varphi)(\bar{\psi}\gamma_a\chi) \\ & + \tfrac{1}{8}(\bar{\lambda}M\gamma^{ab}N\varphi)(\bar{\psi}\gamma_{ab}\chi) + \tfrac{1}{4}(\bar{\lambda}M\gamma^a\gamma_5 N\varphi)(\bar{\psi}\gamma_a\gamma_5\chi) \\ & - \tfrac{1}{4}(\bar{\lambda}M\gamma_5 N\varphi)(\bar{\psi}\gamma_5\chi).\end{aligned} \tag{B.95}$$

B.1.10 Five dimensions

There are no Majorana representations in $d = 5$, but only pairs of (complex) symplectic-Majorana spinors that can be combined into a single unconstrained Dirac spinor.

Using any representation of the four-dimensional gamma matrices γ^a, $a = 0, 1, 2, 3$, we can construct a five-dimensional representation (which is necessarily complex, even if the four-dimensional gamma matrices are purely imaginary)

$$\hat{\gamma}^a = \gamma^a, \quad a = 0, 1, 2, 3, \qquad \hat{\gamma}^4 = -i\gamma_5 = \gamma_0\gamma_1\gamma_2\gamma_3. \tag{B.96}$$

Then the product of all the five-dimensional gammas is

$$\hat{\gamma}^0 \cdots \hat{\gamma}^4 = +1. \tag{B.97}$$

B.1.11 Six dimensions

With the above representation of the five-dimensional gamma matrices, we can construct a six-dimensional representation

$$\hat{\hat{\gamma}}^{\hat{a}} = \hat{\gamma}^{\hat{a}} \otimes \sigma^1, \quad \hat{a} = 0, 1, 2, 3, 4, \qquad \hat{\hat{\gamma}}^5 = \mathbb{I}_{4\times 4} \otimes i\sigma^2, \tag{B.98}$$

which is a Weyl representation, since the chirality matrix $\hat{\hat{\gamma}}_7$ is

$$\hat{\hat{\gamma}}_7 = \hat{\hat{\gamma}}_0 \cdots \hat{\hat{\gamma}}_5 = \mathbb{I}_{4\times 4} \otimes \sigma^3. \tag{B.99}$$

A useful formula is

$$\hat{\hat{\gamma}}^{\hat{\hat{a}}_1 \cdots \hat{\hat{a}}_n} = \frac{(-1)^{[n/2]}}{(6-n)!} \hat{\hat{\epsilon}}^{\hat{\hat{a}}_1 \cdots \hat{\hat{a}}_n \hat{\hat{b}}_1 \cdots \hat{\hat{b}}_{6-n}} \hat{\hat{\gamma}}_{\hat{\hat{b}}_1 \cdots \hat{\hat{b}}_{6-n}} \hat{\hat{\gamma}}_7. \tag{B.100}$$

B.2 Spaces with arbitrary signatures

We now want to generalize our results on spinors and gamma matrices to d-dimensional spaces with signatures $(+^t, -^s)$, where t is the number of timelike dimensions and s is the number of spacelike dimensions. The essential reference is [642] and other useful references are [404, 828, 878], which we roughly follow.

The general setup is the same as in the signature-$(1, d-1)$ case: we consider the generators of the Clifford algebra associated with the metric $\eta_{ab} = \mathrm{diag}(+^t, -^s)$, where the indices are $a, b = -(t-1), -(t-2), \ldots, 0, 1, \ldots, s$, which is the metric of $\mathrm{SO}(t, s)$. These are the $2^{[d/2]} \times 2^{[d/2]}$ gamma matrices Γ^a which satisfy the usual anticommutation relations and out of which one can build the generators of $\mathrm{so}(t, s)$ in the spinorial representation in the usual form. They are unique up to similarity transformations. The complex $2^{[d/2]}$-component vectors in the representation space are Dirac spinors. We consider unitary representations and, therefore, all timelike (spacelike) gamma matrices are Hermitian (anti-Hermitian):

$$\Gamma^{a\,\dagger} = +\Gamma^a, \quad a \leq 0, \qquad \Gamma^{a\,\dagger} = -\Gamma^a, \quad a > 0. \tag{B.101}$$

A representation can be constructed by "Wick-rotating" the signature-$(1, d-1)$ matrices, multiplying them by factors of i if necessary. Given a d-even representation, one can construct the chirality matrix $\mathcal{Q} = \Gamma_{d+1}$,

$$\Gamma_{d+1} = \varphi(s,t)\Gamma^{-(t-1)}\Gamma^{-(t-2)} \cdots \Gamma^{-1}\Gamma^0\Gamma^1 \cdots \Gamma^s, \qquad \varphi(s,t) = -e^{\frac{\pi i}{4}(s-t)}, \tag{B.102}$$

which is unitary and Hermitian and anticommutes with all the Γ^as. Using it, we can construct a representation of the $(d+1)$-dimensional gamma matrices: if the signature is $(t, s+1)$ we define

$$\Gamma^{s+1} = -i\Gamma_{d+1}, \tag{B.103}$$

and, if the signature is $(t+1, s)$, we simply define

$$\Gamma^{-t} = \Gamma_{d+1}. \tag{B.104}$$

Thus, in even dimensions the gamma matrices are independent and in odd dimensions they are not: the product of all the gamma matrices is a power of the imaginary unit.

In even dimensions, $\{\pm\Gamma^a\}$, $\{\pm\Gamma^{a\,\dagger}\}$, $\{\pm\Gamma^{a\,\mathrm{T}}\}$, and $\{\pm\Gamma^{a\,*}\}$ generate equivalent representations. In odd dimensions, however, due to the above-mentioned constraint, only one sign gives an equivalent representation. The matrices of the corresponding similarity transformations are the chirality matrix $\mathcal{Q}$ (Γ_{d+1}), the Dirac matrix $\mathcal{D}_\pm$, the charge conjugation matrix $\mathcal{C}_\pm$, and the $\mathcal{B}_\pm$ matrix:

$$\begin{aligned} \mathcal{Q}\,\Gamma^a\,\mathcal{Q}^{-1} &= -\Gamma^a, & \mathcal{D}_\pm\,\Gamma^a\,\mathcal{D}_\pm^{-1} &= \pm\Gamma^{a\,\dagger}, \\ \mathcal{C}_\pm\,\Gamma^a\,\mathcal{C}_\pm^{-1} &= \pm\Gamma^{a\,\mathrm{T}}, & \mathcal{B}_\pm\,\Gamma^a\,\mathcal{B}_\pm^{-1} &= \pm\Gamma^{a\,*}. \end{aligned} \tag{B.105}$$

In even dimensions all these matrices exist and, evidently,

$$\mathcal{D}_\pm = \mathcal{D}_\mp\mathcal{Q}, \qquad \mathcal{C}_\pm = \mathcal{C}_\mp\mathcal{Q}, \qquad \mathcal{B}_\pm = \mathcal{B}_\mp\mathcal{Q}. \tag{B.106}$$

In odd dimensions $\mathcal{Q}$ does not exist and only one of the $\mathcal{C}_\pm$ and $\mathcal{B}_\pm$ exists.

In general, $\mathcal{D}$ is defined (up to a phase α) by

$$\mathcal{D} = \alpha\Gamma^0\Gamma^{-1}\cdots\Gamma^{-(t-1)}. \tag{B.107}$$

In our conventions we find

$$\mathcal{D}\,\Gamma^a\,\mathcal{D}^{-1} = (-1)^{t+1}\Gamma^{a\,\dagger}, \tag{B.108}$$

and, thus,

$$\mathcal{D} = \mathcal{D}_+, \quad \text{for odd } t, \qquad \mathcal{D} = \mathcal{D}_-, \quad \text{for even } t, \tag{B.109}$$

and then one has the relations

$$\mathcal{C}_\pm = \mathcal{B}_\pm^{\mathrm{T}}\mathcal{D}, \quad \text{for odd } t, \qquad \mathcal{C}_\pm = \mathcal{B}_\mp^{\mathrm{T}}\mathcal{D}, \quad \text{for even } t, \tag{B.110}$$

so the existence and properties of $\mathcal{C}_\pm$ are determined by the existence and properties of $\mathcal{B}_\pm$. The main result is[11]

$$\mathcal{B}_\pm^{\mathrm{T}} = \varepsilon_\pm(t,s)\mathcal{B}_\pm, \qquad \varepsilon_\pm(t,s) = \mathrm{sqcos}\left[\frac{\pi}{4}(s-t\pm 1)\right]. \tag{B.111}$$

When $\varepsilon = \pm 1$, $\mathcal{B}$ is symmetric or antisymmetric. When $\varepsilon_\pm = 0$, $\mathcal{B}_\pm$ does not exist. The value depends on $(s-t) \bmod 8$ and it is represented in Table B.2, from [878]. Observe that, since these matrices are assumed to be unitary, we also have

$$\mathcal{B}_\pm^*\mathcal{B}_\pm = \varepsilon_\pm(t,s). \tag{B.112}$$

Thus, for instance, when $s = t$ only $\mathcal{B}_-$ exists and is symmetric, whereas for $s = t+1$, both $\mathcal{B}_+$ and $\mathcal{B}_-$ exist and are, respectively, antisymmetric and symmetric.

---

[11] The function $\mathrm{sqcos}\,\theta$ is defined as the projection on the x axis of the line that forms an angle θ with the x axis and joins the origin to a square centered on the origin and with sides of length 2. Then $\mathrm{sqcos}(-\pi/4, 0, \pi/4) = +1$, $\mathrm{sqcos}(3\pi/4, \pi, 5\pi/4) = -1$, and $\mathrm{sqcos}(\pi/2, 3\pi/2) = 0$.

Table B.2. Possible values of $\varepsilon_\pm(\pm)$.

	$s-t$							
	1	2	3	4	5	6	7	8
ε_+	0	-1	-1	-1	0	$+1$	$+1$	$+1$
ε_-	$+1$	$+1$	0	-1	-1	-1	0	$+1$

The symmetry of $\mathcal{B}_\pm$ determines that of the charge-conjugation matrix $\mathcal{C}_\pm$

$$\begin{aligned} \mathcal{C}_\pm^{\mathrm{T}} &= \varepsilon_\pm(\pm 1)^t(-1)^{\frac{t(t-1)}{2}}\mathcal{C}_\pm, \quad \text{for odd } t, \\ \mathcal{C}_\pm^{\mathrm{T}} &= \varepsilon_\mp(-1)^{\frac{t(t-1)}{2}}\mathcal{C}_\pm, \quad \text{for even } t. \end{aligned} \tag{B.113}$$

We can now define the Dirac $\bar{\psi}$ and Majorana ψ^c conjugates of a spinor ψ:

$$\bar{\psi} = \psi^\dagger\mathcal{D}, \qquad \psi^c = \psi^{\mathrm{T}}\mathcal{C}. \tag{B.114}$$

The existence of these conjugation operations is due to the equivalence of the Hermitian conjugate and transposed representations of the gamma matrices.

Now we can proceed to define various types of constrained spinors.

Weyl spinors. In any even dimension these are eigenspinors of the chirality matrix, which has only eigenvalues $+1$ and -1 because $\mathcal{Q}^2 = 1$. Since it is traceless, half of the eigenvalues are $+1$ and half are -1. $\mathcal{Q}$ commutes with all the so(t, s) generators, which means that the (Dirac) spinorial representation is reducible to the direct sum of the two Weyl spinorial representations.

(Pseudo-)Majorana spinors. They satisfy the reality constraint

$$\bar{\psi} = \psi^c, \tag{B.115}$$

which, using the relation among $\mathcal{D}$, $\mathcal{C}$, and $\mathcal{B}$, can be rewritten in the form

$$\psi^* = \mathcal{B}\psi. \tag{B.116}$$

On taking the complex conjugate of this equation, we find the consistency condition

$$\mathcal{B}\mathcal{B}^* = \varepsilon = +1. \tag{B.117}$$

Only in this case is it possible to define Majorana spinors. If the equation is satisfied by $\mathcal{B}_-$, the spinors are called Majorana spinors. If it is satisfied by $\mathcal{B}_+$, they are called pseudo-Majorana spinors. Thus, Table B.2 can be reinterpreted in terms of the existence of Majorana (M) or pseudo-Majorana (pM) spinors as in Table B.3.

(Pseudo-)Majorana–Weyl spinors. The Majorana and Weyl conditions are compatible if

$$\mathcal{D}^{-1}\mathcal{Q}^\dagger\mathcal{D} = \mathcal{C}^{-1}\mathcal{Q}^{\mathrm{T}}\mathcal{C}, \tag{B.118}$$

Table B.3. Possible spinors in $d = t + s$ dimensions with signatures $(+^t, -^s)$. M stands for Majorana, pM for pseudo-Majorana, SM for symplectic-Majorana, pSM for pseudo-symplectic-Majorana, MW for Majorana–Weyl, and pMW for pseudo-Majorana–Weyl; * meaning that d has to be even. In addition to this, Weyl spinors are possible for any even d.

$s - t$							
1	2	3	4	5	6	7	8
	pSM	pSM	pSM				
					pM	pM	pM
M	M						M
			SM	SM	SM		
							pMW*
							MW*

which, using the relation among $\mathcal{D}$, $\mathcal{C}$, and $\mathcal{B}$, can be simplified to

$$\mathcal{Q}^* = \mathcal{B}\mathcal{Q}\mathcal{B}^{-1}, \tag{B.119}$$

which is satisfied for $s - t = 0 \bmod 4$.

(Pseudo-)symplectic-Majorana spinors. When $\mathcal{B}\mathcal{B}^* = -1$ the Majorana reality condition cannot be consistently imposed. However, then one can introduce an even number of Dirac spinors labeled by $i = 1, \ldots, 2n$ and impose the reality condition

$$\bar{\psi}^i = \psi_i{}^c \equiv \Omega_{ij} \psi^{j\,c}, \tag{B.120}$$

where Ω is real and satisfies

$$\Omega_{ij}\Omega_{jk} = -\delta_{ik}. \tag{B.121}$$

This condition can be rewritten in the more transparent form

$$\psi^{i\,*} = \Omega_{ij}\mathcal{B}\psi^j, \tag{B.122}$$

which is consistent if $\mathcal{B}\mathcal{B}^* = -1$. The cases in which these spinors can be defined are represented in Table B.3.

Now we are going to use these results in several examples of interest.

B.2.1 AdS$_4$ *gamma matrices and spinors*

The spinor representations of SO(2, 3) (which we also refer to as AdS$_4$) have the same dimension (four) as those of SO(1, 3). The corresponding gamma matrices, which we write with hats, are 4×4 matrices and any representation of them includes a representation of the SO(1, 3) (unhatted) gamma matrices. Furthermore, it is clear that AdS$_4$ spinors transform as Lorentz spinors under the Lorentz subgroup. Our goal now will be to construct an explicit representation of the gamma matrices and the generators of the AdS_4 group. Since $s-t=1$ we expect only a Majorana representation. It appears in two forms, which are equivalent through similarity transformations. We call them *electric* and *magnetic* and denote them by a $-$ or a $+$ subscript or superscript, respectively.

The magnetic representation. It is built by the standard procedure from the Clifford algebra associated with the SO(2, 3) metric $\hat{\eta}_{\hat{a}\hat{b}}$: first we construct five gamma matrices satisfying

$$\left\{\hat{\gamma}_{\hat{a}}, \hat{\gamma}_{\hat{b}}\right\} = +2\hat{\eta}_{\hat{a}\hat{b}}. \tag{B.123}$$

These matrices can be constructed by using the four SO(1, 3) Dirac matrices:

$$\hat{\gamma}_{+\,-1} = \gamma_5 = -i\gamma^0\cdots\gamma^3, \qquad \hat{\gamma}_{+\,a} = \gamma_a, \tag{B.124}$$

and, using purely imaginary Dirac matrices, we obtain a purely imaginary representation of the SO(2, 3) Clifford algebra.

The SO(2, 3) generators in the magnetic spinorial representation are constructed from the Clifford algebra in the usual fashion

$$\Gamma_+\left(\hat{M}_{\hat{a}\hat{b}}\right) = \tfrac{1}{2}\hat{\gamma}_{+\,\hat{a}\hat{b}}, \tag{B.125}$$

and they automatically satisfy the so(2, 3) algebra Eq. (4.152).

SO(2, 3) spinors $\hat{\psi}_+{}^{\alpha}$ transform with the exponential of all these generators and are, in particular, Lorentz spinors.

Since $t=2$, we know that $\mathcal{D}=\mathcal{D}_-$. Furthermore, the only $\mathcal{B}$ leading to consistent Majorana spinors is $\mathcal{B}_-$ and thus we can use only $\mathcal{C}_+$, which is antisymmetric. We can take

$$\mathcal{C}_+ = \mathcal{D}_- = \hat{\gamma}_+^0\hat{\gamma}_+^{-1} = \gamma^0\gamma_5. \tag{B.126}$$

It is easy to check that the charge-conjugation matrix $\mathcal{C}_+$ satisfies

$$\mathcal{C}_+\hat{\gamma}^{\hat{a}}\mathcal{C}_+^{-1} = +\hat{\gamma}^{\hat{a}\,\mathrm{T}} = -\hat{\gamma}^{\hat{a}\,\dagger}. \tag{B.127}$$

Since the Dirac conjugation and charge-conjugation matrices are identical, $\mathcal{B}_- = \mathbb{I}_{4\times 4}$, the Majorana condition $\bar{\hat{\psi}}_+ = \hat{\psi}_+^c$ implies that Majorana spinors are purely real spinors in this representation, $\hat{\psi}_+^* = \hat{\psi}_+$.

In this representation the ten matrices

$$\left[\Gamma_+\left(\hat{M}^{\hat{a}\hat{b}}\right)\mathcal{C}_+^{-1}\right]^{\alpha\beta} \tag{B.128}$$

are real and symmetric. This is necessary in order to build the osp($N/4$) supersymmetry algebra. The six matrices

$$\left(\mathcal{C}_+^{-1}\right)^{\alpha\beta}, \qquad \left(i\hat{\gamma}_+^{\hat{a}}\mathcal{C}_+^{-1}\right)^{\alpha\beta} \tag{B.129}$$

are real and antisymmetric and we will use them to add other ("central") charges in the anticommutator $\{Q^{\alpha\,i}, Q^{\beta\,j}\}$ supersymmetry algebra.

These 16 matrices are a basis of the linear space of real 4×4 matrices. We will later work out the details simultaneously with analogous matrices of the electric representation. For the moment it suffices to observe that other antisymmetrized products of gamma matrices can be related to the above antisymmetrized products of zero, one, and two gamma matrices via

$$\hat{\gamma}_+^{\hat{a}_1\cdots\hat{a}_n} \sim \frac{1}{(5-n)!}\hat{\epsilon}^{\hat{a}_1\cdots\hat{a}_5}\hat{\gamma}_{+\,\hat{a}_{n+1}\cdots\hat{a}_5}, \tag{B.130}$$

with $\hat{\epsilon}^{-10123} = +1$. In particular,

$$\hat{\gamma}_+^{\hat{a}\hat{b}\hat{c}\hat{d}\hat{e}} = i\hat{\epsilon}^{\hat{a}\hat{b}\hat{c}\hat{d}\hat{e}}. \tag{B.131}$$

The electric representation. This is the representation that is used most often. The SO(2, 3) generators in the *electric* spinorial representation Γ_- are built directly from the SO(1, 3) Dirac gamma matrices γ^a:

$$\Gamma_-\left(\hat{M}_{ab}\right) = \tfrac{1}{2}\gamma^{ab}, \qquad \Gamma_-\left(\hat{M}_{a-1}\right) = \frac{i}{2}\gamma^a. \tag{B.132}$$

It can be checked that these matrices satisfy the so(2, 3) algebra Eq. (4.152).

In this representation

$$i\gamma^0\Gamma_-\left(\hat{M}_{\hat{a}\hat{b}}\right)\left(i\gamma^0\right)^{-1} = -\Gamma_-\left(\hat{M}_{\hat{a}\hat{b}}\right)^\dagger = -\Gamma_-\left(\hat{M}_{\hat{a}\hat{b}}\right)^{\mathrm{T}}, \tag{B.133}$$

which implies that we can take as Dirac and charge-conjugation matrices

$$\mathcal{D} = \mathcal{D}_+ = \mathcal{C}_- = i\gamma^0, \tag{B.134}$$

which coincide with the ones we used in $d = 4$ dimensions with signature $(1, 3)$. This may seem contradictory. However, we have not yet identified from which representation of the Clifford algebra the above so(2, 3) representation arises. Actually, it can be constructed by the standard procedure from the following hatted gamma matrices:

$$\hat{\gamma}_{-\,-1} = \gamma_5, \qquad \hat{\gamma}_{-\,a} = i\gamma_a\gamma_5, \tag{B.135}$$

which provide a purely imaginary representation of the Clifford algebra associated with $\hat{\eta}_{\hat{a}\hat{b}}$. Then, we can see that the Dirac conjugation and charge-conjugation matrices are given by the product of the two timelike gammas, as in the *magnetic* case:

$$\mathcal{D}_+ = \mathcal{C}_- = \hat{\gamma}_-^0\hat{\gamma}_-^{-1} = i\gamma^0. \tag{B.136}$$

As is needed for supersymmetry, in the electric representation the ten matrices

$$\left[\Gamma_-\left(\hat{M}^{\hat{a}\hat{b}}\right)\mathcal{C}_-^{-1}\right]^{\alpha\beta} \tag{B.137}$$

are also real and symmetric, and the six matrices

$$\left(\mathcal{C}_-^{-1}\right)^{\alpha\beta}, \qquad \left(i\hat{\gamma}_-^{\hat{a}}\mathcal{C}_-^{-1}\right)^{\alpha\beta} \tag{B.138}$$

are real and antisymmetric.

Although these two representations are built in different ways, they are equivalent: they are related by a complete chiral–dual-type change of basis in spinor space:[12]

$$
\begin{aligned}
\hat{\psi}_+ &= S\hat{\psi}_-, \\
\Gamma_+\left(\hat{M}_{\hat{a}\hat{b}}\right) &= S\Gamma_-\left(\hat{M}_{\hat{a}\hat{b}}\right)S^{-1}, \\
\mathcal{C}_+ &= (S^{-1})^{\mathrm{T}}\mathcal{C}_- S^{-1}, \\
S &= \frac{1}{\sqrt{2}}(1+i\gamma_5).
\end{aligned}
\tag{B.139}
$$

Observe that, given that $\mathcal{C}_\pm$ is always real and antisymmetric and squares to minus the identity, the condition (from now on we suppress $+$ and $-$ subindices)

$$\mathcal{C}\Gamma\left(\hat{M}^{\hat{a}\hat{b}}\right)\mathcal{C}^{-1} = -\Gamma\left(\hat{M}^{\hat{a}\hat{b}}\right)^{\mathrm{T}} \tag{B.140}$$

is equivalent to the statement that the 4×4 matrices $\Gamma\left(\hat{M}^{\hat{a}\hat{b}}\right)^{\alpha}{}_{\beta}$ are at the same time a spinorial representation of the algebra so(2, 3) and a fundamental representation of the algebra sp(4, $\mathbb{R}$). Using $\mathcal{C}_{\alpha\beta}$ as a metric to raise and lower indices,[13] one can construct real, symmetric representations of sp(4, $\mathbb{R}$) [443] $m^{\hat{a}\hat{b}\,\alpha\beta}$, where

$$m^{\hat{a}\hat{b}\,\alpha\beta} = \left[\Gamma\left(\hat{M}^{\hat{a}\hat{b}}\right)\mathcal{C}^{-1}\right]^{\alpha\beta}. \tag{B.141}$$

These objects satisfy the identity

$$m^{\hat{a}\hat{b}\,\alpha\beta}m_{\hat{c}\hat{d}\,\alpha\beta} = 2\delta^{[\hat{a}\hat{b}]}{}_{[\hat{c}\hat{d}]}, \tag{B.142}$$

which simply states that these matrices are an orthonormal basis in the ten-dimensional space of 4×4 real symmetric matrices with the trace of the standard product of matrices as scalar product and, therefore, for any symmetric matrix $O_{\alpha\beta}$,

$$O_{\alpha\beta} = \tfrac{1}{2}m^{\hat{a}\hat{b}\,\gamma\delta}m_{\hat{a}\hat{b}\,\alpha\beta}O_{\gamma\delta}, \quad \Rightarrow m^{\hat{a}\hat{b}\,\gamma\delta}m_{\hat{a}\hat{b}\,\alpha\beta} = 2\delta^{(\gamma\delta)}{}_{(\alpha\beta)}, \tag{B.143}$$

and, by definition of $m^{\hat{a}\hat{b}}$, we obtain the identity

$$m^{\hat{a}\hat{b}\,\alpha\beta}m_{\hat{a}\hat{b}}{}^{\gamma\delta} = \left(\mathcal{C}^{-1}\right)^{\alpha\gamma}\left(\mathcal{C}^{-1}\right)^{\beta\delta} + \left(\mathcal{C}^{-1}\right)^{\alpha\delta}\left(\mathcal{C}^{-1}\right)^{\beta\gamma}, \tag{B.144}$$

which is crucial for the consistency of the osp($4/N$) superalgebra.

The matrices $m^{\hat{a}\hat{b}\,\alpha\beta}$ can also be used to convert objects in the adjoint of so(2, 3) into objects in the fundamental of sp(4, $\mathbb{R}$), which are somewhat easier to deal with.

Let us now consider the six real, antisymmetric matrices $n^{\hat{\hat{a}}\,\alpha\beta}$,

$$
\begin{aligned}
n^{\hat{a}\,\alpha\beta} &= \frac{1}{\sqrt{2}}\left(i\hat{\gamma}^{\hat{a}}\mathcal{C}^{-1}\right)^{\alpha\beta}, \\
n^{4\,\alpha\beta} &= \frac{1}{\sqrt{2}}\left(\mathcal{C}^{-1}\right)^{\alpha\beta},
\end{aligned}
\tag{B.145}
$$

[12] By this we mean a change of basis, not just a rotation of the spinors as in Eqs. (5.84).

[13] Upper-left indices are contracted with adjacent lower-right indices: $\xi_\alpha = \xi^\beta\mathcal{C}_{\beta\alpha} = -\mathcal{C}_{\alpha\beta}\xi^\beta$.

labeled by the index $\hat{\hat{a}} = (\hat{a}, 4)$ which we raise and lower with the SO(2, 4) metric $\hat{\hat{\eta}}_{\hat{\hat{a}}\hat{\hat{b}}} = \mathrm{diag}(++----)$. It can be proved that these matrices are an orthonormal basis in the space of real 4×4 antisymmetric matrices with the trace as scalar product (rasing and lowering indices with $\mathcal{C}$):

$$\begin{aligned} n^{\hat{\hat{a}}\,\alpha\beta} n_{\hat{\hat{b}}\,\beta\alpha} &= 2\delta^{\hat{\hat{a}}}{}_{\hat{\hat{b}}}, \\ n^{\hat{\hat{a}}\,\gamma\delta} n_{\hat{\hat{a}}\,\alpha\beta} &= -2\delta^{[\gamma\delta]}{}_{[\alpha\beta]}. \end{aligned} \tag{B.146}$$

Appendix C
n-Spheres

An n-dimensional unit-radius[1] sphere S^n is the hypersurface of $\mathbb{R}^{n+1}$ defined by $(x^1)^2 + \cdots + (x^{n+1})^2 = 1$. It is usually parametrized in terms of spherical coordinates $\{r, \varphi, \theta_1, \ldots, \theta_{n-1}\}$,

$$\begin{aligned}
x^1 &= \rho_{n-1} \sin\varphi, \\
x^2 &= \rho_{n-1} \cos\varphi, \\
x^3 &= \rho_{n-2} \cos\theta_1, \\
\vdots\ \ & \ \ \vdots \\
x^k &= \rho_{n-k+1} \cos\theta_{k-2}, \qquad 3 \leq k \leq n+1,
\end{aligned} \tag{C.1}$$

where

$$\begin{aligned}
\rho_l &= [(x^1)^2 + \cdots + (x^{n+1-l})^2]^{\frac{1}{2}} = r \prod_{m=1}^{l} \sin\theta_{n-m}, \\
\rho_0 &= r = [(x^1)^2 + \cdots + (x^{n+1})^2]^{\frac{1}{2}},
\end{aligned} \tag{C.2}$$

and $\varphi \in [2, 2\pi]$, $\theta_i \in [0, \pi]$, setting $r = 1$. The metric induced on S^n in spherical coordinates is denoted by $d\Omega^2_{(n)}$ and is implicitly defined in

$$d\vec{x}^{\,2}_{(n+1)} = d\rho_0^2 + \rho_0^2 d\theta_{n-1}^2 + \cdots + \rho_{n-2}^2 d\theta_1^2 + \rho_{n-1}^2 d\varphi^2 \equiv dr^2 + r^2 d\Omega^2_{(n)}. \tag{C.3}$$

In practice, it is convenient to use the recursive formula

$$d\Omega^2_{(n)} = d\theta^2_{n-1} + \sin^2\theta_{n-1}\, d\Omega^2_{(n-1)}, \qquad d\Omega^2_{(1)} = d\varphi^2. \tag{C.4}$$

The spheres equipped with this metric, which is clearly $\mathrm{SO}(n+1)$-invariant, are called *round spheres* (see Appendix A.4.2). Other metrics with less symmetry on the same S^n manifolds are possible, but sometimes a different notation is used to denote the corresponding Riemannian spaces.

[1] For a topological space, the radius is irrelevant, but it becomes relevant when we consider the metric induced from the Euclidean metric of $\mathbb{R}^{n+1}$.

For some purposes, such as the calculation of the curvature in spacetimes with spherical symmetry, it is convenient to rename the coordinates φ and θ_k and use $\psi_i, \ \ i = 1, \ldots, n$ with

$$\psi_i = \theta_{n-i}, \qquad i = 1, \ldots, n-1, \qquad\qquad \psi_n = \varphi, \tag{C.5}$$

and define

$$q_i = r^2 \prod_{k=1}^{i-1} \sin \psi_i, \qquad i = 1, \ldots, n, \tag{C.6}$$

so the metric takes the form

$$d\vec{x}^{\,2}_{(n+1)} = q_0 dr^2 + \sum_{i=1}^{i=n} q_i d\psi_i^2 = dr^2 + r^2 d\Omega^2_{(n)}. \tag{C.7}$$

The volume form on S^n is, in spherical coordinates,

$$d\Omega^n \equiv d\varphi \prod_{i=1}^{n-1} \sin^i \theta_i \ d\theta_i. \tag{C.8}$$

In Cartesian coordinates in the embedding $(n+1)$-dimensional space it takes the form

$$d\Omega^n = \frac{1}{n! r^{n+1}} \epsilon_{\mu_1 \cdots \mu_{n+1}} x^{\mu_{n+1}} dx^{\mu_1} \cdots dx^{\mu_n}. \tag{C.9}$$

Other useful identities are

$$d^{n+1}x = r^n dr d\Omega^n, \qquad\qquad r^n d\Omega^n = d^n y \sqrt{|g|}, \tag{C.10}$$

where the y are coordinates on the n-sphere.

The volume of the unit n-sphere S^n is given by

$$\omega_{(n)} = \int_{S^n} d\Omega^n = \frac{2\pi^{\frac{n+1}{2}}}{\Gamma\left(\frac{n+1}{2}\right)}. \tag{C.11}$$

Using

$$\Gamma(x+1) = x\Gamma(x), \qquad \Gamma(0) = 1, \qquad \Gamma\left(\frac{1}{2}\right) = \pi^{1/2}, \tag{C.12}$$

one obtains $\omega_{(1)} = 2\pi, \ \omega_{(2)} = 4\pi, \ \omega_{(3)} = 2\pi^2$, etc.

The round n-spheres are globally symmetric spaces $SO(n+1)/SO(n)$ (Appendix A.4.2). There is also a description of the round spheres S^3 and S^7 as principal (Hopf) bundles in which both the base space and the fiber are spheres. The S^3 case is based on the description of S^2 as the coset manifold $SO(3)/SO(2) \sim SU(2)/U(1)$ and the general theorem (see page 604) that ensures that G ($SU(2) \sim S^3$) is a principal bundle with base G/H (S^2) and structure group H ($SO(2) \sim S^1$). Let us now study these Hopf fibrations.

C.1 S^3 and S^7 as Hopf fibrations

There is a natural action of U(1) on SU(2)

$$U \to U\begin{pmatrix} u & 0 \\ 0 & \bar{u} \end{pmatrix}, \qquad |u|^2 = 1, \tag{C.13}$$

(i.e. through shifts of ψ), that allows us to take the quotient SU(2)/U(1) that can be identified with S^2. This is why the metric on S^2 is the metric on the coset manifold SU(2)/U(1):

$$d\Omega^2_{(2)} = (e^1)^2 + (e^2)^2. \tag{C.14}$$

We can then view SU(2) (S^3) as a fiber bundle with fiber U(1) (S^1) and base space S^2. From this point of view e^3 is a U(1) connection in the bundle and its curvature coincides with that of the Dirac magnetic monopole [892] (see Section 8.7.2).

This is the simplest case $n = 1$ in the first sequence of Hopf principal fiber bundles [540],

$$S^{2n+1} \overset{U(1)}{\to} \mathbb{CP}^n, \tag{C.15}$$

since $\mathbb{CP}^1$ is nothing but the Riemann sphere S^2. Here S^{2n+1} is described by the equation in $\mathbb{C}^n$, $\bar{z}_0 z_0 + \cdots + \bar{z}_n z_n = 1$. There is another infinite sequence of Hopf fiberings,[2]

$$S^{4n+3} \overset{SU(2)}{\to} \mathbb{HP}^n, \tag{C.16}$$

where $\mathbb{H}$ is the field of quaternions. Here S^{4n+3} is described by the equation in $\mathbb{H}^n$, $\bar{z}_0 z_0 + \cdots + \bar{z}_n z_n = 1$. The first member in this series describes S^7 as a fiber bundle with SU(2) as fiber and S^4 ($\mathbb{HP}^1$) as base space. The S^7 metric can be similarly constructed [892],

$$d\Omega^2_{(7)} = \tfrac{1}{4}\left[d\Omega^2_{(4)} + \sum_{i=1}^{3} (e^i + \mathcal{A}^i)^2 \right], \tag{C.17}$$

where the e^i are the SU(2) Maurer–Cartan 1-forms and $d\Omega^2_{(4)}$ is the metric on S^4, that we construct as before,

$$d\Omega^2_{(4)} = d\chi^2 + \sin^2\chi\, d\Omega^2_{(3)}, \qquad d\Omega^2_{(3)} = \tfrac{1}{4}\sum_{i=1}^{3} (E^i)^2, \tag{C.18}$$

where the E^i are a second set of SU(2) Maurer–Cartan 1-forms and (in different coordinates) and the 1-form with su(2) indices

$$\mathcal{A}^i = -\sin^2(\chi/2)\, E^i \tag{C.19}$$

coincides with the gauge connection of the BPST instanton. This metric is also maximally symmetric (SO(8)-invariant).

[2] The last "sequence" can be defined analogously using octonions, but only the first element is well defined.

C.2 Squashed S^3 and S^7

The metrics of the round S^3 and S^7 associated with their description as Hopf fibrations can easily be deformed to those of *squashed spheres* by introducing a parameter λ:

$$
\begin{aligned}
d\tilde{\Omega}^2_{(3)} &= \tfrac{1}{4}\Big[d\Omega^2_{(2)} + \lambda^2 (e^3)^2\Big], \\
d\tilde{\Omega}^2_{(7)} &= \tfrac{1}{4}\Big[d\Omega^2_{(4)} + \lambda^2 \sum_{i=1}^{3} (e^i + \mathcal{A}^i)^2\Big].
\end{aligned}
\tag{C.20}
$$

Only for certain values of λ does one obtain Einstein metrics: $\lambda = 1$, the *round* spheres (i.e. SO(4)- and SO(8)-invariant), and, for the S^7 case only, $\lambda = 1/\sqrt{5}$. The metric of this squashed S^7 is only $SO(5) \times SO(3)$-invariant, which makes it interesting in Kaluza–Klein compactifications [58].

Appendix D
Palatini's identity

This identity allows us to express the Einstein–Hilbert action in terms of the spin-connection coefficients alone, with no partial derivatives, which are eliminated upon integrating by parts. On substituting the expression for the Ricci scalar,

$$R = 2e_a{}^\mu e_b{}^\nu \partial_{[\mu}\omega_{\nu]}{}^{ab} + \omega_a{}^{ac}\omega_b{}^b{}_c + \omega_b{}^{ac}\omega_{ac}{}^b, \tag{D.1}$$

into the Einstein–Hilbert action and integrating by parts, using the relation between the Levi-Cività connection and the spin connection

$$\partial_a \ln\sqrt{|g|} = \Gamma_{ba}{}^b = \omega_{ba}{}^b + e_a{}^\mu \partial_b e_\mu{}^b, \tag{D.2}$$

we obtain

$$\begin{aligned}\int d^d x\,\sqrt{|g|}\,K\,R = \int d^d x\,\sqrt{|g|}\,K\big\{&-2\partial_{[\mu|}\left(e_a{}^\mu e_b{}^\nu\right)\omega_{|\nu]}{}^{ab} + 2\omega_a{}^{ab}(\partial_b \ln K)\\ &+ 2e_b{}^\mu \partial_c e_\mu{}^c \omega_a{}^{ab} - \omega_b{}^{ba}\omega_c{}^c{}_a - \omega_a{}^{bc}\omega_{cb}{}^a\big\}\,.\end{aligned} \tag{D.3}$$

Simple manipulations of the two terms with explicit Vielbeins lead us to the following generalization of Palatini's identity which is often used:

$$\int d^d x\,\sqrt{|g|}\,K\,R = \int d^d x\,\sqrt{|g|}\,K\big\{-\omega_b{}^{ba}\omega_c{}^c{}_a - \omega_a{}^{bc}\omega_{bc}{}^a + 2\omega_b{}^{ba}(\partial_a \ln K)\big\}\,. \tag{D.4}$$

Observe [836] that the integrand is a scalar under reparametrizations (there are no world indices at all).

Appendix E
Conformal rescalings

If we make the local scale transformation in d dimensions

$$\tilde{g}_{\mu\nu} = \Omega^2 g_{\mu\nu}, \tag{E.1}$$

the determinant of the metric and the Christoffel symbols transform as follows:

$$\sqrt{|\tilde{g}|} = \Omega^d \sqrt{|g|}, \qquad \tilde{\Gamma}_{\mu\nu}{}^{\rho} = \Gamma_{\mu\nu}{}^{\rho} + \left(\delta_\nu{}^\rho \delta_\mu{}^\alpha + \delta_\mu{}^\rho \delta_\nu{}^\alpha - g_{\mu\nu} g^{\rho\alpha}\right) \partial_\alpha \ln \Omega. \tag{E.2}$$

So the covariant derivative of a vector (defined to be invariant with index down) and the Laplacian of a scalar transform as follows:

$$\begin{aligned} \tilde{\nabla}_\mu A_\nu &= \nabla_\mu A_\nu - 2A_{(\mu}\partial_{\nu)} \ln \Omega + g_{\mu\nu} A_\rho\, \partial^\rho \ln \Omega, \\ \tilde{\nabla}^2 s &= \Omega^{-2}\big[\nabla^2 s + (d-2)\, \partial_\mu \ln \Omega\, \partial^\mu s\big], \end{aligned} \tag{E.3}$$

where the formulae are written using only $\tilde{g}$ on the l.h.s. and only g on the r.h.s. The completely antisymmetric tensor $\epsilon^{\mu_1\cdots\mu_d}$ is scale-invariant with our conventions.

The Ricci tensor and scalar and the Einstein tensor transform as follows:

$$\begin{aligned} \tilde{R}_{\mu\nu} = R_{\mu\nu} &- (d-2)\big[\partial_\mu \ln \Omega\, \partial_\nu \ln \Omega - g_{\mu\nu}(\partial \ln \Omega)^2\big] \\ &+ (d-2)\left[\nabla_\mu \partial_\nu \ln \Omega + \frac{1}{d-2} g_{\mu\nu} \nabla^2 \ln \Omega\right], \end{aligned} \tag{E.4}$$

$$\tilde{R} = \Omega^{-2}\big[R + (d-1)(d-2)(\partial \ln \Omega)^2 + 2(d-1)\nabla^2 \ln \Omega\big], \tag{E.5}$$

$$\begin{aligned} \tilde{G}_{\mu\nu} = G_{\mu\nu} &- (d-2)\left[\partial_\mu \ln \Omega\, \partial_\nu \ln \Omega + \frac{d-3}{2} g_{\mu\nu}(\partial \ln \Omega)^2\right] \\ &+ (d-2)\big[\nabla_\mu \partial_\nu \ln \Omega - g_{\mu\nu} \nabla^2 \ln \Omega\big]. \end{aligned} \tag{E.6}$$

Appendix F
Connections and curvature components

F.1 For some $d = 4$ metrics

F.1.1 General static, spherically symmetric metrics (I)

The metric

$$ds^2 = g_{tt}(r)dt^2 + g_{rr}(r)dr^2 - r^2 d\Omega^2_{(2)} \tag{F.1}$$

leads to the Levi-Cività connection components

$$\begin{aligned} \Gamma_{tt}{}^{r} &= -\tfrac{1}{2}\partial_r g_{tt}/g_{rr}, & \Gamma_{tr}{}^{t} &= \tfrac{1}{2}\partial_r g_{tt}/g_{tt}, & \Gamma_{rr}{}^{r} &= \tfrac{1}{2}\partial_r g_{rr}/g_{rr}, \\ \Gamma_{r\theta}{}^{\theta} &= 1/r, & \Gamma_{r\varphi}{}^{\varphi} &= 1/r, & \Gamma_{\theta\theta}{}^{r} &= r/g_{rr}, \\ \Gamma_{\theta\varphi}{}^{\varphi} &= \cos\theta/\sin\theta, & \Gamma_{\varphi\varphi}{}^{r} &= \sin^2\theta\, \Gamma_{\theta\theta}{}^{r}, & \Gamma_{\varphi\varphi}{}^{\theta} &= -\sin\theta\cos\theta, \end{aligned} \tag{F.2}$$

and the Ricci tensor

$$\begin{aligned} R_{tt} &= -\frac{\sqrt{g_{tt}}\kappa'}{\sqrt{-g_{rr}}} + \frac{g'_{tt}}{rg_{rr}}, & R_{rr} &= \frac{\sqrt{-g_{rr}}\kappa'}{\sqrt{g_{tt}}} - \frac{g'_{tt}}{rg_{rr}}, \\ R_{\theta\theta} &= -\frac{rg'_{tt}}{2g_{rr}g_{tt}} + \frac{rg'_{rr}}{2g^2_{rr}} - \left(1 + \frac{1}{g_{rr}}\right), & R_{\varphi\varphi} &= \sin^2\theta\, R_{\theta\theta}, \end{aligned} \tag{F.3}$$

where the prime indicates partial derivatization with respect to r and κ is

$$\kappa = \tfrac{1}{2}\frac{g'_{tt}}{\sqrt{-g_{rr}g_{tt}}}. \tag{F.4}$$

The Ricci scalar is

$$R = 2\frac{\kappa'}{\sqrt{-g_{rr}g_{tt}}} - \frac{2}{rg_{rr}}\left[\ln\left(-\frac{g_{tt}}{g_{rr}}\right)\right]' + \frac{2}{r^2}\left(1 + \frac{1}{g_{rr}}\right). \tag{F.5}$$

If we choose the Vierbein basis

$$e_t{}^0 = \sqrt{g_{tt}}, \qquad e_r{}^1 = \sqrt{-g_{rr}}, \qquad e_\theta{}^2 = r, \qquad e_\varphi{}^3 = r\sin\theta, \tag{F.6}$$

the non-vanishing components of the spin-connection 1-form are

$$\omega_t{}^{01} = \kappa, \quad \omega_\theta{}^{12} = -\frac{1}{\sqrt{-g_{rr}}}, \quad \omega_\varphi{}^{13} = -\frac{\sin\theta}{\sqrt{-g_{rr}}}, \quad \omega_\varphi{}^{23} = -\cos\theta. \tag{F.7}$$

The non-vanishing components of the curvature 2-form are

$$\begin{aligned} R_{tr}{}^{01} &= -\kappa', & R_{t\theta}{}^{02} &= \tfrac{1}{2}g'_{tt}/(g_{rr}\sqrt{g_{tt}}), \\ R_{t\varphi}{}^{03} &= \sin\theta\, R_{t\theta}{}^{02}, & R_{r\theta}{}^{12} &= -\tfrac{1}{2}g'_{rr}/(-g_{rr})^{\frac{3}{2}}, \\ R_{r\varphi}{}^{13} &= \sin\theta\, R_{r\theta}{}^{12}, & R_{\theta\varphi}{}^{23} &= \sin\theta\left(1+\frac{1}{g_{rr}}\right). \end{aligned} \tag{F.8}$$

F.1.2 General static, spherically symmetric metrics (II)

They can be written in the form

$$ds^2 = \lambda(r)dt^2 - \lambda^{-1}(r)dr^2 - R^2(r)d\Omega^2, \tag{F.9}$$

and lead to the non-vanishing Levi-Cività connection coefficients

$$\begin{aligned} \Gamma_{tt}{}^{r} &= \tfrac{1}{2}\lambda\lambda', & \Gamma_{tr}{}^{t} &= \tfrac{1}{2}\lambda^{-1}\lambda', & \Gamma_{rr}{}^{r} &= -\tfrac{1}{2}\lambda^{-1}\lambda', \\ \Gamma_{r\theta}{}^{\theta} &= (\ln R)', & \Gamma_{r\varphi}{}^{\varphi} &= (\ln R)', & \Gamma_{\theta\theta}{}^{r} &= -\tfrac{1}{2}\lambda(R^2)', \\ \Gamma_{\theta\varphi}{}^{\varphi} &= \cos\theta/\sin\theta, & \Gamma_{\varphi\varphi}{}^{r} &= -\tfrac{1}{2}\lambda(R^2)'\sin^2\theta, & \Gamma_{\varphi\varphi}{}^{\theta} &= -\cos\theta\sin\theta. \end{aligned} \tag{F.10}$$

The components of the Ricci tensor are

$$\begin{aligned} R_{tt} &= -\frac{\lambda}{2R^2}\left(R^2\lambda'\right)', & R_{rr} &= -\lambda^{-2}R_{tt} + 2\frac{R''}{R}, \\ R_{\theta\theta} &= \tfrac{1}{2}\left[\lambda(R^2)'\right]' - 1, & R_{\varphi\varphi} &= \sin^2\theta\, R_{\theta\theta}, \end{aligned} \tag{F.11}$$

and the Ricci scalar is

$$R = -\frac{1}{R^2}\left[(R^2\lambda)'' - 2 + 2\lambda R R''\right]. \tag{F.12}$$

If we choose the Vielbein 1-form basis

$$e_t{}^0 = \lambda^{\frac{1}{2}}, \quad e_r{}^1 = \lambda^{-\frac{1}{2}}, \quad e_\theta{}^2 = R, \quad e_\varphi{}^3 = R\sin\theta, \tag{F.13}$$

we obtain the spin-connection 1-form with components

$$\omega_0{}^{01} = \tfrac{1}{2}\lambda^{-\frac{1}{2}}\lambda', \quad \omega_2{}^{21} = (\ln R)'\lambda^{\frac{1}{2}}, \quad \omega_3{}^{31} = (\ln R)'\lambda^{\frac{1}{2}}, \quad \omega_3{}^{32} = (1/R)\cot\theta, \tag{F.14}$$

and with them the curvature 2-form components

$$\begin{aligned} R_{01}{}^{01} &= -\tfrac{1}{2}\lambda'', & R_{02}{}^{02} &= -\tfrac{1}{2}(\ln R)'\lambda', \\ R_{03}{}^{03} &= -\tfrac{1}{2}(\ln R)'\lambda', & R_{12}{}^{12} &= -\frac{1}{2R}(2R''\lambda + R'\lambda'), \\ R_{13}{}^{13} &= -\frac{1}{2R}(2R''\lambda + R'\lambda'), & R_{23}{}^{23} &= -\frac{1}{R^2}\left[(R')^2\lambda - 1\right]. \end{aligned} \tag{F.15}$$

The components of the Ricci-tensor 1-form are

$$R_0{}^0 = \tfrac{1}{2}\nabla^2 \ln\lambda, \qquad R_1{}^1 = \tfrac{1}{2}\nabla^2 \ln\lambda - 2\frac{R''\lambda}{R},$$
$$R_2{}^2 = -\tfrac{1}{2}\nabla^2 \ln\lambda - \frac{(R^2\lambda)'' - 2}{2R^2}, \quad R_3{}^3 = -\tfrac{1}{2}\nabla^2 \ln\lambda - \frac{(R^2\lambda)'' - 2}{2R^2}, \tag{F.16}$$

where the form of the Laplacian of a scalar function $f(r)$ in these coordinates is

$$\nabla^2 f(r) = -R^{-2}\big(R^2\lambda f'\big)'. \tag{F.17}$$

The last two components can also be written in this simpler form:

$$R_2{}^2 = R_3{}^3 = \nabla^2 \ln R + 1/R^2. \tag{F.18}$$

F.1.3 $d = 4$ IWP-type metrics

These are stationary (not necessarily axially symmetric) metrics of the form

$$ds^2 = e^{2U}(dt+\omega)^2 - e^{-2U}d\vec{x}^{\,2}, \qquad \omega = \omega_{\underline{i}}dx^{\underline{i}}, \quad \Rightarrow \sqrt{|g|} = e^{-2U}, \tag{F.19}$$

where U and $\omega_{\underline{i}}$ are functions of $\vec{x}$ only. The components of the inverse metric are

$$g^{tt} = e^{-2U}\big(1 - e^{4U}\omega^2\big), \quad g^{t\underline{i}} = e^{2U}\omega_{\underline{i}}, \quad g^{\underline{i}\underline{j}} = -e^{2U}\delta^{ij}. \tag{F.20}$$

A convenient Vierbein 1-form basis and its dual vector basis are provided by

$$e^0 = e^U(dt+\omega), \quad e^i = e^{-U}dx^{\underline{i}}, \quad e_0 = e^{-U}\partial_t, \quad e_i = e^U\big(-\omega_{\underline{i}}\partial_t + \partial_{\underline{i}}\big), \tag{F.21}$$

and the corresponding spin-connection 1-forms are given by

$$\omega^{0i} = \partial_{\underline{i}}e^U e^0 + e^{3U}\partial_{[\underline{i}}\omega_{\underline{k}]}e^k, \qquad \omega^{ij} = e^{3U}\Big(\partial_{[\underline{i}}\omega_{\underline{j}]}e^0 - \partial_{[\underline{i}}e^{-2U}\delta_{j]k}\Big)e^k. \tag{F.22}$$

The self-dual combinations (in the upper indices) take the form

$$\omega^{+\,0i} = \frac{i}{4}e^{3U}\Big[\partial_{\underline{i}}Ve^0 - i\epsilon_{ijk}\partial_{\underline{j}}Ve^k\Big], \quad \omega^{+\,ij} = -\tfrac{1}{4}e^{3U}\big[\epsilon_{ijk}\partial_{\underline{k}}Ve^0 - 2i\partial_{[\underline{i}}V\delta_{j]k}e^k\big], \tag{F.23}$$

where

$$V = b + ie^{-2U}, \qquad \partial_{[\underline{i}}\omega_{\underline{j}]} = -\tfrac{1}{2}\epsilon_{ijk}\partial_{\underline{k}}b. \tag{F.24}$$

The components of the Ricci tensor are

$$\begin{aligned}
R_{tt} &= e^{8U}\partial_{[\underline{i}}\omega_{\underline{j}]}\partial_{[\underline{i}}\omega_{\underline{j}]} + e^{4U}\partial^2 U = \tfrac{1}{2}(\mathrm{Im}\,V)^{-4}\partial V\partial\bar{V} - \tfrac{1}{2}(\mathrm{Im}\,V)^{-3}\partial^2\mathrm{Im}\,V,\\
R_{t\underline{i}} &= -\tfrac{1}{2}e^{8U}\omega_{\underline{i}}\partial V\partial\bar{V} + \tfrac{1}{2}e^{6U}\omega_{\underline{i}}\partial^2 e^{-2U} + e^{6U}\epsilon_{ijk}\partial_j e^{-2U}\partial_k b,\\
R_{\underline{i}\underline{j}} &= -\tfrac{1}{4}e^{4U}\Big[\Big(e^{4U}\omega_{\underline{i}}\omega_{\underline{j}} + \delta_{ij}\Big)\partial V\partial\bar{V} - \partial_{(\underline{i}}V\partial_{\underline{j})}\bar{V}\Big]\\
&\quad + \tfrac{1}{2}e^{2U}\Big(e^{4U}\omega_{\underline{i}}\omega_{\underline{j}} + \delta_{ij}\Big)\partial^2 e^{-2U} + 4e^{6U}\omega_{(\underline{i}}\epsilon_{j)kl}\partial_{\underline{k}}e^{-2U}\partial_{\underline{l}}b,
\end{aligned} \tag{F.25}$$

and the Ricci scalar is given by

$$R = \tfrac{1}{2}\frac{\partial V \partial \bar{V}}{(\mathrm{Im}\, V)^3} + \frac{\partial^2 \mathrm{Im}\, V}{(\mathrm{Im}\, V)^2}. \tag{F.26}$$

F.2 For some $d > 4$ metrics

F.2.1 $d > 4$ General static, spherically symmetric metrics

We are going to use the metric

$$ds^2 = \lambda(\rho)dt^2 - \mu^{-1}(\rho)d\rho^2 - R^2(\rho)d\Omega^2_{(d-2)}, \tag{F.27}$$

where $d\Omega^2_{(d-2)}$, implicitly defined in Eq. (C.7), is the metric on $\mathrm{S}^{(d-2)}$ (we use the ψ_i coordinates). For considering the most general higher-dimensional static, spherically symmetric metrics it would suffice to take $\mu(\rho) = \lambda(\rho)$ or $R(\rho) = \rho$. We prefer this, however, because it covers both cases and sometimes the components of the metric are simpler if we do not force R to be ρ or μ to be λ. This is the class of metrics we used for single (extreme or non-extreme) BHs (point-like objects with $d-2$ asymptotically flat directions).

The non-vanishing components of the Levi-Cività connection are

$$\begin{aligned}
&\Gamma_{tt}{}^{\rho} = \tfrac{1}{2}\mu\lambda', \qquad \Gamma_{t\rho}{}^{t} = \tfrac{1}{2}\lambda^{-1}\lambda', \qquad \Gamma_{\rho\rho}{}^{r} = -\tfrac{1}{2}\mu^{-1}\mu', \\
&\Gamma_{\rho p}{}^{r} = \delta_q{}^{r}(\ln R)', \quad \Gamma_{qr}{}^{\rho} = -\tfrac{1}{2}\delta_{qr}\mu(R^2)'q_{(r)}/R^2, \\
&\Gamma_{qr}{}^{s} = \left\{\theta_{rq}\delta^{s}_{(r)}\cot\psi_{(q)} + \theta_{qr}\delta^{s}_{(q)}\cot\psi_{(r)} - \theta_{qs}\delta_{(q)r}\cot\psi_{(s)}\, q^{-1}_{(s)}q_{(q)}\right\},
\end{aligned} \tag{F.28}$$

where

$$\theta_{rq} = \begin{cases} 1 & r > q, \\ 0 & r \leq q, \end{cases} \tag{F.29}$$

and where $q, r, s = 1, \ldots, d-2$ label the angular coordinates and, here,

$$g_{qr} = -\delta_{qr}q_{(r)}. \tag{F.30}$$

Using the Laplacian of a scalar function of ρ in this coordinate system,

$$\nabla^2 f(\rho) = -\frac{\left[(\lambda\mu)^{\frac{1}{2}} R^{d-2} f'\right]'}{(\lambda/\mu)^{\frac{1}{2}} R^{d-2}}, \tag{F.31}$$

we can write the components of the Ricci tensor in their simplest form as follows:

$$\begin{aligned}
R_{tt} &= \tfrac{1}{2}\lambda\nabla^2\ln\lambda, \qquad R_{qr} = g_{qr}\left[\nabla^2\ln R + \frac{d-3}{R^2}\right], \\
R_{\rho\rho} &= -\tfrac{1}{2}\mu^{-1}\nabla^2\ln\lambda + \frac{d-2}{R}\left(\tfrac{\lambda}{\mu}\right)^{\frac{1}{2}}\left[R'\left(\tfrac{\lambda}{\mu}\right)^{-\frac{1}{2}}\right]',
\end{aligned} \tag{F.32}$$

and the Ricci scalar is

$$R = \nabla^2 \ln\left(\lambda R^{d-2}\right) + (d-2)(d-3)\frac{1}{R^2} - \frac{d-2}{R}(\lambda\mu)^{\frac{1}{2}}\left[R'\left(\frac{\lambda}{\mu}\right)^{-\frac{1}{2}}\right]'. \tag{F.33}$$

However, if we are interested in finding singular contributions to the curvature, these formulae are not completely appropriate, because in obtaining them we have performed operations in which singular contributions are ignored. The unsimplified formulae are

$$\begin{aligned} R_{qr} = & -\frac{1}{d-2}g_{qr}\left\{\nabla^2 \ln\mu + \frac{1}{R^{d-2}}\left(\frac{\lambda}{\mu}\right)^{-\frac{1}{2}}\left[\left(\frac{\lambda}{\mu}\right)^{\frac{1}{2}}\left(R^{d-2}\mu\right)'\right]'\right. \\ & \left. - \frac{(d-2)(d-3)}{R^2}\right\}, \end{aligned} \tag{F.34}$$

$$\begin{aligned} R = & \nabla^2 \ln\left(\frac{\lambda}{\mu}\right) - \frac{1}{R^{d-2}}\left(\frac{\lambda}{\mu}\right)^{-\frac{1}{2}}\left[\left(\frac{\lambda}{\mu}\right)^{\frac{1}{2}}\left(R^{d-2}\mu\right)'\right]' \\ & - \frac{d-2}{R}(\lambda\mu)^{\frac{1}{2}}\left[R'\left(\frac{\lambda}{\mu}\right)^{\frac{1}{2}}\right]' - \frac{(d-2)(d-3)}{R^2}. \end{aligned} \tag{F.35}$$

F.2.2 A general metric for (single, black) p-branes

This metric can be understood as a generalization of the previous one with translational isometries in p dimensions and it is adequate for describing the gravitational fields of p-branes. Therefore, in general, it is not asymptotically flat in those p dimensions. It is, roughly speaking, the result of adding those p dimensions to the general, static, spherically symmetric $(d-p)$-dimensional metric of the previous section. Thus, it has the general form

$$ds^2 = \lambda(\rho)dt^2 - f(\rho)d\vec{y}_p^{\,2} - \mu^{-1}(\rho)dr^2 - R^2(\rho)d\Omega^2_{(\tilde{p}+2)}, \tag{F.36}$$

where $\vec{y}_p = (y_p^1, \ldots, y_p^p)$ are the coordinates *on the p-brane* that we denote with the indices $i, j, k = 1, \ldots, p$, $\rho^2 = (x^{p+1})^2 + \cdots + (x^{d-1})^2$ is the radial coordinate in the $(d-p-1)$-dimensional, asymptotically flat space transverse to the p-brane, the $d-p-2$ angular coordinates are labeled by $q, r, s = 1, \ldots, d-p-2$, and $\tilde{p} \equiv d-p-4$ is the dimension of the object that is the electric–magnetic dual to the p-brane.

The non-vanishing components of the Levi-Cività connection are

$$\begin{aligned} &\Gamma_{tt}{}^{\rho} = \tfrac{1}{2}\mu\lambda', && \Gamma_{t\rho}{}^{t} = \tfrac{1}{2}\lambda^{-1}\lambda', \quad \Gamma_{\rho\rho}{}^{\rho} = -\tfrac{1}{2}\mu^{-1}\mu', \\ &\Gamma_{\rho q}{}^{r} = \delta_q{}^r(\ln R)', && \Gamma_{ij}{}^{\rho} = -\tfrac{1}{2}\delta_{ij}\mu f', \quad \Gamma_{i\rho}{}^{j} = \tfrac{1}{2}\delta_i{}^j f^{-1}f', \\ &\Gamma_{qr}{}^{\rho} = -\tfrac{1}{2}\delta_{qr}\mu(R^2)'q_{(r)}/R^2, \\ &\Gamma_{qr}{}^{s} = \left\{\theta_{rq}\delta^s_{(r)}\cot\psi_{(q)} + \theta_{qr}\delta^s_{(q)}\cot\psi_{(r)} - \theta_{qs}\delta_{(q)r}\cot\psi_{(s)}\, q^{-1}_{(s)}q_{(q)}\right\}. \end{aligned} \tag{F.37}$$

The non-vanishing components of the Ricci tensor are

$$\begin{aligned} R_{tt} &= R_{tt}^{(d-p)} - \tfrac{1}{4} p\mu (\ln f)'\lambda', & R_{\rho\rho} &= R_{\rho\rho}^{(d-p)} + \tfrac{1}{2} p(\mu f)^{-\frac{1}{2}}\left[(\mu f)^{\frac{1}{2}}(\ln f)'\right]', \\ R_{ij} &= -\tfrac{1}{2}\delta_{ij} f \nabla^2 \ln f, & R_{qr} &= R_{qr}^{(d-p)} - \tfrac{1}{2} p g_{qr} \mu (\ln f)'(\ln R)', \end{aligned} \tag{F.38}$$

where we have indicated with the superscript $(d-p)$ the components of the curvature of the $(d-p)$-dimensional metric that one obtains if the p coordinates $\vec{y}_p$ are suppressed and which are given in Appendix F.2.1.

The Ricci scalar is

$$R = R^{(d-p)} + \tfrac{1}{2} p \left\{ \nabla^2_{(d-p)} \ln f + f^{-1}\nabla^2 f + \mu\left[(\ln f)'\right]^2 \right\}. \tag{F.39}$$

F.2.3 A general metric for (composite, black) p-branes

This metric is a generalization of the general higher-dimensional, static, spherically symmetric metric with translational isometries in $\sum_{n=1}^{N} r_n$ dimensions that split into N blocks. The difference from the metric of Section F.2.2 is that the previous one also had spherical symmetry $SO(p)$ in the p directions associated with the isometries, but in the present metric the spherical symmetry is split into N groups and is, thus, $\prod_{n=1}^{N} SO(r_n)$. This metric is adequate for describing the gravitational field of composite (intersecting etc.) p-branes. It has the general form

$$ds^2 = \lambda(\rho)dt^2 - \sum_{n=1}^{N} f_n(\rho) d\vec{y}_n^{\,2} - \mu^{-1}(\rho)dr^2 - R^2(r)d\Omega^2_{(\delta-2)}, \tag{F.40}$$

where $\vec{y}_n = (y_n^1, \ldots, y_n^{r_n})$ are the coordinates of the nth "block" that we denote with the indices $i_n, j_n, k_n = 1, \ldots, r_n$, $\rho^2 = (x^{p+1})^2 + \cdots + (x^{d-1})^2$ is the radial coordinate in the $(d - \sum_n r_n - 1)$-dimensional, asymptotically flat space transverse to the p-branes, the $\delta - 2$ angular coordinates are labeled by $q, r, s = 1, \ldots, \delta - 2$, and δ is defined by

$$\delta = d - \sum_{n=1}^{N} r_n. \tag{F.41}$$

The non-vanishing components of the Levi-Cività connection are

$$\begin{aligned} \Gamma_{tt}{}^{\rho} &= \tfrac{1}{2}\mu\lambda', & \Gamma_{t\rho}{}^{t} &= \tfrac{1}{2}\lambda^{-1}\lambda', \\ \Gamma_{\rho\rho}{}^{r} &= -\tfrac{1}{2}\mu^{-1}\mu', & \Gamma_{\rho p}{}^{r} &= \delta_q{}^r (\ln R)', \\ \Gamma_{i_n j_m}{}^{\rho} &= -\tfrac{1}{2}\delta_{ij}\delta_{nm}\mu f'_{(n)}, & \Gamma_{i_n \rho}{}^{j_m} &= \tfrac{1}{2}\delta_i{}^j \delta_n{}^m f^{-1} f'_{(n)}, \\ \Gamma_{qr}{}^{\rho} &= -\tfrac{1}{2}\delta_{qr}\mu (R^2)' q_{(r)}/R^2, \\ \Gamma_{qr}{}^{s} &= \left\{ \theta_{rq}\delta^s_{(r)} \cot\psi_{(q)} + \theta_{qr}\delta^s_{(q)} \cot\psi_{(r)} - \theta_{qs}\delta_{(q)r} \cot\psi_{(s)}\, q^{-1}_{(s)} q_{(q)} \right\}. \end{aligned} \tag{F.42}$$

The non-vanishing components of the Ricci tensor are

$$\begin{aligned}
R_{tt} &= R_{tt}^{(\delta)} - \tfrac{1}{4}\mu \textstyle\sum_{n=1}^{N} r_n (\ln f_n)'\lambda', \\
R_{i_n j_m} &= -\tfrac{1}{2}\delta_{ij}\delta_{nm}\left\{\nabla^2 f_n + \mu f_{(n)}\left[(\ln f_n)'\right]^2\right\}, \\
R_{\rho\rho} &= R_{\rho\rho}^{(\delta)} + \tfrac{1}{2}\textstyle\sum_{n=1}^{N} r_n (\mu f_{(n)})^{-\frac{1}{2}}\left[(\mu f_{(n)})^{\frac{1}{2}}\left(\ln f_{(n)}\right)'\right]', \\
R_{qr} &= R_{qr}^{(\delta)} - \tfrac{1}{2}g_{qr}\mu(\ln R)'\sum_{n=1}^{N} r_n (\ln f_n)',
\end{aligned} \tag{F.43}$$

where we have indicated with the superscript (δ) the components of the curvature of the δ-dimensional metric that one obtains if the coordinates $\vec{y}_n$ are suppressed.

The Ricci scalar is

$$R = R^{(\delta)} + \tfrac{1}{2}\sum_{n=1}^{N} r_n \left\{\nabla^2_{(\delta)} \ln f_{(n)} + f_{(n)}^{-1}\nabla^2 f_{(n)} + \mu\left[(\ln f_{(n)})'\right]^2\right\}. \tag{F.44}$$

F.2.4 A general metric for extreme p-branes

d-dimensional metrics of the general form

$$ds^2 = H^{2x}\eta_{ij}dy^{\underline{i}}dy^{\underline{j}} + H^{-2y}\eta_{mn}dx^{\underline{m}}dx^{\underline{n}}, \tag{F.45}$$

where $i, j = 0, 1, \ldots, p$ and $m, n = p+1, \ldots, d-1$ and H is a function solely of the $x^{\underline{m}}$s often occur in the study of p-branes. The coordinates $y^{\underline{i}}$ correspond to the p-brane worldvolume and the coordinates $x^{\underline{m}}$ are transverse to the p-brane. Observe that, with our conventions, $\eta_{mn} = -\delta_{mn}$. The non-vanishing components of the Levi-Cività connection are

$$\begin{aligned}
&\Gamma_{\underline{i}\underline{j}}{}^{\underline{m}} = x\eta_{ij}H^{2(x+y)-1}\partial_{\underline{m}}H, \qquad \Gamma_{\underline{i}\underline{m}}{}^{\underline{j}} = x\delta_i{}^j H^{-1}\partial_{\underline{m}}H, \\
&\Gamma_{\underline{m}\underline{n}}{}^{\underline{p}} = -yH^{-1}\left\{\delta_{pm}\partial_{\underline{n}}H + \delta_{pn}\partial_{\underline{m}}H - \delta_{mn}\partial_{\underline{p}}H\right\}.
\end{aligned} \tag{F.46}$$

The non-vanishing components of the Ricci tensor are

$$\begin{aligned}
R_{\underline{i}\underline{j}} &= g_{\underline{i}\underline{j}}\nabla^2 \ln H^x, \\
R_{\underline{m}\underline{n}} &= g_{\underline{m}\underline{n}}\nabla^2 \ln H^{-y} + zH^{-1}\partial_{\underline{m}}\partial_{\underline{n}}H \\
&\quad + H^{-2}\partial_{\underline{m}}H\partial_{\underline{n}}H\left\{x^2(p+1) + y^2(\tilde{p}+1) + (2y-1)z\right\},
\end{aligned} \tag{F.47}$$

where we have used the fact that

$$\sqrt{|g|} = H^{z-2y}, \qquad z = x(p+1) - y(\tilde{p}+1), \tag{F.48}$$

and the fact that, for a scalar function of the $x^{\underline{m}}$s in this metric,

$$\begin{aligned}
\nabla^2 f(x^m) &= -H^{2y-1}\left[z\partial_{\underline{m}}H\partial_{\underline{m}}f + H\partial^2 f\right], \qquad \partial^2 \equiv +\partial_{\underline{m}}\partial_{\underline{m}}, \\
\nabla^2 \ln H &= (1-z)H^{2y-2}(\partial H)^2 - H^{2y-1}\partial^2 H.
\end{aligned} \tag{F.49}$$

The Ricci scalar is

$$R = \nabla^2 \ln H^{2(z-y)} - H^{2(y-1)}(\partial H)^2\big[x^2(p+1) + y^2(\tilde{p}+1) - z(z-2y)\big]. \tag{F.50}$$

The simplest choice of Vielbein is

$$e_{\underline{i}}{}^{j} = \delta_i{}^{j} H^{x}, \qquad e_{\underline{m}}{}^{n} = \delta_m{}^{n} H^{-y}, \tag{F.51}$$

and it gives the following non-vanishing components of the spin connection:

$$\omega_{\underline{m}}{}^{np} = 2yH^{-1}\partial_{\underline{q}} H \eta_m{}^{[n}\eta^{p]q}, \quad \omega_{\underline{i}}{}^{mj} = -2xH^{(x+y)-1}\partial_{\underline{q}} H \eta_i{}^{[m}\eta^{j]q}. \tag{F.52}$$

F.2.5 Brinkmann metrics

The metric of any spacetime admitting a covariantly constant null Killing vector can always be put into the Brinkmann-metric form Eq. (10.4), that we rewrite here for convenience:

$$ds^2 = 2du(dv + Kdu + A_{\underline{i}}dx^i) + \tilde{g}_{\underline{i}\underline{j}}dx^i dx^j, \tag{F.53}$$

where all the functions in the metric are independent of v. Either K or $A_{\underline{i}}$ can be removed by a coordinate transformation that preserves the above form of the metric, but here we work with the most general form.

Using also light-cone coordinates in tangent space, a natural Vielbein basis is

$$\begin{aligned} &e^u = du, && e^v = dv + Kdu + A_{\underline{i}}dx^i, && e^i = \tilde{e}_{\underline{j}}{}^{i} dx^j, \\ &e_u = \partial_{\underline{u}} - K\partial_{\underline{v}}, && e_v = \partial_{\underline{v}}, && e_i = \tilde{e}_i{}^{\underline{j}}\Big[\partial_{\underline{j}} - A_{\underline{j}}\partial_{\underline{v}}\Big], \end{aligned} \tag{F.54}$$

where the $\tilde{e}^i{}_{\underline{j}}$ are Vielbeins in the $(d-2)$-dimensional wavefront space. The associated components of the spin connection are

$$\begin{aligned} &\omega_{uiu} = \tilde{e}_i{}^{\underline{j}}\Big[\partial_{\underline{j}}K - \partial_{\underline{u}}A_{\underline{j}}\Big], && \omega_{uij} = \tfrac{1}{2}\tilde{F}_{ij} - \tilde{e}_{[i|}{}^{\underline{k}}\partial_{\underline{u}}\tilde{e}_{a|\underline{k}]}, \\ &\omega_{ijk} = \tilde{\omega}_{ijk}, && \omega_{iju} = -\tfrac{1}{2}\tilde{F}_{ij} - \tilde{e}_{(i|}{}^{\underline{k}}\partial_{\underline{u}}\tilde{e}_{\underline{k}|j)}, \end{aligned} \tag{F.55}$$

where $F_{\underline{i}\underline{j}} = 2\partial_{[\underline{i}}A_{\underline{j}]}$.

The components of the Ricci tensor are

$$\begin{aligned} R_{\underline{i}\underline{j}} &= \tilde{R}_{\underline{i}\underline{j}}, \\ 2R_{\underline{i}\underline{u}} &= \tilde{\nabla}^{\underline{j}}F_{\underline{j}\underline{i}} + \tilde{\nabla}_{\underline{i}}\Big(\tilde{g}^{\underline{j}\underline{k}}\partial_{\underline{u}}\tilde{g}_{\underline{j}\underline{k}}\Big) - \tilde{\nabla}_{\underline{j}}\big(\tilde{g}^{\underline{j}\underline{k}}\partial_{\underline{u}}\tilde{g}_{\underline{k}\underline{i}}\big), \\ R_{\underline{u}\underline{u}} &= \tilde{\nabla}_{\underline{i}}\partial^{\underline{i}}K - \tfrac{1}{4}\tilde{F}^2 + \tfrac{1}{2}\tilde{g}^{\underline{i}\underline{j}}\partial^2_{\underline{u}}\tilde{g}_{\underline{i}\underline{j}} + \tfrac{1}{4}\partial_{\underline{u}}\tilde{g}^{\underline{i}\underline{j}}\partial_{\underline{u}}\tilde{g}_{\underline{i}\underline{j}} - \tilde{g}^{\underline{i}\underline{j}}\tilde{\nabla}_{\underline{i}}\Big(\partial_{\underline{u}}A_{\underline{j}}\Big), \end{aligned} \tag{F.56}$$

where all the objects with tildes are calculated from the metric $\tilde{g}_{\underline{i}\underline{j}}$, treating u as some constant. The Ricci scalar is just $R = \tilde{R}$.

Appendix G
The harmonic operator on $\mathbb{R}^3 \times \mathrm{S}^1$

This section is based on [476].

We want to relate the solutions of the Laplace equation in $\mathbb{R}^3 \times \mathrm{S}^1$ and in $\mathbb{R}^3$. We denote the corresponding Laplacians by $\Delta_{(4)}$ and $\Delta_{(3)}$ and we have

$$\Delta_{(4)} = \Delta_{(3)} + \partial_z^2. \tag{G.1}$$

The Laplacian, being a local operator, has the same form in $\mathbb{R}^3 \times \mathrm{S}^1$ and in $\mathbb{R}^4$. Clearly, the difference is in the periodicity conditions that the solutions must satisfy in the first case. This observation will help us to construct them starting with harmonic functions on $\mathbb{R}^4$.

The solution of the Laplace equation (more precisely, it is the Green function of the Laplacian) in $\mathbb{R}^4$ is $1/|\vec{x}_4 - \vec{x}_{4(0)}|^2$, where $\vec{x}_4 = (x^1, x^2, x^3, x^4)$. In particular, it satisfies

$$\Delta_{(4)} \frac{1}{|\vec{x}_4 - \vec{x}_{4(0)}|^2} = -4\pi^2 \delta^{(4)}(\vec{x}_4 - \vec{x}_{4(0)}). \tag{G.2}$$

This harmonic function has a singularity at $\vec{x}_4 = \vec{x}_{4(0)}$. We are in general interested in harmonic functions that go to 1 at infinity and with a different coefficient for the pole (h):

$$H_{\mathbb{R}^4} = 1 + \frac{h}{|\vec{x}_4 - \vec{x}_{4(0)}|^2}. \tag{G.3}$$

Since the Laplacian is linear, we can combine linearly harmonic functions to construct one with singularities placed at regular intervals on the x^4 axis. The resulting harmonic function will have the periodicity required for it to be a harmonic function on $\mathbb{R}^3 \times \mathrm{S}^1$. More explicitly,

$$H_{\mathbb{R}^3 \times \mathrm{S}^1} = 1 + \sum_{n \in \mathbb{Z}} \frac{h}{|\vec{x}_3 - \vec{x}_{3(0)}|^2 + (z - z_{(0)} - 2\pi n \ell)^2}. \tag{G.4}$$

This series can be summed:

$$H_{\mathbb{R}^3 \times \mathrm{S}^1} = 1 + \frac{h}{2\ell |\vec{x}_3 - \vec{x}_{3(0)}|^2} \frac{\sinh |\vec{x}_3 - \vec{x}_{3(0)}|/\ell}{\cosh |\vec{x}_3 - \vec{x}_{3(0)}|/\ell - \cos(z - z_{(0)})/\ell}. \tag{G.5}$$

In this form it is evident that we have a function with the right periodicity and one can also immediately check that it is a solution of the Laplace equation. On the other hand, near the singularity $\vec{x}_3 \to \vec{x}_{3(0)}$, $z \to z_{(0)}$, or in the equivalent limit $\ell \to 0$ in which the periodicity of the fourth coordinate is irrelevant, $H_{\mathbb{R}^3\times\mathrm{S}^1}$ becomes exactly $H_{\mathbb{R}^4}$ plus subdominant terms.

Now we want to expand in Fourier series the periodic harmonic function $H_{\mathbb{R}^3\times\mathrm{S}^1}$:

$$H_{\mathbb{R}^3\times\mathrm{S}^1} = \sum_{n\in\mathbb{Z}} H_{\mathbb{R}^3\times\mathrm{S}^1,n}(\vec{x}_3 - \vec{x}_{3(0)}) e^{\frac{inz}{\ell}}. \tag{G.6}$$

The Fourier modes are

$$H_{\mathbb{R}^3\times\mathrm{S}^1,n}(\vec{x}_3 - \vec{x}_{3(0)}) = \delta_{n,0} + \frac{h/(2\ell)}{|\vec{x}_3 - \vec{x}_{3(0)}|} e^{-\frac{|n|(|\vec{x}_3 - \vec{x}_{3(0)}| - inz_{(0)})}{\ell}}. \tag{G.7}$$

If we consider only the zero mode, we find

$$H_{\mathbb{R}^3\times\mathrm{S}^1,0} = 1 + \frac{h/(2\ell)}{|\vec{x}_3 - \vec{x}_{3(0)}|}, \tag{G.8}$$

which is a harmonic function on $\mathbb{R}^3$, satisfying

$$\Delta_{(3)} H_{\mathbb{R}^3\times\mathrm{S}^1,0} = -\frac{2\pi h}{\ell} \delta^{(3)}(\vec{x}_3 - \vec{x}_{3(0)}). \tag{G.9}$$

Using

$$\Delta_{(3)} \frac{1}{|\vec{x}_3 - \vec{x}_{3(0)}|} = -4\pi \delta^{(3)}(\vec{x}_3 - \vec{x}_{3(0)}), \tag{G.10}$$

it is easy to see that the higher (KK) modes satisfy the massive three-dimensional Laplace equation

$$\left[\Delta_{(3)} - \frac{|n|^2}{\ell^2}\right] H_{\mathbb{R}^3\times\mathrm{S}^1,n} = -\frac{2\pi h}{\ell} \delta^{(3)}(\vec{x}_3 - \vec{x}_{3(0)}) e^{\frac{inz_{(0)}}{\ell}}. \tag{G.11}$$

References

[1] L. F. Abbott and S. Deser, *Nucl. Phys.* **B195** (1982) 76–96.

[2] L. F. Abbott and S. Deser, *Phys. Lett.* **116B** (1982) 259.

[3] M. Abou Zeid and C. M. Hull, *Phys. Lett.* **B428** (1998) 277.

[4] M. Abraham, *Lincei Atti* **20** (1911) 678.

[5] M. Abraham, *Phys. Z.* **13** (1912) 1.

[6] M. Abraham, *Phys. Z.* **13** (1912) 4.

[7] M. Abraham, *Phys. Z.* **13** (1912) 176.

[8] M. Abraham, *Phys. Z.* **13** (1912) 310.

[9] M. Abraham, *Phys. Z.* **13** (1912) 311.

[10] M. Abraham, *Phys. Z.* **13** (1912) 793.

[11] M. Abraham, *Nuovo Cim.* **4** (1912) 459.

[12] M. Abraham, *Jahrbuch der Radioaktivität und Elektronik* **11** (1914) 470–520.

[13] A. Achúcarro, J. Evans, P. K. Townsend, and D. Wiltshire, *Phys. Lett.* **B198** (1987) 441.

[14] A. Achúcarro, P. Kapusta, and K. S. Stelle, *Phys. Lett.* **B232** (1989) 302.

[15] M. Aganagic, J. Park, C. Popescu, and J. H. Schwarz, *Nucl. Phys.* **B496** (1997) 191.

[16] M. Aganagic, C. Popescu, and J. H. Schwarz, *Phys. Lett.* **B393** (1997) 311.

[17] M. Aganagic, C. Popescu, and J. H. Schwarz, *Nucl. Phys.* **B495** (1997) 99.

[18] A. G. Agnese and M. La Camera, *Phys. Rev.* **D31** (1985) 1280–1286.

[19] A. G. Agnese and M. La Camera, *Phys. Rev.* **D49** (1994) 2126–2128.

[20] Y. Aharonov and D. Bohm, *Phys. Rev.* **115** (1959) 485.

[21] Y. Aharonov and D. Bohm, *Phys. Rev.* **D3** (1961) 1511.

[22] O. Aharony, *Nucl. Phys.* **B476** (1996) 470–483.

[23] O. Aharony, S. S. Gubser, J. Maldacena, H. Ooguri, and Y. Oz, *Phys. Rep.* **323** (2000) 183.

[24] P. Aichelburg and R. Sexl, *Gen. Relativ. Gravit.* **2** (1971) 303.

[25] N. Alonso-Alberca, E. Lozano-Tellechea, and T. Ortín, *Class. Quant. Grav.* **19** (2002) 6009–6024.

[26] N. Alonso-Alberca, E. Lozano-Tellechea, and T. Ortín, *Class. Quant. Grav.* **20** (2003) 423–430.

[27] N. Alonso-Alberca, P. Meessen, and T. Ortín, *Class. Quant. Grav.* **17** (2000) 2783–2798.

[28] N. Alonso-Alberca, P. Meessen, and T. Ortín, *Nucl. Phys.* **B602** (2001) 329–345.

[29] N. Alonso-Alberca and T. Ortín, *Nucl. Phys.* **B651** (2003) 263.

[30] E. Álvarez, *Rev. Mod. Phys.* **61** (1989) 561–604.

[31] E. Álvarez, *Nucl. Phys.* **B269** (1986) 596.

[32] E. Álvarez, L. Álvarez-Gaumé, and I. Bakas, *Nucl. Phys.* **B457** (1995) 3–26.

[33] E. Álvarez and M. Á. Osorio, *Phys. Rev.* **D40** (1989) 1150.

[34] E. Álvarez, L. Álvarez-Gaumé, and I. Bakas, *Nucl. Phys.* **B457** (1995) 3–26.

[35] E. Álvarez, L. Álvarez-Gaumé, and Y. Lozano, *Nucl. Phys. Proc. Suppl.* **41** (1995) 1–20.

[36] E. Álvarez, J. L. F. Barbón, and J. Borlaf, *Nucl. Phys.* **B479** (1996) 218.

[37] E. Álvarez and P. Meessen, *JHEP* **9902** (1999) 015.

[38] E. Álvarez and T. Ortín, *Mod. Phys. Lett.* **A7** (1992) 2889–2894.

[39] L. Álvarez-Gaumé and M. A. Vázquez-Mozo, *Les Houches Summer School on Gravitation and Quantizations*, Session 57, Amsterdam: North-Holland, (1995).

[40] S. P. de Alwis, *Phys. Rev.* **D56** (1997) 7963–7977.

[41] D. Amati and C. Klimcik, *Phys. Lett.* **B219** (1989) 443.

[42] S. Åminneborg, I. Bengtsson, S. Holst, and P. Peldán, *Class. Quant. Grav.* **13** (1996) 2707–2714.

[43] L. Andrianopoli and R. D'Auria, `hep-th/0203076`.

[44] C. Angelantonj and A. Sagnotti, *Phys. Rep.* **371** (2002) 1.

[45] T. Applequist, A. Chodos, and P. G. O. Freund, *Modern Kaluza–Klein Theories*, Menlo Park, California: Addison-Wesley (1987).

[46] C. Aragone and S. Deser, *Nuovo Cim.* **3A** (1971) 709.

[47] C. Aragone and S. Deser, *Nuovo Cim.* **57B** (1980) 33.

[48] I.Y. Aref'eva, M. G. Ivanov, O. A. Rytchkov, and I. V. Volovich, *Class. Quant. Grav.* **15** (1998) 2923.

[49] I. Y. Aref'eva, M. G. Ivanov, and I. V. Volovich, *Phys. Lett.* **B406** (1997) 44.

[50] I. Ya. Aref'eva and A. Volovich, *Class. Quant. Grav.* **14** (1997) 2991–3000.

[51] I. Ya. Aref'eva, K. S. Viswanathan, and I. V. Volovich, *Phys. Rev.* **D55** (1997) 4748–4755.

[52] R. Argurio, F. Englert, and L. Houart, *Phys. Lett.* **B398** (1997) 61.

[53] R. Argurio, F. Englert, L. Houart, and P. Windey, *Phys. Lett.* **B408** (1997) 151.

[54] R. Arnowitt, S. Deser, and C. Misner, in *Gravitation: An Introduction to Current Research*, Ed. L. Witten, New York: Wiley (1962).

[55] A. Ashtekar, *Lectures on Non-perturbative Quantum Gravity* Singapore: World Scientific (1991).

[56] M. F. Atiyah and N. J. Hitchin, *Phys. Lett.* **107A** (1985) 21–25. (Reprinted in [439].)

[57] R. D'Auria and P. Fré, in *Classical and Quantum Black Holes*, Eds. P. Fré, V. Gorini, G. Magli, and U. Moschella, Bristol: Insitute of Physics Publishing (1999), pp. 137–272.

[58] M. A. Awada, M. J. Duff, and C. N. Pope, *Phys. Rev. Lett.* **50** (1983) 294.

[59] J. A. de Azcárraga, J. P. Gauntlett, J. M. Izquierdo, and P. K. Townsend, *Phys. Rev. Lett.* **63** (1989) 2443.

[60] J. A. de Azcárraga and J. Lukierski, *Phys. Lett.* **B113** (1982) 170.

[61] J. A. de Azcárraga and J. Lukierski, *Phys. Rev.* **D28** (1983) 1337.

[62] B. E. Baaquie and L. C. Kwek, *Int. J. Mod. Phys.* **A16** (2001) 2605.

[63] S. V. Babak and L. P. Grishchuk, *Phys. Rev.* **D61** (2000) 024038.

[64] C. Bachas, in *Cambridge 1997, Duality and Supersymmetric Theories*, Eds. D. I. Olive and P. C. West, Cambridge: Cambridge University Press (1999), pp. 414–473.

[65] C. P. Bachas, M. R. Douglas, and M. B. Green, *JHEP* **9707** (1997) 002.

[66] C. P. Bachas, M. B. Green, and A. Schwimmer, *JHEP* **9801** (1998) 006.

[67] I. Bakas, *Phys. Rev.* **D54** (1996) 6424–6434.

[68] I. Bakas, *Phys. Lett.* **B343** (1995) 103–112.

[69] I. Bakas and K. Sfetsos, *Phys. Lett.* **B349** (1995) 448–457.

[70] A. P. Balachandran, P. Salomonson, B. S. Skagerstam, and J. O. Winnberg, *Phys. Rev.* **D15** (1977) 2308.

[71] H. Balasin and H. Nachbagauer, *Class. Quant. Grav.* **10** (1993) 2271.

[72] H. Balasin and H. Nachbagauer, *Class. Quant. Grav.* **11** (1994) 1453–1462.

[73] H. Balasin and H. Nachbagauer, *Class. Quant. Grav.* **12** (1995) 707–713.

[74] H. Balasin and H. Nachbagauer, *Class. Quant. Grav.* **13** (1996) 731–737.

[75] V. Balasubramanian, in *Cargese 1997, Strings, Branes and Dualities,* Eds. L. Baulieu, P. Di Francesco, M. Douglas, V. Kazakov, M. Picco, and P. Windey, Dordrecht: Kluwer (1999), pp. 399–410.

[76] V. Balasubramanian and F. Larsen, *Nucl. Phys.* **B478** (1996) 199.

[77] V. Balasubramanian, F. Larsen, and R. G. Leigh, *Phys. Rev.* **D57** (1998) 3509.

[78] I. A. Bandos, J. A. de Azcárraga, J. M. Izquierdo, and J. Lukierski, *Phys. Rev.* **D65** (2002) 021901.

[79] I. A. Bandos, J. A. de Azcárraga, J. M. Izquierdo, and J. Lukierski, Invited talk at 3rd International Sakharov Conference on Physics, Moscow, Russia, 24–29 June 2002, `hep-th/0211065`.

[80] I. Bandos, K. Lechner, A. Nurmagambetov, P. Pasti, D. P. Sorokin, and M. Tonin, *Phys. Rev. Lett.* **78** (1997) 4332.

[81] I. A. Bandos, D. P. Sorokin, and M. Tonin, *Nucl. Phys.* **B497**, (1997) 275.

[82] M. Bañados, C. Teitelboim, and J. Zanelli, *Phys. Rev. Lett.* **69** (1992) 1849–1851.

[83] A. Barajas, *Proc. Nat. Acad. Sci.* **30** (1944) 54.

[84] A. Barajas, G. D. Birkhoff, C. Graef, and M. Sandoval Vallarta, *Phys. Rev.* **66** (1944) 138–143.

[85] J. M. Bardeen, B. Carter, and S. W. Hawking, *Commun. Math. Phys.* **31** (1973) 161–170.

[86] A. Barducci, R. Casalbuoni, and L. Lusanna, *Nuovo Cim.* **A35** (1976) 377.

[87] A. Barducci, R. Casalbuoni, and L. Lusanna, *Nucl. Phys.* **B124** (1977) 93.

[88] A. Barducci, R. Casalbuoni, and L. Lusanna, *Nucl. Phys.* **B124** (1977) 521.

[89] A. O. Barut and R. Raczka, *Theory of Group Representations and Applications*, Singapore: World Scientific (1986).

[90] H. Bauer, *Phys. Z.* **19** (1918) 163.

[91] K. Bautier, S. Deser, M. Henneaux, and D. Seminara, *Phys. Lett.* **B406** (1997) 49.

[92] K. Behrndt and E. Bergshoeff, *Phys. Lett.* **B383** (1996) 383–389.

[93] K. Behrndt, contribution to *Strings '96*, `hep-th/9611237`.

[94] R. Beig and W. Simon, *Commun. Math. Phys.* **144** (1992) 373.

[95] J. Bekenstein, Ph.D. Thesis, Princeton University (unpublished).

[96] J. D. Bekenstein, *Phys. Rev.* **D5** (1972) 1239–1246.

[97] J. D. Bekenstein, *Lett. Nuovo Cim.* **4** (1972) 737.

[98] J. D. Bekenstein, *Phys. Rev.* **D9** (1973) 2333.

[99] J. D. Bekenstein, *Phys. Rev.* **D9** (1974) 3292.

[100] J. D. Bekenstein *Phys. Rev.* **D51** (1995) R6608–R6611.

[101] J. D. Bekenstein, in *2nd International Sakharov Conference on Physics*, Eds. I. M. Dremin and A. M. Semikhatov, Singapore: World Scientific (1997), pp. 216–219.

[102] A. A. Belavin, A. M. Polyakov, A. S. Schwarz, and Yu. S. Tyupkin, *Phys. Lett.* **59B** (1975) 85–87.

[103] F. J. Belinfante, *Physica* **VI** (1939) 887.

[104] V. A. Belinskii, G. W. Gibbons, D. N. Page, and C. N. Pope, *Phys. Lett.* **B76** (1978) 433–435.

[105] F. A. Berezin and M. S. Marinov, *JETP Lett.* **21** (1975) 320.

[106] F. A. Berezin and M. S. Marinov, *Ann. Phys.* **104** (1977) 336.

[107] O. Bergman, M. R. Gaberdiel, and G. Lifschytz, *Nucl. Phys.* **B509** (1998) 194.

[108] O. Bergmann, *Am. J. Phys.* **24** (1956) 38.

[109] P. G. Bergmann, *Introduction to the Theory of Relativity*, New York: Prentice-Hall (1942) and New York: Dover (1976).

[110] P. G. Bergmann and R. Schiller, *Phys. Rev.* **89** (1953) 4.

[111] E. Bergshoeff, H.-J. Boonstra, and T. Ortín, *Phys. Rev.* **D53** 7206–7212.

[112] E. Bergshoeff, P. M. Cowdall, and P. K. Townsend, *Phys. Lett.* **410B** (1997) 13.

[113] E. Bergshoeff and M. De Roo, *Phys. Lett.* **B380** (1996) 265.

[114] E. Bergshoeff, M. de Roo, B. de Wit, and P. van Nieuwenhuizen, *Nucl. Phys.* **B195** (1982) 97.

[115] E. Bergshoeff, M. de Roo, E. Eyras, B. Janssen, and J. P. van der Schaar, *Nucl. Phys.* **B494** (1997) 119.

[116] E. Bergshoeff, M. de Roo, E. Eyras, B. Janssen, and J. P. van der Schaar, *Class. Quant. Grav.* **14** (1997) 2757.

[117] E. Bergshoeff, M. de Roo, B. Janssen, and T. Ortín, *Nucl. Phys.* **B550** (1999) 289–302.

[118] E. Bergshoeff, M. de Roo, M. B. Green, G. Papadopoulos, and P. K. Townsend, *Nucl. Phys.* **B470** (1996) 113–135.

[119] E. Bergshoeff, M. de Roo, and T. Ortín, *Phys. Lett.* **B386** (1996) 85–90.

[120] E. Bergshoeff, T. de Wit, U. Gran, R. Linares, and D. Roest, *JHEP* **10** (2002) 61.

[121] E. Bergshoeff, I. Entrop, and R. Kallosh, *Phys. Rev.* **D49** (1994) 6663.

[122] E. Bergshoeff, E. Eyras, and Y. Lozano, *Phys. Lett.* **B430** (1998) 77–86.

[123] E. Bergshoeff, E. Eyras, R. Halbersma, J. P. van der Schaar, C. M. Hull, and Y. Lozano, *Nucl. Phys.* **B564** (2000) 29–59.

[124] E. Bergshoeff, J. Gomis, and P. K. Townsend, *Phys. Lett.* **B421** (1998) 109.

[125] E. Bergshoeff, C. M. Hull, and T. Ortín, *Nucl. Phys.* **B451** (1995) 547–578.

[126] E. Bergshoeff, B. Janssen, and T. Ortín, *Class. Quant. Grav.* **13** (1996) 321–343.

[127] E. Bergshoeff, B. Janssen, and T. Ortín, *Phys. Lett.* **B410** (1997) 131–141.

[128] E. Bergshoeff, R. Kallosh, T. Ortín, and G. Papadopoulos, *Nucl. Phys.* **B502** (1997) 149.

[129] E. Bergshoeff, R. Kallosh, and T. Ortín, *Phys. Rev.* **D47** (1993) 5444.

[130] E. Bergshoeff, R. Kallosh, and T. Ortín, *Phys. Rev.* **D50**, (1994) 5188–5192.

[131] E. Bergshoeff, R. Kallosh, and T. Ortín, *Phys. Rev.* **D51**, (1995) 3009–3016.

[132] E. Bergshoeff, R. Kallosh, and T. Ortín, *Nucl. Phys.* **B478** (1996) 156–180.

[133] E. Bergshoeff, R. Kallosh, T. Ortín, D. Roest, and A. Van Proeyen, *Class. Quant. Grav.* **18** (2001) 3359.

[134] E. Bergshoeff, I. G. Koh, and E. Sezgin, *Phys. Lett.* **B155** (1985) 71.

[135] E. Bergshoeff, L. A. J. London, and P. K. Townsend, *Class. Quant. Grav.* **9** (1992) 2545.

[136] E. Bergshoeff, Y. Lozano, and T. Ortín, *Nucl. Phys.* **B518** (1998) 363.

[137] E. Bergshoeff, E. Sezgin, and P. K. Townsend, *Phys. Lett.* **B169** (1986) 191.

[138] E. Bergshoeff, E. Sezgin, and P. K. Townsend, *Phys. Lett.* **189B** (1987) 75.

[139] E. Bergshoeff, E. Sezgin, and P. K. Townsend, *Ann. Phys.* **185** (1988) 330.

[140] E. Bergshoeff and P. K. Townsend, *Nucl. Phys.* **B490** (1997) 145.

[141] E. Bergshoeff and P. K. Townsend, *Nucl. Phys.* **B531** (1998) 226.

[142] E. Bergshoeff and J. P. van der Schaar, *Class. Quant. Grav.* **16** (1999) 23.

[143] M. Berkooz, M. R. Douglas, and R. G. Leigh, *Nucl. Phys.* **B480** (1996) 265.

[144] M. Bertolini, P. Fré, and M. Trigiante, *Class. Quant. Grav.* **16** (1999) 2987.

[145] M. Bertolini and M. Trigiante, *Nucl. Phys.* **B582** (2000) 393.

[146] B. Bertotti, *Phys. Rev.* **116** (1959) 1331.

[147] M. Bianchi and A. Sagnotti, *Phys. Lett.* **B247** (1990) 517.

[148] M. Bianchi and A. Sagnotti, *Nucl. Phys.* **B361** (1991) 519.

[149] J. Bičák, *Selected Solutions of Einstein's Field Equations: Their Role in General Relativity and Astrophysics*, Heidelberg: Springer-Verlag (2000).

[150] A. Bilal, *Introduction to Supersymmetry*, Lectures given at the summer school *Gif 2000*, `hep-th/0101055`.

[151] G. D. Birkhoff, *Proc. Nat. Acad. Sci.* **29** (1943) 231–239. (Reprinted in [154].)

[152] G. D. Birkhoff, *Proc. Nat. Acad. Sci.* **30** (1944) 324. (Reprinted in [154].)

[153] G. D. Birkhoff, *Bol. Soc. Mat. mexicana* **1** (No. 4, 5) (1944) 1. (Reprinted in [154].)

[154] G. D. Birkhoff, *Collected Mathematical Papers*, 3 volumes, New York: American Mathematical Society (1950).

[155] G. D. Birkhoff, *Relativity and Modern Physics*, Cambridge, Massachussets: Harvard University Press (1923).

[156] D. Birmingham, *Topological Black Holes in Anti-de Sitter Space*, Report `hep-th/9808032.`

[157] N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space*, Cambridge: Cambridge University Press (1989).

[158] M. Blau, J. M. Figueroa-O'Farrill, and G. Papadopoulos, *Class. Quant. Grav.* **19** (2002) 4753.

[159] M. Blau, J. M. Figueroa-O'Farrill, C. M. Hull, and G. Papadopoulos, *JHEP* **0201** (2002) 047.

[160] M. Blau, J. M. Figueroa-O'Farrill, C. M. Hull, and G. Papadopoulos, *Class. Quant. Grav.* **19** (2002) L87.

[161] M. Blau and M. O'Loughlin, *Nucl. Phys.* **B525** (1998) 182.

[162] M. Blau, W. Thirring, and G. Landi in *25th Schladmig Conference*, *Concepts and Trends in Particle Physics*, Eds. H. Latal and H. Mitter, Berlin: Springer-Verlag (1987), p. 1.

[163] E. Bogomol'nyi, *Sov. J. Nucl. Phys.* **24** (1976) 449.

[164] L. Bombelli, R. K. Koul, G. Kunstatter, J. Lee, and R. D. Sorkin, *Nucl. Phys.* **B299** (1987) 735–756.

[165] W. B. Bonnor, *Proc. Phys. Soc. London* **A67** (1954) 225.

[166] W. B. Bonnor, *Z. Phys.* **190** (1966) 444.

[167] W. B. Bonnor, *Proc. Camb. Phil. Soc.* **66** (1969) 145.

[168] W. B. Bonnor, *Gen. Relativ. Gravit.* **24** (1992) 551.

[169] W. B. Bonnor, J. B. Griffiths, and M. A. H. MacCallum, *Gen. Relativ. Gravit.* **26** (1994) 687.

[170] H. J. Boonstra, K. Skenderis, and P. K. Townsend, *JHEP* **9901** (1999) 003.

[171] M. Born, *Proc. Roy. Soc. London* **A143** (1934) 410.

[172] M. Born, *Ann. Inst. Poincaré* **7** (1939) 155.

[173] M. Born and L. Infeld, *Proc. Roy. Soc. London* **A144** (1934) 425.

[174] N. Boulanger and L. Gualtieri, *Class. Quant. Grav.* **18** (2001) 1485.

[175] D. G. Boulware, *Phys. Rev.* **D8** (1973) 2363.

[176] B. D. Boulware and S. Deser, *Ann. Phys.* **89** (1975) 193–240.

[177] R. H. Boyer, *Proc. Roy. Soc. London* **A311** (1969) 245.

[178] M. Bradley, G. Fodor, L. Á. Gergely, M. Marklund, and Z. Perjés, *Class. Quant. Grav.* **16** (1999) 1667.

[179] D. Brecher and M. J. Perry, *Nucl. Phys.* **B566** (2000) 151.

[180] J. C. Breckenridge, D. A. Lowe, R. C. Myers, A. W. Peet, A. Strominger, and C. Vafa, *Phys. Lett.* **B381** (1996) 423.

[181] J. C. Breckenridge, G. Michaud, and R. C. Myers, *Phys. Rev.* **D55** (1997) 6438–6446.

[182] J. C. Breckenridge, G. Michaud, and R. C. Myers, *Phys. Rev.* **D56** (1997) 5172.

[183] J. C. Breckenridge, R. C. Myers, A. W. Peet, and C. Vafa, *Phys. Lett.* **B391** (1997) 93–98.

[184] D. Brill, *Phys. Rev.* **133** (1964) B845–B848.

[185] D. R. Brill and R. W. Lindquist, *Phys. Rev.* **131** (1963) 471–476.

[186] D. Brill, J. Louko, and P. Peldán, *Phys. Rev.* **D56** (1997) 3600.

[187] D. Brill and J. A. Wheeler, *Rev. Mod. Phys.* **29** (1957) 465.

[188] L. Brink, S. Deser, B. Zumino, P. Di Vecchia, and P. S. Howe, *Phys. Lett.* **B64** (1976) 435.

[189] L. Brink, P. Di Vecchia, and P. Howe, *Phys. Lett.* **65B** (1976) 471–474.

[190] L. Brink, P. Di Vecchia, and P. Howe, *Nucl. Phys.* **B118** (1977) 76.

[191] L. Brink and J. H. Schwarz, *Phys. Lett.* **B100** (1981) 310.

[192] L. Brink, J. H. Schwarz, and J. Scherk, *Nucl. Phys.* **B121** (1977) 77.

[193] H. W. Brinkmann, *Proc. Nat. Acad. Sci.* **9** (1923) 1.

[194] H. W. Brinkmann, *Math. Annal.* **94** (1925) 119.

[195] R. Brout, S. Massar, R. Parentani, and P. Spindel, *Phys. Rep.* **260** (1995) 329–454.

[196] J. D. Brown, *Phys. Rev.* **D51** (1995) 5725.

[197] L. Burko and A. Ori, in *Internal Structure of Black Holes and Spacetime Singularities*, Eds. L. Burko and A. Ori, Bristol: Institute of Physics Publishing and Jerusalem: The Israel Physical Society (1997).

[198] T. Buscher, *Phys. Lett.* **159B** (1985) 127.

[199] T. Buscher, *Phys. Lett.* **194B** (1987) 59.

[200] T. Buscher, *Phys. Lett.* **201B** (1988) 466.

[201] M. Cahen and N. Wallach, *Bull. Am. Math. Soc.* **76** (1970) 585–591.

[202] R.-G. Cai and Y.-Z. Zhang, *Phys. Rev.* **D54** (1996) 4891–4898.

[203] M. M. Caldarelli and D. Klemm, *Nucl. Phys.* **B545** (1999) 434

[204] C. G. Callan Jr, S. Coleman, and R. Jackiw, *Ann. Phys.* **59** (1970) 42–73.

[205] C. G. Callan, E. J. Martinec, M. J. Perry, and D. Friedan, *Nucl. Phys.* **B262** (1985) 593.

[206] C. G. Callan, J. A. Harvey, and A. Strominger, *Nucl. Phys.* **B359** (1991) 611–634.

[207] C. G. Callan, J. A. Harvey, and A. Strominger, in *String Theory and Quantum Gravity '91*, Ed. J. A. Harvey, Singapore: World Scientific (1992), pp. 208–244.

[208] C. G. Callan Jr and J. M. Maldacena, *Nucl. Phys.* **B472** (1996) 591–608.

[209] C. G. Callan Jr and J. M. Maldacena, *Nucl. Phys.* **B513** (1998) 198–212.

[210] C. Callan and L. Thorlacius, in *Particles, Strings and Supernovae*, vol. 2, Eds. A. Jevicki and C. I. Tan, Singapore: World Scientific (1989), p. 795–878.

[211] I. C. G. Campbell and P. C. West, *Nucl. Phys.* **B243** (1984) 112.

[212] G. L. Cardoso, B. de Wit, and T. Mohaupt, *Class. Quant. Grav.* **17** (2000) 1007.

[213] S. Carlip, *Nucl. Phys. Proc. Suppl.* **88** (2000) 10–16.

[214] B. J. Carr, in *General Relativity*, Eds. G. S. Hall and J. R. Pulham, London: Institute of Physics Publishing (1996), pp. 143–202.

[215] E. Cartan, *J. Math. Pure Appl.* **1** (1922) 141–203.

[216] B. Carter, *Phys. Rev. Lett.* **26** (1971) 331–333.

[217] B. Carter, in *Black Holes*, Eds. C. Dewitt and B. S. DeWitt, New York: Gordon and Breach (1973).

[218] B. Carter, *Phys. Rev. Lett.* **33** (1974) 558–561.

[219] R. Casalbuoni, *Phys. Lett.* **B62** (1976) 49.

[220] L. Castellani, A. Ceresole, R. D'Auria, S. Ferrara, P. Fré, and M. Trigiante, *Nucl. Phys.* **B527** (1998) 142.

[221] L. Castellani, R. D'Auria, and P. Fré, *Supergravity and Superstrings, A Geometric Perspective*, 3 volumes, Singapore: World Scientific (1991).

[222] L. Castellani, R. D'Auria, and S. Ferrara, *Class. Quant. Grav.* **1** (1990) 317.

[223] M. Cederwall, A. von Gussich, B. E. Nilsson, and A. Westerberg, *Nucl. Phys.* **B490** (1997) 163.

[224] M. Cederwall, A. von Gussich, B. E. Nilsson, P. Sundell, and A. Westerberg, *Nucl. Phys.* **B490** (1997) 179.

[225] A. Ceresole, R. D'Auria, S. Ferrara, and A. Van Proeyen, *Nucl. Phys.* **444** (1995) 92.

[226] A. Chamseddine, *Nucl. Phys.* **B185** (1981) 403–415.

[227] A. H. Chamseddine, *Ann. Phys.* **113** (1978) 219.

[228] A. H. Chamseddine, S. Ferrara, G. W. Gibbons, and R. Kallosh, *Phys. Rev.* **D55** (1997) 3647.

[229] A. H. Chamseddine, J. M. Figueroa-O'Farrill, and W. Sabra, *Six-Dimensional Supergravity Vacua and Anti-Selfdual Lorentzian Lie Groups*, `hep-th/0306278`.

[230] A. H. Chamseddine and H. Nicolai, *Phys. Lett.* **B96** (1980) 89.

[231] A. H. Chamseddine and P. C. West, *Nucl. Phys.* **B129** (1977) 39.

[232] S. Chandrasekhar, *The Mathemetical Theory of Black Holes*, Oxford: Clarendon Press (1983).

[233] G. F. Chapline and N. S. Manton, *Phys. Lett.* **B120** (1983) 105.

[234] J. E. Chase, *Commun. Math. Phys.* **19** (1970) 276.

[235] J. Chazy, *Bull. Soc. Math. France* **52** (1924) 17.

[236] S. S. Chern, *Ann. Math.* **46** (1945) 674–684.

[237] Y. M. Cho, *Phys. Rev.* **D14** (1976) 2521.

[238] Y. M. Cho and S. W. Zoh, *Phys. Rev.* **D46** (1992) R2290.

[239] Y. M. Cho and S. W. Zoh, *Phys. Rev.* **D46** (1992) 3483.

[240] Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds and Physics*, Amsterdam: Elsevier (1977).

[241] D. Christodoulou, Ph.D. Thesis, Princeton University (unpublished).

[242] J. Ciufolini and J. A. Wheeler, *Gravitation and Inertia*, Princeton, New Jersey: Princeton University Press (1995).

[243] C. J. S. Clarke, *The Analysis of Space-Time Singularities*, Cambridge: Cambridge University Press (1993).

[244] C. J. S. Clarke, *Class. Quant. Grav.* **10** (1993) 1375.

[245] G. Clément and D. V. Gal'tsov, *Phys. Rev.* **D54** (1996) 6136.

[246] S. Coleman, in *Gauge Theories in High Energy Physics*, Eds. M. K. Gaillard and R. F. Stora, Amsterdam: North-Holland (1983).

[247] S. Coleman and R. Jackiw, *Ann. Phys.* **67** (1971) 552.

[248] S. Coleman, S. Parke, A. Neveu, and C. M. Sommerfield, *Phys. Rev.* **D15** (1977) 544–545.

[249] S. Coleman, J. Preskill, and F. Wilczek, *Nucl. Phys.* **B378** (1992) 175.

[250] A. A. Coley, *Phys. Rev. Lett.* **89** (2002) 281601.

[251] P. A. Collins and R. W. Tucker, *Nucl. Phys.* **B121** (1977) 307.

[252] R. Coquereaux and A. Jadczyk, *Riemannian Geometry, Fiber Bundles, Kaluza–Klein Theories and All That.* Singapore: World Scientific, (1988).

[253] N. J. Cornish and J. W. Moffat, *Phys. Rev.* **D47** (1993) 4421–4424.

[254] M. S. Costa, *Nucl. Phys.* **B490** (1997) 202–216.

[255] M. S. Costa, *Nucl. Phys.* **B495** (1997) 195–205.

[256] M. S. Costa and G. Papadopoulos, *Nucl. Phys.* **B510** (1998) 217–231.

[257] M. S. Costa and M. J. Perry, *Nucl. Phys.* **B591** (2000) 469–487.

[258] P. M. Cowdall, `hep-th/0009016`.

[259] J. D. E. Creighton and R. B. Mann, *Phys. Rev.* **D52** (1995) 4569.

[260] J. D. E. Creighton and R. B. Mann, `gr-qc/9511012`.

[261] E. Cremmer, in *Superspace & Supergravity*, Eds. S. W. Hawking and M. Roček, Cambridge: Cambridge University Press (1981), p. 267.

[262] E. Cremmer and S. Ferrara, *Phys. Lett.* **B91** (1980) 61.

[263] E. Cremmer and B. Julia, *Nucl. Phys.* **B159** (1979) 141.

[264] E. Cremmer, B. Julia, and J. Scherk, *Phys. Lett.* **76B** (1978) 409.

[265] E. Cremmer, C. Kounnas, A. Van Proeyen, J. P. Derendinger, S. Ferrara, B. de Wit, and L. Girardello, *Nucl. Phys.* **250** (1985) 385.

[266] E. Cremmer, J. Scherk, and S. Ferrara, *Phys. Lett.* **74B** (1978) 64.

[267] J. F. Cornwell, *Group Theory in Physics* volumes 1, 2, and 3, London: Academic Press (1989).

[268] P. T. Chruściel, *Contemp. Math.* **170** (1994) 23–49.

[269] H. E. J. Curzon, *Proc. London Math. Soc.* **23** (1924) 477.

[270] C. Cutler and R. M. Wald, *Class. Quant. Grav.* **4** (1987) 1267.

[271] M. Cvetič and C. M. Hull, *Nucl. Phys.* **B480** (1996) 296.

[272] M. Cvetič and F. Larsen, *Nucl. Phys.* **B531** (1998) 239.

[273] M. Cvetič and H. Soleng, *Phys. Rep.* **282** (1997) 159–223.

[274] M. Cvetič and A. A. Tseytlin, DAMTP Report DAMTP-R-96-27 and hep-th/9606033.

[275] M. Cvetič and A. A. Tseytlin, *Phys. Lett.* **B366** (1996) 95.

[276] M. Cvetič and A. A. Tseytlin, *Phys. Rev.* **D53** (1996) 5619–5633.

[277] M. Cvetič and D. Youm, *Phys. Lett.* **B359** (1995) 87–92.

[278] A. Dabholkar, *Phys. Lett.* **B357** (1995) 307–312.

[279] A. Dabholkar, *Phys. Lett.* **B402** (1997) 53.

[280] A. Dabholkar, in *High Energy Physics and Cosmology 1997*, Eds. E. Gava *et al.*, Singapore: World Scientific (1998), p. 128.

[281] A. Dabholkar, G. W. Gibbons, J. Harvey, and F. Ruiz-Ruiz, *Nucl. Phys.* **B340** (1990) 33.

[282] A. Dabholkar and J. Harvey, *Phys. Rev. Lett.* **63** (1989) 478.

[283] J. Dai, R. G. Leigh, and J. Polchinski, *Mod. Phys. Lett.* **A4** (1989) 2073.

[284] T. Damour, *Ann. Phys.* **11** (2000) 1.

[285] T. Damour, S. Deser, and J. McCarthy, *Phys. Rev.* **D45** (1992) R3289–R3291.

[286] T. Damour, S. Deser, and J. McCarthy, *Phys. Rev.* **D47** (1993) 1541–1556.

[287] T. Damour and A. M. Polyakov, *Nucl. Phys.* **B423** (1994) 532.

[288] T. Damour and A. M. Polyakov, *Gen. Relativ. Gravit.* **26** (1994) 1171.

[289] T. Damour and R. Ruffini, *Phys. Rev. Lett.* **35** (1975) 463.

[290] U. Danielsson, G. Ferretti, and I. R. Klebanov, *Phys. Rev. Lett.* **79** (1997) 1984.

[291] A. Das and D. Z. Freedman, *Nucl. Phys.* **B120** (1977) 221.

[292] S. R. Das, G. W. Gibbons, and S. D. Mathur, *Phys. Rev. Lett.* **78** (1997) 417.

[293] S. R. Das and S. D. Mathur, *Nucl. Phys.* **B478** (1996) 561.

[294] S. R. Das and S. D. Mathur, *Ann. Rev. Nucl. Part. Sci.* **50** (2000) 153–206.

[295] A. Davidson and E. Gedalin, *Phys. Lett.* **B339** (1994) 304.

[296] T. De Donder, *La gravique Einsteinienne*, Paris: Gauthier-Villars (1921).

[297] M. Demiański and E. T. Newman, *Bull. Acad. Pol. Sci. Ser. Sci. Math. Astron. Phys.* **14** (1966) 653.

[298] M. de Roo, *Nucl. Phys.* **B255** (1985) 515.

[299] S. Deser, *Gen. Relativ. Gravit.* **1** (1970) 9–18.

[300] S. Deser, *The Gravitational Field*, Lectures given at Brandeis University 1971–1972 (unpublished).

[301] S. Deser, *Class. Quant. Grav.* **4** (1987) L99–L105.

[302] S. Deser, in *Proceedings of the 7th Mexican School of Particles and Fields and 1st Latin American Symposium on High-Energy Physics,* Eds. J. C. D'Oliva, M. Klein-Kreisler, and H. Méndez, New York: American Institute of Physics (1997), pp. 437–445.

[303] S. Deser, in *Santiago 1997, Black Holes and the Structure of the Universe*, Eds. C. Teitelboim and J. Zanelli, Singapore: World Scientific (2000), pp. 70–80.

[304] S. Deser, in *College Station 1998, Relativity, Particle Physics and Cosmology*, Ed. R. E. Allen, Singapore: World Scientific (1999), pp. 1–12.

[305] S. Deser, A. Gomberoff, M. Henneaux, and C. Teitelboim, *Nucl. Phys.* **B520** (1998) 179–204.

[306] S. Deser and L. Halpern, *Gen. Relativ. Gravit.* **1** (1970) 131–136.

[307] S. Deser, M. Hennaaux, and C. Teitelboim, *Phys. Rev.* **D55** (1997) 826–828.

[308] S. Deser and B. E. Laurent, *Ann. Phys.* **50** (1968) 76–101.

[309] S. Deser and M. Soldate, *Nucl. Phys.* **B311** (1988/89) 739–750.

[310] S. Deser and C. Teitelboim, *Phys. Rev. Lett.* **39** (1977) 249.

[311] S. Deser, H.-S. Tsao, and P. van Nieuwenhuizen, *Phys. Lett.* **B50** (1974) 491.

[312] S. Deser, H.-S. Tsao, and P. van Nieuwenhuizen, *Phys. Rev.* **D10** (1974) 3337.

[313] S. Deser and P. van Nieuwenhuizen, *Phys. Rev.* **D10** (1974) 401.

[314] S. Deser and P. van Nieuwenhuizen, *Phys. Rev.* **D10** (1974) 411.

[315] S. Deser and B. Zumino, *Phys. Lett.* **62B** (1976) 335.

[316] S. Deser and B. Zumino, *Phys. Lett.* **65B** (1976) 369–373.

[317] B. de Wit, P. G. Lauwers, and A. Van Proeyen, *Nucl. Phys.* **B255** (1985) 569.

[318] B. S. DeWitt, *Phys. Rev.* **162** (1967) 1195–1239.

[319] B. S. DeWitt, *Phys. Rev.* **162** (1967) 1239–1256.

[320] E. D'Hoker and D. H. Phong, *Rev. Mod. Phys.* **60** (1988) 917.

[321] D. E. Diaconescu, *Nucl. Phys.* **B503** (1997) 220.

[322] M. Dine, P. Huet, and N. Seiberg, *Nucl. Phys.* **B322** (1989) 301.

[323] P. A. M. Dirac, *Proc. Roy. Soc. London* **A113** (1931) 60.

[324] L. Dolan and M. J. Duff, *Phys. Rev. Lett.* **52** (1984) 14.

[325] F. Dowker, J. P. Gauntlett, G. W. Gibbons, and G. T. Horowitz, *Phys. Rev.* **D53** (1996) 7115.

[326] F. Dowker, J. P. Gauntlett, S. B. Giddings, and G.T. Horowitz, *Phys. Rev.* **D50** (1994) 2662.

[327] F. Dowker, J. P. Gauntlett, D. A. Kastor, and J. Traschen, *Phys. Rev.* **D49** (1994) 2909.

[328] J. S. Dowker, *Gen. Relativ. Gravit.* **5** (1974) 603–613.

[329] J. S. Dowker and J. A. Roche, *Proc. Phys. Soc. London* **92** (1967) 1.

[330] E. Dudas and J. Mourad, *Phys. Lett.* **B514** (2001) 173.

[331] M. J. Duff, in *Oskar Klein Centenary Nobel Symposium*, `hep-th/9410046`.

[332] M. J. Duff, *Nucl. Phys.* **B442** (1995) 47–63.

[333] M. J. Duff, *Supermembranes*, `hep-th/9611203`.

[334] M. J. Duff, G. W. Gibbons, and P. K. Townsend, *Phys. Lett.* **B332** (1994) 321–328.

[335] M. J. Duff, P. S. Howe, T. Inami, and K. S. Stelle, *Phys. Lett.* **B191** (1987) 70.

[336] M. J. Duff and R. R. Khuri, *Nucl. Phys.* **B411** (1994) 473.

[337] M. J. Duff, R. R. Khuri, and J. X. Lu, *Phys. Rep.* **259** (1995) 213–326.

[338] M. J. Duff, J. T. Liu, and J. Rahmfeld, *Nucl. Phys.* **B459** (1996) 125.

[339] M. J. Duff and J. X. Lu, *Nucl. Phys.* **B390** (1993) 276.

[340] M. J. Duff, H. Lü, and C. N. Pope, *Phys. Lett.* **B409** (1997) 136.

[341] M. J. Duff, H. Lü, C. N. Pope, and E. Sezgin, *Phys. Lett.* **B371** (1996) 206–214.

[342] M. J. Duff, B. E. W. Nilsson, and C. N. Pope, *Phys. Rep.* **130** (1986) 1–142.

[343] M. J. Duff and C. N. Pope, in *School on Supergravity and Supersymmetry*, Eds. S. Ferrara, J. G. Taylor, and P. van Nieuwenhuizen, Singapore: World Scientific (1983).

[344] M. J. Duff and K. S. Stelle, *Phys. Lett.* **B253** (1991) 113.

[345] A. S. Eddington, *Nature* **113** (1924) 192.

[346] J. D. Edelstein, L. Tataru, and R. Tatar, *JHEP* **9806** (1998) 003.

[347] T. Eguchi, P. B. Gilkey, and A. J. Hanson, *Phys. Rep.* **66** (1980) 213.

[348] T. Eguchi and A. J. Hanson, *Phys. Lett.* **74B** (1978) 249–251.

[349] J. Ehlers and W. Rindler, *Gen. Relativ Gravit.* **29** (1997) 519–529.

[350] A. Einstein, *Jahrbuch der Radioaktivität und Elektronik* **4** (1908) 411–462.

[351] A. Einstein, *Ann. Phys.* **35** (1911) 898–908.

[352] A. Einstein, *Ann. Phys.* **35** (1912) 355.

[353] A. Einstein, *Ann. Phys.* **35** (1912) 443.

[354] A. Einstein, *Phys. Z.* **14** (1913) 1251.

[355] A. Einstein, *Sitzungsber. Preuß. Akad. Wiss., phys.-math. Kl.* (1915) 778.

[356] A. Einstein, *Sitzungsber. Preuß. Akad. Wiss., phys.-math. Kl.* (1918) 154.

[357] A. Einstein, *Sitzungsber. Preuß. Akad. Wiss., phys.-math. Kl.* (1918) 448.

[358] A. Einstein, *Sitzungsber. Preuß. Akad. Wiss.* (1925) 414.

[359] A. Einstein, *Sitzungsber. Preuß. Akad. Wiss., phys.-math. Kl.* (1928) 217.

[360] A. Einstein, *Sitzungsber. Preuß. Akad. Wiss., phys.-math. Kl.* (1929) 2.

[361] A. Einstein, *Sitzungsber. Preuß. Akad. Wiss., phys.-math. Kl.* (1929) 156.

[362] A. Einstein, *Sitzungsber. Preuß. Akad. Wiss., phys.-math. Kl.* (1930) 401.

[363] A. Einstein, *Math. Ann.* **102** (1930) 685–697.

[364] A. Einstein, *The Meaning of Relativity. Including the Relativistic Theory of the Non-symmetric Field*, Fifth Edition, Princeton, New Jersey: Princeton University Press (1955).

[365] A. Einstein and A. D. Fokker, *Ann. Phys.* **44** (1914) 321–328.

[366] A. Einstein, L. Infeld, and B. Hoffmann, *Ann. Math.* **39** (1938) 65.

[367] A. Einstein and B. Kaufman, *Ann. Math.* **62** (1955) 128.

[368] A. Einstein and W. R. P. Mayer, *Sitzungsber. Preuß. Akad. Wiss., phys.-math. Kl.* (1930) 110–120.

[369] A. Einstein and E. G. Straus, *Ann. Math.* **47** (1946) 731.

[370] S. Elitzur, A. Giveon, and D. Kutasov, *Phys. Lett.* **B400** (1997) 269.

[371] R. Emparan, *Phys. Rev.* **D61** (2000) 104009.

[372] R. Emparan and H. S. Reall, *Phys. Rev.* **D65** (2002) 084025.

[373] R. Emparan and H. S. Reall, *Phys. Rev. Lett.* **88** (2002) 101101.

[374] R. Emparan and E. Teo, *Nucl. Phys.* **B610** (2001) 190–214.

[375] E. Eyras, B. Janssen, and Y. Lozano, *Nucl. Phys.* **B531** (1998) 275–301.

[376] E. Eyras and Y. Lozano, *Nucl. Phys.* **B546** (1999) 197–218.

[377] E. Eyras and Y. Lozano, *Nucl. Phys.* **B573** (2000) 735–767.

[378] J. Fang and C. Fronsdal, *J. Math. Phys.* **20** (1979) 2264.

[379] A. Fayyazuddin and D. J. Smith, *JHEP* **9904** (1999) 030.

[380] S. Ferrara and P. van Nieuwenhuizen, *Phys. Rev. Lett.* **37** (1976) 1669–1671.

[381] S. Ferrara and R. Kallosh, *Phys. Rev.* **D54** (1996) 1525.

[382] S. Ferrara, R. Kallosh, and A. Strominger, *Phys. Rev.* **D52** (1995) 5412–5416.

[383] S. Ferrara and J. M. Maldacena, *Class. Quant. Grav.* **15** (1998) 749.

[384] S. Ferrara, C. A. Savoy, and B. Zumino, *Phys. Lett.* **100B** (1981) 393.

[385] V. Ferrari and P. Pendenza, *Gen. Relativ. Gravit.* **22** (1990) 1105–1117.

[386] R. P. Feynman, *Feynman Lectures on Gravitation*, Eds. F. B. Morinigo, W. G. Wagner, and B. Hatfield Reading, Massachussets: Addison-Wesley (1995).

[387] R. P. Feynman, *Acta Phys. Polon.* **24** (1963) 697–722.

[388] M. Fierz and W. Pauli, *Proc. Roy. Soc. London*, **A173** (1939) 211.

[389] J. M. Figueroa-O'Farrill, `hep-th/9807149`.

[390] J. M. Figueroa-O'Farrill, *Class. Quant. Grav.* **16** (1999) 2043.

[391] J. M. Figueroa-O'Farrill, *Phys. Lett.* **B471** (1999) 128.

[392] J. M. Figueroa-O'Farrill and G. Papadopoulos, *JHEP* **0108** (2001) 036.

[393] J. M. Figueroa-O'Farrill and G. Papadopoulos, *JHEP* **0303** (2003) 048.

[394] D. Finkelstein, *Phys. Rev.* **110** (1958) 965–967.

[395] W. Fischler and L. Susskind, *Phys. Lett.* **B171** (1986) 383.

[396] V. A. Fock, *J. Phys. USSR* **1** (1939) 81.

[397] A. Font, L. E. Ibáñez, D. Lüst, and F. Quevedo, *Phys. Lett.* **B249** (1990) 35.

[398] E. S. Fradkin and D.M. Gitman, *Phys. Rev.* **D44** (1991) 3230.

[399] E. S. Fradkin and M. A. Vasiliev, Lebedev Institute preprint N 197 (1976).

[400] P. Fré, *Nucl. Phys. Proc. Suppl.* **45BC** (1996) 59–114.

[401] P. Fré, Lectures given at the school *Recent Advances in M-Theory*, `hep-th/0102114`.

[402] P. Fré and P. Soriani, *The $N = 2$ Wonderland: From Calabi–Yau Manifolds to Topological Field Theories*, Singapore: World Scientific (1995).

[403] D. Z. Freedman, P. van Nieuwenhuizen, and S. Ferrara *Phys. Rev.* **D13** (1976) 3214.

[404] P. G. O. Freund, *Introduction to Supersymmetry*, Cambridge: Cambridge University Press (1986).

[405] P. G. O. Freund and Y. Nambu, *Phys. Rev.* **174** (1968) 1741.

[406] P. O. Freund and M. A. Rubin, *Phys. Lett.* **B97** (1980) 233.

[407] J. L. Friedman, K. Schleich, and D. M. Witt, *Phys. Rev. Lett.* **71** (1993) 1486–1489. *Erratum Phys. Rev. Lett.* **75** (1995) 1872.

[408] M. K. Gaillard and B. Zumino, *Nucl. Phys.* **B193** (1981) 221.

[409] D. V. Gal'tsov, A. A. Garcia, and O. V. Kechkin, *J. Math. Phys.* **36** (1995) 5023–5041.

[410] D. V. Gal'tsov and O.V. Kechkin, *Phys. Rev.* **D50** (1994) 7394.

[411] D. V. Gal'tsov and O. V. Kechkin, *Phys. Rev.* **D54** (1996) 1656–1666.

[412] D. V. Gal'tsov and P. S. Letelier, *Phys. Rev.* **D55** (1997) 3580.

[413] D. V. Gal'tsov and S. A. Sharakin, *Phys. Lett.* **B399** (1997) 250–257.

[414] R. Gambini and J. Pullin, *Loops, Knots, Gauge Theories and Quantum Gravity*, Cambridge: Cambridge University Press (1996).

[415] A. Garcia, D. V. Gal'tsov, and O. V. Kechkin, *Phys. Rev. Lett.* **74** (1995) 1276–1279.

[416] D. Garfinkle, G. Horowitz and A. Strominger, *Phys. Rev.* **D43** (1991) 3140. *Erratum Phys. Rev.* **D45** (1992) 3888.

[417] J. P. Gauntlett, in *Dualities of Gauge and String Theories*, Eds. Y. M. Cho and S. Nam, Singapore: World Scientific (1998), pp. 146–193.

[418] J. P. Gauntlett, G. W. Gibbons, G. Papadopoulos, and P. K. Townsend, *Nucl. Phys.* **B500** (1997) 133.

[419] J. P. Gauntlett, J. Gomis, and P. K. Townsend, *JHEP* **9801** (1998) 003.

[420] J. P. Gauntlett, J. B. Gutowski, C. M. Hull, S. Pakis, and H. S. Reall, `hep-th/0209114`.

[421] J. P. Gauntlett, D. A. Kastor, and J. Traschen, *Nucl. Phys.* **B478** (1996) 544.

[422] J. P. Gauntlett, R. C. Myers, and P. K. Townsend, *Phys. Rev.* **D59** (1999) 025001.

[423] J. P. Gauntlett and S. Pakis, *JHEP* **0304** (2003) 039.

[424] S. P. Gavrilov and D. M. Gitman, *Int. J. Mod. Phys.* **A15** (2000) 4499.

[425] H. Georgi and S. L. Glashow, *Phys. Rev. Lett.* **28** (1972) 1494.

[426] J. Gheerardyn and P. Meessen, *Phys. Lett.* **B525** (2002) 322.

[427] F. Giani and M. Pernici, *Phys. Rev.* **D30** (1984) 325.

[428] G. W. Gibbons, *Commun Math. Phys.* **35** (1974) 13–23.

[429] G. W. Gibbons, *Commun. Math. Phys.* **44** (1975) 245–264.

[430] G. W. Gibbons, *Proc. Roy. Soc. London* **A372** (1980) 535.

[431] G. W. Gibbons, Notes of Lectures given at the 1980 Summer Institute of the Ecole Normale Supérieure, LPT, Paris, France, August 4–23, 1980 (unpublished).

[432] G. W. Gibbons, *Nucl. Phys.* **B207** (1982) 337.

[433] G. W. Gibbons, in *Supersymmetry, Supergravity and Related Topics*, Eds. F. del Águila, J. de Azcárraga and L. Ibáñez, Singapore: World Scientific (1985), p. 147.

[434] G. W. Gibbons, *Nucl. Phys.* **B514** (1998) 603.

[435] G. W. Gibbons, M. B. Green, and M. J. Perry. *Phys. Lett.* **B370** (1996) 37–44.

[436] G. W. Gibbons and S. W. Hawking, *Phys. Rev.* **D15** (1977) 2752. (Reprinted in [439].)

[437] G. W. Gibbons and S. W. Hawking, *Phys. Lett.* **78B** (1978) 430–432. (Reprinted in [439].)

[438] G. W. Gibbons and S. W. Hawking, *Commun. Math. Phys.* **66** (1979) 291. (Reprinted in [439].)

[439] G. W. Gibbons and S. W. Hawking (Eds.), *Euclidean Quantum Gravity*, Singapore: World Scientific (1993).

[440] G. W. Gibbons and C. A. R. Herdeiro, *Class. Quant. Grav.* **16** (1999) 3619–3652.

[441] G. W. Gibbons, G. T. Horowitz, and P. K. Townsend, *Class. Quant. Grav.* **12** (1995) 297–318.

[442] G. W. Gibbons and C. M. Hull, *Phys. Lett.* **109B** (1982) 190.

[443] G. W. Gibbons, C. M. Hull, and N. P. Warner, *Nucl. Phys.* **B218** (1983) 173–190.

[444] G. W. Gibbons, D. Ida, and T. Shiromizu, *Phys. Rev. Lett.* **89** (2002) 041101.

[445] G. W. Gibbons and R. E. Kallosh, *Phys. Rev.* **D51** (1995) 2839–2862.

[446] G. W. Gibbons, R. Kallosh, and B. Kol, *Phys. Rev. Lett.* **77** (1996) 4992–4995.

[447] G. W. Gibbons and K. Maeda, *Nucl. Phys.* **B298** (1988) 741.

[448] G. W. Gibbons and M. J. Perry, *Phys. Rev.* **D22** (1980) 313.

[449] G. W. Gibbons and C. N. Pope, *Commun. Math. Phys.* **66** (1979) 267–290. (Reprinted in [439].)

[450] G. W. Gibbons and D. A. Rasheed, *Nucl. Phys.* **B454** (1995) 185.

[451] G. W. Gibbons and D. A. Rasheed, *Phys. Lett.* **B365** (1996) 46.

[452] G. W. Gibbons and P. K. Townsend, *Phys. Rev. Lett.* **71** (1993) 3754–3757.

[453] S. B. Giddings and A. Strominger, *Phys. Lett.* **B230** (1989) 46.

[454] R. Gilmore, *Lie Groups, Lie Algebras and Some of Their Applications*, New York: Wiley (1974).

[455] P. Ginsparg, in *Fields, Strings and Critical Phenomena*, Eds. E. Brézin and J. Zinn-Justin, Amsterdam: North Holland (1990), pp. 1–168.

[456] A. Giveon, M. Porrati, and E. Rabinovici, *Phys. Rep.* **244** (1994) 77–202.

[457] F. Gliozzi, J. Scherk, and D. Olive, *Nucl. Phys.* **B122** (1977) 253–290.

[458] P. Goddard, J. Goldstone, C. Rebbi, and C. B. Thorn, *Nucl. Phys.* **B56** (1973) 109.

[459] P. Goddard and D. I. Olive, *Rep. Prog. Phys.* **41** (1978) 91.

[460] M. Godina and P. Matteucci, math.DG/0201235.

[461] A. Gomberoff, D. Kastor, D. Marolf, and J. Traschen, *Phys. Rev.* **D61** (2000) 024012.

[462] M. H. Goroff and A. Sagnotti, *Nucl. Phys.* **B266** (1986) 709.

[463] T. Goto, *Prog. Theor. Phys.* **46** (1971) 1560–1569.

[464] J. Govaerts, *Phys. Lett.* **B220** (1989) 77.

[465] J. C. Graves and D. Brill, *Phys. Rev.* **120** (1960) 1507.

[466] M. B. Green, *Phys. Lett.* **B329** (1994) 435.

[467] M. B. Green, *Phys. Lett.* **B354** (1995) 271.

[468] M. B. Green, Introduction to string and superstring theory 1, and M. E. Peskin, Introduction to String and Superstring Theory 2, in *From the Planck Scale to the Weak Scale: Towards a Theory of the Universe*, Ed. H. E. Haber, Singapore: World Scientific (1987).

[469] M. B. Green and M. Gutperle, *Phys. Lett.* **B377** (1996) 28.

[470] M. B. Green, C. M. Hull, and P. K. Townsend, *Phys. Lett.* **B382** (1996) 65.

[471] M. B. Green and J. H. Schwarz, *Phys. Lett.* **B136** (1984) 367.

[472] M. B. Green and J. H. Schwarz, *Nucl. Phys.* **B243** (1984) 285.

[473] M. B. Green, J. H. Schwarz, and E. Witten, *Superstring Theory* (two volumes), Cambridge: Cambridge University Press (1987).

[474] B. R. Greene, A. Shapere, C. Vafa, and S.-T. Yau, *Nucl. Phys.* **B337** (1990) 1.

[475] R. Gregory and J. A. Harvey, *Phys. Rev.* **D47** (1993) 2411–2422.

[476] R. Gregory, J. A. Harvey, and G. Moore, *Adv. Theor. Math. Phys.* **1** (1997) 283–297.

[477] R. Gregory and R. Laflamme, *Phys. Rev. Lett.* **70** (1993) 2837.

[478] R. Gregory and R. Laflamme, *Nucl. Phys.* **B428** (1994) 399.

[479] R. Gregory and R. Laflamme, *Phys. Rev.* **D51** (1995) 305.

[480] M. Grisaru, *Phys. Lett.* **73B** (1978) 207.

[481] M. T. Grisaru, P. S. Howe, L. Mezincescu, B. Nilsson, and P. K. Townsend, *Phys. Lett.* **B162** (1985) 116.

[482] F. Gronwald and F. W. Hehl, in *School of Cosmology and Gravitation on Quantum Gravity*, Eds. P. G. Bergmann, V. de Sabbata, and H.-J. Treder, Singapore: World Scientific (1996), pp. 148–198.

[483] D. J. Gross and M. J. Perry, *Nucl. Phys.* **B226** (1983) 29.

[484] S. S. Gubser, lectures given at TASI 2001, hep-th/0201114.

[485] S. S. Gubser and I. Mitra, hep-th/0009126.

[486] S. S. Gubser and I. Mitra, *JHEP* **0108** (2001) 018.

[487] M. Günaydin and N. Marcus, *Class. Quant. Grav.* **2** (1985) L11.

[488] S. N. Gupta, *Phys. Rev.* **96** (1954) 1683–1685.

[489] S. N. Gupta, *Rev. Mod. Phys.* **29** (1957) 334–336.

[490] M. Gurses and F. Gursey, *Phys. Rev.* **D11** (1975) 967.

[491] F. Gürsey, *Ann. Phys.* **24** (1963) 211.

[492] R. Güven, *Phys. Lett.* **158B** (1985) 468–472.

[493] R. Güven, *Phys. Lett.* **B191** (1987) 275.

[494] R. Güven, *Phys. Lett.* **276B** (1992) 49.

[495] R. Güven, *Phys. Lett.* **B482** (2000) 255–263.

[496] R. Haag, J. T. Lopuszański, and M. Sohnius. *Nucl. Phys.* **B88** (1975) 257.

[497] A. Hanany and E. Witten, *Nucl. Phys.* **B492** (1997) 152–190.

[498] B. J. Harrington and H. K. Shepard, *Phys. Rev.* **D17** (1978) 2122.

[499] B. K. Harrison, *J. Math. Phys.* **9** (1968) 1744.

[500] J. B. Hartle and S. W. Hawking, *Commun. Math. Phys.* **26**, (1972) 87.

[501] J. A. Harvey and A. Strominger, *Nucl. Phys.* **B449** (1995) 535–552. *Erratum Nucl. Phys.* **B458** (1996) 456–473.

[502] A. Hashimoto, *Phys. Rev.* **D57** (1998) 6441.

[503] A. Hashimoto, *JHEP* **9901** (1999) 018.

[504] S. F. Hassan, *Nucl. Phys.* **B460** (1996) 362–378.

[505] S. F. Hassan, *Nucl. Phys.* **B568** (2000) 145.

[506] S. F. Hassan, *Nucl. Phys.* **B583** (2000) 431.

[507] S. W. Hawking, *Commun. Math. Phys.* **25** (1972) 152.

[508] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time*, Cambridge: Cambridge University Press (1973).

[509] S. W. Hawking, in *Black Holes*, Eds. C. DeWitt and B. S. DeWitt, New York: Gordon and Breach (1973).

[510] S. W. Hawking, *Nature* **248** (1974) 30.

[511] S. W. Hawking, *Commun. Math. Phys.* **43** (1975) 199–220. (Reprinted in [439].)

[512] S. W. Hawking, *Phys. Rev.* **D13** (1976) 191–197.

[513] S. W. Hawking, *Phys. Lett.* **60A** (1977) 81.

[514] S. W. Hawking, in *General Relativity, An Einstein Centenary Survey*, Eds. S. W. Hawking and W. Israel, Cambridge: Cambridge University Press (1979).

[515] S. W. Hawking and G. T. Horowitz, *Class. Quant. Grav.* **13** (1996) 1487–1498.

[516] S. W. Hawking, G. T. Horowitz, and S. F. Ross, *Phys. Lett.* **B383** (1996) 383–389.

[517] S. W. Hawking and C. J. Hunter, *Phys. Rev.* **D59** (1999) 044025.

[518] S. W. Hawking and C. J. Hunter, *Class. Quant. Grav.* **13** (1996) 2735–2752.

[519] S. W. Hawking, C. J. Hunter and D. N. Page, `hep-th/9809035`.

[520] S. W. Hawking and S. F. Ross, *Phys. Rev.* **D52** (1995) 5865–5876.

[521] K. Hayashi and T. Nakano, *Prog. Theor. Phys.* **38** (1967) 491–507.

[522] K. Hayashi and T. Shirafuji, *Phys. Rev.* **D19** (1979) 3524–3553.

[523] F. W. Hehl, P. von der Heyde, G. D. Kerlick, and J. M. Nester, *Rev. Mod. Phys.* **48** (1976) 393–416.

[524] F. W. Hehl, J. D. McCrea, E. W. Mielke, and Y. Ne'eman, *Phys. Rep.* **258** (1995) 1–171.

[525] K. Heiderich and W. Unruh, *Phys. Rev.* **D38** (1988) 490–497.

[526] K. R. Heiderich and W. G. Unruh, *Phys. Rev.* **D42** (1990) 2057–2068.

[527] S. Helgason, *Differential Geometry, Lie Groups and Symmetric Spaces*, Academic Press, New York (1978).

[528] M. Henneaux and L. Mezincescu, *Phys. Lett.* **B152** (1985) 340.

[529] M. Henneaux and C. Teitelboim, *Ann. Phys.* **143** (1982) 127.

[530] M. Henneaux, C. Teitelboim, and J. Zanelli *Nucl. Phys.* **B332** (1990) 169–188.

[531] M. Heusler, *Black Holes Uniqueness Theorems*, Cambridge: Cambridge University Press (1996).

[532] M. Heusler, *Helv. Phys. Acta* **69** (1996) 501.

[533] M. Heusler, *Living Rev. Relativ.* **1** (1998) 6.

[534] M. Heusler and N. Straumann, *Class. Quant. Grav.* **10** (1993) 1299.

[535] D. Hilbert, *Konigl. Gesell. Wiss. Göttingen Nachr., math.-phys. Kl* 295–407.

[536] J. W. van Holten, in *From Field Theory to Quantum Groups*, Eds. B. Jancewicz and J. Sobczyk, Singapore: World Scientific (1996), pp. 173–189.

[537] J. W. van Holten, *Nucl. Phys.* **B457** (1995) 375.

[538] J. W. van Holten and A. van Proeyen, *J. Phys.* **A15** (1982) 3763.

[539] C. F. Holzhey and F. Wilczek, *Nucl. Phys.* **B380** (1992) 447–477.

[540] H. Hopf, *Math. Ann.* **104** (1931) 637.

[541] P. Hořava, *Nucl. Phys.* **B327** (1989) 461.

[542] P. Hořava, *Phys. Lett.* **B231** (1989) 251.

[543] P. Hořava and E. Witten, *Nucl. Phys.* **B460** (1996) 506–524.

[544] P. Hořava and E. Witten, *Nucl. Phys.* **B475** (1996) 94.

[545] J. H. Horne and G. T. Horowitz, *Nucl. Phys.* **B399** (1993) 169–196.

[546] G. T. Horowitz, in *String Theory and Quantum Gravity*, Eds. J. A. Harvey, R. Iengo, K. S. Narain, S. Randjbar-Daemi, and H. L. Verlinde, Singapore: World Scientific (1993), pp. 55–99.

[547] G. T. Horowitz, in *Proceedings of the Pacific Conference on Gravitation and Cosmology*, Eds. Y. M. Cho, C. H. Lee, and S. W. Kim, Singapore: World Scientific (1999), pp. 46–63.

[548] G. T. Horowitz, in *Black Holes and Relativistic Stars*, Chicago, Illinois (1998), pp. 241–266.

[549] G. T. Horowitz, D. A. Lowe, and J. M. Maldacena, *Phys. Rev. Lett.* **77** (1996) 430.

[550] G. T. Horowitz, J. M. Maldacena, and A. Strominger, *Phys. Lett.* **B383** (1996) 151.

[551] G. T. Horowitz and K. Maeda, *Phys. Rev. Lett.* **87** (2001) 131301.

[552] G. T. Horowitz and K. Maeda, *Phys. Rev.* **D65** (2002) 104028.

[553] G. T. Horowitz and J. Polchinski, *Phys. Rev.* **D55** (1997) 6189.

[554] G. T. Horowitz and J. Polchinski, *Phys. Rev.* **D57** (1998) 2557.

[555] G. T. Horowitz and S. F. Ross, *Phys. Rev.* **D56** (1997) 2180–2187.

[556] G. T. Horowitz and S. F. Ross, *Phys. Rev.* **D57** (1998) 1098–1107.

[557] G. T. Horowitz and A. Strominger, *Nucl. Phys.* **B360** (1991) 197.

[558] G. T. Horowitz and A. R. Steif, *Phys. Rev. Lett.* **64** (1990) 260.

[559] G. T. Horowitz and A. R. Steif, *Phys. Rev.* **D42** (1990) 1950.

[560] G. T. Horowitz and A. A. Tseytlin, *Phys. Rev.* **D51** (1995) 2896.

[561] K. Hosomichi, *JHEP* **0006** (2000) 004.

[562] Y. Hosotani, *Phys. Lett.* **B126** (1983) 309.

[563] Y. Hosotani, *Phys. Lett.* **B129** (1983) 193.

[564] Y. Hosotani, *Ann. Phys.* **190** (1989) 233.

[565] M. Hotta and M. Tanaka, *Class. Quant. Grav.* **10** (1993) 307–314.

[566] P. S. Howe, N. D. Lambert, and P. C. West, *Phys. Lett.* **B416** (1998) 303.

[567] P. S. Howe, N. D. Lambert, and P. C. West, *Nucl. Phys.* **B515** (1998) 203.

[568] P. S. Howe, N. D. Lambert, and P. C. West, *Phys. Lett.* **B419** (1998) 79.

[569] P. S. Howe and E. Sezgin, *Phys. Lett.* **B390** (1997) 133.

[570] P. S. Howe and E. Sezgin, *Phys. Lett.* **B394** (1997) 62.

[571] P. Howe and P. C. West, *Nucl. Phys.* **B238** (1984) 181–219.

[572] C.-G. Huang and C.-B. Liang, *Phys. Lett.* **A201** (1995) 27–32.

[573] V. E. Hubený, *Phys. Rev.* **D59** (1999) 064013.

[574] J. Hughes, J. Liu, and J. Polchinski, *Phys. Lett.* **B180** (1986) 370.

[575] C. M. Hull, *Commun. Math. Phys.* **90** (1983) 545–561.

[576] C. M. Hull, in *Super Field Theories*, Eds. H. C. Lee, V. Elias, G. Kunstatter, R. B. Mann, and K. S. Viswanathan, New York: Plenum Press (1987).

[577] C. M. Hull, *Phys. Lett.* **B357** (1995) 545–551.

[578] C. M. Hull, *Nucl. Phys.* **B509** (1998) 216–251.

[579] C. M. Hull, *JHEP* **9807** (1998) 018.

[580] C. M. Hull, *Phys. Lett.* **B462** (1999) 271–276.

[581] C. M. Hull, Report `hep-th/0203146`.

[582] C. M. Hull, `hep-th/0305039`

[583] C. M. Hull and P. K. Townsend, *Nucl. Phys.* **B438** (1995) 109–137.

[584] C. J. Hunter, *Phys. Rev.* **D59** (1999) 024009.

[585] M. Huq and M. A. Namazie, *Class. Quant. Grav.* **2** (1985) 293–308.

[586] D. J. Hurley and M. A. Vandyck, *J. Phys.* **A27** (1994) 4569.

[587] C. A. Hurst, *Ann. Phys.* **50** (1968) 51.

[588] T. Z. Husain, Ph.D. Thesis, `hep-th/0304143`.

[589] S. Hwang, *Geometriae Dedicata* **71** (1998) 5.

[590] S. Hyun, `hep-th/9704055`.

[591] A. Yu. Ignatev and G. C. Joshi, *Phys. Rev.* **D53** (1996) 984.

[592] E. Inönü and E. P. Wigner, *Proc. Nat. Acad. Sci.* **39** (1953) 510.

[593] W. Israel, *Phys. Rev.* **164** (1967) 1776–1779.

[594] W. Israel, *Commun. Math. Phys.* **8** (1968) 245.

[595] W. Israel and K. A. Khan, *Nuovo Cim.* **33** (1964) 331.

[596] W. Israel and J. M. Nester, *Phys. Lett.* **85A**, (1981) 259.

[597] W. Israel and G. A. Wilson, *J. Math. Phys.* **13**, (1972) 865.

[598] N. Itzhaki, A. A. Tseytlin, and S. Yankielowicz, *Phys. Lett.* **B432** (1998) 298.

[599] C. Itzykson and B. Zuber, *Quantum Field Theory*, New York: McGraw-Hill (1980).

[600] J. M. Izquierdo, N. D. Lambert, G. Papadopoulos, and P. K. Townsend, *Nucl. Phys.* **B460** (1996) 560–578.

[601] R. Jackiw, C. Nohl, and C. Rebbi, *Phys. Rev.* **D15** (1977) 1642–1646.

[602] B. Janssen, *JHEP* **0001** (2000) 044.

[603] B. Janssen, P. Meessen, and T. Ortín, *Phys. Lett.* **B453** (1999) 229.

[604] C. V. Johnson, lectures given at ICTP, TASI, and BUSSTEPP, `hep-th/0007170`.

[605] C. V. Johnson, *D-Branes*, Cambridge: Cambridge University Press (2002).

[606] C. V. Johnson, R. R. Khuri, and R. C. Myers, *Phys. Lett.* **B378** (1996) 78–86

[607] A. I. Janis, E. T. Newman, and J. Winicour, *Phys. Rev. Lett.* **20** (1968) 878.

[608] B. Julia, in *Superspace & Supergravity*, Eds. S. W. Hawking and M. Roček, Cambridge: Cambridge University Press (1981), p. 331.

[609] M. Kaku, *Introduction to Superstrings*, New York: Springer-Verlag (1988).

[610] R. Kallosh, D. Kastor, T. Ortín, and T. Torma, *Phys. Rev.* **D50** (1994) 6374.

[611] R. Kallosh and B. Kol, *Phys. Rev.* **D53** (1996) 5344–5348.

[612] R. Kallosh, A. Linde, T. Ortín, A. Peet, and A. Van Proeyen, *Phys. Rev.* **D46** (1992) 5278–5302.

[613] R. Kallosh and T. Ortín, *Phys. Rev.* **D48** (1993) 742–747.

[614] R. Kallosh, A. Rajaraman, and W. K. Wong, *Phys. Rev.* **D55** (1997) 3246.

[615] T. Kaluza, *Sitzungsber. Preuß. Akad. Wiss., phys.-math. kl.* (1921) 966 (translated into English in [45]).

[616] P. F. Kelly, *Class. Quant. Grav.* **6** (1991) 1217–1229. *Erratum Class. Quant. Grav.* **8** (1992) 1423(E).

[617] R. P. Kerr, *Phys. Rev. Lett.* **11** (1963) 237.

[618] R. R. Khuri, *Phys. Lett.* **B470** (1999) 73.

[619] R. R. Khuri, *Nucl. Phys.* **B588** (2000) 253.

[620] R. R. Khuri and T. Ortín, *Nucl. Phys.* **B467** (1996) 355–382.

[621] R. R. Khuri and T. Ortín, *Phys. Lett.* **B373** (1996) 56–60.

[622] K. Kikkawa and M. Yamasaki, *Phys. Lett.* **B149** (1984) 357.

[623] H. J. Kim, L. J. Romans, and P. van Nieuwenhuizen, *Phys. Rev.* **D32** (1985) 389.

[624] E. Kiritsis, `hep-th/9708130`.

[625] E. Kiritsis, *Leuven Notes in Mathematical and Theoretical Physics* **B9**, Leuven: Leuven University Press (1998).

[626] O. Klein, *Z. Phys.* **37** (1926) 895.

[627] F. Klein, *Konigl. Gesell. Wiss. Göttinger Nachr., math.-phys. Kl.* (1918) 394.

[628] D. Klemm, V. Moretti, and L. Vanzo, *Phys. Rev.* **D57** (1998) 6127

[629] D. Klemm and L. Vanzo, *Phys. Rev.* **D58** (1998) 104025.

[630] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, volumes 1 and 2, New York: Interscience (1969).

[631] A. Komar, *Phys. Rev.* **113** (1954) 934.

[632] Y. Kosmann, *Annali Mat. Pura Appl.* **(IV) 91** (1972) 317–395.

[633] Y. Kosmann, *C. R. Acad. Sci. Paris Sér. A*. **262** (1966) A289–A292.

[634] V. A. Kostelecky and M. J. Perry, *Phys. Lett.* **B371** (1996) 191–198.

[635] F. Kottler, *Encyklopädie der mathematikalischen Wissenschaft*, volume 6, part II, Leipzig: B. G. Teubner (1922).

[636] J. Kowalski-Glikman, *Phys. Lett.* **B134** (1984) 194–196.

[637] J. Kowalski-Glikman, *Phys. Lett.* **B150** (1985) 125–126.

[638] R.H. Kraichnan, *Phys. Rev.* **98** (1955) 1118–1122. *Erratum Phys. Rev.* **99** (1955) 1906.

[639] R. H. Kraichnan, *Phys. Rev.* **107** (1957) 1485–1490.

[640] D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations*, Cambridge: Cambridge University Press (1980).

[641] M. Kruskal, *Phys. Rev.* **119** (1960) 1743–1745.

[642] T. Kugo and P. K. Townsend, *Nucl. Phys.* **B221** (1983) 357–380.

[643] H. P. Kunzle, *Proc. Roy. Soc. London* **A297** (1967) 244.

[644] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, Oxford: Pergamon Press (1975).

[645] M. von Laue *Jahrbuch der Radioaktivität und Elektronik* **14** (1917) 263.

[646] S. M. Lee, A. Peet, and L. Thorlacius, *Nucl. Phys.* **B514** (1998) 161.

[647] R. G. Leigh, *Mod. Phys. Lett.* **A4** (1989) 2767.

[648] J. P. S. Lemos, *Class. Quant. Grav.* **12** (1995) 1081–1086.

[649] J. P. S. Lemos, *Phys. Lett.* **B353** (1995) 46–51.

[650] J. P. S. Lemos and V. T. Zanchin, *Phys. Rev.* **D54** (1996) 3840–3853.

[651] T. Levi-Cività, *Sitzungsber. Preuß. Akad. Wiss., phys.-math. Kl.* (1929) 137–153.

[652] S. Liberati and G. Pollifrone, *Phys. Rev.* **D56** (1997) 6558.

[653] S. Liberati, T. Rothman, and S. Sonego, *Phys. Rev.* **D62** (2000) 024005.

[654] A. Lichnerowicz, *Théories relativistes de la gravitation et de l'électromagnétisme*, Paris: Masson (1955).

[655] A. Lichnerowicz, *C. R. Acad. Sci. Paris* **257** (1963) 7–9.

[656] L. Lindblom and Masood-ul-Alam, *Commun. Math. Phys.* **162** (1994) 123.

[657] U. Lindström and R. von Unge, *Phys. Lett.* **B403** (1997) 233.

[658] A. Loewy, *Phys. Lett.* **B463** (1999) 41.

[659] A. A. Logunov, *Phys. Part. Nucl.* **29** (1) (1998) 1–32. (*Fiz. Élem. Chast. Atom Yadra* **29** (1998) 5–81.)

[660] H. A. Lorentz, *Amsterd. Versl.* **25** (1916) 468.

[661] C. O. Lousto and N. Sánchez, *Nucl. Phys.* **B383** (1992) 377–394.

[662] D. Lovelock, *J. Math. Phys.* **12** (1971) 498–501.

[663] Y. Lozano, *Phys. Lett.* **B414** (1997) 52.

[664] E. Lozano-Tellechea, P. Meessen, and T. Ortín, *Class. Quant. Grav.* **19** (2002) 5921–5934.

[665] E. Lozano-Tellechea and T. Ortín, *Nucl. Phys.* **B569** (2000) 435–450.

[666] E. Lozano-Tellechea and T. Ortín, *Nucl. Phys.* **B607** (2001) 213–236.

[667] H. Lü, C. N. Pope, and J. Rahmfeld, *J. Math. Phys.* **40** (1999) 4518.

[668] H. Lü, C. N. Pope, E. Sezgin, and K. S. Stelle, *Nucl. Phys.* **B456** (1995) 669–698.

[669] H. Lü, C. N. Pope, T. A. Tran, and K. W. Xu, *Nucl. Phys.* **B511** (1998) 98.

[670] J. X. Lu and Shibaji Roy, *Phys. Lett.* **B428** (1998) 289–296.

[671] J. X. Lu and Shibaji Roy, *Nucl. Phys.* **B560** (1999) 181–206.

[672] J. X. Lu and Shibaji Roy, *Phys. Rev.* **D60** (1999) 126002.

[673] D. Lüst and S. Theisen, *Lectures on String Theory*, Heidelberg: Springer-Verlag (1989).

[674] S. W. MacDowell and F. Mansouri, *Phys. Rev. Lett.* **38** (1977) 739–742. *Erratum Phys. Rev. Lett.* **38** (1977) 1376.

[675] J. Maharana and J. H. Schwarz, *Nucl. Phys.* **B390** (1993) 3–32.

[676] S. D. Majumdar, *Phys. Rev.* **72** (1947) 390–398.

[677] J. M. Maldacena, Ph.D. Thesis, Princeton University, `hep-th/9607235`.

[678] J. M. Maldacena, *Nucl. Phys. Proc. Suppl.* **61A** (1998) 111–123.

[679] J. Maldacena, *Adv. Theor. Math. Phys.* **2** (1998) 231–252.

[680] J. M. Maldacena and A. Strominger, *Phys. Rev. Lett.* **77** (1996) 428.

[681] J. M. Maldacena and A. Strominger, *Phys. Rev.* **D55** (1997) 861.

[682] G. Mangano, `gr-qc/9511027`.

[683] R. B. Mann, *Class. Quant. Grav.* **14** (1997) L109–L114.

[684] R. B. Mann, in *Internal Structure of Black Holes and Spacetime Singularities*, Eds. L. Burko and A. Ori, Bristol: Institute of Physics Publishing and Jerusalem: The Israel Physical Society (1997), pp. 311–342.

[685] N. Marcus and J. H. Schwarz, *Phys. Lett.* **115B** (1982) 111–114.

[686] M. A. Markov and V. P. Frolov, *Teor. Mat. Fiz.* **3** (1970) 3.

[687] D. Marolf, in *Proceedings of the E.S. Fradkin Memorial Conference*, `hep-th/0006117`.

[688] D. Marolf and A. Peet, *Phys. Rev.* **D60** (1999) 105007.

[689] A. E. Mayo and J. D. Bekenstein, *Phys. Rev.* **D54** (1996) 5059–5069.

[690] P. Meessen, *Phys. Rev.* **D65** (2002) 087501.

[691] P. Meessen and T. Ortín, *Nucl. Phys.* **B541** (1999) 195–245.

[692] M. A. Melvin, *Phys. Lett.* **8** (1964) 65.

[693] K. Menou, E. Quataert, and R. Narayan, in *The Eighth Marcel Grossman Meeting: Proceedings* Ed. T. Piran, Singapore: World Scientific (1999), pp. 204–224.

[694] R. R. Metsaev, *Nucl. Phys.* **B625** (2002) 70

[695] R. R. Metsaev and A. A. Tseytlin, *Phys. Rev.* **D63** (2001) 046002.

[696] R. R. Metsaev and A. A. Tseytlin, *Phys. Rev.* **D65** (2002) 126004.
[697] C. Misner, *J. Math. Phys.* **4**, 924 (1963).
[698] C. Misner, *Phys. Rev. Lett.* **28** (1972) 994.
[699] J. W. Moffat, *Phys. Rev.* **D19** (1979) 3554.
[700] C. Møller, *Ann. Phys.* **4** (1958) 347.
[701] C. Møller, *Ann. Phys.* **12** (1961) 118.
[702] C. Møller, *Mat. Fys. Skr. Dan. Vid. Selsk.* **1** nr. 10 (1961).
[703] S. Moriyama, *Phys. Lett.* **B522** (2001) 177.
[704] U. Muench, F. Gronwald, and F. W. Hehl, `gr-qc/9801036`.
[705] H. Müller zum Hagen and H. J. Seifert, *Int. J. Theor. Phys.* **8** (1973) 443.
[706] R. C. Myers and M. J. Perry, *Ann. Phys.* **172** (1986) 304.
[707] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*, New York: W. H. Freeman and Co. (1973).
[708] T. Mohaupt, *Class. Quant. Grav.* **17** (2000) 3429.
[709] T. Mohaupt, *Fortschr. Phys.* **49** (2001) 3–161.
[710] D. R. Morrison and N. Seiberg, *Nucl. Phys.* **B483** (1997) 229.
[711] M. Moshinski, *Phys. Rev.* **80** (1950) 514.
[712] R. C. Myers, *Phys. Rev.* **D35** (1987) 455.
[713] R. C. Myers, Summary of Lectures given at Fourth Mexican School on Gravitation and Mathematical Physics, `gr-qc/0107034`.
[714] Y. Nambu, Lectures at the 1970 Copenhagen Symposium.
[715] M. Nakahara, *Geometry, Topology and Physics*, London: Institute of Physics Publishing (1990).
[716] T. Nakamura and H. Sato, *Phys. Lett.* **B61** (1976) 371–374.
[717] C. Nash and S. Sen, *Topology and Geometry for Physicists*, London: Academic Press (1983).
[718] R. I. Nepomechie, *Phys. Rev.* **D31** (1985) 1921–1924.
[719] J. M. Nester, *Phys. Lett.* **83A**, (1981) 241.
[720] J. M. Nester and H.-J. Yo, `gr-qc/9809049`.
[721] G. Neugebauer, Habilitationsschrift, Jena (1969).
[722] A. Neveu and J. H. Schwarz, *Nucl. Phys.* **D31** (1971) 86.
[723] E. T. Newman, E. Couch, K. Chinnapared, A. Exton, A. Prakash, and R. Torrence, *J. Math. Phys.* **6** (1965) 918.
[724] E. Newman, L. Tambourino, and T. Unti, *J. Math. Phys.* **4** (1963) 915.
[725] A. I. Nikishov, *Phys. Part. Nucl.* **32** (2001) 1–14 (*Fiz. Élem. Chast. Atom. Yadra* **32** (2001) 5–30).
[726] H. Nishino and S. Rajpoot, *Phys. Lett.* **B546** (2002) 261–272.
[727] H. Nishino and E. Sezgin, *Nucl. Phys.* **B278** (1986) 353.
[728] G. Nordström, *Phys. Z.* **15** (1914) 504 (translated into English in [45]).
[729] G. Nordström, *Phys. Z.* **13** (1912) 1126.
[730] G. Nordström, *Ann. Phys.* **40** (1913) 856.
[731] G. Nordström, *Ann. Phys.* **42** (1913) 533.
[732] G. Nordström, *Ann. Phys.* **43** (1914) 1101.
[733] G. Nordström, *Phys. Z.* **15** (1914) 375.
[734] G. Nordström, *Ann. Acad. Sci. Fenn.* **57** (1914, 1915).
[735] G. Nordström, *Proc. Kon. Ned. Akad. Wet.* **20** (1918) 1238.
[736] J. D. Norton, *Archive for the History of Exact Sciences* **45** (1993) 17.
[737] I. Novikov and V. P. Frolov, *Physics of Black Holes*, Dordrecht: Kluwer Academic Publishers (1989).
[738] N. A. Obers, B. Pioline, and E. Rabinovici, *Nucl. Phys.* **B525** (1998) 163.
[739] V. I. Ogievetsky and I. V. Polubarinov, *Ann. Phys.* **35** (1965) 167–208.
[740] D. W. Olson and W. G. Unruh, *Phys. Rev. Lett.* **33** (1974) 1116–1119.
[741] B. O'Neill, *The Geometry of Kerr Black Holes*, Wellesley: A. K. Peters (1994).
[742] L. O'Raifeartaigh, *The Dawning of Gauge Theory*, Princeton, New Jersey: Princeton University Press (1997).
[743] T. Ortín, *Phys. Rev.* **D47** (1993) 3136–3143.
[744] T. Ortín, *Phys. Rev.* **D52** (1995) 3392–3405.

[745] T. Ortín, *Nucl. Phys. Proc. Suppl.* **61A** (1998) 131–134.

[746] T. Ortín, *Phys. Lett.* **B415** (1997) 39–44.

[747] T. Ortín, *Phys. Lett.* **B422** (1998) 93–100.

[748] T. Ortín, *Class. Quant. Grav.* **19** (2002) L1–L7.

[749] H. Ooguri and Z. Yin, `hep-th/9612254`.

[750] D. N. Page, *Phys. Rev.* **D16** (1977) 2402.

[751] D. N. Page, *Phys. Lett.* **B78** (1978) 249.

[752] A. Palatini, *Red. Circ. Mat. Palermo* **43** (1919) 203–212.

[753] G. Papadopoulos and P. K. Townsend, *Phys. Lett.* **B380** (1996) 273–279.

[754] A. Papapetrou, *Proc. Roy. Irish Acad.* **A51** (1947) 191.

[755] A. Papapetrou, *Proc. Roy. Soc. London* **209A** (1951) 248.

[756] P. Pasti, D. Sorokin, and M. Tonin, `hep-th/9611100`.

[757] P. Pasti, D. P. Sorokin, and M. Tonin, *Phys. Lett.* **B398** (1997) 41.

[758] A. W. Peet, *Class. Quant. Grav.* **15** (1998) 3291.

[759] A. W. Peet, `hep-th/0008241`.

[760] A. W. Peet, *Class. Quant. Grav.* **17** (2000) 1235.

[761] C. Pellegrini and J. Plebański, *Mat. Fys. Skr. Dan. Vid. Selsk.* **2** nr. 4 (1963).

[762] R. Penrose, *Riv. Nuovo Cim.* **1** (1969) 252–276.

[763] R. Penrose, in *Battelle Rencontres*, Eds. C. M. Will and J. A. Wheeler, New York: Benjamin.

[764] R. Penrose, in *Differential Geometry and Relativity*, Dordrecht: Reidel (1976), pp. 271–275.

[765] R. Penrose, in *General Relativity: An Einstein Centenary Survey*, Eds. S. W. Hawking and W. Israel, Cambridge: Cambridge University Press (1979).

[766] R. Penrose, in *Essays in General Relativity*, Ed. F. J. Tipler, New York: Academic Press (1980).

[767] R. Penrose, in *Black Holes and Relativistic Stars*, Chicago, Illinois: The University of Chicago Press (1998).

[768] R. Penrose and W. Rindler, *Spinors and Space-Time*, volumes 1 and 2, Cambridge: Cambridge University Press (1984).

[769] Z. Perjés, *Phys. Rev. Lett.* **27** (1971) 1668.

[770] A. N. Petrov and J. Katz, *Proc. Roy. Soc. London* **458** (2002) 319–337.

[771] K. Pilch, P. van Nieuwenhuizen, and P. K. Townsend, *Nucl. Phys.* **B242** (1984) 377.

[772] J. F. Plebański, *Ann. Phys.* **90** (1975) 196.

[773] J. F. Plebański and M. Demiański, *Ann. Phys.* **98** (1976) 98.

[774] J. Podolský, in *Gravitation: Following the Prague Inspiration*, Eds. O. Semerak, J. Podolský, and M. Zefka, Singapore: World Scientific (2002), pp. 205–246.

[775] J. Podolský and J. B. Griffiths, *Phys. Rev.* **D58** (1998) 124024.

[776] J. Polchinski, *Phys. Rev.* **D50** (1994) 6041.

[777] J. Polchinski, *Phys. Rev. Lett.* **75** (1995) 4724–4727.

[778] J. Polchinski, `hep-th/9611050`.

[779] J. Polchinski, *String Theory*, volumes 1 and 2, Cambridge: Cambridge University Press (1998).

[780] J. Polchinski, S. Chaudhuri, and C. V. Johnson, `hep-th/9602052`.

[781] J. Polchinski, *Nucl. Phys.* **B303** (1988) 226–236.

[782] J. Polchinski and E. Witten, *Nucl. Phys.* **B460** (1996) 525.

[783] A. M. Polyakov, *Sov. Phys. JETP Lett.* **20** (1974) 194.

[784] A. M. Polyakov, *Phys. Lett.* **B103** (1981) 207.

[785] A. M. Polyakov, *Phys. Lett.* **B103** (1981) 211.

[786] G. Pradisi and A. Sagnotti, *Phys. Lett.* **B216** (1989) 59.

[787] M. K. Prasad, *Phys. Lett.* **B83** (1979) 310.

[788] M. K. Prasad and C. M. Sommerfield, *Phys. Rev. Lett.* **35** (1975) 760–762.

[789] V. Pravda, A. Pravdova, A. Coley, and R. Milson, *Class. Quant. Grav.* **19** (2002) 6213.

[790] J. Preskill, P. Schwarz, A. Shapere, S. Trivedi, and F. Wilczek, *Mod. Phys. Lett.* **A6** (1991) 2353–2361.

[791] R. H. Price, *Phys. Rev.* **D5** (1972) 2419–2438.

[792] R. H. Price, *Phys. Rev.* **D5** (1972) 2439–2454.

[793] H. Quevedo, *Fortschr. Phys.* **38** (1990) 733.

[794] A. Rajaraman, *JHEP* **0109** (2001) 018.

[795] P. Ramond, *Field Theory: A Modern Primer*, Reading, Massachussets: Benjamin/Cummings (1981).

[796] P. Ramond, *Phys. Rev.* **D3** (1971) 2415.

[797] L. Randall and R. Sundrum, *Phys. Rev. Lett.* **83** (1999) 3370.

[798] L. Randall and R. Sundrum, *Phys. Rev. Lett.* **83** (1999) 4690.

[799] J. Rayski, *Acta Phys. Polon* **27** (1965) 89.

[800] H. S. Reall, *Phys. Rev.* **D64** (2001) 044005.

[801] M. Rees, in *Black Holes and Relativistic Stars*, Ed. R. M. Wald, Chicago, Illinois: The University of Chicago Press (1998).

[802] H. Reissner, *Ann. Phys.* **50** (1916) 106.

[803] S. J. Rey, *Phys. Rev.* **D43**, 526 (1991).

[804] F. Riccioni, `hep-th/0301021`.

[805] M. Rocek and E. Verlinde, *Nucl. Phys.* **B373** (1992) 630

[806] M. Rogatko, *Class. Quant. Grav.* **11** (1994) 689–693.

[807] M. Rogatko, *Class. Quant. Grav.* **12** (1995) 3115–3118.

[808] M. Rogatko, *Phys. Rev.* **D58** (1998) 044011.

[809] M. Rogatko, *Class. Quant. Grav.* **19** (2002) 875.

[810] L. J. Romans, *Phys. Lett.* **B169** (1986) 374.

[811] L. J. Romans, *Nucl. Phys.* **B383** (1992) 395–415.

[812] I. Robinson, *Bull. Acad. Polon. Sci.* **7** (1959) 351.

[813] L. Rosenfeld, *Mém. Acad. Roy. Belgique* **6** (1930) 30.

[814] R. Ruffini and J. A. Wheeler, *Physi Today*, **24** (1971) 30–36.

[815] H. Rumpf, *Z. Naturforsch.* **33a** (1978) 1224–1225.

[816] H. Rumpf, *Gen. Relativ. Gravit.* **14** (1982) 873.

[817] V. de Sabbata and C. Sivaram, *Spin and Torsion in Gravitation*, Singapore: World Scientific (1994).

[818] V. de Sabbata and M. Gasperini, *Introduction to Gravitation*, Singapore: World Scientific (1985).

[819] N. Sakai and I. Senda, *Prog. Theor. Phys.* **75** (1986) 692. *Erratum Prog. Theor. Phys.* **77** (1987) 773.

[820] J. H. Schwarz, *Nucl. Phys.* **B226** (1983) 269–288.

[821] J. H. Schwarz, in *Rome String Theory Workshop* (1992), pp. 503–520.

[822] J. H. Schwarz, *Phys. Lett.* **B360** (1995) 13–18. *Erratum Phys. Lett.* **B364** (1995) 252.

[823] J. H. Schwarz and P. C. West, *Phys. Lett.* **B126** (1983) 301.

[824] J. Schwinger, *Phys. Rev.* **82** (1951) 664–679.

[825] J. Schwinger, *Phys. Rev.* **144** (1966) 1087.

[826] J. Schwinger, *Phys. Rev.* **173** (1968) 1536.

[827] A. Sagnotti, in *Nonperturbative Quantum Field Theory*, Eds. G. 't Hooft, A. Jaffe, G. Mack, P. K. Mitter, and R. Stora, New York: Plenum Press (1988), pp. 521–528.

[828] *Supergravities in Diverse Dimensions*, volumes 1 and 2, Eds. A. Salam and E. Sezgin, Amsterdam/Singapore: North Holland/World Scientific (1989).

[829] A. Salam and J. Strathdee, *Ann. Phys. (NY)* **141** (1982) 316.

[830] R. Schoen and S. -T. Yau, *Commun. Math. Phys.* **65** (1979) 45–76.

[831] J. H. Schwarz, *Phys. Rep.* **89** (1982) 223–322.

[832] J. Scherk, *Rev. Mod. Phys.* **47** (1975) 123–164.

[833] J. Scherk and J. H. Schwarz, *Nucl. Phys.* **B81** (1974) 118.

[834] J. Scherk and J. H. Schwarz, *Phys. Lett.* **B52** (1974) 347.

[835] J. Scherk and J. H. Schwarz, *Phys. Lett.* **82B** (1979) 60.

[836] J. Scherk and J. H. Schwarz, *Nucl. Phys.* **B153** (1979) 61–88.

[837] E. Schrödinger, *Proc. Roy Irish Acad.* **51A** (1947) 163.

[838] E. Schrödinger, *Space-Time Structure*, Cambridge: Cambridge University Press (1954).

[839] E. Schrödinger, *Phys. Z.* **19** (1918) 4.

[840] K. Schwarzschild, *Sitzungsber. deutsch. Akad. Wiss. Berlin, Kl. Math. Phys. Technik* (1916) 189–196.

[841] J. Schwinger, *Particles, Sources and Fields*, volume 1, New York: Addison-Wesley (1970).

[842] A. Sen, *Nucl. Phys.* **B404** (1993) 109–126.

[843] A. Sen, *Phys. Lett.* **303B** (1993) 22–26.

[844] A. Sen, *Mod. Phys. Lett.* **A8** (1993) 2023.

[845] A. Sen, *Int. J. Mod. Phys.* **A9** (1994) 3707–3750.

[846] A. Sen, *Mod. Phys. Lett.* **A10** (1995) 2081.

[847] N. Seiberg, *Nucl. Phys.* **B303** (1988) 286.

[848] J.-P. Serre, *Cours d'arithmétique*, Paris: Presses Universitaires de France (1970).

[849] K. Sfetsos and K. Skenderis, *Nucl. Phys.* **B517** (1998) 179.

[850] A. Shapere, S. Trivedi, and F. Wilczek, *Mod. Phys. Lett.* **A6**, (1991) 2677.

[851] I. L. Shapiro, *Phys. Rep* **357** (2001) 113.

[852] M. M. Sheikh Jabbari, *Phys. Lett.* **B420** (1998) 279.

[853] K. Shiraishi, *J. Math. Phys.* **34** (1993) 1480.

[854] W. Siegel, *Phys. Lett.* **B128** (1983) 397.

[855] T. P. Singh, `gr-qc/9805066`.

[856] K. Skenderis, *Lect. Notes Phys.* **541** (2000) 325–364.

[857] L. Smarr, *Phys. Rev. Lett.* **30** (1973) 71–73. *Erratum Phys. Rev. Lett.* **30** (1973) 521.

[858] D. J. Smith, `hep-th/0210157`.

[859] W. L. Smith and R. B. Mann, *Phys. Rev.* **D56** (1997) 4942–4947.

[860] R. D. Sorkin, *Phys. Rev. Lett.* **51** (1983) 87.

[861] A. A. Starobinsky, *Zh. Éksp. Teor. Fiz.* **64** (1973) 48.

[862] K. S. Stelle, in *Trieste 1996 Summer School in High Energy Physics and Cosmology*, Eds. E. Gava, A. Masiero, K. S. Narain, S. Randjbar-Daemi, and Q. Shafi, Singapore: World Scientific (1997), pp. 287–339.

[863] K. S. Stelle, in *Trieste 1997 Summer School in High Energy Physics and Cosmology*, Eds. E. Gava, A. Masiero, K. S. Narain, S. Randjbar-Daemi, G. Senjanovic, A. Smirnov, and Q. Shafi, Singapore: World Scientific (1998), pp. 29–127.

[864] K. Stelle and P. C. West, *Phys. Rev.* **D21** (1980) 1466.

[865] R. Steinbauer, `gr-qc/9809054`.

[866] C. R. Stephens, *Ann. Phys.* **193** (1989) 255–286.

[867] N. Straumann, `astro-ph/0006423`.

[868] A. Strominger, *Commun. Math. Phys.* **133** (1990) 163–180.

[869] A. Strominger, *Phys. Lett.* **B383** (1996) 44.

[870] A. Strominger and C. Vafa, *Phys. Lett.* **B370** (1996) 99.

[871] E. C. G. Stückelberg, *Helv. Phys. Acta* **11** (1938) 225.

[872] D. Sudarsky and T. Zannias, *Phys. Rev.* **D58** (1998) 087502.

[873] S. Sugimoto, *Prog. Theor. Phys.* **102** (1999) 685.

[874] S. Surya and D. Marolf, *Phys. Rev.* **D58** (1998) 124013.

[875] L. Susskind, `hep-th/9309145`.

[876] G. Szekeres, *Pbl. Mat. Debrecen* **7** (1960) 285–301.

[877] F. Tangherlini, *Nuovo Cim.* **77** (1963) 636.

[878] Y. Tanii, `hep-th/9802138`.

[879] A. H. Taub, *Ann. Math.* **53** (1951) 472.

[880] C. Teitelboim, *Phys. Rev.* **D5** (1972) 2941–2954.

[881] C. Teitelboim, *Phys. Lett.* **167B** (1986) 63–68.

[882] C. Teitelboim, *Phys. Lett.* **167B** (1986) 69–72.

[883] C. Teitelboim, *Phys. Rev.* **D51** (1995) 4315–4318. *Erratum Phys. Rev.* **D52** (1995) 6201.

[884] C. Teitelboim, D. Villarroel, and Ch. G. van Weert, *Riv. Nuovo Cim.* **3**, N. 9 (1980) 1–63.

[885] K. S. Thorne, *Black Holes & Time Warps*, New York: W. W. Norton and Co. (1994).

[886] Y. Thiry, *Acad. Sci. Paris* **226** (1948) 216.

[887] W. Thirring, *Acta Phys. Austriaca Suppl.* **9** (1972) 256–271.

[888] W. E. Thirring, *Ann. Phys.* **16** (1964) 96–117.

[889] G. 't Hooft, *Nucl. Phys.* **B62** (1973) 444–460.

[890] G. 't Hooft, *Nucl. Phys.* **B79** (1974) 276–284.

[891] G. 't Hooft and M. Veltman, *Ann. Inst. Henri Poincaré* **20** (1974) 69.

[892] A. Trautman, *Int. J. Theor. Phys.* **16** (1977) 561.

[893] K. P. Tod, *Phys. Lett.* **121B**, (1981) 241.

[894] K. P. Tod, *Class. Quant. Grav.* **12** (1995) 1801–1820.

[895] M. A. Tonnelat, *La théorie du champ unifié d'Einstein*, Paris: Gauthier-Villars (1955).

[896] P. K. Townsend, *Phys. Lett.* **B277** (1992) 285.

[897] P. K. Townsend, *Phys. Lett.* **B350** (1995) 184.

[898] P. K. Townsend, in *PASCOS/Johns Hopkins 1995*, Eds. J. A. Bagger, G. K. Domokos, A. F. Falk, and S. Kövesi-Domokos, Singapore: World Scientific (1996), pp. 271–286.

[899] P. K. Townsend, *Phys. Lett.* **B373** (1996) 68.

[900] P. K. Townsend, *Nucl. Phys. Proc. Suppl.* **B475** (1996).

[901] P. K. Townsend, in *Trieste 1996 Summer School in High Energy Physics and Cosmology*, Eds. E. Gava, A. Masiero, K. S. Narain, S. Randjbar-Daemi, and Q. Shaft, Singapore: World Scientific (1997), pp. 385–438.

[902] P. K. Townsend, *Phys. Lett.* **B409** (1997) 131.

[903] P. K. Townsend, `gr-qc/9707012`.

[904] P. K. Townsend, *Nucl. Phys. Proc. Suppl.* **67** (1998) 88.

[905] P. K. Townsend, in *Strings, Branes and Dualities*, Eds. L. Baulieu, P. Di Francesco, M. Douglas, V. Kazakov, M. Picco, and P. Windey, Dordrecht: Kluwer (1999), pp. 141–177.

[906] P. K. Townsend, `hep-th/9901102`.

[907] J. Traschen, in *Mathematical Methods of Physics*, Eds. A. Bytsenko and F. Williams, Singapore: World Scientific (2000).

[908] A. A. Tseytlin, *Nucl. Phys.* **B475** (1996) 149.

[909] A. A. Tseytlin, in *The Many Faces of the Superworld*, Ed. M. A. Shifman, Singapore: World Scientific (2000), pp. 417–452.

[910] H. J. de Vega, M. Ramón Medrano, and N. Sánchez, *Nucl. Phys.* **B374** (1992) 425.

[911] A. E. M. van de Ven, *Nucl. Phys.* **B378** (1992) 309–366.

[912] P. van Nieuwenhuizen, *Phys. Rep.* **68** (1981) 189–398.

[913] P. van Nieuwenhuizen, in *Supergravity '81*, Cambridge: Cambridge University Press (1982), pp. 151–196.

[914] P. van Nieuwenhuizen, in *1975 Marcel Grossmann Meeting on General Relativity*, Oxford: Oxford University Press (1977), pp. 1–26.

[915] A. van Proeyen, in *Spring School on Quantum Field Theory: Supersymmetry and Superstrings*.

[916] A. van Proeyen, `hep-th/0301005`.

[917] W. G. Unruh, *Phys. Rev.* **D10** (1974) 3194–3205.

[918] R. Utiyama, *Phys. Rev.* **101** (1956) 1597.

[919] M. A. Vandyck, *Gen. Relativ. Gravit.* **20** (1988) 261.

[920] M. A. Vandyck, *Gen. Relativ. Gravit.* **20** (1988) 905.

[921] L. Vanzo, *Phys. Rev.* **D56** (1997) 6475–6483.

[922] M. A. G. Veltman, in *Methods in Field Theory*, Eds. R. Balian and J. Zinn-Justin, Amsterdam: North-Holland (1976), p. 265.

[923] M. A. G. Veltman, *Nucl. Phys.* **B319** (1989) 253–270.

[924] H. Vermeil, *Nachr. Ges. Wiss. Göttingen* (1917) 334.

[925] A. Volovich, *Nucl. Phys.* **B492** (1997) 235–248.

[926] P. von der Heyde, *Z. Naturforsch.* **31a** (1976) 1725–1726.

[927] S. Wadia, `hep-th/0011286`.

[928] R. M. Wald, *Phys. Rev. Lett.* **26** (1971) 1653–1655.

[929] R. M. Wald, *Phys. Rev.* **D6** (1972) 1476–1479.

[930] R. M. Wald, *Ann. Phys.* **83** (1974) 548.

[931] R. M. Wald, in *Black Holes, Gravitational Radiation and the Universe*, Eds. B. R. Iyer, B. Bhawal, and C. V. Vishveshwara, Dordrecht: Kluwer (1998), pp. 69–85.

[932] R. Wald, *General Relativity*, Chicago, Illinois: The University of Chicago Press (1984).

[933] R. M. Wald, *Phys. Rev.* **D33** (1986) 3613–3625.

[934] R. M. Wald, *Class. Quant. Grav.* **4** (1987) 1279.

[935] R. M. Wald, in *College Park 1993, Directions in General Relativity*, volume 1, Eds. B. L. Hu, M. P. Ryan Jr, C. V. Vishveshwara, and T. A. Jacobson, Cambridge: Cambridge University Press (1993), pp. 358–366.

[936] R. M. Wald, *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*, Chicago, Illinois: The University of Chicago Press (1994).

[937] R. M. Wald, *Phys. Rev.* **D56** (1997) 6467–6474.

[938] R. M. Wald, *Living Rev. Rel.* **4** (2001) 6.

[939] S. Weinberg, *Gravitation and Cosmology*, New York: Wiley (1972).

[940] S. Weinberg, *Rev. Mod. Phys.* **61** 1–22.

[941] S. Weinberg, *Phys. Lett.* **9** (1964) 357–359.

[942] S. Weinberg, *Phys. Rev.* **135** (1964) B1049–B1056.

[943] S. Weinberg, *Phys. Rev.* **138** (1965) B988–B1002.

[944] R. Weitzenböck, *Invariantentheorie*, Groningen: Noordhoff (1923).

[945] R. Weitzenböck, *Sitzungsber. Konigl. Preuß. Akad. Wiss., phys.-math. Kl.* (1928) 466.

[946] J. Wess and J. Bagger, *Supersymmetry and Supergravity*, Princeton, New Jersey: Princeton University Press (1992).

[947] P. C. West, in *Duality and Supersymmetric Theories*, Eds. D. I. Olive and P. C. West, Cambridge: Cambridge University Press (1999), pp. 147–266.

[948] P. C. West, *An Introduction to Supersymmetry and Supergravity*, extended second edition, Singapore: World Scientific (1990).

[949] H. Weyl, *Ann. Phys.* **54** (1917) 117.

[950] H. Weyl, *Ann. Phys.* **54** (1917) 185.

[951] H. Weyl, *Am. J. Math.* **66** (1944) 591. (Reprinted in [953].)

[952] H. Weyl, *Space–Time–Matter*, New York: Dover (1922).

[953] H. Weyl, *Gesammelte Abhandlungen*, 4 volumes, Berlin: Springer-Verlag (1968).

[954] H. Weyl, *Z. Phys.* **56** (1929) 330. Translated in [742].

[955] F. Wilczek, *Phys. Rev. Lett.* **80** (1998) 4851–4854.

[956] E. Witten, *Phys. Lett.* **86B** (1979) 283.

[957] E. Witten, *Nucl. Phys.* **B186** (1981) 412.

[958] E. Witten, *Commun. Math. Phys.* **80**, (1981) 381–402.

[959] E. Witten, *Nucl. Phys.* **195** (1982) 481–492.

[960] E. Witten, in *Shelter Island Conference on Quantum Field Theory and the Fundamental Problems of Physics II*, Eds. R. Jackiw, N. N. Khuri, S. Weinberg, and E. Witten, Cambridge, Massachussetts: MIT Press (1985), p. 227.

[961] E. Witten, *Nucl. Phys.* **B266**, 245 (1986).

[962] E. Witten, *Nucl. Phys.* **B443** (1995) 85–126.

[963] E. Witten and D. Olive, *Phys. Lett.* **78B** (1978) 97.

[964] T. T. Wu and C. N. Yang, *Phys. Rev.* **D12** (1975) 3845.

[965] W. Wyss, *Helv. Phys. Acta* **38** (1965) 469.

[966] D. Youm, *Phys. Rep.* **316** (1999) 1–232.

[967] D. Youm, *Nucl. Phys.* **B556** (1999) 222.

[968] Ya. B. Zel'dovich, *Pis'ma Zh. Éskp. Teor. Fiz* **14** (1970) 270.

[969] B. Zumino, *Ann. N.Y. Acad. Sci.* **302** (1977) 545.

[970] D. Zwanziger, *Phys. Rev.* **176**, (1968) 1480.

[971] D. Zwanziger, *Phys. Rev.* **176**, (1968) 1489–1495.

Index

Page numbers in italic are those on which the definition or a very specific discussion of the subject can be found. 4d means four dimensions or four-dimensional, etc.